

# MOLECULAR STUDIES OF COVID-19 CHEMISTRY

EDITED BY: Emilia Pedone, Domenica Capasso, Sonia Di Gaetano and  
Chandrabose Selvaraj

PUBLISHED IN: Frontiers in Chemistry





# frontiers

## Frontiers eBook Copyright Statement

The copyright in the text of individual articles in this eBook is the property of their respective authors or their respective institutions or funders. The copyright in graphics and images within each article may be subject to copyright of other parties. In both cases this is subject to a license granted to Frontiers.

The compilation of articles constituting this eBook is the property of Frontiers.

Each article within this eBook, and the eBook itself, are published under the most recent version of the Creative Commons CC-BY licence.

The version current at the date of publication of this eBook is CC-BY 4.0. If the CC-BY licence is updated, the licence granted by Frontiers is automatically updated to the new version.

When exercising any right under the CC-BY licence, Frontiers must be attributed as the original publisher of the article or eBook, as applicable.

Authors have the responsibility of ensuring that any graphics or other materials which are the property of others may be included in the CC-BY licence, but this should be checked before relying on the CC-BY licence to reproduce those materials. Any copyright notices relating to those materials must be complied with.

Copyright and source acknowledgement notices may not be removed and must be displayed in any copy, derivative work or partial copy which includes the elements in question.

All copyright, and all rights therein, are protected by national and international copyright laws. The above represents a summary only. For further information please read Frontiers' Conditions for Website Use and Copyright Statement, and the applicable CC-BY licence.

ISSN 1664-8714

ISBN 978-2-88971-505-3

DOI 10.3389/978-2-88971-505-3

## About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

## Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

## Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.

Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

## What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: [frontiersin.org/about/contact](https://frontiersin.org/about/contact)

# MOLECULAR STUDIES OF COVID-19 CHEMISTRY

Topic Editors:

**Emilia Pedone**, National Research Council (CNR), Italy

**Domenica Capasso**, University of Naples Federico II, Italy

**Sonia Di Gaetano**, National Research Council (CNR), Italy

**Chandrabose Selvaraj**, Alagappa University, India

**Citation:** Pedone, E., Capasso, D., Di Gaetano, S., Selvaraj, C., eds. (2021).  
Molecular Studies of COVID-19 Chemistry. Lausanne: Frontiers Media SA.  
doi: 10.3389/978-2-88971-505-3

# Table of Contents

- 05 Editorial: Molecular Studies of Covid-19 Chemistry**  
Domenica Capasso, Sonia Di Gaetano, Chandrabose Selvaraj and Emilia Pedone
- 08 Hijacking SARS-CoV-2/ACE2 Receptor Interaction by Natural and Semi-synthetic Steroidal Agents Acting on Functional Pockets on the Receptor Binding Domain**  
Adriana Carino, Federica Moraca, Bianca Fiorillo, Silvia Marchianò, Valentina Sepe, Michele Biagioli, Claudia Finamore, Silvia Bozza, Daniela Francisci, Eleonora Distrutti, Bruno Catalanotti, Angela Zampella and Stefano Fiorucci
- 23 Identification of a New Potential SARS-COV-2 RNA-Dependent RNA Polymerase Inhibitor via Combining Fragment-Based Drug Design, Docking, Molecular Dynamics, and MM-PBSA Calculations**  
Mahmoud A. El Hassab, Aly A. Shoun, Sara T. Al-Rashood, Tarfah Al-Warhi and Wagdy M. Eldehna
- 34 Supporting SARS-CoV-2 Papain-Like Protease Drug Discovery: In silico Methods and Benchmarking**  
Tamer M. Ibrahim, Muhammad I. Ismail, Matthias R. Bauer, Adnan A. Bekhit and Frank M. Boeckler
- 51 Repurposing Known Drugs as Covalent and Non-covalent Inhibitors of the SARS-CoV-2 Papain-Like Protease**  
Pietro Delre, Fabiana Caporuscio, Michele Saviano and Giuseppe Felice Mangiatordi
- 68 Screening of Natural Products Targeting SARS-CoV-2–ACE2 Receptor Interface – A MixMD Based HTVS Pipeline**  
Krishnasamy Gopinath, Elmeri M. Jokinen, Sami T. Kurkinen and Olli T. Pentikäinen
- 81 Identification of Small Molecule Inhibitors of the Deubiquitinating Activity of the SARS-CoV-2 Papain-Like Protease: in silico Molecular Docking Studies and in vitro Enzymatic Activity Assay**  
Eleni Pitsillou, Julia Liang, Katherine Ververis, Kah Wai Lim, Andrew Hung and Tom C. Karagiannis
- 96 Host DDX Helicases as Possible SARS-CoV-2 Proviral Factors: A Structural Overview of Their Hijacking Through Multiple Viral Proteins**  
Flavia Squeglia, Maria Romano, Alessia Ruggiero, Giovanni Maga and Rita Berisio
- 113 Computational Prediction of Potential Inhibitors of the Main Protease of SARS-CoV-2**  
Renata Abel, María Paredes Ramos, Qiaofeng Chen, Horacio Pérez-Sánchez, Flaminia Coluzzi, Monica Rocco, Paolo Marchetti, Cameron Mura, Maurizio Simmaco, Philip E. Bourne, Robert Preissner and Priyanka Banerjee
- 132 Elucidating Interactions Between SARS-CoV-2 Trimeric Spike Protein and ACE2 Using Homology Modeling and Molecular Dynamics Simulations**  
Sugunadevi Sakkiiah, Wenjing Guo, Bohu Pan, Zuowei Ji, Gokhan Yavas, Marli Azevedo, Jessica Hawes, Tucker A. Patterson and Huixiao Hong

- 143** *Structural Insight Into the SARS-CoV-2 Nucleocapsid Protein C-Terminal Domain Reveals a Novel Recognition Mechanism for Viral Transcriptional Regulatory Sequences*  
Mei Yang, Suhua He, Xiaoxue Chen, Zhaoxia Huang, Ziliang Zhou, Zhechong Zhou, Qiuyue Chen, Shoudeng Chen and Sisi Kang
- 155** *Microsecond MD Simulation and Multiple-Conformation Virtual Screening to Identify Potential Anti-COVID-19 Inhibitors Against SARS-CoV-2 Main Protease*  
Chandrabose Selvaraj, Umesh Panwar, Dhurvas Chandrasekaran Dinesh, Evzen Boura, Poonam Singh, Vikash Kumar Dubey and Sanjeev Kumar Singh
- 170** *Computational Characterizations of the Interactions Between the Pontacyl Violet 6R and Exoribonuclease as a Potential Drug Target Against SARS-CoV-2*  
Rangika Munaweera and Ying S. Hu
- 180** *In silico Exploration of Inhibitors for SARS-CoV-2's Papain-Like Protease*  
Tien Huynh, Wendy Cornell and Binqun Luan
- 189** *In Silico Study of Coumarins and Quinolines Derivatives as Potent Inhibitors of SARS-CoV-2 Main Protease*  
Osvaldo Yañez, Manuel Isaías Osorio, Eugenio Uriarte, Carlos Areche, William Tiznado, José M. Pérez-Donoso, Olimpo García-Beltrán and Fernando González-Nilo
- 204** *Viroporins vs. Other Pore-Forming Proteins: What Lessons Can We Take?*  
Eva Žerovnik
- 210** *Structural Basis of Potential Inhibitors Targeting SARS-CoV-2 Main Protease*  
Hylemariam Mihiretie Mengist, Tebelay Dilnessa and Tengchuan Jin
- 229** *Pomegranate Peel Extract as an Inhibitor of SARS-CoV-2 Spike Binding to Human ACE2 Receptor (in vitro): A Promising Source of Novel Antiviral Drugs*  
Annalisa Tito, Antonio Colantuono, Luciano Pirone, Emilia Pedone, Daniela Intartaglia, Giuliana Giamundo, Ivan Conte, Paola Vitaglione and Fabio Apone
- 240** *Pharmacophore Modelling-Based Drug Repurposing Approaches for SARS-CoV-2 Therapeutics*  
Shailima Rampogu and Keun Woo Lee



# Editorial: Molecular Studies of Covid-19 Chemistry

Domenica Capasso<sup>1,2</sup>, Sonia Di Gaetano<sup>3</sup>, Chandrabose Selvaraj<sup>4</sup> and Emilia Pedone<sup>3\*</sup>

<sup>1</sup>CIRPEB, University of Naples "Federico II", Naples, Italy, <sup>2</sup>CESTEV, University of Naples "Federico II", Naples, Italy, <sup>3</sup>Institute of Biostructures and Bioimaging, CNR, Naples, Italy, <sup>4</sup>Department of Bioinformatics, Alagappa University, Karaikudi, India

**Keywords:** repurposing drugs, SARS-CoV-2, inhibitor, host-virus interaction, design

## Editorial on the Research Topic

### Molecular Studies of Covid-19 Chemistry

This Research Topic on "Molecular Studies of Covid-19 Chemistry" aims to highlight current advances carried out at the molecular level, providing insight into the mechanism of infection of SARS-CoV-2. In particular, this Research Topic is devoted to studying unknown aspects of the SARS-CoV-2 infection process in order to identify new therapeutic targets and validate potential inhibitors. The published articles do so using multidisciplinary approaches that involve computational chemistry, chemical-physical, and biochemical methodologies. The researchers who contributed to this Research Topic present 18 themed articles that show the latest advances in newly identified targets against SARS-CoV-2.

Different authors have searched for novel leads from the available natural substances and clinically available drugs by targeting the SARS-CoV-2 Spike's protein receptor binding domain (RBD), which remains by far the main target against the virus. Indeed, Carino et al. specifically target the druggable pockets located in the central  $\beta$ -sheet core in the RBD through virtual screening methods. This strategy suggests several triterpenoid/steroidal agents that block the protein-protein interactions between the RBD with human Angiotensin-Converting Enzyme 2 (ACE2). The findings of the computational screening are experimentally validated through ACE2/SARS-CoV-2 Spike Inhibitor Screening Assay Kit and from the analysis of Anti-SARS-CoV-2 IgG Antibodies. Overall, the authors have come up with strong potential triterpenoids, primary, secondary bile acids, and semi-synthetic derivatives for the RBD/ACE2 binding blockers reporting them as SARS-CoV-2 inhibitors. Gopinath et al. also identified new inhibitors for blocking the RBD-ACE2 interface, applying a unique protocol that combines mixed solvent MD simulations (MixMD) with high-throughput virtual screening (HTVS). MixMD are employed to identify the stable binding conformations of drug-like probes in the S-protein-ACE2 interface and those stable sites are applied for use in the molecular level screening. Among the compounds identified by MixMD, ZINC000002128789 showed a potential binding affinity to block the protein-protein RBD-ACE2 interface binding and prevent SARS-CoV-2 infection.

On the other hand, Tito et al. explored Plant-derived secondary metabolites as a source of potential inhibitors able to prevent and counteract the rapid spread of SARS-CoV-2 infections. By using *in vitro* approaches, the authors investigated the role of a pomegranate peel extract in attenuating the interaction between the SARS-CoV-2 Spike and ACE2, and on the activity of the virus 3CL protease. Although further studies are required to assess the efficacy of this extract *in vivo*, the results here reported open up new promising opportunities to employ natural extracts for the development of effective and innovative therapies in the fight against SARS-CoV-2.

Sakkiah et al. focus their research on interactions between the Spike and ACE2 building, a complex structure by homology modeling and by molecular dynamics simulations elucidating the interactions. In particular, the twenty interacting residues, responsible for binding, were

## OPEN ACCESS

### Edited and reviewed by:

John D. Wade,  
University of Melbourne, Australia

### \*Correspondence:

Emilia Pedone  
empedone@unina.it

### Specialty section:

This article was submitted to  
Chemical Biology,  
a section of the journal  
Frontiers in Chemistry

**Received:** 22 June 2021

**Accepted:** 30 June 2021

**Published:** 31 August 2021

### Citation:

Capasso D, Di Gaetano S, Selvaraj C  
and Pedone E (2021) Editorial:  
Molecular Studies of Covid-  
19 Chemistry.  
Front. Chem. 9:729142.  
doi: 10.3389/fchem.2021.729142

characterized using Molecular Mechanics Poisson-Boltzmann Surface Area (MM/PBSA) and in silico alanine scanning.

One of the main targets against SARS-CoV-2 is the main viral protease (Mpro), a key protein involved in the replication process of the virus and therefore a relevant target for identifying high affinity molecules capable of selectively inhibiting it. This special issue contains different articles reporting in silico studies on potential inhibitors against Mpro. In particular, Yañez et al. used a computational model of Mpro built in complex with different synthetic ligands derived from coumarins and quinolones to identify new potential inhibitors. By an experimental approach based on molecular dynamics and molecular docking of the models, six compounds were selected as putative candidates. However, further biological studies are required to confirm the function of the selected compounds and enable the development of novel drugs that can be employed in SARS-CoV-2 therapy.

Mengist et al. have reviewed the potential inhibitors targeting SARS-CoV-2 Mpro, providing insights into the Mpro mechanism and explaining its role as a drug target for inhibition of SARS-CoV-2. They review the active site structure, morphology for identifying the Structure-based Design of drugs. They discuss the design of the inhibitor based on their competitive binding to the active site, to identify the best candidates. Along with this, they examined all the reported compounds tested against the SARS-CoV-2 Mpro such as  $\alpha$ -ketoamide inhibitors, peptide-based inhibitors, anilid-based inhibitors, screened molecules from the TCM (Traditional Chinese Medicine) database, plant compounds, and indole lactam-based inhibitors.

Selvaraj et al., reported conformational changes of SARS-CoV-2 Mpro in the long-range time scale event of 1  $\mu$ s MD simulations. They have reported the crucial amino acids involved in the mechanism and protein stability through the 1  $\mu$ s MD simulations and also reported a multiple conformation-based ensemble docking approach to screening potential SARS-CoV-2 Mpro inhibitors from the TCM database. Selvaraj et al. found the screened compounds from the TCM database that cause functional distortion of the oxyanion hole in the Mpro reaction mechanism, and the new leads show direct interactions with His41, Gly143, and Cys145. Through the induced fit docking, they found possible binding conformations, that have the ability to interact with residues and disturb the formation of the oxyanion hole, leading to its inhibition. Abel et al. have performed the combined structure and ligand based virtual screening methodology, which includes molecular fingerprints and molecular docking methods for identifying the SARS-CoV-2 Mpro inhibitors Super Natural II and TCM database. Mass evaluation of 80 docked complex was examined in detail, and among the four compounds tested for toxicity, cytochrome inhibition profiles and dynamically simulated for understanding its stability and interactions.

The papain-like protease (PLpro) is also considered another important drug target that plays the imperial role in viral maturation and although it is responsible for the essential mechanism, it is a less studied protein in comparison to Mpro. Three papers report on its inhibitors, including Pitsillou et al., who focus on two different aspects: 1) investigating the

binding characteristics of previously identified, by a high-throughput screening, Naphthalene-based inhibitors against the SARS-CoV-2 papain-like proteases (PLpro) and 2) evaluating their effect on PLpro deubiquitinating activity. These findings are in accordance with the mechanisms and potential antiviral effects of the naphthalene-based, GRL-0617 inhibitor, which is currently progressing in preclinical trials. Their findings indicate further suitable candidates such as PLpro inhibitors, which are considered potential lead compounds.

Ibrahim et al. used three publicly docking tools, AutoDock Vina, PLANTS, and FRED against SARS-CoV-2 PLpro. In particular, the authors start from the assumption that SARS-CoV PLpro and SARS-CoV-2 PLpro share a 100% sequence identity for the binding site of small molecules. Moreover, they consider that the residues (Tyr269 and Gln270), important for recognition site, are present in both proteins, and generate a high-quality DEKOIS 2.0 benchmark set. In fact, since the co-crystal structure of SARS-CoV-2 PLpro with the commonly held small-molecule inhibitor is not reported, the authors make a homology model for SARS-CoV-2 PLpro complexed with a small-molecule ligand based on the ligand-bound SARS-CoV. FRED performed best against the built model, thus its screening performance and chemotype enrichment were comparable to the built model demonstrating the high quality of the built model. Therefore, they employed FRED in a VS campaign using the FDA-reported drugs (from DrugBank) against SARS-CoV-2 PLpro.

Huynh et al. investigated computational screening and the repurposing of FDA approved drugs through molecular modeling studies. The results of their study through MD simulations clarify that such compounds are not appropriate for the PLpro. Long range MD simulations suggest that the known inhibitor rac5c is bound stably inside the PLpro substrate binding pocket, and expose the molecular mechanism of the rac5c-PLpro complex. Detailed molecular level insights of rac5c are elucidated in the dynamic state by quoting the pyridine fragment (with attached -OCH<sub>3</sub> group) loosely bound in the PLpro substrate binding pocket. From this, Huynh et al. suggested optimizing the loosely bound pyridine fragment with the alternative functional ground for the enhancement of binding affinity. Delre et al., have also analyzed the SARS-CoV-2 PLpro by repurposing the known compounds (688 phase III and 1,702 phase IV clinical trial drugs from the ChEMBL database) for Covalent and Non-covalent Inhibitors with desirable poly-pharmacology profiles. Delre et al. applied protein-ligand interaction fingerprint similarities, conventional docking scores, and MM-GBSA-binding free energies for executing the repurpose drugs through a covalent inhibition of PLpro.

Besides better known targets, there are also other equally important targets that can be used to fight the virus, as explored by several articles in this special issue. Yang et al. determined at a resolution of 2.0 Å the crystal structure of the SARS-CoV-2 nucleocapsid protein C-terminal domain (CTD). The CTD was shown to have a comparable distinct electrostatic potential surface to the equivalent domains of other reported CoVs, suggesting that the CTD has novel roles in viral RNA binding and transcriptional regulation. In particular, the crystal structure of the nucleocapsid CTD was analyzed, the potential

self-interaction formation of SARS-CoV-2 N-CTD was studied, and the self-interaction characteristics of the single-point mutant were verified. By studying the recognition mechanism of SARS-CoV-2 N-CTD protein to viral genomic intergenic transcriptional regulatory sequences (TRSs), what emerges is that the nucleocapsid protein CTD is responsible for the discontinuous viral transcription mechanism by recognizing the different patterns of viral TRS during transcription and revealing a new method of viral transcription sequences mechanism.

El Hassab et al. employed Computer-aided drug design (CADD) techniques for the identification of a novel inhibitor for SARS-CoV-2 RNA-dependent RNA polymerase. Such an approach indicates MAW-22 as a potential new inhibitor. MAW-22 demonstrated a strong binding affinity and energy profile for SARS-CoV-2 polymerase even better than the known antiviral drug remdesivir so to suggest that it could be used as an effective agent for the management of SARS-CoV-2 infection. Moreover, this study indicates that CADD is an efficient tool to develop drugs for treating SARS-CoV-2 infection. The aim of this study was not only to design a potential inhibitor but also to establish guidance for future drug development for COVID-19 infection.

Munaweera and Hu employed computational techniques to identify molecules that are able to inhibit SARS-CoV-2 nsp14 at the exoribonuclease (ExoN) site. Many nucleoside analogues (NuA) are known to lead to lethal mutagenesis for the viruses. The success of their activity can be made in vain by the proofreading activity of the ExoN. Thus, the simultaneous use of NuA and nsp14 inhibitors could enhance lethal mutagenesis in SARS-CoV-2. With this in mind, the authors built a homology model using the nsp14 of SARS-CoV as a template by molecular-docking, identified a potential lead molecule, PV6R, which belongs to DEDDh/DEEDh subfamily nuclease inhibitors, and can bind to the ExoN binding site of nsp14. Moreover, PV6R was computationally characterized and its molecular features were extracted and used to perform a virtual screening, by which different molecules were identified and successively optimized by computational strategies.

Squeglia et al. focused their attention on the host DEAD-box (DDX) RNA helicases, hijacked by coronaviruses to play key roles in viral replication steps. The highly conserved viral proteins responsible for DDX interactions probably use common pathways to exploit host proteins for their replication. In this review, the authors produce structural and functional data for considering DDXs as the possible key factors involved in SARS-CoV-2 hijacking mechanisms, exploring possible interactions between human DDX and coronavirus proteins, by integrating the available structural information with homology modeling studies. Furthermore, they hypothesize a double role related to the DDX helicases hijacking by coronaviruses: by enhancing key steps of the virus RNA replication/transcription and, simultaneously, by repressing the host innate immune response. Finally, DDX helicases could be considered novel targets for antiviral therapy also against SARS-CoV-2, as already validated for other RNA viruses.

Rampogu and Lee provided the structure-based pharmacophore modeling approaches for finding suitable inhibitors for SARS-CoV-2 2'-O-methyltransferase (nsp16/nsp10 complex) enzyme with FDA approved drug library compounds. The structure-based pharmacophore screening yields the three best compounds, which are better than the natural substrate S-adenosyl methionine (SAM), and inhibitors like remdesivir and hydroxychloroquine. From the pharmacophore based molecular docking and MD simulation approaches, Rampogu and Lee suggested that framycetin, kanamycin, and tobramycin could be strong and potential SARS-CoV-2 2'-O-methyltransferase inhibitors.

Žerovnik addresses a mini review on the Pore-forming proteins (PFPs) that virtually appear in all organisms and which cause ion dis-balance, small substances, or even protein efflux/influx, influencing a cell's signaling routes and fate by disrupting cellular membranes, depending on the pore size. In particular, these proteins are considered possible ways for therapy of channelopathies and/or modulating immunity relevant to the new threat of SARS-CoV-2 infections. Žerovnik summarized the current knowledge regarding the comparative features and mechanisms of pore formation by amyloid-forming proteins (AFPs), anti-microbial peptides (AMPs) and viroporins, and transmembrane short viral envelope proteins (E protein), helping spread certain viruses, among them the coronavirus SARS-CoV-2. The paper stresses that finding common mechanisms could be useful to design common means of defense and augment anti-viral and anti-amyloid therapies.

This Research Topic on "*Molecular Studies of Covid-19 Chemistry*" provides an overview of current knowledge and highlights interesting new insights into recognizing new targets and identifying or repurposing compounds that can combat the virus.

## AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct, and intellectual contribution to the work, and approved it for publication.

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Capasso, Di Gaetano, Selvaraj and Pedone. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Hijacking SARS-CoV-2/ACE2 Receptor Interaction by Natural and Semi-synthetic Steroidal Agents Acting on Functional Pockets on the Receptor Binding Domain

Adriana Carino<sup>1†</sup>, Federica Moraca<sup>2,3†</sup>, Bianca Fiorillo<sup>2</sup>, Silvia Marchianò<sup>1</sup>, Valentina Sepe<sup>2</sup>, Michele Biagioli<sup>1</sup>, Claudia Finamore<sup>2</sup>, Silvia Bozza<sup>4</sup>, Daniela Francisci<sup>4</sup>, Eleonora Distrutti<sup>5</sup>, Bruno Catalanotti<sup>2\*</sup>, Angela Zampella<sup>2</sup> and Stefano Fiorucci<sup>1</sup>

## OPEN ACCESS

### Edited by:

Chandrabose Selvaraj,  
Alagappa University, India

### Reviewed by:

Dhanachandra Singh Khurajam,  
Lerner Research Institute,  
United States  
Demetres D. Leonidas,  
University of Thessaly, Greece  
Dongqing Wei,  
Shanghai Jiao Tong University, China

### \*Correspondence:

Bruno Catalanotti  
bruno.catalanotti@unina.it

<sup>†</sup>These authors have contributed  
equally to this work

### Specialty section:

This article was submitted to  
Chemical Biology,  
a section of the journal  
Frontiers in Chemistry

Received: 15 June 2020

Accepted: 12 August 2020

Published: 23 October 2020

### Citation:

Carino A, Moraca F, Fiorillo B,  
Marchianò S, Sepe V, Biagioli M,  
Finamore C, Bozza S, Francisci D,  
Distrutti E, Catalanotti B, Zampella A  
and Fiorucci S (2020) Hijacking  
SARS-CoV-2/ACE2 Receptor  
Interaction by Natural and  
Semi-synthetic Steroidal Agents  
Acting on Functional Pockets on the  
Receptor Binding Domain.  
Front. Chem. 8:572885.  
doi: 10.3389/fchem.2020.572885

<sup>1</sup> Department of Surgical and Biomedical Sciences, University of Perugia, Perugia, Italy, <sup>2</sup> Department of Pharmacy, University of Naples Federico II, Naples, Italy, <sup>3</sup> Net4Science S.r.l., University "Magna Graecia" of Catanzaro, Campus Universitario "S. Venuta", Catanzaro, Italy, <sup>4</sup> Microbiology Section, Department of Medicine, University of Perugia, Perugia, Italy, <sup>5</sup> SC di Gastroenterologia ed Epatologia, Azienda Ospedaliera di Perugia, Perugia, Italy

The coronavirus disease 2019 (COVID-19) is a respiratory tract infection caused by the severe acute respiratory syndrome coronavirus (SARS)-CoV-2. In light of the urgent need to identify novel approaches to be used in the emergency phase, we have embarked on an exploratory campaign aimed at repurposing natural substances and clinically available drugs as potential anti-SARS-CoV2-2 agents by targeting viral proteins. Here we report on a strategy based on the virtual screening of druggable pockets located in the central  $\beta$ -sheet core of the SARS-CoV-2 Spike's protein receptor binding domain (RBD). By combining an *in silico* approach and molecular *in vitro* testing we have been able to identify several triterpenoid/steroidal agents that inhibit interaction of the Spike RBD with the carboxypeptidase domain of the Angiotensin Converting Enzyme (ACE2). In detail, we provide evidence that potential binding sites exist in the RBD of the SARS CoV-2 Spike protein and that occupancy of these pockets reduces the ability of the RBD to bind to the ACE2 consensus *in vitro*. Naturally occurring and clinically available triterpenoids such as glycyrrhetic and oleanolic acids, as well as primary and secondary bile acids and their amidated derivatives such as glyco-ursodeoxycholic acid and semi-synthetic derivatives such as obeticholic acid reduces the RBD/ACE2 binding. In aggregate, these results might help to define novel approaches to COVID-19 based on SARS-CoV-2 entry inhibitors.

**Keywords: SARS-CoV-2, COVID-19, virtual screening, nutraceuticals, drug repurposing and repositioning, bile acids, spike protein**

## INTRODUCTION

The coronavirus disease 2019 (COVID-19) is a respiratory tract infection caused severe acute respiratory syndrome (SARS)-CoV-2, a newly emerged coronavirus first identified in the city of Wuhan in China in December 2019 (Zhu et al., 2020). Globally, as of June 9, 2020 there have been more than ~7 million confirmed cases of COVID-19, including 404,396 deaths

(World Health Organization, 2020) in 216 countries (Fauci et al., 2020). Genetic sequencing SARS-CoV-2 demonstrates that the virus is a betacoronavirus sharing  $\sim 80\%$  genetic identity with SARS-CoV and MERS-CoV, identified in 2003 and 2012, respectively, and  $\sim 96\%$  identity with bat SARS-related CoV (SARS-CoV) RaTG13 (Wang et al., 2020b; Wrapp et al., 2020; Yan et al., 2020). Similarly to the 2003 and 2012 pandemics caused by these viruses (De Wit et al., 2016), the human infection caused by SARS-CoV-2 induces respiratory symptoms whose severity ranges from asymptomatic/poorly symptomatic to life threatening pneumonia and a cytokine related syndrome that might be fatal (Guan et al., 2020; Zou et al., 2020).

It is well-established that, similarly to SARS-CoV, SARS-CoV-2 enters the host cells by hijacking the human angiotensin converting enzyme receptor (ACE2) (Gui et al., 2017; Yuan et al., 2017; Walls et al., 2019, 2020; Wang et al., 2020b; Yan et al., 2020). The interaction of the virus with ACE2 is mediated by the transmembrane spike (S) glycoprotein, which shares 80% of the amino acid sequence identity with SARS-CoV and 97.2% of sequence homology with the bat SARS-CoV-RaTG13. In the case of SARS-CoV and SARS-CoV-2, the spike glycoprotein (S protein) on the virion surface mediates receptor recognition and membrane fusion (Lu et al., 2020). In the intact virus, the S protein assembles in a trimeric structure protruding from the viral surface. Each monomer of the trimeric S protein has a molecular weight of  $\approx 180$  kDa and contains two functional subunits, S1 and S2 that mediate, respectively, the attachment to ACE2 and the membrane fusion. The S1 binds to the carboxypeptidase domain of ACE2 with a dissociation constant (Kd) of  $\sim 15$  nM (Hoffmann et al., 2020).

Structural analysis has demonstrated that the N- and C-terminal portions of S1 fold as two independent domains, N-terminal domain (NTD) and C-terminal domain (CTD), with the latter corresponding to the receptor-binding domain (RBD) (Wang et al., 2020b). According to the high-resolution crystal structure information available so far, the RBD moves like a hinge between two conformations (“up” or “down”) to expose or hide the residues binding the ACE2. Within the RBD, there is a receptor binding motif (RBM), containing two binding loops separated by a short  $\beta$ -sheet, which makes the primary contact with the carboxypeptidase domain of ACE2. Importantly, while amino acid alignment studies have shown that the RBD of SARS-CoV-2 shares 73.5% homology with SARS-CoV, the identity of RBM, the most variable region of RBD, is significantly lower ( $\sim 50\%$ ) making it unclear whether the RBMs of the two viruses can induce cross-reactive antibodies. The region outside the RBM is thought to play an important role in maintaining the structural stability of the RBD.

The entry of SARS-CoV-2 in the host cells requires the cleavage of the S protein, a process that takes place in two steps. After binding to ACE2, the S protein is cleaved between the S1 and S2 subunits by a camostat-sensitive transmembrane serine protease, TMPRSS2 (Li et al., 2003; Lan et al., 2020; Shang et al., 2020; Wang et al., 2020b). Unlike SARS-CoV, SARS-CoV-2 has a distinct furin cleavage site (Arg-Arg-Ala-Arg) between the S1 and S2 domains, at residues 682–685, which may explain some of the biological differences. This furin cleavage site expands the

versatility of SARS-CoV-2 for cleavage by cellular proteases and potentially increases the tropism and transmissibility owing to the wide cellular expression of furin proteases especially in the respiratory tract (Belouzard et al., 2009; Ou et al., 2020). Cleavage at the S1/S2 site is essential to unlock the S2 subunit, which, in turn, drives the membrane fusion. Importantly, a second S2 site of cleavage has been identified at the S2' site which is thought essential to activate the protein for membrane fusion.

The spreading of the COVID-19 pandemic and the lack of effective therapies targeting the viral replication have prompted an impressive amount of investigations aimed at targeting several aspects of SARS-CoV-2 biology and viral interaction with ACE2. In this scenario, drug repurposing is a well-established strategy to quickly move already approved or shelved drugs to novel therapeutic targets, bypassing the time-consuming stages of drug development (Ghosh et al., 2020; Khan et al., 2020; Micholas and Jeremy, 2020). This accelerated drug development and validation strategy has led to numerous clinical trials for the treatment of COVID-19 (Li and De Clercq, 2020; Liu et al., 2020). Despite several encouraging results, however, treatment of SARS-CoV-2 infection remains suboptimal and there is an urgent need to identify novel approaches to be used in clinical settings.

One of such approaches is to prevent the S protein/ACE2 interaction as a strategy to prevent SARS-CoV-2 entry into target cells. Several virtual screening campaigns have already identified small molecules able to bind residues at the interface between the RBD of SARS-CoV-2 S protein and the ACE2 receptor (Ghosh et al., 2020; Micholas and Jeremy, 2020; Senathilake et al., 2020; Utomo et al., 2020; Wang et al., 2020a; Yan et al., 2020; Zhou et al., 2020). In this paper, we have expanded on this area. Our results demonstrate that several potential binding sites exist in the SARS CoV-2 S protein and that the occupancy of these pockets reduces the ability of the S protein RBD to bind to the ACE2 consensus *in vitro*. Together, these results might help to define novel treatments by using SARS-CoV-2 entry inhibitors.

## MATERIALS AND METHODS

### Virtual Screening

The electron microscopy (EM) model of SARS-CoV-2 Spike glycoprotein was downloaded from the Protein Data Bank (PDB ID: 6VSB). Missing loops were added from the Swiss-Model web-site (Wrapp et al., 2020). The obtained model was submitted to the Protein Preparation Wizard tool implemented into Maestro ver. 2019 (Schrödinger, 2019) to assign bond orders, adding all hydrogen atoms and adjusting disulfide bonds. The pocket search was performed by using the Fpocket website (Schmidtke et al., 2010).

The AutoDock4.2.6 suite (Morris et al., 2009) and the Raccoon2 graphical interface (Forli et al., 2016) were employed to carry out the virtual screening approach using the Lamarckian genetic algorithm (LGA). This hybrid algorithm combines two conformational search methods, the genetic algorithm and the local search. For the first low-accuracy screening, for each of the 2906 drugs, 3 poses were generated using 250,000 steps of genetic algorithm and 300 steps of local search, while in the second high-accuracy screening protocol, we generated 20 poses

for each ligand, increasing the number of genetic algorithm steps to 25,000,000. The MGLTools were used to convert both ligands and each pocket into appropriate pdbqt files. Virtual screening was performed on a hybrid CPU/GPU HPC cluster equipped with 2 NVIDIA® Tesla® V100 GPUs and 560 Intel® Xeon® Gold and 64 AMD® EPYC® processors.

Each of the six selected RBD pockets were submitted to the AutoGrid4 tool, which calculates, for each bonding pocket, maps (or grids) of interaction, considering the different ligands and receptor-atom types through the definition of a cubic box. Subsequently, for each grid AutoDock4 calculates interaction energies (ADscore) that express the affinity of a given ligand for the receptor.

The library of FDA approved drugs has been obtained both from DrugBank (2106 compounds) (Drugbank, 2020) and from the Selleckchem website (FDA-approved Drug Library, 2020) (tot. 2638). Each database was converted to 3D and prepared with the LigPrep tool (Schrödinger, 2019) considering a protonation state at a physiological pH of 7.4. Subsequently, the two libraries were merged and deduplicated with Open Babel (O'Boyle et al., 2011), giving a total amount of 2,906 drugs. The bile acids (BA) focused library was prepared with the same protocol described above. All the images are rendered using UCSF Chimera (Pettersen et al., 2004).

## Molecular Dynamics (MD)

MD simulations were performed using the CUDA version of the AMBER18 suite (Lee et al., 2018) on NVIDIA Titan Xp and K20 GPUs, using the Amber ff14SB force field (Maier et al., 2015) to treat the protein. RBD was then immersed in a pre-equilibrated octahedral box of TIP3P water and the system was neutralized. The system was then minimized using energy gradient convergence criterion set to 0.01 kcal/mol Å<sup>2</sup> in four steps involving: (i) an initial 5,000 minimization steps (2,500 with the steepest descent and 2,500 with the conjugate gradient) of only hydrogen atoms, (ii) 20,000 minimization steps (10,000 with the steepest descent and 10,000 with the conjugate gradient) of water and hydrogen atoms, keeping the solute restrained, (iii) 50,000 minimization steps (25,000 with the steepest descent and 25,000 with the conjugate gradient) of protein side chains, water and hydrogen atoms, (iv) 100,000 (50,000 with the steepest descent and 50,000 with the conjugate gradient) of complete minimization. Successively, the water, ions and protein side chains were thermally equilibrated in three steps: (i) 5 ns of NVT equilibration with the Langevin thermostat by gradually heating from 0K to 300K, while gradually rescaling solute restraints from a force constant of 10 to 1 kcal/mol Å<sup>2</sup>, (ii) 5 ns of NPT equilibration at 1 atm with the Berendsen thermostat, gradually rescaling restraints from 1.0 to 0.1 kcal/mol Å<sup>2</sup>, (iii) 5 ns of NPT equilibration with no restraints. Finally, a production run of 500 ns was performed using a timestep of 2 fs. The SHAKE algorithm was used for those bonds containing hydrogen atoms in conjunction with periodic boundary conditions at constant pressure and temperature, particle mesh Ewald for the treatment of long range electrostatic interactions, and a cutoff of 10 Å for nonbonded interactions.

## Dynamical Network Analysis

The Dynamical Network Analysis was performed on 500 ns long MD trajectories of the RBD domain using the plugin Carma ver. 0.8 (Glykos, 2006) implemented in VMD 1.9.2 (Humphrey et al., 1996). The optimal community distribution is calculated by using the Girvan–Newman algorithm (Girvan and Newman, 2002). Edges between each node (here defined as Cα atoms) were drawn between those nodes whose residues were within a default cut-off distance (4.5 Å) for at least 75% of our MD trajectories. Communities map analysis and representation were obtained using the NetworkView tool, implemented in VMD 1.9.2.

## Chemistry

OCA, BAR704, BAR501, and BAR502 were synthesized as previously described (Festa et al., 2014; Sepe et al., 2016).

## ACE2/SARS-CoV-2 Spike Inhibitor Screening Assay Kit

We tested the selected compounds (UDCA, T-UDCA, G-UDCA, CDCA, G-CDCA, OCA, BAR501, BAR502, BAR704, betulinic acid, oleanolic acid, glycyrrhetic acid, potassium canrenoate) using the ACE2: SARS-CoV-2 Spike Inhibitor Screening Assay Kit (BPS Bioscience Cat. number #79936) according to the manufacturer's instructions. All compounds were tested at different concentrations in a range from 0.01 to 100 μM. In addition, a concentration-response curve for the Spike protein (0.1–100 nM) was constructed to confirm a concentration-dependent increase in luminescence. A spike concentration of 5 nM was used for the screening of the compounds. Briefly, thaw ACE2 protein on ice and dilute to 1 μg/ml in PBS. Use 50 μL of ACE solution to coat a 96-well nickel-coated plate and incubate 1 h at room temperature with slow shaking. Wash the plate 3 times and incubate for 10 min with a Blocking Buffer. Next, add 10 μL of inhibitor solution containing the selected compound and incubate for 1 h at room temperature with slow shaking. For the “Positive Control” and “Blank,” add 10 μL of inhibitor buffer (5% DMSO solution). After the incubation, thaw SARS-CoV-2 Spike (RBD)-Fc on ice and dilute to 0.25 ng/μL (~5 nM) in Assay Buffer 1. Add the diluted Spike protein to each well, except to the blank. Incubate the reaction for 1 h at room temperature, with slow shaking. After 3 washes and incubation with a Blocking Buffer (10 min), treat the plate with an Anti-mouse-Fc-HRP and incubate for 1 h at room temperature with slow shaking. Finally, add an HRP substrate to the plate to produce chemiluminescence, which then can be measured using FluoStar Omega microplate reader.

In another experimental setting, we have tested the selected compounds using the ACE2: SARS-CoV-2 Spike Inhibitor Screening Assay Kit with a slight modification to the protocol. In particular, tested compounds were pre-incubated for 2 h with the Spike-RBD, and immediately afterwards the mix was incubated with ACE2 coated on the 96-well plate.

## Quantitative Analysis of the Anti-SARS-CoV-2 IgG Antibodies

To confirm the validity of the assay used in this study, five remnants of plasma samples used to test levels of anti-SARS CoV2 IgG in post COVID-19 patients were used. The original samples were collected at the blood bank of Azienda Ospedaliera of Perugia from post COVID-19 donors who participate to a program of plasma biobanking. An informed and written consent was signed by donors recruited in this program. The program's protocol included the quantitative analysis of the anti-SARS-CoV-2 IgG antibodies directed against the subunits (S1) and (S2) of the virus spike protein. IgGs were therefore measured by chemiluminescence immunoassay (CLIA) technology (LIAISON®SARS-CoV-2 IgG kit, DiaSorin®, Saluggia, Italy). Leftovers of five samples from this assay of  $\approx 40\text{--}50\ \mu\text{L}$  whose destiny was to be discharged were used to validate the SARS-CoV-2/ACE2 assay used in our study. While donors have provided a written informed consent for plasma donation as mentioned above, and no blood samples were taken specifically for this study, we (SB and DF) have contacted the five donors whose serum leftovers were used in this study by a phone call and asked the permission to use the sample remnants. The permission was granted by all five donors. We wish to thank all of them for the kind collaboration.

## RESULTS

### Virtual Screening of the FDA-Approved Drug Library

With the aim to identify chemical scaffolds capable of inhibiting ACE2/Spike interaction by targeting the RBD of the S1 domain of the SARS-CoV-2 (**Figure 1B**), we carried out a virtual screening campaign on an FDA-approved drug library, using the RBD 3D structure obtained from the Protein Data Bank (PDB ID 6SVB; Chain A, residues N331-A520) (Wrapp et al., 2020). Missing regions in the structure were built through the SwissModel webserver (Bertoni et al., 2017). A pocket search was performed with the Fpocket web-server (Le Guilloux et al., 2009), resulting in the identification of  $\approx 300$  putative pockets on the whole trimeric structure of the S protein. This search was further refined to identify selected pockets in the RBD according to three main factors: (i) the potential druggability, i.e., the possibility of interfering, directly or through an allosteric mechanism, with the interaction with ACE2; (ii) the flexibility degree of the pockets, i.e., excluding pockets defined, even partially, by highly flexible loops, whose coordinates were not defined in the experimental structure; (iii) sequence conservation with respect to SARS-CoV RBD (**Figure 1A**). On these bases, 6 pockets were selected on the RBD and numbered according to the Fpocket ranking (**Figures 1A,C**).

First, these pockets were used for the virtual screening of 2,906 FDA-approved drugs from the DrugBank and the Selleckchem websites, using the AutoDock4.2.6 program (Morris et al., 2009) and the Raccoon2 graphical user interface (Forli et al., 2016). This step was followed by a high-accuracy screening, based on the

binding affinity predicted by AutoDock4 (ADscore), with a focus on the results showing an ADscore lower than  $-6\text{ kcal/mol}$ .

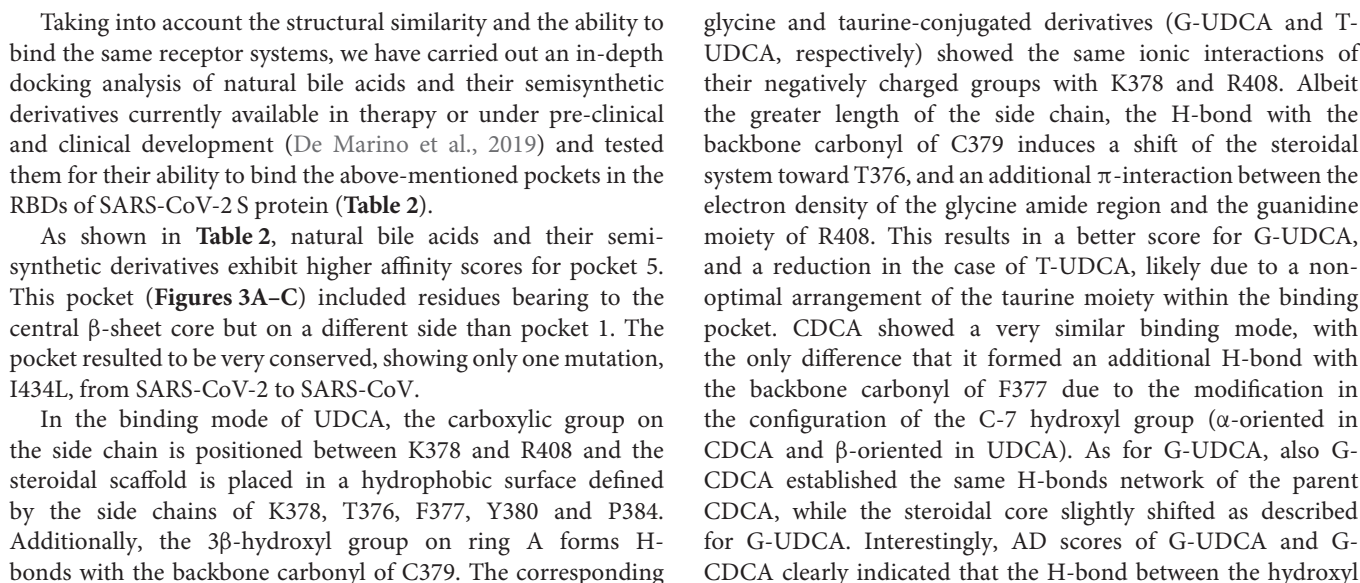
These studies allowed the identification of several compounds with steroidal and triterpenoid scaffold, including glycyrrhetic acid, betulinic acid and the corresponding alcohol (betulin), canrenone and the corresponding open form on the  $\gamma$ -lactone ring as potassium salt (potassium canrenoate), spironolactone and oleanolic acid, showing robust binding selectivity toward the RBD's pocket 1 (**Table 1**).

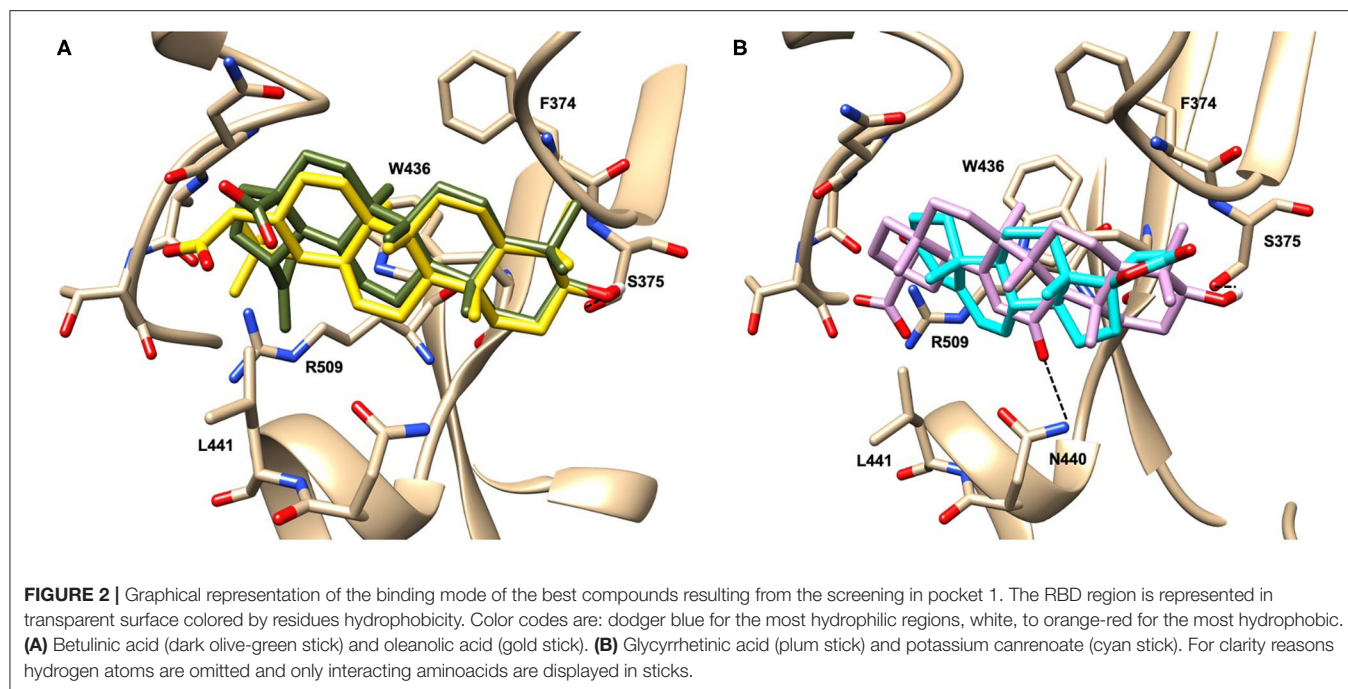
Pocket 1, located on the  $\beta$ -sheet in the central core of the RBD, is the less conserved among the screened, presenting five conservative (R346K, S438T, L440I, S442A) and two non-conservative (G445T and L451K) mutations from SARS-CoV-2 to SARS-CoV.

Glycyrrhetic acid, the best compound according to the AD score, binds the pocket through both hydrophobic and polar interactions. The triterpenoid scaffold relied between the hydrophobic side of the  $\beta$ -sheet core of RBD, defined by W436, F374 and the side chain of R509, and L441 on the other side, engaging hydrophobic contacts. In addition, the binding is reinforced by ionic contacts between the carboxyl group with R509, and by hydrogen bonds between the carbonyl group with N440 and the hydroxyl group with S375. Oleanolic acid and betulinic acid showed similar binding modes with the main difference in the carboxylic groups oriented toward the solvent. Finally, potassium canrenoate showed a different orientation of the steroidal system within the binding site, with the carboxylic function weakly bonded to S375 ( $3.1\ \text{\AA}$ ), and the  $\pi$ -system of rings A and B stacked between W436 and L441 (**Figure 2**).

Because the above mentioned triterpenoids have been identified as natural ligands for two bile acid activated receptors, the Farnesoid-X-Receptor (FXR) and G protein Bile Acid Receptor (GPBAR)-1 (Sepe et al., 2015; De Marino et al., 2019; Fiorucci and Distrutti, 2019), we have further investigated whether mammalian ligands of these receptors were also endowed with the ability to bind the above mentioned RBD's pockets. More specifically, oleanolic, betulinic and ursolic acids have been proved to act as selective and potent GPBAR1 agonists (Sato et al., 2007; Genet et al., 2010; Lo et al., 2016), while glycyrrhetic acid, the major metabolic component of licorice, and its corresponding saponin, glycyrrhizic acid, have been shown to act as dual FXR and GPBAR1 agonists in transactivation assay (Distrutti et al., 2015), also promoting GLP-1 secretion in type 1-like diabetic rats (Wang et al., 2017).

Bile acids are steroidal molecules generated in the liver from cholesterol breakdown (Fiorucci and Distrutti, 2019). Primary bile acids include cholic acid (CA) and chenodeoxycholic acid (CDCA), which have been recognized as functioning as the main FXR ligands in humans (Fiorucci and Distrutti, 2019). Secondary bile acids, deoxycholic acid and lithocholic acid (DCA and LCA) generated by intestinal microbiota, are preferential ligands for GPBAR1 (Maruyama et al., 2002; Fiorucci and Distrutti, 2019). Ursodeoxycholic acid (UDCA), which is a primary bile acid in mice, but a "tertiary" bile acid found in trace in humans, is, along with CDCA, the only bile acid approved for clinical use, and is a weak agonist for GPBAR1 and considered a neutral or weak antagonist toward FXR (Carino et al., 2019).





group at C-7 and F377 does not contribute significantly to the binding mode.

With respect to CDCA, the introduction of the ethyl group at the C-6 position as in OCA and in BAR704 improves the internal energy of the ligand ( $-0.27$  for CDCA vs.  $-0.59$  and  $-0.60$  kcal/mol for OCA and BAR704, respectively) and further favors the binding (**Figure 3B**), even if, albeit in close proximity of P384 and Y369, the 6-ethyl group did not show any particular contact within the RBD region.

BAR501, a neutral UDCA derivative, with an alcoholic side-chain end group and the ethyl group at C-6  $\beta$ -oriented showed a very similar binding mode compared to the parent compound, with the side chain hydroxyl group H-bonded to R408. Finally, BAR502, with a one carbon less on the side chain positioned the steroidal core as for G-CDCA, thus allowing the C-23 OH group H-bonding with the side chain hydroxyl group of T376.

## Dynamical Network Analysis

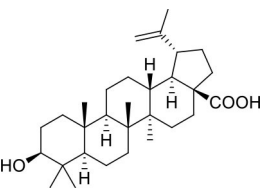
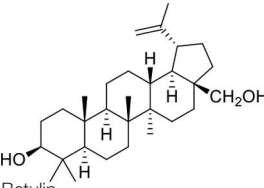
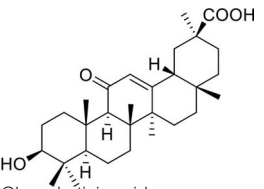
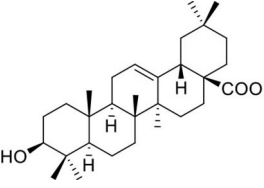
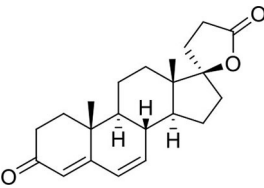
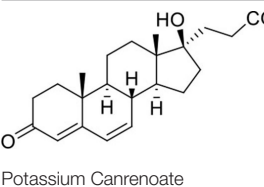
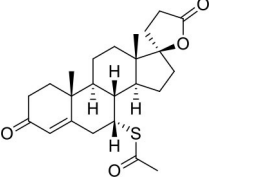
To support our hypothesis about the allosteric inhibitory potential of the identified pockets, we performed a dynamical network and community map analysis on 500 ns of molecular dynamics (MD) simulations of the RBD domain. Overall, the network analysis found 12 communities (Com1-Com12) (**Figures 4A–C** and **Table 3**). Each community corresponds to a set of residues in the RBD domain that move in concert with each other. By definition, nodes (defined here as the  $C\alpha$  atoms) belonging to the same community are highly interconnected, however, few nodes (called “critical”) may also connect to the edge of different communities by a metric called betweenness (**Figure 4C**). In our network analysis, the 12 communities identified are distributed as follows: the RBM region resulted in a split into three communities (Com4, Com6, and Com7),

with Com4 including the short  $\beta$ -sheet, while Com6 and Com7 include residues of the binding loops G496–Y505 and F456–F490 (**Table 3**), respectively. Pocket 1 and pocket 5 residues lie mainly in Com11 (**Table 3**), but few residues are included in other communities, in particular pocket 1 residue Y451 in Com4 and residues S438 and D442 in Com12, while pocket 5 residues T376, K378, C379, R408 in Com8 and Y380 Com10. In order to highlight the potential allosteric communication among the different communities, we analyzed the edge betweenness (**Figure 4C**), which is a measure of the shortest paths between pairs of nodes belonging to two different communities. We found that communities including residues of pocket 1 and pocket 5 indirectly communicate with Com6 and Com7, through Com4. In particular, Com8, Com10, Com11, and Com12, including most of the residues in both pockets 1 and 5, were connected to Com4, which in turn was strongly connected to Com6 and weakly to Com7, thus indicating at least a strong potential allosteric communication among the pockets and the loops at the receptor interface.

## In vitro Screening

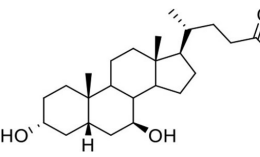
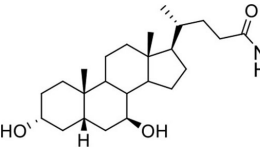
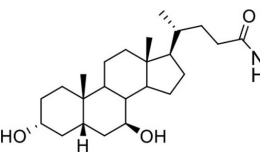
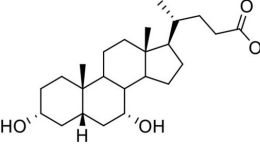
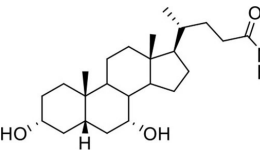
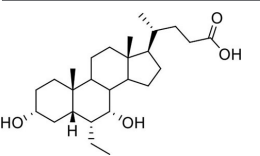
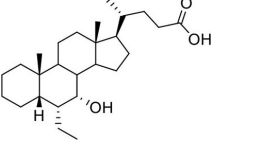
Given the results of the virtual screening, we have then investigated whether the agents mentioned in **Tables 1, 2** impact on the binding of S protein to the ACE2 receptor. For this purpose, a Spike/ACE2 Inhibitor Screening Assay Kit was used. The assay is designed for screening and profiling inhibitors for RBD/ACE2 interaction. To validate the assay, we first performed a concentration-response curve by adding increasing concentrations of the Spike RBD (0.1–100 nM) and confirmed a concentration-dependent increase of luminescence ( $n = 5$  experiments, **Figure 5A**). Since the curve was linear in the range from 0.1 to 10 nM, we have used the concentration of

**TABLE 1** | Results of the screening of FDA approved drugs on the RBD region of the Spike protein of SARS-CoV-2 with the Autodock 4.2.6 program.

Compound	ADscore	Pocket
	-8.1	1
Betulinic Acid		
	-7.4	1
Betulin		
	-8.6	1
Glycyrrhétinic acid		
	-8.2	1
Oleanolic acid		
	-7.9	1
Canrenone		
	-6.9	1
Potassium Canrenoate		
	-6.2	1
Spironolactone		

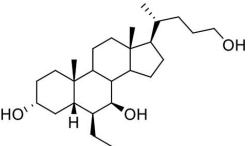
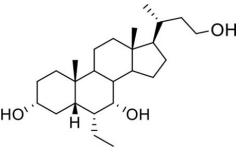
Binding affinity values (ADscore) are expressed in kcal/mol.

**TABLE 2** | Results of the screening of natural bile acids on the RBD region of the Spike protein of SARS-CoV-2 with the Autodock 4.2.6 program.

Compound	ADscore	Pocket
	-7.0	5
Ursodeoxycholic acid (UDCA)		
	-7.0	5
Tauro-ursodeoxycholic Acid (T-UDCA)		
	-7.3	5
Glyco-ursodeoxycholic Acid (G-UDCA)		
	-7.3	5
Chenodeoxycholic acid (CDCA)		
	-7.6	5
Glyco-chenodeoxycholic acid (G-CDCA)		
	-7.6	5
Obeticholic acid (OCA)		
	-7.2	5
BAR704		

(Continued)

TABLE 2 | Continued

Compound	ADscore	Pocket
 BAR501	-6.9	5
 BAR502	-7.3	5

Binding affinity values (ADscore) are expressed in kcal/mol.

5 nM for all the following assays. As illustrated in **Figure 5**, we found that incubating the Spike RBD with betulinic acid, glycyrrhetic acid, oleanolic acid, and potassium canrenoate (the active metabolite of spironolactone) results in concentration-dependent reductions of the binding of S Spike RBD to the ACE2 receptor. While all agents effectively reversed the binding at a concentration of 10  $\mu$ M, betulinic acid and oleanolic acid showed a significant inhibition at a concentration of 0.1 and 1  $\mu$ M, respectively ( $n = 3$  replicates).

Because these data demonstrate that betulinic acid and oleanolic acid were effective in inhibiting the binding of the S protein RBD to ACE2, and the two triterpenoids were known for their ability to modulate GPBAR1, we then tested whether natural GPBAR1 bile acids ligands were also effective in reducing the SARS-CoV-2-ACE2 interaction. As illustrated in **Figure 6**, the secondary bile acid UDCA and its taurine conjugate, T-UDCA, caused a slight and dose dependent inhibition of the bind of the S protein RBD to the ACE2 receptor (**Figures 6A,B**). G-UDCA, i.e., the main metabolite of UDCA in humans, inhibits the RBD binding to the ACE2 receptor by  $\sim 20\%$  in a concentration dependent manner. Similar concentration dependent effects were observed with CDCA and to a greater extent with its metabolite, G-CDCA (**Figure 6D**). A combination of UDCA and G-CDCA exerted a slight additive effect, confirming that UDCA itself has a very limited inhibitory activity.

Continuing the *in vitro* screening, we investigated whether the semisynthetic bile acid derivatives obeticholic acid (OCA), BAR704, BAR501, and BAR502, exerted comparable or better effects than G-CDCA. As illustrated in **Figure 7**, adding OCA to the incubation mixture reduced the binding of SARS-CoV-2 S spike to ACE2 by  $\approx 20\%$ . In contrast, BAR704, a 3-deoxy 6-ethyl derivative of CDCA, and a highly selective and potent FXR agonist, was significantly more effective and reduced the binding by  $\sim 40\%$  at the dose of 10  $\mu$ M. On the other hand, BAR501 and BAR502, alcoholic derivatives of UDCA and CDCA, respectively,

were only slightly effective in reducing the binding of S protein RBD to ACE2.

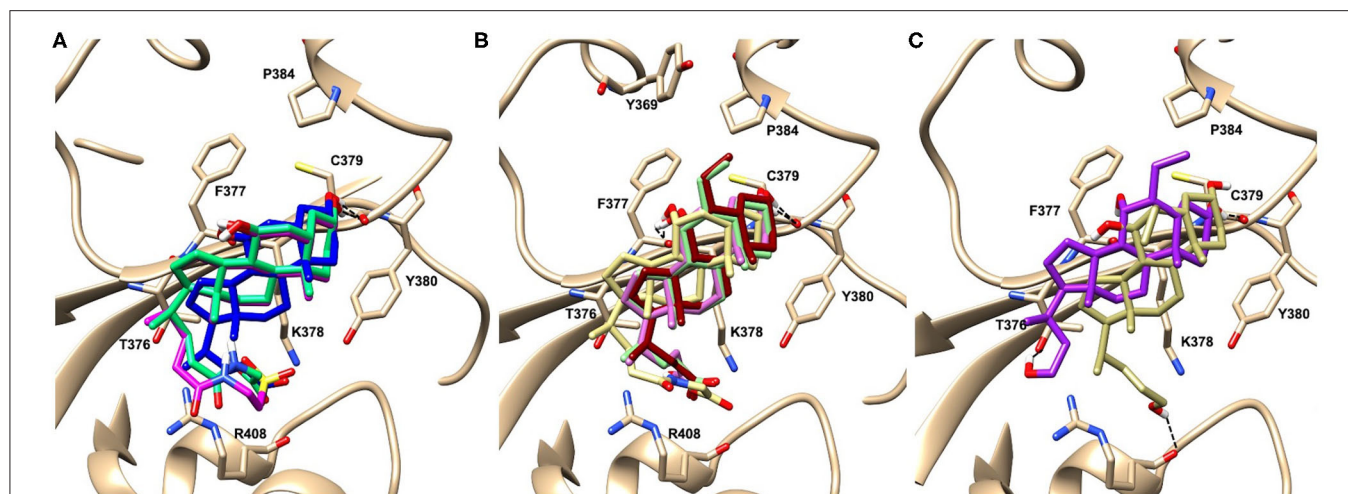
To further confirm our results, additional *in vitro* experiments were carried by pre-incubating the Spike RBD alone with 10  $\mu$ M of selected compound. As shown in **Figure 8**, several of the compounds exhibited a greater ability to reduce the interaction between Spike and ACE2 when pre-incubated with Spike-RBD compared with the standard incubation performed in the same experiment (**Figures 8A–M**,  $*p < 0.05$ ). In particular, we found that oleanolic and glycyrrhetic acid reduced the binding of Spike-RBD to ACE2 by 40% when pre-incubated with the RBD, whereas betulinic acid and potassium canrenoate showed no additional gain (**Figures 8A–D**,  $*p < 0.05$ ). Several natural bile acids, such as UDCA, T-UDCA, CDCA and G-CDCA, exerted a greater inhibitory effect when preincubated with Spike reaching  $\sim 45$ –50% of binding inhibition (**Figures 8E–I**,  $*p < 0.05$ ). Among the semisynthetic bile acid derivatives, their pre-incubation with Spike-RBD improved the efficacy of OCA (40%) and BAR502 (45%) (**Figures 8J,K**,  $*p < 0.05$ ) and BAR704 that reduced the interaction ACE2/Spike-RBD by 55% (**Figure 8L**,  $*p < 0.05$ ). These results suggested that the reduction of Spike-ACE2 interaction is actually due to the binding of tested compounds with the residues of Spike-RBD, thus confirming the molecular docking results.

## Effects of Plasma Samples From Post-COVID-19 Convalescent Patients on Spike RBD –ACE2 Interaction

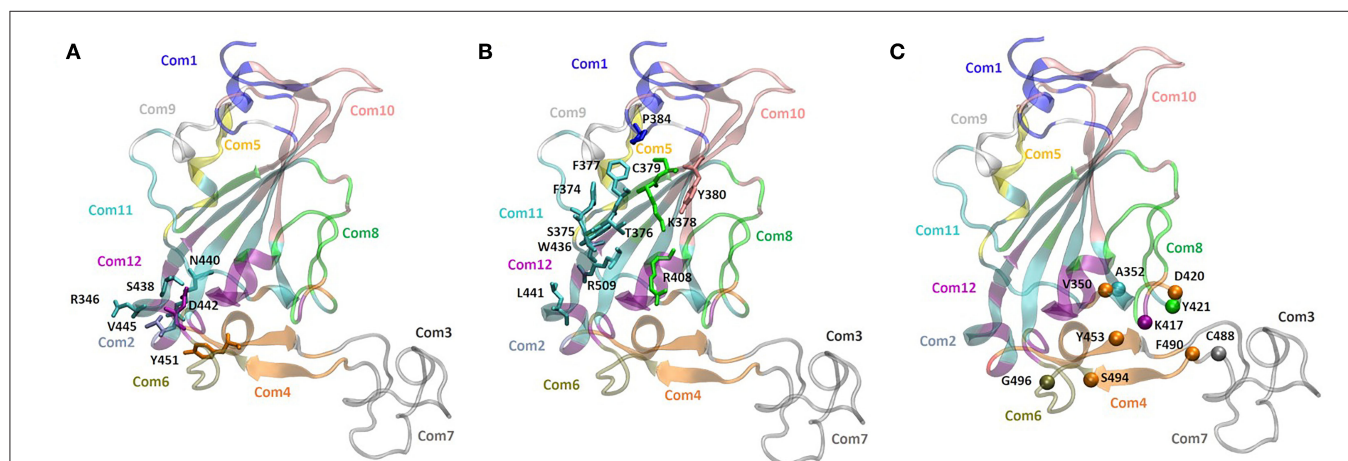
To confirm the concept that binding the pockets in the central  $\beta$ -sheet core of Spike RBD effectively prevents its interaction with the consensus of ACE2 receptor, we then carried out a set of control experiments using remnants of the plasma samples from five donors that have recovered from COVID-19. These donors had a slightly different title of anti SARS-CoV-2 antibodies (See Material and Methods, **Table 4**), but all the dilutions tested effectively inhibited the Spike RBD binding to ACE2 in our assay system by more than 95%. These data highlight that the test used in this paper correctly identify the binding of SARS-CoV-2 RBD to ACE2, but the levels of inhibition, were, as expected, significantly lower than those that could be reached by anti-SARS-CoV-2 antibodies.

## DISCUSSION

In this study we report the results of a virtual screening campaign designed to identify natural and clinically available compounds that might have utility in the prevention/treatment of the SARS-CoV-2 infection. In the light of the need of effective therapies to be rapidly tested for preventing or treating COVID-19, we initiated an *in silico* campaign to identify putative molecular targets that could be exploited to prevent the interaction of the SARS-CoV-2 Spike protein with the cellular machinery hijacked by the virus to enter target cells. To this end, we identified the Spike RBD as a potential pharmacological target. Accordingly, we developed the concept that putative pockets on



**FIGURE 3** | Graphical representation of the binding mode of the best compounds resulting from the screening in pocket 5. The RBD region is represented in tan cartoon, while the pocket 5 residues as transparent surface colored by residues hydrophobicity. Color codes are: dodger blue for the most hydrophilic regions, white, to orange-red for the most hydrophobic. **(A)** UDCA (blue stick), T-UDCA (magenta stick) and G-UDCA (spring-green stick); **(B)** CDCA (orchid stick), OCA (light-green stick); **(C)** BAR501 (gold stick) and BAR502 (purple stick). For clarity reasons hydrogen atoms are omitted and only interacting aminoacids are displayed in sticks.



**FIGURE 4** | Community network representation of the RBD domain and community residue members of **(A)** pocket 1 (N440, S438, R346, D442, V445, and Y451), **(B)** pocket 5 (Y380, K378, F377, R408, C379, T376, P384, F374, S375, W436, L441, and R509). **(C)** Highest score edge connectivity residues retrieved on the basis of the betweenness matrix. Spheres indicate the C $\alpha$  atoms of residues that occur in a majority of shortest paths connecting nodes in different communities.

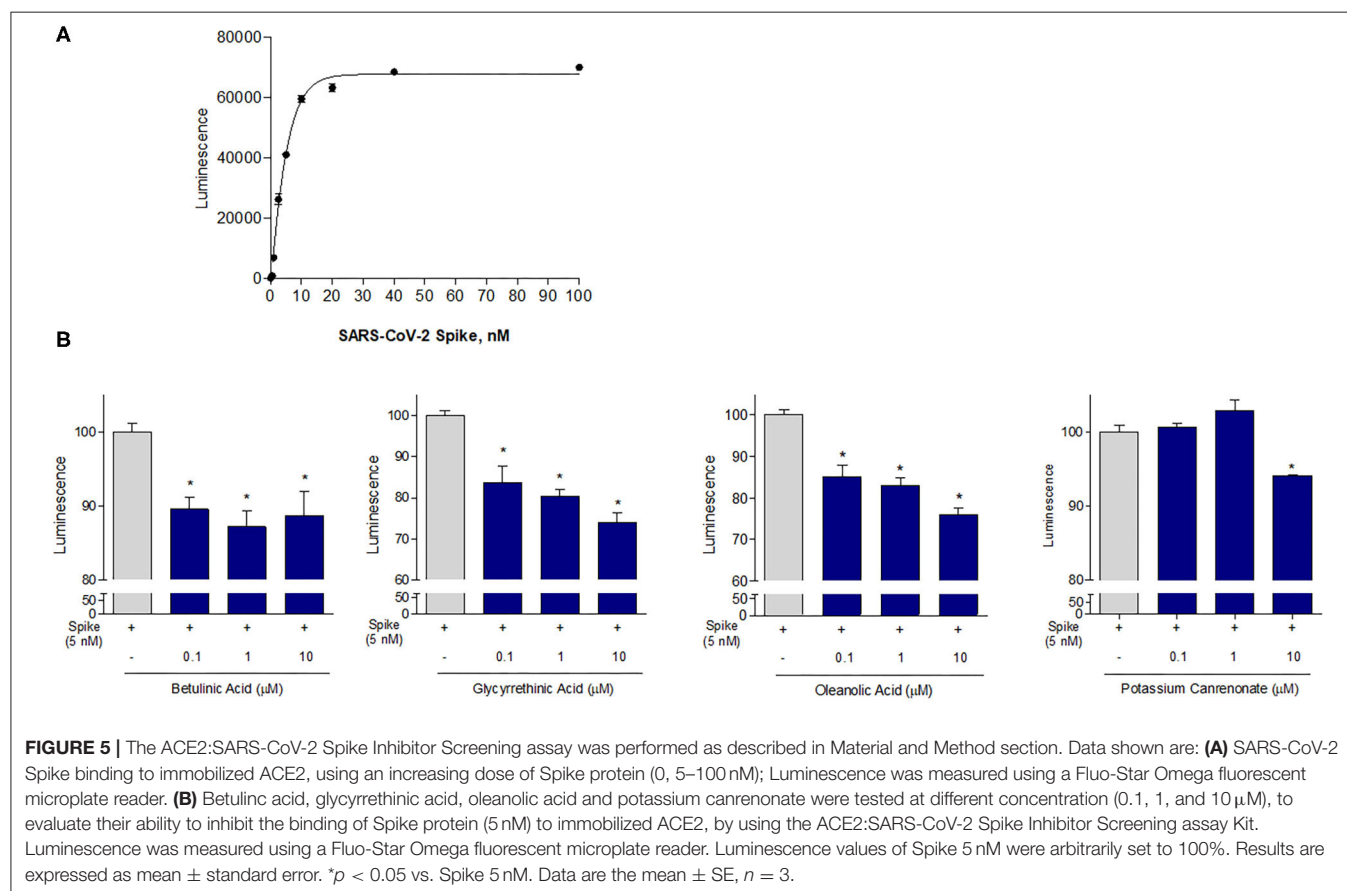
the surface of the central  $\beta$ -sheet core of the S protein RBD could be exploited eventually to prevent the binding of the virus to ACE2.

Our *in silico* screening has allowed the identification of six potentially druggable pockets and the virtual screening of the FDA-approved drug library identified steroidal compounds as potential hits against two pockets, namely pocket 1 and pocket 5. Interestingly, high accuracy docking demonstrated that flat steroidal scaffolds (i.e., A/B rings junction *in trans* configuration **Table 1**) prefer pocket 1, while compounds with the A/B junction in *cis* configuration (**Table 2**, such as bile acids) show greater affinity for pocket 5.

Our *in vitro* testing has largely confirmed the functional relevance of the two main pockets identified by *in silico* analyses. One important finding of this study has been that several steroidal molecules were effective inhibitors of the binding of the RBD to ACE2 *in vitro*. In particular, the most interesting compounds in **Table 1**, glycyrrhetic and oleanolic acid, showed good agreements in terms of docking AD score and in their ability to inhibit the spike/ACE2 interaction *in vitro*. The results also suggested that the main determinant for the inhibition efficacy is the hydrophobicity, as demonstrated by oleanolic acid, lacking any charge interaction within the pocket and resulting the most effective inhibitor in the series.

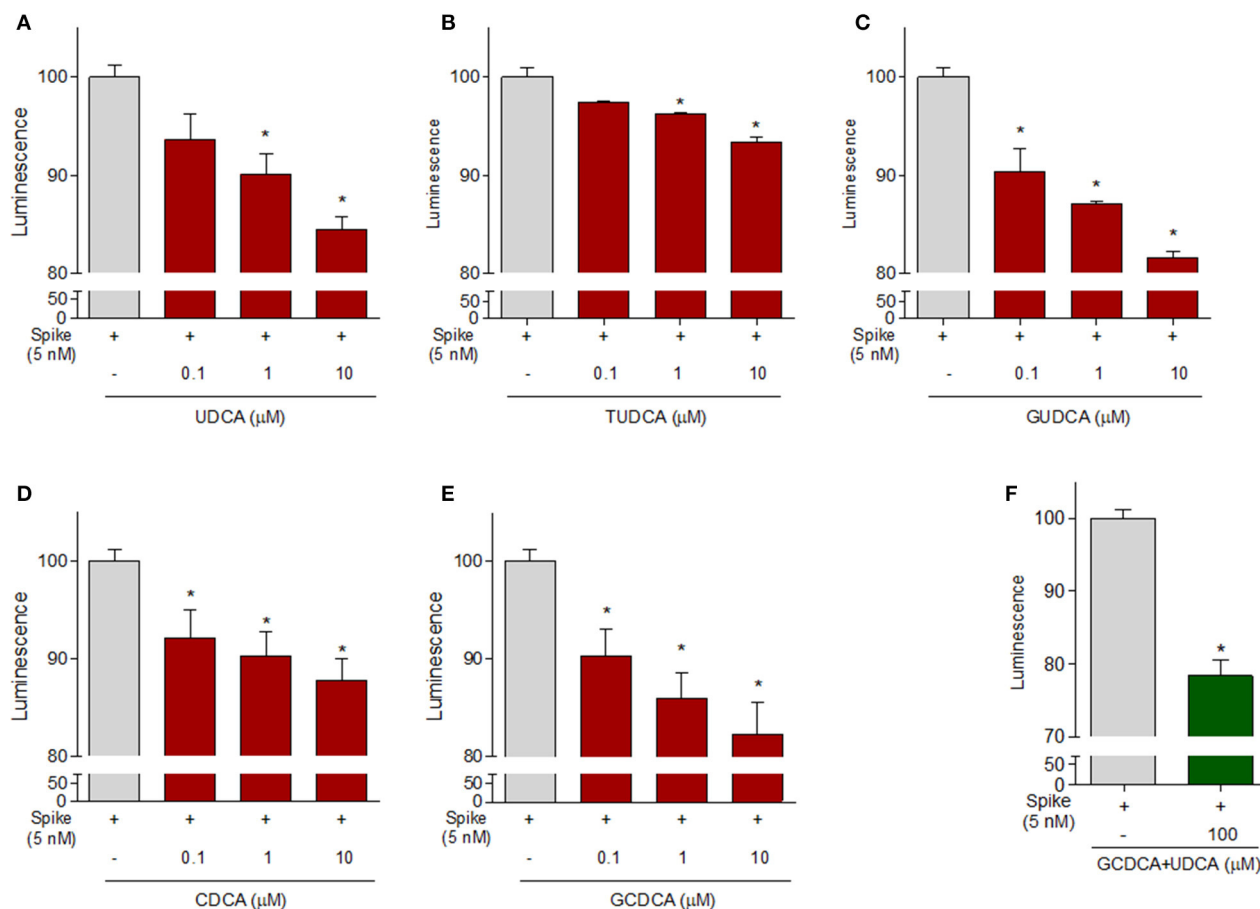
**TABLE 3** | Community map distribution of the RBD domain, retrieved after 500 ns-long MD simulation.

Community	N. of members	Residues	Color code cartoon
Com1	14	N334;C361;V382;P384;T385;L387;D389;V524-K529;	Blue
Com2	1	V445	Ice-blue
Com3	1	G476	Dark-gray
Com4	18	V350; G416; D420; G446-R454; F456; F490-S494;	Orange
Com5	7	C336; E340; F342; A344	Yellow
Com6	9	Y495-G502; G504	Tan
Com7	33	L455;R457-A475; S477-Y489	Light-Gray
Com8	20	T376;K378;C379;R408;I410-T415;I418;A419;Y421;Y423-P426;D428;T430;V511	Green
Com9	9	A363-Y369; S371; S383	White
Com10	30	L335;R355-N360;V362;Y380;G381;K386;L390-V395;D427-F429; L513-T523	Pink
Com11	38	V341;N343;T345;R346;Y351-N354;N370;A372-S375;F377;Y396-F400;N422;G431-A435;N437;N439-L441;S443;P507-V510;V512	Cyan
Com12	16	V401-V407;Q409;K417;W436;S438;D442;K444;V503;Y505;Q506	Purple

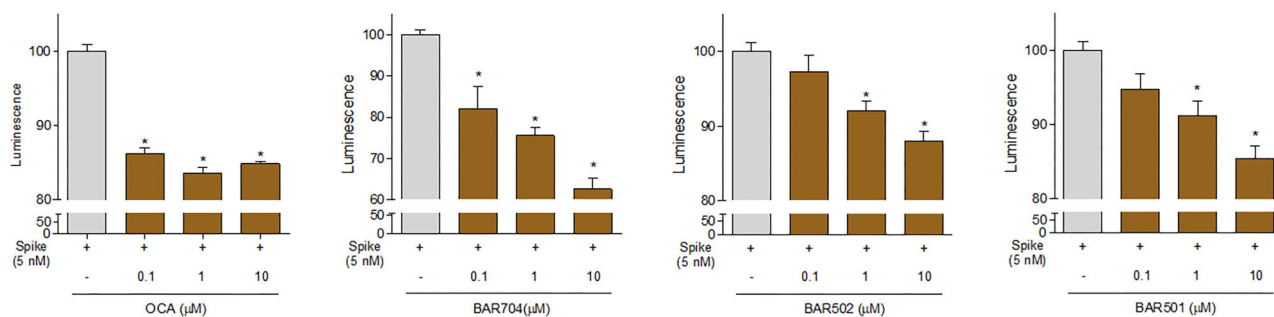


Hydrophobicity is also the main determinant of the activity of the bile acids and their semisynthetic derivatives, as demonstrated by CDCA, the corresponding glyco-conjugated derivative (G-CDCA) and its semisynthetic derivatives OCA, BAR704, and BAR502. Indeed, comparing the binding mode and the inhibition efficacy of CDCA and OCA with the related 6-ethyl derivative BAR704 highlighted the critical effect of the

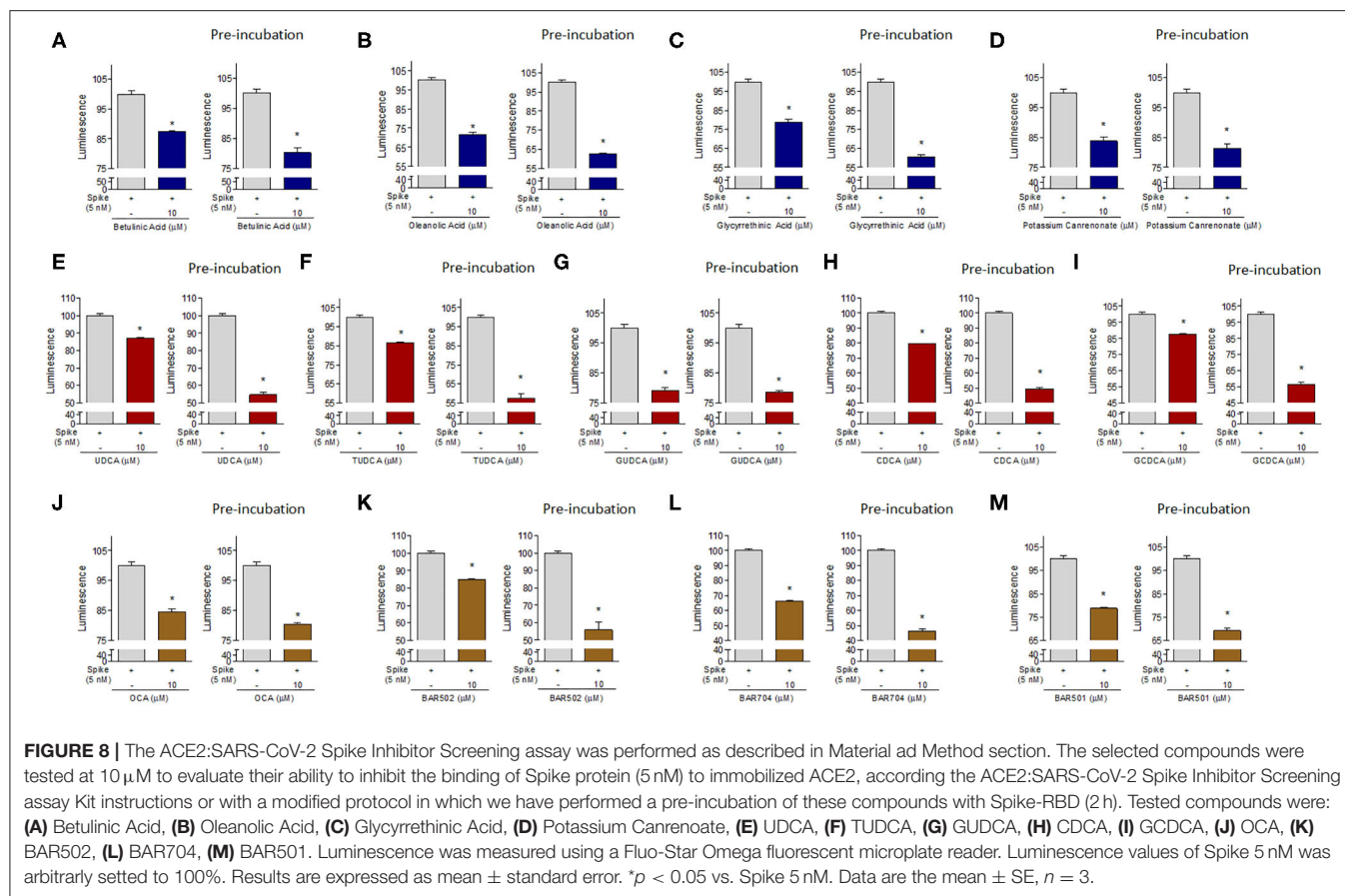
6 $\alpha$ -ethyl group in the inhibition activity and the negligible contribution of the 3 $\beta$ -hydroxyl group. The above positive effect could be explained considering the internal energy contribution of these ligands to the AD score, as well as the possibility of engaging more hydrophobic contacts. Indeed, the AD score internal energy contribution, significantly higher for the 6-ethyl derivatives, represents a measure of the conformational energy



**FIGURE 6 |** The ACE2:SARS-CoV-2 Spike Inhibitor Screening assay was performed as described in Material and Method section. Natural bile acids **(A)** UDCA, **(B)** TUDCA, **(C)** GUDCA, **(D)** CDCA, **(E)** GCDCA (0.1, 1 and 10  $\mu\text{M}$ ) and **(F)** a combination of GCDCA + UDCA (100  $\mu\text{M}$ ), were tested to evaluate their ability to inhibit the binding of Spike protein (5 nM) to immobilized ACE2, by using the ACE2:SARS-CoV-2 Spike Inhibitor Screening assay Kit. Luminescence was measured using a Fluo-Star Omega fluorescent microplate reader. Luminescence values of Spike 5 nM were arbitrarily set to 100%. Results are expressed as mean  $\pm$  standard error. \* $p < 0.05$  vs. Spike 5 nM. Data are the mean  $\pm$  SE,  $n = 3$ .



**FIGURE 7 |** The ACE2:SARS-CoV-2 Spike Inhibitor Screening assay was performed as described in Materials and Methods section. The semi-synthetic bile acid receptor agonists OCA, BAR704, BAR502, and BAR501, were tested at different concentration (0.1, 1, and 10  $\mu\text{M}$ ) to evaluate their ability to inhibit the binding of Spike protein (5 nM) to immobilized ACE2, by using the ACE2:SARS-CoV-2 Spike Inhibitor Screening assay Kit. Luminescence was measured using a Fluo-Star Omega fluorescent microplate reader. Luminescence values of Spike 5 nM were arbitrarily set to 100%. Results are expressed as mean  $\pm$  standard error. \* $p < 0.05$  vs. Spike 5 nM. Data are the mean  $\pm$  SE,  $n = 3$ .



**TABLE 4 |** Percentage of inhibition of the Spike:ACE2 binding.

Patient ID	Antibody Title	% of Binding Inhibition		
		5 $\mu$ L of Serum	10 $\mu$ L of Serum	20 $\mu$ L of Serum
1	96.6 AU/mL	98.6	99.5	99.6
2	170 AU/mL	99.3	99.4	99.3
3	89.4 AU/mL	98.1	99.3	99.4
4	125 AU/mL	98.8	99.3	99.4
5	146 AU/mL	95.7	96.9	97.3

Serum efficacy has been calculated in ACE2:SARS-CoV-2 Spike Inhibitor Screening Assay Kit as percent of inhibition of Spike RBD binding to ACE2 binding obtained using SPIKE at 5 nM, arbitrarily set as 100%.

of the bound vs. unbound state of the ligand, thus indicating that the ethyl group facilitates the assumption of the bioactive conformation. Moreover, the analysis of the binding mode of this compound highlighted that the 6-ethyl in the  $\alpha$ -position could establish hydrophobic contacts with P384 and Y369, positioned at a slightly longer distance than the optimal admitted for VdW interactions. However, it should be noted that the docking approach considers the protein receptor as rigid and didn't allow for mutual adaptation, which is an important process in ligand-receptor binding. In agreement with docking results, the lower efficacy observed for BAR502 could be explained with a slight

change in the binding mode, with a different position of the compound in the pocket in order to allow the hydroxyl group on a shortened side chain to interact with the side chain hydroxyl group of T376.

Moreover, also the comparison of the binding modes for G-CDCA and G-UDCA supported the hypothesis that the main determinant for the activity should be related to the network of hydrophobic interactions more than to the lack of a punctual hydrogen bond. Indeed, unlike the weakly active UDCA, the steroid core of G-UDCA is shifted to T376, and the resulting binding mode looks very similar to G-CDCA's. Finally, the better inhibitory efficacy of BAR501 with respect to UDCA, further confirmed the not-essential effect of the charged group on the side chain in terms of inhibition activity. Interestingly, the analysis of the binding mode of BAR501 also suggested that the stereochemistry of the ethyl group at C-6 is not pharmacophoric, being the 6 $\beta$ -ethyl group still able to potentially interact with P384 and Y369.

In the present study, we have developed a strategy to target the interaction of SARS-CoV-2 S protein RBD with the ACE2 receptor. As described in the introduction, SARS-CoV-2 enters the target cells by binding the carboxypeptidase domain of the ACE2 receptor, exposing a cleavage site, a hinge region between S1 and S2, to TMPRSS2, which in turn allows the S2 subunit of the Spike protein to bind with the cell membrane, leading to the

virus/host cells membrane fusion and SARS-CoV-2 penetration in to host cells.

The two pockets we have identified in the  $\beta$ -sheet core of the Spike RBD appear to be targetable by steroidal molecules and, importantly, we found that both naturally occurring bile acids and their metabolites in humans reduce the binding of Spike's RBD to ACE2. Of interest, natural bile acids, such as UDCA, T-UDCA, CDCA, and G-CDCA, exerted a greater inhibitory effect when preincubated with Spike reaching ~45-50% of binding inhibition. Importantly, we found that most of the agents tested in this study were agonists of two main bile acid activated receptors, i.e., the Farnesoid-x-Receptor (FXR) and a cell membrane receptor known as GPBAR1. Thus, betulinic acid and oleanolic acid, along with UDCA and its metabolites, BAR501 and BAR502 are effective ligands for GPBAR1. In contrast, glycyrrhethinic acid, CDCA, G-CDCA and T-CDCA, OCA and BAR704 are known for their ability to bind FXR (Festa et al., 2014). The fact that FXR/GPBAR1 agonists bind the SARS-CoV-2 RBD is of general interest and deserve further investigations.

Of interest, some of these agents have been reported for the potential use as anti-HIV agents (Rezanka et al., 2009), and oleanolic acid has been reported as a broad spectrum entry inhibitor of influenza viruses (Yang et al., 2018). On the other side, betulinic acid has been demonstrated to be useful in reducing inflammation and pulmonary edema induced by influenza virus (Hong et al., 2015), and potassium canrenoate, the main metabolite of spironolactone *in vivo*, is an anti-aldosteronic/diuretic used in the treatment of hypertensive patients. Finally, several GPBAR1 and FXR ligands, as bile acid derivatives, have been proved to exert beneficial effects in immune disorders (Fiorucci et al., 2018) and among these, BAR501, the first example of a C-6 $\beta$ -substituted UDCA derivative with potent and selective GPBAR1 activity, has been recently demonstrated as a promising lead in attenuating inflammation and immune dysfunction by shifting the polarization of colonic macrophages from the inflammatory phenotype M1 to the anti-inflammatory phenotype M2, increasing the expression of IL-10 gene transcription in the intestine and enhanced secretion of IL-10 by macrophages (Biagioli et al., 2017).

One important observation we have made in this study is that, while two different pockets of Spike RBD are potentially druggable, these are contiguous, and indeed, when we attempted drug combinations, none of these combinations effectively increased the anti-adhesive efficacy in comparison to the single agent.

This study has several limitations. First of all, we observed that the anti-adhesive efficacy of hyperimmune plasmas obtained from donors who have recovered from COVID-19 and containing high titles of neutralizing antibodies, in inhibiting the Spike RBD/ACE2 interaction, is close to 99%. This percentage is significantly higher than what we measured with our compounds. One possible explanation of this different efficacy can be found in terms of difference in affinity of our compounds with respect

to the antibodies but could also be related to the mechanism of allosteric connections suggested by dynamical network and community map analysis. Indeed pockets 1 and 5 resulted tightly connected with the loop G496-Y505, and weakly with the larger loop F456-F490. This suggests that small molecules binding the hydrophobic pockets are less effective than a neutralizing antibody. This also suggests that our pharmacological approach will likely be poorly effective in the presence of a high viral load, and the approach we have developed might have some efficacy only in the case of low viral load. Nevertheless, the mild inhibition efficacy showed by bile acids and their derivatives could pave the way for a further optimization of the binding mode in order to identify additional potential interactions, particularly in pocket 5, which has been demonstrated the least exposed to mutations.

Another limitation is that we have not tested the effect of these treatments on viral replication and further studies are needed to clarify this point.

In conclusion, in this paper, we report the identification of several potential binding sites in the RBD of the SARS-CoV-2 S protein. Several triterpenoids, such as glycyrrhethinic and oleanolic acids, and natural bile acids and their semisynthetic derivatives have been proven effective in reducing the Spike RBD's adhesion to its ACE2 consensus *in vitro*. Altogether, these results might help to define novel approaches to COVID-19 by using SARS-CoV-2 entry inhibitors.

## DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

## ETHICS STATEMENT

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. The patients/participants provided their written informed consent to participate in this study.

## AUTHOR CONTRIBUTIONS

SB and DF provided serum samples. BF, FM, and BC performed virtual screening and analyzed the data. CF and VS performed chemical synthesis. AC, SM, and MB generated the *in vitro* data and performed the data analysis. AZ, BC, ED, and SF conceived the study. All authors drafted the manuscript and wrote the final submission.

## ACKNOWLEDGMENTS

Authors wish to thank all the donors for the kind collaboration. This manuscript has been released as a PrePrint (Carino et al., 2020).

## REFERENCES

- Belouzard, S., Chu, V. C., and Whittaker, G. R. (2009). Activation of the SARS coronavirus spike protein via sequential proteolytic cleavage at two distinct sites. *Proc. Natl. Acad. Sci. U.S.A.* 106, 5871–5876. doi: 10.1073/pnas.0809524106
- Bertoni, M., Kiefer, F., Biasini, M., Bordoli, L., and Schwede, T. (2017). Modeling protein quaternary structure of homo- and hetero-oligomers beyond binary interactions by homology. *Sci. Rep.* 7:10480. doi: 10.1038/s41598-017-09654-8
- Biagioli, M., Carino, A., Cipriani, S., Francisci, D., Marchianò, S., Scarpelli, P., et al. (2017). The bile acid receptor GPBAR1 regulates the M1/M2 phenotype of intestinal macrophages and activation of GPBAR1 rescues mice from murine colitis. *J. Immunol.* 199, 718–733. doi: 10.4049/jimmunol.1700183
- Carino, A., Biagioli, M., Marchianò, S., Fiorucci, C., Zampella, A., Monti, M. C., et al. (2019). Ursodeoxycholic acid is a GPBAR1 agonist and resets liver/intestinal FXR signaling in a model of diet-induced dysbiosis and NASH. *Biochim. Biophys. Acta Mol. Cell Biol. Lipids* 1864, 1422–1437. doi: 10.1016/j.bbalip.2019.07.006
- Carino, A., Moraca, F., Fiorillo, B., Marchianò, S., Sepe, V., Biagioli, M., et al. (2020). Hijacking SARS-Cov-2/ACE2 receptor interaction by natural and semi-synthetic steroidal agents acting on functional pockets on receptor binding region. *bioRxiv [Preprint]*. doi: 10.1101/2020.06.10.144964
- De Marino, S., Festa, C., Sepe, V., and Zampella, A. (2019). “Chemistry and pharmacology of GPBAR1 and FXR selective agonists, dual agonists, and antagonists,” in *Handbook of Experimental Pharmacology* (New York, NY: Springer), 137–165. doi: 10.1007/164\_2019\_237
- De Wit, E., Van Doremalen, N., Falzarano, D., and Munster, V. J. (2016). SARS and MERS: recent insights into emerging coronaviruses. *Nat. Rev. Microbiol.* 14, 523–534. doi: 10.1038/nrmicro.2016.81
- Distrutti, E., Santucci, L., Cipriani, S., Renga, B., Schiaroli, E., Ricci, P., et al. (2015). Bile acid activated receptors are targets for regulation of integrity of gastrointestinal mucosa. *J. Gastroenterol.* 50, 707–719. doi: 10.1007/s00535-015-1041-8
- Drugbank (2020). Available online at: <https://www.drugbank.ca/releases/latest#structures> (accessed March 5, 2020).
- Fauci, A. S., Lane, H. C., and Redfield, R. R. (2020). Covid-19 - navigating the uncharted. *N. Engl. J. Med.* 382, 1268–1269. doi: 10.1056/NEJMe2002387
- FDA-approved Drug Library. (2020). Available online at: <https://www.selleckchem.com/screening/fda-approved-drug-library.html> (accessed March 4, 2020).
- Festa, C., Renga, B., D'Amore, C., Sepe, V., Finamore, C., De Marino, S., et al. (2014). Exploitation of choline scaffold for the discovery of potent and selective farnesoid X receptor (FXR) and G-protein coupled bile acid receptor 1 (GPBAR1) ligands. *J. Med. Chem.* 57, 8477–8495. doi: 10.1021/jm501273r
- Fiorucci, S., Biagioli, M., Zampella, A., and Distrutti, E. (2018). Bile acids activated receptors regulate innate immunity. *Front. Immunol.* 9:1853. doi: 10.3389/fimmu.2018.01853
- Fiorucci, S., and Distrutti, E. (2019). “The pharmacology of bile acids and their receptors,” in *Handbook of Experimental Pharmacology* (New York, NY: Springer), 3–18. doi: 10.1007/164\_2019\_238
- Forli, S., Huey, R., Pique, M. E., Sanner, M. F., Goodsell, D. S., and Olson, A. J. (2016). Computational protein-ligand docking and virtual drug screening with the AutoDock suite. *Nat. Protoc.* 11, 905–919. doi: 10.1038/nprot.2016.051
- Genet, C., Strehle, A., Schmidt, C., Boudjelal, G., Lobstein, A., Schoonjans, K., et al. (2010). Structure - activity relationship study of betulinic acid, a novel and selective TGR5 agonist, and its synthetic derivatives: potential impact in diabetes. *J. Med. Chem.* 53, 178–190. doi: 10.1021/jm900872z
- Ghosh, A. K., Brindisi, M., Shahabi, D., Chapman, M. E., and Mesecar, A. D. (2020). Drug development and medicinal chemistry efforts toward SARS-Coronavirus and Covid-19 therapeutics. *ChemMedChem* 15, 907–932. doi: 10.1002/cmdc.202000223
- Girvan, M., and Newman, M. E. J. (2002). Community structure in social and biological networks. *Proc. Natl. Acad. Sci. U.S.A.* 99, 7821–7826. doi: 10.1073/pnas.122653799
- Glykos, N. M. (2006). Software news and updates carma: a molecular dynamics analysis program. *J. Comput. Chem.* 27, 1765–1768. doi: 10.1002/jcc.20482
- Guan, W., Ni, Z., Hu, Y., Liang, W., Ou, C., He, J., et al. (2020). Clinical characteristics of coronavirus disease 2019 in China. *N. Engl. J. Med.* 382, 1708–1720. doi: 10.1056/NEJMoa2002032
- Gui, M., Song, W., Zhou, H., Xu, J., Chen, S., Xiang, Y., et al. (2017). Cryo-electron microscopy structures of the SARS-CoV spike glycoprotein reveal a prerequisite conformational state for receptor binding. *Cell Res.* 27, 119–129. doi: 10.1038/cr.2016.152
- Hoffmann, M., Kleine-Weber, H., Schroeder, S., Krüger, N., Herrler, T., Erichsen, S., et al. (2020). SARS-CoV-2 Cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. *Cell* 181, 271–280.e8. doi: 10.1016/j.cell.2020.02.052
- Hong, E. H., Song, J. H., Kang, K. B., Sung, S. H., Ko, H. J., and Yang, H. (2015). Anti-influenza activity of betulinic acid from Zizyphus Jujuba on influenza A/PR/8 virus. *Biomol. Ther.* 23, 345–349. doi: 10.4062/biomolther.2015.019
- Humphrey, W., Dalke, A., and Schulten, K. (1996). VMD: visual molecular dynamics. *J. Mol. Graph.* 14, 33–38. doi: 10.1016/0263-7855(96)00018-5
- Khan, A., Ali, S. S., Khan, M. T., Saleem, S., Ali, A., Suleman, M., et al. (2020). Combined drug repurposing and virtual screening strategies with molecular dynamics simulation identified potent inhibitors for SARS-CoV-2 main protease (3CLpro). *J. Biomol. Struct. Dyn.* 18, 1–12. doi: 10.1080/07391102.2020.1779128
- Lan, J., Ge, J., Yu, J., Shan, S., Zhou, H., Fan, S., et al. (2020). Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *Nature* 581, 215–220. doi: 10.1038/s41586-020-2180-5
- Le Guilloux, V., Schmidtke, P., and Tuffery, P. (2009). Fpocket: an open source platform for ligand pocket detection. *BMC Bioinformatics* 10:168. doi: 10.1186/1471-2105-10-168
- Lee, T. S., Cerutti, D. S., Mermelstein, D., Lin, C., Legrand, S., Giese, T. J., et al. (2018). GPU-accelerated molecular dynamics and free energy methods in amber18: performance enhancements and new features. *J. Chem. Inf. Model.* 58, 2043–2050. doi: 10.1021/acs.jcim.8b00462
- Li, G., and De Clercq, E. (2020). Therapeutic options for the 2019 novel coronavirus (2019-nCoV). *Nat. Rev. Drug Discov.* 19, 149–150. doi: 10.1038/d41573-020-00016-0
- Li, W., Moore, M. J., Vaslieva, N., Sui, J., Wong, S. K., Berne, M. A., et al. (2003). Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus. *Nature* 426, 450–454. doi: 10.1038/nature02145
- Liu, C., Zhou, Q., Li, Y., Garner, L. V., Watkins, S. P., Carter, L. J., et al. (2020). Research and development on therapeutic agents and vaccines for Covid-19 and related human coronavirus diseases. *ACS Cent. Sci.* 6, 315–331. doi: 10.1021/acscentsci.0c00272
- Lo, S. H., Cheng, K. C., Li, Y. X., Chang, C. H., Cheng, J. T., and Lee, K. S. (2016). Development of betulinic acid as an agonist of TGR5 receptor using a new *in vitro* assay. *Drug Des. Devel. Ther.* 10, 2669–2676. doi: 10.2147/DDDT.S113197
- Lu, R., Zhao, X., Li, J., Niu, P., Yang, B., Wu, H., et al. (2020). Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *Lancet* 395, 565–574. doi: 10.1016/S0140-6736(20)30251-8
- Maier, J. A., Martinez, C., Kasavajhala, K., Wickstrom, L., Hauser, K. E., and Simmerling, C. (2015). ff14SB: improving the accuracy of protein side chain and backbone parameters from ff99SB. *J. Chem. Theory Comput.* 11, 3696–3713. doi: 10.1021/acs.jctc.5b00255
- Maruyama, T., Miyamoto, Y., Nakamura, T., Tamai, Y., Okada, H., Sugiyama, E., et al. (2002). Identification of membrane-type receptor for bile acids (M-BAR). *Biochem. Biophys. Res. Commun.* 298, 714–719. doi: 10.1016/S0006-291X(02)02550-0
- Micholas, S., and Jeremy, C. S. (2020). Repurposing therapeutics for COVID-19: supercomputer-based docking to the SARS-CoV-2 viral spike protein and viral spike protein-human ACE2 interface. *ChemRxiv [Preprint]*. doi: 10.26434/chemrxiv.11871402.v4
- Morris, G. M., Huey, R., Lindstrom, W., Sanner, M. F., Belew, R. K., Goodsell, D. S., et al. (2009). AutoDock4 and AutoDockTools4: automated docking with selective receptor flexibility. *J. Comput. Chem.* 30, 2785–2791. doi: 10.1002/jcc.21256
- O'Boyle, N. M., Banck, M., James, C. A., Morley, C., Vandermeersch, T., and Hutchison, G. R. (2011). Open Babel. *J. Cheminform.* 3, 1–14. doi: 10.1186/1758-2946-3-33
- Ou, X., Liu, Y., Lei, X., Li, P., Mi, D., Ren, L., et al. (2020). Characterization of spike glycoprotein of SARS-CoV-2 on virus entry and its immune cross-reactivity with SARS-CoV. *Nat. Commun.* 11:1620. doi: 10.1038/s41467-020-15562-9

- Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., et al. (2004). UCSF Chimera - a visualization system for exploratory research and analysis. *J. Comput. Chem.* 25, 1605–1612. doi: 10.1002/jcc.20084
- Rezanka, T., Siristova, L., and Sigler, K. (2009). Sterols and triterpenoids with antiviral activity. *Antimicrob. Agents Med. Chem.* 8, 193–210. doi: 10.2174/187152109788680207
- Sato, H., Genet, C., Strehle, A., Thomas, C., Lobstein, A., Wagner, A., et al. (2007). Anti-hyperglycemic activity of a TGR5 agonist isolated from *Olea europaea*. *Biochem. Biophys. Res. Commun.* 362, 793–798. doi: 10.1016/j.bbrc.2007.06.130
- Schmidtke, P., Le Guilloux, V., Maupetit, J., and Tufféry, P. (2010). fpocket: online tools for protein ensemble pocket detection and tracking. *Nucleic Acids Res.* 38, W582–W589. doi: 10.1093/nar/gkq383
- Schrödinger (2019). *Schrödinger Release 2019-1: Maestro*. New York, NY: Schrödinger, LLC.
- Senathilake, K., Samarakoon, S., and Tennekoon, K. (2020). Virtual screening of inhibitors against spike glycoprotein of 2019 Novel Corona Virus: A Drug Repurposing Approach. 1. doi: 10.20944/preprints202003.0042.v1
- Sepe, V., Distrutti, E., Limongelli, V., Fiorucci, S., and Zampella, A. (2015). Steroidal scaffolds as FXR and GPBAR1 ligands: from chemistry to therapeutical application. *Future Med. Chem.* 7, 1109–1135. doi: 10.4155/fmc.15.54
- Sepe, V., Festa, C., Renga, B., Carino, A., Cipriani, S., Finamore, C., et al. (2016). Insights on FXR selective modulation. Speculation on bile acid chemical space in the discovery of potent and selective agonists. *Sci. Rep.* 6:19008. doi: 10.1038/srep19008
- Shang, J., Ye, G., Shi, K., Wan, Y., Luo, C., Aihara, H., et al. (2020). Structural basis of receptor recognition by SARS-CoV-2. *Nature*. 581, 221–224. doi: 10.1038/s41586-020-2179-y
- Utomo, R. Y., Ikawati, M., and Meiyanto, E. (2020). Revealing the potency of citrus and galangal constituents to halt SARS-CoV-2 infection. *Preprints. Org.* 2, 1–8. doi: 10.20944/preprints202003.0214.v1
- Walls, A. C., Park, Y. J., Tortorici, M. A., Wall, A., McGuire, A. T., and Veasler, D. (2020). Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein. *Cell* 181, 281–292.e6. doi: 10.1016/j.cell.2020.02.058
- Walls, A. C., Xiong, X., Park, Y.-J., Rey, A., Corti, D., and Correspondence, D. V. (2019). Unexpected receptor functional mimicry elucidates activation of coronavirus fusion. *Cell* 176, 1026–1039.e15. doi: 10.1016/j.cell.2018.12.028
- Wang, C., Wang, S., Li, D., Wei, D.-Q., Zhao, J., and Wang, J. (2020a). Human intestinal defensin 5 inhibits SARS-CoV-2 invasion by cloaking ACE2. *Gastroenterology* 159, 1145–1147.e4. doi: 10.1053/j.gastro.2020.05.015
- Wang, L. Y., Cheng, K. C., Li, Y., Niu, C. S., Cheng, J. T., and Niu, H. S. (2017). Glycyrrhizic acid increases glucagon like peptide-1 secretion via TGR5 activation in type 1-like diabetic rats. *Biomed. Pharmacother.* 95, 599–604. doi: 10.1016/j.biopha.2017.08.087
- Wang, Q., Zhang, Y., Wu, L., Niu, S., Song, C., Zhang, Z., et al. (2020b). Structural and functional basis of SARS-CoV-2 entry by using human ACE2. *Cell* 181, 894–904.e9. doi: 10.1016/j.cell.2020.03.045
- World Health Organization (2020). Available online at: <https://www.who.int/emergencies/diseases/novel-coronavirus-2019> (accessed July 9, 2020).
- Wrapp, D., Wang, N., Corbett, K. S., Goldsmith, J. A., Hsieh, C.-L., Abiona, O., et al. (2020). Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science* 367, 1260–1263. doi: 10.1126/science.abb2507
- Yan, R., Zhang, Y., Li, Y., Xia, L., Guo, Y., and Zhou, Q. (2020). Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2. *Science* 367, 1444–1448. doi: 10.1126/science.abb2762
- Yang, Y., He, H. J., Chang, H., Yu, Y., Yang, M. B., He, Y., et al. (2018). Multivalent oleanolic acid human serum albumin conjugate as nonglycosylated neomucin for influenza virus capture and entry inhibition. *Eur. J. Med. Chem.* 143, 1723–1731. doi: 10.1016/j.ejmech.2017.10.070
- Yuan, Y., Cao, D., Zhang, Y., Ma, J., Qi, J., Wang, Q., et al. (2017). Cryo-EM structures of MERS-CoV and SARS-CoV spike glycoproteins reveal the dynamic receptor binding domains. *Nat. Commun.* 8:15092. doi: 10.1038/ncomms15092
- Zhou, Y., Hou, Y., Shen, J., Huang, Y., Martin, W., and Cheng, F. (2020). Network-based drug repurposing for novel coronavirus 2019-nCoV/SARS-CoV-2. *Cell Discov.* 6:14. doi: 10.1038/s41421-020-0153-3
- Zhu, N., Zhang, D., Wang, W., Li, X., Yang, B., Song, J., et al. (2020). A novel coronavirus from patients with pneumonia in China, 2019. *N. Engl. J. Med.* 382, 727–733. doi: 10.1056/NEJMoa2001017
- Zou, L., Ruan, F., Huang, M., Liang, L., Huang, H., Hong, Z., et al. (2020). SARS-CoV-2 viral load in upper respiratory specimens of infected patients. *N. Engl. J. Med.* 382, 1177–1179. doi: 10.1056/NEJMc2001737

**Conflict of Interest:** This paper was supported by a research grant by BAR Pharmaceuticals S.r.l. to the Department of Pharmacy of the University of Napoli Federico II and to the Department of Surgical and Biomedical Sciences, University of Perugia. The authors declare the following competing financial interest(s): SF, AZ, and BC have filed an Italian patent application no.102020000011092 in the name of BAR Pharmaceuticals S.r.l. on the compounds described in this paper.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Carino, Moraca, Fiorillo, Marchianò, Sepe, Biagioli, Finamore, Bozza, Francisci, Distrutti, Catalanotti, Zampella and Fiorucci. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Identification of a New Potential SARS-COV-2 RNA-Dependent RNA Polymerase Inhibitor via Combining Fragment-Based Drug Design, Docking, Molecular Dynamics, and MM-PBSA Calculations

## OPEN ACCESS

### Edited by:

Emilia Pedone,  
Italian National Research Council, Italy

### Reviewed by:

Muhammad Tahir Ul Qamar,  
Guangxi University, China  
Tatsuya Takagi,  
Osaka University, Japan

### \*Correspondence:

Wagdy M. Eldehna  
wagdy2000@gmail.com  
Mahmoud A. El Hassab  
mahmoud65582@pharm.tanta.edu.eg

<sup>†</sup>These authors have contributed  
equally to this work

### Specialty section:

This article was submitted to  
Medicinal and Pharmaceutical  
Chemistry,  
a section of the journal  
Frontiers in Chemistry

**Received:** 18 July 2020

**Accepted:** 02 September 2020

**Published:** 30 October 2020

### Citation:

El Hassab MA, Shoun AA,  
Al-Rashood ST, Al-Warhi T and  
Eldehna WM (2020) Identification of a  
New Potential SARS-COV-2  
RNA-Dependent RNA Polymerase  
Inhibitor via Combining  
Fragment-Based Drug Design,  
Docking, Molecular Dynamics, and  
MM-PBSA Calculations.  
Front. Chem. 8:584894.  
doi: 10.3389/fchem.2020.584894

Mahmoud A. El Hassab<sup>1\*†</sup>, Aly A. Shoun<sup>2†</sup>, Sara T. Al-Rashood<sup>3†</sup>, Tarfah Al-Warhi<sup>4†</sup> and  
Wagdy M. Eldehna<sup>5\*†</sup>

<sup>1</sup> Department of Pharmaceutical Chemistry, School of Pharmacy, Badr University in Cairo (BUC), Cairo, Egypt, <sup>2</sup> Department of Microbiology and Immunology, Faculty of Pharmacy, Damanhour University, Damanhour, Egypt, <sup>3</sup> Department of Pharmaceutical Chemistry, College of Pharmacy, King Saud University, Riyadh, Saudi Arabia, <sup>4</sup> Department of Chemistry, College of Science, Princess Nourah Bint Abdulrahman University, Riyadh, Saudi Arabia, <sup>5</sup> Department of Pharmaceutical Chemistry, Faculty of Pharmacy, Kafrelsheikh University, Kafrelsheikh, Egypt

The world has recently been struck by the SARS-Cov-2 pandemic, a situation that people have never before experienced. Infections are increasing without reaching a peak. The WHO has reported more than 25 million infections and nearly 857,766 confirmed deaths. Safety measures are insufficient and there are still no approved drugs for the COVID-19 disease. Thus, it is an urgent necessity to develop a specific inhibitor for COVID-19. One of the most attractive targets in the virus life cycle is the polymerase enzyme responsible for the replication of the virus genome. Here, we describe our Structure-Based Drug Design (SBDD) protocol for designing of a new potential inhibitor for SARS-COV-2 RNA-dependent RNA Polymerase. Firstly, the crystal structure of the enzyme was retrieved from the protein data bank PDB ID (7bv2). Then, Fragment-Based Drug Design (FBDD) strategy was implemented using Discovery Studio 2016. The five best generated fragments were linked together using suitable carbon linkers to yield compound **MAW-22**. Thereafter, the strength of the binds between compound **MAW-22** and the SARS-COV-2 RNA-dependent RNA Polymerase was predicted by docking strategy using docking software. **MAW-22** achieved a high docking score, even more so than the score achieved by Remdesivir, indicating very strong binding between **MAW-22** and its target. Finally, three molecular dynamic simulation experiments were performed for 150 ns to validate our concept of design. The three experiments revealed that **MAW-22** has a great potentiality to inhibit the SARS-COV-2 RNA-dependent RNA Polymerase compared to Remdesivir. Also, it is thought that this study has proven SBDD to be the most suitable avenue for future drug development for the COVID-19 infection.

**Keywords:** COVID-19, polymerase inhibitors, fragment-based drug design, molecular dynamics, MM-PBSA calculations

## HIGHLIGHTS

- Structure-Based drug design approach suggested **MAW-22** as a potential SARS-CoV-2 polymerase.
- **MAW-22** demonstrated strong binding affinity and energy profile for SARS-CoV-2 polymerase better than Remdesivir.
- **MAW-22** could be used as an effective agent for management of SARS-CoV-2 infection.
- Computer aided drug design is an efficient tool to develop drugs for SARS-CoV-2 infection.

## INTRODUCTION

The World Health Organization (WHO) declared COVID-19 as a pandemic disease on 11th March 2020. Since that time, COVID-19 has continued to expand and spread. The severe acute respiratory syndrome (SARS CoV-2) virus causing the COVID-19 infection has now affected 216 countries, with 25,803,688 confirmed cases and 857,766 deaths reported globally (as of 1st September 2020) (Carlos et al., 2020; Perlman, 2020; She et al., 2020). The prefix “Corona” comes from the Latin word for “crown,” this named for the virus’ t crown-like appearance under electron microscope (Almeida et al., 1968; Tyrrell and Fielder, 2002). Coronaviruses, a family of RNA viruses in the order known as Nidovirales, are a positive-sense single-stranded RNA virus with medium size viruses ranging from 26 to 32 kb in length (Su et al., 2016). They are significant viral pathogens that affect animals and humans, causing viral pneumonia. There are four genera of coronaviruses (CoV): alpha, beta, gamma, and delta (Coronaviridae Study Group of the International Committee on Taxonomy of Viruses, 2020). Alpha and beta are responsible for seven coronaviruses that cause diseases in humans, while gamma and delta are pathogens in animals, not humans. SARS-COV-2 is the seventh coronavirus found to cause diseases in humans and is a novel Beta coronavirus (group 2B) (Zhu et al., 2020). Through genetic analysis, SARS-COV-2 has shown similarity to SARS-COV, with an 80% sequence homology similarity (Lu et al., 2020).

The first COVID-19 infections were estimated to be as early as November, 2019. Although studies are underway, it is still uncertain from where the SARS-CoV-2 pandemic began, however since December 2019, it has caused an outbreak of Acute Respiratory Distress syndrome (ARDS) and spread globally. On 30, January 2020, WHO declared a public health emergency due to SARS-COV-2. Studies showed that SARS-COV-2 originated from bats and translated to humans. Although the complete mechanism of transmission is unknown, it is likely that bats are the original source (Cavanagh, 2007). The subsequent cases involved family members of patients, healthcare workers, and finally human-to-human transmission (Chan et al., 2020). Moreover, transmission can be from symptomatic patients and also from asymptomatic individuals who are carrying the virus. Transmission could be conveyed through different modes of transmission. Although the exact modes are still unknown, it could be transmitted through respiratory droplets, and through exposure to sneezing or coughing from COVID-19 patients, through faces, and from close contact <2 m (Chan et al.,

2020). COVID-19 disease and the common cold share common symptoms, including cough, fever, shortness of breath, and fatigue, which makes it confusing to differentiate between the two (Armstrong et al., 2019; Dong et al., 2020; Wrapp et al., 2020). SARS-CoV-2 attacks the lower respiratory system, like SARS-CoV and MERS, causing pneumonia, moreover it can also attack the central nervous system, kidney, liver, and gastrointestinal system, causing multiple organ damage (Zhu et al., 2020).

Two other coronaviruses from the last two decades, middle east respiratory syndrome (MERS) and severe acute respiratory syndrome (SARS), which happened in 2012 (Zaki et al., 2012) and 2003 (Chan-Yeung and Xu, 2003), respectively, had much lower health impacts compared to the contagious SARS-Cov-2 (Drosten et al., 2003; Ksiazek et al., 2003; Chan et al., 2020). This devastating virus could be the most threatening outbreak the world has faced as it could threaten the life we know. To date, numbers of cases are growing with no approved antiviral agent specific for coronaviruses in humans. Although attempts have been carried out using different approved antivirals and immunomodulatory agents, they have not demonstrated acceptable efficacy in randomized clinical trials (Wang et al., 2020).

The SARS-CoV-2 genome is about ~30,000 nucleotides. It encodes four structural proteins—Nucleocapsid (N) protein, Membrane (M) protein, Spike (S) protein, and Envelop (E) protein—and several non-structural proteins (nsp) that are essential in the virus replication cycle (21 new). The virus enters human cells by attaching its spike protein to the human angiotensin converting enzyme 2 (ACE2) receptors present at the surface of numerous cells, like those of the lungs and GIT. Then, the Spike protein is subjected to proteolytic cleavages by host proteases to release the fusion peptide. This is followed by a cascade of cellular processes that ends with virus entry into the cytoplasm. After that, the virus is uncoated, releasing its single-stranded RNA genome into cytoplasm where the replication and transcription take place by the aid of the virus’ several non-structural proteins. Finally, the resulting proteins from the replication and transcription processes are assembled into new virions ready to infect new cells (Boopathi et al., 2020; Naqvi et al., 2020).

It is a difficult process to effectively develop specific COVID-19 direct acting antivirals (DAAs), however it is urgent to do so. From the above-mentioned life cycle of SARS-COV-2, several enzymes are considered potential targets for developing specific drugs against COVID-19, including non-structural protein 12 (nsp12), RNA Dependent RNA Polymerase (RdRp), 3C-like protease (3CL<sup>Pro</sup>), Papin like protease (PL<sup>Pro</sup>), and human angiotensin converting enzyme 2 receptor (ACE2) (Wang et al., 2020). 3C-like protease (3CL<sup>Pro</sup>) and Papin like protease (PL<sup>Pro</sup>) play crucial roles in the SARS-COV-2 replication cycle by processing the resulting polyprotein from the transcription stage into functioning subunits. Inhibiting any of the two proteases is believed to cease the virus replication cycle; blocking the angiotensin converting enzyme 2 receptor is also believed to prevent the virus entry by preventing the attachment of the SARS-COV-2 spike protein to the cell surface. Another attractive target to develop specific drugs for COVID-19 infection

is the multi-subunit machinery SARS-COV-2 RNA-dependent RNA polymerase (Ziebuhr, 2005). The nsp12 RNA dependent RNA polymerase plays a central role in the transcription and replication process by catalyzing the synthesis of viral RNA of COVID-19 with the assistance of two essential cofactors: non-structural protein 7 (nsp7) and non-structural protein 8 (nsp8) (Subissi et al., 2014). This is the reason that the broad-spectrum antiviral Remdesivir, targeting nsp12, was given great attention as an RNA-dependent RNA polymerase inhibitor for COVID-19 infection, and studies show good results for treating COVID-19 viral disease with this. In early 2020, Remdesivir was tested as a potential treatment for COVID-19. It has been approved for use in many countries like the USA, Japan, the UK, and Singapore (Grein et al., 2020; Lamb, 2020; Mehta et al., 2020). It is highly recommended in severe symptoms and in emergency cases, as it would reduce the time needed for full recovery from the virus (Drożdżal et al., 2020; Mehta et al., 2020).

Most of the conducted studies to discover inhibitors for COVID-19 infection aimed either to repurpose already approved drugs or to design novel compounds for the above-mentioned potential targets, and many of those studies obtained good outcomes by implementing various Computer-aided drug discovery techniques like docking, molecular dynamics, and Fragment-based drug design (FBDD), proving the power and importance of these techniques in the process of drug discovery. In this work, we report on the employment of Computer-aided drug design studies into the identification of a novel inhibitor for RNA-dependent RNA polymerase inhibitor of SARS-COV-2 (Alamri et al., 2020; Choudhury, 2020; ul Qamar et al., 2020).

## MATERIALS AND METHODS

### Fragment Based Drug Design (FBDD)

The crystal structure of SARS-COV-2 RNA-dependent RNA polymerase in complex with Remdesivir was retrieved from the protein data bank ID (7bv2) (Yin et al., 2020). A cavity surrounding the co-crystallized Remdesivir was constructed by Discovery studio 2016 (Dassault Systèmes BIOVIA, 2016). The cavity was extended to make good use of the entire active site. Then, a fragment-based approach was implemented using the *de novo* receptor strategy to fetch fragments from a library of fragments that fit properly into the active site of the enzyme (Böhm, 1992). The default Ludi fragments library found in the discovery studio, which contains 1,053 diverse fragments with molecular weights less than (300 KD), was used as the input source of fragments. The strength of binding of the retrieved fragments from the search was evaluated by docking to the receptor cavity using MCSS (Multiple Copy Simultaneous Search) (Cafilisch et al., 1993; Evensen et al., 1997). Finally, the successful fragments were linked together with suitable carbon linkers to produce compound **MAW-22** ([[(2S)-3-[2-amino-5-[(1R)-1-[6-amino-4-(dihydroxymethylene)-1H-pyridin-2-yl]propyl]-4-pyridyl]-2-carboxy-2-[(2-carboxybenzoyl)amino]propyl]-[(1S)-1-carboxypropyl]ammonium]). The binding strength of the generated compound was evaluated from the docking stage.

### *In silico* ADME and toxicity Calculations

The online Swiss ADME server (<http://www.swissadme.ch/index.php>) and Preadmet server (<https://preadmet.bmdrc.kr/>) were implemented to calculate the compound's physicochemical properties and toxicity, respectively.

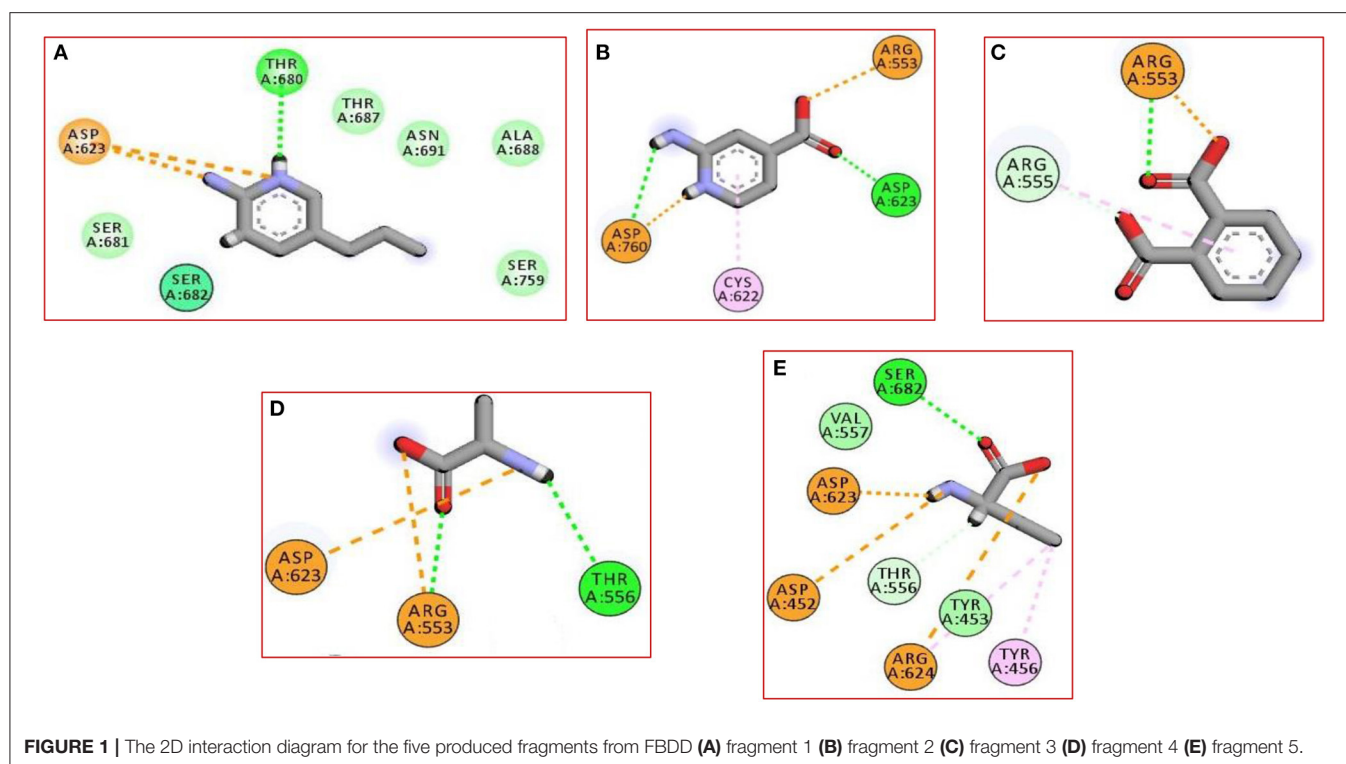
### Docking

The SARS-COV-2 RNA-dependent RNA polymerase retrieved from the protein data bank was utilized to conduct the docking study using Discovery Studio 2016. The active site was determined from the binding of Remdesivir by constructing a cavity surrounding the binding domain of the co-crystallized Remdesivir. The receptor was prepared by protein preparation wizard, while the ligands were prepared by the ligand preparation wizard. The receptor was energy minimized and equilibrated for 10 Nano seconds (ns) under GROMOS96 43a1 force field (see molecular dynamic section). Docking was commenced by C-Docker software found in the discovery Studio 2016 package in two steps (Wu et al., 2003). Step one involved the re-docking of Remdesivir to its corresponding receptor and the second step was the docking of the generated compound **MAW-22** from a fragment-based drug design stage. The docking results were visualized and analyzed by the Discovery studio visualizer available from Biovia Inc<sup>1</sup>.

### Molecular Dynamics

In the current study, we performed three molecular dynamic simulation experiments to support our concept of design. One experiment was conducted using the polymerase alone (none complexed with any ligand). The second and third experiments were for SARS-COV-2 RNA-dependent RNA polymerase enzyme in complex with Remdesivir and with the generated compound **MAW-22**, respectively. The entire MD simulation experiments were conducted using the latest version of GROningen MAchine for Chemical Simulations (GROMACS 2020.3) (Abraham et al., 2015). The receptor topology was obtained by the “pdb2gmx” script, while the ligand topologies were obtained by the CHARMM General Force Field CGENFF server and converted to the gromacs format using the “cgenff\_charmm2gmx\_py3\_nx2.py” script (Phillips et al., 2005). Each of the generated ligand topologies was rejoined to the processed receptor structure to construct the ligand-protein complex. GROMOS96 43a1 force field was used to obtain the energy minimized conformations of all the processed complexes (Chiu et al., 2009). After that, those complexes were solvated with a single point charge (SPC) water model to add water molecules to the cubic simulation boxes. Neutralization of the system net charges was done by adding counter-ions using the “gmX genion” script. Energy minimization of the unbound enzyme and the two complexes was achieved by employing the steepest descent minimization algorithm with a maximum of 50,000 steps and <10.0 kJ/mol force. Then, the solvated energy minimized structures were equilibrated with two consecutive steps. Firstly, NVT ensemble with a constant number of particles, volume, and temperature was done for 2 ns, followed by NPT ensemble with

<sup>1</sup><https://3dsbiovia.com/resource-center/downloads/>



a constant number of particles, pressure, and temperature for 8 ns. In the two systems, only the solvent molecules were allowed free movement to ensure its equilibration in the system, while other atoms were restrained. A constant temperature of 310 K and constant pressure of 1 atm were maintained through the entire MD simulation. The long-range electrostatic interactions were obtained by the particle mesh Ewald method with a 12 Å cut-off and 12 Å Fourier spacing (Bhardwaj et al., 2020). Finally, the three well-equilibrated systems (one empty protein and two protein-ligand complexes) entered the production stage without any restraints for 150 ns with a time step of 2 fs, and after every 5 ps the structural coordinates were saved. The root mean square deviation (RMSD) was calculated from the generated trajectories of the MD simulations as well as the distances of the formed hydrogen bonds between the receptor and the ligands by various scripts of GROMACS.

## MM-PBSA calculation

A common application in MD simulations and thermodynamic calculations is to determine the binding free energy of a protein-ligand complex. Generally, the binding free energy of protein and ligand complexes can be calculated using the molecular Mechanic/Poisson-Boltzmann Surface Area (MM-PBSA) alongside MD simulations using the following equation:

$$\Delta G_{\text{(Binding)}} = G_{\text{(Complex)}} - G_{\text{(Receptor)}} - G_{\text{(Ligand)}}$$

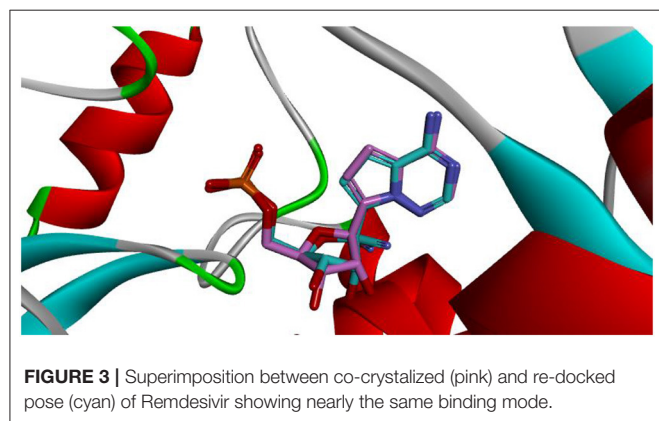
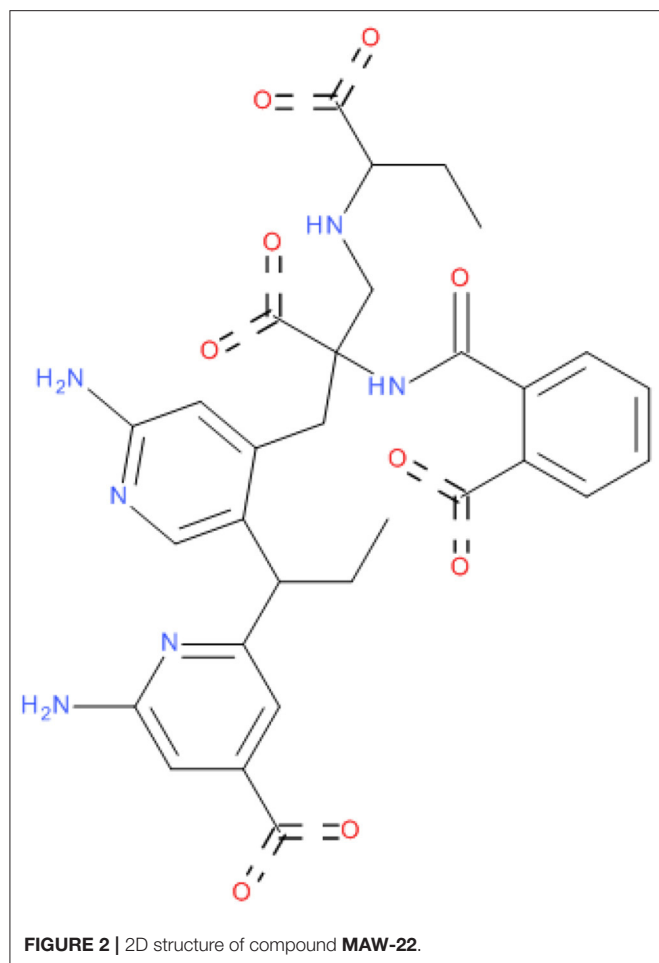
Where  $G_{\text{(complex)}}$  is the total free energy of the protein-ligand complex and  $G_{\text{(receptor)}}$  and  $G_{\text{(ligand)}}$  are total free energies of the isolated protein and ligand in solvent,

respectively. The total free energy of any of the three mentioned entities (complex or receptor or ligand) could be calculated from its molecular mechanics potential energy plus the energy of solvation. Thus, the “g\_mmpbsa” (Kumari and Kumar, 2014) package of GROMACS was used to perform MM-PBSA calculations through all the MD trajectories.

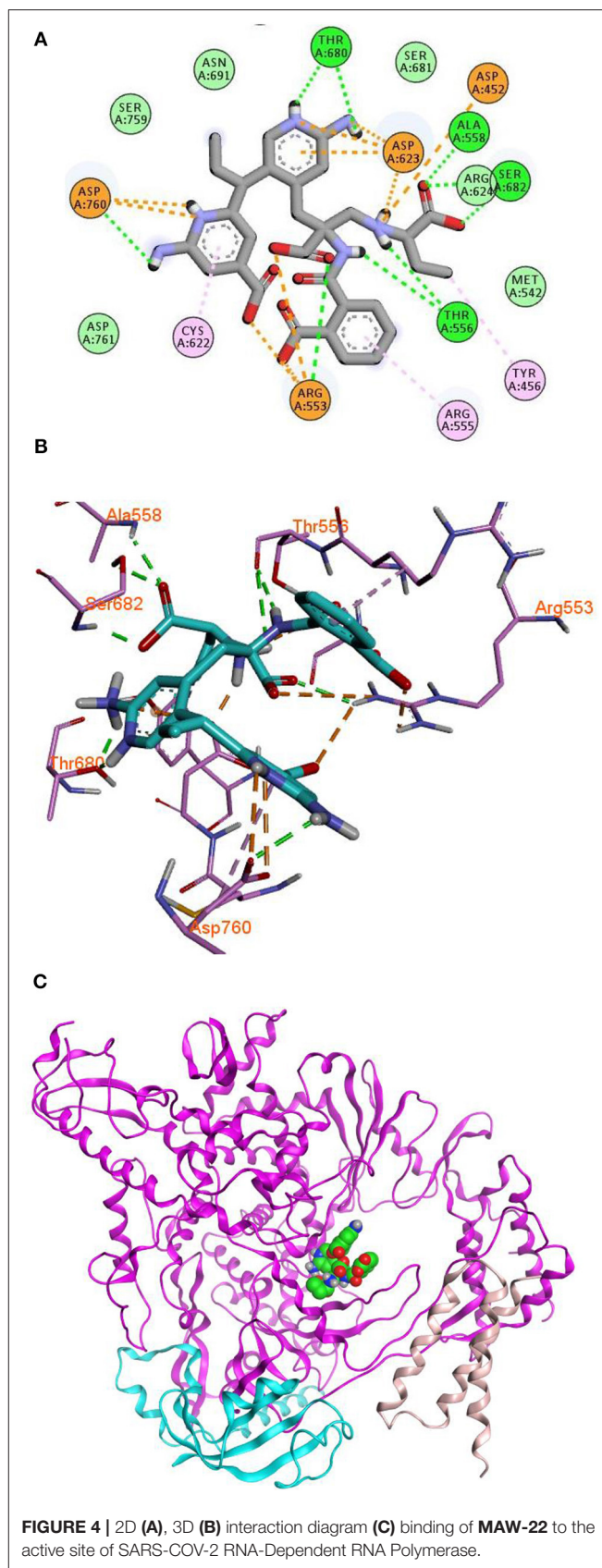
## RESULTS AND DISCUSSION

### FBDD

Fragment-based drug discovery is a very important technique in the field of drug design (de Kloe et al., 2009). The technique enables the discovery of novel drugs through the screening of fragments databases. The fragments should always be low in molecular weight, less than (300 KD), and chemically diverse (Kumar et al., 2012). FBDD is advantageous over other high throughput screening (HTS) methods because of three points. The first advantage is that the active site is better covered and fully explored in FBDD. This is attributed to small fragments that could be easily inserted in any space in the active site in contrast to large molecules from other HTS methods (Blum and Raymond, 2009; Roughley and Hubbard, 2011). The second point is that screening a fragment library achieves higher hit rates as compared to conventional HTS (Mortenson and Murray, 2011). However, in this study our focus was to design one compound with the maximum binding strength. The last advantage of FBDD over HTS is that compounds designed by FBDD achieve higher binding affinity than compounds designed by HTS (Kumar et al., 2012).



It is well-established that there are three strategies in FBDD: (a) Fragment growing (Kirsch et al., 2019), in which a fragment that obeys the role of three is increased in size to optimize the interaction of the proposed target; (b) Fragment merging (Kirsch et al., 2019), in which two fragments bound to the same regions in the binding site of the target are merged to give one compound; and (c) Fragment linking (Kirsch et al., 2019), in



which two or more fragments bound to different regions in the binding site of the target are linked together by a suitable linker to yield one compound. The active site was determined by a cavity surrounding the binding of Remdesivir. The *de novo* receptor wizard available in the discovery studio 2016 was employed to screen the fragment library. The technique subdivides the cavity, and all the fragments are screened through the entire binding domain (cavity). Energy estimate 3 was used as a scoring function. Successful fragments should achieve good interaction with the receptor and also make a negative change in the free energy upon binding to the receptor. The conducted search resulted in 821 fragments. Those fragments were further filtered through docking to the receptor cavity using the MCSS protocol. The protocol involves the following steps (Evensen et al., 1997): generating conformations for each fragment, screening of all the fragments through the binding cavity, CHARMM fragment minimization, and clustering and removing fragments that converge to similar positions. The resulted top fragment in each region in the binding site was selected to construct **MAW-22**.

In the current work, we did not have a lead compound or a fragment to apply the Fragment-growing approach, whereas the generated compound was designed from scratch. Also, all generated fragments were apart from each other, so the fragment linking strategy was more suitable than the fragment merging strategy. The binding of the selected fragments was strong and involved many types of interactions (Figure 1). Those fragments were then linked *via* a carbon linker to yield **MAW-22** (Figure 2).

### **In silico ADME and toxicity Calculations**

Any compound to be considered as a potential drug candidate should have acceptable pharmacokinetic and pharmacodynamic profiles, as well as a high safety margin. Thus, both Swiss ADME and Preadmet servers were employed to predict the physicochemical properties and potential toxicity of **MAW-22**. In general, both servers predicted excellent safety profiles of compound **MAW-22**, with no mutagenicity or carcinogenicity as predicted by the Preadmet server. On the other hand, Swiss ADME revealed that **MAW-22** has no inhibitory effect toward any hepatic Cytochrome enzyme, and thus it has neither hepatotoxicity nor drug-drug interactions and could be used concurrently with any drugs in a COVID-19 treatment protocol. Furthermore, both servers predicted no BBB penetration for **MAW-22**, so it could be safely used with no concerns about potential neurotoxicity. Another important advantage is that **MAW-22** would mostly have no teratogenic effect. This assumption is evidenced by that fact that placental penetration requires compounds to have molecular weights <500 Dalton, high Lipophilicity, and predominance of the non-ionized form (Griffiths and Campbell, 2015). **MAW-22** is a polar compound with logP equals -0.13 and molecular weight of 630 Dalton, as well as four ionized carboxylate groups. Thus, it would probably have no penetration with significant concentration through the placenta and could be safely given to pregnant woman with a COVID-19 infection. **MAW-22** violated the Lipinski rule only in the Molecular weight and the number of hydrogen bond donors. But when it comes to medicinal chemistry friendliness, **MAW-22**

has no alertness for either PAINS or Brenk (Veber et al., 2002; Baell and Holloway, 2010).

### **Docking**

Docking is the most widely used technique in drug design. In the current study, docking was inevitable as it is the only computational method able to predict the exact binding between the generated compound **MAW-22** and the SARS-COV-2 RNA-dependent RNA Polymerase. Yet there are two issues; the first is that docking is vulnerable to error and needs proper validation, and the second is that the docking results need to be compared to an experimental reference. These two points were addressed by re-docking Remdesivir to the SARS-COV-2 RNA-dependent RNA Polymerase. The calculated RMSD between the docked and co-crystallized poses of Remdesivir was 0.48, indicating a valid docking approach (Figure 3).

Remdesivir achieved a—C-Docker \_Energy score of 87.3, whereas **MAW-22** achieved a—C-Docker \_Energy score of 135.7. The high score achieved by **MAW-22** is well-matched with its strong binding and interactions with the SARS-COV-2 RNA-dependent RNA Polymerase. As shown in Figure 4, **MAW-22** was able to engage in a number of diverse interaction types with its target. For instance, the proposed compound was involved in hydrogen-bond interactions with residues THR556, ALA558, ARG553, THR680, SER682, and ASP760, in addition to ionic interactions with residues ASP452, ASP623, ASP760, and ARG553. Moreover, **MAW-22** achieved several hydrophobic interactions with residues CY5622, ARG555, and TYR456. The detailed interactions of **MAW-22** within the SARS-COV-2 RNA-dependent RNA Polymerase active site were summarized in Table 1.

### **Molecular Dynamics**

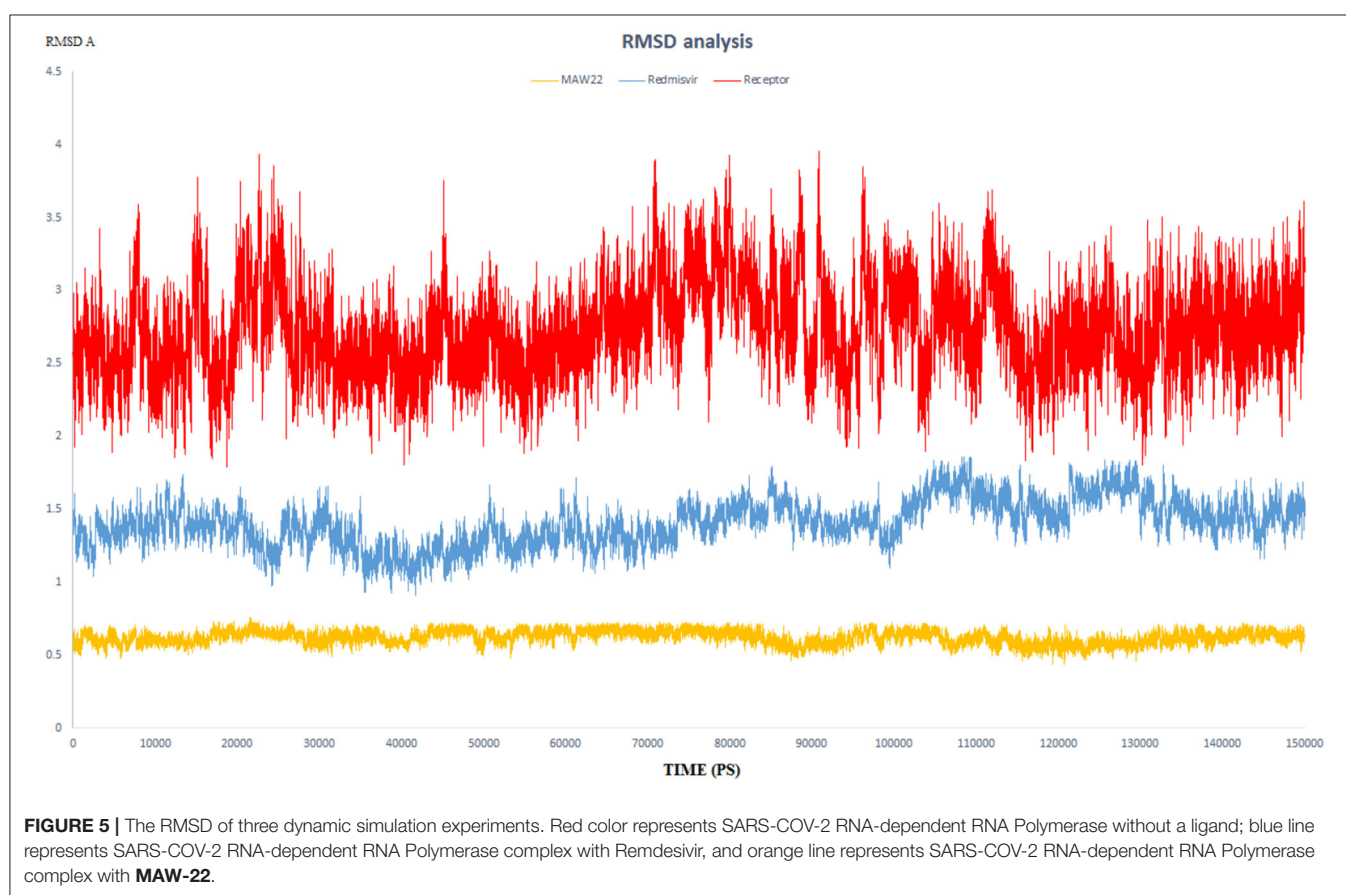
Molecular dynamic simulations, such as identification of potential inhibitors for promising targets, studying the nature of macromolecules, or interpretations of drug resistances, have been implemented in many drug discovery applications (El-Hasab et al., 2018; El-Hassab et al., 2019; Alamri et al., 2020; Nagarajan et al., 2020). Despite the good outcomes from the docking study that supported our rationale of design, molecular dynamic simulation experiments were conducted for extra confirmation and validation for the entire work. Also aiming to identify and study the nature of the SARS-COV-2 RNA-dependent RNA Polymerase to give insights for future lead optimization, we performed three dynamic simulations, one for the free SARS-COV-2 RNA-dependent RNA Polymerase and the two others for the enzyme with Remdesivir and compound **MAW-22**.

### **RMSD and RMSF Analysis and Hydrogen Bond Monitoring**

The endeavor of any viral polymerase enzyme is to replicate the virus genome or polyproteins, a function that needs a wide range of flexibility at least in the active site to accommodate both the template and the replicate (Kennedy et al., 2007). All the reported polymerases are relatively dynamic and have a wide active site (Bose-Basu et al., 2004). So, the conducted simulation experiment aimed to determine the extent of dynamicity of the

**TABLE 1** | Types of interactions of **MAW-22** within COVID-19 polymerase active site, and distance (Å°).

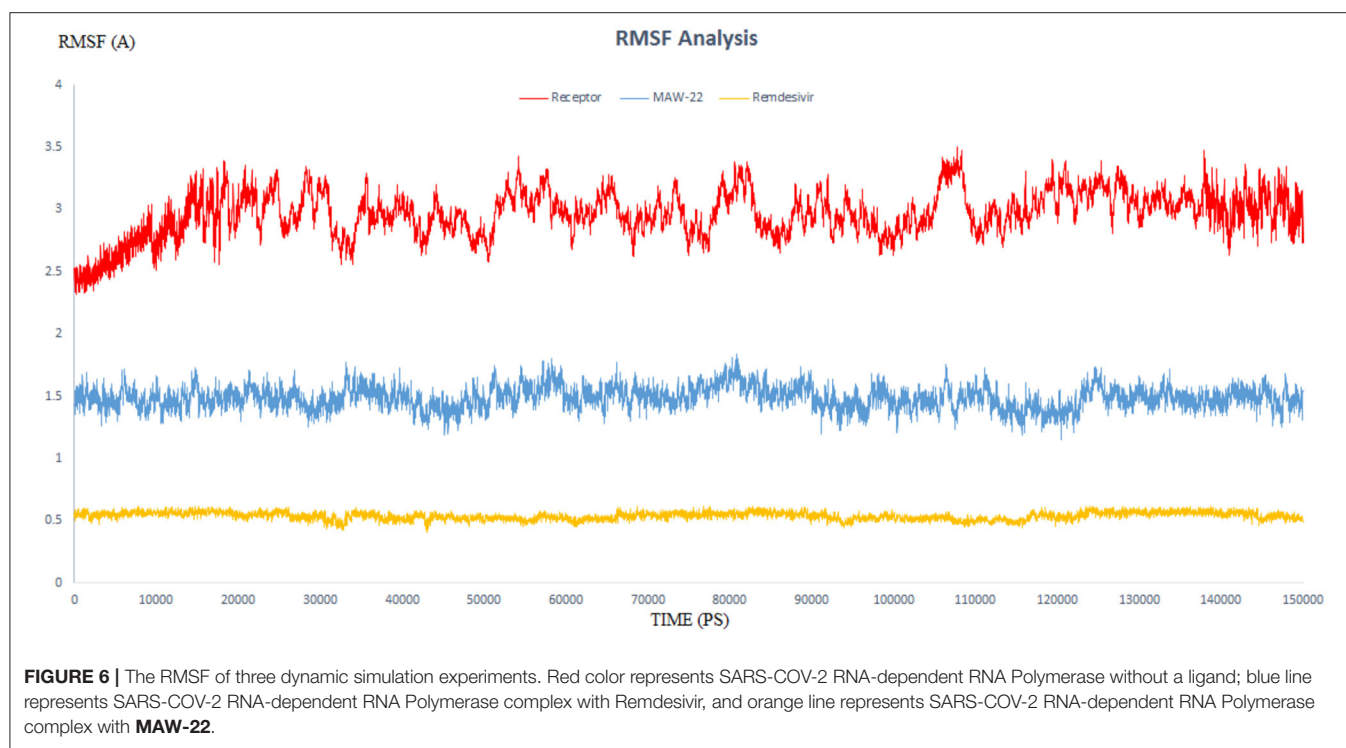
Bond type	Distance Å°	Bond type	Distance Å°
Hydrogen bond with Alanine 558	2.13	Ionic interaction with Aspartic acid 760	3.20
Hydrogen bond with Serine 682	2.04	Ionic interaction with Aspartic acid 623	1.78
Hydrogen bond with Serine 682	1.94	Ionic interaction with Aspartic acid 623	3.90
Hydrogen bond with Threonine 556	2.27	Ionic interaction with Aspartic acid 623	4.08
Hydrogen bond with Threonine 556	2.96	Ionic interaction with Aspartic acid 623	4.36
Hydrogen bond with Arginine 553	2.36	Ionic interaction with Arginine 553	2.12
Hydrogen bond with Aspartic acid 760	2.30	Ionic interaction with Arginine 553	4.70
Hydrogen bond with Threonine 680	1.68	Ionic interaction with Arginine 553	2.97
Hydrogen bond with Threonine 680	1.86	Pi-Alkyl interaction with Arginine 555	4.92
Ionic interaction with Aspartic acid 452	4.92	Pi-Alkyl interaction with Cysteine 622	5.04
Ionic interaction with Aspartic acid 760	3.07	Pi-Alkyl interaction with Tyrosine 456	4.57



SARS-COV-2 RNA-dependent RNA Polymerase enzyme as well as to be a reference for comparison with the other two simulation experiments. The calculated RMSD and RMSF of all the residues of the unbound enzyme reached 3.96 and 3.50 Å°, respectively, revealing the high dynamic properties of the SARS-COV-2 RNA-dependent RNA Polymerase (**Figures 5, 6**).

This dynamicity and instability of the enzyme fit perfectly for its intended role to duplicate the entire RNA genome of the virus. This means that the active site of the SARS-COV-2 RNA-dependent RNA Polymerase needs to accommodate not

only for the template RNA but also for the generated duplicate before releasing them and starting another replication process. Also, this dynamicity gives us a reliable parameter to evaluate the efficiency of a proposed inhibitor, as potential potent inhibitors should have the ability to bind strongly to the enzyme and form stable non-dynamic complex. Thus, it was important to monitor the dynamic behavior for both SARS-COV-2 RNA-dependent RNA Polymerase complex with Remdesivir and SARS-COV-2 RNA-dependent RNA Polymerase complex with **MAW-22** through measuring the RMSD and RMSF for both the complexes.



This is far more reliable than the static binding image or the energy score provided from the docking. RMSD and RMSF from MD simulations are more reliable indicators to monitor the stability of the protein-ligand complex and could verify the predicted binding mode.

RMSD values for both **MAW-22** and Remdesivir reached their maximum dynamicity peaks at 0.76 and 1.86 Å°, respectively (Figure 5), whereas the RMSF values were 0.62 Å° for **MAW-22** and 1.84 Å° for Remdesivir at their maximum fluctuation (Figure 6). This indicates that **MAW-22** has a higher ability to inhibit SARS-COV-2 RNA-dependent RNA Polymerase than Remdesivir. The potential inhibitory activity of **MAW-22** may be attributed to its strong binding ability and many involved interactions formed between **MAW-22** and its target COVID-19 polymerase. Thus, it was worthy to monitor the stability of those interactions through an MD experiment. GROMACS has built-in commands that were used to measure the distances of the formed hydrogen bonds between **MAW-22** and COVID-19 polymerase. The distance between the hydrogen bond donor and the hydrogen bond acceptor in a valid hydrogen bond should always be <3.5 Å°. This criterion was fulfilled in all the formed hydrogen bonds between **MAW-22** and COVID-19 polymerase, indicating a stable and valid binding between **MAW-22** and its target (Table 2).

## MM-PBSA Binding Free Energy Calculations

The `g_mmpbsa` package was used to calculate the MM-PBSA binding free energy for the two complexes—**MAW-22** and Remdesivir—bound to SARS-COV-2 RNA-dependent RNA

**TABLE 2 |** The average distances of all the hydrogen bond formed between the **MAW-22** and Covid-19 polymerase through the entire 150 ns MD simulation.

Hydrogen bond name	Average distance (Å°) ± SD
Hydrogen bond with Alanine 558	2.15 ± 0.11
Hydrogen bond with Serine 682	2.0 ± 0.07
Hydrogen bond with Serine 682	2.05 ± 0.15
Hydrogen bond with Threonine 556	2.21 ± 0.1
Hydrogen bond with Threonine 556	3.0 ± 0.09
Hydrogen bond with Arginine 553	2.32 ± 0.12
Hydrogen bond with Aspartic acid 760	2.37 ± 0.1
Hydrogen bond with Threonine 680	1.85 ± 0.2
Hydrogen bond with Threonine 680	1.8 ± 0.13

Polymerase enzyme by the employment of `MmPbSaStat.py` python script. `n` script allows the package to calculate the total free energy for each component of the complex, i.e., the energy of the complex, receptor, and the ligand. Furthermore, the free energy for each component could be calculated using the cumulative sum of its molecular mechanics' potential energy in a vacuum and the free energy of solvation. Molecular mechanics' potential energy includes the energy of both bonded as well as non-bonded interactions (Vanderwal's and electrostatic interaction energies), while the free energy of solvation includes the polar solvation energy (electrostatic) and non-polar solvation energy (non-electrostatic) (Kumari and Kumar, 2014). One of the most widely used non-polar models is the solvent accessible surface area (SASA) (Kumari and Kumar, 2014). All those types

**TABLE 3 |** MM-PBSA calculations of the binding free energy for the two complexes; **MAW-22** and Remdesivir.

Complex	$\Delta E$ binding (kJ/mol)	$\Delta E$ Bonded interaction (kJ/mol)	$\Delta E$ Electrostatic (kJ/mol)	$\Delta E$ Vander Waal(kJ/mol)	$\Delta E$ polar solvation (kJ/mol)	SASA (kJ/mol)
<b>MAW-22</b>	$-390.24 \pm 20.501$	$1.35 \pm 0.125$	$-134.703 \pm 20.105$	$-343.405 \pm 25.42$	$117.722 \pm 16.104$	$-31.204 \pm 1.092$
Remdesivir	$-328.447 \pm 22.334$	$3.547 \pm 0.232$	$-118.958 \pm 22.612$	$-294.110 \pm 26.713$	$106.352 \pm 18.001$	$-25.278 \pm 2.08$

of energies were calculated by the g\_mmpbsa package, along with the values standard deviation, and then summed together to yield the average total free energy of each component. Finally, the binding free energy could be calculated by subtracting the total free energy of the receptor and the total free energy of the ligand from the total free energy of the complex. As a general fact, the lesser the binding free energy the more stable the complex and the stronger the binding between the ligand and the receptor. **Table 3** summarizes the interaction energies and the binding free energy for both the complexes.

Generally, **MAW-22**–Covid-19 complex was better than Remdesivir complex in all the calculated energy formats, except for polar solvation energy; its average binding free energy reached  $-390.24$  KJ/mol, while Remdesivir's average binding free energy reached  $-328.447$  KJ/mol. The overall results of the three dynamic simulations supported our concept of design and validated the entire virtual screening approach; also, they emphasized and assured the potential inhibitory effect of **MAW-22** on COVID-19 polymerase enzyme.

## FUTURE OUTLOOK

The aim of our study was not only to design a potential inhibitor for the devastating COVID-19 infection but also to establish guidance that could open a new era for the development of an effective treatment for COVID-19. Despite our success in designing a potential inhibitor for SARS-COV-2 RNA-dependent RNA Polymerase, a lot of work is still needed and further optimization for the proposed hit compound should be done before reaching the clinic. Regardless, this study established the key residues and the required type of bonds for inhibiting the SARS-COV-2 RNA-dependent RNA Polymerase enzyme. For instance, compound **MAW-22** has a good combination of basic and acidic groups complementary to the active site of the enzyme that has been found to be rich in basic amino acids like Arginine and Lysine and acidic amino acids like Aspartic acid. Thus, compounds that have negatively and/or positively ionizable groups would have a great chance to bind strongly and inhibit the SARS-COV-2 RNA-dependent RNA Polymerase enzyme. So, drug design techniques that identify the most suitable function groups, such as 3D Pharmacophore, should be applied for further improvement of the generated hit compound. Also, further optimization of the ADMET profile and drug likeness should be considered. Based on the outcomes from this study, our future work will involve the implementation of more structure-based drug design strategies to furnish a lead compound that is worthy of entering clinical trials for the treatment of COVID-19.

## CONCLUSION

In this study we employed a protocol of structure-based drug design with the primary aim of designing a potential specific inhibitor for SARS-COV-2 RNA-dependent RNA Polymerase enzyme. The crystal structure of SARS-COV-2 RNA-dependent RNA Polymerase was retrieved from the protein data bank PDB ID (7bv2) in complex with Remdesivir. Firstly, FBDD strategy was implemented using *de novo* Receptor wizard found in the Discovery Studio 2016. The default fragment database of the software was used to identify potential fragments that could interact strongly with the active site of SARS-COV-2 RNA-dependent RNA Polymerase enzyme. The five best generated fragments were linked together using suitable carbon linkers to yield compound **MAW-22**. Thereafter, the strength of binding between compound **MAW-22** and the SARS-COV-2 RNA-dependent RNA Polymerase enzyme was predicted by Docking strategy using C-Docker software. Compound **MAW-22** achieved a high score of docking, even more so than the score achieved by Remdesivir, which indicates a very strong binding between **MAW-22** and the SARS-COV-2 RNA-dependent RNA Polymerase enzyme. Finally, three molecular dynamic simulation experiments were performed for 150 ns to validate and augment our concept of design. The three experiments revealed that compound **MAW-22** has a great potentiality to inhibit the SARS-COV-2 RNA-dependent RNA Polymerase enzyme, even more so than Remdesivir. The aim of this study was not only to design a potential inhibitor but also to establish guidance for future drug development for the COVID-19 infection.

## DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: Protein Data Bank ID (7bv2).

## AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

## FUNDING

The authors acknowledge financial support from the Researchers Supporting Project number (RSP-2020/103), King Saud University, Riyadh, Saudi Arabia. Also, this research was funded by the Deanship of Scientific Research at Princess Nourah Bint Abdulrahman University through the Fast-track Research Funding Program.

## REFERENCES

- Abraham, M. J., Murtola, T., Schulz, R., Pall, S., Smith, J. C., Hess, B., et al. (2015). GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* 1–2, 19–25. doi: 10.1016/j.softx.2015.06.001
- Alamri, M. A., Tahir ul Qamar, M., Mirza, M. U., Bhadane, R., Alqahtani, S. M., Muneer, I., et al. (2020). Pharmacoinformatics and molecular dynamics simulation studies reveal potential covalent and FDA-approved inhibitors of SARS-CoV-2 main protease 3CLpro. *J. Biomol. Struct. Dyn.* 23, 1–3. doi: 10.1080/07391102.2020.1782768
- Almeida, J. D., Berry, D. M., Cunningham, C. H., Hamre, D., Hofstad, M. S., Mallucci, L., et al. (1968). Virology: coronaviruses. *Nature* 220:650. doi: 10.1038/220650b0
- Armstrong, G. L., MacCannell, D. R., Taylor, J., Carleton, H. A., Neuhaus, E. B., Bradbury, R. S., et al. (2019). Pathogen genomics in public health. *N. Engl. J. Med.* 381, 2569–2580. doi: 10.1056/NEJMSr1813907
- Baell, J. B., and Holloway, G. A. (2010). New substructure filters for removal of pan assay interference compounds (PAINS) from screening libraries and for their exclusion in bioassays. *J. Med. Chem.* 53, 2719–40. doi: 10.1021/jm901137j
- Bhardwaj, V. K., Singh, R., Sharma, J., Rajendran, V., Purohit, R., and Kumar, S. (2020). Identification of bioactive molecules from Tea plant as SARS-CoV-2 main protease inhibitors. *J. Biomol. Struct. Dyn.* 8, p1–13. doi: 10.1080/07391102.2020.1766572
- Blum, L. C., and Reymond, J. L. (2009). 970 million druglike small molecules for virtual screening in the chemical universe database GDB-13. *J. Am. Chem. Soc.* 131, 8732–8733. doi: 10.1021/ja902302h
- Böhm, H. J. (1992). The computer program LUDI: a new method for the de novo design of enzyme inhibitors. *J. Comput. Aided Mol. Des.* 6, 61–78. doi: 10.1007/BF00124387
- Boopathi, S., Poma, A. B., and Kolandaivel, P. (2020). Novel 2019 coronavirus structure, mechanism of action, antiviral drug promises and rule out against its treatment. *J. Biomol. Struct. Dyn.* 29, 1–10. doi: 10.1080/07391102.2020.1758788
- Bose-Basu, B., DeRose, E. F., Kirby, T. W., Mueller, G. A., Beard, W. A., Wilson, S. H., et al. (2004). Dynamic characterization of a DNA repair enzyme: NMR studies of [methyl-13C] methionine-labeled DNA polymerase  $\beta$ . *Biochemistry* 43, 8911–8922. doi: 10.1021/bi049641n
- Cafilisch, A., Miranker, A., and Karplus, M. (1993). Multiple copy simultaneous search and construction of ligands in binding sites: application to inhibitors of HIV-1 aspartic proteinase. *J. Med. Chem.* 36, 2142–2167. doi: 10.1021/jm00067a013
- Carlos, W. G., Dela Cruz, C. S., Cao, B., Pasnick, S., and Jamil, S. (2020). COVID-19 disease due to SARS-CoV-2 (novel coronavirus). *Am. J. Respir. Crit. Care Med.* 201, P7–8. doi: 10.1164/rccm.2014P7
- Cavanagh, D. (2007). Coronavirus avian infectious bronchitis virus. *Vet. Res.* 38, 281–297. doi: 10.1051/vetres:2006055
- Chan, J. F., Yuan, S., Kok, K. H., To, K. K., Chu, H., Yang, J., et al. (2020). A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster. *Lancet* 395, 514–523. doi: 10.1016/S0140-6736(20)30154-9
- Chan-Yeung, M., and Xu, R. H. (2003). SARS: epidemiology. *Respirology* 8, S9–14. doi: 10.1046/j.1440-1843.2003.00518.x
- Chiu, S. W., Pandit, S. A., Scott, H. L., and Jakobsson, E. (2009). An improved united atom force field for simulation of mixed lipid bilayers. *J. Phys. Chem. B* 113, 2748–2763. doi: 10.1021/jp807056c
- Choudhury, C. (2020). Fragment tailoring strategy to design novel chemical entities as potential binders of novel corona virus main protease. *J. Biomol. Struct. Dyn.* 18, 1–5. doi: 10.1080/07391102.2020.1771424
- Coronaviridae Study Group of the International Committee on Taxonomy of Viruses (2020). The species Severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. *Nat. Microbiol.* 5, 536–544. doi: 10.1038/s41564-020-0695-z
- Dassault Systèmes BIOVIA, BIOVIA Workbook Release (2016). *BIOVIA Pipeline Pilot*. San Diego, CA: Dassault Systèmes.
- de Kloe, G. E., Bailey, D., Leurs, R., and de Esch, I. J. (2009). Transforming fragments into candidates: small becomes big in medicinal chemistry. *Drug Discov. Today* 14, 630–646. doi: 10.1016/j.drudis.2009.03.009
- Dong, Y., Mo, X., Hu, Y., Qi, X., Jiang, F., Jiang, Z., et al. (2020). Epidemiological characteristics of 2143 pediatric patients with 2019 coronavirus disease in China. *Pediatrics* 145:e20200702. doi: 10.1542/peds.2020-0702
- Drosten, C., Günther, S., Preiser, W., Van Der Werf, S., Brodt, H. R., Becker, S., et al. (2003). Identification of a novel coronavirus in patients with severe acute respiratory syndrome. *N. Engl. J. Med.* 348, 1967–1976. doi: 10.1056/NEJMoa030747
- Drożdżal, S., Rosik, J., Lechowicz, K., Machaj, F., Kotfis, K., Ghavami, S., et al. (2020). FDA approved drugs with pharmacotherapeutic potential for SARS-CoV-2 (COVID-19) therapy. *Drug Resist. Updat.* 53:100719.
- El-Hasab, M. A., El-Bastawissy, E. E., and El-Moselhy, T. F. (2018). Identification of potential inhibitors for HCV NS3 genotype 4a by combining protein–ligand interaction fingerprint, 3D pharmacophore, docking, and dynamic simulation. *J. Biomol. Struct. Dyn.* 36, 1713–1727. doi: 10.1080/07391102.2017.1332689
- El-Hassab, M. A., El-Bastawissy, E. E., and El-Moselhy, T. F. (2019). Identification of potential inhibitors for HCV NS5b of genotype 4a by combining dynamic simulation, protein–ligand interaction fingerprint, 3D pharmacophore, docking and 3D QSAR. *J. Biomol. Struct. Dyn.* 38, 4521–4535. doi: 10.1080/07391102.2019.1685005
- Evensen, E., Joseph-McCarthy, D., and Karplus, M. (1997). *MCSS Version 2.1*. Cambridge, MA: Harvard University.
- Grein, J., Ohmagari, N., Shin, D., Diaz, G., Asperges, E., Castagna, A., et al. (2020). Compassionate use of remdesivir for patients with severe Covid-19. *N. Engl. J. Med.* 382, 2327–2336.
- Griffiths, S. K., and Campbell, J. P. (2015). Placental structure, function and drug transfer. *Continu. Educ. Anaesth. Crit. Care Pain* 15, 84–89. doi: 10.1093/bjaceaccp/mku013
- Kennedy, W. P., Momand, J. R., and Yin, Y. W. (2007). Mechanism for *de novo* RNA synthesis and initiating nucleotide specificity by  $\tau 7$  RNA polymerase. *J. Mol. Biol.* 370, 256–268. doi: 10.1016/j.jmb.2007.03.041
- Kirsch, P., Hartman, A. M., Hirsch, A. K., and Empting, M. (2019). Concepts and core principles of fragment-based drug design. *Molecules* 24:4309. doi: 10.3390/molecules24234309
- Ksiazek, T. G., Erdman, D., Goldsmith, C. S., Zaki, S. R., Peret, T., Emery, S., et al. (2003). A novel coronavirus associated with severe acute respiratory syndrome. *N. Engl. J. Med.* 348, 1953–1966. doi: 10.1056/NEJMoa030781
- Kumar, A., Voet, A., and Zhang, K. Y. (2012). Fragment based drug design: from experimental to computational approaches. *Curr. Med. Chem.* 19, 5128–5147. doi: 10.2174/092986712803530467
- Kumari, R., and Kumar, R. C. (2014). Open source drug discovery and A. Lynn. *J. Chem. Inf. Model.* 54, 1951–1962. doi: 10.1021/ci500020m
- Lamb, Y. N. (2020). Remdesivir: First Approval. *Drugs* 80, 1355–1363.
- Lu, R., Zhao, X., Li, J., Niu, P., Yang, B., Wu, H., et al. (2020). Genomic characterization and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *Lancet* 395, 565–574. doi: 10.1016/S0140-6736(20)30251-8
- Mehta, N., Mazer-Amirshahi, M., Alkindi, N., and Pourmand, A. (2020). Pharmacotherapy in COVID-19; a narrative review for emergency providers. *Am. J. Emerg. Med.* 38, 1488–1493. doi: 10.1016/j.ajem.2020.04.035
- Mortenson, P. N., and Murray, C. W. (2011). Assessing the lipophilicity of fragments and early hits. *J. Comput. Aided Mol. Des.* 25, 663–667. doi: 10.1007/s10822-011-9435-z
- Nagarajan, H., Narayanaswamy, S., and Vetrivel, U. (2020). Mutational landscape screening of methylene tetrahydrofolate reductase to predict homocystinuria associated variants: an integrative computational approach. *Mutat. Res. Fundam. Mol. Mech. Mutagen.* 819:111687. doi: 10.1016/j.mrfmmm.2020.111687
- Naqvi, A. A., Fatima, K., Mohammad, T., Fatima, U., Singh, I. K., Singh, A., et al. (2020). Insights into SARS-CoV-2 genome, structure, evolution, pathogenesis and therapies: structural genomics approach. *Biochim. Biophys. Acta Mol. Basis Dis.* 13:165878. doi: 10.1016/j.bbdis.2020.165878
- Perlman, S. (2020). Another decade, another coronavirus. *N. Engl. J. Med.* 382, 760–762. doi: 10.1056/NEJMe2001126
- Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., et al. (2005). Scalable molecular dynamics with NAMD. *J. Comput. Chem.* 26, 1781–1802. doi: 10.1002/jcc.20289

- Roughley, S. D., and Hubbard, R. E. (2011). How well can fragments explore accessed chemical space? A case study from heat shock protein 90: miniperspective. *J. Med. Chem.* 54, 3989–4005. doi: 10.1021/jm200350g
- She, J., Jiang, J., Ye, L., Hu, L., Bai, C., and Song, Y. (2020). 2019 novel coronavirus of pneumonia in Wuhan, China: emerging attack and management strategies. *Clin. Transl. Med.* 9, 1–7. doi: 10.1186/s40169-020-00271-z
- Su, S., Wong, G., Shi, W., Liu, J., Lai, A. C., Zhou, J., et al. (2016). Epidemiology, genetic recombination, and pathogenesis of coronaviruses. *Trends Microbiol.* 24, 490–502. doi: 10.1016/j.tim.2016.03.003
- Subissi, L., Posthuma, C. C., Collet, A., Zevenhoven-Dobbe, J. C., Gorbalenya, A. E., Decroly, E., et al. (2014). One severe acute respiratory syndrome coronavirus protein complex integrates processive RNA polymerase and exonuclease activities. *Proc. Natl. Acad. Sci. U. S. A.* 111, E3900–E3909. doi: 10.1073/pnas.1323705111
- Tyrrell, D. A., and Fielder, M. (2002). *Cold Wars: The Fight against the Common Cold*. New York, NY: Oxford University Press.
- ul Qamar, M. T., Alqahtani, S. M., Alamri, M. A., and Chen, L. L. (2020). *Structural basis of SARS-CoV-2 3CLpro and anti-COVID-19 drug discovery from medicinal plants*. *J. Pharm. Anal.* 10, 313–319. doi: 10.1016/j.jpha.2020.03.009
- Veber, F., Johnson, S. R., Cheng, H. Y., Smith, B. R., Ward, K. W., and Kopple, K. D. (2002). Molecular properties that influence the oral bioavailability of drug candidate. *J. Med. Chem.* 45:2615–23. doi: 10.1021/jm020017n
- Wang, M., Cao, R., Zhang, L., Yang, X., Liu, J., Xu, M., et al. (2020). Remdesivir and chloroquine effectively inhibit the recently emerged novel coronavirus (2019-nCoV) *in vitro*. *Cell. Res.* 30, 269–271. doi: 10.1038/s41422-020-0282-0
- Wrapp, D., Wang, N., Corbett, K. S., Goldsmith, J. A., Hsieh, C. L., Abiona, O., et al. (2020). Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science* 367, 1260–1263. doi: 10.1126/science.abb2507
- Wu, G., Robertson, D. H., Brooks, C. L. III, and Vieth, M. (2003). Detailed analysis of grid-based molecular docking: a case study of CDOCKER—A CHARMM-based MD docking algorithm. *J. Comput. Chem.* 24, 1549–1562. doi: 10.1002/jcc.10306
- Yin, W., Mao, C., Luan, X., Shen, D. D., Shen, Q., Su, H., et al. (2020). Structural basis for inhibition of the RNA-dependent RNA polymerase from SARS-CoV-2 by remdesivir. *Science* 1:eabc1560. doi: 10.1101/2020.04.08.032763
- Zaki, A. M., Van Boheemen, S., Bestebroer, T. M., Osterhaus, A. D., and Fouchier, R. A. (2012). Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia. *N. Engl. J. Med.* 367, 1814–1820. doi: 10.1056/NEJMoa1211721
- Zhu, N., Zhang, D., Wang, W., Li, X., Yang, B., Song, J., et al. (2020). A novel coronavirus from patients with pneumonia in China, 2019. *N. Engl. J. Med.* 382, 727–733. doi: 10.1056/NEJMoa2001017
- Ziebuhr, J. (2005). “The coronavirus replicase,” in *Coronavirus Replication and Reverse Genetics. Current Topics in Microbiology and Immunology*, ed L. Enjuanes (Berlin; Heidelberg: Springer), 287, 57–94. doi: 10.1007/3-540-26765-4\_3

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 El Hassab, Shoun, Al-Rashood, Al-Warhi and Eldehna. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Supporting SARS-CoV-2 Papain-Like Protease Drug Discovery: *In silico* Methods and Benchmarking

Tamer M. Ibrahim<sup>1\*</sup>, Muhammad I. Ismail<sup>2</sup>, Matthias R. Bauer<sup>3,4</sup>, Adnan A. Bekhit<sup>5,6</sup> and Frank M. Boeckler<sup>4</sup>

<sup>1</sup> Department of Pharmaceutical Chemistry, Faculty of Pharmacy, Kafrelsheikh University, Kafrelsheikh, Egypt, <sup>2</sup> Department of Pharmaceutical Chemistry, Faculty of Pharmacy, The British University in Egypt, Cairo, Egypt, <sup>3</sup> Structure, Biophysics and Fragment-Based Lead Generation, Discovery Sciences, R&D, AstraZeneca, Cambridge, United Kingdom, <sup>4</sup> Department of Pharmacy, Eberhard-Karls University, Tuebingen, Germany, <sup>5</sup> Department of Pharmaceutical Chemistry, Faculty of Pharmacy, Alexandria University, Alexandria, Egypt, <sup>6</sup> Pharmacy Program, Allied Health Department, College of Health and Sport Sciences, University of Bahrain, Zallaq, Bahrain

## OPEN ACCESS

### Edited by:

Domenica Capasso,  
University of Naples Federico II, Italy

### Reviewed by:

Ramar Vanajothi,  
Bharathidasan University, India  
Giuseppe Felice Mangiardi,  
Italian National Research Council, Italy

### \*Correspondence:

Tamer M. Ibrahim  
tamer\_mohamad@pharm.kfs.edu.eg;  
tamer.ibrahim2@gmail.com

### Specialty section:

This article was submitted to  
Medicinal and Pharmaceutical  
Chemistry,  
a section of the journal  
Frontiers in Chemistry

**Received:** 10 August 2020

**Accepted:** 29 September 2020

**Published:** 05 November 2020

### Citation:

Ibrahim TM, Ismail MI, Bauer MR,  
Bekhit AA and Boeckler FM (2020)  
Supporting SARS-CoV-2 Papain-Like  
Protease Drug Discovery: *In silico*  
Methods and Benchmarking.  
Front. Chem. 8:592289.  
doi: 10.3389/fchem.2020.592289

The coronavirus disease 19 (COVID-19) is a rapidly growing pandemic caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). Its papain-like protease (SARS-CoV-2 PLpro) is a crucial target to halt virus replication. SARS-CoV PLpro and SARS-CoV-2 PLpro share an 82.9% sequence identity and a 100% sequence identity for the binding site reported to accommodate small molecules in SARS-CoV. The flexible key binding site residues Tyr269 and Gln270 for small-molecule recognition in SARS-CoV PLpro exist also in SARS-CoV-2 PLpro. This inspired us to use the reported small-molecule binders to SARS-CoV PLpro to generate a high-quality DEKOIS 2.0 benchmark set. Accordingly, we used them in a cross-benchmarking study against SARS-CoV-2 PLpro. As there is no SARS-CoV-2 PLpro structure complexed with a small-molecule ligand publicly available at the time of manuscript submission, we built a homology model based on the ligand-bound SARS-CoV structure for benchmarking and docking purposes. Three publicly available docking tools FRED, AutoDock Vina, and PLANTS were benchmarked. All showed better-than-random performances, with FRED performing best against the built model. Detailed performance analysis via pROC-Chemotype plots showed a strong enrichment of the most potent bioactives in the early docking ranks. Cross-benchmarking against the X-ray structure complexed with a peptide-like inhibitor confirmed that FRED is the best-performing tool. Furthermore, we performed cross-benchmarking against the newly introduced X-ray structure complexed with a small-molecule ligand. Interestingly, its benchmarking profile and chemotype enrichment were comparable to the built model. Accordingly, we used FRED in a prospective virtual screen of the DrugBank<sup>1</sup> database. In conclusion, this study provides an example of how to harness a custom-made DEKOIS 2.0 benchmark set as an approach to enhance the virtual screening success rate against a vital target of the rapidly emerging pandemic.

**Keywords:** COVID-19, docking, VS, benchmarking, DEKOIS 2.0, PLpro

<sup>1</sup> <https://www.drugbank.ca/>

## INTRODUCTION

The latest situation report of the World Health Organization (WHO), of May 6, 2020, reported that COVID-19 is highly spreading worldwide in over 184 countries and responsible so far for >3.6 million cases and >260,000 fatalities. Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is the causative virus for COVID-19 and was recognized in Wuhan, China (Li et al., 2020b; Qian et al., 2020; Rabi et al., 2020; Tilocca et al., 2020). Coronaviruses belong to a large family of enveloped single-stranded RNA genome (ssRNA) that belong to the Coronaviridae family and divided into four genera: alpha, beta, gamma, and delta coronaviruses (Yang and Leibowitz, 2015). Among coronaviruses, some instigated several respiratory diseases, such as SARS-CoV (Drosten et al., 2003), Middle East respiratory syndrome coronavirus (MERS-CoV) (Zaki et al., 2012), and the pandemic COVID-19 (Rabi et al., 2020). SARS-CoV-2 are beta coronaviruses (Li et al., 2020a; Rabi et al., 2020) with symptoms usually resembling other respiratory virus infections like influenza and rhinovirus (Hsieh et al., 2020).

Upon the virion entry to the host cell, translation of 5'-terminal open reading frames (ORF1a and ORF1ab) is initiated to produce two large polyproteins, pp1a and pp1ab, which are then processed by papain-like protease (PLpro) and 3C-like protease (3CLpro), also called main protease (Mpro) (Barretto et al., 2005; Mielech et al., 2015). This processing is crucial for the release of 16 non-structural proteins (nsp1–16). The formation of the replicase complex essential for viral genome replication is dependent on nsps (Fehr and Perlman, 2015). PLpro plays an essential role for the release of nsp1–3 from the viral polyprotein which are indispensable for viral replication. Also, PLpro has been observed to negatively regulate the host innate immune response toward the viral infection by its deubiquitinating and deISGylating effect (Báez-Santos et al., 2015; Clemente et al., 2020). As a result, PLpro has been recognized as an important target for viral replication suppression endeavors in SARS-CoV and SARS-CoV-2 (Báez-Santos et al., 2015; Freitas et al., 2020).

Structure-based virtual screening (SBVS) remained a crucial technique in modern drug discovery (Schapira et al., 2003; Schneider, 2010; Santiago et al., 2012; Scior et al., 2012). Molecular docking is widely employed in SBVS campaigns, which exploits the structural information of the molecular targets binding sites to assess large molecular databases and predict the preferred binding of compounds prior to the biological screening. Nevertheless, the docking tool and the VS workflow selection must be assessed using benchmarking molecular sets. The benchmarking depends on challenging the VS workflow to enrich known bioactives within a set of decoys (Bauer et al., 2013; Ibrahim et al., 2015a).

**Abbreviations:** WHO, World Health Organization; COVID-19, corona virus disease 19; SARS-CoV-2, severe acute respiratory syndrome coronavirus 2; ssRNA, single stranded RNA genome; MERS-CoV, middle east respiratory syndrome coronavirus; PLpro, papain-like protease; 3CLpro, 3C-like protease; Mpro, main protease; DEKOIS, Demanding Evaluation Kits for Objective *In silico* Screening; VS, virtual screening.

In this study, we benchmark three publicly available docking tools, AutoDock Vina, PLANTS, and FRED against SARS-CoV-2 PLpro. One challenge comprises the absence of small molecules known to inhibit SARS-CoV-2 PLpro and consequently to generate a matching decoy set. Another challenge encompasses the absence of structural conformation of the SARS-CoV-2 PLpro binding site when complexed with conventional small molecules. To overcome these challenges, we conducted a cross benchmark by generating a DEKOIS 2.0 benchmark set of known SARS-CoV PLpro bioactives for SARS-CoV-2 PLpro, using the advantage of the high similarity between both enzymes and identical binding site residues. Furthermore, we modeled the conformation of SARS-CoV-2 complexed with small molecule based on its co-crystal structure homolog, SARS-CoV PLpro. Guided by the benchmarking outcome, we performed a VS effort against the DrugBank database and discuss the most promising hits. This study offers an example of how to employ a DEKOIS 2.0 benchmark set to enhance virtual screening success against a vital target of SARS-CoV-2. This procedure may facilitate virtual finding also against other rapidly resolved protein structures of SARS-CoV-2.

## RESULTS AND DISCUSSIONS

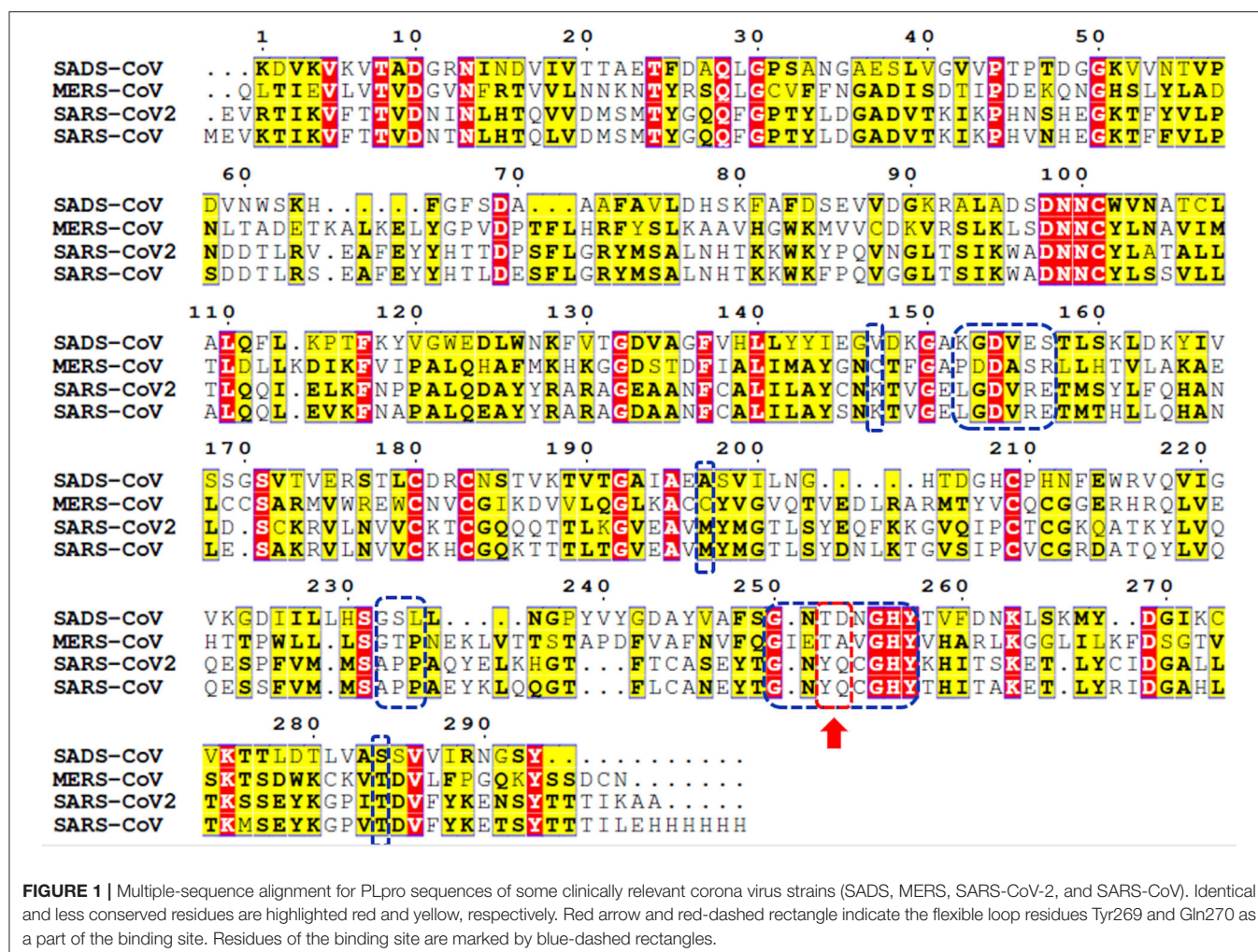
### Multiple-Sequence Alignment and Modeling

Genome sequencing showed an 80% similarity between SARS-CoV and SARS-CoV-2 genetic sequences (Rabaan et al., 2020; Rabi et al., 2020; Zhou et al., 2020). The multiple-sequence alignment (MSA) of PLpro from the most clinically relevant human corona viruses, e.g., SADS, MERS, SARS-CoV-2, and SARS-CoV PLpro, is portrayed in **Figure 1**. Comparing the percentage sequence identity of SARS-CoV-2 PLpro with other human corona viruses reveals that SARS-CoV PLpro is the closest strain to the SARS-CoV-2 PLpro with 82.9% identity. Interestingly, SARS-CoV PLpro and SARS-CoV-2 PLpro share identical binding site residues for small-molecule binding, as marked by the blue-dashed rectangle in **Figure 1**. Residues Tyr269 and Gln270 in SARS-CoV, marked by the red-dashed rectangle, play an important role in small molecule-protein binding event (Ratia et al., 2008; Ghosh et al., 2009). They encompass a flexible loop capable of accommodating different backbone and side chain conformations. Interestingly, it was reported that small-molecule inhibitors of SARS-CoV PLpro were not able to recognize and specifically inhibit MERS-PLpro (Lee et al., 2015). This is attributed to many factors among which is the lack of the key residues Tyr269 and Gln270 of SARS-CoV PLpro in MERS-CoV PLpro (Lee et al., 2015), as shown by the red arrow of **Figure 1**. Interestingly, such key residues are present in SARS-CoV-2 PLpro (Tyr268 and Gln269).

### Structural Aspects

#### SARS-CoV PLpro vs. SARS-CoV-2 PLpro

Binding site residues of both SARS-CoV and SARS-CoV-2 PLpro are 100% identical. The PDB structures of both (SARS-CoV and SARS-CoV-2 PLpro) proteins appear to show a comparable fold and do not deviate substantially in backbone conformations.



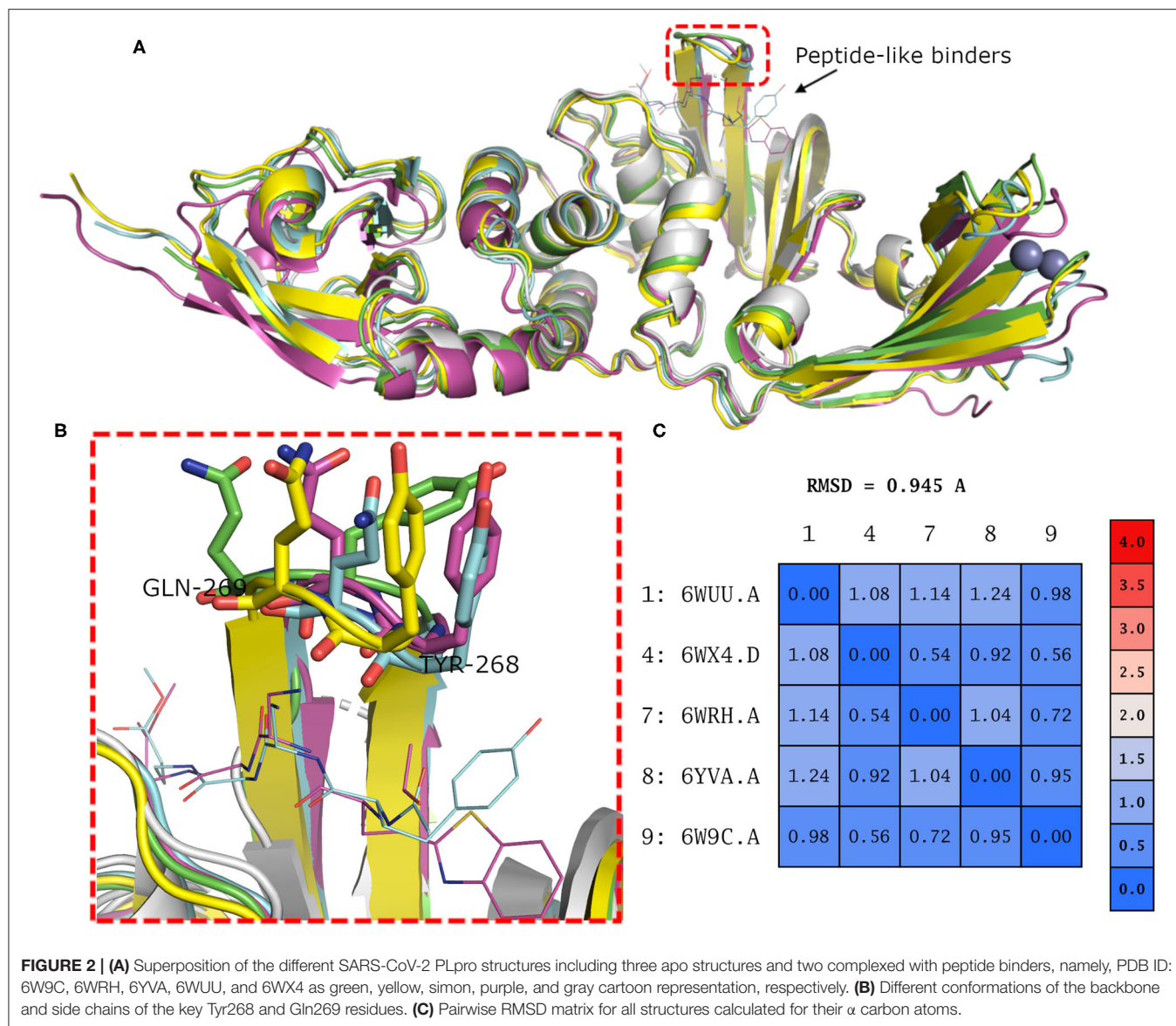
For SARS-CoV-2 PLpro, three crystal structures are available in the apo form, with an average pairwise RMSD matrix for the backbone around 1 Å, as seen in **Figure 2**. The binding site exhibits more diverse conformations for the backbone and side chains of the flexible loop (i.e., Tyr268 and Gln269) where the side chains mostly appear to point outward to the solvent exposed area, as shown in **Figure 2B**. These conformations represent only the unbound state (apo state) for the binding site. It is noteworthy that some SARS-CoV-2 PLpro structures complexed with peptide-like binders were introduced in the PDB, while none complexed with conventional small molecules are available yet at the time of manuscript submission. **Figure 2** shows a depiction of these SARS-CoV-2 PLpro structures.

To have a clue on the possible rearrangement of SARS-CoV-2 PLpro upon conventional small-molecule binding, we investigated its homolog, the SARS-CoV PLpro co-crystal structures complexed with small molecules. For this, we retrieved 11 high-quality crystal structures of SARS-CoV PLpro for both small-molecule complexes and the unbound structures to small molecules (referred to as apo structures in this study). Like SARS-CoV-2, the apo SARS-CoV PLpro (seven crystal structures) displayed a wide range of conformations for the

backbone and side chains of the flexible loop (residues Tyr269 and Gln270 in **Figure 3A**) with an average pairwise RMSD values for the backbone <2 Å (data not shown). On the other hand, the co-crystal structures with small molecules (four crystal structures) showed more ordered rearrangement of Tyr269 for all of them. This is likely to offer a hydrophobic wall for optimum interactions with the aromatic substructure of the bound small molecule (**Figure 3B**; Lee et al., 2015), while Gln270 appeared to adapt more conformations depending on ligand topology and size. Based on the previous, it is likely that SARS-CoV-2 PLpro would behave in a similar fashion to its analog, SARS-CoV PLpro, upon small-molecule binding. Therefore, for docking and benchmarking purposes, and due to the lack of co-crystallized structures of SARS-CoV-2 PLpro with small molecules, we constructed a homology model for SARS-CoV-2 PLpro complexed with small molecule, based on its co-crystallized homolog SARS-CoV PLpro.

### Homology Model

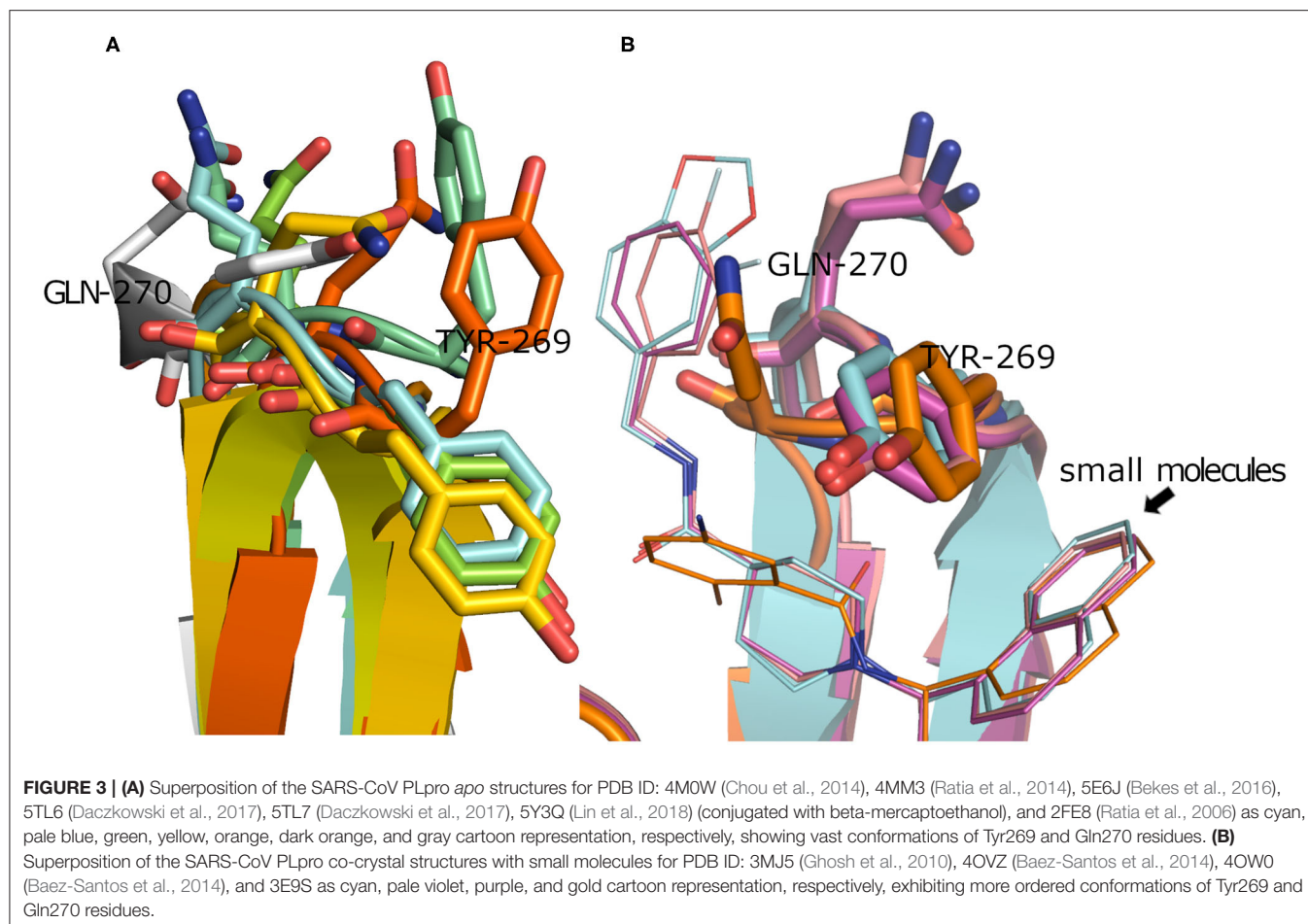
A model of SARS-CoV-2 PLpro (317 residues) complexed with a small-molecule ligand (compound TTT, “5-amino-2-methyl-N-[(1R)-1-naphthalen-1-ylethyl]benzamide”) is built by



the aid of the automated homology modeling, SWISS-MODEL (Waterhouse et al., 2018) web server, using SARS-CoV PLpro (PDB ID: 3E9S, chain A) as a template, as shown in **Figure 4**. The model has a high sequence identity (82.9%) to the template. Quality estimates for the built model indicated high reliability of the model, with a QMEAN (Benkert et al., 2011) value of  $-0.22$  and GMQE (Global Model Quality Estimation) (Waterhouse et al., 2018) value of  $0.95$ . The Ramachandran plot, in **Figure 4B**, shows that 100% of the residues are in the allowed regions. Also, it displayed that 94.9% of the residues, including the binding site residues, are in the most favored region. In addition, the validation web servers (SAVES, 2020) presented that 98.73% of the residues have averaged 3D-1D score  $\geq 0.2$  on the Verify 3D module. The overall quality factor of ERRAT is 92.8 %. Globally, these values indicate a valid and a high-quality homology model.

**Figure 4C** exhibits a noticeable difference in the side chain conformations of the key Tyr298 and Gln269 between the model and the X-ray structure complexed with a peptide-like inhibitor (**Figure 2B**). Unlike the model, both key residues of the latter structure (i.e., PDB ID: 6WUU) appear to point outward to the solvent-exposed area.

While our manuscript was under review, new X-ray co-crystal structures with impact on our study were released. Thus, we closely investigated an example of these structures in comparison to the homology model we built. For instance, we considered the recently introduced X-ray co-crystal structure complexed with compound TTT (PDB ID: 7JRN). This structure is for the wild type and with best resolution available for a SARS-CoV2 PLpro-TTT complex. We did not observe a significant difference between both the model and the X-ray structure (average RMSD for the whole proteins =  $0.98 \text{ \AA}$ ). Interestingly, unlike the



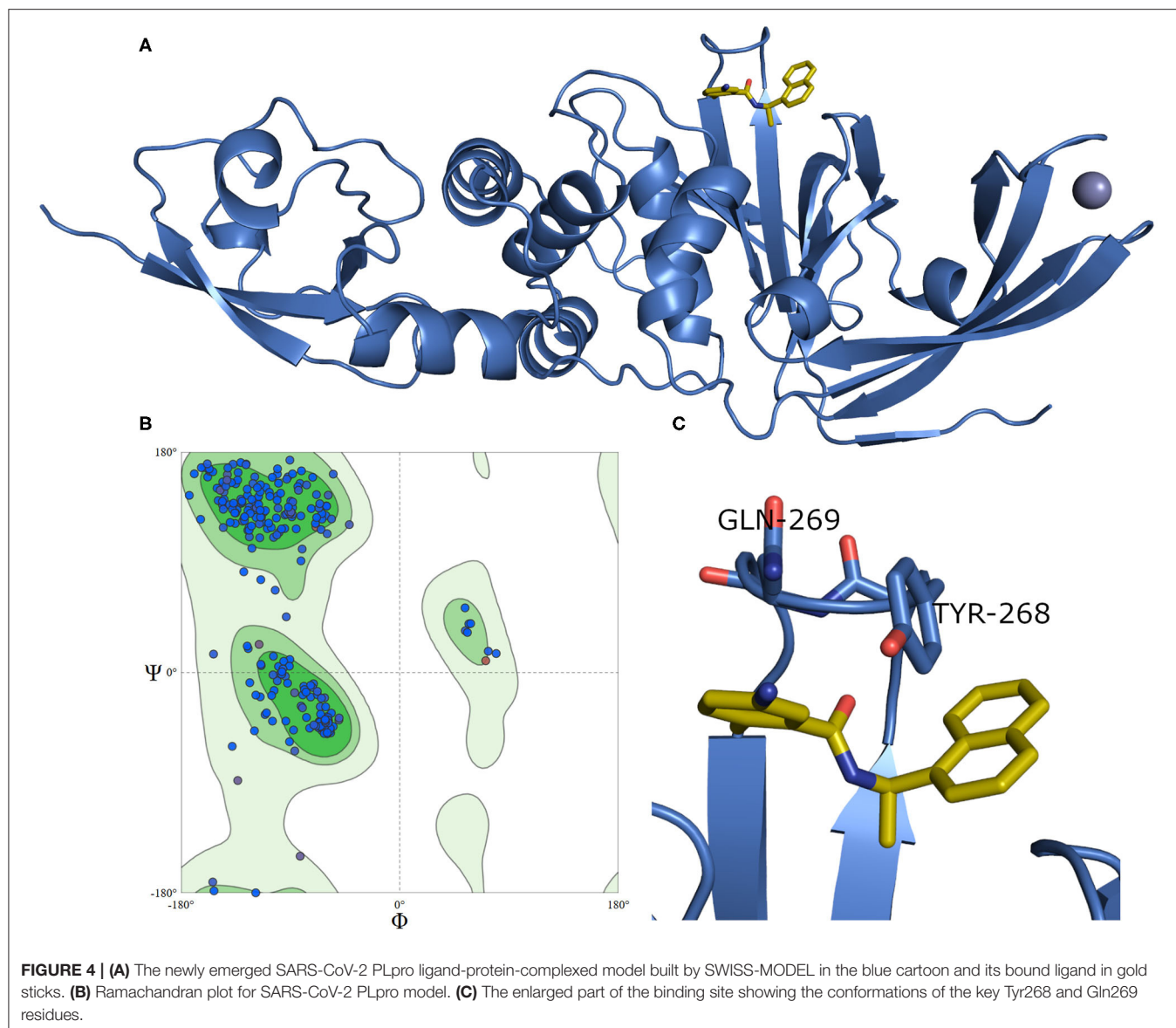
X-ray structure complexed with the peptide-like inhibitor (PDB ID: 6WUU), both the homology model and the SARS-CoV2 PLpro-TTT complex (PDB ID: 7JRN) exhibited almost similar conformations for the key residues Tyr298 and Gln269, as well as for the pose of TTT, as shown in **Figure 5**. This reflects the high reliability and quality of our predicted model.

## Benchmarking

Generally, it was reported that VS performance depends strongly on the respective target properties (Bauer et al., 2013). Accordingly, diverse docking methods and scoring schemes may work better on some targets than others. To avert delays and unnecessary efforts on unproductive VS strategies, it is crucial to evaluate the performance of different VS setups in order to select the most effective workflow (Ibrahim et al., 2015a). Screening performance can be assessed using molecular benchmark sets, such as DEKOIS 2.0 (Vogel et al., 2011; Bauer et al., 2013; Boeckler et al., 2014) and DUD-E (Mysinger et al., 2012). The idea aims at recognizing the suitable docking tool that can efficiently differentiate between the bioactive ligands and the generated challenging decoys. The higher the number of bioactives at the top of the score-ordered list of screened molecules, the better is the respective screening performance.

Due to the lack of small-molecule ligands for SARS-CoV-2 PLpro, and the high similarity of both SARS-CoV and SARS-CoV-2 PLpro enzymes, we performed cross benchmarking of SARS-CoV-2 PLpro based on SARS-CoV PLpro reported small-molecule ligands. We generated a challenging decoy set by our DEKOIS 2.0 (Vogel et al., 2011; Bauer et al., 2013; Boeckler et al., 2014) protocol from the available bioactives of SARS-CoV PLpro (retrieved from BindingDB; Liu et al., 2007). Then, we conducted a benchmarking study using three publicly available docking tools, namely, AutoDock Vina, PLANTS, and FRED. Pleasingly, a recent study by Freitas et al. confirmed our cross benchmarking approach since the naphthalene-based SARS-CoV PLpro inhibitors showed inhibitory activities against SARS-CoV-2 PLpro and stopped the SARS-CoV-2 replication (Freitas et al., 2020).

The SARS-CoV-2 PLpro homology model benchmarking results revealed that FRED screening performance exhibited the best performance with a pROC-AUC value of 2.15, compared to pROC-AUC values of 1.35 and 0.98 for AutoDock Vina and PLANTS, respectively (**Figure 6A**). Interestingly, the screening performance against the co-crystal SARS-CoV PLpro structure (PDB ID: 3E9S) yielded a comparable outcome for FRED, and non-significant differences (i.e.,  $\Delta$ pROC-AUC values  $\leq 0.05$ ; Bauer et al., 2013) for AutoDock Vina and PLANTS docking



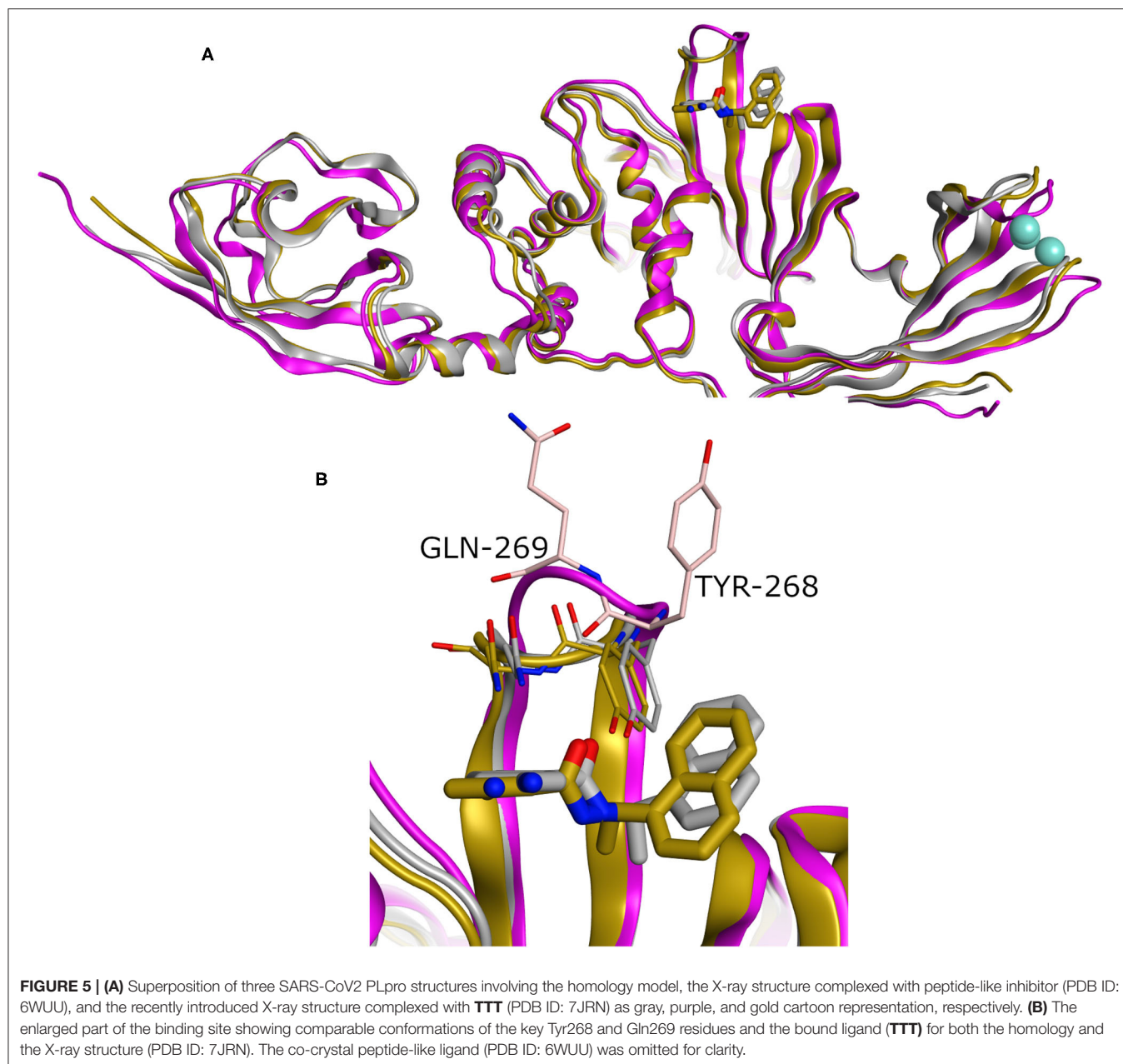
tools (**Figure 6B**). Therefore, these results emphasize also the druggability of the SARS-CoV-2 PLpro homology model by the SARS-CoV PLpro benchmark set. Interestingly, all docking tools exhibited better-than-random performance, i.e., pROC-AUC value >0.43.

In addition, benchmarking results of the co-crystal X-ray structure of SARS-CoV2 with a peptide-like inhibitor (e.g., PDB ID: 6WUU) emphasizes that FRED screening performance appeared to be superior to AutoDock Vina and PLANTS, with pROC-AUC values of 0.95, 0.61, and 0.51, respectively (**Figure 6C**). Nonetheless, in this case, the three docking tools exhibited significant lower performances compared to the homology model. This is likely attributed to the differences in the backbone and side-chain conformations of the key Tyr298 and Gln269 between the model and the X-ray structure complexed with a peptide-like inhibitor (as shown earlier

in **Figure 5**). We also assessed the *in silico* druggability of the unbound conformation (i.e., apo form) of the binding site of SARS-CoV-2 PLpro (PDB ID: 6W9C) and SARS-CoV PLpro (PDB ID: 2FE8) using the generated DEKOIS 2.0 benchmark set. Most of the docking tools showed significant lower performance compared to the bound state as shown in **Figures 6D,E**.

The benchmarking outcome of the newly introduced X-ray complexed with TTT (PDB ID: 7JRN) displayed non-significant differences from the homology model for FRED and PLANTS, and a slightly improved performance for AutoDock Vina, as seen in **Figure 6F**. This is likely attributed to the comparable conformations of the key residues (Tyr298 and Gln269) for both protein structures.

We analyzed the chemotype enrichment with the “pROC-Chemotype” (Ibrahim et al., 2014, 2015b) plot (see **Figure 7**) for

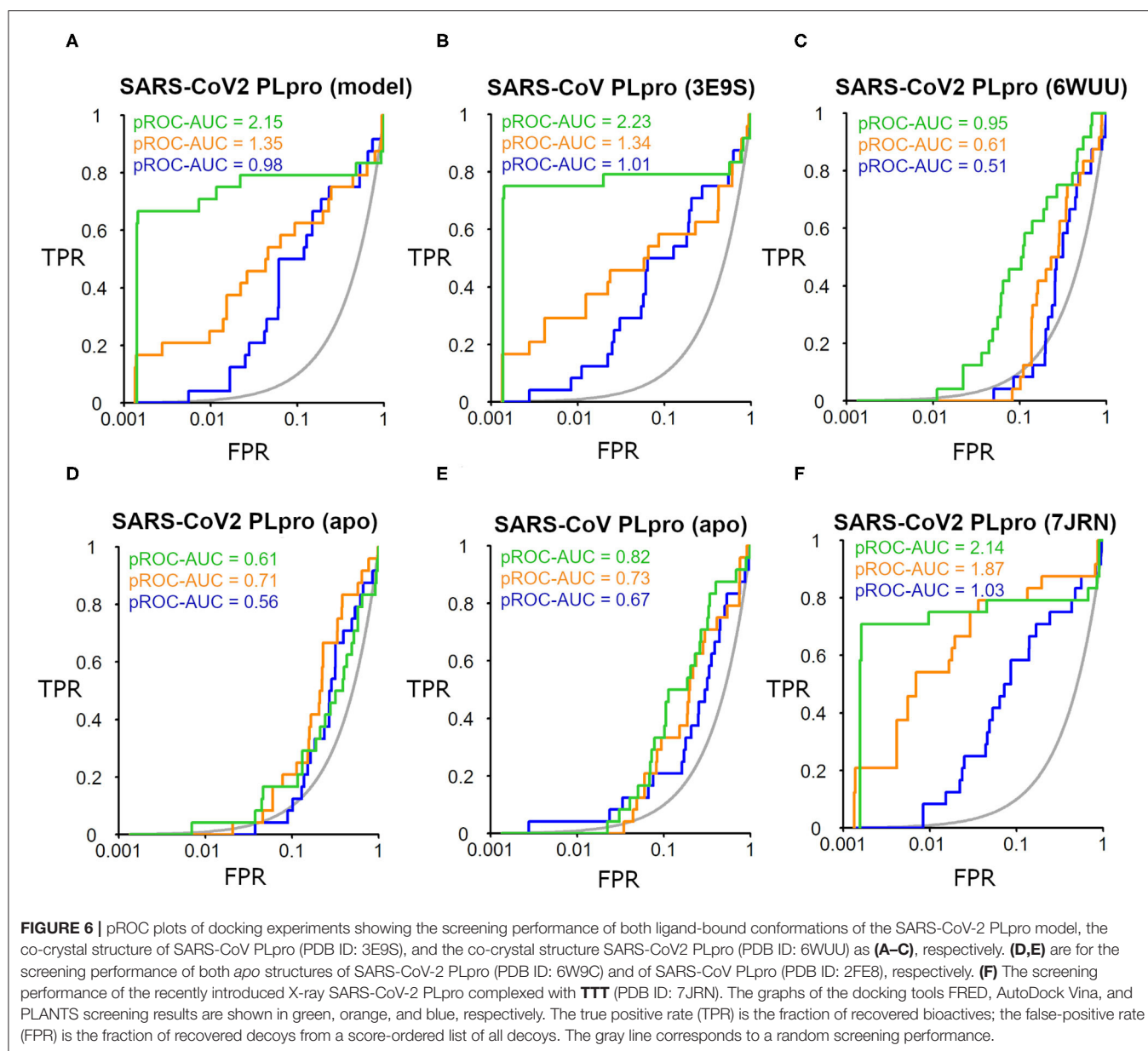


the benchmarking of the SARS-CoV-2 PLpro model using the FRED docking tool. Only 32 small-molecule binders to SARS-CoV PLpro were introduced and collected by the BindingDB repository (Liu et al., 2007) when searching with the keyword “SARS coronavirus papain-like protease.” These molecules were collected mainly from Ghosh et al. (2009) and Lee et al. (2015). This ended up to 24 small-molecule binders after we curated them and removed the duplicates.

Maximum common substructure (MCS) (Ibrahim et al., 2015b) chemotype clustering demonstrates 3 main clusters representing different chemotype classes. Clusters 1 and 2 represent singletons (i.e., a compound per cluster), while cluster 3 (methyl naphthalene substructure) represents the rest of the

bioactive compounds. Therefore, the average *Tanimoto similarity* (*Ts*) was determined by using definition 1 for clusters 1 and 2, while showing *Ts* < 1 for cluster 3, as shown in the relative intercluster (dis)similarity (**Figure 7E**). Generally, such MCS clustering behavior reflects the narrow diversity of the known chemotypes, emphasizing the need of developing more diverse small-molecule inhibitors for SARS-CoV PLpro and eventually for SARS-CoV-2 PLpro. The bioactivity data are represented by *level of activity* (LOA) ranging from  $10^{-4}$  to  $10^{-7}$  M and recorded as  $IC_{50}$ ,  $K_i$ , or  $K_d$  as a *type of data* (TOD), as seen in **Figure 7**.

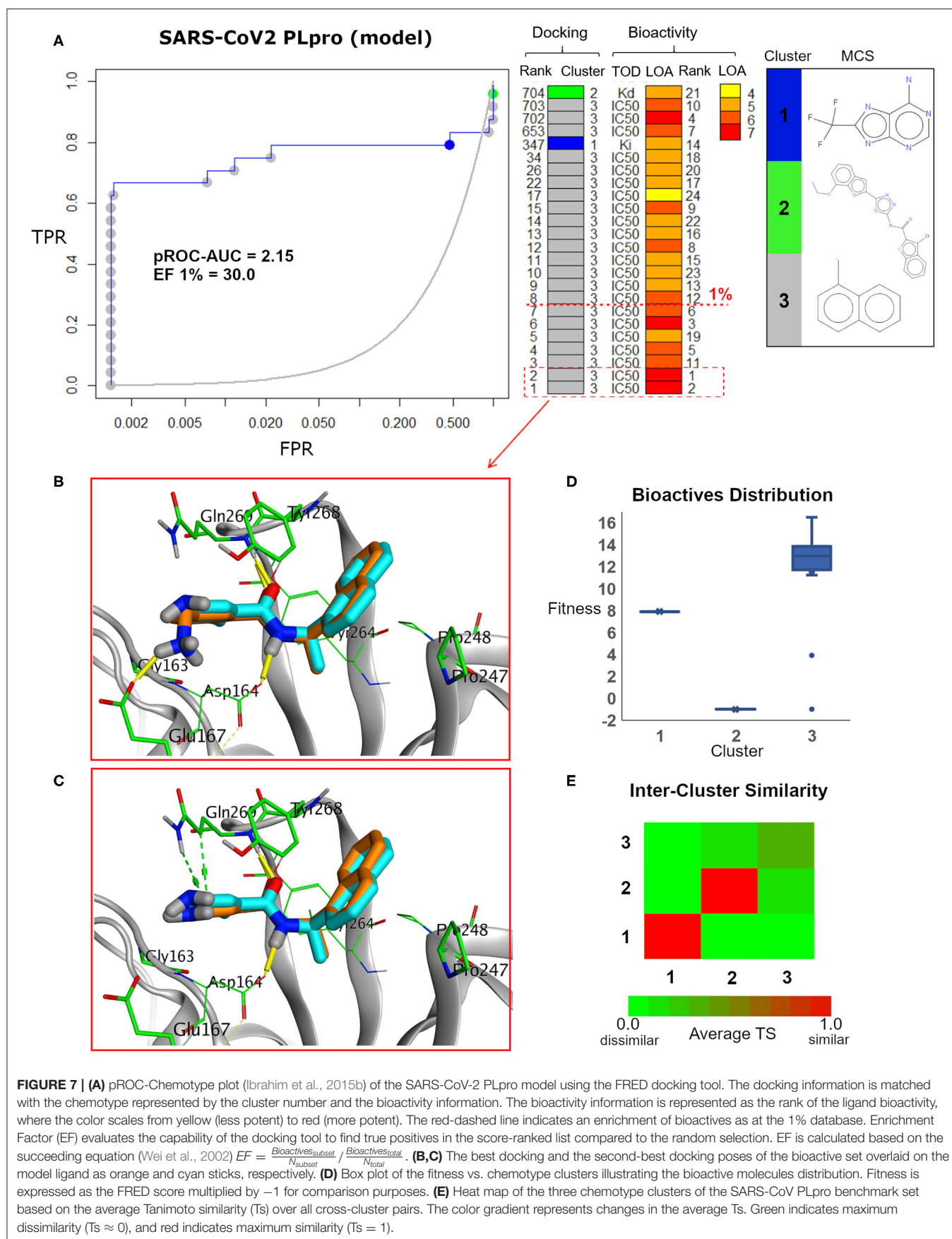
The pROC-Chemotype plot visualized that the applied docking protocol is likely capable of detecting high-affinity binders at early enrichment, as seen in **Figure 7**. For instance,

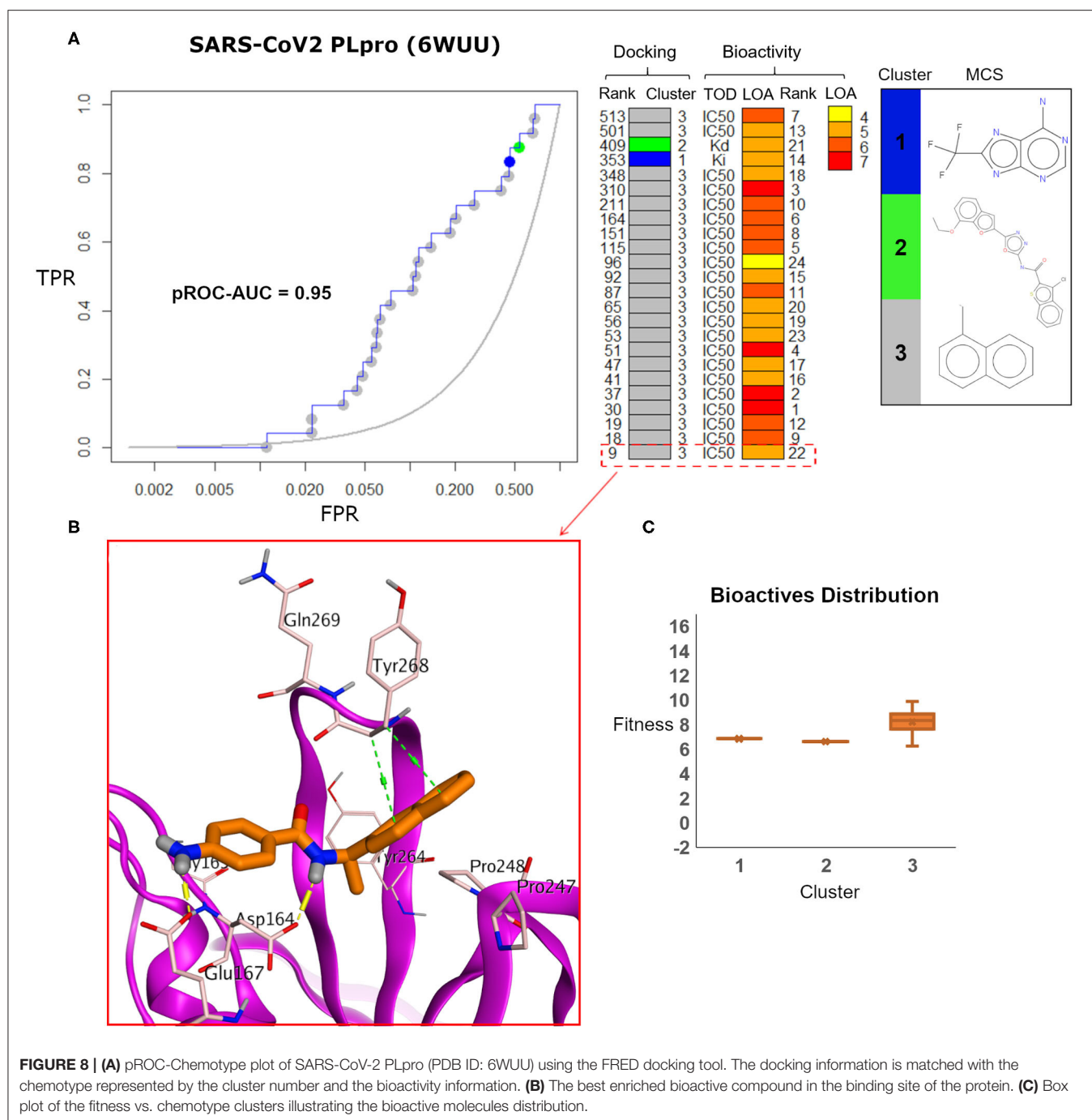


the best two docked active molecules (docking rank 1 and 2) are also the highest in bioactivity (i.e., with bioactivity rank 2 and 1, respectively, **Figure 7A**) with  $IC_{50}$  values of 230 and 460 nM against SARS-CoV PLpro (Ghosh et al., 2009; Lee et al., 2015). Visualizing their docking poses emphasizes that they reproduced the key interactions of the model ligand, as shown in **Figures 7B,C**. It is worth mentioning that such model ligand (TTT) is included in the bioactive set with bioactivity rank 1 and docking rank 2, as shown in **Figure 7C**. Furthermore, at 1% of the score-ordered database, only bioactive molecules were enriched and none of the decoys were recognized, resulting in an Enrichment Factor (EF 1%) of 30.0. This highlights promising enrichment power for the tool under investigation for such a target.

**Figure 7D** shows the docking fitness distribution of the bioactive compounds. The docking score ranges from  $-16.51$  (best score) to  $1.00$  (worse score) and presented as fitness values of  $16.51$  to  $-1.00$  in **Figure 7D**. Also, the majority of cluster 3 compounds lie in the superior region of fitness (i.e., fitness  $> 12$ ). Such superior scores can be attributed mainly to the fact that the naphthyl substructures of their docking poses are involved in hydrophobic interactions and packed between the side chain of the key residue Tyr268 and the side chains of Pro247 and Pro248, as seen in e.g., **Figures 7B,C**.

Visualizing the benchmarking results for the experimental X-ray co-crystal structure (e.g., PDB ID: 6WUU), **Figure 8** displays the pROC-Chemotype plot using FRED docking. Unlike the high value of EF 1% for the SARS-CoV2 model (**Figure 7**), the



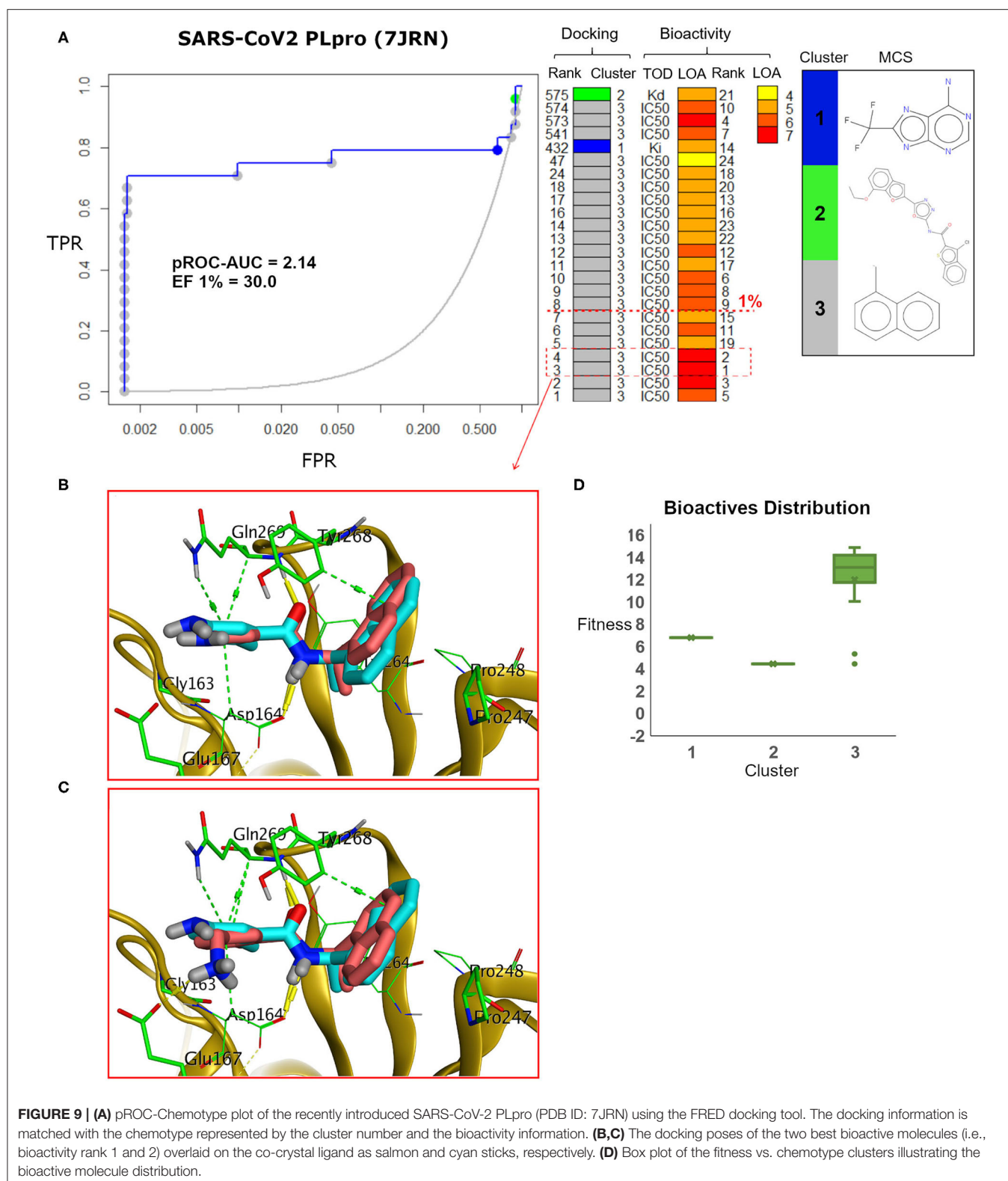


screening performance of the X-ray co-crystal structure did not enrich any bioactive compounds at 1% of the database. Unlike the best enriched bioactive compounds for the SARS-CoV2 model (Figures 7B,C), the best enriched bioactive (Figure 8B) appeared to lose some contacts with the side chains of the key Tyr268 and Gln269 where their side chains appear to be solvent-exposed and directed outward.

Furthermore, the docking fitness distribution of the bioactive compounds in this case is narrower with inferior score range

compared to the model performance. For instance, the docking score ranges from  $-9.89$  (best score) to  $-6.23$  (worst score) and presented as fitness values of  $9.89-6.23$  in Figure 8C. In this case, molecules of cluster 3 did not gain significant advantage since side chains of the key residues Tyr268 and Gln269 are not likely able to optimally interact with their naphthyl substructures.

The pROC-Chemotype plot of the recently introduced X-ray SARS-CoV2 complexed with TTT (PDB ID: 7JRN) for



FRED docking displayed comparable results to the homology model, as seen in **Figure 9**. Both protein structures exhibited similar pROC-AUC and EF 1% values. Additionally, similar

bioactive molecules (6 out of 7 molecules) were enriched at EF 1% for both protein structures. Also, the best two bioactive compounds in the bioactive set (i.e., with bioactivity rank 1

and 2) exhibited similar poses in the binding site (**Figures 9B,C**) compared to their respective poses in the homology model (**Figures 7C,B**). The docking fitness distribution of the bioactive compounds in this case (**Figure 9D**) appeared to be related to their distribution in the case of the homology model (**Figure 7D**). Generally, such behavior is not surprising since both the X-ray SARS-CoV2 PLpro (PDB ID: 7JRN) and the model protein structures exhibit similar conformations for the key residues of the binding site, as discussed earlier (see **Figure 5**).

## Virtual Screening of the DrugBank Database

These promising benchmarking outcomes encouraged us to employ FRED in a virtual screening campaign to screen the FDA-approved drugs from the DrugBank (Wishart et al., 2018) database against SARS-CoV-2 PLpro. We used the homology model, the X-ray co-crystal structure with a peptide inhibitor (PDB ID: 6WUU), as well as the recently introduced X-ray structure complexed with TTT (PDB ID: 7JRN). We utilized these three structures as an approach to target diverse conformations of the ligand-bound state of the binding site and to extract consensus ranking of the screened drugs. The results of the best enriched 1% of the DrugBank database are shown in **Table 1**.

As a consensus, both the model and the X-ray structure complexed with TTT (PDB ID: 7JRN) enriched similar 10 out of 25 drugs at 1% of the DrugBank database, as seen in **Tables 1A,C**. However, only 2 drugs out of 25 drugs were enriched together for the model and the X-ray structure complexed with peptide-like inhibitor (PDB ID: 6WUU), as shown in **Tables 1A,B**. Interestingly, as a consensus for all the three SARS-CoV2 PLpro structures, two drugs appeared to be commonly enriched, namely: Benserazide and Midodrine. However, the latter is in its prodrug form and therefore is not considered in our investigation.

It is worth mentioning that Perphenazine, Benserazide, and Isocarboxazid appeared to be the best-ranked drugs for the three SARS-CoV2 PLpro structures: the model, PDB ID: 6WUU and PDB ID: 7JRN, respectively.

Elucidating the postulated binding interactions of a consensus binder from the DrugBank to the three PLpro protein structures, **Figure 10** shows the binding pose of Benserazide in the binding site of the X-ray SARS-CoV2 PLpro structure (PDB ID: 7JRN). Benserazide is a decarboxylase inhibitor usually combined with levodopa to treat Parkinson's disease. Also, benserazide has been conferred by European Medicines Agency as an orphan designation since 2015 for its potential to be used as a therapy for beta thalassaemia. It was marketed since 1977 by Hoffmann La Roche. Its postulated binding pose in the SARS-CoV-2 PLpro binding site exhibited H-bonding interactions via its hydrazide group with side chains of Asp164 and the key residue Gln296, as seen in **Figure 10**. Also, its trihydroxy phenyl group appeared to be packed in the hydrophobic cleft (green surface in **Figure 10A**) formed by the key residue Tyr268 with residues Pro247 and Pro248. It is worthy to mention that this binding pose of Benserazide is reproduced for the homology model, while

**TABLE 1 |** The best-ranked 1% of the VS efforts for FDA-approved drugs (DrugBank—release March 2020) against the SARS-CoV-2 PLpro homology model, the co-crystal structure (PDB ID: 6WUU), and the recently introduced co-crystal structure (PDB ID: 7JRN) for (A–C), respectively.

Docking rank	Drug <sup>a</sup>	Docking score <sup>b</sup>	Molecular weight	DrugBank ID	Status
<b>(A) SARS-CoV2 PLpro (MODEL)</b>					
1	Perphenazine	−12.94	404.0	DB00850	Approved
2	Zuclopenthixol	−12.57	401.0	DB01624	Approved; investigational
3	Benznidazole	−12.51	260.3	DB11989	Approved; investigational
4	Acetohexamide	−12.08	324.4	DB00414	Approved; investigational; withdrawn
5	Metoclopramide	−11.65	299.8	DB01233	Approved; investigational
6	Tolazamide	−11.54	311.4	DB00839	Approved; investigational
7	Chlorpropamide	−11.53	276.7	DB00672	Approved; investigational
8	Pericizine	−11.52	365.5	DB01608	Approved; investigational
9	Pantothenic acid	−11.25	219.2	DB01783	Approved; nutraceutical; vet_approved
10	Dexpanthenol	−11.19	205.3	DB09357	Approved
11	Agomelatine	−11.11	243.3	DB06594	Approved; investigational
12	Lomustine	−11.11	233.7	DB01206	Approved; investigational
13	Isocarboxazid	−11.04	231.3	DB01247	Approved
14	Practolol	−10.96	266.3	DB01297	Approved
15	Vaborbactam	−10.95	297.1	DB12107	Approved; investigational
16	Salsalate	−10.94	258.2	DB01399	Approved
17	Erdosteine	−10.93	249.3	DB05057	Approved; investigational
18	Sulpiride	−10.79	341.4	DB00391	Approved; investigational
19	Cephalexin	−10.79	347.4	DB00567	Approved; investigational; vet_approved
20	<b>Midodrine<sup>c</sup></b>	−10.79	254.3	DB00211	Approved
21	Nadolol	−10.76	309.4	DB01203	Approved
22	Fluphenazine	−10.74	437.5	DB00623	Approved
23	Acetophenazine	−10.71	411.6	DB01063	Approved
24	Paroxetine	−10.68	329.4	DB00715	Approved; investigational
25	<b>Benserazide</b>	−10.56	257.2	DB12783	Approved; investigational
Average (SD) <sup>d</sup> = −11.26 (±0.64)					
<b>(B) SARS-CoV2 PLpro (PDB ID: 6WUU)</b>					
1	<b>Benserazide</b>	−10.12	257.2	DB12783	Approved; investigational

(Continued)

TABLE 1 | Continued

Docking rank	Drug <sup>a</sup>	Docking score <sup>b</sup>	Molecular weight	DrugBank ID	Status
2	5-O-Phosphono-alpha-D-ribofuranosyl diphosphate	-10.01	390.1	DB01632	Approved; experimental; investigational
3	Omeprazole	-9.77	345.4	DB00338	Approved; investigational; vet_approved
4	N-Acetylglucosamine	-9.73	221.2	DB00141	Approved; investigational; nutraceutical
5	Losartan	-9.57	422.9	DB00678	Approved
6	Melatonin	-9.50	232.3	DB01065	Approved; nutraceutical; vet_approved
7	<b>Midodrine</b>	-9.34	254.3	DB00211	Approved
8	Pyrophosphoric acid	-9.15	178.0	DB04160	Approved; experimental
9	Lactulose	-9.12	342.3	DB00581	Approved
10	Mycophenolic acid	-9.06	320.3	DB01024	Approved
11	Glasdegib	-8.99	374.4	DB11978	Approved; investigational
12	Unoprostone	-8.87	382.5	DB06826	Approved; investigational
13	Calcium glucoheptonate	-8.78	490.4	DB00326	Approved
14	Magnesium gluconate	-8.70	450.6	DB13749	Approved; investigational
15	Calcium gluconate	-8.70	430.4	DB11126	Approved; vet_approved
16	Potassium gluconate	-8.70	234.2	DB13620	Approved
17	Ferrous gluconate	-8.70	446.1	DB14488	Approved
18	Chromium gluconate	-8.70	637.4	DB14528	Approved
19	Copper gluconate	-8.70	453.8	DB11246	Approved; investigational
20	Zinc gluconate	-8.70	455.7	DB11248	Approved; vet_approved
21	Aminohippuric acid	-8.66	194.2	DB00345	Approved; investigational
22	Mannitol busulfan	-8.57	338.3	DB12097	Approved; investigational
23	Tipiracil	-8.53	242.7	DB09343	Approved; investigational
24	Indacaterol	-8.53	392.5	DB05039	Approved
25	Naftazone	-8.52	215.2	DB13680	Approved

Average (SD)<sup>d</sup> = -9.03 (±0.49)**(C) SARS-CoV2 PLpro (PDB ID: 7JRN)**

1	Isocarboxazid	-11.84	231.3	DB01247	Approved
2	Procainamide	-11.68	235.3	DB01035	Approved
3	Metoclopramide	-11.50	299.8	DB01233	Approved; investigational

(Continued)

TABLE 1 | Continued

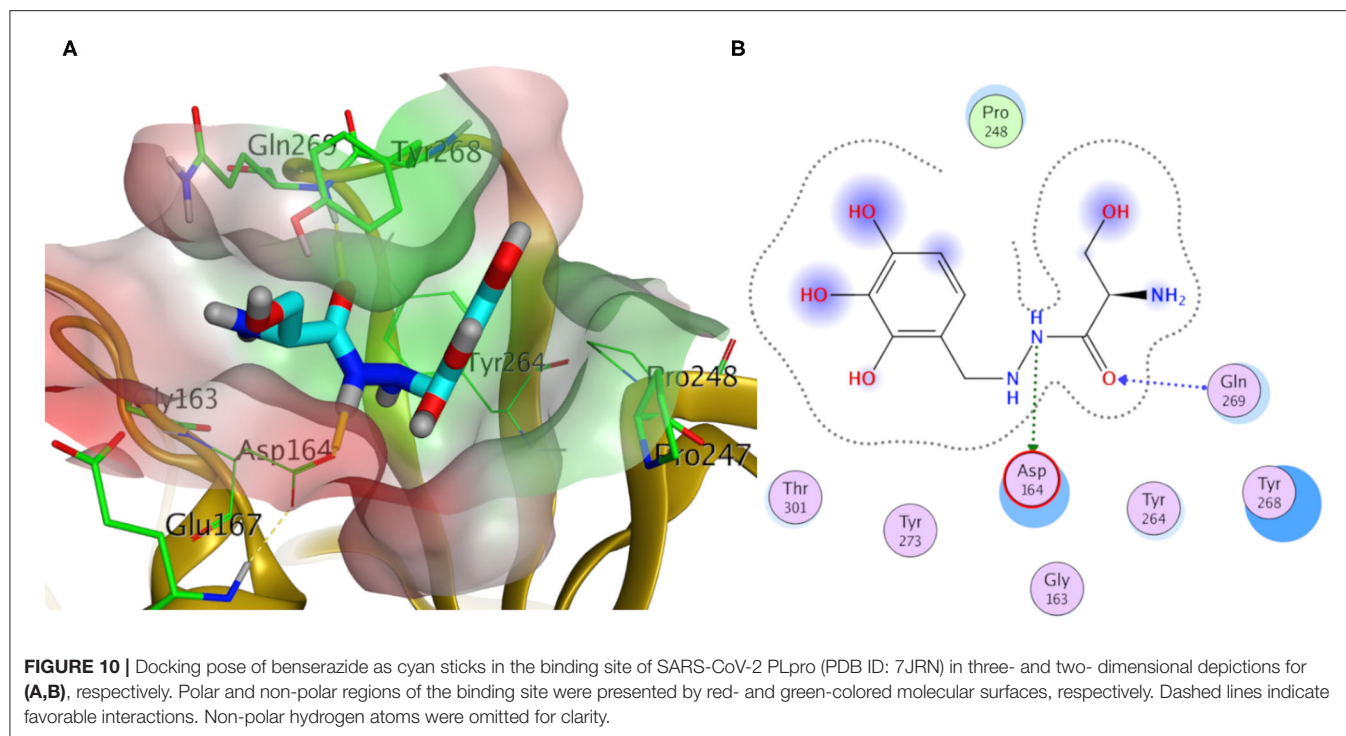
Docking rank	Drug <sup>a</sup>	Docking score <sup>b</sup>	Molecular weight	DrugBank ID	Status
4	Sulpiride	-11.44	341.4	DB00391	Approved; investigational
5	<b>Benserazide</b>	-11.41	257.2	DB12783	Approved; investigational
6	Erdosteine	-11.10	249.3	DB05057	Approved; investigational
7	Pyrophosphoric acid	-11.10	178.0	DB04160	Approved; experimental
8	Remoxipride	-11.09	371.3	DB00409	Approved; withdrawn
9	Dexpanthenol	-10.99	205.3	DB09357	Approved
10	<b>Midodrine</b>	-10.98	254.3	DB00211	Approved
11	Agomelatine	-10.88	243.3	DB06594	Approved; investigational
12	Dobutamine	-10.49	301.4	DB00841	Approved
13	Pantothenic acid	-10.36	219.2	DB01783	Approved; nutraceutical; vet_approved
14	Sulfabenzamide	-10.28	276.3	DB09355	Approved
15	5-O-phosphono-alpha-D-ribofuranosyl diphosphate	-10.22	390.1	DB01632	Approved; experimental; investigational
16	Cefadroxil	-10.21	363.4	DB01140	Approved; vet_approved; withdrawn
17	Nialamide	-10.16	298.3	DB04820	Approved; withdrawn
18	Pergolide	-10.11	314.5	DB01186	Approved; investigational; vet_approved; withdrawn
19	Chlorthalidone	-10.07	338.8	DB00310	Approved
20	Salsalate	-10.03	258.2	DB01399	Approved
21	Pirbuterol	-10.02	240.3	DB01291	Approved
22	Fenoterol	-10.00	303.4	DB01288	Approved; investigational
23	Mefenamic acid	-9.99	241.3	DB00784	Approved
24	Eslicarbazepine acetate	-9.99	296.3	DB09119	Approved
25	Tolnaftate	-9.96	307.4	DB00525	Approved; investigational; vet_approved

Average (SD)<sup>d</sup> = -10.64 (±0.62)<sup>a</sup>Drug: is the generic name of the drug.<sup>b</sup>Docking score is expressed as "FRED Chemgauss4 score".<sup>c</sup>Consensus drugs resulting from both protein structures VS are bold-formatted.<sup>d</sup>Average and standard deviation (SD) are for the docking scores.

differences were observed for 6WUU (data not shown). Again, this is not surprising due to the high similarity of key residues conformations between the model and 7JRN.

## CONCLUSION

MSA and protein structure superposition revealed high sequence identity between SARS-CoV PLpro and SARS-CoV-2 PLpro



with 82.9 and 100% identity for the binding site. The key residues Tyr269 and Gln270 of the binding site of SARS-CoV PLpro for small-molecule recognition are also present in SARS-CoV-2 PLpro. This encouraged us to use the reported small-molecule binders to SARS-CoV PLpro to generate a high-quality DEKOIS 2.0 benchmark set. Accordingly, we performed a cross-benchmarking study using the SARS-CoV PLpro benchmark set against SARS-CoV-2 PLpro. There is no reported co-crystal structure of SARS-CoV-2 PLpro with the conventional small-molecule inhibitor; hence, there is a lack of information for the binding site in a ligand-protein conformation. Thus, we built a homology model for SARS-CoV-2 PLpro complexed with a small-molecule ligand for benchmarking and docking purposes. Three publicly available docking tools were employed in the benchmarking study against the model, FRED, AutoDock Vina, and PLANTS. All showed better-than-random performances with pROC-AUC values of 2.34 for FRED, compared to pROC-AUC values of 1.35 and 0.98 for AutoDock Vina and PLANTS, respectively. Visualizing the FRED performance via the pROC-Chemotype plot emphasizes that this docking tool can enrich the best bioactivity in the early docking rank. Cross-benchmarking against the X-ray co-crystal structure with a peptide-like inhibitor (PDB ID: 6WUU) confirmed that FRED is the best-performing tool. Furthermore, we performed cross-benchmarking against the recently introduced X-ray structure complexed with a small-molecule ligand (PDB ID: 7JRN). Interestingly, its screening performance and chemotype enrichment were comparable to the built model signifying the high quality of the built model. This encourages us to employ FRED in a VS campaign using the FDA-reported drugs (from DrugBank) against SARS-CoV-2 PLpro. In general, this study offers an example of how to employ a DEKOIS 2.0 benchmark set against a vital target of SARS-CoV-2. This

can help improve the success rate for many virtual screening campaigns against the rapidly resolved protein structures of SARS-CoV-2, for fighting the quickly emerging COVID-19.

## METHODS

### Multiple-Sequence Alignment and Homology Modeling

The protein sequences of SARS, MERS, SARS-CoV-2, and SARS-CoV PLpro were retrieved as FASTA format from the Protein Data Bank (PDB) using the PDB IDs: 6L5T, 5W8U, 6W9C, and 2FE8, respectively. The multiple sequence alignment is performed using Clustal Omega (Sievers et al., 2011) and presented by ESript v3.0 (Robert and Gouet, 2014) web server.

SWISS-MODEL (Waterhouse et al., 2018) web server is used to build a homology model for the small-molecule-bound conformation of SARS-CoV-2 PLpro using its automated mode. The template (PDB ID: 3E9S, chain A for SARS-CoV PLpro) was the best recommended for ligand-bound conformation via quality estimate metrics of SWISS-MODEL (Benkert et al., 2011; Waterhouse et al., 2018). The template X-ray crystal structure is with 2.5 Å resolution and R-value free of 0.261. The small-molecule co-crystal ligand is with chemical name “5-amino-2-methyl-N-[(1R)-1-naphthalen-1-ylethyl]benzamide” and involved in the bioactive set for benchmarking with bioactivity rank 1 and IC<sub>50</sub> value 230 nM (Lee et al., 2015). This small molecule is included in the built homology model. The Ramachandran plot of SWISS-MODEL was used to test the validity of the model. Furthermore, the structure analysis and verification server (SAVES, 2020) of the University of California Los Angeles (UCLA) is used to assess the

model, using PROCHECK (Laskowski et al., 1993), Verify 3D (Bowie et al., 1991), PROVE (Pontius et al., 1996), and ERRAT (Colovos and Yeates, 1993).

## Benchmarking and Virtual Screening

### Preparation of Protein Structures

Molecular Operating Environment (MOE) was used to prepare the protein structures for docking experiments, including (i) the homology model complex of SARS-CoV-2 PLpro, (ii) the apo forms of SARS-CoV-2 PLpro (PDB ID: 6W9C) and SARS-CoV PLpro (PDB ID: 2FE8), (iii) the co-crystal structure of SARS-CoV PLpro (PDB ID: 3E9S), (iv) the co-crystal structure SARS-CoV2 PLpro (PDB ID: 6WUU), and (v) the recently introduced co-crystal structure SARS-CoV2 PLpro (PDB ID: 7JRN). Module “Quickprep” of MOE was used at default settings after removing the redundant chains, irrelevant ions, molecules of crystallization, and solvent atoms (if any). Briefly, these settings include using the “Protonate 3D” function to optimize the H-bonding network and allow ASN/GLN/HIS to flip during protonation. Also, these settings involve refining the ligand and binding site atoms via energy minimization to an RMS gradient of 0.1 kcal/mol/Å, while a force constant (strength = 10) was applied for the restraints of receptor atoms. The rest of the receptor atoms outside the binding site were kept fixed. These settings produced a non-significant change of the binding site/ligand coordinates. Also, none of the HIS residues were inspected in the binding site which can be affected by certain protonation/tautomerization state. Conformations of GLN and ASN can be depicted in the respective Figures (in the Results and Discussion section) of the binding site. The prepared structures were saved as mol2 for docking experiments.

### Preparation of the DEKOIS 2.0 Benchmark Set and DrugBank-Approved Drugs

The DEKOIS 2.0 (Bauer et al., 2013) protocol was applied on 24 SARS-CoV PLpro bioactives, which were extracted from BindingDB, to generate 720 challenging decoys (1:30 ratio). Then, all molecules were prepared by MOE with comparable settings to the previous report (Bekhit et al., 2019). Only one conformer was retrieved, and one protonation state was generated at pH 7.0 for each molecule. The specified stereo configuration of all bioactives, decoys, and DrugBank molecules was retained. All prepared molecules were saved as SD files. The SD files were converted and split into PDBQT files by OpenBabel (O’Boyle et al., 2011) for AutoDock Vina docking experiments and into mol2 files for PLANTS docking experiments.

### Docking Experiments

For AutoDock Vina (version 1.1.2) (Trott and Olson, 2010) docking, the protein files were converted to PDBQT files by employing a python script (*prepare\_receptor4.py*) provided by the MGLTools package (version 1.5.4) (Sanner, 1999). The search efficiency of the docking algorithm was kept at default level, while the size of the docking grid was 22.5 Å × 22.5 Å × 22.5 Å, with a grid spacing of 1 Å to make sure to cover all geometries of the docked compounds. For PLANTS (Korb et al., 2009) docking, the scoring function used was “ChemPLP” with the “screen”

mode selected. The binding site was defined within 5 Å of the coordinates of the complexed ligand, and the apo structures were superposed on the complexed ones to extract similar binding site surroundings. For the OEDocking v3.2.0.2 docking (McGann, 2011, 2012), the FRED docking module (McGann, 2011, 2012) was used at default settings. MakeReceptor GUI of OpenEye was used to define the binding site as a search box around the complexed ligand with 19.69 Å × 16 Å × 15.67 Å dimensions.

### pROC Calculations

The docking rank was used in calculating the pROC-AUC employing “R-Snippet” component of KNIME (Berthold et al., 2007) according to the following equation (Clark and Webster-Clark, 2008):

$$pROC\ AUC = \frac{1}{n} \sum_i^n [-\log_{10}(D_i)] = \frac{1}{n} \sum_i^n \log_{10}\left(\frac{1}{D_i}\right)$$

where  $n$  is the number of bioactives and  $D_i$  is the fraction of decoys ranked higher than the  $i$ th bioactive found.

The pROC-Chemotype plots were generated by the “pROC-Chemotype plot” tool which is available in <http://www.dekois.com/> (Ibrahim et al., 2014, 2015b).

Protein structure Figures were rendered using Pymol<sup>2</sup> and MOE.

## DATA AVAILABILITY STATEMENT

The SARS-CoV PLpro active and decoy sets (DEKOIS 2.0 set) can be found in the **Supplementary Material**. The rest of raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

## AUTHOR CONTRIBUTIONS

TI designed the experiments. TI and MI carried out all experiments. All authors have given approval to the final version of the manuscript and agree to be accountable for the content of the work.

## ACKNOWLEDGMENTS

We thank Marina M. Habashy, Noha Galal, Reem Ghazal, and Mohammed A. Elalmawy, undergraduate students in the Faculty of Pharmacy, Kafrelsheikh University, for their literature evaluations and discussions about COVID-19.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fchem.2020.592289/full#supplementary-material>

**Supplementary Table 1** | The active set of SARS-CoV PLpro.

**Supplementary Table 2** | The challenging DEKOIS 2.0 decoy set.

<sup>2</sup>The PyMOL Molecular Graphics System, Schrödinger, LLC.

## REFERENCES

- Baez-Santos, Y. M., Barraza, S. J., Wilson, M. W., Agius, M. P., Mielech, A. M., Davis, N. M., et al. (2014). X-ray structural and biological evaluation of a series of potent and highly selective inhibitors of human coronavirus papain-like proteases. *J. Med. Chem.* 57, 2393–2412. doi: 10.1021/jm401712t
- Báez-Santos, Y. M., John, S. E. S., and Mesecar, A. D. (2015). The SARS-coronavirus papain-like protease: structure, function and inhibition by designed antiviral compounds. *Antiviral Res.* 115, 21–38. doi: 10.1016/j.antiviral.2014.12.015
- Barretto, N., Jukneliene, D., Ratia, K., Chen, Z., Mesecar, A. D., and Baker, S. C. (2005). The papain-like protease of severe acute respiratory syndrome coronavirus has deubiquitinating activity. *J. Virol.* 79, 15189–15198. doi: 10.1128/JVI.79.24.15189-15198.2005
- Bauer, M. R., Ibrahim, T. M., Vogel, S. M., and Boeckler, F. M. (2013). Evaluation and optimization of virtual screening workflows with DEKOIS 2.0—a public library of challenging docking benchmark sets. *J. Chem. Inf. Model.* 53, 1447–1462. doi: 10.1021/ci400115b
- Bekes, M., van der Heden van Noort, G. J., Ekkebus, R., Ovaa, H., Huang, T. T., and Lima, C. D. (2016). Recognition of Lys48-linked Di-ubiquitin and deubiquitinating activities of the SARS coronavirus papain-like protease. *Mol. Cell* 62, 572–585. doi: 10.1016/j.molcel.2016.04.016
- Bekhit, A. A., Saudi, M. N., Hassan, A. M. M., Fahmy, S. M., Ibrahim, T. M., Ghareeb, D., et al. (2019). Synthesis, *in silico* experiments and biological evaluation of 1,3,4-trisubstituted pyrazole derivatives as antimalarial agents. *Eur. J. Med. Chem.* 163, 353–366. doi: 10.1016/j.ejmech.2018.11.067
- Benkert, P., Biasini, M., and Schwede, T. (2011). Toward the estimation of the absolute quality of individual protein structure models. *Bioinformatics* 27, 343–350. doi: 10.1093/bioinformatics/btq662
- Berthold, M. R., Cebon, N., Dill, F., Gabriel, T. R., Kötter, T., Meinel, T., et al. (2007). *KNIME: The Konstanz Information Miner. Studies in Classification, Data Analysis, and Knowledge Organization (GfKL 2007)*. Heidelberg; Berlin: Springer-Verlag. doi: 10.1007/978-3-540-78246-9\_38
- Boeckler, F. M., Bauer, M. R., Ibrahim, T. M., and Vogel, S. M. (2014). Use of DEKOIS 2.0 to gain insights for virtual screening. *J. Cheminform* 6(Suppl.):O24. doi: 10.1186/1758-2946-6-S1-O24
- Bowie, J. U., Luthy, R., and Eisenberg, D. (1991). A method to identify protein sequences that fold into a known three-dimensional structure. *Science* 253, 164–170. doi: 10.1126/science.1853201
- Chou, C. Y., Lai, H. Y., Chen, H. Y., Cheng, S. C., Cheng, K. W., and Chou, Y. W. (2014). Structural basis for catalysis and ubiquitin recognition by the severe acute respiratory syndrome coronavirus papain-like protease. *Acta Crystallogr. D Biol. Crystallogr.* 70(Pt 2), 572–581. doi: 10.1107/S1399004713031040
- Clark, R. D., and Webster-Clark, D. J. (2008). Managing bias in ROC curves. *J. Comput. Aided Mol. Des.* 22, 141–146. doi: 10.1007/s10822-008-9181-z
- Clemente, V., D'Arcy, P., and Bazzaro, M. (2020). Deubiquitinating enzymes in coronaviruses and possible therapeutic opportunities for COVID-19. *Int. J. Mol. Sci.* 21:3492. doi: 10.3390/ijms21103492
- Colovos, C., and Yeates, T. O. (1993). Verification of protein structures: patterns of nonbonded atomic interactions. *Protein Sci.* 2, 1511–1519. doi: 10.1002/pro.5560020916
- Daczkowski, C. M., Dzimiński, J. V., Clasmann, J. R., Goodwin, O., Mesecar, A. D., and Pegan, S. D. (2017). Structural insights into the interaction of coronavirus papain-like proteases and interferon-stimulated gene product 15 from different species. *J. Mol. Biol.* 429, 1661–1683. doi: 10.1016/j.jmb.2017.04.011
- Drosten, C., Günther, S., Preiser, W., van Der Werf, S., Brodt, H.-R., Becker, S., et al. (2003). Identification of a novel coronavirus in patients with severe acute respiratory syndrome. *N Engl. J. Med.* 348, 1967–1976. doi: 10.1056/NEJMoa030747
- Fehr, A. R., and Perlman, S. (2015). “Coronaviruses: an overview of their replication and pathogenesis,” in *Coronaviruses: Methods and Protocols*, eds H. J. Maier, E. Bickerton and P. Britton (New York, NY: Springer New York), 1–23. doi: 10.1007/978-1-4939-2438-7\_1
- Freitas, B. T., Durie, I. A., Murray, J., Longo, J. E., Miller, H. C., Crich, D., et al. (2020). Characterization and noncovalent inhibition of the deubiquitinase and deISGylase activity of SARS-CoV-2 papain-like protease. *ACS Infect. Dis.* 6, 2099–2109. doi: 10.1021/acsinfectdis.0c00168
- Ghosh, A. K., Takayama, J., Aubin, Y., Ratia, K., Chaudhuri, R., Baez, Y., et al. (2009). Structure-based design, synthesis, and biological evaluation of a series of novel and reversible inhibitors for the severe acute respiratory syndrome-coronavirus papain-like protease. *J. Med. Chem.* 52, 5228–5240. doi: 10.1021/jm900611t
- Ghosh, A. K., Takayama, J., Rao, K. V., Ratia, K., Chaudhuri, R., Mulhearn, D. C., et al. (2010). Severe acute respiratory syndrome coronavirus papain-like novel protease inhibitors: design, synthesis, protein-ligand X-ray structure and biological evaluation. *J. Med. Chem.* 53, 4968–4979. doi: 10.1021/jm1004489
- Hsieh, W.-H., Cheng, M.-Y., Ho, M.-W., Chou, C.-H., Lin, P.-C., Chi, C.-Y., et al. (2020). Featuring COVID-19 cases via screening symptomatic patients with epidemiologic link during flu season in a medical center of central Taiwan. *J. Microbiol. Immunol. Infect.* 53, 459–466. doi: 10.1016/j.jmii.2020.03.008
- Ibrahim, T. M., Bauer, M. R., and Boeckler, F. M. (2014). Probing the impact of protein and ligand preparation procedures on chemotype enrichment in structure-based virtual screening using DEKOIS 2.0 benchmark sets. *J. Cheminform.* 6(Suppl. 1):P19. doi: 10.1186/1758-2946-6-S1-P19
- Ibrahim, T. M., Bauer, M. R., and Boeckler, F. M. (2015a). Applying DEKOIS 2.0 in structure-based virtual screening to probe the impact of preparation procedures and score normalization. *J. Cheminform.* 7:21. doi: 10.1186/s13321-015-0074-6
- Ibrahim, T. M., Bauer, M. R., Dorr, A., Veyisoglu, E., and Boeckler, F. M. (2015b). pROC-chemotype plots enhance the interpretability of benchmarking results in structure-based virtual screening. *J. Chem. Inf. Model.* 55, 2297–2307. doi: 10.1021/acs.jcim.5b00475
- Korb, O., Stutzle, T., and Exner, T. E. (2009). Empirical scoring functions for advanced protein-ligand docking with PLANTS. *J. Chem. Inf. Model.* 49, 84–96. doi: 10.1021/ci800298z
- Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* 26, 283–291. doi: 10.1107/S0021889892009944
- Lee, H., Lei, H., Santarsiero, B. D., Gatuz, J. L., Cao, S., Rice, A. J., et al. (2015). Inhibitor recognition specificity of MERS-CoV papain-like protease may differ from that of SARS-CoV. *ACS Chem. Biol.* 10, 1456–1465. doi: 10.1021/cb500917m
- Li, X., Geng, M., Peng, Y., Meng, L., and Lu, S. (2020a). Molecular immune pathogenesis and diagnosis of COVID-19. *J. Pharm. Anal.* 10, 102–108. doi: 10.1016/j.jpha.2020.03.001
- Li, X., Wang, W., Zhao, X., Zai, J., Zhao, Q., Li, Y., et al. (2020b). Transmission dynamics and evolutionary history of 2019-nCoV. *J. Med. Virol.* 92, 501–511. doi: 10.1002/jmv.25701
- Lin, M. H., Moses, D. C., Hsieh, C. H., Cheng, S. C., Chen, Y. H., Sun, C. Y., et al. (2018). Disulfiram can inhibit MERS and SARS coronavirus papain-like proteases via different modes. *Antiviral Res.* 150, 155–163. doi: 10.1016/j.antiviral.2017.12.015
- Liu, T., Lin, Y., Wen, X., Jorissen, R. N., and Gilson, M. K. (2007). BindingDB: a web-accessible database of experimentally determined protein-ligand binding affinities. *Nucleic Acids Res.* 35, D198–D201. doi: 10.1093/nar/gkl999
- McGann, M. (2011). FRED pose prediction and virtual screening accuracy. *J. Chem. Inf. Model.* 51, 578–596. doi: 10.1021/ci100436p
- McGann, M. (2012). FRED and HYBRID docking performance on standardized datasets. *J. Comput. Aided Mol. Des.* 26, 897–906. doi: 10.1007/s10822-012-9584-8
- Mielech, A. M., Deng, X., Chen, Y., Kindler, E., Wheeler, D. L., Mesecar, A. D., et al. (2015). Murine coronavirus ubiquitin-like domain is important for papain-like protease stability and viral pathogenesis. *J. Virol.* 89, 4907–4917. doi: 10.1128/JVI.00338-15
- Molecular Operating Environment (2018). Montreal, QC: Chemical Computing Group Inc. Available online at: <http://www.chemcomp.com> (accessed 2019).
- Mysinger, M. M., Carchia, M., Irwin, J. J., and Shoichet, B. K. (2012). Directory of useful decoys, enhanced (DUD-E): better ligands and decoys for better benchmarking. *J. Med. Chem.* 55, 6582–6594. doi: 10.1021/jm300687e
- O'Boyle, N. M., Banck, M., James, C. A., Morley, C., Vandermeersch, T., and Hutchison, G. R. (2011). Open babel: an open chemical toolbox. *J. Cheminform.* 3:33. doi: 10.1186/1758-2946-3-33
- Pontius, J., Richelle, J., and Wodak, S. J. (1996). Deviations from standard atomic volumes as a quality measure for protein crystal structures. *J. Mol. Biol.* 264, 121–136. doi: 10.1006/jmbi.1996.0628

- Qian, X., Ren, R., Wang, Y., Guo, Y., Fang, J., Wu, Z.-D., et al. (2020). Fighting against the common enemy of COVID-19: a practice of building a community with a shared future for mankind. *Infect. Dis. Poverty* 9:34. doi: 10.1186/s40249-020-00650-1
- Rabaan, A. A., Al-Ahmed, S. H., Haque, S., Sah, R., Tiwari, R., Malik, Y. S., et al. (2020). SARS-CoV-2, SARS-CoV, and MERS-CoV: a comparative overview. *Infez. Med.* 28, 174–184. [Epub ahead of print].
- Rabi, F. A., Al Zoubi, M. S., Kasasbeh, G. A., Salameh, D. M., and Al-Nasser, A. D. (2020). SARS-CoV-2 and coronavirus disease 2019: what we know so far. *Pathogens* 9:231. doi: 10.3390/pathogens9030231
- Ratia, K., Kilianski, A., Baez-Santos, Y. M., Baker, S. C., and Mesecar, A. (2014). Structural basis for the ubiquitin-linkage specificity and deISGylating activity of SARS-CoV papain-like protease. *PLoS Pathog.* 10:e1004113. doi: 10.1371/journal.ppat.1004113
- Ratia, K., Pegan, S., Takayama, J., Sleeman, K., Coughlin, M., Baliji, S., et al. (2008). A noncovalent class of papain-like protease/deubiquitinase inhibitors blocks SARS virus replication. *Proc. Natl. Acad. Sci. U.S.A.* 105, 16119–16124. doi: 10.1073/pnas.0805240105
- Ratia, K., Saikatendu, K. S., Santarsiero, B. D., Barretto, N., Baker, S. C., Stevens, R. C., et al. (2006). Severe acute respiratory syndrome coronavirus papain-like protease: structure of a viral deubiquitinating enzyme. *Proc. Natl. Acad. Sci. U.S.A.* 103, 5717–5722. doi: 10.1073/pnas.0510851103
- Robert, X., and Gouet, P. (2014). Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res.* 42, W320–W324. doi: 10.1093/nar/gku316
- Sanner, M. F. (1999). Python: a programming language for software integration and development. *J. Mol. Graph. Model.* 17, 57–61.
- Santiago, D. N., Pevzner, Y., Durand, A. A., Tran, M., Scheerer, R. R., Daniel, K., et al. (2012). Virtual target screening: validation using kinase inhibitors. *J. Chem. Inf. Model.* 52, 2192–2203. doi: 10.1021/ci300073m
- SAVES (2020). *Structural Analysis and Verification Server Website*. SAVES v5.0. Available online at: <https://servicesn.mbi.ucla.edu/SAVES/>.
- Schapira, M., Abagyan, R., and Totrov, M. (2003). Nuclear hormone receptor targeted virtual screening. *J. Med. Chem.* 46, 3045–3059. doi: 10.1021/jm0300173
- Schneider, G. (2010). Virtual screening: an endless staircase? *Nat. Rev. Drug Discov.* 9, 273–276. doi: 10.1038/nrd3139
- Scior, T., Bender, A., Tresadern, G., Medina-Franco, J. L., Martínez-Mayorga, K., Langer, T., et al. (2012). Recognizing pitfalls in virtual screening: a critical review. *J. Chem. Inf. Model.* 52, 867–881. doi: 10.1021/ci200528d
- Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., et al. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using clustal Omega. *Mol. Syst. Biol.* 7:539. doi: 10.1038/msb.2011.75
- Tilocca, B., Soggiu, A., Sanguinetti, M., Musella, V., Britti, D., Bonizzi, L., et al. (2020). Comparative computational analysis of SARS-CoV-2 nucleocapsid protein epitopes in taxonomically related coronaviruses. *Microbes Infect.* 22, 188–194. doi: 10.1016/j.micinf.2020.04.002
- Trott, O., and Olson, A. J. (2010). AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* 31, 455–461. doi: 10.1002/jcc.21334
- Vogel, S. M., Bauer, M. R., and Boeckler, F. M. (2011). DEKOIS: demanding evaluation kits for objective *in silico* screening—a versatile tool for benchmarking docking programs and scoring functions. *J. Chem. Inf. Model.* 51, 2650–2665. doi: 10.1021/ci2001549
- Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., et al. (2018). SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res.* 46, W296–W303. doi: 10.1093/nar/gky427
- Wei, B. Q., Baase, W. A., Weaver, L. H., Matthews, B. W., and Shoichet, B. K. (2002). A model binding site for testing scoring functions in molecular docking. *J. Mol. Biol.* 322, 339–355. doi: 10.1016/S0022-2836(02)00777-5
- Wishart, D. S., Feunang, Y. D., Guo, A. C., Lo, E. J., Marcu, A., Grant, J. R., et al. (2018). DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Res.* 46, D1074–D1082. doi: 10.1093/nar/gkx1037
- Yang, D., and Leibowitz, J. L. (2015). The structure and functions of coronavirus genomic 3' and 5' ends. *Virus Res.* 206, 120–133. doi: 10.1016/j.virusres.2015.02.025
- Zaki, A. M., Van Boheemen, S., Bestebroer, T. M., Osterhaus, A. D., and Fouchier, R. A. (2012). Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia. *N Engl. J. Med.* 367, 1814–1820. doi: 10.1056/NEJMoa1211721
- Zhou, P., Yang, X.-L., Wang, X.-G., Hu, B., Zhang, L., Zhang, W., et al. (2020). A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* 579, 270–273. doi: 10.1038/s41586-020-2012-7g

**Conflict of Interest:** MB was employed by the company AstraZeneca (Cambridge, UK).

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Ibrahim, Ismail, Bauer, Bekhit and Boeckler. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Repurposing Known Drugs as Covalent and Non-covalent Inhibitors of the SARS-CoV-2 Papain-Like Protease

Pietro Delre<sup>1,2†</sup>, Fabiana Caporuscio<sup>3†</sup>, Michele Saviano<sup>2</sup> and Giuseppe Felice Mangiatordi<sup>2\*</sup>

<sup>1</sup> Department of Chemistry, University of Bari "Aldo Moro", Bari, Italy, <sup>2</sup> National Research Council (CNR) – Institute of Crystallography, Bari, Italy, <sup>3</sup> Department of Life Sciences, University of Modena and Reggio Emilia, Modena, Italy

## OPEN ACCESS

### Edited by:

Chandrabose Selvaraj,  
Alagappa University, India

### Reviewed by:

Umer Rashid,  
COMSATS University  
Islamabad, Pakistan  
Letizia Crocetti,  
University of Florence, Italy  
Naresh Gunaganti,  
Northeastern University, United States

### \*Correspondence:

Giuseppe Felice Mangiatordi  
giuseppe.mangiatordi@ic.cnr.it

<sup>†</sup>These authors have contributed  
equally to this work

### Specialty section:

This article was submitted to  
Medicinal and Pharmaceutical  
Chemistry,  
a section of the journal  
Frontiers in Chemistry

**Received:** 12 August 2020

**Accepted:** 08 October 2020

**Published:** 16 November 2020

### Citation:

Delre P, Caporuscio F, Saviano M and  
Mangiatordi GF (2020) Repurposing  
Known Drugs as Covalent and  
Non-covalent Inhibitors of the  
SARS-CoV-2 Papain-Like Protease.  
Front. Chem. 8:594009.  
doi: 10.3389/fchem.2020.594009

In the absence of an approved vaccine, developing effective severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) antivirals is essential to tackle the current pandemic health crisis due to the coronavirus disease 2019 (COVID-19) spread. As any traditional drug discovery program is a time-consuming and costly process requiring more than one decade to be completed, *in silico* repurposing of existing drugs is the preferred way for rapidly selecting promising clinical candidates. We present a virtual screening campaign to identify covalent and non-covalent inhibitors of the SARS-CoV-2 papain-like protease (PLpro) showing potential multitarget activities (i.e., a desirable polypharmacology profile) for the COVID-19 treatment. A dataset including 688 phase III and 1,702 phase IV clinical trial drugs was downloaded from ChEMBL (version 27.1) and docked to the recently released crystal structure of PLpro in complex with a covalently bound peptide inhibitor. The obtained results were analyzed by combining protein–ligand interaction fingerprint similarities, conventional docking scores, and MM-GBSA–binding free energies and allowed the identification of some interesting candidates for further *in vitro* testing. To the best of our knowledge, this study represents the first attempt to repurpose drugs for a covalent inhibition of PLpro and could pave the way for new therapeutic strategies against COVID-19.

**Keywords:** drug repurposing, SARS-CoV-2, papain-like cysteine protease, molecular docking, molecular interaction fingerprints

## INTRODUCTION

In the last two decades, three zoonotic spillovers of a coronavirus to humans have caused major epidemics, namely, the severe acute respiratory syndrome (SARS) epidemic of 2003 (more than 8,000 human infections and about 800 deaths) (Lu et al., 2020), the Middle East respiratory syndrome (MERS) outbreak of 2012 (about 2,500 confirmed cases and 858 deaths) (Lu et al., 2020), and the current Coronavirus Disease 2019 (COVID-19) pandemic (more than 15,700,000 confirmed cases and 637,810 deaths to date (WHO, COVID-19 daily report of July 26, 2020), the latter being the most devastating one. Coronaviruses are enveloped, single-strand, positive-sense RNA viruses infecting vertebrates and causing respiratory, enteric, and systemic diseases. The causative agent of COVID-19 has been named severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2)

(Gorbalenya et al., 2020) and belongs to the *Sarbecovirus* subgenus of the *Betacoronavirus* genus, which in turn belongs to the Coronaviridae family (Wu et al., 2020). SARS-CoV-2 RNA genome is about 79% identical to that of the highly pathogenic SARS coronavirus (SARS-CoV), which belongs to the *Sarbecovirus* subgenus as well, and 50% identical to that of the more recently emerged MERS-CoV, a member of the *Merbecovirus* subgenus of the *Betacoronavirus* genus (Llanes et al., 2020; Lu et al., 2020).

The most common manifestation of SARS-CoV-2 infection is pneumonia flanked by dry cough, dyspnea, and fever. Other manifestations include, e.g., gastrointestinal symptoms, leukopenia, fatigue, and/or loss of taste and smell. In the most severe cases, respiratory failure may occur and needs to be treated in an intensive care unit through mechanical ventilation. Life-threatening outcomes are frequently associated with elderly patients with concomitant diseases such as hypertension and cardiovascular diseases, chronic obstructive pulmonary disease (COPD), or diabetes. Finally, neurological complications, acute respiratory distress syndrome (ARDS), coagulation dysfunction, septic shock, and multiple organ dysfunction may follow, unfortunately leading to death (Lupia et al., 2020; Prezioso et al., 2020). In particular, ARDS arises as a result of hyperinflammation that is triggered by the viral infection and causes lung tissue damage (Freeman and Swartz, 2020). Hyperinflammation is characterized by the activation of the innate immune response, including the so-called cytokine storm, i.e., an excessive or uncontrolled release of proinflammatory cytokines such as interferons, tumor necrosis factor  $\alpha$ , interleukin 6 (IL-6), and IL-1 $\beta$  (Tisoncik et al., 2012).

SARS-CoV-2 genome contains 14 open reading frames encoding (i) the spike (S), envelope (E), membrane (M), and nucleocapsid (N) structural proteins; (ii) the replicase/transcriptase polyproteins, which self-cleave to form 16 non-structural proteins (NSP1–NSP16); and (iii) accessory proteins. Non-structural proteins assemble into the replicase–transcriptase complex and include the papain-like protease (NSP3, PLpro), the main protease (NSP5, Mpro), the NSP7–NSP8 primase complex, the primary RNA-dependent RNA polymerase (NSP12), the helicase–triphosphatase (NSP13), the exoribonuclease (NSP14), the endonuclease (NSP15), and the N7- and 2' O-methyltransferases (NSP10 and NSP16) (Gordon et al., 2020). As in the case of SARS-CoV, SARS-CoV-2 entry into human cells is driven by the interaction of the viral S glycoprotein with the angiotensin-converting enzyme II (ACE2) receptor, which is highly expressed in alveoli, heart, and brain, whereas MERS-CoV uses dipeptidyl peptidase 4 (DPP4) to enter the host cells (Llanes et al., 2020; Zhou et al., 2020). Moreover, SARS-CoV-2 interacts with many different human proteins expressed in lung tissue, including, e.g., innate immune signaling proteins, histone deacetylase 2, epigenetic readers such as bromodomain proteins, proteins of the translational machinery, etc. (Gordon et al., 2020). Therefore, drugs able to disrupt the SARS-CoV-2 interactome, as well as drugs targeting viral proteins, may represent a feasible strategy to treat COVID-19.

Neither antiviral drugs nor a vaccine has been approved so far for SARS-CoV, MERS-CoV, and SARS-CoV-2. Treatments

for COVID-19 are daily experimented by clinicians, and several clinical trials are ongoing. In the early stages of viral infection, therapies with antivirals designed for other viruses showed some beneficial effects. They include remdesivir, an anti-Ebola virus agent targeting viral RNA transcription; HIV-1 protease inhibitors such as the combination of lopinavir and ritonavir; and ribavirin, a molecule targeting the RNA polymerase and protein synthesis of different RNA viruses. On the contrary, in the advanced stages of COVID-19, antivirals are replaced by immunomodulatory agents targeting the host immune response such as the IL-6 receptor inhibitors tocilizumab, sarilumab, and siltuximab that are able to contain the cytokine storm (Song et al., 2020). Considering that developing an effective vaccine or a specific SARS-CoV-2 antiviral agent starting from scratch may take years, repurposing of approved drugs seems to be the quickest and most straightforward way to limit the burden of COVID-19 (Pinzi et al., 2020; Singh et al., 2020; Yamamoto et al., 2020). In this scenario, *in silico* drug-design tools can aid in the selection of the most suitable candidates. Moreover, at this stage of COVID-19 drug discovery research, structure-based approaches, which do not require a dataset of known active ligands to build a predictive model, are to be preferred. Indeed, since February 2020 the Protein Data Bank has collected up to 282 apo or holo structures of SARS-CoV-2 targets (accessed on July 14) that, together with the structures of human proteins entangled by the virus or responsible for some of its pathogenic effects, can be used to prioritize drugs for COVID-19 therapy. In particular, drugs can be repurposed as inhibitors of SARS-CoV-2 proteins, inhibitors of host proteins such as those involved in the immune response, or disruptors of virus–host interactions.

In this study, we present a structure-based virtual screening (VS) campaign to potentially repurpose 688 phase III and 1,702 phase IV clinical trial drugs from ChEMBL (version 27.1) (Gaulton et al., 2017) as covalent or non-covalent inhibitors of PLpro. To this aim, the recently released crystal structure of PLpro in complex with a covalently bound peptide inhibitor (PDB ID 6WX4) (Rut et al., 2020) was used for the first time. PLpro is a cysteine protease, and its activity consists in (i) the recognition of the LXGG motif and the subsequent hydrolysis of the peptide bond on the carboxyl side of glycine in the P1 position that results in the release of the NSP1, NSP2, and NSP3 proteins; (ii) deubiquitination; and (iii) deISGylation, i.e., the removal of the ubiquitin-like protein interferon-induced gene 15 from host proteins. It is noteworthy that these latter two activities interfere with the innate immune response to viral infection (Rut et al., 2020).

The 6WX4 cocrystallized inhibitor (VIR251; **Figure 1A**) is accommodated in the S4–S1 pockets of the catalytic site and makes a covalent bond with the catalytic C111, as a result of a Michael addition reaction that involves the  $\beta$  carbon of the vinyl group belonging to the VIR251 vinylmethyl ester moiety and the C111 thiol. Moreover, hydrogen bonds are formed with the backbone of G163, Y268, and G271 and with the side chains of W106, D164, and Y264, whereas inhibitor moieties at the P4 position are engaged in hydrophobic interactions (**Figure 1B**).

Therefore, a valuable candidate for drug repurposing should ideally mimic such interactions. Furthermore, other regions flanking the S4 pocket in the proximity of D164, Y273, and T301, not involved in accommodating VIR251, may be explored as well for the design of SARS-CoV-2 PLpro inhibitors (Rut et al., 2020).

## MATERIALS AND METHODS

### Dataset Preparation

Candidate compounds were retrieved from ChEMBL (version 27.1), an open large-scale bioactivity database (Gaulton et al., 2017). In particular, all the compounds that reached phase III or phase IV clinical trials were selected. Subsequently, all duplicates were removed, and only molecules with a molecular weight (MW) in the range of 200 to 700 Da were retained. In this way, 2,390 chemicals were collected, including 688 phase III and 1,702 phase IV clinical trial drugs. All SMILES strings and compound names were extracted and collected in a smi file that was submitted to ligand preparation through the LigPrep tool, available from the Schrödinger Suite 2019-4 (LigPrep | Schrödinger, 2019), to build the 3D structures retaining the correct chirality specified in each SMILE string, desalt and generate all the tautomers and ionization states at a pH value of  $7.0 \pm 2.0$  (LigPrep | Schrödinger, 2019).

### Non-covalent Docking Simulations

Molecular docking simulations were performed on the recently deposited X-ray structure of PLpro in complex with the peptide inhibitor VIR251 (PDB ID 6WX4, resolution: 1.66 Å) (Rut et al., 2020). The structure was preliminarily pretreated by using the Protein Preparation Wizard (PPW) tool (Protein Preparation Wizard | Schrödinger, 2019). More specifically, PPW added missing hydrogen atoms, reconstructed incomplete side chains, assigned the ionization states at physiological pH, set the orientation of any misoriented groups (N, Q, and H residues), removed water molecules farther than 3 Å from any atom of the cognate ligand, optimized the hydrogen bond network, and performed a restrained minimization using PPW default settings. Finally, before docking, all the water molecules were removed from the minimized protein structure (Madhavi Sastry et al., 2013). A cubic grid centered on the centroid of the VIR251 cognate ligand was generated, after breaking the bond between VIR251 and the C111 residue. An inner box of  $10 \text{ Å} \times 10 \text{ Å} \times 10 \text{ Å}$  and an automatic outerbox of  $29 \text{ Å} \times 29 \text{ Å} \times 29 \text{ Å}$  were built. Molecular docking simulations were carried out by using the Extra Precision (XP) protocol (Friesner et al., 2006). All docking simulations were performed using the default force field OPLS\_2005, and during the docking process, the receptor protein was fixed, whereas full conformational flexibility was allowed for the ligands. Importantly, such a protocol was validated by redocking the cognate ligand VIR251 (RMSD = 0.79 Å).

### Covalent Docking Simulations

Covalent docking simulations were performed by using the covalent docking (CovDock, Maestro 12.2.012; Schrödinger LLC) workflow implemented in the Schrödinger suite 2019-4 (Zhu

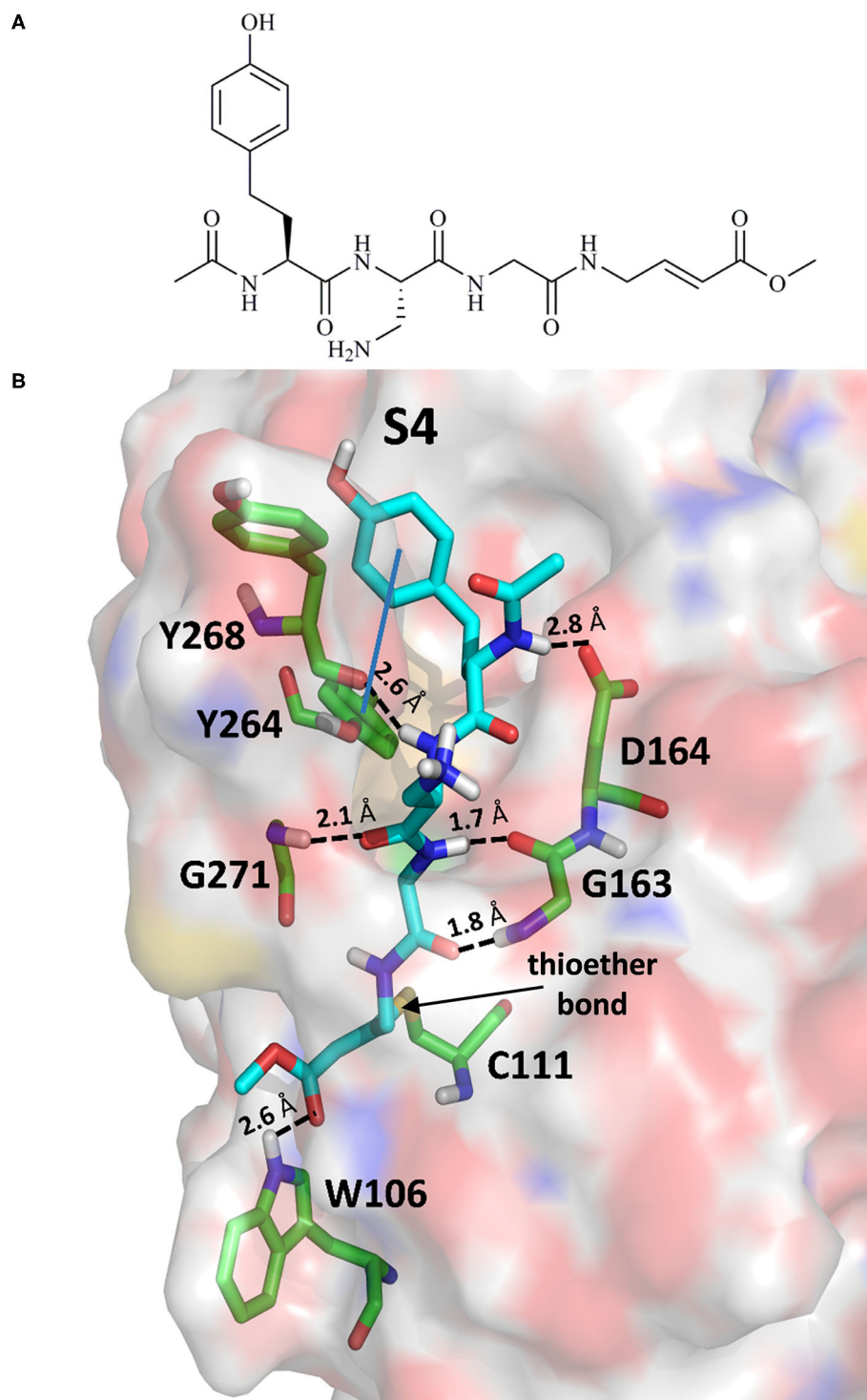
et al., 2014) and the previously pretreated 6WX4 crystal structure. A cubic grid, with an inner box and an automatic outer box having a side equal to 10 and 29 Å, respectively, was generated on the centroid of the VIR251 cognate ligand, C111 was selected as the reactive residue, and the Michael addition reaction was selected as the reaction type. Dataset compounds (total number equal to 2,390) were filtered in order to include only ligands matching the SMART pattern:  $[C,c]=[C,c]-[C,c,S,s]=[O]$ . Only 263 ligands potentially able to be engaged in a Michael addition reaction with C111 were retrieved and subsequently submitted to covalent docking simulations. Covalent docking consisted of five automatic steps (Zhu et al., 2014):

- (i) for each molecule, conformations were generated by the ConfGen utility (Watts et al., 2010), and only the first three with the lowest conformational energies were submitted to a preliminary docking simulation in which the C111 residue was mutated to alanine to avoid steric clashes with the protein;
- (ii) the C111 residue was restored, and docking poses in which the two atoms involved in the formation of the covalent bond are farther than 5 Å were discarded;
- (iii) the covalent bond was then formed, and all changes in bond order, ionization state, or chirality were adjusted;
- (iv) all covalent ligand–protein complexes were refined in order to restore standard bond lengths and avoid steric clashes. The obtained prime energy was used to rank the poses and select the most favorable binding geometry;
- (v) finally, a docking score was assigned to the poses selected in the previous step. This score is defined as the average between the glide/docking score of the binding mode of the pre-reactive ligand and the glide/docking score of the ligand in the final covalent complex (Zhu et al., 2014).

CovDock returned poses for 27 ligands that were ranked by docking score and analyzed by visual inspection. Importantly, such a protocol reproduced the binding mode of the VIR251 ligand (RMSD 1.5 Å).

### Protein–Ligand Interaction Fingerprints Generation

To generate the interaction fingerprints, a common binding site for all compounds was identified. In this regard, the ligand-binding site (BS) was defined using a cutoff radius of 6 Å from all the atoms of VIR251. Subsequently, the SIFt [Interaction Fingerprints (IFPs)] tool of Maestro (version 12.2.012, Schrödinger LLC) was applied to the selected docking poses, as well as to the VIR251 crystallographic coordinates, for computing the IFPs (Deng et al., 2004; Singh et al., 2006). Notably, the selected PLpro BS consists of 34 residues, each of which could potentially establish different chemical interactions with ligands. In particular, the presence of nine possible types of contacts have been verified: (i) any contact, (ii) backbone interactions, (iii) side-chain interactions, (iv) contacts with polar residues, (v) contacts with hydrophobic residues, (vi) formation of hydrogen bonds with H-bond acceptors of the BS, (vii) formation of hydrogen bonds with H-bond donors of the BS, (viii) contacts with aromatic residues, and (ix) contacts with



**FIGURE 1 | (A)** 2D sketch of VIR251; **(B)** X-ray coordinates of VIR251 within the PLpro binding site (PDB ID 6WX4). VIR251 and important residues are rendered as sticks, whereas the protein is represented as a surface. H-bonds are represented by dotted black lines, whereas the pi-stacking interaction between VIR251 and Y264 is itemized by a blue line. For the sake of clarity, only polar hydrogen atoms are shown.

charged residues. Therefore, each residue was represented by a nine-bit-long string for a total of 306 bits per string. A value equal to 1 means that an atom of the ligand is within

the distance required to establish a specific interaction with a specific residue of the BS; on the contrary, a value equal to 0 indicates no contacts. The Tanimoto coefficient (TC) was used

as a quantitative measure of the bit string similarity (Willett et al., 1998). The TC between two strings A and B is defined as follows:

$$Tc = \frac{|A \cap B|}{|A \cup B|}$$

where  $|A \cap B|$  is the number of bits equal to 1 common to both A and B, and  $|A \cup B|$  is the number of bits equal to 1 present in either A or B; the value of TC can range between 0 and 1, with 1 corresponding to two identical fingerprints. By selecting the VIR251 IFP as reference, TC was then calculated for all the docking poses generated in the previous step.

## MM-GBSA Calculations

Docking poses were submitted to a postdocking minimization using the MM-GBSA method (Genheden and Ryde, 2015), by allowing the flexibility of the residues at a maximum distance of 5 Å from the ligand. Default dielectric constants, the OPLS3 force field and the VSGB solvation model were used (Li et al., 2011). Prime MM-GBSA (Prime MM-GBSA | Schrödinger, 2019) outputs were ranked according to the Prime MM-GBSA  $\Delta G$  (Bind) calculated as follows:

$$\text{MM-GBSA } \Delta G_{\text{bind}} = \text{complex} - \text{ligand} - \text{receptor}$$

where *complex* is the energy contribution calculated from the optimized ligand–receptor complex, and *ligand* and *receptor* are the energy contributions calculated from the optimized free ligand and free receptor, respectively. More negative values of  $\Delta G$  (bind) indicate a stronger binding.

## RESULTS AND DISCUSSION

### Candidates for Non-covalent PLpro Inhibition

A database of 1,702 approved drugs (i.e., currently in postmarketing surveillance trial) and 688 compounds that have reached phase III clinical trials was docked to the crystal structure of PLpro [PDB ID 6WX4 (Rut et al., 2020)]. Noteworthy, the high resolution (1.66 Å) and the presence of a cocrystallized peptide inhibitor make this PLpro structure, released on the May 20, 2020, particularly suitable for docking-based VS campaigns that to date have been only performed with PLpro homology models (Amin et al., 2020; Contreras-Puentes and Alvíz-Amador, 2020) or apo structures (Quimque et al., 2020). First, all the compounds were ranked according to their docking scores, and the 500 top-ranking molecules were kept for further evaluation. In particular, in order to overcome possible scoring function deficiencies (Marcou and Rognan, 2007), IFPs were computed for each ligand and compared to those obtained from the crystallographic coordinates of the peptide inhibitor VIR251 by computing a TC (hereinafter referred to as TC-IFP). Noteworthy, it has been shown that accounting for TC-IFPs in a VS campaign yields higher receiver operating characteristic curves and enrichments than ranking compounds based on the docking score only. In particular, compounds returning high TC-IFPs (i.e.,  $\geq 0.6$ ) are more likely to be active

(Marcou and Rognan, 2007) with respect to others with similar docking scores. Furthermore, all the top-500 compounds were submitted to MM-GBSA calculations in order to compute the binding free energies of the relative protein–ligand complexes. In order to properly estimate the energetic contribution of all the protein–ligand interactions, protein flexibility was incorporated during the calculations (see *Materials and Methods* for details). The selection of the most promising candidates was performed by considering the computed docking scores, TC-IFPs, and MM-GBSA–binding free energies. Furthermore, docking poses were carefully visually inspected in order to discard those with solvent-exposed hydrophobic groups or conformational artifacts. Special attention was given to the occupancy of regions in the proximity of the S4 pocket of the enzyme. As recently reported (Rut et al., 2020), regions flanking such a hydrophobic subcavity, although not involved in the VIR251 accommodation, are worth to be explored for designing PLpro inhibitors. Therefore, compounds protruding toward these regions were not discarded, albeit not fully mimicking the VIR251 binding mode reported in **Figure 1B**. Finally, a review of the available literature allowed us to privilege those compounds whose original therapeutic indication may be responsible for a desired polypharmacology to treat COVID-19 patients (Pinzi et al., 2020). Among the selected compounds, some inhibit the same protein (factor Xa) or the same family of proteins such as protein kinases (PKs) or viral/host proteases suggesting some similarity among the BSs. Moreover, other candidates belong to the same pharmacological class (e.g., antidiabetes, antihypertensives). **Table 1** shows the 22 candidates selected for non-covalent PLpro inhibition along with their computed docking scores, TC-IFPs, and MM-GBSA–binding free energies.

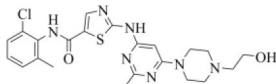
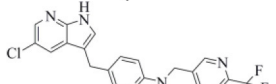
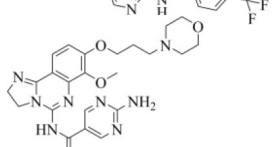
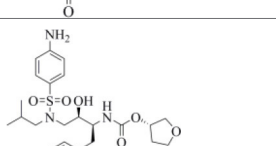
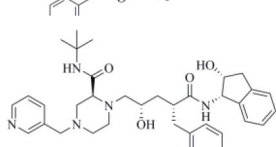
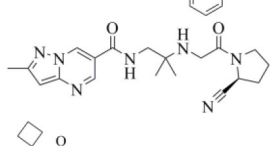
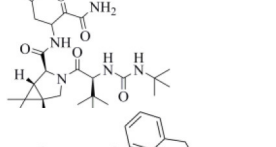
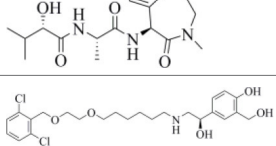
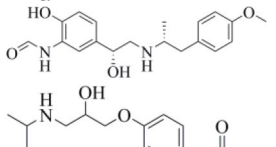
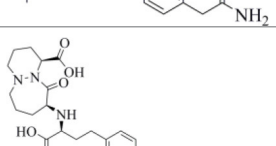
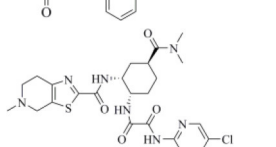
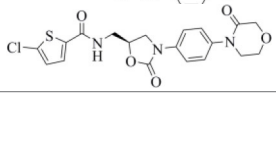

The presence, among the 22 selected compounds, of five inhibitors of other proteases (i.e., amprenavir, indinavir, anagliptin, boceprevir, and semagacestat) supports the reliability of the employed VS protocol. Importantly, amprenavir and indinavir have been already tested as inhibitors of SARS-CoV-2 replication returning EC<sub>50</sub> values in the micromolar range (Yamamoto et al., 2020).

### PK Inhibitors

Dasatinib was approved by the Food and Drug Administration (FDA) about 15 years ago and is used for the treatment of chronic myelogenous leukemia (CML) and acute lymphoblastic leukemia (Keskin et al., 2016). It acts as an ATP-competitive inhibitor of different tyrosine kinases such as Bcr-Abl and the Src PK family (Keskin et al., 2016). The obtained data suggested that this drug could efficiently bind to PLpro. Among all the screened compounds, dasatinib returned the best docking score (−10.46 kcal/mol; **Table 1**), outperforming VIR251 (−9.213 kcal/mol). As confirmed by the computed TC-IFP (0.648), the predicted binding mode mimics that of VIR251. Dasatinib was predicted to establish four well-oriented H-bond interactions with the PLpro BS (**Figure 2A**), in particular with the G163 backbone C=O and G271 backbone NH (as observed for VIR251; **Figure 1B**), as well as with the backbone of N109 and C270.

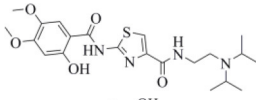
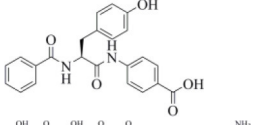
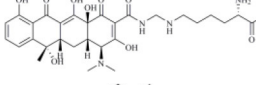
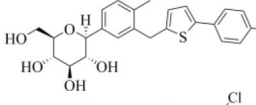
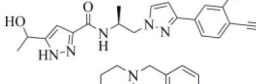
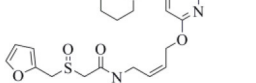
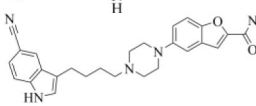
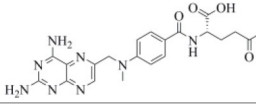
Interestingly, in 2014, dasatinib showed promising antiviral activities against other coronaviruses such as MERS-CoV

**TABLE 1** | Candidate drugs for non-covalent PLpro inhibition.

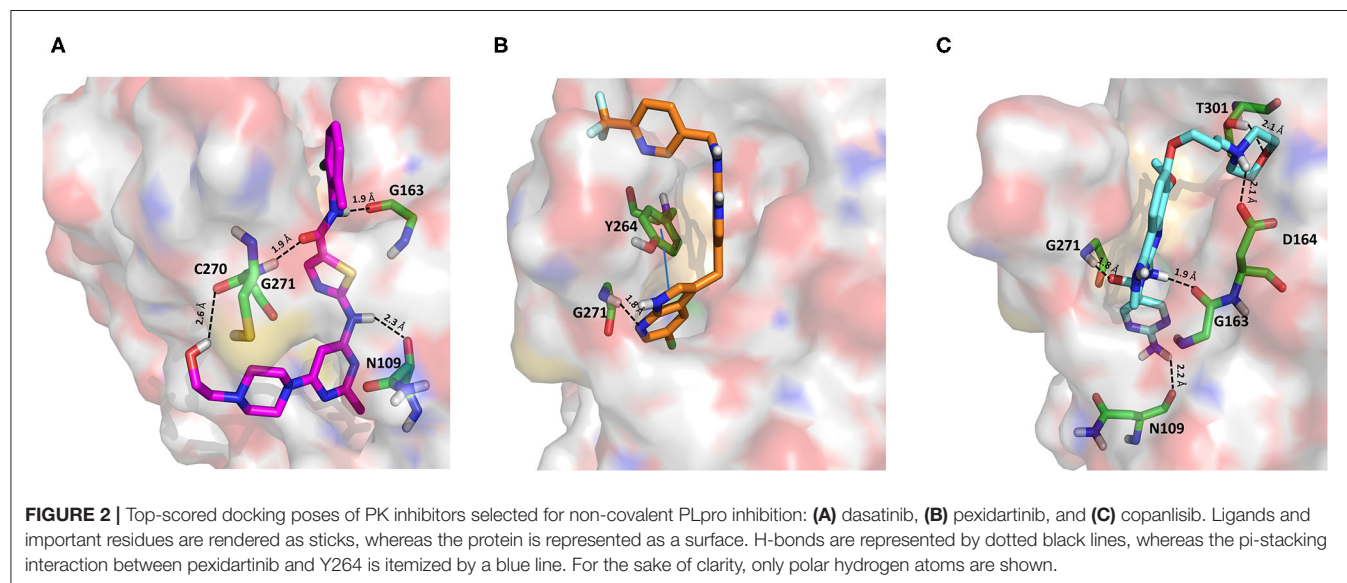
CHEMBL ID	Compound	2D structure	Docking score (rank)	TC-IFP (rank)	MM-GBSA score (rank)	Original mechanism of action
/	<b>VIR251</b>		<b>−9.21</b>	<b>—</b>	<b>−91.27</b>	<b>PLpro inhibitor</b>
1421	Dasatinib		−10.46 (1)	0.648 (68)	−75.72 (71)	PK inhibitors
3813873	Pexidartinib		−6.24 (225)	0.582 (166)	−75.96 (68)	
3218576	Copanlisib		−6.90 (145)	0.494 (299)	−81.32 (40)	
116	Amprenavir		−5.50 (412)	0.677 (31)	−76.18 (67)	Protease inhibitors
115	Indinavir		−6.31 (212)	0.513 (274)	−75.77 (70)	
1929396	Anagliptin		−6.08 (252)	0.687 (27)	−54.16 (258)	
218394	Boceprevir		−6.01 (271)	0.507 (283)	−73.95 (83)	
520733	Semagacestat		−9.36 (14)	0.591 (151)	−67.41 (128)	
1198857	Vilanterol		−5.78 (327)	0.653 (60)	−100.57 (1)	Adrenergic receptor modulators
1363	Arformoterol		−5.61 (370)	0.662 (49)	−72.08 (93)	
24	Atenolol		−5.47 (423)	0.613 (108)	−71.26 (97)	
515606	Cilazaprilat		−5.79 (312)	0.682 (28)	−51.74 (275)	ACE inhibitors and direct oral anticoagulants
1269025	Edoxaban		−7.97 (59)	0.762 (5)	−77.28 (61)	
198362	Rivaroxaban		−5.26 (477)	0.587 (161)	−68.85 (115)	

(Continued)

**TABLE 1** | Continued

CHEMBL ID	Compound	2D structure	Docking score (rank)	TC-IFP (rank)	MM-GBSA score (rank)	Original mechanism of action
2107723	Acotiamide		−8.07 (55)	0.706 (19)	−82.16 (35)	Drugs belonging to other classes
1200368	Bentiromide		−8.80 (27)	0.781 (2)	−59.55 (200)	
2103929	Lymecycline		−6.78 (161)	0.701 (22)	−94.45 (8)	
2103841	Canagliflozin		−5.87 (297)	0.530 (243)	−83.73 (28)	
4297185	Darolutamide		−10.03 (2)	0.698 (24)	−83.45 (30)	
1742461	Lafutidine		−8.20 (54)	0.662 (43)	−77.86 (57)	
439849	Vilazodone		−5.98 (279)	0.742 (10)	−70.76 (101)	
34259	Methotrexate		−5.41 (433)	0.458 (345)	−38.42 (408)	

Docking scores and MM-GBSA-binding free energy values are reported in kcal/mol. The numbers in brackets indicate the corresponding rank positions. Docking and MM-GBSA scores obtained by redocking VIR251 (non-covalent docking protocol) are also reported.



( $EC_{50} = 17.6 \mu M$ ) and SARS-CoV ( $EC_{50} = 2.1 \mu M$ ) (Dyall et al., 2014) and has been recently used to treat a CML patient with a concomitant SARS-CoV-2 infection (Abruzzese et al., 2020). Therefore, it has been hypothesized that

the ABL1 pathway could have an important role in viral replication. Our findings suggested an alternative explanation; i.e., the detected antiviral activity may be the result of PLpro inhibition.

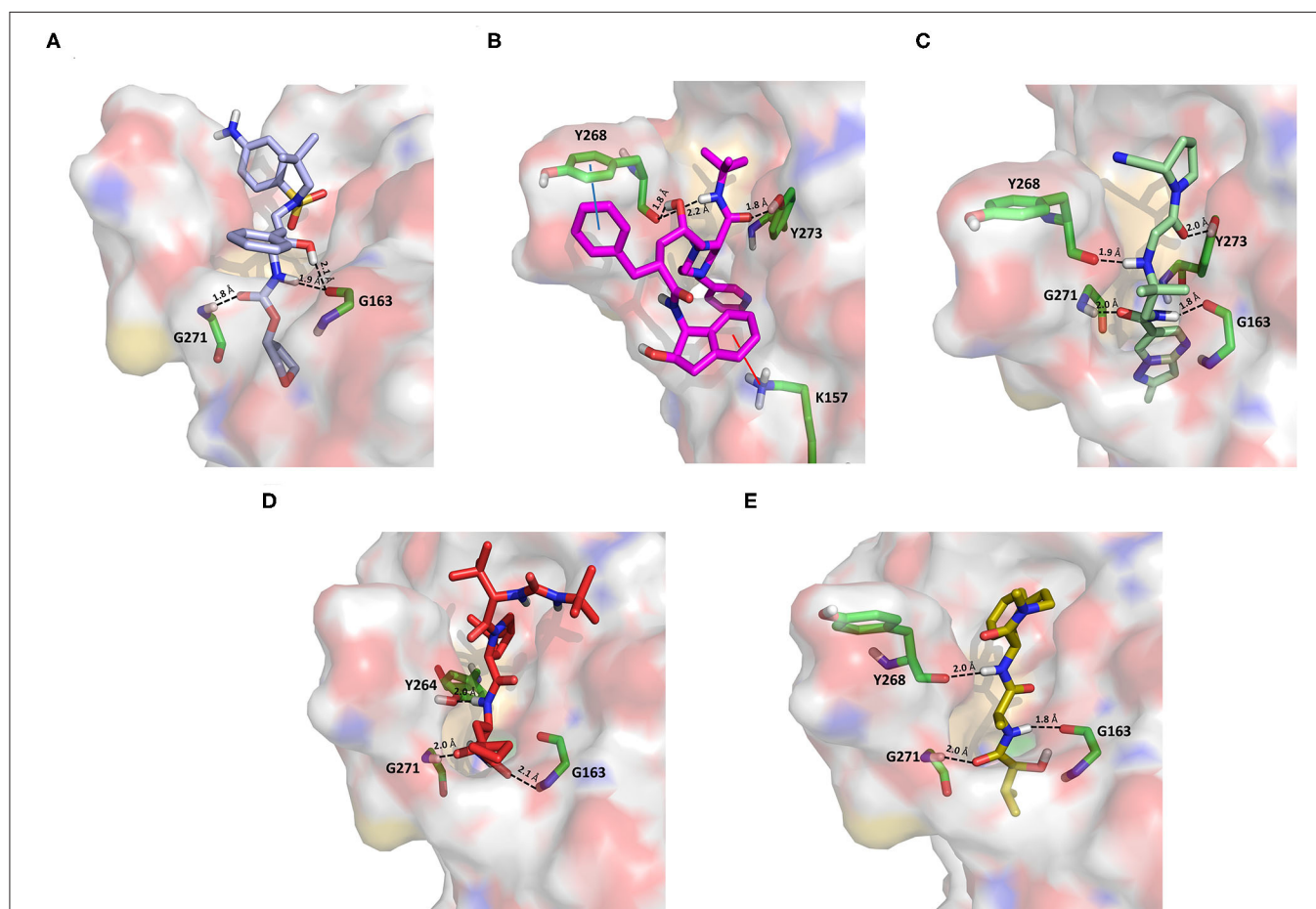
Pexidartinib is a tyrosine kinase inhibitor recently approved by the FDA for the treatment of adults with symptomatic tenosynovial giant cell tumor (Benner et al., 2020). In particular, it works by inhibiting the colony-stimulating factor 1 receptor (Benner et al., 2020). Interestingly, pexidartinib returned an MM-GBSA-binding free energy comparable to that of dasatinib and a top-scored docking pose showing a binding mode in good agreement with that of VIR251 (TC-IFP = 0.582). In particular, as observed for the co-crystallized inhibitor, this compound interacts with the G271 backbone NH and the side chain of Y264, although by establishing a T-shaped pi-stacking rather than an H-bond interaction. It also shares with VIR251 the same orientation within the hydrophobic S4 pocket, as is evident by comparing **Figures 1B, 2B**. Noteworthy, pexidartinib is able to cross the blood-brain barrier (Butowski et al., 2016), a desirable property for treating patients because SARS-CoV-2 particles in the central nervous system (CNS) may be responsible for COVID-19 neurological manifestations (Baig et al., 2020; Zubair et al., 2020).

Copanlisib is a selective phosphoinositide 3-kinase (PI3K) inhibitor approved by the FDA for treating follicular lymphoma

(Tarantelli et al., 2020). As reported in **Table 1**, it returned one of the best MM-GBSA scores (−81.32 kcal/mol). Copanlisib is able to mimic the binding mode observed for VIR251, in particular, by establishing H-bond interactions with the backbone of G163 and G271, as well as with the side chain of D164. An exception is represented by its morpholin group exploring a subpocket alternative to S4 where it is involved in an H-bond interaction with the T301 side chain. This evidence justifies, at least in part, the low TC-IFP value returned by this compound (0.482). Finally, an H-bond interaction with the N109 backbone C=O was also detected (**Figure 2C**). Noteworthy, Kindrachuk et al. showed that PK inhibitors targeting the PI3K/AKT/mTOR pathway are able to *in vitro* inhibit MERS-CoV replication, thus suggesting these compounds as promising tools for the treatment of coronavirus infections (Kindrachuk et al., 2015).

### Protease Inhibitors

Amprenavir is an HIV-1 protease inhibitor approved by the FDA in 1999 (Fung et al., 2000). Herein it was predicted as an efficient PLpro binder based on the computed MM-GBSA score (−76.18 kcal/mol). Notably, the obtained docking pose well-reproduces



**FIGURE 3 |** Top-scored docking poses of protease inhibitors selected for noncovalent PLpro inhibition: **(A)** amprenavir, **(B)** indinavir, **(C)** anagliptin; **(D)** boceprevir, and **(E)** semagacestat. Ligands and important residues are rendered as sticks, whereas the protein is represented as a surface. H-bonds are represented by dotted black lines, whereas the cation pi (pi-stacking) interaction between indinavir and K157 (Y268) is itemized by a red line. For the sake of clarity, only polar hydrogen atoms are shown.

the binding mode of VIR251, as also shown by the high TC-IFP value (0.677). Amprenavir engages interactions with the backbone of both G271 and G163 and projects its P4 moiety in the S4 pocket as VIR251 (**Figure 3A**).

Noteworthy, this compound has been recently tested as a potential inhibitor of another SARS-CoV-2 protease (i.e., main protease, Mpro) and has been shown to be unable to inhibit Mpro at 20  $\mu\text{M}$  (Ma et al., 2020). Finally, it has already been tested as an inhibitor of viral replication in SARS-CoV-2-infected VeroE6/TMPRSS2 cells (cells constitutively expressing the serine protease TMPRSS2, which confers a high susceptibility to SARS-CoV-2 infection) returning an  $\text{EC}_{50}$  value equal to 31.32  $\mu\text{M}$  (Yamamoto et al., 2020).

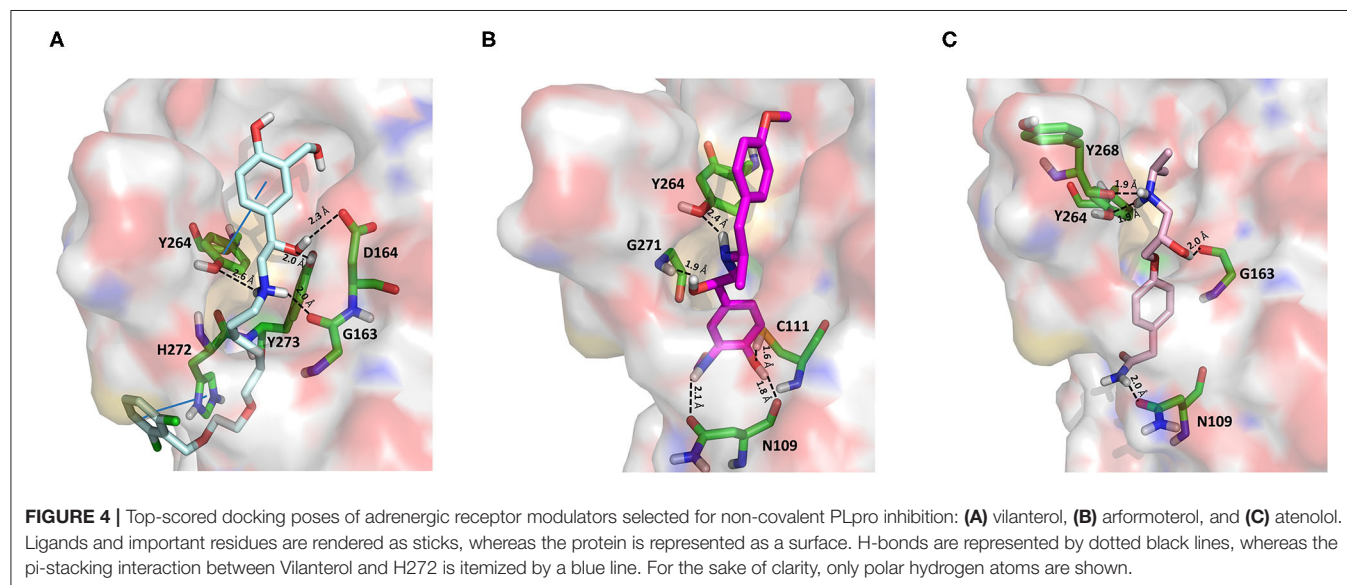
Approved by the FDA in 1996 for the treatment of AIDS, indinavir is a selective HIV-1 protease inhibitor with good oral bioavailability (Plosker and Noble, 1999). Herein it was predicted to bind to the PLpro catalytic site with a high affinity (MM-GBSA score:  $-75.77$  kcal/mol) by making two H-bonds with the backbone of Y268 and one with the side chain of Y273 (**Figure 3B**). In addition, a T-shaped pi-stacking interaction with Y268 and a cation-pi interaction with K157 were also detected. As amprenavir, indinavir has been recently proved to be an inhibitor of SARS-CoV-2 replication in cells ( $\text{EC}_{50} = 59.14$   $\mu\text{M}$ ) (Yamamoto et al., 2020).

Anagliptin belongs to the class of “gliptins” (i.e., DPP4 inhibitors), which are antidiabetes drugs currently used by millions of patients and known to have a high safety profile (Nishio et al., 2015). It is still in phase III development in the United States and European Union, whereas in Japan it has been recently approved for use. Based on its docking pose, well-mimicking the VIR251 interaction pattern (TC-IFP: 0.687) and the predicted binding affinities, anagliptin may efficiently bind to PLpro. As shown in **Figure 3C**, this drug is able to make some interactions also established by the co-crystallized inhibitor, namely, the H-bonds with the backbone of G163, G271, and Y268. In addition, it forms also an H-bond with the side chain

of Y273. It is worth to note that the administration of gliptins is expected to have beneficial effects on COVID-19 patients, with or without type 2 diabetes, because DPP4 is supposed to facilitate the entrance of SARS-CoV-2 in the airway tract (Solerte et al., 2020). These literature evidences, combined with the herein discussed results, put forward anagliptin as a drug that could modulate different relevant targets for COVID-19 therapy.

Boceprevir is an inhibitor of the non-structural protein 3/4A protease of the hepatitis C virus approved by the FDA in 2011 (Tungol et al., 2011). According to our results, boceprevir showed a favorable binding affinity to PLpro (MM-GBSA score:  $-73.95$  kcal/mol). This finding is in agreement with a previous computational screening performed on a PLpro homology model (Elfiky and Ibrahim, 2020). **Figure 3D** shows the obtained top-scored docking pose. Boceprevir was predicted to establish H-bond interactions with the backbone of G163 and G271, as well as with the side chain of Y264, thus mimicking the binding mode of VIR251. This drug has recently been reported to be an inhibitor of SARS-CoV-2 Mpro ( $\text{IC}_{50}$ : 4.13  $\mu\text{M}$ ) and a potent inhibitor of SARS-CoV-2 replication in cell culture ( $\text{EC}_{50}$ : 1.31  $\mu\text{M}$ ) (Ma et al., 2020). Herein we hypothesize that its antiviral activity might be the result of a synergistic effect on the two SARS-CoV-2 proteases.

Semagacestat is a drug in phase III clinical trials for Alzheimer disease (AD) treatment and its activity is related to the inhibition of a multisubunit protease complex named  $\gamma$ -secretase (Doody et al., 2013). Interestingly, this compound outperformed VIR251 in terms of docking score ( $-9.36$  vs.  $-9.21$  kcal/mol) and returned a good MM-GBSA score ( $-67.41$  kcal/mol), as well as a top-scored docking pose almost mimicking the binding mode of the cocrystallized inhibitor (TC-IFP: 0.591). In particular, as VIR251, semagacestat makes H-bonds with the G163 backbone C=O, the G271 backbone NH, and the Y268 side chain OH (**Figure 3E**). Being developed to treat AD, this compound is able to efficiently cross the blood-brain barrier, a required property to treat neurological manifestations in COVID-19 patients.



## Adrenergic Receptor Modulators

As an agonist of the  $\beta_2$ -adrenoreceptor, vilanterol was approved by the FDA in 2013 for the treatment of COPD (Ramadan et al., 2019). Herein, it was predicted to bind to PLpro with a high binding affinity as it returned the best MM-GBSA score ( $-100.57$  kcal/mol) among all the screened compounds (Table 1), exceeding the predicted binding free energy of VIR251 ( $-91.27$  kcal/mol). Importantly, its docking pose was predicted to be consistent with the binding mode of the cognate ligand (TC-IFP: 0.653), as shown in Figure 4A, making vilanterol all the important interactions for VIR251 recognition, such as H-bonds with the G163 backbone C=O and the side chains of D164 and Y264.

Additional interactions involve an H-bond with the side chain of Y273 and a T-shaped pi-stacking with H272 and Y264. As recently emphasized by Deslée et al. (2020), several evidences suggest a link between COVID-19 infection and COPD. More specifically, a higher expression of the ACE2 receptor has been observed in COPD patients. Therefore, vilanterol may show potential polypharmacological effects of interest for treating COVID-19 patients with COPD.

As vilanterol, arformeterol is a  $\beta_2$ -adrenoreceptor agonist approved for COPD treatment (King, 2008). It showed a good predicted binding affinity to PLpro (MM-GBSA score:  $-72.08$  kcal/mol) and a binding mode consistent with that of VIR251 (TC-IFP: 0.662). As reported in Figure 4B, arformeterol interacts with the G271 backbone NH and the side chain of Y264, as VIR251. Other H-bond interactions involve the side chain of C111, as well as the side chain and backbone of N109.

Atenolol is a second-generation cardioselective  $\beta_1$ -adrenergic antagonist, approved by the FDA in 1981. This drug is widely used for the management of hypertension, angina pectoris, cardiac dysrhythmias, and myocardial infarction (Rehman et al., 2020). In particular, atenolol has reached, in the United States, more than 20 million prescriptions in 2017 The Top 300 of 2020. Although relatively small (MW = 266 Da) compared to VIR251 (MW = 480 Da), this compound returned a good MM-GBSA score ( $-71.26$  kcal/mol) and TC-IFP (0.613). Interestingly, atenolol makes the majority of the interactions observed for VIR251, namely, H-bond interactions with the backbones of G163 and Y268, as well as with the side chain of Y264. In addition, a well-oriented H-bond with the side chain of N109 was also detected (Figure 4C). Noteworthy, the beneficial effect of  $\beta$ -adrenergic blockers for the treatment of COVID-19 patients has been recently hypothesized by Vasanthakumar on the basis of their ability to reduce “the mortality in respiratory failure, ARDS, and septic shock conditions” (Vasanthakumar, 2020).

## ACE Inhibitors and Direct Oral Anticoagulants

Cilazapril is a prodrug and is converted by carboxylesterases to cilazaprilat, a member of the class of angiotensin-converting enzyme inhibitors (ACE-is), i.e., drugs blocking the conversion of angiotensin I to angiotensin II (Deget and Brogden, 1991). Because of their ability to reduce cytokine production, ACE-is have been proposed as a possible therapeutic intervention to decrease the intensity of the host response to SARS-CoV-2 infection. Even if some authors have suggested that the

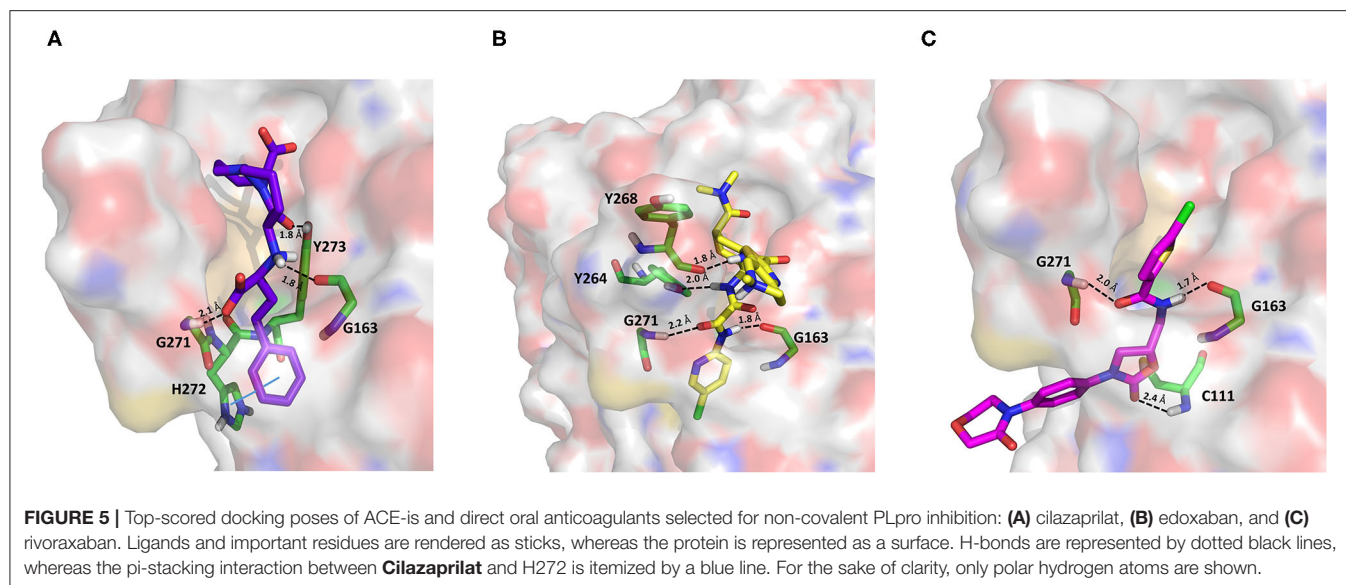
upregulation of ACE2 expression induced by a chronic use of ACE-is may be linked to the most severe outcomes associated with COVID-19, this hypothesis has not yet been experimentally confirmed. Moreover, some studies have shown lower IL-6 plasma levels, a lower rate of progression to severe complications, and a reduced mortality in COVID-19 patients treated with ACE-is (Braga et al., 2020). Cilazapril was among the ChEMBL docked compounds and was found to bind to the catalytic site of PLpro with a good stereoelectronic complementarity. Therefore, the active metabolite cilazaprilat was docked with the same docking protocol, and a docking pose similar to that of cilazapril was retrieved. Cilazaprilat makes H-bonds with the G163 backbone C=O and the G271 backbone NH, as the VIR251 cocrystallized inhibitor, and with the side chain OH of Y273 (Figure 5A). Moreover, it forms a pi-stacking interaction with H272 (docking score:  $-5.791$  kcal/mol, MM-GBSA score:  $-51.74$  kcal/mol, TC-IFP: 0.682).

Edoxaban and rivaroxaban are direct oral anticoagulants (DOACs) targeting factor Xa activity, commonly used in the therapy of patients with atrial fibrillation (Trujillo and Dobesh, 2014; Stacy et al., 2016). As DOACs are reported to interact with the P-glycoprotein and/or cytochrome P450-based metabolic pathways, many drugs such as antivirals administered to COVID-19 patients may interfere with their anticoagulant action. Therefore, for patients regularly assuming DOACs, clinicians have recommended to replace DOACs with heparin to avoid drug-drug interactions (Testa et al., 2020). However, in COVID-19 patients, the coagulation function is heavily unbalanced leading to hypercoagulation and the development of life-threatening coagulopathies, which may negatively affect the prognosis (Han et al., 2020; Pryzdial et al., 2020). Indeed, the alteration of blood clotting and inflammation are two frequently coupled manifestations of viral infections. Therefore, DOACs such as apixaban have shown an antiviral activity on herpes simplex virus type 1 and have been proposed as a possible therapeutic strategy to control COVID-19 (Pryzdial et al., 2020).

In our VS campaign, both rivaroxaban and edoxaban were shown to make favorable interactions within the PLpro catalytic site. As VIR251, rivaroxaban (docking score:  $-5.262$  kcal/mol, MM-GBSA score:  $-68.85$  kcal/mol, TC-IFP = 0.587) is hydrogen bonded to the G163 backbone C=O and the G271 backbone NH, and in addition, it forms an H-bond with the C111 NH (Figure 5C), whereas edoxaban (docking score:  $-7.973$  kcal/mol, MM-GBSA score:  $-77.28$  kcal/mol, TC-IFP: 0.762) is characterized by a high TC-IFP score mimicking VIR251 H-bonds with the backbone carbonyl group of G163, the backbone NH of G271, and the Y264 side chain OH. Furthermore, edoxaban is at hydrogen bond distance from the backbone C=O of Y268 and part of the compound projects toward the deep pocket flanking the S4 pocket, thus potentially exploring other still unexplored regions of the BS (Figure 5B).

## Drugs Belonging to Other Classes

Acotiamide is an acetylcholinesterase inhibitor approved in Japan for treating dyspepsia and functional dyspepsia (Bhalla, 2017). In Europe and United States, it is undergoing phase III clinical trials with promising results. Our data suggested that acotiamide



may be a strong inhibitor of PLpro. As reported in **Table 1**, this hypothesis is supported by the computed docking ( $-8.07$  kcal/mol) and MM-GBSA ( $-82.16$  kcal/mol) scores, being among the best scores returned by all the screened compounds. This compound interacts with the side chains of D164 and Y264, as well as with the backbones of G163 and G271 (**Figure 6A**), thus reproducing the binding mode of VIR251, as also confirmed by the computed TC-IFP, being equal to 0.706.

Our data, combined with its proved high safety profile (Tack et al., 2018), make this drug an ideal candidate for further testing.

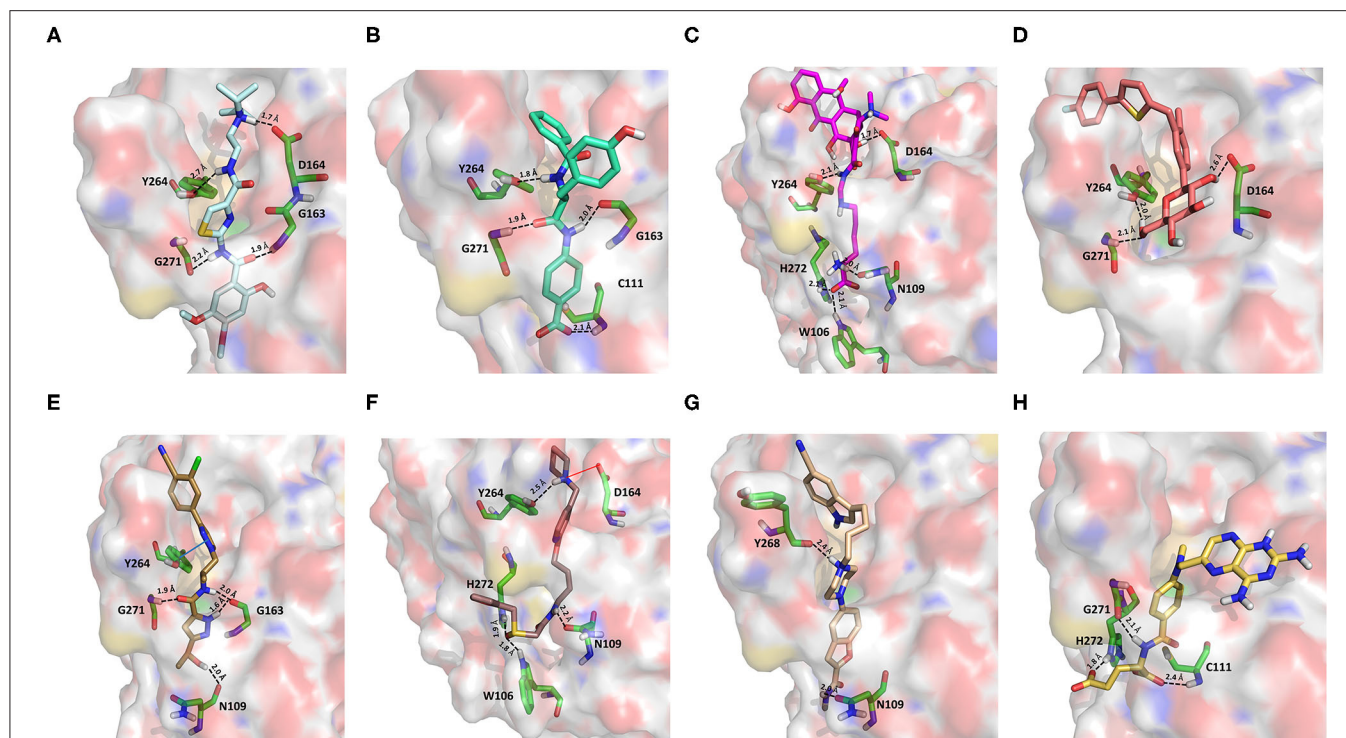
Bentiromide is an orally administrated dipeptide used in the so-called “bentiromide test” for the evaluation of the pancreatic exocrine function (Weizman et al., 1985). According to our results, bentiromide may efficiently bind to PLpro. This is mainly supported by the computed docking score ( $-8.80$  kcal/mol), close to that returned by redocking VIR251 ( $-9.21$  kcal/mol) and by the very high TC-IFP (0.781). Indeed, bentiromide interacts *via* H-bonds with the backbones of G163, G271, and Y264, as in the case of VIR251. In addition, an H-bond interaction with the backbone of C111 and a pi-stacking interaction with the side chain of Y264 were also observed (**Figure 6B**).

Lymecycline is a broad-spectrum antibiotic belonging to the tetracycline class and approved for use in Europe (Stratford, 1965). As reported in **Table 1**, it was predicted to have a good affinity to PLpro, with an MM-GBSA score ( $-94.45$  kcal/mol) better than that returned by VIR251 ( $-91.27$  kcal/mol) and a high value of TC-IFP (0.701). Lymecycline is hydrogen bonded to the side chains of D164, Y264, and W106 as VIR251. In addition, H-bond interactions with the side chains of N109 and H272 were also observed (**Figure 6C**). Because of their immunomodulatory and anti-inflammatory properties, combined with their well-known safety profile, tetracyclines are considered ideal candidates for repurposing against SARS-CoV-2, as recently highlighted by Singh et al. (2020). Importantly, this class of antibiotics has shown a potential efficacy in patients

with ARDS, one of the most common clinical manifestations in COVID-19 patients (Singh et al., 2020).

Canagliflozin is a drug approved by the FDA for the treatment of type 2 diabetes, one of the main risk factors for severe COVID-19 outcomes (Jakher et al., 2019). In particular, it acts as an inhibitor of the sodium glucose co-transporter-2 (SGLT2). According to our data, this drug binds to PLpro with a good affinity (MM-GBSA score:  $-83.73$  kcal/mol) and mimics the binding mode of VIR251, being able to make H-bonds with the side chains of D164 and Y264 and with the backbone of G271 (**Figure 6D**). Moreover, SGLT2 inhibitors have recently been proved to have a protective effect on the heart, kidney, and lung, and their potential benefits in COVID-19 patients have been hypothesized on the basis of a clinical trial showing the impact of dapagliflozin, a parent compound of canagliflozin, in patients with respiratory failure (Fernandez-Fernandez et al., 2020).

Approved by the FDA in 2019 for the treatment of metastatic castration-resistant prostate cancer (Research CDE, 2019), darolutamide is a non-steroidal antagonist of the androgen receptor. Our data indicated that this drug may efficiently bind to PLpro. Indeed, its docking pose scored better than that of VIR151 ( $-10.03$  vs.  $-9.21$  kcal/mol) and showed a high TC-IFP (0.698) and a good MM-GBSA score ( $-83.45$  kcal/mol). Darolutamide makes H-bond interactions with the backbones of G163, G271, and N109, as well as a well-oriented pi-stacking interaction with Y264 (**Figure 6E**). As highlighted by Sugawara et al., darolutamide is responsible for a reduced expression of TMPRSS2 (Sugawara et al., 2019), the serine protease proved to be implicated in the replication of SARS-CoV and MERS-CoV infections (Montopoli et al., 2020). Notably, the TMPRSS2 involvement in SARS-CoV-2 host cell entry has recently been hypothesized on the basis of epidemiological studies indicating that the development of a serious infection is less frequent in patients treated with androgen receptor antagonists (Montopoli et al., 2020). All these evidences combined with our *in-silico*



**FIGURE 6 |** Top-scored docking poses of candidates for non-covalent PLpro inhibition belonging to different classes: **(A)** acotiamide, **(B)** bentiromide, **(C)** lymecycline, **(D)** canagliflozin, **(E)** darolutamide, **(F)** lafutidine, **(G)** vilazodone, and **(H)** methotrexate. Ligands and important residues are rendered as sticks, whereas the protein is represented as a surface. H-bonds are represented by dotted black lines, whereas the salt bridge interaction between lafutidine and D164 is itemized by a red line. For the sake of clarity, only polar hydrogen atoms are shown.

findings make darolutamide an ideal candidate for further *in vitro* testing.

Lafutidine is a new second-generation histamine H2 receptor antagonist (H2RA) with gastroprotective actions. It acts by inhibiting the daytime secretion of gastric acid by acting both directly on the H2 receptors and indirectly by increasing gastric nitric oxide production (Nakano et al., 2011). Although lafutidine has already been approved and marketed in Japan and India to treat gastric ulcers, it is still in phase III development in European Union and United States. This compound returned one of the best docking scores ( $-8.20$  kcal/mol) among the selected molecules (Table 1) and a TC-IFP value (0.662), indicating a binding mode similar to that of the VIR251 inhibitor. As shown in Figure 6F, lafutidine was predicted to make H-bonds with the side chains of H272, W106, N109, and Y264. A T-shaped pi-stacking with Y264 and a salt bridge with D164 were also observed. It is worth to note that the administration of another H2RA (famotidine) has been associated to a reduced risk of death in COVID-19 patients (Freedberg et al., 2020). Moreover, Aguila et al. speculated that H2RAs could be a good option for COVID-19 treatment because of their ability to interfere with the gastric pH (Aguila and Cua, 2020). Therefore, its potential multitarget activity could show promise for COVID-19 treatment.

Approved by the FDA in 2011 to treat major depressive disorder, a condition that affects approximately 200 million people worldwide (Mirzaei et al., 2019), vilazodone acts as a

5-HT<sub>1A</sub> receptor partial agonist and is the only drug currently defined as a serotonin partial agonist-reuptake inhibitor (SPARI) (Stahl, 2014). As reported in Table 1, it returned a promising MM-GBSA score ( $-70.76$  kcal/mol) and a very high TC-IFP (0.742). In particular, this compound interacts with PLpro via H-bond interactions with the backbone of Y268, as VIR251, and the side chain of N109. Remarkably, the orientation of its indole substituent within the S4 cavity is almost superimposable with the crystallographic coordinates of VIR251, as shown in Figure 6G. If confirmed by experiments, vilazodone ability to inhibit PLpro would be particularly appealing for the treatment of COVID-19 patients with neurological manifestations (Armocida et al., 2020), because of its ability to efficiently cross the blood-brain barrier (Bundgaard et al., 2016).

Methotrexate is a well-known antineoplastic, immunosuppressive, and anti-inflammatory agent that inhibits dihydrofolate reductase preventing the formation of tetrahydrofolate, which is required for DNA synthesis (Hannood and Mittal, 2020). Because of its anti-inflammatory effects, it has been recently proposed for the treatment of COVID-19 hyperinflammation (Safavi and Nath, 2020). In our study, we found that methotrexate (docking score:  $-5.417$  kcal/mol, MM-GBSA score:  $-38.42$  kcal/mol, TC-IFP: 0.458) is able to bind to the PLpro catalytic site, albeit exploring a partially different hydrogen bond network compared to VIR251. Indeed, methotrexate makes hydrogen bond interactions with the G271

backbone C=O, the backbone NH of C111, and the side chains of C111 and H272 (**Figure 6H**).

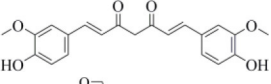
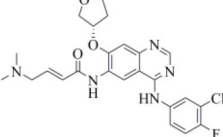
## Candidates for Covalent PLpro Inhibition

All the compounds belonging to the developed dataset and including Michael acceptor (MA) groups, potentially able to alkylate C111 as the VIR251 cocrystallized inhibitor, were submitted to covalent docking simulations. As reviewed some years ago by Santos et al., MA groups can be responsible for an irreversible and very effective inactivation of cysteine proteases (Santos and Moreira, 2007). A covalent mechanism of action, indeed, leads to several advantages in terms of potency, duration of action, and selectivity (Singh et al., 2011) and therefore is highly desirable for COVID-19 treatment. On the basis of the obtained docking scores and visual inspection, two compounds

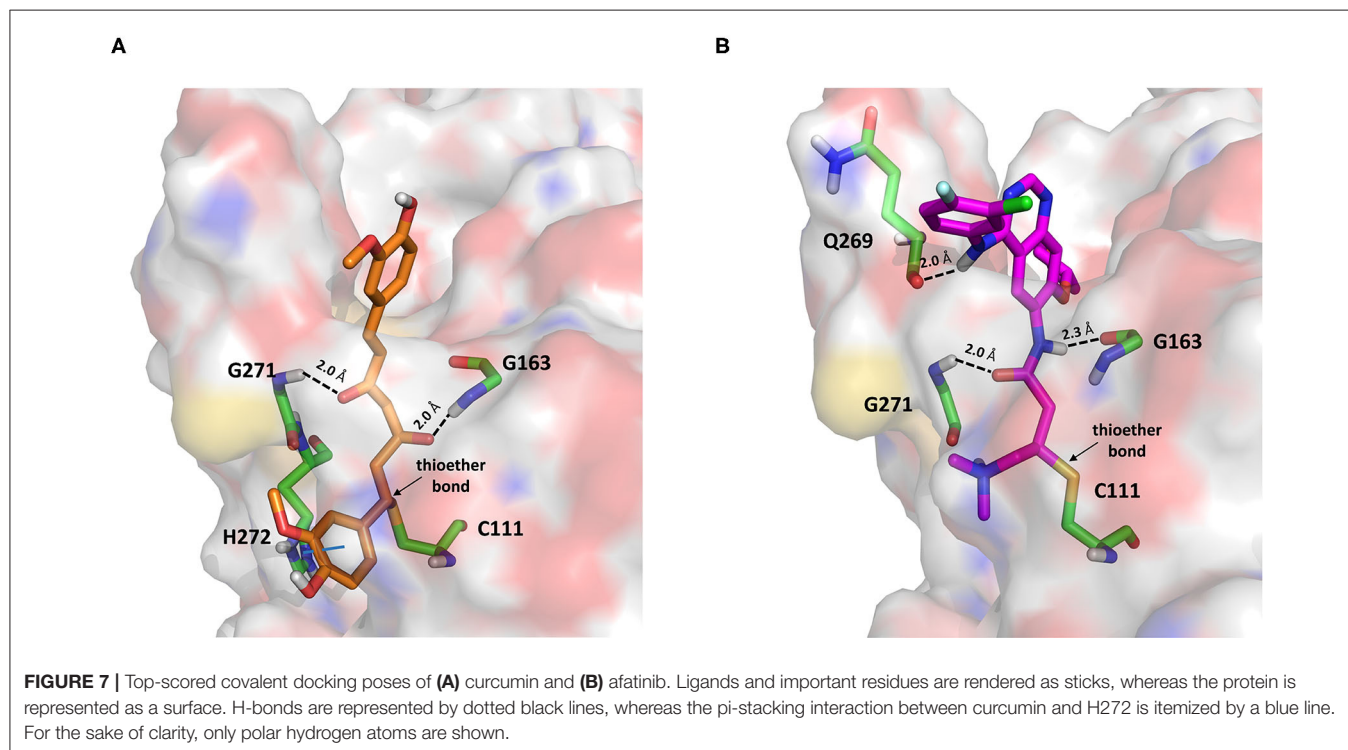
were selected as the best candidates for a covalent inhibition of PLpro (**Table 2**).

Curcumin is a polyphenol extracted from an East Indian plant *Curcuma longa* that reached a phase III clinical trial for the treatment of inoperable pancreatic cancer (Hatcher et al., 2008). It has a long history of use as a food additive due to its potent anti-inflammatory and antioxidative properties (Praditya et al., 2019), whereas no toxicity concerns are associated with its administration. This compound returned one of the best docking scores among all the screened molecules (−8.051 kcal/mol) and its top-scored docking pose mimics some key interactions observed for VIR251 such as those with the backbones of G163 and G271 (TC-IFP: 0.609). Finally, a pi-stacking with H272 was also observed (**Figure 7A**).

**TABLE 2** | Candidate drugs for covalent PLpro inhibition.

CHEMBL ID	Compound	2D structure	Docking score (rank)	TC-IFP (rank)
/	VIR251		−10.08	—
116438	Curcumin		−8.05 (4)	0.609 (7)
1173655	Afatinib		−5.80 (10)	0.603 (9)

Docking scores are reported in kcal/mol. The numbers in brackets indicate the corresponding rank positions. The docking score obtained by redocking VIR251 (covalent docking protocol) is also reported.



Importantly, the ability of curcumin to interact, *via* a Michael addition, with a cysteine residue has been well-documented by studies on other pharmacological targets such as the myeloid differentiation protein 2 (Gradišar et al., 2007) and the transcription factor STAT-3 (Hahn et al., 2018). Furthermore, the antiviral properties of this compound against several viruses have recently been reviewed by Praditya et al. (2019). In particular, curcumin exhibited a significant inhibitory effect against SARS-CoV (Wen et al., 2007). Last but not least, a recent study has confirmed that this polyphenol exerts a protective effect on the lung in case of severe pneumonia caused by SARS-CoV-2, decreasing the expression of proinflammatory cytokines (IL-6, IL-8, and IL-10) (Liu and Ying, 2020). All these evidences, combined with our findings, make curcumin an ideal candidate for further investigations.

Afatinib is a tyrosine kinase inhibitor approved by the FDA in 2013 for the treatment of advanced non-small cell lung cancer (Deeks and Keating, 2018). Interestingly, it is a well-known covalent inhibitor of different proteins belonging to the ErbB family such as, e.g., the epidermal growth factor receptor (Deeks and Keating, 2018). Indeed, the presence of an MA group allows the reaction with a conserved cysteine residue of the catalytic cleft (Yu et al., 2018). According to the obtained docking score ( $-5.799$  kcal/mol) and pose, herein we hypothesize that afatinib could bind to PLpro by means of the same mechanism observed in different tyrosine kinases. Afatinib is well-accommodated in the catalytic site (TC-IFP: 0.603) making H-bond interactions with the backbone of G171, G163, and Q269 (Figure 7B). Noteworthy, ErbB receptors have recently been hypothesized to have an important role in different stages of viral infections (Ho et al., 2017), such as host cell entry and proliferation; hence, afatinib can be considered as a drug with the potential of targeting both host proteins engaged by the virus and viral targets.

## CONCLUSIONS

In this article, we selected 24 known drugs as promising non-covalent (22) and covalent (2) inhibitors of the SARS-CoV-2 papain-like protease for the treatment of COVID-19 patients. All the compounds were selected through a structure-based computational screening performed by using, for the first time,

the crystal structure of PLpro in complex with an inhibitor. This study differs from other *in silico* screenings performed for repurposing drugs on SARS-CoV-2 protein targets as we (i) extended our investigation to compounds that have reached phase III clinical trial; (ii) rescored the obtained docking poses on the basis of their computed IFPs; and (iii) performed both non-covalent and covalent docking simulations. The selected compounds, belonging to different pharmacological classes, such as that of protease inhibitors (amprenavir, indinavir, anagliptin, boceprevir, and semagacestat), adrenergic receptor modulators (vilanterol, arformeterol, atenolol), anticoagulants (edoxaban and rivaroxaban), ACE-is (cilazapril), antidiabetes (anagliptin, canagliflozin), PK inhibitors (dasatinib, pexidartinib, copanlisib, and afatinib), and antiandrogens (darolutamide), can be considered as promising candidates for further *in vitro* testing to select or discard them as SARS-CoV-2 papain-like protease inhibitors. Importantly, according to the available literature data and well-reported clinical trials, all the proposed compounds have a known safety profile, and for the majority of them, polypharmacological effects highly desirable to treat COVID-19 patients can be hypothesized because of the concomitant inhibition of viral and host proteins involved in viral infection. Therefore, once their antiviral activity could be confirmed, these drugs may represent a ready-to-use treatment for hindering SARS-CoV-2 devastating effects.

## DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

## AUTHOR CONTRIBUTIONS

PD performed the calculations, supervised by FC and GM. All authors contributed to the design and implementation of the research, to the analysis of the results and to the writing of the manuscript.

## ACKNOWLEDGMENTS

The authors thank Dr. Candida Giovannelli for technical support.

## REFERENCES

- Abruzzese, E., Luciano, L., D'Agostino, F., Trawinska, M. M., Pane, F., and De Fabritiis, P. (2020). SARS-CoV-2 (COVID-19) and chronic myeloid leukemia (CML): a case report and review of ABL kinase involvement in viral infection. *Mediterr. J. Hematol. Infect. Dis.* 12:e2020031. doi: 10.4084/mjhid.2020.031
- Aguila, E. J. T., and Cua, I. H. Y. (2020). Repurposed GI drugs in the treatment of COVID-19. *Dig Dis. Sci.* 65, 2452–2453. doi: 10.1007/s10620-020-06430-z
- Amin, S. A., Ghosh, K., Gayen, S., and Jha, T. (2020). Chemical-informatics approach to COVID-19 drug discovery: monte carlo based QSAR, virtual screening and molecular docking study of some in-house molecules as papain-like protease (PLpro) inhibitors. *J. Biomol. Struct. Dynam.* 1–10. doi: 10.1080/07391102.2020.1780946
- Armocida, D., Palmieri, M., Frati, A., Santoro, A., and Pesce, A. (2020). How SARS-Cov-2 can involve the central nervous system. A systematic analysis of literature of the department of human neurosciences of sapienza university, Italy. *J. Clin. Neurosci.* 79, 231–236. doi: 10.1016/j.jocn.2020.07.007
- Baig, A. M., Khaleeq, A., Ali, U., and Syeda, H. (2020). Evidence of the COVID-19 virus targeting the CNS: tissue distribution, host-virus interaction, and proposed neurotropic mechanisms. *ACS Chem. Neurosci.* 11, 995–998. doi: 10.1021/acscchemneuro.0c00122
- Benner, B., Good, L., Quiroga, D., Schultz, T. E., Kassem, M., Carson, W. E., et al. (2020). Pexidartinib, a novel small molecule CSF-1R inhibitor

- in use for tenosynovial giant cell tumor: a systematic review of pre-clinical and clinical development. *Drug Des. Devel. Ther.* 14, 1693–1704. doi: 10.2147/DDDT.S253232
- Bhalla, A. (2017). Acotiamide: a novel drug for the treatment of patients with functional dyspepsia. *Int. J. Basic Clin. Pharmacol.* 6, 1238–1241. doi: 10.18203/2319-2003.ijbcp20172220
- Braga, C. L., Silva-Aguiar, R. P., Battagliani, D., Peruchetti, D. B., Robba, C., Pelosi, P., et al. (2020). The renin-angiotensin-aldosterone system: role in pathogenesis and potential therapeutic target in COVID-19. *Pharmacol. Res. Perspect.* 8:e00623. doi: 10.1002/prp2.623
- Bundgaard, C., Eneberg, E., and Sánchez, C. (2016). P-glycoprotein differentially affects escitalopram, levomilnacipran, vilazodone and vortioxetine transport at the mouse blood-brain barrier *in vivo*. *Neuropharmacology* 103, 104–111. doi: 10.1016/j.neuropharm.2015.12.009
- Butowski, N., Colman, H., De Groot, J. F., Omuro, A. M., Nayak, L., Wen, P. Y., et al. (2016). Orally administered colony stimulating factor 1 receptor inhibitor PLX3397 in recurrent glioblastoma: an Ivy foundation early phase clinical trials consortium phase II study. *Neuro-Oncol.* 18, 557–564. doi: 10.1093/neuonc/nov245
- Contreras-Puentes, N., and Alvíz-Amador, A. (2020). Virtual screening of natural metabolites and antiviral drugs with potential inhibitory activity against 3CL-PRO and PL-PRO. *Biomed. Pharmacol. J.* 13, 933–941. doi: 10.13005/bpj/1962
- Coronavirus Disease (COVID-19). Available online at: <https://www.who.int/emergencies/diseases/novel-coronavirus-2019> (accessed July 28, 2020).
- Deeks, E. D., and Keating, G. M. (2018). Afatinib in advanced NSCLC: a profile of its use. *Drugs Ther. Perspect.* 34, 89–98. doi: 10.1007/s40267-018-0482-6
- Deget, F., and Brogden, R. N. (1991). Cilazapril. A review of its pharmacodynamic and pharmacokinetic properties, and therapeutic potential in cardiovascular disease. *Drugs* 41, 799–820. doi: 10.2165/00003495-199141050-00008
- Deng, Z., Chuaqui, C., and Singh, J. (2004). Structural interaction fingerprint (SIFt): A novel method for analyzing three-dimensional protein–ligand binding interactions. *J. Med. Chem.* 47, 337–344. doi: 10.1021/jm030331x
- Deslée, G., Zysman, M., Burgel, P.-R., Perez, T., Boyer, L., Gonzalez, J., et al. (2020). Chronic obstructive pulmonary disease and the COVID-19 pandemic: reciprocal challenges. *Respir. Med. Res.* 78:100764. doi: 10.1016/j.resmer.2020.100764
- Doody, R. S., Raman, R., Farlow, M., Iwatsubo, T., Vellas, B., Joffe, S., et al. (2013). A phase 3 trial of semagacestat for treatment of Alzheimer's disease. *N. Engl. J. Med.* 369, 341–350. doi: 10.1056/NEJMoa1210951
- Dyall, J., Coleman, C. M., Hart, B. J., Venkataraman, T., Holbrook, M. R., Kindrachuk, J., et al. (2014). Repurposing of clinically developed drugs for treatment of Middle East respiratory syndrome coronavirus infection. *Antimicrob. Agents Chemother.* 58, 4885–4893. doi: 10.1128/AAC.03036-14
- Elfiky, A., and Ibrahim, N. S. (2020). Anti-SARS and anti-HCV drugs repurposing against the Papain-like protease of the newly emerged coronavirus (2019-nCoV). *Res. Square [Preprint]*. doi: 10.21203/rs.2.23280/v1
- Fernandez-Fernandez, B., D'Marco, L., Górriz, J. L., Jacobs-Cachá, C., Kanbay, M., Luis-Lima, S., et al. (2020). Exploring sodium glucose Co-Transporter-2 (SGLT2) inhibitors for organ protection in COVID-19. *J. Clin. Med.* 9:2030. doi: 10.3390/jcm9072030
- Freedberg, D. E., Conigliaro, J., Sobieszczyk, M. E., Markowitz, D. D., Gupta, A., O'Donnell, M. R., et al. (2020). Famotidine use is associated with improved clinical outcomes in hospitalized COVID-19 patients: a propensity score matched retrospective Cohort study. *medRxiv*. doi: 10.1053/j.gastro.2020.05.053
- Freeman, T. L., and Swartz, T. H. (2020). Targeting the NLRP3 inflammasome in severe COVID-19. *Front. Immunol.* 11:1518. doi: 10.3389/fimmu.2020.01518
- Friesner, R. A., Murphy, R. B., Repasky, M. P., Frye, L. L., Greenwood, J. R., Halgren, T. A., et al. (2006). Extra precision glide: docking and scoring incorporating a model of hydrophobic enclosure for protein–ligand complexes. *J. Med. Chem.* 49, 6177–6196. doi: 10.1021/jm051256o
- Fung, H. B., Kirschenbaum, H. L., and Hameed, R. (2000). Amprenavir: a new human immunodeficiency virus type 1 protease inhibitor. *Clin. Ther.* 22, 549–572. doi: 10.1016/S0149-2918(00)80044-2
- Gaulton, A., Hersey, A., Nowotka, M., Bento, A. P., Chambers, J., Mendez, D., et al. (2017). The ChEMBL database in 2017. *Nucleic Acids Res.* 45, D945–D954. doi: 10.1093/nar/gkw1074
- Genheden, S., and Ryde, U. (2015). The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. *Expert Opin. Drug Discov.* 10, 449–461. doi: 10.1517/17460441.2015.1032936
- Gorbalenya, A. E., Baker, S. C., Baric, R. S., de Groot, R. J., Drosten, C., Gulyaeva, A. A., et al. (2020). The species severe acute respiratory syndrome-related coronavirus : classifying 2019-nCoV and naming it SARS-CoV-2. *Nat. Microbiol.* 5, 536–544. doi: 10.1038/s41564-020-0695-z
- Gordon, D. E., Jang, G. M., Bouhaddou, M., Xu, J., Obernier, K., White, K. M., et al. (2020). A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. *Nature* 583, 459–468. doi: 10.1038/s41586-020-2286-9
- Gradišar, H., Keber, M. M., Pristovšek, P., and Jerala, R. (2007). MD-2 as the target of curcumin in the inhibition of response to LPS. *J. Leukocyte Biol.* 82, 968–974. doi: 10.1189/jlb.1206727
- Hahn, Y.-I., Kim, S.-J., Choi, B.-Y., Cho, K.-C., Bandu, R., Kim, K. P., et al. (2018). Curcumin interacts directly with the Cysteine 259 residue of STAT3 and induces apoptosis in H-Ras transformed human mammary epithelial cells. *Sci. Rep.* 8:6409. doi: 10.1038/s41598-018-23840-2
- Han, H., Yang, L., Liu, R., Liu, F., Wu, K.-L., Li, J., et al. (2020). Prominent changes in blood coagulation of patients with SARS-CoV-2 infection. *Clin. Chem. Lab. Med.* 58, 1116–1120. doi: 10.1515/cclm-2020-0188
- Hannood, M., and Mittal, M. (2020). “Methotrexate,” in *StatPearls* (Treasure Island, FL: StatPearls Publishing). Available online at: <http://www.ncbi.nlm.nih.gov/books/NBK556114/> (accessed July 28, 2020).
- Hatcher, H., Planalp, R., Cho, J., Torti, F. M., and Torti, S. V. (2008). Curcumin: From ancient medicine to current clinical trials. *Cell. Mol. Life Sci.* 65, 1631–1652. doi: 10.1007/s00018-008-7452-4
- Ho, J., Moyes, D. L., Tavassoli, M., and Naglik, J. R. (2017). The role of ErbB receptors in Infection. *Trends Microbiol.* 25, 942–952. doi: 10.1016/j.tim.2017.04.009
- Jakher, H., Chang, T. I., Tan, M., and Mahaffey, K. W. (2019). Canagliflozin review - safety and efficacy profile in patients with T2DM. *Diabetes Metab. Syndr. Obes.* 12, 209–215. doi: 10.2147/DMSO.S184437
- Keskin, D., Sadri, S., and Eskazan, A. E. (2016). Dasatinib for the treatment of chronic myeloid leukemia: patient selection and special considerations. *Drug Des. Devel. Ther.* 10, 3355–3361. doi: 10.2147/DDDT.S85050
- Kindrachuk, J., Ork, B., Hart, B. J., Mazur, S., Holbrook, M. R., Frieman, M. B., et al. (2015). Antiviral potential of ERK/MAPK and PI3K/AKT/mTOR signaling modulation for Middle East respiratory syndrome Coronavirus infection as identified by temporal kinome analysis. *Antimicrob. Agents Chemother.* 59, 1088–1099. doi: 10.1128/AAC.03659-14
- King, P. (2008). Role of arformoterol in the management of COPD. *Int. J. Chron. Obstruct Pulmon. Dis.* 3, 385–392. doi: 10.2147/COPD.S753
- Li, J., Abel, R., Zhu, K., Cao, Y., Zhao, S., and Friesner, R. A. (2011). The VSGB 2.0 model: A next generation energy model for high resolution protein structure modeling. *Proteins* 79, 2794–2812. doi: 10.1002/prot.23106
- LigPrep (2019). New York, NY: Schrödinger, LLC. Available online at: <https://www.schrodinger.com/ligprep> (accessed July 28, 2020).
- Liu, Z., and Ying, Y. (2020). The inhibitory effect of curcumin on virus-induced cytokine storm and its potential use in the associated severe pneumonia. *Front. Cell Dev. Biol.* 8:479. doi: 10.3389/fcell.2020.00479
- Llanes, A., Restrepo, C. M., Caballero, Z., Rajeev, S., Kennedy, M. A., and Lleona, R. (2020). Betacoronavirus genomes: how genomic information has been used to deal with past outbreaks and the COVID-19 pandemic. *Int. J. Mol. Sci.* 21:4546. doi: 10.20944/preprints202005.0448.v1
- Lu, R., Zhao, X., Li, J., Niu, P., Yang, B., Wu, H., et al. (2020). Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *Lancet* 395, 565–574. doi: 10.1016/S0140-6736(20)30251-8
- Lupia, T., Scabini, S., Mornese Pinna, S., Di Perri, G., De Rosa, F. G., and Corcione, S. (2020). 2019 novel coronavirus (2019-nCoV) outbreak: a new challenge. *J. Glob. Antimicrob. Resist.* 21, 22–27. doi: 10.1016/j.jgar.2020.02.021
- Ma, C., Sacco, M. D., Hurst, B., Townsend, J. A., Hu, Y., Szeto, T., et al. (2020). Boceprevir, GC-376, and calpain inhibitors II, XII inhibit SARS-CoV-2 viral replication by targeting the viral main protease. *Cell Res.* 30, 678–692. doi: 10.1101/2020.04.20.051581
- Madhavi Sastry, G., Adzhigirey, M., Day, T., Annabhimoju, R., and Sherman, W. (2013). Protein and ligand preparation: parameters, protocols, and influence

- on virtual screening enrichments. *J. Comput. Aided Mol. Des.* 27, 221–234. doi: 10.1007/s10822-013-9644-8
- Marcou, G., and Rognan, D. (2007). Optimizing fragment and scaffold docking by use of molecular interaction fingerprints. *J. Chem. Inf. Model.* 47, 195–207. doi: 10.1021/ci600342e
- Mirzaei, M., Yasini Ardekani, S. M., Mirzaei, M., and Dehghani, A. (2019). Prevalence of depression, anxiety and stress among adult population: results of Yazd health study. *Iran J. Psychiatr.* 14, 137–146. doi: 10.18502/ijps.v14i2.993
- Montopoli, M., Zumerle, S., Vettor, R., Rugge, M., Zorzi, M., Catapano, C. V., et al. (2020). Androgen-deprivation therapies for prostate cancer and risk of infection by SARS-CoV-2: a population-based study (N = 4532). *Ann Oncol.* 31, 1040–1045. doi: 10.1016/j.annonc.2020.04.479
- Nakano, M., Kitano, S., Nanri, M., and Kuniwa, M. (2011). Lafutidine, a unique histamine H2-receptor antagonist, inhibits distention-induced gastric acid secretion through an H2 receptor-independent mechanism. *Eur. J. Pharmacol.* 658, 236–241. doi: 10.1016/j.ejphar.2011.02.007
- Nishio, S., Abe, M., and Ito, H. (2015). Anagliptin in the treatment of type 2 diabetes: safety, efficacy, and patient acceptability. *Diabetes Metab. Syndr. Obes.* 8, 163–171. doi: 10.2147/DMSO.S54679
- Pinzi, L., Tinivella, A., Caporuscio, F., and Rastelli, G. (2020). Drug repurposing and polypharmacology to fight SARS-CoV-2 through the inhibition of the main protease. *Nat. Commun. [Preprint]*. doi: 10.21203/rs.3.rs-27222/v1
- Plosker, G. L., and Noble, S. (1999). Indinavir: a review of its use in the management of HIV infection. *Drugs* 58, 1165–1203. doi: 10.2165/00003495-199958060-00011
- Praditya, D., Kirchhoff, L., Brüning, J., Rachmawati, H., Steinmann, J., and Steinmann, E. (2019). Anti-infective properties of the golden spice curcumin. *Front. Microbiol.* 10:912. doi: 10.3389/fmicb.2019.00912
- Prezioso, C., Marcocci, M. E., Palamara, A. T., De Chiara, G., and Pietropaolo, V. (2020). The Three Italy of the COVID-19 epidemic and the possible involvement of SARS-CoV-2 in triggering complications other than pneumonia. *J. Neurovirol.* 26, 311–323. doi: 10.1007/s13365-020-00862-z
- Prime MM-GBSA (2019). New York, NY: Schrödinger, LLC. Available online at: <https://www.schrodinger.com/kb/1484> (accessed August 5, 2020).
- Protein Preparation Wizard (2019). New York, NY: Epik; Schrödinger, LLC. Available online at: <https://www.schrodinger.com/protein-preparation-wizard> (accessed July 28, 2020).
- Prydzial, E. L. G., Sutherland, M. R., Lin, B. H., and Horwitz, M. (2020). Antiviral anticoagulation. *Res. Pract. Thromb. Haemost.* 4, 774–788. doi: 10.1002/rth2.12406
- Quimque, M. T. J., Notarte, K. I. R., Fernandez, R. A. T., Mendoza, M. A. O., Liman, R. A. D., Lim, J. A. K., et al. (2020). Virtual screening-driven drug discovery of SARS-CoV2 enzyme inhibitors targeting viral attachment, replication, post-translational modification and host immunity evasion infection mechanisms. *J. Biomol. Struct. Dyn.* 1–18. doi: 10.1080/07391102.2020.1776639
- Ramadan, W. H., Al Masri, S., and Rizk, J. (2019). Fixed-dose combination of umecidinium and vilanterol for patients with chronic obstructive pulmonary disease: a systematic review. *Clin. Respir. J.* 13, 663–673. doi: 10.1111/crj.13073
- Rehman, B., Sanchez, D. P., and Shah, S. (2020). “Atenolol,” in *StatPearls* (Treasure Island, FL: StatPearls Publishing). Available online at: <http://www.ncbi.nlm.nih.gov/books/NBK539844/> (accessed July 28, 2020).
- Research CDE (2019). *FDA Approves Darolutamide For Non-Metastatic Castration-Resistant Prostate Cancer*. (FDA). Available online at: <https://www.fda.gov/drugs/resources-information-approved-drugs/fda-approves-darolutamide-non-metastatic-castration-resistant-prostate-cancer> (accessed July 28, 2020).
- Rut, W., Lv, Z., Zmudzinski, M., Patchett, S., Nayak, D., Snipas, S. J., et al. (2020). Activity profiling and structures of inhibitor-bound SARS-CoV-2-PLpro protease provides a framework for anti-COVID-19 drug design. *bioRxiv [Preprint]*. doi: 10.1101/2020.04.29.068890
- Safavi, F., and Nath, A. (2020). Silencing of immune activation with methotrexate in patients with COVID-19. *J. Neurol. Sci.* 415:116942. doi: 10.1016/j.jns.2020.116942
- Santos, M. M. M., and Moreira, R. (2007). Michael acceptors as cysteine protease inhibitors. *Mini Rev. Med. Chem.* 7, 1040–1050. doi: 10.2174/138955707782110105
- Singh, H., Kakkar, A. K., and Chauhan, P. (2020). Repurposing minocycline for COVID-19 management: mechanisms, opportunities, and challenges. *Expert Rev. Anti Infect. Ther.* 18, 997–1003. doi: 10.1080/14787210.2020.1782190
- Singh, J., Deng, Z., Narale, G., and Chuaqui, C. (2006). Structural interaction fingerprints: a new approach to organizing, mining, analyzing, and designing protein-small molecule complexes. *Chem. Biol. Drug Design* 67, 5–12. doi: 10.1111/j.1747-0285.2005.00323.x
- Singh, J., Petter, R. C., Baillie, T. A., and Whitty, A. (2011). The resurgence of covalent drugs. *Nat. Rev. Drug Discov.* 10, 307–317. doi: 10.1038/nrd3410
- Solerte, S. B., Di Sabatino, A., Galli, M., and Fiorina, P. (2020). Dipeptidyl peptidase-4 (DPP4) inhibition in COVID-19. *Acta Diabetol.* 57, 779–783. doi: 10.1007/s00592-020-01539-z
- Song, Y., Zhang, M., Yin, L., Wang, K., Zhou, Y., Zhou, M., et al. (2020). COVID-19 treatment: close to a cure? a rapid review of pharmacotherapies for the novel coronavirus (SARS-CoV-2). *Int. J. Antimicrob. Agents* 56:106080. doi: 10.1016/j.ijantimicag.2020.106080
- Stacy, Z. A., Call, W. B., Hartmann, A. P., Peters, G. L., and Richter, S. K. (2016). Edoxaban: a comprehensive review of the pharmacology and clinical data for the management of atrial fibrillation and venous thromboembolism. *Cardiol. Ther.* 5, 1–18. doi: 10.1007/s40119-016-0058-2
- Stahl, S. M. (2014). Mechanism of action of the SPARI vilazodone: serotonin 1A partial agonist and reuptake inhibitor. *CNS Spectrums* 19, 105–109. doi: 10.1017/S1092852914000169
- Stratford, B. C. (1965). Lymecycline. *Br. Med. J.* 1:922. doi: 10.1136/bmj.1.5439.922
- Sugawara, T., Baumgart, S. J., Nevedomskaya, E., Reichert, K., Steuber, H., Lejeune, P., et al. (2019). Darolutamide is a potent androgen receptor antagonist with strong efficacy in prostate cancer models. *Int. J. Cancer* 145, 1382–1394. doi: 10.1002/ijc.32242
- Tack, J., Pokrotnieks, J., Urbonas, G., Banciu, C., Yakusevich, V., Bunganic, I., et al. (2018). Long-term safety and efficacy of acotiamide in functional dyspepsia (postprandial distress syndrome)-results from the European phase 3 open-label safety trial. *Neurogastroenterol. Motil.* 30:e13284. doi: 10.1111/nmo.13284
- Tarantelli, C., Lange, M., Gaudio, E., Cascione, L., Spriano, F., Kwee, I., et al. (2020). Copanlisib synergizes with conventional and targeted agents including venetoclax in B- and T-cell lymphoma models. *Blood Adv.* 4, 819–829. doi: 10.1182/bloodadvances.2019000844
- Testa, S., Paoletti, O., Giorgi-Pierfranceschi, M., and Pan, A. (2020). Switch from oral anticoagulants to parenteral heparin in SARS-CoV-2 hospitalized patients. *Intern. Emerg. Med.* 15, 751–753. doi: 10.1007/s11739-020-02331-1
- The Top 300 of 2020. Available online at: <https://clincalc.com/DrugStats/Top300Drugs.aspx> (accessed July 28, 2020).
- Tisoncik, J. R., Korth, M. J., Simmons, C. P., Farrar, J., Martin, T. R., and Katze, M. G. (2012). Into the eye of the cytokine storm. *Microbiol. Mol. Biol. Rev.* 76, 16–32. doi: 10.1128/MMBR.05015-11
- Trujillo, T., and Dobesh, P. P. (2014). Clinical use of Rivaroxaban: Pharmacokinetic and pharmacodynamic rationale for dosing regimens in different indications. *Drugs* 74, 1587–1603. doi: 10.1007/s40265-014-0278-5
- Tungol, A., Rademacher, K., and Schafer, J. A. (2011). Formulary management of the protease inhibitors boceprevir and telaprevir for chronic hepatitis C virus. *J. Manag. Care Pharm.* 17, 685–694. doi: 10.18553/jmcp.2011.17.9.685
- Vasanthakumar, N. (2020). Can beta-adrenergic blockers be used in the treatment of COVID-19? *Med. Hypotheses* 142:109809. doi: 10.1016/j.mehy.2020.109809
- Watts, K. S., Dalal, P., Murphy, R. B., Sherman, W., Friesner, R. A., and Shelley, J. C. (2010). ConfGen: a conformational search method for efficient generation of bioactive conformers. *J. Chem. Inf. Model.* 50, 534–546. doi: 10.1021/ci100015j
- Weizman, Z., Forstner, G. G., Gaskin, K. J., Kopelman, H., Wong, S., and Durie, P. R. (1985). Bentiromide test for assessing pancreatic dysfunction using analysis of para-aminobenzoic acid in plasma and urine. Studies in cystic fibrosis and Shwachman's syndrome. *Gastroenterology* 89, 596–604. doi: 10.1016/0016-5085(85)90456-1
- Wen, C.-C., Kuo, Y.-H., Jan, J.-T., Liang, P.-H., Wang, S.-Y., Liu, H.-G., et al. (2007). Specific plant terpenoids and lignoids possess potent antiviral activities against severe acute respiratory syndrome coronavirus. *J. Med. Chem.* 50, 4087–4095. doi: 10.1021/jm070295s
- Willett, P., Barnard, J. M., and Downs, G. M. (1998). Chemical similarity searching. *J. Chem. Inf. Comput. Sci.* 38, 983–996. doi: 10.1021/ci9800211

- Wu, F., Zhao, S., Yu, B., Chen, Y.-M., Wang, W., Song, Z.-G., et al. (2020). A new coronavirus associated with human respiratory disease in China. *Nature* 579, 265–269. doi: 10.1038/s41586-020-2008-3
- Yamamoto, N., Matsuyama, S., Hoshino, T., and Yamamoto, N. (2020). Nelfinavir inhibits replication of severe acute respiratory syndrome coronavirus 2 *in vitro*. *bioRxiv [Preprint]*. doi: 10.1101/2020.04.06.026476
- Yu, C.-H., Chou, C.-C., Tu, H.-F., Huang, W.-C., Ho, Y.-Y., Khoo, K.-H., et al. (2018). Antibody-assisted target identification reveals afatinib, an EGFR covalent inhibitor, down-regulating ribonucleotide reductase. *Oncotarget* 9, 21512–21529. doi: 10.18632/oncotarget.25177
- Zhou, P., Yang, X.-L., Wang, X.-G., Hu, B., Zhang, L., Zhang, W., et al. (2020). A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* 579, 270–273. doi: 10.1038/s41586-020-2012-7
- Zhu, K., Borrelli, K. W., Greenwood, J. R., Day, T., Abel, R., Farid, R. S., et al. (2014). Docking covalent inhibitors: a parameter free approach to pose prediction and scoring. *J. Chem. Inf. Model.* 54, 1932–1940. doi: 10.1021/ci500118s
- Zubair, A. S., McAlpine, L. S., Gardin, T., Farhadian, S., Kuruvilla, D. E., and Spudich, S. (2020). Neuropathogenesis and neurologic manifestations of the coronaviruses in the age of coronavirus disease 2019: a review. *JAMA Neurol.* 77, 1018–1027. doi: 10.1001/jamaneurol.2020.2065

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Delre, Caporuscio, Saviano and Mangiatordi. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Screening of Natural Products Targeting SARS-CoV-2–ACE2 Receptor Interface – A MixMD Based HTVS Pipeline

Krishnasamy Gopinath, Elmeri M. Jokinen, Sami T. Kurkinen and Olli T. Pentikäinen\*

Faculty of Medicine, Integrative Physiology and Pharmacology, Institute of Biomedicine, University of Turku, Turku, Finland

## OPEN ACCESS

### Edited by:

Chandrabose Selvaraj,  
Alagappa University, India

### Reviewed by:

Vidhi Thakkar,  
Independent Researcher, Cordova,  
United States  
Sneha Ray,  
University of Texas Southwestern  
Medical Center, United States

### \*Correspondence:

Olli T. Pentikäinen  
olli.pentikainen@utu.fi

### Specialty section:

This article was submitted to  
Medicinal and Pharmaceutical  
Chemistry,  
a section of the journal  
Frontiers in Chemistry

Received: 31 July 2020

Accepted: 20 October 2020

Published: 19 November 2020

### Citation:

Gopinath K, Jokinen EM, Kurkinen ST  
and Pentikäinen OT (2020) Screening  
of Natural Products Targeting  
SARS-CoV-2–ACE2 Receptor  
Interface – A MixMD Based HTVS  
Pipeline. *Front. Chem.* 8:589769.  
doi: 10.3389/fchem.2020.589769

The COVID-19 pandemic, caused by novel severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), is a severe global health crisis now. SARS-CoV-2 utilizes its Spike protein receptor-binding domain (S-protein) to invade human cell through binding to Angiotensin-Converting Enzyme 2 receptor (ACE2). S-protein is the key target for many therapeutics and vaccines. Potential S-protein–ACE2 fusion inhibitor is expected to block the virus entry into the host cell. In many countries, traditional practices, based on natural products (NPs) have been in use to slow down COVID-19 infection. In this study, a protocol was applied that combines mixed solvent molecular dynamics simulations (MixMD) with high-throughput virtual screening (HTVS) to search NPs to block SARS-CoV-2 entry into the human cell. MixMD simulations were employed to discover the most promising stable binding conformations of drug-like probes in the S-protein–ACE2 interface. Detected stable sites were used for HTVs of 612093 NPs to identify molecules that could interfere with the S-protein–ACE2 interaction. In total, 19 NPs were selected with rescoring model. These top-ranked NP–S-protein complexes were subjected to classical MD simulations for 300 ns (3 replicates of 100 ns) to estimate the stability and affinity of binding. Three compounds, ZINC000002128789, ZINC000002159944 and SN00059335, showed better stability in all MD runs, of which ZINC000002128789 was predicted to have the highest binding affinity, suggesting that it could be effective modulator in RBD–ACE2 interface to prevent SARS-CoV-2 infection. Our results support that NPs may provide tools to fight COVID-19.

**Keywords:** COVID-19, mixed solvent molecular dynamics simulation, natural product, spike protein, ACE2

## INTRODUCTION

The severe acute respiratory syndrome Coronavirus-2 (SARS-CoV-2), has led to a global pandemic, as it had spread rapidly around the world. The entire scientific community is in an urge to find a therapeutic solution to reduce the spread and severity of COVID-19 infections. In many countries, traditional medicines have been in use to fight against COVID-19. The traditional, complementary and alternative medicines have many benefits (World Health Organization, 2020). The traditional Chinese medicine (TCM) can effectively contribute to COVID-19 treatment as an alternative measure. In China, The TCM has been in use along with the conventional Western antiviral medicine for the treatment of COVID-19, many clinical trials are in progress to test the efficacy

and safety in COVID-19 treatment (Lim, 2020; Ling, 2020). Yang et al. (2020) reported that 85% of SARS-CoV-2 infected patients in China are receiving TCM treatment (Yang et al., 2020). A clinical trial is registered in India to study the effect of nutraceutical formulations to fight against SARS-CoV-2. Another clinical trial is proposed to study the effect of natural product-based oral spray with curcumin and artemisinin for the treatment of COVID-19. Food supplements also contribute to a therapeutic solution for COVID-19. Researchers from Australia, Egypt and Saudi Arabia proposed to clinically test the Zinc, Natural Honey and oral nutrition supplements respectively to fight against SARS-CoV-2. Besides that, in the US and Spain, a clinical trial studies the effect of micronutrients and even resistant potato starch in clinical recovery (Koe, 2020a,b).

SARS-CoV-2 has four structural and seven non-structural proteins (Shereen et al., 2020). S-protein is one of the structural proteins and is present on the surface of the coronavirus. S-proteins are highly glycosylated. These S-proteins are the protrusions on the virus, and the spike is clove shaped with three receptor-binding S1 heads attaching to the top of a trimeric membrane fusion S2 stalk, these protrusions on the surface resemble a crown. The S-protein plays a vital role in invading the host cells (Walls et al., 2020). Previously, it has been reported that angiotensin-converting enzyme 2 (ACE2) is essential for SARS-CoV-2 infection, acting as its effective host receptor (Kuba et al., 2005). Both SARS-CoV and SARS-CoV-2 share similar cell entry mechanism by binding with ACE2 located on the surface of the host cell. The ACE2 receptor is widely expressed in lungs, guts, kidney, cardiovascular system, central nervous system and adipose tissue (Mahmoud et al., 2020). The S1 subunit contains a receptor-binding domain (RBD) that recognizes ACE2 (Lin et al., 2020). The RBD of S1 subunit undergoes conformational changes upon recognizing ACE2. The RBD stands up and keeps the S1 domain in an open conformational state to initiate the attachment of the virus to the human ACE2. This conformational change of RBD results in the fusion with the host cell membrane through the S2 subunit of the Spike protein. The S2 subunit helps the virus in conformational changes during the fusion process of the virus after endocytosis, by then the pH levels on the surface of the host cell gets reduced and helps in the intervention of virus. (Laha et al., 2020; Shang et al., 2020). This functional mechanism of RBD of the S-protein provides the framework for the design of the inhibitors to prevent the entry of the virus into a host cell, and thus, curbs further infections in the host.

The S-protein is considered as the potential target for therapeutic intervention and vaccination. The main interest of this study was to search for NPs that could inhibit the SARS-CoV-2 interaction with human cells, and thus, prevent the replication of the virus. NP-databases have numerous bioactive compounds with known antiviral properties. NPs have been practically always used for the treatment of various infections (Gopinath et al., 2020). In the context of SARS-CoV-2, NPs have contributed effectively in the past to treat severe acute respiratory syndrome caused by SARS-CoV and MERS-CoV. Accordingly, in the present alarming situation, NPs would be an obvious resource to identify treatments against SARS-CoV-2 (Antonio et al., 2020). For example, NPs have the

potential to block the recognition site of the HSPA5 cell-surface and compete for the viral spike recognition (Antonio et al., 2020; Elfiky, 2020), two natural products, thioflexibilolide A and candidine, potentially interact with RBD domain of S-protein (Chen et al., 2020), and herb-derived naturally occurring compounds sinigrin, indigo, aloe-emodin, hesperetin, quercetin, epigallocatechin gallate, herbacetin, rhoifolin and pectolinarin have potential to inhibit the SARS 3CLpro activity in SARS-CoV-2 (Chen et al., 2020; Elfiky, 2020; Yang et al., 2020). Also, NPs have been reported to have effect potential in the treatment of SARS-CoV-2 (Ling, 2020; Steele, 2020). Computer-aided drug screening has been performed to find effective molecules for fighting against SARS-CoV-2 that is causing current pandemic (Chen et al., 2020).

Supercomputers have been in use across the world to accelerate the COVID-19 drug search. Recent advancement in computational facilities has increased the efficient usage of various MD simulation techniques in drug search (Amaro et al., 2008, 2018; Salmaso and Moro, 2018). Mixed solvent molecular dynamics simulations (MixMD) is a cosolvent simulation technique for identification of binding hotspots such as orthosteric, allosteric and cofactor binding sites as well as protein-protein interaction (PPI) sites (Ghanakota and Carlson, 2016; Ung et al., 2016). In addition, the conformational changes of these sites can be observed to find suitable protein conformations for docking. Our natural product search pipeline combines MixMD with high-throughput virtual screening (HTVS) in the initial screening. Here, we used MixMD to detect potential inhibitor binding sites on RBD—ACE2 protein-protein interaction (PPI) interface. The identified binding site was used for the HTVS with various natural product databases. Furthermore, rescoring of docking results jointly with MD simulations, and binding energy calculations were carried out to identify potent NPs that could block the S-protein—ACE2 interaction.

## METHODS

### MixMD Simulations

In MixMD simulations, drug-like organic probe molecules are added to the solvent, and their localization during simulations is observed to detect possible small molecule binding sites on the protein surface (Ghanakota and Carlson, 2016; Ung et al., 2016). Crystal structure of SARS-CoV-2 S-protein complexed with ACE2 (PDB: 6M0J) was obtained from Protein Data Bank in 2.45 Å resolution (Lan et al., 2020). ACE2, ions and crystal waters were deleted. Three different probes were used in MixMD simulations: pyrimidine (1P3), acetonitrile (ACN) and isopropanol (IPA). RBD was solvated in 5% v/v ratio of the probe to water with each probe. Tleap in AMBER18 (Case et al., 2020) was used for simulation setup. Hydrogens were added, and disulfide bonds and histidine protonation were assigned. A layer of probe molecules was added after which the system was solvated with enough TIP3P (Jorgensen et al., 1983) water to obtain the correct probe-water ratio. Protein was parameterized with ff14SB force field (Maier et al., 2015). Parameters from GLYCAM\_06j-1 force field were used for N-acetyl-D-glucosamine and glycosylated

asparagine (Kirschner et al., 2008). Parameters validated for MixMD simulations in TIP3P water were used for 1P3, ACN and IPA (Lexa et al., 2014).

CUDA implementation of PMEMD was used in all simulations (Götz et al., 2012; Salomon-Ferrer et al., 2013). NPT simulations were run using 2 fs timestep. Temperature and pressure were maintained at 300 K and 1 atm, respectively, using Andersen thermostat (Andersen, 1980). SHAKE algorithm was used to restrain bonds to hydrogen atoms (Ryckaert et al., 1976). Short-range electrostatics cutoff was set to 10 Å, and Particle Mesh Ewald was used for long-range electrostatics (Darden et al., 1993; Essmann et al., 1995). Energy minimization and gradual heating were conducted for all systems. Protein backbone was restrained during equilibration runs. After gradually decreasing the backbone restraints, 1.4 ns equilibrations were run without restraints. Production simulations were continued until 100 ns and repeated ten times per probe, resulting in 3 μs of total simulation time.

## Probe Pose Clustering

Last 25 ns of each simulation run was included in probe pose clustering. Probe occupancy maps were generated to identify sites on the RBD surface where probe molecules have high residence times. Grid function in cpptraj (Roe and Cheatham, 2013) was used with 0.5 Å spacing to generate the occupancy maps. The maps of different probes were visualized and combined in Pymol 2.3.0 (The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC).  $\sigma$  values of probe densities were adjusted as described by Ghanakota and Carlson (2016). Higher  $\sigma$  value enables visualization of spots with the highest probe residence times while the sites where probes are rapidly exchanged with water get disposed. Based on the occupancy map visualization, the residues that are close to or in contact with the probe densities were identified and used in probe pose clustering.

The most common probe poses at the high occupancy sites were obtained by root-mean-square deviation (RMSD)-based clustering of probes using cpptraj. Probe molecules within <3 Å of residues around the observed probe densities were extracted from all simulation snapshots. Each probe molecule was written in its PDB file, and all the probes were given the same residue and atom numbers to enable clustering in cpptraj. The probe files were loaded to cpptraj, and clustering was performed with the average-linkage method using epsilon value 4 Å. After this, RBD structures related to the obtained centroid probes were searched from the simulation trajectories to evaluate their use in molecular docking.

## Ligand Preparation

The natural product database used in HTVS consists compounds from ZINC Biogenic (<http://zinc.docking.org/substances/subsets/biogenic>), FooDB Version 1.0 (<https://foodb.ca>), Molport Natural Compound and Natural-Like Compound Database ([www.molport.com](http://www.molport.com)) and Super Natural II database (Banerjee et al., 2015). All compounds were downloaded as 2D structural data and were converted to 3D format with OPLS3 charges and tautomeric states at pH 7.4 using LIGPREP in MAESTRO 2020-1 (Schrödinger, LLC, New York, NY,

**TABLE 1 |** The selected compounds from HTVS and their PANTHER/ShaEP-based rescoring.

Database ID	ESP similarity score	Shape similarity score
ZINC000002155511	0.210	0.697
MolPort-002-515-240	0.214	0.692
SN00059335	0.203	0.682
ZINC000002151580	0.203	0.675
ZINC000072325799	0.205	0.666
SN00236224	0.209	0.661
ZINC000002159944	0.219	0.649
MolPort-021-745-932	0.203	0.648
ZINC000002108239	0.222	0.647
ZINC000096296967	0.206	0.638
MolPort-027-852-900	0.203	0.637
ZINC000002108298	0.212	0.635
ZINC000095559555	0.203	0.629
ZINC000002102314	0.209	0.628
SN00341524	0.210	0.621
FDB023015	0.300	0.619
MolPort-027-852-870	0.209	0.618
ZINC000002114285	0.204	0.616
ZINC000002128789	0.217	0.615

Compound ID prefix depicts its source (ZINC0000-ZINC biogenic; FDB- FooDB; SN- Super Natural II database; MolPort- Molport database).

United States, 2020). Molecules that possess more than ten rotatable bonds or molecular weight that exceeds the range of 150–550 g/mol were excluded from the dataset with LIGFILTER in MAESTRO. Molport, ZINC and Super Natural II databases were also filtered from the molecules exceeding partition coefficient (logP) of 5.7 calculated with QIKPROP in MAESTRO. Conversion of molecules to SYBYL MOL2 format was done with MOL2CONVERT in MAESTRO.

## Virtual Screening and Rescoring

Protein conformation from probe pose clustering of MixMD was used for HTVS. HTVS of NPs was performed with PLANTS software (Korb et al., 2009). The docking site was defined as a sphere with 12 Å radius from Arg403 of RBD, Chemp1p scoring function and search speed “speed1” were used in screening. Five best-scored conformations of each compound were kept with cluster RMSD 3.0.

Docking results were rescored with the negative image-based (NIB) rescoring method using the programs PANTHER (Niinivehmas et al., 2015) (version 0.18.19) and ShaEP (Vainio et al., 2009) (version 1.3.1). PANTHER was used to generate a NIB-model that describes the shape and electrostatic properties of an optimal ligand for the binding pocket. Extremely fast molecular similarity screening was then performed with ShaEP with the option “noOptimization” to score the ligand poses generated by the docking program. PANTHER rescoring has resulted in significant enrichment of active compounds in VS with multiple protein targets (Kurkinen et al., 2018, 2019). Furthermore, 1P3 molecules

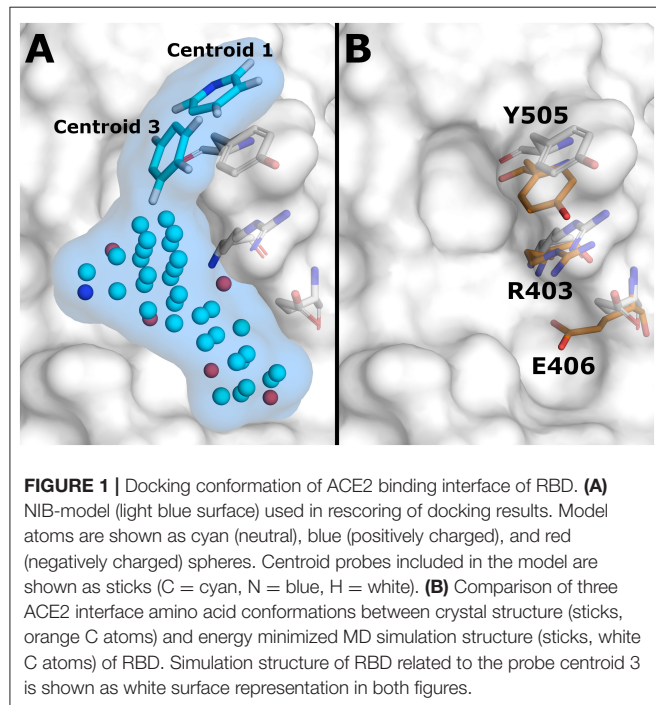
obtained by probe pose clustering were incorporated to the NIB-model to favor compounds forming similar interactions in the high occupancy areas. This fragment-based approach for NIB-screening has been previously utilized in the discovery

of active inhibitor molecules (Jokinen et al., 2019). In NIB-model generation, suitable coordinate points were used to obtain small molecule like entities into the identified cavity. Face-centered cubic packing method was used. In models where 1P3 molecules were included, their partial charges were assigned similarly as for the rest of the model. Those nitrogen atoms of 1P3 that were buried were replaced with neutral carbon atoms.

Top ranking compounds from rescoring were inspected in Maestro (Schrödinger Release 2020-1: Maestro, Schrödinger, LLC, New York, NY, 2020). Top ranking natural compounds with satisfying ESP and Shape similarity score from rescoring calculation were selected for further computational validation.

## Molecular Dynamics (MD) Simulation and Post-MD Analysis

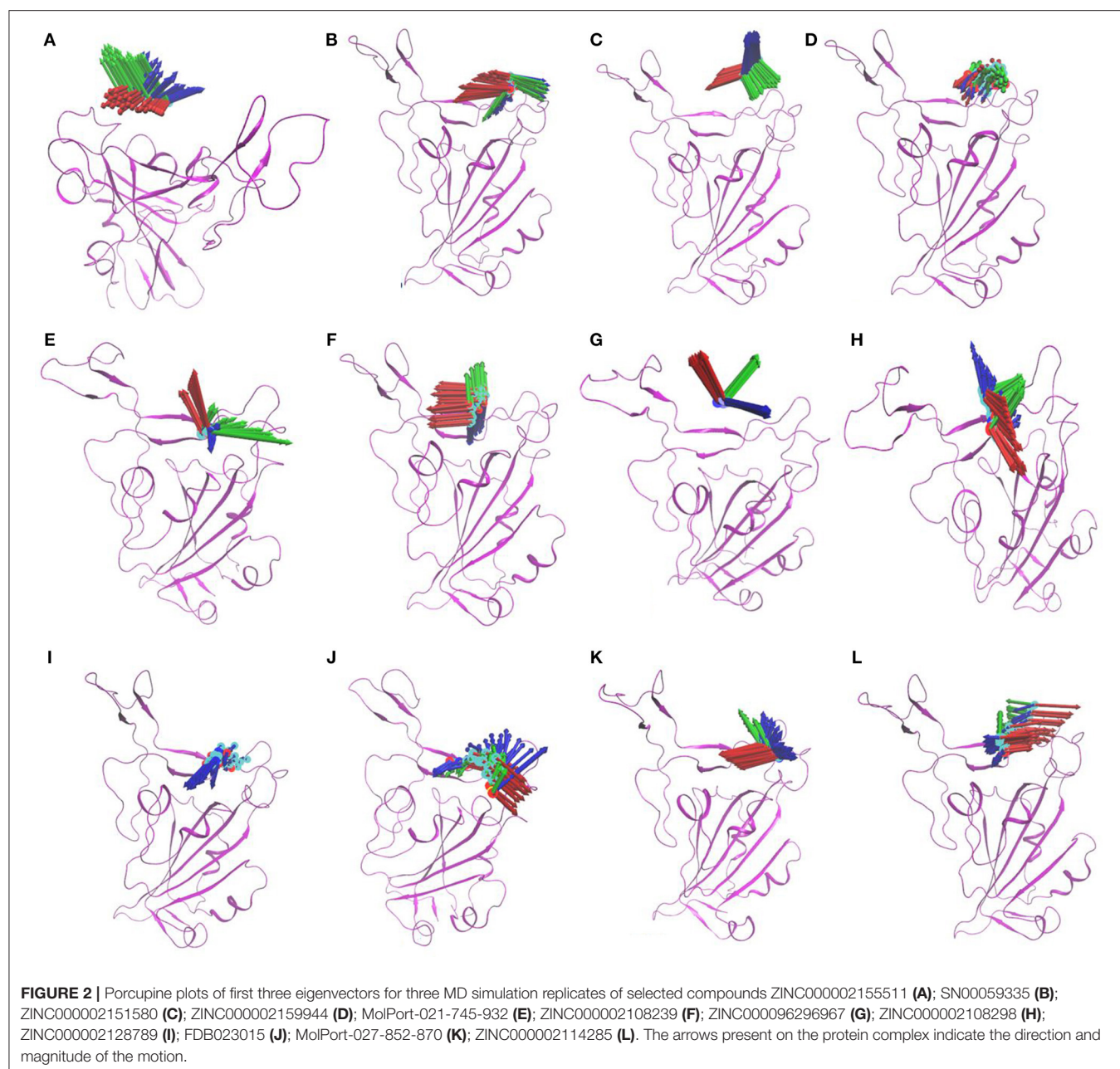
Classical MD simulations were performed with AMBER18 to ensure the binding stability of 19 top-ranked compounds from rescoring (Table 1). Protein-ligand complexes were solvated in a cubic box of TIP3P water extending 10 Å from protein atoms in each dimension. Otherwise, the same simulation protocol and settings were used as in the MixMD simulations. General AMBER force field was used to obtain parameters for the ligands (Wang et al., 2004). Ligand topology files and atomic charges were generated with antechamber, using AM1-BCC charge method (Wang et al., 2006). For each protein-ligand complex, three 100 ns repeats were performed, and post-MD analysis was performed with the snapshots of the trajectory ensembles. Principal component analysis (PCA) was performed using CPPTRAJ on three replicates of MD



**TABLE 2 |** Amino acid contacts of natural compounds in the S-protein–ACE2 interface region.

Database ID	Hydrogen bond	Pi-Pi	Pi-Cation	Salt bridge
ZINC000002155511	Arg403 (1), Gly496 (1), Asn501 (1)	Tyr505 (1)	Arg403 (2)	
MolPort-002-515-240	Gly496 (1)	Tyr505 (1)	Arg403 (1)	
SN00059335		Tyr505 (2)	Arg403 (2)	Arg403 (1)
ZINC000002151580	Gly496 (1), Asn501 (1)			
ZINC000072325799	Gln409 (1), Lys417 (1), Tyr453 (1), Asn501 (1)	Tyr505 (1)	Arg403 (1)	
SN00236224	Arg403 (1), Gln409 (1), Gly496 (1)			
ZINC000002159944	Arg403 (1), Gly496 (1), Tyr453 (1), Asn501 (1)	Tyr505 (2)	Arg403 (1)	
MolPort-021-745-932	Arg403 (1), Gln409 (1), Lys417 (1), Tyr453 (1)	Tyr505 (1)	Arg403 (1)	
ZINC000002108239	Arg403 (1)	Tyr505 (1)	Arg403 (1)	Arg403 (1)
ZINC000096296967	Arg403 (1), Gly496 (1)			
MolPort-027-852-900	Arg403 (1), Gly496 (1)			
ZINC000002108298	Arg403 (1)	Tyr505 (1)		Arg403 (1)
ZINC000095559555	Gly496 (1)			
ZINC000002102314	Arg403 (1), Tyr453 (1), Asn501 (1)	Tyr505 (3)	Arg403 (2)	
SN00341524	Gln409 (1), Lys417 (1), Gly496 (1), Tyr505 (1)	Tyr505 (1)		
FDB023015	Arg403 (1), Glu409 (1)			
MolPort-027-852-870	Gly496 (1), Asn 501 (1)	Tyr505 (1)	Arg403 (1)	
ZINC000002114285	Arg403 (1), Tyr453 (1), Asn501 (1)	Tyr505 (2)	Arg403 (2)	
ZINC000002128789	Arg403 (1)	Tyr505 (2)	Arg403 (2)	

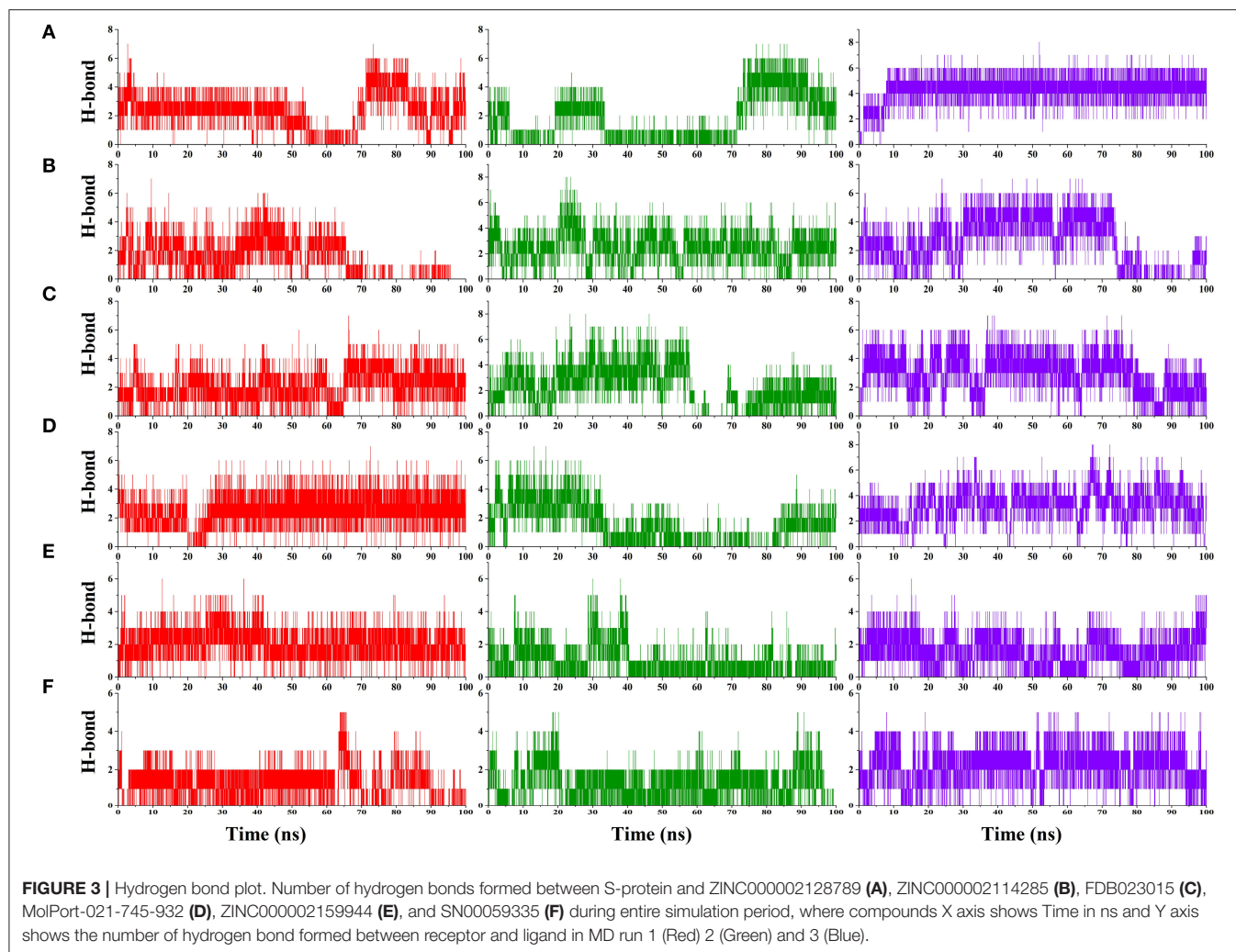
Values in the parenthesis represent the number of contacts with each residue.



simulation of each complex (Galindo-Murillo et al., 2014). The coordinate covariance matrix of all-atom was calculated for the raw trajectory of all three simulations for each complex, and first three eigenvectors were obtained using the matrix, and principal component data was visualized with the Normal Mode Wizard (NMWiz) plugin of VMD. Variation in the ligand location and its motion was examined using the porcupine plot of the first three eigenvectors of the entire simulation of each complex. Further, the stability of compound binding was also checked by hydrogen bond analysis, and binding affinity calculation were performed using MMPBSA.py (Miller et al., 2012).

## Protein-Protein Interaction Analysis

Protein-protein Interaction analysis was performed to examine the effect of the selected compound in the ACE2-Spike protein interface region. Ligand bound RBD structure was obtained from the final frame of three replicates. Ligand bound structures were superimposed with the crystal structure of ACE2-RBD using atom-pairs method and ACE2-ligand-bound RBD complex was obtained by replacing the RBD with ligand-bound RBD. The structures were further prepared, and atom clashes were removed using the protein preparation wizard. The entire preparation was done in Maestro and Optimized Potentials for Liquid Simulations (OPLS)-2005 force field was



used for minimization (Schrödinger Release 2020-3: Maestro, Schrödinger, LLC, New York, NY, 2020). Molecular dynamics simulations were performed for ACE2 with RBD domain, and ACE2 with ligand-bound RBD. Each complex was minimized and solvated with TIP3P water model in an orthorhombic box with a distance of 10 Å. The whole system was neutralized by adding Na<sup>+</sup> ions. The system was prepared using OPLS-2005 force field, and subjected to MD simulation using Desmond (Desmond Molecular Dynamics System, D. E. Shaw Research, New York, NY, 2020 Maestro-Desmond Interoperability Tools, Schrödinger, New York, NY, 2020) under the NPT ensemble with the temperature of 300 K and pressure of 1.01325 bar followed by relaxation (Ponder and Case, 2003; Bowers et al., 2006). 100 ns MD simulations were carried out for each protein-ligand bound complex. Finally, the snapshots of the complex from the trajectory were extracted using “trj2mae.py” script. Protein-Protein interaction fingerprint analysis was performed with BioLuminate by using a distance cutoff of 4 Å (Schrödinger Release 2020-3: BioLuminate, Schrödinger, LLC, New York, NY, 2020).

## RESULTS AND DISCUSSION

### Detection of the Binding Site in the S-Protein—ACE2 Interface

To identify the suitable druggable site with binding hot spots from the S-protein—ACE2 interface, MixMD simulations with different probes were used. The identified stable probe poses were used as seeds in the HTVS. Probe pose clustering analysis was focused on 1P3 as it showed the highest occupation at the ACE2 binding interface of RBD (Jokinen et al., submitted). Two 1P3 poses were identified that overlapped with the region known to be occupied by Lys353 of ACE2 (centroids 1 and 3), suggesting a possible mechanism for small molecule inhibition of ACE2 binding. Centroid 1 was the second and centroid 3 the fourth most common 1P3 pose at the ACE2 interface, and both were used in the HTVS rescoring model (Figure 1A). Both centroid probes showed a favorable binding mode by forming  $\pi$ -stacking interaction with Tyr505 in the area predicted to be crucial for inhibition. RBD structure related to centroid 3 was chosen for molecular docking as it had a groove-like shape at the PPI interface that could accommodate binding

**TABLE 3** | Calculated average hydrogen bond per frame for natural compounds in three MD runs.

Compounds	MD run 1	MD run 2	MD run 3
ZINC000002128789	2.39 ± 1.32	1.74 ± 1.70	4.36 ± 1.10
ZINC000002114285	1.39 ± 1.32	2.53 ± 1.25	2.57 ± 1.76
FDB023015	2.09 ± 1.10	2.27 ± 1.60	2.97 ± 1.39
MolPort-021-745-932	2.51 ± 1.10	1.48 ± 1.37	3.22 ± 1.31
ZINC000002159944	2.12 ± 0.88	1.09 ± 0.89	1.63 ± 1.01
SN00059335	1.25 ± 0.90	1.20 ± 0.90	2.23 ± 1.01
ZINC000002155511	2.00 ± 1.0	0.59 ± 0.78	2.46 ± 1.22
ZINC000002108239	2.72 ± 0.95	0.08 ± 0.32	2.68 ± 1.10
ZINC000002102314	2.48 ± 1.36	1.41 ± 1.03	0.46 ± 0.77
ZINC000002151580	1.52 ± 0.79	2.19 ± 2.00	0.78 ± 0.93
ZINC000002108298	0.78 ± 0.93	0.99 ± 0.96	2.23 ± 1.01
MolPort-027-852-870	1.27 ± 0.92	0.72 ± 0.67	1.44 ± 1.14
MolPort-002-515-240	1.57 ± 1.57	0.88 ± 0.99	0.13 ± 0.58
MolPort-027-852-900	0.77 ± 0.8	0.55 ± 0.80	0.10 ± 0.46
SN00236224	0.79 ± 0.85	0.29 ± 0.75	1.52 ± 1.19
SN00341524	2.51 ± 1.08	0.01 ± 0.80	1.75 ± 1.28
ZINC000072325799	2.78 ± 1.20	0.43 ± 0.94	0.26 ± 0.71
ZINC000095559555	0.68 ± 0.92	0.80 ± 0.97	0.35 ± 0.66
ZINC000096296967	0.77 ± 0.86	0.18 ± 0.51	1.44 ± 0.96

Hydrogen bond counts values are shown with standard deviation.

small molecules. Comparison with the crystal structure (PDB: 6M0J) showed that the residues Tyr505, Arg403, and Glu406 adopted conformations that increased the depth of this groove and expanded the space available for a possible small molecule inhibitor (**Figure 1B**).

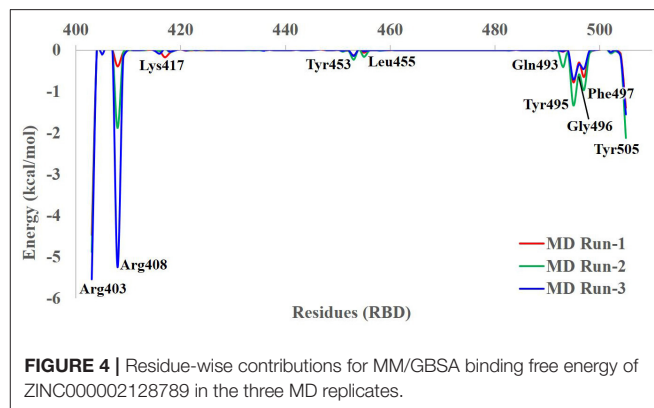
## Screening of Potential Natural Products

Here, the goal was to identify NPs that would bind into S-protein and interfere with SARS-CoV-2 attachment to the host cell. The screened NP-library library consisted of ZINC biogenic (206,800 compounds) FooDB (18,477 compounds), Molport Natural Compound and Natural-Like Compound Database (119,054 compounds), and Super Natural II database (267,762). HTVS-docking was performed with this NP-database against the detected binding site in S-protein. Docking results were filtered by using rescoring with PANTHER/ShaEP-based NIB rescoring where both the shape and electrostatic potential (ESP) were compared between cavity-based NIB model and the docked molecule (Vainio et al., 2009; Niinivehmas et al., 2015; Kurkinen et al., 2018, 2019). Identified stable 1P3 poses were used in the NIB-models as they show the area where drug-like compounds could bind.

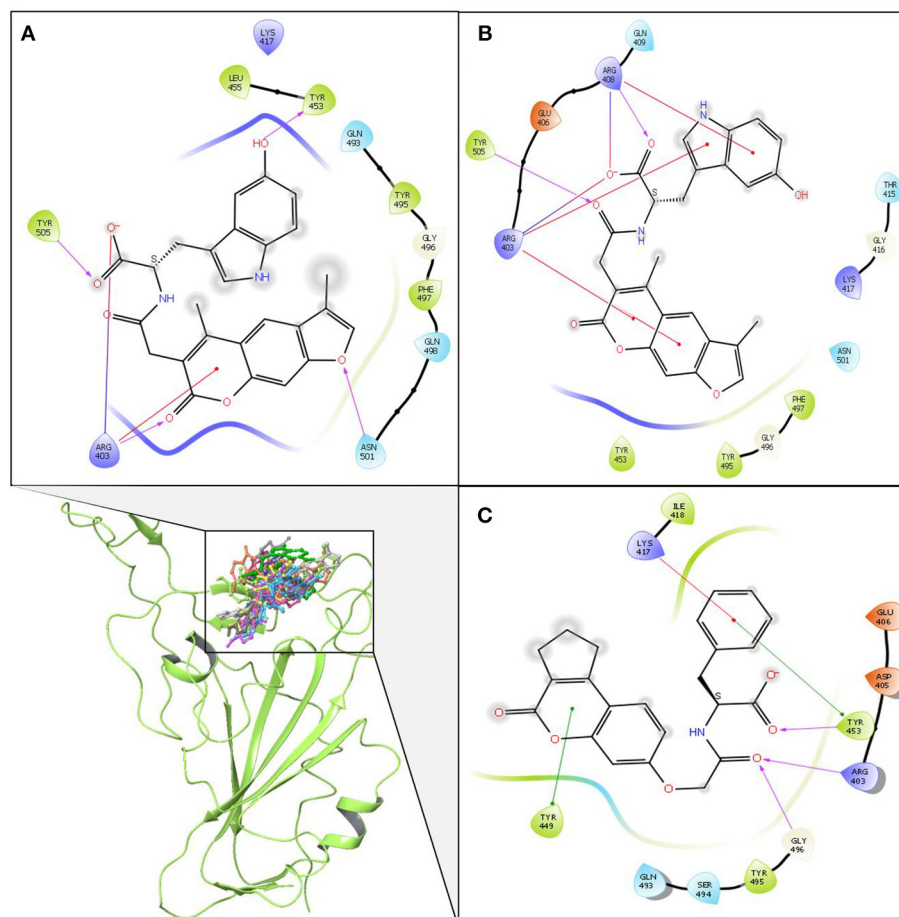
Interaction analysis was performed for the filtered compounds having ESP similarity score > 0.2 and Shape similarity score > 0.6, in NIB-model based rescoring, by using Ligand Interaction Diagram Panel in Maestro (Schrödinger Release 2020-2: Maestro, Schrödinger, LLC, New York, NY, 2020). Further, compounds interacting with residues in the interface region and that occupy the cavity region next to Tyr505 and Arg403 were selected (**Table 1**). Recent studies

**TABLE 4** | Hydrogen-bonding residues obtained by hydrogen bond-lifetime analysis.

Compounds	Hydrogen-bonding residues in RBD
ZINC000002128789	Arg403, Arg408, Gln409, Lys417, Tyr453, Tyr473, Tyr489, Gln493, Tyr495, Gly496, Gln498, Asn501, Gly502, Tyr505
ZINC000002114285	Arg346, Ser349, Arg403, Arg408, Lys417, Tyr421, Lys444, Gly446, Tyr449, Tyr453, Arg466, Thr470, Tyr473, Tyr489, Gln493, Ser494, Gly496, Gln498, Asn501, Tyr505
FDB023015	Ser375, Thr376, Lys378, Arg403, Arg408, Gln414, Lys417, Asn437, Tyr449, Tyr453, Tyr489, Gln493, Ser494, Gly496, Gln498, Thr500, Asn501, Gly502, Val503, Gly504, Tyr505, Tyr508
MolPort-021-745-932	Ser375, Thr376, Lys378, Arg403, Arg408, Gln414, Lys417, Asn437, Tyr449, Tyr453, Tyr489, Gln493, Ser494, Gly496, Gln498, Thr500, Asn501, Gly502, Val503, Gly504, Tyr505, Tyr508
ZINC000002159944	Arg403, Gly502, Asn501, Gly496, Tyr495, Gln498, Gln493, Unk198, Tyr505, Lys417, Tyr453, Tyr449, Ser494, Arg408, Thr500
SN00059335	Arg403, Arg408, Lys417, Tyr421, Lys444, Gly446, Tyr449, Tyr453, Lys458, Tyr473, Tyr489, Gln493, Gly496, Gln498, Thr500, Asn501, Gly502, Val503, Gly504, Tyr505

**FIGURE 4** | Residue-wise contributions for MM/GBSA binding free energy of ZINC000002128789 in the three MD replicates.

imply that the following residues Arg403, Asp405, Lys417, Asn439, Val445, Gly446, Tyr449, Tyr453, Lys455, Phe456, Tyr473, Ala475, Gly476, Glu484, Asn487, Tyr489, Gln493, Gln498, Gln493, Gln498, Thr500, Asn501, Gly502, Val503, Tyr505 mediates the fusion of SARS-CoV-2 with the cellular membrane through RBD-ACE2 interface (Brielle et al., 2020; Lan et al., 2020; Sun et al., 2020). Herein, selected NPs interacts with the surrounding amino acids by hydrogen bonds,  $\pi$ -stacking interactions, and salt bridges (**Table 2**). Further computational validation was extended to check the compound's affinity and stability from classical MD simulation and Principal component analysis and hydrogen bond analysis.

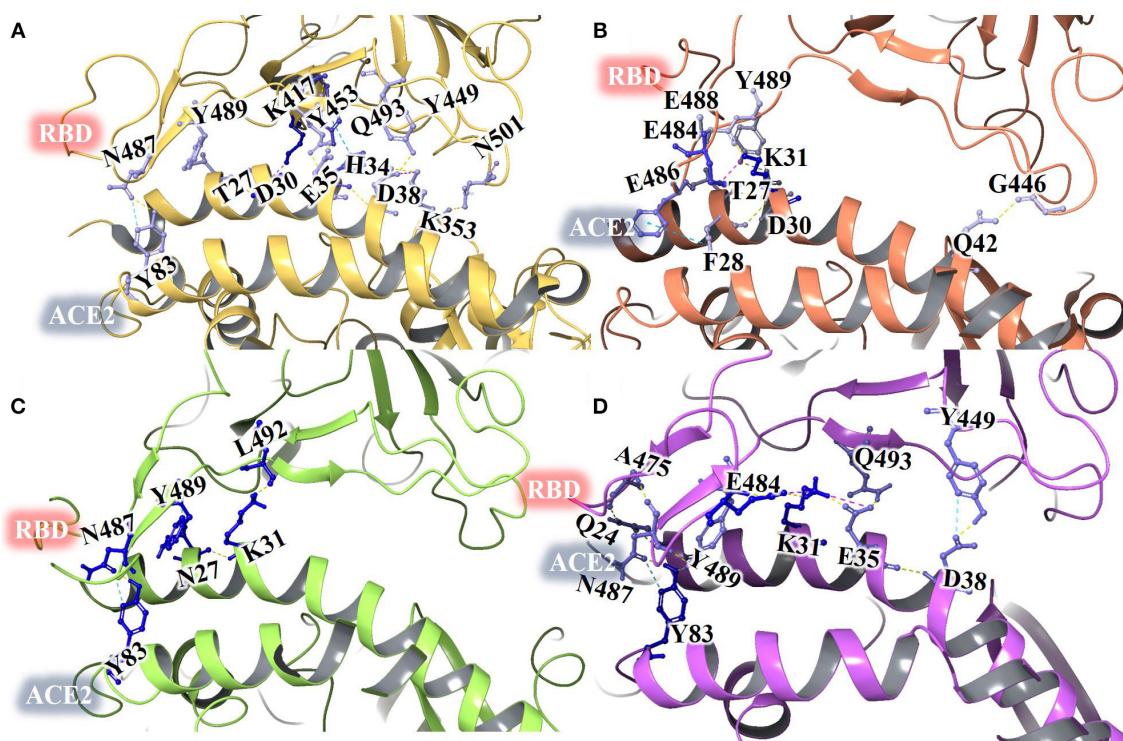


**FIGURE 5 |** Schematic representation of detailed atom interactions of ZINC000002128789 at 100 ns of MD replicate 1 **(A)** 2 **(B)** and 3 **(C)**. Structure were obtained from all three replicates at 25 ns interval and shown as compounds cluster in ribbon model. Ligand interaction diagram is obtained for snapshot obtained from final frame. Pink, red oval and green oval arrows represent Hydrogen bond,  $\pi$ -Cation and  $\pi$ - $\pi$  stacking interaction. Interacting amino acids are differentiated with various spheres. Charged, hydrophobic, polar residues are shown in blue, green and cyan while Glycine is represented with ivory spheres, respectively.

## Affinity and Stability of Selected Natural Products

MD simulations were performed for selected NP—protein complexes for 300 ns (3 replicates of 100 ns) using Amber18, and trajectories of all replicates were used for PCA analysis. PCA systematically reduces the dimensionality of a complex system, and can characterize the cumulative and overall motion of the protein-ligand system (Bhutani et al., 2015). PCA is used to check the dominant modes of motion in a trajectory and variance in the data (Haider et al., 2008). First three PCs for each complex was obtained by diagonalizing the coordinate covariance matrix. Further, the analysis was focused on ensuring the stability using obtained PCs. A porcupine plot was drawn with first three eigenvector of each complex, and differences in the ligand location on generated averaged coordinates and motion was examined in VMD. The arrows in the porcupine plot represent the ligand direction and magnitude

of the motion in three PC modes. Among all compounds, ZINC000002159944 and ZINC000002128789 had clearly lower magnitude of motion, and they remained located on the binding site in the generated averaged coordinates. The compounds SN00059335, ZINC000002108239, and FDB023015 were found to be in the binding site of the protein with less motion. Another five compounds, ZINC000002151580, MolPort-021-745-932, ZINC000002108298, MolPort-027-852-870, and ZINC000002114285, were located in the binding pocket with higher magnitude of the motion in all modes. Compounds ZINC000096296967 and ZINC000002155511 (**Figures 2A–L**) are slightly off from the binding pocket in the generated averaged coordinates, and they show a higher magnitude of the motion in all modes. In contrast, ZINC000096296967, SN00341524, ZINC000095559555, MolPort-002-515-240, ZINC000072325799, MolPort-027-852-900, and SN00236224 show higher deviation, and they moved out of binding site in the



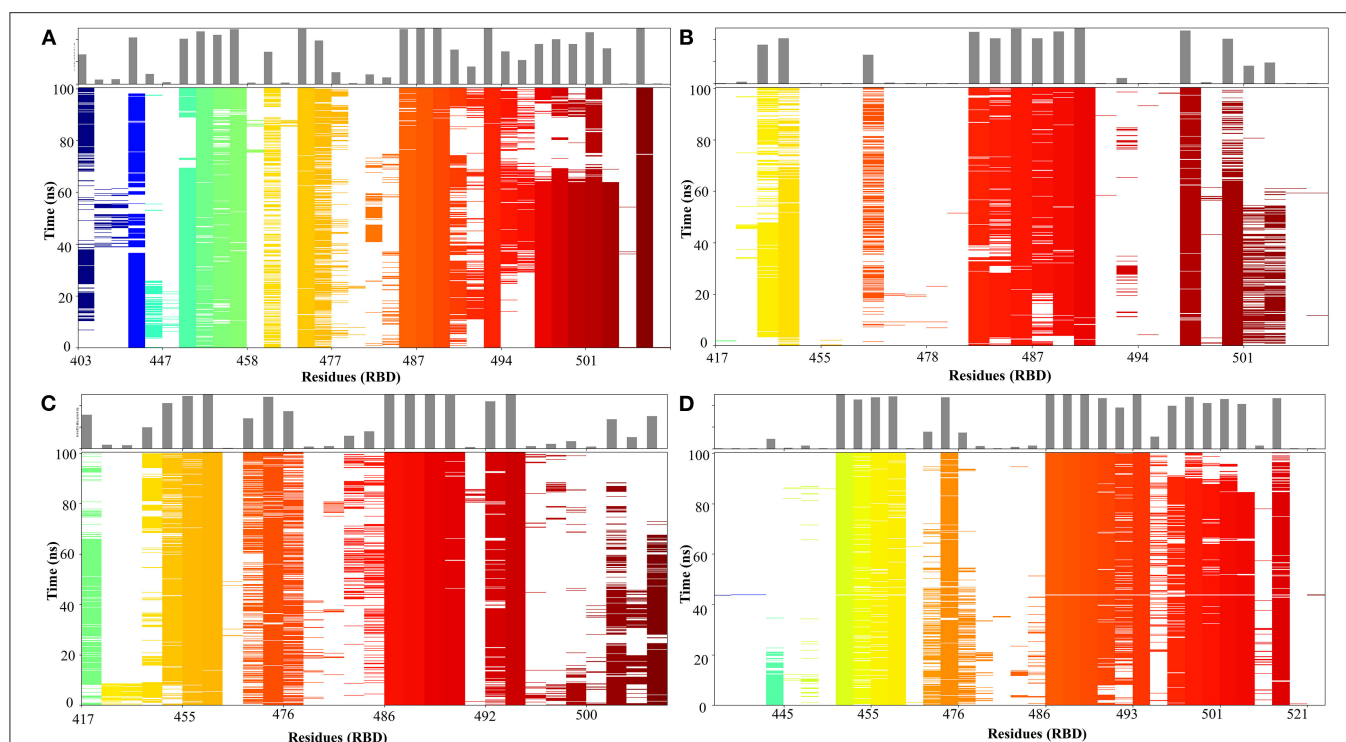
**FIGURE 6 |** Protein interface interactions of crystal structure of ACE2-RBD of SARS-CoV-2 spike protein (A) and ACE2-ZINC000002128789 bound RBD structure obtained from MD run 1 (B) run 2 (C) and run 3 (D).

generated averaged coordinates. Thus, these seven compounds do not likely form strong enough interactions with S-protein to remain bound (**Supplementary Figure 1**).

Hydrogen bond analysis was carried out to understand the stability of the compound binding to RBD. Time-dependent behavior of the hydrogen bonds was monitored, and the number of hydrogen bonds per frame was plotted. Compounds ZINC000002128789, ZINC000002114285, FDB023015, MolPort-021-745-932, ZINC000002159944, and SN00059335 (**Figures 3A–F**) showed sustained hydrogen bonds with RBD domain, and they all had more than 1 average hydrogen bond per frame in all three MD runs (**Table 3**). The porcupine plot also shows that these compounds were in the binding interface of the RBD domain. Remaining compounds, MolPort-002-515-240, MolPort-027-852-870, MolPort-027-852-900, SN00236224, SN00341524, ZINC000002102314, ZINC000002108239, ZINC000002108298, ZINC000002151580, ZINC000002155511, ZINC0000027325799, ZINC000095559555, and ZINC000096296967 (**Supplementary Figures 2, 3**) do not show sustained hydrogen bond connections, and the average number of hydrogen bonds per frame is more than 1 in only one or two MD runs. Hydrogen bond-lifetime analysis was focused for top six stably binding compounds. This analysis depicts that during the simulations ZINC000002128789, ZINC000002159944 and SN00059335 form hydrogen bonds with ACE2-Spike protein binding interface residues, whereas, ZINC000002114285, FDB023015, and MolPort-021-745-932 also form hydrogen

bonds with residues that are located outside of the binding site (**Table 4**).

The Porcupine plot and hydrogen bond analysis of the simulations confirms that ZINC000002128789, ZINC000002159944, and SN00059335 remain stably bound in RBD interface, and they maintain contacts with ACE2-Spike protein binding interface residues. Furthermore, MM/GBSA calculations were performed with an ensemble of all three replicates to estimate the binding affinity of these compounds. The snapshots were recorded at 50 ps intervals, and these snapshots were used for ensemble-average MM/GBSA calculation to estimate the binding affinity of selected NPs. Based on MM/GBSA, the ensemble-average total binding affinities ( $\Delta G_{\text{bind}}$ ) of ZINC000002128789, ZINC000002159944 and SN00059335 are  $-20.82 \pm 5.10$  kcal/mol,  $-13.88 \pm 6.87$  kcal/mol and,  $-11.60 \pm 10.77$  kcal/mol, respectively (**Supplementary Table 1**). Mean  $\Delta G_{\text{bind}}$  of all three replicates shows that compound ZINC000002128789 has highest binding affinity ( $\Delta G_{\text{bind}} = -20.82 \pm 5.10$  kcal/mol) among three compounds and residues Arg403, Arg408, and Tyr505 largely contribute to the binding free energy of ZINC000002128789 (**Figure 4**). ZINC000002128789 maintains sustained contacts with Arg403, Arg408, and Tyr505 in all replicates of ZINC000002128789-RBD simulation (**Figure 5**). Arg403 play a role in stabilizing the interface through water mediated indirection with Asn33/His34/Glu37/Asp38 of ACE2. Arg408 and Tyr505 have



**FIGURE 7 |** Time-dependent Protein-protein interaction matrix of crystal structure of ACE2-RBD of SARS-CoV-2 spike protein **(A)** and ACE2- ZINC000002128789 bound RBD structure obtained from MD run 1 **(B)** 2 **(C)** and 3 **(D)**. The colored plot shows the presence of interactions as a function of the time and residue number and the top plot shows interactions count for each residue in the RBD.

been reported to interact with human ACE2 and contributes to higher affinities (Ali and Vijayan, 2020; Mittal et al., 2020).

Many natural products have shown to inhibit coronavirus with the unknown mechanism of action (Xiu et al., 2020). Few computational studies have also identified promising natural compounds to block viral entry by targeting spike protein. Wahedi et al. reported that resveratrol ( $\Delta G_{\text{bind}}$  MM-GBSA =  $-23.88$  kcal/mol) can be promising anti-COVID-19 drug candidates acting through disruption of the spike protein among other stilbenoid analogs (Piceatannol, Pinosylvlin, Pterostilbene, and Chloroquine) (Wahedi et al., 2020). Chen et al. screened Thioflexibilolide A (Binding Energy:  $-9.2$  kcal/mol) and Candidine (Binding Energy:  $-9.0$  kcal/mol) as best compounds from 2000 natural compounds (Chen et al., 2020). Kar et al. performed docking study with natural compounds from *Clerodendrum* spp., and reported that Taraxerol ( $\Delta G_{\text{bind}}$  prime MM-GBSA =  $-45.19$  kcal/mol) as most promising inhibitory candidate against the SARS-CoV-2 spike protein (Kar et al., 2020). All the above reported compounds have shown to interact with amino acids in the ACE2-RBD interface region. In our study, selected compound ZINC000002128789 also shown stable interaction with interface residues which was confirmed by three MD simulation replicates. ZINC000002128789 and top ranked compounds from phytochemical database shares similar interactions with residues

in RBD domain (Jani et al., 2020). In our study, the selected compound ZINC000002128789 maintain stable contact with Arg403, Arg408 and Tyr 505 in all three MD runs. Calculated ensemble-average  $\Delta G_{\text{bind}}$  of two replicates (MD run 2 =  $-22.89 \pm 10.93$  kcal/mol; MD run 3 =  $-24.57 \pm 22.96$  kcal/mol) are higher than  $\Delta G_{\text{bind}}$  of reported natural product resveratrol (Supplementary Table 1).

### ZINC000002128789 on ACE2-RBD Interface

A 100 ns MD simulation was performed with crystal structure of ACE2-RBD and ACE2- ZINC000002128789 bound RBD. Three different ZINC000002128789 conformers were obtained from the final snapshot of MD replicates, and the crystal structure (PDB: 6M0J) coordinates were used to get similar RBD bound ACE2 with docked ZINC000002128789 in interface region. A sum of 500 snapshots were obtained at 200 ps interval and structures were used to analyze the interactions at the ACE2-RBD interface region using BioLuminate. Total number of contacts in ACE2-RBD interface region is decreased in the ZINC000002128789 bound RBD complexes (Figure 6). Further, stability of contacts was checked. Interaction fingerprints were generated for any contacts that are possibly formed between interface residues, and interaction matrix was obtained to examine the sustained interaction between ACE2-RBD interface regions (Figure 7). Ali and Vijayan reported that Lys417, Tyr449, Phe456, Tyr473, Ala475, Phe486, Tyr489, Gln493, and Gln498 interact with ACE2

through stable hydrogen bonds, hydrophobic interactions and salt bridges (Ali and Vijayan, 2020). Herein, Tyr453, Phe456, Ala475, Phe486, Asn487, Tyr489, Gln493, Asn501, Tyr 505 in RBD maintains stable contacts (> 90% of simulation time) with ACE2. While ZINC000002128789 bound complex shows only two (Phe486, Tyr489) stable contacts with ACE2 in three MD runs. A large number of stable contacts could be associated with a higher binding affinity of SARS-CoV-2 (Ali and Vijayan, 2020). ZINC000002128789 has stable interactions with Arg403, Arg408, and Tyr505, and contact matrix shows that these residues have not maintained contacts with ACE2 residues. This suggests that ZINC000002128789 blocks the major contacts between ACE2 and RBD, and thus, might be able to interfere with ACE2-RBD complex formation.

Overall, Compounds ZINC000002128789, ZINC000002159944, and SN00059335 are stable in RBD interface and ZINC000002128789 top ranked with the highest binding affinity prediction. These compounds maintain stable contacts with residues that mediate the entry of SARS-CoV-2 through RBD-ACE2 interface. Accordingly, these compounds could be potent S-protein—ACE2 interaction modulators, and based on our results, the effect of at least these three compounds should be experimentally tested.

## CONCLUSION

Over the years, NPs have shown a remarkable effect in the treatment of SARS-CoV-1, MERS-Cov, HIV, Influenza, Dengue and other viruses. NPs are considered as a safe and effective source in the treatment of SARS-CoV-2 and its related symptoms. Receptor binding domain in the S-protein mediates the fusion of SARS-CoV-2 with the cellular membrane through RBD-ACE2 interface. S-protein is a potential target for preventing SARS-CoV-2 entry into the human cell. In our study, we screened NPs targeting S-protein to block SARS-CoV-2 entry.

## REFERENCES

- Ali, A., and Vijayan, R. (2020). Dynamics of the ACE2-SARS-CoV-2/SARS-CoV spike protein interface reveal unique mechanisms. *Sci. Rep.* 10:14214. doi: 10.1038/s41598-020-71188-3
- Amaro, R. E., Baron, R., and McCammon, J. A. (2008). An improved relaxed complex scheme for receptor flexibility in computer-aided drug design. *J. Comput. Aided. Mol. Des.* 22, 693–705. doi: 10.1007/s10822-007-9159-2
- Amaro, R. E., Baudry, J., Chodera, J., Demir, Ö., McCammon, J. A., Miao, Y., et al. (2018). Ensemble docking in drug discovery. *Biophys. J.* 114, 2271–2278. doi: 10.1016/j.bpj.2018.02.038
- Andersen, H. C. (1980). Molecular dynamics simulations at constant pressure and/or temperature. *J. Chem. Phys.* 72, 2384–2393. doi: 10.1063/1.439486
- Antonio, A. D. S., Wiedemann, L. S. M., and Veiga-Junior, V. F. (2020). Natural products' role against COVID-19. *RSC Adv.* 10, 23379–23393. doi: 10.1039/D0RA03774E
- Banerjee, P., Erehman, J., Gohlke, B. O., Wilhelm, T., Preissner, R., and Dunkel, M. (2015). Super natural II-a database of natural products. *Nucleic Acids Res.* 43, D935–D939. doi: 10.1093/nar/gku886
- Bhutani, I., Loharch, S., Gupta, P., Madathil, R., and Parkesh, R. (2015). Structure, dynamics, and interaction of *Mycobacterium tuberculosis* (Mtb) DprE1 and DprE2 examined by molecular modeling, simulation, and electrostatic studies. *PLoS ONE* 10:e0119771. doi: 10.1371/journal.pone.0119771
- Bowers, K. J., Sacerdoti, F. D., Salmon, J. K., Shan, Y., Shaw, D. E., Chow, E., et al. (2006). "Scalable algorithms for molecular dynamics simulations on commodity clusters," in *Proceedings of the 2006 ACM/IEEE Conference on Supercomputing - SC '06* (New York, NY: ACM Press), 84. doi: 10.1145/1188455.1188544
- Brielle, E. S., Schneidman-Duhovny, D., and Linial, M. (2020). The SARS-CoV-2 exerts a distinctive strategy for interacting with the ACE2 human receptor. *Viruses* 12:497. doi: 10.3390/v12050497
- Case, D. A., Belfon, K., Ben-Shalom, I. Y., Brozell, S. R., Cerutti, D. S., III, T. E. C., et al. (2020). *AMBER 2020*. San Francisco, CA: University of California, San Francisco. Available online at: <https://ambermd.org/>
- Chen, G.-Y., Yao, T.-Y., Ahmed, A., Pan, Y.-C., Yang, J.-C., and Wu, Y.-C. (2020). The discovery of potential natural products for targeting SARS-CoV-2 spike protein by virtual screening. *bioRxiv [Preprint]*. 1–42. doi: 10.1101/2020.06.25.170639
- Darden, T., York, D., and Pedersen, L. (1993). Particle mesh ewald: an N-log(N) method for ewald sums in large systems. *J. Chem. Phys.* 98, 10089–10092. doi: 10.1063/1.464397
- Elfiky, A. A. (2020). Natural products may interfere with SARS-CoV-2 attachment to the host cell. *J. Biomol. Struct. Dyn.* doi: 10.1080/07391102.2020.1761881. [Epub ahead of print].

We used a protocol that combines MixMD simulation with HTVS. Selected compounds from PANTHER/ShaEP based NIB rescoring were subjected to classical MD simulations to verify the stability and affinity of binding. This protocol suggests that ZINC000002128789, ZINC000002159944, and SN00059335 would bind to RBD, and especially, ZINC000002128789 is predicted to be a potent NP to hinder the entry of SARS-CoV-2 by blocking the S-protein—ACE2 interaction.

## DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

## AUTHOR CONTRIBUTIONS

OP, KG, and EJ designed the study. KG performed all calculations, except MixMD simulations and their analysis, which were performed by EJ. Preparation of NP-library was performed by SK. All authors contributed to the manuscript and approved the submitted version.

## ACKNOWLEDGMENTS

We acknowledge the CSC-IT Center for Science (Project 2002823) for offering prioritized computational resources for this project. This study was financially supported by the Academy of Finland (OP; Grant No. 323413), and by The Instrumentarium Science Foundation (EJ).

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fchem.2020.589769/full#supplementary-material>

- Essmann, U., Perera, L., Berkowitz, M. L., Darden, T., Lee, H., and Pedersen, L. G. (1995). A smooth particle mesh ewald method. *J. Chem. Phys.* 103, 8577–8593. doi: 10.1063/1.470117
- Galindo-Murillo, R., Roe, D. R., and Cheatham, T. E. (2014). On the absence of intrahelical DNA dynamics on the  $\mu$ s to ms timescale. *Nat. Commun.* 5:5152. doi: 10.1038/ncomms6152
- Ghanakota, P., and Carlson, H. A. (2016). Moving beyond active-site detection: MixMD applied to allosteric systems. *J. Phys. Chem. B* 120, 8685–8695. doi: 10.1021/acs.jpcc.6b03515
- Gopinath, K., Subbiah, N., and Karthikeyan, M. (2020). Isolation of natural compounds from *syzygium densiflorum* fruits and exploring its chemical property, therapeutic role in diabetic management. *Nat. Prod. J.* 10, 168–176. doi: 10.2174/2210315508666180622113414
- Götz, A. W., Williamson, M. J., Xu, D., Poole, D., Le Grand, S., and Walker, R. C. (2012). Routine microsecond molecular dynamics simulations with AMBER on GPUs. 1. generalized born. *J. Chem. Theory Comput.* 8, 1542–1555. doi: 10.1021/ct200909j
- Haider, S., Parkinson, G. N., and Neidle, S. (2008). Molecular dynamics and principal components analysis of human telomeric quadruplex multimers. *Biophys. J.* 95, 296–311. doi: 10.1529/biophysj.107.120501
- Jani, V., Koulgi, S., Uppuladinne, V. N., M., Sonavane, U., and Joshi, R. (2020). Computational drug repurposing studies on the ACE2-spike (RBD) interface of SARS-CoV-2. *ChemRxiv [Preprint]*. 1–52. doi: 10.26434/chemrxiv.12652832
- Jokinen, E. M., Postila, P. A., Ahinko, M., Niinivehmas, S., and Pentikäinen, O. T. (2019). Fragment- and negative image-based screening of phosphodiesterase 10A inhibitors. *Chem. Biol. Drug Des.* 94, 1799–1812. doi: 10.1111/cbdd.13584
- Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. (1983). Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* 79, 926–935. doi: 10.1063/1.445869
- Kar, P., Sharma, N. R., Singh, B., Sen, A., and Roy, A. (2020). Natural compounds from *Clerodendrum* spp. as possible therapeutic candidates against SARS-CoV-2: an *in silico* investigation. *J. Biomol. Struct. Dyn.* doi: 10.1080/07391102.2020.1780947. [Epub ahead of print].
- Kirschner, K. N., Yongye, A. B., Tschampel, S. M., González-Outeirino, J., Daniels, C. R., Foley, B. L., et al. (2008). GLYCAM06: a generalizable biomolecular force field. *Carbohydrates. J. Comput. Chem.* 29, 622–655. doi: 10.1002/jcc.20820
- Koe, T. (2020a). “Can honey, omega-3, resistant potato starch help fight COVID-19? Researchers to examine via clinical trials,” in *News & Analysis on Supplements, Health and Nutrition – Asia Pacific*. Available online at: <https://www.nutraingredients-asia.com/Article/2020/04/20/Can-honey-omega-3-resistant-potato-starch-help-fight-COVID-19-Researchers-to-examine-via-clinical-trials>
- Koe, T. (2020b). “COVID-19 natural product trial: New curcumin, artemisinin supplement to be tested on patients,” in *News & Analysis on Supplements, Health and Nutrition – Asia Pacific*. Available online at: <https://www.nutraingredients-asia.com/Article/2020/04/23/COVID-19-natural-product-trial-New-curcumin-artemisinin-supplement-to-be-tested-on-patients>
- Korb, O., Stützel, T., and Exner, T. E. (2009). Empirical scoring functions for advanced protein-ligand docking with PLANTS. *J. Chem. Inf. Model.* 49, 84–96. doi: 10.1021/ci800298z
- Kuba, K., Imai, Y., Rao, S., Gao, H., Guo, F., Guan, B., et al. (2005). A crucial role of angiotensin converting enzyme 2 (ACE2) in SARS coronavirus-induced lung injury. *Nat. Med.* 11, 875–879. doi: 10.1038/nm1267
- Kurkinen, S. T., Lätti, S., Pentikäinen, O. T., and Postila, P. A. (2019). Getting docking into shape using negative image-based rescoring. *J. Chem. Inf. Model.* 59, 3584–3599. doi: 10.1021/acs.jcim.9b00383
- Kurkinen, S. T., Niinivehmas, S., Ahinko, M., Lätti, S., Pentikäinen, O. T., and Postila, P. A. (2018). Improving docking performance using negative image-based rescoring. *Front. Pharmacol.* 9:260. doi: 10.3389/fphar.2018.00260
- Laha, S., Chakraborty, J., Das, S., Manna, S. K., Biswas, S., and Chatterjee, R. (2020). Characterizations of SARS-CoV-2 mutational profile, spike protein stability and viral transmission. *Infect. Genet. Evol.* 85:104445. doi: 10.1016/j.meegid.2020.104445
- Lan, J., Ge, J., Yu, J., Shan, S., Zhou, H., Fan, S., et al. (2020). Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *Nature* 581, 215–220. doi: 10.1038/s41586-020-2180-5
- Lexa, K. W., Goh, G. B., and Carlson, H. A. (2014). Parameter choice matters: validating probe parameters for use in mixed-solvent simulations. *J. Chem. Inf. Model.* 54, 2190–2199. doi: 10.1021/ci400741u
- Lim, G. Y. (2020). “TCM and COVID-19: China conducting trials to test efficacy and safety of traditional herbal interventions,” in *News & Analysis on Supplements, Health and Nutrition – Asia Pacific*. Available online at: <https://www.nutraingredients-asia.com/Article/2020/04/06/TCM-and-COVID-19-China-conducting-trials-to-test-efficacy-and-safety-of-traditional-herbal-interventions>
- Lin, L., Ting, S., Yufei, H., Wendong, L., Yubo, F., and Jing, Z. (2020). Epitope-based peptide vaccines predicted against novel coronavirus disease caused by SARS-CoV-2. *Virus Res.* 288:198082. doi: 10.1016/j.virusres.2020.198082
- Ling, C. Q. (2020). Traditional Chinese medicine is a resource for drug discovery against 2019 novel coronavirus (SARS-CoV-2). *J. Integr. Med.* 18, 87–88. doi: 10.1016/j.joim.2020.02.004
- Mahmoud, I. S., Jarrar, Y. B., Alshaer, W., and Ismail, S. (2020). SARS-CoV-2 entry in host cells-multiple targets for treatment and prevention. *Biochimie* 175, 93–98. doi: 10.1016/j.biochi.2020.05.012
- Maier, J. A., Martinez, C., Kasavajhala, K., Wickstrom, L., Hauser, K. E., and Simmerling, C. (2015). ff14SB: improving the accuracy of protein side chain and backbone parameters from ff99SB. *J. Chem. Theory Comput.* 11, 3696–3713. doi: 10.1021/acs.jctc.5b00255
- Miller, B. R., McGee, T. D., Swails, J. M., Homeyer, N., Gohlke, H., and Roitberg, A. E. (2012). MMPBSA.py: an efficient program for end-state free energy calculations. *J. Chem. Theory Comput.* 8, 3314–3321. doi: 10.1021/ct300418h
- Mittal, A., Manjunath, K., Ranjan, R. K., Kaushik, S., Kumar, S., and Verma, V. (2020). COVID-19 pandemic: insights into structure, function, and hACE2 receptor recognition by SARS-CoV-2. *PLoS Pathog.* 16:e1008762. doi: 10.1371/journal.ppat.1008762
- Niinivehmas, S. P., Salokas, K., Lätti, S., Raunio, H., and Pentikäinen, O. T. (2015). Ultrafast protein structure-based virtual screening with panther. *J. Comput. Aided. Mol. Des.* 29, 989–1006. doi: 10.1007/s10822-015-9870-3
- Ponder, J. W., and Case, D. A. (2003). Force fields for protein simulations. *Advances* 66, 27–85. doi: 10.1016/S0065-3233(03)66002-X
- Roe, D. R., and Cheatham, T. E. (2013). PTRAJ and CPPTRAJ: software for processing and analysis of molecular dynamics trajectory data. *J. Chem. Theory Comput.* 9, 3084–3095. doi: 10.1021/ct400341p
- Ryckaert, J.-P., Ciccotti, G., and Berendsen, H. (1976). Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of *n*-alkenes. *J. Comp. Phys.* 23, 327–341. doi: 10.1016/0021-9991(77)90098-5
- Salmaso, V., and Moro, S. (2018). Bridging molecular docking to molecular dynamics in exploring ligand-protein recognition process: an overview. *Front. Pharmacol.* 9:923. doi: 10.3389/fphar.2018.00923
- Salomon-Ferrer, R., Götz, A. W., Poole, D., Le Grand, S., and Walker, R. C. (2013). Routine microsecond molecular dynamics simulations with AMBER on GPUs. 2. explicit solvent particle mesh ewald. *J. Chem. Theory Comput.* 9, 3878–3888. doi: 10.1021/ct400314y
- Shang, J., Wan, Y., Luo, C., Ye, G., Geng, Q., Auerbach, A., et al. (2020). Cell entry mechanisms of SARS-CoV-2. *Proc. Natl. Acad. Sci. U. S. A.* 117, 11727–11734. doi: 10.1073/pnas.2003138117
- Shereen, M. A., Khan, S., Kazmi, A., Bashir, N., and Siddique, R. (2020). COVID-19 infection: origin, transmission, and characteristics of human coronaviruses. *J. Adv. Res.* 24, 91–98. doi: 10.1016/j.jare.2020.03.005
- Steele, J. (2020). *Lab Finds 125 Naturally Occurring Compounds with Potential Against COVID-19*. 1–3. Available online at: <https://medicalxpress.com/news/2020-06-lab-naturally-compounds-potential-covid-.html>
- Sun, C., Chen, L., Yang, J., Luo, C., Zhang, Y., Li, J., et al. (2020). SARS-CoV-2 and SARS-CoV spike-RBD structure and receptor binding comparison and potential implications on neutralizing antibody and vaccine development. *bioRxiv [Preprint]*. doi: 10.1101/2020.02.16.951723
- Ung, P. M. U., Ghanakota, P., Graham, S. E., Lexa, K. W., and Carlson, H. A. (2016). Identifying binding hot spots on protein surfaces by mixed-solvent molecular dynamics: HIV-1 protease as a test case. *Biopolymers* 105, 21–34. doi: 10.1002/bip.22742
- Vainio, M. J., Puranen, J. S., and Johnson, M. S. (2009). ShaEP: molecular overlay based on shape and electrostatic potential. *J. Chem. Inf. Model.* 49, 492–502. doi: 10.1021/ci800315d

- Wahedi, H. M., Ahmad, S., and Abbasi, S. W. (2020). Stilbene-based natural compounds as promising drug candidates against COVID-19. *J. Biomol. Struct. Dyn.* doi: 10.1080/07391102.2020.1762743. [Epub ahead of print].
- Walls, A. C., Park, Y. J., Tortorici, M. A., Wall, A., McGuire, A. T., and Vesler, D. (2020). Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein. *Cell* 181, 281–292.e6. doi: 10.1016/j.cell.2020.02.058
- Wang, J., Wang, W., Kollman, P. A., and Case, D. A. (2006). Automatic atom type and bond type perception in molecular mechanical calculations. *J. Mol. Graph. Model.* 25, 247–260. doi: 10.1016/j.jmgl.2005.12.005
- Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A., and Case, D. A. (2004). Development and testing of a general amber force field. *J. Comput. Chem.* 25, 1157–1174. doi: 10.1002/jcc.20035
- World Health Organization (2020). *WHO Supports Scientifically-Proven Traditional Medicine*. Brazzaville: World Health Organization – Regional Office for Africa. Available online at: <https://www.afro.who.int/news/who-supports-scientifically-proven-traditional-medicine>
- Xiu, S., Dick, A., Ju, H., Mirzaie, S., Abdi, F., Cocklin, S., et al. (2020). Inhibitors of SARS-CoV-2 entry: current and future opportunities. *J. Med. Chem.* 15:acs.jmedchem.0c00502. doi: 10.1021/acs.jmedchem.0c00502
- Yang, Y., Islam, M. S., Wang, J., Li, Y., and Chen, X. (2020). Traditional Chinese medicine in the treatment of patients infected with 2019-new coronavirus (SARS-CoV-2): a review and perspective. *Int. J. Biol. Sci.* 16, 1708–1717. doi: 10.7150/ijbs.45538

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Gopinath, Jokinen, Kurkinen and Pentikäinen. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Identification of Small Molecule Inhibitors of the Deubiquitinating Activity of the SARS-CoV-2 Papain-Like Protease: *in silico* Molecular Docking Studies and *in vitro* Enzymatic Activity Assay

Eleni Pitsillou<sup>1,2</sup>, Julia Liang<sup>1,2</sup>, Katherine Ververis<sup>1</sup>, Kah Wai Lim<sup>1,3</sup>, Andrew Hung<sup>2</sup> and Tom C. Karagiannis<sup>1,4\*</sup>

<sup>1</sup> Epigenomic Medicine, Department of Diabetes, Central Clinical School, Monash University, Melbourne, VIC, Australia,

<sup>2</sup> School of Science, College of Science, Engineering and Health, RMIT University, Melbourne, VIC, Australia, <sup>3</sup> Department of Microbiology and Immunology, The University of Melbourne, Parkville, VIC, Australia, <sup>4</sup> Department of Clinical Pathology, The University of Melbourne, Parkville, VIC, Australia

## OPEN ACCESS

### Edited by:

Emilia Pedone,  
Institute of Biostructure and  
Bioimaging, Italy

### Reviewed by:

Pádraig D'Arcy,  
Linköping University, Sweden  
Muhammad Tahir Ul Qamar,  
Guangxi University, China

### \*Correspondence:

Tom C. Karagiannis  
tom.karagiannis@monash.edu

### Specialty section:

This article was submitted to  
Chemical Biology,  
a section of the journal  
Frontiers in Chemistry

**Received:** 30 October 2020

**Accepted:** 16 November 2020

**Published:** 08 December 2020

### Citation:

Pitsillou E, Liang J, Ververis K,  
Lim KW, Hung A and Karagiannis TC  
(2020) Identification of Small Molecule  
Inhibitors of the Deubiquitinating  
Activity of the SARS-CoV-2  
Papain-Like Protease: *in silico*  
Molecular Docking Studies and *in vitro*  
Enzymatic Activity Assay.  
Front. Chem. 8:623971.  
doi: 10.3389/fchem.2020.623971

COVID-19 is an ongoing pandemic caused by the SARS-CoV-2 virus with important political, socio-economic, and public health consequences. Inhibiting replication represents an important antiviral approach, and in this context two viral proteases, the SARS-CoV-2 main and papain-like proteases (PL<sup>pro</sup>), which cleave pp1a and pp1ab polypeptides, are critical. Along with protease activity, the PL<sup>pro</sup> possesses deubiquitinating activity, which is important in immune regulation. Naphthalene-based inhibitors, such as the well-investigated GRL-0617 compound, have been shown to possess dual effects, inhibiting both protease and deubiquitinating activity of the PL<sup>pro</sup>. Rather than binding to the canonical catalytic triad, these type of non-covalent inhibitors target an adjacent pocket, the naphthalene-inhibitor binding site. Using a high-throughput screen, we have previously identified the dietary hypericin, rutin, and cyanidin-3-O-glucoside compounds as potential protease inhibitors targeting the naphthalene-inhibitor binding site. Here, our aim was to investigate the binding characteristics of these compounds to the PL<sup>pro</sup>, and to evaluate deubiquitinating activity, by analyzing seven different PL<sup>pro</sup> crystal structures. Molecular docking highlighted the relatively high affinity of GRL-0617 and dietary compounds. In contrast binding of the small molecules was abolished in the presence of ubiquitin in the palm subdomain of the PL<sup>pro</sup>. Further, docking the small molecules in the naphthalene-inhibitor binding site, followed by protein-protein docking revealed displacement of ubiquitin in a conformation inconsistent with functional activity. Finally, the deubiquitinating activity was validated *in vitro* using an enzymatic activity assay. The findings indicated that the dietary compounds inhibited deubiquitinase activity in the micromolar range with an order of activity of GRL-0167, hypericin >> rutin, cyanidin-3-O-glucoside > epigallocatechin gallate, epicatechin gallate, and cefotaxime. Our findings are in accordance with

mechanisms and potential antiviral effects of the naphthalene-based, GRL-0617 inhibitor, which is currently progressing in preclinical trials. Further, our findings indicate that in particular hypericin, rutin, and cyanidin-3-O-glucoside, represent suitable candidates for subsequent evaluation as PL<sup>pro</sup> inhibitors.

**Keywords:** coronavirus, COVID-19, SARS-CoV-2, papain-like protease, deubiquitinase inhibitors, molecular docking

## INTRODUCTION

Coronavirus disease (COVID-19) was declared a pandemic on the 11th of March 2020 (World Health Organization, 2020). The first reported cases of this disease came from Wuhan, China in late 2019, and the infectious agent responsible for causing this disease was identified as severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) (Coronaviridae Study Group of the International Committee on Taxonomy of Viruses, 2020). Since the start of the year, the scientific literature on COVID-19 has increased and the findings from these studies have formed an integral part of the public health response.

In regards to treatment options, a number of vaccine trials have been established and there is also a focus on drug repositioning (Bar-Zeev and Moss, 2020; Folegatti et al., 2020). The U.S Food and Drug Administration (FDA) has recently approved remdesivir, which is an RNA-dependent RNA polymerase inhibitor, as a COVID-19 treatment for hospitalized patients (FDA, 2020). Interestingly, the WHO Solidarity trial has produced contradicting findings regarding the effectiveness of remdesivir (Pan H. et al., 2020). In addition to antiviral drugs, the efficacy of compounds that have immunomodulating properties are also being investigated (de la Rica et al., 2020).

Further research is required to establish the precise mechanisms of action of potential therapeutic drugs and determine their biological targets (Zhou Y. et al., 2020). The virus replication cycle is comprised of several stages and compounds that inhibit key proteins involved in these steps may have antiviral properties (Jeong et al., 2020; Pandey et al., 2020). Papain-like protease (PL<sup>pro</sup>) is a cysteine protease enzyme that is encoded by the multi-domain non-structural protein 3 (nsp3) and is required for polypeptide processing (Báez-Santos et al., 2015; Folegatti et al., 2020). The role of the SARS-CoV-2 PL<sup>pro</sup> in viral replication and the regulation of the innate immune response is being explored (Shin et al., 2020).

Ubiquitin and ubiquitin-like proteins, such as interferon-stimulated gene 15 (ISG15), are important effector molecules of the antiviral immune response (Jiang and Chen, 2011; Perng and Lenschow, 2018). Through binding to target proteins, various cellular pathways can be modulated (Jiang and Chen, 2011; Perng and Lenschow, 2018). Viruses have developed mechanisms to evade detection and destruction by the host's immune response, and these strategies continue to evolve (Nelemans and Kikkert, 2019). In terms of SARS-CoV-2, the deubiquitinating and deISGylating activities of the PL<sup>pro</sup> enzyme have been described (Bosken et al., 2020; Klemm et al., 2020; Rut et al., 2020; Shin et al., 2020). GRL-0617 is a naphthalene-based inhibitor that has

been found to interfere with the protease, deubiquitinase and deISGylating activities of the SARS-CoV and SARS-CoV-2 PL<sup>pro</sup> enzymes (Ratia et al., 2008; Freitas et al., 2020; Gao et al., in press). The SARS-CoV-2 PL<sup>pro</sup> has consequently been identified as an attractive drug target and in this study, the deubiquitinase activity of this viral protein was of interest (McClain and Vabret, 2020).

The health-promoting properties of dietary compounds have been extensively explored over the years and in response to the COVID-19 pandemic, bioactive compounds are being investigated further (Dhama et al., 2020; Mani et al., 2020). There is a growing body of literature on the antiviral and immunomodulating properties of plant-based compounds, and their potential use as therapeutic agents against SARS-CoV-2 (Tiwari et al., 2018; Panyod et al., 2020; Tahir Ul Qamar et al., 2020). This includes traditional Chinese and traditional Indian medicinal compounds, vitamins, curcumin, glycyrrhizic acid, tea polyphenols and compounds derived from *Allium sativum* to name a few (Chen et al., 2020; Divya et al., 2020; Donma and Donma, 2020; Tripathi et al., 2020). Enhancing the bioavailability of natural compounds continues to be a challenge however, their structures can be used as scaffolds for the development of novel drugs (Ngwa et al., 2020).

*In silico* methods were used to compare the binding mode of naphthalene-based inhibitors (GRL-0617 and 3k) to dietary compounds including hypericin, rutin, cyanidin-3-O-glucoside and (-)-epigallocatechin gallate. The antimicrobial, anti-inflammatory and antioxidant activities of these phytochemicals have been reported and their structures may even be used as scaffolds in the drug development process (Mohammadi Pour et al., 2019). The aim was to determine whether the dietary compounds were able to bind in a similar manner as the positive control GRL-0617, and potentially interfere with the binding of ubiquitin. The results were validated further using *in vitro* assays.

## MATERIALS AND METHODS

### Protein Structures and Ligands

Several crystal structures of the SARS-CoV-2 PL<sup>pro</sup> were obtained from the RCSB Protein Data Bank (PDB ID: 6xaa, 6w9c, 6wu, 6wx4, and 7jrn) (Berman et al., 2000; Klemm et al., 2020; Osipiuk et al., 2020; Rut et al., 2020; Sacco et al., 2020). The SARS-CoV PL<sup>pro</sup> (PDB ID: 4mm3) and MERS-CoV PL<sup>pro</sup> (PDB ID: 4rf0) were also used for comparison in this study (Bailey-Elkin et al., 2014; Ratia et al., 2014). Crystallographic waters were removed and the native zinc ions were retained. A ubiquitin

chain was present in the structures of 6xaa, 4mm3, and 4rf0, which was used to generate two sets of docking data for each protein: PL<sup>Pro</sup> in complex with ubiquitin and, and apo PL<sup>Pro</sup> in the absence of ubiquitin. The ubiquitin in each complex was also isolated for protein-protein docking. The ligands that were used in this *in silico* study were the naphthalene inhibitors GRL-0617 and 3k, and the dietary compounds (-)-epigallocatechin gallate, cyanidin-3-O-glucoside, rutin and hypericin. The structures of the dietary compounds were obtained from the National Center for Biotechnology Information PubChem (Kim et al., 2019). GRL-0617 and 3k were drawn using Chem3D 19.0 (Perkin Elmer, Massachusetts, USA).

## Molecular Docking Using the Schrödinger Suite

Molecular docking was performed using the Schrödinger Suite (Schrödinger, 2020a). The protein structures were prepared using the Protein Preparation Wizard, while the compounds were prepared using the LigPrep tool (Madhavi Sastry et al., 2013; Schrödinger, 2020b). The default settings were used for both of these steps and the optimized potentials for liquid simulations 3e (OPLS3e) force field was selected (Jorgensen and Tirado-Rives, 1988; Jorgensen et al., 1996; Shivakumar et al., 2010; Harder et al., 2016). The top ranking ligand conformation was selected for the subsequent molecular docking stage.

The residues that were within 5 Å of the co-crystallized ligand GRL-0617 in the 7jrn structure were used to generate the receptor grid for each SARS-CoV-2 PL<sup>Pro</sup> (Friesner et al., 2004, 2006; Halgren et al., 2004; Schrödinger, 2020c). These residues were E167, K157, Y273, D164, G163, L162, C270, Q269, Y268, N267, G266, Y264, P248, P247, M208, and T301. To determine the corresponding residues in the SARS-CoV and MERS-CoV crystal structures, pairwise alignment was performed in the Multiple Sequence Viewer tool (Schrödinger, 2020a). The residues used to form the grid for SARS-CoV PL<sup>Pro</sup> were E168, K158, Y274, D165, G164, L163, C271, Q270, Y269, N268, G267, Y265, P249, P248, and M209. For MERS-CoV PL<sup>Pro</sup>, the receptor grid was generated based on the residues R1649, C1639, Y1760, D1646, D1645, P1644, V1757, A1756, T1755, E1754, G1752, F1750, P1731, T1730, V1691, and T1789.

The receptor grids were 20 × 20 × 20 Å in size and the OPLS3e force field was utilized. The compounds were then docked to each protein structure using the quantum-polarized ligand docking (QPLD) protocol, as previously described (Liang et al., 2020).

## Blind Docking and the Prediction of Ligand-Binding Sites

The PL<sup>Pro</sup> crystal structures and the compounds were prepared as macromolecules and ligands in PyRx, respectively (Dallakyan and Olson, 2015). The corresponding.pdbqt files were obtained and the receptor grid was generated around the entire surface of the protein. The exhaustiveness was set to 2048. AutoDock Vina was used for molecular docking and the jobs were run on the cloud-computing server Galileo (Hypernet Labs) (Trott and Olson, 2010; Hypernet Labs Galileo, 2020). In addition to blind docking, the PrankWeb server was used to identify potential

binding pockets that were conserved in the 6xaa, 4mm3, and 4rf0 crystal structures (Jendele et al., 2019).

## Protein-Protein Docking

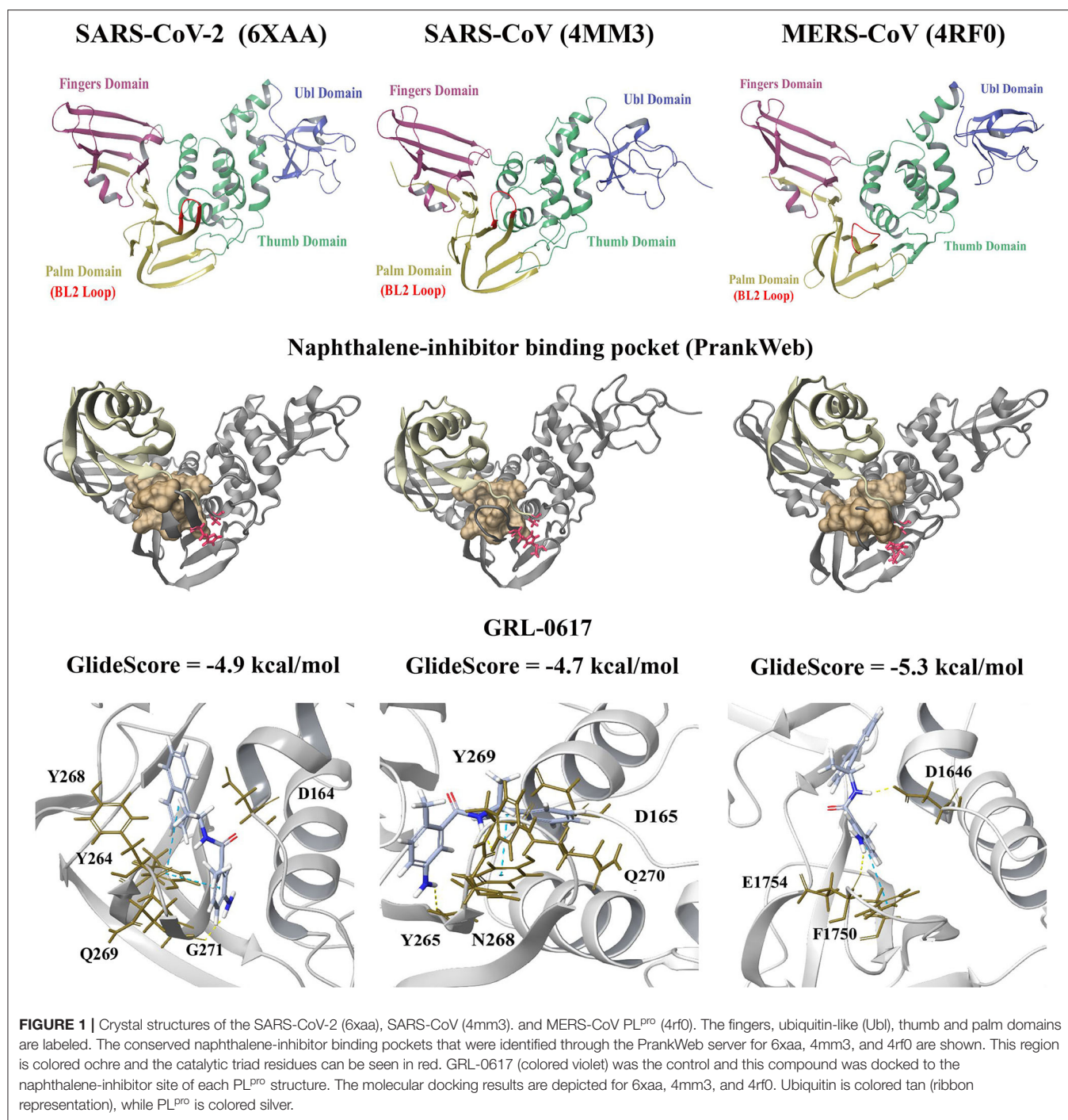
The HDock server was used for *ab initio* template free protein-protein docking in this study (Yan et al., 2017, 2020). For the SARS-CoV-2, SARS-CoV, and MERS-CoV complexes, the main chain of PL<sup>Pro</sup> was defined as the receptor molecule, whereas the ubiquitin chain was the ligand. Protein-protein docking was also performed with PL<sup>Pro</sup> in the presence of compounds that were docked to the naphthalene-inhibitor site in the Schrödinger Suite. This was done for the 6xaa, 4mm3, and 4rf0 crystal structures in order to determine whether the presence of these ligands in the naphthalene-inhibitor region would affect the ability of ubiquitin to bind to PL<sup>Pro</sup>.

## Enzymatic Activity Assay

*In vitro* inhibition of the PL<sup>Pro</sup> deubiquitinase activity was measured using a commercially available enzymatic activity assay (BP Bioscience, San Diego, CA, USA). The experiment was performed according to the manufacturer's instructions and all samples were assayed in triplicate. An excitation wavelength of 360 nm was used, and fluorescence was measured at an emission wavelength of 460 nm on the basis of the presence of a ubiquitinated fluorogenic substrate. The non-covalent inhibitor GRL-0617 was provided as an internal positive control and was used at concentration of 100 μM in the assay. We tested the following compounds for potential inhibition of PL<sup>Pro</sup> deubiquitinase activity: hypericin (89%, HWI pharma services GmbH, Germany), cyanidin-3-O-glucoside (reference standard, PhytoLab, Germany), and rutin (>94%), (-)-epicatechin gallate (>98%), (-)-epigallocatechin gallate (>95%), and cefotaxime (European pharmacopeia reference standard) from Sigma-Aldrich (St Luis, MO, USA); 20 mM stock solutions of each compound were stored at -80°C until use. Serial doubling dilutions were performed to yield a final concentration of 3.1 to 200 μM for assaying each compound. Readings (absolute fluorescence values at 1,200 gain), were made using a CLARIOstar® Plus fluorescence microplate reader (BMG Biotech, Otertberg, Germany). Where appropriate the % PL<sup>Pro</sup> deubiquitinase activity was calculated as the ratio of activity in the presence of inhibitor and total activity, and taking into account background readings.

## RESULTS

Molecular docking was performed to examine the potential inhibitor behavior of dietary compounds and naphthalene inhibitors GRL-0617 and 3k on the SARS-CoV-2 PL<sup>Pro</sup>. Like SARS-CoV-2, SARS-CoV, and MERS-CoV are classified as betacoronaviruses and they predominantly affect the respiratory tract (Abdelrahman et al., 2020; Petersen et al., 2020). The PL<sup>Pro</sup> from the novel SARS-CoV-2 was consequently compared to the SARS-CoV and MERS-CoV PL<sup>Pro</sup> structures (Figure 1). In terms of the SARS-CoV and MERS-CoV PL<sup>Pro</sup>, the pairwise alignment revealed they had a sequence similarity of 89 and 50% with the SARS-CoV-2 PL<sup>Pro</sup>,



respectively. The pairwise alignment was used to identify the corresponding residues of the naphthalene-inhibitor binding site for the SARS-CoV and MERS-CoV crystal structures. The three crystal structures (6xaa, 4mm3, and 4rf0) were analyzed using PrankWeb, a binding site prediction tool, which identified the naphthalene-inhibitor binding site as a conserved ligand binding pocket. This region was ranked as pocket

2, pocket 3 and pocket 2 in the SARS-CoV-2, SARS-CoV and MERS-CoV PL<sup>pro</sup> structures, respectively (**Figure 1** and **Supplementary Table 1**).

GRL-0617 and 3k, as well as the dietary compounds were docked to the naphthalene-inhibitor binding site of the SARS-CoV-2 and SARS-CoV PL<sup>pro</sup> structures. In the apo 6xaa crystal structure of SARS-CoV-2, the control compound

GRL-0617 docked to the naphthalene-inhibitor binding site with a GlideScore of  $-4.9$  kcal/mol (**Figure 1**). It was predicted to form inter-atomic contacts with G271 (H-bond) and Y264 ( $\pi$ - $\pi$  interactions). The GlideScore for the naphthalene-based inhibitor 3k was  $-2.4$  kcal/mol and it also formed a  $\pi$ - $\pi$  interaction with the PL<sup>Pro</sup> residue Y264 (**Figure 2**). In addition to Y264, a hydrogen bond with L162 was present in the ligand-interaction diagram. The molecular docking results demonstrated that the dietary compounds (-)-epigallocatechin gallate, cyanidin-3-O-glucoside, rutin and hypericin had strong binding affinities (**Figure 2**). These ligands formed a bond with Y268 (cyanidin-3-O-glucoside: H-bond, (-)-epigallocatechin gallate: H-bond, hypericin:  $\pi$ - $\pi$  interaction, and rutin: H-bond). The protein residues Y264, T301, D164, R166, E167, K157, L162, G163, and Y273 were also involved in interactions with the ligands (**Figure 2**).

The GlideScores of the naphthalene-based inhibitors and the dietary compounds for the SARS-CoV and MERS-CoV PL<sup>Pro</sup> structures can be seen in **Figures 1, 2**. Similar to docking to the same site in the SARS-CoV-2 PL<sup>Pro</sup>, the dietary ligands also had stronger binding affinities and formed a greater number of inter-atomic contacts with surrounding residues compared to GRL-0617 and 3k. GRL-0617 formed bonds with Y265 (H-bond and  $\pi$ - $\pi$  interaction) and N268 (H-bond) in the SARS-CoV PL<sup>Pro</sup>, whereas these interactions were absent in docking with 3k (**Figures 1, 2**). The residues Y265, Y269, Q270, Y274, D165, G164, and N268 were the most prominent amino acids involved in binding with (-)-epigallocatechin gallate, cyanidin-3-O-glucoside, hypericin and rutin (**Figure 2**). In regards to MERS-CoV, both GRL-0617 and 3k formed inter-atomic contacts with E1754 (GRL-0617: H-bond, 3k: H-bond), F1750 (GRL-0617:  $\pi$ - $\pi$  interaction, 3k:  $\pi$ - $\pi$  cation) and D1646 (GRL-0617: H-bond, 3k: salt bridge) (**Figures 1, 2**). The dietary compounds also interacted with these residues, as well as D1645, G1758, A1756, G1729, and Y1760 (**Figure 2**).

Furthermore, blind docking was conducted on the 6xaa (SARS-CoV-2), 4mm3 (SARS-CoV), and 4rf0 (MERS-CoV) crystal structures in the absence of ubiquitin (**Figure 3** and **Supplementary Figures 1, 2**). The blind docking results for the SARS-CoV-2 PL<sup>Pro</sup> can be seen in **Figure 3**. All six compounds had poses within the naphthalene-inhibitor binding site. For the SARS-CoV-2 and MERS-CoV PL<sup>Pro</sup>, all poses generated for (-)-epigallocatechin gallate were predicted to bind to this region (**Figure 3** and **Supplementary Figure 2**). The highest ranking pose that was positioned in the naphthalene-inhibitor binding pocket of the SARS-CoV-2 PL<sup>Pro</sup> was pose 1 for GRL-0617 ( $-8.0$  kcal/mol), pose 4 for 3k ( $-7.5$  kcal/mol), pose 1 for cyanidin-3-O-glucoside ( $-7.4$  kcal/mol), pose 1 for (-)-epigallocatechin gallate ( $-8.1$  kcal/mol), pose 1 for hypericin ( $-8.5$  kcal/mol), and pose 1 for rutin ( $-7.6$  kcal/mol). The binding affinities of the poses that were present in the naphthalene-inhibitor binding region for the 6xaa, 4mm3, and 4rf0 crystal structures can be found in **Supplementary Table 2**.

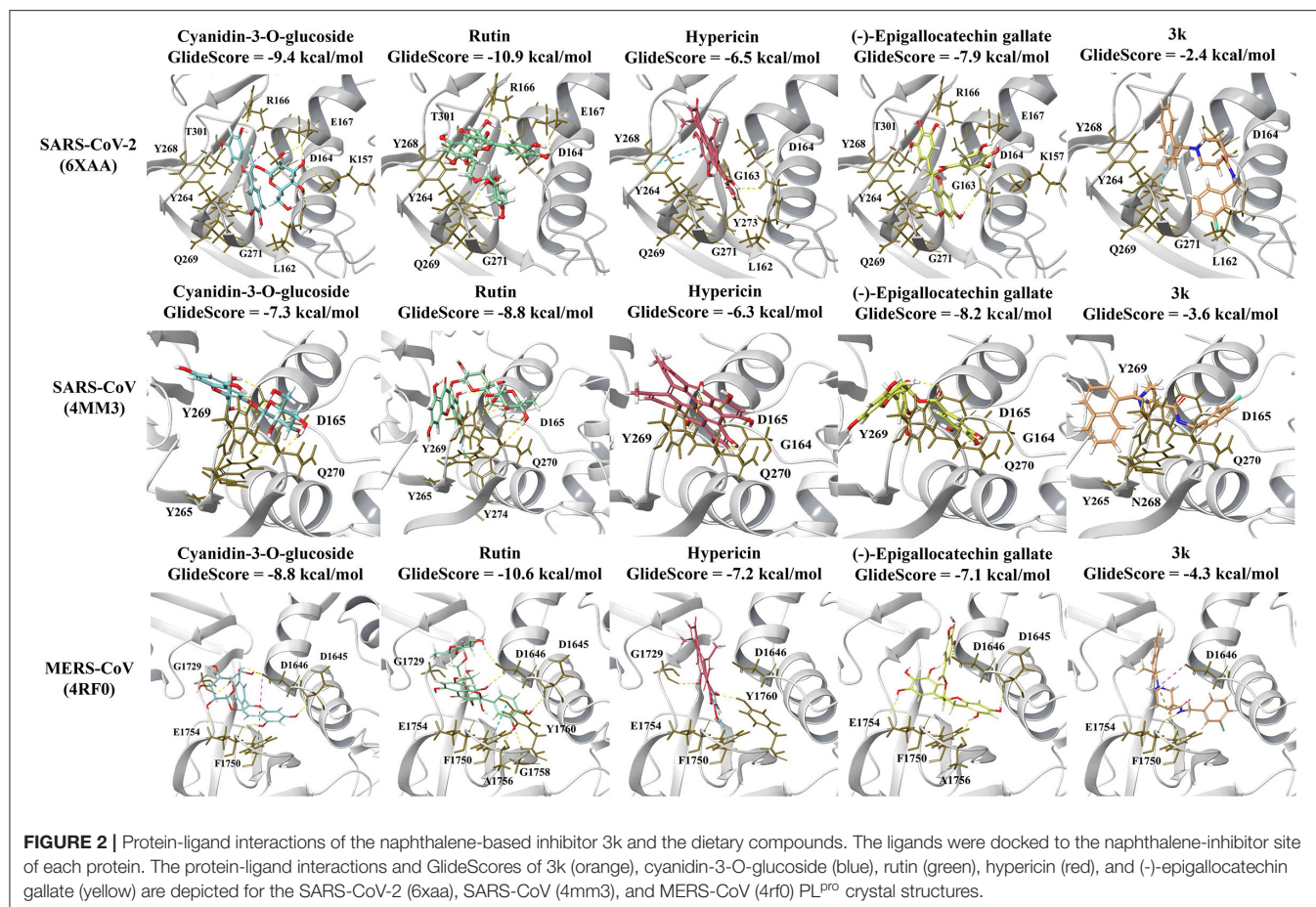
Moreover, the six compounds were docked to the 6wu, 6w9c, 6wx4, and 7jrn crystal structures of the SARS-CoV-2 PL<sup>Pro</sup> (**Supplementary Figures 3–6**). With the exception of the 7jrn PL<sup>Pro</sup> structure, all of the dietary compounds had stronger

GlideScores than the naphthalene-based inhibitors. Hypericin, rutin and (-)-epigallocatechin gallate were predicted to bind more strongly than GRL-0617 to the 7jrn PL<sup>Pro</sup> structure, while cyanidin-3-O-glucoside had a similar GlideScore to the control compound. In the 6w9c, 6wu, and 7jrn SARS-CoV-2 PL<sup>Pro</sup> structures, 3k formed bonds with Y264 (6wu: H-bond, 6w9c: H-bond, and salt bridge, 7jrn: salt bridge). In 7jrn, 3k formed a hydrogen bond with Y268. The intermolecular bonds that GRL-0617 formed with the protein residues varied amongst the crystal structures. GRL-0617 formed two hydrogen bonds with G266 and N267 in 6w9c, whereas a hydrogen bond was present with D164 in 6wu. In 6wx4, there were two  $\pi$ - $\pi$  interactions with Y264 and hydrogen bonds with G271 and Y268. There were  $\pi$ - $\pi$  interactions with Y268 in 7jrn and a hydrogen bond with Q269.

Compared to GRL-0617 and 3k, the ligands rutin, cyanidin-3-O-glucoside and (-)-epigallocatechin gallate formed a greater number of inter-atomic contacts with the protein residues (**Supplementary Figures 3–6**). In 6w9c, the dietary compounds formed hydrogen bonds with D164. Hydrogen bonds were also present with D164 for (-)-epigallocatechin gallate and rutin in the 6wu and 7jrn structures. The amino acid Y268 was also frequently involved in intermolecular bonds with the dietary compounds. This included (-)-epigallocatechin gallate (6w9c: H-bond), cyanidin-3-O-glucoside (6wu: H-bond, 7jrn:  $\pi$ - $\pi$  interaction), rutin (6wu: H-bond and 6wx4: H-bond), and hypericin (6wx4:  $\pi$ - $\pi$  interaction and 7jrn:  $\pi$ - $\pi$  interactions). In the 7jrn PL<sup>Pro</sup>, hypericin also formed a hydrogen bond with Q269 and this was similar to the control compound. The blind docking results for the 6xaa, 6w9c, 6wx4, and 7jrn SARS-CoV-2 PL<sup>Pro</sup> crystal structures can be found in **Supplementary Table 3**.

Molecular docking was also performed on the SARS-CoV-2, SARS-CoV and MERS-CoV PL<sup>Pro</sup> structures with the ubiquitin chain present (**Figure 4**). When GRL-0617, 3k, (-)-epigallocatechin gallate, rutin, cyanidin-3-O-glucoside, and hypericin were docked to PL<sup>Pro</sup>, it was evident that they were binding distant from the naphthalene-inhibitor binding site. The position and orientation of these ligands, as well as their GlideScores, differed to when ubiquitin was absent (**Figure 4** and **Supplementary Table 4**). GRL-0617 in the SARS-CoV-2 PL<sup>Pro</sup>, for example, was found to form hydrogen bonds with P248 and G266. The control compound was no longer binding to the pocket that the C-terminal chain of ubiquitin extends into. Interestingly, hypericin was not able to bind in the presence of ubiquitin. Similarly, the blind docking results for PL<sup>Pro</sup> in complex with ubiquitin revealed that the compounds that originally had poses within the naphthalene-inhibitor binding site were displaced from this region.

The PL<sup>Pro</sup>-ubiquitin complexes were investigated further using the HDock server for protein-protein docking (**Figure 5**). The ubiquitin chain from each complex was isolated and was re-docked to the main PL<sup>Pro</sup> structure. There was a clear overlap in the position between the top ranked model of the docked ubiquitin and the crystallographic ubiquitin. The root-mean square deviation (RMSD) values that were generated from protein-protein docking were 0.4, 0.6, and 0.5 Å for the SARS-CoV-2, SARS-CoV, and MERS-CoV complexes, respectively, and importantly, the C-terminal tail of ubiquitin extended



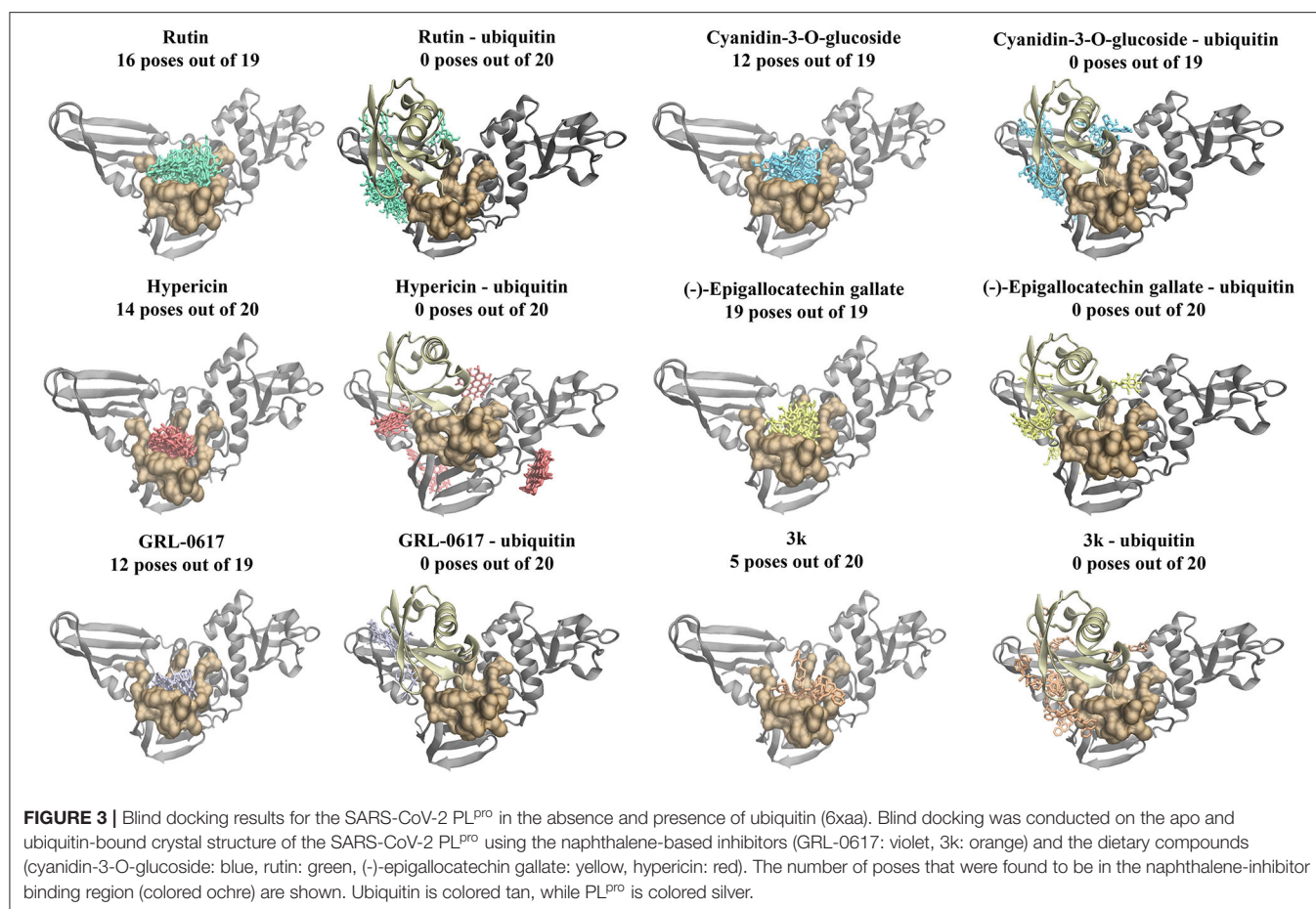
into the catalytic and naphthalene binding region of PL<sup>pro</sup>. The C-terminus of ubiquitin in the SARS-CoV-2, SARS-CoV and MERS-CoV PL<sup>pro</sup> consists of residues R72, L73, R74, and G75. The C-terminal residue of ubiquitin is AYE76, GLZ76, and 3CN101 in the SARS-CoV-2, SARS-CoV and MERS-CoV structures, respectively. These three residues form a covalent bond with the catalytic cysteine residues and this is critical for deubiquitinase activity (C111 in SARS-CoV-2: 1.67 Å, C112 in SARS-CoV: 1.72 Å and C1592 in MERS-CoV: 1.45 Å). The 1.67, 1.72, and 1.45 Å correspond to the distances between the sulfur atom of the catalytic cysteine residue and the C-terminal ubiquitin residue in the crystal structures. Although non-covalent protein-protein docking was performed in the present study, the distance between AYE76 and C111 in SARS-CoV-2 was 1.94 Å. The distance between GLZ76 and residue C112 in SARS-CoV was 2.05 Å. Likewise, the distance between 3CN101 and C1592 in MERS-CoV was 1.48 Å.

In order to examine the ability of ubiquitin to bind to PL<sup>pro</sup> when compounds are present in the naphthalene-inhibitor binding site, the ligands that were docked using the Schrödinger Suite were retained in the protein structures (Figure 5). When GRL-0617, 3k, (-)-epigallocatechin gallate, rutin, cyanidin-3-O-glucoside, and hypericin were bound to this region, dramatic

differences were observed in the binding mode of ubiquitin. The C-terminal chain of ubiquitin was no longer extending into the pocket that is located just above the catalytic triad. There were consequently changes in the position and orientation of ubiquitin for each ligand. When comparing the docked ubiquitin chain to the ubiquitin present in the original crystal structure, the RMSD values and docking scores were different (Supplementary Table 5). This was apparent for the SARS-CoV-2, SARS-CoV and MERS-CoV complexes.

### Hypericin, Rutin, and Cyanidin-3-O-Glucoside Inhibit PL<sup>pro</sup> Deubiquitinase Activity in a Concentration-Dependent Manner

Inhibition of PL<sup>pro</sup> deubiquitinase activity by small molecules *in vitro* was investigated using an enzymatic assay. Overall, the findings indicate that hypericin, rutin, and cyanidin-3-O-glucoside resulted in a concentration-dependent inhibition of PL<sup>pro</sup> deubiquitinase activity, with hypericin clearly highlighting the most potent inhibition of the test ligands (Figure 6A). At the higher concentrations (>50 μM) hypericin inhibited PL<sup>pro</sup> deubiquitinase activity to a level akin to the internal



positive control (GRL-0617), which was used at 100  $\mu$ M in the assay. Indeed, calculation of the percentage inhibition of PL<sup>Pro</sup> deubiquitinase activity at 100  $\mu$ M for each ligand highlighted the equivalence of GRL-0617 and hypericin in the assay (inhibition of activity by  $\sim$ 90% by both compounds, **Figure 6B**). The findings indicate that at 100  $\mu$ M both rutin ( $\sim$ 50% inhibition) and cyanidin-3-O-glucoside ( $\sim$ 42% inhibition), can also be considered as potentially useful inhibitors of PL<sup>Pro</sup> deubiquitinase activity. The epicatechins (epigallocatechin gallate,  $\sim$ 14% and epicatechin gallate,  $\sim$ 20%), and cefotaxime ( $\sim$ 2%), yielded more modest effects.

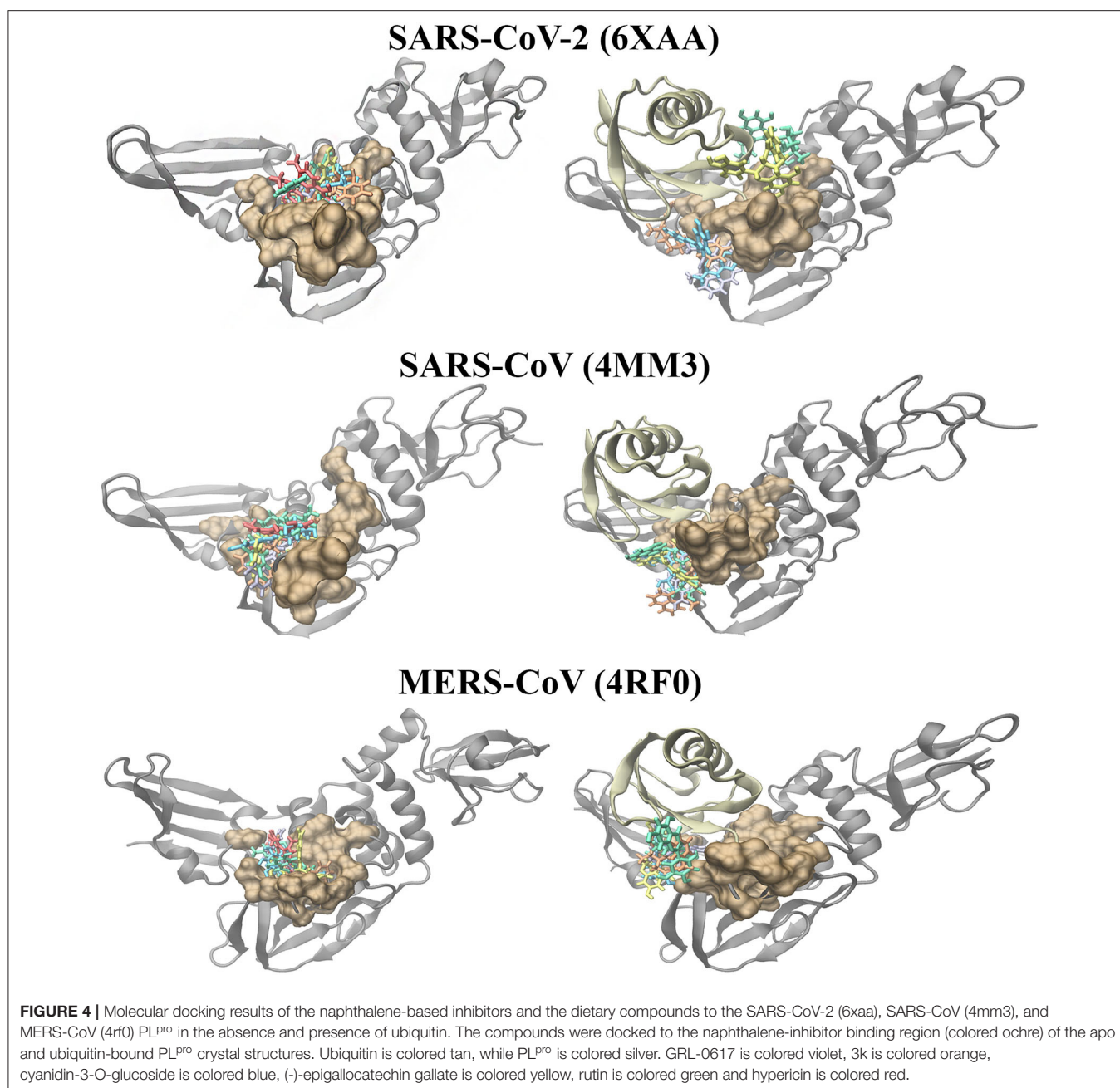
## DISCUSSION

The innate immune system is the first line of defense against foreign pathogens and various cellular and molecular components are involved in this process (Liu et al., 2016; Huang et al., 2019). Post-translational modifications are also important regulators of immunity and this includes ubiquitination (Liu et al., 2016). Ubiquitin is a 76-amino acid polypeptide that can covalently interact with target proteins and ubiquitin itself can undergo ubiquitination at certain residues (Ciechanover et al., 2000; Pickart, 2001; Jiang and Chen, 2011). This results in

the formation of lysine-linked polyubiquitin chains or linear polyubiquitin chains (Heaton et al., 2015). It is well-known that lysine 48 (K48)-linked polyubiquitylation promotes the proteasomal degradation of target proteins (Ciechanover et al., 2000). Conversely, lysine 63 (K63)-linked polyubiquitylation has been implicated in cellular processes such as the DNA damage response, inflammation and endocytosis (Panier and Durocher, 2009; Erpapazoglou et al., 2014; Zhou Z. et al., 2020). It is also important to note that other types of polyubiquitin chain linkages are being explored and that ubiquitin modifications can lead to different cellular outcomes (Komander and Rape, 2012; Ohtake et al., 2018).

Moreover, human deubiquitinases (DUBs) are enzymes that remove ubiquitin modifications and they contribute to homeostasis (Li et al., 2016). Viruses are dependent on host cells for their survival and in order to complete their life cycle, they have developed strategies to evade the antiviral immune response (Nelemans and Kikkert, 2019). Interestingly, several viral proteins have been found to possess deubiquitinating activity and they can be used to antagonize or modulate the antiviral immune signaling pathway (Kumari and Kumar, 2018).

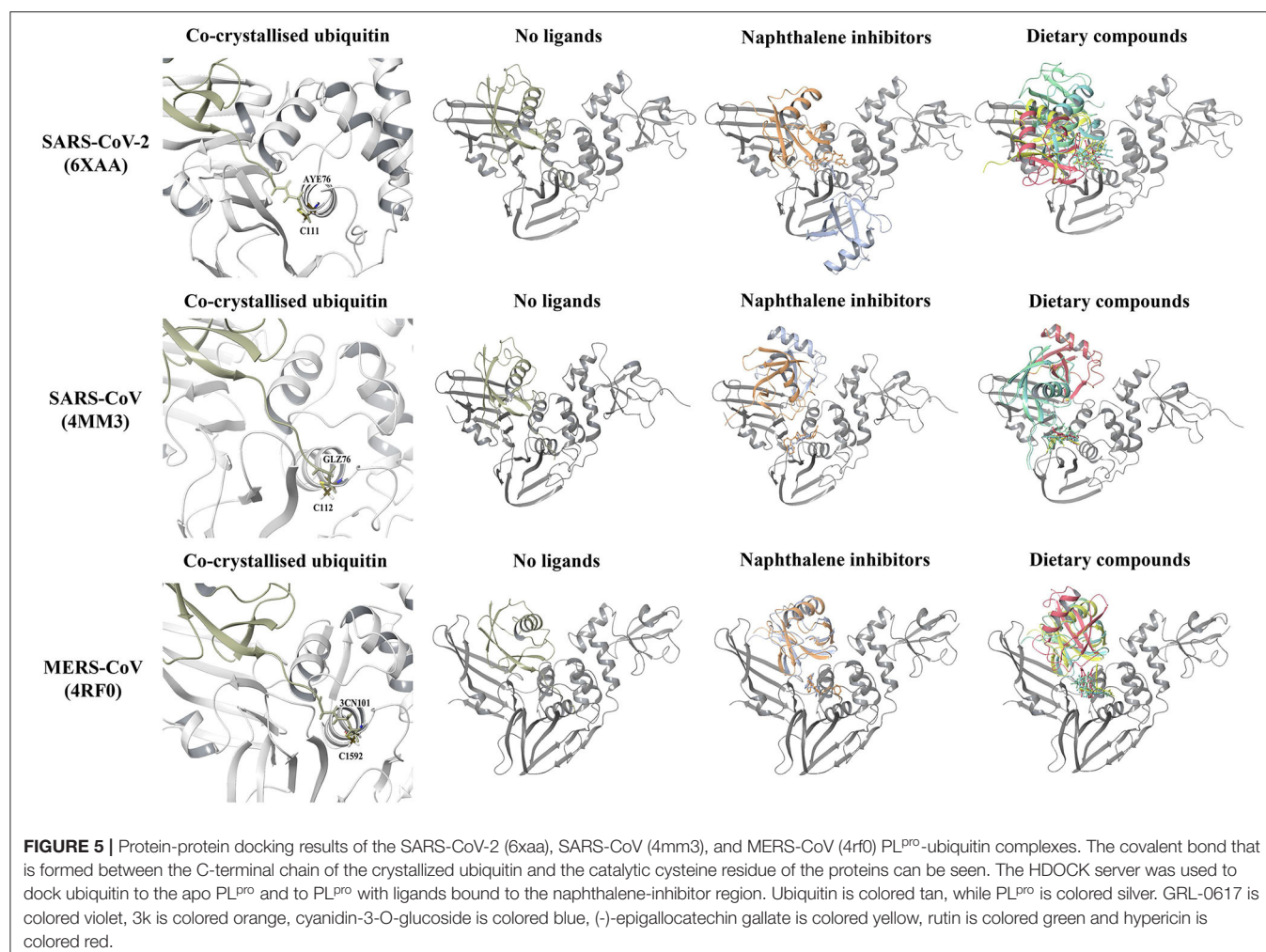
The deubiquitinating activity of the SARS-CoV-2 PL<sup>Pro</sup> was the main focus of this study and the crystal structure of



PL<sup>Pro</sup> in complex with ubiquitin propargylamide was utilized (Klemm et al., 2020). In addition to this, four other crystal structures of SARS-CoV-2 that were available on the RCSB PDB were evaluated. The crystal structures of the SARS-CoV PL<sup>Pro</sup>-ubiquitin aldehyde and MERS-CoV PL<sup>Pro</sup>-ubiquitin complexes were used for comparison (Bailey-Elkin et al., 2014; Ratia et al., 2014). Molecular docking allowed for the binding properties of compounds to the known target site of naphthalene-based inhibitors to be predicted and examined.

GRL-0617 was the control and this has previously been found to potently inhibit the SARS-CoV and SARS-CoV-2 PL<sup>Pro</sup> in a non-covalent manner (Ratia et al., 2008; Freitas

et al., 2020; Shin et al., 2020). The GRL-0617 inhibitor occupies the S3 and S4 pockets of the SARS-CoV and SARS-CoV-2 PL<sup>Pro</sup> (Ratia et al., 2008; Gao et al., in press). Based on the molecular docking results from the current study, GRL-0617 was predominantly surrounded by hydrophobic residues in the SARS-CoV and SARS-CoV-2 crystal structures (**Figure 1** and **Supplementary Figures 3–6**) (Ratia et al., 2008). This ligand was predicted to form inter-atomic contacts with the protein residues and this included D164 in the 6wu SARS-CoV-2 structure, as well as Q269 in the 7jrn SARS-CoV-2 structure. In the crystal structure determined by Gao et al., GRL-0617, which was the co-crystallized ligand, was found to form hydrogen bonds with these



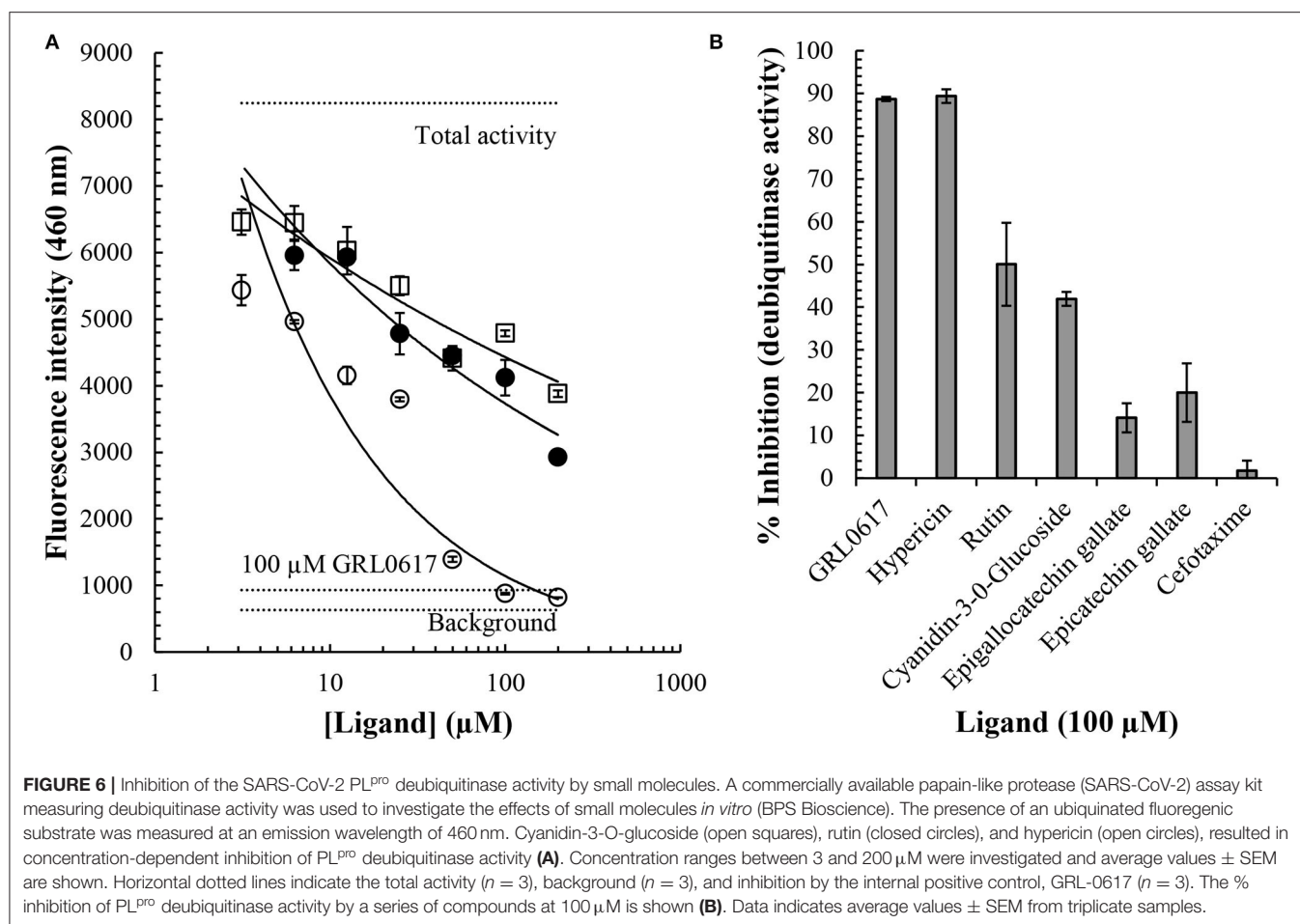
critical residues (Gao et al., in press). In saying this, D164 and Q269 were found to surround GRL-0617 in the 6xaa, 6w9c, and 6wx4 structures of the SARS-CoV-2 PL<sup>pro</sup>.

GRL-0617 also interacted with Y268 in the SARS-CoV-2 PL<sup>pro</sup> and inter-atomic contacts were present with this residue in some structures. Interestingly, the naphthalene-based inhibitor 3k was found to consistently form inter-atomic contacts with Y264 in all of the SARS-CoV-2 crystal structures (Figure 2 and Supplementary Figures 3–6) (Bosken et al., 2020). Intermolecular bonds were also formed between 3k and D164 in three of the SARS-CoV-2 papain-like proteases, as well as Y268 in the 6wx4 structure (Bosken et al., 2020). Bosken et al. have identified these residues as playing an important role in the binding mode of 3k (Bosken et al., 2020).

The differences observed in the intermolecular bonds may be due to the conformations of the fingers domain and BL2 loop in the PL<sup>pro</sup> crystal structures (Figure 1) (Báez-Santos et al., 2015). In the SARS-CoV-2 and SARS-CoV PL<sup>pro</sup> structures, the BL2 loop corresponds to residues 267–272 (Lee et al., 2015; Gao et al., in press). In the MERS-CoV PL<sup>pro</sup> structure used in this study, the BL2 loop is comprised of residues 1,752–1,758 (Bailey-Elkin

et al., 2014; Lee et al., 2015). The structural significance of the BL2 loop (blocking loop) has been discussed in a number of papers and its flexibility has been highlighted (Báez-Santos et al., 2015; Bosken et al., 2020; Klemm et al., 2020). Conformational changes have been observed in the BL2 loop and “open” or “closed” conformations have been reported in the literature (Báez-Santos et al., 2015). In regards to MERS-CoV, there are significant structural differences in the BL2 loop and it has been suggested that this affects inhibitor recognition specificity (Lee et al., 2015).

GRL-0617 is ineffective against MERS-CoV and in the study by Shin et al., it was discussed that this may be due to the presence of a threonine residue instead of tyrosine at a conserved position (Lee et al., 2015; Shin et al., 2020). In the SARS-CoV-2 and SARS-CoV PL<sup>pro</sup> sequences, the corresponding residues are Y268 and Y269, respectively (Shin et al., 2020). Y268 is required for the inhibitory effect of GRL-0617 and Shin et al. demonstrated that mutating this residue strongly reduces its potency (Shin et al., 2020). While GRL-0617 and 3k were predicted to bind to the MERS-CoV PL<sup>pro</sup> (4rf0) in this study, further docking to additional crystal structures may be required for comparison (Figures 1, 2). The naphthalene-based inhibitors



were surrounded by the residues D165, Y269, and Q270 in the SARS-CoV PL<sup>Pro</sup>. Ratia et al. and Báez-Santos et al., have also described the importance of these residues in the mechanisms of action of these ligands (Ratia et al., 2008; Báez-Santos et al., 2014).

Most notably, the dietary compounds (-)-epigallocatechin gallate, hypericin, rutin and cyanidin-3-O-glucoside were predicted to bind more strongly to the naphthalene-inhibitor site of the SARS-CoV-2, SARS-CoV, and MERS-CoV papain-like proteases than the known inhibitors (Figure 2 and Supplementary Figures 3–6). They also formed multiple interactions with the key protein residues compared to GRL-0617 and 3k. Similarly, the blind docking results on the main PL<sup>Pro</sup> chains showed that these natural ligands had multiple poses within this region (Figure 3 and Supplementary Figures 1, 2). (-)-Epigallocatechin gallate, rutin and cyanidin-3-O-glucoside are flavonoids, a biologically active class of the phenolic compounds (Bonvino et al., 2018). In a recent literature review conducted by Verma et al., the flavonoids were found to be the largest class of compounds with potential activity against coronaviruses (Verma et al., 2020).

In 2005, Li et al. published a study about the antiviral activities of natural compounds against SARS-CoV (Li et al.,

2005). Lycorine was identified as a potent antiviral compound and potentially a candidate for the development of new medicines (Li et al., 2005). Natural compounds have been screened for their ability to target SARS-CoV-2 proteins. This includes extracts of medicinal herbs and several studies have focused on their inhibitory effects on key proteins, such as the spike glycoprotein and the main protease (M<sup>Pro</sup>) (Mani et al., 2020; Pitsillou et al., 2020; Russo et al., 2020; Smith and Smith, 2020). In a recent paper by Alamri et al. a structured-based computational approach was utilized to identify compounds that may act as pan-PL<sup>Pro</sup> inhibitors and could be developed further as antiviral agents (Alamri et al., in press). Given the current situation, *in silico* methods have made it possible for large libraries of existing approved compounds to be screened in a relatively fast manner (Ojha et al., 2020). The structures of the hits identified from these computational studies could be optimized as part of the drug discovery process (Ojha et al., 2020). In addition to synthetic pharmacological compounds, herbal constituents can be screened in the same manner and this method has been described in many papers (Bhowmik et al., 2020; Chikhale et al., 2020; Ghosh et al., 2020; Gupta et al., 2020; Jena et al., 2020; Krupanidhi et al., 2020; Muhseen et al., 2020; Sinha et al.,

2020; Subbaiyan et al., 2020). A number of studies that can be found on the World Health Organization's International Clinical Trials Registry Platform also involve plant-based compounds, particularly flavonoids.

Ratia et al. determined the crystal structure of SARS-CoV in complex with ubiquitin aldehyde and they described how this polypeptide interacts with the palm and fingers regions of PL<sup>Pro</sup> (Ratia et al., 2014). They emphasized that a significant amount of the binding energy of ubiquitin can be attributed to its C-terminal residues (R72-G76) and that this portion of ubiquitin forms an extensive number of intermolecular hydrogen bonds with PL<sup>Pro</sup> (Ratia et al., 2014). The results from their study also indicated that the SARS-CoV PL<sup>Pro</sup> had a preference for K48-linked ubiquitin and ISG15, over K63-polyubiquitin chains and mono-ubiquitin (Ratia et al., 2014). Most notably, two recognition sites on the surface of PL<sup>Pro</sup> were characterized and were defined as either SUB1 or SUB2 (Ratia et al., 2014).

In the SARS-CoV-2 PL<sup>Pro</sup> complex, ubiquitin propargylamide sits on the same subdomains as the SARS-CoV structure (palm and fingers regions), and the C-terminus extends into the active site (Klemm et al., 2020). Like SARS-CoV, the SARS-CoV-2 PL<sup>Pro</sup> was also found to have a second ubiquitin binding site (SUB2) that is important for the binding of polyubiquitin (K48-diubiquitin) and ISG15 (Klemm et al., 2020). The MERS-CoV (space group P6<sub>5</sub>22) PL<sup>Pro</sup>-ubiquitin complex was solved by Bailey-Elkin et al. and in their paper, they refer to this structure as the closed conformation since the fingers domain is shifted toward the ubiquitin (Bailey-Elkin et al., 2014).

When ubiquitin was present in the SARS-CoV-2, SARS-CoV and MERS-CoV PL<sup>Pro</sup> structures, the results from Schrödinger and blind docking showed that compounds were displaced from the naphthalene-inhibitor binding pocket (Figures 3, 4 and Supplementary Figures 1, 2). Compared to the apo PL<sup>Pro</sup>, hypericin was unable to produce molecular docking poses for the PL<sup>Pro</sup>-ubiquitin complexes. Likewise, the protein-protein docking results with the ligands already bound to this region in PL<sup>Pro</sup> revealed that ubiquitin was binding in different conformations and that the position of the C-terminus was altered (Figure 5). This suggests that the dietary compounds may be able to interfere with the deubiquitinase activity of PL<sup>Pro</sup> and in terms of *in silico* methods, this can be evaluated further using molecular dynamics (MD) simulations.

A commercially available PL<sup>Pro</sup> enzymatic assay was used to measure deubiquitinase activity using GRL-0617 as an internal positive control; in this specific assay GRL-0617 has been shown to have an IC<sub>50</sub> value of 1.7 μM for inhibition of PL<sup>Pro</sup> deubiquitinase activity (BP Bioscience). Overall, our findings indicated inhibition with an order of potency of GRL-0617 and hypericin > rutin and cyanidin-3-O-glucoside > epigallocatechin gallate and epicatechin gallate >> cefotaxime. GRL-0617 and hypericin > rutin > cyanidin-3-O-glucoside and epicatechin gallate > cefotaxime and epigallocatechin gallate (Figure 6). The potent inhibition of PL<sup>Pro</sup> deubiquitinase activity by hypericin which, at higher concentrations, was analogous to GRL-0617, is particularly encouraging. Hypericin is an anthraquinone derivative that

can be found in the flowering plant *Hypericum perforatum*, which is also commonly known as St. John's Wort (Napoli et al., 2018). It has been identified as a lead compound for the SARS-CoV-2 spike protein and its antiviral properties have been the subject of numerous papers in the past (Jacobson et al., 2001; Shih et al., 2018; Chen et al., 2019). In addition to its antiviral effects, St. John's Wort is also being investigated for its antidepressant properties and synthetic hypericin (SGX301) has gained attention for its use as a photodynamic agent in the treatment of cutaneous T-cell lymphoma (Rook et al., 2010; Montoya et al., 2015; Apaydin et al., 2016).

While specific compounds were selected for use in this study, it would be important to expand this in the future to incorporate a greater number of phytochemicals that are present in various plant extracts. Network pharmacology is also being increasingly used in drug discovery and this systematic approach can assist with identifying potential protein targets and lead compounds, as well as understanding their mechanisms of action (Zhang et al., 2019; Pan H. D. et al., 2020). Nonetheless, the antiviral, antioxidant and anti-inflammatory properties of the compounds used in this study had been previously reported in the literature and were consequently suitable candidates. Molecular docking was used for virtual screening and although the scoring functions produced from docking aren't absolute binding energies, it allowed for predictions to be made about the protein-ligand interactions. Docking was performed using the Glide (XP) protocol of the Schrödinger Suite and in a study conducted by Wang et al., this was found to have a 90% success rate in identifying the correct binding poses of ligands (Wang et al., 2016). In this study, the inhibitory activities of the compounds were subsequently measured using an enzymatic activity assay. *In silico* tools are currently being utilized for the early stages of the drug discovery pipeline however, it is important to note that the pipeline involves multiple steps and is a time consuming process (Agostino et al., 2019). Potential drugs must be explored further in pre-clinical trials using a combination of techniques and clinical trials (Agostino et al., 2019).

## CONCLUSION

Overall, on the basis of our *in silico* and *in vitro* evaluations, hypericin, rutin, and cyanidin-3-O-glucoside can be considered potential lead compounds. In particular, further clarification of the molecular mechanisms and antiviral properties of hypericin, which displayed high potency in the *in vitro* assay and favorable binding properties in the *in silico* studies, is warranted.

## DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding author/s.

## AUTHOR CONTRIBUTIONS

TK and AH conceptualized the aims and methodology and were involved in supervision. TK was involved in the production of the first draft of the manuscript. EP performed data analysis, data curation, and was involved in production of the first draft of the manuscript. JL was involved in data analysis and curation and was involved in production of the first draft of the manuscript. KV performed formal data analysis and was involved in data curation. KL performed formal data analysis and validation. All authors contributed to editing and reviewing the manuscript.

## ACKNOWLEDGMENTS

We would like to acknowledge intellectual and financial support by McCord Research (Iowa, USA). JL was supported by an Australian Government Research Training Program Scholarship. We are indebted to Alfonso Perez Escudero and the team at Crowdfight COVID-19 for enabling access to supercomputing facilities, and to Matthew Gasperetti and the team at Hypernet Labs; Galileo, for enabling cloud computing for this project. We thank the National Computing Infrastructure (NCI), and the Pawsey Supercomputing Centre in Australia (funded by the Australian Government). Further, we thank the Spartan High Performance Computing service (University of Melbourne), and the Partnership for Advanced Computing in Europe (PRACE) for awarding the access to Piz Daint, hosted at the Swiss National Supercomputing Centre (CSCS), Switzerland.

## REFERENCES

- Abdelrahman, Z., Li, M., and Wang, X. (2020). Comparative review of SARS-CoV-2, SARS-CoV, MERS-CoV, and influenza a respiratory viruses. *Front. Immunol.* 11:2309. doi: 10.3389/fimmu.2020.552909
- Agostino, B., Gabriele, C., Luca, S., and Marco, R. (2019). The *in silico* drug discovery toolbox: applications in lead discovery and optimization. *Curr. Med. Chem.* 26, 3838–3873. doi: 10.2174/0929867324666171107101035
- Alamri, M. A., Tahir ul Qamar, M., Mirza, M. U., Alqahtani, S. M., Froeyen, M., and Chen, L. L. (in press). Discovery of human coronaviruses pan-papain-like protease inhibitors using computational approaches. *J. Pharm. Anal.* doi: 10.1016/j.jpah.2020.08.012
- Apaydin, E. A., Maher, A. R., Shanman, R., Booth, M. S., Miles, J. N. V., Sorbero, M. E., et al. (2016). A systematic review of St. John's wort for major depressive disorder. *Syst. Rev.* 5:12. doi: 10.1186/s13643-016-0325-2
- Báez-Santos, Y. M., Barraza, S. J., Wilson, M. W., Agius, M. P., Mielech, A. M., Davis, N. M., et al. (2014). X-ray structural and biological evaluation of a series of potent and highly selective inhibitors of human coronavirus papain-like proteases. *J. Med. Chem.* 57, 2393–2412. doi: 10.1021/jm401712t
- Báez-Santos, Y. M., St John, S. E., and Mesecar, A. D. (2015). The SARS-coronavirus papain-like protease: structure, function and inhibition by designed antiviral compounds. *Antiviral Res.* 115, 21–38. doi: 10.1016/j.antiviral.2014.12.015
- Bailey-Elkin, B. A., Knaap, R. C. M., Johnson, G. G., Dalebout, T. J., Ninaber, D. K., van Kasteren, P. B., et al. (2014). Crystal structure of the Middle East respiratory syndrome coronavirus (MERS-CoV) papain-like protease bound to ubiquitin facilitates targeted disruption of deubiquitinating activity

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fchem.2020.623971/full#supplementary-material>

**Supplementary Figure 1** | Blind docking of the naphthalene-based inhibitors and dietary compounds to the SARS-CoV PL<sup>pro</sup>.

**Supplementary Figure 2** | Blind docking of the naphthalene-based inhibitors and dietary compounds to the MERS-CoV PL<sup>pro</sup>.

**Supplementary Figure 3** | Molecular docking results of the naphthalene-based inhibitors and dietary compounds for the SARS-CoV-2 PL<sup>pro</sup> structure 6w9c.

**Supplementary Figure 4** | Molecular docking results of the naphthalene-based inhibitors and dietary compounds for the SARS-CoV-2 PL<sup>pro</sup> structure 6wu4.

**Supplementary Figure 5** | Molecular docking results of the naphthalene-based inhibitors and dietary compounds for the SARS-CoV-2 PL<sup>pro</sup> structure 6wx4.

**Supplementary Figure 6** | Molecular docking results of the naphthalene-based inhibitors and dietary compounds for the SARS-CoV-2 PL<sup>pro</sup> structure 7jrn.

**Supplementary Table 1** | PrankWeb conservation analysis results for the SARS-CoV-2 (6xaa), SARS-CoV (4mm3), and MERS-CoV PL<sup>pro</sup> (4rf0).

**Supplementary Table 2** | Blind docking of the naphthalene-based inhibitors and dietary compounds to the SARS-CoV-2 (6xaa), SARS-CoV (4mm3), and MERS-CoV (4rf0) PL<sup>pro</sup> structures.

**Supplementary Table 3** | Blind docking of the naphthalene-based inhibitors and dietary compounds to the SARS-CoV-2 6w9c, 6wu4, 6wx4, and 7jrn PL<sup>pro</sup> crystal structures. No poses were found in the naphthalene inhibitor site for GRL-0617 and hypericin in the 6wu4 structure.

**Supplementary Table 4** | The GlideScores (kcal/mol) of the naphthalene-based inhibitors and dietary compounds for the SARS-CoV-2, SARS-CoV, and MERS-CoV crystal structures with ubiquitin are listed.

**Supplementary Table 5** | The docking scores and RMSD values for the top-ranked PL<sup>pro</sup>-ubiquitin complexes from the HDock server are provided.

- to demonstrate its role in innate immune suppression. *J. Biol. Chem.* 289, 34667–34682. doi: 10.1074/jbc.M114.609644
- Bar-Zeev, N., and Moss, W. J. (2020). Encouraging results from phase 1/2 COVID-19 vaccine trials. *Lancet* 396, 448–449. doi: 10.1016/S0140-6736(20)31611-1
- Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., et al. (2000). The protein data bank. *Nucleic Acids Res.* 28, 235–242. doi: 10.1093/nar/28.1.235
- Bhowmik, D., Nandi, R., and Kumar, D. (2020). Evaluation of flavonoids as 2019-nCoV cell entry inhibitor through molecular docking and pharmacological analysis. *ChemRxiv [Preprint]*. doi: 10.26434/chemrxiv.12071508
- Bonvino, N. P., Liang, J., McCord, E. D., Zafiris, E., Benetti, N., Ray, N. B., et al. (2018). OliveNet<sup>TM</sup>: a comprehensive library of compounds from *Olea europaea*. *Database* 2018:bay016. doi: 10.1093/database/bay016
- Bosken, Y. K., Cholko, T., Lou, Y. C., Wu, K. P., and Chang C. E. A. (2020). Insights into dynamics of inhibitor and ubiquitin-like protein binding in SARS-CoV-2 papain-like protease. *Front. Mol. Biosci.* 7:174. doi: 10.3389/fmolb.2020.00174
- Chen, H., Muhammad, I., Zhang, Y., Ren, Y., Zhang, R., Huang, X., et al. (2019). Antiviral activity against infectious bronchitis virus and bioactive components of *Hypericum Perforatum* L. *Front. Pharmacol.* 10:1272. doi: 10.3389/fphar.2019.01272
- Chen, L., Hu, C., Hood, M., Zhang, X., Zhang, L., Kan, J., et al. (2020). A novel combination of vitamin C, curcumin and glycyrrhizic acid potentially regulates immune and inflammatory response associated with coronavirus infections: a perspective from system biology analysis. *Nutrients* 12:1193. doi: 10.3390/nu12041193
- Chikhale, R. V., Gurav, S. S., Patil, R. B., Sinha, S. K., Prasad, S. K., Shakya, A., et al. (2020). Sars-cov-2 host entry and replication inhibitors from Indian ginseng: an *in-silico* approach.

- J. *Biomol. Struct. Dyn.* 2020, 1–12. doi: 10.1080/07391102.2020.1778539
- Ciechanover, A., Orian, A., and Schwartz, A. L. (2000). Ubiquitin-mediated proteolysis: biological regulation via destruction. *Bioessays* 22, 442–451. doi: 10.1002/(SICI)1521-1878(200005)22:5<442::AID-BIES6>3.0.CO;2-Q
- Coronaviridae Study Group of the International Committee on Taxonomy of Viruses (2020). The species severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. *Nat. Microbiol.* 5, 536–544. doi: 10.1038/s41564-020-0695-z
- Dallakyan, S., and Olson, A. J. (2015). “Small-molecule library screening by docking with PyRx,” in *Chemical Biology: Methods and Protocols*, eds J. E. Hempel, C. H. Williams, and C. C. Hong (New York, NY: Springer New York), 243–250.
- de la Rica, R., Borges, M., and Gonzalez-Freire, M. (2020). COVID-19: in the eye of the cytokine storm. *Front. Immunol.* 11:2313. doi: 10.3389/fimmu.2020.558898
- Dhama, K., Khan, S., Tiwari, R., Sircar, S., Bhat, S., Malik, Y. S., et al. (2020). Coronavirus disease 2019–COVID-19. *Clin. Microbiol. Rev.* 33:e00028-20. doi: 10.1128/CMR.00028-20
- Divya, M., Vijayakumar, S., Chen, J., Vaseeharan, B., and Durán-Lara, E. F. (2020). South Indian medicinal plants can combat deadly viruses along with COVID-19? - a review. *Microb. Pathog.* 2020:104277. doi: 10.1016/j.micpath.2020.104277
- Donma, M. M., and Donma, O. (2020). The effects of allium sativum on immunity within the scope of COVID-19 infection. *Med. Hypotheses* 144:109934. doi: 10.1016/j.mehy.2020.109934
- Erpapazoglou, Z., Walker, O., and Haguenaue-Tsapis, R. (2014). Versatile roles of k63-linked ubiquitin chains in trafficking. *Cells* 3, 1027–1088. doi: 10.3390/cells3041027
- FDA (2020). *Approves First Treatment For Covid-19*.
- Folegatti, P. M., Ewer, K. J., Aley, P. K., Angus, B., Becker, S., Belij-Rammerstorfer, S., et al. (2020). Safety and immunogenicity of the ChAdOx1 nCoV-19 vaccine against SARS-CoV-2: a preliminary report of a phase 1/2, single-blind, randomised controlled trial. *Lancet* 396, 467–478. doi: 10.1016/S0140-6736(20)31604-4
- Freitas, B. T., Durie, I. A., Murray, J., Longo, J. E., Miller, H. C., Crich, D., et al. (2020). Characterization and noncovalent inhibition of the deubiquitinase and deISGylase activity of SARS-CoV-2 papain-like protease. *ACS Infect. Dis.* 6, 2099–2109. doi: 10.1021/acinfeddis.0c00168
- Friesner, R. A., Banks, J. L., Murphy, R. B., Halgren, T. A., Klicic, J. J., Mainz, D. T., et al. (2004). Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J. Med. Chem.* 47, 1739–1749. doi: 10.1021/jm0306430
- Friesner, R. A., Murphy, R. B., Repasky, M. P., Frye, L. L., Greenwood, J. R., Halgren, T. A., et al. (2006). Extra precision glide: docking and scoring incorporating a model of hydrophobic enclosure for protein–ligand complexes. *J. Med. Chem.* 49, 6177–6196. doi: 10.1021/jm051256o
- Gao, X., Qin, B., Chen, P., Zhu, K., Hou, P., Wojdyla, J. A., et al. (in press). Crystal structure of SARS-CoV-2 papain-like protease. *Acta Pharmaceut. Sin. B*. doi: 10.1016/j.apsb.2020.08.014
- Ghosh, R., Chakraborty, A., Biswas, A., and Chowdhuri, S. (2020). Evaluation of green tea polyphenols as novel corona virus (SARS CoV-2) main protease (Mpro) inhibitors - an in silico docking and molecular dynamics simulation study. *J. Biomol. Struct. Dyn.* 2020, 1–13. doi: 10.1080/07391102.2020.1779818
- Gupta, S., Singh, A. K., Kushwaha, P. P., Prajapati, K. S., Shuaib, M., Senapati, S., et al. (2020). Identification of potential natural inhibitors of SARS-CoV2 main protease by molecular docking and simulation studies. *J. Biomol. Struct. Dyn.* 2020, 1–12. doi: 10.1080/07391102.2020.1776157
- Halgren, T. A., Murphy, R. B., Friesner, R. A., Beard, H. S., Frye, L. L., Pollard, W. T., et al. (2004). Glide: a new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening. *J. Med. Chem.* 47, 1750–1759. doi: 10.1021/jm030644s
- Harder, E., Damm, W., Maple, J., Wu, C., Reboul, M., Xiang, J. Y., et al. (2016). OPLS3: a force field providing broad coverage of drug-like small molecules and proteins. *J. Chem. Theory Comput.* 12, 281–296. doi: 10.1021/acs.jctc.5b00864
- Heaton, S. M., Borg, N. A., and Dixit, V. M. (2015). Ubiquitin in the activation and attenuation of innate antiviral immunity. *J. Exp. Med.* 213, 1–13. doi: 10.1084/jem.20151531
- Huang, Y., Dai, H., and Ke, R. (2019). Principles of effective and robust innate immune response to viral infections: a multiplex network analysis. *Front. Immunol.* 10:1736. doi: 10.3389/fimmu.2019.01736
- Hypernet Labs and Galileo (2020). Available online at: <https://galileoapp.io/>
- Jacobson, J. M., Feinman, L., Liebes, L., Ostrow, N., Koslowski, V., Tobia, A., et al. (2001). Pharmacokinetics, safety, and antiviral effects of hypericin, a derivative of st. john's wort plant, in patients with chronic hepatitis C virus infection. *Antimicrob. Agents Chemother.* 45, 517–524. doi: 10.1128/AAC.45.2.517-524.2001
- Jena, A., Kanungo, N., Nayak, V., Chainy, G. B. N., and Dandapat, J. (2020). Catechin and curcumin interact with corona (2019-nCoV/SARS-CoV2) viral S protein and ACE2 of human cell insights from computational study and implication for intervention. *Res. Square [Preprint]*. doi: 10.21203/rs.3.rs-22057/v1
- Jendele, L., Krivak, R., Skoda, P., Novotny, M., and Hoksza, D. (2019). PrankWeb: a web server for ligand binding site prediction and visualization. *Nucleic Acids Res.* 47, W345–W349. doi: 10.1093/nar/gkz424
- Jeong, G. U., Song, H., Yoon, G. Y., Kim, D., and Kwon, Y. C. (2020). Therapeutic strategies against COVID-19 and structural characterization of SARS-CoV-2: a review. *Front. Microbiol.* 11:1723. doi: 10.3389/fmicb.2020.01723
- Jiang, X., and Chen, Z. J. (2011). The role of ubiquitylation in immune defence and pathogen evasion. *Nat. Rev. Immunol.* 12, 35–48. doi: 10.1038/nri3111
- Jorgensen, W. L., Maxwell, D. S., and Tirado-Rives, J. (1996). Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *J. Am. Chem. Soc.* 118, 11225–11236. doi: 10.1021/ja962176o
- Jorgensen, W. L., and Tirado-Rives, J. (1988). The OPLS [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin. *J. Am. Chem. Soc.* 110, 1657–1666. doi: 10.1021/ja00214a001
- Kim, S., Chen, J., Cheng, T., Gindulyte, A., He, J., He, S., et al. (2019). PubChem 2019 update: improved access to chemical data. *Nucleic Acids Res.* 47, D1102–D1109. doi: 10.1093/nar/gky1033
- Klemm, T., Ebert, G., Calleja, D. J., Allison, C. C., Richardson, L. W., Bernardini, J. P., et al. (2020). Mechanism and inhibition of the papain-like protease, PLP<sup>pro</sup>, of SARS-CoV-2. *EMBO J.* 39:e106275. doi: 10.15252/emboj.2020106275
- Komander, D., and Rape, M. (2012). The ubiquitin code. *Annu. Rev. Biochem.* 81, 203–229. doi: 10.1146/annurev-biochem-060310-170328
- Krupanidhi, S., Abraham Peele, K., Venkateswarulu, T. C., Ayyagari, V. S., Nazneen Bobby, M., John Babu, D., et al. (2020). Screening of phytochemical compounds of tinospora cordifolia for their inhibitory activity on SARS-CoV-2: an in silico study. *J. Biomol. Struct. Dyn.* 2020, 1–5. doi: 10.1080/07391102.2020.1787226
- Kumari, P., and Kumar, H. (2018). Viral deubiquitinases: role in evasion of anti-viral innate immunity. *Crit. Rev. Microbiol.* 44, 304–317. doi: 10.1080/1040841X.2017.1368999
- Lee, H., Lei, H., Santarsiero, B. D., Gatuz, J. L., Cao, S., Rice, A. J., et al. (2015). Inhibitor recognition specificity of MERS-CoV papain-like protease may differ from that of SARS-CoV. *ACS Chem. Biol.* 10, 1456–1465. doi: 10.1021/cb500917m
- Li, J., Chai, Q. Y., and Liu, C. H. (2016). The ubiquitin system: a critical regulator of innate immunity and pathogen–host interactions. *Cell. Mol. Immunol.* 13, 560–576. doi: 10.1038/cmi.2016.40
- Li, S. Y., Chen, C., Zhang, H. Q., Guo, H. Y., Wang, H., Wang, L., et al. (2005). Identification of natural compounds with antiviral activities against SARS-associated coronavirus. *Antiviral Res.* 67, 18–23. doi: 10.1016/j.antiviral.2005.02.007
- Liang, J., Pitsillou, E., Karagiannis, C., Darmawan, K. K., Ng, K., Hung, A., et al. (2020). Interaction of the prototypical alpha-ketoamide inhibitor with the SARS-CoV-2 main protease active site in silico: molecular dynamic simulations highlight the stability of the ligand–protein complex. *Comput. Biol. Chem.* 87:107292. doi: 10.1016/j.compbiolchem.2020.107292
- Liu, J., Qian, C., and Cao, X. (2016). Post-translational modification control of innate immunity. *Immunity* 45, 15–30. doi: 10.1016/j.immuni.2016.06.020
- Madhavi Sastry, G., Adzhigirey, M., Day, T., Annabhimoju, R., and Sherman, W. (2013). Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments. *J. Comput. Aided Mol. Des.* 27, 221–234. doi: 10.1007/s10822-013-9644-8

- Mani, J. S., Johnson, J. B., Steel, J. C., Broszczak, D. A., Neilsen, P. M., Walsh, K. B., et al. (2020). Natural product-derived phytochemicals as potential agents against coronaviruses: a review. *Virus Res.* 284:197989. doi: 10.1016/j.virusres.2020.197989
- McClain, C. B., and Vabret, N. (2020). SARS-CoV-2: the many pros of targeting PLpro. *Signal Transduct Target Ther.* 5:223. doi: 10.1038/s41392-020-00335-z
- Mohammadi Pour, P., Fakhri, S., Asgary, S., Farzaei, M. H., and Echeverria, J. (2019). The signaling pathways, and therapeutic targets of antiviral agents: focusing on the antiviral approaches and clinical perspectives of anthocyanins in the management of viral diseases. *Front. Pharmacol.* 10:1207. doi: 10.3389/fphar.2019.01207
- Montoya, A., Daza, A., Muñoz, D., Ríos, K., Taylor, V., Cedeño, D., et al. (2015). Development of a novel formulation with hypericin to treat cutaneous leishmaniasis based on photodynamic therapy in *in vitro* and *in vivo* studies. *Antimicrob. Agents Chemother.* 59, 5804–5813. doi: 10.1128/AAC.00545-15
- Muhseen, Z. T., Hameed, A. R., Al-Hasani, H. M. H., Tahir ul Qamar, M., and Li, G. (2020). Promising terpenes as SARS-CoV-2 spike receptor-binding domain (RBD) attachment inhibitors to the human ACE2 receptor: integrated computational approach. *J. Mol. Liq.* 320:114493. doi: 10.1016/j.molliq.2020.114493
- Napoli, E., Siracusa, L., Ruberto, G., Carrubba, A., Lazzara, S., Speciale, A., et al. (2018). Phytochemical profiles, phototoxic and antioxidant properties of eleven hypericum species – a comparative study. *Phytochemistry* 152, 162–173. doi: 10.1016/j.phytochem.2018.05.003
- Nelemans, T., and Kikkert, M. (2019). Viral innate immune evasion and the pathogenesis of emerging RNA virus infections. *Viruses* 11:961. doi: 10.3390/v11100961
- Ngwa, W., Kumar, R., Thompson, D., Lyerly, W., Moore, R., Reid, T. E., et al. (2020). Potential of flavonoid-inspired phytomedicines against COVID-19. *Molecules* 25:2707. doi: 10.3390/molecules25112707
- Ohtake, F., Tsuchiya, H., Saeki, Y., and Tanaka, K. (2018). K63 ubiquitylation triggers proteasomal degradation by seeding branched ubiquitin chains. *Proc. Natl. Acad. Sci. U.S.A.* 115:E1401. doi: 10.1073/pnas.1716673115
- Ojha, P. K., Kar, S., Krishna, J. G., Roy, K., and Leszczynski, J. (2020). Therapeutics for COVID-19: from computation to practices—where, where we are heading to. *Mol. Divers.* doi: 10.1007/s11030-020-10134-x. [Epub ahead of print].
- Osiptuk, J., Jedrzejczak, R., Tesar, C., Endres, M., Stols, L., Babnigg, G., et al. (2020). The crystal structure of papain-like protease of SARS CoV-2. 6W9C. doi: 10.2210/pdb6W9C/pdb
- Pan, H., Peto, R., Karim, Q. A., Alejandria, M., Henao-Restrepo, A. M., García, C. H., et al. (2020). Repurposed antiviral drugs for COVID-19 –interim WHO SOLIDARITY trial results. *medRxiv [Preprint]*. doi: 10.1101/2020.10.15.20209817
- Pan, H. D., Yao, X. J., Wang, W. Y., Lau, H. Y., and Liu, L. (2020). Network pharmacological approach for elucidating the mechanisms of traditional Chinese medicine in treating COVID-19 patients. *Pharmacol. Res.* 159:105043. doi: 10.1016/j.phrs.2020.105043
- Pandey, A., Nikam, A. N., Shreya, A. B., Mutalik, S. P., Gopalan, D., Kulkarni, S., et al. (2020). Potential therapeutic targets for combating SARS-CoV-2: drug repurposing, clinical trials and recent advancements. *Life Sci.* 256:117883. doi: 10.1016/j.lfs.2020.117883
- Panier, S., and Durocher, D. (2009). Regulatory ubiquitylation in response to DNA double-strand breaks. *DNA Rep.* 8, 436–443. doi: 10.1016/j.dnarep.2009.01.013
- Panyod, S., Ho, C. T., and Sheen, L. Y. (2020). Dietary therapy and herbal medicine for COVID-19 prevention: a review and perspective. *J. Tradition. Complement. Med.* 10, 420–427. doi: 10.1016/j.jtcm.2020.05.004
- Perng, Y. C., and Lenschow, D. J. (2018). ISG15 in antiviral immunity and beyond. *Nat. Rev. Microbiol.* 16, 423–39. doi: 10.1038/s41579-018-0020-5
- Petersen, E., Koopmans, M., Go, U., Hamer, D. H., Petrosillo, N., Castelli, F., et al. (2020). Comparing SARS-CoV-2 with SARS-CoV and influenza pandemics. *Lancet Infect. Dis.* 20, e238–e244. doi: 10.1016/S1473-3099(20)30484-9
- Pickart, C. M. (2001). Mechanisms underlying ubiquitination. *Annu. Rev. Biochem.* 70, 503–533. doi: 10.1146/annurev.biochem.70.1.503
- Pitsillou, E., Liang, J., Karagiannis, C., Ververis, K., Darmawan, K. K., Ng, K., et al. (2020). Interaction of small molecules with the SARS-CoV-2 main protease *in silico* and *in vitro* validation of potential lead compounds using an enzyme-linked immunosorbent assay. *Comput. Biol. Chem.* 2020:107408. doi: 10.1016/j.compbiolchem.2020.107408
- Ratia, K., Kilianski, A., Baez-Santos, Y. M., Baker, S. C., and Mesecar, A. (2014). Structural basis for the ubiquitin-linkage specificity and deISGylating activity of SARS-CoV papain-like protease. *PLoS Pathog.* 10:e1004113. doi: 10.1371/journal.ppat.1004113
- Ratia, K., Pegan, S., Takayama, J., Sleeman, K., Coughlin, M., Baliji, S., et al. (2008). A noncovalent class of papain-like protease/deubiquitinase inhibitors blocks SARS virus replication. *Proc. Natl. Acad. Sci. U.S.A.* 105:16119. doi: 10.1073/pnas.0805240105
- Rook, A. H., Wood, G. S., Duvic, M., Vonderheid, E. C., Tobia, A., and Cabana, B. (2010). A phase II placebo-controlled study of photodynamic therapy with topical hypericin and visible light irradiation in the treatment of cutaneous T-cell lymphoma and psoriasis. *J. Am. Acad. Dermatol.* 63, 984–990. doi: 10.1016/j.jaad.2010.02.039
- Russo, M., Moccia, S., Spagnuolo, C., Tedesco, I., and Russo, G. L. (2020). Roles of flavonoids against coronavirus infection. *Chem. Biol. Interact.* 328:109211. doi: 10.1016/j.cbi.2020.109211
- Rut, W., Lv, Z., Zmudzinski, M., Patchett, S., Nayak, D., Snipas, S. J., et al. (2020). Activity profiling and crystal structures of inhibitor-bound SARS-CoV-2 papain-like protease: a framework for anti-COVID-19 drug design. *Sci. Adv.* 6:eabd4596. doi: 10.1126/sciadv.abd4596
- Sacco, M., Ma, C., Wang, J., and Chen, Y. (2020). Crystal structure of the native SARS-CoV-2 papain-like protease (PL<sup>Pro</sup>) with inhibitor GRL0617. *7JRN*. doi: 10.2210/pdb7JRN/pdb
- Schrödinger (2020a). *Schrödinger Release 2020-2: Maestro*. New York, NY: Schrödinger, LLC.
- Schrödinger (2020b). *Schrödinger Release 2020-2: LigPrep*. New York, NY: Schrödinger, LLC.
- Schrödinger (2020c). *Schrödinger Release 2020-2: Glide*. New York, NY: Schrödinger, LLC.
- Shih, C. M., Wu, C. H., Wu, W. J., Hsiao, Y. M., and Ko, J. L. (2018). Hypericin inhibits hepatitis C virus replication via deacetylation and down-regulation of heme oxygenase-1. *Phytomedicine* 46, 193–198. doi: 10.1016/j.phymed.2017.08.009
- Shin, D., Mukherjee, R., Grewe, D., Bojkova, D., Baek, K., Bhattacharya, A., et al. (2020). Papain-like protease regulates SARS-CoV-2 viral spread and innate immunity. *Nature* 587, 657–662. doi: 10.1038/s41586-020-2601-5
- Shivakumar, D., Williams, J., Wu, Y., Damm, W., Shelley, J., and Sherman, W. (2010). Prediction of absolute solvation free energies using molecular dynamics free energy perturbation and the OPLS force field. *J. Chem. Theory Comput.* 6, 1509–1519. doi: 10.1021/ct900587b
- Sinha, S. K., Prasad, S. K., Islam, M. A., Gurav, S. S., Patil, R. B., AlFaris, N. A., et al. (2020). Identification of bioactive compounds from glycyrrhiza glabra as possible inhibitor of SARS-CoV-2 spike glycoprotein and non-structural protein-15: a pharmacoinformatics study. *J. Biomol. Struct. Dyn.* 2020, 1–15. doi: 10.1080/07391102.2020.1779132
- Smith, M., and Smith, J. (2020). Repurposing therapeutics for COVID-19: supercomputer-based docking to the SARS-CoV-2 viral spike protein and viral spike protein-human ACE2 interface. *ChemRxiv [Preprint]*. doi: 10.26434/chemrxiv.11871402
- Subbaiyan, A., Ravichandran, K., Singh, S., Sankar, M., Thomas, P., Dhama, K., et al. (2020). *in silico* molecular docking analysis of targeting SARS-CoV-2 spike protein and selected herbal constituents. *J. Pure Appl. Microbiol.* 14, 775–788. doi: 10.22207/JJPM.14.SPL1.37
- Tahir Ul Qamar, M., Alqahtani, S. M., Alamri, M. A., and Chen, L. L. (2020). Structural basis of SARS-CoV-2 3CL(pro) and anti-COVID-19 drug discovery from medicinal plants. *J. Pharm. Anal.* 10, 313–319. doi: 10.1016/j.jpha.2020.03.009
- Tiwari, R., Latheef, S., Ahmed, I., Iqbal, H., Bule, M., Dhama, K., et al. (2018). Herbal immunomodulators, a remedial panacea for the designing and developing effective drugs and medicines: current scenario and future prospects. *Curr. Drug Metab.* 19, 264–301. doi: 10.2174/1389200219666180129125436
- Tripathi, M. K., Singh, P., Sharma, S., Singh, T. P., Ethayathulla, A. S., and Kaur, P. (2020). Identification of bioactive molecule from Withania somnifera (Ashwagandha) as SARS-CoV-2 main protease inhibitor. *J. Biomol. Struct. Dyn.* 2020, 1–14. doi: 10.1080/07391102.2020.1790425
- Trott, O., and Olson, A. J. (2010). AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization,

- and multithreading. *J. Comput. Chem.* 31, 455–461. doi: 10.1002/jcc.21334
- Verma, S., Twilley, D., Esmear, T., Oosthuizen, C. B., Reid, A. M., Nel, M., et al. (2020). Anti-SARS-CoV natural products with the potential to inhibit SARS-CoV-2 (COVID-19). *Front. Pharmacol.* 11:1514. doi: 10.3389/fphar.2020.561334
- Wang, Z., Sun, H., Yao, X., Li, D., Xu, L., Li, Y., et al. (2016). Comprehensive evaluation of ten docking programs on a diverse set of protein–ligand complexes: the prediction accuracy of sampling power and scoring power. *Phys. Chem. Chem. Phys.* 18, 12964–12975. doi: 10.1039/C6CP01555G
- World Health Organization (2020). *Coronavirus Disease 2019 (COVID-19): Situation Report 51*.
- Yan, Y., Tao, H., He, J., and Huang, S. Y. (2020). The HDock server for integrated protein–protein docking. *Nat. Protoc.* 15, 1829–1852. doi: 10.1038/s41596-020-0312-x
- Yan, Y., Zhang, D., Zhou, P., Li, B., and Huang, S. Y. (2017). HDock: a web server for protein–protein and protein–DNA/RNA docking based on a hybrid strategy. *Nucleic Acids Res.* 45, W365–W373. doi: 10.1093/nar/gkx407
- Zhang, R., Zhu, X., Bai, H., and Ning, K. (2019). Network pharmacology databases for traditional chinese medicine: review and assessment. *Front. Pharmacol.* 10:123. doi: 10.3389/fphar.2019.00123
- Zhou, Y., Hou, Y., Shen, J., Huang, Y., Martin, W., and Cheng, F. (2020). Network-based drug repurposing for novel coronavirus 2019-nCoV/SARS-CoV-2. *Cell Discov.* 6:14. doi: 10.1038/s41421-020-0153-3
- Zhou, Z., Zhu, X., Yin, R., Liu, T., Yang, S., Zhou, L., et al. (2020). K63 ubiquitin chains target NLRP3 inflammasome for autophagic degradation in ox-LDL-stimulated THP-1 macrophages. *Aging* 12, 1747–1759. doi: 10.18632/aging.102710
- Conflict of Interest:** Epigenomic Medicine Program (TCK) is supported financially by McCord Research (Iowa, USA), which has a financial interest in dietary compounds described in this work. However, there is no conflict of interest with respect to the inhibition of the SARS-CoV-2 papain-like protease.
- The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.
- Copyright © 2020 Pitsillou, Liang, Ververis, Lim, Hung and Karagiannis. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Host DDX Helicases as Possible SARS-CoV-2 Proviral Factors: A Structural Overview of Their Hijacking Through Multiple Viral Proteins

Flavia Squeglia<sup>1\*</sup>, Maria Romano<sup>1†</sup>, Alessia Ruggiero<sup>1†</sup>, Giovanni Maga<sup>2</sup> and Rita Berisio<sup>1\*</sup>

<sup>1</sup> Institute of Biostructures and Bioimaging (IBB-CNR), Naples, Italy, <sup>2</sup> Institute of Molecular Genetics (IGM-CNR), Pavia, Italy

## OPEN ACCESS

### Edited by:

Sonia Di Gaetano,  
Italian National Research Council, Italy

### Reviewed by:

Zufeng Guo,  
Johns Hopkins University,  
United States  
Kathleen Boris-Lawrie,  
University of Minnesota Twin Cities,  
United States

### \*Correspondence:

Flavia Squeglia  
flavia.squeglia@cnr.it  
Rita Berisio  
rita.berisio@cnr.it

<sup>†</sup>These authors have contributed  
equally to this work

### Specialty section:

This article was submitted to  
Chemical Biology,  
a section of the journal  
Frontiers in Chemistry

**Received:** 02 September 2020

**Accepted:** 10 November 2020

**Published:** 10 December 2020

### Citation:

Squeglia F, Romano M, Ruggiero A,  
Maga G and Berisio R (2020) Host  
DDX Helicases as Possible  
SARS-CoV-2 Proviral Factors: A  
Structural Overview of Their Hijacking  
Through Multiple Viral Proteins.  
Front. Chem. 8:602162.  
doi: 10.3389/fchem.2020.602162

As intracellular parasites, viruses hijack the host cell metabolic machinery for their replication. Among other cellular proteins, the DEAD-box (DDX) RNA helicases have been shown to be hijacked by coronaviruses and to participate in essential DDX-mediated viral replication steps. Human DDX RNA helicases play essential roles in a broad array of biological processes and serve multiple roles at the virus-host interface. The viral proteins responsible for DDX interactions are highly conserved among coronaviruses, suggesting that they might also play conserved functions in the SARS-CoV-2 replication cycle. In this review, we provide an update of the structural and functional data of DDX as possible key factors involved in SARS-CoV-2 hijacking mechanisms. We also attempt to fill the existing gaps in the available structural information through homology modeling. Based on this information, we propose possible paths exploited by the virus to replicate more efficiently by taking advantage of host DDX proteins. As a general rule, sequestration of DDX helicases by SARS-CoV-2 is expected to play a pro-viral role in two ways: by enhancing key steps of the virus life cycle and, at the same time, by suppressing the host innate immune response.

**Keywords:** SARS-CoV-2, COVID19, protein structure, viral infection, DDX helicases

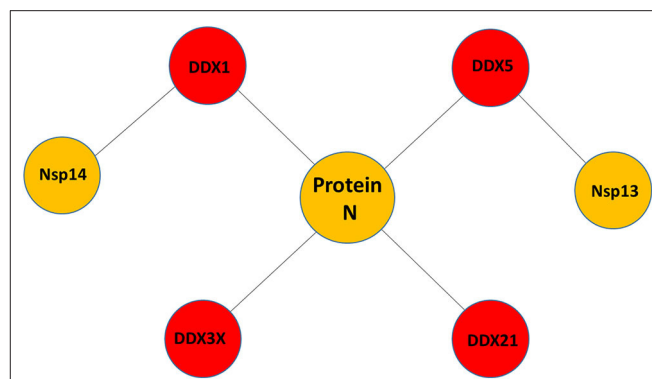
## INTRODUCTION

COVID-19 is a respiratory disease caused by a novel enveloped, positive-sense, single-stranded RNA betacoronavirus, designated as SARS-CoV-2. Mechanistically, SARS-CoV-2 enters the cell through the binding of the spike protein to the ACE2 receptors, as previously observed for SARS-CoV (Luan et al., 2020; Romano et al., 2020a; Wrapp et al., 2020). Then, the human transmembrane protease serine 2 (TMPRSS2) hydrolyses and activates the spike protein (Hoffmann et al., 2020). An additional protease, possibly furin, is also involved in this process (Lukassen et al., 2020). Spike proteolysis allows SARS-CoV-2 to enter the cells by endocytosis or by direct fusion of viral and host membranes (Xia et al., 2020; Yang and Shen, 2020). The infecting RNA produces messenger RNA (mRNA), which will be then translated by host ribosomes into protein products (Walsh et al., 2013; Romano et al., 2020b). Using the genetic information encoded in mRNA, the virus takes advantage of the host cell to produce all the components needed for the generation of new viral particles.

Host-pathogen interactions form the basis of the pathogenicity of viruses, including highly pathogenic emerging viruses such as Ebola, SARS-CoV, MERS-CoV, and SARS-CoV-2. By analogy with other known coronaviruses, the RNA replication machinery of SARS-CoV-2 is expected to be regulated by a diversity of host factors, including cellular RNA helicases involved in key events of viral infection (van Hemert et al., 2008; Ranji and Boris-Lawrie, 2010; Sharma and Boris-Lawrie, 2012). Among those, DEAD-box (DDX) RNA helicases are emerging as key players in the host-pathogen interaction network, by modulating innate immunity and viral proliferation in multiple ways. DDX helicases are involved in many steps of the RNA metabolism, including RNA-RNA and RNA-protein remodeling in an ATP-dependent manner (Hilbert et al., 2009; Linder and Jankowsky, 2011). Given the essential role of DDX proteins in many biological processes in humans, their mutation or mis-regulation correlate with an increasing number of pathological processes, including oncogenesis, inflammation, viral replication, and immune response (Steimer and Klostermeier, 2012).

In host-pathogen interactions, DEAD-box helicases display diverse functions upon viral invasion, where they can act as positive or negative regulators of viral replication at different levels (Taschuk and Cherry, 2020). Specifically, DDX enzymes have shown to promote interferon (IFN) induction or other inflammatory signaling, leading to antiviral immunity (Soulat et al., 2008). Some of them function as cytoplasmic sensors of viral RNA, such as the canonical DDX58/RIG-1, while some others act in concert with other proteins (Zhang et al., 2011; Yoo et al., 2014). DDX proteins are required for replication of a number of human viral pathogens, such as HIV-1, HCV, Influenza A, Dengue, Infectious Bronchitis Virus (IBV)-CoV, and SARS-CoV (Fang et al., 2004; Goh et al., 2004; Chen J. Y. et al., 2009; Xu et al., 2010; Diot et al., 2016). Among these, HIV-1 is the most well-described case where the virus hijacks host helicases, such as DDX1, DDX3X, DDX5, and DDX17, to support reverse transcription, transcription, and nuclear export (Yasuda-Inoue et al., 2013).

Little is known about the potential role of DDX helicases in SARS-CoV-2 replication. Although coronaviruses carry their own RNA helicases, they hijack DDX proteins to positively modulate genome transcription and virus proliferation, acting as pro-viral effectors (Chen J. Y. et al., 2009; Xu et al., 2010). Specifically, SARS-CoV has been shown to recruit host helicases, such as DDX1 and DDX5, a feature observed in other coronaviruses, such as IBV-CoV (Chen Y. et al., 2009; Xu et al., 2010). A number of RNA viruses hijack DDX5 and DDX3X helicases, thus facilitating various steps of their replication cycles, although details of interactions and on the interference of host-pathogen interaction on the DDX functional states are lacking (Cheng et al., 2018). Gathering functional data from several coronaviruses, it is clear that the nucleocapsid (N) protein, a crucial protein in the coronaviral life cycle, plays a central role in the hijacking mechanism, although other molecular actors, like Nsp13 and Nsp14, have been identified in SARS-CoV (Figure 1, Table 1). The high sequence identities of the molecular players involved in these interactions suggests that viral replication of SARS-CoV-2 can be modulated by the host cell machinery in a



**FIGURE 1** | Pattern of putative interactions between host DDX helicases and SARS-CoV-2 proteins. These interactions were experimentally observed in other coronaviruses sharing highly conserved viral proteins with SARS-CoV-2.

similar manner. This review summarizes structural information on the host-pathogen interaction processes mediated by DDX helicases in coronaviruses, with an emphasis on SARS-CoV-2. In fact, since the molecular processes involved in these events are highly conserved among coronaviruses, it is likely that similar mechanisms operate in the replication of the novel SARS-CoV-2.

## A STRUCTURAL OVERVIEW OF SARS-CoV-2 PROTEINS LIKELY INVOLVED IN DDX HIJACKING

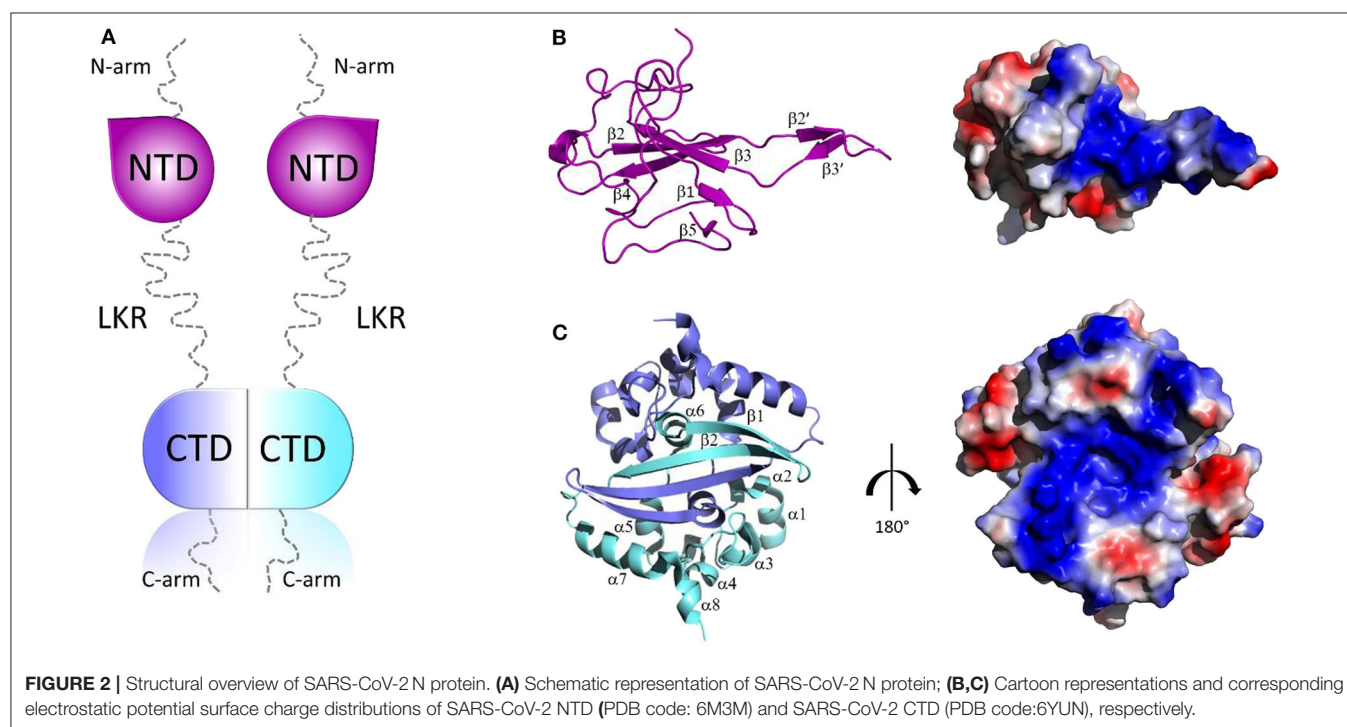
### Nucleocapsid (N) Protein, a Crucial Protein in Coronaviral Life Cycle

Like the other coronaviruses, SARS-CoV-2 has four major structural proteins: the spike (S), the membrane (M), the envelope (E), and the nucleocapsid (N) protein. While the S, M, and E proteins make up the virion envelope, the N protein is located inside the virus particle where it binds with viral RNA (Siu et al., 2008). The N protein plays a primary role in protecting genomic RNA by forming the ribonucleoprotein (RNP) complex, which subsequently condenses through the interaction with the M protein (Narayanan et al., 2000).

In addition to its role in RNP formation, the N protein is a multifunctional phosphoprotein with pivotal roles in several events of the viral life cycle, such as regulating viral RNA synthesis (Chang et al., 2014; McBride et al., 2014; Cong et al., 2020). In fact, a number of studies showed that the N protein co-localizes with components of replication-transcription complex (RTC) at the early stage of CoV infection. Also, it interacts with the transmembrane protein Nsp3 of SARS-CoV to stimulate the infectivity of genomic RNA (Hurst et al., 2010). The N protein is post-translationally phosphorylated, a process that allows discrimination in the binding viral vs. non-viral mRNA, suggesting a pleiotropic effect in RNA regulation (Spencer et al., 2008). So far, multiple investigations underlined the regulatory role of the N protein in viral replication or transcription, with the common view that the N protein has RNA-binding and chaperone activities (Cong et al., 2020). The

**TABLE 1** | Viral proteins known to interact with host DDX helicases in CoVs.

Viral protein	CoV	Host DDX helicase	References	Hypothetic advantages of protein-mediated DDX hijacking in SARS-CoV-2
Protein N	IBV-CoV	DDX3X	Emmott et al., 2013	Inhibition of DDX3X-mediated antiviral immune response
	IBV-CoV	DDX1/DDX21	Emmott et al., 2013;	Evasion of innate immune response through the inhibition of the viral sensor DDX1-DDX21-DHX36
	SARS-CoV-2	DDX21	Gordon et al., 2020	
Nsp14	IBV-CoV, SARS-CoV	DDX1	Xu et al., 2010	DDX1-mediated enhancement of the catalytic activity of Nsp14
Nsp13	SARS-CoV	DDX5	Chen J. Y. et al., 2009	Enhancement of viral RNA transcription and repression innate immunity



N protein is abundantly expressed during infection and is highly immunogenic, capable of inducing a protective immune response against CoV (He Y. et al., 2004). Furthermore, it is also involved in the modulation of the host cellular machinery, by perturbing cellular events such as gene transcription, interferon production, actin reorganization, host cell cycle progression, and apoptosis (McBride et al., 2014). Consistently, the N protein of CoV has been reported to interact with numerous host cell proteins, including hnRNP-A1 (Wang and Zhang, 1999), B23 phosphoprotein (Zeng et al., 2008), Smad3 (Zhao et al., 2008), chemokine Cxcl16 (Zhang et al., 2010), translation elongation factor-1 alpha (Zhou et al., 2008), pyruvate kinase (Wei et al., 2012), and 14-3-3 (Surjit et al., 2005).

Structurally, all CoV N proteins share the same modular organization (Jayaram et al., 2006; Saikatendu et al., 2007), consisting of two structural and independently folded domains, noted as the N-terminal domain (NTD) and C-terminal domain (CTD), flanked by three intrinsically disordered regions (IDR): an N-terminal arm (N-arm), a central linker region (LKR), and a C-terminal tail (C-tail) (Figure 2A). Although

the full-length structure at atomic resolution is lacking, 3D structural information of the structured domains, together with biochemical data, provide crucial insights into the RNP formation process as well as other regulatory functions. Several studies have shown that both the NTD and CTD domains are responsible for the binding of viral RNA genome, whereas CTD also contributes to N protein oligomerisation (Huang et al., 2004; Chang et al., 2005, 2013; Yu et al., 2005; Chen et al., 2007; Chen I. J. et al., 2013; Lo et al., 2013). Interestingly, the LKR linker can modulate the process of viral genome packing, allowing the two structured domains to adopt a wide range of conformations (Chang et al., 2009) and is crucial for N protein oligomerisation (He R. et al., 2004; Luo et al., 2006). The LKR includes a Ser/Arg-rich (SR rich) region that contains a number of putative phosphorylation sites, which regulates N protein function during the early replication step of the viral RNA synthesis (Surjit et al., 2005; Lin et al., 2007; Peng et al., 2008; Wu et al., 2009). The kinase responsible for the phosphorylation of this SR-rich motif is the glycogen synthase kinase-3 (GSK-3), which is conserved in both JHMV and SARS-CoV (Wu et al.,

2014). Consistently, treatment with GSK-3 inhibitor reduces N protein phosphorylation and the viral titer and cytopathic effects (Wu et al., 2014).

The 3D structures of the NTD and the CTD of SARS-CoV-2 were recently released (PDB codes: 6M3M and 6YUN, respectively), revealing high similarity with those of other CoVs (Ye et al., 2020). The NTD domain adopts a U-shaped  $\beta$ -platform structure containing a five stranded antiparallel  $\beta$ -sheet with the topology  $\beta 4$ - $\beta 2$ - $\beta 3$ - $\beta 1$ - $\beta 5$  (Figure 2B), with the two strands  $\beta 2'$  and  $\beta 3'$  located on a protruding  $\beta$ -hairpin (Figure 2B). This fold creates a positively charged pocket between which represents the RNA binding site, as confirmed by NMR studies on SARS-CoV NTD (Chang et al., 2014). Different features characterize the CTD domain, which forms a tightly intertwined homodimer (Figure 2C). Each CTD monomer presents an  $\alpha$ - $\beta$  fold with  $\alpha 1$ - $\alpha 2$ - $\alpha 3$ - $\alpha 4$ - $\alpha 5$ - $\alpha 6$ - $\beta 1$ - $\beta 2$ - $\alpha 7$ - $\alpha 8$  topology and the  $\beta 1$ - $\beta 2$  hairpin of the two monomers form an antiparallel  $\beta$ -sheet at the dimer interface (Yu et al., 2005; Lo et al., 2013), stabilized by extensive hydrogen bonding (Figure 2C). The dimer is also strongly stabilized by hydrophobic interactions between the helix  $\alpha 7$  of one monomer and the  $\beta$ -sheet of an adjacent monomer and between helices  $\alpha 5$  and  $\alpha 6$  of the two monomers (Figure 2C). Solution NMR studies have confirmed that CTD exists predominantly as a dimer in the absence of nucleic acids (Chen et al., 2007). Nevertheless, X-ray crystallography and biochemical studies have shown that CTD is able to form further transient self-associations, conferring to the N protein the ability to form high-order oligomers (Chen et al., 2007; Chang et al., 2013). It has been hypothesized that N protein oligomerisation allows the optimal packaging of the RNA genome during RNP formation process (Chen et al., 2007; Takeda et al., 2008; Chang et al., 2009, 2014). As in the case with NTD, RNA-binding sites on the CTD have strongly positive electrostatic potential surfaces (Figures 2B,C) (Huang et al., 2004; Takeda et al., 2008; Chang et al., 2014), indicating non-specific electrostatic interactions between the N protein and the viral RNA (Chang et al., 2014), as expected for a protein that has to bind to diverse RNA sequences with reasonable affinity during encapsidation.

### Nsp13 Helicase, a Multifunctional Enzyme Involved in Genome Unwinding and the First Step in mRNA Capping

Helicases are ubiquitous motor proteins, also present in (+) RNA viruses with genomes larger than 7 kb. They are nucleic acid-dependent ATPases capable of unwinding DNA or RNA duplex substrates during nucleic acid replication, transcription, DNA repair, RNA maturation, and splicing (Patel and Donmez, 2006). Viral helicases belong to three out of the six currently recognized super-families: SF1, SF2, and SF3 (Lehmann et al., 2015). Sequence analysis suggests that Nsp13 of SARS-CoV-2 is part of the SF1 superfamily and can exert multiple enzymatic activities. Consistently, biochemical studies have shown that SARS-CoV Nsp13 unwinds both RNA and DNA duplexes in the 5' to 3' direction and is able to hydrolyze deoxyribonucleotide and ribonucleotide triphosphates (Ivanov and Ziebuhr, 2004; Adedeji et al., 2012). This reaction characterizes the first step of

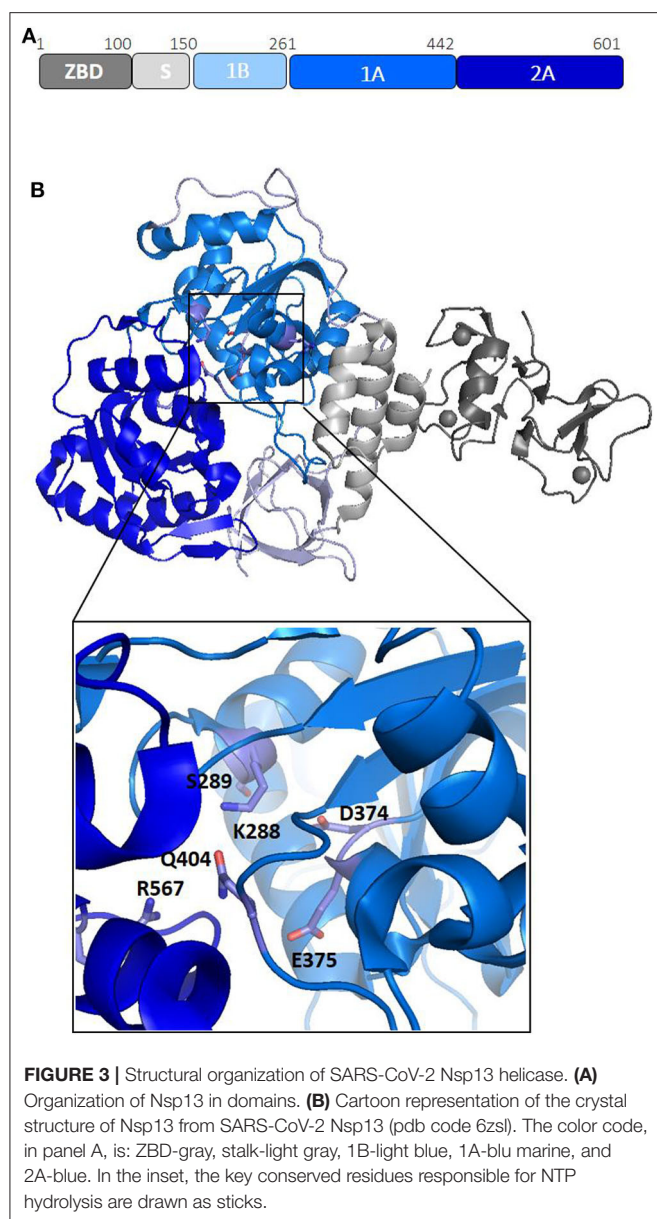
an important process of viral RNA synthesis, the protection of nascent mRNAs at their 5' ends by a cap structure, which makes viral mRNA more stable and able to evade the host immune response. Therefore, not only is Nsp13 involved in genome unwinding, but, due to its RNA 5'-triphosphatase activity, it is an essential enzyme in the mRNA capping (Ivanov and Ziebuhr, 2004; Adedeji et al., 2012).

Nsp13 is conserved in all coronaviruses and is key to viral replication (van Dinten and van Tol H, 2000; Lehmann et al., 2015). Given the high sequence conservation and indispensability across all CoV species (Lehmann et al., 2015), Nsp13 is a promising target for the development of anti-viral drugs (Shum and Tanner, 2008; Adedeji et al., 2014). Structures of Nsp13 helicases from coronaviruses SARS-CoV and MERS and, more recently, SARS-CoV-2, have been hitherto reported (Hao et al., 2017; Jia et al., 2019). The Nsp13 structure is composed of five domains, organized in a triangular pyramid shape (Figure 3). At the apex of the pyramid are the N-terminal zinc binding (ZBD) and the stalk (S) domains (Figure 3B). Domains 1A and 2A present the RecA-like structure typical of helicases, with the catalytic site located at their interface. The catalytic pocket of the enzyme is located between the two RecA domains 1A and 2A and include the six residues K288, S289, D374, E375, Q404, and R567 (Figure 3B). Interestingly, the unwinding activity of SARS-CoV Nsp13 is stimulated by its interactions with the RNA dependent RNA polymerase (RdRP, Nsp12), the main enzyme responsible for CoV RNA polymerisation, with this interaction mediated by ZBD and 1A domains (Figure 3B) (Jia et al., 2019). The high sequence conservation of Nsp12 and Nsp13 in all CoVs (Supplementary Table 1) suggests their association as a common feature across CoVs (Jia et al., 2019).

### Nsp14, a Key Enzyme Involved in mRNA Proofreading and Final Capping

Nsp14 of SARS-CoV-2 is a peculiar bi-functional enzyme. It is composed of two different functional domains: an N-terminal exoribonuclease (ExoN) domain and a C-terminal N7-guanine MTase domain (N7-MTase). The ExoN domain of Nsp14 hydrolyses single-stranded and double-stranded RNAs and is critical for proofreading function in coronaviruses (Denison et al., 2011; Ogando et al., 2019), a property that is missing in other RNA viruses. Due to the presence in the ExoN domain of three conserved motifs, motif I (DE), II (E), and III (D), Nsp14 is classified as a "DEED outlier" among DEDD exonucleases (Ma et al., 2015; Ogando et al., 2019). Knockout mutants of ExoN in Murine Hepatitis Virus (MHV), SARS-CoV, Alphacoronaviruses HCoV-229E, and Transmissible Gastroenteritis Virus (TGEV) were shown to display defects in the synthesis of genome and sub-genome length RNAs and a reduction in replication efficiency, thus emphasizing the supporting role of Nsp14 in CoV replication (Minskaia et al., 2006; Eckerle et al., 2007, 2010). In this context, Nsp14, through its ExoN domain, is part of the RNA replication machinery.

The carboxy-terminal region of Nsp14, containing N7-guanine MTase activity, instead plays a key role in capping of viral both genomic and sub-genomic mRNAs. The cap synthesis



starts with the hydrolysis of the 5' end of a nascent RNA by the RNA 5'-triphosphatase Nsp13 to yield pp-RNA (Ivanov and Ziebuhr, 2004). Subsequently, a still unknown GTase transfers a GMP molecule onto the pp-RNA to yield Gppp-RNA. The cap structure is then methylated at the N7 position by the N7-MTase domain of Nsp14 (Chen J. Y. et al., 2009; Chen Y. et al., 2013). In addition to being important for the stability of mRNAs, the cap structure is essential to avoid the host immune response and allows the ribosomal complex to recognize mRNAs to ensure their efficient translation (Decroly et al., 2012).

Based on the crystal structure of SARS-CoV Nsp14-Nsp10 complex (PDB code: 5C8S), we have generated, using Modeler, a homology model of the complex between Nsp14 and Nsp10 of SARS-CoV-2 (Figure 4) (Ma et al., 2015; Ferron et al., 2018). In this heterodimer, the co-factor Nsp10 forms numerous

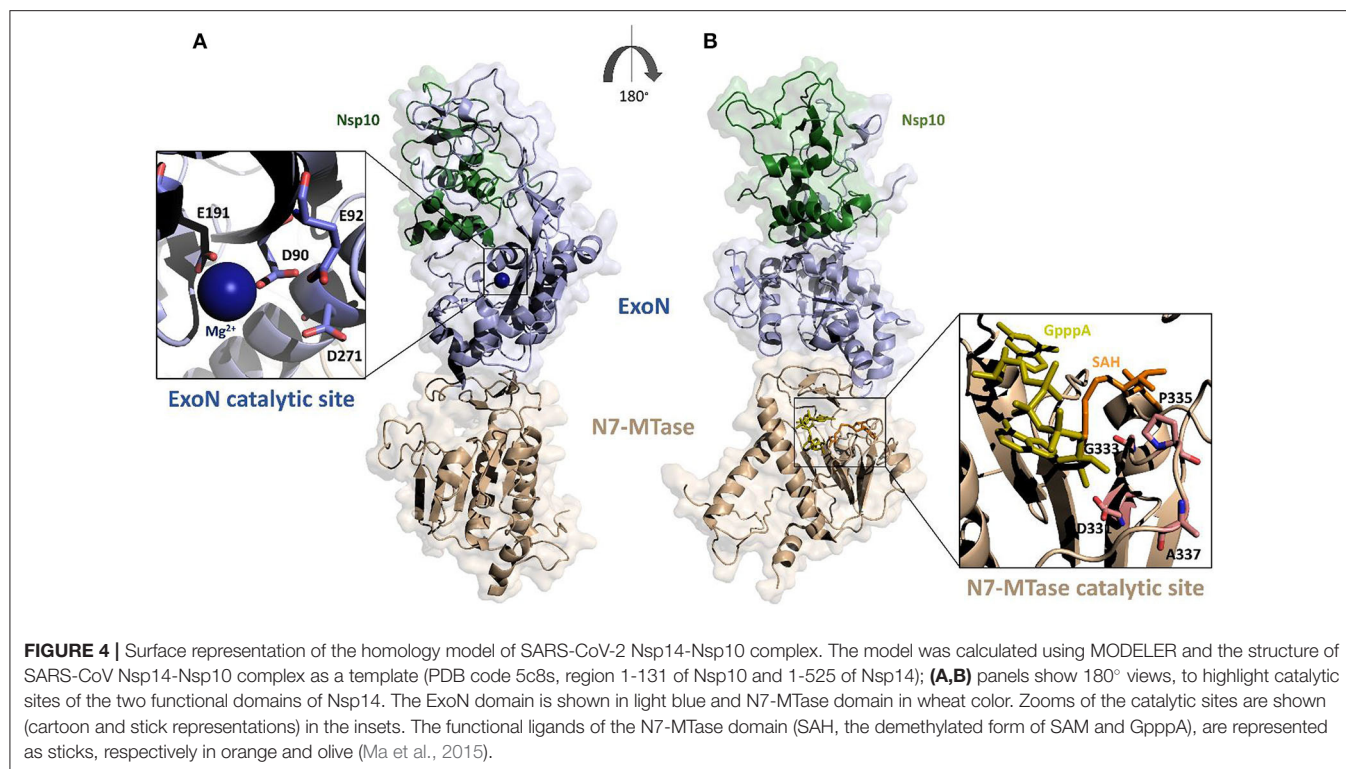
interactions with the ExoN domain of Nsp14 (Bouvet et al., 2014). In particular, the first 25 residues of Nsp14 form a clamp that accommodates the first 10 residues of Nsp10. This structural relationship is the key to the ExoN activity. Indeed, this tight interaction between Nsp10 and the N-terminus of Nsp14 allows the exonuclease active site to adopt a stable and highly active conformation. These structural findings, together with experimental evidence, underline that the interaction with Nsp10 strongly influences the nucleolytic activity of Nsp14 (Bouvet et al., 2012). Consistently, as shown by SAXS experiments, in the absence of Nsp10, Nsp14 from SARS-CoV shows large conformational changes at its N terminus, which affect the overall structural architecture of the ExoN domain, leading to weak ExoN activity (Ferron et al., 2018).

The ExoN domain of Nsp14 is essentially composed of a central twisted  $\beta$ -sheet which is formed by five  $\beta$ -strands flanked by  $\alpha$ -helices (Ma et al., 2015; Ferron et al., 2018). This architecture is typical of exonucleases of the DEED superfamily, like RNase T from *E. coli* and RNase AS from *M. tuberculosis* (Derbyshire et al., 1991; Romano et al., 2014, 2015). The catalytic tetrad was identified using the structural alignment with SARS-CoV Nsp14 and include the DEED residues Asp90, Glu92, Glu191, and Asp272 (Figure 4A).

A flexible hinge region connects the ExoN domain with the N7-MTase domain of Nsp14. This region, consisting of a loop and three strands, is highly conserved across CoVs and permits significant movements of the two domains (Figure 4) (Ma et al., 2015; Ogando et al., 2019). As mentioned before, the N7-MTase domain participates in mRNA capping by adding a methyl at the N7 position of the guanosine of the mRNA cap structure. To complete this enzymatic reaction, it uses the co-enzyme S-Adenosyl Methionine (SAM) as methyl donor (Figure 4). The N7-MTase domain does not belong to any of the classes of SAM-dependent MTases (Schubert et al., 2003; Byszewska et al., 2014; Chouhan et al., 2019) and shows a non-canonical MTase fold different from the Rossmann fold of the virus RNA MTase (Rao and Rossmann, 1973; Ferron et al., 2018). Indeed, whereas the canonical fold is formed by a seven-strand  $\beta$ -sheet with at least three parallel  $\alpha$ -helices on each side, Nsp14 contains a central  $\beta$ -sheet composed by only a five-strand (Figure 4B) (Rao and Rossmann, 1973). The N7-MTase domain presents two clusters of residues key for its enzymatic activity; the first cluster is a canonical SAM-binding motif I (DxGxPxG/A), including Asp331, Gly333, Pro335, and Ala337 (Figure 4B). A second cluster forms a pocket that holds the GTP of the mRNA cap structure close to the methyl donor SAM.

## A STRUCTURAL OVERVIEW OF HUMAN DDX HELICASES INVOLVED IN COV INTERACTION

Human DDX helicases belong to the DEAD-box protein family, which is the largest family of the superfamily 2 (SF2) helicases (Byrd and Raney, 2012). They share a highly conserved helicase core consisting of two RecA-like domains (D1 and D2) connected via a flexible linker region (Hilbert et al., 2009). These two



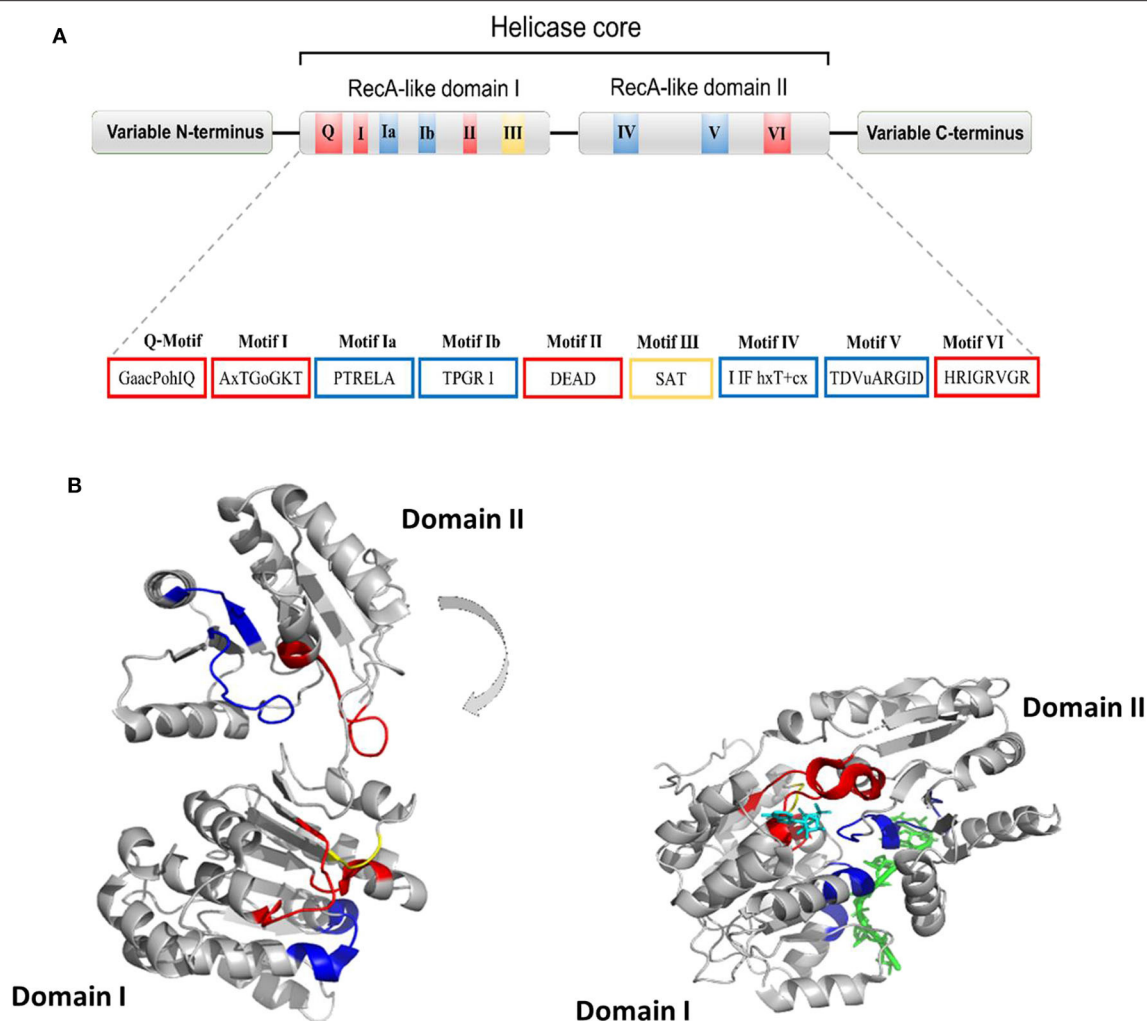
domains, also known as DEAD domain and helicase domain, are characterized by the presence of nine conserved motifs, Q, 1, 1a, 1b, II, III, IV, V, and VI, and the common tetrapeptide (Asp-Glu-Ala-Asp) in motif II (**Figure 5A**). In many DEAD box helicases, the D1D2 domains are linked to distinct flanking regions or ancillary domains, ranging from a few to several hundred amino acids in length (Rudolph and Klostermeier, 2015). These variable N-terminal and C-terminal domains contribute to the functional diversity of this protein family, and direct individual helicases to their targets via protein or RNA interaction or by modulating the activity of the helicase core (Del Campo and Lambowitz, 2009; Mallam et al., 2011; Ferrage et al., 2012; Rudolph and Klostermeier, 2015). Structural studies on DDX helicases, reported in **Tables 2, 3**, are mainly focused on the helicase core or isolated ancillary domains, with few structures of full-length helicases. Nevertheless, the conserved patch in the helicases core provides important insights on their activity. In DDX proteins, the two RecA-like domains are both responsible for RNA binding and ATP hydrolysis (Hilbert et al., 2009). Specifically, the conserved motifs Q, I (P-loop), II, and VI participate in ATP binding and hydrolysis; motif III is responsible for coupling NTP hydrolysis to nucleic acid unwinding, and motifs Ia, Ib, IV, and V are involved in RNA binding (**Figure 5**) (Hilbert et al., 2009).

Comparative studies on the helicase core highlighted a structural mechanism of open-close conformation of the two RecA-like domains, suggesting a link between the binding of ATP and activation of the RNA binding site (Theissen et al., 2008; Schutz et al., 2010). In the absence of a ligand, the D1

and D2 domains are separated and more flexible, resulting in an open conformation (**Figure 5B**). Upon ATP and RNA binding, they approach each other in a more compact arrangement, thus forming a cleft which aligns the catalytic site for ATP hydrolysis (**Figure 5B**). In particular, the P-loop and motif II coordinate the nucleotide phosphates and the magnesium ion, whereas residues of the Q-motif bind and recognize the adenine moiety (Schutz et al., 2010) (**Figure 5B**). The domain closure mechanism is consistent with biochemical studies, indicating that ATP and nucleic acid binding is highly cooperative in DDX proteins, most of which are unable to bind or hydrolyze ATP without RNA (Jankowsky and Fairman, 2007; Theissen et al., 2008). Importantly, in all DEAD-box protein structures, conserved amino acids make contact exclusively with the sugar-phosphate backbone of the RNA, including several interactions with the 2'-OH groups (Jarmoskaite and Russell, 2011). This finding explains why RNA binding is not sequence specific but DDX can still distinguish RNA from DNA.

## The Two Homologs DDX3X and DDX5

Like all DDXs, DDX3X contains a helicase core composed of two RecA-like domains, constituting the D1D2 core. This core is flanked by N- or C-terminal unstructured tails (**Figure 6A**). Despite the lack of structural information on the N (NTE) and C terminal tails (CTE), evidence has accumulated showing that NTE contains a nuclear export sequence and is key for DDX3X nuclear export (Heaton et al., 2019). On the other hand, CTE is essential for DDX3X oligomerisation, a process which is required for optimal helicase activity (Putnam et al., 2015).



**FIGURE 5 | (A)** Typical domain organization of DDX helicases. The motifs are shown with colors according to their primary function (red, ATP binding and hydrolysis; yellow, coordination between NTP- and RNA-binding sites; blue, RNA binding). **(B)** Cartoon representation showing domain movements in DDX helicases upon RNA binding of the DEAD-box protein eIF4AIII without RNA (left) and in complex with RNA (right). An RNA fragment and AMPPNP are shown in stick representation in green and in cyan, respectively.

As for the D1D2 core, DDX3X is one of the most structurally characterized DDX helicases. The crystal structure D1D2 core of DDX3X has been reported in several functional states: the apo state; the pre-unwound state, consisting of a complex with a 2-turn dsRNA; and a post-unwound state, in which DDX3X is complexed with a single strand RNA (**Figure 6**) (Song and Ji, 2019). These studies have shown that the binding of dsRNA to apoDDX3X triggers dramatic conformational changes of the enzyme to form the pre-unwound DDX:dsRNA:DDX complex (**Figure 6B**). This structure, also confirmed in solution, resembles a clamp that locks dsRNA among four RecA-like domains (**Figure 6B**), with one D1D2 core mainly binding RNA Strand 1 and the other D1D2 core binding RNA Strand 2 (Song and Ji, 2019). In the transition from a pre-unwound to a post-unwound state, RNA unwinding is driven by ATP hydrolysis and finally leads to a monomeric complex of DDX3X and single strand RNA (**Figure 6B**). In this unwinding model, the two dsRNA-bound

D1D2 cores undergo dramatic conformational changes upon the binding of MgATP and thus unwind dsRNA (**Figure 6**). This mechanism has been proposed as a general unwinding mechanism of DDX helicases (Song and Ji, 2019). DDX3X has a confirmed role in anti-viral immune signaling pathways leading to type I IFN induction (Schroder et al., 2008; Soulat et al., 2008; Gu et al., 2013, 2017). Notably, DDX3X has already been validated as a target for broad-spectrum antiviral molecules against a number of RNA viruses (HIV, HCV, DENV, and WNV).

DDX5 plays fundamental roles in transcriptional regulation and in viral replication (Cheng et al., 2018). Like other DEAD-box helicases, DDX5 shares the modular domain architecture with the nine conserved motifs constituting the core region (**Figure 7**). Moreover, DDX5 helicase contains a Arg-Gly-Ser-Arg-Gly-Gly (RGS-RGG) motif and an Ile-Gln (IQ) motif, which is localized in the C-terminal region and can act either as RNA-binding site or protein interaction module (Rajyaguru and

**TABLE 2 |** Structural similarity of SARS-CoV-2 Nsp13 with human DDX helicases, as computed by DALI.

	PDB code	Z score	Seqid (%)	Rmsd (Å)	Length of alignment
DDX20	2oxc	9.8	12	3.1	149
DDX53	3iuy	9.5	10	3.0	213
DDX10	2pl3	9.3	14	3.1	147
DDX47	3ber	9.1	10	3.0	146
DDX52	3dkp	9.0	11	3.3	148
DDX18	3ly5	8.9	8	3.4	150
DDX17	6uv1	8.7	11	3.6	200
DDX5	4a4d	8.4	11	3.4	155
DDX3X	6o5f	7.8	8	3.8	155
DDX41	2p6n	7.7	11	3.0	131
DDX25	2rb4	7.1	8	3.2	128

Parker, 2012) (**Figure 7**). Indeed, there is evidence that DDX5 may interact with other DEAD-box helicases, such as DDX3X and its close homolog DDX17; an interaction between DDX5 and DDX3X was proposed as a novel combined mechanism of action for DEAD-box RNA helicases involved in RNP remodeling and splicing (Ogilvie et al., 2003; Choi and Lee, 2012).

The only structural information available for DDX5 include domain I and part of the variable N-terminal region (Schutz et al., 2010) (**Table 3**). However, DDX5 shares a high sequence identity with DDX17 (83.7%, region 35-476), whose structure is known (Ngo et al., 2019). Therefore, we computed the homology modeling structure of a major portion of DDX5, including an N-terminal region of unknown function and the D1D2 core (**Figure 7A**). In this structure, the N-terminal region crosses the entire length of the D1 domain, reaching the RNA substrate binding region (**Figure 7**). In DDX17, the N-terminal extension can modulate its ATPase and unwinding activities (Ngo et al., 2019). The high sequence similarity of DDX5 with that of DDX17 suggests a similar autoregulatory role of the N-terminal region in controlling the ATPase activity, through the observed intramolecular interaction between N-terminal tail and the D1 domain (Ngo et al., 2019) (**Figure 7**). As for the D1D2 core, structural similarity analysis by DALI server shows that DDX5 D1D2 (PDB entry 3fe2) possesses high structural similarity with DDX3X, as well as with DDX17 and DDX41 (z-scores 31.4, 35.9, 33.5, respectively), thus indicating similar roles for all these DDX.

## DDX1, a Structurally Unique DDX

In DEAD-box helicases, different insertions, N- or C-terminal appendages, or additional domains to the standard modular architecture of RecA-like domains may be responsible for the diversity of DEAD-box protein functions. As an example, DDX1 is unique among DEAD-box proteins because of a domain insertion in the amino-terminal helicase domain (Kellner et al., 2015). This extra domain, denoted as SPRY (SP1a and the RY anodine Receptor), is inserted between the P-loop and the conserved motif Ia of the RecA domain D1 (**Figure 8A**). The SPRY domain is the sole structural information hitherto available

for DDX1 and shows a typical  $\beta$ -sandwich fold, with a variability in the loop regions (Kellner and Meinhart, 2015). However, using structural information from its homologous structures, we computed a homology model of the entire DDX1, for a better description of its structural and binding properties (**Figure 8B**). In this structure, the SPRY domain protrudes from the structure of the helicase and is located on the opposite side of the enzyme catalytic site. Consistently, SPRY is a protein interaction module implicated in important biological pathways, including innate immunity (D'Cruz et al., 2013), regulating DDX1 assembly into multiprotein complexes. However, the interacting partners of SPRY remain unknown, as do the molecular determinants of binding specificity.

## INVOLVEMENT OF COV PROTEINS IN DDX HIJACKING

### The N Protein Is a Central Hub for DDX Interactions

The multifunctional role of the N protein in the virus life cycle, from regulation of replication and transcription and genome packaging to modulation of host cell processes, strongly relies on interactions with host cell proteins (Emmott et al., 2013). Consistently, this protein has been shown to be involved in interactions with multiple host proteins in IBV-CoV and SARS-CoV, including proteins of the DDX family (Emmott et al., 2013) (**Table 1**). An interaction of N protein has also been observed directly in SARS-CoV-2 with the helicase DDX21 (Gordon et al., 2020). However, the effect of this interaction on viral replication is currently unknown.

Among DDX proteins interacting with the N protein, DDX3X is massively involved in immune reactions against the viral infection. In particular, it is involved in the interferon (IFN)-mediated innate immune cascade (Gu et al., 2013) and is a positive regulator of the TBK1/IKK $\epsilon$  signaling cascade, as it interacts with and acts as a substrate of the TBK1 and IKK $\epsilon$  kinases, activating the IRF3 transcription factor (Brai et al., 2020a). DDX3X is also recruited to the IFN promoter through interaction with IRF3 itself (Gu et al., 2013). In addition, DDX3X interacts with IPS-1 and TRAF3, components of the RIG-I sensor complex, for the recognition of viral RNA and activation of the IFN response (Oshiumi et al., 2010; Gu et al., 2017) and modulates the NF- $\kappa$ B inflammatory response through activation of the IKK $\beta$ /PPA2C complex (Wang et al., 2017). Therefore, it is possible to hypothesize that the N protein plays a role in modulating the immune response mediated by DDX3X and that binding of N to DDX3X results in the inhibition of these antiviral pathways. Indeed, a similar mechanism acting on DDX3X has been already demonstrated for other viruses (Schroder et al., 2008).

DDX1 and DDX21 were also shown to belong to the interactome of the N protein in IBV-CoV (Emmott et al., 2013) (**Table 1**). Interestingly, these two DDX proteins have been shown to interact with each other to form a large complex with another RNA helicase, DHX36, and with the adaptor molecule TRIF, forming a sensor of viral dsRNA in dendritic cells (Zhang

**TABLE 3 |** Structures of DDX helicases.

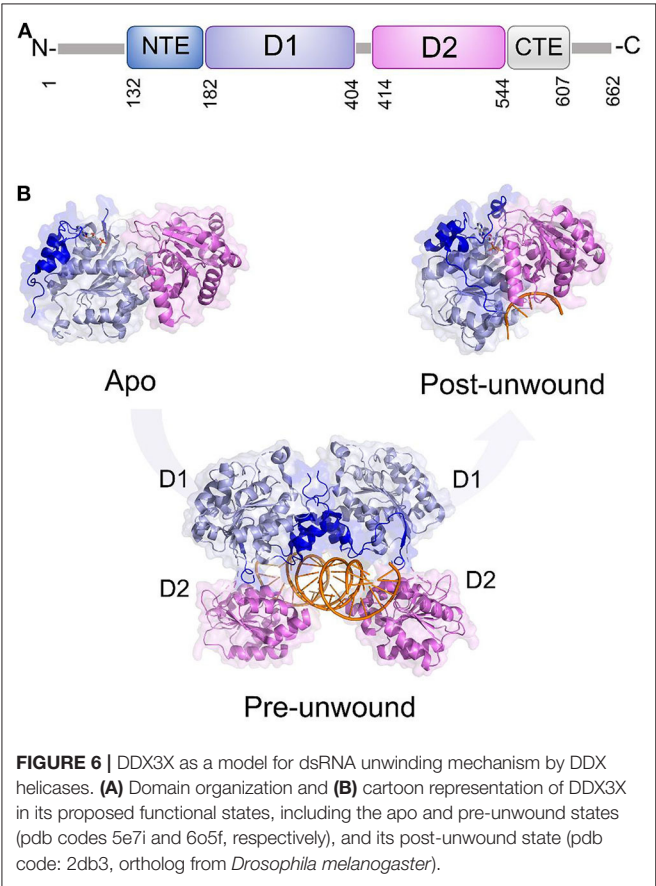
Protein	PDB code	Sequence information/residue	Ligands
DDX1	4XW3	SPRY domain	None
DDX3X	2JGN	Domain II 409–580	None
	4PX9	NTE and Domain I 135–407	ADP
	4PXA	NTE and D1D2 core 135–582 (D354V)	ADP, PO <sub>4</sub>
	2I4I	D1D2 core 18–582	AMP
	5E7I, 5E7M, 5E7J	NTE and D1D2 core 133–584	None – AMPPNP – AMP
	6O5F	NTE, CTE and D1D2 core 132–607	23-bp dsRNA
	6CZ5	NTE, CTE and D1D2 core 132–607 (S228)	AMP-acrylamide
	3JRV	N-terminal peptide 71–90	Protein K7 ( <i>Vaccinia virus</i> )
	4O2C, 4O2E, 4O2F	N-terminal peptides SL9 NAc, SL9, HL8	HLA-B*3901 (HLA class I histocompatibility antigen, B-39 alpha chain)
DDX5/p68	4A4D	Domain I 52–304	None
	3FE2	Domain I 68–307	ADP, SO <sub>4</sub>
DDX6/p54	1VEC	Domain I 94–299	Zn <sup>+2</sup> , C <sub>4</sub> H <sub>6</sub> O <sub>6</sub>
	2WAY, 2WAX	Domain II 296–483	EDC3 FDF peptide
	4GRW	Domain II 307–483	CNOT1 MIF4G domain (fragment)
	6F9S	Domain II 301–469	LSM14 peptide, SO <sub>4</sub>
	6S8S	Domain II 295–483	EDC3 FDF peptide, PO <sub>4</sub>
	4CT5, 4CT4, 5ANR	D1D1 core 95–469	None - CNOT1 MIF4G domain, Mg <sup>+2</sup> - 4E-T/CNOT1
DDX10	2PL3	Domain I 47–280	ADP, Mg <sup>+2</sup>
DDX17	6UV0	D1D2 core 111–556	Mg <sup>+2</sup>
	6UV1	D1D2 core 111–556	rU10 RNA, ADP, Mg <sup>+2</sup>
	6UV2	D1D2 core 111–556	pri-125a-oligo1 RNA, ADP, Mg <sup>+2</sup>
	6UV3	D1D2 core 111–556	pri-125a-oligo2 RNA, ADP, Mg <sup>+2</sup>
	6UV4	D1D2 core 111–556	pri-18a-oligo1 RNA, ADP, Mg <sup>+2</sup>
DDX18	3LY5	Domain I 149–387	PO <sub>4</sub> , Mg <sup>+2</sup>
DDX19B	3FHC	Domain I 68–302	Nup214
	3FHT	D1D2 core 68–479 (ΔN67)	AMPPNP, U10 ssRNA, Mg <sup>+2</sup>
	3FMO - 3FMP	NTE and Domain I 1–300 - Full length	Nup214, ADP
	3EWS, 3G0H	D1D2 core 54–475 (ΔN53)	ADP - AMPPNP, U <sub>10</sub> RNA, Mg <sup>+2</sup>
	6B4K - 6B4J - 6B4I	D1D2 core 54–479 (ΔN53)	AMPPNP, Mg <sup>+2</sup> - AMPPNP/Gle1 <sup>CTD</sup> -Nup42 <sup>GBM</sup> , Mg <sup>+2</sup> - ADP/Gle1 <sup>CTD</sup> -Nup42 <sup>GBM</sup>
DDX20	2OXC	Domain I 41–268	ADP
	3B7G	Domain I 41–268	AMPPNP
DDX21	2M3D	GUCT domain	None
DDX23/Prp28	4NHO	D1D2 core 338–820 (ΔN337)	SO <sub>4</sub> , Hg <sup>+2</sup>
	3JCR	Full length	U4/U6.U5 tri-snRNP spliceosomal complex
	6AH0	Full length	Precursor of pre-catalytic spliceosome
	6QX9	Full length	Fully assembled pre-catalytic spliceosome
	6QW6	Full length	U4/U6.U5 spliceosomal tri-snRNP complex
DDX25	2RB4	Domain II 307–479	SO <sub>4</sub> , Zn <sup>+2</sup>
DDX41	2P6N	Domain II 402–569	None
	5GVR - 5GVS	Domain I 169–402 closed form - open form	None
	5H1Y	Domain I 153–410	SO <sub>4</sub> , Mg <sup>+2</sup>
DDX47	3BER	Domain I 5–230	AMP, PO <sub>4</sub>
DDX48/eIF4AIII	2HXY	D1D2 core 23–411	None
	2HYI - 2XB2 - 3EX7	D1D2 core 1–411	ECJ (Exon junction complex), RNA, AMPPNP, Mg <sup>+2</sup> - UPF3b ECJ, AMPPNP, Mg <sup>+2</sup> - ECJ transition state, ADP, Mg <sup>+2</sup>
	4C9B	D1D2 core 1–411	CWC22, PO <sub>4</sub>
	2J0Q, 2J0S	D1D2 core 2–411	ECJ, AMPPNP, Mg <sup>+2</sup>
	2J0U	D1D2 core 39–411	CASC3

(Continued)

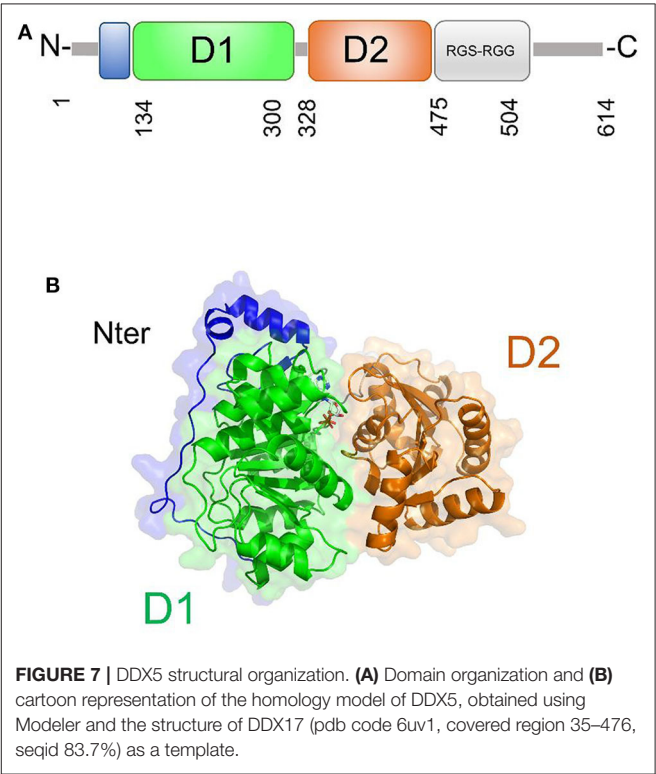
TABLE 3 | Continued

Protein	PDB code	Sequence information/residue	Ligands
DDX50	5YZG, 5MQF, 5XJC, 6ICZ, 6QDV	D1D2 core 1–411	C complex, step I - C complex, step II - P complex (human spliceosome)
	2E29	GUCT domain	None
	3DKP	Domain I <sub>139–381</sub>	ADP, Mg <sup>2+</sup>
	3IUY	Domain I <sub>204–430</sub>	AMP
	2LWD, 2LWE	CARD 2 domain <sub>95–190</sub> WT and T170E variant	None
	4NQK - 4P4H	CARD 1 and 2 domains <sub>1–200</sub>	Ubiquitin - CARD <sup>MAVS</sup> , UBA-52
	2RMJ - 3NCU	CTD <sub>792–925</sub>	None - 12bp blunt-end 5'-pp dsRNA, Zn <sup>2+</sup>
	2QFB - 2QFD - 3LRN - 3LRR - 3OG8	CTD <sub>802–925</sub>	Zn <sup>2+</sup> - Hg <sup>2+</sup> - 14bp GC rich 5' ppp dsRNA, Zn <sup>2+</sup> - 12bp AU rich 5' ppp dsRNA, Zn <sup>2+</sup> - 14bp blunt-end dsRNA, Zn <sup>2+</sup>
	4ON9	D1D2 core <sub>230–793</sub>	SO <sub>4</sub>
	5E3H, 6GPG	CTD and D1D2 core <sub>232–925</sub> (ΔCARDs) - C268F	14bp dsRNA, ADP, Mg <sup>2+</sup> , Zn <sup>2+</sup> - 14bp dsRNA, Mg <sup>2+</sup> , Zn <sup>2+</sup>
DDX58/RIG-1	2YKG, 4BPB, 3ZD6, 3ZD7	CTD and D1D2 core <sub>230–925</sub> (ΔCARDs)	5'-OH-GC10 dsRNA, Zn <sup>2+</sup> , SO <sub>4</sub> - 5'-OH-GC10 dsRNA, Zn <sup>2+</sup> - 5'-OH-GC10 dsRNA, ADP, Mg <sup>2+</sup> , Zn <sup>2+</sup>
	4AY2, 5F9F, 5F9H, 5F98	CTD and D1D2 core <sub>232–925</sub> (ΔCARDs)	5'ppp 8-bp HP RNA, ADP, Mg <sup>2+</sup> , Zn <sup>2+</sup> - 5'-OH HP RNA, Mg <sup>2+</sup> , Zn <sup>2+</sup> - 5'ppp HP RNA, Mg <sup>2+</sup> , Zn <sup>2+</sup> - Cap-0 HP RNA

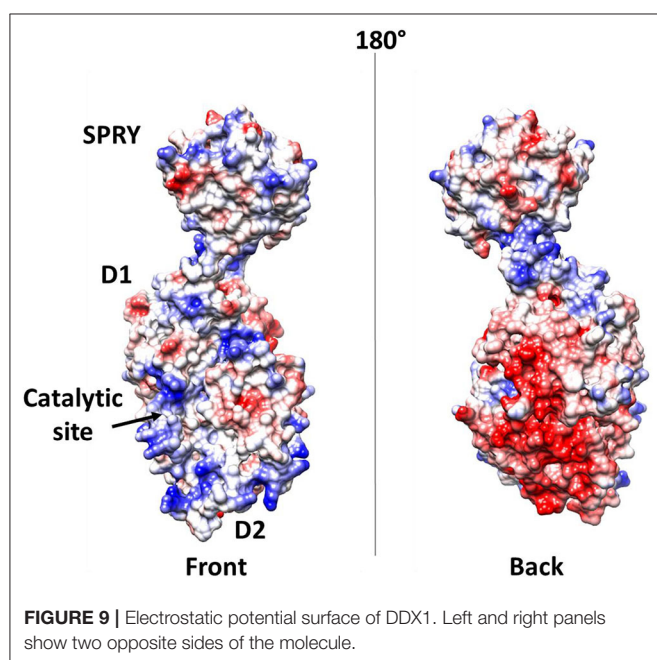
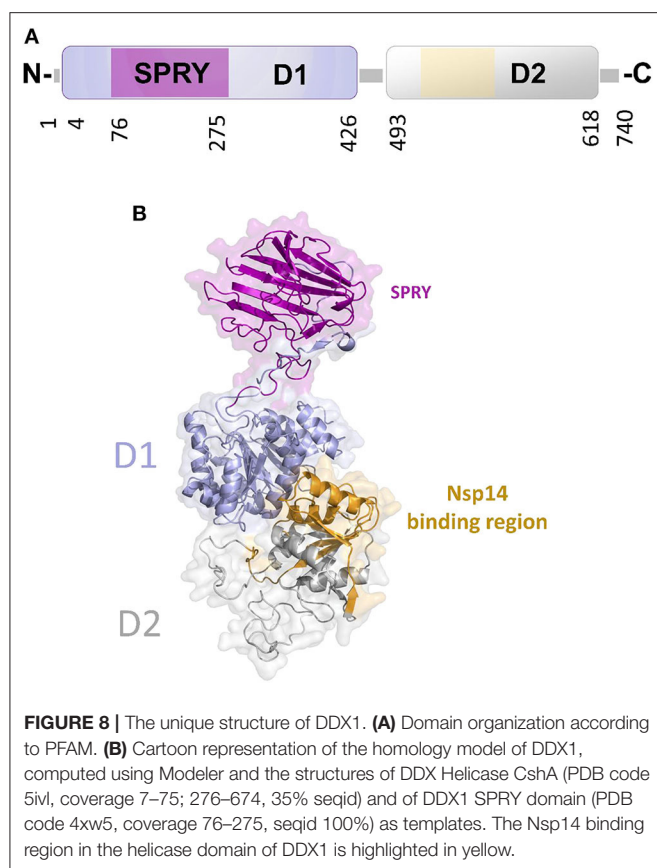
NTE (N-terminal extension), CTE (C-terminal extension), CTD(C-Terminal regulatory Domain), CARD (Caspase recruitment domain).



et al., 2011). This DDX1-DDX21-DHX36 complex uses DDX1 to bind dsRNA and DDX21, and DHX36 to bind TRIE, to activate the NF-κB pathway and type I IFN responses in dendritic cells



(Zhang et al., 2011). In this context, the DDX1-DDX21-DHX36 complex helps the innate immune system to detect viral infection by sensing viral nucleic acids cells (Zhang et al., 2011). Targeting of this complex by the N protein plays a role in immune evasion, possibly interfering with the formation of the sensor complex or by inhibiting the recognition of dsRNA by DDX1. However,



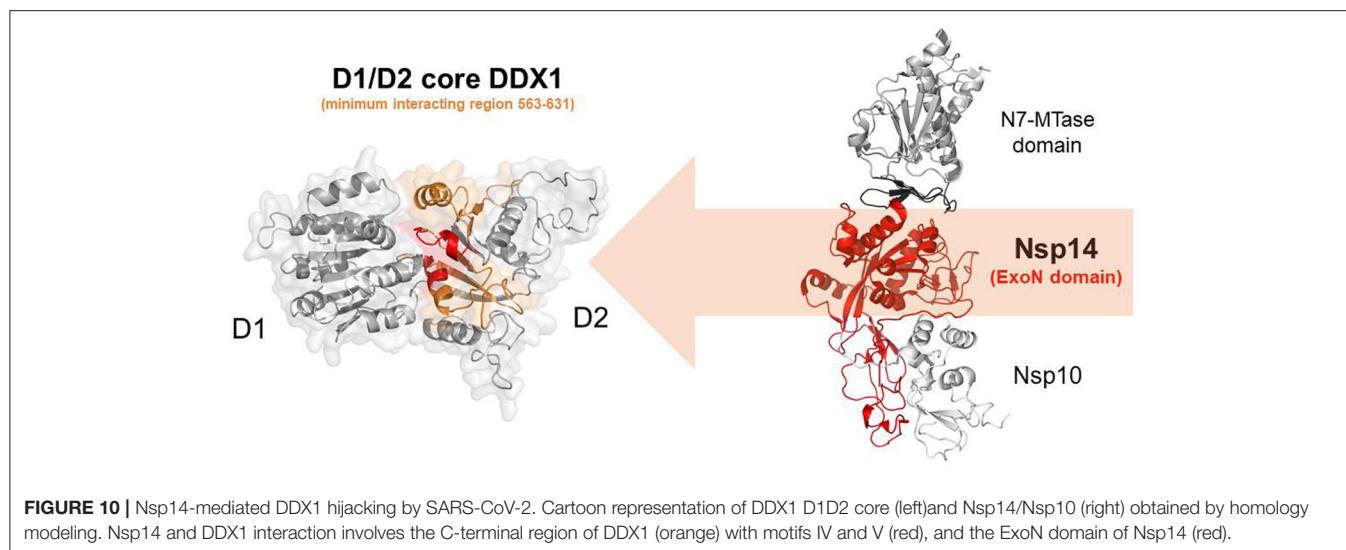
research studies are still needed to clarify this aspect, as details of these interactions are currently unknown.

Further elements are known on the interaction of DDX1 with the N protein (de Haan and Rottier, 2005; McBride et al., 2014; Wu et al., 2014). As discussed above, the N protein is a highly basic protein with phosphorylation modifications, predominantly found in serine residues clustered within the central Ser-arginine (SR)-rich motif located in the LKR linker, which connects the NTD and CTD domains (Wu et al., 2009). This crucial protein also regulates the discontinuous transcription of sub-genomic mRNAs, since its depletion reduces the synthesis of subgenomic mRNA, but not of genomic RNA (Zuniga et al., 2010). Interestingly, the phosphorylation of the N protein by host kinase GSK-3 allows the recruitment of DDX1 helicase, as DDX1 specifically interacts with the N protein of JHMV phosphorylated at Ser197 (pS197-N) belonging to the LKR linker (Wu et al., 2014). In this process, DDX1 plays a role in the regulation of viral RNA transcription by increasing the synthesis of longer viral RNAs. Indeed, DDX1 knockdown markedly reduces the synthesis of longer RNAs of JHMV but minimally reduces the synthesis of shorter sub-genomic mRNA (Wu et al., 2014). Consistently, overexpression of the wild-type, but not of an inactive mutant of DDX1, increases the synthesis of longer viral RNAs, showing that the effect of DDX1 on viral RNA synthesis is linked to its enzymatic function (Wu et al., 2014). This evidence suggests that the involvement of host DDX1, through interactions with the phosphorylated N protein, is a unique viral strategy to support the transition from discontinuous to continuous transcription (Wu et al., 2014). Mechanistic details of this regulatory interaction are unknown. However, the key involvement of phosphorylated serine residues in DDX1 binding, as discussed above, suggests an electrostatic interaction between the LKR linker and the highly charged DDX1 surface to regulate its function (Figure 9).

### Nsp14 and DDX1, a Molecular Liaison to Repress Viral Induced IFN- $\beta$ Immune Reaction

In addition to the N protein, Nsp14 has also been shown to interact with DDX1 in both IBV-CoV and SARS-CoV (Xu et al., 2010; Zhou et al., 2017). Importantly, this interaction enhances coronavirus replication, as confirmed by manipulating DDX1 expression, either by small interfering RNA-induced knockdown or by overexpression of a mutant DDX1 protein (Xu et al., 2010). As previously explained, Nsp14 is involved in replication and transcription of the viral genomic and subgenomic RNAs. Therefore, a possibility for the observed enhancement of CoV replication is that DDX1 binding to the Exo domain may facilitate this process. Structurally, the interaction between Nsp14 and DDX1 involves the C-terminal portion of the DDX1 helicase, containing motifs V and VI, and the N-terminal Exo domain of Nsp14, although details of the interaction are currently unknown (Xu et al., 2010) (Figure 10). Further studies are also needed to elucidate the effect of DDX1 interaction with Nsp14 on the exoribonuclease and N7-guanine MTase activities catalyzed by Nsp14.

It is interesting to note that, in IBV-infected cells, DDX1 relocates from the nucleus to the cytoplasm with a predominant



staining pattern in the viral RNA replication site, suggesting that the interaction of Nsp14 could alter the subcellular localization of DDX1 (Xu et al., 2010). Also, because phosphorylation protein N does not affect the interaction between DDX1 and Nsp14, it is likely that the two interaction patterns follow independent roots (Wu et al., 2014).

Interestingly, an opposite effect of the interaction between Nsp14 and DDX1 was observed in TGEV, with DDX1 showing antiviral activity against TGEV replication (Zhou et al., 2017). Nsp14 was shown to induce a DDX1-dependent IFN- $\beta$  production via NF- $\kappa$ B pathway in TGEV infection (Zhou et al., 2017). Indeed, knockdown of DDX1 significantly decreased Nsp14-induced IFN- $\beta$  production and NF- $\kappa$ B activation (Zhou et al., 2017). As discussed above, DDX1 and helicases DDX21 and DHX36 are involved in sensing viral dsRNA and inducing IFN- $\beta$  production. Therefore, Nsp14 of TGEV may be recognized as a Pathogen Associated Molecular Pattern molecule (PAMP) by the DDX1 portion of the DDX1-DDX21-DHX36 viral sensor (Zhang et al., 2011). This antiviral effect of DDX1 in TGEV witnesses a different regulation of CoVs compared to other bacteria.

## Nsp13 and DDX5, a Trick to Enhance Viral RNA Unwinding?

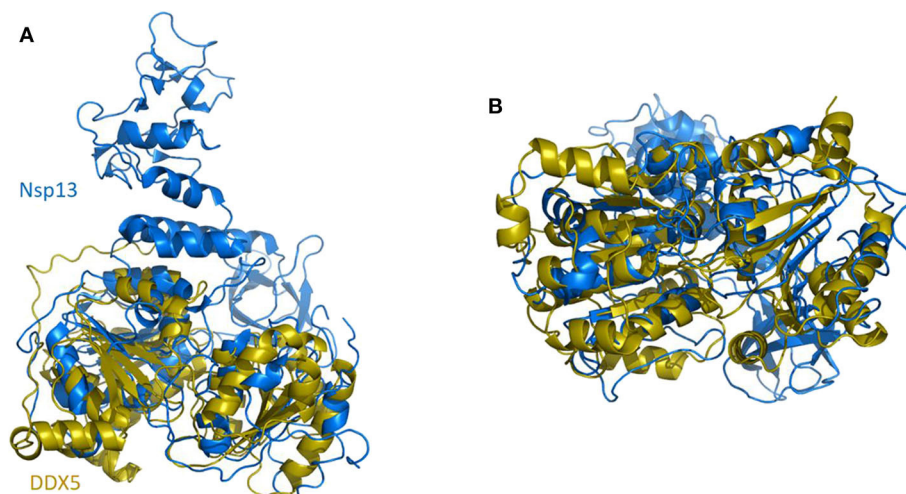
In SARS-CoV replication, the helicase Nsp13 interacts with DDX5 (Chen J. Y. et al., 2009). Interestingly, inhibition of DDX5 by RNA interference results in the suppression of viral replication (Chen J. Y. et al., 2009). This finding indicates a proviral function of DDX5 in coronavirus infection through Nsp13 binding, suggesting that the host helicase may act as a coactivator to enhance viral genome transcription and virus proliferation. A search in the DALI database shows that Nsp13 structurally resembles several homologous human DDX helicases, as judged by the high values of Z scores and low root mean square deviation (rmsd) between backbone atoms (Table 2). This structural similarity occurs despite the low sequence identity between

Nsp13 and these helicases (ranging between 8 and 12%, Table 2), which include DDX5 (Table 2, Figure 11).

As previously described, crystallographic and solution structural studies of DDX3X have proven that two DDX molecules function cooperatively to unwind dsRNA, in a mechanism that has been proposed as a general unwinding mechanism of DDX helicases (Song and Ji, 2019), and therefore applying to DDX5. Although which regions of Nsp13 and DDX5 are involved in binding is unknown, it is tempting to speculate that Nsp13 may form a pre-unwound dimer with DDX5, as observed for DDX3X and proposed for all DDX helicases (Song and Ji, 2019). Consistent with this hypothesis, recent studies have reported that Nsp13 of SARS-CoV has a strong affinity for duplex RNA as a substrate for unwinding, in a reaction which requires high ATP concentrations to unwind duplex RNA (Jang et al., 2020). By crosslinking dsRNA in conjunction with DDX5, Nsp13 would be able to efficiently unwind viral RNA even in the early stages of infection, when the concentration of Nsp13 is still low, by exploiting its interaction with DDX5.

## Sequestration of DDX5 as Possible Viral Evasion Mechanism From Inflammatory Response

Recently, DDX5 has been shown to interact with the diacylglycerol kinase  $\zeta$  (DGK $\zeta$ ), an activator of the NF- $\kappa$ B transcription factor, an essential innate immunity/inflammation modulator. Knockdown of DDX5 only repressed NF- $\kappa$ B transcriptional activity, thus attenuating this essential branch of the innate immune response (Tanaka et al., 2020). Thus, similarly to the case of DDX3X and DDX1 reported above, binding of Nsp13 to DDX5 may play a dual role: enhancing viral RNA transcription and repressing innate immunity.



**FIGURE 11 |** Superposition of SARS-CoV-2 Nsp13 on DDX5 homology model. Cartoon representation of DALI superposed Nsp13 (blue) and the D1D2 core of DDX5 homology model (gold). **(A)** Side and **(B)** bottom view.

## CONCLUSIONS

While human coronaviruses usually cause mild symptoms, three highly pathogenic coronaviruses have emerged in the last years, causing serious diseases: SARS-CoV in 2002, MERS-CoV in 2012, and SARS-CoV-2 in 2019, responsible for the current COVID-19 pandemic. A detailed understanding of the SARS-CoV-2 life cycle and its interactions with the host cell proteome is mandatory to develop effective therapeutic strategies and is currently being actively pursued. However, while such studies are being carried out for SARS-CoV-2, precious information can also be derived from previous studies on the other coronaviruses. In fact, both structural and functional studies have identified host cell proteins as molecular players of coronaviral hijacking mechanisms and described ways exploited by the virus to replicate more efficiently by taking advantage of its host.

As discussed in this review, various DDX helicases, including DDX3X, DDX5, and DDX1, are known to play important roles in the replication cycle of coronaviruses. The viral proteins responsible for DDX-mediated hijacking mechanisms are highly conserved among coronaviruses, an observation that suggests common pathways used by these viruses to exploit host proteins for their own advantage. Here, based on available structural data integrated with homology modeling, we explore possible interactions between human DDX and coronavirus proteins. It appears that the explored mechanisms work to modulate different aspects of viral life. In many instances, sequestration of DDX helicases can have a dual pro-viral role: enhancing key steps of the virus life cycle such as RNA replication/transcription, and, at the same time, repressing the innate immune response at various levels. Based on the structural similarity of the involved viral proteins among different coronaviruses, we thus hypothesize that similar mechanisms are operating also for SARS-CoV-2. Thus, we suggest that DDX helicases could represent useful novel targets for antiviral therapy also against

SARS-CoV-2, as already validated for other RNA viruses (Brai et al., 2016, 2019, 2020b). Indeed, proteomic studies have identified interactions between SARS-CoV-2 proteins and various DDX helicases. However, work needs to be done to understand the roles that these host proteins play in the viral life cycle. Since DDX helicases are not essential for cell viability, gene silencing/disruption approaches could be used to measure the infection cycle of SARS-CoV-2 in cells ablated for specific DDX proteins. In addition, biophysical studies could be used to map the interaction domains between DDX and the cognate viral proteins. Also, for those viral proteins endowed with enzymatic activities, such as Nsp13 and Nsp14, biochemical studies could reveal the effects of DDX interactions on their individual catalytic activities.

## AUTHOR CONTRIBUTIONS

RB, FS, and AR: conceptualization. RB, MR, and AR: methodology and software. MR, FS, and AR: formal analysis. FS and RB: data curation. RB: original draft preparation, supervision, and funding acquisition. RB, FS, and GM: writing review and editing. All authors contributed to the article and approved the submitted version.

## FUNDING

We acknowledge funding by Regione Campania, project “RECOVER-COVID19” (RicErCa e sviluppOVERsus COVID19 in Campania) POR FESR CAMPANIA 2014–2020.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fchem.2020.602162/full#supplementary-material>

## REFERENCES

- Adedeji, A. O., Marchand, B., Te Velhuis, A. J., Snijder, E. J., Weiss, S., Eoff, R. L., et al. (2012). Mechanism of nucleic acid unwinding by SARS-CoV helicase. *PLoS ONE* 7:e36521. doi: 10.1371/journal.pone.0036521
- Adedeji, A. O., Singh, K., Kassim, A., Coleman, C. M., Elliott, R., Weiss, S. R., et al. (2014). Evaluation of SSYA10-001 as a replication inhibitor of severe acute respiratory syndrome, mouse hepatitis, and middle east respiratory syndrome coronaviruses. *Antimicrob. Agents Chemother.* 58, 4894–4898. doi: 10.1128/AAC.02994-14
- Bouvet, M., Imbert, I., Subissi, L., Gluais, L., Canard, B., and Decroly, E. (2012). RNA 3'-end mismatch excision by the severe acute respiratory syndrome coronavirus nonstructural protein nsp10/nsp14 exoribonuclease complex. *Proc. Natl. Acad. Sci. U.S.A.* 109, 9372–9377. doi: 10.1073/pnas.1201130109
- Bouvet, M., Lugari, A., Posthuma, C. C., Zevenhoven, J. C., Bernard, S., Betzi, S., et al. (2014). Coronavirus Nsp10, a critical co-factor for activation of multiple replicative enzymes. *J. Biol. Chem.* 289, 25783–25796. doi: 10.1074/jbc.M114.577353
- Brai, A., Bocuto, A., Monti, M., Marchi, S., Vicenti, I., Saladini, F., et al. (2020a). Exploring the implication of DDX3X in DENV infection: discovery of the first-in-class DDX3X fluorescent inhibitor. *ACS Med. Chem. Lett.* 11, 956–962. doi: 10.1021/acsmchemlett.9b00681
- Brai, A., Fazi, R., Tintori, C., Zamperini, C., Bugli, F., Sanguinetti, M., et al. (2016). Human DDX3 protein is a valuable target to develop broad spectrum antiviral agents. *Proc. Natl. Acad. Sci. U.S.A.* 113, 5388–5393. doi: 10.1073/pnas.1522987113
- Brai, A., Martelli, F., Riva, V., Garbelli, A., Fazi, R., Zamperini, C., et al. (2019). DDX3X helicase inhibitors as a new strategy to fight the west Nile virus infection. *J. Med. Chem.* 62, 2333–2347. doi: 10.1021/acs.jmedchem.8b01403
- Brai, A., Riva, V., Saladini, F., Zamperini, C., Trivisani, C. I., Garbelli, A., et al. (2020b). DDX3X inhibitors, an effective way to overcome HIV-1 resistance targeting host proteins. *Eur. J. Med. Chem.* 200:112319. doi: 10.1016/j.ejmech.2020.112319
- Byrd, A. K., and Raney, K. D. (2012). Superfamily 2 helicases. *Front. Biosci.* 17, 2070–2088. doi: 10.2741/4038
- Byszewska, M., Smietanski, M., Purta, E., and Bujnicki, J. M. (2014). RNA methyltransferases involved in 5' cap biosynthesis. *RNA Biol.* 11, 1597–1607. doi: 10.1080/15476286.2015.1004955
- Chang, C. K., Chen, C. M., Chiang, M. H., Hsu, Y. L., and Huang, T. H. (2013). Transient oligomerization of the SARS-CoV N protein—implication for virus ribonucleoprotein packaging. *PLoS ONE* 8:e65045. doi: 10.1371/journal.pone.0065045
- Chang, C. K., Hou, M. H., Chang, C. F., Hsiao, C. D., and Huang, T. H. (2014). The SARS coronavirus nucleocapsid protein—forms and functions. *Antiviral Res.* 103, 39–50. doi: 10.1016/j.antiviral.2013.12.009
- Chang, C. K., Hsu, Y. L., Chang, Y. H., Chao, F. A., Wu, M. C., Huang, Y. S., et al. (2009). Multiple nucleic acid binding sites and intrinsic disorder of severe acute respiratory syndrome coronavirus nucleocapsid protein: implications for ribonucleocapsid protein packaging. *J. Virol.* 83, 2255–2264. doi: 10.1128/JVI.02001-08
- Chang, C. K., Sue, S. C., Yu, T. H., Hsieh, C. M., Tsai, C. K., Chiang, Y. C., et al. (2005). The dimer interface of the SARS coronavirus nucleocapsid protein adapts a porcine respiratory and reproductive syndrome virus-like structure. *FEBS Lett.* 579, 5663–5668. doi: 10.1016/j.febslet.2005.09.038
- Chen, C. Y., Chang, C. K., Chang, Y. W., Sue, S. C., Bai, H. I., Riag, L., et al. (2007). Structure of the SARS coronavirus nucleocapsid protein RNA-binding dimerization domain suggests a mechanism for helical packaging of viral RNA. *J. Mol. Biol.* 368, 1075–1086. doi: 10.1016/j.jmb.2007.02.069
- Chen, I. J., Yuann, J. M., Chang, Y. M., Lin, S. Y., Zhao, J., Perlman, S., et al. (2013). Crystal structure-based exploration of the important role of Arg106 in the RNA-binding domain of human coronavirus OC43 nucleocapsid protein. *Biochim. Biophys. Acta* 1834, 1054–1062. doi: 10.1016/j.bbapap.2013.03.003
- Chen, J. Y., Chen, W. N., Poon, K. M., Zheng, B. J., Lin, X., Wang, Y. X., et al. (2009). Interaction between SARS-CoV helicase and a multifunctional cellular protein (Ddx5) revealed by yeast and mammalian cell two-hybrid systems. *Arch. Virol.* 154, 507–512. doi: 10.1007/s00705-009-0323-y
- Chen, Y., Cai, H., Pan, J., Xiang, N., Tien, P., Ahola, T., et al. (2009). Functional screen reveals SARS coronavirus nonstructural protein nsp14 as a novel cap N7 methyltransferase. *Proc. Natl. Acad. Sci. U.S.A.* 106, 3484–3489. doi: 10.1073/pnas.0808790106
- Chen, Y., Tao, J., Sun, Y., Wu, A., Su, C., Gao, G., et al. (2013). Structure-function analysis of severe acute respiratory syndrome coronavirus RNA cap guanine-N7-methyltransferase. *J. Virol.* 87, 6296–6305. doi: 10.1128/JVI.00061-13
- Cheng, W., Chen, G., Jia, H., He, X., and Jing, Z. (2018). DDX5 RNA helicases: emerging roles in viral infection. *Int. J. Mol. Sci.* 19:1122. doi: 10.3390/ijms19041122
- Choi, Y. J., and Lee, S. G. (2012). The DEAD-box RNA helicase DDX3 interacts with DDX5, co-localizes with it in the cytoplasm during the G2/M phase of the cycle, and affects its shuttling during mRNP export. *J. Cell. Biochem.* 113, 985–996. doi: 10.1002/jcb.23428
- Chouhan, B. P. S., Maimaiti, S., Gade, M., and Laurino, P. (2019). Rossmann-fold methyltransferases: taking a “beta-turn” around their cofactor, S-adenosylmethionine. *Biochemistry* 58, 166–170. doi: 10.1021/acs.biochem.8b00994
- Cong, Y., Ulasli, M., Schepers, H., Mauthe, M., V’Kovski, P., Kriegenburg, F., et al. (2020). Nucleocapsid protein recruitment to replication-transcription complexes plays a crucial role in coronavirus life cycle. *J. Virol.* 94, e01925–19. doi: 10.1128/JVI.01925-19
- D’Cruz, A. A., Babon, J. J., Norton, R. S., Nicola, N. A., and Nicholson, S. E. (2013). Structure and function of the SPRY/B30.2 domain proteins involved in innate immunity. *Protein Sci.* 22, 1–10. doi: 10.1002/pro.2185
- de Haan, C. A., and Rottier P. J. (2005). Molecular interactions in the assembly of coronaviruses. *Adv. Virus Res.* 64, 165–230. doi: 10.1016/S0065-3527(05)64006-7
- Decroly, E., Ferron, F., Lescar, J., and Canard, B. (2012). Conventional and unconventional mechanisms for capping viral mRNA. *Nat. Rev. Microbiol.* 10, 51–65. doi: 10.1038/nrmicro2675
- Del Campo, M., and Lambowitz, A. M. (2009). Structure of the Yeast DEAD box protein Mss116p reveals two wedges that crimp RNA. *Mol. Cell.* 35, 598–609. doi: 10.1016/j.molcel.2009.07.032
- Denison, M. R., Graham, R. L., Donaldson, E. F., Eckerle, L. D., and Baric, R. S. (2011). Coronaviruses: an RNA proofreading machine regulates replication fidelity and diversity. *RNA Biol.* 8, 270–279. doi: 10.4161/rna.8.2.15013
- Derbyshire, V., Grindley, N. D., and Joyce, C. M. (1991). The 3'-5' exonuclease of DNA polymerase I of Escherichia coli: contribution of each amino acid at the active site to the reaction. *EMBO J.* 10, 17–24. doi: 10.1002/j.1460-2075.1991.tb07916.x
- Diot, C., Fournier, G., Dos Santos, M., Magnus, J., Komarova, A., van der Werf, S. et al. (2016). Influenza A virus polymerase recruits the RNA helicase DDX19 to promote the nuclear export of viral mRNAs. *Sci. Rep.* 6:33763. doi: 10.1038/srep33763
- Eckerle, L. D., Becker, M. M., Halpin, R. A., Li, K., Venter, E., Lu, X., et al. (2010). Infidelity of SARS-CoV Nsp14-exonuclease mutant virus replication is revealed by complete genome sequencing. *PLoS Pathog.* 6:e1000896. doi: 10.1371/journal.ppat.1000896
- Eckerle, L. D., Lu, X., Sperry, S. M., Choi, L., and Denison, M. R. (2007). High fidelity of murine hepatitis virus replication is decreased in nsp14 exoribonuclease mutants. *J. Virol.* 81, 12135–12144. doi: 10.1128/JVI.01296-07
- Emmott, E., Munday, D., Bickerton, E., Britton, P., Rodgers, M. A., Whitehouse, A., et al. (2013). The cellular interactome of the coronavirus infectious bronchitis virus nucleocapsid protein and functional implications for virus biology. *J. Virol.* 87, 9486–9500. doi: 10.1128/JVI.00321-13
- Fang, J. H., Kubota, S., Yang, B., Zhou, N. M., Zhang, H., Godbout, R., et al. (2004). A DEAD box protein facilitates HIV-1 replication as a cellular co-factor of Rev. *Virology* 330, 471–480. doi: 10.1016/j.virol.2004.09.039
- Ferrage, F., Dutta, K., Nistal-Villan, E., Patel, J. R., Sanchez-Aparicio, M. T., De Ioannes, P., et al. (2012). Structure and dynamics of the second CARD of human RIG-I provide mechanistic insights into regulation of RIG-I activation. *Structure* 20, 2048–2061. doi: 10.1016/j.str.2012.09.003
- Ferron, F., Subissi, L., De Morai, A. T. S., Le, N. T. T., Sevajol, M., Gluais, L., et al. (2018). Structural and molecular basis of mismatch correction and ribavirin excision from coronavirus RNA. *Proc. Natl. Acad. Sci. U.S.A.* 115, E162–E171. doi: 10.1073/pnas.1718806115
- Goh, P. Y., Tan, Y. J., Lim, S. P., Tan, Y. H., Lim, S. G., Fuller-Pace, F., et al. (2004). Cellular RNA helicase p68 relocation and interaction with the hepatitis

- C virus (HCV) NS5B protein and the potential role of p68 in HCV RNA replication. *J. Virol.* 78, 5288–5298. doi: 10.1128/JVI.78.10.5288-5298.2004
- Gordon, D. E., Jang, G. M., Bouhaddou, M., Xu, J., Obernier, K., White, K. M., et al. (2020). A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. *Nature* 583, 459–468. doi: 10.1038/s41586-020-2286-9
- Gu, L., Fullam, A., Brennan, R., and Schroder, M. (2013). Human DEAD box helicase 3 couples IkappaB kinase epsilon to interferon regulatory factor 3 activation. *Mol. Cell. Biol.* 33, 2004–2015. doi: 10.1128/MCB.01603-12
- Gu, L., Fullam, A., McCormack, N., Hohn, Y., and Schroder, M. (2017). DDX3 directly regulates TRAF3 ubiquitination and acts as a scaffold to co-ordinate assembly of signalling complexes downstream from MAVS. *Biochem. J.* 474, 571–587. doi: 10.1042/BCJ20160956
- Hao, W., Wojdyla, J. A., Zhao, R., Han, R., Das, R., Zlatev, I., et al. (2017). Crystal structure of Middle East respiratory syndrome coronavirus helicase. *PLoS Pathog.* 13:e1006474. doi: 10.1371/journal.ppat.1006474
- He, R., Dobie, F., Ballantine, M., Leeson, A., Li, Y., Bastien, N., et al. (2004). Analysis of multimerization of the SARS coronavirus nucleocapsid protein. *Biochem. Biophys. Res. Commun.* 316, 476–483. doi: 10.1016/j.bbrc.2004.02.074
- He, Y., Zhou, Y., Wu, H., Kou, Z., Liu, S., and Jiang, S. (2004). Mapping of antigenic sites on the nucleocapsid protein of the severe acute respiratory syndrome coronavirus. *J. Clin. Microbiol.* 42, 5309–5314. doi: 10.1128/JCM.42.11.5309-5314.2004
- Heaton, S. M., Atkinson, S. C., Sweeney, M. N., Yang, S. N. Y., Jans, D. A., and Borg, N. A. (2019). Exportin-1-dependent nuclear export of DEAD-box helicase DDX3X is central to its role in antiviral immunity. *Cells* 8:1181. doi: 10.3390/cells8101181
- Hilbert, M., Karow, A. R., and Klostermeier, D. (2009). The mechanism of ATP-dependent RNA unwinding by DEAD box proteins. *Biol. Chem.* 390, 1237–1250. doi: 10.1515/BC.2009.135
- Hoffmann, M., Kleine-Weber, H., Schroeder, S., Krüger, N., Herrler, T., Erichsen, S., et al. (2020). SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. *Cell* 181, 271–280.e8. doi: 10.1016/j.cell.2020.02.052
- Huang, Q., Yu, L., Petros, A. M., Gunasekera, A., Liu, Z., Xu, N., et al. (2004). Structure of the N-terminal RNA-binding domain of the SARS CoV nucleocapsid protein. *Biochemistry* 43, 6059–6063. doi: 10.1021/bi036155b
- Hurst, K. R., Ye, R., Goebel, S. J., Jayaraman, P., and Masters, P. S. (2010). An interaction between the nucleocapsid protein and a component of the replicase-transcriptase complex is crucial for the infectivity of coronavirus genomic RNA. *J. Virol.* 84, 10276–10288. doi: 10.1128/JVI.01287-10
- Ivanov, K. A., and Ziebuhr, J. (2004). Human coronavirus 229E nonstructural protein 13: characterization of duplex-unwinding, nucleoside triphosphatase, and RNA 5'-triphosphatase activities. *J. Virol.* 78, 7833–7838. doi: 10.1128/JVI.78.14.7833-7838.2004
- Jang, K. J., Jeong, S., Kang, D. Y., Sp, N., Yang, Y. M., and Kim, D. E. (2020). A high ATP concentration enhances the cooperative translocation of the SARS coronavirus helicase nsP13 in the unwinding of duplex RNA. *Sci. Rep.* 10:4481. doi: 10.1038/s41598-020-61432-1
- Jankowsky, E., and Fairman, M. E. (2007). RNA helicases—one fold for many functions. *Curr. Opin. Struct. Biol.* 17, 316–324. doi: 10.1016/j.sbi.2007.05.007
- Jarmoskaite, I., and Russell, R. (2011). DEAD-box proteins as RNA helicases and chaperones. *Wiley Interdiscip. Rev. RNA* 2, 135–152. doi: 10.1002/wrna.50
- Jayaram, H., Fan, H., Bowman, B. R., Ooi, A., Jayaram, J., Collisson, E. W., et al. (2006). X-ray structures of the N- and C-terminal domains of a coronavirus nucleocapsid protein: implications for nucleocapsid formation. *J. Virol.* 80, 6612–6620. doi: 10.1128/JVI.00157-06
- Jia, Z., Yan, L., Ren, Z., Wu, L., Wang, J., Guo, J., et al. (2019). Delicate structural coordination of the Severe Acute Respiratory Syndrome coronavirus Nsp13 upon ATP hydrolysis. *Nucleic Acids Res.* 47, 6538–6550. doi: 10.1093/nar/gkz409
- Kellner, J. N., and Meinhart, A. (2015). Structure of the SPRY domain of the human RNA helicase DDX1, a putative interaction platform within a DEAD-box protein. *Acta Crystallogr. F Struct. Biol. Commun.* 71, 1176–1188. doi: 10.1107/S2053230X15013709
- Kellner, J. N., Reinstein, J., and Meinhart, A. (2015). Synergistic effects of ATP and RNA binding to human DEAD-box protein DDX1. *Nucleic Acids Res.* 43, 2813–2828. doi: 10.1093/nar/gkv106
- Lehmann, K. C., Snijder, E. J., Posthuma, C. C., and Gorbalenya, A. E. (2015). What we know but do not understand about nidovirus helicases. *Virus Res.* 202, 12–32. doi: 10.1016/j.virusres.2014.12.001
- Lin, L., Shao, J., Sun, M., Liu, J., Xu, G., Zhang, X., et al. (2007). Identification of phosphorylation sites in the nucleocapsid protein (N protein) of SARS-coronavirus. *Int. J. Mass Spectrom.* 268, 296–303. doi: 10.1016/j.ijms.2007.05.009
- Linder, P., and Jankowsky, E. (2011). From unwinding to clamping - the DEAD box RNA helicase family. *Nat. Rev. Mol. Cell Biol.* 12, 505–516. doi: 10.1038/nrm3154
- Lo, Y. S., Lin, S. Y., Wang, S. M., Wang, C. T., Chiu, Y. L., Huang, T. H., et al. (2013). Oligomerization of the carboxyl terminal domain of the human coronavirus 229E nucleocapsid protein. *FEBS Lett.* 587, 120–127. doi: 10.1016/j.febslet.2012.11.016
- Luan, J., Lu, Y., Jin, X., and Zhang, L. (2020). Spike protein recognition of mammalian ACE2 predicts the host range and an optimized ACE2 for SARS-CoV-2 infection. *Biochem. Biophys. Res. Commun.* 526, 165–169. doi: 10.1016/j.bbrc.2020.03.047
- Lukassen, S., Chua, R. L., Trefzer, T., Kahn, N. C., Schneider, M. A., Muley, T., et al. (2020). SARS-CoV-2 receptor ACE2 and TMPRSS2 are primarily expressed in bronchial transient secretory cells. *EMBO J.* 39:e105114. doi: 10.15252/embj.20105114
- Luo, H., Chen, J., Chen, K., Shen, X., and Jiang, H. (2006). Carboxyl terminus of severe acute respiratory syndrome coronavirus nucleocapsid protein: self-association analysis and nucleic acid binding characterization. *Biochemistry* 45, 11827–11835. doi: 10.1021/bi0609319
- Ma, Y., Wu, L., Shaw, N., Gao, Y., Wang, J., Sun, Y., et al. (2015). Structural basis and functional analysis of the SARS coronavirus nsp14-nsp10 complex. *Proc. Natl. Acad. Sci. U.S.A.* 112, 9436–9441. doi: 10.1073/pnas.1508686112
- Mallam, A. L., Jarmoskaite, I., Tijerina, P., Del Campo, M., Seifert, S., Guo, L., et al. (2011). Solution structures of DEAD-box RNA chaperones reveal conformational changes and nucleic acid tethering by a basic tail. *Proc. Natl. Acad. Sci. U.S.A.* 108, 12254–12259. doi: 10.1073/pnas.1109566108
- McBride, R., van Zyl, M., and Fielding, B. C. (2014). The coronavirus nucleocapsid is a multifunctional protein. *Viruses* 6, 2991–3018. doi: 10.3390/v6082991
- Minskaia, E., Hertzog, T., Gorbalenya, A. E., Campanacci, V., Cambillau, C., Canard, B., et al. (2006). Discovery of an RNA virus 3'->5' exoribonuclease that is critically involved in coronavirus RNA synthesis. *Proc. Natl. Acad. Sci. U.S.A.* 103, 5108–5113. doi: 10.1073/pnas.0508200103
- Narayanan, K., Maeda, A., Maeda, J., and Makino, S. (2000). Characterization of the coronavirus M protein and nucleocapsid interaction in infected cells. *J. Virol.* 74, 8127–8134. doi: 10.1128/JVI.74.17.8127-8134.2000
- Ngo, T. D., Partin, A. C., and Nam, Y. (2019). RNA Specificity and Autoregulation of DDX17, a Modulator of MicroRNA Biogenesis. *Cell Rep.* 29, 4024–4035.e5. doi: 10.1016/j.celrep.2019.11.059
- Ogando, N. S., Ferron, F., Decroly, E., Canard, B., Posthuma, C. C., and Snijder, E. J. (2019). The curious case of the nidovirus exoribonuclease: Its Role in RNA synthesis and replication fidelity. *Front. Microbiol.* 10:1813. doi: 10.3389/fmicb.2019.01813
- Ogilvie, V. C., Wilson, B. J., Nicol, S. M., Morrice, N. A., Saunders, L. R., Barber, G. N., et al. (2003). The highly related DEAD box RNA helicases p68 and p72 exist as heterodimers in cells. *Nucleic Acids Res.* 31, 1470–1480. doi: 10.1093/nar/gkg236
- Oshiumi, H., Sakai, K., Matsumoto, M., and Seya, T. (2010). DEAD/H BOX 3 (DDX3) helicase binds the RIG-I adaptor IPS-1 to up-regulate IFN-beta-inducing potential. *Eur. J. Immunol.* 40, 940–948. doi: 10.1002/eji.200940203
- Patel, S. S., and Donmez, I. (2006). Mechanisms of helicases. *J. Biol. Chem.* 281, 18265–18268. doi: 10.1074/jbc.R60008200
- Peng, T. Y., Lee, K. R., and Tarn, W. Y. (2008). Phosphorylation of the arginine/serine dipeptide-rich motif of the severe acute respiratory syndrome coronavirus nucleocapsid protein modulates its multimerization, translation inhibitory activity and cellular localization. *FEBS J.* 275, 4152–4163. doi: 10.1111/j.1742-4658.2008.06564.x
- Putnam, A. A., Gao, Z., Liu, F., Jia, H., Yang, Q., and Jankowsky, E. (2015). Division of labor in an oligomer of the DEAD-Box RNA helicase ded1p. *Mol. Cell.* 59, 541–552. doi: 10.1016/j.molcel.2015.06.030
- Rajyaguru, P., and Parker, R. (2012). RGG motif proteins: modulators of mRNA functional states. *Cell Cycle* 11, 2594–2599. doi: 10.4161/cc.20716

- Ranji, A., and Boris-Lawrie, K. (2010). RNA helicases: emerging roles in viral replication and the host innate response. *RNA Biol.* 7, 775–787. doi: 10.4161/rna.7.6.14249
- Rao, S. T., and Rossmann, M. G. (1973). Comparison of super-secondary structures in proteins. *J. Mol. Biol.* 76, 241–256. doi: 10.1016/0022-2836(73)90388-4
- Romano, M., Ruggiero, A., Squeglia, F., and Berisio, R. (2020a). An engineered stable mini-protein to plug SARS-CoV-2 spikes. *bioRxiv [Preprint]*. doi: 10.1101/2020.04.29.067728
- Romano, M., Ruggiero, A., Squeglia, F., Maga, G., and Berisio, R. (2020b). A structural view of SARS-CoV-2 RNA replication machinery: RNA synthesis, proofreading and final capping. *Cells* 9:1267. doi: 10.3390/cells9051267
- Romano, M., Squeglia, F., and Berisio, R. (2015). Structure and function of RNase AS: a novel virulence factor from mycobacterium tuberculosis. *Curr. Med. Chem.* 22, 1745–1756. doi: 10.12174/0929867322666150417125301
- Romano, M., van de Weerd, R., Brouwer, F. C., Roviello, G. N., Lacroix, R., Sparrius, M., et al. (2014). Structure and function of RNase AS, a polyadenylate-specific exoribonuclease affecting mycobacterial virulence *in vivo*. *Structure* 22, 719–730. doi: 10.1016/j.str.2014.01.014
- Rudolph, M. G., and Klostermeier, D. (2015). When core competence is not enough: functional interplay of the DEAD-box helicase core with ancillary domains and auxiliary factors in RNA binding and unwinding. *Biol. Chem.* 396, 849–865. doi: 10.1515/hsz-2014-0277
- Saikatendu, K. S., Joseph, J. S., Subramanian, V., Neuman, B. W., Buchmeier, M. J., Stevens, R. C., et al. (2007). Ribonucleocapsid formation of severe acute respiratory syndrome coronavirus through molecular action of the N-terminal domain of N protein. *J. Virol.* 81, 3913–3921. doi: 10.1128/JVI.02236-06
- Schroder, M., Baran, M., and Bowie, A. G. (2008). Viral targeting of DEAD box protein 3 reveals its role in TBK1/IKKepsilon-mediated IRF activation. *EMBO J.* 27, 2147–2157. doi: 10.1038/emboj.2008.143
- Schubert, H. L., Blumenthal, R. M., and Cheng, X. D. (2003). Many paths to methyltransferase: a chronicle of convergence. *Trends Biochem. Sci.* 28, 329–335. doi: 10.1016/S0968-0004(03)00090-2
- Schutz, P., Karlberg, T., van den Berg, S., Collins, R., Lehtio, L., Hogbom, M., et al. (2010). Comparative structural analysis of human DEAD-box RNA helicases. *PLoS ONE* 5:e12791. doi: 10.1371/journal.pone.0012791
- Sharma, A., and Boris-Lawrie, K. (2012). Determination of host rna helicases activity in viral replication. *Method Enzymol.* 511, 405–435. doi: 10.1016/B978-0-12-396546-2.00019-X
- Shum, K. T., and Tanner, J. A. (2008). Differential inhibitory activities and stabilisation of DNA aptamers against the SARS coronavirus helicase. *ChemBiochem* 9, 3037–3045. doi: 10.1002/cbic.200800491
- Siu, Y. L., Teoh, K. T., Lo, J., Chan, C. M., Kien, F., Escρίου, N., et al. (2008). The M, E, and N structural proteins of the severe acute respiratory syndrome coronavirus are required for efficient assembly, trafficking, and release of virus-like particles. *J. Virol.* 82, 11318–11330. doi: 10.1128/JVI.01052-08
- Song, H., and Ji, X. (2019). The mechanism of RNA duplex recognition and unwinding by DEAD-box helicase DDX3X. *Nat. Commun.* 10:3085. doi: 10.1038/s41467-019-11083-2
- Soulat, D., Burckstummer, T., Westermayer, S., Goncalves, A., Bauch, A., Stefanovic, A., et al. (2008). The DEAD-box helicase DDX3X is a critical component of the TANK-binding kinase 1-dependent innate immune response. *EMBO J.* 27, 2135–2146. doi: 10.1038/emboj.2008.126
- Spencer, K. A., Dee, M., Britton, P., and Hiscox, J. A. (2008). Role of phosphorylation clusters in the biology of the coronavirus infectious bronchitis virus nucleocapsid protein. *Virology* 370, 373–381. doi: 10.1016/j.virol.2007.08.016
- Steimer, L., and Klostermeier, D. (2012). RNA helicases in infection and disease. *RNA Biol.* 9, 751–771. doi: 10.4161/rna.20090
- Surjit, M., Kumar, R., Mishra, R. N., Reddy, M. K., Chow, V. T., and Lal, S. K. (2005). The severe acute respiratory syndrome coronavirus nucleocapsid protein is phosphorylated and localizes in the cytoplasm by 14-3-3-mediated translocation. *J. Virol.* 79, 11476–11486. doi: 10.1128/JVI.79.17.11476-11486.2005
- Takeda, M., Chang, C. K., Ikeya, T., Guntert, P., Chang, Y. H., Hsu, Y. L., et al. (2008). Solution structure of the c-terminal dimerization domain of SARS coronavirus nucleocapsid protein solved by the SAIL-NMR method. *J. Mol. Biol.* 380, 608–622. doi: 10.1016/j.jmb.2007.11.093
- Tanaka, K., Tanaka, T., Nakano, T., Hozumi, Y., Yanagida, M., Araki, Y., et al. (2020). Knockdown of DEAD-box RNA helicase DDX5 selectively attenuates serine 311 phosphorylation of NF-kappaB p65 subunit and expression level of anti-apoptotic factor Bcl-2. *Cell. Signal.* 65:109428. doi: 10.1016/j.cellsig.2019.109428
- Taschuk, F., and Cherry, S. (2020). DEAD-box helicases: sensors, regulators, and effectors for antiviral defense. *Viruses* 12:181. doi: 10.3390/v12020181
- Theissen, B., Karow, A. R., Kohler, J., Gubaev, A., and Klostermeier, D. (2008). Cooperative binding of ATP and RNA induces a closed conformation in a DEAD box RNA helicase. *Proc. Natl. Acad. Sci. U.S.A.* 105, 548–553. doi: 10.1073/pnas.0705488105
- van Dinten, L. C., and van Tol H, Gorbelenya AE, Snijder EJ. (2000). The predicted metal-binding region of the arterivirus helicase protein is involved in subgenomic mRNA synthesis, genome replication, virion biogenesis. *J. Virol.* 74, 5213–23. doi: 10.1128/74.11.5213-5223.2000
- van Hemert, M. J., van den Worm, S. H. E., Knoops, K., Mommaas, A. M., Gorbelenya, A. E., and Snijder, E. J. (2008). SARS-coronavirus replication/transcription complexes are membrane-protected and need a host factor for activity *in vitro*. *PLoS Pathog.* 4:e1000054. doi: 10.1371/journal.ppat.1000054
- Walsh, D., Mathews, M. B., and Mohr, I. (2013). Tinkering with translation: protein synthesis in virus-infected cells. *Cold Spring Harb. Perspect. Biol.* 5:a012351. doi: 10.1101/cshperspect.a012351
- Wang, X., Wang, R., Luo, M., Li, C., Wang, H. X., Huan, C. C., et al. (2017). (DEAD)-box RNA helicase 3 modulates NF-kappaB signal pathway by controlling the phosphorylation of PP2A-C subunit. *Oncotarget* 8, 33197–33213. doi: 10.18632/oncotarget.16593
- Wang, Y., and Zhang, X. (1999). The nucleocapsid protein of coronavirus mouse hepatitis virus interacts with the cellular heterogeneous nuclear ribonucleoprotein A1 *in vitro* and *in vivo*. *Virology* 265, 96–109. doi: 10.1006/viro.1999.0025
- Wei, W. Y., Li, H. C., Chen, C. Y., Yang, C. H., Lee, S. K., Wang, C. W., et al. (2012). SARS-CoV nucleocapsid protein interacts with cellular pyruvate kinase protein and inhibits its activity. *Arch. Virol.* 157, 635–645. doi: 10.1007/s00705-011-1221-7
- Wrapp, D., Wang, N., Corbett, K. S., Goldsmith, J. A., Hsieh, C. L., Abiona, O., et al. (2020). Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science* 367, 1260–1263. doi: 10.1126/science.abb2507
- Wu, C. H., Chen, P. J., and Yeh, S. H. (2014). Nucleocapsid phosphorylation and RNA helicase DDX1 recruitment enables coronavirus transition from discontinuous to continuous transcription. *Cell Host Microbe* 16, 462–472. doi: 10.1016/j.chom.2014.09.009
- Wu, C. H., Yeh, S. H., Tsay, Y. G., Shieh, Y. H., Kao, C. L., Chen, Y. S., et al. (2009). Glycogen synthase kinase-3 regulates the phosphorylation of severe acute respiratory syndrome coronavirus nucleocapsid protein and viral replication. *J. Biol. Chem.* 284, 5229–5239. doi: 10.1074/jbc.M805747200
- Xia, S., Zhu, Y., Liu, M., Lan, Q., Xu, W., Wu, Y., et al. (2020). Fusion mechanism of 2019-nCoV and fusion inhibitors targeting HR1 domain in spike protein. *Cell. Mol. Immunol.* 17, 765–767. doi: 10.1038/s41423-020-0374-2
- Xu, L., Khadijah, S., Fang, S., Wang, L., Tay, F. P., and Liu, D. X. (2010). The cellular RNA helicase DDX1 interacts with coronavirus nonstructural protein 14 and enhances viral replication. *J. Virol.* 84, 8571–8583. doi: 10.1128/JVI.00392-10
- Yang, N., and Shen, H. M. (2020). Targeting the endocytic pathway and autophagy process as a novel therapeutic strategy in COVID-19. *Int. J. Biol. Sci.* 16, 1724–1731. doi: 10.7150/ijbs.45498
- Yasuda-Inoue, M., Kuroki, M., and Ariumi, Y. (2013). Distinct DDX DEAD-box RNA helicases cooperate to modulate the HIV-1 rev function. *Biochem. Biophys. Res. Commun.* 434, 803–808. doi: 10.1016/j.bbrc.2013.04.016
- Ye, Q., West, A. M. V., Silletti, S., and Corbett, K. D. (2020). Architecture and self-assembly of the SARS-CoV-2 nucleocapsid protein. *bioRxiv [Preprint]*. doi: 10.1002/pro.3909
- Yoo, J. S., Kato, H., and Fujita, T. (2014). Sensing viral invasion by RIG-I like receptors. *Curr. Opin. Microbiol.* 20, 131–138. doi: 10.1016/j.mib.2014.05.011
- Yu, I. M., Gustafson, C. L., Diao, J., Burgner, J. W., Li, Z., Zhang, J., et al. (2005). Recombinant severe acute respiratory syndrome (SARS) coronavirus nucleocapsid protein forms a dimer through its C-terminal domain. *J. Biol. Chem.* 280, 23280–23286. doi: 10.1074/jbc.M501015200

- Zeng, Y., Ye, L., Zhu, S., Zheng, H., Zhao, P., Cai, W., et al. (2008). The nucleocapsid protein of SARS-associated coronavirus inhibits B23 phosphorylation. *Biochem. Biophys. Res. Commun.* 369, 287–291. doi: 10.1016/j.bbrc.2008.01.096
- Zhang, Y. P., Zhang, R. W., Chang, W. S., and Wang, Y. Y. (2010). Cxcl16 interact with SARS-CoV N protein in and out cell. *Virol. Sin.* 25, 369–374. doi: 10.1007/s12250-010-3129-x
- Zhang, Z., Kim, T., Bao, M., Facchinetti, V., Jung, S. Y., Ghaffari, A. A., et al. (2011). DDX1, DDX21, and DHX36 helicases form a complex with the adaptor molecule TRIF to sense dsRNA in dendritic cells. *Immunity* 34, 866–78. doi: 10.1016/j.immuni.2011.03.027
- Zhao, X., Nicholls, J. M., and Chen, Y. G. (2008). Severe acute respiratory syndrome-associated coronavirus nucleocapsid protein interacts with Smad3 and modulates transforming growth factor-beta signaling. *J. Biol. Chem.* 283, 3272–3280. doi: 10.1074/jbc.M708033200
- Zhou, B., Liu, J., Wang, Q., Liu, X., Li, X., Li, P., et al. (2008). The nucleocapsid protein of severe acute respiratory syndrome coronavirus inhibits cell cytokinesis and proliferation by interacting with translation elongation factor 1alpha. *J. Virol.* 82, 6962–6971. doi: 10.1128/JVI.00133-08
- Zhou, Y., Wu, W., Xie, L., Wang, D., Ke, Q., Hou, Z., et al. (2017). Cellular RNA helicase DDX1 is involved in transmissible gastroenteritis virus nsp14-induced interferon-beta production. *Front. Immunol.* 8:940. doi: 10.3389/fimmu.2017.500940
- Zuniga, S., Cruz, J. L., Sola, I., Mateos-Gomez, P. A., Palacio, L., and Enjuanes, L. (2010). Coronavirus nucleocapsid protein facilitates template switching and is required for efficient transcription. *J. Virol.* 84, 2169–2175. doi: 10.1128/JVI.02011-09

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Squeglia, Romano, Ruggiero, Maga and Berisio. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Computational Prediction of Potential Inhibitors of the Main Protease of SARS-CoV-2

Renata Abel<sup>1</sup>, María Paredes Ramos<sup>2</sup>, Qiaofeng Chen<sup>1</sup>, Horacio Pérez-Sánchez<sup>3</sup>, Flaminia Coluzzi<sup>4,5</sup>, Monica Rocco<sup>5,6</sup>, Paolo Marchetti<sup>6</sup>, Cameron Mura<sup>7</sup>, Maurizio Simmaco<sup>8,9</sup>, Philip E. Bourne<sup>7</sup>, Robert Preissner<sup>10</sup> and Priyanka Banerjee<sup>1\*</sup>

<sup>1</sup> Institute of Physiology, Charité—University Medicine Berlin, Berlin, Germany, <sup>2</sup> METMED Research Group, Physical Chemistry Department, Universidade da Coruña (UDC), A Coruña, Spain, <sup>3</sup> Structural Bioinformatics and High-Performance Computing (BIO-HPC) Research Group, Universidad Católica de Murcia (UCAM), Murcia, Spain, <sup>4</sup> Department of Medical and Surgical Sciences and Biotechnologies, Sapienza University of Rome, Latina, Italy, <sup>5</sup> Unit of Anesthesia and Intensive Care Medicine, Sant' Andrea University Hospital, Rome, Italy, <sup>6</sup> Department of Clinical and Surgical Translational Medicine, Sapienza University of Rome, Rome, Italy, <sup>7</sup> Department of Biomedical Engineering and School of Data Science, University of Virginia, Charlottesville, VA, United States, <sup>8</sup> Department of Neurosciences, Mental Health and Sensory Organs, Sapienza University of Rome, Rome, Italy, <sup>9</sup> Advanced Molecular Diagnostic Unit, Sant' Andrea University Hospital, Rome, Italy, <sup>10</sup> Institute of Physiology and Science-IT, Charité—Universitätsmedizin Berlin, Corporate Member of Freie Universität Berlin, Humboldt-Universität zu Berlin, and Berlin Institute of Health, Berlin, Germany

## OPEN ACCESS

### Edited by:

Chandrabose Selvaraj,  
Alagappa University, India

### Reviewed by:

Kshitij Verma,  
Genentech, Inc., United States  
Mahmoud A. A. Ibrahim,  
Minia University, Egypt

### \*Correspondence:

Priyanka Banerjee  
priyanka.banerjee@charite.de

### Specialty section:

This article was submitted to  
Medicinal and Pharmaceutical  
Chemistry,  
a section of the journal  
Frontiers in Chemistry

**Received:** 03 August 2020

**Accepted:** 13 November 2020

**Published:** 23 December 2020

### Citation:

Abel R, Paredes Ramos M, Chen Q, Pérez-Sánchez H, Coluzzi F, Rocco M, Marchetti P, Mura C, Simmaco M, Bourne PE, Preissner R and Banerjee P (2020) Computational Prediction of Potential Inhibitors of the Main Protease of SARS-CoV-2. *Front. Chem.* 8:590263. doi: 10.3389/fchem.2020.590263

The rapidly developing pandemic, known as coronavirus disease 2019 (COVID-19) and caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), has recently spread across 213 countries and territories. This pandemic is a dire public health threat—particularly for those suffering from hypertension, cardiovascular diseases, pulmonary diseases, or diabetes; without approved treatments, it is likely to persist or recur. To facilitate the rapid discovery of inhibitors with clinical potential, we have applied ligand- and structure-based computational approaches to develop a virtual screening methodology that allows us to predict potential inhibitors. In this work, virtual screening was performed against two natural products databases, Super Natural II and Traditional Chinese Medicine. Additionally, we have used an integrated drug repurposing approach to computationally identify potential inhibitors of the main protease of SARS-CoV-2 in databases of drugs (both approved and withdrawn). Roughly 360,000 compounds were screened using various molecular fingerprints and molecular docking methods; of these, 80 docked compounds were evaluated in detail, and the 12 best hits from four datasets were further inspected *via* molecular dynamics simulations. Finally, toxicity and cytochrome inhibition profiles were computationally analyzed for the selected candidate compounds.

**Keywords:** virtual screening (VS), drug repurposing and molecular docking, SARS-CoV-2, COVID-19, computational drug discovery, molecular dynamics

## INTRODUCTION

A novel coronavirus (CoV), known as the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), began spreading among humans in December 2019 in the city of Wuhan, China, causing a major outbreak of often-fatal pneumonia (Wu et al., 2020). The rapid expansion of SARS-CoV-2 has been labeled a pandemic by the World Health Organization (WHO), and the global crisis

has continued to devastate both human health and national economies (WHO<sup>1</sup>). The symptoms associated with most instances of this infection include fever, dry cough, fatigue, shortness of breath and respiratory issues (Wu et al., 2020), and deterioration of some sensory modalities (e.g., taste, smell); a smaller fraction of cases also present with other symptoms, e.g., conjunctivitis (presumably another mode of transmission, too) (Scasso et al., 2018). With SARS-CoV-2, aggressive human–human transmission has occurred, yielding exponential growth in the number of detected cases. The disease has now been termed as “coronavirus disease 2019” (COVID-19) (Zhang L. et al., 2020). At present, the number of confirmed cases reported internationally has reached 15,581,009, with 635,173 deaths reported<sup>2</sup>. As of yet, no potent drug or vaccine has been reported (or approved) to treat individuals infected with SARS-CoV-2; only symptomatic treatment has been given to the most critically ill patients. A surge in activity among the scientific community has advanced research efforts toward the development of therapeutic intervention and finding viral drug targets; currently, 36 repurposed drugs are already used in experimental (unapproved) treatments for COVID-19, and 432 drugs are being tested in ongoing clinical trials<sup>3</sup>. In addition, there are 23 candidate vaccines in clinical evaluation and 140 candidate vaccines in preclinical evaluation<sup>4</sup>. Initial results from a phase 1 clinical trial are already available for a vaccine known as mRNA-1273 (Jackson et al., 2020). Recent reports suggest that some U.S. Food & Drug Administration (FDA)-approved drugs, specifically remdesivir (which inhibits viral RNA polymerase) (Al-Tawfiq et al., 2020) and lopinavir and ritonavir (HIV protease inhibitors) (Cao et al., 2020), may be effective against SARS-CoV-2. Remdesivir exhibits an antiviral activity with an EC<sub>50</sub> of 0.77  $\mu$ M against SARS-CoV-2, and shorter recovery times (vs. a placebo group) were found for adults hospitalized with COVID-19 and treated with remdesivir; additionally, those patients showed fewer infections of the respiratory tract (Beigel et al., 2020). In March 2020, the WHO launched a “solidarity clinical trials” of repurposed drugs and experimental candidates, wherein testing of the three aforementioned drugs was supplemented with testing of the antimalarial compounds chloroquine and

hydroxychloroquine<sup>5</sup>. In July 2020, WHO decided to discontinue the hydroxychloroquine and lopinavir/ritonavir trials, as these compounds yielded little to no reduction in the mortality of hospitalized COVID-19 patients when compared to standard of care<sup>5</sup>. Currently, dexamethasone—an anti-inflammatory drug approved to treat COVID-19 patients in the UK and Japan—also has been reported, in an unpublished study, to reduce mortality among COVID-19 patients hospitalized with severe infection (Horby et al., 2020).

The earliest-discovered CoVs do not correspond to those strains that are the causative infectious agents in recent outbreaks, including COVID-19 (Khedkar and Patzak, 2020). The first “coronavirus” (to be termed as such) was isolated from chicken in 1937; human CoVs were identified years later, in the mid-1960s<sup>6</sup>. These viruses belong to the taxonomic family *Coronaviridae*, which are single-stranded, positive-sense RNA viruses of ~29.9 Kb genomic length (Khedkar and Patzak, 2020). The CoVs encode more than a dozen proteins, some of which have been identified as critical for viral entry and replication (Muramatsu et al., 2016). Among the structural proteins encoded by the CoV genome, four proteins are of special interest from the perspective of therapeutics and drug design—namely, the *spike* (S), *envelope* (E), *membrane* (M), and *nucleocapsid* (N) proteins. The S, E, and M proteins are housed in the membranes of these enveloped virions. The M and E proteins are actively involved in viral coat assembly, while the N protein is involved in compacting the RNA genome. The most-studied proteins thus far have been a papain-like protease (PLpro), a 3C-like protease (3CL<sup>pro</sup>), an endosomal protease, and the spike protein (Yang and Wang, 2020).

At the molecular level, CoVs are known to gain cellular entry *via* the S protein (Anand et al., 2003). Viral entry depends on the binding of the surface unit S1 of the S protein to a surface-exposed cellular receptor in the host, thereby supporting the process of viral attachment to target cell surfaces (Muramatsu et al., 2016). The 3CL protease (3CL<sup>pro</sup>), also known as M<sup>pro</sup>, is the main protease produced by the CoV; it plays a key role in viral replication (Wu et al., 2020). Most of the functional proteins of CoVs are encoded by specific genes, which are first translated into polypeptides that are then cleaved by the viral 3CL<sup>pro</sup> or by PLpro. This stage of the viral replication cycle yields the RNA-dependent RNA polymerase (RdRp), along with multiple other proteins that play roles in virus replication, transcription, and translation. Inhibiting the activity of the main CoV protease would presumably block viral replication (Yang and Wang, 2020). Thus, 3CL<sup>pro</sup> is considered a potential drug target for COVID-19. In addition, targeting the main protease for inhibition is an appealing strategy because it may well be thought that this would inactivate the virus in different cell types and

**Abbreviations:** SARS-CoV-2, severe acute respiratory syndrome coronavirus 2; WHO, World Health Organization; COVID-19, coronavirus disease 2019; FDA, Food & Drug Administration; EMA, European Medicines Agency; TCM, Traditional Chinese Medicine; MD, molecular dynamics; VS, virtual screening; 2D, two-dimensional; 3D, three-dimensional; LBVS, ligand-based virtual screening; SBVS, structure-based virtual screening; ADMET, absorption, distribution, metabolism, elimination, and toxicity; LD, lethal dose; RMSD, root-mean-square deviation; CYP, cytochrome P450; ACE, angiotensin-converting-enzyme.

<sup>1</sup>[https://www.who.int/health-topics/coronavirus#tab=tab\\_1](https://www.who.int/health-topics/coronavirus#tab=tab_1) (accessed July 28, 2020).

<sup>2</sup>WHO Coronavirus Disease (COVID-19) Dashboard | WHO Coronavirus Disease (COVID-19) Dashboard. Available online at: [https://covid19.who.int/?gclid=CjwKCAjwsO\\_4BRBBEiwAyagRTezBE06lkOleffHglbg0emog5Zo38YH6hOcVz3YyPcI5LkKi0DCIwxwhoCkGgQAvD\\_BwE](https://covid19.who.int/?gclid=CjwKCAjwsO_4BRBBEiwAyagRTezBE06lkOleffHglbg0emog5Zo38YH6hOcVz3YyPcI5LkKi0DCIwxwhoCkGgQAvD_BwE) (accessed July 28, 2020).

<sup>3</sup>DrugBank. Available online at: <https://www.drugbank.ca/covid-19> (accessed July 28, 2020).

<sup>4</sup>Draft landscape of COVID-19 candidate vaccines. Available online at: <https://www.who.int/publications/m/item/draft-landscape-of-covid-19-candidate-vaccines> (accessed July 28, 2020).

<sup>5</sup>Solidarity clinical trial for COVID-19 treatments. Available online at: <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/global-research-on-novel-coronavirus-2019-ncov/solidarity-clinical-trial-for-covid-19-treatments> (accessed July 28, 2020).

<sup>6</sup>Coronavirus Disease 2019 (COVID-19) | CDC. Available online at: <https://www.cdc.gov/coronavirus/2019-ncov/index.html> (accessed July 30, 2020).

in different organs—regardless of the various matches between receptors/host proteases (on the cell membrane) that underlie viral entry in a cell- or tissue-specific manner (Zhang L. et al., 2020).

Because of its mechanistic significance, 3CL<sup>Pro</sup> is now a central target for the development of effective inhibitors (antiviral drugs) against both SARS-CoV-2 and other known CoVs (Anand et al., 2003). The X-ray crystal structure of 3CL<sup>Pro</sup>(M<sup>Pro</sup>) from SARS-CoV-2 (PDB code: 6LU7) reveals a protein comprised of three primary domains (Jin et al., 2020). The first domain (Domain I) consists of amino acid residues 8–101; the second domain (Domain II) maps to residues 102–184; and the third domain (Domain III) mainly consists of residues 201–306, largely as a cluster of  $\alpha$ -helical conformations (Jin et al., 2020). The substrate-binding region of 3CL<sup>Pro</sup>, located between Domains I and II, includes residues His41 and Cys145. Visual inspection of the binding site structure confirms that peptide-type inhibitors attach to the active site cysteine: a previously identified peptidomimetic inhibitor, “N3,” interacts irreversibly with this site and engages in supporting interactions with subsites (S1, S2, and S4) (Jin et al., 2020). The S1 subsite contains residues His163, Glu166, Cys145, Gly143, His172, and Phe140, while the S2 subsite consists mainly of Cys145, His41, and Thr25; these amino acid types are compatible with favorable non-bonded contacts such as electrostatic interactions and van der Waals (apolar/dispersive) forces. There are two additional subsites (S3–S5), consisting of Thr190, Gln192, Glu166, Met49, Leu167, Gln189, and Met165 (Anand et al., 2003). A ligand interaction diagram drawn from the 3D crystal structure (**Figure 1**) illustrates that this particular N3 inhibitor engages in multidentate hydrogen bonds with Glu166 (as both donor and acceptor). In addition, there are close—and presumably energetically favorable—contacts between moieties of N3 and Gly143 and the catalytic Cys145 (both of the S1 subsite).

The traditional drug discovery and development pipeline is generally a quite time-consuming endeavor, taking upward of  $\approx 10$ –15 years (Turanli et al., 2019; Kupferschmidt and Cohen, 2020). Computational drug “repurposing” is an effective approach to accelerate this timescale by finding new uses for existing (and already approved) drugs (McNamee et al., 2017). Computational approaches to drug discovery, particularly as part of a repurposing strategy/framework, can fasten the drug development process and alleviate the burdens of traditional approaches—features that are especially critical in the context of a pandemic. Such computational approaches have been used to identify candidate drugs for several infectious diseases, including Ebola, Zika, dengue, and influenza (Cha et al., 2018). Many computational methods are available to examine the key interrelationships between chemical structure, biological/physiological systems, interactions between chemicals and (bio) molecular targets, and, finally, the ultimate therapeutic endpoints and diseases (Metushi et al., 2015).

For COVID-19, several recent studies have reported the computational screening of inhibitors for specific single targets, such as either the main protease (the aforementioned 3CL<sup>Pro</sup>/M<sup>Pro</sup>) or the spike protein (Shiryaev et al., 2017; Botta et al., 2018; Pizzorno et al., 2019; Ton et al., 2020; Wang, 2020;

Zhang D.-h. et al., 2020). To the best of our knowledge, the work reported here is the first study to report screening results including databases like Super Natural II<sup>7</sup>, as well as the TCM (Chen, 2011)<sup>8</sup>, and repurposing-focused databases, such as SuperDRUG2<sup>9</sup> and WITHDRAWN database<sup>10</sup>. Additionally, our final selection of lead compounds is based on visual inspection and analysis of ligand interactions with crucial residues of the target (3CL<sup>Pro</sup>), thereby implicitly incorporating human expertise and clinical insights into our workflow (once the number of candidates becomes manageable for manual analysis). The final 12 best candidates were further evaluated using molecular dynamics (MD) simulation studies, and clinical feasibility for repurposed drugs was further investigated.

The general approach used in this study is based on an integrated pipeline: we include a virtual screening of cheminformatics-driven databases, and we employ molecular similarity and molecular docking to identify promising drug target pairs. Additionally, MD simulation studies of selected compounds were performed to select the best candidates for main protease inhibitors and evaluate their stability and strength of interactions. Our screening is aimed at dual objectives: (i) first, to find potential new candidates in the Super Natural II<sup>7</sup> and TCM (Chen, 2011)<sup>8</sup>, and (ii) second, to identify promising repurposing candidates in the SuperDRUG2<sup>8</sup> and WITHDRAWN databases<sup>10</sup>.

## MATERIALS AND METHODS

To facilitate the rapid discovery of target inhibitors with real clinical potential, we have employed a prediction strategy that is based on the fundamental principle of “neighborhood behavior” (Patterson et al., 1996) implemented as a computational pipeline that utilizes both *ligand-based* [two-dimensional (2D) chemical space of small molecules] and *structure-based* [protein three-dimensional (3D) spaces and features]. Our computational predictions use the first resolved crystal structure of SARS-CoV-2 main protease (at a resolution 2.16 Å). Currently, several crystal structures of the SARS-CoV-2 main protease have been experimentally determined (approximately 175 structures)<sup>11</sup>. The methodological details of our approach and computational methods are described below.

### Databases

In order to broadly screen, we utilized four different database resources: (i) SuperNatural II, a database of natural products<sup>7</sup>; (ii) The Traditional Chinese Medicine (SuperTCM) database, an in-house (unpublished) database that is manually created

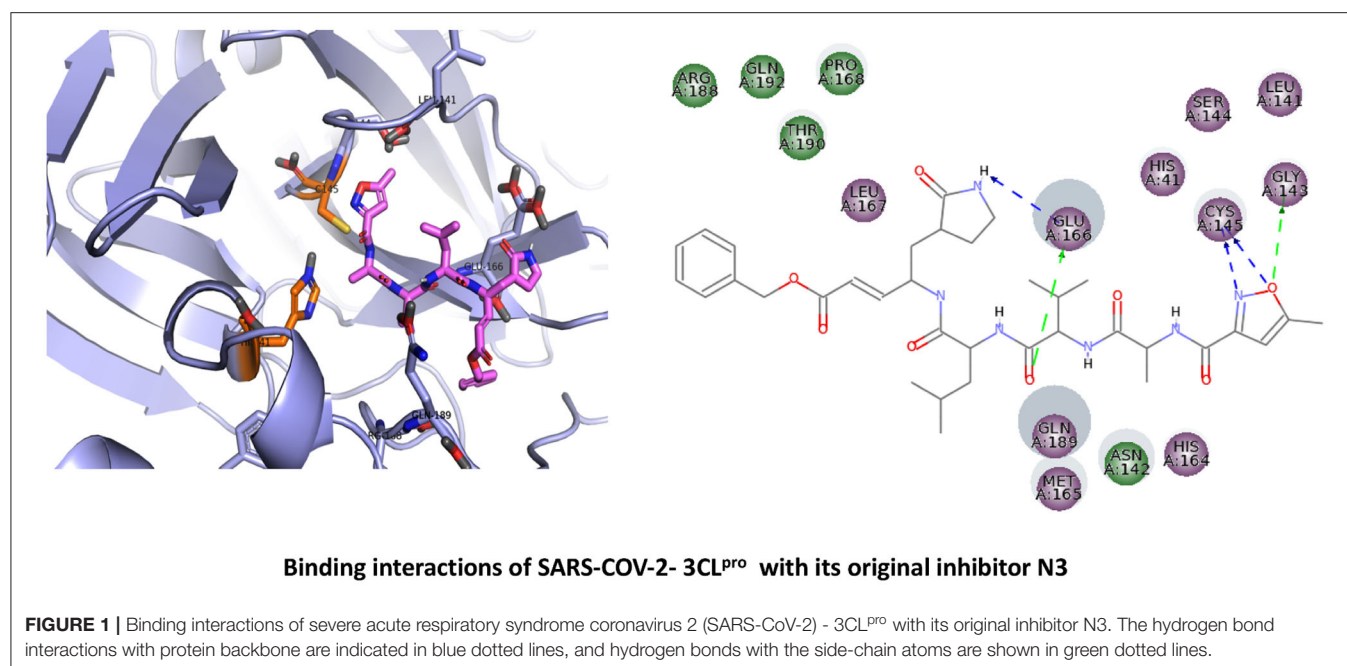
<sup>7</sup>Super Natural II—a database of natural products - PubMed. Available at: <https://pubmed.ncbi.nlm.nih.gov/25300487/> (accessed July 28, 2020).

<sup>8</sup>ETCM: an encyclopaedia of traditional Chinese medicine. Available online at: <https://pubmed.ncbi.nlm.nih.gov/30365030/> (accessed July 28, 2020).

<sup>9</sup>SuperDRUG2: a one stop resource for approved/marketed drugs. Available online at: <https://pubmed.ncbi.nlm.nih.gov/29140469/> (accessed July 28, 2020).

<sup>10</sup>WITHDRAWN—a resource for withdrawn and discontinued drugs - PubMed. Available online at: <https://pubmed.ncbi.nlm.nih.gov/26553801/> (accessed July 28, 2020).

<sup>11</sup>RCSB PDB: Homepage. Available online at: <https://www.rcsb.org/> (accessed July 31, 2020).



from other databases and Chinese literature (Chen, 2011)<sup>8</sup>; this comprehensive database built at the Charité–University of Medicine<sup>12,13</sup> covers all aspects of traditional Chinese medicine mainly derived from medicinal plants, and it encompasses pharmaceutical recipes up to molecular ingredients. The database was manually curated by domain experts, and Chinese plant-based drugs were mapped to their plant of origin, common non-scientific names and scientific names, targets, Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways<sup>14</sup>, as well as the traditional Chinese traditional recipes. (iii) We also used SuperDRUG2, a one-stop resource for approved/marketed drugs<sup>8</sup>; (iv) And, finally, we also used WITHDRAWN, a resource for withdrawn and discontinued drugs<sup>10</sup>. Overall, we utilized more than 360,000 compounds for virtual screening and initial filtering purposes (Table 1).

The potential inhibitors are screened mainly from phytochemical databases because the literature suggests that seven out of 10 synthetic agents approved by the U.S. FDA are modeled on a natural product parent (Newman and Cragg, 2016). There is an urgent need to identify novel active chemotypes as leads for effective antiviral therapy for COVID-19 infections. Similarly, several thousands of plant extracts have been shown to possess *in vitro* antiviral activity with little overlap in species between studies (Chen, 2011). Promising docking outcomes have been executed in this study, which evidenced the

**TABLE 1 |** Databases and total number of compounds used in this study.

Database	Compounds	Total number of compounds used in this study
SuperDRUG2	Approved and marketed drugs	3,992
WITHDRAWN	Withdrawn or discontinued drugs	626
TCM	In-house database of compounds related to Traditional Chinese Medicine	28,974
Super Natural II	Natural compounds	325,508

worth of these selected chemical compounds from Super Natural II and TCM databases for future drug development to combat CoV diseases.

Additionally, drug databases from both approved and withdrawn chemical spaces were screened to support reuse of already available drugs for new indications such as COVID-19 therapy when they have been originally developed for specific diseases.

## Ligand-Based Screening

We screened the molecular libraries from these databases based on ligand similarity, which rests upon the assumption that “structurally similar compounds might have similar biological properties” (Stumpfe and Bajorath, 2011). As many chemical-based fingerprint methods are in widespread use, based on the performances of various structural similarity measures, a 2D similarity screening protocol was designed, and three different types of molecular fingerprints were initially chosen for comparison and performance evaluation: MACCS (Durant et al.,

<sup>12</sup>Structural Bioinformatics Group. Available online at: <http://bioinf-apache.charite.de/main/index.php> (accessed July 31, 2020).

<sup>13</sup>Traditional Chinese Medicine: Institut für Theorie, Geschichte, Ethik Chinesischer Lebenswissenschaften - Charité – Universitätsmedizin Berlin Available online at: [https://icl.charite.de/forschung/buch\\_details\\_2011\\_2020/traditional\\_chinese\\_medicine](https://icl.charite.de/forschung/buch_details_2011_2020/traditional_chinese_medicine) (accessed July 31, 2020).

<sup>14</sup>KEGG PATHWAY Database. Available online at: <https://www.genome.jp/kegg/pathway.html> (accessed July 31, 2020).

2002), Extended Connectivity Fingerprints (ECFP-4) (Rogers and Hahn, 2010), and E-state (Hall and Kier, 1995). We found that the performance of the MACCS fingerprint surpasses that of ECFP-4 and E-state in an approved drug dataset, with a higher Tanimoto score threshold (from 0.73 to 0.83) than ECFP-4 (**Supplementary Table 1**). MACCS ranked the similarity scores higher for co-crystallized ligands and drugs like lopinavir, angiotensin II, and indinavir (compounds that were previously reported as potential inhibitors of the main protease 3CL<sup>Pro</sup>) (Contini, 2020; Nutho et al., 2020). Using this small set of active compounds, we tried to find the optimal similarity cutoff as well as optimal chemical fingerprint to yield the best balance of precision vs. recall.

Chemical-based fingerprints were calculated using RDKit<sup>15</sup> nodes in KNIME (Berthold et al., 2008), and pairwise similarities were calculated for all the datasets.

The lead compounds used in the similarity search for targeting the 3CL<sup>Pro</sup> of SARS-CoV-2 were obtained from the Protein Data Bank (PDB)<sup>5</sup>. The original ligands from PDB structures 6LU7 and 6Y2F were considered as lead compounds. The first ligand (“N3”) is a peptidomimetic irreversible inhibitor and was found covalently bonded to Cys145. This interaction is reported to be essential for preserving the protease’s S1 pocket in the right shape and also for the active conformation of the enzyme (Zhang L. et al., 2020). The N3 ligand interacts with the catalytic center of the target proteases through two hydrogen bond interactions. It was observed that the pyridine ring might have some steric clash with the side chain of Gln 189 (Zhang L. et al., 2020). The reported  $\alpha$ -ketoamide ligand-bound X-ray crystal structure of SARS-CoV-2 Mpro (PDB ID: 6Y2F) forms hydrogen bonds with the Ser of chain B and Glu166 of chain A (Zhang L. et al., 2020). This interaction is reported to be essential for keeping the S1 pocket in the right shape for ligand–receptor interactions and also for the active conformation of the enzyme (Zhang L. et al., 2020). It interacts with the catalytic center of the target proteases through two hydrogen bond interactions. It was observed that the pyridine ring might have some steric clash with the side chain of Gln189 (Zhang L. et al., 2020). The compounds that showed the highest similarity considering structural properties were finally chosen for molecular docking studies. For each of the dataset and each of the lead compounds, the 10 best hits were chosen (**Supplementary Tables 7, 8**). The Tanimoto score values for selected compounds range between 0.63 and 0.83. It is also possible that a “false similar” or “false active” pair of molecules could occur, featuring structural similarity but dissimilarity in terms of their biological activities; to assess this possibility regarding activity profiles, further molecular docking and MD simulation studies were conducted.

## Structure-Based Screening

To further refine the list of candidate compounds and select the top hits, molecular docking calculations were carried out using the GOLD software (version 5.7.2) (Jones et al., 1997). This code uses a genetic algorithm to sample the ligand’s conformational space, making it particularly suitable for docking flexible ligands

with numerous rotational degrees of freedom (Jones et al., 1997). In addition, the GOLD scoring function was used to rank the compounds, with the number of docked poses to be generated set to 10. In the present study, the co-crystal structures of 3CL<sup>Pro</sup> of SARS-CoV-2 (PDB 6LU7) was selected as the starting structure for docking calculations. Residues that are proximal (within a 10 Å radius) to the original ligand co-crystallized in 6LU7, along with binding site residues (as defined in the literature) were taken to be the active site for docking calculations. Thus, the final docking protocol incorporates information from the successful re-docking of the original ligand to the target.

A total of 80 compounds (top 10 screened compounds for each lead compound, selected using ligand-based screening from four different databases) were docked into the main protease protein 3CL<sup>Pro</sup>. Based on interactions with key residues (His41 or Cys145), present in the binding cavity of the 6LU7 crystal structure (and based on visual inspection of the ligand–receptor binding interactions), the top 3 best candidates were selected per database. Therefore, our final list includes the 12 best candidates based on this computational screening protocol. Two-dimensional ligand–receptor binding interaction maps were computed using Accelrys Discovery Studio (version 4.5)<sup>16</sup>. The 3D interactions and structural illustrations were created using PyMOL (version 2.3.5)<sup>17</sup>. After the molecular docking analyses, MD simulations were performed for the best 12 screened candidates, enabling us to evaluate the dynamical stability of the bound/docked complexes, at least on the timescale of the MD trajectories.

## Molecular Dynamics Simulation

Analyzing the dynamic evolution of the ligand-bound system helps us predict the stability of those interactions that we first detected via ligand-based virtual screening (LBVS) and structure-based virtual screening (SBVS); ideally, the MD faithfully recapitulates the real (physiological) environment of such interactions. Ligand–protein interactions at the binding site can be monitored for a period of time, so that ligands with more temporally stable poses can be detected and proposed as better candidates for 3CL<sup>Pro</sup> inhibition. Thus, after our database screening stages, MD simulations of the different complexes were computed using the GPU version of Desmond included with Maestro suite 2019.4 (Schrödinger LLC)<sup>18</sup> on a workstation with a NVIDIA QUADRO 5000. The various drug/receptor complexes were solvated in an aqueous environment in a cubic box with a minimal distance of 10 Å between the biomolecule and the box boundary (for periodic boundary conditions). Next, systems were neutralized and maintained in 0.15 M NaCl. The OPLS3 force field and the TIP3P-TIP4P water model were employed (Mark and Nilsson, 2001). Initially, the systems were simply

<sup>16</sup>Free Download: BIOVIA Discovery Studio Visualizer - Dassault Systèmes. Available online at: <https://discover.3ds.com/discovery-studio-visualizer-download> (accessed July 31, 2020).

<sup>17</sup>PyMOL | pymol.org. Available online at: <https://pymol.org/2/> (accessed July 31, 2020).

<sup>18</sup>Maestro | Schrödinger. Available online at: <https://www.schrodinger.com/maestro> (accessed July 31, 2020).

<sup>15</sup>RDKit. Available online at: <https://www.rdkit.org/> (accessed July 31, 2020).

energy-minimized for 2,000 time steps. Next, systems were allowed to execute free dynamics in the NPT ensemble; pressure was controlled using the Martyna–Tobias–Klein methodology, and the Nose–Hoover thermostat was employed to maintain the system near 310 K. Production-grade MD trajectories were extended to a total duration of 100 ns per system. MD trajectories were characterized in terms of the root-mean-square deviation (RMSD) of fluctuations of ligand and enzyme, particularly in terms of the main interactions with the top interacting residues. The trajectories were also used to assess the stabilities of the protein secondary structures (in complex with potential inhibitor) by plotting RMSDs. Additionally, to estimate the relative binding free energies of the 12 final compounds and also N3 ligand to the macromolecule, molecular mechanics–generalized Born surface area (MM-GBSA) method was applied. The MM-GBSA method is based on the difference between the free energies of the protein, ligand, and the complex in solution. The free energy for each species involved in the reaction (ligand, protein, and ligand–protein complex) is described as a sum of a gas-phase energy, polar and non-polar solvation terms, and an entropy term. In our computational protocol, the MM-GBSA method is used to calculate the free energy (dG) related to all poses obtained in the MD simulation by using the OPLS3 as implemented in the Small-Drug Design Suite of Schrodinger (Kollman et al., 2000; Greenidge et al., 2013).

## Absorption, Distribution, Metabolism, Elimination, and Toxicity Properties

The docked compounds were further filtered using the standard ADMET (Absorption, Distribution, Metabolism, Elimination, and Toxicity) pharmacokinetic properties. Computational toxicity analysis was performed using the ProTox-II methods<sup>19</sup>. The ProTox-II web server currently holds 40 different predictive models, incorporating chemical similarity, fragment propensities, most frequent features, pharmacophores, and machine learning for toxicity prediction. The acute toxicity value of the ProTox-II method is divided into six classes based on a globally harmonized system of classification of labeling of chemicals (GHS). The classes are described as: Class 1: fatal if swallowed ( $LD_{50} \leq 5$ ); Class 2: fatal if swallowed ( $5 < LD_{50} \leq 50$ ); Class 3: toxic if swallowed ( $50 < LD_{50} \leq 300$ ); Class 4: harmful if swallowed ( $300 < LD_{50} \leq 2,000$ ); Class 5: may be harmful if swallowed ( $2,000 < LD_{50} \leq 5,000$ ); Class 6: non-toxic ( $LD_{50} > 5,000$ )<sup>38</sup>. Additionally, cytochrome (CYP) inhibition profiles of each compound were computed using the SuperCYPsPred web server<sup>20</sup>. Currently, the SuperCYPsPred web server includes 10 models for five major CYPs isoforms (including 3A4, 2C9, 2C19, 2D6, and 1A2). These cytochrome predictive models are based on machine learning methods (see **Tables 2–4** in the Results section).

<sup>19</sup>ProTox-II: a webserver for the prediction of toxicity of chemicals. Available online at: <https://pubmed.ncbi.nlm.nih.gov/29718510/> (accessed July 28, 2020).

<sup>20</sup>SuperCYPsPred—a web server for the prediction of cytochrome activity. Available online at: <https://academic.oup.com/nar/article/48/W1/W580/5809167> (accessed July 28, 2020).

## RESULTS

### Overview

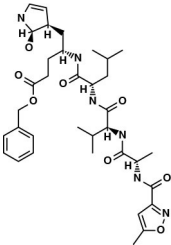
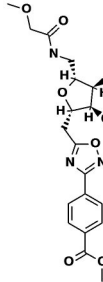
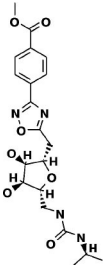
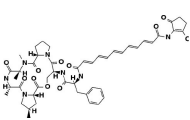
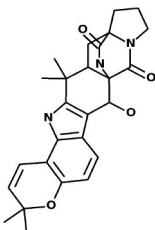
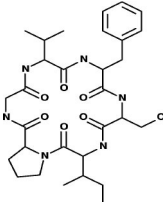
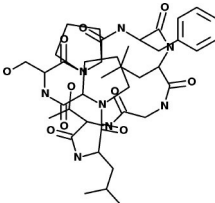
A similarity-based approach, using structural chemical fingerprints, was used for initial screening of compounds from different databases. The crystal structure of the main protease was used as the molecular target for computational docking and protein–ligand interaction analyses; in total, visual inspections were performed for 80 compounds. The top 40 compounds for the N3 inhibitor-based docking are shown in **Supplementary Table 7**, and the top 40 compounds for the 6OK inhibitor-based efforts are given in **Supplementary Table 8**. Using our integrated approach—i.e., chemical similarity with molecular docking—a list of 12 top lead compounds were identified from four different databases. Visual inspection and selection of best candidates were based on analyses of the interaction with at least one of the two catalytic residues Cys154 or His41. The results for these compounds are shown in **Tables 2, 3**. Besides the selected compounds, the tables also include information on therapeutic endpoints, toxicity endpoints, and cytochrome activity profiles of the compounds; interactive residues in the protease receptor are also given. For compounds from the Traditional Chinese Medicine database, information on their respective plants (as curated from our database) is also provided, with additional information on their indications and effects (**Supplementary Table 6**). Furthermore, to elucidate potential molecular mechanisms in terms of ligand interactions, the top 3 compounds from each database were selected and examined via all-atom MD simulations. The 12 best candidates are described below, including the RMSD plots obtained from the MD trajectories. For the validation of the MD protocol, simulation of N3 ligand was also performed. Results are presented in **Supplementary Figures 5, 6**. Additional information on binding interactions from molecular docking studies (for a selected subset of 56 compounds) is provided in **Supplementary Table 9**, and docking scores for the N3 and O6K ligands are provided in **Supplementary Table 10**.

### Super Natural II Database

The top 3 compounds we identified in the Super Natural II database are SN00017653, SN00019468, and SN00303378. The compound SN00017653 interacts *via* hydrogen bonds with the side-chain atoms of Glu166, Ser144, Leu141, Gly143, and Cys145 (**Figure 2A**). The second compound, SN00019468, interacts with Cys145, Gly143, and Ser144 and hydrogen-bonds with the backbone of His41 (**Figure 2B**). The third compound, SN00303378, interacts with His41, Asn142, Thr190, Gln192, and Ser144 (**Figure 2C**). In addition, the top 10 compounds from the Super Natural database are reported in **Supplementary Table 2**.

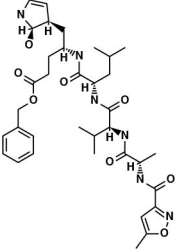
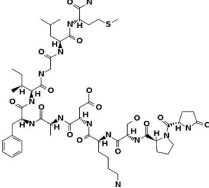
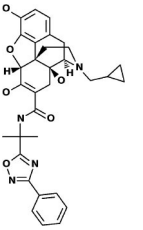
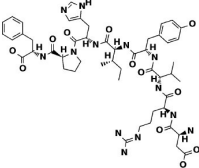
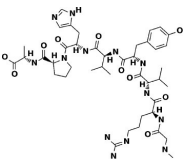
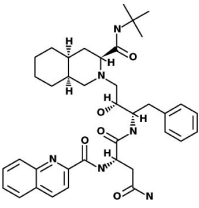
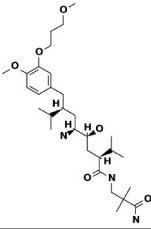
MD simulations reveal that SN00017653 exhibits a stable pose when interacting with 3CL<sup>Pro</sup>, at least in terms of RMSD values of  $\approx 2.0$  and  $2.8$  Å for ligand and protein atoms, respectively. This compound exhibits quite stable contacts with the S2 subsite, where it interacts with Thr26 and His41. The time evolution of protein–ligand contacts shows also a weak interaction with Cys145, but this contact is highly transient. Furthermore, SN00017653 interacts with the S3 and S5 sub-pockets by binding

**TABLE 2 |** Potential inhibitors for the main protease of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) from Super Natural II and SuperTCM databases.

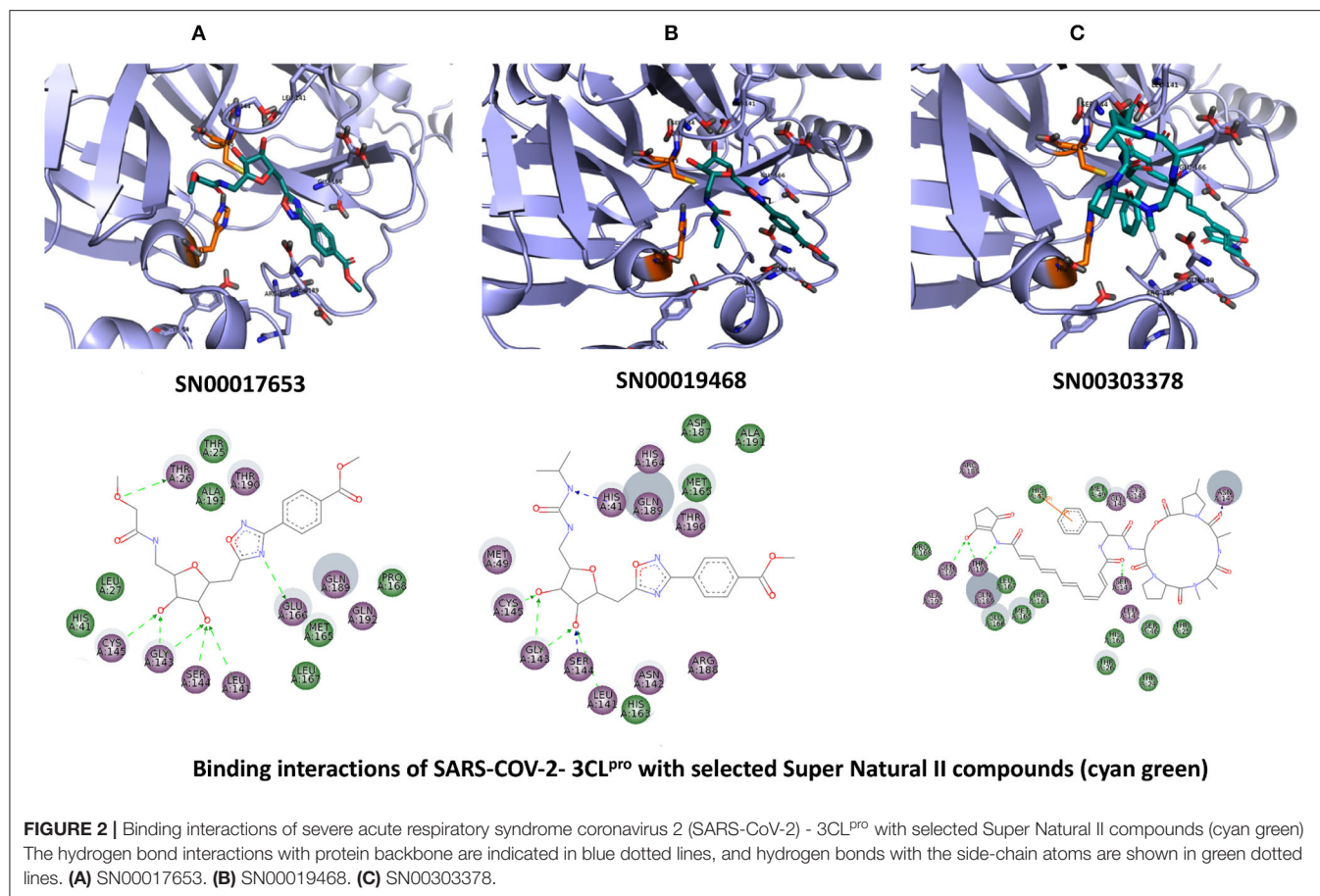
Compound	Interacting residues (3CL)	Acute toxicity	Toxicity endpoints	CYP activity	Tanimoto score	Structure*
N3-inhibitor	Glu166, Cys145, Gly143	Class 5	NA	CYP3A4	1	
<b>Super Natural II database</b>						
SN00017653	Cys145, Gly143, Glu166, Ser144, Leu141, Thr26	Class 4	NA	NA	0.75	
SN00019468	His41, Cys145, Gly143, Ser144, Leu141	Class 4	NA	NA	0.74	
SN00303378	His41, Thr190, Gln192, Ser144	Class 3	Immunotoxic	CYP3A4	0.73	
<b>TCM database</b>						
Notoamide R	Cys145, His41, His164, Gln189	Class 4	Immunotoxic	3A4	0.75	
Dianthin E	Cys145, Gln189, Glu166, Ser144, Gly143	Class 4	None	None	0.74	
Pseudostellarin C	Cys145, His164, His41, Glu166	Class 4	Immunotoxic	None	0.77	

\*Structures of compounds generated with PubChem Sketcher V2.4 (Ihlenfeldt et al., 2009).

**TABLE 3 |** Potential inhibitors for the main protease of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) from SuperDRUG2 and WITHDRAWN drug databases.

Compound	Interacting residues (3CL)	Therapeutic endpoints	Acute toxicity	Toxicity endpoints	CYP activity profile	Tanimoto score	Structure*
N3-inhibitor	Glu166, Cys145, Gly143		Class 5	NA	CYP3A4	1	
<b>Approved drugs database</b>							
Eledoisin	Ser144, Arg188, Asn142, Cys145, His41	Vasodilator	Class 5	None	None	0.72	
Naldemedine	His41, Cys145, Gln192, His164	Alimentary tract and metabolism	Class 4	Immunotoxic	3A4	0.72	
Angiotensin II	Cys145, Gln189, Asn142	Cardiac therapy	Class 5	None	None	0.73	
<b>Withdrawn drugs database</b>							
Saralasin	Cys145, Met165, Gln189, Arg188	Cardiac therapy	Class 5	None	None	0.71	
Saquinavir	Cys145, Gly143, His41, Glu166	Antiviral	Class 4	None	2C8, 2C9, 2C19, 2D6, 3A4, 3A5	0.67	
Aliskiren	Tyr54, Cys145, Ser144	Cardiac therapy	Class 5	Immunotoxic	3A4	0.63	

\*Structures of compounds generated with PubChem Sketcher V2.4 (Ihlenfeldt et al., 2009).



Glu166, Gln189, and Thr190. Interactions with Glu166 persist after 100 ns, and it shows a high interaction score in comparison with the other key residue, so it is considered the most important interaction (**Figure 6**, **Supplementary Figure 1**).

Compound SN00019468 shows high ligand RMSD fluctuations, especially during 50–100-ns trajectory. Examination of protein-ligand interaction contacts reveals that SN00019468 samples two distinct positions. During the first 50 ns, this ligand interacts with Thr26, His164, and Gln189, being only contact important for 3CL<sup>pro</sup> inhibition, so it is part of the S3–S5 subsite. During the last 50 ns, these contacts evolve to an interaction with the S2 subsite by means of His41, but this interaction is not persistent, maintained during 50–90 ns, but lost after the 100-ns simulation. The poor stability of SN0019468 in the binding site suggests it is less likely to be an effective 3CL<sup>pro</sup> inhibitor (**Figure 6**).

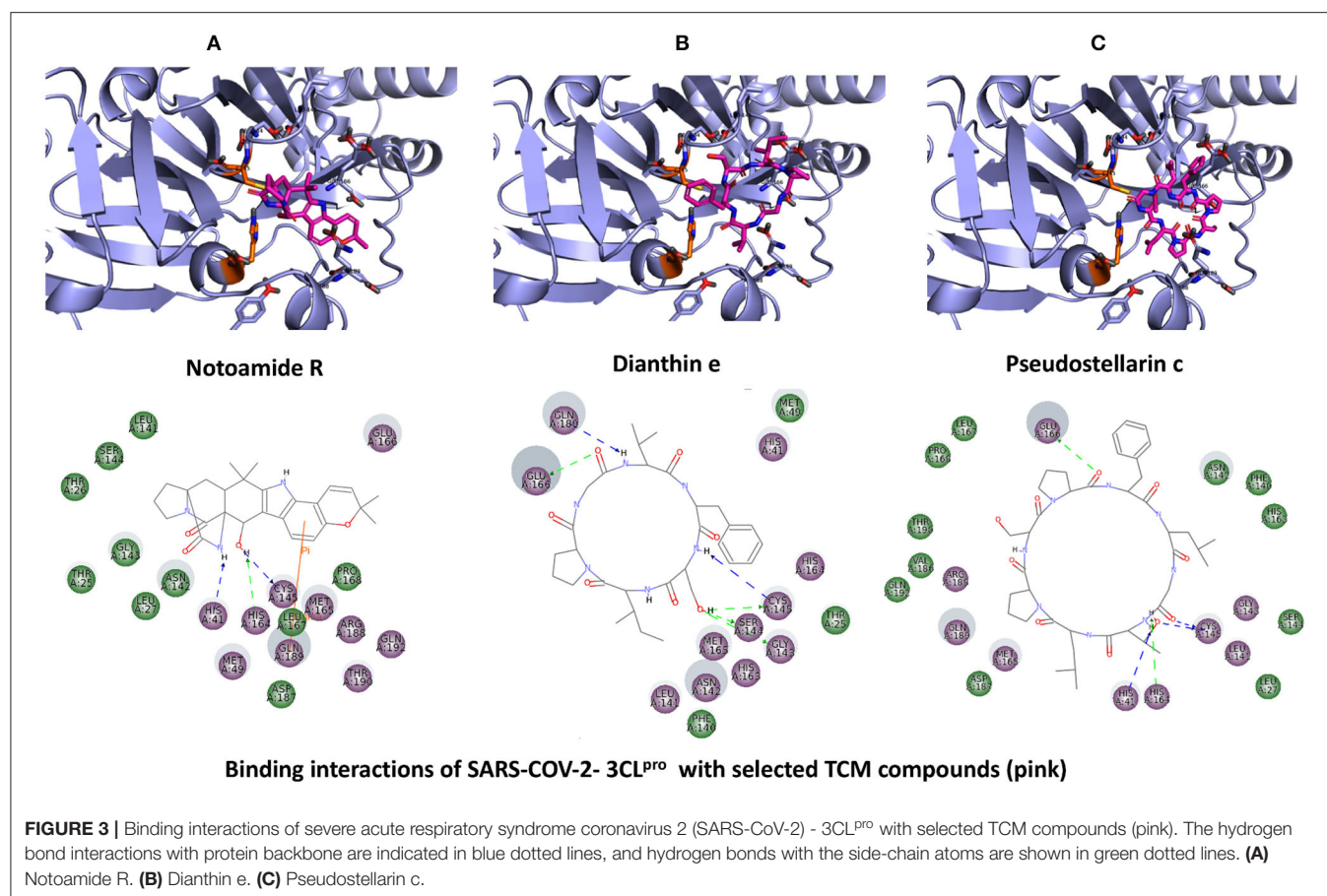
The ligand and the protein RMSDs are stable for compound SN00303378, indicating a stable position of this compound when bound to the protein. The pattern of protein contacts (**Figure 6**) shows an interaction with His41, but it fluctuates along the 100-ns simulation (being more important in the  $\approx$ 50–70-ns range). Also, the RMSD is stabilized during this period, with closely matched values for protein and ligand atoms. Glu166 engages in the clearest interactions, persisting along the full trajectory. Nevertheless, SN00303378 is not considered a strong candidate

to inhibit 3CL<sup>pro</sup> because of its paucity of contacts with other key residues of the main protease (**Figure 6**).

## Traditional Chinese Medicine Database (SuperTCM)

The top 3 compounds from the SuperTCM database are notoamide R, dianthin E, and pseudostellarin C. Notoamide R interacts with Cys145, His41, His164, and Gln189 (**Figure 3A**). Dianthin E interacts with the backbone atoms of Cys145 and Gln189, and the side-chain atoms of Glu166, Ser144, and Gly143 chiefly *via* hydrogen bonds (**Figure 3B**). Pseudostellarin C interacts with Cys145, His164, His41, and Glu166 (**Figure 3C**). Further information on the top 10 compounds from the SuperTCM database is reported in **Supplementary Table 3**.

MD simulations do not suggest that Notoamide R engages in notable interactions with the binding site and, in fact, reveals an unstable pattern of protein–ligand contacts. Regarding the time evolution of the RMSDs and protein–ligand contacts, this ligand appears to adopt rather different conformations along the trajectory. During the first 30 ns, it occupies a stable conformation and nominally interacts with His41 and, to a lesser extent, with Cys145. A conformational change then occurs, and the positions adopted by the ligand from 30 to 100 ns preclude contacts with binding-site residues. Thus, this ligand



is not considered as a suitable candidate for 3CL<sup>Pro</sup> inhibition (Figure 7).

The RMSD graphic of Dianthin E shows a highly stable pose along the first  $\approx 40$  ns. After that, a peak in the plot indicates a shift in the pose; nevertheless, the simulation concludes with stable RMSD values for protein and ligand (granted, these values are higher than for the other ligands simulated here). The patterns of protein–ligand contacts are in concordance with the RMSD fluctuations. The first  $\approx 40$  ns saw some interaction with Cys145 and, in a clearer way, with Glu166 and Gln189. The ligand reoriented in the next  $\approx 60$  ns, corresponding to a weak bond with His41 and to a stabilization of the Gln189 bond (Figure 7).

The RMSD trace for Pseudostellarin C shows a high stable pose from the first 40 ns. After that, there is a peak that indicates a pose change, but the simulation ends with stable RMSD values for protein and ligand. The protein–ligand contacts are aligned with the RMSD fluctuation. During the first 40 ns, there was some interaction with Cys145 and with higher interaction score values with Glu166 and Gln189 (Supplementary Figure 2). The ligand reorientation after the following 60 ns was translated to an intermittent bond with His41 and to a persistent bond with Gln189 (Figure 7).

## SuperDrug2 Drug Database

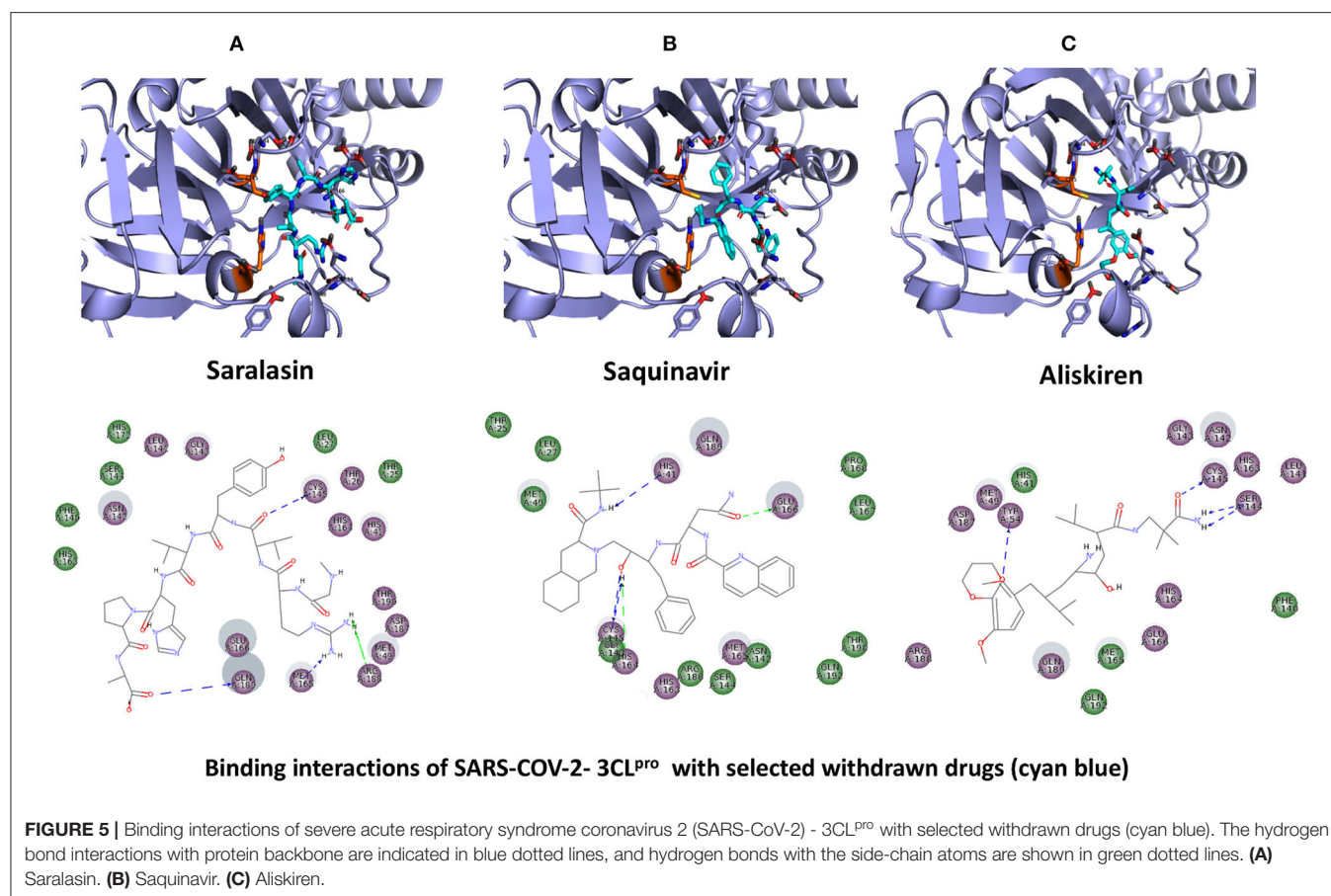
The top three selected compounds from the approved drug database include naldemedine, eldeoisin, and angiotensin II.

Naldemedine interacts with Ser144, Arg188, and Asn142 (Figure 5A). It also interacts with the catalytic residues Cys145 and His41. Eldeoisin also interacts with the catalytic residues His41 and Cys145 and with two other residues, Gln192 and His164 (Figure 5B). Angiotensin II interacts with Cys145, Gln189, and Asn142 (Figure 5C). More information on the 10 best drug candidates from the SuperDRUG2 database is reported in Supplementary Table 4.

Naldemedine shows close values for protein and ligand RMSDs, and the ligand pose is considered stable. Regarding protein–ligand contacts, the residues Glu166 and Gln189 were stable in contact throughout the simulation. Initially, also Cys145 showed a continuous interaction, so the S2 sub-pocket would be the main location responsible for 3CL<sup>Pro</sup> inhibition. After the first 60 ns, this interaction becomes less stable and the contacts with Glu166 and Gln189 become more prominent, being the S3–S5 subunits responsible for the inhibition. Hence, an inhibition mode that is initially dominated by interactions at the S2 sub-pocket (and a relatively minor S3–S5 presence) evolves over the course of the trajectory to feature dominant S3–S5 inhibition (Figure 8).

Eldeoisin shows persistent interaction with Thr26. Also, it interacts with His41 and, to a lesser extent, with Cys145. This means that this ligand would efficiently bind the S2 subsite of 3CL<sup>Pro</sup>. In addition, it interacts with Glu166 and Gln189. However, the glutamine contacts, despite being a





## Calculation of Relative Protein–Ligand Binding Free Energies Using the Molecular Mechanics–Generalized Born Surface Area Method

Notoamide R, which showed an inefficient performance to inhibit 3CL<sup>Pro</sup> during the MD analysis due to poor contacts with binding site residues and high instability, also showed poor MM-GBSA values, with  $-21.4$  kcal/mol.

SN00017653, which is considered a strong inhibitor considering the RMSD analysis from MD results, shows a strong affinity, with  $-54.5$  kcal/mol, and low deviation values. This MM-GBSA value is comparable to SN000303378, but this ligand, despite having a strong affinity value, does not interact with key binding site residues. SN00019468 shows also a strong affinity value, but this ligand does not interact significantly with the binding site residues and also has a high instability, which is translated to an elevated standard deviation value ( $16.5$  kcal/mol), which is comparable to notoamide R; these two ligands are less suitable for inhibition of 3CL<sup>Pro</sup>.

Both dianthin E and pseudostellarin C suffer a pose rearrangement during the 100-ns simulation, but the interaction within the binding site persists during the whole simulation despite of interacting with different residues. Thus, they were considered possible 3CL<sup>Pro</sup> inhibitors, and the MM-GBSA values indicate their strong affinity.

Aliskiren showed a stable conformation during the 100-ns simulation, but with intermittent contacts with binding site residues. MM-GBSA values show a strong interaction, but this is not relevant due to the discontinuous contacts.

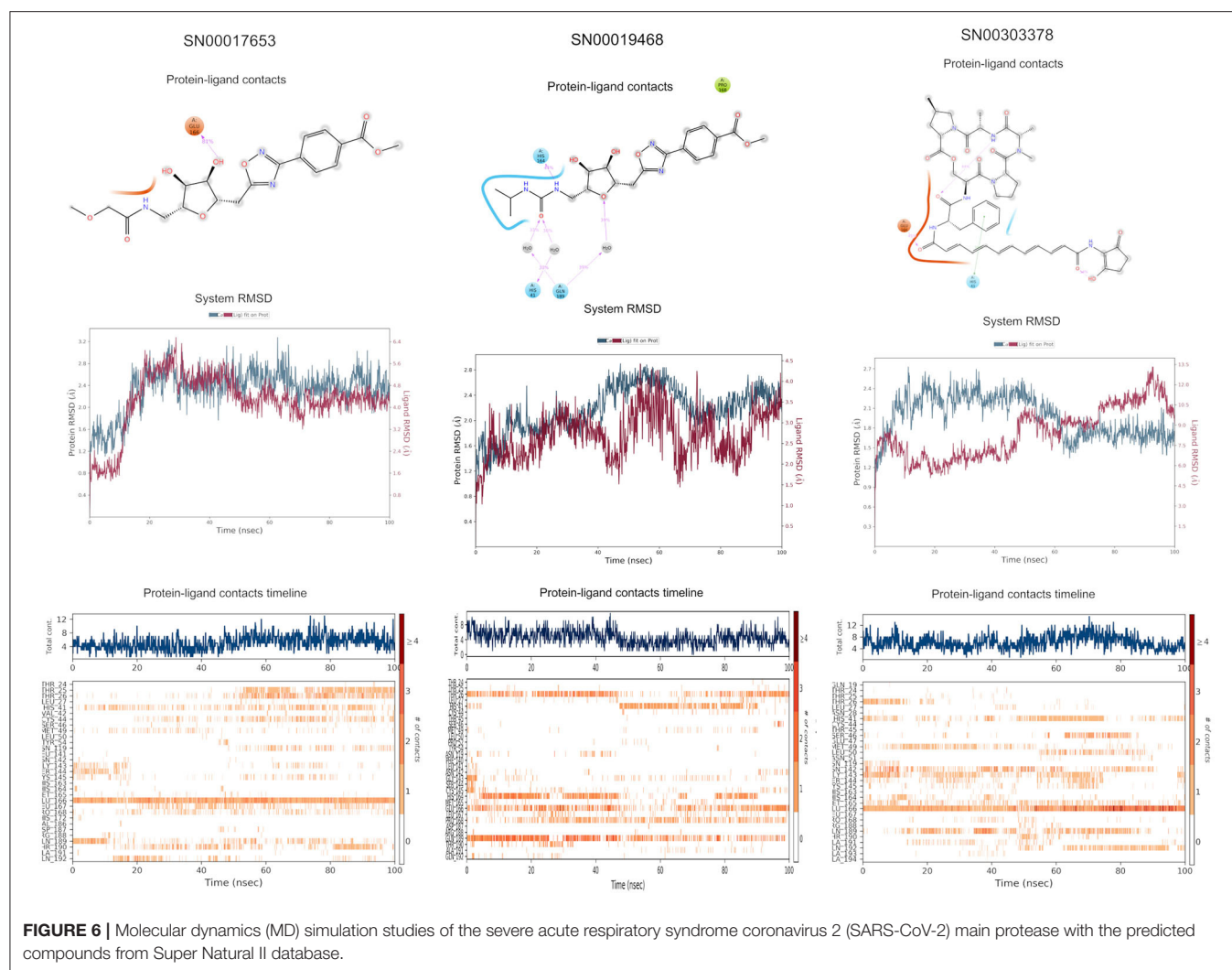
Both saquinavir and saralasin showed a good performance during the MD analysis, and the MM-GBSA values also indicate a strong affinity of these two ligands within the binding site.

Eledoisin showed an important inhibitory potential during the MD analysis, and this is correlated to the high MM-GBSA affinity value, which shows the best score for all studied ligands ( $-93.3$  kcal/mol).

Angiotensin II also showed persistent contacts with the main key residues of the binding site during the 100-ns simulation, and this is also extrapolated to the MM-GBSA values, which show a high-affinity energy for this ligand ( $-74.9$  kcal/mol).

Naldemedine, which interacts within the key residues of the binding site, but has less persistent contacts than eledoisin and angiotensin II, also shows a lower MM-GBSA affinity ( $-64.2$  kcal/mol) but comparable to the N3 ligand ( $-64.3$  kcal/mol).

Excluding notoamide R, all these ligands showed good MM-GBSA affinity values in comparison to the previously identified N3 inhibitor. Nevertheless, the MD analysis showed poor interaction of SN00019468, SN000303378, and aliskiren within the binding site, so their corresponding MM-GBSA affinities are not considered relevant. Dianthin E, pseudostellarin C, saquinavir, saralasin, eledoisin, angiotensin II, and naldemedine



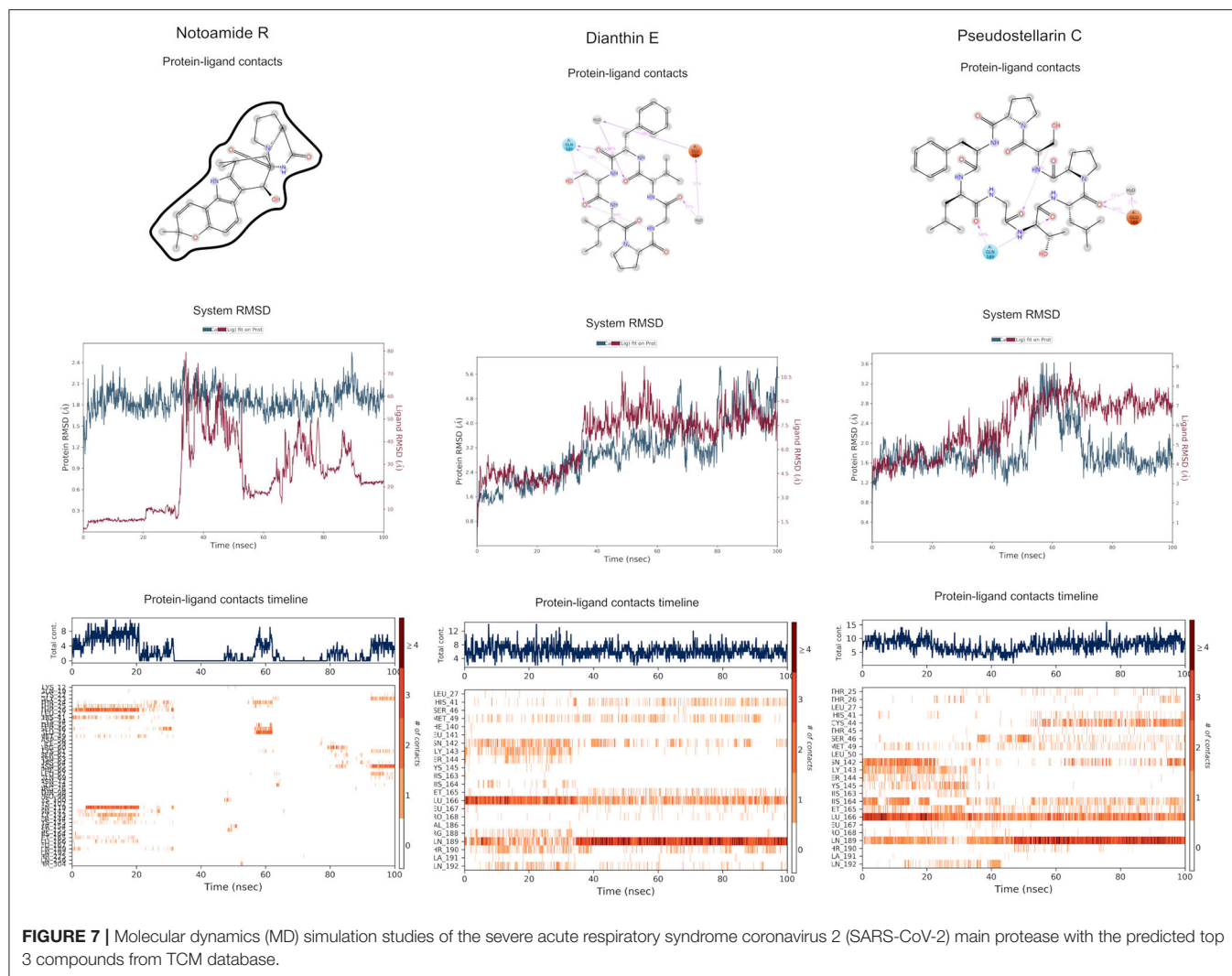
are considered as potential 3CL<sup>PRO</sup> inhibitors, though eldeoisin and angiotensin II are the most promising ones due to their lower MM-GBSA values (Table 4). The obtained results (Table 4) indicate the mean of the energy calculated for each pose of the simulation and its corresponding standard deviation.

## CLINICAL INSIGHTS

Our ultimate goal is to identify putative drug compounds that can be used safely and efficaciously while mitigating risks—deleterious side effects likely could not be reliably or robustly tolerated in severe COVID-19 cases. The general strategy of “drug repurposing” involves identifying existing compounds (both approved and withdrawn drugs) via their biological plausibility/rationale (e.g., mechanism-based inhibitors), via *in vitro*, *in vivo*, and *in silico* studies, or via serendipitous clinical observations. Much clinical pharmacological data, and clinical trial knowledge, are required in order to really elucidate (and extend) the use of a given chemical for a new indication; such efforts can stem from clinical expertise or

smaller-scale studies (before a fully systematic, population-wide study). A consideration of the possible strengths and weaknesses of the drug candidates predicted herein will require expertise in translational drug development, including clinical pharmacologists and infectious disease specialists conducting clinical trials related to COVID-19. As an alternative to traditional drug development strategies, which are often slow, financially costly, and failure-prone, drug repositioning approaches, though computationally intricate, can be especially useful in emergency situations such as the COVID-19 pandemic. Identifying and selecting molecular candidates for drug repositioning entails numerous factors—e.g., pharmacokinetics, clinical indications, drug-related adverse events, drug–drug interactions, toxicity profiles, and available formulations.

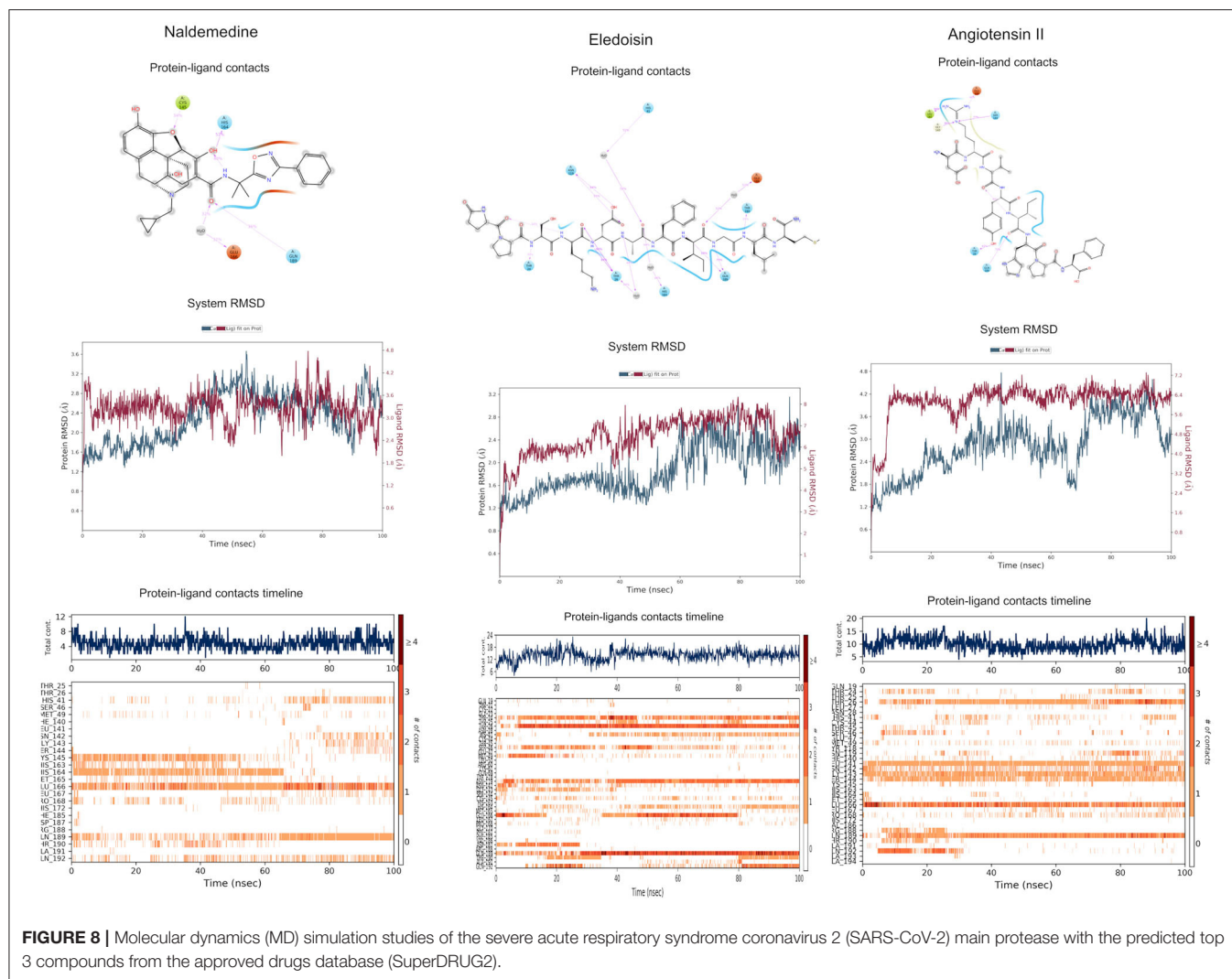
With regard to pharmacokinetics, metabolism plays a key role in this selection. The Phase I metabolism, through the CYP450 family, significantly increases the risk of drug–drug interactions when coadministered with CYP inhibitors or inducers. Moreover, idiosyncratic genetic variability in the CYP gene family may affect an individual’s response to the



**FIGURE 7 |** Molecular dynamics (MD) simulation studies of the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) main protease with the predicted top 3 compounds from TCM database.

administered treatment, both in terms of effectiveness and tolerability. By these considerations, we suggest that eledoisin, daptomycin, and angiotensin II, which have no interaction with the CYP system (**Supplementary Table 4**), could be potential inhibitors of the SARS-CoV-2 main protease. However, the vasodilatory activity of eledoisin may render it unsuitable in critically ill patients because severe cases (e.g., septic shock) often require vasopressor support; in contrast, angiotensinamide, which is an oligopeptide used to increase blood pressure by vasoconstriction, could be an interesting option in septic patients. Drugs not specifically indicated for cardiovascular disease may also affect hemodynamics. For instance, dihydroergocornine, a dopamine agonist used as an anti-Parkinson agent, presents large hypotensive effects; therefore, it should not be considered as a potential medication for COVID-19-infected patients. A consensus seems to be emerging that cardiovascular drugs (or other medications with significant hemodynamic effects) should not be considered promising candidates for targeting SARS-CoV-2, as their overall efficacy (and potential side effects) is too coupled to other clinical aspects of a patient's condition.

There are also additional considerations in the drug selection process, such as the environments of use and how a patient will interact with the drug, and these factors may influence the chosen administration route. Most infected COVID-19 patients have been managed in non-critical areas; therefore, in this context, the oral route of administration can be considered feasible, alongside with intravenous injections. Naldemedine, which is an oral peripherally acting  $\mu$ -opioid receptor antagonist (PAMORA), indicated for opioid-induced constipation (Coluzzi et al., 2020), may represent a practicable alternative in less severe COVID-19 patients. Its use, in healthy subjects, was associated with a slight increase in the incidence of diarrhea (Fukumura et al., 2018). Conversely, in critically ill patients, even when the enteral nutrition is guaranteed, oral formulations could be unsuitable if they cannot be crushed or dissolved and administered through the enteral feeding tube, as is the case with [CM1] naldemedine. From a clinical point of view, the approved drug indication could be a relevant criterion for selection, analogous to what holds true for antimicrobials and antiviral drugs. Indeed, the COVID-19 pandemic has seen



**FIGURE 8 |** Molecular dynamics (MD) simulation studies of the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) main protease with the predicted top 3 compounds from the approved drugs database (SuperDRUG2).

remdesivir, an established drug with broad-spectrum antiviral activity, receive emergency use authorization from the FDA and the European Medicines Agency (EMA) (Grein et al., 2020). Similarly, the antiviral medication telaprevir, a hepatitis C virus protease inhibitor, could represent an alternative. Among antibacterial agents, daptomycin, which is a lipopeptide antibiotic with *in vitro* bactericidal activity against Gram-positive bacteria, could be interesting potential inhibitors for SARS-CoV-2 targets. Daptomycin remains one of the main treatment options for methicillin-resistant *Staphylococcus aureus* (MRSA) infections; however, sporadic cases of resistance have been noted (Barros et al., 2019). In drug repositioning approaches, even withdrawn drugs can become reborn—particularly if the reason for market withdrawal was commercial (i.e., not safety issues). The top three withdrawn compounds that we have identified here include two cardiovascular drugs and an antiviral. Saralasin, an old partial agonist of angiotensin II receptors, has been withdrawn from sale for commercial reasons. Similarly, aliskiren, a direct renin inhibitor—which failed to show

benefit over angiotensin-converting-enzyme (ACE) inhibitors in heart failure<sup>21</sup>—was withdrawn from the European market, without intention to market it in the future. Saquinavir, indicated for treatment of HIV-1-infected adult patients, also could be an appealing potential drug for COVID-19 patients. However, apart from its withdrawal from the market, the CYP activity profile causes significant potential clinically relevant drug-drug interactions with a number of coadministered drugs. When considering the top 10 withdrawn compounds (Supplementary Table 5)—excluding cardiovascular drugs (for concerns expressed above) and antineoplastic agents (for reasons of toxicity)—we suggest that antibacterial drugs could be considered. Azlocillin, a wide-spectrum acylated form of ampicillin with antibacterial activity, has been recently proposed as a potential drug candidate for Lyme disease (targeting drug-tolerant *Borrelia burgdorferi*). Besides its efficacy, the safety of

<sup>21</sup> End of the road for Aliskiren in heart failure. Available online at: <https://academic.oup.com/eurheartj/article/38/5/312/2990021> (accessed July 28, 2020).



**TABLE 4 |** Molecular mechanics-generalized Born surface area (MM-GBSA) calculation.

Drug	dG (kcal/mol)	SD (kcal/mol)
SN00019468	−54.0	13.5
SN00017653	−54.5	5.8
SN00303378	−54.3	6.4
Pseudostellarin C	−53.6	12.1
Notoamide R	−21.4	17.5
Dianthin E	−61.70	10.6
Angiotensin II	−74.9	6.8
Eledoisin	−93.3	8.9
Naldemedine	−64.2	6.7
Saquinavir	−54.3	6.4
Aliskiren	−61.2	9.9
Saralasin	−61.8	10.6
N3 inhibitor	−64.3	12.3

dG, free energy; SD, standard deviation. The obtained results indicate the mean of the energy calculated for each pose of the simulation and its corresponding standard deviation.

plant (*Pseudostellaria heterophylla*), is known medicinally for its application in dry cough arising from “lung dryness” (Hu et al., 2019).

Naturally occurring compounds are a rich resource for drug innovation and development. We suggest that the COVID-19-related leads reported here can support the discovery and development of high-potency inhibitors *in vitro* and *in vivo*. New leads from the Super Natural II databases (Figure 2, Table 2) are novel promising candidates, as they have similar binding interaction profiles [including overall good structural similarity (0.73 and above)] as compared to the main N3 inhibitors. Besides that, the compounds from the TCM database have also shown good interactions with the main protease (Figures 3, 7, Table 2). Specially, compound pseudostellarin C is a promising candidate from the TCM chemical space, sharing similar binding interactions as the N3 ligand and showing a high stable pose from the first 40 ns (Figure 7) on MD simulation studies.

Additionally, the repurposed drugs (as shown in Table 3, Figures 4, 5) have shown good interactions; in particular, naldemedine from the approved drug set, which is a PAMORA recently approved for the treatment of opioid-induced constipation in adult patients (Hu and Bridgeman, 2018). This drug is also supported as a clinically valid alternative due to its safe profile. Furthermore, based on the MD simulation studies, naldemedine shows close values for protein and ligand RMSD, so the ligand pose is considered stable. Another interesting candidate is saquinavir from the withdrawn dataset, which interacts with the main protease in a similar manner as compared to its original ligand (Figure 6). It has a structural similarity of 0.67 with the original N3 ligand. Saquinavir is an antiretroviral protease (peptidomimetic) inhibitor that is used in the therapy and prevention of human immunodeficiency virus (HIV) infection and the acquired immunodeficiency

syndrome (AIDS). This drug is discontinued in Europe<sup>10</sup>, and due to its CYP activity profile, which causes significant potential clinically relevant drug–drug interactions with a number of coadministered drugs, this drug becomes clinically less preferable as a COVID-19 potential drug candidate. Saquinavir is also reported as a potential repurposed drug for COVID-19 disease by other studies (Montenegro et al., 2020). On the other hand, drugs such as angiotensin II, aliskiren, and phytochemicals like notoamide R, SN00019468, dianthin E, and SN00303378 cannot be considered as optimal candidates based on this study, as some of these compounds were stable but showed poor contacts with crucial residues of the main protease, as well as some showed higher RMSD values on MD simulation studies, and sometimes both poor contacts and high ligand RMSD fluctuations were observed (see Results section). Additionally, in this study, we have also addressed the toxicity and cytochrome activity of the reported compounds. Most of the resulting compounds predicted to be immunotoxic (that is, cytotoxicity of the B and T cells). Thus, we believe gaining insight into the molecular mechanism responsible for protein–ligand recognition through this study will facilitate the development of drugs for the treatment of COVID-19 disease.

The work reported here addresses an important concern and urgent need for drugs for the treatment of SARS-CoV-2 infection. As demonstrated via this integrated approach, computational prediction of approved drugs, withdrawn drugs, and phytochemicals for inhibition of SARS-CoV-2 main protease has resulted in some promising leads for further experimental validation. We hope that the *in silico* results and predictions obtained in this study, including the potential clinical insights, could facilitate the discovery of highly potent inhibitors of the SARS-CoV-2 main protease. Overall, our computational drug repositioning strategy predicts some promising drug candidates that, if borne out *via* experimental and clinical approaches, could contribute toward resolving the global crisis of the COVID-19 pandemic.

## DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

## AUTHOR CONTRIBUTIONS

RP, MS, and PB conceived the project. RA, MP, HP-S, QC, FC, MR, RP, and PB designed the project workflow. RA, QC, MP, and PB prepared the databases. QC performed SuperTCM data curations. RA, MP, HP-S, and PB implemented the project. RA performed molecular docking. RA and PB analyzed and selected the compounds. MP and HP-S performed the molecular dynamics. FC, MP, PC, and MS contributed to the clinical insights and clinical trial study. RA, MP, and PB provided the figures. RA and QC provided the **Supplementary Material**. PB, RA, MP,

FC, CM, and PB contributed to manuscript writing. All authors proofread the final manuscript.

## FUNDING

This work has been partially funded by grants from the Spanish Ministry of Economy and Competitiveness (CTQ2017-87974-R) and by the Fundación Séneca del Centro de Coordinación de la Investigación de la Región de Murcia under Project 20988/PI/18. Supercomputing resources in this work have been supported by the infrastructures of Poznan Supercomputing Center, the e-infrastructure program of the Research Council of Norway, and the supercomputer center of UiT-the Arctic University of Norway, by the Plataforma Andaluza de Bioinformática of the University of Málaga, by the supercomputing infrastructure of the NLHPC (ECM-02, Powered@NLHPC), and by the Extremadura Research Centre for Advanced Technologies (CETA-CIEMAT), funded by the European Regional Development Fund (ERDF). CETA-CIEMAT belongs to CIEMAT and the Government of Spain. Portions of this work were also supported by the University of Virginia School of Data Science and by NSF Career award MCB-1350957. We acknowledge support from the German Research Foundation

(DFG) (KFO 339 and TRR 295) and the Open Access Publication Fund of Charité-Universitätsmedizin Berlin.

## ACKNOWLEDGMENTS

Currently, clinical study is aimed and approved at evaluating the activity of drug Naldemedine in asymptomatic subjects positive for SARS-CoV-2 in terms of reducing the number of subjects presenting symptoms or reducing the number of positive days of the test and a second study in symptomatic patients to assess the reduction in the number and extent of symptoms and the time to recovery. Both studies will be phase II studies, with a control group consisting of patients treated according to the clinical practice of the various participating centers in the project. The clinical study is approved by the Shionogi pharmaceutical study, and will be conducted at Sant'Andrea University Hospital, Rome, Italy.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fchem.2020.590263/full#supplementary-material>

## REFERENCES

- Al-Tawfiq, J. A., Al-Homoud, A. H., and Memish, Z. A. (2020). Remdesivir as a possible therapeutic option for the COVID-19. *Travel Med. Infect. Dis.* 34, 101615. doi: 10.1016/j.tmaid.2020.101615
- Anand, K., Ziebuhr, J., Wadhwani, P., Mesters, J. R., and Hilgenfeld, R. (2003). Coronavirus main proteinase (3CLpro) structure: basis for design of anti-SARS drugs. *Science* 300, 1763–1767. doi: 10.1126/science.1085658
- Barros, E. M., Martin, M. J., Selleck, E. M., Lebreton, F., Sampaio, J. L. M., and Gilmore, M. S. (2019). Daptomycin resistance and tolerance due to loss of function in *Staphylococcus aureus* *dsp1* and *asp23*. *Antimicrob. Agents Chemother.* 63:e01542–18. doi: 10.1128/AAC.01542-18
- Beigel, J. H., Tomashek, K. M., Dodd, L. E., Mehta, A. K., Zingman, B. S., Kalil, A. C., et al. (2020). Remdesivir for the treatment of Covid-19 — preliminary report. *N. Engl. J. Med.* 383, 1813–1826. doi: 10.1056/nejmoa2007764
- Berthold, M. R., Cebon, N., Dill, F., Gabriel, T. R., Kötter, T., Meinl, T., et al. (2008). *KNIME: The Konstanz Information Miner*. Berlin: Springer. doi: 10.1145/1656274.1656280
- Botta, L., Rivara, M., Zuliani, V., and Radi, M. (2018). Drug repurposing approaches to fight Dengue virus infection and related diseases. *Front. Biosci.* 23, 997–1019. doi: 10.2741/4630
- Cao, B., Wang, Y., Wen, D., Liu, W., Wang, J., Fan, G., et al. (2020). A trial of lopinavir–ritonavir in adults hospitalized with severe covid-19. *N. Engl. J. Med.* 382, 1787–1799. doi: 10.1056/NEJMoa2001282
- Cha, Y., Erez, T., Reynolds, I. J., Kumar, D., Ross, J., Koytiger, G., et al. (2018). Drug repurposing from the perspective of pharmaceutical companies. *Brit. J. Pharmacol.* 175, 168–180. doi: 10.1111/bph.13798
- Chen, C. Y. C. (2011). TCM Database@Taiwan: The world's largest traditional Chinese medicine database for drug screening *in silico*. *PLoS ONE* 6:e15939. doi: 10.1371/journal.pone.0015939
- Coluzzi, F., Scerpa, M. S., and Pergolizzi, J. (2020). Naldemedine: a new option for OIBD. *J. Pain Res.* 13, 1209–1222. doi: 10.2147/JPR.S243435
- Contini, A. (2020). Virtual screening of an FDA approved drugs database on two COVID-19 coronavirus proteins. *ChemRxiv*. doi: 10.26434/CHEMRXIV.11847381.V1
- Durant, J. L., Leland, B. A., Henry, D. R., and Nourse, J. G. (2002). Reoptimization of MDL keys for use in drug discovery. *J. Chem. Inform. Comput. Sci.* 42, 1273–1280. doi: 10.1021/ci010132r
- Fukumura, K., Yokota, T., Baba, Y., and Arjona Ferreira, J. C. (2018). Phase 1, randomized, double-blind, placebo-controlled studies on the safety, tolerability, and pharmacokinetics of naldemedine in healthy volunteers. *Clin. Pharmacol. Drug Dev.* 7, 474–483. doi: 10.1002/cpdd.387
- Greenidge, P. A., Kramer, C., Mozziconacci, J.-C., and Wolf, R. M. (2013). MM/GBSA binding energy prediction on the pdbbind data set: successes, failures, and directions for further improvement. *J. Chem. Inform. Model.* 53, 201–209. doi: 10.1021/ci300425v
- Grein, J., Ohmagari, N., Shin, D., Diaz, G., Asperges, E., Castagna, A., et al. (2020). Compassionate use of remdesivir for patients with severe covid-19. *N. Engl. J. Med.* 382, 2327–2336. doi: 10.1056/NEJMoa2007016
- Hall, L. H., and Kier, L. B. (1995). Electrotopological state indices for atom types: a novel combination of electronic, topological, and valence state information. *J. Chem. Inform. Comput. Sci.* 35, 1039–1045. doi: 10.1021/ci00028a014
- Horby, P., Lim, W. S., Emberson, J., Mafham, M., Bell, J., Linsell, L., et al. (2020). Dexamethasone for COVID-19-preliminary report effect of dexamethasone in hospitalized patients with COVID-19 - preliminary report. *medRxiv*. doi: 10.1101/2020.06.22.20137273
- Hu, D. J., Shakerian, F., Zhao, J., and Li, S. P. (2019). Chemistry, pharmacology and analysis of *Pseudostellaria heterophylla*: a mini-review. *Chin. Med.* 14, 1–8. doi: 10.1186/s13020-019-0243-z
- Hu, K., and Bridgeman, M. B. (2018). Naldemedine (Symproic) for the treatment of opioid-induced constipation. *P T* 43, 601–605.
- Ihlenfeldt, W. D., Bolton, E. E., and Bryant, S. H. (2009). The PubChem chemical structure sketcher. *J. Cheminform.* 1:20. doi: 10.1186/1758-2946-1-20
- Jackson, L. A., Anderson, E. J., Rouphael, N. G., Roberts, P. C., Makhene, M., Coler, R. N., et al. (2020). An mRNA vaccine against SARS-CoV-2 — preliminary report. *N. Engl. J. Med.* 383, 1920–1931. doi: 10.1056/NEJMoa2022483
- Jin, Z., Du, X., Xu, Y., Deng, Y., Liu, M., Zhao, Y., et al. (2020). Structure of Mpro from SARS-CoV-2 and discovery of its inhibitors. *Nature* 582, 289–293. doi: 10.1038/s41586-020-2223-y

- Jones, G., Willett, P., Glen, R. C., Leach, A. R., and Taylor, R. (1997). Development and validation of a genetic algorithm for flexible docking. *J. Mol. Biol.* 267, 727–748. doi: 10.1006/jmbi.1996.0897
- Khedkar, P. H., and Patzak, A. (2020). SARS-CoV-2: what do we know so far? *Acta Physiol.* 229:e13470. doi: 10.1111/apha.13470
- Kollman, P. A., Massova, I., Reyes, C., Kuhn, B., Huo, S., Chong, L., et al. (2000). Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. *Acc. Chem. Res.* 33, 889–897. doi: 10.1021/ar000033j
- Kupferschmidt, K., and Cohen, J. (2020). Race to find COVID-19 treatments accelerates. *Science* 367, 1412–1413. doi: 10.1126/science.367.6485.1412
- Mark, P., and Nilsson, L. (2001). Structure and dynamics of the TIP3P, SPC, and SPC/E water models at 298 K. *J. Phys. Chem. A* 105, 9954–9960. doi: 10.1021/jp003020w
- McNamee, L. M., Walsh, M. J., and Ledley, F. D. (2017). Timelines of translational science: from technology initiation to FDA approval. *PLoS ONE* 12. doi: 10.1371/journal.pone.0177371
- Metushi, I. G., Wriston, A., Banerjee, P., Gohlke, B. O., English, A. M., Lucas, A., et al. (2015). Acyclovir has low but detectable influence on HLA-B\*57:01 specificity without inducing hypersensitivity. *PLoS ONE* 10:e0124878. doi: 10.1371/journal.pone.0124878
- Montenegro, M. F., Al-Abed, Y., He, M., Tracey, K. J., and Billiar, T. R. (2020). HIV protease inhibitors saquinavir and nelfinavir are potent inhibitors of cathepsin L activity: a potential treatment for COVID-19 patients. doi: 10.21203/rs.3.rs-37258/v1
- Muramatsu, T., Takemoto, C., Kim, Y. T., Wang, H., Nishii, W., Terada, T., et al. (2016). SARS-CoV 3CL protease cleaves its C-terminal autoproteolytic site by novel subsite cooperativity. *Proc. Natl. Acad. Sci. U.S.A.* 113, 12997–13002. doi: 10.1073/pnas.1601327113
- Newman, D. J., and Cragg, G. M. (2016). Natural products as sources of new drugs from 1981 to 2014. *J. Nat. Prod.* 79, 629–661. doi: 10.1021/acs.jnatprod.5b01055
- Nutho, B., Mahalapbutr, P., Hengphasatporn, K., Pattarangoon, N. C., Simanon, N., Shiget, Y., et al. (2020). Why are lopinavir and ritonavir effective against the newly emerged coronavirus 2019? Atomistic insights into the inhibitory mechanisms. *Biochemistry* 59, 1769–1779. doi: 10.1021/acs.biochem.0c00160
- Patterson, D. E., Cramer, R. D., Ferguson, A. M., Clark, R. D., and Weinberger, L. E. (1996). Neighborhood behavior: a useful concept for validation of “molecular diversity” descriptors. *J. Med. Chem.* 39, 3049–3059. doi: 10.1021/jm960290n
- Pizzorno, A., Padey, B., Terrier, O., and Rosa-Calatrava, M. (2019). Drug repurposing approaches for the treatment of influenza viral infection: reviving old drugs to fight against a long-lived enemy. *Front. Immunol.* 10:531. doi: 10.3389/fimmu.2019.00531
- Pothineni, V. R., Potula, H. S. K., Ambati, A., Mallajosyula, V. V. A., Sridharan, B., Inayathullah, M., et al. (2020). Azlocillin can be the potential drug candidate against drug-tolerant *Borrelia burgdorferi* sensu stricto JLB31. *Sci. Rep.* 10, 1–15. doi: 10.1038/s41598-020-59600-4
- Rogers, D., and Hahn, M. (2010). Extended-connectivity fingerprints. *J. Chem. Inform. Model.* 50, 742–754. doi: 10.1021/ci100050t
- Scasso, F., Ferrari, G., de Vincentiis, G. C., Arosio, A., Bottero, S., Carretti, M., et al. (2018). Emerging and re-emerging infectious disease in otorhinolaryngology. *Acta Otorhinolaryngol. Ital.* 38, S1–S106. doi: 10.14639/0392-100X-suppl.1-38-2018
- Shiryaev, S. A., Mesci, P., Pinto, A., Fernandes, I., Sheets, N., Shrestha, S., et al. (2017). Repurposing of the anti-malaria drug chloroquine for Zika Virus treatment and prophylaxis. *Sci. Rep.* 7:15771. doi: 10.1038/s41598-017-15467-6
- Stumpfe, D., and Bajorath, J. (2011). Similarity searching. *WIREs Comput. Mol. Sci.* 1, 260–282. doi: 10.1002/wcms.23
- Ton, A.-T., Gentile, F., Hsing, M., Ban, F., and Cherkasov, A. (2020). Rapid identification of potential inhibitors of SARS-CoV-2 main protease by deep docking of 1.3 billion compounds. *Mol. Inform.* 39:2000028. doi: 10.1002/minf.202000028
- Turanli, B., Zhang, C., Kim, W., Benfeitas, R., Uhlen, M., Arga, K. Y., et al. (2019). Discovery of therapeutic agents for prostate cancer using genome-scale metabolic modeling and drug repositioning. *EBioMedicine* 42, 386–396. doi: 10.1016/j.ebiom.2019.03.009
- Wang, J. (2020). Fast identification of possible drug treatment of Coronavirus Disease-19 (COVID-19) through computational drug repurposing study. *J. Chem. Inform. Model.* 60, 3277–3286. doi: 10.1021/acs.jcim.0c00179
- Wu, F., Zhao, S., Yu, B., Chen, Y. M., Wang, W., Song, Z. G., et al. (2020). A new coronavirus associated with human respiratory disease in China. *Nature* 579, 265–269. doi: 10.1038/s41586-020-2008-3
- Yang, P., and Wang, X. (2020). COVID-19: a new challenge for human beings. *Cell. Mol. Immunol.* 17, 555–557. doi: 10.1038/s41423-020-0407-x
- Zhang D.-h., Wu K.-l., Zhang X., Deng S.-q., and Peng B. (2020). *In silico* screening of Chinese herbal medicines with the potential to directly inhibit 2019 novel coronavirus. *J. Integr. Med.* 18, 152–158. doi: 10.1016/j.joim.2020.02.005
- Zhang, L., Lin, D., Sun, X., Curth, U., Drosten, C., Sauerhering, L., et al. (2020). Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved a-ketoamide inhibitors. *Science* 368, 409–412. doi: 10.1126/science.abb3405

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The reviewer KV declared a past co-authorship with one of the authors HP-S to the handling Editor.

Copyright © 2020 Abel, Paredes Ramos, Chen, Pérez-Sánchez, Coluzzi, Rocco, Marchetti, Mura, Simmaco, Bourne, Preissner and Banerjee. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Elucidating Interactions Between SARS-CoV-2 Trimeric Spike Protein and ACE2 Using Homology Modeling and Molecular Dynamics Simulations

Sugunadevi Sakkiah, Wenjing Guo, Bohu Pan, Zuowei Ji, Gokhan Yavas, Marli Azevedo, Jessica Hawes, Tucker A. Patterson and Huixiao Hong\*

National Center for Toxicological Research, U.S. Food and Drug Administration, Jefferson, AR, United States

## OPEN ACCESS

### Edited by:

Domenica Capasso,  
University of Naples Federico II, Italy

### Reviewed by:

Kalaimathy Parthiban,  
University of Turku, Finland  
Yun Tang,  
East China University of Science and  
Technology, China

### \*Correspondence:

Huixiao Hong  
Huixiao.hong@fda.hhs.gov

### Specialty section:

This article was submitted to  
Theoretical and Computational  
Chemistry,  
a section of the journal  
Frontiers in Chemistry

**Received:** 28 October 2020

**Accepted:** 09 December 2020

**Published:** 05 January 2021

### Citation:

Sakkiah S, Guo W, Pan B, Ji Z,  
Yavas G, Azevedo M, Hawes J,  
Patterson TA and Hong H (2021)  
Elucidating Interactions Between  
SARS-CoV-2 Trimeric Spike Protein  
and ACE2 Using Homology Modeling  
and Molecular Dynamics Simulations.  
*Front. Chem.* 8:622632.  
doi: 10.3389/fchem.2020.622632

Severe Acute Respiratory Syndrome Coronavirus-2 (SARS-CoV-2) causes coronavirus disease 2019 (COVID-19). As of October 21, 2020, more than 41.4 million confirmed cases and 1.1 million deaths have been reported. Thus, it is immensely important to develop drugs and vaccines to combat COVID-19. The spike protein present on the outer surface of the virion plays a major role in viral infection by binding to receptor proteins present on the outer membrane of host cells, triggering membrane fusion and internalization, which enables release of viral ssRNA into the host cell. Understanding the interactions between the SARS-CoV-2 trimeric spike protein and its host cell receptor protein, angiotensin converting enzyme 2 (ACE2), is important for developing drugs and vaccines to prevent and treat COVID-19. Several crystal structures of partial and mutant SARS-CoV-2 spike proteins have been reported; however, an atomistic structure of the wild-type SARS-CoV-2 trimeric spike protein complexed with ACE2 is not yet available. Therefore, in our study, homology modeling was used to build the trimeric form of the spike protein complexed with human ACE2, followed by all-atom molecular dynamics simulations to elucidate interactions at the interface between the spike protein and ACE2. Molecular Mechanics Poisson-Boltzmann Surface Area (MMPBSA) and *in silico* alanine scanning were employed to characterize the interacting residues at the interface. Twenty interacting residues in the spike protein were identified that are likely to be responsible for tightly binding to ACE2, of which five residues (Val445, Thr478, Gly485, Phe490, and Ser494) were not reported in the crystal structure of the truncated spike protein receptor binding domain (RBD) complexed with ACE2. These data indicate that the interactions between ACE2 and the tertiary structure of the full-length spike protein trimer are different from those between ACE2 and the truncated monomer of the spike protein RBD. These findings could facilitate the development of drugs and vaccines to prevent SARS-CoV-2 infection and combat COVID-19.

**Keywords:** SARS-CoV-2, spike protein, molecular dynamics simulations, homology modeling, COVID-19

## INTRODUCTION

The Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) was first identified in Hubei, China, and causes the severe respiratory syndrome known as COVID-19 in humans. Seven strains of human coronaviruses have been identified, which include human coronavirus HKU1 (HCoV-HKU1), human coronavirus 229E (HCoV-229E), human coronavirus NL63 (HCoV-NL63), human coronavirus OC43 (HCoV-OC43), severe acute respiratory syndrome coronavirus (SARS-CoV), Middle East respiratory syndrome-related coronavirus (MERS-CoV), and SARS-CoV-2 (Malik, 2020). Coronaviruses are composed of four genera belonging to the *coronaviridae* family (Zhang and Liu, 2020), wherein SARS-CoV-2, SARS-CoV, and MERS-CoV belong to the  $\beta$ -coronavirus genus (Petrosillo et al., 2020). Despite a high degree of structural homology with SARS-CoV and MERS-CoV (Jaimes et al., 2020), SARS-CoV-2 is much more transmissible than its predecessors, which may be attributable to unique differences in the spike protein (Rabaan et al., 2020). As of October 21, 2020, more than 1.1 million deaths and 41.4 million infected cases have been confirmed (<https://www.worldometers.info/coronavirus/>), with the COVID-19 pandemic remaining a significant global threat for eight continuous months and counting.

SARS-CoV-2 is a positive-sense, single-stranded RNA virus that encodes 16 non-structural proteins (NSP1-NSP16), four structural proteins (spike, membrane, envelope, and nucleocapsid), and nine accessory proteins (Figure 1; Romano et al., 2020). Spike proteins present on the virion surface are responsible for targeting host cells and triggering fusion of viral and host cell membranes, which are critical steps in initiating infection and enabling the transfer of viral RNA into host cells. Membrane and envelope proteins are responsible for the virus shape and assembly/budding, respectively. The nucleocapsid protein enters the host cell along with the SARS-CoV-2 genetic material, which serves to facilitate RNA transcription, replication, virus assembly, and release (Kang et al., 2020; Zeng et al., 2020). Since the spike protein plays a major role in initializing viral infection through binding to ACE2, inhibiting the binding of the spike protein to ACE2 is an attractive strategy for developing drugs to block the spread of SARS-CoV-2 infection and treat COVID-19 (Das et al., 2020; Wu et al., 2020). Therefore, understanding interactions between the spike protein and ACE2 may facilitate the development of drugs that target binding of the spike protein to ACE2.

The spike protein contains 1273 amino acids and is composed of two subunits, S1 (amino acids 14-685) and S2 (amino acids 686-1273), which are responsible for receptor binding and membrane fusion with the host cell, respectively, preceded by a short signal peptide (amino acids 1-13). The S1 subunit consists of three domains: an N-terminal domain (NTD; amino acids 14-305), a receptor binding domain (RBD; amino acids 319-541), and a carboxy-terminal domain (CTD) which has two subdomains (SD1 and SD2) (Henderson et al., 2020; Huang et al., 2020; Tang et al., 2020a,b). The S2 subunit consists of a fusion peptide (FP; amino acids 788-806) composed of hydrophobic residues, heptapeptide repeat 1 (HR1; amino acids

912-984), heptapeptide repeat 2 (HR2; amino acids 1163-1213), a transmembrane domain (TM; amino acids 1213-1237), and a cytoplasmic domain (CP; amino acids 1237-1273) (Astuti and Ysrafil, 2020; Huang et al., 2020; Tang et al., 2020a,b). In its native form, the spike protein is present as a trimer on the surface of the virion, with the S1 and S2 subunits forming the extracellular stalk and bulbous “crown,” for which the Latin translation is “corona” (Huang et al., 2020; Tang et al., 2020a,b; Walls et al., 2020). The crown of the trimeric spike protein undergoes hinge-like conformational changes between a closed/down conformation and a less-stable open/up conformation (Huang et al., 2020; Tang et al., 2020a,b; Walls et al., 2020; Wrapp et al., 2020). In the open/up conformation, the RBD is accessible for binding to the ACE2 receptor; whereas, in the closed/down conformation, the RBD cannot interact with the ACE2 receptor (Ortega et al., 2020; Wrapp et al., 2020). Upon binding to the ACE2 receptor, the spike protein trimer undergoes a conformational change resulting in the accessibility of the S1 and S2 cleavage sites to host proteases (Lan et al., 2020; Shang et al., 2020; Xia et al., 2020). Cleavage of the S1 and S2 subunits primes the spike protein for membrane fusion by enabling insertion of the S2 FP domain into the host cell membrane, which enables subsequent interactions between the HR1 and HR2 coiled-coil domains to form a six helical bundle (6-HB). This bundle stabilizes another S2 subunit conformational change in which viral and host membranes are close enough in proximity to trigger membrane fusion (Huang et al., 2020; Tang et al., 2020a,b).

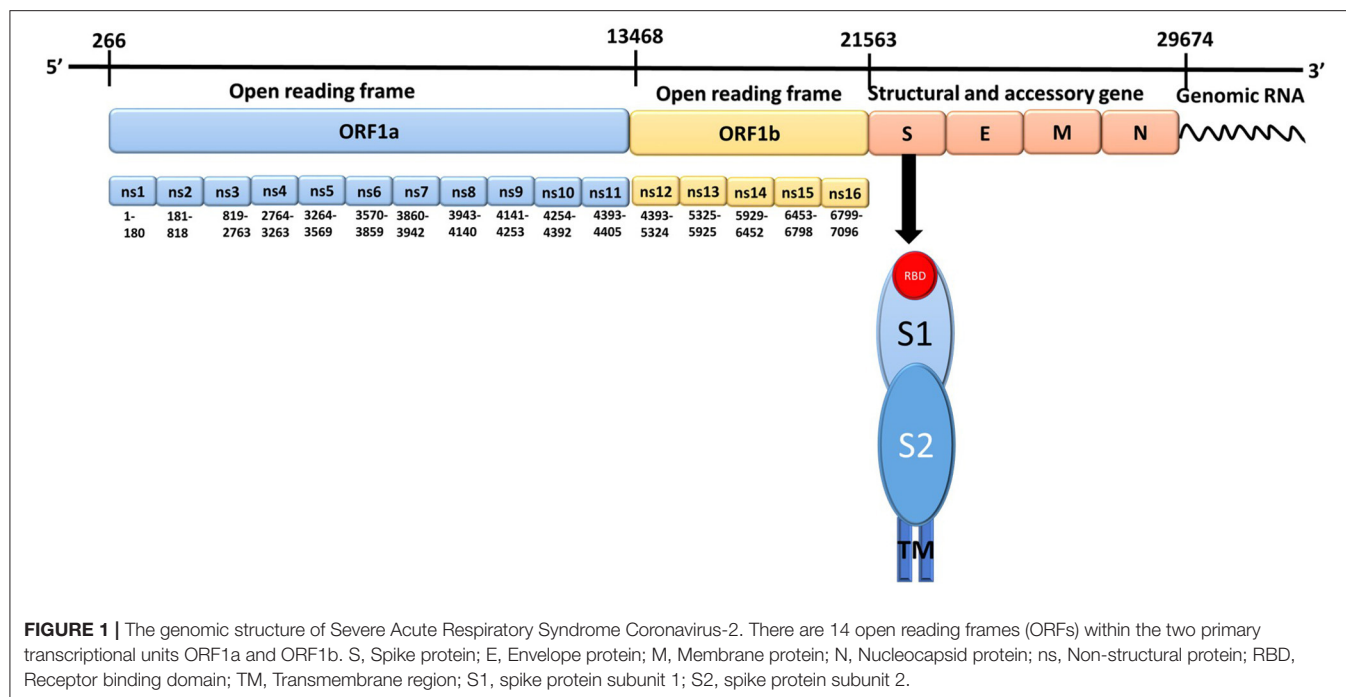
Recently, structures of the trimeric form of mutant or truncated spike proteins complexed with antibodies were reported. Walls et al. determined atomic models of SARS-CoV-2 spike protein in the closed/down (6VXX) trimeric conformation and a one-up (6VYB) trimeric conformation, in which a single S unit is in the open/up conformation and two S units are in the closed/down conformation (Walls et al., 2020). However, the full-length wild-type S protein was not determined due to missing residues and loops, and spike protein trimers bound to the ACE2 receptor were not evaluated. Thus, a crystal structure of the wild-type trimeric form of spike protein bound to ACE2 has not yet been determined (Song et al., 2018).

In this study, a complex structure of the wild-type trimeric spike protein with ACE2 was constructed using homology modeling based on the atomic details of the trimeric form of the spike protein with the one-up conformation and the RBD of the spike protein with ACE2. The homology models were evaluated using a Ramachandran plot and the highest quality model was subjected to molecular dynamics simulations to identify interactions between the trimeric spike protein and ACE2.

## METHODS AND MATERIALS

### Homology Modeling

The primary sequence of the spike protein (ID: P0DTC2) was retrieved from the UniProt database (<https://www.uniprot.org>) and used as a template sequence. The mutant trimeric spike protein of SARS-CoV-2 with the one-up conformation (PDB ID: 6VYB) (Walls et al., 2020) and the complex of spike protein RBD



of SARS-CoV-2 with ACE2 (PDB ID: 6M0J) (Lan et al., 2020) were selected as the template structures. Sequence alignment between templates and targets was performed using the EBI-Clustal Omega Server (Madeira et al., 2019). Homology modeling was performed using the Modeler v9.24 (Sali and Blundell, 1993) program. Ten models were generated, and one model was selected based on the DOPE energy value. To avoid steric clashes, the selected model was energy minimized using the Schrödinger suite ([www.schrodinger.com](http://www.schrodinger.com)) and evaluated for its stereo-chemical quality using a Ramachandran plot (<https://swissmodel.expasy.org/assess>).

## Molecular Dynamics Simulations

The validated homology model of the trimeric form of the spike protein complexed with ACE2 was used as the starting structure for molecular dynamics simulations using Amber 18 (<https://ambermd.org/CiteAmber.php>). The *tleap* from AmberTools was used to prepare topologies and coordination files for the protein using protein.ff18SB forcefield (Ponder and Case, 2003; Maier et al., 2015) by adding the force field and hydrogen atoms. The prepared system was then placed inside an octahedral box 10 Å away from the protein surface, solvated with the TIP3P (Jorgensen et al., 1983) water model, and subjected to energy minimization. The counter ions were added to neutralize the unbalanced charge of the system. The whole process was divided into 2 phases. In phase 1, the solute molecules were constrained and only the solvents were minimized and equilibrated. Subsequently, the steepest descent method (1,000 steps) followed by conjugate gradient (4,000 steps) methods were applied to minimize the whole system (solute and solvent). The minimized system was gradually heated from 0 to 310 K, and the entire system was equilibrated without any constraints.

During the simulations, the system temperature and pressure were maintained at 310.5 K and 1 atm, respectively. In Phase 2, a 100 nanosecond (ns) unrestrained production run was applied to the system. The SHAKE (Ryckaert et al., 1977) algorithm and Particle Mesh Ewald (PME) (Darden et al., 1993) were applied for the hydrogen bonds and the long-range electrostatic interactions, respectively. Coordinate files were saved for every 5 picosecond (ps). The resultant trajectory was processed using the CPPTRAJ (Roe and Cheatham, 2013) from AmberTool package, Visual Molecular Dynamics (Humphrey et al., 1996), and PyMol (<https://pymol.org>). The kclust algorithm from MMTSB toolset (<http://www.mmtsb.org/>) was used to cluster the resultant trajectory. The representative structure from the cluster analysis was used for hydrogen bond and energy analysis. The hydrogen bond between the trimeric form of the spike protein and ACE2 was determined using a distance cut-off value of 4 Å or less.

## Binding Free Energy Calculation

The spike protein, ACE2, and spike protein-ACE2 complex structures were used to calculate binding free energy. The *tleap* from AmberTools18 (<https://ambermd.org/AmberTools.php>) was used to generate the gas phase, solvated complex topology (prmtop); and coordination (inpcrd) files for the spike protein, ACE2, and spike protein-ACE2 complex. The MMPBSA.py (Miller et al., 2012) script from the Amber package was used to calculate the binding free energy using the Molecular Mechanism-Generalized Born Surface Area (MMGBSA) approach. The representative structure from clustering analysis was used to calculate binding free energies using Equation (1)

$$\Delta G_{BFE} = G_{S-ACE2} - G_S - G_{ACE2} \quad (1)$$

where  $\Delta G_{BFE}$  represents the binding free energy between the spike protein and ACE2. The terms  $G_{S-ACE2}$ ,  $G_S$ , and  $G_{ACE2}$  represent the free energy of the spike protein-ACE2 complex, spike protein, and ACE2, respectively. The MMPBSA.py script was used to calculate interaction and solvation free energies for the spike protein, ACE2, and spike protein-ACE2 complex. The energy values were calculated using Equation (2)

$$\Delta G = \Delta G_{gas} + \Delta G_{sol} - T\Delta S \quad (2)$$

where  $\Delta G_{gas}$ ,  $\Delta G_{sol}$ , and  $T\Delta S$  represent the gas phase molecular mechanism component, solvation of binding free energies, and changes in entropy due to ACE2 binding, respectively. Because our goal was to obtain an estimated free energy rather than an absolute value, and since the computational cost was high, entropy changes in the free energy calculation were ignored.

The gas phase molecular mechanism ( $\Delta G_{gas}$ ) and solvation free energy ( $\Delta G_{sol}$ ) were calculated using Equations (3–5)

$$\Delta G_{gas} = \Delta G_{ele} + \Delta G_{vdW} \quad (3)$$

$$\Delta G_{sol} = \Delta G_{pol.sol} + \Delta G_{nonpol.sol} \quad (4)$$

$$\Delta G_{non-polar} = \gamma * SASA + \beta \quad (5)$$

where  $\Delta G_{vdW}$  and  $\Delta G_{ele}$  correspond to van der Waals and electrostatic interactions, respectively, and  $\Delta G_{pol.sol}$  and  $\Delta G_{nonpol.sol}$  represent polar solvation and non-solvation terms from the MMGBSA approach. The terms SASA,  $\gamma$ , and  $\beta$  denote solvent-accessible surface area, surface tension, and regression offset of the linear relationship, respectively. Default values were used to calculate the estimated binding free energy between the spike protein and ACE2.

## Interaction Free Energy of Residues

The representative structure from clustering analysis was used to analyze interaction free energy for residues at the interface of the spike protein and ACE2 using the *decomp* (Miller et al., 2012) module from the MMPBSA.py program. Contributions of the interacting residues were calculated based on the following energy terms: electrostatic contribution, non-polar solvation contribution, van der Waals contribution, and polar solvation, using Equation (6).

$$\Delta G_{residue} = \Delta G_{ele} + \Delta G_{nonpol.sol} + \Delta G_{vdW} + \Delta G_{polar} \quad (6)$$

## Alanine Scanning

The selected interacting residues were mutated individually in the representative structure from the cluster analysis. The free energies of the mutated trimeric spike protein and ACE2 complexes were calculated by MMPBSA.py using Equation (7).

$$\Delta G_{BFE} = G_{S_{mut}-ACE2} - G_{S_{mut}} - G_{ACE2} \quad (7)$$

where  $\Delta G_{BFE}$  represents the estimated binding free energy for mutated spike protein and ACE2. The  $G_{S_{mut}-ACE2}$ ,  $G_{S_{mut}}$ , and  $G_{ACE2}$  terms represent the free energy components estimated for

**TABLE 1 |** The DOPE energy values for homology models of the trimeric form of the Severe Acute Respiratory Syndrome Coronavirus-2 spike protein bound to ACE2.

Model number	DOPE score
Model_1	−453061.63
Model_2	−452078.13
Model_3	−452595.94
Model_4	−453871.81*
Model_5	−453290.16
Model_6	−452068.91
Model_7	−453414.09
Model_8	−452842.60
Model_9	−452185.50
Model_10	−452439.75

\*The lowest energy.

the mutated spike protein-ACE2 complex, mutated spike protein, and ACE2, respectively. The MMPBSA.py script was used to calculate the free energies for the mutated spike protein, ACE2, and mutated spike protein-ACE2 complex.

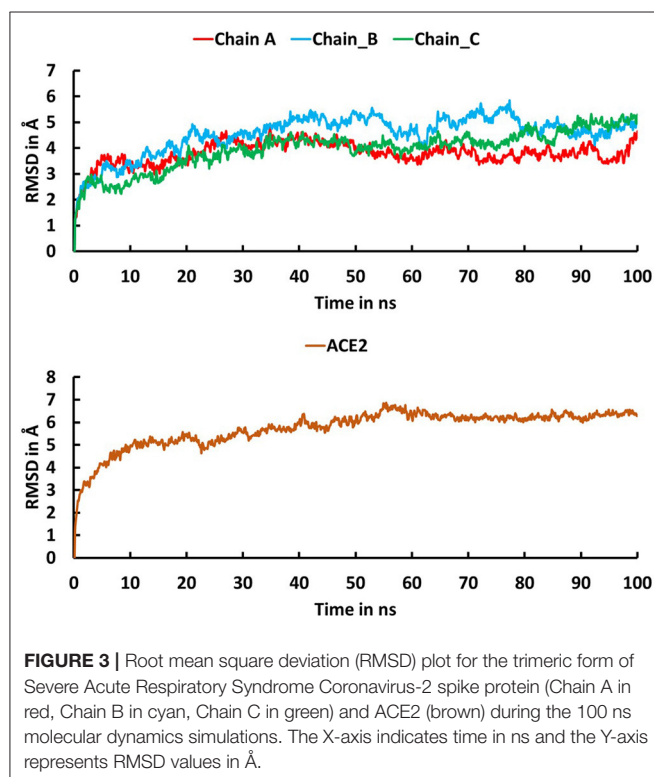
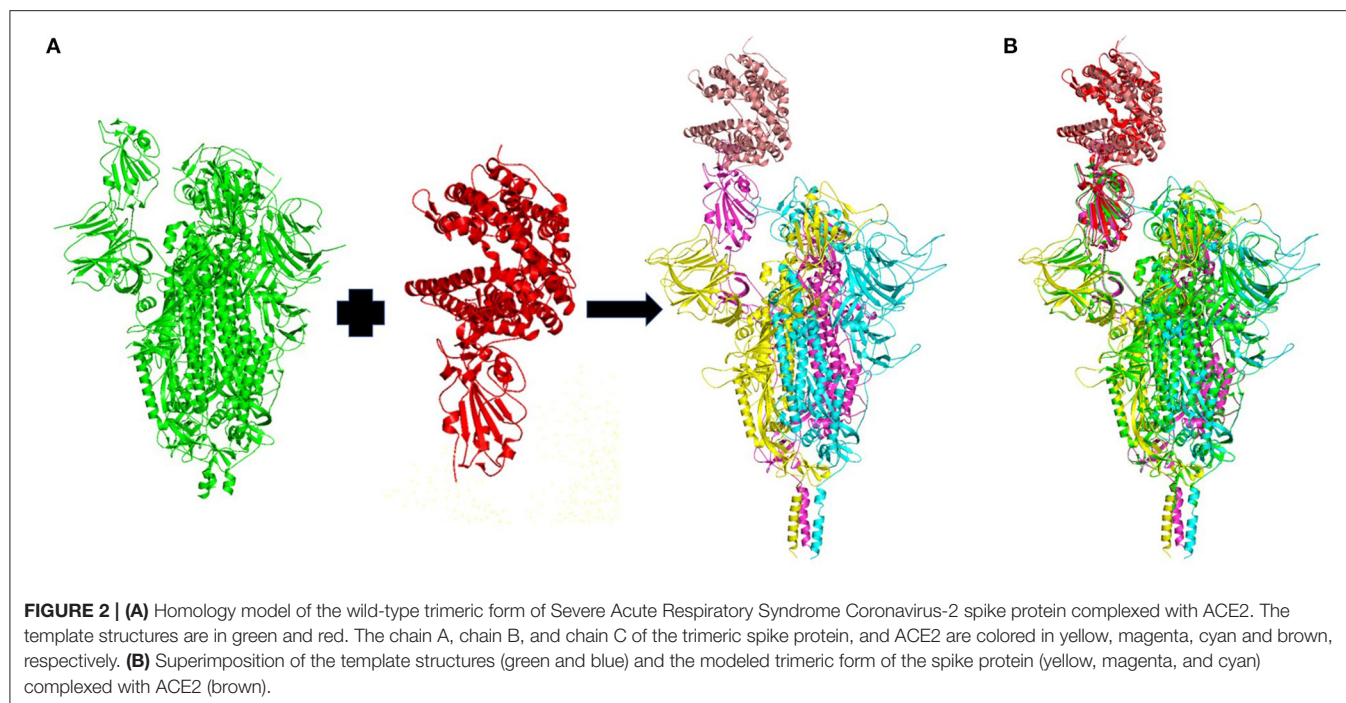
## RESULTS AND DISCUSSION

### Homology Modeling of the Trimeric Form of the Spike Protein

Ten homology models of the trimeric form of the spike protein bound to ACE2 were generated and their DOPE energy values were calculated (Table 1). Model\_4 showed the lowest DOPE energy value and was selected as the best structure for the trimeric form of spike protein complexed with ACE2. Model\_4 was evaluated for stereochemical chemical quality using a Ramachandran plot. According to the Ramachandran plot (Supplementary Figure 1), 93.57 and 1.53% of residues were in the favored and outlier regions, respectively. Chain A, chain B, and chain C showed 94, 93.39, and 92.95% residues in the favored regions, respectively. Superimpositions between Model\_4 and the reported structure of the trimeric form of the spike protein with the one-up conformation, and between Model\_4 and the crystal structure of the truncated spike protein RBD bound with ACE2 (Figure 2), showed RMSD values of 0.28 Å and 0.744 Å, respectively. This finding indicates that the homology model was consistent with the empirical structures. Therefore, Model\_4 was used for subsequent molecular dynamics simulations.

### Molecular Dynamics Simulations

The trajectory obtained from the molecular dynamics simulations was analyzed to identify critical residues at the interface between spike protein and ACE2. Stabilities of the trimeric form of the spike protein complexed with ACE2 and fluctuation of residues in the simulations were examined using root mean squared deviation (RMSD) of Cα atoms and root mean square fluctuation (RMSF), respectively. The RMSD plot (Figure 3) indicates that the trimeric form of wild-type spike protein was stable in the last 20 ns with a converged RMSD of



~5 Å. The averaged RMSD values for the chain A, chain B, and chain C were 3.36 Å, 4.51 Å, and 4.27 Å, respectively. ACE2 was stable in the last 40 ns with an average RMSD value of 5.74 Å. The high RMSD values are caused by the loop regions in the

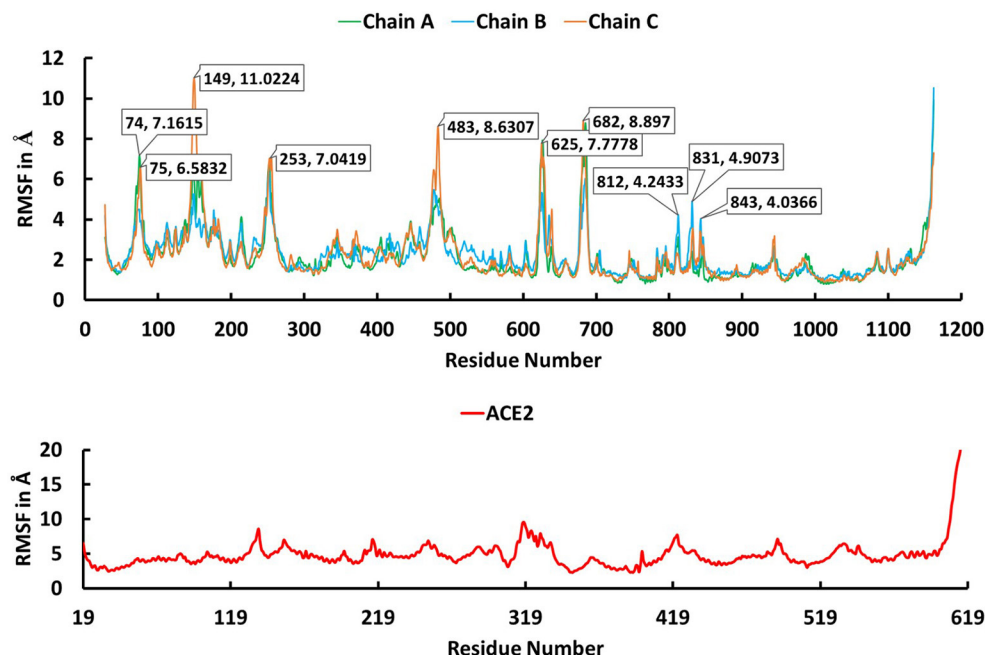
spike protein and ACE2. The RMSF plot (**Figure 4**) suggests that the residues in the secondary structure regions were quite stable, but the residues in the loop regions showed large fluctuations with RMSF values >3Å.

## Clustering Analysis

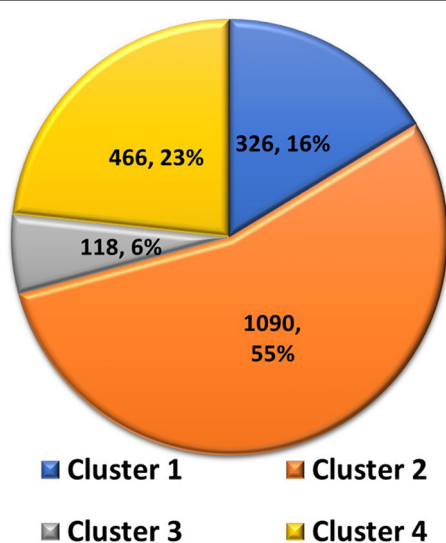
Clustering analysis was used to select a representative structure from the trajectory for binding free energy analysis. The kclust algorithm from MMTSB was used to cluster the structures in the trajectory. The kclust is a fast and sensitive algorithm and is widely used to cluster large size proteins. First, structures were extracted from the trajectory file with an interval of five frames and saved as PDB files for clustering analysis. The extracted structures were then put into the kclust algorithm with the radius set to 3 Å. Four clusters were obtained from kclust that resulted in 326, 1,090, 118, and 466 structures, as shown in **Figure 5**. As Cluster two was the largest, the structure with the smallest RMSD value in Cluster two exhibited the shortest distances to other structures and, thus, was selected as the representative structure for subsequent binding free energy analysis via alanine scanning.

## Interacting Residues

The representative structure from the cluster analysis was used to identify the residues between the trimeric spike protein and ACE2 at distances of 4 Å or less (Lan et al., 2020). Twenty residues of the trimeric spike protein were within 4 Å to ACE2 and were defined as interacting residues. Whereas, Lan et al. reported 17 spike protein RBD interacting residues within 4 Å to ACE2 when the truncated RBD monomer was analyzed. The superimposition of both complexes, the truncated spike protein RBD monomer complexed with ACE2 and the full-length trimeric spike protein complexed with ACE2, is depicted



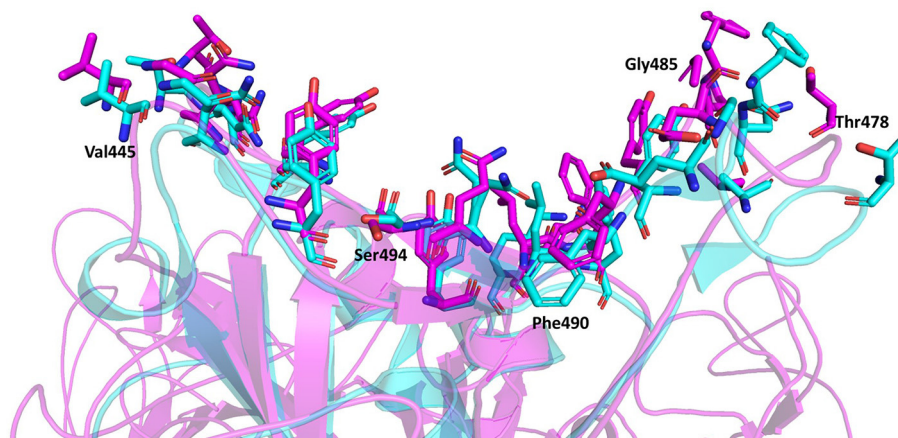
**FIGURE 4 |** The root mean square fluctuation (RMSF) plot for C $\alpha$  atoms in chain A (green), B (cyan), and C (brown) of the trimeric form of Severe Acute Respiratory Syndrome Coronavirus-2 spike protein and ACE2 (red) during the 100 ns molecular dynamics simulations. The X axis indicates residue number and Y axis represents RMSF value in Å.



**FIGURE 5 |** Pie chart of statistics of the 4 clusters of Severe Acute Respiratory Syndrome Coronavirus-2 spike protein complexed with ACE2. Number of structures and percentages are indicated in each pie segment.

in **Figure 6**. Both approaches to identify interacting residues, using the full-length trimer described here and the truncated RBD monomer (Lan et al., 2020), resulted in identification of 14 conserved residues with distances to ACE2 of  $\leq 4$  Å. Lan et al.

reported three additional interacting residues (Lys417, Tyr453, and Ala475) that were within 4 Å to ACE2 when using the truncated spike protein RBD monomer, but were slightly  $>4$  Å to ACE2 in assessments using the full-length trimeric spike protein. However, given that their distances to ACE2 are very close to 4 Å using the full-length trimeric spike protein, it is acceptable to consider all three as interacting residues based on both approaches. Conversely, Glu484 was slight  $>4$  Å from ACE2 using the truncated spike protein RBD monomer with a distance 4.39 Å but was  $<4$  Å to ACE2 using the full-length trimeric spike protein. Thus, it is also acceptable to include Glu484 as an interacting residue according to both approaches. Therefore, a total of 18 ACE2-interacting residues may be considered to be conserved between both approaches using either the full-length trimer or the truncated RBD monomer. However, five additional interacting residues (Val445, Thr478, Gly485, Phe490, and Ser494) with distances  $<4$  Å from ACE2 were identified using the full-length trimeric spike protein but were too far away to interact with ACE2 (5.69 Å, 7.62 Å, 5.72 Å, 5.34 Å, and 6.25 Å, respectively) using the truncated spike protein RBD monomer. Among these five new residues, four are in loop regions and the remaining one, Ser494, is at the end of a beta sheet in the spike protein RBD. Altogether, these data suggest that there are likely to be 23 spike protein residues capable of interacting with ACE2. These findings further suggest that the specific interacting residues may change based on the tertiary conformation, as indicated by the conformation difference between the full-length trimeric spike protein and its truncated RBD.



**FIGURE 6 |** Superimposition of the spike protein-ACE2 complexes using the full-length trimeric spike protein (Magenta) and the truncated spike protein RBD monomer (Cyan). The interacting residues are represented by stick model illustrations, while the rest of the ACE2 proteins are depicted in ribbon model form. The five new interacting residues identified using the full-length trimeric spike protein complexed with ACE2 are labeled.

**TABLE 2 |** The residues involved in the hydrogen bond interactions between the trimeric form of the Severe Acute Respiratory Syndrome Coronavirus-2 spike protein and ACE2 complex.

Residue in spike	Residue in ACE2	Distance Å
Tyr489*	Glu23	2.715
Gln493*	Leu29	3.033
Phe490	Asp30*	3.184
Thr500*	Leu351	2.938
Asn501*	Leu351	2.909
Gly502*	Gly352	2.834

\*indicates the residues reported in the crystal structure of receptor binding domain of spike protein complexed with ACE2.

## Hydrogen Bond Interactions

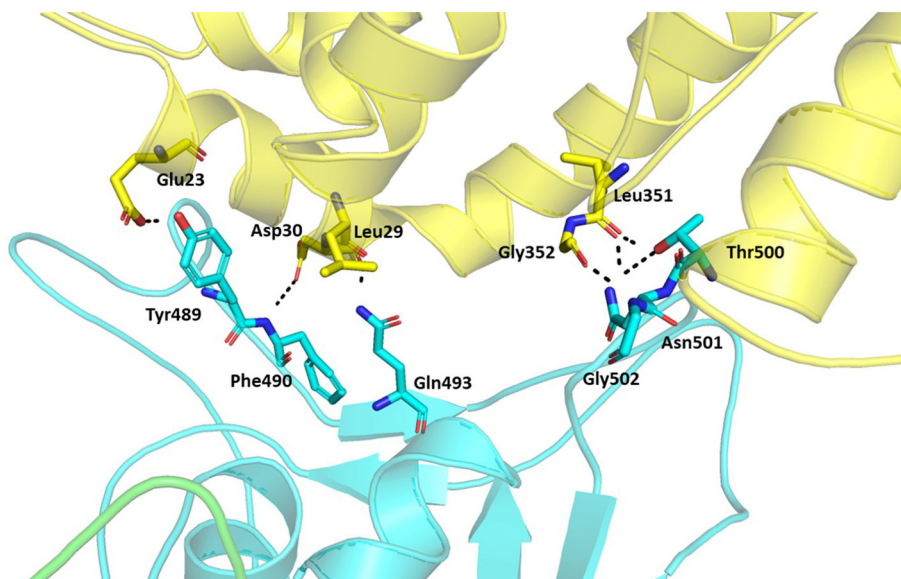
Hydrogen bonds are major interactions between proteins. Hence, hydrogen bonds formed between the spike protein and ACE2 in the representative structure from the molecular dynamics simulations were identified using CPPTRAJ from AmberTools. A hydrogen bond was formed if the distance between two hydrophilic atoms (O or N) near a hydrogen atom was  $<4$  Å and the angle from the hydrogen to the two hydrophilic atoms was  $>135^\circ$ . Using this criterion, hydrogen bonds were formed for six residues of the spike protein and five residues of ACE2 at the interface region (Table 2, Figure 7). Among these, five spike residues and one ACE2 residue were reported to form hydrogen bonds in the crystal structure of the truncated spike protein RBD complexed with ACE2 (Lan et al., 2020). Two residues that were reported to form hydrogen bonds in the crystal structure of the spike protein RBD bound with ACE2 (Lan et al., 2020), Leu455 and Phe456, did not form hydrogen bonds in our structure of the trimeric form of spike protein complexed with ACE2. On the other hand, Phe490 involved hydrogen bonding in our structure, but was not reported to form hydrogen bonds in the crystal structure of the truncated RBD of spike protein

monomer complexed with ACE2. Thus, these findings suggest that the comprehensive hydrogen bonding network between the spike protein and ACE2 is likely to depend on the trimer of full-length spike protein complexed with ACE2.

## Binding Free Energy Calculation

MMPBSA.py script from AmberTools is a fast method to compute binding free energy compared to other methods (Marimuthu et al., 2020), such as Replica-Exchange Free-Energy Perturbation (Fratesi and Sirimulla, 2019) and umbrella sampling (Kumar et al., 1992). MMPBSA.py script was used to calculate binding free energy values between the spike protein and ACE2 for the representative structure and its mutated structures from alanine screening. The estimated binding free energy between the spike protein and ACE2 was  $-60.54$  kcal/mol, indicating that the spike protein tightly binds with ACE2. The contributions from different energy terms (van der Waals, electrostatic, polar and non-polar solvation, solvation, and gas phase) to the free energy of the spike protein-ACE2 complex, the spike protein, and ACE2 were calculated and are shown in Table 3. Electrostatic interactions contributed more to the total binding free energy than van der Waals. To investigate contributions of individual residues to the binding between the spike protein and ACE2, decomposition of free energies to individual residues was conducted. The energy components (van der Waals, electrostatic, and polar and non-polar solvation energy) of the 20 interacting residues with distances  $<4$  Å from ACE2 are depicted in Table 4 and indicate that van der Waals interactions were the major force by which the interacting residues interacted with ACE2.

The 20 interacting residues with distances  $<4$  Å from ACE2 were used for alanine scanning. The current version of alanine scanning in MMPBSA script does not support energy calculation for the mutation of Gly to Ala. Hence, the four Gly residues were excluded from the alanine scanning calculation. Each of the remaining 16 interacting residues were mutated to alanine to generate a complex, then the free energy values were calculated



**FIGURE 7 |** Hydrogen bonds between the trimeric spike protein and ACE2. The residues involved in the hydrogen bond formation are shown in stick. The trimeric spike (Chain A–Green and Chain B–Cyan) and ACE2 (Yellow) are shown in ribbons.

**TABLE 3 |** MMPBSA energy values for the trimeric form of Severe Acute Respiratory Syndrome Coronavirus-2 spike protein bound with ACE2.

Energy component	Spike-ACE2 complex	Spike	ACE2	$\Delta G$ kCal/mol
vdW	–31291.7513	–26670.0359	–4534.5528	–87.1626
EEL	280175.1068	–238713.6906	–40760.4637	–700.9525
EGB	–43799.2943	–33677.4334	–10860.3462	738.4853
ESURF	1395.7146	1141.7286	264.8911	–10.9051
G gas	–36997.8641	–35045.8034	–1163.9455	–788.1152
G solv	–42403.5797	–32535.7048	–10595.4551	727.5802
Total	–79401.4438	–67581.5082	–11759.4006	–60.5350

vdW, van der Waals; EEL, Electrostatic energy; EGB, Polar solvation energy; ESURF, Non-polar solvation energy; G solv, Total solvation free energy; G gas, Total gas phase free energy.

for the complex using MMPBSA.py script. The binding free energy values of the wild-type and 16 mutated trimeric spike protein-ACE2 complexes were calculated and are reported in **Table 5**. Of the 16 alanine mutations, 15 resulted in structures with higher binding free energy than the representative wild-type structure, indicating that these 15 residues are likely to be important for spike protein binding to ACE2. The remaining one residue (Glu: 484) led to a structure with a slightly lower binding free energy, indicating that this residue is less important for the interaction between the spike protein and ACE2. Overall, the alanine screening analysis confirmed that the 20 interacting residues identified using the full-length trimeric spike protein play key roles in the binding interaction between the spike protein and ACE2.

Recently, 35,000 *de novo* hACE2 decoys were designed and CTC-445.2 was found to tightly bind with the RDB of the trimeric spike protein in four different states (state 1: 1 RBD up, state 2: 2 RBD up, state 3: 1 RBD up and 1 RBD down, and state 4: 2 RBD up and 1 RBD down) (Linsky et al., 2020) which were deposited in the electron microscopy databank and protein data

bank, but the structures are not yet available for the public. Eleven residues of the state four trimeric spike protein were determined to interface with CTC-442.2. We compared the 11 residues with the interacting residues identified in our structure and found nine that are interacting residues, Tyr489, Phe486, Gln493, Asn501, Thr500, Tyr449, Phe456, Asn487, and Gln498. The remaining two, Tyr495 and Ala475, do not meet the criterion of 4 Å for interacting residues, but are located in the interface between the trimeric spike protein and ACE2 in our structure, at distances of 4.13 Å and 4.28 Å to ACE2 residues, respectively. Therefore, all the interface residues were confirmed in our structure.

## CONCLUSIONS

The structure of the trimeric form of the full-length wild-type spike protein complexed with ACE2 was generated using homology modeling. The homology model was of good quality, as determined by Ramachandran plot evaluation. Using molecular dynamics simulations of the homology model, the

**TABLE 4 |** The 20 interacting residues from the decomposition analysis of Severe Acute Respiratory Syndrome Coronavirus-2 spike protein—ACE2 complex and their energy.

Residue	van der Waals	Electrostatic	Polar solvation	Non-polar solvation	Total
Val445	−0.311	−2.402	2.477	−0.113616	−0.349616
Gly446	−0.385	1.558	−1.172	−0.1263888	−0.1253888
Leu455	−0.971	−0.862	1.079	−0.1799856	−0.9339856
Thr478	−0.419	−0.473	0.702	−0.1375488	−0.3275488
Glu484	−0.822	100.71	−99.548	−0.2132928	0.1267072
Gly485	−1.044	0.466	0.272	−0.1362312	−0.4422312
Phe:486	−6.396	−3.798	4.073	−1.11034	−7.23134
Tyr:489	−4.958	−11.682	10.557	−0.55015	−6.63315
Gly496	−0.412	0.483	−0.12	−0.0960624	−0.1450624
Gln498	−1.528	2.017	−0.13	−0.29088	0.06812
Thr:500	−3.922	−4.18	4.737	−0.71958	−4.08458
Asn:501	−2.079	−9.038	7.484	−0.12496	−3.75796
Tyr:505	−3.503	−2.343	3.597	−0.57291	−2.82191
Gln:493	−2.628	−2.711	3.95	−0.52306	−1.91206
Gly:502	−0.508	−3.913	2.715	−0.19863	−1.90463
Phe:456	−1.989	−0.354	1.033	−0.3534	−1.6634
Tyr:449	−2.24	−2.808	4.02	−0.53238	−1.56038
Phe:490	−0.879	0.908	−1.492	−0.09493	−1.55793
Asn:487	−1.824	−4.005	4.825	−0.16806	−1.17206
Ser:494	−0.428	−0.841	0.254	−0.10336	−1.11836

**TABLE 5 |** Binding free energy values for the representative and mutated Severe Acute Respiratory Syndrome Coronavirus-2 spike protein-ACE2 complexes.

Residue mutated	Spike	ACE2	Spike-ACE2 complex	ΔG
Tyr:489:Ala	−67555.8	−11759.4	−79366.7	−51.48
Phe:486:Ala	−67585.2	−11759.4	−79396.3	−51.64
Tyr:505:Ala	−67567.4	−11759.4	−79382.4	−55.61
Gln:493:Ala	−67521.5	−11759.4	−79337.1	−56.21
Asn:501:Ala	−67505.7	−11759.4	−79321.5	−56.41
Thr:500:Ala	−67555	−11759.4	−79372	−57.53
Tyr:449:Ala	−67557.1	−11759.4	−79374.2	−57.74
Ser:494:Ala	−67557.1	−11759.4	−79374.2	−57.74
Phe:456:Ala	−67572.5	−11759.4	−79387.6	−57.76
Lys:455:Ala	−67567.5	−11759.4	−79385.9	−58.95
Thr:478:Ala	−67557.5	−11759.4	−79376.6	−59.65
Asn:487:Ala	−67502.3	−11759.4	−79321.3	−59.65
Gln:498:Ala	−67520.7	−11759.4	−79340.1	−60.04
Phe:490:Ala	−67588.4	−11759.4	−79408.2	−60.41
Val:445:Ala	−67572.2	−11759.4	−79391.8	−60.25
Glu:484:Ala	−67522.4	−11759.4	−79342.9	−61.09

ΔG: Binding free energy; ΔΔG: binding energy difference between the representative and mutated complex structure.

residues in the spike protein that are key for tightly binding ACE2 were identified. Most of the interacting residues reported in the crystal structure of the monomeric truncated spike protein RBD bound with ACE2 were among the interacting residues identified in this study with interaction distances defined as  $\leq 4$  Å, and all were included when considering those just slighter  $>4$  Å. In addition, one previously excluded residue and five new interacting residues were identified as likely to be important

contributors to ACE2 binding. Of the 16 interacting residues analyzed by alanine screening, 15 were confirmed to be important for the binding between the spike protein and ACE2. The binding interactions between the spike protein and ACE2 that were identified using the trimeric form of full-length wild-type spike protein are different from those reported in the crystal structure of the monomeric form of the spike protein RBD, indicating that the binding interface between the spike protein and ACE2

receptor is likely to be dependent on the tertiary structure of the spike protein used in the analysis. Together, the constructed structure of the trimeric full-length wild-type spike protein bound with ACE2 and the key binding residues identified in this study provide new insights into understanding mechanisms of SARS-CoV-2 infection of host cells, which could facilitate the development of drugs and vaccines to prevent SARS-CoV-2 infection and to combat COVID-19.

## DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

## AUTHOR CONTRIBUTIONS

SS, HH, MA, JH, and TP: conceived the experiment. SS and WG: conducted the experiment. SS, BP, GY, and ZJ: analyzed the result. SS, HH, MA, JH, and TP: wrote the manuscript. All authors contributed to the article and approved the submitted version.

## REFERENCES

- Astuti, I., and Ysrafil. (2020). Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2): an overview of viral structure and host response. *Diabetes Metab. Syndr.* 14, 407–412. doi: 10.1016/j.dsx.2020.04.020
- Darden, T., York, D., and Pedersen, L. (1993). Particle mesh ewald: an N-log(N) method for ewald sums in large systems. *J. Chem. Phys.* 98, 10089–10092. doi: 10.1063/1.464397
- Das, G., Ghosh, S., Garg, S., Ghosh, S., Jana, A., Samat, R., et al. (2020). An overview of key potential therapeutic strategies for combat in the COVID-19 battle. *RSC Adv.* 10, 28243–28266. doi: 10.1039/D0RA05434H
- Frater, F., and Sirimulla, S. (2019). An improved free energy perturbation FEP+ sampling protocol for flexible ligand-binding domains. *Sci. Rep.* 9:16829. doi: 10.1038/s41598-019-53133-1
- Henderson, R., Edwards, R. J., Mansouri, K., Janowska, K., Stalls, V., Gobeil, S. M. C., et al. (2020). Controlling the SARS-CoV-2 spike glycoprotein conformation. *Nat. Struct. Mol. Biol.* 27, 925–933. doi: 10.1038/s41594-020-0479-4
- Huang, C., Wang, Y., Li, X., Ren, L., Zhao, J., Hu, Y., et al. (2020). Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet* 395, 497–506. doi: 10.1016/S0140-6736(20)30183-5
- Humphrey, W., Dalke, A., and Schulten, K. (1996). VMD: visual molecular dynamics. *J. Mol. Graph.* 14, 33–38. doi: 10.1016/0263-7855(96)00018-5
- Jaimes, J. A., André, N. M., Chappie, J. S., Millet, J. K., and Whittaker, G. R. (2020). Phylogenetic analysis and structural modeling of SARS-CoV-2 spike protein reveals an evolutionary distinct and proteolytically sensitive activation loop. *J. Mol. Biol.* 432, 3309–3325. doi: 10.1016/j.jmb.2020.04.009
- Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. (1983). Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* 79, 926–935. doi: 10.1063/1.445869
- Kang, S., Yang, M., Hong, Z., Zhang, L., Huang, Z., Chen, X., et al. (2020). Crystal structure of SARS-CoV-2 nucleocapsid protein RNA binding domain reveals potential unique drug targeting sites. *Acta Pharm. Sin. B* 10, 1228–1238. doi: 10.1016/j.apsb.2020.04.009
- Kumar, S., Rosenberg, J. M., Bouzida, D., Swendsen, R. H., and Kollman, P. A. (1992). The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *J. Comput. Chem.* 13, 1011–1021. doi: 10.1002/jcc.540130812
- Lan, J., Ge, J., Yu, J., Shan, S., Zhou, H., Fan, S., et al. (2020). Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *Nature* 581, 215–220. doi: 10.1038/s41586-020-2180-5
- Linsky, T. W., Vergara, R., Codina, N., Nelson, J. W., Walker, M. J., Su, W., et al. (2020). *De novo* design of potent and resilient hACE2 decoys to neutralize SARS-CoV-2. *Science* 370, 1208–1214. doi: 10.1126/science.abe0075
- Madeira, F., Park, Y. M., Lee, J., Buso, N., Gur, T., Madhusoodanan, N., et al. (2019). The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Res.* 47, W636–W641. doi: 10.1093/nar/gkz268
- Maier, J. A., Martinez, C., Kasavajhala, K., Wickstrom, L., Hauser, K. E., and Simmerling, C. (2015). ff14SB: improving the accuracy of protein side chain and backbone parameters from ff99SB. *J. Chem. Theory Comput.* 11, 3696–3713. doi: 10.1021/acs.jctc.5b00255
- Malik, Y. A. (2020). Properties of coronavirus and SARS-CoV-2. *Malays. J. Pathol.* 42, 3–11.
- Marimuthu, P., Razzokov, J., Singaravelu, K., and Bogaerts, A. (2020). Predicted hotspot residues involved in allosteric signal transmission in pro-apoptotic peptide-Mcl1 complexes. *Biomolecules* 10:1114. doi: 10.3390/biom10081114
- Miller, B. R. 3rd, McGee, T. D. Jr, Swails, J. M., Homeyer, N., Gohlke, H., and Roitberg, A. E. (2012). MMPBSA.py: an efficient program for end-state free energy calculations. *J. Chem. Theory Comput.* 8, 3314–3321. doi: 10.1021/ct300418h
- Ortega, J. T., Serrano, M. L., Pujol, F. H., and Rangel, H. R. (2020). Role of changes in SARS-CoV-2 spike protein in the interaction with the human ACE2 receptor: an *in silico* analysis. *EXCLI J.* 19, 410–417. doi: 10.17179/excli2020-1167
- Petrosillo, N., Viceconte, G., Ergonul, O., Ippolito, G., and Petersen, E. (2020). COVID-19, SARS and MERS: are they closely related? *Clin. Microbiol. Infect.* 26, 729–734. doi: 10.1016/j.cmi.2020.03.026
- Ponder, J. W., and Case, D. A. (2003). Force fields for protein simulations. *Adv. Protein Chem.* 66, 27–85. doi: 10.1016/S0065-3233(03)66002-X
- Rabaan, A. A., Al-Ahmed, S. H., Haque, S., Sah, R., Tiwari, R., Malik, Y. S., et al. (2020). SARS-CoV-2, SARS-CoV, and MERS-CoV: a comparative overview. *Infez. Med.* 28, 174–184. doi: 10.1016/j.scitotenv.2020.138862
- Roe, D. R., and Cheatham, T. E. 3rd (2013). PTRAJ and CPPTRAJ: software for processing and analysis of molecular dynamics trajectory data. *J. Chem. Theory Comput.* 9, 3084–3095. doi: 10.1021/ct400341p
- Romano, M., Ruggiero, A., Squeglia, F., Maga, G., and Berisio, R. (2020). A structural view of SARS-CoV-2 RNA replication machinery: RNA synthesis, proofreading and final capping. *Cells* 9:1267. doi: 10.3390/cells9051267
- Ryckaert, J.-P., Ciccotti, G., and Berendsen, H. J. C. (1977). Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys.* 23, 327–341. doi: 10.1016/0021-9991(77)90098-5

## ACKNOWLEDGMENTS

The authors thank Dr. Mike Mikailov from the Center for Devices and Radiological Health (CDRH) for his technical assistance to run the calculations in FDA's HPC. We thank Dr. William Slikker, the Director of the National Center for Toxicological Research, for his review and constructive comments on this manuscript, and Gwenn Merry for her outstanding editing. This research was supported in part by an appointment to the Research Participation Program at the National Center for Toxicological Research (WG, ZJ) administered by the Oak Ridge Institute for Science and Education through an interagency agreement between the U.S. Department of Energy and the U.S. Food and Drug Administration.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fchem.2020.622632/full#supplementary-material>

- Sali, A., and Blundell, T. L. (1993). Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* 234, 779–815. doi: 10.1006/jmbi.1993.1626
- Shang, J., Wan, Y., Luo, C., Ye, G., Geng, Q., Auerbach, A., et al. (2020). Cell entry mechanisms of SARS-CoV-2. *Proc. Natl. Acad. Sci. U.S.A.* 117, 11727–11734. doi: 10.1073/pnas.2003138117
- Song, W., Gui, M., Wang, X., and Xiang, Y. (2018). Cryo-EM structure of the SARS coronavirus spike glycoprotein in complex with its host cell receptor ACE2. *PLoS Pathog.* 14:e1007236. doi: 10.1371/journal.ppat.1007236
- Tang, T., Bidon, M., Jaimes, J. A., Whittaker, G. R., and Daniel, S. (2020a). Coronavirus membrane fusion mechanism offers a potential target for antiviral development. *Antiviral Res.* 178:104792. doi: 10.1016/j.antiviral.2020.104792
- Tang, X., Wu, C., Li, X., Song, Y., Yao, X., Wu, X., et al. (2020b). On the origin and continuing evolution of SARS-CoV-2. *Natl. Sci. Rev.* 7, 1012–1023. doi: 10.1093/nsr/nwaa036
- Walls, A. C., Park, Y.-J., Tortorici, M. A., Wall, A., McGuire, A. T., and Veesler, D. (2020). Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein. *Cell* 181, 281–292.e286. doi: 10.1016/j.cell.2020.02.058
- Wrapp, D., Wang, N., Corbett, K. S., Goldsmith, J. A., Hsieh, C.-L., Abiona, O., et al. (2020). Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science* 367, 1260–1263. doi: 10.1126/science.abb2507
- Wu, C., Liu, Y., Yang, Y., Zhang, P., Zhong, W., Wang, Y., et al. (2020). Analysis of therapeutic targets for SARS-CoV-2 and discovery of potential drugs by computational methods. *Acta Pharm. Sin. B* 10, 766–788. doi: 10.1016/j.apsb.2020.02.008
- Xia, S., Liu, M., Wang, C., Xu, W., Lan, Q., Feng, S., et al. (2020). Inhibition of SARS-CoV-2 (previously 2019-nCoV) infection by a highly potent pan-coronavirus fusion inhibitor targeting its spike protein that harbors a high capacity to mediate membrane fusion. *Cell Res.* 30, 343–355. doi: 10.1038/s41422-020-0305-x
- Zeng, W., Liu, G., Ma, H., Zhao, D., Yang, Y., Liu, M., et al. (2020). Biochemical characterization of SARS-CoV-2 nucleocapsid protein. *Biochem. Biophys. Res. Commun.* 527, 618–623. doi: 10.1016/j.bbrc.2020.04.136
- Zhang, L., and Liu, Y. (2020). Potential interventions for novel coronavirus in China: a systematic review. *J. Med. Virol.* 92, 479–490. doi: 10.1002/jmv.25707

**Disclaimer:** The views presented in this article do not necessarily reflect those of the US Food and Drug Administration.

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Sakkiah, Guo, Pan, Ji, Yavas, Azevedo, Hawes, Patterson and Hong. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Structural Insight Into the SARS-CoV-2 Nucleocapsid Protein C-Terminal Domain Reveals a Novel Recognition Mechanism for Viral Transcriptional Regulatory Sequences

Mei Yang<sup>1,2</sup>, Suhua He<sup>1</sup>, Xiaoxue Chen<sup>1</sup>, Zhaoxia Huang<sup>1</sup>, Ziliang Zhou<sup>1</sup>, Zhechong Zhou<sup>1</sup>, Qiuyue Chen<sup>1</sup>, Shoudeng Chen<sup>1,2\*</sup> and Sisi Kang<sup>1,2\*</sup>

<sup>1</sup> Guangdong Provincial Key Laboratory of Biomedical Imaging, Molecular Imaging Center, The Fifth Affiliated Hospital, Sun Yat-sen University, Zhuhai, China, <sup>2</sup> Molecular Imaging Center, The Fifth Affiliated Hospital, Sun Yat-sen University, Zhuhai, China

## OPEN ACCESS

### Edited by:

Emilia Pedone,  
Institute of Biostructure and  
Bioimaging, Italy

### Reviewed by:

Dayi Zhang,  
Tsinghua University, China  
Wai-Lung Ng,  
The Chinese University of  
Hong Kong, China

### \*Correspondence:

Shoudeng Chen  
chenshd5@mail.sysu.edu.cn  
Sisi Kang  
kangss@mail2.sysu.edu.cn

### Specialty section:

This article was submitted to  
Chemical Biology,  
a section of the journal  
Frontiers in Chemistry

Received: 01 November 2020

Accepted: 08 December 2020

Published: 12 January 2021

### Citation:

Yang M, He S, Chen X, Huang Z, Zhou Z, Zhou Z, Chen Q, Chen S and Kang S (2021) Structural Insight Into the SARS-CoV-2 Nucleocapsid Protein C-Terminal Domain Reveals a Novel Recognition Mechanism for Viral Transcriptional Regulatory Sequences. *Front. Chem.* 8:624765. doi: 10.3389/fchem.2020.624765

Coronavirus disease 2019 (COVID-19) has caused massive disruptions to society and the economy, and the transcriptional regulatory mechanisms behind the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) are poorly understood. Herein, we determined the crystal structure of the SARS-CoV-2 nucleocapsid protein C-terminal domain (CTD) at a resolution of 2.0 Å, and demonstrated that the CTD has a comparable distinct electrostatic potential surface to equivalent domains of other reported CoVs, suggesting that the CTD has novel roles in viral RNA binding and transcriptional regulation. Further *in vitro* biochemical assays demonstrated that the viral genomic intergenic transcriptional regulatory sequences (TRSs) interact with the SARS-CoV-2 nucleocapsid protein CTD with a flanking region. The unpaired adeno dinucleotide in the TRS stem-loop structure is a major determining factor for their interactions. Taken together, these results suggested that the nucleocapsid protein CTD is responsible for the discontinuous viral transcription mechanism by recognizing the different patterns of viral TRS during transcription.

**Keywords:** COVID-19, coronavirus, SARS-CoV-2, nucleocapsid protein, C terminal domain, crystal structure, transcription regulating sequences

## INTRODUCTION

The coronavirus disease 2019 (COVID-19) pandemic, caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), has led to a total of 45,140,131 confirmed cases and 1,182,747 deaths across 216 countries and regions as of October 31, 2020 (World Health Organization, <https://covid19.who.int>). Despite remarkable efforts to study the pathological roles of the SARS-CoV-2 virus, there are still many mysteries about the life cycle of SARS-CoV-2.

Similar to other pathogenic betacoronaviruses (order *Nidovirales*, family *Coronaviridae*, subfamily *Coronavirinae*), SARS-CoV-2 was discovered as an enveloped, positive-sense, single-stranded RNA virus ~30 kilobase in genomic size (Wu et al., 2020; Zhou et al., 2020).

Coronaviruses (CoVs) have the largest genomes among the RNA virus families and a conserved 5' leader sequence (Lai and Stohlman, 1981; Sola et al., 2015). In the viral life cycle, the positive-sense RNA genome is replicated and transcribed by the viral RNA-dependent RNA polymerase (RdRp) protein (Sola et al., 2015; Snijder et al., 2016). The replication of the genome requires continuous RNA synthesis since a full-length complementary negative-strand (–) RNA is used as the template for the production of genomic RNA (gRNA) copies. In contrast, CoV transcription requires a unique discontinuous synthesis of (–) subgenomic RNA (sgRNA). The RdRp complex utilizes the template switching mechanism of the nascent (–) RNA fused with the genomic 5' leader sequence to generate a nested set of subgenomic mRNAs (sgmRNAs) that are identical to the 5' and 3' termini of the viral genome (Pasternak et al., 2006; Sawicki et al., 2007). The discontinuous transcription process is controlled by a conserved transcriptional regulatory sequence (TRS), which is located after the conserved 5' leader sequence (termed leader TRS, TRS-L) and in front of each ORF gene (termed body TRS, TRS-B). A prevailing model suggests that base pairing between the TRS-L and the complementary TRS-B occurs during (–) strand RNA synthesis (Sola et al., 2015). The pairing leads to template-switching events that generate discontinuous (–) strand RNAs, which serve as templates for the transcription of large amounts of discontinuous nested (+) strand sgmRNAs. These sgmRNAs encode conserved structural proteins (spike (S), envelope (E), membrane (M), and nucleocapsid proteins) and several accessory proteins (Sola et al., 2015; Wu et al., 2020). The SARS-CoV-2 nucleocapsid protein consists of five domains: an N-terminal tail region (residues 1 to 40), an N-terminal RNA binding domain (residues 41 to 173, termed N-NTD), a Ser/Arg-rich linker region (residues 174 to 249, termed LKR), a C-terminal dimerization domain (residues 250 to 364, termed N-CTD), and a C-terminal intrinsically disordered region (residues 365 to 419, termed IDR) (**Supplementary Figure 1A**).

According to a recent transcriptome study, Vero cells infected with SARS-CoV-2 produce 92.6% canonical full-length gRNA and nine sgRNAs, as well as 7.4% other non-canonical transcripts because of numerous discontinuous transcription events (Kim et al., 2020). Among the top transcribed sgRNAs, the sgRNA-encoded nucleocapsid is the most abundant transcript. Of note, a previous study suggested that the coronavirus nucleocapsid protein participates in the discontinuous transcription process of sgRNAs since depletion of the nucleocapsid-encoded region from the replicon reduces the synthesis of sgmRNAs but not gRNAs (Zúñiga et al., 2010). Importantly, the phosphorylation of the SARS-CoV nucleocapsid results in the recruitment of the RNA helicase DDX1 and then enables the transcription in the transition from a discontinuous process to a continuous process (Wu et al., 2014). The SARS-CoV-2 nucleocapsid protein is a multifunctional protein with potential primary functions of binding to the viral RNA genome and packing it into a long helical nucleocapsid structure or RNP complex (Masters and Sturman, 1990; McBride et al., 2014). Recent systematic proteomic results indicate that SARS-CoV-2 nucleocapsid proteins expressed in HEK293T/17 cells associate with host mRNA binding proteins and stress granule proteins using

affinity purification mass spectrometry (AP-MS) with little bound RNA information (Gordon et al., 2020). Studies have investigated several cis-regulating elements and trans-regulating factors involved in discontinuous transcription processes (Sola et al., 2011), but the molecular mechanisms of the nucleocapsid protein involved in this process in SARS-CoV-2 remain unclear.

In previous work, we solved the crystal structure of the nucleocapsid N-terminal RNA binding domain, suggesting a potential leading compound binding pocket in antiviral agent screening applications (Kang et al., 2020). Herein, we continued our work on the structural studies of the nucleocapsid protein by solving the crystal structure of the nucleocapsid protein C-terminal domain (termed as N-CTD) at a resolution of 2.0 Å. By combining structural comparisons and *in vitro* interacting assays, we sought to investigate the potential molecular mechanisms of interplay between the N-CTD and the conserved SARS-CoV-2 TRS and to provide detailed insight into the function of the nucleocapsid protein in discontinuous transcription.

## MATERIALS AND METHODS

### Cloning, Expression, and Purification

The SARS-CoV-2 N-FL plasmid was a gift from Guangdong Medical Laboratory Animal Center. We designed several constructs, including the SARS-CoV-2 N-CTD domain (residues 250 to 365), SARS-CoV-2 CTD+IDR domain (residues 250 to 419), and SARS-CoV-2 LKR+CTD domain (residues 183 to 365), which were designed using secondary structure predictions and sequence conservation characteristics. The above gene fragments were obtained by PCR. These sequences were then cloned into the pRSF-Duet-1 vector with an N-terminal 6xHis-SUMO tag using BamHI and XhoI. All the constructs were expressed in the *E. coli* Rosetta strain. When the OD<sub>600</sub> of the culture reached 0.8–1.0, IPTG (final concentration of 0.1 mM) was used to induce the expression for 18–20 h at 16°C. The culture was collected by a Beckman high-speed centrifuge at 4,500 rpm for 15 min and disrupted by ultrahigh pressure treatment, and the supernatant was separated by centrifugation at 18,000 rpm for 90 min. After nickel column chromatography, followed by Ulp1 protease digestion for tag removal, the SARS-CoV-2 N-CTD proteins were further purified with size-exclusion chromatography (with a buffer consisting of 20 mM Tris-HCl (pH 8.0), 150 mM sodium chloride, and 1 mM dithiothreitol) and then concentrated by ultrafiltration to a final concentration of 45, 0.965, and 4.7 mg/mL. The SARS-CoV-2 N-CTD-S327C and SARS-CoV-2 N-CTD-S289C mutants were constructed using designed primers and PCR and purified with the same protocol as the wild-type N-CTD protein. Details of all oligonucleotide sequences are available from the authors upon request.

### Crystallization and Data Collection

Crystals were grown from a solution containing 100 mM CHES (pH = 9.3) (Hampton research: HR2-256) and 40% PEG6000 (Sigma-Aldrich) by the hanging drop vapor diffusion method at 16°C. Crystals were frozen in liquid nitrogen in reservoir solutions supplemented without a cryoprotectant. X-ray diffraction data were collected at the South China Sea

Institute of Oceanology, Chinese Academy of Sciences with the Rigaku X-ray diffraction (XRD) instrument XtaLAB P200 007HF. The structure of SARS-CoV-2 N-CTD was determined by molecular replacement using the SARS-CoV N-CTD structure (PDB ID: 2GIB) as the search model (Saikatendu et al., 2007) with the PHENIX software suite. The X-ray diffraction and structure refinement statistics are summarized in **Table 1**.

## Biolayer Interferometry Assays

Biolayer interferometry (BLI) experiments were performed using an Octet RED96e instrument from ForteBio. All assays were run at 25°C with continuous shaking at 100 rpm. PBS with 0.02% Tween 20 was used as the assay buffer. For RNA binding assays, we designed and synthesized six RNA oligomers with following sequences: RNA1, 5' biotin-ACGAAC-3'; RNA2, 5' biotin-AAACGAAC-3'; RNA3, 5' biotin-AAACGAACUU-3'; RNA4, 5' biotin-GUUCGU-3'; RNA5, 5' biotin-GUUCGUUU-3'; and RNA6, 5' biotin-AAGUUCGUUU-3'. All RNA oligomers are dissolved in enzyme-free water. The paired or impaired duplexes of TRSs were obtained with a ratio of 1:1 via the process of heating and annealing. Duplex TRS-1 was obtained by an annealing process of RNA1 with RNA4; Duplex TRS-2 was obtained by an annealing process of RNA2 with RNA5. Duplex TRS-3 was obtained by an annealing process of RNA3 with RNA6. Unpaired Duplex TRS1 was obtained by an annealing process of RNA1 with RNA5; Unpaired Duplex TRS2 was obtained by an annealing process of RNA2 with RNA4. Biotinylated RNA was tethered on super streptavidin (SSA) biosensors (ForteBio) by dipping sensors into 100 nmol RNA solution. Average saturation response levels of 0.8 nm were achieved in 1 min for all samples. Sensors with tethered RNA were washed in assay buffer for 10 min to eliminate non-specifically bound protein molecules and to establish stable baselines before starting association-dissociation cycles with different concentrations of CTD proteins. The collected raw kinetic data were processed in the data analysis software provided by the manufacturer using double reference subtraction in which both the 0.02% Tween-20-only reference and the inactive reference were subtracted. The resulting data were analyzed based on a 1:1 binding model from which  $K_{on}$  and  $K_{off}$  values were obtained, and then the  $K_d$  values were calculated.

## RESULTS

### The Crystal Structure of the SARS-CoV-2 Nucleocapsid C-Terminal Domain Reveals a Conserved Stable Dimer Formation Mechanism

To determine the precise structural information of the novel coronavirus SARS-CoV-2 N-CTD, we solved the N-CTD structure at a resolution of 2.0 Å with X-ray crystallography. The final structure was refined to R-factor and R-free values of 0.1744 and 0.2221, respectively. The complete statistics for the data collection, phasing, and refinement are presented in **Table 1**. One N-CTD monomer is composed of three  $3_{10}$ -helices, five  $\alpha$ -helices, and two  $\beta$ -strands (**Figure 1A**), with an additional N-terminal  $\alpha_0$ -helix in the partial electron density of the traced

**TABLE 1 |** Data collection and refinement statistics.

Item	SARS-CoV-2 N-CTD*
Protein Data Bank ID	7DE1
Wavelength (Å)	1.5418
Resolution range	18.48–2.0 (2.071–2.0)
Space group	<i>P</i> 2 <sub>1</sub> 2 <sub>1</sub> 2
Unit cell	
a, b, c (Å)	45.214, 101.438, 59.5261
$\alpha$ , $\beta$ , $\gamma$ (°)	90, 90, 90
Total reflections	67,575 (6134)
Unique reflections	19,027 (1805)
Multiplicity	3.6 (3.4)
Completeness (%)	99.27 (96.27)
Mean I/sigma (I)	18.58 (6.46)
Wilson B-factor	15.13
R-merge**	0.0588 (0.2126)
R-meas***	0.06906 (0.2513)
R-pim****	0.03568 (0.132)
CC1/2	0.997 (0.921)
CC	0.999 (0.979)
Reflections used in refinement	19,026 (1806)
Reflections used for R-free	1,903 (180)
R-work#	0.1744 (0.1961)
R-free##	0.2221 (0.2925)
CC (work)	0.954 (0.893)
CC (free)	0.918 (0.842)
Number of non-hydrogen atoms	2,031
Macromolecules	1,775
Solvent	242
Protein residues	223
RMS (bonds) (Å)	0.010
RMS (angles) (°)	1.01
Ramachandran favored (%)	98.17
Ramachandran allowed (%)	1.83
Ramachandran outliers (%)	0.00
Rotamer outliers (%)	0.54
Clashscore	8.75
Average B-factor	16.39
Macromolecules	15.33
Solvent	23.57

\*Statistics for the highest-resolution shell are shown in parentheses.

\*\* $R_{merge} = \sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} \sum_i I_i(hkl)$ , where  $I_i(hkl)$  is the intensity measured for the  $i$ th reflection and  $\langle I(hkl) \rangle$  is the average intensity of all reflections with indices  $hkl$ .

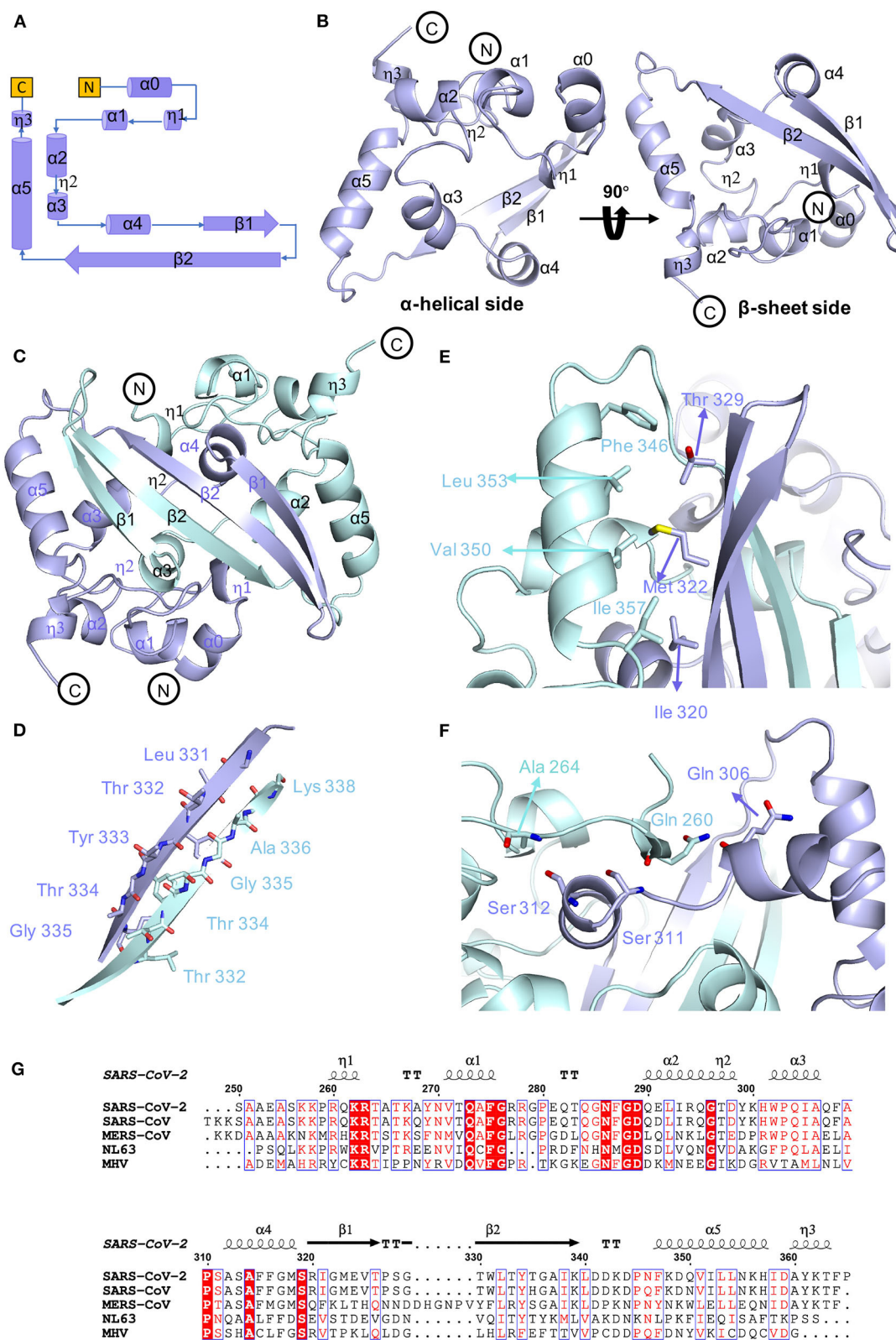
\*\*\* $R_{meas}$ , redundancy-independent (multiplicity-weighted)  $R_{merge}$  (Evans, 2006, 2011).

\*\*\*\* $R_{pim}$ , precision-indicating (multiplicity-weighted)  $R_{merge}$  (Diederichs and Karplus, 1997; Weiss, 2001).

# $R_{work} = \sum_{hkl} |F_{obs}(hkl) - |F_{calc}(hkl)|| / \sum_{hkl} |F_{obs}(hkl)|$ .

## $R_{free}$  is calculated in an identical manner using 10% of randomly selected reflections that were not included in the refinement.

molecule. And SARS-CoV-2 N-CTD consists of two “C-shaped” monomers to form a stable dimer (**Figures 1A,B**). As shown in **Figure 1C**, two N-CTD monomers utilize three regions to form a stable symmetrical dimer, with a buried surface area



**FIGURE 1 |** Conserved stable dimer of the C-terminal domain of the SARS-CoV-2 nucleocapsid protein. **(A)** Topological style illustration of the SARS-CoV-2 N-CTD structure. **(B)** Monomer structure of SARS-CoV-2 N-CTD. **(C)** The interaction between SARS-CoV-2 N-CTD monomers. **(D)** The hydrogen bond interactions between SARS-CoV-2 N-CTD monomer B chains. **(E,F)** The hydrophobic interactions between SARS-CoV-2 N-CTD molecules. **(G)** Sequence alignment of SARS-CoV-2 N-CTD, SARS-CoV N-CTD (UniProtKB: P59595), MERS-CoV N-CTD (UniProtKB: R9UM87), HCoV-NL63 N-CTD (UniProtKB: P33469), and MHV N-CTD (UniProtKB: NP\_040838.1). Red arrows indicate conserved residues for ribonucleotide binding sites, and dashed boxes indicate variable residues in the structural comparisons.

of 2618 Å<sup>2</sup> (of the 8026 Å<sup>2</sup> monomer surface) (calculated by the online PISA server). First, the most distinctive feature is the antiparallel four-stranded  $\beta$ -sheet that has domain swapping interactions between the two monomers. Within the  $\beta$ -sheet,  $\beta$ 2 strands of monomers interact with each other via a wide range of stable hydrogen bonds (Figure 1D). Second, residues F346, L353, V350, and I357 of the longest  $\alpha$ 5-helix form multiple intermolecular hydrophobic interactions with residues T329, M322, and I320 in the  $\beta$ 1-strand (Figure 1E). Last, the  $\alpha$ 1 helix forms a strong hydrophobic interaction with other monomer  $\alpha$ 4 helices, like two clip-bars, to firmly fix the two monomers (i.e., A264-S312, Q260-S311, and Q260-Q306 interactions, as shown in Figure 1F). Further gel filtration results also support the dimerization of N-CTD in solution (Supplementary Figure 1B). Therefore, the stable dimerization of SARS-CoV-2 N-CTD shows a conserved stable dimerization status in both crystal packing and solution conditions.

As shown in the sequence alignments of betacoronaviruses, the amino acid sequences of the SARS-CoV-2 N-CTD and the counterpart proteins of the highly pathogenic SARS-CoV and MERS-CoV and the low pathogenic HCoV-NL63 were quite different, with sequence identities of 89.74, 48.59, and 35.71% (Figure 1G), respectively. However, the overall structure of SARS-CoV-2 N-CTD is similar to the N-CTD of previously reported coronaviruses (including SARS-CoV; Yu et al., 2006; Chen et al., 2007; Takeda et al., 2008; HCoV-NL63; Szelazek et al., 2017; MERS-CoV; Nguyen et al., 2019; mouse hepatitis virus (MHV); Ma et al., 2010, and infectious bronchitis virus (IBV); Jayaram et al., 2006).

## Potential Self-Interactions of the SARS-CoV-2 N-CTD Dimer

To investigate the potential self-interactions of SARS-CoV-2 N-CTD, we next analyzed the symmetry molecules of crystal packing. As shown in Figure 2A, SARS-CoV-2 N-CTDs form repeating cylindrical high-order structures with six dimers, which is slightly different from the SARS-CoV N-CTD octamer X-shaped high-order oligomer pattern. There are three key features of the SARS-CoV-2 N-CTD's potential self-interactions in our study (Figure 2A). To validate these potential self-interaction features in solution, we performed *in vitro* disulfide trapping assays by engineering single-site cysteine mutations at the feature I, II, and III regions. Since wild-type SARS-CoV-2 N-CTD does not contain any cysteine residues, the exotic cysteine residues will form disulfide bonds within suitable distances (Bass et al., 2007). The first remarkable feature is that residues P326 and T329 in the  $\beta$ 5– $\beta$ 6 loop interact with symmetric molecules in the same position, forming a hand-in-hand-like structure (Figure 2B). The second loop is another  $\beta$ 5– $\beta$ 6 loop of the same N-CTD dimer that interacts with the  $\alpha$ 2-helix and C-terminal tail of the other side of the symmetric molecule via salt bridges or hydrogen bonds (Figure 2C). As shown in Figures 2B,C, S327 is located at a favorable position of the  $\beta$ 5– $\beta$ 6 loop, which mediates self-interactions within the feature I and II regions. The last feature is a salt bridge interaction between the side chains of Q289 and R294 (Figure 2D). Q289 and R294 are located at the  $\alpha$ 1-helix

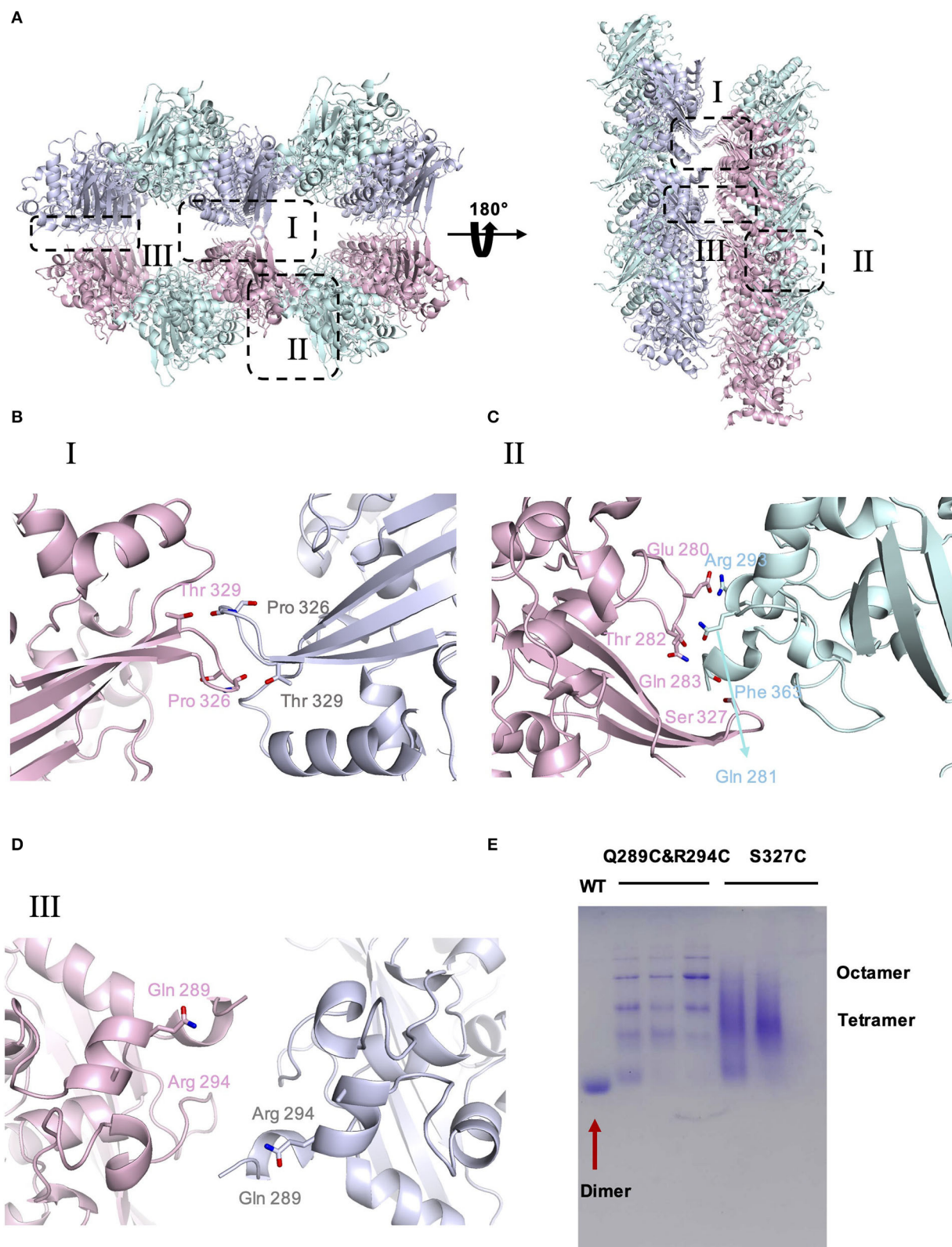
region, which mediates self-interactions within the feature III region. Therefore, the S327C single mutation and the Q289C with R294C double mutation were expressed and tested through size-exclusion chromatography. As shown in Figure 2E, compared with the wild-type SARS-CoV-2 N-CTD, the positions of the S327C and Q289C/R294C mutant proteins shifted up to varying degrees, and these mutants are in a higher polymerization state than the dimer. These results suggest that SARS-CoV-2 N-CTD has potential self-interactions via the  $\beta$ 5– $\beta$ 6 loop and  $\alpha$ 1-helix regions. In summary, the  $\beta$ 5– $\beta$ 6 loop and  $\alpha$ 1-helix regions are important for the self-association of SARS-CoV-2 N-CTD in the crystal packing.

## Surface Electrostatic Potential Characteristics of SARS-CoV-2 N-CTD

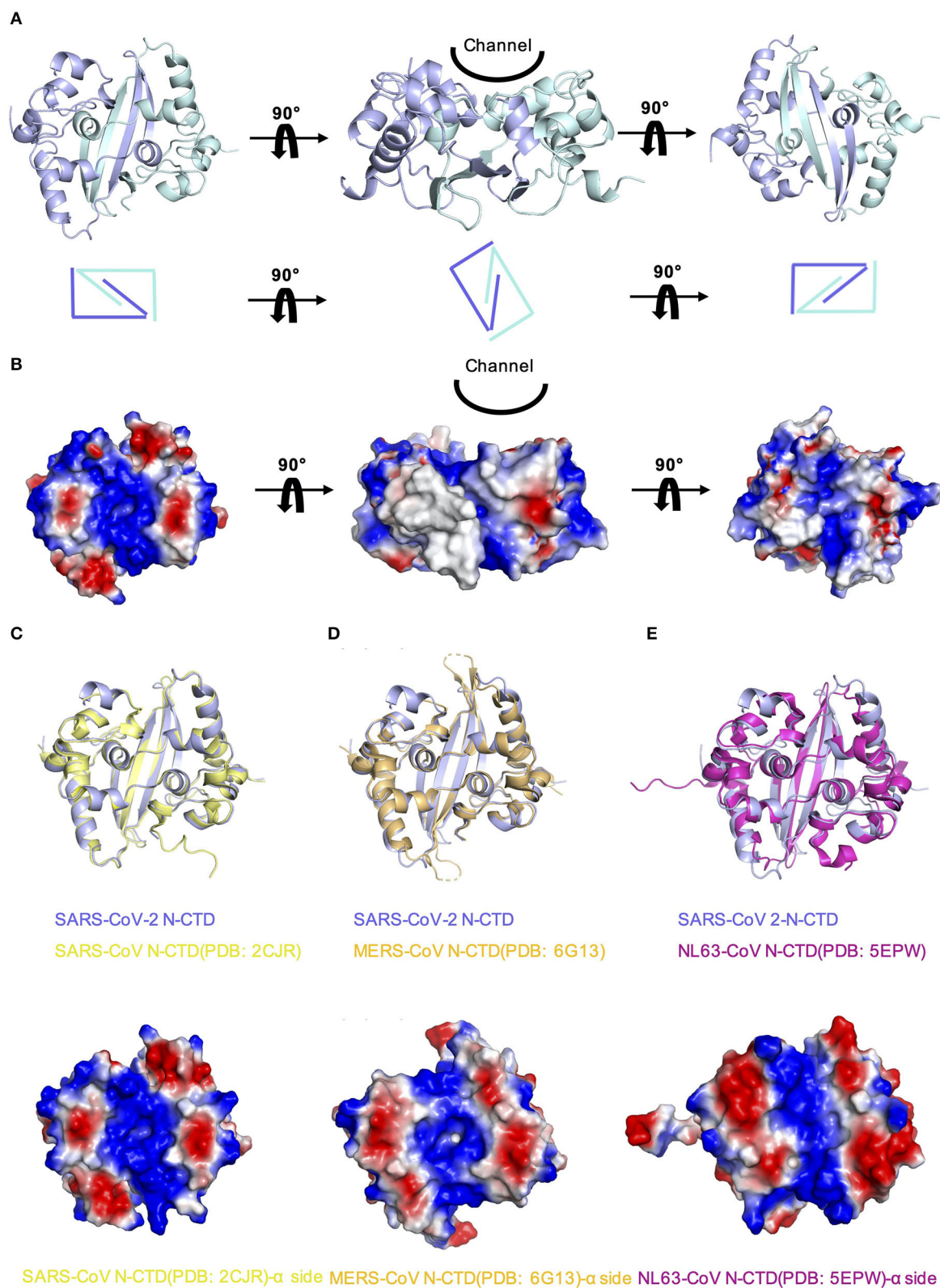
In order to explore whether there are other RNA binding domains that exist in the rest of the nucleocapsid in SARS-CoV-2, we analyzed the surface electrostatic potential characteristics of the SARS-CoV-2 N-CTD based on the structure. The dimer of SARS-CoV-2 N-CTD is shown as a cuboid shape, as described above, containing an  $\alpha$ -helix-rich side and a  $\beta$ -sheet side (Figure 3A). The electrostatic surface suggests that the  $\alpha$ -helix-rich side contains a positively charged channel (Figure 3B, the left panel), whereas the  $\beta$ -sheet side is a neutral surface (Figure 3B, the right panel). Its shape is the same as the positively charged channel rich in  $\alpha$ -helix side in SARS-CoV and MERS-CoV (Figures 3C,D). Intriguingly, although low pathogenic CoV-NL63 has a similarly positively charged channel in its nucleocapsid CTD, the shape of the channel is quite different (Figure 3E) (Szelazek et al., 2017). The surface charge of another low pathogenic IBV is also different from that of SARS-CoV-2 (Supplementary Figure 2) (Jayaram et al., 2006). Nevertheless, the conserved positively charged channel of the  $\alpha$ -helix-rich side is thought to be a potential RNA binding site in the SARS-CoV-2 N-CTD.

## SARS-CoV-2 N-CTD With Flanking Regions Recognizes Transcriptional Regulatory Sequences

We hypothesize that SARS-CoV-2 N-CTD is capable of binding to viral RNA, especially the most conserved transcriptional regulatory sequences (TRSs) of the viral genome. Recent genomic data suggested that there are 10 TRSs in the SARS-CoV-2 genome, with one TRS in the 5' leader region (TRS-L) and nine TRSs in the 3' region (TRS-B) of the genome (Kim et al., 2020) (Figure 4A). To study the mechanisms of SARS-CoV-2 N-CTD protein recognition of TRS, we next synthesized three TRSs in the leader region for *in vitro* binding assays, named TRS-1 (5'-ACGAAC-3', 6 nucleotides), TRS-2 (5'-AAACGAAC-3', 8 nucleotides), and TRS-3 (5'-AAACGAACUU-3', 10 nucleotides) (Figure 4B). For the shortest sequence, TRS-1, SARS-CoV-2 N-CTD showed a very weak binding affinity of 320  $\mu$ M via biolayer interferometry assays (Figure 4C, the left panel). However, SARS-CoV-2 N-CTD with a flanking internal disorder region, regardless of the middle LKR motif (residues 183 to 365, termed LKR+CTD) or the C-terminal IDR (residues 250 to 419, termed



**FIGURE 2 |** The crystal packing of SARS-CoV-2 N-CTD. **(A)** SARS-CoV-2 N-CTD high-order oligomers in the crystal packing. The interaction regions of putative high-order oligomers are highlighted with dashed boxes I, II, and III. **(B)** Zoomed-in detailed view of the interaction of symmetric molecules in dashed box I. **(C)** Zoomed-in detailed view of the interaction of symmetric molecules in dashed box II. **(D)** Zoomed-in detailed view of the interaction of symmetric molecules in dashed box III. **(E)** Native PAGE analysis of SARS-CoV-2-N-CTD and mutants in disulfide trapping assays.



**FIGURE 3 |** Comparison of SARS-CoV-2 N-CTD with related viral N-CTD structures. **(A)** Structure of SARS-CoV-2 N-CTD. **(B)** Electrostatic surface of SARS-CoV-2 N-CTD. Blue denotes a positive charge potential, while red indicates a negative charge potential. **(C)** Overall structural comparison of SARS-CoV-2 N-CTD with related viral N-CTD structures. (Continued)

**FIGURE 3 |** SARS-CoV N-CTD. Top panel: superimposition of SARS-CoV-2 N-CTD (blue) with SARS-CoV N-CTD (yellow). Bottom panel: electrostatic surface of SARS-CoV N-CTD. **(D)** Overall structural comparison of SARS-CoV-2 N-CTD with MERS-CoV N-CTD. Top panel: superimposition of SARS-CoV-2 N-CTD (blue) with MERS-CoV N-CTD (orange). Bottom panel: electrostatic surface of MERS-CoV N-CTD. **(E)** Overall structural comparison of SARS-CoV-2 N-CTD with CoV-NL63 N-CTD. Top panel: superimposition of SARS-CoV-2 N-CTD (blue) with SARS-CoV N-CTD (magenta). Bottom panel: electrostatic surface of the CoV-NL63 N-CTD.

CTD+IDR), showed up to a 20-fold elevated binding affinity with  $K_d$  values of 14 and 33  $\mu\text{M}$ , respectively (**Figure 4C**, the middle and right panels). As the length of the TRS increased, the binding affinities were enhanced (**Figures 4D,E**). TRS-2 interacted with CTD, CTD+IDR, and LKR+CTD with binding affinities of 180, 9.1, and 17  $\mu\text{M}$ , respectively. Furthermore, we found that the tightest interaction is CTD+IDR bound to TRS-3, with a  $K_d$  value of 5.9  $\mu\text{M}$  (**Figure 4E**, right panel), whereas CTD and LKR+CTD bind to TRS-3 with a  $K_d$  value of 150 and 61  $\mu\text{M}$ , respectively.

In the single positive strand of the viral RNA genome, TRS-L has a stem-loop structure. Along with the discontinuous transcription process, the TRS has two other states: the single-stranded TRS-B and the double-stranded fully paired cTRS (**Figure 5A**). Therefore, we next continued to explore if there are any differences in CTD+IDR with different TRSs. The CTD+IDR protein binds to Duplex TRS-1 (TRS1-cTRS1, 6 bp), Duplex TRS-2 (TRS2-cTRS2, 8 bp), and Duplex TRS-3 (TRS3-cTRS3, 10 bp) with double-stranded TRS-paired RNA with  $K_d$  values of 29, 18, and 11  $\mu\text{M}$ , respectively (**Figures 5B–D**). The binding affinities are approximately or slightly weaker than those of its single-stranded counterpart. For the imperfectly paired RNA composed of TRS1-cTRS2(unpaired TRS-1) or TRS2-cTRS-3(unpaired TRS-2), the binding affinities were similar to those of double-stranded RNA substrates, with  $K_d$  values of 23 and 33  $\mu\text{M}$ , respectively (**Figures 5E,F**). To summarize the binding results, we determined that the unpaired adeno dinucleotides in the 5' regions of the TRS, which exist in TRS-3 and cTRS-3 but not in the paired RNA or imperfectly paired RNA substrates, have a micro-molar binding affinity to nucleocapsid protein CTD-IDR constructs, whereas the unpaired adeno dinucleotides in the 3' region of the TRS (i.e., inside the TRS2-cTRS-3 imperfectly paired RNA) have a weaker binding ability to the same protein. Combined with the TRS-L model, these data suggest that the 5' unpaired adeno dinucleotide in the stem-loop region of TRS-L is a key factor involved in the binding of the nucleocapsid protein.

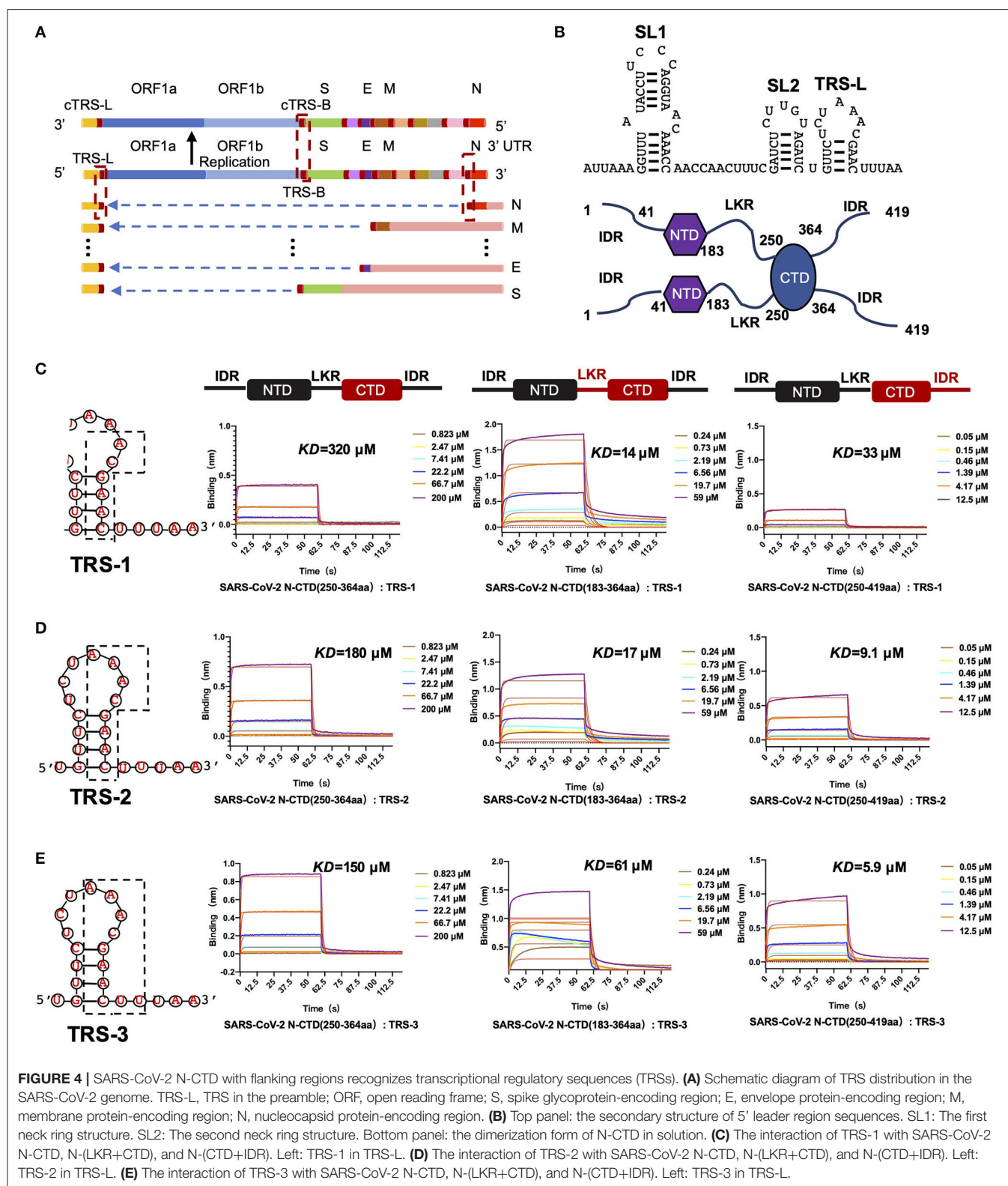
## DISCUSSION

The N protein mediates ribonucleoprotein (RNP) complex formation via two key steps: packaging of the viral RNA genome and self-assembly of oligomerizations. Studies on coronavirus N-CTD suggest that the multiple packing modes of N-CTD dimers probably lead to the formation of rigid helically symmetric nucleocapsids, an unusual feature that is supported by various biochemical assays, including the disulfide trapping technique (Jayaram et al., 2006; Chen et al., 2007; Chang et al., 2013; Gui

et al., 2017). Currently, the SARS-CoV N-CTD domain self-association has been widely studied for viral RNP assembly (Surjit et al., 2004; Yu et al., 2005; Luo et al., 2006). However, the role of N-CTD in the self-association of SARS-CoV-2 remains unclear. Our structural data suggest that SARS-CoV-2 N-CTD possesses conserved dimerization mechanisms via multiple hydrophilic and hydrophobic interactions, similar to the CTD of other coronavirus nucleocapsid proteins. Intriguingly, the higher-order self-association of SARS-CoV-2 N-CTD seems different from that of SARS-CoV N-CTD in our studies. Previous studies showed that SARS-CoV N-CTD packs into octamers and forms a twin helix in the crystal packing (Chang et al., 2014); however, SARS-CoV-2 N-CTD packs into a cylindrical shape in the crystal packing. To further verify these observations, *in vitro* disulfide trapping assays combined with size-exclusion chromatography were performed to illustrate the status of SARS-CoV-2 N-CTD in solution. Our data suggest that the observed potential self-interactions via the  $\beta 5$ – $\beta 6$  loop and  $\alpha 1$ -helix regions in the crystal actually exist in solution, which may serve as the first step of the RNP assembly process.

Previous studies suggest that the coronavirus nucleocapsid contains multiple RNA binding sites, including the NTD, CTD, and C-terminal IDR regions (Chang et al., 2014). Our previous work demonstrated that the N-terminal domain of the nucleocapsid is capable of binding to viral single-stranded 32-mer RNA. Our structural data suggest that SARS-CoV-2 N-CTD contains a positively charged channel similar to MERS-CoV N-CTD and SARS-CoV N-CTD. These surface electrostatic potential characteristics are conserved among the highly pathogenic viral nucleocapsid proteins (Chen et al., 2007; Nguyen et al., 2019). These positively charged channels in the  $\alpha$ -helix-rich side are considered as potential RNA binding sites in SARS-CoV-2 N-CTD.

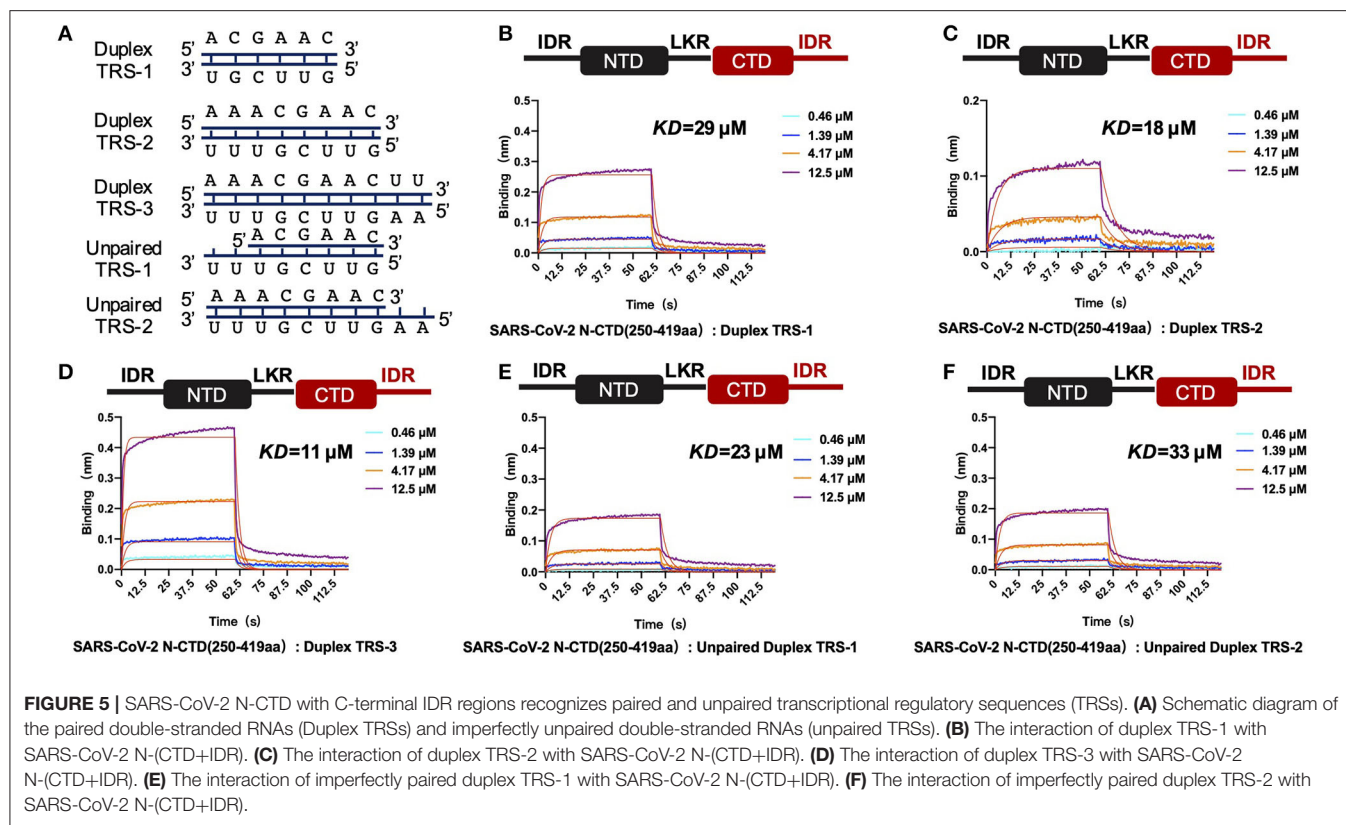
Previous studies demonstrated the role of the CTD in the recognition of the packaging signal in coronavirus nucleocapsid proteins in CoVs, such as MHV (Kuo et al., 2014), MERS-CoV (Hsin et al., 2018), and SARS-CoV (Chang et al., 2009), but HCoV-NL63 N-CTD fails to bind RNA (Zuwała et al., 2015). The nucleocapsid proteins of coronaviruses are homologous and possess a conserved modular composition comprising five domains, represented as the N-tail domain, NTD, LKR, CTD, and C-IDR. The N-NTD, N-CTD, and C-IDR were all reported to bind viral RNA in SARS-CoV (Huang et al., 2004; Chen et al., 2007; Takeda et al., 2008). However, the roles of these domains in RNA binding remain to be elucidated in the SARS-CoV-2 N protein. Our early work suggests that SARS-CoV-2 N-NTD displays a modest binding affinity to viral transcriptional regulatory sequence (TRS) RNA, with a  $K_d$  value of 140  $\mu\text{M}$  (Kang et al., 2020). In our study, although SARS-CoV-2 N-CTD binds to TRS RNA with a relatively weaker binding affinity ( $K_d$



**FIGURE 4 | SARS-CoV-2 N-CTD with flanking regions recognizes transcriptional regulatory sequences (TRSs).** (A) Schematic diagram of TRS distribution in the SARS-CoV-2 genome. TRS-L, TRS in the preamble; ORF, open reading frame; S, spike glycoprotein-encoding region; E, envelope protein-encoding region; M, membrane protein-encoding region; N, nucleocapsid protein-encoding region. (B) Top panel: the secondary structure of 5' leader region sequences. SL1: The first neck ring structure. SL2: The second neck ring structure. Bottom panel: the dimerization form of N-CTD in solution. (C) The interaction of TRS-1 with SARS-CoV-2 N-CTD, N-(LKR+CTD), and N-(CTD+IDR). Left: TRS-1 in TRS-L. (D) The interaction of TRS-2 with SARS-CoV-2 N-CTD, N-(LKR+CTD), and N-(CTD+IDR). Left: TRS-2 in TRS-L. (E) The interaction of TRS-3 with SARS-CoV-2 N-CTD, N-(LKR+CTD), and N-(CTD+IDR). Left: TRS-3 in TRS-L.

value of  $320 \mu\text{M}$ ), SARS-CoV-2 N-CTD with flanking regions (either LKR or C-IDR) demonstrated interactions with the same RNA template in a micro-molar binding affinity (the highest  $K_d$

value was  $5.9 \mu\text{M}$ ). The flanking regions of CTD are rich in positively charged amino acids (seven arginines and four lysines out of 69 total residues in the N-terminal flanking region, with



one arginine and nine lysines out of 55 residues in C-terminal flanking region.). These characteristics of the SARS-CoV-2 N-CTD may explain how the flanking regions are beneficial to the binding of RNA. To our knowledge, the function of the SARS-CoV-2 N LKR motif, which enhances CTD binding to the TRS RNA sequence beyond its potential phosphorylation function of oligomerization (Peng et al., 2008), is reported here for the first time.

In conclusion, in this paper, we analyzed the crystal structure of the nucleocapsid C-terminal domain, studied the potential self-interaction formation of SARS-CoV-2 N-CTD, and verified the self-interaction characteristics of the single-point mutant. By studying the recognition mechanism of SARS-CoV-2 N-CTD protein to TRS, it is found that the 5' unpaired adeno dinucleotide in the stem-loop region of TRS-L is a key factor involved in the binding of nucleocapsid protein. Altogether, these results reveal a new method of viral transcription sequences mechanism.

## DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found at: Protein Data Bank, 7DE1.

## AUTHOR CONTRIBUTIONS

SC and SK contributed to the conception of the study and performed the structural determination and

validation. SC, SK, and MY constructed the article, designed the experiments, created the figures, and wrote the manuscript. SK, MY, and SH performed the protein purification, crystallization, *in vitro* protein-protein interaction, and biochemical experiments. XC, ZH, ZiZ, ZhZ, and QC assisted in analyzing the experimental results of the protein-protein interactions. All authors contributed to the article and approved the submitted version.

## FUNDING

This work was supported by the National Natural Science Foundation of China (31770801), Natural Science Foundation of Guangdong Province, China (2018B030306029 and 2017A030313145), and COVID-19 Emerging Prevention Research Special Fund of Zhuhai City (ZH22036302200016PWC and ZH22046301200011PWC) to SC.

## ACKNOWLEDGMENTS

We thank Guangdong Medical Laboratory Animal Center for providing the N-protein encoding gene plasmids, Dr. Yongzhi Lu from the Guangzhou Institutes of Biomedicine and Health (Chinese Academy of Sciences) for the initial crystals X-ray diffraction screening, and Dr. Xuan Ma from the South China Sea Institute of Oceanology (Chinese Academy of Sciences)

for his support in offering a home source X-ray diffraction facility. We thank the staff from BL18U1 beamline of the National Facility for Protein Science in Shanghai (NFPS) at the Shanghai Synchrotron Radiation Facility for assistance during data collection.

## REFERENCES

- Bass, R. B., Butler, S. L., Chervitz, S. A., Gloor, S. L., and Falke, J. J. (2007). Use of site-directed cysteine and disulfide chemistry to probe protein structure and dynamics: applications to soluble and transmembrane receptors of bacterial chemotaxis. *Two Component Signal. Syst.* 423, 25–51. doi: 10.1016/S0076-6879(07)23002-2
- Chang, C. K., Chen, C. M. M., Chiang, M. H., Hsu, Y. L., and Huang, T. H. (2013). Transient oligomerization of the SARS-CoV N protein - implication for virus ribonucleoprotein packaging. *PLoS ONE* 8:e65045. doi: 10.1371/journal.pone.0065045
- Chang, C. K., Hou, M. H., Chang, C. F., Hsiao, C. D., and Huang, T. H. (2014). The SARS coronavirus nucleocapsid protein - forms and functions. *Antiviral Res.* 103, 39–50. doi: 10.1016/j.antiviral.2013.12.009
- Chang, C. K., Hsu, Y. L., Chang, Y. H., Chao, F. A., Wu, M. C., Huang, Y. S., et al. (2009). Multiple nucleic acid binding sites and intrinsic disorder of severe acute respiratory syndrome coronavirus nucleocapsid protein: implications for ribonucleocapsid protein packaging. *J. Virol.* 83, 2255–2264. doi: 10.1128/JVI.02001-08
- Chen, C. Y., Chang, C. K., Chang, Y. W., Sue, S. C., Bai, H. I., et al. (2007). Structure of the SARS coronavirus nucleocapsid protein RNA-binding dimerization domain suggests a mechanism for helical packaging of viral RNA. *J. Mol. Biol.* 368, 1075–1086. doi: 10.1016/j.jmb.2007.02.069
- Diederichs, K., and Karplus, P. A. (1997). Improved R-factors for diffraction data analysis in macromolecular crystallography. *Nat. Struct. Biol.* 4, 269–275. doi: 10.1038/nsb0497-269
- Evans, P. (2006). Scaling and assessment of data quality. *Acta Crystallogr. D Biol. Crystallogr.* 62(Pt. 1), 72–82. doi: 10.1107/S0907444905036693
- Evans, P. R. (2011). An introduction to data reduction: space-group determination, scaling and intensity statistics. *Acta Crystallogr. D Biol. Crystallogr.* 67(Pt. 4), 282–292. doi: 10.1107/S090744491003982X
- Gordon, D. E., Jang, G. M., Bouhaddou, M., Xu, J., Obernier, K., White, K. M., et al. (2020). A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. *Nature* 583, 459–468. doi: 10.1038/s41586-020-2286-9
- Gui, M., Liu, X., Guo, D. Y., Zhang, Z., Yin, C. C., Chen, Y., et al. (2017). Electron microscopy studies of the coronavirus ribonucleoprotein complex. *Protein Cell* 8, 219–224. doi: 10.1007/s13238-016-0352-8
- Hsin, W. C., Chang, C. H., Chang, C. Y., Peng, W. H., Chien, C. L., Chang, M. F., et al. (2018). Nucleocapsid protein-dependent assembly of the RNA packaging signal of Middle East respiratory syndrome coronavirus. *J. Biomed. Sci.* 25:47. doi: 10.1186/s12929-018-0449-x
- Huang, Q., Yu, L., Petros, A. M., Gunasekera, A., Liu, Z., Xu, N., et al. (2004). Structure of the N-terminal RNA-binding domain of the SARS CoV nucleocapsid protein. *Biochemistry* 43, 6059–6063. doi: 10.1021/bi036155b
- Jayaram, H., Fan, H., Bowman, B. R., Ooi, A., Jayaram, J., Collisson, E. W., et al. (2006). X-ray structures of the N- and C-terminal domains of a coronavirus nucleocapsid protein: implications for nucleocapsid formation. *J. Virol.* 80, 6612–6620. doi: 10.1128/JVI.00157-06
- Kang, S., Yang, M., Hong, Z., Zhang, L., Huang, Z., Chen, X., et al. (2020). Crystal structure of SARS-CoV-2 nucleocapsid protein RNA binding domain reveals potential unique drug targeting sites. *Acta Pharm. Sin. B* 10, 1228–1238. doi: 10.1016/j.apsb.2020.04.009
- Kim, D., Lee, J. Y., Yang, J. S., Kim, J. W., Kim, V. N., and Chang, H. (2020). The architecture of SARS-CoV-2 transcriptome. *Cell* 181, 914–921 e910. doi: 10.1016/j.cell.2020.04.011
- Kuo, L., Koetzner, C. A., Hurst, K. R., and Masters, P. S. (2014). Recognition of the murine coronavirus genomic RNA packaging signal depends on the second RNA-binding domain of the nucleocapsid protein. *J. Virol.* 88, 4451–4465. doi: 10.1128/JVI.03866-13
- Lai, M. M., and Stohlman, S. A. (1981). Comparative analysis of RNA genomes of mouse hepatitis viruses. *J. Virol.* 38, 661–670. doi: 10.1128/JVI.38.2.661-670.1981
- Luo, H., Chen, J., Chen, K., Shen, X., and Jiang, H. (2006). Carboxyl terminus of severe acute respiratory syndrome coronavirus nucleocapsid protein: self-association analysis and nucleic acid binding characterization. *Biochemistry* 45, 11827–11835. doi: 10.1021/bi0609319
- Ma, Y., Tong, X., Xu, X., Li, X., Lou, Z., and Rao, Z. (2010). Structures of the N- and C-terminal domains of MHV-A59 nucleocapsid protein corroborate a conserved RNA-protein binding mechanism in coronavirus. *Protein Cell* 1, 688–697. doi: 10.1007/s13238-010-0079-x
- Masters, P. S., and Sturman, L. S. (1990). Background paper: functions of the coronavirus nucleocapsid protein. *Adv. Exp. Med. Biol.* 276, 235–238. doi: 10.1007/978-1-4684-5823-7\_32
- McBride, R., van Zyl, M., and Fielding, B. C. (2014). The coronavirus nucleocapsid is a multifunctional protein. *Viruses* 6, 2991–3018. doi: 10.3390/v6082991
- Nguyen, T. H. V., Lichiere, J., Canard, B., Papageorgiou, N., Attoumani, S., Ferron, F., et al. (2019). Structure and oligomerization state of the C-terminal region of the Middle East respiratory syndrome coronavirus nucleoprotein. *Acta Crystallogr. Sect. D Struct. Biol.* 75, 8–15. doi: 10.1107/S2059798318014948
- Pasternak, A. O., Spaan, W. J. M., and Snijder, E. J. (2006). Nidovirus transcription: how to make sense...? *J. Gen. Virol.* 87, 1403–1421. doi: 10.1099/vir.0.81611-0
- Peng, T. Y., Lee, K. R., and Tarn, W. Y. (2008). Phosphorylation of the arginine/serine dipeptide-rich motif of the severe acute respiratory syndrome coronavirus nucleocapsid protein modulates its multimerization, translation inhibitory activity and cellular localization. *FEBS J.* 275, 4152–4163. doi: 10.1111/j.1742-4658.2008.06564.x
- Saikatendu, K. S., Joseph, J. S., Subramanian, V., Neuman, B. W., Buchmeier, M. J., Stevens, R. C., et al. (2007). Ribonucleocapsid formation of severe acute respiratory syndrome coronavirus through molecular action of the N-terminal domain of N protein. *J. Virol.* 81, 3913–3921. doi: 10.1128/JVI.02236-06
- Sawicki, S. G., Sawicki, D. L., and Siddell, S. G. (2007). A contemporary view of coronavirus transcription. *J. Virol.* 81, 20–29. doi: 10.1128/JVI.01358-06
- Snijder, E. J., Decroly, E., and Ziebuhr, J. (2016). The nonstructural proteins directing coronavirus RNA synthesis and processing. *Adv. Virus Res.* 96, 59–126. doi: 10.1016/bs.aivir.2016.08.008
- Sola, I., Almazan, F., Zuniga, S., and Enjuanes, L. (2015). Continuous and discontinuous RNA synthesis in coronaviruses. *Annu Rev Virol.* 2, 265–288. doi: 10.1146/annurev-virology-100114-055218
- Sola, I., Mateos-Gomez, P. A., Almazan, F., Zuniga, S., and Enjuanes, L. (2011). RNA-RNA and RNA-protein interactions in coronavirus replication and transcription. *RNA Biol.* 8, 237–248. doi: 10.4161/rna.8.2.14991
- Surjit, M., Liu, B., Kumar, P., Chow, V. T. K., and Lal, S. K. (2004). The nucleocapsid protein of the SARS coronavirus is capable of self-association through a C-terminal 209 amino acid interaction domain. *Biochem. Biophys. Res. Commun.* 317, 1030–1036. doi: 10.1016/j.bbrc.2004.03.154
- Szelazek, B., Kabala, W., Kus, K., Zdzalik, M., Twarda-Clapa, A., Golik, P., et al. (2017). Structural characterization of human coronavirus NL63 N protein. *J. Virol.* 91:e02503-16. doi: 10.1128/JVI.02503-16
- Takeda, M., Chang, C. K., Ikeya, T., Güntert, P., Chang, Y. H., Hsu, Y. L., et al. (2008). Solution structure of the C-terminal dimerization domain of SARS coronavirus nucleocapsid protein solved by the SAIL-NMR method. *J. Mol. Biol.* 380, 608–622. doi: 10.1016/j.jmb.2007.11.093
- Weiss, M. S. (2001). Global indicators of X-ray data quality. *J. Appl. Crystallogr.* 34, 130–135. doi: 10.1107/S0021889800018227
- Wu, C. H., Chen, P. J., and Yeh, S. H. (2014). Nucleocapsid phosphorylation and RNA helicase DDX1 recruitment enables coronavirus transition from discontinuous to continuous transcription. *Cell Host Microbe* 16, 462–472. doi: 10.1016/j.chom.2014.09.009

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fchem.2020.624765/full#supplementary-material>

- Wu, F., Zhao, S., Yu, B., Chen, Y. M., Wang, W., Song, Z. G., et al. (2020). A new coronavirus associated with human respiratory disease in China. *Nature* 579, 265–269. doi: 10.1038/s41586-020-2008-3
- Yu, I. M., Gustafson, C. L., Diao, J., Burgner, J. W. II, Li, Z., Zhang, J., et al. (2005). Recombinant severe acute respiratory syndrome (SARS) coronavirus nucleocapsid protein forms a dimer through its C-terminal domain. *J. Biol. Chem.* 280, 23280–23286. doi: 10.1074/jbc.M501015200
- Yu, I. M., Oldham, M. L., Zhang, J., and Chen, J. (2006). Crystal structure of the severe acute respiratory syndrome (SARS) coronavirus nucleocapsid protein dimerization domain reveals evolutionary linkage between Corona- and Arteriviridae. *J. Biol. Chem.* 281, 17134–17139. doi: 10.1074/jbc.M602107200
- Zhou, P., Yang, X. L., Wang, X. G., Hu, B., Zhang, L., Zhang, W., et al. (2020). A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* 579, 270–273. doi: 10.1038/s41586-020-2012-7
- Zúñiga, S., Cruz, J. L. G., Sola, I., Mateos-Gómez, P. A., Palacio, L., and Enjuanes, L. (2010). Coronavirus nucleocapsid protein facilitates template switching and is required for efficient transcription. *J. Virol.* 84, 2169–2175. doi: 10.1128/JVI.02011-09
- Zuwała, K., Golda, A., Kabala, W., Burmistrz, M., Zdzalik, M., Nowak, P., et al. (2015). The nucleocapsid protein of human coronavirus NL63. *PLoS ONE* 10:e0117833. doi: 10.1371/journal.pone.0117833

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Yang, He, Chen, Huang, Zhou, Zhou, Chen, Chen and Kang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Microsecond MD Simulation and Multiple-Conformation Virtual Screening to Identify Potential Anti-COVID-19 Inhibitors Against SARS-CoV-2 Main Protease

Chandrabose Selvaraj<sup>1\*</sup>, Umesh Panwar<sup>1</sup>, Dhurvas Chandrasekaran Dinesh<sup>2</sup>, Evzen Boura<sup>2</sup>, Poonam Singh<sup>3</sup>, Vikash Kumar Dubey<sup>4</sup> and Sanjeev Kumar Singh<sup>1\*</sup>

<sup>1</sup> Computer Aided Drug Design and Molecular Modeling Lab, Department of Bioinformatics, Alagappa University, Karaikudi, India, <sup>2</sup> Section of Molecular Biology and Biochemistry, Institute of Organic Chemistry and Biochemistry AS CR, v.v.i., Prague, Czechia, <sup>3</sup> Corrosion and Materials Protection Division, Council of Scientific and Industrial Research (CSIR)-Central Electrochemical Research Institute, Karaikudi, India, <sup>4</sup> School of Biochemical Engineering, Indian Institute of Technology (BHU), Varanasi, India

## OPEN ACCESS

### Edited by:

Simone Brogi,  
University of Pisa, Italy

### Reviewed by:

Sergio F. Sousa,  
University of Porto, Portugal  
Claudio Norberto Cavasotto,  
Universidad Austral, Argentina  
Marco Tutone,  
University of Palermo, Italy

### \*Correspondence:

Chandrabose Selvaraj  
selnikraj@bioclues.org  
Sanjeev Kumar Singh  
skysanjeev@gmail.com

### Specialty section:

This article was submitted to  
Medicinal and Pharmaceutical  
Chemistry,  
a section of the journal  
Frontiers in Chemistry

**Received:** 15 August 2020

**Accepted:** 19 November 2020

**Published:** 13 January 2021

### Citation:

Selvaraj C, Panwar U, Dinesh DC,  
Boura E, Singh P, Dubey VK and  
Singh SK (2021) Microsecond MD  
Simulation and Multiple-Conformation  
Virtual Screening to Identify Potential  
Anti-COVID-19 Inhibitors Against  
SARS-CoV-2 Main Protease.  
Front. Chem. 8:595273.  
doi: 10.3389/fchem.2020.595273

The recent pandemic outbreak of COVID-19, caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), raised global health and economic concerns. Phylogenetically, SARS-CoV-2 is closely related to SARS-CoV, and both encode the enzyme main protease ( $M^{pro}/3CL^{pro}$ ), which can be a potential target inhibiting viral replication. Through this work, we have compiled the structural aspects of  $M^{pro}$  conformational changes, with molecular modeling and 1- $\mu$ s MD simulations. Long-scale MD simulation resolves the mechanism role of crucial amino acids involved in protein stability, followed by ensemble docking which provides potential compounds from the Traditional Chinese Medicine (TCM) database. These lead compounds directly interact with active site residues (His41, Gly143, and Cys145) of  $M^{pro}$ , which plays a crucial role in the enzymatic activity. Through the binding mode analysis in the S1, S1', S2, and S4 binding subsites, screened compounds may be functional for the distortion of the oxyanion hole in the reaction mechanism, and it may lead to the inhibition of  $M^{pro}$  in SARS-CoV-2. The hit compounds are naturally occurring compounds; they provide a sustainable and readily available option for medical treatment in humans infected by SARS-CoV-2. Henceforth, extensive analysis through molecular modeling approaches explained that the proposed molecules might be promising SARS-CoV-2 inhibitors for the inhibition of COVID-19, subjected to experimental validation.

**Keywords:** SARS-CoV-2 main protease, COVID-19, TCM, natural products, molecular dynamics, ensemble sampling

## INTRODUCTION

The novel coronavirus “Severe Acute Respiratory Syndrome Coronavirus-2” (SARS-CoV-2) became a disease of interest with the initial alert of several pneumonia cases on 31 December 2019 by China (Lai et al., 2020). Since then, the virus has spread globally, representing a major threat to public health, and eventually, on 11 March 2020, WHO announced it as a pandemic

(Vannabouathong et al., 2020). As of 19<sup>th</sup> November 2020, there are over 55.6 million cases globally, with a 2.6% case-mortality rate (<https://ourworldindata.org/mortality-risk-covid>). From the coronavirus family, SARS-CoV-2 is the seventh virus that infects humans, and like other viruses in that family, like SARS-CoV and MERS-CoV, SARS-CoV-2 has a high lethality (Li et al., 2020; Zhu et al., 2020). COVID-19 can be either asymptomatic or symptomatic, and severe cases result in pneumonia, multiple-organ failure, and eventual death (Ayres, 2020). The SARS-CoV-2 belongs to the family *Coronaviridae* and subfamily *Coronavirinae*, which contains enveloped, positive-sense single-stranded RNA (+ssRNA) viruses whose external glycoprotein spikes from the envelope appear as a “corona,” which is Latin for crown or halo-like, hence the virus family name (Pal et al., 2020). The SARS-CoV-2 whole-genome sequence data (29,903 nucleotides) shows an overall 82% sequence identity with SARS-CoV. Sequence analyses confirm SARS-CoV-2's possible natural reservoirs are horseshoe bats (*Rhinolophus sp.*), though the intermediate host is still not clear (Zhao et al., 2020; Zheng, 2020). Available metagenomics data and sequence similarities of CoV from animals suggest that Malayan pangolins (*Manis javanica*, long-snouted, ant-eating mammals) could be a possible intermediate host (Rabi et al., 2020; Wahba et al., 2020). These lethal ssRNA viruses are highly flexible to adapt, acquiring new mutations which enable them to move into a new host and to elude available conventional drugs and making vaccine development challenging (Hanney et al., 2020). For temporary solutions, repurposing FDA-approved drugs, especially low-molecular-weight drugs, and using available cutting-edge techniques like CRISPR/Cas13d targeting SARS-CoV-2 viral RNA can also be rapid approaches to fight COVID-19 (Gao et al., 2020; Mohammadi et al., 2020; Singh et al., 2020). In this pandemic, it is mandatory to analyze the drug targets of SARS-CoV-2 with available FDA-approved compounds and also to find new inhibitors against those targets (Gil et al., 2020; Meyer-Almes, 2020; Saul and Einav, 2020). Therefore, structural biology and computational virtual screening of antiviral molecules that target key viral proteins can be comparatively faster and more effective than developing vaccines or therapeutic antibodies to fight against SARS-CoV-2 (Battisti et al., 2020; Frances-Monerris et al., 2020; Shyr et al., 2020). The host dependency factors mediating virus infection will be the key to understanding effective molecular targets for developing broadly acting antiviral therapeutics against SARS-CoV-2 (Tharappel et al., 2020).

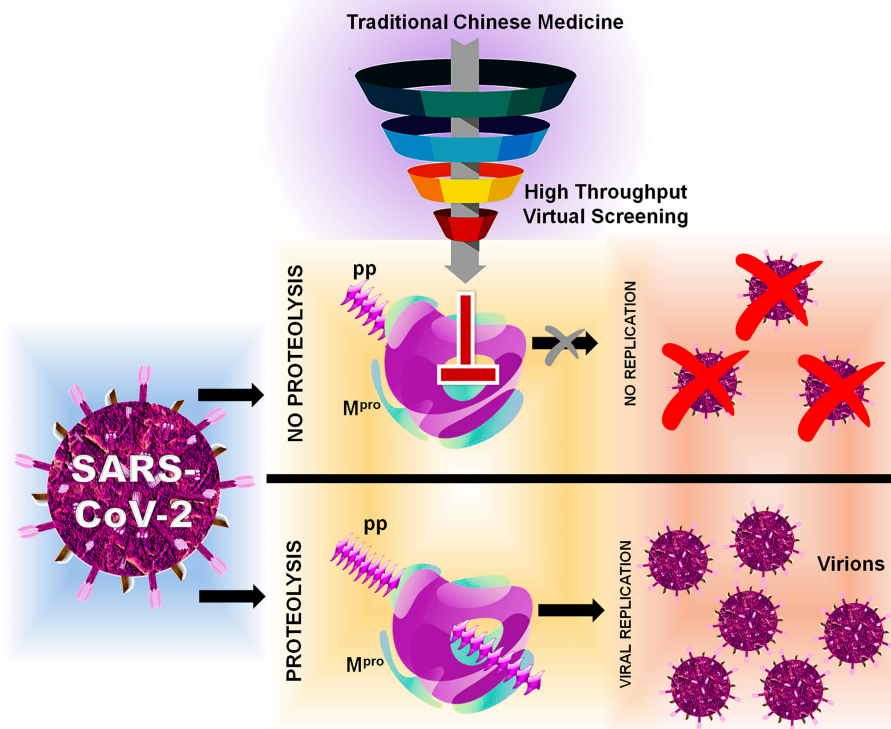
In this, SARS-CoV-2 encodes two different proteases that are crucial for viral replication, polyprotein processing, and immune regulation, namely, 3-chymotrypsin-like protease or main protease (3CL<sup>pro</sup>/M<sup>pro</sup>) and papain-like protease (PL<sup>pro</sup>) (Pillaiyar et al., 2016; Jeong et al., 2020; Luan et al., 2020). These two enzymes process the large polyprotein (pp) or replicase 1a and 1ab, which are translated from the viral RNA. M<sup>pro</sup>/3CL<sup>pro</sup> has eleven cleavage sites in pp1a and pp1ab, with similar processing pathways to other coronaviruses, generating many of the non-structural proteins which are important in viral replication (Fang et al., 2008). These proteases possess a characteristic active-site architecture: glutamine in the P1

position of the substrate and (small)-X-(L/F/M)-Q↓(G/A/S)-X as a cleavage pattern (X-any amino acid). No known human protease functions with similar specificity, making M<sup>pro</sup>/3CL<sup>pro</sup> unique and also avoid side effects or toxicity (Hilgenfeld, 2014; Eleftheriou et al., 2020). M<sup>pro</sup> contains a catalytic Cys...His dyad which serves as a functionally active dimer, making it unique from the other enteroviral 3C proteases (Ullrich and Nitsche, 2020). Therefore, M<sup>pro</sup> can be one of the key drug targets because of its strikingly high sequence and 3D structural similarity with M<sup>pro</sup> from several other coronaviruses, for which several recently experimentally solved crystal structures for SARS-CoV-2 apo and holo forms are available in Protein Data Bank (PDB) (Gahlawat et al., 2020; Goyal and Goyal, 2020). Based on the initial structural analyses, it is already clear that drug-binding pockets of these enzymes are highly conserved as per the genome sequence analyses. Considering the functional importance, several researchers have put a mass effort via *in silico* and *in vitro* approaches, for the FDA-approved compounds (Frances-Monerris et al., 2020; Touret et al., 2020). Currently, big pharmaceutical companies worldwide are taking advantage of the available traditional Chinese medicine (TCM) databases for performing virtual screening and developing novel lead compounds for a wide range of diseases (Chen, 2011). Therefore, we have also adapted computer-aided drug discovery (CADD) approaches to screen and identify possible novel small-molecule drug-like inhibitors against SARS-CoV-2 M<sup>pro</sup>/3CL<sup>pro</sup> available from the TCM Database@Taiwan (<http://tcm.cmu.edu.tw>) which include over 20,000 compounds isolated from TCMs shown in **Figure 1** indicating the inhibition mechanism elucidated in this study.

## MATERIALS AND METHODS

### Protein and Ligand Preparation

For the molecular modeling calculations, the crystal structure SARS-CoV-2 main protease (M<sup>pro</sup>/3CL<sup>pro</sup>) with the PDB ID: 6LU7 is prepared utilizing the standard protocol of the protein preparation wizard (Jin et al., 2020). The co-crystal ligand N3 is covalently bound with an active site amino acid Cys145, and in preparation we manually break the covalent bond and fill the open valance (Culletta et al., 2020; Wang, 2020). Here, the missing atoms and partial charges (charge alteration shown in **Supplementary Figures 1a,b**) are added, and the missing side-chain atoms and bond orders are refined (Sastri et al., 2013). Absolute side-chain angles of amino acids (Asn, Gln, and His) are obtained by flip, which can influence the formation of H-bonds, generating tautomers/ionized states. The intramolecular H-bonds were optimized and minimized by using the OPLS-3e force field (FF) till the RMSD threshold reaches 0.30 Å for all atoms (Selvaraj et al., 2020a). Similarly, the ligands from the TCM (Traditional Chinese Medicine) database@Taiwan (<http://tcm.cmu.edu.tw/>) is prepared using the LigPrep module using the OPLS-3e FF. TCM database holds extensive sources of medicinal benefits with long history, and we believe that TCM can provide the potent leads for SARS-CoV-2 M<sup>pro</sup> inhibition. Ligand ionization states at pH 7.0 ± 2.0 and stereoisomers are generated for each ligand structure. Conformations up



**FIGURE 1** | Flowchart of HTVS with SARS-CoV-2 with protein M<sup>pro</sup>/3CL<sup>pro</sup> polypeptide processing and inhibition.

to 10 poses are generated based on the available rotatable bonds and subject to molecular modeling calculations (Sastry et al., 2013).

### Molecular Dynamics Simulation: Stage I

The prepared structure of SARS-CoV-2 M<sup>pro</sup>/3CL<sup>pro</sup> with the presence and absence of peptidomimetic inhibitor N3 is subject to molecular dynamic (MD) simulations using the GROMINGEN MACHINE for Chemical Simulations (GROMACS 5.1.4: <http://www.gromacs.org/>) (Van Der Spoel et al., 2005; Selvaraj et al., 2018). This stage I MD simulation is performed for the timescale of 1 microsecond for understanding the structural variance occurring between the apo and ligand-bound complex. The apo and complex are subject to simple point charge (SPC) water molecules within a cubic period box of 1.0 nm distance, fixed in position between the protein and cubic box system is prepared with GROMOS96 54a7 FF (Pronk et al., 2013). The molecular topology of the N3 inhibitor is generated externally using the PRODRG web server (<http://prodr1.dyndns.org/>) and then merged with protein topology files prepared by the GROMACS (Van Aalten et al., 1996; Zeng et al., 2015). The compiled system is neutralized by adding an accurate concentration of (Na<sup>+</sup>/Cl<sup>-</sup>) ions and subject to energy minimization, to remove initial steric clashes using 1,000 steps of the steepest descent algorithm via a tolerance of 10 kJ/mol/nm. Literature evidence suggests adaptation of coronaviruses to host species with different body temperatures. SARS-CoV-2 M<sup>pro</sup> is a viral protein, which can

adopt the human body temperature of 300–310 K, and thus we have chosen the default Berendsen thermostat with 300 K as reference temperature and Parrinello-Rahman pressure-coupling with a 1.0-bar reference pressure. The LINCS algorithm is applied as length constraints of covalent bonds, and particle-mesh Ewald is applied for computing the long-range electrostatic interactions (Chinnasamy et al., 2020a,b). The vdW and Coulomb energy cutoff values are set to 1.0 nm, and the time step is defined as 2 fs recorded in the intervals of 10 ps (Umesh et al., 2020). The minimized systems are well-equilibrated for 1,000 ps at 300 K and 1 bar pressure in NVT and NPT ensembles. Furthermore, the MD simulation is processed for the stability and time-dependent behavior of both apo and ligand-bound complexes for the long run simulation time of 1  $\mu$ s (Childers and Daggett, 2018). Final trajectories are analyzed using UCSF Chimera, Visual Molecular Dynamics (VMD), GROMACS tools, and the secondary structure and ligand interactions are predicted through PDBsum.

### Preparation of Multiple Conformation-Based GRIDs

For the molecular docking, multiple grids are prepared from different conformations obtained from the 0-ns to 1- $\mu$ s MD simulations. From the MD trajectories, the measure of similarity/dissimilarity is checked for conformational changes. From this, we have evidently observed the backbone stability in SARS-CoV-2 M<sup>pro</sup> for the long-scale MD simulations

(Evangelista Falcon et al., 2019). Thus, the conformations obtained from each 50-ns interval snapshots from 0 ns to 1  $\mu$ s along with an average conformation (21 conformations) are extracted and subject to Glide Grid generation. Even though the snapshot interval of 50 ns is a bit high, the long-scale 1- $\mu$ s MD simulation shows that the backbone is stable and side residues are contributing to the conformational changes (Amaro et al., 2018; Salmaso and Moro, 2018). Thus, the 50-ns interval conformations from 0-ns to 1- $\mu$ s MD simulations are extracted and subjected to the grid generation method, by matching the ligand-bound pose, and the grid for glide docking is prepared (Lorber and Shoichet, 1998; Halgren et al., 2004). The box positioned on the interacting amino acids Phe140, Gly143, Cys145, His163, His164, Glu166, Gln189, and Thr190 coordinates is constructed to perform a docking analysis by including the positional region focused around 2 Å in the grid generation (Friesner et al., 2006). This process is repeatedly performed for all the 21 conformations obtained through the MD simulation.

## Virtual Screening From the TCM Database

Using the multiple-grid input in the virtual screening workflow and the prepared TCM database, high-throughput virtual screening is carried out. Within this framework, the top 10% of ligands linked to each conformation are processed from HTVS (high-throughput virtual screening) to SP docking (standard precision) and then to XP docking (extra precision) (Seifert et al., 2007). Final compounds on XP docking are processed with molecular mechanics with the generalized born surface area (MM/GBSA) method for gauging the effectiveness of interactions between a docked protein-ligand complex (Zoete et al., 2010; Selvaraj et al., 2020a). Here the average binding free energy ( $\Delta G_{\text{bind}}$ ) is calculated based on the equation ( $\Delta G_{\text{bind}} = \Delta E_{\text{MM}} + \Delta G_{\text{Solv}} + \Delta G_{\text{SA}}$ ), in which  $\Delta E_{\text{MM}}$  denotes minimized energies of protein and ligand,  $\Delta G_{\text{Solv}}$  represents solvation-free energy, and  $\Delta G_{\text{SA}}$  is the surface area energy (Tripathi et al., 2012). Best compounds from XP docking and MM/GBSA are filtered with the criteria of  $< -8$  and  $< -30$  kcal/mol of docking score and  $\Delta G_{\text{bind}}$ , respectively. Compounds surpassing the abovementioned criteria are again redocked with average conformation of the 1- $\mu$ s MD simulations, using the induced fit docking (IFD) method. The IFD approach provided the selections to fix both the ligand and protein as flexible, and this furnishes the final possible best pose to describe (Selvaraj et al., 2020b). The scaling factor is defined as 0.5 for softening the potentials of protein and ligand, and the final complex pose up to twenty poses is saved. The IFD scores, which accounts for both the protein-ligand interaction energy and the total energy of the system, were calculated (Mizutani et al., 2006). For the hit compounds, theoretical validation is performed using the enrichment studies from the decoy set of the known 1,000 active sets, available in the Schrödinger database (Gani et al., 2013). This is performed for evaluating the power of hit compounds, accuracy ranking of hit compounds, and model reliability (Kalyanamoorthy and Chen, 2014; Perez-Regidor et al., 2016).

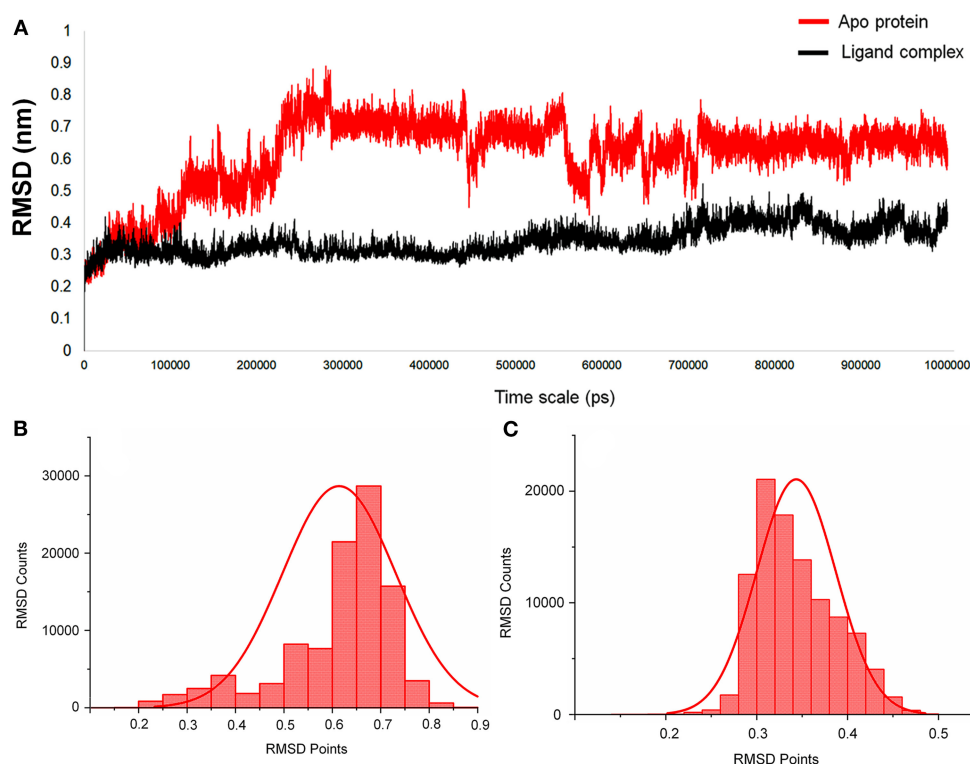
## Molecular Dynamics Simulation: Stage II

The final best pose from the IFD docking is subject to MD simulations of the complex for the timescale of 100 ns for understanding the dynamic behavior of hit compounds with SARS-CoV-2 M<sup>Pro</sup> (Umesh et al., 2020). The ligand complex MD simulations are also performed as per protocol provided in the stage 1 MD simulations, and for the analysis, the root mean square deviation (RMSD), solvent-accessible surface area (SASA), and ligand-binding energy using MM-PBSA are calculated (Klimovich and Mobley, 2015; Selvaraj et al., 2018). The molecular mechanics' potential energy along with the free energy of solvation of individual complexes is analyzed using the equation ( $\Delta G_{\text{binding}} = E_{\text{gas}} + G_{\text{sol}} - T\Delta S$ ) by considering all the frames from 100 ns of MD simulation trajectories. For this calculation, the "bondi" was considered as type of radius (-rad), with an assigned default value 1, the inner dielectric constant value is quoted as 2, and for solvent, the value is assigned as 80 (Aldeghi et al., 2017). The detailed methodology of the MM/PBSA analysis is provided in the **Supplementary Information**.

## RESULTS

### Main Protease Structure and Substrate-Binding Site

The architecture of SARS-CoV-2 M<sup>Pro</sup>/3CL<sup>Pro</sup> comprises three domains I (1–101 residues), II (102–184 residues), and domain III (201–301 residues), which is required for enzymatic activity (Bzowska et al., 2020). The co-crystal ligand (N3 inhibitor) bound inside the active site located in between the two anti-parallel  $\beta$ -barrel domains (domain I and domain II), and this domain contains the catalytic Cys-His dyad, which is mechanistically crucial for cutting the polyprotein precursors. Domain III is a five-helix bundle and is primarily responsible for the dimerization and linked with domain II by a long loop (185–200 AAs). We have inspected the interactions of ligands, available with recently solved crystal structures of M<sup>Pro</sup> (PDB ID: 6XMK, 6Y2G, 6Y2F, 6XFN, 6ZRT, 6ZRU, 7BQY, 7BRP, 7BUY, 6LZE, and 6M0K), and we noticed that all the co-crystal ligands have a similar binding mode (details provided in **Supplementary Table 1**). Among the structures, the structures reported by Dai et al. (2020) provide the inhibitors 11a and 11b, which strongly show the activity in the cell culture, but 11a tends to have a strong pharmacokinetic property and comes out as a strong drug candidate, by showing interactions with core important residues (Dai et al., 2020). The co-crystal ligand (PRD\_002214)-based FDA repurpose screening with docking and consensus ranking provides a similar binding mode of ritonavir with a SARS-CoV-2 M<sup>Pro</sup> active site (Cavasotto and Di Filippo, 2020). Likewise, Wang (2020) has reported that the FDA-approved compounds, namely, carfilzomib, eravacycline, valrubicin, lopinavir, and elbasvir, have the features to bind inside the SARS-CoV-2 M<sup>Pro</sup> active site (Wang, 2020). Ferraz et al. (2020) have reported three FDA-approved compounds, namely, bedaquiline, glibenclamide, and miconazole, through ligand and structure-based methods. They additionally stated



**FIGURE 2 | (A)** The RMSD graph for the entire timescale of the 1-microsecond ( $\mu$ s) molecular dynamics (MD) simulation shown for the apo protein (red) and the protein-ligand complex (black) of SARS-CoV-2 main protease ( $M^{Pro}$ /3CL $^{Pro}$ ) co-crystal structure (PDB ID:6LU7), for exploring the conformational landscapes. **(B,C)** Histogram representing the deviation points that occur in 1-microsecond ( $\mu$ s) molecular dynamics (MD) simulation, and here **(B)** represents the apo SARS-CoV-2  $M^{Pro}$ , while **(C)** represents the SARS-CoV-2  $M^{Pro}$  in complex with an inhibitor N3, PRD\_002214 from 6LU7 co-crystal structure.

that conformational changes occur in the S2-binding pocket, which must be considered for the drug design approach (Ferraz et al., 2020). The active site architecture includes four subsites (S1', S1, S2, and S4) which can accommodate the substrate recognition sequence (N-terminal—P4-P3-P2-P1↓P1'-P2'-P3'-C-terminal) and the cleavage between P1 and P1' residues resulting in proteolytic cleavage, thereby processing the viral polyproteins. The Gln residue in the P1 position plays a pivotal role in substrate recognition (S1 subsite) and highly specific only for the main protease of coronavirus and 3C protease of enterovirus. The active site includes the catalytic dyad (Cys145 and His41) and several other key residues (His163, His172, and Glu166) involved in opening gates to the active site (Zhang et al., 2020). The S2 and S4 subsites form deep and shallow hydrophobic pockets, respectively, accommodating residues in P2 and P4 positions with varied specificities. The residue in the P3 position is mostly solvent exposed, and because of the absence of the S3 subsite, it can tolerate any amino acid residue. The amides from amino acids Gly143, Cys145, and Ser144 form the cysteine protease's canonical "oxyanion hole". Recently, a report on another antiviral drug showed carmofur to inhibit the viral replication with the  $EC_{50}$  value of 24.30  $\mu$ M, by binding with Gly143 and Cys145. Here carmofur shows to be bound covalently with Cys145 and the ligand fatty acid tail region readily occupies the hydrophobic S2 subsite (Jin et al., 2020).

## Molecular Dynamics Simulation: Stage I

For this  $M^{Pro}$  structure in both apo and holo forms, MD simulations are performed for the long-scale simulation time of 1 microsecond (1  $\mu$ s). The holo form bound with the peptidomimetic N3 inhibitor (PRD\_002214) remains stable throughout the MD simulations for 1  $\mu$ s as shown in **Figure 2A** (black), and their RMSD value ranges between  $\sim$ 0.30 and  $\sim$ 0.35 nm. The absence of the peptidomimetic N3 inhibitor in the apo form shows high deviations and fluctuations in 1- $\mu$ s MD simulations. The initial 100 ns shows matching with holo forms, and from the 110<sup>th</sup> ns, the changes are seen in the loop regions of the apo form. At the 230<sup>th</sup> ns, the RMSD values reached up to  $\sim$ 0.8 nm and high fluctuations are seen in the 440<sup>th</sup> ns to the 700<sup>th</sup> ns. The apo form stabilizes at the 720<sup>th</sup> ns, and the RMSD values range between  $\sim$ 0.6 and  $\sim$ 0.8 nm until the end of the 1- $\mu$ s MD simulation as shown in **Figure 2A** (red). The values of RMSD for both apo and holo forms are plotted in histograms provided in **Figures 2B,C**. For the apo form (**Figure 2B**), the major count of the RMSD lies from 0.6 to 0.75 nm in between the range count of 2,000 and 3,000 times. However, for the holo form (**Figure 2C**), the major count of the RMSD lies from 0.28 to 0.36 nm in between the range count of 1,000 to 2,000 times approximately. This histogram result along with an RMSD graph shows that the apo state is highly flexible than the holo form of SARS-CoV-2  $M^{Pro}$ . For understanding the causative

**TABLE 1** | Statistical analysis of RMSD values obtained from the 1  $\mu$ s MD simulation for apo and holo forms of SARS-CoV-2 M<sup>Pro</sup>.

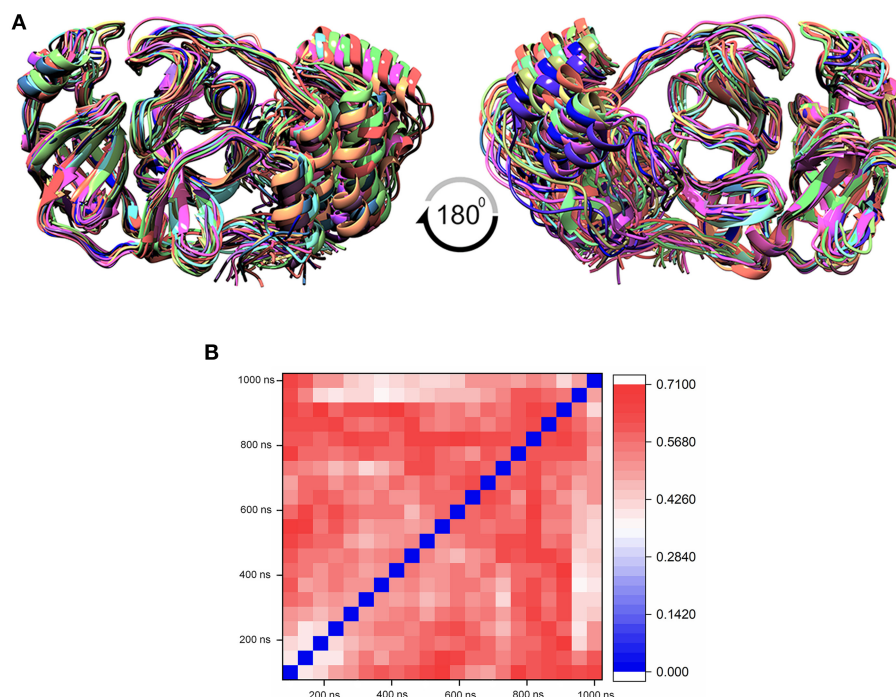
Content	Mean	Median	SD	Variance	Coefficient of variation	Min	Max
Apo protein (PDB ID: 6LU7)	0.61	0.64	0.11	0.01	0.19	0	0.89
Holo protein N3-bound complex	0.34	0.33	0.04	0.00	0.12	0	0.52

fluctuations with residue-wise participation in MD simulations, the root mean square fluctuation (RMSF) is calculated, and the values are plotted in **Supplementary Figure 2**. In addition, the active site residue values are plotted with 2D interactions in **Supplementary Figure 3** and the secondary structure is provided in **Supplementary Figure 4**. The secondary structure along with the RMSF plot shows the high fluctuations of residues in loop regions of the apo form, while those residues show few fluctuations in holo forms. The data suggest that the amino acids Phe140, Gly143, Cys145, His163, His164, Glu166, Gln189, and Thr190 are responsible for the stable RMSD of the holo form.

In addition, the statistical values are calculated to show the impact of ligand binding in SARS-CoV-2 M<sup>Pro</sup> and the values are provided in **Table 1**. The statistical data shows the mean difference of 0.27 and median difference of 0.31 between the apo and holo forms. Separating the higher half from the lower half of the data sample shows the median values of 0.64 and 0.33 for the apo and holo forms, respectively. The standard deviation (SD) of the holo form shows 0.04, while the apo form is 0.11, demonstrating the ligand-binding effect with the difference of 0.07. This statistical data narrates that the RMSD values in apo forms are widespread, and the holo forms show to be linear and stable and tend to be close to the mean of the set. In terms of variance, the apo form with high fluctuations leads to a spread-out with value 0.01, while the holo form shows fewer fluctuations leading to a spread close to value 0.001. The statistical variance difference of 0.011 shows that the apo form is more flexible and the ligand bound in M<sup>Pro</sup> tends to arrest the flexibility in the 1- $\mu$ s MD simulations. The range of apo lies between 0 (minimum) and 0.89 (maximum), and in this, 68% of the data lies in the interval between 0.49 and 0.73 and about 95% of the data lie in the interval between 0.37 and 0.84. However, for the holo form, the range lies between 0 (minimum) and 0.52 (maximum), about 68% of the data lies in the interval between 0.29 and 0.38, and about 95% of the data lie in the interval between 0.25 and 0.43. Since mode = 3 median – 2 mean and from **Table 1**, the mean and median are approximately equal for apo protein, and so the distribution for apo protein can be assumed to be approximately symmetrical. Similarly, the value distributions of holo forms are symmetrical but show more stability than the apo form. The coefficient of variation for apo form and holo form are 0.19 and 0.12, respectively, which shows that the ligand-bound form is consistent and the apo state is flexible.

## Ensemble Docking With Multiple Conformations of the Protein Structure

The MD simulation states that the loop regions are physically playing the role of protein flexibility. The structure visualization shows that the active site residues Phe140, Gly143, Cys145, Gln189, and Thr190 are present in the loop regions and those are mechanistically essential for protein function. Especially the residues between the 132 and 146 AA regions are unstructured loops and may undergo partial folding with the binding of the substrate to initiate the transition state. The conformational changes from SARS-CoV-2 M<sup>Pro</sup> are a core mechanism, which makes us consider the ensemble docking methods for virtual screening. For the ensemble docking approach, 21 conformations are taken as stated in the materials and methods, and those conformations are aligned, as shown in **Figure 3A**. Each conformation of the RMSD values is plotted in a heat map by applying the identity matrix method, as provided in **Figure 3B**. The heat map generated from 20  $\times$  20-based RMSD values visualizes all the pairwise correlations showing the RMSD values in the 0.36–0.71-nm range. The color codes of heat map in red indicate the values between 0.42 and 0.71 nm; blue color indicates values of 0–0.28 nm and the remaining mid regions in white. The heat map in **Figure 3B** shows a deep red color range from 0.55 to 0.71 nm, and the light red color below 0.50 nm indicates that a variety of conformations are adopted, illustrating the degree of conformational diversity. Thus, with a variety of conformations, along with average conformation, the 21 conformation-based grids are prepared and allowed to dock with a prepared TCM database. The crystal structure of SARS-CoV-2 M<sup>Pro</sup> complexed with an N3 inhibitor yields a docking score of –7.63 kcal/mol and binding energy of –52.38 kcal/mol. Based on this, we have assigned the scrutinization filter of choosing the compounds showing a minimum docking score of –8.00 kcal/mol and binding energy of –53.00 kcal/mol with all 21 conformations extracted from MD simulations. The final XP docking with the TCM database provides the compounds TCM 12495, TCM 24045, TCM 17404, and TCM 43709 with these filtering criteria, and another compound TCM 18935 shows potential but with one conformation, which shows the docking score –7.86 kcal/mol. Apart from these five compounds, few other hit compounds are eliminated due to lack of interactions with the core active site residues. The final average scoring values (mean value of 21 poses) of hit compounds and co-crystal ligand are provided in **Supplementary Table 2**. For the known compound, the average docking score is –7.99 kcal/mol, and the average binding energy is –60.172 kcal/mol, but the hit compounds have the tendency to surpass those values. Docking vs binding comparison is fitted in the linear fit model, and the  $R^2$  of the predicted model clearly represents that the compounds TCM 12495, TCM 24045, TCM 17404, TCM 43709, TCM 18935, and PRD\_002214 have the values of 0.0954, 0.0014, 0.2554, 0.2231, 0.2915, and 0.0839, respectively. There is a clear correlation seen with a docking score and binding energy, especially if the binding energy is below –60 kcal/mol, when the docking score ranges between –7 and –10 kcal/mol. For representation of the docking vs binding correlation, the 2D kernel display plot is provided in



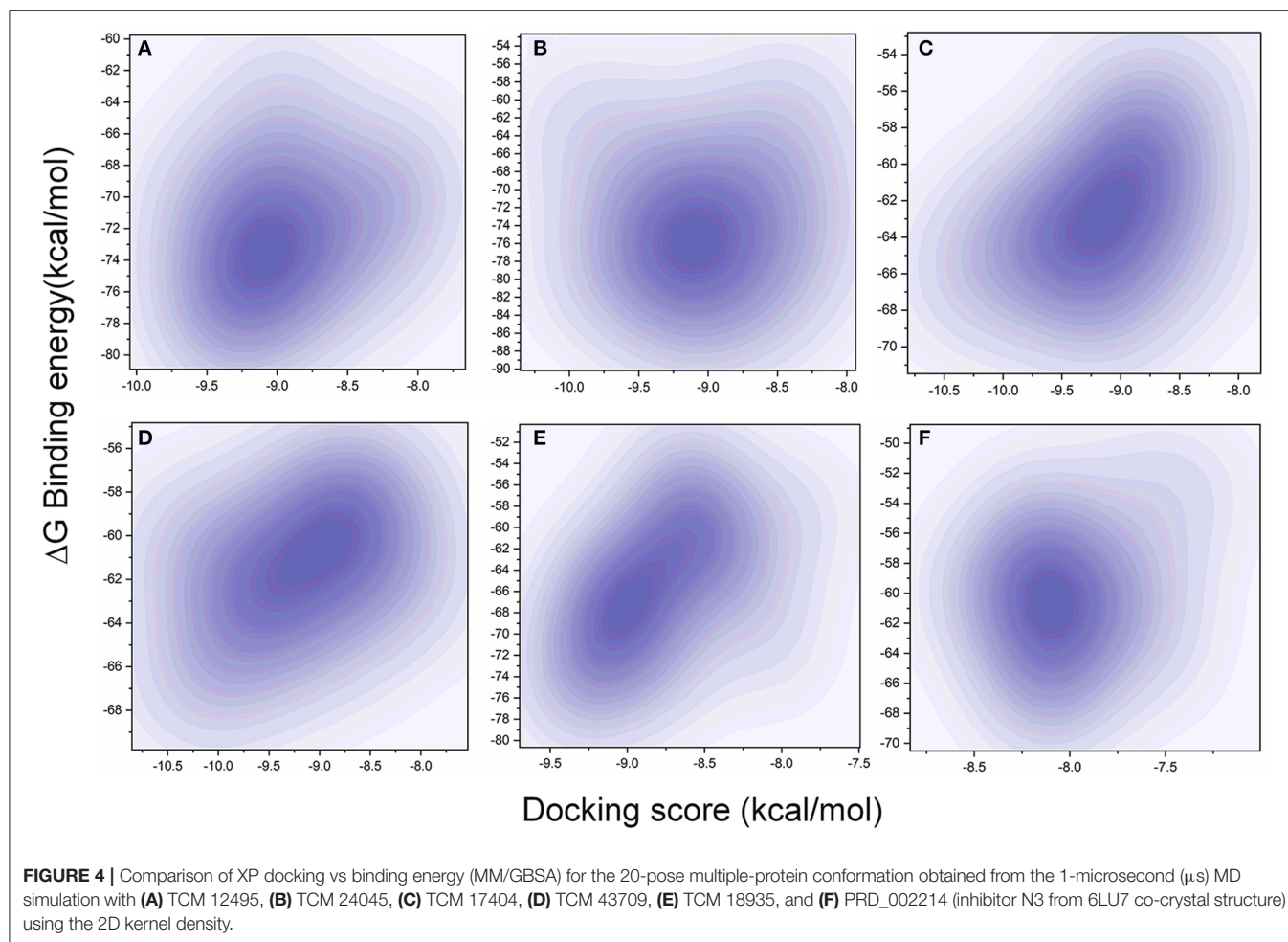
**FIGURE 3 | (A)** Aligned structure of the SARS-CoV-2 main protease ( $M^{pro}/3CL^{pro}$ ) obtained from various conformations of the MD simulation for the timescale of 1 microsecond ( $\mu$ s), and these multiple conformations are used for multiple-grid-based virtual screening. **(B)** Heatmap matrix for RMSD variation poses obtained from each 50-ns interval for the 1- $\mu$ s MD simulation for SARS-CoV-2  $M^{pro}$ .

**Figure 4.** The deep-blue regions in **Figures 4A–F** represent the zone of higher density (ZHD) for the correlation of docking score and binding energy. The ZHDs of hit compounds TCM 12495, TCM 24045, TCM 17404, TCM 43709, and TCM 18935 range between  $-8.9$  and  $-9.3$  kcal/mol,  $-8.7$  and  $-9.3$  kcal/mol,  $-8.8$  and  $9.4$  kcal/mol,  $-8.7$  and  $-9.4$  kcal/mol, and  $-8.9$  and  $-9.2$  kcal/mol, which directly impose the binding energy density zone with the range of  $-72$  to  $-76$  kcal/mol,  $-72$  to  $-78$  kcal/mol,  $-61$  to  $-65$  kcal/mol,  $-59$  to  $-63$  kcal/mol, and  $-64$  to  $-71$  kcal/mol, respectively. While the N3 bound holo forms show the ZHD range of  $-7.9$  to  $-8.2$  kcal/mol in docking score and  $-59$  to  $-63$  kcal/mol in binding energy, it clearly represents the hit compounds ZHD is energetic than the available co-crystal ligand.

### Induced Fit Docking (IFD)

While the ensemble docking approach provides better hit compounds than the co-crystal ligand, those hit compounds are redocked with the IFD method, to obtain the best pose among the multiple conformations. This IFD is purposefully performed with the average conformation (**Supplementary Figure 5**) of the 1- $\mu$ s MD simulation, for avoiding the limitation of partial flexibility in XP docking. The IFD approach provides up to 30 possible poses per ligand and ranked based on best poses. For that, the best hit compounds (2D information available in **Supplementary Figure 6**) and co-crystal ligand are flexibly allowed to interact with average conformation of SARS-CoV-2  $M^{pro}$ . The IFD results seem to be interesting by showing

strong bonding interactions with functionally important active sites, with strong scoring values. Like XP docking, the IFD provides the scoring values, represented in **Table 2**, indicating that the hit compounds surpass the available co-crystal ligand. The IFD-based docking score for the compounds TCM 12495, TCM 24045, TCM 17404, TCM 43709, TCM 18935, and PRD\_002214 are  $-16.81$ ,  $-12.40$ ,  $-17.73$ ,  $-14.95$ ,  $-15.07$ , and  $-10.05$  kcal/mol, respectively. Similarly, the IFD scores are above  $-660$  kcal/mol for both co-crystal and hit compounds, and the best pose of interactions is shown in **Figures 5a–f**. The changes between XP docking and IFD are seen in scoring values along with the interactions. The co-crystal compounds show 8 hydrogen interactions with SARS-CoV-2  $M^{pro}$  with the main support of residues Asn142 and Gly143. The hit compound, namely, TCM 17404, can form 11 hydrogen bond interactions with Asn142, Gly143, and Cys145. The other hit compound interactions also show a strong binding with the active sites; the details of interactions are provided in **Supplementary Table 3**. We expect the hit compounds to interact with Phe140, Gly143, Cys145, His163, His164, Glu166, Gln189, and Thr190 residues in resembling the co-crystal ligand binding. However, the hit compounds interact with core functional residues His41, Gly143, and Cys145, along with Thr26, Ser46, Tyr54, Asn119, Gly138, Phe140, Leu141, Asn142, His163, His164, Glu166, Pro168, Gly170, His172, Gln189, and Thr190. For conforming these residues' involvement in experimental structures, these residues are cross-checked with available crystal structure interactions



**TABLE 2 |** IFD scores of new and known compounds with an average structure of  $M^{Pro}/3CL^{Pro}$  obtained from 1  $\mu$ s MD simulation.

Compound name/ID*	Docking score (kcal/mol)	IFD score (kcal/mol)	No of H. bonds	Atomic interactions	$\pi$ - $\pi$ interaction
TCM 12495	-16.8	-671.8	07	Supplementary Table 3	0
TCM 24045	-12.4	-669.0	08		1
TCM 20302	-17.7	-667.6	11		0
TCM 43709	-14.9	-666.4	06		2
TCM 18935	-15.0	-662.2	09		0
PRD_002214	-10.0	-667.1	08		0

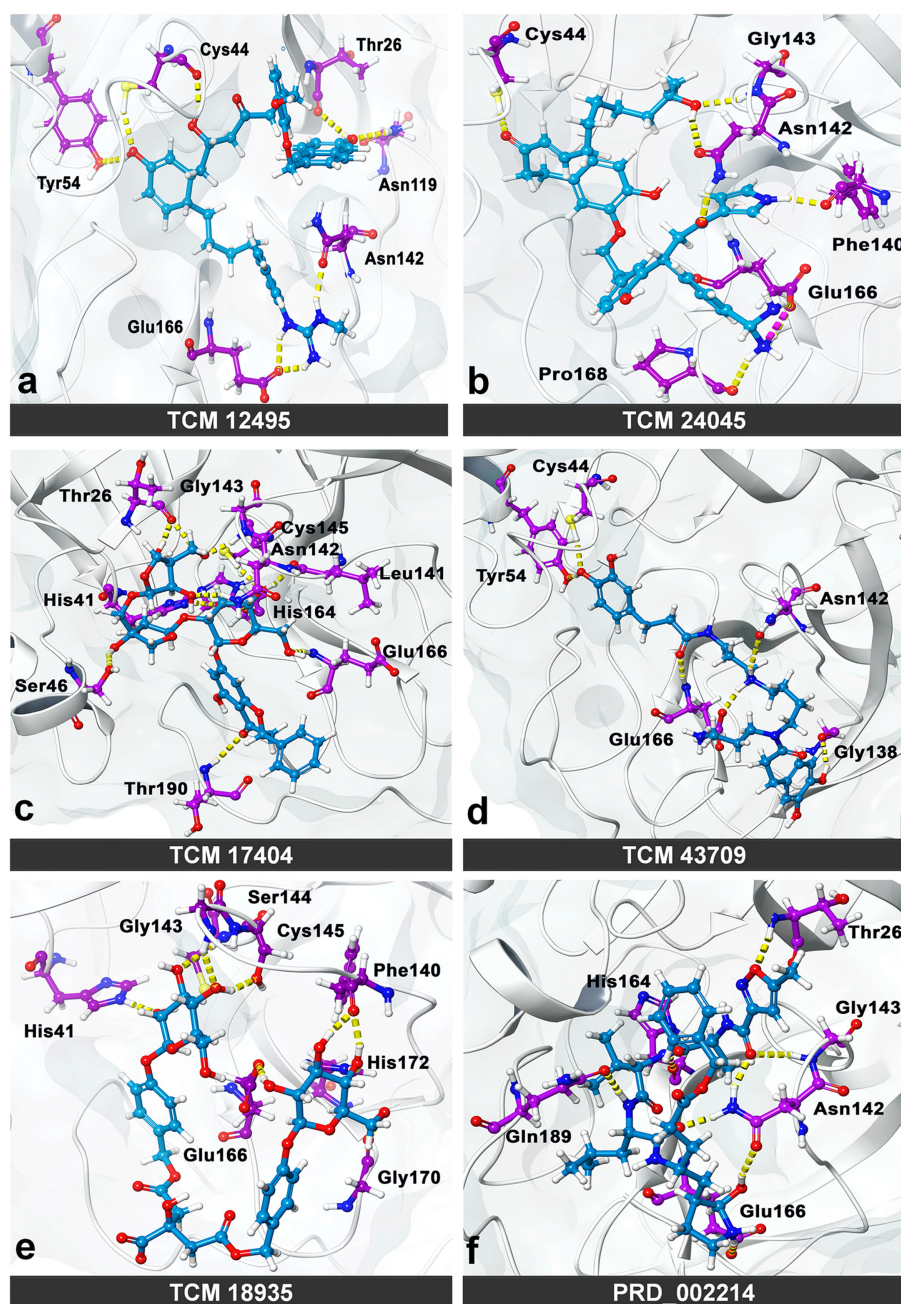
provided in **Supplementary Table 1**. Interestingly, the residues interacted with hit compounds are experimentally reported to have the bonding interactions.

## Molecular Dynamics Simulation: Stage II

The IFD complex shows prominence in binding and scoring parameters, and those final poses are simulated for 100 ns (stage II MD) for understanding the ligand stability in dynamic

state. Here the IFD best pose for hit compound and co-crystal ligand stability is concerned, as the whole workflow relies on multiple protein conformations. This ligand stability in the 100-ns MD simulations can conform to the hit compound binding efficiency in the dynamic state. The RMSD plot for the ligand-bound complex is provided in **Supplementary Figure 7**, which shows that only the TCM 17404 complex shows the RMSD above 0.4 nm. Except the TCM 17404 complex, all the other hit compounds and co-crystal bound ligands show the RMSD values below 0.4 nm. Each snapshot deviation is calculated and plotted in **Figure 6**; the color codes of blue represent the values below 0.25 nm, those in red color represent the values above 0.35 nm, and the range between 0.25 and 0.34 nm is represented in white. The hit compound TCM 17404, showing the deviations from the 5<sup>th</sup> to 100<sup>th</sup> ns with dominant red color, indicates higher RMSD values. The hit compounds TCM 24045 and TCM 43709 and the co-crystal ligand (PRD\_002214) complex show a moderate red color, as the RMSD lies between 0.25 and 0.4 nm. The other hit compound, TCM 18935 complex, has shown limited deviations, with blue and white color lines indicating the RMSD ranges between 0.24 and 0.36 nm.

For understanding these RMSD values, the statistical methods are applied and incorporated in **Table 3**. Even though the range

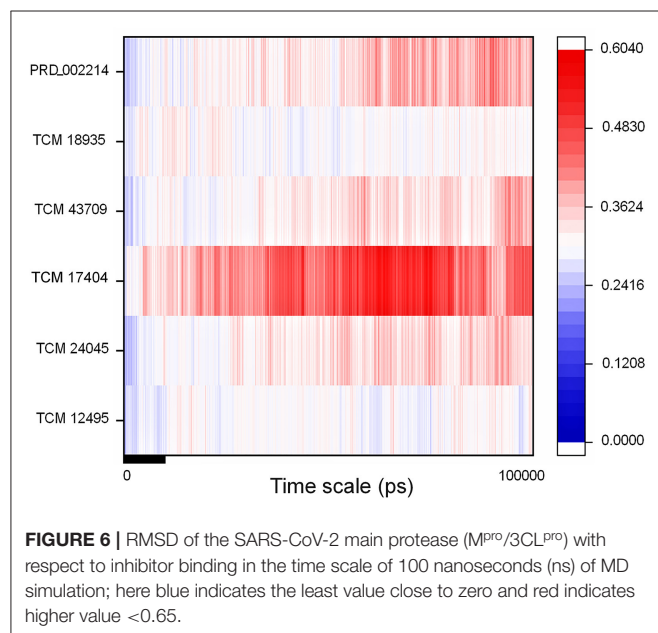


**FIGURE 5 |** Molecular interaction of compounds screened from the TCM database: (a) TCM 12495, (b) TCM 24045, (c) TCM 17404, (d) TCM 43709, (e) TCM 18935, and (f) PRD\_002214 (inhibitor N3 from 6LU7) in interaction with average conformation obtained from 1 microsecond ( $\mu$ s) of MD simulations through the induced fit docking (IDF) method.

(min to max) of compounds, namely, TCM 12495 (a), TCM 24045 (b), TCM 17404 (c), TCM 43709 (d), TCM 18935 (e), and PRD\_002214 (f) are 0.1–0.38, 0.–0.43, 0.1–0.60, 0.1–0.47, 0.1–0.38, and 0.1–0.49, respectively, about 95% of the data of compounds a–f, lies in the interval 0.25–0.33, 0.26–0.39, 0.30–0.57, 0.25–0.39, 0.26–0.34, and 0.25–0.43, respectively. The mean, median, and mode are approximately equal for all

compounds except for compound c. This shows that, except compound c, all the value distributions of the compounds can be assumed to be approximately symmetrical. The compound c mode is lesser than the mean and median, so that the value distribution of compound c is skewed to the high deviation. On analyzing the standard deviation of the RMSD values of the protein–ligand complex, all the compounds show

the values approximately equal to zero. Through this, we understood that 100-ns MD simulations produced RMSD values approximately identical, which is equal to its mean value. The coefficients of variations for compounds a, b, c, d, e, and f are 0.06, 0.10, 0.15, 0.10, 0.06, and 0.13 nm, respectively. In this, compounds a (TCM 12495) and e (TCM 18935) are holding smaller coefficients of variations than the others. So,



**TABLE 3 |** Statistical analysis of RMSD values obtained from the 100 ns of MD simulation for screened compounds for comparing the stability with co-crystal compound PRD\_002214.

Compound	Mean	Median	Mode	SD	Variance	Coefficient of variation	Min	Max
TCM 12495	0.29	0.30	0.28	0.02	0.00	0.06	0.1	0.38
TCM 24045	0.32	0.33	0.33	0.03	0.00	0.10	0.1	0.43
TCM 20302	0.43	0.44	0.32	0.06	0.00	0.15	0.1	0.60
TCM 43709	0.32	0.32	0.28	0.03	0.00	0.10	0.1	0.47
TCM 18935	0.30	0.30	0.28	0.01	0.00	0.06	0.1	0.38
PRD_002214	0.34	0.33	0.27	0.04	0.00	0.13	0.1	0.49

**TABLE 4 |** Energy values obtained from the MM/PBSA script for the screened compounds and co-crystal compound for the simulation timescale of 100 ns.

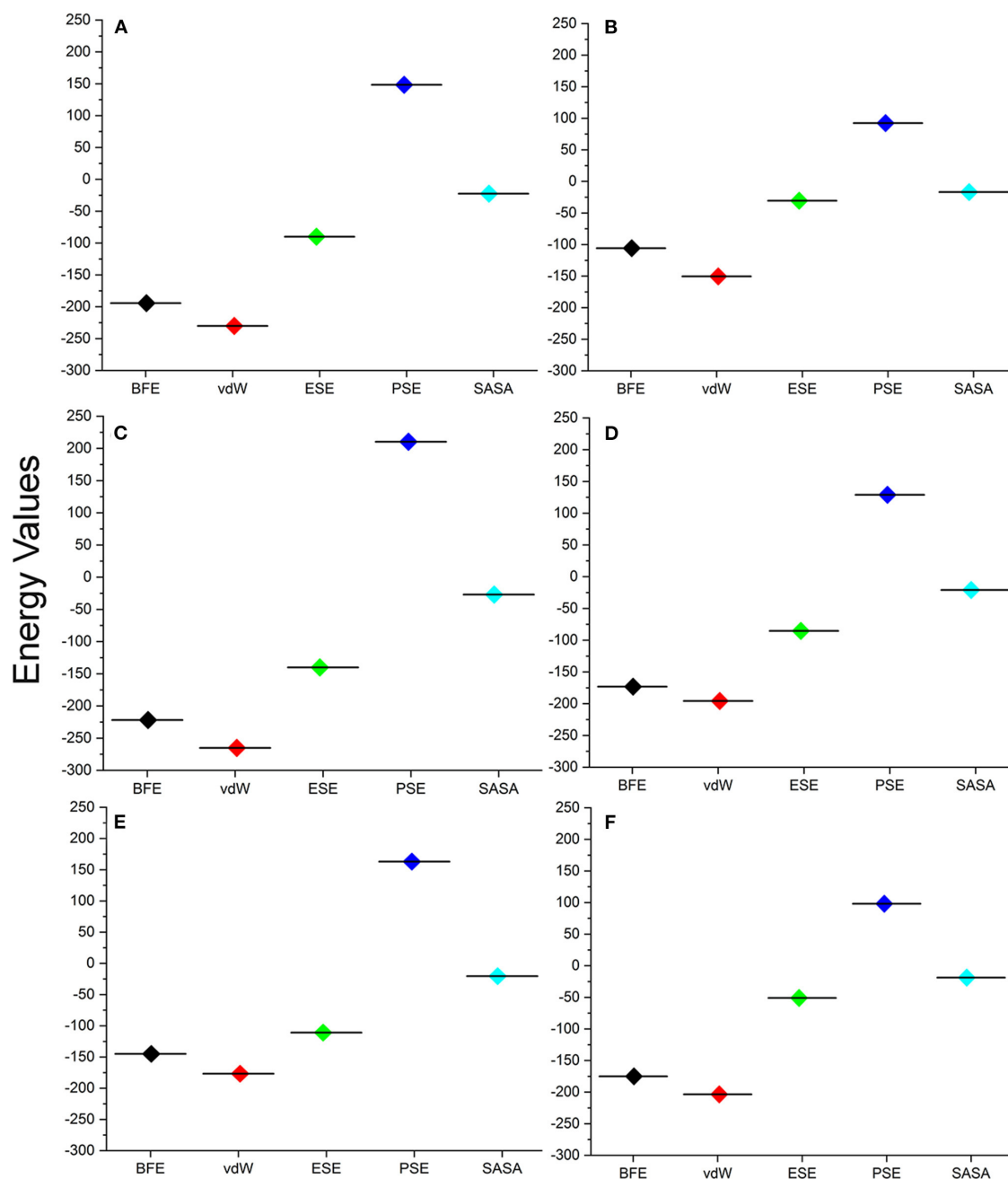
S. No.	Compound ID	Binding energy (kJ/mol)	van der Waal energy (kJ/mol)	Electrostatic energy (kJ/mol)	Polar solvation energy (kJ/mol)	SASA energy (kJ/mol)
1	TCM12495	$-194.2 \pm 33.0$	$-230.0 \pm 27.7$	$-90.0 \pm 33.8$	$148.4 \pm 20.4$	$-22.5 \pm 2.4$
2	TCM24045	$-105.7 \pm 25.6$	$-150.4 \pm 22.8$	$-30.6 \pm 22.5$	$92.3 \pm 24.3$	$-16.9 \pm 2.9$
3	TCM20302	$-221.9 \pm 35.5$	$-265.1 \pm 26.5$	$-140.2 \pm 67.4$	$210.3 \pm 45.5$	$-26.9 \pm 2.1$
4	TCM43709	$-173.1 \pm 51.3$	$-195.7 \pm 50.2$	$-85.3 \pm 48.6$	$128.9 \pm 31.9$	$-20.9 \pm 4.7$
5	TCM18935	$-144.9 \pm 47.5$	$-176.6 \pm 24.6$	$-110.9 \pm 63.6$	$163.0 \pm 41.0$	$-20.4 \pm 2.6$
6	PRD_002214	$-175.1 \pm 20.2$	$-203.5 \pm 17.6$	$-50.9 \pm 19.8$	$98.1 \pm 15.1$	$-18.7 \pm 1.6$

Units (kJ/mol).

compounds a and e can be more consistent than the others. For validation of protein–ligand interaction, various energy values are analyzed which contributed to the efficient binding. Through MM/PBSA calculations, binding energy (BFE), van der Waals (vdW), electrostatic energy (ESE), polar solvent energy (PSE), and SASA energy are predicted and the values are provided in **Table 4** and **Figures 7A–F**. From the screened compounds, TCM 12495, TCM 17404, and TCM 43709 show lower binding energy values of  $-194.262$  kJ/mol,  $-221.947$  kJ/mol, and  $-173.163$  kJ/mol, respectively. This shows that the compounds TCM 12495, TCM 17404, and TCM 43709 have a higher stability in the 100-ns MD simulations. Other compounds, namely, TCM 24045 and TCM 18935, also show moderate binding energy that resembles known co-crystal ligands. For attaining these energy levels, the other energy parameters like van der Waals (vdW), electrostatic energy (ESE), and SASA energy have contributed to supporting the ligand stability. In comparing all the energies, the polar solvation energy (PSE) is the only energy which contributed positively to the total binding free energy. Thus, the predicted binding energy strongly supports the binding interactions of an effective compound with the targeted protein. The validation of the virtual screening-based docking protocol was evaluated using the enrichment calculation method with EF, receiver operating characteristic (ROC), and Boltzmann-Enhanced Discrimination of the receiver operating characteristic (BEDROC) metrics. ROC scores is received with 0.98 RIE of 15.18, respectively, which represents the higher-ranking order of active compounds based on quality. Because ROC score with “ $\geq 0.7$ ” signifies a satisfied metric value to define the highest precision and predicting skill of virtual screening and docking protocol defined. **Supplementary Table 4** shows the enrichment metric value, and the graphical representation enrichment curve in **Supplementary Figure 8** represents the quality of retrieval of known actives from the external database, which were ranked to decoys for evaluation.

## DISCUSSION

The enzyme SARS-CoV-2  $M^{pro}$  is known to control the activity of the replication mechanism and is designated as an attractive target for inhibition. We have applied the MD simulations for elucidating the functional active site role in  $M^{pro}$  along with



**FIGURE 7 |** Energy values obtained from the MM/PBSA script for the screened compounds [(A) TCM 12495, (B) TCM 24045, (C) TCM 17404, (D) TCM 43709, (E) TCM 18935] and from inhibitor N3 (F) from the SARS-CoV-2 co-crystal structure, PDB ID:6LU7 for the simulation timescale of 100 ns.

conformational changes, ligand-binding effect, and stability with the apo and holo forms. Long-scale MD simulation shows that the apo form of SARS-CoV-2 M<sup>Pro</sup> shows higher fluctuations than the holo form. The whole conformational heterogeneity shown by the binding site loop in the apo form persists to a large degree of deviations in the 1- $\mu$ s MD simulations. The RMSF and statistical analysis shows that the functional residues in apo form are key to representing the fluctuations in the MD simulations. For the holo forms, those functional residues grasp the ligand

molecule, steady its positions, and show stability throughout the MD simulations. Yoshino *et al.* have also reported the high-end MD simulations, to show the importance of the roles of His41, Gly143, and Gly166 in peptide functional groups (Yoshino *et al.*, 2020). Similarly, the co-crystal ligand (PRD\_002214) shows the binding sites of Phe140, Gly143, Cys145, His163, His164, Glu166, Gln189, and Thr190, and from this, except for Phe140 and Gly143, all the other amino acids are polar and charged. These charged residues play a vital role in electron transfer

and modulate the overall electrostatic balance of the protein and its binding. The functional behavior of apo protein which leads to multiple conformations is controlled by binding of a suitable inhibitor between the two anti-parallel  $\beta$ -barrel domains that result in stable protein-ligand MD simulations. This may be due to structural changes occurring in the active site loop, also considering that we applied ensemble docking methods to employ the identification of new molecules.

Thereby the MD trajectories are processed to obtain an ensemble of distant conformations rather than a minimally fluctuating single global-minima structure. Perhaps an active site is located in between the two anti-parallel  $\beta$ -barrel domains, especially the loop regions as a target region of action for inhibitory molecules, which contains the catalytic Cys-His dyad. Here the electrostatic, stabilized Cys145 is a potent nucleophile linked with the nearby histidine amino acid along with the amide backbone of Cys145 and Gly143 which stabilize the oxyanion hole in the transition state formation. The His41 imidazole ring places the active site linked with Cys145 through the stabilization of thiolate ion. Nucleophilic attack of the anionic cysteine S (thiolate ion) occurs on the peptide carbonyl carbon. In this step, a fragment of the substrate is released with an (amino) amine group in the terminus, the histidine residue in the protease is restored to its deprotonated form, and a thioester intermediate linking the new carboxy-terminus of the substrate to the cysteine thiol is formed. Thus, finding an appropriate compound, which binds in the binding site residues of His41, Gly143, and Cys145, will block the substrate-binding activity. Considering the catalytic activity of His41, Gly143, and Cys145, the multiple conformation-based ensembles docking with filtering criteria of docking score of  $-8.00$  kcal/mol and binding energy of  $-53.00$  kcal/mol readily which threw most of the compounds from the TCM database are executed. From the whole TCM database, our screening scrutinization filters actively pass only five compounds, namely, TCM 12495, TCM 24045, TCM 17404, TCM 43709, and TCM 18935. Ensemble conformations obtained from different intervals show variations in pose and space in the binding site and may also have a functional difference. The successive compounds TCM 12495, TCM 24045, TCM 17404, TCM 43709, and TCM 18935 hold the tendency to bind multiple conformations which adapt well inside the protease active site located in between the two anti-parallel  $\beta$ -barrel domains, like co-crystal ligand (PRD\_002214) binding. Thus, we believe that these compounds can affect the  $M^{pro}$  functional mechanism, by interacting with core active site residues.

The residual interactions of new leads in the flexible environment are shown using the re-docking method of IFD, which provides the strong support for new compounds, which are bound more efficiently than the co-crystal ligand. Normal XP docking with multiple protein conformations followed by allowing adjustments to the receptor conformation through the soft receptor approach using flexible side chains or IFD shows huge improvement. Through IFD, we have seen new compounds from the screening able to bind with functional catalytic activity residues His41, Gly143, and Cys145. Ensemble docking with multiple conformation proteins followed by IFD-derived ensembles provides high benefit of success in virtual

screening. Through this, the shortlisted TCM compounds from this study and co-crystal ligand bind to the substrate-binding site in similar binding mode and will efficiently inhibit SARS-CoV-2  $M^{pro}$  activity as shown in **Supplementary Figure 9** (Dai et al., 2020; Jin et al., 2020; Zhang et al., 2020).

The screened compounds dynamic behavior is analyzed for low energy profiles using MM/PBSA, which is used for post-processing of docked structures along with the reliability of compound binding inside the flexible binding pocket. The 100 ns of protein-ligand complex simulation along with MM-PBSA binding free energy suggests that lead molecules perfectly fit in the binding site and are structurally stable with a low energy profile. Available literature suggests that the compounds with lower energy profiles in MM/PBSA are suitable candidates for further experimental analysis. In addition, the external validation with random decoy set along with the screened compounds favors the top screened compounds by positioning those compounds in the top with the ROC curve score of 0.98, which clearly shows that the screened compounds are best among the random decoy set. The successive compounds from this work are readily available as the TCM herbal compounds are believed to be non-toxic or less toxic and have been used to treat numerous kinds of diseases for more than 2,000 years in eastern Asian countries (Yuan et al., 2016). By interest, we searched the source of the screened compounds and found that the compound TCM 12495 is from the herb *Lantana camara*, and compound TCM 17404 is isolated from the herb *Viscum angulatum* (Xu et al., 2019). However, the source of TCM 18935, TCM 24045, and TCM 43709 are not available in the literature.

## CONCLUSION

Overall, this study provides a comprehensive structural analysis of the SARS-CoV-2 main protease ( $M^{pro}$ /3CL $^{pro}$ ) substrate-binding pocket for the purpose of inhibitor screening and design. Through this work, we have disclosed a long-range 1- $\mu$ s MD simulation for the apo and holo forms of SARS-CoV-2  $M^{pro}$ . The MD simulations revealed conformational changes in the active site loop regions and, accounting those changes into the screening, provide a strong, potential compound against SARS-CoV-2  $M^{pro}$ . Small molecules targeting this binding pocket should have the ability to interact with residues His41, Gly143, and Cys145; disturbing the formation of the oxyanion hole can lead to its inhibition. The final antiviral inhibitors screened from the world's largest traditional Chinese medicine database (TCM@Taiwan) have robust scoring values that have been evaluated from the theoretical, statistical, and internal motion of atoms in dynamic status. These prospective compounds are from natural resources and used as medicine for several years, and this incorporating international effort can bring these compounds to light as suitable drug candidates against the COVID-19.

## DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories

and accession number(s) can be found in the article/**Supplementary Material**.

## AUTHOR CONTRIBUTIONS

CS, VKD, and SKS conceived and designed the study. CS and UP performed the screening and MD simulations. CS, DCD, EB, PS, VKD, and SKS analyzed the results. CS and DCD prepared the manuscript. All authors read and approved the final manuscript.

## FUNDING

This study was supported by a RUSA-Phase 2.0 Policy (TNmulti-Gen), Dept. of Edn, Govt. of India (Grant No. F.24-51/2014-U), and part of the work is supported by a DST-sponsored SERB project (CVD/2020/000031), India.

## ACKNOWLEDGMENTS

CS and SKS thankfully acknowledge RUSA-Phase 2.0 Policy (TNmulti-Gen), Dept. of Edn, Govt. of India (Grant No:

F.24-51/2014-U). DCD and EB acknowledge the European Regional Development Fund; OP RDE; Project: Chemical biology for drugging undruggable targets (ChemBioDrug) (No. CZ.02.1.01/0.0/0.0/16\_019/0000729); Czech Academy of Sciences Post-doctoral Fellowship (No. L200551951 Programu podpory perspektivních lidských zdroju–postdoktorandu); and Academy of Sciences of the Czech Republic (RVO: 61388963). VKD thankfully acknowledged DST, India, for the SERB project grant (CVD/2020/000031). The authors thank Mr. G. Suresh Kumar, Assistant Professor, Madurai Kamaraj University, for his immense support in mathematical calculation. Special thanks are given to Marisol Vierra and Dr. Satish Selvaraj Venkatesh (Prifer, India) for proofreading the manuscript. UP thanks to Indian Council of Medical Research (ISRM/11/(19)/2017, dated: 09.08.2018) for providing ICMR-SRF (Senior Research Fellowship).

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fchem.2020.595273/full#supplementary-material>

## REFERENCES

- Aldeghi, M., Heifetz, A., Bodkin, M. J., Knapp, S., and Biggin, P. C. (2017). Predictions of ligand selectivity from absolute binding free energy calculations. *J. Am. Chem. Soc.* 139, 946–957. doi: 10.1021/jacs.6b11467
- Amaro, R. E., Baudry, J., Chodera, J., Demir, O., Mccammon, J. A., Miao, Y., et al. (2018). Ensemble docking in drug discovery. *Biophys. J.* 114, 2271–2278. doi: 10.1016/j.bpj.2018.02.038
- Ayres, J. S. (2020). A metabolic handbook for the COVID-19 pandemic. *Nat Metab.* 2, 572–585. doi: 10.1038/s42255-020-0237-2
- Battisti, V., Wieder, O., Garon, A., Seidel, T., Urban, E., and Langer, T. (2020). A Computational approach to identify potential novel inhibitors against the coronavirus SARS-CoV-2. *Mol. Inform.* 39:e2000090. doi: 10.1002/minf.202000090
- Bzowka, M., Mitusinska, K., Raczynska, A., Samol, A., Tuszyński, J. A., and Gora, A. (2020). Structural and evolutionary analysis indicate that the SARS-CoV-2 mpro is a challenging target for small-molecule inhibitor design. *Int. J. Mol. Sci.* 21:3099. doi: 10.3390/ijms21093099
- Cavasotto, C. N., and Di Filippo, J. I. (2020). *In silico* drug repurposing for COVID-19: targeting SARS-CoV-2 proteins through docking and consensus ranking. *Mol. Inform.* doi: 10.1002/minf.202000115
- Chen, C. Y. (2011). TCM Database@Taiwan: the world's largest traditional Chinese medicine database for drug screening *in silico*. *PLoS ONE* 6:e15939. doi: 10.1371/journal.pone.0015939
- Childers, M. C., and Daggett, V. (2018). Validating molecular dynamics simulations against experimental observables in light of underlying conformational ensembles. *J. Phys. Chem. B* 122, 6673–6689. doi: 10.1021/acs.jpcc.8b02144
- Chinnasamy, S., Selvaraj, G., Kaushik, A. C., Kalliamurthi, S., Chandrabose, S., Singh, S. K., et al. (2020a). Molecular docking and molecular dynamics simulation studies to identify potent AURKA inhibitors: assessing the performance of density functional theory, MM-GBSA and mass action kinetics calculations. *J. Biomol. Struct. Dyn.* 38, 4325–4335. doi: 10.1080/07391102.2019.1674695
- Chinnasamy, S., Selvaraj, G., Selvaraj, C., Kaushik, A. C., Kalliamurthi, S., Khan, A., et al. (2020b). Combining *in silico* and *in vitro* approaches to identification of potent inhibitor against phospholipase A2 (PLA2). *Int. J. Biol. Macromol.* 144, 53–66. doi: 10.1016/j.ijbiomac.2019.12.091
- F.24-51/2014-U). DCD and EB acknowledge the European Regional Development Fund; OP RDE; Project: Chemical biology for drugging undruggable targets (ChemBioDrug) (No. CZ.02.1.01/0.0/0.0/16\_019/0000729); Czech Academy of Sciences Post-doctoral Fellowship (No. L200551951 Programu podpory perspektivních lidských zdroju–postdoktorandu); and Academy of Sciences of the Czech Republic (RVO: 61388963). VKD thankfully acknowledged DST, India, for the SERB project grant (CVD/2020/000031). The authors thank Mr. G. Suresh Kumar, Assistant Professor, Madurai Kamaraj University, for his immense support in mathematical calculation. Special thanks are given to Marisol Vierra and Dr. Satish Selvaraj Venkatesh (Prifer, India) for proofreading the manuscript. UP thanks to Indian Council of Medical Research (ISRM/11/(19)/2017, dated: 09.08.2018) for providing ICMR-SRF (Senior Research Fellowship).
- Culetta, G., Gulott, M. R., Perricone, U., Zappal, M., Almerico, A. M., and Tutone, M. (2020). Exploring the SARS-CoV-2 Proteome in the search of potential inhibitors via structure-based pharmacophore modeling/docking approach. *Computation* 8:77. doi: 10.3390/computation8030077
- Dai, W., Zhang, B., Jiang, X. M., Su, H., Li, J., Zhao, Y., et al. (2020). Structure-based design of antiviral drug candidates targeting the SARS-CoV-2 main protease. *Science* 368, 1331–1335. doi: 10.1126/science.abb4489
- Eleftheriou, P., Amanatidou, D., Petrou, A., and Geronikaki, A. (2020). *In silico* evaluation of the effectivity of approved protease inhibitors against the main protease of the novel SARS-CoV-2 virus. *Molecules* 25:2529. doi: 10.3390/molecules25112529
- Evangelista Falcon, W., Ellingson, S. R., Smith, J. C., and Baudry, J. (2019). Ensemble docking in drug discovery: how many protein configurations from molecular dynamics simulations are needed to reproduce known ligand binding? *J. Phys. Chem. B* 123, 5189–5195. doi: 10.1021/acs.jpcc.8b11491
- Fang, S. G., Shen, H., Wang, J., Tay, F. P., and Liu, D. X. (2008). Proteolytic processing of polyproteins 1a and 1ab between non-structural proteins 10 and 11/12 of Coronavirus infectious bronchitis virus is dispensable for viral replication in cultured cells. *Virology* 379, 175–180. doi: 10.1016/j.virol.2008.06.038
- Ferraz, W. R., Gomes, R. A., Al, S. N., and Goulart Trossini, G. H. (2020). Ligand and structure-based virtual screening applied to the SARS-CoV-2 main protease: an *in silico* repurposing study. *Future Med. Chem.* 12, 1815–1828. doi: 10.4155/fmc-2020-0165
- Frances-Monerris, A., Hognon, C., Miclot, T., Garcia-Iriepa, C., Iriepa, I., Terenzi, A., et al. (2020). Molecular basis of SARS-CoV-2 infection and rational design of potential antiviral agents: modeling and simulation approaches. *J. Proteome Res.* 19, 4291–4315. doi: 10.1021/acs.jproteome.0c00779
- Friesner, R. A., Murphy, R. B., Repasky, M. P., Frye, L. L., Greenwood, J. R., Halgren, T. A., et al. (2006). Extra precision glide: docking and scoring incorporating a model of hydrophobic enclosure for protein-ligand complexes. *J. Med. Chem.* 49, 6177–6196. doi: 10.1021/jm051256o
- Gahlawat, A., Kumar, N., Kumar, R., Sandhu, H., Singh, I. P., Singh, S., et al. (2020). Structure-based virtual screening to discover potential lead molecules for the SARS-CoV-2 main protease. *J. Chem. Inf. Model.* 4:acs.jcim.0c00546. doi: 10.1021/acs.jcim.0c00546

- Gani, O. A., Narayanan, D., and Engh, R. A. (2013). Evaluating the predictivity of virtual screening for ABL kinase inhibitors to hinder drug resistance. *Chem. Biol. Drug Des.* 82, 506–519. doi: 10.1111/cbdd.12170
- Gao, J., Zhang, L., Liu, X., Li, F., Ma, R., Zhu, Z., et al. (2020). Repurposing low-molecular-weight drugs against the main protease of severe acute respiratory syndrome coronavirus 2. *J. Phys. Chem. Lett.* 11, 7267–7272. doi: 10.1021/acs.jpclett.0c01894
- Gil, C., Ginex, T., Maestro, I., Nozal, V., Barrado-Gil, L., Cuesta-Geijo, M. A., et al. (2020). COVID-19: drug targets and potential treatments. *J. Med. Chem.* doi: 10.1021/acs.jmedchem.0c00606
- Goyal, B., and Goyal, D. (2020). Targeting the dimerization of the main protease of coronaviruses: a potential broad-spectrum therapeutic strategy. *ACS Comb. Sci.* 22, 297–305. doi: 10.1021/acscmbosci.0c00058
- Halgren, T. A., Murphy, R. B., Friesner, R. A., Beard, H. S., Frye, L. L., Pollard, W. T., et al. (2004). Glide: a new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening. *J. Med. Chem.* 47, 1750–1759. doi: 10.1021/jm030644s
- Hanney, S. R., Wooding, S., Sussex, J., and Grant, J. (2020). From COVID-19 research to vaccine application: why might it take 17 months not 17 years and what are the wider lessons? *Health Res. Policy Syst.* 18:61. doi: 10.1186/s12961-020-00571-3
- Hilgenfeld, R. (2014). From SARS to MERS: crystallographic studies on coronavirus proteases enable antiviral drug design. *FEBS J.* 281, 4085–4096. doi: 10.1111/febs.12936
- Jeong, G. U., Song, H., Yoon, G. Y., Kim, D., and Kwon, Y. C. (2020). Therapeutic strategies against COVID-19 and structural characterization of SARS-CoV-2: a review. *Front. Microbiol.* 11:1723. doi: 10.3389/fmicb.2020.01723
- Jin, Z., Du, X., Xu, Y., Deng, Y., Liu, M., Zhao, Y., et al. (2020). Structure of M(pro) from SARS-CoV-2 and discovery of its inhibitors. *Nature* 582, 289–293. doi: 10.1038/s41586-020-2223-y
- Kalyaanamoorthy, S., and Chen, Y. P. (2014). A steered molecular dynamics mediated hit discovery for histone deacetylases. *Phys. Chem. Chem. Phys.* 16, 3777–3791. doi: 10.1039/c3cp5511h
- Klimovich, P. V., and Mobley, D. L. (2015). A Python tool to set up relative free energy calculations in GROMACS. *J. Comput. Aided Mol. Des.* 29, 1007–1014. doi: 10.1007/s10822-015-9873-0
- Lai, C. C., Shih, T. P., Ko, W. C., Tang, H. J., and Hsueh, P. R. (2020). Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and coronavirus disease-2019 (COVID-19): the epidemic and the challenges. *Int. J. Antimicrob. Agents* 55:105924. doi: 10.1016/j.ijantimicag.2020.105924
- Li, H., Liu, S. M., Yu, X. H., Tang, S. L., and Tang, C. K. (2020). Coronavirus disease 2019 (COVID-19): current status and future perspectives. *Int. J. Antimicrob. Agents* 55:105951. doi: 10.1016/j.ijantimicag.2020.105951
- Lorber, D. M., and Shoichet, B. K. (1998). Flexible ligand docking using conformational ensembles. *Protein Sci.* 7, 938–950. doi: 10.1002/pro.5560070411
- Luan, B., Huynh, T., Cheng, X., Lan, G., and Wang, H. R. (2020). Targeting proteases for treating COVID-19. *J. Proteome Res.* 19, 4316–4326. doi: 10.1021/acs.jproteome.0c00430
- Meyer-Almes, F. J. (2020). Repurposing approved drugs as potential inhibitors of 3CL-protease of SARS-CoV-2: virtual screening and structure based drug design. *Comput. Biol. Chem.* 88:107351. doi: 10.1016/j.compbiolchem.2020.107351
- Mizutani, M. Y., Takamatsu, Y., Ichinose, T., Nakamura, K., and Itai, A. (2006). Effective handling of induced-fit motion in flexible docking. *Proteins* 63, 878–891. doi: 10.1002/prot.20931
- Mohammadi, E., Benfeitas, R., Turkez, H., Boren, J., Nielsen, J., Uhlen, M., et al. (2020). Applications of genome-wide screening and systems biology approaches in drug repositioning. *Cancers* 12:2694. doi: 10.3390/cancers12092694
- Pal, M., Berhanu, G., Desalegn, C., and Kandi, V. (2020). Severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2): an update. *Cureus* 12:e7423. doi: 10.7759/cureus.7423
- Perez-Regidor, L., Zariw, M., Ortega, L., and Martin-Santamaria, S. (2016). Virtual screening approaches towards the discovery of toll-like receptor modulators. *Int. J. Mol. Sci.* 17:1508. doi: 10.3390/ijms17091508
- Pillaiyar, T., Manickam, M., Namasivayam, V., Hayashi, Y., and Jung, S. H. (2016). An overview of severe acute respiratory syndrome-coronavirus (SARS-CoV) 3CL protease inhibitors: peptidomimetics and small molecule chemotherapy. *J. Med. Chem.* 59, 6595–6628. doi: 10.1021/acs.jmedchem.5b01461
- Pronk, S., Pall, S., Schulz, R., Larsson, P., Bjelkmar, P., Apostolov, R., et al. (2013). GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* 29, 845–854. doi: 10.1093/bioinformatics/btt055
- Rabi, F. A., Al Zoubi, M. S., Kasasbeh, G. A., Salameh, D. M., and Al-Nasser, A. D. (2020). SARS-CoV-2 and coronavirus disease 2019: what we know so far. *Pathogens* 9:231. doi: 10.3390/pathogens9030231
- Salmaso, V., and Moro, S. (2018). Bridging molecular docking to molecular dynamics in exploring ligand-protein recognition process: an overview. *Front. Pharmacol.* 9:923. doi: 10.3389/fphar.2018.00923
- Sastry, G. M., Adzhigirey, M., Day, T., Annabhimoju, R., and Sherman, W. (2013). Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments. *J. Comput. Aided Mol. Des.* 27, 221–234. doi: 10.1007/s10822-013-9644-8
- Saul, S., and Einav, S. (2020). Old drugs for a new virus: repurposed approaches for combating COVID-19. *ACS Infect. Dis.* 6, 2304–2318. doi: 10.1021/acsinfecdis.0c00343
- Seifert, M. H., Kraus, J., and Kramer, B. (2007). Virtual high-throughput screening of molecular databases. *Curr. Opin. Drug Discov. Dev.* 10, 298–307
- Selvaraj, C., Dinesh, D. C., Panwar, U., Boura, E., and Singh, S. K. (2020b). High-throughput screening and quantum mechanics for identifying potent inhibitors against Mac1 Domain of SARS-CoV-2 Nsp3. *Chem. Biol.* 11, 1445–1453. doi: 10.1016/j.chembiol.2004.08.011
- Selvaraj, C., Sakkiath, S., Tong, W., and Hong, H. (2018). Molecular dynamics simulations and applications in computational toxicology and nanotoxicology. *Food Chem. Toxicol.* 112, 495–506. doi: 10.1016/j.fct.2017.08.028
- Selvaraj, C., Dinesh, D. C., Panwar, U., Abhirami, R., Boura, E., and Singh, S. K. (2020a). Structure-based virtual screening and molecular dynamics simulation of SARS-CoV-2 Guanine-N7 methyltransferase (nsp14) for identifying antiviral inhibitors against COVID-19. *J. Biomol. Struct. Dyn.* 1–12. doi: 10.1080/07391102.2020.1778535
- Shyr, Z. A., Gorshkov, K., Chen, C. Z., and Zheng, W. (2020). Drug discovery strategies for SARS-CoV-2. *J. Pharmacol. Exp. Ther.* 375, 127–138. doi: 10.1124/jpet.120.000123
- Singh, T. U., Parida, S., Lingaraju, M. C., Kesavan, M., Kumar, D., and Singh, R. K. (2020). Drug repurposing approach to fight COVID-19. *Pharmacol. Rep.* 5, 1–30. doi: 10.1007/s43440-020-00155-6
- Tharappel, A. M., Samrat, S. K., Li, Z., and Li, H. (2020). Targeting crucial host factors of SARS-CoV-2. *ACS Infect. Dis.* 6, 2844–2865. doi: 10.1021/acsinfecdis.0c00456
- Touret, F., Gilles, M., Barral, K., Nougaiere, A., Van Helden, J., Decroly, E., et al. (2020). *In vitro* screening of a FDA approved chemical library reveals potential inhibitors of SARS-CoV-2 replication. *Sci. Rep.* 10:13093. doi: 10.1038/s41598-020-70143-6
- Tripathi, S. K., Singh, S. K., Singh, P., Chellaperumal, P., Reddy, K. K., and Selvaraj, C. (2012). Exploring the selectivity of a ligand complex with CDK2/CDK1: a molecular dynamics simulation approach. *J. Mol. Recognit.* 25, 504–512. doi: 10.1002/jmr.2216
- Ullrich, S., and Nitsche, C. (2020). The SARS-CoV-2 main protease as drug target. *Bioorg. Med. Chem. Lett.* 30:127377. doi: 10.1016/j.bmcl.2020.127377
- Umesh, Kundu, D., Selvaraj, C., Singh, S. K., and Dubey, V. K. (2020). Identification of new anti-nCoV drug chemical compounds from Indian spices exploiting SARS-CoV-2 main protease as target. *J. Biomol. Struct. Dyn.* 1–9. doi: 10.1080/07391102.2020.1763202
- Van Aalten, D. M., Bywater, R., Findlay, J. B., Hendlich, M., Hooft, R. W., and Vriend, G. (1996). PRODRG, a program for generating molecular topologies and unique molecular descriptors from coordinates of small molecules. *J. Comput. Aided Mol. Des.* 10, 255–262. doi: 10.1007/BF00355047
- Van Der Spoel, D., Lindahl, E., Hess, B., Groenhof, G., Mark, A. E., and Berendsen, H. J. (2005). GROMACS: fast, flexible, and free. *J. Comput. Chem.* 26, 1701–1718. doi: 10.1002/jcc.20291
- Vannabouathong, C., Devji, T., Ekhtiari, S., Chang, Y., Phillips, S. A., Zhu, M., et al. (2020). Novel coronavirus COVID-19: current evidence and evolving strategies. *J. Bone Joint Surg. Am* 102, 734–744. doi: 10.2106/JBJS.20.00396
- Wahba, L., Jain, N., Fire, A. Z., Shoura, M. J., Artiles, K. L., McCoy, M. J., et al. (2020). An Extensive meta-metagenomic search identifies

- sars-cov-2-homologous sequences in pangolin lung viromes. *mSphere* 5, e00160–20. doi: 10.1128/mSphere.00160-20
- Wang, J. (2020). Fast identification of possible drug treatment of coronavirus disease-19 (covid-19) through computational drug repurposing study. *J. Chem. Inf. Model.* 60, 3277–3286. doi: 10.1021/acs.jcim.0c00179
- Xu, H. Y., Zhang, Y. Q., Liu, Z. M., Chen, T., Lv, C. Y., Tang, S. H., et al. (2019). ETCM: an encyclopaedia of traditional Chinese medicine. *Nucleic Acids Res.* 47, D976–D982. doi: 10.1093/nar/gky987
- Yoshino, R., Yasuo, N., and Sekijima, M. (2020). Identification of key interactions between SARS-CoV-2 main protease and inhibitor drug candidates. *Sci. Rep.* 10:12493. doi: 10.1038/s41598-020-69337-9
- Yuan, H., Ma, Q., Ye, L., and Piao, G. (2016). The traditional medicine and modern medicine from natural products. *Molecules* 21:559. doi: 10.3390/molecules21050559
- Zeng, L., Guan, M., Jin, H., Liu, Z., and Zhang, L. (2015). Integrating pharmacophore into membrane molecular dynamics simulations to improve homology modeling of G protein-coupled receptors with ligand selectivity: A2A adenosine receptor as an example. *Chem. Biol. Drug Des* 86, 1438–1450. doi: 10.1111/cbdd.12607
- Zhang, L., Lin, D., Sun, X., Curth, U., Drosten, C., Sauerhering, L., et al. (2020). Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved alpha-ketoamide inhibitors. *Science* 368, 409–412. doi: 10.1126/science.abb3405
- Zhao, J., Cui, W., and Tian, B. P. (2020). The potential intermediate hosts for SARS-CoV-2. *Front. Microbiol.* 11:580137. doi: 10.3389/fmicb.2020.580137
- Zheng, J. (2020). SARS-CoV-2: an emerging coronavirus that causes a global threat. *Int. J. Biol. Sci.* 16, 1678–1685. doi: 10.7150/ijbs.45053
- Zhu, Z., Lian, X., Su, X., Wu, W., Marraro, G. A., and Zeng, Y. (2020). From SARS and MERS to COVID-19: a brief summary and comparison of severe acute respiratory infections caused by three highly pathogenic human coronaviruses. *Respir. Res.* 21:224. doi: 10.1186/s12931-020-01479-w
- Zoete, V., Irving, M. B., and Michielin, O. (2010). MM-GBSA binding free energy decomposition and T cell receptor engineering. *J. Mol. Recognit.* 23, 142–152. doi: 10.1002/jmr.1005

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Selvaraj, Panwar, Dinesh, Boura, Singh, Dubey and Singh. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Computational Characterizations of the Interactions Between the Pontacyl Violet 6R and Exoribonuclease as a Potential Drug Target Against SARS-CoV-2

Rangika Munaweera and Ying S. Hu\*

Department of Chemistry, College of Liberal Arts and Sciences, University of Illinois at Chicago, Chicago, IL, United States

## OPEN ACCESS

### Edited by:

Sonia Di Gaetano,  
Institute of Biostructure and  
Bioimaging, Italian National Research  
Council, Italy

### Reviewed by:

Milan Sencanski,  
Vinča Institute of Nuclear Science,  
University of Belgrade, Serbia  
Prashasti Kumar,  
Icahn School of Medicine at Mount  
Sinai, United States

### \*Correspondence:

Ying S. Hu  
yshu@uic.edu

### Specialty section:

This article was submitted to  
Theoretical and Computational  
Chemistry,  
a section of the journal  
Frontiers in Chemistry

**Received:** 09 November 2020

**Accepted:** 21 December 2020

**Published:** 21 January 2021

### Citation:

Munaweera R and Hu YS (2021)  
Computational Characterizations of the  
Interactions Between the Pontacyl  
Violet 6R and Exoribonuclease  
as a Potential Drug Target Against  
SARS-CoV-2.  
Front. Chem. 8:627340.  
doi: 10.3389/fchem.2020.627340

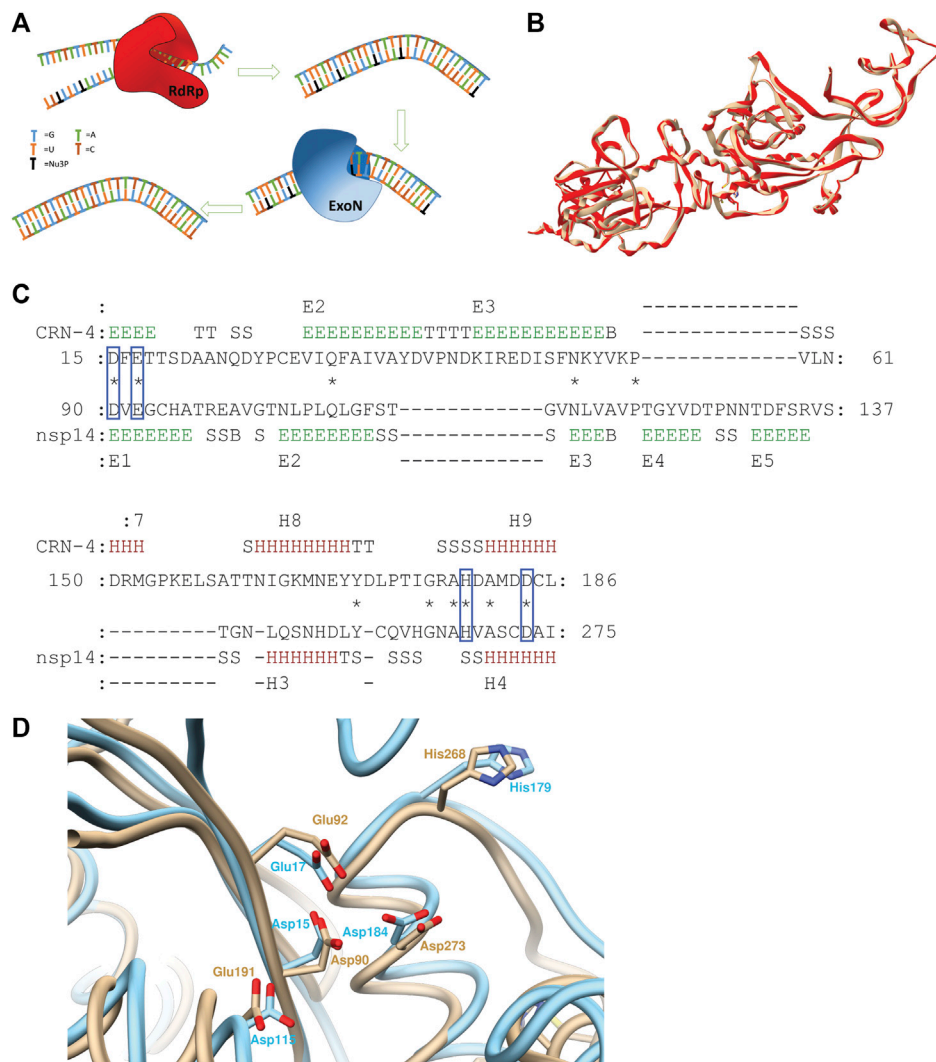
We report a molecular-docking and virtual-screening-based identification and characterization of interactions of lead molecules with exoribonuclease (ExoN) enzyme in severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). From previously identified DEDDh/DEEDh subfamily nuclease inhibitors, our results revealed strong binding of pontacyl violet 6R (PV6R) at the catalytic active site of ExoN. The binding was found to be stabilized via two hydrogen bonds and hydrophobic interactions. Molecular dynamics simulations further confirmed the stability of PV6R at the active site showing a shift in ligand to reach a more stabilized binding. Using PV6R as the lead molecule, we employed virtual screening to identify potential molecular candidates that form strong interactions at the ExoN active site. Our study paves ways for evaluating the ExoN as a novel drug target for antiviral treatment against SARS-CoV-2.

**Keywords:** exoribonuclease, SARS-CoV-2, pontacyl violet 6R, DEDDh exonucleases, RNA-dependent RNA polymerase, orthocoronavirinae

## INTRODUCTION

Due to the highly contagious nature of the virus and the high cost associated with the production of synthetic RNA molecules, the traditional high throughput screening has not shown sufficient effectivity in identifying new drugs against the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). To this end, computational studies play critical roles by narrowing down potential hits through accurate theoretical predictions, such as using molecular docking and molecular dynamics simulations. This study employed computational techniques to identify lead molecules that competitively bind at the exoribonuclease (ExoN) site of non-structural protein 14 (nsp14) from SARS-CoV-2.

The RNA-dependent RNA polymerase (RdRp) is a critical enzyme that enables the error-prone replication of RNA viruses and facilitates the rapid adaptation of these viruses to changing circumstances (Crotty et al., 2001; Elena and Sanjuan, 2005). RNA synthesis by the RdRp is one of the main drug targets in the formulation of antiviral agents for genetically diverse *Orthocoronavirinae* viruses (CoVs) (Brown et al., 2019). *In vitro* studies have demonstrated the efficacy of nucleoside analogs (NuA) in CoV infected cells (Warren et al., 2016; Jordan et al., 2018; Pruijssers and Denison, 2019; Tchesnokov et al., 2019). Specifically, intracellular kinases metabolize NuA to the corresponding 5'-triphosphates of the NuA (Nu3P), thus perturbing endogenous nucleoside triphosphate pools (Warren et al., 2016). RdRp incorporates Nu3Ps to the growing RNA



**FIGURE 1 | (A)** Schematic illustration of ExoN activity countering the drug effect of nucleoside analogs. **(B)** Superimposed structures of 5NFY template (beige) and homology model (red) **(C)** Pairwise 3D structure alignment of active sites of CRN-4 and homology model of nsp14 indicating the alignment of homologues catalytic residues of two enzymes in blue boxes (regions indicated by letter E show extended strands which participates in beta ladders; regions indicated by letter T show hydrogen bonded turns; regions indicated by letter S show bends; regions indicated by letter B show residues in isolated beta-bridges; regions indicated by letter H show alpha helices) **(D)** CRN-4 (cyan) and nsp14 (beige) superimposed highlighting active site residues. 90th aspartic acid, 92nd glutamic acid, 191th glutamic acid, 268th histidine, 273rd aspartic acid residues of nsp14 aligns with 15th aspartic acid, 17th glutamic acid, 115th aspartic acid, 179th histidine and 184 aspartic residues of CRN-4 respectively. SARS-CoV-2 has a 191th glutamic acid residue while 115th CRN-4 has an aspartic acid residue.

strands, resulting in mutated RNA products. Accumulation of mutated RNA products leads to lethal mutagenesis for the viruses, thereby achieving the desired antiviral effect (Vignuzzi et al., 2005).

However, the effectiveness of NuAs, such as the nucleotide prodrug remdesivir (GS-5734), against SARS-CoV-2 can be countered by the presence of exoribonuclease (ExoN) enzymes. In particular, the ExoN enzymes proofread erroneously added Nu3Ps in the RNA products (Figure 1A) (Minskaia et al., 2006; Eckerle et al., 2010; Smith et al., 2013; Agostini et al., 2018). This mechanism counters the effect of NuAs and may lead to the requirement of a higher dosage of

NuAs with increased systemic side-effects (Agostini et al., 2018; Dong et al., 2020).

CoVs contain one of the largest viral genomes ranging from 27 to 34 kB. The viral genome encodes for both structural (sp) and nonstructural proteins (nsp). Out of the 16 nsps, nsp12 is known as the RdRp. The RdRp forms a complex with two other nsps (nsp7 and nsp8) to function in the cellular environment (Kirchdoerfer and Ward, 2019; Ogando et al., 2019). Nsp 14, also known as the ExoN, is classified into the superfamily of DEDD exonucleases. DEDD exonucleases contain the proofreading domains of many DNA polymerases as well as other eukaryotic and prokaryotic exonucleases (Ogando et al.,

2019). Further, structural studies of SARS-CoV nsp14 showed the presence of five catalytic residues. These features reveal that nsp14 is a member of DEDDh/DEEDh subfamily (Huang et al., 2016; Ogando et al., 2019). Interestingly, nsp14 in SARS-CoV exhibits both the (N7-guanine)-methyltransferase (N7-MTase) activity and ExoN activity. N7-MTase domain is located in the C-terminal domain of nsp14. Studies confirmed that these two enzymatic activities (ExoN and N7-MTase) of the nsp14 are functionally distinct and physically independent, although they are structurally interconnected (Ogando et al., 2019).

Importantly, the proofreading activities of ExoN may counter the drug effect of Nu3Ps (Ogando et al., 2019). A recent report argued that an effective anti-viral drug design has to consider the balancing act between nsp12 and nsp14 (Shannon et al., 2020). Despite being a potential drug target, few compounds have been identified to inhibit the nsp14 activity *in vitro* (Huang et al., 2016). Simultaneous attack of RdRp and nsp14 using a combination of a nucleoside analog and a specific nsp14 inhibitor may enhance lethal mutagenesis in SARS-CoV-2 (Ogando et al., 2019). This strategy may enable the use of a range of NuAs that have been identified as possible drugs but failed to show efficacy *in vitro* and *in vivo*.

Since the nsp14 structure of SARS CoV-2 has not been solved at the time of this study, a homology model was constructed using the nsp14 of SARS CoV as the template. Candidate inhibitor molecules were computationally docked to the modeled nsp14 protein to identify potential lead molecules. Identified molecules were further evaluated for their stability using molecular dynamics (MD) simulations. Interactions of the strongly binding ligand molecule were computationally characterized and its molecular features were extracted. Virtual screening was performed using the extracted molecular features from lead molecule. Further, we optimized lead molecules using a number of computational strategies.

## MATERIALS AND METHODS

### Homology Modeling and Active Site Alignment

The newly-emerged SARS-CoV-2 nucleotide gene (YP\_009725309.1) was retrieved from the National Center for Biotechnology Information (NCBI) nucleotide database ([https://www.ncbi.nlm.nih.gov/protein/YP\\_009725309.1](https://www.ncbi.nlm.nih.gov/protein/YP_009725309.1)). A homology model for the SARS-CoV-2 nsp14 was built using the Swiss Model web server (Waterhouse et al., 2018). The structure of the SARS-CoV nsp10/nsp14 complex (PDB ID: 5NFY) was the most sequelogenous (99% sequence identity) to the SARS-CoV-2 nsp14 with a resolution of 3.38 Å. The quality of the homology model was analyzed using QMEAN score and Ramachandran plot generated by Swiss Model server itself (Waterhouse et al., 2018). It was further analyzed using the ProSA server (Wiederstein and Sippl, 2007). The active site pairwise structure alignment of CRN4 and homology model was carried out using TM align server and MATRAS protein 3D structure comparison tool (Kawabata, 2003; Zhang and Skolnick, 2005).

### Molecular Docking

Molecular docking studies were performed using the AutoDock Vina software package. Results were analyzed using Autodock Tools and Chimera packages (Forli et al., 2016; Anil, 2019). Blind dock studies were carried out using previously identified three DEDDh/DEEDh subfamily nuclease inhibitors, pontacyl Violet 6R (PV6R), aurintricarboxylic acid (ATA) and 5,5'-dithiobis (2-nitrobenzoic acid) (DTNB) (Huang et al., 2016). The homology model of the SARS-CoV-2 nsp14 was used as the macromolecule.

### Molecular Dynamics Simulations

MD simulations were performed using GROMACS software package. The RMSD of the PV6R molecule at the ExoN active site was calculated. Docking pose of the PV6R molecule was used to carry out the MD simulation. Ligands topology was prepared using CGenFF server and protein topology was prepared using Gromacs software itself. MD simulations were done using CHARMM36 all-atom force field. All other molecule preparations were done using Chimera software package unless otherwise stated.

Prepared complexes were immersed in a cubic box with a 1.5 nm distance between the protein surface and the boundary of the cubic box. The selected cubic box was filled with SPC/E water molecules to solvate the system and periodic boundary conditions were applied from all sides. To neutralize the protein and to maintain the net ionic strength of the system to a value closer to the physiological conditions  $\text{Na}^+$  and  $\text{Cl}^-$  ions were added replacing the solvent molecules. Energy minimization was done using the steepest descent algorithm with an energy minimization step of  $0.01 \text{ kJ mol}^{-1}$  for 50,000 steps setting the maximum force value for  $1,000 \text{ kJ mol}^{-1} \text{ nm}^{-1}$ . NVT and NPT equilibrations were carried out for a total of 100 ps for each with 2 fs steps. Product MD run was performed for a total of 21.0 ns consist of 2 fs steps. Temperature and pressure coupling were performed using the modified Berendsen thermostat and Parrinello-Rahman method. Ewald particle mesh method was used to calculate long-range electrostatic interactions. Results were analyzed using the VMD software package.

### Virtual Screening, Lead-like Molecule Identification, Physiochemical and Pharmacokinetics Studies

The compound selected from molecular docking (lead compound) was virtually screened using the OpenEye Scientific vROCS application (ROCS 3.3.2.2: OpenEye Scientific Software, Santa Fe, NM. <http://www.eyesopen.com>; Hawkins et al., 2007). Virtual screening was carried out using eMolecule and Kishida databases (ROCS 3.3.2.2: OpenEye Scientific Software, Santa Fe, NM. <http://www.eyesopen.com>). Best 20 molecules from each database were docked with homology model and average structure obtained from MD simulation using Autodock vina and the binding energies of selected compounds were calculated. Compounds that bind at the ExoN active site were selected. Control docks were carried out with endogenous ribonucleotides. All of the compounds were prepared to be optimized in their active forms in physiological

conditions using Avagadro software (López, 2012). Binding of selected compounds at the ExoN active site of nsp14 was further examined using LigPlot application (Wallace et al., 1995). ADME parameters, pharmacokinetic properties, druglike nature and medicinal chemistry friendliness was evaluated using SwissADME server (Daina et al., 2017).

The lead molecule was used as the seed compound to generate novel compounds through generative shape-based neural network decoding using the LigDream web application (Skalic et al., 2019). The PV6R molecule was used as the seed compound to generate novel compounds from generative shape-based neural network decoding. Novel compounds generated by LigDream were arranged using RDock software based on the RDock score (Ruiz-Carmona et al., 2014). Five compounds with the best RDock score were directed to molecular docking using AutoDock vina for comparison purposes.

## RESULTS AND DISCUSSIONS

### Homology Modeling and Active Site Alignment

The SWISS MODEL server was used to build the homology model for the SARS-CoV-2 nsp14 using SARS-CoV nsp14/nsp10 complex (PDB ID: 5NFY) as the template. The template had a 99% sequence alignment with SARS-CoV-2 nsp14. The built homology model showed a QMEAN score of  $-3.18$ , GMQE score of  $0.98$ , and MolProbity results were better compared to the template (**Supplementary Table S1**). ProSA server results showed a Z-score of  $-9.15$ , suggesting an accurate homology model. The Produced model was further evaluated using the TM align server and MATRAS pairwise 3D alignment web server shown in **Supplementary Data 1** (Kawabata, 2003; Zhang and Skolnick, 2005). A 98% structure alignment (512 residues align with template out of 523 residues in the entire protein) of the homology model was obtained within a five Å range compared to the template and a root mean square deviation (RMSD) of  $0.15$  Å (**Figure 1B**). These results suggest a high-quality homology model, based on which our studies have been performed. The Ramchandran plots of the Homology model and the template are provided in **Supplementary Figures S1, S2**.

PV6R, ATA and DTNB have been previously identified as CRN-4 of and RNase T inhibitors. CRN-4 and RNase T are DEDDh exonucleases found in *C. elegans* and *E. coli*, respectively (Huang et al., 2016). Structure alignment of CRN-4 and nsp14 of SARS-CoV-2 using MATRAS pairwise 3D structure alignment and Chimera structure comparison shows close proximity (**Figures 1C,D**) of active site residues. Specifically, all DEDDh exonucleases share a conserved active site with a common set of residues (Huang et al., 2016; Ogando et al., 2019). In the nsp14 of SARS-CoV-2, residues Asp 90, Glu 92 (motif I), Glu 191 (motif II), Asp 273 (motif III) and His 268 (motif III) constitute the active site while the equivalent of E191 alternates between E and D in different nidovirus taxa. Nsp14 has Asp 90, Glu 92, Glu 191, His 268 and Asp 273 residues as the active site residues while CRN-4 having Asp15, Glu 17, Asp 115, His 179 and Asp 184 as active site residues (Huang et al., 2016; Ogando et al., 2019). This

observation suggests that previously identified inhibitors for the DEDDh exonuclease may serve as effective inhibitors of nsp14.

### Molecular Docking

Using the selected DEDDh exonucleases inhibitors, molecular docking revealed that PV6R binds to the ExoN binding site of nsp14 with a preferable calculated binding energy of  $-8.3$  kcal/mol. We further analyzed the interactions between active site residues and PV6R. **Figure 2A** shows that the PV6R molecule is stabilized *via* two hydrogen bonds at the active site including a very short hydrogen bond between Ala 187 and fifth oxygen of PV6R along with hydrophobic interactions significantly contributing to the stabilization. Interestingly, most of these hydrophobic interactions are formed with five residues (Asp 90, Glu 92, Glu 191, His 268 and Asp 273) that are involved in the catalytic activity. Protonated state of PV6R shows a similar calculated binding energy ( $-8.3$  kcal/mol) forming hydrogen bonds with Gly 93, His 268, Asp 273 and Asn 266N.

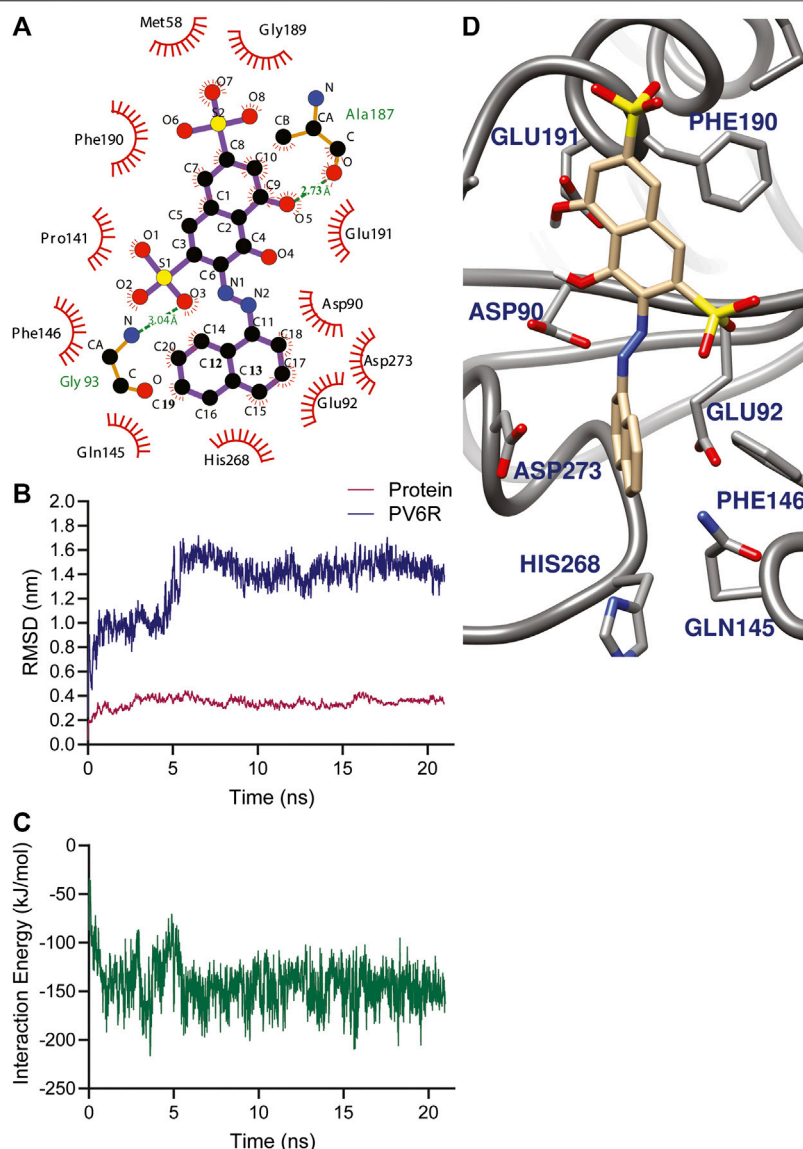
### Molecular Dynamics Simulations

The stability of the PV6R and ExoN complex in an aqueous solution was evaluated using the parameter RMSD. MD simulation was performed for 21.0 ns using the docked structure of PV6R to ExoN. The average RMSD of the trajectories for bound protein backbone atoms showed relative stability (**Figure 2B**). During the simulation PV6R molecules shows a shift in binding mode with structural conformational change to achieve more stabilized interaction at the ExoN active site showing more stable calculated interaction energy in the later part of the simulation (**Figures 2B–D**). This flexibility is observed due to the presence of more polar groups in PV6R that can form stronger interaction with acidic residue rich active site of ExoN.

### Virtual Screening and Lead-like Molecule Identification

Next, virtual screening was performed using the vROCS application (OpenEye scientific software). This software package identifies the molecular features of a given molecule based on the structure, shape and electrostatics. To identify structurally and electrostatically similar molecules, the PV6R molecule was compared with the eMolecules database and Kishida database from the OpenEye software. Parameters were set to provide 20 best fits for each run. From obtained hits, molecules that bind at the active site with a calculated binding energy of less than  $-7.0$  kcal/mol were selected and shown in **Table 1**. Two molecules from the eMolecules database (Molecule reference numbers: 22722115\_4 and 6826969\_6) and three molecules from the Kishida database (Molecule reference numbers: KS122-0741955\_27, KS122-0742530\_28, KS122-0742095\_14) were selected for further studies as shown in **Figure 3**.

**Figure 4** shows the molecular docking studies of these five selected compounds identified in **Figure 3** (22722115\_4, 6826969\_6 from eMolecules database and KS122-0741955\_27, KS122-0742530\_28, KS122-0742095\_14 from Kishida database). Each compound was further analyzed by the LigPlot application



**FIGURE 2 | (A)** Amino acid environment of the PV6R molecule bound to the active site of nsp 14. PV6R forms two hydrogens bonds with 187th Alanine residue and 93rd Glycine residue. It is further stabilized at the binding site through hydrophobic interactions including all the active site residues; Asp 90, Glu 92, Glu 191, His 268 and Asp 273. MD simulations results showing **(B)** RMSD of protein backbone and PV6R at ExoN site during the simulation time (21.0 ns) showing a shift in binding type achieving better interaction energy according to **(C)** Interaction energy (kJ/mol) plot of PV6R with active site residues during 21.0 ns simulation **(D)** Snapshot of simulation taken at 10 ns.

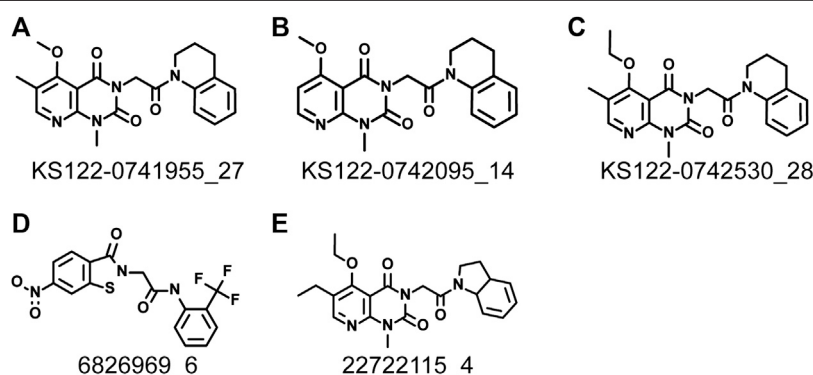
to identify the interactions that make these molecules stable at the binding site. **Table 2** shows the interacting amino acids with the PV6R and the five selected compounds at the ExoN binding site. KS122-0741955\_27, KS122-0742095\_14 and KS122-0742530\_28 exhibit a similar pattern of interactions at the active site forming one hydrogen bond with the Gly 93 residue and several hydrophobic interactions. All three molecules form hydrophobic interactions with Asp 90, Val 91, Glu 92, Pro 141, Phe 146, Phe 190, Val 191, His 268 and Asp 273 while KS122-0741955\_27 and KS122-0742530\_28 displays interactions with Asn 104 and KS122-0742530\_28 form an extra hydrophobic

interaction with Leu 149. Calculated Binding free energies of KS122-0741955\_27, KS122-0742095\_14 and KS122-0742530\_28 were  $-8.2$  kcal/mol,  $-7.9$  kcal/mol and  $-8.7$  kcal/mol, respectively.

From the calculated binding energies, KS122-0742530\_28 represents the most favorable candidate due to the presence of extra hydrophobic interactions along with close H bond acceptor and donor localizations (2.95 nm). Calculated binding free energy of  $-8.7$  kcal/mol suggests KS122-0742530\_28 as a better ExoN inhibitor than PV6R. Molecules identified from eMolecules database showed higher binding free energies due to the

**TABLE 1** | Calculated binding energies of selected ligands at ExoN active site.

	Ligand	TanimotoCombo score	Calculated binding energy with homology model (kcal/mol)	Calculated binding energy with average structure obtained from MD simulation (kcal/mol)
Lead molecules	DTNB	n/a	>−7.0	n/a
	ATA	n/a	>−7.0	n/a
	PV6R	n/a	−8.3	n/a
Virtual screening - OpenEye scientific eMolecules database	22722115_4	0.970	−7.0	>−7.0
	6826969_6	0.957	−7.6	−7.4
Virtual screening - OpenEye scientific kishida database	KS122-	1.031	−8.2	−8.3
	0741955_27			
	KS122-	0.999	−8.7	−9.1
	0742530_28			
	KS122-	0.994	−7.9	−8.0
	0742095_14			
Positive control runs	ATP	n/a	−6.7	n/a
	GTP	n/a	−7.2	n/a
	CTP	n/a	−6.9	n/a
	UTP	n/a	−7.0	n/a
5 selected compounds based on Rdoc score from generative modeling (Please refer <b>Supplementary Data</b> for structures)	18	n/a	−7.7	n/a
	40	n/a	−7.9	n/a
	48	n/a	−7.8	n/a
	61	n/a	−7.8	n/a
	79	n/a	−7.9	n/a

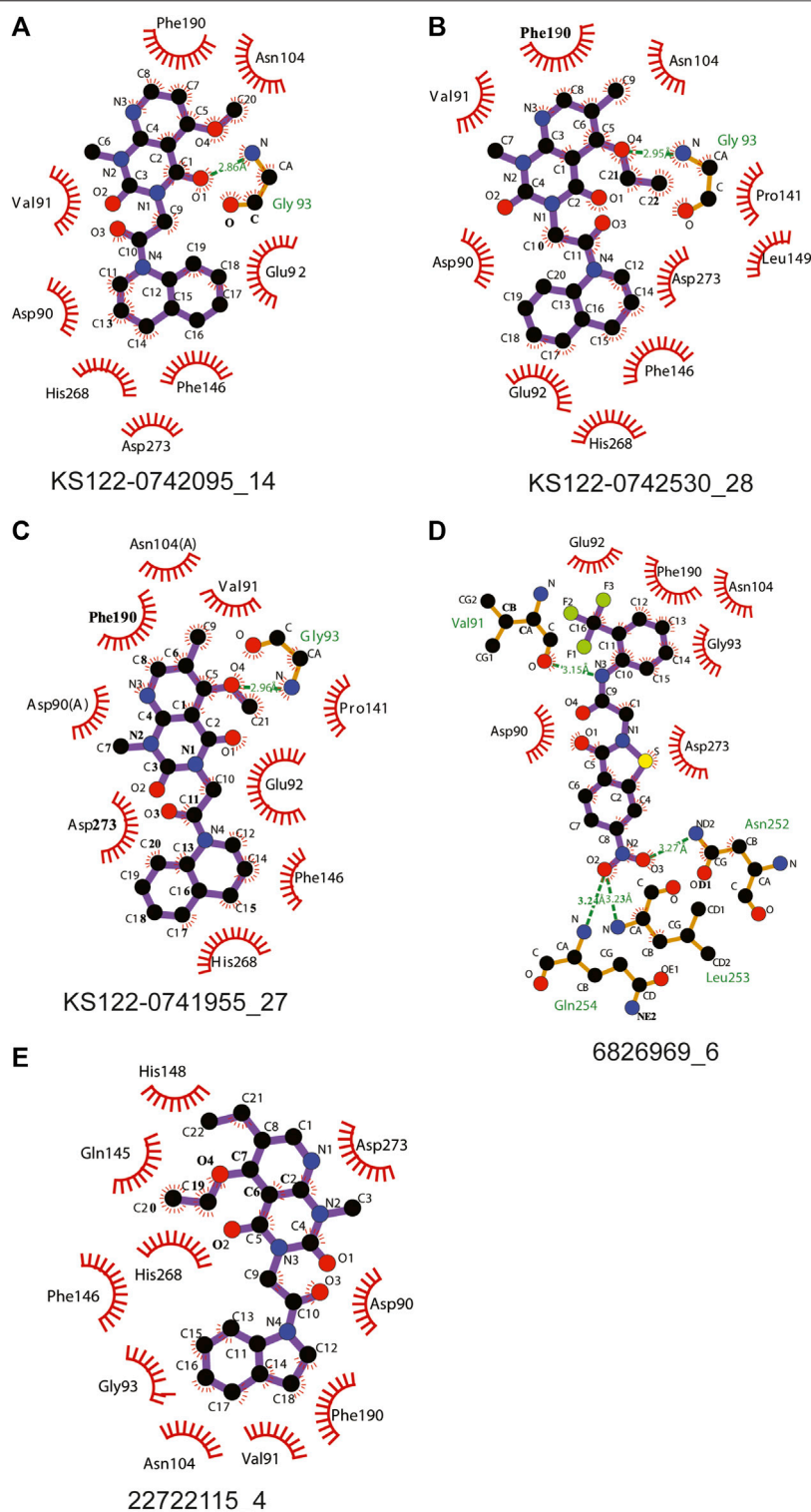
**FIGURE 3** | Identified molecules that show favorable binding at nsp14 active site from virtual screening (A) KS122-0741955\_27 (B) KS122-0742095\_14 (C) KS122-0742530\_28 (D) 6826969\_6 (E) 22722115\_4.

presence of high electronegative groups and the formation of weak hydrogen bonds in 6826969\_6 and the lack of strong intermolecular hydrogen bonding in 22722115\_4.

Molecular docking studies between compounds identified with virtual screening and average structure of ExoN obtained from MD simulation showed further decrease for the KS122-0741955\_27, KS122-0742530\_28 and KS122-0742095\_14 molecules with a calculated binding energy value of −8.3 kcal/mol, −9.1 kcal/mol and 8.0 kcal/mol respectively. Molecules from eMolecules database showed an increase in energy, as shown in **Table 1**. The drastic stability enhancement of KS122-0742530\_28 can be due to its structural similarity to the PV6R molecule

showing similar trend as did the PV6R molecule during the MD simulation. In general, the enhanced binding strength of molecules obtained from Kishida database shows their effectiveness and can be promising lead molecules for further experimental validation.

Next, a technique based on machine learning was used to generate novel molecules using PV6R as the seed compound. Three-dimensional shape and pharmacophoric features of the seed molecule (lead molecule) were used to produce lead-like compounds (Skalic et al., 2019). 92 identified novel compounds were further arranged using the RDock binding scores at the ExoN active site of the nsp14 (SMILES of 92 compounds have



**FIGURE 4 |** Interactions of molecules identified from virtual screening **(A)** KS122-0742095\_14 **(B)** KS122-0742530\_28 **(C)** KS122-0741955\_27 **(D)** 6826969\_6 **(E)** 22722115\_4 at nsp14 ExoN active site with amino acid environment.

been provided in **Supplementary Table S2**). Five compounds (**Supplementary Figure S3**) with the best RDock scores were further docked using AutoDock vina. Their calculated binding

energies at the ExoN site were tabulated in **Table 1**. Together with the virtual screening results, these identified molecules with preferable binding at the Exon active site serve as promising

**TABLE 2 |** Interacting amino acids in the docking poses of the ligands.

Compound	Calculated binding energy (kcal/mol)	Interacting amino acids in the active site	H Bonds
PV6R	−8.3	Met 58, Asp 90, Glu 92, Gly 93, Pro 141, Gln 145, Phe 146, Ala 187, Gly 189, Phe 190, Glu191, His 268 and Asp 273	2
PV6R (Protonated)	−8.3	Met 58, Glu 92, Gly 93, Asn 104, Pro 141, Phe 146, Leu 149, Phe 190, His 268, Asp 273, Asn 266	4
KS122-0741955_27	−8.2	Asp 90, Val 91, Glu 92, Gly 93, Asn 104, Pro 141, Phe 146, Phe 190, Val 191, His 268 and Asp 273	1
KS122-0742530_28	−8.7	Asp 90, Val 91, Glu 92, Gly 93, Asn 104, Pro 141, Phe 146, Leu 149, Phe 190, His 268 and Asp 273	1
KS122-0742095_14	−7.9	Asp 90, Val 91, Glu 92, Gly 93, Pro 141, Phe 146, Phe 190, Val 191, His 268 and Asp 273	1
22722115_4	−7.0	Asp 90, Val 91, Glu 92, Gly 93, Asn 104, Phe 190, Phe 191, Asn 252, Leu 253, Gln 254, Asp 273	4
6826969_6	−7.6	Asp 90, Val 91, Gly 93, Asn 104, Gln 145, Phe 146, His 148, Phe 190, His 268, Asp 273	0

**TABLE 3 |** Physiochemical and general properties of ligands.

Compound	Oral bioavailability: TPSA (Å²)	MW (g/mol)	Log P	H Bonds		Solubility (ESOL)		
				No. of H bond acceptors	No. of H bond donors	Log S	Solubility (mg/ml)	Class
PV6R	196.34	472.45	2.76	10	2	−5.19	3.04 × 10 <sup>−3</sup>	Moderately soluble
KS122-0741955_27	88.43	394.42	2.06	5	0	−3.42	9.45 × 10 <sup>−2</sup>	Soluble
KS122-0742530_28	86.43	408.45	2.38	5	0	−3.85	5.71 × 10 <sup>−2</sup>	Soluble
KS122-0742095_14	86.43	380.4	1.73	5	0	−3.31	1.84 × 10 <sup>−1</sup>	Soluble
22722115_4	86.43	410.47	2.14	5	0	−3.61	1.01 × 10 <sup>−1</sup>	Soluble
6826969_6	129	398.34	1.33	7	2	−3.97	4.28 × 10 <sup>−2</sup>	Soluble

**TABLE 4 |** ADME, pharmacokinetics, drug-likeness and medicinal friendliness of ligands.

Compound	GI absorption	BBB permeant	P-gp substrate	Log Kp (skin permeation) cm/s	Druglikeness					Bioavailability score	PAINS	Brenk	Synthetic accessibility
					Lipinski	Ghose	Veber	Egan	Muegge				
PV6R	Low	No	No	−6.68	Yes	No	No	No	No	No	1	2	3.62
KS122-0741955_27	High	No	No	−7.36	Yes	Yes	Yes	Yes	Yes	0.55	0	0	3.11
KS122-0742530_28	High	No	No	−7.19	Yes	Yes	Yes	Yes	Yes	0.55	0	0	3.24
KS122-0742095_14	High	No	No	−7.54	Yes	Yes	Yes	Yes	Yes	0.55	0	0	2.97
22722115_4	High	No	No	−7.26	Yes	Yes	Yes	Yes	Yes	0.55	0	0	4.48
6826969_6	Low	No	Yes	−6.88	Yes	Yes	Yes	Yes	Yes	0.55	0	1	3.16

candidates for further investigations of their competitive inhibition of nsp14.

## Physiochemical and Pharmacokinetics Studies

Physiochemical studies of the selected compounds (Table 3) showed that the calculated TPSA values were within acceptable limits. All selected compounds possess a large number of hydrogen bond acceptors and 6826969\_6 has two hydrogen bond donors. The presence of moieties that enable the

hydrogen bond formation increases the binding affinity at the active site of the ExoN. A higher polar surface area indicates the drug's ability to permeate the cell membrane. Lipophilicity was calculated as an average value of iLOGP, XLOGP3, WLOGP, MLOGP and SILICOS-IT (Daina et al., 2017) indicated as log P, which can be implicated in blood-brain barrier penetration and permeability prediction. Further, log P along with the molecular weight (MW) can be used to predict the metabolism and excretion of xenobiotic compounds from the human body. All compounds possess better aqueous solubility compared to PV6R according to this analysis (Table 3).

High gastrointestinal (GI) absorption and impermeability of blood-brain barrier (BBB) make these promising candidates to be used as drugs against respiratory illness (Table 4). The pharmacokinetic parameters suggest that these compounds follow drug-likeness rules (Lipinski, Ghose, Veber, Egan and Muegge) confirming that these molecules are drug-like and have a strong possibility to be directed for further studies to be used as drugs (Ghose et al., 1999; Egan et al., 2000; Muegge et al., 2001; Veber et al., 2002; Lipinski et al., 2012). Two medicinal chemistry filters, pan assay interference compounds (PAINS) and Brenk, showed 0 alerts for four of the selected compounds, suggesting their lead likeness.

In summary, our molecular docking and molecular dynamics studies show that PV6R binds at the ExoN catalytic active site of SARS-CoV-2 with a strong calculated free binding energy of -8.3 kcal/mol. Virtual screening using PV6R as the lead molecule identified KS122-0742530\_28 from the Kishida database with a better binding free energy. Both PV6R and KS122-0742530\_28 confirmed to have drug-like physiochemical and pharmacokinetic properties. These molecules may serve as lead molecules in further experimental validation for the development of ExoN inhibitors against SARS-CoV-2, including profiling their 50% inhibitory concentration (IC<sub>50</sub>) values and virus-inhibitory effects.

## REFERENCES

- Agostini, M. L., Andres, E. L., Sims, A. C., Graham, R. L., Sheahan, T. P., Lu, X., et al. (2018). Coronavirus susceptibility to the antiviral remdesivir (GS-5734) is mediated by the viral polymerase and the proofreading exoribonuclease. *MBio*. 9, 1–15. doi:10.1128/mBio.00221-18
- Anil, K. T. J. W. (2019). Autodock vina: improving the speed and accuracy of docking. *J. Comput. Chem.* 31 (2), 455–461. doi:10.1002/jcc.21334
- AutoDock Brown, A. J., Won, J. J., Graham, R. L., Dinno, K. H., Sims, A. C., Feng, J. Y., et al. (2019). Broad spectrum antiviral remdesivir inhibits human endemic and zoonotic deltacoronaviruses with a highly divergent RNA dependent RNA polymerase. *Antivir. Res.* 169, 104541. doi:10.1016/j.antiviral.2019.104541
- Crotty, S., Cameron, C. E., and Andino, R. (2001). RNA virus error catastrophe: direct molecular test by using ribavirin. *Proc. Natl. Acad. Sci. U.S.A.* 98 (12), 6895–6900. doi:10.1073/pnas.111085598
- Daina, A., Michielin, O., and Zoete, V. (2017). SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. *Sci. Rep.* 7, 42717–42813. doi:10.1038/srep42717
- Dong, L., Hu, S., and Gao, J. (2020). Discovering drugs to treat coronavirus disease 2019 (COVID-19). *Drug Discov. Ther.* 14 (1), 58–60. doi:10.5582/ddt.2020.01012
- Eckerle, L. D., Becker, M. M., Halpin, R. A., Li, K., Venter, E., Lu, X., et al. (2010). Infidelity of SARS-CoV Nsp14-exonuclease mutant virus replication is revealed by complete genome sequencing. *PLoS Pathog.* 6, e1000896. doi:10.1371/journal.ppat.1000896
- Egan, W. J., Merz, K. M., and Baldwin, J. J. (2000). Prediction of drug absorption using multivariate statistics. *J. Med. Chem.* 43 (21), 3867–3877. doi:10.1021/jm000292e
- Elena, S. F., and Sanjuán, R. (2005). Adaptive value of high mutation rates of RNA viruses: separating causes from consequences. *J. Virol.* 79, 11555–11558. doi:10.1128/jvi.79.18.11555-11558.2005
- Forli, S., Huey, R., Pique, M. E., Sanner, M., Goodsell, D. S., and Olson, A. J. (2016). Computational protein-ligand docking and virtual drug screening with the AutoDock suite. *Nat. Protoc.* 11, 905–919. doi:10.1038/nprot.2016.051

## DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding author.

## AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

## ACKNOWLEDGMENTS

The authors thank the support from the Department of Chemistry from College of Liberal Arts and Sciences at the University of Illinois at Chicago and Chicago Biomedical Consortium (CR-002).

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fchem.2020.627340/full#supplementary-material>.

- Ghose, A. K., Viswanadhan, V. N., and Wendoloski, J. J. (1999). A knowledge-based approach in designing combinatorial or medicinal chemistry libraries for drug discovery. 1. A qualitative and quantitative characterization of known drug databases. *J. Comb. Chem.* 1 (1), 55–68. doi:10.1021/cc9800071
- Hawkins, P. C., Skillman, A. G., and Nicholls, A. (2007). Comparison of shape-matching and docking as virtual screening tools. *J. Med. Chem.* 50 (1), 74–82. doi:10.1021/jm0603365
- Huang, K. W., Hsu, K. C., Chu, L. Y., Yang, J. M., Yuan, H. S., and Hsiao, Y. Y. (2016). Identification of inhibitors for the DEDDh family of exonucleases and a unique inhibition mechanism by crystal structure analysis of CRN-4 bound with 2-Morpholin-4-ylethanesulfonate (MES). *J. Med. Chem.* 59 (17), 8019–8029. doi:10.1021/acs.jmedchem.6b00794
- Jordan, P. C., Liu, C., Raynaud, P., Lo, M. K., Spiropoulou, C. F., Symons, J. A., et al. (2018). Initiation, extension, and termination of RNA synthesis by a paramyxovirus polymerase. *PLoS Pathog.* 14 (2), 1–23. doi:10.1371/journal.ppat.1006889
- Kawabata, T. (2003). MATRAS: a program for protein 3D structure comparison. *Nucleic Acids Res.* 31 (13), 3367–3369. doi:10.1093/nar/gkg581
- Kirchdoerfer, R. N., and Ward, A. B. (2019). Structure of the SARS-CoV nsp12 polymerase bound to nsp7 and nsp8 co-factors. *Nat. Commun.* 10, 2342–2349. doi:10.1038/s41467-019-10280-3
- Lipinski, C. A., Lombardo, F., Dominy, B. W., and Feeney, P. J. (2012). Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Deliv. Rev.* 64 (1–3), 4–17. doi:10.1016/j.addr.2012.09.019
- López, R. (2012). Avogadro: an advanced semantic chemical editor, visualization, and analysis platform. *J. Cheminf.* 4 (1), 17. doi:10.1186/1758-2946-4-17
- Minskaia, E., Hertzog, T., Gorbalenya, A. E., Campanacci, V., Cambillau, C., Canard, B., et al. (2006). Discovery of an RNA virus 3'→5' exoribonuclease that is critically involved in coronavirus RNA synthesis. *Proc. Natl. Acad. Sci. U.S.A.* 103, 5108–5113. doi:10.1073/pnas.0508200103
- Muegge, I., Heald, S. L., and Brittelli, D. (2001). Simple selection criteria for drug-like chemical matter. *J. Med. Chem.* 44 (12), 1841–1846. doi:10.1021/jm015507e
- Ogando, N. S., Ferron, F., Decroly, E., Canard, B., Posthuma, C. C., and Snijder, E. J. (2019). The curious case of the nidovirus exoribonuclease: its role in RNA synthesis and replication fidelity. *Front. Microbiol.* 10, 1–17. doi:10.3389/fmicb.2019.01813

- Protein (1988). Bethesda (MD): National Library of Medicine (US), Bethesda, MA: National Center for Biotechnology Information. Available from: [https://www.ncbi.nlm.nih.gov/protein/YP\\_009725309.1](https://www.ncbi.nlm.nih.gov/protein/YP_009725309.1).
- Pruijssers, A. J., and Denison, M. R. (2019). Nucleoside analogues for the treatment of coronavirus infections. *Curr. Opin. Virol.* 35, 57–62. doi:10.1016/j.coviro.2019.04.002
- Protein (1998). ROCS 3.3.2.2: OpenEye Scientific Software, Santa Fe, NM. Available at: <http://www.eyesopen.com>.
- Ruiz-Carmona, S., Alvarez-Garcia, D., Foloppe, N., Garmendia-Doval, A. B., Juhos, S., Schmidtke, P., et al. (2014). rDock: a fast, versatile and open source program for docking ligands to proteins and nucleic acids. *PLoS Comput. Biol.* 10, e1003571–e1003577. doi:10.1371/journal.pcbi.1003571
- Shannon, A., Le, N. T., Selisko, B., Eydoux, C., Alvarez, K., Guillemot, J. C., et al. (2020). Remdesivir and SARS-CoV-2: structural requirements at both nsp12 RdRp and nsp14 Exonuclease active-sites. *Antivir. Res.* 178, 104793. doi:10.1016/j.antiviral.2020.104793
- Skalic, M., Jiménez, J., Sabbadin, D., and De Fabritiis, G. (2019). Shape-based generative modeling for de Novo drug design. *J. Chem. Inf. Model.* 59 (3), 1205–1214. doi:10.1021/acs.jcim.8b00706
- Smith, E. C., Blanc, H., Surdel, M. C., Vignuzzi, M., and Densson, M. R., (2013). Coronaviruses lacking exoribonuclease activity are susceptible to lethal mutagenesis: evidence for proofreading and potential therapeutics. *PLoS Pathog.* 9 (8), e1003565. doi:10.1371/journal.ppat.1003565
- Tchesnokov, E. P., Feng, J. Y., Porter, D. P., and Götte, M. (2019). Mechanism of inhibition of ebola virus RNA-dependent RNA polymerase by remdesivir. *Viruses.* 11 (4), 1–16. doi:10.3390/v11040326
- Veber, D. F., Johnson, S. R., Cheng, H. Y., Smith, B. R., Ward, K. W., and Kopple, K. D. (2002). Molecular properties that influence the oral bioavailability of drug candidates. *J. Med. Chem.* 45, 2615–2623. doi:10.1021/jm020017n
- Vignuzzi, M., Stone, J. K., and Andino, R. (2005). Ribavirin and lethal mutagenesis of poliovirus: molecular mechanisms, resistance and biological implications. *Virus Res.* 107, 173–181. doi:10.1016/j.virusres.2004.11.007
- Wallace, A. C., Laskowski, R. A., and Thornton, J. M. (1995). LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. *Protein Eng.* 8, 127–134. doi:10.1093/protein/8.2.127
- Warren, T. K., Jordan, R., Lo, M. K., Ray, A. S., Mackman, R. L., Soloveva, V., et al. (2016). Therapeutic efficacy of the small molecule GS-5734 against Ebola virus in rhesus monkeys. *Nature.* 531, 381–385. doi:10.1038/nature17180
- Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., et al. (2018). SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res.* 46, W296–W303. doi:10.1093/nar/gky427
- Wiederstein, M., and Sippl, M. J. (2007). ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Res.* 35, W407–W410. doi:10.1093/nar/gkm290
- Zhang, Y., and Skolnick, J. (2005). TM-align: a protein structure alignment algorithm based on the TM-score. *Nucleic Acids Res.* 33 (7), 2302–2309. doi:10.1093/nar/gki524

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Munaweera and Hu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# *In silico* Exploration of Inhibitors for SARS-CoV-2's Papain-Like Protease

Tien Huynh, Wendy Cornell and Binqun Luan\*

Computational Biological Center, IBM Thomas J. Watson Research, New York, NY, United States

Coronavirus disease 2019 (COVID-19) is an ongoing global pandemic caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), with very limited treatments so far. Demonstrated with good druggability, two major proteases of SARS-CoV-2, namely main protease (Mpro) and papain-like protease (PLpro) that are essential for viral maturation, have become the targets for many newly designed inhibitors. Unlike Mpro that has been heavily investigated, PLpro is not well-studied so far. Here, we carried out the *in silico* high-throughput screening of all FDA-approved drugs *via* the flexible docking simulation for potential inhibitors of PLpro and explored the molecular mechanism of binding between a known inhibitor rac5c and PLpro. Our results, from molecular dynamics simulation, show that the chances of drug repurposing for PLpro might be low. On the other hand, our long (about 450 ns) MD simulation confirms that rac5c can be bound stably inside the substrate-binding site of PLpro and unveils the molecular mechanism of binding for the rac5c-PLpro complex. The latter may help perform further structural optimization and design potent leads for inhibiting PLpro.

**Keywords:** COVID-19, SARS-CoV-2, papain-like protease, docking, molecular dynamics simulation (MD)

## OPEN ACCESS

### Edited by:

Chandrabose Selvaraj,  
Alagappa University, India

### Reviewed by:

Giuseppe Felice Mangiatordi,  
Italian National Research Council, Italy

Ahmed H. E. Hassan,  
Mansoura University, Egypt

### \*Correspondence:

Binqun Luan  
bluan@us.ibm.com

### Specialty section:

This article was submitted to  
Medicinal and Pharmaceutical  
Chemistry,  
a section of the journal  
Frontiers in Chemistry

**Received:** 02 November 2020

**Accepted:** 08 December 2020

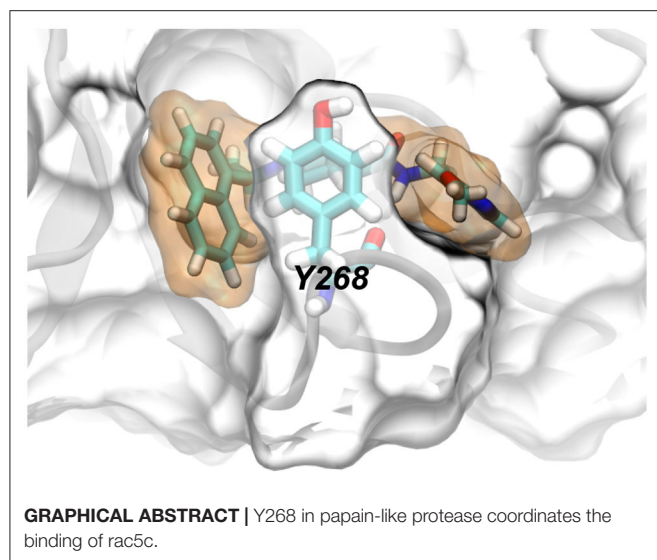
**Published:** 04 February 2021

### Citation:

Huynh T, Cornell W and Luan B (2021)  
*In silico* Exploration of Inhibitors for  
SARS-CoV-2's Papain-Like Protease.  
Front. Chem. 8:624163.  
doi: 10.3389/fchem.2020.624163

## 1. INTRODUCTION

The ongoing COVID-19 pandemic, which emerged in China at the end of 2019, is caused by the third severely pathogenic novel coronavirus since the beginning of the twenty-first century. This new virus known as severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is similar to, yet distinct from, the previous two zoonotic coronaviruses, SARS-CoV and MERS-CoV, which also caused serious and atypical pneumonia (Kuiken et al., 2003; Bermingham et al., 2012; Wu et al., 2020). COVID-19 is a highly contagious disease that can spread very easily between people through liquid droplets (or aerosols) which come out of infected persons when they sneeze, cough, or even talk. People with COVID-19 can spread the virus before they show any symptoms or realize that they are sick, and become especially infectious when they start showing symptoms, including fever, chills, cough, sore throat, congestion, shortness of breath, headache, muscle pain, diarrhea, vomiting, and loss of smell or taste. Most COVID-19 infected individuals will only have mild symptoms (or no symptoms at all). However, those with certain underlying health conditions (such as cardiovascular disease, chronic respiratory disease, or diabetes) can become severely sick with fatal outcomes. So far, the COVID-19 pandemic has unleashed an unprecedented crisis in a very short period of time. Beyond the spread of the disease itself, which has resulted in over 39 million laboratory-confirmed cases of infection globally with over 1 million reported deaths as of October 17, 2020 (<https://covid19.who.int>), efforts to quarantine it have caused far-reaching impact on human physical well-being in societies, on economic development and prosperity, as well as on political systems around the globe. As the entire world battles this pandemic, scientists and



researchers are racing against time to search for a vaccine, drug, or therapy to get rid of the novel coronavirus. Despite all of these efforts, it is believed that the COVID-19 pandemic might not be under full control anytime soon and might play out until next year and beyond (Scudellari, 2020).

Currently, many existing drugs (including remdesivir) have been tested or are being tested in clinics to expedite the discovery of drugs for treating COVID-19 patients. With a known safety profile and bioavailability, an existing FDA-approved drug which shows a strong inhibitory effect on key targets (such as the viral proteases or RNA-dependent RNA polymerase) of SARS-CoV-2 could be quickly repurposed for COVID-19. For example, remdesivir which can inhibit the biological function of RNA-dependent RNA polymerase of SARS-CoV-2 was shown to shorten the hospital stay (Beigel et al., 2020) and was recently approved by the FDA. Meanwhile, papain-like protease (PLpro) and 3-chymotrypsin-like protease (3CLPro or Mpro) are proteins encoded in the SARS-CoV-2 genome that are known to play an important role in the viral replication process and hence are attractive antiviral drug targets to fight against COVID-19 (Luan et al., 2020). So far, Mpro has been extensively investigated to identify potent inhibitors (Huynh et al., 2020a,b; Jin et al., 2020; Zhang et al., 2020), however PLpro was much less studied mainly due to the lack of crystal structures. Previous studies (Barretto et al., 2005; Ratia et al., 2008; Báez-Santos et al., 2015; Shin et al., 2020) on SARS-CoV showed that besides working together with Mpro to process the 16 non-structural proteins (nsps) for assembly of the viral replicase complex to initiate the replication and transcription of the viral genome, PLpro also has significant functional implications on the host innate immune responses. Therefore, it is warranted to search for antiviral drugs to target PLpro, inhibiting not only the viral replication but also the dysregulation of signaling cascades in infected cells to protect the surrounding healthy ones.

Recently, a number of crystal structures (Gao et al., 2020; Rut et al., 2020; Shin et al., 2020) for SARS-CoV-2's PLpro became

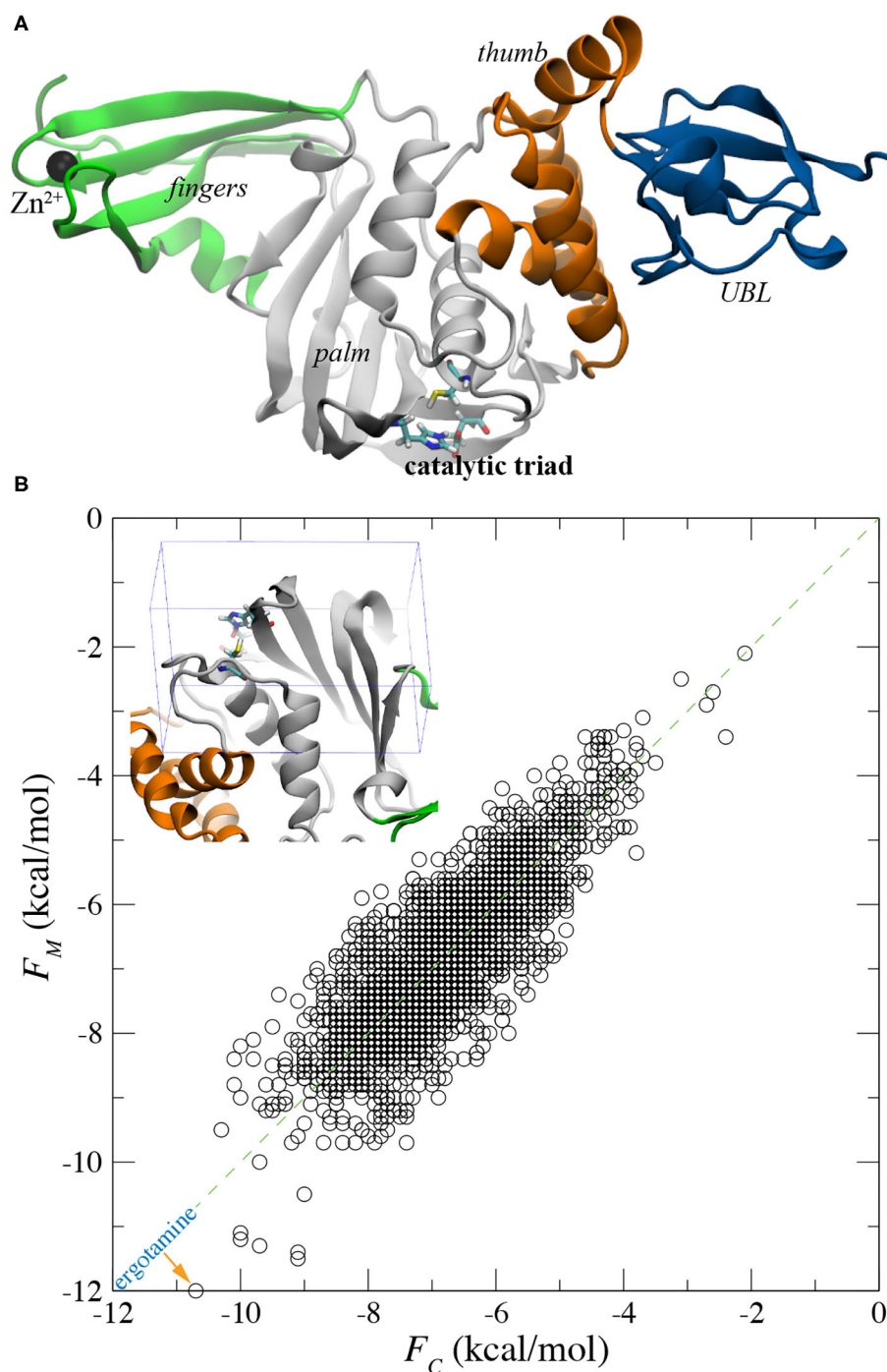
available, which helps unveil its structure-related activities. Similar to PLpro of previous SARS-CoV, PLpro of emerging SARS-CoV-2 contains four domains linked with flexible cords. Namely, fingers domain, palm domain, thumb domain and Ubiquitin-like (UBL) domain (**Figure 1**). Notably, there is a bound  $\text{Zn}^{2+}$  ion in the fingers domain and a catalytic triad (in the active site) in the palm domain. Near the catalytic triad, there exists a groove-like pocket that can harbor a protein ligand and possibly guide the ligand toward the triad for cleavage. For example, the C-terminal region of the protein M15G15 can be bound inside the pocket (PDB entry: 6YVA) (Shin et al., 2020). Due to the good druggability, recent research efforts have been focused on the discovery of inhibitors targeting this substrate binding pocket. So far, some covalently bound peptide ligands [such as VIR250 and VIR251, Rut et al., 2020] as well as small molecules [such as GRL-0617 (Gao et al., 2020; Shin et al., 2020) and rac5c (Klemm et al., 2020)] have been found to occupy the pocket, inhibiting the function of PLpro.

The combination of experimental and computational approaches is an important strategy for the discovery of novel and promising compounds. Molecular docking is a powerful method to predict the conformations of ligands within a macromolecule's substrate-binding site and roughly estimate the corresponding receptor-ligand binding affinity, which is usually applied at the initial stage of structure-based drug design process. While this paper was in the process of review, two docking-based *in silico* studies for repurposing existing drugs to inhibit SARS-CoV-2's PLpro were published (Delre et al., 2020; Ibrahim et al., 2020). Additionally, molecular dynamics (MD) simulations have been shown to yield predictions [such as the binding affinity for a ligand-protein complex Huynh et al., 2020a; Luan and Huynh, 2020] consistent with experimental results, with the ever-improving force fields. Motivated to find promising leads for SARS-CoV-2's PLpro, we carried out *in silico* modeling of all FDA-approved drugs inside the substrate-binding pocket for potential drug repurposing and also investigated the molecular mechanism of a known inhibitor rac5c bound in the same pocket (no crystal structure yet). Thus, this work provides insights derived from the computational studies intended to inform discovery and optimization of lead molecules and to manage expectations for the potential to repurpose approved drugs against this relatively new and less well-studied target.

## 2. RESULTS AND DISCUSSIONS

### 2.1. High-Throughput Screening of All FDA Approved Drugs

For the docking study, we first obtained the atomic (or crystal) structure of SARS-CoV-2's PLpro from the protein data bank (PDB) with the entry code 6WX4, after removing its covalently bound ligand (VIR251). The apo structure of PLpro was either directly utilized in the docking study or was further equilibrated in a 0.15 M NaCl electrolyte (see **Supplementary Figure 1A**) using the MD simulation before its use in the docking study. After about 50 ns of MD simulation, the PLpro's structure was properly equilibrated in the physiology-relevant environment



**FIGURE 1 |** Flexible docking for screening all FDA-approved drugs for binding with PLpro. **(A)** Illustration of PLpro in the cartoon representation. The entire protein contains the ubiquitin-like (UBL), thumb, palm, and fingers domains that are colored in blue, orange, gray and green, respectively. In the active site, the catalytic triad C111-H272-D286 is in the stick representation. The bound  $Zn^{2+}$  ion in the fingers domain is shown as a black sphere. **(B)** A scatter plot showing affinity scores  $F_M$  and  $F_C$  for all FDA approved drugs docked in the active sites of the MD-equilibrated PLpro structure and the crystal PLpro structure, respectively. *Inset*: an illustration of the box around the active site within which all drug molecules were docked.

as evidenced by root-mean-square-deviation (RMSD) values of the protein backbone (**Supplementary Figure 1B**). Due to the presence of multiple domains in flexible PLpro (**Figure 1A**),

RMSD for the backbone of entire PLpro fluctuated between 1.5 and 3.5 Å during the rest of 150 ns of simulation (**Supplementary Figure 1B**). However, for the backbone in the

palm domain (harboring the substrate-binding site), RMSD values saturated at about 1.3 Å (**Supplementary Figure 1B**), indicating that the palm domain is structurally stable. During the simulation, we observed that the  $\text{Zn}^{2+}$  ions were stably coordinated with four deprotonated cysteines (C189, C192, C224, and C226) in the fingers domain (see **Supplementary Video 2**).

With the crystal and MD-equilibrated structures of PLpro, we performed flexible-docking-based screening of the entire FDA-approved drug library (see section 4 for details). The inset of **Figure 1B** shows the pocket-containing simulation box within which conformations of each drug were searched for. Employing 22 IBM power nodes (with twenty 8-way cores in each node), the docking of all FDA-approved drugs in each PLpro structure took about 22 h, compared with 45 min for the rigid docking using same modeling parameters (see section 4). With the binding affinities  $F_M$  and  $F_C$  for drugs docked in MD-equilibrated and crystal structures respectively, we highlight the docking results in a scatter plot shown in **Figure 1B**. Two sets of data ( $F_M$  and  $F_C$ ) appear to be well-correlated, with a majority of data falling on the line defined by  $y = x$  in the plot. The broadening of data indicates the variation of docking scores for two PLpro structures. Remarkably, with the flexible docking enabled, we found the same top candidate (ergotamine) for both PLpro structures, with the affinity scores  $-12.0$  and  $-10.7$  kcal/mol for the MD-equilibrated and crystal structures, respectively. The best conformation of ergotamine in the MD-equilibrated PLpro is illustrated in **Figures 2A,B**, which highlights that the hydrophobic interaction is dominant. Furthermore, we carried out MD simulation to evaluate the stability of ergotamine inside the pocket (see below).

It is worth mentioning that when applying the rigid docking approach outcomes can be different for two PLpro structures. Overall, more data points in the scatter plot lie above the  $y = x$  line (**Supplementary Figure 2**), suggesting that generally  $F_M$  is larger than  $F_C$ . The correlation coefficient  $R^2$  for rigid-docking results is 0.58, which is less than the  $R^2$  value (0.66) for the flexible docking results (**Figure 1B**). Overall, results from the flexible docking are less dependent on the different PLpro structure than the ones from rigid docking. Additionally, these affinity scores from the rigid docking (**Supplementary Figure 2**) are about 2–3 kcal/mol larger (i.e., weaker binding) than those obtained from the flexible docking. The rigid docking yielded two different top candidates for the two PLpro structures, with aprepitant ( $-9.2$  kcal/mol) for the crystal structure and lomitapide ( $-8.7$  kcal/mol) for the MD-equilibrated structure.

## 2.2. Molecular Dynamics Simulations for Ergotamine and rac5c in PLpro' Pocket

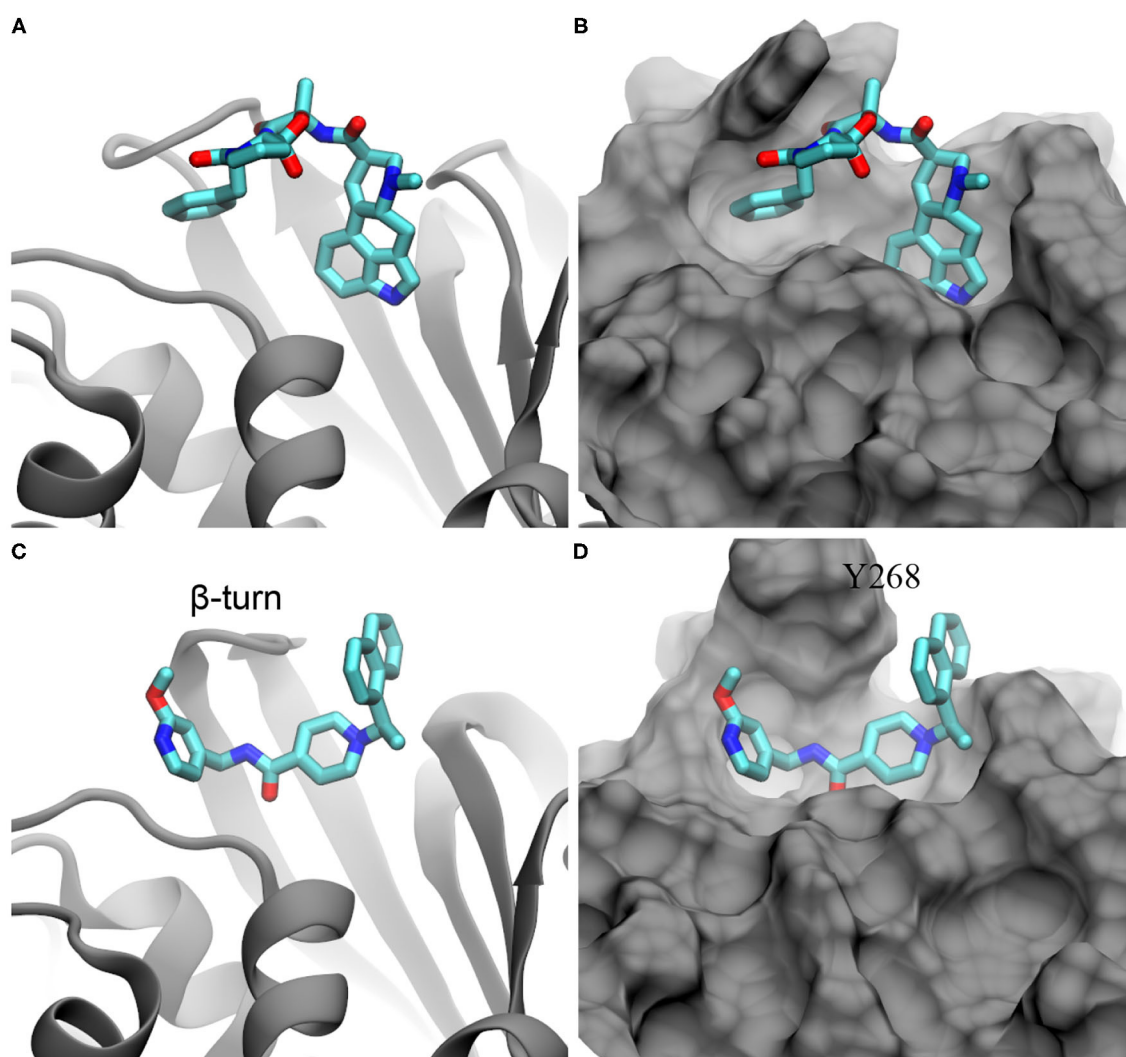
Recently, it was demonstrated in an experiment that rac5c, a known inhibitor of SARS-CoV's PLpro, can also inhibit SARS-CoV-2's PLpro with an  $\text{IC}_{50}$  value of  $0.81 \mu\text{M}$  (Klemm et al., 2020). So far, the crystal structure for the rac5c-PLpro complex is still not available, which hinders the understanding of its binding mechanism. Previously, rac3j (with a molecular structure similar to the one of rac5c) was co-crystallized with SARS-CoV's PLpro (Báez-Santos et al., 2014). However, it is noted

that in that crystal environment PLpro forms a dimer structure (**Supplementary Figure 3**) where two substrate-binding sites in the PLpro dimer are close to each other. Consequently, two rac3j ligands in their respective PLpro's pockets are in contact with each other and thus the rac3j's conformation can be affected by the proximity of a neighboring copy in the crystal structure. Additionally, each PLpro's pocket is occupied not only by a rac3j ligand but also a DMSO molecule (a co-crystallization agent), as shown in **Supplementary Figure 3**. It is not clear whether a rac3j ligand is stabilized due to the presence of the DMSO molecule. To reveal the molecular mechanism for rac5c's binding with PLpro, we carried out MD simulation for the rac5c-PLpro complex in a 0.15 M NaCl electrolyte. The initial complex structure before equilibration in MD simulation is shown in **Figures 2C,D**, built according to the rac3j's conformation in the SARS-CoV's PLpro. It is noted that rac5c is the R-enantiomer for the best fit inside the pocket (Báez-Santos et al., 2014). In addition, ergotamine discovered from the flexible docking occupies a slightly different location inside the pocket, from the one occupied by rac5c in the pocket of PLpro (**Figure 2**).

Following our MD simulation protocol (see section 4) that was calibrated to investigate the binding between drug molecules and SARS-CoV-2's Mpro (Huynh et al., 2020a), we performed all-atom MD simulations to investigate the stability of ergotamine and rac5c inside the PLpro's substrate-binding pocket. **Figure 3A** illustrates the simulation system that PLpro with a bound inhibitor (either ergotamine or rac5c) was solvated in a 0.15 M NaCl electrolyte.

During the MD simulation of the ergotamine-PLpro complex (see Movie PLpro-ergotamine.mpg), we observed that ergotamine was only stable inside the pocket for about 40 ns. During this time, ergotamine formed hydrophobic interactions with Y264, P247, P248, and M208, mainly through its functional group 4,6,6a,7,8,9-hexahydroindolo[4,3-fg]quinoline. The benzene ring on the other end of molecule only weakly interacted with the hydrophobic residue L162. From the trajectory analysis, after aligning all PLpro according to their backbone structures, we calculated RMSD values for ergotamine (non-hydrogen atoms) against its starting conformation. Consistently, the RMSD values remained at 3.5 Å for the first 40 ns of MD simulation (**Figure 3B**). After that, the benzene ring started drifting away from its original position, while the fragment 4,6,6a,7,8,9-hexahydroindolo[4,3-fg]quinoline remained to be stably bound inside the pocket. In this stage (from 40 to 80 ns), the average RMSD value is about 8.0 Å (**Figure 3B**). During the remaining MD simulation, ergotamine gradually diffused away from the pocket as reflected by large RMSD values (**Figure 3B**).

Thus, despite the good docking score ( $-12.0$  kcal/mol), we found that in the physiology-like environment ergotamine was not stable inside the PLpro's pocket. We also carried out MD simulation for aprepitant (the best candidate from the rigid docking study) in the PLpro's pocket and it was not stably bound too (diffusing away in a few nanoseconds). From these MD simulations, it is premature to state that the chance of drug repurposing for PLpro is little, however in recent high-throughput experimental screening (Klemm et al., 2020) it was concluded that the repurposing of existing drug molecules is



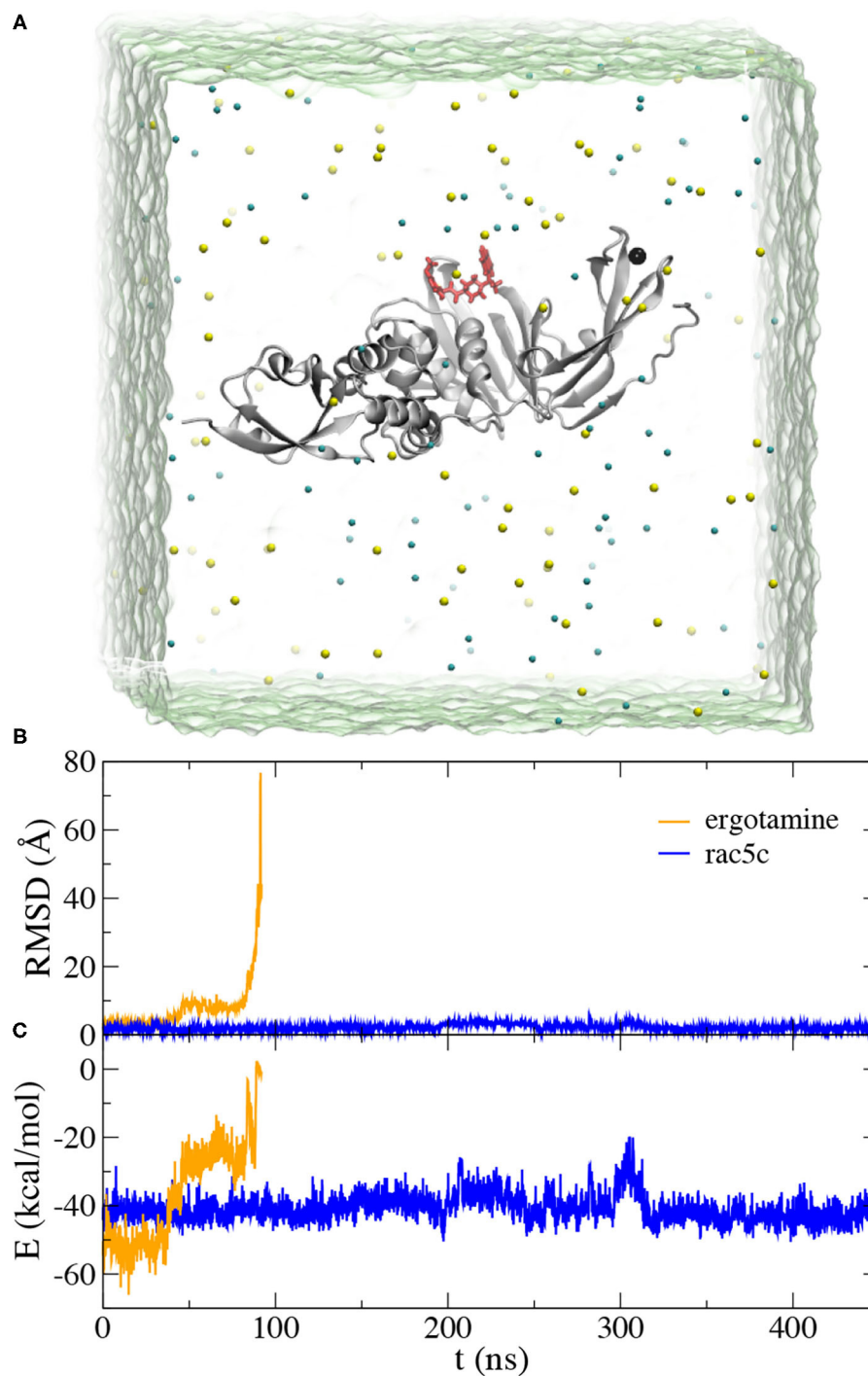
**FIGURE 2 |** Illustration of ergotamine and rac5c (in stick representation) inside the substrate-binding site of PLpro. **(A,B)** The docked pose of ergotamine inside the PLpro's active site. **(C,D)** The hypothesized pose of rac5c inside the PLpro's active site, according to rac3j in the substrate-binding site of PLpro from SARS-CoV (PDB entry: 4OVZ). PLpro is in the cartoon representation in **(A,C)** and is in the molecular surface representation in **(B,D)**.

unlikely to yield drug candidates for PLpro. Therefore, our simulation results are somewhat consistent with the conclusion from the experiment and it might be more promising by exploring/optimizing other known inhibitors for PLpro.

As expected, in MD simulation of the rac5c-PLpro complex we found that rac5c was stably bound inside the PLpro's pocket for the entire ~450 ns simulation (see Movie PLpro-rac5c.mpg). Notably, RMSD values for rac5c kept at about 1.5 Å, corroborating its stable binding inside the PLpro's pocket. Additionally, we calculated interaction energies including van der Waals and electrostatic (dielectric constant~4) ones between rac5c and PLpro (**Figure 3B**). Overall, interaction energies are nearly constant (indicating stable binding) except for one instability event occurring around 300 ns (see more discussion below). For comparison, interaction energies between

ergotamine and PLpro quickly increased to about zero (i.e., no interaction) when ergotamine escaped from the PLpro's pocket (**Figure 3C**). Therefore, our simulation results are consistent with experimental ones showing that rac5c can be an efficacious inhibitor for PLpro (Klemm et al., 2020).

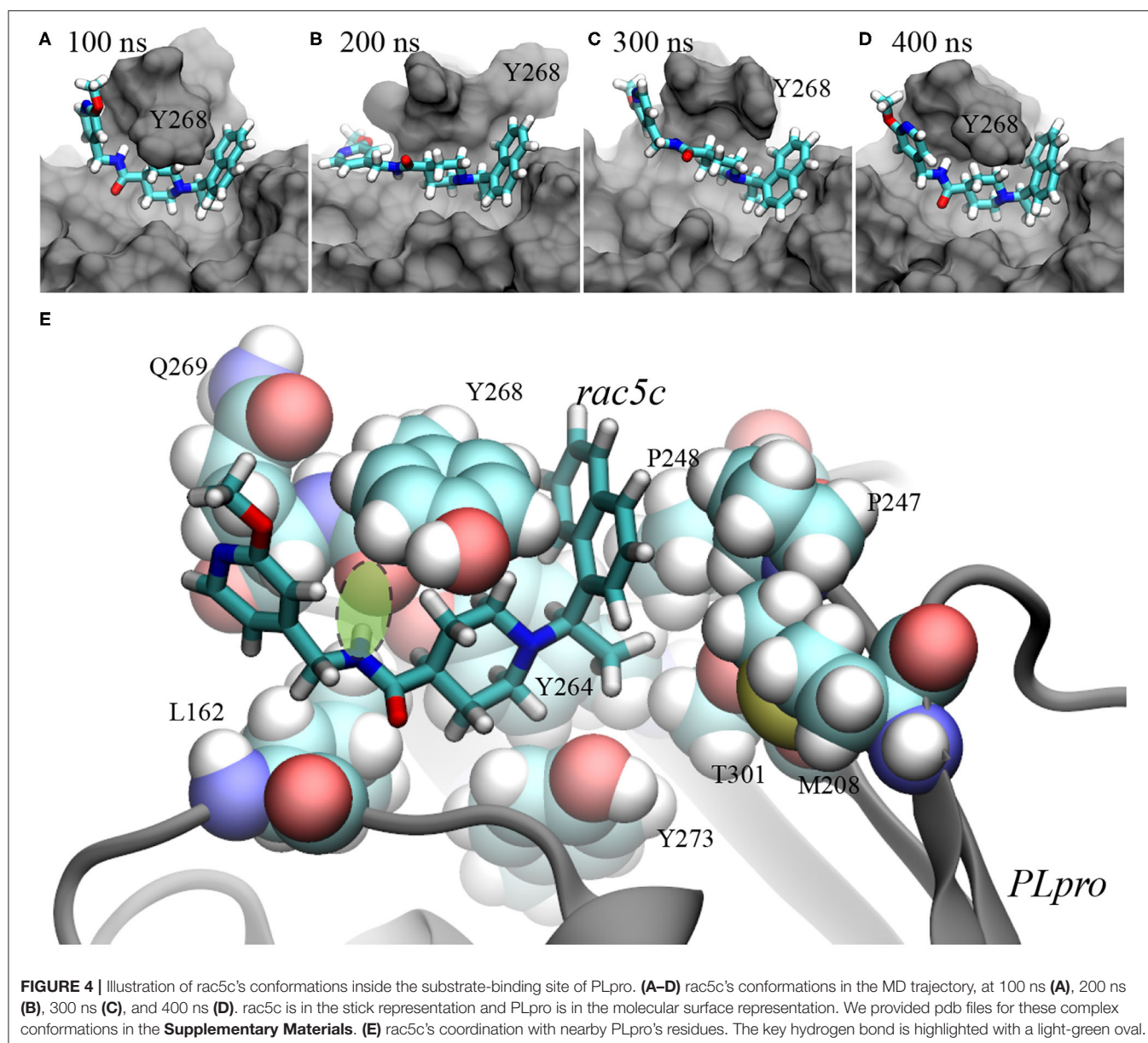
From the simulation trajectory, we observed an interesting dynamic interaction between rac5c and PLpro, especially the  $\beta$ -turn (including residues from 264 to 271 with the sequence YTGNYQCG) as shown in **Figure 2C**. At the beginning (**Figure 2D**), Y268 of PLpro (modeled from the crystal structure with the PDB entry 6WX4) was in an "up" conformation solvated in water. Within 100 ns, **Figure 4A** shows that Y268 moved into a new conformation (the "down" conformation) surrounded by a U-shaped rac5c molecule, which highlights the convergence into a complex structure similar



**FIGURE 3 |** MD simulation for ergotamine and rac5c inside the active site of PLpro. **(A)** Illustration of MD simulation system. PLpro (gray) is in the cartoon representation and the bound ligand (red) is in the stick representation.  $\text{Na}^+$  and  $\text{Cl}^-$  ions are shown as yellow and cyan spheres, respectively. Water (green) is transparent. **(B)** Root-mean-square-deviation (RMSD) values for the bound ligands, ergotamine (orange) and rac5c (blue), when the PLpro's backbones in the simulation trajectory were aligned. **(C)** Interaction energies  $E$  (including van der Waals and electrostatic ones) between the ligand (ergotamine or rac5c) and PLpro calculated from simulation trajectories.

to that of the rac3j-PLpro (SARS-CoV) complex, as shown in **Supplementary Figure 3**. However, the local structure was still not optimized and Y268 rearranged itself back into water at

about 200 ns (**Figure 4B**). The pyridine fragment (with the -OCH<sub>3</sub> group) in rac5c was flexible in the PLpro's pocket. Note that the similar fragment in rac3j was partly stabilized



by a co-crystallization agent DMSO (**Supplementary Figure 3**). Meanwhile, the other fragment naphthalene in *rac5c* was pinned at its original position, preventing the escape of *rac5c* from the pocket.

At 300 ns, Y268 moved into contact with *rac5c* again and was also stacked with the neighboring residue Q269 (**Figure 4C**). Consequently, interaction energies between *rac5c* and PLpro temporarily increased (**Figure 3C**), indicating a weaker binding. After that, Y268 quickly moved into a full contact with *rac5c* again (**Figures 3C, 4D**) and remained at that pose for the rest of MD simulation. For comparison, interaction energies near the end of simulation were more negative (by a few kcal/mol) than those near the beginning (**Figure 3C**), suggesting that the *rac5c*-PLpro complex was structurally improved/optimized during the MD simulation.

The molecular mechanism of binding between *rac5c* and PLpro is illustrated in **Figure 4E**. First, the naphthalene fragment along with an attached methyl group is highly hydrophobic and is favorably surrounded by Y268, P248, P247, T301 and M208 (**Figure 4E**), which accounts for its stability during the entire MD simulation (**Figures 4A–D**). Second, the piperidine fragment in the middle of *rac5c* also interacts hydrophobically with surrounding Y273, Y264, and Y268. Finally, the pyridine fragment (with attached -OCH<sub>3</sub> group) only interacts weakly with nearby alkyl-group-containing side-chains of Q269 and L162, and thus is flexible. Additionally, the peptide-backbone fragment forms a hydrogen bond with the oxygen atom in the backbone of Y268, further stabilizing Y268 in its energetically favorable “down” conformation. All these atomic coordination's allowed *rac5c* to be bound stably inside the PLpro's pocket.

### 3. CONCLUSIONS

In summary, leveraging our powerful HPC cluster, we applied the flexible docking approach to screen the entire FDA-approved drug library for drug candidates that can inhibit PLpro. Due to the urgency of the pandemic, drug repurposing might provide an immediate and temporary solution with already known safety and PK profiles of each approved drug. Unfortunately, both our *in silico* and previous *in vitro* (Klemm et al., 2020) studies suggest that drug repurposing might not be a viable way to discover efficacious inhibitors for PLpro. We further studied rac5c which is known to inhibit PLpros of both SARS-CoV and SARS-CoV-2. Without crystal structures for the rac5c-PLpro complex, the molecular mechanism of rac5c's binding in the PLpro's pocket was not well-understood. Through a long (~450) MD simulation, we confirmed that rac5c can be stably bound inside the PLpro's pocket, which prevents PLpro from performing its biological function of binding and processing its protein targets. From the atomic coordination's between rac5c and PLpro that was revealed from MD simulation, we found that in rac5c only the pyridine fragment (with attached -OCH<sub>3</sub> group) was bound loosely in the pocket, while the remaining part provides essential interactions (both hydrophobic and hydrogen-bond) with PLpro. Thus, it is feasible to optimize the pyridine fragment as well as the attached functional group to enhance the binding affinity.

Supplementary to existing experimental studies for PLpro, our *in silico* investigation provided invaluable insights for drug discovery, from the point of view of atomic-level structures. With collaborative *in silico* and *in vitro/in vivo* efforts, it can be expected to accelerate the discovery of a drug candidate targeting PLpro of SARS-CoV-2.

## 4. METHODS

### 4.1. MD Simulations

MD simulations were performed to investigate the molecular insight and binding stability for ergotamine and rac5c in the active site of PLpro. The crystal structure of SARS-CoV-2's PLpro (Figure 1A) was obtained from the Protein Data Bank (PDB) entry 6WX4. The pdb file of ergotamine (Drugbank accession number DB00696) was converted from the mol2 file downloaded from ZINC 15 database (ZINC ID ZINC000052955754) (Sterling and Irwin, 2015). To get the pdb file for rac5c, we modified the rac3c's pdb file obtained from the crystal structure (PDB entry 4OVZ). We employed the NAMD 2.13 package (Phillips et al., 2005) running on an IBM Power cluster for the MD simulations of PLpro with or without a bound ligand. All systems were solvated in 107 × 107 × 107 Å<sup>3</sup> water (TIP3P model, Jorgensen et al., 1983; Neria et al., 1996) boxes with the NaCl concentration 0.15 M, and the periodic boundary conditions (PDB) were applied in all three dimensions. The CHARMM36 force field (MacKerell et al., 1998) was used for PLpro whereas the standard force field was used for the ions (Beglov and Roux, 1994). As for ergotamine and rac5c, we generated their corresponding force fields using the SwissParam tool (Zoete et al., 2011). The electrostatic interactions were calculated with the particle-mesh Ewald (PME) method with the grid size of 1 Å set in all dimensions, while the van der Waals (VDW) interactions

were handled with a smooth cutoff distance of 10–12 Å. We applied the Langevin thermostat (Allen and Tildesley, 1987) to maintain the temperature T at 300 K as well as the Nosé-Hoover method (Martinetz and Schulten, 1994) to keep the pressure constant at 1 bar. All production runs were carried out in the NVT ensemble with the simulation time-step set to 2 fs (rigid bonds for all).

### 4.2. Docking Simulations

We used AutoDock Vina (Trott and Olson, 2010), the most commonly used open-source software tool, to carry out the molecular docking simulations. In addition to rigid docking, we also used the flexible docking option which allows limited flexibility of selected receptor side-chains aiming to reflect a more realistic ligand-protein interaction environment under reasonable computer processing time. The set of residues allowed to move in the flexible docking simulation includes: L162, D164, R166, M208, S245, Y264, Y268, Y273, T301, and D302. These mobile residues in the binding pocket were selected due to their proximity to the tested ligands. The search box size was set to 22 × 14 × 32 Å, which includes the entire active site of PLpro (see inset in Figure 1B). We set the exhaustiveness and num\_modes to be 8 and 1,000, respectively. To prepare the ligands and target protein for the docking simulations, we used the scripts provided with the AutoDock Tools (Morris et al., 2009) suite such as prepare\_ligand4.py, prepare\_receptor4.py and prepare\_flexreceptor4.py to generate the corresponding input files required by AutoDock Vina. The list of small molecules (FDA-approved drugs) used in this study was downloaded from DrugBank (<https://zinc15.docking.org/catalogs/dbfda>) which contains 1,426 different compounds. Including different protonation states, totally we have 2,173 variants of mol2 files.

## DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding author/s.

## AUTHOR CONTRIBUTIONS

BL designed the project. TH and BL carried out all docking and MD simulations, respectively. TH, WC, and BL discussed the results. TH and BL wrote paper together with inputs from WC. All authors contributed to the article and approved the submitted version.

## FUNDING

The authors declare that this study received funding from IBM. The funder was not involved in the study design, collection, analysis, interpretation of data, the writing of this article or the decision to submit it for publication.

## ACKNOWLEDGMENTS

TH and BL gratefully acknowledge the computing resource from the IBM Cognitive Computing Program.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fchem.2020.624163/full#supplementary-material>

MD simulation of apo PLpro; results from the rigid docking approach; crystal contacts (PDB entry: 4OVZ)

## REFERENCES

- Allen, M. P., and Tildesley, D. J. (1987). *Computer Simulation of Liquids*. New York, NY: Oxford University Press.
- Báez-Santos, Y. M., Barraza, S. J., Wilson, M. W., Agius, M. P., Mielech, A. M., Davis, N. M., et al. (2014). X-ray structural and biological evaluation of a series of potent and highly selective inhibitors of human coronavirus papain-like proteases. *J. Medicin. Chem.* 57, 2393–2412. doi: 10.1021/jm401712t.
- Báez-Santos, Y. M., John, S. E. S., and Mesecar, A. D. (2015). The SARS-coronavirus papain-like protease: structure, function and inhibition by designed antiviral compounds. *Antiviral Res.* 115, 21–38. doi: 10.1016/j.antiviral.2014.12.015.
- Barretto, N., Jukneliene, D., Ratia, K., Chen, Z., Mesecar, A. D., and Baker, S. C. (2015). The papain-like protease of severe acute respiratory syndrome coronavirus has deubiquitinating activity. *J. Virol.* 79, 15189–15198. doi: 10.1128/JVI.79.24.15189-15198.2005.
- Beglov, D., and Roux, B. (1994). Finite representation of an infinite bulk system: solvent boundary potential for computer simulations. *J. Chem. Phys.* 100, 9050–9063.
- Beigel, J. H., Tomashek, K. M., Dodd, L. E., Mehta, A. K., Zingman, B. S., Kalil, A. C., et al. (2020). Remdesivir for the treatment of Covid-19—preliminary report. *N. Engl. J. Med.* 383, 1813–1826. doi: 10.1056/NEJMoa2007764.
- Bermingham, A., Chand, M., Brown, C., Aarons, E., Tong, C., Langrish, C., et al. (2012). Severe respiratory illness caused by a novel coronavirus, in a patient transferred to the United Kingdom from the Middle East, September 2012. *Eurosurveillance* 17:20290. doi: 10.2807/ese.17.40.20290-en
- Delre, P., Caporuscio, F., Saviano, M., and Mangiatordi, G. F. (2020). Repurposing known drugs as covalent and non-covalent inhibitors of the SARS-CoV-2 papain-like protease. *Front. Chem.* 8:594009. doi: 10.3389/fchem.2020.594009.
- Gao, X., Qin, B., Chen, P., Zhu, K., Hou, P., Wojdyla, J. A., et al. (2020). Crystal structure of SARS-CoV-2 papain-like protease. *Acta Pharm. Sin. B.* doi: 10.1016/j.apsb.2020.08.014. [Epub ahead of print].
- Huynh, T., Wang, H., and Luan, B. (2020a). *In silico* exploration of molecular mechanism of clinically oriented drugs for inhibiting SARS-CoV-2's main protease. *J. Phys. Chem. Lett.* 11, 4413–4420. doi: 10.1021/acs.jpclett.0c00994.
- Huynh, T., Wang, H., and Luan, B. (2020b). Structure-based lead optimization of herbal medicine rutin for inhibiting SARS-CoV-2's main protease. *Phys. Chem. Chem. Phys.* 22, 25335–25343. doi: 10.1039/D0CP03867A
- Ibrahim, T. M., Ismail, M. I., Bauer, M. R., Bekhit, A. A., and Boeckler, F. M. (2020). Supporting SARS-CoV-2 papain-like protease drug discovery: *in silico* methods and benchmarking. *Front. Chem.* 8:592289. doi: 10.3389/fchem.2020.592289.
- Jin, Z., Du, X., Xu, Y., Deng, Y., Liu, M., Zhao, Y., et al. (2020). Structure of Mpro from COVID-19 virus and discovery of its inhibitors. *Nature* 582, 289–293. doi: 10.1038/s41586-020-2223-y.
- Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. (1983). Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* 79, 926–935.
- Klemm, T., Ebert, G., Calleja, D. J., Allison, C. C., Richardson, L. W., Bernardini, J. P., et al. (2020). Mechanism and inhibition of the papain-like protease, PLpro, of SARS-CoV-2. *EMBO J.* 39:e106275. doi: 10.15252/embj.2020106275.
- Kuiken, T., Fouchier, R. A., Schutten, M., Rimmelzwaan, G. F., Van Amerongen, G., van Riel, D., et al. (2003). Newly discovered coronavirus as the primary cause of severe acute respiratory syndrome. *Lancet* 362, 263–270. doi: 10.1016/S0140-6736(03)13967-0.
- Luan, B., and Huynh, T. (2020). *In silico* antibody mutagenesis for optimizing its binding to spike protein of severe acute respiratory syndrome coronavirus 2. *J. Phys. Chem. Lett.* 11, 9781–9787. doi: 10.1021/acs.jpclett.0c02706.
- Luan, B., Huynh, T., Cheng, X., Lan, G., and Wang, H.-R. (2020). Targeting proteases for treating COVID-19. *J. Proteome Res.* 19, 4316–4326. doi: 10.1021/acs.jproteome.0c00430.
- MacKerell, A. Jr., Bashford, D., Bellott, M., Dunbrack, R. L. Jr., Evanseck, J., Field, M. J., et al. (1998). All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* 102, 3586–3616.
- Martinetz, T., and Schulten, K. (1994). Topology representing networks. *Neural Netw.* 7, 507–522.
- Morris, G. M., Huey, R., Lindstrom, W., Sanner, M. F., Belew, R. K., Goodsell, D. S., et al. (2009). AutoDock4 and AutoDockTools4: automated docking with selective receptor flexibility. *J. Comput. Chem.* 30, 2785–2791. doi: 10.1002/jcc.21256.
- Neria, E., Fischer, S., and Karplus, M. (1996). Simulation of activation free energies in molecular systems. *J. Chem. Phys.* 105, 1902–1921.
- Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., et al. (2005). Scalable molecular dynamics with NAMD. *J. Comp. Chem.* 26, 1781–1802. doi: 10.1002/jcc.20289
- Ratia, K., Pegan, S., Takayama, J., Sleeman, K., Coughlin, M., Baliji, S., et al. (2008). A noncovalent class of papain-like protease/deubiquitinase inhibitors blocks SARS virus replication. *Proc. Natl. Acad. Sci. U.S.A.* 105, 16119–16124. doi: 10.1073/pnas.0805240105.
- Rut, W., Lv, Z., Zmudzinski, M., Patchett, S., Nayak, D., Snipas, S. J., et al. (2020). Activity profiling and structures of inhibitor-bound SARS-CoV-2-PLpro protease provides a framework for anti-COVID-19 drug design. *Sci. Advances* 6:eabd4596. doi: 10.1126/sciadv.abd4596
- Scudellari, M. (2020). How the pandemic might play out in 2021 and beyond. *Nature* 584, 22–25. doi: 10.1038/d41586-020-02278-5
- Shin, D., Mukherjee, R., Grewe, D., Bojkova, D., Baek, K., Bhattacharya, A., et al. (2020). Papain-like protease regulates SARS-CoV-2 viral spread and innate immunity. *Nature* 587, 657–662. doi: 10.1038/s41586-020-2601-5.
- Sterling, T., and Irwin, J. J. (2015). ZINC 15—ligand discovery for everyone. *J. Chem. Inf. Model* 55, 2324–2337. doi: 10.1021/acs.jcim.5b00559.
- Trott, O., and Olson, A. J. (2010). AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* 31, 455–461. doi: 10.1002/jcc.21334.
- Wu, F., Zhao, S., Yu, B., Chen, Y.-M., Wang, W., Song, Z.-G., et al. (2020). A new coronavirus associated with human respiratory disease in China. *Nature* 579, 265–269. doi: 10.1038/s41586-020-2008-3.
- Zhang, L., Lin, D., Sun, X., Curth, U., Drosten, C., Sauerhering, L., et al. (2020). Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved  $\alpha$ -ketoamide inhibitors. *Science* 368, 409–412. doi: 10.1126/science.abb3405.
- Zoete, V., Cuendet, M. A., Grosdidier, A., and Michielin, O. (2011). SwissParam: a fast force field generation tool for small organic molecules. *J. Comput. Chem.* 32, 2359–2368. doi: 10.1002/jcc.21816.

**Conflict of Interest:** TH, WC, and BL were employed by IBM.

Copyright © 2021 Huynh, Cornell and Luan. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# In Silico Study of Coumarins and Quinolines Derivatives as Potent Inhibitors of SARS-CoV-2 Main Protease

Osvaldo Yañez<sup>1,2,3</sup>, Manuel Isaías Osorio<sup>2,4</sup>, Eugenio Uriarte<sup>5,6</sup>, Carlos Areche<sup>7</sup>, William Tiznado<sup>1</sup>, José M. Pérez-Donoso<sup>2</sup>, Olimpo García-Beltrán<sup>8\*</sup> and Fernando González-Nilo<sup>2\*</sup>

<sup>1</sup>Computational and Theoretical Chemistry Group, Departamento de Ciencias Químicas, Facultad de Ciencias Exactas, Universidad Andres Bello, Santiago, Chile, <sup>2</sup>Center for Bioinformatics and Integrative Biology (CBIB), Facultad de Ciencias de la Vida, Universidad Andres Bello, Santiago, Chile, <sup>3</sup>Center of New Drugs for Hypertension (CENDHY), Santiago, Chile, <sup>4</sup>Facultad de Medicina, Universidad Diego Portales, Santiago, Chile, <sup>5</sup>Departamento Química Orgánica, Facultad de Farmacia, Universidad de Santiago de Compostela, Santiago de Compostela, Spain, <sup>6</sup>Instituto de Ciencias Químicas Aplicadas, Universidad Autónoma de Chile, Santiago de Chile, Chile, <sup>7</sup>Departamento de Química, Facultad de Ciencias, Universidad de Chile, Santiago, Chile, <sup>8</sup>Facultad de Ciencias Naturales y Matemáticas, Universidad de Ibagué, Ibagué, Colombia

## OPEN ACCESS

### Edited by:

Sonia Di Gaetano,  
Italian National Research Council, Italy

### Reviewed by:

Gabriella Gabriella,  
University of Florence, Italy  
Chandrabose Selvaraj,  
Alagappa University, India

### \*Correspondence:

Olimpo García-Beltrán  
jose.garcia@unibague.edu.co  
Fernando González-Nilo  
fernando.gonzalez@unab.cl

### Specialty section:

This article was submitted to  
Medicinal and  
Pharmaceutical Chemistry,  
a section of the journal  
Frontiers in Chemistry

Received: 15 August 2020

Accepted: 16 December 2020

Published: 08 February 2021

### Citation:

Yañez O, Osorio MI, Uriarte E,  
Areche C, Tiznado W,  
Pérez-Donoso JM, García-Beltrán O  
and González-Nilo F (2021) In Silico  
Study of Coumarins and Quinolines  
Derivatives as Potent Inhibitors of  
SARS-CoV-2 Main Protease.  
Front. Chem. 8:595097.  
doi: 10.3389/fchem.2020.595097

The pandemic that started in Wuhan (China) in 2019 has caused a large number of deaths, and infected people around the world due to the absence of effective therapy against coronavirus 2 of the severe acute respiratory syndrome (SARS-CoV-2). Viral maturation requires the activity of the main viral protease ( $M^{pro}$ ), so its inhibition stops the progress of the disease. To evaluate possible inhibitors, a computational model of the SARS-CoV-2 enzyme  $M^{pro}$  was constructed in complex with 26 synthetic ligands derived from coumarins and quinolines. Analysis of simulations of molecular dynamics and molecular docking of the models show a high affinity for the enzyme ( $\Delta E_{binding}$  between  $-5.1$  and  $7.1$  kcal mol<sup>-1</sup>). The six compounds with the highest affinity show  $K_d$  between  $6.26 \times 10^{-6}$  and  $17.2 \times 10^{-6}$ , with binding affinity between  $-20$  and  $-25$  kcal mol<sup>-1</sup>, with ligand efficiency less than 0.3 associated with possible inhibitory candidates. In addition to the high affinity of these compounds for SARS-CoV-2  $M^{pro}$ , low toxicity is expected considering the Lipinski, Veber and Pfizer rules. Therefore, this novel study provides candidate inhibitors that would allow experimental studies which can lead to the development of new treatments for SARS-CoV-2.

**Keywords:** SARS-CoV-2, coumarins, quinolines, protease, molecular dynamics

## INTRODUCTION

In recent years, different viruses have emerged in around of the world. These diseases are a generation of respiratory diseases in infected patients, also due to the rapid dissemination of the diseases. These kinds of viruses include the severe acute respiratory syndrome coronavirus (SARS-CoV), Middle East respiratory syndrome coronavirus (MERS-CoV), avian influenza A/H<sub>7</sub>N<sub>9</sub> and H<sub>5</sub>N<sub>1</sub> viruses, and Nipah virus (Yuen et al., 1998; Peiris et al., 2003; MacNeil and Rollin, 2012; Marsh and Wang, 2012; To et al., 2012; To et al., 2013; Zaki et al., 2012; Chan et al., 2015). The capacity of these viruses to evolve and infect humans has been associated with the close interaction occurring

between human populations and different animal species in markets of densely populated areas (Chan et al., 2015). In December 2019, cases of atypical pneumonia began to be observed in the city of Wuhan (China) (Lu et al., 2020; Zhu et al., 2020). By January 2020, the etiological agent was classified as a new member of the family Coronaviridae and genus  $\beta$ -coronavirus (2019-nCoV) that differ from SARS-CoV and MERS-CoV. The genome of 2019-nCoV shares an 82% sequence identity to SARS-CoV (Elfiky, 2020; Hui et al., 2020; Rothan and Byrareddy, 2020; World Health Organization, 2020). As a matter of fact, its genome has high similarity with the genome of a bat coronavirus (96.2% identity), which has allowed the virus to be associated with a zoonotic origin (Lu et al., 2020; Wu et al., 2020; Zhou et al., 2020). According to the International Committee Virus Taxonomy the new  $\beta$ -coronavirus was called severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and on February 11 the set of symptoms associated with this new virus was designated as COVID-19 by the World Health Organization (WHO) (Zhang et al., 2020b). The evolution of infections at a global level increased as the number of cases and deaths, with the most affected countries being the USA, India, Brazil, Russia, France and United Kingdom (COVID-19 Map - Johns Hopkins Coronavirus Resource Center, 2020), which together have presented more than 50% of the global cases in more than 180 countries, declaring it a pandemic on March 11 (WHO Director-General's opening remarks at the media briefing on COVID-19-20 March, 2020).

Consequently, developed and developing countries are working on the generation of vaccines or antivirals find a solution in the short or medium term. In this context, many governments, medical institutions, and scientists have tried various treatments used for other diseases with promising but so far inconclusive results. These treatments include Chloroquine, Hydroxychloroquine, Camostat, Nafamostat, Umifenovir, Tenofovir, Remdesivir, Sofosbuvir, Galidesivir, Lopinavir, and indinavir, which are used to treat other diseases but have shown a degree of inhibitory activity of SARS-CoV-2 (Chu et al., 2004; Sissoko et al., 2016; Yamamoto et al., 2016; Mulangu et al., 2019; Cortegiani et al., 2020; Deng et al., 2020; Elfiky, 2020; Grein et al., 2020; Hirota et al., 2020; Hoffmann et al., 2020). Each of these compounds has different modes of action and targets such as antiviral drug (RNA-dependent RNA polymerase (RdRp), viral proteases and membrane fusion clathrin-mediated endocytosis (CME), antimalarial drug (elevation of the endosomal pH and ACE2) and serine protease inhibitor (TMPRSS2) (Chu et al., 2004; Sissoko et al., 2016; Yamamoto et al., 2016; Mulangu et al., 2019; Cortegiani et al., 2020; Deng et al., 2020; Grein et al., 2020; Hirota et al., 2020; Hoffmann et al., 2020; McKee et al., 2020).

Like all other coronaviruses, SARS-CoV-2 is composed by single-stranded RNA as their genetic material with an approximate length of 29,891 nucleotides and a 5'-cap structure and 3'-poly-A tail, encoding 9,860 amino acids (Chan et al., 2020). This RNA encodes both the structural and non-structural proteins of the virus. Among the structural proteins, there is the Spike (S) (present in all coronaviruses),

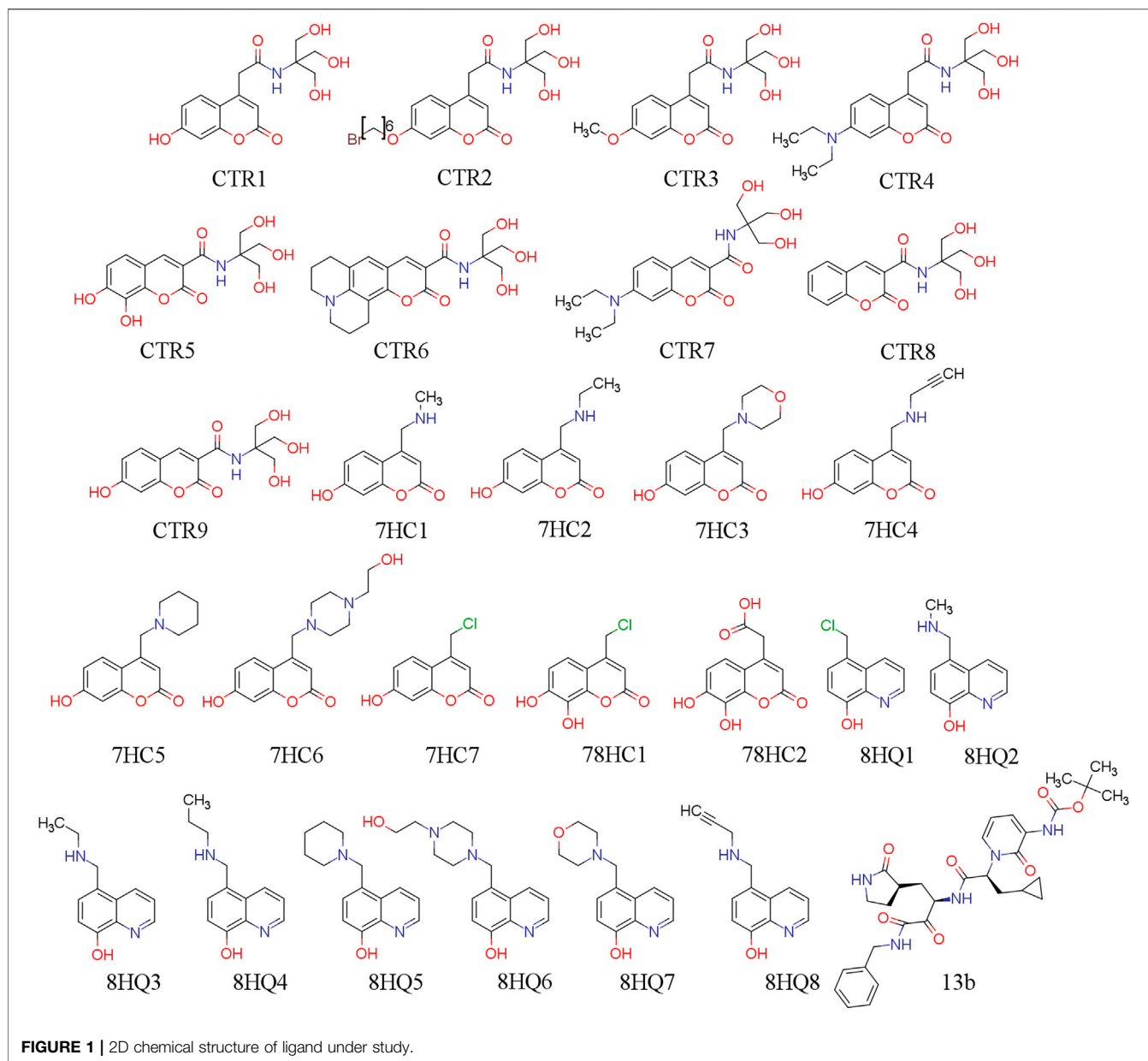
Nucleocapsid (N), Matrix (M), and Envelope (E) (Chan et al., 2020; Walls et al., 2020; Wrapp et al., 2020). Proteases and RNA-dependent RNA polymerase constitute the non-structural proteins of the virus. The genome contains at least six open reading frames (ORFs), the first of these ORF occupies about 60% of the length of the genome and translates two polyproteins known as pp1a and pp1ab, which are processed by the main protease ( $M^{Pro}$ , also called  $3CL^{Pro}$ ) and papain-like proteases (PLPs) (Jin et al., 2020a; Zhang et al., 2020b; Chen et al., 2020). Consequently, inhibiting  $M^{Pro}$  activity blocks virus replication and thus affects the life cycle of SARS-CoV-2. Compounds derived from coumarins and quinoline have been tested against various viruses (McKee et al., 2020; Mishra et al., 2020). Quinolines have recently been used in experimental treatments for SARS-CoV-2 infected persons in several countries (Arshad et al., 2020; Lopez et al., 2020; Mori et al., 2020) and it has been proposed that coumarins inhibit the replication of several viruses including influenza (Pavurala et al., 2018), HIV (Jesumoroti et al., 2019), Dengue (Coulerie et al., 2013), Chikungunya (Hwu et al., 2019), hepatitis (Hwu et al., 2008) and filoviruses (EBOLA, Marburgvirus (MARV) and Cuevavirus) (Liu et al., 2019).

In this study, we evaluated twenty-sixth molecules derived from coumarins and quinolines as promising SARS-CoV-2  $M^{Pro}$  inhibitors, and so, by using computational biochemistry protocols we tried to find the most appropriate molecules that can act as potential anti-SARS-CoV-2 activity drugs. Six of the compounds evaluated are highlighted, which are CTR9, 7HC6, CTR6, 7HC5, 7HC3 and 8HQ6. We performed molecular docking (rigid), efficiency calculations of ligands, pharmacological and toxicological property predictions (ADMET), and molecular dynamics simulations (MD) simulations, together with MM-GBSA binding free energy predictions to identify the binding characteristics for identifying the inhibitors of SARS-CoV-2  $M^{Pro}$ .

## COMPUTATIONAL METHODS

### Compounds Set

In this work, we use twenty-sixth ligands selected for their possible capability to inhibit SARS-CoV-2 protease  $M^{Pro}$ . The coumarins and quinolines derivatives were extracted from two compound series that have been synthesized by the laboratories of applied chemistry of the Universidad de Ibagué, in Ibagué-Colombia. The work was based on strategies that the authors have reported in the literature (García-Beltrán et al., 2013; Mena et al., 2015; Aguirre et al., 2017; García-Beltrán et al., 2017). 13b is our reference molecule and was obtained from the Protein Data Bank (PDB) (Zhang et al., 2020b) (PDB id: 6Y2F). These molecules were designed *in silico* and evaluated using docking methodologies and physicochemical and pharmacokinetic descriptors, and to predict ADME parameters. Their chemical structures are shown in **Figure 1**. Their molecular conformations were optimized using PM6-D3H4 semi-empirical method (Stewart, 2007; Řezáč and Hobza, 2012) as implemented in MOPAC2016 (Stewart, 2016) software. The optimized



**FIGURE 1** | 2D chemical structure of ligand under study.

molecules were used for molecular docking simulations in order to study the interactions established by these compounds in the SARS-CoV-2  $M^{pro}$  pocket.

## Molecular Docking

We made molecular docking analyses to examine the potential binding modes of ligands to the main protease  $M^{pro}$  of SARS-CoV-2, as potential inhibitors. Then, based on structural information obtained from the crystal structures of  $M^{pro}$  in complex with other ligands we established the binding site of the proposed inhibitors for SARS-CoV-2  $M^{pro}$ . In addition to delimit the binding region of possible inhibitors to  $M^{pro}$  (Dai et al., 2020; Jin et al., 2020a; ul Qamar, 2020; Xu et al., 2020; Zhang et al., 2020a, 2020b). We established the key residues for catalysis according to experimental and theoretical data.

Finally, we used AutoDock (v 4.2.1) and AutoDock Vina (Trott and Olson, 2010) for all dockings in this study. The initial 3D inhibitors structures were drawn using Discovery Studio (Dassault Systèmes BIOVIA, 2017) 3.1 (Accelrys, CA) which were optimized (considering the RMS gradient of 0.001 kcal/mol) using the PM6-D3H4 semi-empirical method (Stewart, 2007; Řezáč and Hobza, 2012) implemented in the MOPAC2016 (Stewart, 2016) software. PM6-D3H4 introduces dispersion and hydrogen-bonded corrections to the PM6 method. The ligand files were prepared using the AutoDockTools package (Sanner, 1999) provided by AutoDock through accepting all rotatable bonds; moreover, the atomic charges are computed toward the PM6-D3H4 procedure, and non-polar hydrogen atoms are merged. The semi-empirical method has shown to increase significantly

docking accuracy and cluster population of the most accurate docking (Bikadi and Hazai, 2009; Fanfrlík et al., 2010; Hou et al., 2013). The crystal structure of SARS-CoV-2 M<sup>Pro</sup> (PDB Code: 6YB7, <https://www.rcsb.org/structure/6YB7>), was downloaded from the PDB (Berman et al., 2000). The resolution of the retrieved structure was 1.25 Å. The SARS-CoV-2 M<sup>Pro</sup> was treated with the Schrödinger's (Schrödinger, 2020) Protein Preparation Wizard (Madhavi Sastry et al., 2013); polar hydrogen atoms were added, non-polar hydrogen atoms were merged, and charges were assigned. Docking was treated as rigid and carried out using the empirical free energy function and the Lamarckian Genetic Algorithm provided by AutoDock Vina (Morris et al., 1998). The docking grid dimensions were 30.75 × 30.75 × 30.75 Å, making the binding pocket of SARS-CoV-2 M<sup>Pro</sup> the center of mass between amino-acid residues (Cys145 and His41) of the catalytic site. All other parameters were set as defined by default through AutoDock Vina. Dockings were repeated 50 times with space search exhaustiveness set to 100. The best interaction binding energy (kcal·mol<sup>-1</sup>) was selected for evaluation and analyzed according to the potential intermolecular interactions (protein/ligand), such as hydrogen bonding, hydrophobic interactions and the cation-π, π-π stacking. Docking results 3D representations were used. VMD molecular graphics system (Humphrey et al., 1996).

## Ligand Efficiency Approach

Ligand efficiency (*LE*) calculations were performed using one parameter  $K_d$ . The  $K_d$  parameter corresponds to the dissociation constant between a ligand/protein, and their value indicates the bond strength between the ligand/protein (Abad-Zapatero, 2007; Abad-Zapatero, 2013; Abad-Zapatero et al., 2010). Low values indicate strong binding of the molecule to the protein.  $K_d$  calculations were done using the following equations:

$$\Delta G^0 = -2.303RT \log(K_d) \quad (1)$$

$$K_d = 10^{\frac{\Delta G^0}{2.303RT}} \quad (2)$$

where  $\Delta G^0$  corresponds to binding energy (kcal mol<sup>-1</sup>) obtained from docking experiments,  $R$  is the gas constant whose value is 1.987207 cal/mol K and  $T$  is the temperature in degrees Kelvin. At standard conditions of aqueous solution at 298.15 K, neutral pH and remaining concentrations of 1 M. The ligand efficiency (*LE*) allows us to compare molecules according to their average binding energy (Reynolds et al., 2008; Abad-Zapatero, 2013). Thus, it determined as the ratio of binding energy per non-hydrogen atom, as follows (Abad-Zapatero, 2007; Abad-Zapatero, 2013; Abad-Zapatero et al., 2010; Cavalluzzi et al., 2017):

$$LE = \frac{2.303RT}{HAC} \log(K_d) \quad (3)$$

where  $K_d$  is obtained from eq. (2) and *HAC* denotes the heavy atom count (i.e., number of non-hydrogen atoms) in a ligand.

**TABLE 1** | Empirical rules for predicting oral availability and toxicity of a compound.

Properties	Oral availability	Toxicity	Pfizer 3/75 rules
	Lipinski rules	Veber rules	
<i>MW</i>	≤500	–	–
<i>LogP</i>	≤5	–	≤3
<i>HBA</i>	≤10	–	–
<i>HBD</i>	≤5	–	–
<i>TPSA</i>	–	≤140	≥75
<i>RB</i>	–	≤10	–

*MW*: Molecular weight, *LogP*: octanol/water partition coefficient, *HBA*: Hydrogen Bond Acceptor, *HBD*: Hydrogen Bond Donor, *TPSA*: Topological Polar Surface Area and *RB*: Rotatable Bond.

## ADMET Properties

The purpose of calculating ADMET profiles is to supply, with reasonable accuracy, a preliminary prediction of the *in vivo* behavior of a compound to assess its potential to become a drug (Yu and Adedoyin, 2003). The molecules used in this study were submitted to the calculation of their absorption, distribution, metabolism, excretion and toxicological properties (ADMET). Also, the physicochemical properties such as molecular hydrogen bond acceptor (*HBA*), hydrogen bond donor (*HBD*), molecular weight (*MW*), topological polar surface area (*TPSA*), rotatable bond count (*RB*) and octanol/water partition coefficient (*LogP*) were calculated using SwissADME webserver (Daina et al., 2017). Compound toxicological properties were analyzed taking into account the Lipinski, Veber and Pfizer toxicity empirical rules, see Table 1. (MacLeod-Carey et al., 2020).

## Molecular Dynamics Simulations

MDs calculations were performed for the lowest six binding energy docking and the compound 13b, which is our reference ligand. These calculations were also obtained from Protein Data Bank (PDB id: 6Y2E). The ligands were bound to SARS-CoV-2 M<sup>Pro</sup> protein (PDB ID:6YB7) in aqueous solutions with an explicit solvent TIP3P water model (Neria et al., 1996) (~16,000 water molecules). Protonation states of ionizable residues corresponding to pH 7.0 were determined by H++ web interface for computes pK values of ionizable groups in macromolecules and adds missing hydrogen atoms according to the specified pH of the environment (Anandakrishnan et al., 2012). Besides, NaCl ions were modeled to neutralize the systems and maintain an ionic concentration of 0.15 mol/L. The compounds were parameterized by GAFF Force Field for organic molecules. (Wang et al., 2004; Özpınar et al., 2010), using the Antechamber module in AmberTools18 with AM1-BCC charges, (Jakalian et al., 2000). The protein structures were modeled with the force field ff14SB. (Salomon-Ferrer et al., 2013). The simulations were carried out using a standard MD protocol: (I) Minimization and structural relaxation of water molecules with 2000 steps of minimization (downward step) and MD simulation with an NPT (300 K) assembly by 1,000 ps using harmonic restrictions of 10 kcal mol<sup>-1</sup> Å<sup>-2</sup> for protein and ligand; (II) minimization of the complete

**TABLE 2 |** Molecular docking study between selected ligands and SARS-CoV-2 M<sup>Pro</sup>. Intermolecular docking values, presented with their interaction energy ( $\Delta E_{\text{binding}}$ ), H-bond residues, interacting residues are shown and Ligand efficiency calculation for SARS-CoV-2 M<sup>Pro</sup> complexes.

Compound	Docking results			Ligand Efficiency	
	$\Delta E_{\text{binding}}$ (kcal mol <sup>-1</sup> ) <sup>a</sup>	H-bonds <sup>b</sup>	Residue interactions <sup>b</sup>	$K_d$	LE (kcal mol <sup>-1</sup> )
13b <sup>c</sup>	-7.2	Leu167; Glu166	Arg188; Asn142; Asp187; <b>Cys145</b> ; Gln189; Glu166; Gly143; His164; His163; <b>His41</b> ; Leu167; Leu27; Met165; Met49; Phe140; Pro168; Ser144; Thr25; Thr26	$5.29 \times 10^{-6}$	0.167
CTR6	-7.1	Met49; Gln189; Glu166; Gln192	Arg188; Asn142; <b>Cys145</b> ; Gln189; Gln192; Glu166; Gly143; <b>His41</b> ; Leu167; Met49; Pro168; Thr190; Thr25	$6.26 \times 10^{-6}$	0.253
7HC6	-6.7	Glu166; Ser46	Arg188; <b>Cys145</b> ; Cys44; Gln189; Gln192; Glu166; <b>His41</b> ; Met165; Met49; Ser46; Thr25; Thr45	$12.3 \times 10^{-6}$	0.304
7HC5	-6.6	–	<b>Cys145</b> ; Glu166; His164; His163; <b>His41</b> ; Met165; Met49; Phe140; Ser144; Ser305	$14.6 \times 10^{-6}$	0.347
CTR9	-6.6	Met165; Asn142	Asn142; <b>Cys145</b> ; Cys44; Glu166; Gly143; His164; His163; Leu141; Met165; Met49; Phe140; Ser144; Ser46; Thr25; Thr45	$14.6 \times 10^{-6}$	0.300
7HC3	-6.6	–	<b>Cys145</b> ; Glu166; His164; His163; <b>His41</b> ; Leu141; Met165; Met49; Phe140; Ser144; Ser305	$14.6 \times 10^{-6}$	0.347
8HQ6	-6.5	–	<b>Cys145</b> ; Cys44; Glu166; His164; <b>His41</b> ; Leu141; Met165; Met49; Phe140; Ser144; Ser305; Thr25	$17.2 \times 10^{-6}$	0.309
CTR4	-6.5	His163; Glu166; Met49	Asn142; <b>Cys145</b> ; Cys44; Glu166; His163; His172; Leu141; Met165; Met49; Phe140; Ser144; Ser305; Ser46; Thr25; Thr45	$17.2 \times 10^{-6}$	0.240
CTR5	-6.4	His163; Thr26; Gly143; Thr25	Asn142; <b>Cys145</b> ; Glu166; Gly143; His163; Leu141; Met165; Phe140; Ser144; Thr24; Thr25; Thr26	$20.4 \times 10^{-6}$	0.278
CTR1	-6.3	His41; Glu166; Arg188; Leu27; Thr25	Arg188; Asn142; <b>Cys145</b> ; Gln189; Gln192; Glu166; Gly143; <b>His41</b> ; Leu27; Met165; Met49; Thr190; Thr25; Thr26	$24.1 \times 10^{-6}$	0.273
CTR3	-6.3	Glu166; His163; Asn142	Asn142; <b>Cys145</b> ; Cys44; Glu166; His163; His172; Leu141; Met165; Met49; Phe140; Ser144; Ser305; Ser46; Thr25; Thr45	$24.1 \times 10^{-6}$	0.262
8HQ5	-6.3	Gln189	<b>Cys145</b> ; Gln189; Glu166; His163; Met165; Met49; Phe140; Ser144; Ser305	$24.1 \times 10^{-6}$	0.350
8HQ7	-6.2	–	<b>Cys145</b> ; Gln189; Glu166; His163; Met165; Met49; Phe140; Ser144; Ser305	$28.6 \times 10^{-6}$	0.344
78HC2	-6.1	His163	Asn142; <b>Cys145</b> ; Gln189; Glu166; His163; Met165; Phe140; Ser144	$33.8 \times 10^{-6}$	0.358
CTR7	-6.1	Thr25; Cys44	Arg188; Cys44; Gln189; Gln192; Glu166; <b>His41</b> ; Leu167; Met165; Met49; Pro168; Ser46; Thr190; Thr25; Thr45	$33.8 \times 10^{-6}$	0.234
CTR2	-6.1	Leu141; Thr25; His163; Asn142	Asn142; <b>Cys145</b> ; Glu166; Gly143; His163; Leu141; Leu27; Met165; Met49; Phe140; Ser144; Ser46; Thr25	$33.8 \times 10^{-6}$	0.203
CTR8	-6.0	Gly143	Asn142; <b>Cys145</b> ; Glu166; Gly143; His163; Leu141; Met165; Phe140; Ser144; Thr25; Thr26	$40.1 \times 10^{-6}$	0.285
78HC1	-5.7	–	<b>Cys145</b> ; Glu166; His163; Leu141; Met165; Phe140; Ser144; Ser305	$66.5 \times 10^{-6}$	0.380
7HC2	-5.7	–	<b>Cys145</b> ; Glu166; His164; His163; Leu141; Met165; Phe140; Ser144; Ser305	$66.5 \times 10^{-6}$	0.356
7HC4	-5.7	Phe140; Asn142	Asn142; <b>Cys145</b> ; Glu166; Gly143; His163; <b>His41</b> ; Leu141; Leu27; Met165; Phe140; Ser144	$66.5 \times 10^{-6}$	0.335
7HC1	-5.6	–	<b>Cys145</b> ; Glu166; His164; His163; <b>His41</b> ; Met165; Phe140; Ser144; Ser305	$78.6 \times 10^{-6}$	0.373
8HQ4	-5.6	Glu166	Arg188; <b>Cys145</b> ; Gln189; Glu166; His163; Leu141; Met165; Met49; Phe140; Ser144; Ser305	$78.7 \times 10^{-6}$	0.350
7HC7	-5.5	His163; Phe140	<b>Cys145</b> ; Glu166; His163; Leu141; Met165; Phe140; Ser144	$93.2 \times 10^{-6}$	0.392
8HQ8	-5.4	Met165	Asn142; <b>Cys145</b> ; Glu166; <b>His41</b> ; Met165; Met49	$110 \times 10^{-6}$	0.337
8HQ2	-5.3	Glu166	<b>Cys145</b> ; Glu166; His163; Met165; Phe140; Ser144; Ser305	$130 \times 10^{-6}$	0.378
8HQ3	-5.3	Glu166	<b>Cys145</b> ; Gln189; Glu166; His163; Met165; Met49; Phe140; Ser144; Ser305	$130 \times 10^{-6}$	0.353
8HQ1	-5.1	–	Asn142; Glu166; His164; His163; Leu141; Met165; Phe140; Ser144; Ser305	$182 \times 10^{-6}$	0.392

<sup>a</sup>In each site, the energy was calculated to see which site had the highest degree of union with the ligand.

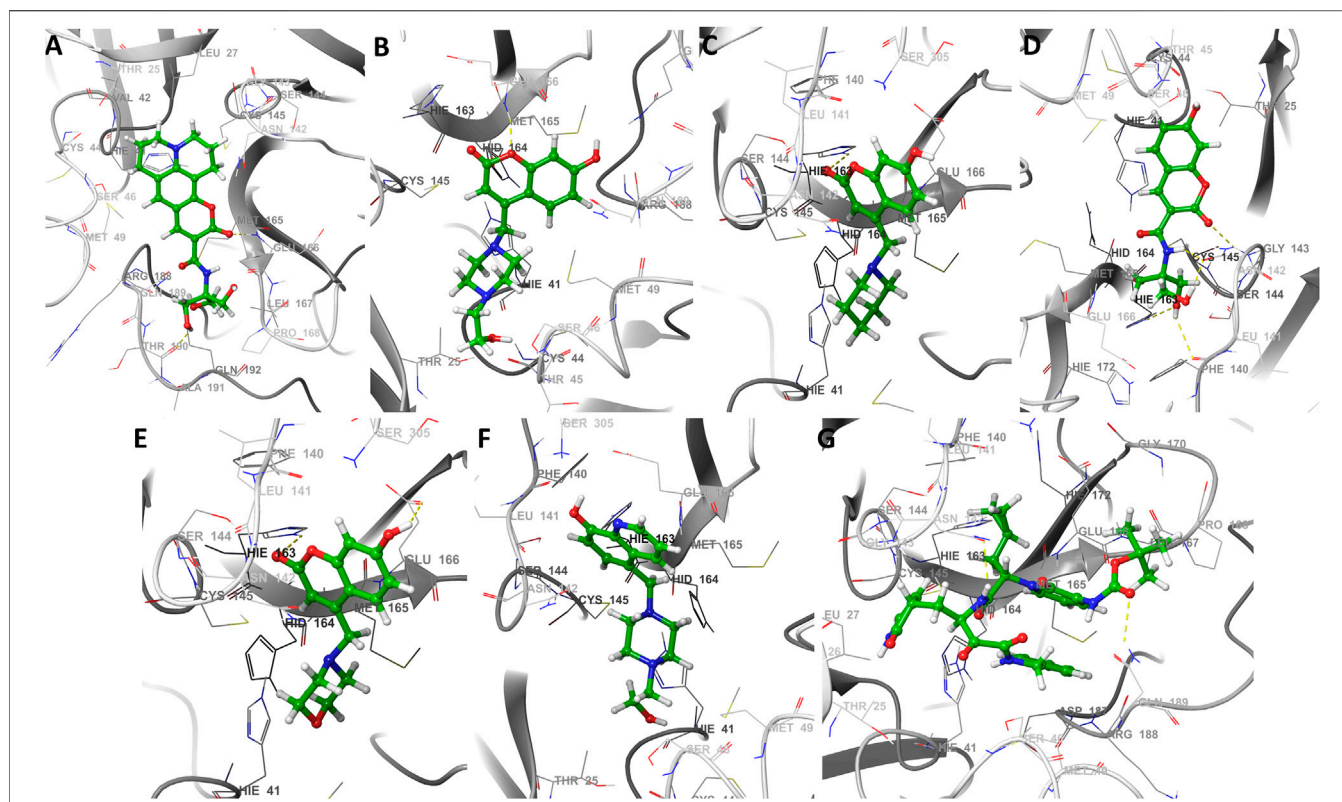
<sup>b</sup>The reason of 3 Å was the length of the Hydrogen bond ranges from 2.6 Å to 3.1 Å based on observations from the PDB.

<sup>c</sup>The ligand 13b is our reference ligand and was obtained from PDB (id: 6Y2E).

His41 and Cys145 residues of the catalytic site are highlighted with bold font.

structure considering 2000 downstream minimization steps and 6,500 steps of conjugate gradient minimization; (III) the minimized systems were progressively heated to 300 K, with harmonic restrictions of 10 kcal mol<sup>-2</sup> Å<sup>-2</sup> in the carbon skeleton and ligand during 0.5 ns; (IV) the system was then balanced by

0.5 ns maintaining the restrictions and then by 5 ns without restrictions to 300 K in a canonical assembly (NVT); and (5) finally, a production dynamic was carried out with an isothermal isobaric assembly (NPT) without restrictions for 200 ns at 310 K and 1 atm with a temporary passage of 2 fs. In the MD



**FIGURE 2 |** The best seven docking poses of different ligands in SARS-CoV-2 M<sup>pro</sup> binding pocket. Snapshots of (A) CTR6, (B) 7HC6, (C) 7HC5, (D) CTR9, (E) 7HC3, (F) 8HQ6 and (G) 13b during docking simulations. The yellow dotted line usually represents intermolecular interactions, like hydrogen bonds.

simulation, the temperature was controlled by the Langevin dynamics with a collision frequency of  $1 \text{ ps}^{-1}$  (NVT) and the pressure with the Berendsen barostat (NPT). Besides, the Particle Mesh Ewald (PME) method with a cut-off value of  $10 \text{ \AA}$  was used to treat nonbonding and long-range electrostatic interactions. All MD simulation calculations were performed using the AMBER-GPU Implementations18 (Mermelstein et al., 2018). Molecular visualization of the systems and MD trajectory analysis was carried out with the VMD software package (Humphrey et al., 1996).

## Free Energy Calculation

The molecular MM/GBSA method was employed to estimate the binding free energy of the protease-ligand complexes. For calculations from a total of 200 ns of MD, the last 50 ns were extracted for analysis, and the explicit water molecules and ions were removed. The MM/GBSA analysis was performed on three subsets of each system: the protein alone, the ligand alone, and the complex (protein-ligand). For each of these subsets, the total free energy ( $\Delta G_{\text{tot}}$ ) was calculated as follows:

$$\Delta G_{\text{tot}} = H_{\text{MM}} + G_{\text{solv}} - T\Delta S_{\text{conf}} \quad (4)$$

where  $H_{\text{MM}}$  is the bonded and Lennard-Jones energy terms;  $G_{\text{solv}}$  is the polar contribution of solvation energy and non-polar contribution to the solvation energy;  $T$  is the temperature; and  $\Delta S_{\text{conf}}$  corresponds to the conformational

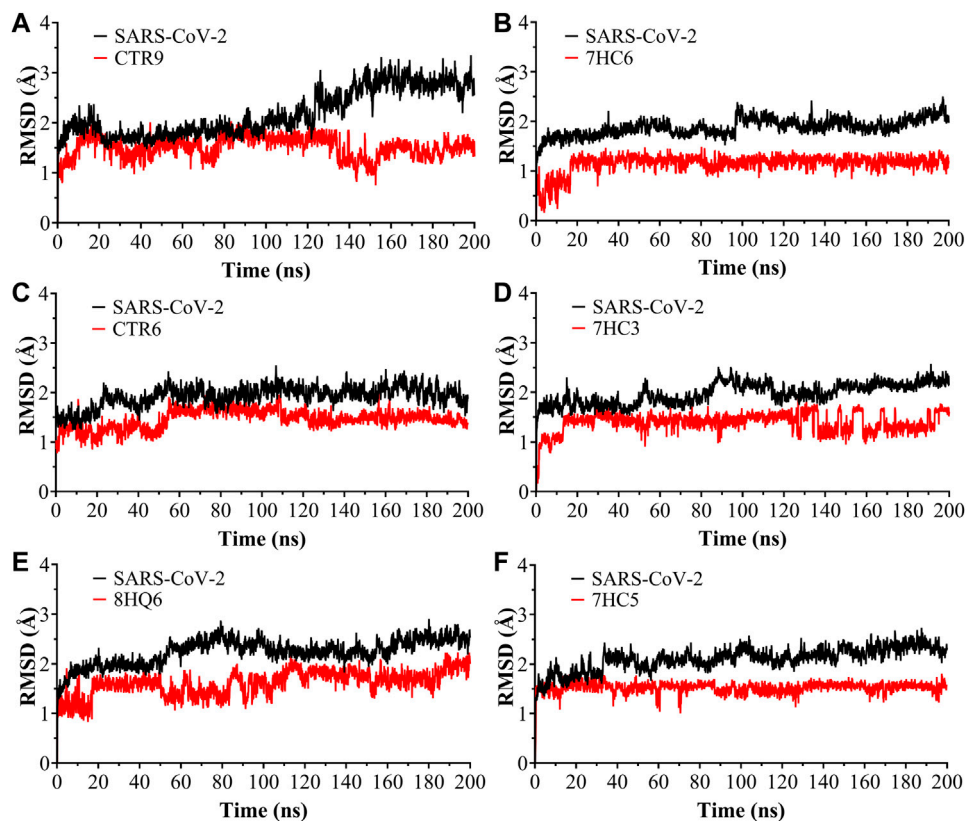
entropy (Hayes and Archontis, 2012). Both  $H_{\text{MM}}$  and  $G_{\text{solv}}$  were calculated using AMBER 18 program with the generalized Born implicit solvent model (Götz et al., 2012; Song et al., 2019).  $\Delta G_{\text{tot}}$  was calculated as a linear function of the solvent-accessible surface area, which was calculated with a probe radius of  $1.4 \text{ \AA}$  (Abroshan et al., 2010). The binding free energy of SARS-CoV-2 M<sup>pro</sup> and ligand complexes ( $\Delta G_{\text{bind}}$ ) were calculated by the difference where  $G_{\text{tot}}$  values are the averages over the simulation.

$$\Delta G_{\text{bind}} = G_{\text{tot}}(\text{complex}) - G_{\text{tot}}(\text{protein}) - G_{\text{tot}}(\text{ligand}) \quad (5)$$

## Non-Covalent Interactions

The principal cluster of main component analysis of trajectory were analyzed with the non-covalent interaction index (NCI) (Johnson et al., 2010; Contreras-García et al., 2011) using NCIPLOT program (Contreras-García et al., 2011) to identify and map non-covalent interactions, such hydrogen bonds, steric repulsion, and van der Waals interactions, using the promolecular densities ( $\rho^{\text{pro}}$ ), computed as the sum of all atomic contributions. The NCI is based on the electron density ( $\rho$ ), its derivatives and the reduced density gradient ( $s$ ). The reduced density gradient is given by:

$$s = \frac{1}{2(3\pi^2)^{1/3} \frac{\nabla \rho}{\rho^{4/3}}} \quad (6)$$



**FIGURE 3** | Root Mean Square Deviation (RMSD) as a function of simulated times for the complexes formed between SARS-CoV-2 M<sup>pro</sup> and (A) CTR9, (B) 7HC6, (C) CTR6, (D) 7HC3, (E) 8HQ6 and (F) 7HC5 molecules.

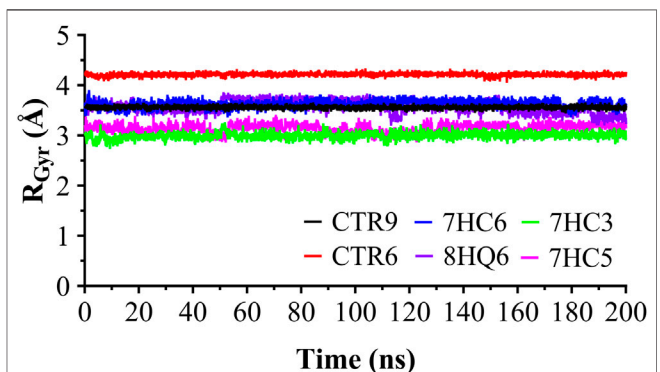
These interactions are local and manifest in real space as low-gradient isosurfaces with low densities which are interpreted and colored according to the corresponding values of  $\text{sign}(\lambda_2)\rho$ . The surfaces are colored on a blue-green-red scale according to the strength and type of interaction. Blue indicates strong attractive interactions, green indicates weak van der Waals interactions, and red indicates a strong non bonded overlap.

## RESULTS AND DISCUSSION

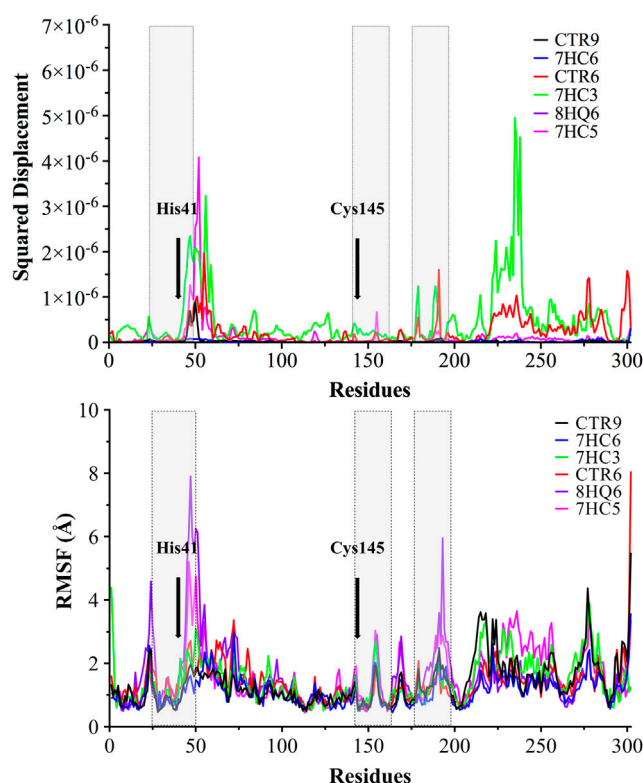
### Molecular Docking Analysis

The docking results, which were conducted to estimate the possible binding sites of potential inhibitors on SARS-CoV-2 M<sup>pro</sup>. The genetic material of SARS-CoV-2 expresses multiple proteins (more than 20 proteins), among these proteins the main protease (M<sup>pro</sup>) is identified, a molecule similar to 3 chymotrypsin (3CL<sup>pro</sup>) that shows a similarity of 96.1% with the 3CL<sup>pro</sup> of SARS-CoV. 3CL<sup>pro</sup> plays a very important role in replication and transcription processes of the virus genome (Hui et al., 2020). Therefore, 3CL<sup>pro</sup> is a strategic drug target in the inhibition of the SARS-CoV cycle. The protease is active as a homodimer, structured by the dimerization of two protomers designated as monomer A and monomer B, and the catalytic dyad in each protomer is defined by

Cys145 and its residues (Zhang et al., 2020b). This has led to the development of multiple studies with experimental and computational approaches in search of possible inhibitors that can effectively block the activity of this protease (Balaramnavar et al., 2020; Dai et al., 2020; Gao et al., 2020; Jin et al., 2020a; Jin et al., 2020b; Ngo et al., 2020; Zhang et al., 2020b). Work with the 3C-like proteinase from SARS coronavirus revealed that the Cys145 residue is key at the active site of 3CL<sup>pro</sup> (Huang et al.,



**FIGURE 4** | Radius of gyration for the SARS-CoV-2 M<sup>pro</sup> in complex with CTR9, 7HC6, CTR6, 7HC3, 8HQ6 and 7HC5, during 200 ns simulation time.



**FIGURE 5 |** Normal Mode Analysis and RMSF of the  $\alpha$ -carbon. A main component analysis was carried out using the last 100 ns trajectories, and the main normal mode of movement was obtained. The displacement was plotted for each residue of SARS-CoV-2 M<sup>Pro</sup> in complex with CTR9, 7HC6, CTR6, 7HC3, 8HQ6 and 7HC5. In grey boxes represented pocket site residues. His41 and Cys145 residues of the catalytic site are highlighted with bold font.

2004), this advance allowed the mentioned residue to be an attractive target for covalent ligands to bind acting as inhibitors of 3CL<sup>Pro</sup>. This amino acid residue is also a popular target for covalent inhibitors because of its intrinsic reactivity at physiological pH (Cuesta et al., 2020). Tung Ngo et al. report in recent studies that additionally Glu166 residue has a prominent and important role in binding ligands to SARS-CoV-2 M<sup>Pro</sup> (Ngo et al., 2020).

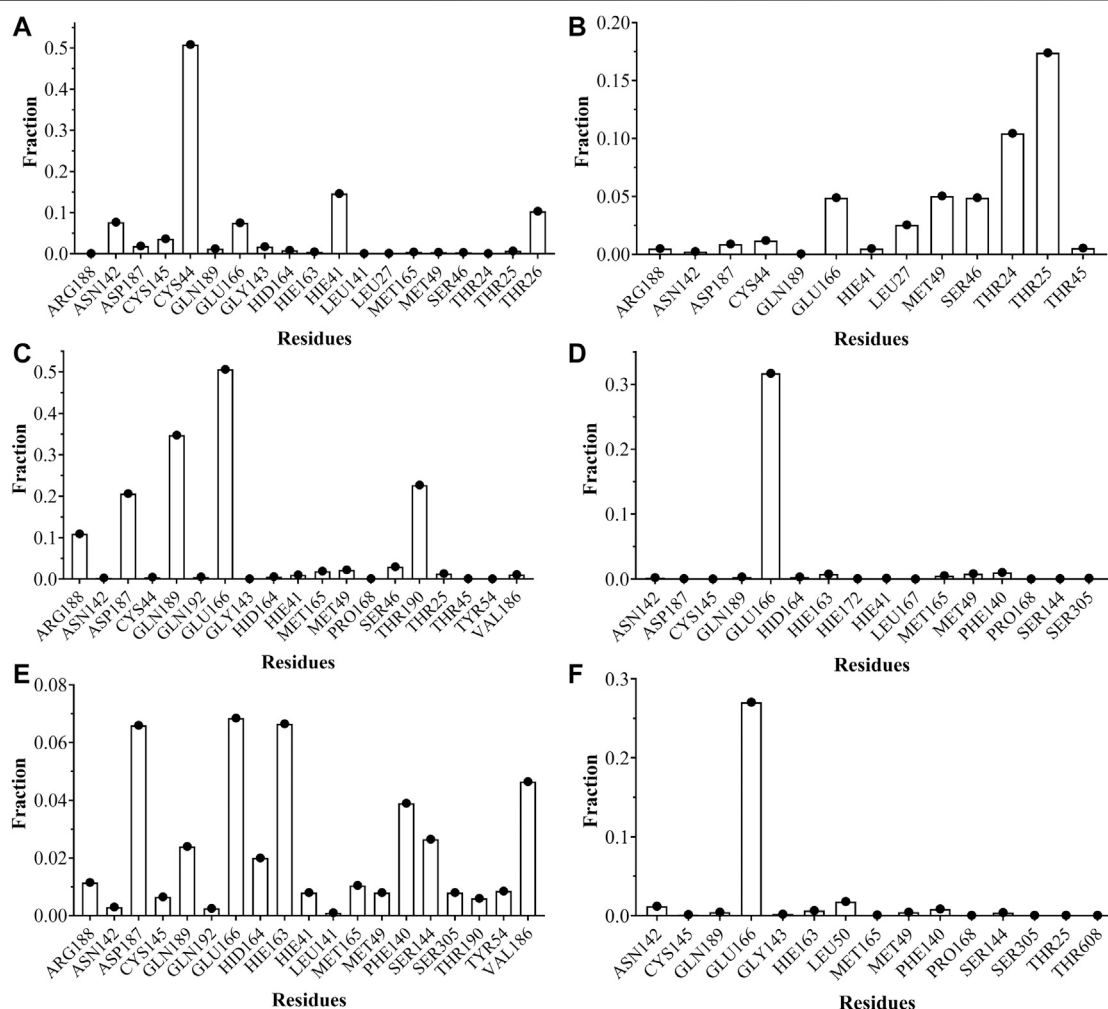
Twenty-seven inhibitors, including the reference ligand 13b were evaluated *in silico* anti-SARS-CoV-2 activity. The results of this study of molecular docking calculations indicate the strong interactions of molecules derived from coumarins and quinolones targets the Cys-His catalytic dyad (Cys145 and His41) in the binding pocket of SARS-CoV-2 M<sup>Pro</sup>. The results of the binding of these molecules are presented in **Table 2**. Meanwhile, the binding to various amino acid residues due to their presence in the conserved region of the active site in all compounds is seen and presents a very important role in enzymatic catalysis.

Docking results with SARS-CoV-2 M<sup>Pro</sup>, indicated that all ligands present binding energies between  $-5.3$  and  $-7.2$  kcal mol<sup>-1</sup> (**Table 2**), with a difference of 0.1 kcal mol<sup>-1</sup> between 13b and CTR6, which means that the level of stability is very similar between these two protease-complex ligands. AutoDock Vina, presents that the ligand 13b attached to the co-crystal, and the re-coupled ligand 13b presents an RMSD value of 3.1 Å, suggesting a partially acceptable value of the

coupling method. Furthermore, it shows that the compound CTR6 gives the lowest energy ( $-7.1$  kcal mol<sup>-1</sup>) in complex with the protease, which is the best score when compared to other docked compounds used in this study. CTR6 gives better score than 7HC6 ( $-6.7$  kcal mol<sup>-1</sup>), 7HC5 ( $-6.6$  kcal mol<sup>-1</sup>), CTR9 ( $-6.6$  kcal mol<sup>-1</sup>), 7HC3 ( $-6.6$  kcal mol<sup>-1</sup>), 8HQ6 ( $-6.5$  kcal mol<sup>-1</sup>) and the other compounds. The coumarins and quinolines derivatives are located inside the protein pocket, by means of electrostatic and hydrophobic interactions with the residues Arg188, Asn142, Asp187, Cys145, Gln189, Glu166, Gly143, His164, His163, His41, Leu167, Leu27, Met165, Met49, Phe140, Pro168, Ser144, Thr25 and Thr26. In relation to 13b, it shows two hydrogen bonding interactions with Leu167 and Glu166 residues. In the case of CTR6, 7HC6 and CTR9 in complex with SARS-CoV-2 M<sup>Pro</sup> (**Figure 2** and **Table 2**) show possible hydrogen-bonding interactions with the residues Ser46, Met49, Asn142, Met165, Glu166, Gln189 and Gln192 (H-donor). This allows us to conclude that the hydrogen bonds and hydrophobic forces, are the majority interactions that dominate these complexes. Interactions between the rest of the compounds and SARS-CoV-2 M<sup>Pro</sup> are reported in **Table 2**.

### Ligand Efficiency Analysis

The parameters dissociation constant ( $K_d$ ) and ligand efficiency ( $LE$ ) were used to compare the affinity of the molecules studied



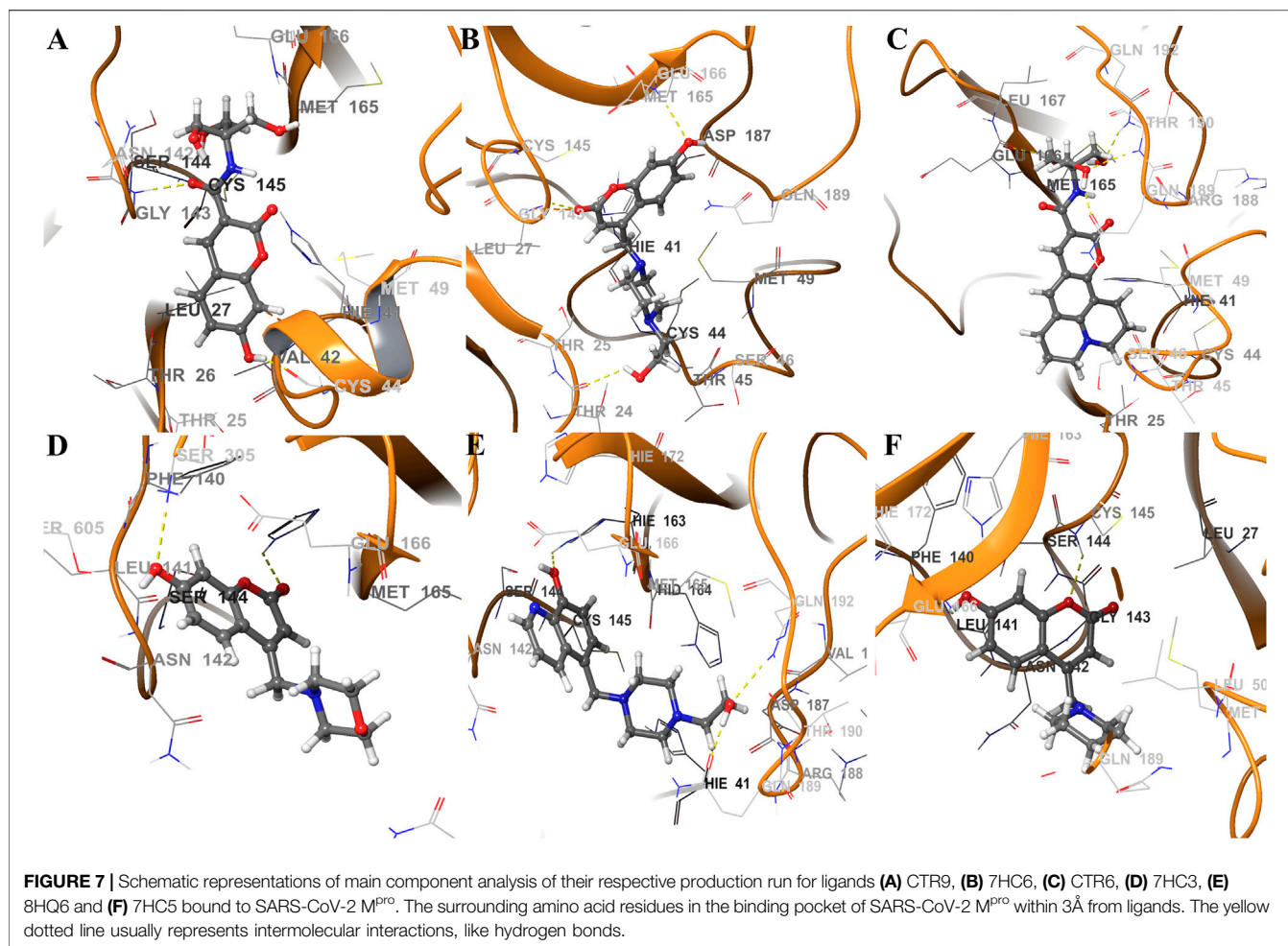
**FIGURE 6 |** Fraction of intermolecular hydrogen bonds for SARS-CoV-2 M<sup>pro</sup> interacting with (A) CTR9, (B) 7HC6, (C) CTR6, (D) 7HC3, (E) 8HQ6 and (F) 7HC5. The graph bar shows the most common hydrogen bonds formed between the residues on the pocket and the inhibitors. Values obtained from CPPTRAJ script in AMBER.

and SARS-CoV-2 M<sup>pro</sup>. The  $K_d$  of a ligand-protease complex, the values shown indicate the strength of the protein-ligand interaction. Very low values are an indicator that the ligand has a very close bond to the protein. LE represents the average bonding energy per non-hydrogen atom, giving standardized values allowing to compare the molecules derived from coumarins and quinolines of different sizes, see **Table 2**. The six best ligands obtained from the docking exhibit low  $K_d$  values, which leads to the conclusion that these complexes are the most stable of the series presented, these ligands are CTR6, 7HC6, 7HC5, CTR9, 7HC3 and 8HQ6, including the reference ligand 13b. The results are coherent with those obtained in the molecular docking, where these complexes according to the values of  $\Delta E_{\text{binding}}$  showed greater stability. Default tolerable values of LE of inhibitor candidate compounds should show LE values  $>0.3 \text{ kcal}\cdot\text{mol}^{-1}$ . According to the values, the compounds 7HC6, 7HC5, CTR9, 7HC3 and 8HQ6 are excellent prospects to be used as SARS-CoV-2 M<sup>pro</sup> inhibitors.

## Molecular Dynamics Simulation and MM/GBSA Analysis

Molecular dynamics simulations were performed in 200 ns to analyze the steady nature and conformations stability of ligand-SARS-CoV-2 M<sup>pro</sup> complexes (ligands: CTR9, 7HC6, CTR6, 7HC3, 8HQ6 and 7HC5).

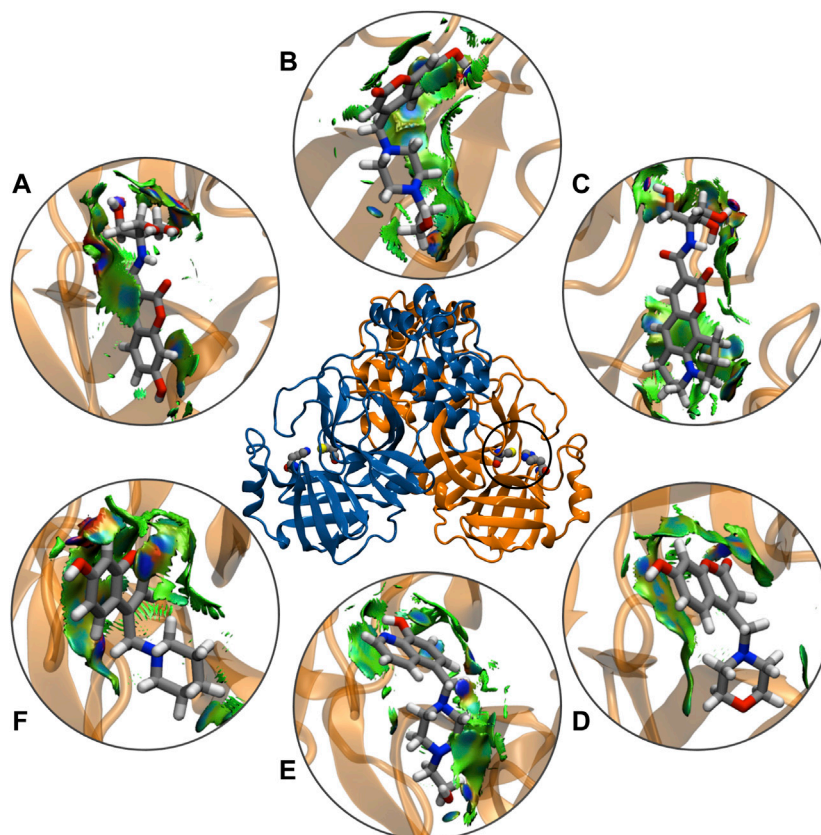
The RMSD was used to estimate the stability of protein-ligand systems. RMSD trajectories of SARS-CoV-2 M<sup>pro</sup>-ligand complexes during 200 ns simulation indicated that the complexes formed with the ligands during the DM simulations have a high stability during the simulation time (**Figure 3**). The structure does not show significant changes, in this case there is an increase of the RMSD until reaching a point in which the values fluctuate around values of 0.5 and 1.8 Å of RMSD. After a Molecular dynamics of 200 ns the structures remain within the parameter that considers the system to be in equilibrium, therefore, no complex suffered structural destabilization during the simulation. Deviations with a maximum difference of 3.0 Å of



RMSD (Carugo, 2003) indicate that the system is in equilibrium, situation that is fulfilled for the simulation of the possible SARS-CoV-2 M<sup>pro</sup> indicating equilibrium states of the ligands within the active site of protease. Also, RMSD curves for 7HC5 and CTR6 are remarkably more stable than those of CTR9, 7HC6, 7HC3 and 8HQ6. To complement the analysis carried out from the RMSD, the study of the Radius of Gyration ( $R_{Gyr}$ ) was carried out for the same runs. From  $R_{Gyr}$  analysis **Figure 4**, we can conclude that the  $R_{Gyr}$  of ligands 7HC5 and 7HC3 have values that oscillate in an interval close to 3.0 Å, for the case of CTR9, 8HQ6 and 7HC6 have values fluctuating in an interval close to 3.5 Å and for the CTR6 ligand it has values higher than 4.0 Å. The stable values during the 200 ns simulation for  $R_{Gyr}$  indicate that ligand binding at the active site of the protein does not induce major conformational changes in the protein structure.

With the purpose of identifying the deviation of the ligand from respect to its initial position and the movement of proteins residues, the Normal Mode Analysis (NMA) and Root Mean Square Fluctuations (RMSF) values were calculated averaging over all the conformations sampled during the last 100 ns simulation. The NMA and RMSF were calculated using the Ca atom of each amino acid residue as a reference and the graph was used to represent the fluctuations in the residue level. NMA plot

in **Figure 5** shows a similar trend of residue fluctuation profile for the complex with an average NMA of  $1.0 \times 10^{-6}$  fluctuations. The two CTR6-protease and 7HC3-protease complexes showed a comparatively higher fluctuation in some residues. This trend in the quadratic displacement figure of the complex suggests that the binding of the six compounds to the protein showed stability and no effect on the flexibility of the protein was observed in the whole range of the simulations. The N-terminus to C-terminus of the proteins normally present great fluctuations and in most cases their movement does not represent importance. As shown in **Figure 5**, the RMSF graph, all the compounds made the residues present the same fluctuation, except for the two 8HQ6-protease and 7HC5-protease complexes showed a comparatively higher fluctuation in some residues. We consider and from the theoretical point of view, that these ligands present movements in the active site to settle in the best orientation, reason why part of the amino acidic residues fluctuate more than normal. However, other ligands exhibit good behavior and their RMSF values show that they can handle the fluctuations of the residues. On the active site of the protease, the fluctuation values of the main residues (Zhang et al., 2020a) (His41, His163, His164, Phe140 and Cys145) of the six selected molecules were similar among them. The results of the MD simulation indicate that two



**FIGURE 8 |** Schematic representations of main component analysis of their respective production run for NCIPLT isosurface gradient (0.6 au) of ligands (A) CTR9, (B) 7HC6, (C) CTR6, (D) 7HC3, (E) 8HQ6 and (F) 7HC5 on the structure of SARS-CoV-2 M<sup>Pro</sup>. The color scale is  $-2.0 < \rho < 2.0$  au.

**TABLE 3 |** Predicted binding free energies ( $\Delta G_{\text{binding}}$ ) calculated from molecular dynamics simulation through the MM/GBSA protocol for SARS-CoV-2 M<sup>Pro</sup> complexes.

Ligand	$\Delta G_{\text{binding}}$ (kcal·mol <sup>-1</sup> )
13b <sup>a</sup>	$-29.1 \pm 0.12$
CTR9	$-24.8 \pm 0.12$
7HC6	$-24.7 \pm 0.07$
CTR6	$-22.8 \pm 0.10$
7HC5	$-22.6 \pm 0.08$
7HC3	$-20.5 \pm 0.11$
8HQ6	$-20.5 \pm 0.11$

<sup>a</sup>The ligand 13b is our reference ligand and was obtained from PDB (id: 6Y2E).

of the ligands obtained from the coupling analysis (CTR9 and 7HC6) remain close to their initial locations even in uncontrolled simulations, which points to the constitution of stable complexes. From these results it can be clearly deduced that it is likely that the molecules CTR9 and 7HC6 play the same role in inhibiting SARS-CoV-2 M<sup>Pro</sup> as 13b.

The analyses of trajectories indicate that during most of the simulation the ligands CTR9, 7HC6, CTR6, 7HC5, 7HC3 and 8HQ6 maintain hydrogen bonds with residues of the active site of SARS-CoV-2 M<sup>Pro</sup>. However, the number of hydrogen bridges formed was different for each ligand (Figure 6). CTR9 formed

three hydrogen bridges between the residues Cys44, His41 and Thr26, highlighting the participation of the residues Glu166, Cys145, Asn142 and Asp187. 7HC6 formed two hydrogen bridges between the residues Thr24 and Thr25, highlighting the participation of the residues Leu27, Ser46, Met49 and Glu166. In the case of CTR6, three hydrogen bridges between the residues Glu166, Gln189 and Thr190 were determined, highlighting the participation of the residues Arg188 and Ser46. 7HC3 formed one hydrogen bridge with the residue Glu166, highlighting the participation of the residues His163, Met49 and Phe140. Three hydrogen bridges are formed between 8HQ6 and the residues Asp187, Glu166 and His163, highlighting the participation of the residues Gln189, Phe140, Ser144 and Val186. Finally, 7HC5 formed one hydrogen bridge with the Glu166 residue, highlighting the participation of the residues Leu50, Asn142 and Phe140. These residues, see Figures 6 and 7, are consistent with previous theoretical-experimental studies carried out by Dai et al. (Dai et al., 2020), where they detail the interaction that some of the synthesized compounds have with the active site of the protease.

The noncovalent interactions analysis labeled all the hydrogen-bonding interactions in total agreement with the molecular dynamics simulations, providing a qualitative confirmation of these interactions, using a topological and

**TABLE 4 |** ADME molecular descriptors of compounds designed to inhibit SARS-CoV-2 M<sup>Pro</sup>.

Compound	MW (g/mol)	LogP	HBA	HBD	TPSA (Å <sup>2</sup> )	RB
13b	593.67	2.27	7	4	164.70	17
CTR6	388.41	0.94	6	4	123.24	6
7HC6	304.34	0.93	6	2	77.15	4
7HC5	259.3	2.27	4	1	53.68	2
CTR9	309.27	-0.16	7	5	140.23	6
7HC3	261.27	1.40	5	1	62.91	2
8HQ6	287.36	1.16	5	2	59.83	4
CTR4	378.42	0.65	6	4	123.24	10
CTR5	325.27	-0.71	8	6	160.46	6
CTR1	323.3	-0.31	7	5	140.23	7
CTR3	337.32	0.02	7	4	129.23	8
8HQ5	242.32	2.52	3	1	36.36	2
8HQ7	244.29	1.65	4	1	45.59	2
78HC2	236.18	0.62	6	3	107.97	2
CTR7	364.39	0.83	6	4	123.24	9
CTR2	486.35	1.99	7	4	129.23	14
CTR8	293.27	0.31	6	4	120.00	6
78HC1	226.61	1.54	4	2	70.67	1
7HC2	219.24	1.63	4	2	62.47	3
7HC4	229.23	1.61	4	2	62.47	3
7HC1	205.21	1.29	4	2	62.47	2
8HQ4	216.28	2.22	3	2	45.15	4
7HC7	210.61	2.03	3	1	50.44	1
8HQ8	212.25	1.85	3	2	45.15	3
8HQ2	188.23	1.53	3	2	45.15	2
8HQ3	202.25	1.85	3	2	45.15	3
8HQ1	193.63	2.30	2	1	33.12	1

visual analysis of a scalar field related to the electron density (Figure 8). These results suggest that the better inhibitors character is due to direct mechanisms.

Finally, the binding free energy (MM-GBSA) was estimated subsequent to the MD simulation; the last 50 ns for all the complexes and the results are given in Table 3. CTR9 and 7HC6 compounds depicted the lowest binding free energy ( $-24.8$  and  $-24.7$  kcal mol<sup>-1</sup>) with SARS-CoV-2 M<sup>Pro</sup>, while the compounds CTR6, 7HC5, 7HC3 and 8HQ6 showed relatively higher binding energy ( $-24.7$ ,  $-22.8$ ,  $-22.6$ ,  $-20.5$  and  $-20.5$  kcal mol<sup>-1</sup>). On the other hand, the reference compound 13b showed the lowest binding free energy ( $-29.1$  kcal mol<sup>-1</sup>) with the SARS-CoV-2 M<sup>Pro</sup> in comparison to compound CTR9, with a slight difference of 4.3 kcal mol<sup>-1</sup>. Although compound 13b has a lowest binding free energy, it presents the problem of breaking the rules of Lipinski and Veber rules, however compound CTR9 does not.

## ADMET Properties

In the search for new drugs, safety is very important and the regulations related to ADMET (Absorption, Distribution, Metabolism, Excretion and Toxicity), most of the time are the cause for a drug to fail. Therefore, it is of utmost importance to identify aspects such as the toxicity of compounds in early stages of development and thus avoid the loss of resources and time. (Sivamani et al., 2012). To evaluate the best ligands as potential anti-SARS-CoV-2 activity drugs; we have calculated some pharmacokinetic properties (Table 4). These results were compared to Veber's (Veber et al., 2002), Pfizer's (Hughes

et al., 2008) and Lipinski's rule (Lipinski et al., 2001). in the development of new drugs if a molecule complies only with one of the Lipinski's rule is not an appropriate candidate and it is not relevant to continue with the study, however, by presenting a greater number of rules the probabilities of being a candidate begin to increase and deepen their study. In accordance with Veber's rule, if a compound does not satisfy at least two parameters, it is not a candidate for further development. In addition, Pfizer 3/75 toxicity rules have also been taken into account in this study, concluding that if any of the proposed ligands do not meet the established parameters, then it is not a suitable candidate.

ADME prediction showed that in most cases, all the compounds proposed in this study satisfy with the Veber's, Pfizer's and Lipinski's rule. This suggests that these ligands could be safe molecules for use as anti-SARS-CoV-2 activity drugs. In the case of reference ligand 13b presents a violation of Lipinski's and Veber's rule, due to their molecular weight, topological polar surface area and rotatable bond count. The values of these properties are higher than the admissible limit, making this substance fat-soluble which indicates a tendency to be more toxic and less selective to their target. In the case of the six best compounds found in the docking simulations (CTR6, 7HC6, 7HC5, CTR9, 7HC3 and 8HQ6), all of them satisfactorily meet the Veber's, Pfizer's and Lipinski's rule. These compounds represent most promising compounds to molecular dynamics simulation and MM-GBSA.

## CONCLUSION

This paper predicts that compounds derived from coumarins and quinolines that can be successfully potential drugs to treat viral diseases such as COVID-19. Herein, we used a computational chemistry protocol to identify the ligands most promising candidates that may inhibit main protease of SARS-CoV-2 activity determined by means of this protocol involves of molecular dockings, molecular dynamics simulations, MM-GBSA, NCI and ADMET properties to predict whether these compounds are appropriate to be utilized in an anti-COVID-19 therapy. We identified six compounds (CTR9, 7HC6, CTR6, 7HC5, 7HC3, 8HQ6) that are already synthesized (García-Beltrán et al., 2013; Mena et al., 2015; Aguirre et al., 2017; García-Beltrán et al., 2017) with a potential inhibition of main protease of SARS-CoV-2. These compounds might be repurposed against COVID-19. These hits were described as drug-like compounds and showed harmless ADMET properties and may aid in developing and optimizing more efficient and potent COVID-19 inhibitors. Trajectory analysis showed that the studied complexes display structural stability during the MD runs. These results encourage further *in vitro* and *in vivo* investigations and also preventively boost the traditional use of coumarins and quinolines derivatives preventively. We anticipate that the insights obtained in the present study may prove valuable for researching and developing novel anti-COVID-19 therapeutic agents in the future.

## DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding author.

## AUTHOR CONTRIBUTIONS

OY, OG-B, CA, WT and FG-N contributed to the conception and design of the study; OY, MIO and JPD performed the theoretical calculations; OY, EU, MO and JP-D organized the database; OY, MO, JP-D, CA and OG-B wrote the first draft of the manuscript; OY, OG-B, and FG-N wrote sections of the

manuscript. All authors contributed to manuscript revision, read and approved the submitted version.

## ACKNOWLEDGMENTS

The authors express their thanks to Grant RC-FP44842-212-2018 Colombia científica and Universidad de Ibagué and for the financial support of the ANID/PIA/ACT192144. OG-B Thank funding from the Ministry of Science, Technology and Innovation, the Ministry of Education, the Ministry of Industry, Commerce and Tourism, and ICETEX, Programme Ecosistema Científico-Colombia Científica, from the Francisco José de Caldas Fund, Grant RC-FP44842-212-2018.

## REFERENCES

- Abad-Zapatero, C. (2013). *Ligand efficiency indices for drug Discovery*, in Editor C. Abad-Zapatero (San Diego: Academic Press). doi:10.1016/B978-0-12-404635-1.00009-8
- Abad-Zapatero, C. (2007). Ligand efficiency indices for effective drug discovery. *Expet Opin. Drug Discov.* 2, 469–488. doi:10.1517/17460441.2.4.469
- Abad-Zapatero, C., Perišić, O., Wass, J., Bento, A. P., Overington, J., Al-Lazikani, B., et al. (2010). Ligand efficiency indices for an effective mapping of chemico-biological space: the concept of an atlas-like representation. *Drug Discov. Today* 15, 804–811. doi:10.1016/J.DRUDIS.2010.08.004
- Abroshan, H., Akbarzadeh, H., and Parsafar, G. A. (2010). Molecular dynamics simulation and MM-PBSA calculations of sickle cell hemoglobin in dimer form with Val, Trp, or Phe at the lateral contact. *J. Phys. Org. Chem.* 23, 866–877. doi:10.1002/poc.1679
- Aguirre, P., García-Blrán, O., Tapia, V., Muñoz, Y., Cassels, B. K., and Núñez, M. T. (2017). Neuroprotective effect of a new 7,8-Dihydroxycoumarin-based Fe<sup>2+</sup>/Cu<sup>2+</sup> chelator in cell and animal models of Parkinson's disease. *ACS Chem. Neurosci.* 8, 178–185. doi:10.1021/acschemneuro.6b00309
- Anandakrishnan, R., Aguilar, B., and Onufriev, A. V. (2012). H++ 3.0: automating pK prediction and the preparation of biomolecular structures for atomistic molecular modeling and simulations. *Nucleic Acids Res.* 40, W537–W541. doi:10.1093/nar/gks375
- Arshad, S., Kilgore, P., Chaudhry, Z. S., Jacobsen, G., Wang, D. D., Huitsing, K., et al. (2020). Treatment with hydroxychloroquine, azithromycin, and combination in patients hospitalized with COVID-19. *Int. J. Infect. Dis.* 97, 396–403. doi:10.1016/j.ijid.2020.06.099
- Balaramnavar, V. M., Ahmad, K., Saeed, M., Ahmad, I., Kamal, M., and Jawed, T. (2020). Pharmacophore-based approaches in the rational repurposing technique for FDA approved drugs targeting SARS-CoV-2 Mpro. *RSC Adv.* 10, 40264–40275. doi:10.1039/D0RA06038K
- Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., et al. (2000). The protein Data Bank. *Nucleic Acids Res.* 28, 235–242. doi:10.1093/nar/28.1.235
- Bikadi, Z., and Hazai, E. (2009). Application of the PM6 semi-empirical method to modeling proteins enhances docking accuracy of AutoDock. *J. Cheminf.* 1, 1–16. doi:10.1186/1758-2946-1-15
- Carugo, O. (2003). How root-mean-square distance (r.m.s.d.) values depend on the resolution of protein structures that are compared. *J. Appl. Crystallogr.* 36, 125–128. doi:10.1107/S0021889802020502
- Cavalluzzi, M. M., Mangiatordi, G. F., Nicolotti, O., and Lentini, G. (2017). Ligand efficiency metrics in drug discovery: the pros and cons from a practical perspective. *Expet Opin. Drug Discov.* 12, 1087–1104. doi:10.1080/17460441.2017.1365056
- Chan, J. F.-W., Kok, K.-H., Zhu, Z., Chu, H., To, K. K.-W., Yuan, S., et al. (2020). Genomic characterization of the 2019 novel human-pathogenic coronavirus isolated from a patient with atypical pneumonia after visiting Wuhan. *Emerg. Microb. Infect.* 9, 221–236. doi:10.1080/22221751.2020.1719902
- Chan, J. F. W., Lau, S. K. P., To, K. K. W., Cheng, V. C. C., Woo, P. C. Y., and Yuen, K.-Y. (2015). Middle East respiratory syndrome coronavirus: another zoonotic betacoronavirus causing SARS-like disease. *Clin. Microbiol. Rev.* 28, 465–522. doi:10.1128/CMR.00102-14
- Chen, Y., Liu, Q., and Guo, D. (2020). Emerging coronaviruses: genome structure, replication, and pathogenesis. *J. Med. Virol.* 92, 418–423. doi:10.1002/jmv.25681
- Chu, C. M., Cheng, V. C. C., Hung, I. F. N., Wong, M. M. L., Chan, K. H., Chan, K. S., et al. (2004). Role of lopinavir/ritonavir in the treatment of SARS: initial virological and clinical findings. *Thorax* 59, 252–256. doi:10.1136/thorax.2003.012658
- Contreras-García, J., Johnson, E. R., Keinan, S., Chaudret, R., Piquemal, J.-P., Beratan, D. N., et al. (2011). NCIPLOT: a program for plotting noncovalent interaction regions. *J. Chem. Theor. Comput.* 7, 625–632. doi:10.1021/ct100641a
- Cortegiani, A., Ingoglia, G., Ippolito, M., Giarratano, A., and Einav, S. (2020). A systematic review on the efficacy and safety of chloroquine for the treatment of COVID-19. *J. Crit. Care* 21, 684. doi:10.1016/j.jccr.2020.03.005
- Coulerie, P., Maciuk, A., Lebouvier, N., Hnawia, E., Guillemot, J. C., Canard, B., et al. (2013). Phytochemical study of myrtopsis corymbosa, perspectives for anti-dengue natural compound research. *Record Nat. Prod.* 7, 250–253.
- COVID-19 Map - Johns Hopkins Coronavirus Resource Center (2020). <https://coronavirus.jhu.edu/map.html>. [Accessed]
- Cuesta, A., Wan, X., Burlingame, A. L., and Taunton, J. (2020). Ligand conformational bias Drives enantioselective modification of a surface-exposed lysine on Hsp90. *J. Am. Chem. Soc.* 142, 3392–3400. doi:10.1021/jacs.9b09684
- Dai, W., Zhang, B., Su, H., Li, J., Zhao, Y., Xie, X., et al. (2020). Structure-based design of antiviral drug candidates targeting the SARS-CoV-2 main protease. *Science* (80-). 4489, 1–10. doi:10.1126/science.abb4489
- Daina, A., Michielin, O., and Zoete, V. (2017). SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. *Sci. Rep.* 7, 1–13. doi:10.1038/srep42717
- Dassault Systèmes BIOVIA. (2017). Discovery Studio modeling environment. *BIOVIA*. 42, 717.
- Deng, L., Li, C., Zeng, Q., Liu, X., Li, X., Zhang, H., et al. (2020). Arbidol combined with LPV/r versus LPV/r alone against Corona Virus Disease 2019: a retrospective cohort study. *J. Infect.* 17, 592. doi:10.1016/j.jinf.2020.03.002
- Elfiky, A. A. (2020). Ribavirin, Remdesivir, Sofosbuvir, Galidesivir, and Tenofovir against SARS-CoV-2 RNA dependent RNA polymerase (RdRp): a molecular docking study. *Life Sci.* 253, 117592. doi:10.1016/j.lfs.2020.117592
- Fanfrlík, J., Bronowska, A. K., Rezáč, J., Přenosil, O., Konvalinka, J., and Hobza, P. (2010). A reliable docking/scoring scheme based on the semiempirical quantum mechanical PM6-DH2 method accurately covering dispersion and H-bonding: HIV-1 protease with 22 ligands. *J. Phys. Chem. B* 114, 12666–12678. doi:10.1021/jp1032965
- Gao, K., Nguyen, D. D., Chen, J., Wang, R., and Wei, G.-W. (2020). Repositioning of 8565 existing drugs for COVID-19. *J. Phys. Chem. Lett.* 11, 5373–5382. doi:10.1021/acs.jpclett.0c01579

- García-Beltrán, O., Mena, N. P., Aguirre, P., Barriga-Gonzalez, G., Galdamez, A., Nagles, E., et al. (2017). Development of an iron-selective antioxidant probe with protective effects on neuronal function. *PLoS One* 12, 1–18. doi:10.1371/journal.pone.0189043
- García-Beltrán, O., Mena, N., Yañez, O., Caballero, J., Vargas, V., Nuñez, M. T., et al. (2013). Design, synthesis and cellular dynamics studies in membranes of a new coumarin-based “turn-off” fluorescent probe selective for Fe 2+. *Eur. J. Med. Chem.* 67, 60–63. doi:10.1016/j.ejmech.2013.06.022
- Götz, A. W., Williamson, M. J., Xu, D., Poole, D., Le Grand, S., and Walker, R. C. (2012). Routine microsecond molecular dynamics simulations with AMBER on GPUs. 1. Generalized Born. *J. Chem. Theor. Comput.* 8, 1542–1555. doi:10.1021/ct200909j
- Grein, J., Ohmagari, N., Shin, D., Diaz, G., Asperges, E., Castagna, A., et al. (2020). Compassionate use of Remdesivir for patients with severe covid-19. *N. Engl. J. Med.* 382, 2327–2336. doi:10.1056/NEJMoa2007016
- Hayes, J. M., Archontis, G., and Wang, L. (2012). “MM-GB(PB)SA calculations of protein-ligand binding free energies” in *Molecular dynamics*. (Rijeka: IntechOpen). doi:10.5772/37107
- Hirota, M., Shimosegawa, T., Kitamura, K., Takeda, K., Takeyama, Y., Mayumi, T., et al. (2020). Continuous regional arterial infusion versus intravenous administration of the protease inhibitor nafamostat mesilate for predicted severe acute pancreatitis: a multicenter, randomized, open-label, phase 2 trial. *J. Gastroenterol.* 55, 342–352. doi:10.1007/s00535-019-01644-z
- Hoffmann, M., Kleine-Weber, H., Schroeder, S., Krüger, N., Herrler, T., Erichsen, S., et al. (2020). SARS-CoV-2 cell entry Depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. *Cell* 181, 271–280. doi:10.1016/j.cell.2020.02.052
- Hou, X., Du, J., Zhang, J., Du, L., Fang, H., and Li, M. (2013). How to improve docking accuracy of AutoDock4.2: a case study using different electrostatic potentials. *J. Chem. Inf. Model.* 53, 188–200. doi:10.1021/ci300417y
- Huang, C., Wei, P., Fan, K., Liu, Y., and Lai, L. (2004). 3C-like proteinase from SARS coronavirus catalyzes substrate hydrolysis by a general base mechanism. *Biochemistry* 43, 4568–4574. doi:10.1021/bi036022q
- Hughes, J. D., Blagg, J., Price, D. A., Bailey, S., DeCrescenzo, G. A., Devraj, R. V., et al. (2008). Physicochemical drug properties associated with *in vivo* toxicological outcomes. *Bioorg. Med. Chem. Lett.* 18, 4872–4875. doi:10.1016/j.bmcl.2008.07.071
- Hui, D. S., Azhar, I. E., Madani, T. A., Ntoumi, F., Kock, R., et al. (2020). The continuing 2019-nCoV epidemic threat of novel coronaviruses to global health—the latest 2019 novel coronavirus outbreak in Wuhan, China. *Int. J. Infect. Dis.* 91, 264–266. doi:10.1016/j.ijid.2020.01.009
- Humphrey, W., Dalke, A., and Schulten, K. (1996). VMD: visual molecular dynamics. *J. Mol. Graph.* 14, 33–38. doi:10.1016/0263-7855(96)00018-5
- Hwu, J. R., Huang, W.-C., Lin, S.-Y., Tan, K.-T., Hu, Y.-C., Shieh, F.-K., et al. (2019). Chikungunya virus inhibition by synthetic coumarin-guanosine conjugates. *Eur. J. Med. Chem.* 166, 136–143. doi:10.1016/j.ejmech.2019.01.037
- Hwu, J. R., Singha, R., Hong, S. C., Chang, Y. H., Das, A. R., Vlieghe, I., et al. (2008). Synthesis of new benzimidazole-coumarin conjugates as anti-hepatitis C virus agents. *Antiviral Res.* 77, 157–162. doi:10.1016/j.antiviral.2007.09.003
- Jakalian, A., Bush, B. L., Jack, D. B., and Bayly, C. I. (2000). Fast, efficient generation of high-quality atomic charges. AM1-BCC model: I. Method. *J. Comput. Chem.* 21, 132–146. doi:10.1002/(SICI)1096-987X(20000130)21:2<132::AID-JCC5>3.0.CO;2-P
- Jesumoroti, O. J., Faridoon, M. D., Isaacs, M., Hoppe, H. C., and Klein, R. (2019). Evaluation of novel N'-(3-hydroxybenzoyl)-2-oxo-2H-chromene-3-carbohydrazide derivatives as potential HIV-1 integrase inhibitors. *Med. Chem. Commun.* 10, 80–88. doi:10.1039/C8MD00328A
- Jin, Z., Du, X., Xu, Y., Deng, Y., Liu, M., Zhao, Y., et al. (2020a). Structure of Mpro from SARS-CoV-2 and discovery of its inhibitors. *Nature* 582, 289–293. doi:10.1038/s41586-020-2223-y
- Jin, Z., Zhao, Y., Sun, Y., Zhang, B., Wang, H., Wu, Y., et al. (2020b). Structural basis for the inhibition of SARS-CoV-2 main protease by antineoplastic drug carmofur. *Nat. Struct. Mol. Biol.* 27, 529–532. doi:10.1038/s41594-020-0440-6
- Johnson, E. R., Keinan, S., Mori-Sánchez, P., Contreras-García, J., Cohen, A. J., and Yang, W. (2010). Revealing noncovalent interactions. *J. Am. Chem. Soc.* 132, 6498–6506. doi:10.1021/ja100936w
- Lipinski, C. A., Lombardo, F., Dominy, B. W., and Feeney, P. J. (2001). Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Deliv. Rev.* 46, 3–26. doi:10.1016/s0169-409x(00)00129-0
- Liu, G., Wang, C., Wang, H., Zhu, L., Zhang, H., Wang, Y., et al. (2019). Antiviral efficiency of a coumarin derivative on spring viremia of carp virus *in vivo*. *Virus Res.* 268, 11–17. doi:10.1016/j.virusres.2019.05.007
- Lopez, A., Duclos, G., Pastene, B., Bezulier, K., Guilhaumou, R., Solas, C., et al. (2020). Effects of hydroxychloroquine on covid-19 in intensive care unit patients: preliminary results. *Int. J. Antimicrob. Agents* 56, 106136. doi:10.1016/j.ijantimicag.2020.106136
- Lu, R., Zhao, X., Li, J., Niu, P., Yang, B., Wu, H., et al. (2020). Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *Lancet* 395, 565–574. doi:10.1016/S0140-6736(20)30251-8
- MacLeod-Carey, D., Solis-Céspedes, E., Lamazares, E., and Mena-Ulecia, K. (2020). Evaluation of new antihypertensive drugs designed in silico using Thermolysin as a target. *Saudi Pharm. J.* 28, 582–592. doi:10.1016/j.jsps.2020.03.010
- MacNeil, A., and Rollin, P. E. (2012). Ebola and marburg hemorrhagic fevers: neglected tropical diseases? *PLoS Negl. Trop. Dis.* 6, e1546. doi:10.1371/journal.pntd.0001546
- Madhavi Sastry, G., Adzhigirey, M., Day, T., Annabhimoju, R., and Sherman, W. (2013). Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments. *J. Comput. Aided Mol. Des.* 27, 221–234. doi:10.1007/s10822-013-9644-8
- Marsh, G. A., and Wang, L.-F. (2012). Hendra and Nipah viruses: why are they so deadly? *Curr. Opin. Virol.* 2, 242–247. doi:10.1016/j.coviro.2012.03.006
- McKee, D. L., Sternberg, A., Stange, U., Laufer, S., and Naujokat, C. (2020). Candidate drugs against SARS-CoV-2 and COVID-19. *Pharmacol. Res.* 157, 104859. doi:10.1016/j.phrs.2020.104859
- Mena, N. P., García-Beltrán, O., Lourido, F., Urrutia, P. J., Mena, R., Castro-Castillo, V., et al. (2015). The novel mitochondrial iron chelator 5-((methylamino)methyl)-8-hydroxyquinoline protects against mitochondrial-induced oxidative damage and neuronal death. *Biochem. Biophys. Res. Commun.* 463, 787–792. doi:10.1016/j.bbrc.2015.06.014
- Mermelstein, D. J., Lin, C., Nelson, G., Kretsch, R., McCammon, J. A., and Walker, R. C. (2018). Fast and flexible gpu accelerated binding free energy calculations within the amber molecular dynamics package. *J. Comput. Chem.* 39, 1354–1358. doi:10.1002/jcc.25187
- Mishra, S., Pandey, A., and Manvati, S. (2020). An emerging antiviral agent. *Heliyon*. 6, e03217. doi:10.1016/j.heliyon.2020.e03217
- Mori, N., Katayama, M., and Nukaga, S. (2020). Triple therapy with hydroxychloroquine, azithromycin, and ciclesonide for COVID-19 pneumonia. *J. Microbiol. Immunol. Infect.* 93, 5187. doi:10.1016/j.jmii.2020.09.003
- Morris, G. M., Goodsell, D. S., Halliday, R. S., Huey, R., Hart, W. E., Belew, R. K., et al. (1998). Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *J. Comput. Chem.* 19, 1639–1662. doi:10.1002/(SICI)1096-987X(19981115)19:14<1639::AID-JCC10>3.0.CO;2-B
- Mulangu, S., Dodd, L. E., Davey, R. T., Tshiani Mbaya, O., Proschan, M., Mukadi, D., et al. (2019). A randomized, controlled trial of ebola virus disease therapeutics. *N. Engl. J. Med.* 381, 2293–2303. doi:10.1056/NEJMoa1910993
- Neria, E., Fisher, S., and Karplus, M. (1996). Simulation of activation free energies in molecular systems. *J. Chem. Phys.* 105, 1902–1921. doi:10.1063/1.472061
- Ngo, S. T., Quynh Anh Pham, N., Thi Le, L., Pham, D.-H., and Vu, V. (2020). VComputational Determination of potential inhibitors of SARS-CoV-2 main protease. *J. Chem. Inf. Model.* 4, 91. doi:10.1021/acs.jcim.0c00491
- Özpınar, G. A., Peukert, W., and Clark, T. (2010). An improved generalized AMBER force field (GAFF) for urea. *J. Mol. Model.* 16, 1427–1440. doi:10.1007/s00894-010-0650-7
- Pavurala, S., Vaarla, K., Kesharwani, R., Naesens, L., Liekens, S., and Vedula, R. R. (2018). Bis coumarinyl bis triazolothiadiazinyl ethane derivatives: synthesis, antiviral activity evaluation, and molecular docking studies. *Synth. Commun.* 48, 1494–1503. doi:10.1080/00397911.2018.1455871
- Peiris, J. S. M., Lai, S. T., Poon, L. L. M., Guan, Y., Yam, L. Y. C., Lim, W., et al. (2003). Coronavirus as a possible cause of severe acute respiratory syndrome. *Lancet (London, England)*. 361, 1319–1325. doi:10.1016/s0140-6736(03)13077-2

- Reynolds, C. H., Tounge, B. A., and Bembenek, S. D. (2008). Ligand binding efficiency: trends, physical basis, and implications. *J. Med. Chem.* 51, 2432–2438. doi:10.1021/jm701255b
- Řezáč, J., and Hobza, P. (2012). Advanced corrections of hydrogen bonding and dispersion for semiempirical quantum mechanical methods. *J. Chem. Theor. Comput.* 8, 141–151. doi:10.1021/ct200751e
- Rothan, H. A., and Byrareddy, S. N. (2020). The epidemiology and pathogenesis of coronavirus disease (COVID-19) outbreak. *J. Autoimmun.* 109, 102433. doi:10.1016/j.jaut.2020.102433
- Salomon-Ferrer, R., Case, D. A., and Walker, R. C. (2013). An overview of the Amber biomolecular simulation package. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* 3, 198–210. doi:10.1002/wcms.1121
- Sanner, M. F. (1999). Python: a programming language for software integration and development. *J. Mol. Graph. Model.* 17, 55–84. doi:10.1016/S1093-3263(99)99999-0
- Schrödinger, L. (2020). *Maestro*. New York: Schrödinger.
- Sissoko, D., Laouenan, C., Folkesson, E., M'Lebing, A. B., Beavogui, A. H., Baize, S., et al. (2016). Experimental treatment with favipiravir for ebola virus disease (the JIKI trial): a historically controlled, single-arm proof-of-concept trial in Guinea. *PLoS Med.* 13, 1–36. doi:10.1371/journal.pmed.1001967
- Sivamani, P., Singaravelu, G., Thiagarajan, V., Jayalakshmi, T., and Kumar, G. R. (2012). Comparative molecular docking analysis of essential oil constituents as elastase inhibitors. *Bioinformation* 8, 457–460. doi:10.6026/97320630008457
- Song, L., Lee, T.-S., Zhu, C., York, D. M., and Kajid, M. M. (2019). Validation of AMBER/GAFF for relative free energy calculations. *ChemRxiv*. doi:10.26434/chemrxiv.7653434.v1
- Stewart, J. J. P. (2016). *MOPAC*. Colorado: MOPAC.
- Stewart, J. J. P. (2007). Optimization of parameters for semiempirical methods V: modification of NDDO approximations and application to 70 elements. *J. Mol. Model.* 13, 1173–1213. doi:10.1007/s00894-007-0233-4
- To, K. K., Ng, K. H., Que, T.-L., Chan, J. M., Tsang, K.-Y., Tsang, A. K., et al. (2012). Avian influenza A H5N1 virus: a continuous threat to humans. *Emerg. Microb. Infect.* 1, 1–12. doi:10.1038/emi.2012.24
- To, K. K. W., Chan, J. F. W., Chen, H., Li, L., and Yuen, K.-Y. (2013). The emergence of influenza A H7N9 in human beings 16 years after influenza A H5N1: a tale of two cities. *Lancet Infect. Dis.* 13, 809–821. doi:10.1016/S1473-3099(13)70167-1
- Trott, O., and Olson, A. (2010). AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization and multithreading. *J. Comput. Chem.* 31, 455–461. doi:10.1002/jcc.21334
- ul Qamar, M. T., Alqahtani, S. M., Alamri, M. A., and Chen, L.-L. (2020). Structural basis of SARS-CoV-2 3CLpro and anti-COVID-19 drug discovery from medicinal plants. *J. Pharm. Anal.* 10, 313–319. doi:10.1016/j.jpha.2020.03.009
- Veber, D. F., Johnson, S. R., Cheng, H.-Y., Smith, B. R., Ward, K. W., and Kopple, K. D. (2002). Molecular properties that influence the oral bioavailability of drug candidates. *J. Med. Chem.* 45, 2615–2623. doi:10.1021/jm020017n
- Walls, A. C., Park, Y.-J., Tortorici, M. A., Wall, A., McGuire, A. T., and Veesler, D. (2020). Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein. *Cell* 181, 281–292. doi:10.1016/j.cell.2020.02.058
- Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A., and Case, D. A. (2004). Development and testing of a general amber force field. *J. Comput. Chem.* 25, 1157–1174. doi:10.1002/jcc.20035
- WHO Director-General's opening remarks at the media briefing on COVID-19–20 March (2020). <https://www.who.int/dg/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19--20-march-2020>. [Accessed]
- World Health Organization (2020). *Surveillance case definitions for human infection with novel coronavirus (nCoV)*. Geneva: World Health Organization <https://apps.who.int/iris/handle/10665/330376>
- Wrapp, D., Wang, N., Corbett, K. S., Goldsmith, J. A., Hsieh, C. L., Abiona, O., et al. (2020). Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science (80-)* 367, 1260–1263. doi:10.1126/science.aax0902
- Wu, F., Zhao, S., Yu, B., Chen, Y.-M., Wang, W., Song, Z.-G., et al. (2020). A new coronavirus associated with human respiratory disease in China. *Nature* 579, 265–269. doi:10.1038/s41586-020-2008-3
- Xu, Z., Peng, C., Shi, Y., Zhu, Z., Mu, K., Wang, X., et al. (2020). Nelfinavir was predicted to be a potential inhibitor of 2019-nCoV main protease by an integrative approach combining homology modelling, molecular docking and binding free energy calculation. *bioRxiv* 13, 59. doi:10.1101/2020.01.27.921627.2020.01.27.921627
- Yamamoto, M., Matsuyama, S., Li, X., Takeda, M., Kawaguchi, Y., Inoue, J., et al. (2016). Identification of nafamostat as a potent inhibitor of Middle East respiratory syndrome coronavirus S protein-mediated membrane fusion using the split-protein-based cell-cell fusion assay. *Antimicrob. Agents Chemother.* 60, 6532–6539. doi:10.1128/AAC.01043-16
- Yu, H., and Adedoyin, A. (2003). ADME-Tox in drug discovery: integration of experimental and computational technologies. *Drug Discov. Today* 8, 852–861. doi:10.1016/S1359-6446(03)02828-9
- Yuen, K. Y., Chan, P. K., Peiris, M., Tsang, D. N., Que, T. L., Shortridge, K. F., et al. (1998). Clinical features and rapid viral diagnosis of human disease associated with avian influenza A H5N1 virus. *Lancet (London, England)* 351, 467–471. doi:10.1016/S0140-6736(98)01182-9
- Zaki, A. M., van Boheemen, S., Bestebroer, T. M., Osterhaus, A. D. M. E., and Fouchier, R. A. M. (2012). Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia. *N. Engl. J. Med.* 367, 1814–1820. doi:10.1056/NEJMoa1211721
- Zhang, L., Lin, D., Kusov, Y., Nian, Y., Ma, Q., Wang, J., et al. (2020a).  $\alpha$ -Ketoamides as broad-spectrum inhibitors of coronavirus and enterovirus replication: structure-based design, synthesis, and activity assessment. *J. Med. Chem.* 18, 28. doi:10.1021/acs.jmedchem.9b01828
- Zhang, L., Lin, D., Sun, X., Curth, U., Drosten, C., Sauerherring, L., et al. (2020b). Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved  $\alpha$ -ketoamide inhibitors. *Science (80-)* 3405, eabb3405. doi:10.1126/science.abb3405
- Zhou, P., Yang, X.-L., Wang, X.-G., Hu, B., Zhang, L., Zhang, W., et al. (2020). A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* 579, 270–273. doi:10.1038/s41586-020-2012-7
- Zhu, N., Zhang, D., Wang, W., Li, X., Yang, B., Song, J., et al. (2020). A novel coronavirus from patients with pneumonia in China, 2019. *N. Engl. J. Med.* 382, 727–733. doi:10.1056/NEJMoa2001017

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Yañez, Osorio, Uriarte, Areche, Tiznado, Perez-Donoso, García-Blatrán and González-Nilo. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Viroporins vs. Other Pore-Forming Proteins: What Lessons Can We Take?

Eva Žerovnik \*

Department of Biochemistry and Molecular and Structural Biology, J. Stefan Institute, Ljubljana, Slovenia

Pore-forming proteins (PFPs) exist in virtually all domains of life, and by disrupting cellular membranes, depending on the pore size, they cause ion dis-balance, small substances, or even protein efflux/influx, influencing cell's signaling routes and fate. Such pore-forming proteins exist from bacteria to viruses and also shape host defense systems, including innate immunity. There is strong evidence that amyloid toxicity is also caused by prefibrillar oligomers making "amyloid pores" into cellular membranes. For most of the PFPs, a 2-step mechanism of protein-membrane interaction takes place on the "lipid rafts," membrane microdomains rich in gangliosides and cholesterol. In this mini-review paper, common traits of different PFPs are looked at. Possible ways for therapy of channelopathies and/or modulating immunity relevant to the new threat of SARS-CoV-2 infections could be learnt from such comparisons.

**Keywords:** SARS-CoV-2 E protein, amyloid pore, channel formation, ion conductance activity, drug target

## OPEN ACCESS

### Edited by:

Emilia Pedone,  
Italian National Research Council, Italy

### Reviewed by:

Silvia Dante,  
Italian Institute of Technology (IIT), Italy  
Fernando Teran Arce,  
University of Arizona, United States

### \*Correspondence:

Eva Žerovnik  
eva.zerovnik@ijs.si

### Specialty section:

This article was submitted to  
Chemical Biology,  
a section of the journal  
Frontiers in Chemistry

**Received:** 04 November 2020

**Accepted:** 07 January 2021

**Published:** 18 February 2021

### Citation:

Žerovnik E (2021) Viroporins vs. Other  
Pore-Forming Proteins: What Lessons  
Can We Take?.  
Front. Chem. 9:626059.  
doi: 10.3389/fchem.2021.626059

## WIDELY SPREAD PHENOMENON OF PORE FORMATION

Pore-forming proteins (PFPs) appear in virtually all organisms starting from viruses and bacteria. Bacteria use pore-forming toxins (PFTs) to disrupt plasma membrane of host cells. Even though there are several structural classes of PFTs, they all make pores after oligomerization. They can form  $\alpha$ -helical or  $\beta$ -barrel transmembrane channels. Many reviews of bacterial PFTs have been written; among them is a comprehensive review by Dal Peraro and van der Goot (2016). The mechanism of pore formation by bacterial PFTs and structure of the transmembrane pores have been studied extensively. It was observed that the majority of pore-forming proteins make pores composed from transmembrane  $\beta$ -barrels (Heuck et al., 2001) or from clusters of  $\alpha$ -helices (Kristan et al., 2009; Kagan and Thundimadathil, 2010). The sizes of pores are in range of a few nm up to 40 nm in diameter (Bischofberger et al., 2009). These pores allow uncontrolled permeabilization of ions and small molecules and the larger pores even of proteins. Consequences of pore formation vary and depend on the number of pores present in the plasma membrane, the mechanism of membrane binding, cell type, and so forth.

Viruses, predominantly RNA viruses, also use the so-called "viroporins" to enhance pathogenic response of the host immune system and cause extensive inflammatory response (Nieto-Torres et al., 2015). Already in 2013, the channel activity of the viroporin-E protein (E for envelope) from SARS Coronavirus was studied (Aguilella et al., 2013). The E protein of the actual SARS-CoV-2 virus, similar to the one from SARS-CoV, oligomerizes into homopentamers as obtained by Sankar et al. (2020) by molecular modeling as confirmed by NOE's contacts from heteronuclear NMR (Sarkar and Saha, 2020). Studies of structures, dynamics, and interactions with host cells of viroporins and bacterial PFTs are important as they may help in the search for novel antibacterial and antiviral therapies.

Pore formation is not only an ancient mechanism of attack, such as used by viruses, bacteria, and lower invertebrates but is also used for signaling and defense in higher organisms (Iacovache et al., 2008; Feil et al., 2010). Innate immunity has evolved from invertebrates via fish to mammals (Buchmann, 2014). Innate effector molecules are oxygen and nitrogen species, anti-microbial peptides, lectins, fibrinogen-related peptides, leucine rich repeats, pentraxins, and complement-related proteins (Buchmann, 2014). In this context, anti-microbial peptides (AMPs), also termed host defense peptides, are used by invertebrates and vertebrates, including mammals, in order to kill microbes via membrane perforation.

Amyloid-forming proteins (AFPs), involved in neurodegenerative diseases, with highest prevalence in the aging population of Alzheimer's and Parkinson's disease, also form transmembrane pores/channels when in oligomeric form. *In vitro* several AFPs were shown to interact with membranes and form the so-called amyloid pores (Kagan and Thundimadathil, 2010). *In vivo* situation is a bit less clear as no-one has observed any amyloid pore directly; however, they are implied from a *C. elegans* study where the membrane repair response was observed when animals were fed by human A $\beta$  (Julien et al., 2018). Some functional proteins also can make amyloid fibrils and pores, at least *in vitro*. Such is the case with stefin B (cystatin B) (Ceru et al., 2008; Rabzelj et al., 2008) and might underlie epileptogenesis, as suggested by Surguchov et al. (2017).

There is not much difference between AMPs and amyloid toxicity as pointed out by Jang et al. (2011), who showed that fragment of protegrin forms amyloid fibrils. That a common mechanism may apply was proposed by Last and Miranker (2013). In addition, amyloid-beta (A $\beta$ ) likely possesses anti-microbial activity, which relates AD risk to microbial infection (Moir et al., 2018). Kumar et al. (2016) show a model in which soluble A $\beta$  oligomers first bind to microbial cell wall carbohydrates via heparin-binding domain, after which growing protofibrils inhibit pathogen adhesion to host cells. Similarly, Walsh et al. (2014) report that PrP(106–126) composition is reminiscent of cationic anti-microbial peptide dermaseptin. In agreement with expectation, oligomeric PrP(106–126) inhibited the growth of BL21 *E. coli* cultures (Walsh et al., 2014).

Additional similarity between amyloid oligomers and other PFPs is a multistep mechanism of channel formation, which includes oligomerization at the plane of the membrane (Bischofberger et al., 2009).

As there seem to be some common mechanisms on the side of lipid composition and protein oligomeric structures, I hereby suggest that one needs to study and compare what is known about the pore-forming peptides from amyloid proteins, anti-microbial peptides, bacterial PFTs, and viroporins. I drop out from consideration in the mini-review bacterial PFTs, due to their complexity and different structural classes.

## VIRAL PORE FORMATION—VIROPORINS

RNA of several pathogenic human viruses encodes at least one viroporin. This is the case with pathogenic human influenza A

virus (IAV), human immunodeficiency virus 1 (HIV-1), hepatitis C virus (HCV), and coronaviruses (CoVs), including the one responsible for the severe acute respiratory syndrome (SARS-CoV) and the other causing Middle East respiratory syndrome (MERS-CoV) (Dal Peraro and van der Goot, 2016).

As demonstrated for SARS-CoV E protein, ion conductivity (IC) activity can overstimulate host immune response, leading to cytokine storm, also reported for the SARS-CoV-2. Of importance, when virus was devoid of E protein IC activity, it proved less lethal (Nieto-Torres et al., 2014). Viral IC activity overstimulates inflammatory response by the activation of NLRP3 inflammasome. There are some promising results in search of specific inhibitors of NLRP3 inflammasome in order to reduce inflammatory responses (Coll et al., 2015).

Thus, the envelope E protein can make homo-oligomers and generate an ion channel termed viroporin. Peptides making the transmembrane domain of E protein were synthesized, and their oligomerization was studied. It was shown that E protein can form dimers, trimers, and pentamers. When SARS-CoV E protein was expressed in Sf9 insect cells, it formed multimeric homo-oligomers. By mutations of hydrophobic residues in the TMD with charged residues, monomers were obtained. In more detail, mutations of the TMD residues asparagine 15 (N15) to alanine (N15A) and valine 25 (V25) to phenylalanine (V25F) were found to abolish the IC activity of CoV E viroporin, confirming that this activity depends on its homopentameric conformation. The ability of CoV E protein to assemble into homopentamers is clearly important for the functional CoV E viroporin (Schoeman and Fielding, 2019).

More studies have been performed recently on the structure and potential drug binding sites of the E protein from CoV-2 (Mandala et al., 2020). The orientation of five-helix bundle of the transmembrane region of the E protein in lipid bilayers was determined by solid-state NMR. It provides explanation for how Ca<sup>2+</sup> ions could enter and how to block this activity, which leads to inflammasome activation (Coll et al., 2015).

For big DNA viruses and smaller RNA viruses, host cellular double membrane invaginations from ER, Golgi, and autophagosomes are used for viral transport and replication (the so-called virus factories), in a similar way as for protein aggregates removal by autophagy. There are some parallels here again with amyloid-forming proteins (see the following section). The role of cholesterol and gangliosides rich lipid rafts has been reported for both cases; it is known that disruption of the lipid rafts causes a significant reduction of viral RNA production.

*In silico* approaches to detect inhibitors of the human SARS-CoV-2 E protein ion channel activity have already led to some possible drugs (Gupta et al., 2020).

## ANTI-MICROBIAL PEPTIDES AND AMYLOID TOXINS: TWO SIDES OF THE SAME COIN

On one hand are anti-microbial peptides (AMPs), also termed host defense peptides, used by invertebrates and vertebrates, including mammals. By perforating microbial membranes,

AMPs act as potent, broad spectrum antibiotics against bacteria, fungi, and some (enveloped) viruses. Structurally they can be classified into three major groups: peptides with an  $\alpha$ -helical conformation (e.g., insect cecropins, magainins), cyclic peptides with pairs of cysteine residues (e.g., defensins, protegrins), and peptides rich in some amino acid residues (e.g., proline rich, histidine rich). Most AMPs are proteins of <25 kDa and adopt amphipathic structures, which contribute to their interaction with anionic membranes (Bulet et al., 2004). It was shown that protegrins are able to make channels (Sokolov et al., 1999; Capone et al., 2010).

Pore-forming proteins also play an important role in innate immunity, such as the case with perforin 1 (perforating extracellular bacteria), perforin 2 (perforating bacteria which entered cells by endosomes), and membrane attack complex of the complement, with perforin-like D9 component (Voskoboinik et al., 2006; Rosado et al., 2008).

On the other hand, prefibrillar oligomers of many amyloid-forming proteins can make the so-called amyloid pores into membranes and exert cyto-toxicity. The oligomeric prefibrillar state, either, on the way to amyloid fibrils or sometimes off-pathway, after a temporary  $\alpha$  to  $\beta$  secondary transition, usually adopts  $\beta$ -barrel transmembrane pore conformation. Amyloid pores can disrupt plasma membrane and intracellular membranes among them mitochondrial (Squier, 2001; Pagani and Eckert, 2011). For example, mitochondrial dysfunction in PD may be due to cardiolipin-promoted perforation of mitochondrial membranes by  $\alpha$ -synuclein oligomers (Ghio et al., 2019).

A recent review by Lee et al. (2020) tries to connect properties of AMPs and AFPs, especially shorter fragments or peptides of AFPs, like A $\beta$  and amylin. They conclude: "In fact, a large number of naturally occurring AMPs including LL37, lysozyme, protegrin-1, plant defensins, temporins, etc., form amyloid fibrils, oligomerise, and interact with membranes, causing membrane permeation by similar mechanisms to amyloid pores."

## MORE ON AMYLOID PORES: IS THERE A COMMON MECHANISM FOR AMPS PORES?

Morphologically and structurally amyloid pores are similar to pores formed by other pore-forming proteins (Parker and Feil, 2005; Anderluh and Lakey, 2008). They have been detected in the case of at least 12 amyloid-forming proteins, ranging from typical globular to intrinsically disordered proteins or proteolytic fragments of the amyloidogenic proteins. They are in general quite large (diameter of 3–10 nm) and rather non-selective (Butterfield and Lashuel, 2010; Kagan and Thundimadathil, 2010). Lipid components, such as sphingolipin and cholesterol, part of the lipid rafts, facilitate the conformational change of the amyloid pores from natively unfolded into  $\alpha$ -helix and/or  $\beta$ -sheet-rich structures (Butterfield and Lashuel, 2010).

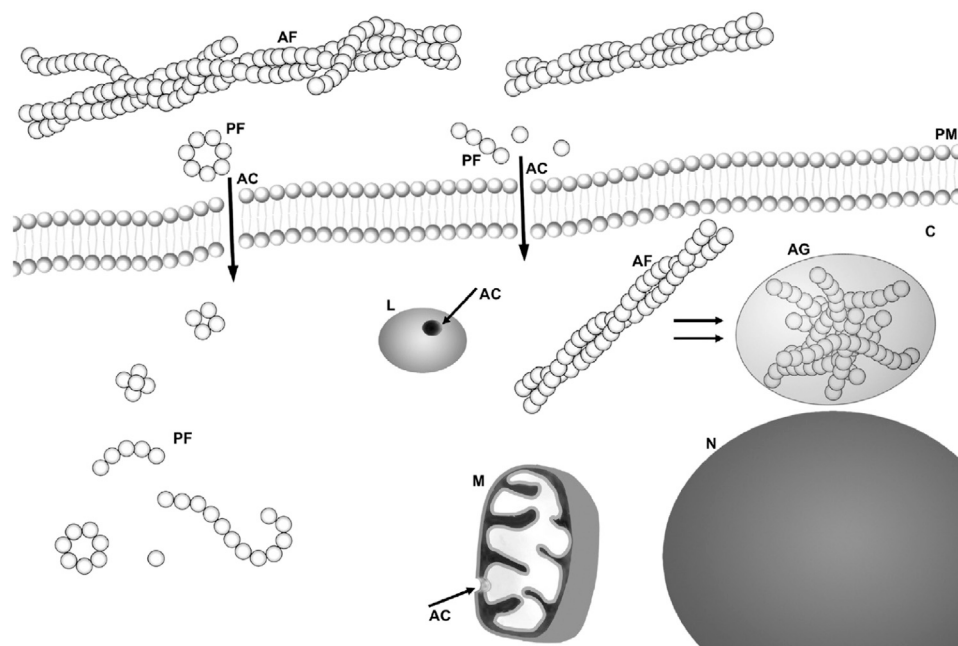
Amyloid pores have been observed by oligomers of  $\alpha$ -synuclein, A $\beta$ , and prion, among others. Even though the

oligomers have not been visualized in interaction with cellular membranes *in vivo*, they are indirectly indicated by pore-like activities such as Ca<sup>2+</sup> entry, mitochondrial ROS increase, and nuclear pore damage. However, recently the channel activity of A $\beta$  was observed in extracted cells membranes (Bode et al., 2017). Moreover, what is even more convincing, an animal model of *C. elegans* showed membrane defense response in this worm when challenged by human A $\beta$  (fed by *E. coli* bacteria expressing the A $\beta$  peptide) (Julien et al., 2018). Current understanding about the relative toxicity of endogenous soluble  $\alpha$ -synuclein oligomers and multimers and their cross-reactivity with Tau and A $\beta$  in different neurodegenerative diseases is reviewed by Kayed et al. (2020).

To determine the structure of oligomers making amyloid pores has also been challenging; however, atomic force microscopy (AFM) has provided some insight on the pore structure (Lin et al., 2001). Amyloid-beta (A $\beta$ ), 40 or 42 long peptide-forming plaques in Alzheimer's disease, has been extensively studied. A $\beta$  (1-42) in a planar lipid bilayer revealed multimeric (tetrameric, pentameric, and hexameric) channel-like structures. In accordance, electrophysiological recordings demonstrated the presence of multiple single channel currents. At the cellular level, A $\beta$  (1-42) incorporation increased calcium influx and induced aberrant neuritic growth (Lin et al., 2001). A very recent paper by Ciudad et al. (2020) described a molecular dynamics study of insertion of A $\beta$  (1-42) tetramers and octamers in lipid bilayers. A mechanism of membrane disruption in which water permeation occurred through lipid-stabilized pores has been revealed.

Di Scala et al. (2016) proposed a common molecular mechanism of amyloid pore formation by A $\beta$  and  $\alpha$ -synuclein ( $\alpha$ S). They have compared a panel of amyloid-forming fragments of the above-mentioned proteins and arrived at conclusion that 2-step mechanism applies, whereas each of the gangliosides and cholesterol components of lipid membranes interacts with specific structural motifs of A $\beta$  and  $\alpha$ S. Whether this is a universal mechanism applying to other amyloid toxins remains to be seen.

Fusco et al. (2016) characterized membrane bound  $\alpha$ S. Despite the biological relevance, the structural details of the membrane-bound oligomer of  $\alpha$ S remain elusive. It is difficult to isolate a well-defined and stable oligomer and also difficult to study it in cells. The authors used solid state NMR and restrained MD simulations to refine the structure of the N-terminal (1-30 a. acids) of  $\alpha$ S bound to synaptic-like membranes. The results indicate that the first 12 residues of  $\alpha$ S are key to anchoring the protein to lipid surface. In order to improve the study bearing in mind that  $\alpha$ S pore could be in the soluble fraction, Fusco et al. (2017) used solid state and solution NMR to determine structural constraints of  $\alpha$ S membrane interaction. The structured region strongly inserted into lipid bilayers and disrupted their integrity, leading finally to cell death. Mutations which prevented membrane interaction also prevented toxicity. The authors reported two types of oligomers; the ones with more  $\beta$ -structure and deeper membrane insertion/disruption proved toxic in distinction to surface bound oligomers.



**FIGURE 1** | Scheme of possible sites for amyloid “toxins”: membrane interaction/perforation where AF stands for amyloid fibrils, PF for annular and other kind of protofibrils, AC for amyloid channel, M for mitochondria, N for nucleus, and L for lysosome. Taken from Žganec and Žerovnik (2014), copyright to Elsevier.

Canale with coworkers studied non-pathological bacterial protein HypF-N as a model for amyloid induced toxicity (Oropesa-Nuñez et al., 2016; Oropesa-Nuñez et al., 2018). They differentiated between toxic and non-toxic HypF-N oligomers and used AFM to observe their interaction with lipid bilayers. Their findings support the notion that GM1 ganglioside mediates the oligomer-membrane interaction.

Scheme was taken from Žganec and Žerovnik (2014).

## DISCUSSION

This mini-review aims to compare features and mechanisms of pore formation by amyloid-forming proteins (AFPs), that is, their membrane perforating oligomers, anti-microbial peptides (AMPs), also called defense peptides of the innate immunity system and viroporins, and transmembrane short viral envelope proteins (E protein), helping spread certain viruses, among them the coronavirus SARS-CoV-2.

One should be able to derive common structural traits and interaction mechanism of some AFPs, AMPs, and viroporins, which would include oligomerization, alignment of  $\alpha$ -helices against lipid surface (on acidic phospholipids, initially driven by electrostatics), and transition of the “pre-pore” into  $\beta$  structure and making a pore (Omersa et al., 2019). By finding common mechanisms, perhaps one could design common means of defense and augment anti-viral and anti-amyloid therapies. By stabilizing membranes, inhibiting the process of pore formation by small drugs/peptides competing with ganglioside and cholesterol binding sites or inhibiting

channel conductance might be a possible therapeutic way to attack such broad spectrum of disease (Scott and Griffin, 2015). Compounds blocking channel activity by A $\beta$  oligomers have been reported for a mouse model of AD (Martinez Hernandez et al., 2018).

Out of curiosity, perhaps, we have previously compared A $\beta$ , part of prion protein and part of our model amyloid-forming protein, stefin B (cystatin B) (Yoichi et al., 2005), which also (when in prefibrillar oligomeric form) makes pores into acidic phospholipid membranes as our *in vitro* studies show (Ceru and Žerovnik, 2008; Rabzelj et al., 2008). In the Clustal alignment, the  $\alpha$ -helix and first  $\beta$ -strand of stefin B showed low similarity with parts of prion and amyloid-beta, including the protease binding site QVVAG. This comparison might seem superficial. However, I suggest to compare the sequences of more pore-forming peptides and to use more sophisticated methods of prediction and sequence comparative analysis, for example, those used in Venko et al., submitted to Frontiers in Mol. Neuroscience.

## AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and has approved it for publication.

## FUNDING

I am part of program 0140 by ARRS. I was informed that contributions will be free of charge if submitted early.

## REFERENCES

- Aguilella, V. M., Verdia-Baguena, C., Nieto-Torres, J. L., Alcaraz, A., Dediego, M. L., and Enjuanes, L. (2013). SARS-CoV E protein ion channel characterization by tuning the protein and lipid charge. *Eur. Biophys. J. Biophys.* 42, S76.
- Anderluh, G., and Lakey, J. H. (2008). Disparate proteins use similar architectures to damage membranes. *Trends Biochem. Sci.* 33 (10), 482–490. doi:10.1016/j.tibs.2008.07.004
- Bischofberger, M., Gonzalez, M. R., and van der Goot, F. G. (2009). Membrane injury by pore-forming proteins. *Curr. Opin. Cell Biol.* 21 (4), 589–595. doi:10.1016/j.ceb.2009.04.003
- Bode, D. C., Baker, M. D., and Viles, J. H. (2017). Ion Channel formation by amyloid- $\beta$ 2 oligomers but not amyloid- $\beta$ 40 in cellular membranes. *J. Biol. Chem.* 292 (4), 1404–1413. doi:10.1074/jbc.M116.762526
- Buchmann, K. (2014). Evolution of innate immunity: clues from invertebrates via fish to mammals. *Front. Immunol.* 5, 459. doi:10.3389/fimmu.2014.00459
- Bulet, P., Stöcklin, R., and Menin, L. (2004). Anti-microbial peptides: from invertebrates to vertebrates. *Immunol. Rev.* 198, 169–184. doi:10.1111/j.0105-2896.2004.0124.x
- Butterfield, S. M., and Lashuel, H. A. (2010). Amyloidogenic protein-membrane interactions: mechanistic insight from model systems. *Angew Chem. Int. Ed. Engl.* 49 (33), 5628–5654. doi:10.1002/anie.200906670
- Capone, R., Mustata, M., Jang, H., Arce, F. T., Nussinov, R., and Lal, R. (2010). Antimicrobial protegrin-1 forms ion channels: molecular dynamic simulation, atomic force microscopy, and electrical conductance studies. *Biophys. J.* 98 (11), 2644–2652. doi:10.1016/j.bpj.2010.02.024
- Ceru, S., Kokalj, S., Rabzelj, S., Skarabot, M., Gutierrez-Aguirre, I., Kopitar-Jerala, N., et al. (2008). Size and morphology of toxic oligomers of amyloidogenic proteins: a case study of human stefin B. *Amyloid.* 15 (3), 147–159. doi:10.1080/13506120802193555
- Ceru, S., and Žerovnik, E. (2008). Similar toxicity of the oligomeric molten globule state and the prefibrillar oligomers. *FEBS Lett.* 582 (2), 203–209. doi:10.1016/j.febslet.2007.12.002
- Ciudad, S., Puig, E., Botzanowski, T., Meigooni, M., Arango, A. S., Do, J., et al. (2020). A $\beta$ (1–42) tetramer and octamer structures reveal edge conductivity pores as a mechanism for membrane damage. *Nat. Commun.* 11 (1), 3014. doi:10.1038/s41467-020-16566-1
- Coll, R. C., Robertson, A. A., Chae, J. J., Higgins, S. C., Muñoz-Planillo, R., Inserra, M. C., et al. (2015). A small-molecule inhibitor of the NLRP3 inflammasome for the treatment of inflammatory diseases. *Nat. Med.* 21 (3), 248–255. doi:10.1038/nm.3806
- Dal Peraro, M., and van der Goot, F. G. (2016). Pore-forming toxins: ancient, but never really out of fashion. *Nat. Rev. Microbiol.* 14 (2), 77–92. doi:10.1038/nrmicro.2015.3
- Di Scala, C., Yahi, N., Boutemur, S., Flores, A., Rodriguez, L., Chahinian, H., et al. (2016). Common molecular mechanism of amyloid pore formation by Alzheimer's beta-amyloid peptide and alpha-synuclein. *Sci. Rep.* 6, 28781. doi:10.1038/srep28781
- Feil, S. C., Polekhina, G., Gorman, M. A., and Parker, M. W. (2010). Proteins: membrane binding and pore formation. introduction. *Adv. Exp. Med. Biol.* 677, 1–13. doi:10.1007/978-1-4419-6327-7\_1
- Fusco, G., Chen, S. W., Williamson, P. T. F., Cascella, R., Perni, M., Jarvis, J. A., et al. (2017). Structural basis of membrane disruption and cellular toxicity by  $\alpha$ -synuclein oligomers. *Science*. 358 (6369), 1440–1443. doi:10.1126/science.aan6160
- Fusco, G., De Simone, A., Arosio, P., Vendruscolo, M., Veglia, G., and Dobson, C. M. (2016). Structural ensembles of membrane-bound  $\alpha$ -synuclein reveal the molecular determinants of synaptic vesicle affinity. *Sci. Rep.* 6, 27125. doi:10.1038/srep27125
- Ghio, S., Camilleri, A., Caruana, M., Ruf, V. C., Schmidt, F., Leonov, A., et al. (2019). Cardiolipin promotes pore-forming activity of alpha-synuclein oligomers in mitochondrial membranes. *ACS Chem. Neurosci.* 10 (8), 3815–3829. doi:10.1021/acchemneuro.9b00320
- Gupta, M. K., Vemula, S., Donde, R., Gouda, G., Behera, L., and Vadde, R. (2020). In-silico approaches to detect inhibitors of the human severe acute respiratory syndrome coronavirus envelope protein ion channel. *J. Biomol. Struct. Dyn.*, 1–11. doi:10.1080/07391102.2020.1751300
- Heuck, A. P., Tweten, R. K., and Johnson, A. E. (2001). Beta-barrel pore-forming toxins: intriguing dimorphic proteins. *Biochemistry*. 40 (31), 9065–9073. doi:10.1021/bi0155394
- Iacovache, I., van der Goot, F. G., and Pernot, L. (2008). Pore formation: an ancient yet complex form of attack. *Biochim. Biophys. Acta*. 1778 (7–8), 1611–1623. doi:10.1016/j.bbamem.2008.01.026
- Jang, H., Arce, F. T., Mustata, M., Ramachandran, S., Capone, R., Nussinov, R., et al. (2011). Antimicrobial protegrin-1 forms amyloid-like fibrils with rapid kinetics suggesting a functional link. *Biophys. J.* 100 (7), 1775–1783. doi:10.1016/j.bpj.2011.01.072
- Julien, C., Tomberlin, C., Roberts, C. M., Akram, A., Stein, G. H., Silverman, M. A., et al. (2018). In vivo induction of membrane damage by  $\beta$ -amyloid peptide oligomers. *Acta Neuropathol. Commun.* 6 (1), 131. doi:10.1186/s40478-018-0634-x
- Kagan, B. L., and Thundimadathil, J. (2010). Amyloid peptide pores and the beta sheet conformation. *Adv. Exp. Med. Biol.* 677, 150–167. doi:10.1007/978-1-4419-6327-7\_13
- Kayed, R., Dettmer, U., and Lesné, S. E. (2020). Soluble endogenous oligomeric  $\alpha$ -synuclein species in neurodegenerative diseases: expression, spreading, and cross-talk. *J. Parkinsons Dis.* 10 (3), 791–818. doi:10.3233/JPD-201965
- Kristan, K. C., Viero, G., Dalla Serra, M., Macek, P., and Anderluh, G. (2009). Molecular mechanism of pore formation by actinoporins. *Toxicon*. 54 (8), 1125–1134. doi:10.1016/j.toxicon.2009.02.026
- Kumar, D. K., Choi, S. H., Washicosky, K. J., Eimer, W. A., Tucker, S., Ghofrani, J., et al. (2016). Amyloid- $\beta$  peptide protects against microbial infection in mouse and worm models of Alzheimer's disease. *Sci. Transl. Med.* 8 (340), 340ra72. doi:10.1126/scitranslmed.aaf1059
- Last, N. B., and Miranker, A. D. (2013). Common mechanism unites membrane poration by amyloid and antimicrobial peptides. *Proc. Natl. Acad. Sci. U.S.A.* 110 (16), 6382–6387. doi:10.1073/pnas.1219059110
- Lee, E. Y., Srinivasan, Y., de Anda, J., Nicastro, L. K., Tükel, Ç., and Wong, G. C. L. (2020). Functional reciprocity of amyloids and antimicrobial peptides: rethinking the role of supramolecular assembly in host defense, immune activation, and inflammation. *Front. Immunol.* 11, 1629. doi:10.3389/fimmu.2020.01629
- Lin, H., Bhatia, R., and Lal, R. (2001). Amyloid beta protein forms ion channels: implications for Alzheimer's disease pathophysiology. *FASEB J.* 15 (13), 2433–2444. doi:10.1096/fj.01-0377com
- Mandala, V. S., McKay, M. J., Shcherbakov, A. A., Dregni, A. J., Kolocouris, A., and Hong, M. (2020). Structure and drug binding of the SARS-CoV-2 envelope protein transmembrane domain in lipid bilayers. *Nat. Struct. Mol. Biol.* 27 (12), 1202–1208. doi:10.1038/s41594-020-00536-8
- Martinez Hernandez, A., Urbanke, H., Gillman, A. L., Lee, J., Ryazanov, S., Agbemenyah, H. Y., et al. (2018). The diphenylpyrazole compound anle138b blocks A $\beta$  channels and rescues disease phenotypes in a mouse model for amyloid pathology. *EMBO Mol. Med.* 10 (1), 32–47. doi:10.15252/emmm.201707825
- Moir, R. D., Lathe, R., and Tanzi, R. E. (2018). The antimicrobial protection hypothesis of Alzheimer's disease. *Alzheimers Dement.* 14 (12), 1602–1614. doi:10.1016/j.jalz.2018.06.3040
- Nieto-Torres, J. L., DeDiego, M. L., Verdiá-Báguena, C., Jimenez-Guardeño, J. M., Regla-Nava, J. A., Fernandez-Delgado, R., et al. (2014). Severe acute respiratory syndrome coronavirus envelope protein ion channel activity promotes virus fitness and pathogenesis. *PLoS Pathog.* 10 (5), e1004077. doi:10.1371/journal.ppat.1004077
- Nieto-Torres, J. L., Verdiá-Báguena, C., Castaño-Rodríguez, C., Aguilella, V. M., and Enjuanes, L. (2015). Relevance of viroporin ion channel activity on viral replication and pathogenesis. *Viruses*. 7 (7), 3552–3573. doi:10.3390/v7072786
- Omersa, N., Podobnik, M., and Anderluh, G. (2019). Inhibition of pore-forming proteins. *Toxins*. 11 (9), 545. doi:10.3390/toxins11090545
- Oropesa-Núñez, R., Keshavan, S., Dante, S., Diaspro, A., Mannini, B., Capitini, C., et al. (2018). Toxic HypF-N oligomers selectively bind the plasma membrane to impair cell adhesion capability. *Biophys. J.* 114 (6), 1357–1367. doi:10.1016/j.bpj.2018.02.003
- Oropesa-Núñez, R., Seghezza, S., Dante, S., Diaspro, A., Cascella, R., Cecchi, C., et al. (2016). Interaction of toxic and non-toxic HypF-N oligomers with lipid bilayers investigated at high resolution with atomic force microscopy. *Oncotarget*. 7 (29), 44991–45004. doi:10.18632/oncotarget.10449

- Pagani, L., and Eckert, A. (2011). Amyloid-Beta interaction with mitochondria. *Int. J. Alzheimer's Dis.* 2011, 925050. doi:10.4061/2011/925050
- Parker, M. W., and Feil, S. C. (2005). Pore-forming protein toxins: from structure to function. *Prog. Biophys. Mol. Biol.* 88 (1), 91–142. doi:10.1016/j.pbiomolbio.2004.01.009
- Rabzelj, S., Viero, G., Gutiérrez-Aguirre, I., Turk, V., Dalla Serra, M., Anderluh, G., et al. (2008). Interaction with model membranes and pore formation by human stefin B: studying the native and prefibrillar states. *FEBS J.* 275 (10), 2455–2466. doi:10.1111/j.1742-4658.2008.06390.x
- Rosado, C. J., Kondos, S., Bull, T. E., Kuiper, M. J., Law, R. H., Buckle, A. M., et al. (2008). The MACPF/CDC family of pore-forming toxins. *Cell Microbiol.* 10 (9), 1765–1774. doi:10.1111/j.1462-5822.2008.01191.x
- Sarkar, M., and Saha, S. (2020). Structural insight into the role of novel SARS-CoV-2 E protein: a potential target for vaccine development and other therapeutic strategies. *PLoS One.* 15 (8), e0237300. doi:10.1371/journal.pone.0237300
- Schoeman, D., and Fielding, B. C. (2019). Coronavirus envelope protein: current knowledge. *Virol. J.* 16 (1), 69. doi:10.1186/s12985-019-1182-0
- Scott, C., and Griffin, S. (2015). Viroporins: structure, function and potential as antiviral targets. *J. Gen. Virol.* 96 (8), 2000–2027. doi:10.1099/vir.0.000201
- Sokolov, Y., Mirzabekov, T., Martin, D. W., Lehrer, R. I., and Kagan, B. L. (1999). Membrane channel formation by antimicrobial protegrins. *Biochim. Biophys. Acta.* 1420 (1-2), 23–29. doi:10.1016/s0005-2736(99)00086-3
- Squier, T. C. (2001). Oxidative stress and protein aggregation during biological aging. *Exp. Gerontol.* 36 (9), 1539–1550. doi:10.1016/s0531-5565(01)00139-5
- Surguchov, A., Surgucheva, I., Sharma, M., Sharma, R., and Singh, V. (2017). Pore-forming proteins as mediators of novel epigenetic mechanism of epilepsy. *Front. Neurol.* 8, 3. doi:10.3389/fneur.2017.00003
- Voskoboinik, I., Smyth, M. J., and Trapani, J. A. (2006). Perforin-mediated target-cell death and immune homeostasis. *Nat. Rev. Immunol.* 6 (12), 940–952. doi:10.1038/nri1983
- Walsh, P., Vanderlee, G., Yau, J., Campeau, J., Sim, V. L., Yip, C. M., et al. (2014). The mechanism of membrane disruption by cytotoxic amyloid oligomers formed by prion protein(106-126) is dependent on bilayer composition. *J. Biol. Chem.* 289 (15), 10419–10430. doi:10.1074/jbc.M113.515866
- Yoichi, M., Eva, Z., Tatsuo, Y., and Vito, T. (2005). Conformational changes preceding amyloid-fibril formation of amyloid- beta, prion protein and stefin B; parallels in pH dependence. *Med. Chem. Rev. Online.* 2 (4), 359–367. doi:10.2174/1567203054637533
- Žganec, M., and Žerovnik, E. (2014). Amyloid fibrils compared to peptide nanotubes. *Biochim. Biophys. Acta Gen. Subj.* 1840 (9), 2944–2952. doi:10.1016/j.bbagen.2014.05.019

**Conflict of Interest:** The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Žerovnik. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Structural Basis of Potential Inhibitors Targeting SARS-CoV-2 Main Protease

Hylemariam Mihiretie Mengist<sup>1,2,3</sup>, Tebelay Dilnessa<sup>3</sup> and Tengchuan Jin<sup>1,2,4\*</sup>

<sup>1</sup>Department of Obstetrics and Gynecology, The First Affiliated Hospital of USTC, Division of Life Sciences and Medicine, University of Science and Technology of China, Hefei, China, <sup>2</sup>Hefei National Laboratory for Physical Sciences at Microscale, CAS Key Laboratory of innate immunity and chronic disease, School of Basic Medical Sciences, Division of Life Sciences and Medicine, University of Science and Technology of China, Hefei, China, <sup>3</sup>Department of Medical Laboratory Science, College of Health Science, Debre Markos University, Debre Markos, Ethiopia, <sup>4</sup>CAS Center for Excellence in Molecular Cell Science, Chinese Academy of Science, Shanghai, China

## OPEN ACCESS

### Edited by:

Chandrabose Selvaraj,  
Department of Bioinformatics,  
Alagappa University, India

### Reviewed by:

Wenshe R Liu,  
Texas A&M University, College  
Station, United States  
Kailin Yang,  
Cleveland Clinic, Cleveland,  
United States  
Tamer M Ibrahim,  
Faculty of Pharmacy, Kafrelsheikh  
University, Egypt

### \*Correspondence:

Tengchuan Jin  
jint@ustc.edu.cn

### Specialty section:

This article was submitted to  
Medicinal and Pharmaceutical  
Chemistry,  
a section of the journal  
Frontiers in Chemistry

**Received:** 30 October 2020

**Accepted:** 04 January 2021

**Published:** 12 March 2021

### Citation:

Mengist HM, Dilnessa T and Jin T  
(2021) Structural Basis of Potential  
Inhibitors Targeting SARS-CoV-2  
Main Protease.  
Front. Chem. 9:622898.  
doi: 10.3389/fchem.2021.622898

The Coronavirus disease-19 (COVID-19) pandemic is still devastating the world causing significant social, economic, and political chaos. Corresponding to the absence of globally approved antiviral drugs for treatment and vaccines for controlling the pandemic, the number of cases and/or mortalities are still rising. Current patient management relies on supportive treatment and the use of repurposed drugs as an indispensable option. Of a crucial role in the viral life cycle, ongoing studies are looking for potential inhibitors to the main protease ( $M^{pro}$ ) of severe acute respiratory syndrome Coronavirus -2 (SARS-CoV-2) to tackle the pandemic. Although promising results have been achieved in searching for drugs inhibiting the  $M^{pro}$ , work remains to be done on designing structure-based improved drugs. This review discusses the structural basis of potential inhibitors targeting SARS-CoV-2  $M^{pro}$ , identifies gaps, and provides future directions. Further, compounds with potential  $M^{pro}$  based antiviral activity are highlighted.

**Keywords:** COVID-19, SARS-CoV-2, crystal structure, main protease, inhibitors

## INTRODUCTION

Since its emergence in Wuhan, China (Huang et al., 2020), COVID-19 (caused by the novel SARS-CoV-2) has been causing significant mortality and morbidity worldwide. The pandemic sparked global attention affecting every corner of the world and is changing the social, economic, and political status of the globe. As of 15 December 2020, the number of confirmed cases is over 73 million and deaths have surpassed 1.63 million (<https://www.worldometers.info/coronavirus/>). Curbing the spread of the virus has been challenging as it has various means of transmission including direct contact, *via* droplets, airborne, fomite, fecal-oral, bloodborne, sexual intercourse, ocular, mother-to-child, and animal-to-human (Patel et al., 2020). Although the virus primarily causes a mild respiratory illness, significant proportions of patients experience severe disease with outcomes of death. Moreover, there is also a significant number of asymptomatic infections that can transmit the virus to others. COVID-19 patients with underlying conditions are known to have a higher risk of developing a severe disease (Chow et al., 2020; Zhang et al., 2020c).

The Remdesivir drug and the Pfizer vaccine have been approved by the USA FDA for emergency use, but there are (at the time of writing this article) no globally approved specific antiviral drugs and vaccines for official use. The primary treatment relies on symptomatic and oxygen therapy to manage respiratory impairment. When there is respiratory failure, mechanical ventilation is recommended to prevent respiratory arrest. In the case of complicated disease, intensive care is needed because of acute respiratory distress syndrome (ARDS) or multiple organ failure (MOF) (Cascella et al., 2020;

Chen et al., 2020a; Gattinoni et al., 2020). Fifteen drugs (chloroquine, hydroxychloroquine, lopinavir, ritonavir, nafamostat, camostat, famotidine, umifenovir, nitazoxanide, ivermectin, corticosteroids, tocilizumab, sarilumab, bevacizumab, and fluvoxamine) are under clinical trial but conducting solid clinical trials is reportedly more difficult with increased public inquiry over readily available drugs (Shaffer, 2020). A combination of drugs could be more effective; for example, a combination of antitussive nescapine and hydroxychloroquine showed a strong binding affinity to SARS-CoV-2 M<sup>pro</sup> (Kumar et al., 2020b). A tremendous number of studies are underway to determine the therapeutic use of antivirals (bemcentinib, chloroquine & hydroxychloroquine, lopinavir boosted with ritonavir and remdesivir) and immune modulators (anakinra and canakinumab, azithromycin, brensocatib, convalescent plasma, corticosteroids, interferon beta, ruxolitinib, mesenchymal stromal cells and sarilumab and tocilizumab) to treat COVID-19 (Connelly, 2020).

Treatment of COVID-19 is medically unmet and designing potential drugs that could halt infection and disease progression is critical. Designing drugs that directly act on conserved enzymes like the main protease or 3C-like protease (M<sup>pro</sup> or 3CL<sup>pro</sup>), papain-like protease (PL<sup>pro</sup>), non-structural protein 12 (nsp12), and RNA-dependent RNA polymerase (RdRP) could be broad-spectrum and effective (Zumla et al., 2016). Remdesivir is one of the antivirals under clinical trial for COVID-19 treatment with probable inhibition of RNA synthesis *via* targeting RdRP (Saha et al., 2020). In a randomized controlled trial of 1,062 patients, compared to a placebo, remdesivir significantly shortened the recovery time of adult COVID-19 patients, suggesting its therapeutic role (Beigel et al., 2020). Its clinical effect on severely ill patients, however, is controversial.

Several studies combining structure-based, virtual, and high-throughput screening methods are currently underway to identify SARS-CoV-2 M<sup>pro</sup> inhibitors (Zhu et al., 2020). Summarizing the results of these studies, identifying their gaps, and appraising critiques are crucial to putting forward strong recommendations and future directions. Therefore, this review discusses recent advancements and prospects of structure-based drug designing activities that target SARS-CoV-2 M<sup>pro</sup>.

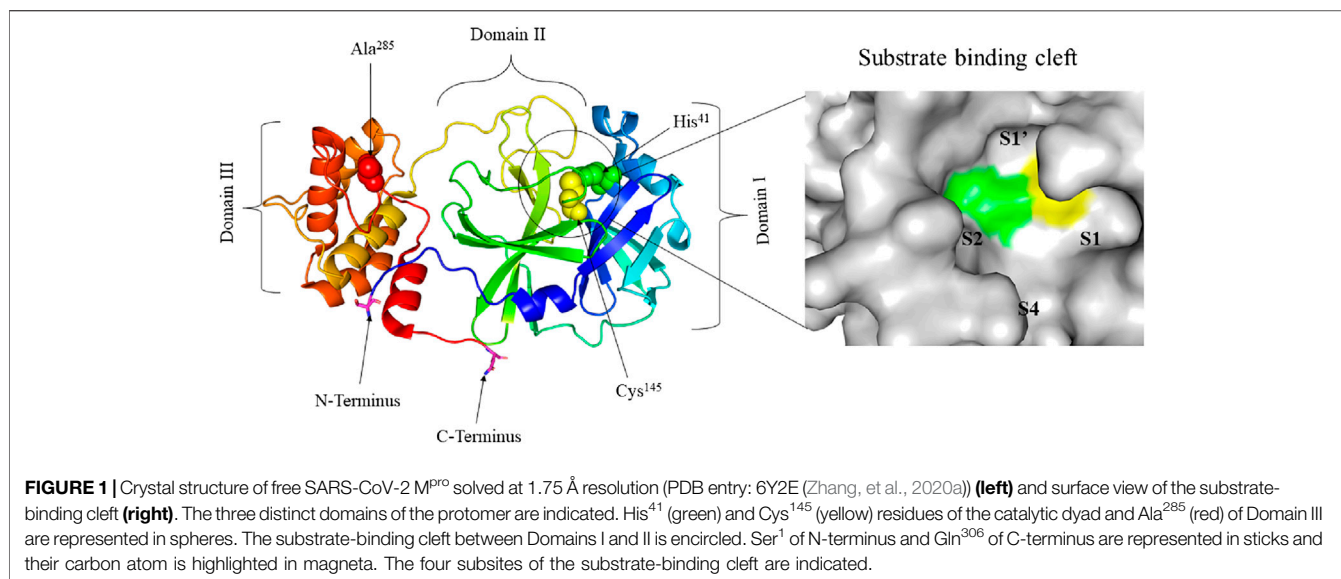
## SARS-COV-2 M<sup>PRO</sup> AS A DRUG TARGET

For many viruses, the protease enzyme plays a critical role in viral protein maturation by cleaning proproteins after their translation into the host cell cytosol. As a result, viral proteases are often potential drug targets. The inhibition of viral protease can reduce the assembly of mature viral particles. To date, many antiviral drugs have been developed against viral infections *via* targeting proteases. For instance, HIV-1 protease inhibitors (tipranavir, darunavir, amprenavir, lopinavir, saquinavir, atazanavir, indinavir, ritonavir, and nelfinavir) (Lv et al., 2015) and hepatitis C virus (HCV) NS3/4A protease inhibitors (boceprevir, telaprevir, ritonavir, asunaprevir, paritaprevir, grazoprevir, glecaprevir, voxilaprevir, and sofosbuvir) (de Leuw and Stephan, 2017) are amongst the FDA approved drugs.

Therefore, formulating antiviral drugs inhibiting SARS-CoV-2 M<sup>pro</sup> could also have potential clinical use.

SARS-COV-2 is one of the seven medically important coronaviruses and has been causing the most catastrophic once in a century disease of pathogens (Cui et al., 2019; Gates, 2020; Gorbalenya et al., 2020). SARS-CoV-2 is an enveloped betacoronavirus with a positive-strand large RNA genome (Holmes and Lai, 1996; Fehr and Perlman, 2015; Lu et al., 2020). Although SARS-CoV-2 has a large RNA genome of about 30 kb, it encodes only a few proteins (Dömling and Gao, 2020). Among these proteins, M<sup>pro</sup>, a cysteine protease, mediates the maturation cleavage of polyproteins during virus replication (Gorbalenya et al., 1989; Lee et al., 1991; Ziebuhr et al., 2000). The M<sup>pro</sup> is a homodimer containing two protomers each, comprising three domains (Domains I, II, and III). Domains I and II, comprised of residues 8–101 and 102–184, respectively, are made up of six antiparallel  $\beta$ -barrels. An antiparallel globular cluster of five  $\alpha$  helices forms domain III (residues 201–303) which is connected to domain II *via* a long loop region (residues 185–200). In the cleft between domains I and II, there is a Cys-His catalytic dyad which, together with N-terminus residues 1 to 7, is thought to have a vital role in proteolytic activity (Anand et al., 2002; Anand et al., 2003; Yang et al., 2003; Shi et al., 2004; Xue et al., 2007). The substrate-binding site is located in the cleft between domains I and II and the protomers, which bind each other through N-terminus residues 1–7, are located between domains II and III with roles in the formation of the substrate-binding site (Yang et al., 2003; Chou et al., 2004; Hsu et al., 2005; ul Qamar et al., 2020; Zhang et al., 2020a). The substrate-binding cleft is comprised of four subsites namely; S1', S1, S2, and S4 (Figure 1). The M<sup>pro</sup> is a conserved protein across all coronaviruses and the amino acids in substrates are numbered as -P4-P3-P2-P1 and P1'-P2'-P3'- from the N-terminus to the C-terminus (Hayden et al., 2003; Yang et al., 2005; Kim et al., 2016). The cleavage site is located between P1 and P1' and a glutamine residue is required in the P1 position (Dai et al., 2020; Jin et al., 2020a; Su et al., 2020; Zhang et al., 2020a).

The interaction between the two protomers determines the activity of the enzyme. The interaction of the N-terminus of one protomer with domain II of the other *via* hydrogen bonding, helps shape the S1 pocket of the active site. Therefore, the dimer is the active form while the monomer is inactive. Dimerization involves intermolecular interactions between the two protomers. Salt bridges between the N-terminus domain III of one protomer and electrostatic and hydrophobic interactions between two different domains III help enzyme dimerization (Fan et al., 2004; Zhang et al., 2020a). Grottesi et al. (2020) used computational approaches to study the structure and function of SARS-CoV-2 M<sup>pro</sup>. The authors demonstrated that when the average volume of the binding pocket increases in one chain, a decrease takes place in the other chain. Moreover, the interactions between the N-terminus and domain III of one monomer stabilizes the residues in the pocket. While dimerization is crucial for enzyme activity, one protomer is active and the other is inactive.



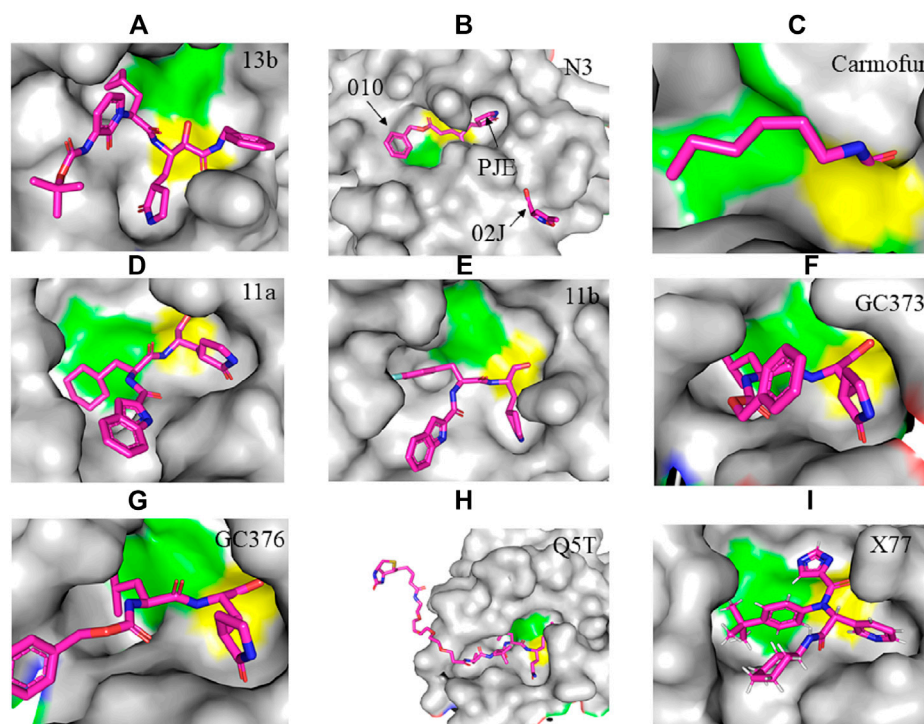
Understanding the atomic-level mechanism of the peptide cleavage, catalyzed by cysteine proteases, is crucial for designing structure-based potent inhibitors. Earlier studies proposed that the imidazole group of histidine polarizes and activates the SH group of the cysteine forming CysS<sup>-</sup>/HisH<sup>+</sup> ion which has a high nucleophilic property that reacts with substrates (Keillor and Brown, 1992). A QM/MM study reported that proteolysis catalyzed by cruzain cysteine protease has acetylation and deacetylation stages. In the acetylation step, cysteine attacks the carbonyl carbon atom of the peptide after which the proton from the protonated HisH<sup>+</sup> is transferred to the nitrogen atom of the scissile peptide bond. Then the deacetylation stage is supposed to be assisted by a water molecule activated by histidine (Arafet et al., 2017; Swiderek and Moliner, 2020).

A similar study by Swiderek and Moliner (2020) presented the cleavage of a polypeptide, Ac-Val-Lys-Leu-Gln-ACC, catalyzed by SARS-CoV-2 M<sup>pro</sup>. First, a proton is transferred from Cys<sup>145</sup> to His<sup>41</sup> with a simultaneous nucleophilic attack of the carbonyl carbon atom of the peptide bond by the sulfur atom of Cys<sup>145</sup> which results in a thiohemiketal intermediate. The transfer of a proton from His<sup>41</sup> to the nitrogen atom of the substrate, which forms an acyl-enzyme complex intermediate, assists the cleavage of the peptide bonds. This reaction produces the first product ACC released from the active site. After the release of ACC, an activated water molecule attacks the carbonyl carbon atom of Gln<sup>5</sup> of the peptide with simultaneous transfer of proton to His<sup>41</sup>. Finally, the second product species is released after the covalent bond between Cys<sup>145</sup> and the peptide in the thiohemiketal intermediate is broken.

A DFM/MM simulation study revealing an equivalent mechanism of the reaction is reported by Ramos-Guzmán et al. (2020) using the peptidomimetic Ac-Ser-Ala-Val-Leu-His-aldehyde inhibitor as a substrate. These studies provide insights into the structure-based design of potential M<sup>pro</sup> inhibitors that can form a stable enzyme-inhibitor complex

similar to the product in the acetylation step of the proteolysis reaction. These studies help the scientific community deeply understand the structure and function of the M<sup>pro</sup> which is important for effectively designing potent inhibitors.

The main protease of coronaviruses is a potential drug target since it is responsible for the maturation of itself and other important polyproteins (Ziebuhr et al., 2000). SARS-CoV-2 has 14 open reading frames (ORFs). The M<sup>pro</sup> (nsp5), encoded by the major ORF1ab, cleaves two overlapping polyproteins (pp1a and pp1ab) into 16 non-structural proteins which are important for viral replication and maturation (Paul, 2006; Ziebuhr et al., 2000; Chen et al., 2020b; Gordon et al., 2020). In addition, it plays a significant role in virus entry to host cells where inhibition of this enzyme halts the viral entry and the subsequent infection (Jain and Mujwar, 2020). These important functions of the viral protease enzyme purpose itself are an interesting therapeutic target for curbing coronavirus associated diseases (Thiel et al., 2003; Naqvi et al., 2020). Structurally optimized broad-spectrum drugs are effectively inhibiting the main protease of coronaviruses pertaining to its relatively conserved nature (Yang et al., 2005; Xue et al., 2007). Similarly, SARS-CoV-2 M<sup>pro</sup> has attracted great attention to the development of drugs to fight the ongoing COVID-19 pandemic. In addition to the M<sup>pro</sup>, other proteins including spike protein (S), RdRP, NTPase/helicase, and papain-like protease are currently alternative drug targets (Wu et al., 2020). The M<sup>pro</sup> is also critically important for the proteolytic release of enzymes essential for viral replication including nsp 13 which has NTPase and RNA helicase activity (Thiel et al., 2003; Shu et al., 2020). Homology modeling studies presented structural similarity between SARS-CoV-2, SARS-CoV, and MERS-CoV main proteases with a conserved active site (Stoermer, 2020; Ullrich and Nitsche, 2020), and it is also noted that the binding of lead compounds is similar in both SARS-CoV and SARS-CoV-2 main proteases regardless of the protonation state of the Cys-His catalytic dyad (Macchiagodena et al., 2020) indicating the



**FIGURE 2 |** Crystal structure of SARS-CoV-2 M<sup>Pro</sup> in complex with potential inhibitors. (A) 13b (PDB entry: 6Y2G, 2.20 Å resolution) (B) Michael acceptor N3 (PDB entry: 6LU7, 2.16 Å resolution) (C) Carmofur (PDB entry: 7BUY, 1.60 Å resolution) (D) 11a (PDB entry: 6LZE, 1.505 Å resolution) (E) 11b (PDB entry: 6M0K, 1.504 Å resolution) (F) GC373 (PDB entry: 6WTK, 2.00 Å resolution) (G) GC376 (PDB entry: 6WTT, 2.15 Å resolution) (H) Q5T (PDB entry: 6Z2E, 1.70 Å resolution) and (I) X77 (PDB entry: 6W63, 2.10 Å resolution). Residues in the catalytic dyad: His<sup>41</sup> is highlighted green and Cys<sup>145</sup> is highlighted yellow. The carbon atoms of each drug are highlighted in magenta.

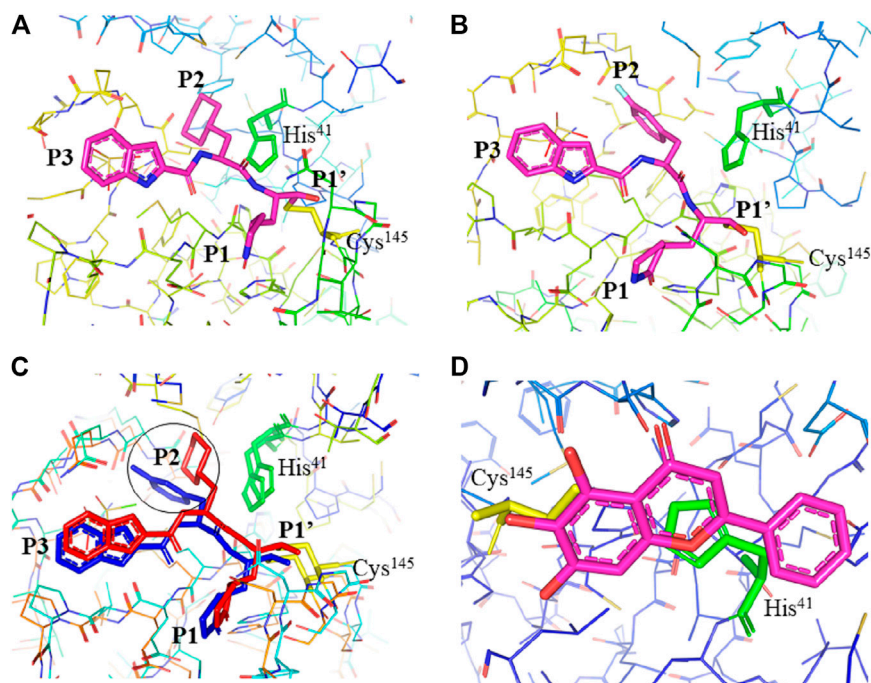
possibility of designing broad-spectrum drugs against these viruses.

The enhanced activity observed in SARS-CoV-2 M<sup>Pro</sup> providing evidence for repurposing it as a potential drug target. Mutations (Ser284Ala, Thr285Ala, and Ile286Ala) in SARS-CoV M<sup>Pro</sup> are reported to result in an enhanced activity where the two similar mutations (Thr285Ala and Ile286Leu) in SARS-CoV-2 M<sup>Pro</sup> caused the higher activity of SARS-CoV-2 M<sup>Pro</sup> over SARS-CoV M<sup>Pro</sup> (Lim et al., 2014; Zhang et al., 2020a). Thus, the design of improved broad-spectrum inhibitors should consider key amino acid differences that occur in SARS-CoV-2 M<sup>Pro</sup> compared to previous viruses. Main proteases also have substrate recognition site preference in cleaving polyproteins which is important when designing specific inhibitors. The main proteases of the three viruses (SARS-CoV-2, SARS-CoV, and MERS) have very similar substrate recognition profiles with heightened preference to glutamine in the P1 of polyproteins, but SARS-CoV M<sup>Pro</sup> demonstrated broader substrate specificity at the P2 position given that the two enzymes (SARS-CoV and SARS-CoV-2 M<sup>Pro</sup>) prefer leucine at this position (Rut et al., 2020a). Interestingly, human host-cell proteases with such similar specificity have not been reported yet causing the anticipated drugs to have reduced off-target activities (Hilgenfeld, 2014; Zhang et al., 2020b) proving that it was the right decision to select SARS-CoV-2 M<sup>Pro</sup> as an outstanding drug target. Therefore, although it needs clinical evidence, M<sup>Pro</sup>

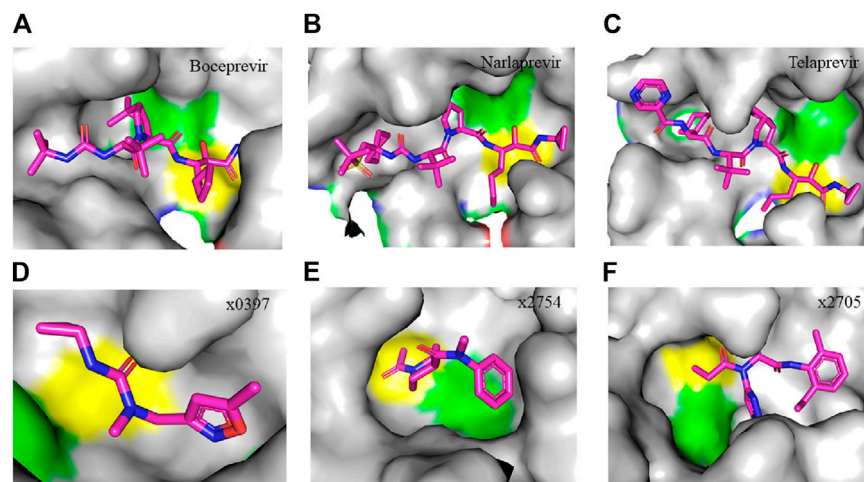
targeting drugs are thought to be suitable for human beings and have fewer side effects.

## STRUCTURE-BASED DESIGN OF DRUGS THAT TARGET SARS-COV-2 M<sup>PRO</sup>

Drugs that specifically bind to and inhibit SARS-CoV-2 M<sup>Pro</sup> could be promising alternatives to fight the pandemic. Gly<sup>143</sup> of SARS-CoV-2 M<sup>Pro</sup> is reported to be the most attractive residue to form hydrogen bonds with ligands followed by Glu<sup>166</sup>, Cys<sup>145</sup>, and His<sup>163</sup> (Nguyen et al., 2020). Therefore, determining the crystal structure of viral proteases in a complex with potential inhibitors is vital as it provides a glimpse into designing improved drugs through the modification of the inhibitors according to the structural dynamics (monomer or dimer, narrow or wide, deep or shallow) of the active site in the target enzymes. For example, AG7088 is a potent inhibitor of Rhinoviruses and other Picornaviral 3C-like proteases (3C<sup>Pro</sup>), but not for SARS-CoV M<sup>Pro</sup> because the latter is a monomer with only two catalytic domains (Binford et al., 2005; Shie et al., 2005; Kuo et al., 2009) indicating the importance of modifying drugs accordingly, as sequence differences and structural alterations significantly affect the specificity of inhibitors. Additionally, the monomer of M<sup>Pro</sup> is principally considered inactive and therefore the dimer is the best alternative drug target (Grum-Tokars et al., 2008; Pillaiyar et al.,



**FIGURE 3** | Interaction of 11a, 11b, and baicalein (PDB entry: 6M2N, 2.198 Å resolution) with SARS-CoV-2 M<sup>pro</sup>. **(A)** Interaction of 11a and **(B)** 11b with residues at the subsites of the substrate-binding cleft. The carbon atoms of 11a and 11b are highlighted in magenta. **(C)** Comparison of the difference in the binding modes of 11a (red) and 11b (blue). The main difference at the P2 position is encircled. **(D)** The unique binding of baicalein perfectly inserted in the core position of the substrate-binding pocket where the S1/S2 subsites and the oxyanion loop shielding the active site from a peptide substrate (Su et al., 2020). Baicalein is highlighted in magenta. His<sup>41</sup> is highlighted green and Cys<sup>145</sup> is highlighted yellow.



**FIGURE 4** | Crystal structure of SARS-CoV-2 M<sup>pro</sup> in complex with clinically approved antiviral drugs and PanDDA analyzed fragments. **(A)** Boceprevir (PDB entry: 7K40, 1.35 Å resolution) **(B)** narlaprevir (PDB entry: 7JYC, 1.79 Å resolution) **(C)** telaprevir (PDB entry: 7K6D, 1.48 Å resolution) **(D)** x0397 (PDB entry: 5RGI, 1.57 Å resolution) **(E)** x2754 (PDB entry: 5RHF, 1.76 Å resolution) and **(F)** x2705 (PDB entry: 5RH7, 1.71 Å resolution). Residues in the catalytic dyad: His<sup>41</sup> is highlighted green and Cys<sup>145</sup> is highlighted yellow. The carbon atoms of each drug are highlighted in magenta. All structures described in the figures in this paper are solved by X-ray crystallography.

2016). Moreover, designing inhibitors based on their competitive binding to the active site, could help in identifying the best inhibitors. The illustration of the binding of different compounds with SARS-CoV-2 M<sup>Pro</sup> is described in **Figures 2–4**.

Earlier, Xue et al. (2008) demonstrated the crystal structure of infectious bronchitis virus (IBV) M<sup>Pro</sup> and an active site mutant, His41Ala, SARS-CoV M<sup>Pro</sup> in complex with N-terminal autocleavage substrate, and described the implications for substrate binding and antiviral drug design. Unlike IBV M<sup>Pro</sup>, the outer wall of SARS-CoV M<sup>Pro</sup> in the S1 subsite is made up of residues 141 to 143. The S1 and S2 pockets of IBV M<sup>Pro</sup> are comparatively larger than SARS-CoV M<sup>Pro</sup>, since Ala<sup>140</sup> of IBV M<sup>Pro</sup> is away from the active site and its Lys<sup>45</sup> is about 2 Å away from the S2 subsite. Here the authors suggested that modification of the P3 position of substrates to have a large side chain is a good choice to design substrate-based inhibitors for the main protease of coronaviruses. New inhibitors, N27 and H16, which have relatively large side chains at the P3 position compared to the previously designed inhibitor N3, showed more potent inhibition of SARS-CoV M<sup>Pro</sup> but similar activity with N3 against IBV M<sup>Pro</sup>. Moreover, the N3 inhibitor inactivated the M<sup>Pro</sup> of IBV *in vitro* and demonstrated potent antiviral activity against IBV in chicken embryos (Xue et al., 2008). Another study (Xue et al., 2007) reported that the addition of residues at the N terminus, not the C-terminus, of SARS-CoV M<sup>Pro</sup> affects the enzyme activity. Based on the crystal structure of wild type SARS-CoV M<sup>Pro</sup> in complex with Michael receptor N3, it has been proven that the first N terminus residues of the enzyme are vital for keeping the inhibitor binding cleft. These studies laid a concrete foundation for the design of broad-spectrum inhibitors for coronaviruses, including SARS-CoV-2.

The crystal structure of SARS-CoV-2 M<sup>Pro</sup> in complex with different potential drugs has been illustrated. Zhang et al. (2020a) reported the X-ray crystal structure of SARS-CoV-2 M<sup>Pro</sup> in complex with peptidomimetic  $\alpha$ -ketoamide inhibitors. The authors modified a previously designed inhibitor (**11r**) by incorporating a P2-P3 amide bond into the pyridone ring (**13a**) with an enhanced half-life in plasma and noticeable tropism to the lung. Moreover, they modified **13a** into a more potent but a narrow-spectrum drug **13b** (**Figure 2A**) by replacing the P2 cyclohexyl moiety with a small cyclopropyl where the drug binds in the shallow substrate-binding pocket at the surface of protomers between domains I and II. Although the improvements (**11r** to **13a**) resulted in a pharmacokinetically better drug, structural modifications negatively affected some inhibitory activities of the drug against SARS-CoV-2 M<sup>Pro</sup> (Mengist et al., 2020; Zhang et al., 2020a) indicating the cautious design of candidate drugs.

In their study, Jin et al. (2020a) (**Figure 2B**) demonstrated the structure of SARS-CoV-2 M<sup>Pro</sup> in complex with N3. Their results showed that N3 binds inside the pocket of the substrate-binding site. The interaction of N3 with the M<sup>Pro</sup> is in such a way that the S $\gamma$  atom of Cys<sup>145</sup> of protomer A forms a covalent bond with the C $\beta$  atom of the vinyl group. The P1 lactam inserts into the S1 subsite forming a hydrogen bond with His<sup>163</sup> of protomer A, whereas the side chain of leucine at the P2 site inserts deeply into the hydrophobic S2 subsite. Additionally, the side chain of valine

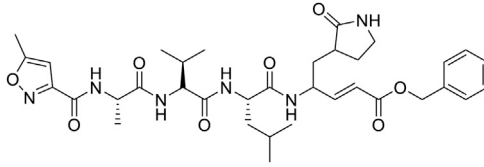
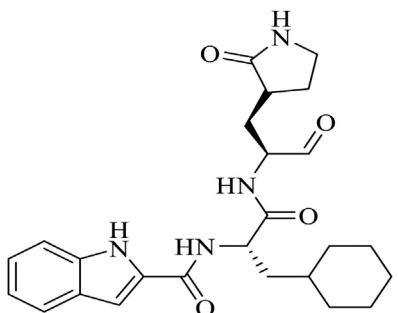
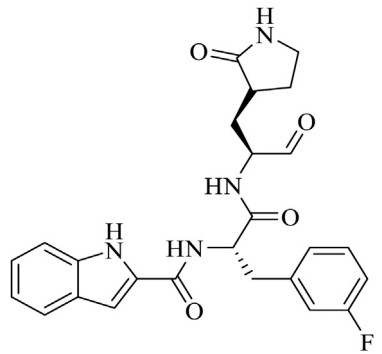
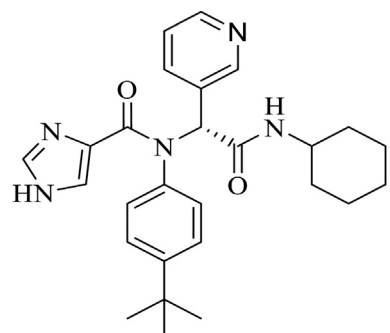
at P3 is solvent-exposed and the side chain of alanine at the P4 side is surrounded by the side chains and main chains of protomer A while the P5 form contacts with Pro<sup>168</sup> of protomer A and with residues 190–191 at the backbone. Overall, the authors illustrated the specific binding of N3 with the main chain of the substrate-binding pocket through multiple hydrogen bonds and then pharmacokinetically exerting a two-step irreversible inactivation of SARS-CoV-2 M<sup>Pro</sup>. Additionally, Jin et al. (2020b) also demonstrated that carmofur, an antineoplastic drug, binds on the Cys<sup>145</sup> catalytic dyad of SARS-CoV-2 M<sup>Pro</sup> (**Figure 2C**) with promising *in vitro* inhibition of virus replication.

Dai et al., (2020) designed two drugs (**11a** and **11b**) against SARS-CoV-2 M<sup>Pro</sup> and described structural-functional considerations in designing potent inhibitors based on the structure of the substrate-binding site. The two drugs were found to be outstanding main protease inhibitors that can also halt SARS-CoV-2 infection. The crystal structure of the complex showed that the aldehyde group of both drugs covalently bind to Cys<sup>145</sup> with an *in vivo* auspicious pharmacokinetic property. The antiviral activity of the drugs was maintained by covalent anchoring from the thiol of a cysteine residue in the S1' subsite of the substrate-binding pocket. In case of **11a** (**Figure 2D**), the carbon atom of the aldehyde group and Cys<sup>145</sup> of SARS-CoV-2 M<sup>Pro</sup> form a C–S covalent bond. **11b** exhibits a similar inhibitory binding mode with **11a**, with a small difference probably due to the 3-fluorophenyl group of **11b** at P2, which experiences a downward rotation (**Figure 2E**). The oxygen atom in the aldehyde group in **11a** stabilizes the conformation of the drug by forming a hydrogen bond with the backbone of Cys<sup>145</sup> in the S1' subsite while its (S)- $\gamma$ -lactam ring at P1 fits in the S1 subsite. The differences in the binding modes of **11a** and **11b** are illustrated in **Figures 3A–C**. Here, the authors demonstrated the pros and cons of modifying drugs at relevant positions (P1, P2, P3, P4, or P5) through detailed structural-functional explanations.

It has been reported that baicalein exhibited a unique binding mode with SARS-CoV-2 M<sup>Pro</sup> as it does not have direct contact with the 12 amino acids which differed SARS-CoV and SARS-CoV-2 main proteases (Su et al., 2020). The binding of baicalein on the S1 and S2 subsites of the active site is possible through multiple hydrogen bonds between three phenolic hydroxyl groups of the ligand and Leu<sup>141</sup>/Gly<sup>143</sup> at the main chains in addition to Ser<sup>144</sup>/His<sup>163</sup> at the side chains. Glu<sup>166</sup> at the main chain form hydrogen bonding with the carbonyl group whereas the insertion of the free phenyl group into the sub-pocket of S2 was maintained by hydrophobic interactions with Gln<sup>189</sup>, Arg<sup>188</sup>, Met<sup>49</sup>, Cys<sup>44</sup>, and His<sup>41</sup>. Furthermore, the aromatic ring of baicalein forms S- $\pi$  and  $\pi$ - $\pi$  interactions with His<sup>41</sup> and Cys<sup>145</sup>, respectively (**Figure 3D**).

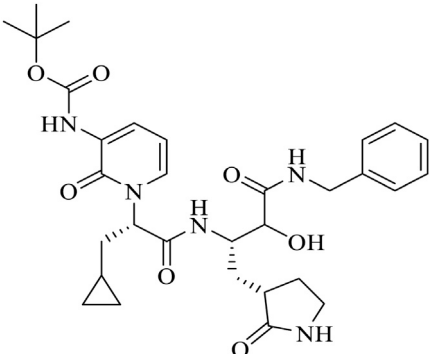
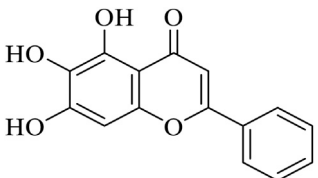
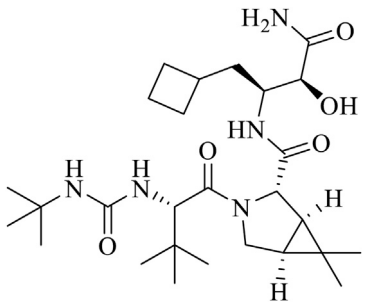
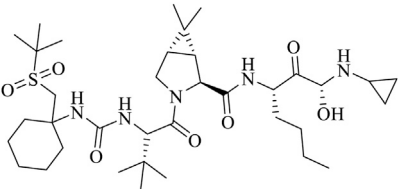
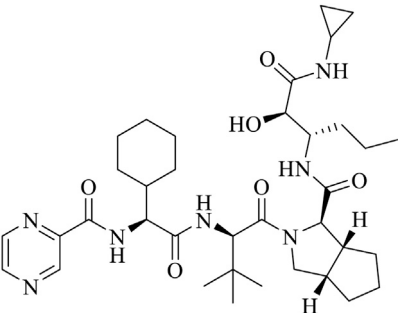
According to Vuong et al. (2020), GC376 was converted into GC373 upon incubation with SARS-CoV-2 M<sup>Pro</sup> which formed a covalent bond with Cys<sup>145</sup> (**Figure 2F**). Accordingly, residues of domain II form contacts supporting drug binding where P2 was inserted into the hydrophobic pocket consisting of His<sup>41</sup>, while the S2 subsite was represented by Met<sup>49</sup> and Met<sup>165</sup>. His<sup>163</sup> and Glu<sup>166</sup> side chains form a hydrogen bond with the glutamine surrogate in the P1 position whereas a hydrophobic interaction was noticed with His<sup>172</sup> while hydrogen bond connects the backbone amide of Glu<sup>166</sup> with carbonyl in the P3, suggesting the strong binding capacity of the drugs on the catalytic site of the

**TABLE 1** | Details of compounds with their complex structure with SARS-CoV-2 M<sup>pro</sup> solved, which have potential subsequent antiviral activity.

Name	PDBChem code	PDB entry	Chemical formula	Molecule name	Chemical structure
N3		6LU7	C35H48N6O8	benzyl (3S,6R,9S,E)-9-isobutyl-6-isopropyl-3-methyl-1-(5-methylisoxazol-3-yl)-1,4,7,10-tetraoxo-12-((2-oxopyrrolidin-3-yl)methyl)-2,5,8,11-tetraazapentadec-13-en-15-oate	
11a	FHR	6LZE	C25H32N4O4	(~{N})-[(2~{S})-3-cyclohexyl-1-oxidanylidene-1-[[[(2~{S})-1-oxidanylidene-3-[(3~{S})-2-oxidanylidene-3-yl]propan-2-yl]amino]propan-2-yl]-1~{H})-indole-2-carboxamide	
11b	FJC	6MOK	C25H25FN4O4	~{N})-[(2~{S})-3-(3-fluorophenyl)-1-oxidanylidene-1-[[[(2~{S})-1-oxidanylidene-3-[(3~{S})-2-oxidanylidene-3-yl]propan-2-yl]amino]propan-2-yl]-1~{H})-indole-2-carboxamide	
X77	X77	6W63	C27H33N5O2	N-(4-tert-butylphenyl)-N-[(1R)-2-(cyclohexylamino)-2-oxo-1-(pyridine-3-yl)ethyl]-1H-imidazole-4-carboxamide	

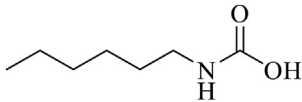
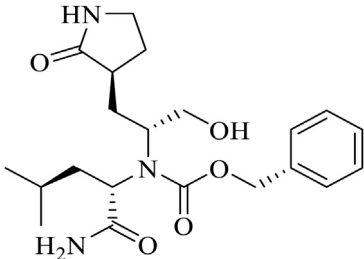
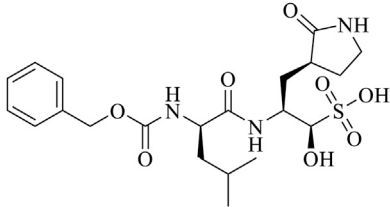
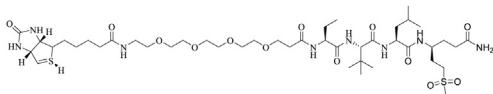
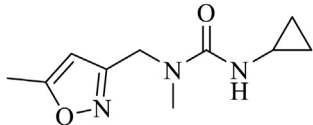
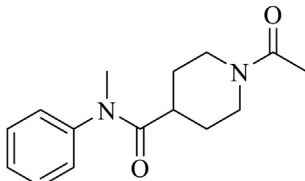
(Continued on following page)

**TABLE 1 |** (Continued) Details of compounds with their complex structure with SARS-CoV-2 M<sup>pro</sup> solved, which have potential subsequent antiviral activity.

Name	PDBChem code	PDB entry	Chemical formula	Molecule name	Chemical structure
13b	O6K	6Y2G	C31H41N5O7	(tert)-butyl ~{N}-[1-[(2~{S})-3-cyclopropyl-1-oxidanylidene-1-[[[2~{S},3~{R})-3-oxidanyl-4-oxidanylidene-1-[(3~{S})-2-oxidanylidene-pyrrolidin-3-yl]-4-[(phenylmethyl)amino]butan-2-yl]amino]propan-2-yl]-2-oxidanylidene-pyridin-3-yl]carbamate	
Baicalein	3WL	6M2N	C15H10O5	5,6,7-trihydroxy-2-phenyl-4H-chromen-4-one	
Boceprevir	U5G	7K40	C27H47N5O5	(1R,2S,5S)-N-[(2S,3R)-4-amino-1-cyclobutyl-3-hydroxy-4-oxobutan-2-yl]-3-[N-(tert-butylcarbamoyl)-3-methyl-L-valyl]-6,6-dimethyl-3-azabicyclo[3.1.0]hexane-2-carboxamide	
Narlaprevir	NNA	7JYC	C36H63N5O7S	(1R,2S,5S)-3-[N-[(1-[(tert-butylsulfonyl)methyl]cyclohexyl]carbamoyl)-3-methyl-L-valyl]-N-[(1S)-1-[(1R)-2-(cyclopropyl amino)-1-hydroxy-2-oxoethyl]pentyl]-6,6-dimethyl-3-azabicyclo[3.1.0]hexane-2-carboxamide	
Telaprevir	SV6	7K6D	C36H55N7O6	(1S,3Ar,6As)-2-[(2S)-2-[(2S)-2-cyclohexyl-2-[(pyrazin-2-ylcarbonyl)amino]acetyl]amino)-3,3-dimethylbutanoyl]-N-[(2R,3S)-1-(cyclopropylamino)-2-hydroxy-1-oxohexan-3-yl]octahydrocyclopenta[c]pyrrole-1-carboxamide	

(Continued on following page)

**TABLE 1** | (Continued) Details of compounds with their complex structure with SARS-CoV-2 M<sup>pro</sup> solved, which have potential subsequent antiviral activity.

Name	PDBChem code	PDB entry	Chemical formula	Molecule name	Chemical structure
Carmofur	JRY	7BUY	C7H15NO2	hexylcarbamic acid	
GC373	UED	6WTK	C21H31N3O5	N~2~-[(benzyloxy)carbonyl]-N-[(2S)-1-hydroxy-3-[(3S)-2-oxopyrrolidin-3-yl]propan-2-yl]-L-leucinamide	
GC376	K36	6WTT	C21H31N3O8S	(1S,2S)-2-[(N-[(benzyloxy)carbonyl]-L-leucyl)amino]-1-hydroxy-3-[(3S)-2-oxopyrrolidin-3-yl]propane-1-sulfonic acid	
biotin-PEG(4)-Abu-Tle-Leu-Gln-vinylsulfone	Q5T	6Z2E	C44H80N8O13S2	(4~{S})-4-[[[(2~{S})-2-[[[(2~{S})-2-[[[(2~{S})-2-[3-[2-[2-[2-[5-[(3~{a})~{S}),4~{R}),6~{a})~{R})]-2-oxidanylidene-3,3~{a},4,6~{a}-tetrahydro-1~{H}-thieno[3,4-d]imidazol-4-yl]]pentanoylamino]ethoxy]ethoxy]ethoxy]ethoxy]propanoylamino]butanoyl]amino]-3,3-dimethyl-butanoyl]amino]-4-methyl-pentanoyl]amino]-6-methylsulfonyl-hexanamide	
x0397	UOP	5RGI	C10H15N3O2	N'-cyclopropyl-N-methyl-N-[(5-methyl-1,2-oxazol-3-yl)methyl]urea	
X2754	UPJ	5RHF	C15H20N2O2	1-acetyl-N-methyl-N-phenylpiperidine-4-carboxamide	

(Continued on following page)

**TABLE 1 |** (Continued) Details of compounds with their complex structure with SARS-CoV-2 M<sup>Pro</sup> solved, which have potential subsequent antiviral activity.

Name	PDBChem code	PDB entry	Chemical formula	Molecule name	Chemical structure
X2705	UJ1	5RH7	C <sub>26</sub> H <sub>33</sub> N <sub>5</sub> O <sub>2</sub>	N-(5-tert-butyl-1H-pyrazol-3-yl)-N-[(1R)-2-[(2-ethyl-6-methylphenyl)amino]-2-oxo-1-(pyridin-3-yl)ethyl]propanamide	

M<sup>Pro</sup>. A simulation study by Jain and Mujwar, (2020) showed that Metocurine, a neuromuscular blocking agent, binds specifically with the substrate-binding cavity of the protease enzyme supported with residues Phe<sup>140</sup>, Leu<sup>141</sup>, Cys<sup>145</sup>, His<sup>163</sup>, His<sup>164</sup>, Met<sup>165</sup>, Glu<sup>166</sup>, Leu<sup>167</sup>, and Pro<sup>168</sup> repurposing the compound as a safe and effective prospective drug.

The crystal structure of three protomers of SARS-CoV-2 M<sup>Pro</sup> complexed with GC376 (Figure 2G) (Ma et al., 2020) showed unique binding configurations, suggesting the potential candidacy of the compound for COVID-19 treatment. The authors reported that GC376 formed numerous hydrogen bonds with the active site supported with covalent bonds which formed with aldehyde bisulfite warhead and Cys<sup>145</sup>. In another study, the oxygen atoms in the vinyl sulfone group of Q5T (Biotin-PEG (4)-Abu-Tle-Leu-Gln-VS (B-QS1-VS)) form hydrogen bonds with the amide groups of Gly<sup>143</sup> and Cys<sup>145</sup> while the catalytic cysteine residue was covalently linked to the C $\beta$  atom of the vinyl group. Although the polar side chains of the P3 form hydrogen bonds with Glu<sup>166</sup>, the authors did not find a well-defined pocket for the P3 moiety (Rut et al., 2021) (Figure 2H). A non-covalent broad-spectrum inhibitor X77 also binds to the substrate-binding cleft of SARS-CoV-2 M<sup>Pro</sup> (Figure 2I).

Clinically approved HCV NS3/4A protease inhibitors (boceprevir, naldaprevir, and telaprevir) showed specific binding on the active site of SARS-CoV-2 M<sup>Pro</sup> secondary to the structural similarity between proteases of the two viruses. Molecular docking revealed that boceprevir formed hydrogen bonding with different residues and hydrophobic interactions with key residues His<sup>41</sup>, Leu<sup>141</sup>, His<sup>164</sup>, Met<sup>165</sup>, Glu<sup>166</sup>, and Asp<sup>187</sup> through unique binding conformation at the active site (Bafna et al., 2020) (Figures 4A–C). Pan-Dataset Density analysis (PanDDA) of SARS-CoV-2 M<sup>Pro</sup> fragment screening showed that several compounds including x0397, x2754, and x2705 bind on the active site with possible inhibitory activities (Figures 4D–F). Crystallographic and electrophilic fragment analysis of SARS-CoV-2 M<sup>Pro</sup> showed the plasticity of the S1' subsite indicating an improved design of prospective potent inhibitors. For example, the side-chain movement of catalytic residues Cys<sup>145</sup> and His<sup>41</sup> was observed upon binding of Z369936976 compared to Z1129283193. Accordingly, the size

and shape of the S1' subsite were altered, resulting in exceptional binding of the compound to both S1 and S1' subsites (Douangamath et al., 2020). The details of compounds whose complex X-ray structure with SARS-CoV-2 M<sup>Pro</sup> is solved and reported to have potential subsequent antiviral activity, are described in Table 1.

## POTENTIAL INHIBITORS OF SARS-COV-2 M<sup>PRO</sup>

Both *in vitro/in vivo*, and *in silico* studies demonstrated that several classes of compounds showed effective binding and inhibition of SARS-CoV-2 M<sup>Pro</sup> (Sharma et al., 2020). In addition to *in vitro* experiments, *in vivo* experiments also illustrated suppression of SARS-CoV-2 infectivity secondary to inhibition of the main protease. For example, significant suppression of multiple coronaviruses by optimized M<sup>Pro</sup> inhibitors was reported in infected mice (Rathnayake et al., 2020). Therefore, pertaining to similar substrate-binding sites across the main proteases of coronaviruses, formulation of broad-spectrum inhibitors could be recommended provided that the structural and functional effect of some key residue differences are well tolerated.

### *In vitro/In Vivo* Inhibitors

Alpha-keto amides **11u** and **11r** demonstrated broad-spectrum inhibition of the main proteases of beta coronaviruses and alpha coronaviruses, and the 3C-proteases of enteroviruses in cell culture (Zhang et al., 2020b). Improved compounds **13a** and **13b** (Zhang et al., 2020a) demonstrated specific binding to SARS-CoV-2 M<sup>Pro</sup> with subsequent enzyme inactivation and reduction of viral infectivity. N3 is an irreversible Michael acceptor inhibitor (Yang et al., 2005) which covalently binds with SARS-CoV-2 M<sup>Pro</sup> through Michael reaction, blocking its active site (Griffin, 2020; Jin et al., 2020a). Cell-based assays showed strong antiviral activity of N3 at 10  $\mu$ M concentration in SARS-CoV-2 infected Vero cells (Jin et al., 2020a). Carmofur has been used to treat several cancers (Nishio et al., 1987; Morimoto and Koh, 2003; Sakamoto et al., 2005) which also

demonstrated a clinical potential inhibition of SARS-CoV-2 through targeting M<sup>Pro</sup> (Jin et al., 2020a; Jin et al., 2020b).

Ebselen is an anti-inflammatory, anti-oxidant, and cytoprotective drug which has been studied for treating multiple diseases including bipolar disorders (Singh et al., 2013) and hearing loss (Lynch and Kil, 2009; Kil et al., 2017). This compound showed low cytotoxicity in rats (Renison et al., 1982) whereas whether it is safe for humans is under investigation (Lynch and Kil, 2009; Masaki et al., 2016; Kil et al., 2017). Ebselen specifically binds on the active site of SARS-CoV-2 M<sup>Pro</sup> and showed strong antiviral activity recommended for treating diseases associated with coronaviruses (Jin et al., 2020a). Sies and Parnham (2020) discussed ebselen as a potential drug for COVID-19 with promising inhibition of SARS-CoV-2 M<sup>Pro</sup> provided that the *in vivo* antiviral activity of the drug is determined.

Cinanserin is a well-characterized serotonin antagonist that showed a strong reduction of SARS-CoV replication through inhibiting viral 3CL<sup>Pro</sup> (Chen et al., 2005). This compound demonstrated a moderate inhibitory activity against SARS-CoV-2 M<sup>Pro</sup> (Jin et al., 2020a) suggesting the potential role of cinanserin in preventing coronavirus diseases following targeted modification. Famotidine is an AG protein-coupled receptor antagonist under clinical trial for COVID-19 treatment which showed weak binding affinity to SARS-CoV-2 M<sup>Pro</sup> and only intravenous administration was suggested to be advantageous (Ortega et al., 2020) given that structural modifications could enhance its binding energy and antiviral activity.

Aldehydes are compounds consisting of -CHO as a functional group and carbonyl center (a carbon double bonded to oxygen) where the carbon atom is also bonded to a hydrogen atom or any generic alkyl or side chain R group (alkyl or saturated hydrocarbon). Dai et al., (2020) designed aldehyde-based drugs and reported that compounds **11a** and **11b** showed high anti-SARSCoV-2 M<sup>Pro</sup> activity with inhibition of 100% and 96% at 1  $\mu$ M, respectively. These compounds specifically bind on the Cys<sup>145</sup> of the catalytic dyad of M<sup>Pro</sup> and block its activity. Peptide aldehydes also inhibit the main protease of Feline coronavirus (FCoV) (Kim et al., 2015). Among these drugs, GC376 (a prodrug) and GC373 (a drug) specifically bind on the catalytic dyad of recombinant SARS-CoV-2 M<sup>Pro</sup> with potent *in vitro* inhibition at the nano-molar level (Vuong et al., 2020). According to Rathnayake et al. (2020), among compounds tested, **6e** showed more potent antiviral activity in SARS-CoV-2 infected Vero E6 cells, while **7j** showed effective binding with SARS-CoV-2 M<sup>Pro</sup>. The authors illustrated effective inhibition of multiple coronaviruses and increased survival of infected mice treated with M<sup>Pro</sup> inhibitors.

Nelfinavir is a protease inhibitor used to treat HIV and is predicted to be a potential inhibitor of SARS-Cov-2 M<sup>Pro</sup> as it showed strong binding affinity to the enzyme (Xu et al., 2020). Further, lopinavir and ritonavir bind to viral main proteases (Nukoolkarn et al., 2008) and have demonstrated effective suppression of the virus through binding and inactivating the M<sup>Pro</sup>, as evidenced by effective activity on SARS-CoV-2 patients (Liu and Wang, 2020). Lopinavir and ritonavir also showed a high binding ability to the active pocket of SARS-CoV-2 M<sup>Pro</sup>

where Thr<sup>24</sup>, Thr<sup>26</sup>, and Asn<sup>119</sup> are the key residues important for binding. Furthermore, several commercial medicines including colistin (antibiotic), valrubicin (antitumor), icatibant (indicated for hereditary angioedema), bepotastine (prescribe for rhinitis), caspofungin (antifungal), and perphenazine (antipsychotic) also bind to the protease even with more tolerance to mutation than lopinavir/ritonavir, suggesting possible candidate drugs (Liu and Wang, 2020). Lopinavir and ritonavir were found to have a poor effect in treating COVID-19 pneumonia in addition to their toxic side effects (Wu et al., 2020).

Thirteen potential inhibitors of recombinant SARS-CoV-2 M<sup>Pro</sup> with IC<sub>50</sub> values ranging from 0.2–23  $\mu$ M were identified through biochemical high throughput screening. Among them, thimerosal, phenylmercuric acetate, and Evans blue demonstrated the highest inhibitory activity with IC<sub>50</sub> values below 1  $\mu$ M (Coelho et al., 2020). Su et al. (2020) reported natural products, baicalin and baicalein (non-covalent, non-peptidomimetic compounds), derived from Chinese traditional medicine as novel inhibitors of the M<sup>Pro</sup>. In Vero E6 cells, baicalin and baicalein showed potent antiviral activities with respective IC<sub>50</sub> values of  $6.41 \pm 0.95$  and  $0.94 \pm 0.20$   $\mu$ M, indicating a better performance of baicalein over baicalin.

## In silico Inhibitors

Computational modeling is an emerging area of research making drug discovery efforts more successful. However, it depends on various factors including protein-ligand geometry, chemical interactions, protonation, hydration, quantum effects, and several other constraints. These complex molecular dynamics computations are expensive and thus molecular docking tools are currently in practice to estimate the binding affinity and stability of protein-ligand interactions (Morris et al., 2009; Cofala et al., 2020). These tools are currently being extensively utilized to discover potential inhibitors that target SARS-CoV-2 M<sup>Pro</sup>.

Both new and known antiviral compounds have been studied for their effective binding to the active site of SARS-CoV-2 M<sup>Pro</sup>. Among these, a molecular docking study showed that HCV NS3/4A protease inhibitors (sofosbuvir, sofosbuvir, glecaprevir, boceprevir, simeprevir, paritaprevir, danoprevir, and grazoprevir) bind effectively on SARS-CoV-2 M<sup>Pro</sup> (Bafna et al., 2020) indicating their possible clinical significance. Another study showed that FDA-approved antiviral drugs lopinavir-ritonavir, tipranavir, and raltegravir showed strong, stable, and flexible binding on the active site of SARS-CoV-2 M<sup>Pro</sup> (Kumar et al., 2020c). Four antiviral molecules (Prulifloxacin, Nelfinavir, Tegobuvir, and Bictegravir) were also reported to bind on the active site of SARS-CoV-2 M<sup>Pro</sup> suggesting their ability to block viral protease and thus infection (Li et al., 2020a). Further, fragment-based drug designing identified 47 target compounds of which **#46** showed strong binding potential. Accordingly, the triazole ring binds to the S1 subsite, the covalent fragment of  $\alpha$ ,  $\beta$ -unsaturated aldehyde binds to S1' subsite,  $\beta$ -lactam ring binds to S2 subsite, and 5,7-dihydroxy chromone binds to S3 subsite of SARS-CoV-2 M<sup>Pro</sup>. The Triazole ring forms a H-bond with His<sup>163</sup>, the fragment of

$\alpha,\beta$ -unsaturated aldehyde forms a covalent bond with Cys<sup>145</sup>, aldehyde carbonyl forms a H-bond with His<sup>41</sup>, while the hydroxyl of chromone at position 7 forms a H-bond with Thr<sup>190</sup> (Tang et al., 2020).

Repurposing existing drugs facilitates the time needed to discover potent compounds for new diseases. Virtual screening of known drugs identified 15 potent inhibitors of SARS-CoV-2 M<sup>Pro</sup> where dipyridamole was the most potent inhibitor followed by candesartan cilexetil, hydroxychloroquine, and chloroquine with respective IC<sub>50</sub> values of  $0.60 \pm 0.01$ ,  $2.8 \pm 0.3$ ,  $2.9 \pm 0.3$ , and  $3.9 \pm 0.2$   $\mu\text{M}$  (Li et al., 2020b). Sixty-six FDA-approved drugs demonstrated higher binding scores in a pharmacophore-based drug activity analysis. Based on this, several classes of drugs such as viz. D2 receptor antagonist, HMG-CoA inhibitors, HIV reverse transcriptase and protease inhibitors, anticancer agents, and folate inhibitors presented potential interaction with SARS-CoV-2 M<sup>Pro</sup>. Among top-scoring compounds, imatinib showed a promising protease inhibition at 9.823  $\mu\text{M}$  (Balaramnavar et al., 2020). In another *in silico* study, four known drugs (remdesivir, simeprevir, nafamostat, and foretinib) were docked to bind on the catalytic dyad of the main protease. Also, drugs including bromocriptine (a dopamine antagonist), ergotamine (antimigraine), bictegavir (antiviral), antibacterial agents (oxytetracycline, tigecycline, ceftolozane), and immune modulators (vinflunine, vindesine, and topotecan) exhibited effective binding on the active site of the M<sup>Pro</sup> (Chakraborti et al., 2020) showing that the repurposing of several classes of known drugs is crucial to identifying the best drugs for COVID-19 treatment.

More *in silico* studies investigating the potential binding of lead compounds on the M<sup>Pro</sup> of SARS-CoV-2 are emerging. A study (Abel et al., 2020) identified 12 best hits from Super Natural II and Traditional Chinese Medicine databases. Selvaraj et al. (2020) also identified potential compounds from the Traditional Chinese Medicine database interacting with active site residues (His<sup>41</sup>, Gly<sup>143</sup>, and Cys<sup>145</sup>) of the M<sup>Pro</sup>. A ligand and virtual screening study (Ferraz et al., 2020) reported on three approved drugs (glibenclamide, bedaquiline, and miconazole) that effectively bind on the active site of SARS-CoV-2 M<sup>Pro</sup> with possible inhibitory activities. In an attempt to predict potential M<sup>Pro</sup> inhibitors from known antivirals, Kanhed et al., (2020) identified ritonavir, nelfinavir, and saquinavir to be potent M<sup>Pro</sup> inhibitors. Structurally, ritonavir formed hydrogen bonds with Gly<sup>143</sup> and Cys<sup>145</sup> with its (thiazoly-5-yl) methylcarbamate of oxygen while the thiazolyl ring forms polar contacts with Thr<sup>25</sup>, Thr<sup>26</sup>, and Leu<sup>27</sup> of the S1' subsite. Nelfinavir stabilized its binding with M<sup>Pro</sup> via hydrogen bonding with Glu<sup>166</sup>, and with His<sup>41</sup> and Tyr<sup>54</sup> in the S2 subsite. Depending on the structure of the pocket, compounds containing oxirane rings are suggested to be good M<sup>Pro</sup> inhibitors (Palese, 2020). Arbutin, terbutaline, barnidipine, tipiracil, and aprepitant were identified as potential hits forming different hydrophilic, hydrophobic, and electrostatic interactions with M<sup>Pro</sup> (Baby et al., 2020). Thioflavonol is a synthetic flavonoid analog that showed a strong binding with the conserved residues in the S1 subsite (Batool et al., 2020).

Fragment-based approaches to identify low molecular weight drugs are also other promising areas of investigation. Gao et al. (2020) repurposed low molecular weight drugs using a fragment-based approach for COVID-19 treatment where the authors identified low molecular weight drugs containing pharmacophores of niacin and hit 1, binding and inhibiting SARS-CoV-2 M<sup>Pro</sup>. A niacin derivative, carmofur, showed strong binding and M<sup>Pro</sup> inhibition with a IC<sub>50</sub> value of  $2.8 \pm 0.2$   $\mu\text{M}$ . Moreover, other low molecular weight analogs of hit 1 including triclabendazole, emedastine, omeprazole, and bendamustine were identified. Carmofur and bendamustine were reported to show potent inhibition whereas omeprazole was suggested for combinational use with another hit 1 analog. Choudhury, (2020) screened 191,678 fragments for their binding ability on the cavity of SARS-CoV-2 M<sup>Pro</sup>. The authors then generated new molecules tailored from those fragments which demonstrated strong binding on the adjacent sub-pockets. Finally, 17 molecules with binding abilities were found from which 15 molecules form a stable binding. Luan and Huynh, (2020) merged three-drug fragments (JFM, U0P, and HWH) into B19 which showed a slightly better free binding energy than the native peptide cleaved by the M<sup>Pro</sup>.

A 1,3-benzodioxolyl sulfonamide fragment from LASSBio-1945 was identified as a potential inhibitor of SARS-CoV-2 M<sup>Pro</sup> by applying a molecular docking and fragment-based pharmacophore model. The compound exhibited a strong binding energy interacting with residues His<sup>41</sup>, His<sup>163</sup>, and Glu<sup>166</sup> and potential inhibitory activity with IC<sub>50</sub> value of 15.97  $\mu\text{M}$ . Here, His<sup>41</sup> forms hydrophobic interactions and His<sup>163</sup> donates hydrogen bonds whereas Glu<sup>166</sup> serves as a hydrogen bond donor or acceptor. A 3-amino-pyridinyl moiety found in several fragments including 1, TRY-UNI-714a760b-6, and EDG-MED-0da5ad92-2 was also found to show effective inhibitory properties with respective IC<sub>50</sub> values of 24.57 and 53.72  $\mu\text{M}$  (Franco et al., 2020). An amino acid decomposition analysis together with a molecular dynamic simulation was also applied when looking for SARS-CoV-2 M<sup>Pro</sup> inhibitors where hydrogen bonds and hydrophobic interactions were found to hold the complex (Choudhury, 2020). For example, Glu<sup>166</sup> formed a permanent hydrogen bond with ZINC\_252512772 while hydrophobic interactions were observed with His<sup>41</sup> (Razzaghi-Asl et al., 2020). In addition, histone deacetylase inhibitors (Mamdouh et al., 2020), Cobalt (III) (Kozak et al., 2020), and Copper (II) (Garza-Lopez et al., 2020) are also reported to bind and inhibit SARS-CoV-2 M<sup>Pro</sup>. These studies (Garza-Lopez et al., 2020; Kozak et al., 2020) suggested that the positively charged metal ions binding on the negatively charged imidazole ring of stable histidine residues at positions 41, 163, 164, and 246, and the thiolate of cysteine residues at positions 44 and 145 could break up the bonds, resulting in inhibition of the M<sup>Pro</sup> activity.

Phytochemicals, extracted from medicinal plants, are now worth studying in the search for inhibitors of SARS-CoV-2 M<sup>Pro</sup>. Several phytochemicals including 5,7,30,40-Tetrahydroxy-2'-(3,3-dimethylallyl) isoflavone, Myricitrin, Methylrosmarinat, licoleafol, and amaranthin have been studied to bind and inhibit SARS-CoV-2 M<sup>Pro</sup>. Among these,

5,7,30,4'-tetrahydroxy-2' (3,3- dimethylallyl) is an isoflavone extracted from *Psoralea argyrea* which showed high binding affinity, forming strong hydrogen bonds with residues in the catalytic dyad (Cys<sup>145</sup> and His<sup>45</sup>). Moreover, this extract showed a significant interaction with receptor-binding residues of SARS-CoV-2 M<sup>Pro</sup> even more than the approved drugs (Nelfinavir, Prulifloxacin, and Colistin) (ul Qamar et al., 2020) but investigating the clinical applicability of these drugs is the next area of scrutiny. Another molecular docking study (Shree et al., 2020) reported that six phytochemicals (Withanoside V, Somniferine, Tinocordiside, Vicenin, Isorientin 40-O-glucoside 200-O-phydroxybenzoate, and Ursolic acid) exhibited strong binding with possible inhibition of SARS-CoV-2 M<sup>Pro</sup>. Chikhale et al. (2020) also reported that Asparoside-C, extracted from *Asparagus racemosus*, binds on the substrate-binding pocket of SARS-CoV-2 M<sup>Pro</sup> suggesting its possible inhibitory effect. Ten ligands from olive and four ligands from turmeric exhibited the best lowest binding energies with SARS-CoV-2 M<sup>Pro</sup> (Saif et al., 2020).

A study also showed that different phytochemicals effectively bind on SARS-CoV-2 M<sup>Pro</sup> suggesting their possible medicinal importance (Srivastav et al., 2020). Ursolic acid, carvacrol, and oleanolic acid showed stable and favorable energies resulting in strong binding of these phytochemicals on the active site of the enzyme (Kumar et al., 2020a). Chrysosplenetin is a phytochemical which showed strong binding affinity to the active site of SARS-CoV-2 M<sup>Pro</sup> interacting with residues Leu<sup>141</sup>, Gly<sup>143</sup>, Ser<sup>144</sup>, and Cys<sup>145</sup> (Ebada et al., 2020). Flavonoids are abundant in plants, fruit, and vegetables. Glycosylated flavonoids were suggested to be good inhibitors where Quercetin-3-O-rhamnoside showed the highest binding affinity. The sugar moiety of these compounds is found to be important for activity as the best compounds have sugar in their target structure (Cherrak et al., 2020). Cucurbitacins from foodstuffs strongly bind on different enzymes of SARS-CoV-2 with cucurbitacin G 2-glucoside and cucurbitacin H showing good drug-likeness properties (Kapoor et al., 2020). Seventeen potent M<sup>Pro</sup> inhibitors were identified from the Marine Natural Product (MNP) library using the pharmacophore model and molecular docking technique. The ligand-enzyme complex at the active site was stabilized by hydrogen bonds with Thr<sup>24</sup>, Ser<sup>46</sup>, Asn<sup>142</sup>, Glu<sup>166</sup>, and Pro<sup>168</sup> whereas  $\pi$ -hydrogen bonds and hydrophobic interactions connect the ligands with His<sup>41</sup>, Gly<sup>143</sup>, and Met<sup>49</sup>, Met<sup>65</sup>, Leu<sup>141</sup>, and Pro<sup>168</sup>, respectively (Gentile et al., 2020) suggesting their potential clinical use. As most studies on phytochemicals report only *in silico* screening results, the actual experimental inhibitory effects are not described requiring future investigation.

## DISCUSSION AND FUTURE PERSPECTIVES

This review discusses the structure-based design of inhibitors targeting SARS-CoV-2 M<sup>Pro</sup> and highlights the antiviral activity of potential candidate drugs of COVID-19. Especially in this time of urgent therapeutic need to treat COVID-19, pharmaceutical

repurposing and structure-based designing of drugs play significant roles in the fast discovery of potent drugs which in turn, apart from reaching the treatment needs of the community, also saves time and resources. Provided this, the structure-based design of drugs requires producing high-quality structures. The desired inhibitors should also have high binding specificity with the target (to minimize off-target binding), be competitive (increased affinity) and flexible (increased efficacy), easy for administration, and have an acceptable plasma half-life.

While the structure-based design of drugs is a robust approach, translation of the structural information into practice is another challenge. Moreover, X-ray structures present a static state of proteins which affects the design of effective drugs as the static structures may not be the most representative conformations of active enzymes. The designed drug might have other clinical shortcomings like high toxicity, teratogenicity, quick metabolism, inability to reach the target site, quick clearance, instability, is difficult to synthesize, and be costly to the general public (Verlinde and Hol, 1994; Craig and Eakin, 2000). This indicates that the successful design of a drug which specifically binds to the target does not mean success, rather, that the structure-based drug design needs to be done cautiously. Another challenge is designing *de novo* drugs using unliganded target proteins alone, nevertheless, computational approaches have significantly overcome this challenge. However, scoring is considered a serious problem since a large number of potential ligands are generated during molecular docking (Kuntz, 1992; Craig and Eakin, 2000). Another important issue in structure-based drug design is the optimization of the compounds based on the pharmacophore requirements of the M<sup>Pro</sup>. In a study, it was reported that only cinanserin showed best binding affinity and inhibitory activity after optimizing 220 compounds (Stoddard et al., 2020).

*In silico* studies are important for a better understanding of the M<sup>Pro</sup> structure and function which is a key factor when designing drugs. Moreover, computational drug design methods have an indispensable role in predicting the best drug, among others. Studies focusing on the *in silico* design of potent drugs targeting SARS-CoV-2 M<sup>Pro</sup> are increasing steadily; however, the clinical use of these desired drugs is questionable, corresponding to the possible limitations of passing clinical trials. Recently, a structural simulation study (Ahamad et al., 2020) screened three malaria-box compounds (MB-241, MB-250, and Mb-266) as the best lead drugs binding on SARS-CoV-2 M<sup>Pro</sup>; however, whether these compounds have experimental and/or clinical inhibitory activities is unknown. The development of drugs with broad-spectrum antiviral activity is considered a long purposed goal in drug discovery (Maurya et al., 2020). Therefore, using previously approved broad-spectrum drugs after appropriate improvements in design and potency could be an alternative solution during urgent times. In this regard, Zhang and colleagues (Zhang et al., 2020a) designed an improved peptidomimetic  $\alpha$ -ketoamide inhibitor (**13a**) from a previously designed broad-spectrum drug (**11r**) (Zhang et al., 2020b). While **13a** is pharmacokinetically improved, some inhibitory activities of **11r** were lost. Further, they modified **13a** into a more potent drug **13b** with compromised broad-spectrum activity, while

removal of the Boc group in **14b** inactivated the drug which provides a big lesson for curious design and/or improvement of drugs.

A study by Jin et al. (2020a) found promising cell-based inhibitory activity of screened drugs. The authors found that ebselen and N3 exhibit the strongest cell-based inhibitory activity against M<sup>Pro</sup>. Although ebselen, N3, carmofur, and PX-12 bind on the catalytic dyad of M<sup>Pro</sup>, carmofur, and PX-12 modified its structure completely. It is noted that the mechanism of covalent modification of M<sup>Pro</sup> by carmofur is different from N3 where N3 modifies Cys<sup>145</sup> by adding a vinyl group (Jin et al., 2020a). Further, unlike N3 which occupies all four subsites, carmofur is restricted only at the S2 subsite (Yang et al., 2005; Jin et al. 2020b) which also showed a promising lead drug to treat COVID-19 as it inhibits viral replication in cells. In-house designed drugs (**11a** and **11b**) bind on the M<sup>Pro</sup> and inhibit SARS-CoV-2 infectivity. Here the authors designed the inhibitors in such a way that the aldehyde group in the P1 serves as a new warhead to bind covalently with Cys<sup>145</sup> while the indole group was added in the P3 to form a hydrogen bond with Ser<sup>4</sup> to enhance its drug-likeness properties. At P2 position, **11a** has cyclohexyl while 3-fluorophenyl is in **11b** which makes **11a** pharmacokinetically better (Dai et al., 2020) (Figure 3). Aldehyde based drugs GC373 and GC376 were also pharmacokinetically effective to inhibit SARS-CoV-2 infection (Vuong et al., 2020). Baicalein also showed unique binding on S1 and S2 subsites of the catalytic dyad with promising *in vitro* inhibition at a IC<sub>50</sub> value below 1  $\mu$ M (Su et al., 2020).

GC376 was developed to treat Feline infectious peritonitis and showed potent antiviral activity against MERS-CoV, FIPV, and the norovirus (Kim et al., 2012; Kim et al., 2016; Pedersen et al., 2018). Boceprevir, GC376, and calpain inhibitors (II and XII) inhibit replication of the SARS-CoV-2 virus *via* targeting M<sup>Pro</sup> in the cell culture, with EC<sub>50</sub> values of 0.49–3.37  $\mu$ M at acceptable cell cytotoxicity. HCV NS3/4A serine protease inhibitors (boceprevir and narlaprevir) strongly bind and inhibit SARS-CoV-2 M<sup>Pro</sup>, with IC<sub>50</sub> values of 4.13 and 4.73  $\mu$ M, respectively. Among the screened drugs, GC376 was the most potent M<sup>Pro</sup> inhibitor (IC<sub>50</sub> = 0.03  $\mu$ M) while an anti-Rhinovirus drug (rupintrivir) failed to inhibit SARS-CoV-2 M<sup>Pro</sup>. Unlike HCV serine protease and SARS-CoV-2 cysteine protease M<sup>Pro</sup>, boceprevir and narlaprevir did not inhibit enterovirus A71 2A and 3C proteases and all four drugs (boceprevir, GC376, calpain inhibitor II and calpain inhibitor XII) did not inhibit the unrelated influenza virus H1N1 because of their specificity (Ma et al., 2020), suggesting that broad-spectrum viral protease inhibitors should be clinically investigated before using them. Additionally, boceprevir and GC376 were found to effectively bind and inhibit SARS-CoV-2 M<sup>Pro</sup> (Fu et al., 2020).

HIV protease inhibitors lopinavir and ritonavir are reported to bind and inhibit SARS-CoV (Nukoolkarn et al., 2008) and SARS-CoV-2 M<sup>Pro</sup> with promising antiviral activities (Liu and Wang, 2020). But a contradicting report by Ma et al. (2020) showed that the two drugs failed to inhibit SARS-CoV-2 M<sup>Pro</sup>. Similarly, these drugs demonstrated lack of efficacy in a clinical trial of severe COVID-19 adult patients (Cao et al., 2020) with unacceptable toxicity in treating COVID-19 related pneumonia (Wu et al.,

2020). This indicates that promising inhibitory activity of drugs either *in silico* or *in vivo* studies does not guarantee clinical efficacy corresponding to complex pharmacodynamics in the human body.

Remdesivir has been approved for COVID-19 treatment in the USA and many vaccine trials are at their last phase while the Pfizer vaccine has been approved for emergency use in some countries (at the time of writing this article). It has been noted that many small steps have been made in discovering clinically applicable drugs to treat COVID-19 (Erlanson, 2020). It is anticipated that vaccines and antibody-based drugs will be discovered before small molecules. However, vaccines might not be 100% effective and antibodies could have immunopathological consequences. Therefore, looking for putative drugs targeting SARS-CoV-2 M<sup>Pro</sup> is necessary. However, there are many challenges in designing drugs that target the proteases of coronaviruses due to poor pharmacokinetic properties of peptidomimetic/high molecular weight compounds and low inhibitory potential of non-peptidomimetic/low molecular weight compounds (Turk, 2006; Drag and Salvesen, 2010). Based on their inhibitory potency and selectivity, focusing on high molecular weight compounds over low molecular weight compounds has been advantageous; however, their drug-likeness property is questionable. On the other hand, 11 residues' long peptide (WWTWTPFHLLV), showed a strong binding affinity compared to  $\alpha$ -keto amide inhibitors with a suggested better inhibitory activity over small molecules (Rossetto and Zhou, 2020). Amin et al., 2020) analyzed the drug-likeness properties of recently reported SARS-CoV-2 M<sup>Pro</sup> inhibitors. The authors reported that only baicalein, disulfiram, carmofur, ebselen, tideglusib, shikonin, and PX-12 passed the drug-likeness evaluation.

A lot has been learned from previous structure-based drug design studies which could help prospective studies succeed fast in discovering effective antivirals for COVID-19 targeting the M<sup>Pro</sup>. Accordingly, atomistic-level mechanisms of peptide cleavage and pharmacophore requirements of the M<sup>Pro</sup>, stability of inhibitor-enzyme complex similar to the native peptide, plasticity of the active site of M<sup>Pro</sup>, the occurrence of mutations at the domains and/or the active site affecting the pocket, and the size and accommodation capacity of the subsites should be considered when designing new drugs or modifying previously known broad-spectrum drugs. This means that the optimization of the inhibitor-enzyme complex is ultimately important. As a significant number of studies solely report the binding affinity and energy of compounds towards the substrate-binding cleft of the M<sup>Pro</sup>, improvements considering the abovementioned points should be considered in the future. Future, structure-based drug design studies should comprehensively consider potency, selectivity, and drug-likeness properties of the candidate drugs in addition to optimizing their binding ability on the active site. Since low molecular weight compounds and non-peptidomimetic drugs have better drug-likeness properties over their counterparts (Amin et al., 2020), fragment-based drug design strategies should be considered to enhance the potency of these compounds. Drug repurposing is also crucial in urgent

times. In this regard, structure-based drug repurposing studies need to determine the dynamics of molecules targeting the M<sup>Pro</sup> (Durdagi et al., 2020). Such strategies may facilitate the efforts of discovering clinically applicable potent drugs for COVID-19.

Several drug targets are available to treat diseases caused by coronaviruses. The essential functions of SARS-CoV-2 M<sup>Pro</sup> in the viral life cycle with a conserved active site and structural suitability of its substrate-binding site for potent drugs, recognize it to be a promising drug target for treating COVID-19. The structural-functional reports so far presented strong pieces of evidence showing the binding specificity and inhibitory roles of compounds against the M<sup>Pro</sup> that could subsequently control SARS-CoV-2 infection. Taken together, their results provide a strong base to design further improved drugs with either limited or broad-spectrum activities with determined potency and pharmacokinetic profiles. Although significant efforts have been made in the search for potent drugs that inhibit SARS-CoV-2 M<sup>Pro</sup>, longitudinal studies on the therapeutic safety and efficacy of candidate drugs are still limited, ongoing, or not yet disseminated.

## CONCLUSION

The main protease of coronaviruses is relatively conserved (Hayden et al., 2003; Yang et al., 2005; Xue et al., 2007; Kim et al., 2016; Stoermer, 2020; Ullrich and Nitsche, 2020) and is what most drug repurposing studies are focusing on. However, mutations at the substrate-binding site and/or other sites due to viral evolution could potentially affect the structure of the M<sup>Pro</sup> substrate-binding pocket. For example, surface loops and helical domains III are variable across different M<sup>Pro</sup>s (Jin et al., 2020a) which affect the conformation of the active site (Xue et al., 2007). Moreover, a mutagenesis study depicted that some specific mutations cause major changes on the structure of the protein (Wolfe et al., 2020). Further, some plasticity is reported on the active site of SARS-CoV-2 M<sup>Pro</sup> compared to SARS-CoV M<sup>Pro</sup> (Palese, 2020) which may hinder the design of broad-spectrum drugs. Therefore, updated designs of potential inhibitors that can suitably bind with the active site of the enzyme are ultimately

necessary. Similarly, caution should be taken when modifying broad-spectrum inhibitors as the modification could affect the inhibitory activity of some drugs (for example, **13a** lost its broad-spectrum activity upon improved into **13b** (Zhang et al., 2020a)) as designing potent individual drugs for every virus strain is a resource, technical and also time-demanding. Considering the size of functional groups while designing improved drugs is also crucial as it affects the binding modes of drugs on the catalytic dyad of M<sup>Pro</sup> (Dai et al., 2020). Several potential classes of drugs are effective against SARS-CoV-2 M<sup>Pro</sup>. Among these,  $\alpha$ -ketoamide inhibitors, peptide-based inhibitors, anilid-based inhibitors, drugs from Chinese traditional medicine, phytochemicals, and indole lactam-based inhibitors are amongst the famous drug classes studied well. Although remdesivir is currently approved by the USA-FDA to treat COVID-19 patients, its clinical efficacy remains debatable. Therefore, improved, well-designed, potent, and structurally and pharmacokinetically effective drugs are urgently needed. Further investigations should focus on validating and finalizing effective drugs for COVID-19 beyond preliminary *in silico* and *in vivo* screening.

## AUTHOR CONTRIBUTIONS

HM: conceived the topic and wrote the original draft. TD: assisted in collecting articles and organizing the draft. TJ: conceived the topic, supervised the work and approved the final draft. All authors read and approved the final draft.

## FUNDING

This study is supported by the Strategic Priority Research Program of the Chinese Academy of Sciences (Grant No. XDB29030104), the National Natural Science Fund (Grant No: 31870731 and 31971129), the Fundamental Research Funds for the Central Universities, and the 100 Talents Program of the Chinese Academy of Sciences. HM is supported by the University of Science and Technology of China scholarship program.

## REFERENCES

- Abel, R., Chen, Q., Paredes Ramos, M., Pérez Sánchez, H., Coluzzi, F., Rocco, M., et al. (2020). Computational prediction of potential inhibitors of the main protease of SARS-CoV-2. *Front. Chem.* 8, 590263. doi:10.3389/fchem.2020.590263
- Ahamad, S., Kanipakam, H., Birla, S., Ali, M. S., and Gupta, D. (2020). Screening Malaria-box compounds to identify potential inhibitors against SARS-CoV-2 Mpro using molecular docking and dynamics simulations studies. *Eur. J. Pharmacol.* 890, 173664. doi:10.1016/j.ejphar.2020.173664
- Amin, S. A., Banerjee, S., Ghosh, K., Gayen, S., and Jha, T. (2020). Protease targeted COVID-19 drug discovery and its challenges: insight into viral main protease (Mpro) and papain-like protease (PLpro) inhibitors. *Bioorg. Med. Chem.* 20, 115860. doi:10.1016/j.bmc.2020.115860
- Anand, K., Palm, G. J., Mesters, J. R., Siddell, S. G., Ziebuhr, J., and Hilgenfeld, R. (2002). Structure of coronavirus main proteinase reveals combination of a chymotrypsin fold with an extra alpha-helical domain. *EMBO J.* 21 (13), 3213–3224. doi:10.1093/emboj/cdf327
- Anand, K., Ziebuhr, J., Wadhwani, P., Mesters, J. R., and Hilgenfeld, R. (2003). Coronavirus main proteinase (3CLpro) structure: basis for design of anti-SARS drugs. *Science* 300 (5626), 1763–1767. doi:10.1126/science.1085658
- Arafet, K., Ferrer, S., and Moliner, V. (2017). Computational study of the catalytic mechanism of the cruzain cysteine protease. *ACS Catal.* 7 (2), 1207–1215. doi:10.1021/acscatal.6b03096
- Baby, K., Maity, S., Mehta, C. H., Suresh, A., Nayak, U. Y., and Nayak, Y. (2020). Targeting SARS-CoV-2 main protease: a computational drug repurposing study. *Archives Medical Res.* 52 (1), 38–47. doi:10.1016/j.arcmed.2020.09.013
- Bafna, K., Krug, R. M., and Montelione, G. T. (2020). Structural similarity of SARS-CoV2 Mpro and HCV NS3/4A proteases suggests new approaches for identifying existing drugs useful as COVID-19 therapeutics. *ChemRxiv*. doi:10.26434/chemrxiv.12153615
- Balaramnavar, V. M., Ahmad, K., Saeed, M., Ahmad, I., Kamal, M., and Jawaid, T. (2020). Correction: pharmacophore-based approaches in the rational

- repurposing technique for FDA approved drugs targeting SARS-CoV-2 Mpro. *RSC Adv.* 10 (70), 40264–40275. doi:10.1039/D0RA90120B
- Batool, F., Mughal, E. U., Zia, K., Sadiq, A., Naeem, N., Javid, A., et al. (2020). Synthetic flavonoids as potential antiviral agents against SARS-CoV-2 main protease. *J. Biomol. Struct. Dyn.*, 1–12. doi:10.1080/07391102.2020.1850359
- Beigel, J. H., Tomashek, K. M., Dodd, L. E., Mehta, A. K., Zingman, B. S., Kalil, A. C., et al. (2020). Remdesivir for the treatment of Covid-19. *New Engl. J. Med.* 383 (19), 1813–1826. doi:10.1056/NEJMoa2007764
- Binford, S., Maldonado, F., Brothers, M., Weady, P., Zalman, L., Meador, J., et al. (2005). Conservation of amino acids in human rhinovirus 3C protease correlates with broad-spectrum antiviral activity of rupintrivir, a novel human rhinovirus 3C protease inhibitor. *Antimicrob. Agents Chemother.* 49 (2), 619–626. doi:10.1128/AAC.49.2.619-626.2005
- Cao, B., Wang, Y., Wen, D., Liu, W., Wang, J., Fan, G., et al. (2020). A trial of lopinavir–ritonavir in adults hospitalized with severe Covid-19. *New Engl. J. Med.* 382 (19), 1787–1799. doi:10.1056/NEJMoa2001282
- Casella, M., Rajnik, M., Cuomo, A., Dulebohn, S. C., and Di Napoli, R. (2020). *Features, evaluation and treatment coronavirus (COVID-19)* Statpearls. Florida: StatPearls Publishing.
- Chakraborti, S., Bheemireddy, S., and Srinivasan, N. (2020). Repurposing drugs against the main protease of SARS-CoV-2: mechanism-based insights supported by available laboratory and clinical data. *Mol. Omics* 16 (5), 474–491. doi:10.1039/D0MO00057D
- Chen, L., Gui, C., Luo, X., Yang, Q., Günther, S., Scandella, E., et al. (2005). Cinanserin is an inhibitor of the 3C-like proteinase of severe acute respiratory syndrome coronavirus and strongly reduces virus replication *in vitro*. *J. Virol.* 79 (11), 7095–7103. doi:10.1128/JVI.79.11.7095-7103.2005
- Chen, N., Zhou, M., Dong, X., Qu, J., Gong, F., Han, Y., et al. (2020a). Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. *The lancet* 395 (10223), 507–513. doi:10.1016/S0140-6736(20)30211-7
- Chen, Y., Liu, Q., and Guo, D. (2020b). Emerging coronaviruses: genome structure, replication, and pathogenesis. *J. Med. Virol.* 92 (4), 418–423. doi:10.1002/jmv.25681
- Cherrak, S. A., Merzouk, H., and Mokhtari-Soulmane, N. (2020). Potential bioactive glycosylated flavonoids as SARS-CoV-2 main protease inhibitors: a molecular docking and simulation studies. *PLoS One* 15 (10), e0240653. doi:10.1371/journal.pone.0240653
- Chikhale, R. V., Sinha, S. K., Patil, R. B., Prasad, S. K., Shakyia, A., Gurav, N., et al. (2020). In-silico investigation of phytochemicals from *Asparagus racemosus* as plausible antiviral agent in COVID-19. *J. Biomol. Struct. Dyn.*. doi:10.1080/07391102.2020.1784289
- Chou, C. Y., Chang, H. C., Hsu, W. C., Lin, T. Z., Lin, C. H., and Chang, G. G. (2004). Quaternary structure of the severe acute respiratory syndrome (SARS) coronavirus main protease. *Biochemistry* 43 (47), 14958–14970. doi:10.1021/bi0490237
- Choudhury, C. (2020). Fragment tailoring strategy to design novel chemical entities as potential binders of novel corona virus main protease. *J. Biomol. Struct. Dyn.* doi:10.1080/07391102.2020.1771424
- Chow, N., Fleming-Dutra, K., Gierke, R., and Hall, A. (2020). Preliminary estimates of the prevalence of selected underlying health conditions among patients with coronavirus disease 2019—United States, February 12–March 28, 2020. *Morbidity Mortality Weekly Rep.* 69 (13), 382. doi:10.15585/mmwr.mm6913e2
- Coelho, C., Gallo, G., Campos, C. B., Hardy, L., and Würtele, M. (2020). Biochemical screening for SARS-CoV-2 main protease inhibitors. *PLoS One* 15 (10), e0240079. doi:10.1371/journal.pone.0240079
- Cofala, T., Elend, L., Mirbach, P., Prellberg, J., Teusch, T., and Kramer, O. (2020). “Evolutionary multi-objective design of SARS-CoV-2 protease inhibitor candidates,” in International conference on parallel problem solving from nature. Cham: Springer. 357–371.
- Connelly, D. (2020). Targeting COVID-19: the drugs being fast-tracked through clinical trials and how they work. *Pharm. J.* 304 (7937), 312–313. doi:10.1211/PJ.2020.2020794
- Craig, S. P., 3rd, and Eakin, A. E. (2000). Structure-based inhibitor design. *Vitam Horm.* 58, 149–169. doi:10.1016/s0083-6729(00)58024-8
- Cui, J., Li, F., and Shi, Z. L. (2019). Origin and evolution of pathogenic coronaviruses. *Nat. Rev. Microbiol.* 17 (3), 181–192. doi:10.1038/s41579-018-0118-9
- Dömling, A., and Gao, L. (2020). Chemistry and biology of SARS-CoV-2. *Chem* 6 (6), 1283–1295. doi:10.1016/j.chempr.2020.04.023
- Dai, W., Zhang, B., Jiang, X. M., Su, H., Li, J., Zhao, Y., et al. (2020). Structure-based design of antiviral drug candidates targeting the SARS-CoV-2 main protease. *Science* 368 (6497), 1331–1335. doi:10.1126/science.abb4489
- de Leuw, P., and Stephan, C. (2017). Protease inhibitors for the treatment of hepatitis C virus infection. *GMS Infect Dis.* 5, Doc08. doi:10.3205/id000034
- Doungamath, A., Fearon, D., Gehrtz, P., Krojer, T., Lukacik, P., Owen, C. D., et al. (2020). Crystallographic and electrophilic fragment screening of the SARS-CoV-2 main protease. *Nat. Commun.* 11 (1), 5047. doi:10.1038/s41467-020-18709-w
- Drag, M., and Salvesen, G. S. (2010). Emerging principles in protease-based drug discovery. *Nat. Rev. Drug Discov.* 9 (9), 690–701. doi:10.1038/nrd3053
- Durdagi, S., Dag, C., Dogan, B., Yigin, M., Avsar, T., Buyukdag, C., et al. (2020). Near-physiological-temperature serial femtosecond X-ray crystallography reveals novel conformations of SARS-CoV-2 main protease active site for improved drug repurposing. *bioRxiv* doi:10.1101/2020.09.09.287987
- Ebada, S. S., Al-Jawabri, N. A., Youssef, F. S., El-Kashef, D. H., Knedel, T.-O., Albohy, A., et al. (2020). Anti-inflammatory, antiallergic and COVID-19 protease inhibitory activities of phytochemicals from the Jordanian hawkbeard: identification, structure–activity relationships, molecular modeling and impact on its folk medicinal uses. *RSC Adv.* 10 (62), 38128–38141. doi:10.1039/d0ra04876c
- Erlanson, D. A. (2020). Many small steps towards a COVID-19 drug. *Nat. Commun.* 11 (1), 5048–5054. doi:10.1038/s41467-020-18710-3
- Fan, K., Wei, P., Feng, Q., Chen, S., Huang, C., Ma, L., et al. (2004). Biosynthesis, purification, and substrate specificity of severe acute respiratory syndrome coronavirus 3C-like proteinase. *J. Biol. Chem.* 279 (3), 1637–1642. doi:10.1074/jbc.M310875200
- Fehr, A. R., and Perlman, S. (2015). *Coronaviruses: an overview of their replication and pathogenesis* Coronaviruses. New York: Springer, 1–23.
- Ferraz, W. R., Gomes, R. A., S Novaes, A. L., and Goulart Trossini, G. H. (2020). Ligand and structure-based virtual screening applied to the SARS-CoV-2 main protease: an in silico repurposing study. *Future Med. Chem.* 12 (20), 1815–1828. doi:10.4155/fmc-2020-0165
- Franco, L. S., Maia, R. C., and Barreiro, E. J. (2020). Identification of LASSBio-1945 as an inhibitor of SARS-CoV-2 main protease (M PRO) through in silico screening supported by molecular docking and a fragment-based pharmacophore model. *RSC Med. Chem.* 12 (1), 110–119. doi:10.1039/d0md00282h
- Fu, L., Ye, F., Feng, Y., Yu, F., Wang, Q., Wu, Y., et al. (2020). Both Boceprevir and GC376 efficaciously inhibit SARS-CoV-2 by targeting its main protease. *Nat. Commun.* 11 (1), 4417. doi:10.1038/s41467-020-18233-x
- Gao, J., Zhang, L., Liu, X., Li, F., Ma, R., Zhu, Z., et al. (2020). Repurposing low-molecular-weight drugs against the main protease of severe acute respiratory syndrome coronavirus 2. *J. Phys. Chem. Lett.* 11 (17), 7267–7272. doi:10.1021/acs.jpclett.0c01894
- Garza-Lopez, R. A., Kozak, J. J., and Gray, H. B. (2020). Copper(II) inhibition of the SARS-CoV-2 main protease. *ChemRxiv. Preprint.* doi:10.26434/chemrxiv.12673436
- Gates, B. (2020). Responding to Covid-19—a once-in-a-century pandemic? *New Engl. J. Med.* 382 (18), 1677–1679. doi:10.1056/NEJMp2003762
- Gattinoni, L., Coppola, S., Cressoni, M., Busana, M., Rossi, S., and Chiumello, D. (2020). COVID-19 does not lead to a “typical” acute respiratory distress syndrome. *Am. J. Respir. Crit. Care Med.* 201 (10), 1299–1300. doi:10.1164/rccm.202003-0817LE
- Gentile, D., Patamia, V., Scala, A., Sciortino, M. T., Piperno, A., and Rescifina, A. (2020). Putative inhibitors of SARS-CoV-2 main protease from a library of marine natural products: a virtual screening and molecular modeling study. *Mar. Drugs* 18 (4), 225. doi:10.3390/md18040225
- Gorbalenya, A., Baker, S., Baric, R., de Groot, R., Drosten, C., Gulyaeva, A., et al. (2020). The species severe acute respiratory syndrome related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. *Nat. Microbiol.* 5, 536–544. doi:10.1038/s41564-020-0695-z

- Gorbalenya, A. E., Koonin, E. V., Donchenko, A. P., and Blinov, V. M. (1989). Coronavirus genome: prediction of putative functional domains in the non-structural polyprotein by comparative amino acid sequence analysis. *Nucleic Acids Res.* 17 (12), 4847–4861. doi:10.1093/nar/17.12.4847
- Gordon, D., Jang, G., Bouhaddou, M., Xu, J., Obernier, K., White, K., et al. (2020). A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. *Nature* 583, 459–468. doi:10.1038/s41586-020-2286-9
- Griffin, J. W. D. (2020). SARS-CoV and SARS-CoV-2 main protease residue interaction networks change when bound to inhibitor N3. *J. Struct. Biol.* 211 (3), 107575. doi:10.1016/j.jsb.2020.107575
- Grottesi, A., Bešker, N., Emerson, A., Manelfi, C., Beccari, A. R., Frigerio, F., et al. (2020). Computational studies of SARS-CoV-2 3CLpro: insights from MD simulations. *Int. J. Mol. Sci.* 21 (15), 5346. doi:10.3390/ijms21155346
- Grum-Tokars, V., Ratia, K., Begaye, A., Baker, S. C., and Mesecar, A. D. (2008). Evaluating the 3C-like protease activity of SARS-Coronavirus: recommendations for standardized assays for drug discovery. *Virus Res.* 133 (1), 63–73. doi:10.1016/j.virusres.2007.02.015
- Hayden, F. G., Turner, R. B., Gwaltney, J. M., Chi-Burris, K., Gersten, M., Hsyu, P., et al. (2003). Phase II, randomized, double-blind, placebo-controlled studies of rupintrivir nasal spray 2-percent suspension for prevention and treatment of experimentally induced rhinovirus colds in healthy volunteers. *Antimicrob. Agents Chemother.* 47 (12), 3907–3916. doi:10.1128/aac.47.12.3907-3916.2003
- Hilgenfeld, R. (2014). From SARS to MERS: crystallographic studies on coronaviral proteases enable antiviral drug design. *FEBS J.* 281 (18), 4085–4096. doi:10.1111/febs.12936
- Holmes, K. V., and Lai, M. (1996). Coronaviridae: the viruses and their replication. *Fields Virol.* 1, 1075–1093.
- Hsu, M. F., Kuo, C. J., Chang, K. T., Chang, H. C., Chou, C. C., Ko, T. P., et al. (2005). Mechanism of the maturation process of SARS-CoV 3CL protease. *J. Biol. Chem.* 280 (35), 31257–31266. doi:10.1074/jbc.M502577200
- Huang, C., Wang, Y., Li, X., Ren, L., Zhao, J., Hu, Y., et al. (2020). Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *The Lancet* 395 (10223), 497–506. doi:10.1016/S0140-6736(20)30183-5
- Jain, R., and Mujwar, S. (2020). Repurposing metocurine as main protease inhibitor to develop novel antiviral therapy for COVID-19. *Struct. Chem.* 36 (1), 2487–2499. doi:10.1007/s11224-020-01605-w
- Jin, Z., Du, X., Xu, Y., Deng, Y., Liu, M., Zhao, Y., et al. (2020a). Structure of Mpro from SARS-CoV-2 and discovery of its inhibitors. *Nature* 582 (7811), 289–293. doi:10.1038/s41586-020-2223-y
- Jin, Z., Zhao, Y., Sun, Y., Zhang, B., Wang, H., Wu, Y., et al. (2020b). Structural basis for the inhibition of SARS-CoV-2 main protease by antineoplastic drug carmofur. *Nat. Struct. Mol. Biol.* 27 (6), 529–532. doi:10.1038/s41594-020-0440-6
- Kanheb, A. M., Patel, D. V., Teli, D. M., Patel, N. R., Chhabria, M. T., and Yadav, M. R. (2020). Identification of potential Mpro inhibitors for the treatment of COVID-19 by using systematic virtual screening approach. *Mol. Divers.* 25 (1), 383–401. doi:10.1007/s11030-020-10130-1
- Kapoor, N., Ghorai, S. M., Kushwaha, P. K., Shukla, R., Aggarwal, C., and Bandichhor, R. (2020). Plausible mechanisms explaining the role of cucurbitacins as potential therapeutic drugs against coronavirus 2019. *Informatics in medicine unlocked* 21, 100484. doi:10.1016/j.imu.2020.100484
- Keillor, J., and Brown, R. (1992). Attack of zwitterionic ammonium thiols on a distorted anilide as a model for the acylation of papain by amides. A simple demonstration of a bell-shaped pH/rate profile. *J. Am. Chem. Soc.* 114 (21), 7983–7989. doi:10.1021/ja00047a004
- Kil, J., Lobarinas, E., Spankovich, C., Griffiths, S. K., Antonelli, P. J., Lynch, E. D., et al. (2017). Safety and efficacy of ebselen for the prevention of noise-induced hearing loss: a randomised, double-blind, placebo-controlled, phase 2 trial. *Lancet* 390 (10098), 969–979. doi:10.1016/S0140-6736(17)31791-9
- Kim, Y., Liu, H., Galasiti Kankanamalage, A. C., Weerasekara, S., Hua, D. H., Groutas, W. C., et al. (2016). Reversal of the progression of fatal coronavirus infection in cats by a broad-spectrum coronavirus protease inhibitor. *PLoS Pathog.* 12 (3), e1005531. doi:10.1371/journal.ppat.1005531
- Kim, Y., Lovell, S., Tiew, K. C., Mandadapu, S. R., Alliston, K. R., Battaile, K. P., et al. (2012). Broad-spectrum antivirals against 3C or 3C-like proteases of picornaviruses, noroviruses, and coronaviruses. *J. Virol.* 86 (21), 11754–11762. doi:10.1128/JVI.01348-12
- Kim, Y., Shivanna, V., Narayanan, S., Prior, A. M., Weerasekara, S., Hua, D. H., et al. (2015). Broad-spectrum inhibitors against 3C-like proteases of feline coronaviruses and feline caliciviruses. *J. Virol.* 89 (9), 4942–4950. doi:10.1128/JVI.03688-14
- Kozak, J. J., Gray, H. B., and Garza-López, R. A. (2020). Structural stability of the SARS-CoV-2 main protease: can metal ions affect function? *J. Inorg. Biochem.* 211, 111179. doi:10.1016/j.jinorgbio.2020.111179
- Kumar, A., Choudhri, G., Shukla, S. K., Sharma, M., Tyagi, P., Bhushan, A., et al. (2020a). Identification of phytochemical inhibitors against main protease of COVID-19 using molecular modeling approaches. *J. Biomol. Struct. Dyn.* doi:10.1080/07391102.2020.1772112
- Kumar, N., Awasthi, A., Kumari, A., Sood, D., Jain, P., Singh, T., et al. (2020b). Antitussive nospicine and antiviral drug conjugates as arsenal against COVID-19: a comprehensive chemoinformatics analysis. *J. Biomol. Struct. Dyn.* 1–16. doi:10.1080/07391102.2020.1808072
- Kumar, Y., Singh, H., and Patel, C. N. (2020c). In silico prediction of potential inhibitors for the Main protease of SARS-CoV-2 using molecular docking and dynamics simulation based drug-repurposing. *J. Infect Public Health* 13 (9), 1210–1223. doi:10.1016/j.jiph.2020.06.016
- Kuntz, I. D. (1992). Structure-based strategies for drug design and discovery. *Science* 257 (5073), 1078–1082. doi:10.1126/science.257.5073.1078
- Kuo, C. J., Liu, H. G., Lo, Y. K., Seong, C. M., Lee, K. I., Jung, Y. S., et al. (2009). Individual and common inhibitors of coronavirus and picornavirus main proteases. *FEBS Lett.* 583 (3), 549–555. doi:10.1016/j.febslet.2008.12.059
- Lee, H. J., Shieh, C. K., Gorbalenya, A. E., Koonin, E. V., La Monica, N., Tuler, J., et al. (1991). The complete sequence (22 kilobases) of murine coronavirus gene 1 encoding the putative proteases and RNA polymerase. *Virology* 180 (2), 567–582. doi:10.1016/0042-6822(91)90071-i
- Li, Y., Zhang, J., Wang, N., Li, H., Shi, Y., Guo, G., et al. (2020a). Therapeutic drugs targeting 2019-nCoV main protease by high-throughput screening. *BioRxiv*. doi:10.1101/2020.01.28.922922
- Li, Z., Li, X., Huang, Y.-Y., Wu, Y., Liu, R., Zhou, L., et al. (2020b). Identify potent SARS-CoV-2 main protease inhibitors via accelerated free energy perturbation-based virtual screening of existing drugs. *Proc. Natl. Acad. Sci.* 117 (44), 27381–27387. doi:10.1073/pnas.2010470117
- Lim, L., Shi, J., Mu, Y., and Song, J. (2014). Dynamically-driven enhancement of the catalytic machinery of the SARS 3C-like protease by the S284-T285-I286/A mutations on the extra domain. *PLoS One*, 9 (7), e101941. doi:10.1371/journal.pone.0101941
- Liu, X., and Wang, X. J. (2020). Potential inhibitors against 2019-nCoV coronavirus M protease from clinically approved medicines. *J. Genet. Genomics* 47 (2), 119–121. doi:10.1016/j.jgg.2020.02.001
- Lu, R., Zhao, X., Li, J., Niu, P., Yang, B., Wu, H., et al. (2020). Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *The Lancet* 395 (10224), 565–574. doi:10.1016/S0140-6736(20)30251-8
- Luan, B., and Huynh, T. (2020). Crystal-structures-guided design of fragment-based drugs for inhibiting the main protease of SARS-CoV-2. *Research square*. doi:10.21203/rs.3.rs-80869/v1
- Lv, Z., Chu, Y., and Wang, Y. (2015). HIV protease inhibitors: a review of molecular selectivity and toxicity. *HIV AIDS (Auckl)* 7, 95–104. doi:10.2147/HIV.S79956
- Lynch, E., and Kil, J. (2009). Development of ebselen, a glutathione peroxidase mimic, for the prevention and treatment of noise-induced hearing loss. *Seminars in Hearing* 30, 47–55. doi:10.1055/s-0028-1111106
- Ma, C., Sacco, M. D., Hurst, B., Townsend, J. A., Hu, Y., Szeto, T., et al. (2020). Boceprevir, GC-376, and calpain inhibitors II, XII inhibit SARS-CoV-2 viral replication by targeting the viral main protease. *Cell Res.* 30 (8), 678–692. doi:10.1038/s41422-020-0356-z
- Macchiagodena, M., Pagliai, M., and Procacci, P. (2020). Identification of potential binders of the main protease 3CLpro of the COVID-19 via structure-based ligand design and molecular modeling. *Chem. Phys. Lett.* 750, 137489. doi:10.1016/j.cplett.2020.137489
- Mamdouh, F. A., Mohamed, F. A. M., El-Din, G., and Abu-Rahma, G. E. D. A. (2020). Molecular docking study reveals the potential repurposing OF histone deacetylase inhibitors against COVID-19. *IJPSR* 11 (9), 4261–4270. doi:10.13040/IJPSR.0975-8232.11(9).4261-70
- Masaki, C., Sharpley, A. L., Cooper, C. M., Godlewska, B. R., Singh, N., Vasudevan, S. R., et al. (2016). Effects of the potential lithium-mimetic, ebselen, on

- impulsivity and emotional processing. *Psychopharmacology (Berl)* 233 (14), 2655–2661. doi:10.1007/s00213-016-4319-5
- Maurya, V. K., Kumar, S., Prasad, A. K., Bhatt, M. L. B., and Saxena, S. K. (2020). Structure-based drug designing for potential antiviral activity of selected natural products from Ayurveda against SARS-CoV-2 spike glycoprotein and its cellular receptor. *Virus Disease* 31 (2), 179–193. doi:10.1007/s13337-020-00598-8
- Mengist, H. M., Fan, X., and Jin, T. (2020). Designing of improved drugs for COVID-19: crystal structure of SARS-CoV-2 main protease Mpro. *Signal Transduct Target Ther.* 5 (1), 67–72. doi:10.1038/s41392-020-0178-y
- Morimoto, K., and Koh, M. (2003). Postoperative adjuvant use of carmofur for early breast cancer. *Osaka City Med. J.* 49 (2), 77–83.
- Morris, G. M., Huey, R., Lindstrom, W., Sanner, M. F., Belew, R. K., Goodsell, D. S., et al. (2009). AutoDock4 and AutoDockTools4: automated docking with selective receptor flexibility. *J. Comput. Chem.* 30 (16), 2785–2791. doi:10.1002/jcc.21256
- Naqvi, A. A. T., Fatima, K., Mohammad, T., Fatima, U., Singh, I. K., Singh, A., et al. (2020). Insights into SARS-CoV-2 genome, structure, evolution, pathogenesis and therapies: structural genomics approach. *Biochim. Biophys. Acta (BBA)-Molecular Basis Dis.* 1866 (10), 165878. doi:10.1016/j.bbadis.2020.165878
- Nguyen, D. D., Gao, K., Chen, J., Wang, R., and Wei, G.-W. (2020). Unveiling the molecular mechanism of SARS-CoV-2 main protease inhibition from 137 crystal structures using algebraic topology and deep learning. *Chem. Sci.* 11 (44), 12036–12046. doi:10.1039/D0SC04641H
- Nishio, S., Kishimoto, T., Maekawa, M., Kawakita, J., Morikawa, Y., Funai, K., et al. (1987). [Study on effectiveness of carmofur (Mifuro) in urogenital carcinoma, especially bladder cancer, as a post-operative adjuvant chemotherapeutic agent]. *Hinyokika Kyo* 33 (2), 295–303.
- Nukoolkarn, V., Lee, V. S., Malaisree, M., Aruksakulwong, O., and Hannongbua, S. (2008). Molecular dynamic simulations analysis of ritonavir and lopinavir as SARS-CoV 3CL(pro) inhibitors. *J. Theor. Biol.* 254 (4), 861–867. doi:10.1016/j.jtbi.2008.07.030
- Ortega, J. T., Serrano, M. L., and Jastrzebska, B. (2020). Class AG protein-coupled receptor antagonist famotidine as a therapeutic alternative against SARS-CoV2: an in silico analysis. *Biomolecules* 10 (6), 954. doi:10.3390/biom10060954
- Palese, L. L. (2020). The structural landscape of SARS-CoV-2 main protease: hints for inhibitor search. *ChemRxiv*. doi:10.26434/chemrxiv.12209744.v1
- Patel, K. P., Vunnam, S. R., Patel, P. A., Krill, K. L., Korbitz, P. M., Gallagher, J. P., et al. (2020). Transmission of SARS-CoV-2: an update of current literature. *Eur. J. Clin. Microbiol. Infect. Dis.* 39 (1), 1–7. doi:10.1007/s10096-020-03961-1
- Paul, S. (2006). The molecular biology of coronavirus. *Adv. Virus Res.* 66 (48), 193–292. doi:10.1016/S0065-3527(06)66005-3
- Pedersen, N. C., Kim, Y., Liu, H., Galasiti Kankanamalage, A. C., Eckstrand, C., Groutas, W. C., et al. (2018). Efficacy of a 3C-like protease inhibitor in treating various forms of acquired feline infectious peritonitis. *J. Feline Med. Surg.* 20 (4), 378–392. doi:10.1177/1098612X17729626
- Pillaiyar, T., Manickam, M., Namasivayam, V., Hayashi, Y., and Jung, S. H. (2016). An overview of severe acute respiratory syndrome-coronavirus (SARS-CoV) 3CL protease inhibitors: peptidomimetics and small molecule Chemotherapy. *J. Med. Chem.* 59 (14), 6595–6628. doi:10.1021/acs.jmedchem.5b01461
- Ramos-Guzmán, C. A., Ruiz-Pernía, J. J., and Tuñón, I. (2020). Unraveling the SARS-CoV-2 main protease mechanism using multiscale methods. *ACS Catal.* 10 (21), 12544–12554. doi:10.1021/acscatal.0c03420
- Rathnayake, A. D., Zheng, J., Kim, Y., Perera, K. D., Mackin, S., Meyerholz, D. K., et al. (2020). 3C-like protease inhibitors block coronavirus replication *in vitro* and improve survival in MERS-CoV-infected mice. *Sci. Transl. Med.* 12 (557), eabc5332. doi:10.1126/scitranslmed.abc5332
- Razzaghi-Asl, N., Ebadi, A., Shahabipour, S., and Gholamin, D. (2020). Identification of a potential SARS-CoV2 inhibitor via molecular dynamics simulations and amino acid decomposition analysis. *J. Biomol. Struct. Dyn.* 1–16. doi:10.1080/07391102.2020.1797536
- Renson, M., Etschenberg, E., and Winkelman, J. (1982). *2-Phenyl-1, 2-benzisoxazolin-3 (2H)-one containing pharmaceutical preparations and process for the treatment of rheumatic diseases: Google Patents*. U.S. Patent No. 4,352,799. Washington, DC: U.S. Patent and Trademark Office.
- Rossetto, A., and Zhou, W. (2020). Novel generated peptides for COVID-19 targets. Paper presented at the Proceedings of the 11th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics; Virtual Event, USA: Association for Computing Machinery, 86.
- Rut, W., Grobocz, K., Zhang, L., Sun, X., Zmudzinski, M., Hilgenfeld, R., et al. (2020a). Substrate specificity profiling of SARS-CoV-2 Mpro protease provides basis for anti-COVID-19 drug design. *Biorxiv*. doi:10.1101/2020.03.07.981928
- Rut, W., Grobocz, K., Zhang, L., Sun, X., Zmudzinski, M., Pawlik, B., et al. (2021). SARS-CoV-2 Mpro inhibitors and activity-based probes for patient-sample imaging. *Nat. Chem. Biol.* 17, 222–228. doi:10.1038/s41589-020-00689-z
- Saha, A., Sharma, A. R., Bhattacharya, M., Sharma, G., Lee, S.-S., and Chakraborty, C. (2020). Probable molecular mechanism of remdesivir for the treatment of COVID-19: need to know more. *Archives Med. Res.* 51 (6), 585–586. doi:10.1016/j.arcmed.2020.05.001
- Saif, R., Raza, M. H., Rehman, T., Zafar, M. O., Zia, S., Qureshi, A. R., et al. (2020). Molecular docking of Olea europaea and Curcuma Longa compounds as potential drug agents for targeting main-protease of SARS-nCoV2. *ChemRxiv*. doi:10.26434/chemrxiv.13246739.v1
- Sakamoto, J., Hamada, C., Rahman, M., Kodaira, S., Ito, K., Nakazato, H., et al. (2005). An individual patient data meta-analysis of adjuvant therapy with carmofur in patients with Curatively resected Colon cancer. *Jpn. J. Clin. Oncol.* 35 (9), 536–544. doi:10.1093/jjco/hy147
- Selvaraj, C., Panwar, U., Dinesh, D. C., Boura, E., Singh, P., Dubey, V. K., et al. (2020). Microsecond MD simulation and multiple-Confirmation virtual screening to identify potential anti-COVID-19 inhibitors against SARS-CoV-2 main protease. *Front. Chem.* 8, 595273. doi:10.3389/fchem.2020.595273
- Shaffer, L. (2020). 15 drugs being tested to treat COVID-19 and how they would work. *Nat. Med. News Feature*. 1–1. doi:10.1038/d41591-020-00019-9
- Sharma, M., Prasher, P., Mehta, M., Zaccaroni, F. C., Singh, Y., Kapoor, D. N., et al. (2020). Probing 3CL protease: rationally designed chemical moieties for COVID-19. *Drug devel. Res.* 81, 21724. doi:10.1002/ddr.21724
- Shi, J., Wei, Z., and Song, J. (2004). Dissection study on the severe acute respiratory syndrome 3C-like protease reveals the critical role of the extra domain in dimerization of the enzyme: defining the extra domain as a new target for design of highly specific protease inhibitors. *J. Biol. Chem.* 279 (23), 24765–24773. doi:10.1074/jbc.M311744200
- Shie, J. J., Fang, J. M., Kuo, T. H., Kuo, C. J., Liang, P. H., Huang, H. J., et al. (2005). Inhibition of the severe acute respiratory syndrome 3CL protease by peptidomimetic alpha,beta-unsaturated esters. *Bioorg. Med. Chem.*, 13(17), 5240–5252. doi:10.1016/j.bmc.2005.05.065
- Shree, P., Mishra, P., Selvaraj, C., Singh, S. K., Chaube, R., Garg, N., et al. (2020). Targeting COVID-19 (SARS-CoV-2) main protease through active phytochemicals of ayurvedic medicinal plants—Withania somnifera (Ashwagandha), Tinospora cordifolia (Giloy) and Ocimum sanctum (Tulsi)—a molecular docking study. *J. Biomol. Struct. Dyn.* 1–14. doi:10.1080/07391102.2020.1810778
- Shu, T., Huang, M., Wu, D., Ren, Y., Zhang, X., Han, Y., et al. (2020). SARS-Coronavirus-2 Nsp13 possesses NTPase and RNA helicase activities that can be inhibited by bismuth salts. *Viral. Sin.* 35 (3), 321–329. doi:10.1007/s12250-020-00242-1
- Sies, H., and Parnham, M. J. (2020). Potential therapeutic use of ebselen for COVID-19 and other respiratory viral infections. *Free Radic. Biol. Med.* 156, 107–112. doi:10.1016/j.freeradbiomed.2020.06.032
- Singh, N., Halliday, A. C., Thomas, J. M., Kuznetsova, O. V., Baldwin, R., Woon, E. C., et al. (2013). A safe lithium mimetic for bipolar disorder. *Nat. Commun.* 4 (1), 1332–1337. doi:10.1038/ncomms2320
- Srivastav, A. K., Gupta, S. K., and Kumar, U. (2020). Computational studies towards identification of lead herbal compounds of medicinal importance for development of nutraceutical against COVID-19. *ChemRxiv*. doi:10.26434/chemrxiv.12581819.v1
- Stoddard, S. V., Stoddard, S. D., Oelkers, B. K., Fitts, K., Whalum, K., Whalum, K., et al. (2020). Optimization rules for SARS-CoV-2 Mpro antivirals: ensemble docking and exploration of the coronavirus protease active site. *Viruses* 12 (9), 942. doi:10.3390/v12090942
- Stoermer, M. (2020). Homology models of coronavirus 2019-nCoV 3CLpro protease. *ChemRxiv*. doi:10.26434/chemrxiv.11637294.v3
- Su, H., Yao, S., Zhao, W., Li, M., Liu, J., Shang, W., et al. (2020). Discovery of baicalin and baicalein as novel, natural product inhibitors of SARS-CoV-2 3CL protease *in vitro*. *bioRxiv*. doi:10.1101/2020.04.13.038687
- Swiderek, K., and Moliner, V. (2020). Revealing the molecular mechanisms of proteolysis of SARS-CoV-2 Mpro from QM/MM computational

- methods. *Chem. Sci.* 11, 10626–10630. doi:10.26434/chemrxiv.12283967
- Tang, B., He, F., Liu, D., Fang, M., Wu, Z., and Xu, D. (2020). AI-aided design of novel targeted covalent inhibitors against SARS-CoV-2. *bioRxiv*. doi:10.1101/2020.03.03.972133
- Thiel, V., Ivanov, K. A., Putics, A., Hertzog, T., Schelle, B., Bayer, S., et al. (2003). Mechanisms and enzymes involved in SARS coronavirus genome expression. *J. Gen. Virol.* 84 (9), 2305–2315. doi:10.1099/vir.0.19424-0
- Turk, B. (2006). Targeting proteases: successes, failures and future prospects. *Nat. Rev. Drug Discov.* 5 (9), 785–799. doi:10.1038/nrd2092
- ul Qamar, T. M., Alqahtani, S. M., Alamri, M. A., and Chen, L. L. (2020). Structural basis of SARS-CoV-2 3CLpro and anti-COVID-19 drug discovery from medicinal plants. *J. Pharm. Anal.* 10 (4), 313–319. doi:10.1016/j.jpba.2020.03.009
- Ullrich, S., and Nitsche, C. (2020). The SARS-CoV-2 main protease as drug target. *Bioorg. Med. Chem. Lett.* 30, 127377. doi:10.1016/j.bmcl.2020.127377
- Verlinde, C. L., and Hol, W. G. (1994). Structure-based drug design: progress, results and challenges. *Structure* 2 (7), 577–587. doi:10.1016/S0969-2126(00)00060-5
- Vuong, W., Khan, M. B., Fischer, C., Arutyunova, E., Lamer, T., Shields, J., et al. (2020). Feline coronavirus drug inhibits the main protease of SARS-CoV-2 and blocks virus replication. *Nat. Commun.* 11, 4282. doi:10.1038/s41467-020-18096-2
- Wolfe, G., Belhoussine, O., Dawson, A., Lisaius, M., and Jagodzinski, F. (2020). Impactful mutations in Mpro of the SARS-CoV-2 proteome. Paper presented at the Proceedings of the 11th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics, Virtual Event, USA: Association for Computing Machinery, 1–3, 108.
- Wu, C., Liu, Y., Yang, Y., Zhang, P., Zhong, W., Wang, Y., et al. (2020). Analysis of therapeutic targets for SARS-CoV-2 and discovery of potential drugs by computational methods. *Acta Pharmaceutica Sinica B* 10, 766–788. doi:10.1016/j.japsb.2020.02.008
- Xu, Z., Peng, C., Shi, Y., Zhu, Z., Mu, K., Wang, X., et al. (2020). Nelfinavir was predicted to be a potential inhibitor of 2019-nCoV main protease by an integrative approach combining homology modelling, molecular docking and binding free energy calculation. *bioRxiv*. doi:10.1101/2020.01.27.921627
- Xue, X., Yang, H., Shen, W., Zhao, Q., Li, J., Yang, K., et al. (2007). Production of authentic SARS-CoV M(pro) with enhanced activity: application as a novel tag-cleavage endopeptidase for protein overproduction. *J. Mol. Biol.* 366 (3), 965–975. doi:10.1016/j.jmb.2006.11.073
- Xue, X., Yu, H., Yang, H., Xue, F., Wu, Z., Shen, W., et al. (2008). Structures of two coronavirus main proteases: implications for substrate binding and antiviral drug design. *J. Virol.* 82 (5), 2515–2527. doi:10.1128/JVI.02114-07
- Yang, H., Xie, W., Xue, X., Yang, K., Ma, J., Liang, W., et al. (2005). Design of wide-spectrum inhibitors targeting coronavirus main proteases. *PLoS Biol.* 3 (10), e324. doi:10.1371/journal.pbio.0030324
- Yang, H., Yang, M., Ding, Y., Liu, Y., Lou, Z., Zhou, Z., et al. (2003). The crystal structures of severe acute respiratory syndrome virus main protease and its complex with an inhibitor. *Proc. Natl. Acad. Sci. USA* 100 (23), 13190–13195. doi:10.1073/pnas.1835675100
- Zhang, L., Lin, D., Sun, X., Curth, U., Drosten, C., Sauerhering, L., et al. (2020a). Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved  $\alpha$ -ketoamide inhibitors. *Science* 368 (6489), 409–412. doi:10.1126/science.abb3405
- Zhang, L., Lin, D., Kusov, Y., Nian, Y., Ma, Q., Wang, J., et al. (2020b).  $\alpha$ -Ketoamides as broad-spectrum inhibitors of coronavirus and enterovirus replication: structure-based design, synthesis, and activity assessment. *J. Med. Chem.* 63, 4562–4578. doi:10.1021/acs.jmedchem.9b01828
- Zhang, W., Du, R.-H., Li, B., Zheng, X.-S., Yang, X.-L., Hu, B., et al. (2020c). Molecular and serological investigation of 2019-nCoV infected patients: implication of multiple shedding routes. *Emerging microbes & infections* 9 (1), 386–389. doi:10.1080/22221751.2020.1729071
- Zhu, G., Zhu, C., Zhu, Y., and Sun, F. (2020). Minireview of progress in the structural study of SARS-CoV-2 proteins. *Curr. Res. Microb. Sci.* 1, 53–61. doi:10.1016/j.crmicr.2020.06.003
- Ziebuhr, J., Snijder, E. J., and Gorbalenya, A. E. (2000). Virus-encoded proteinases and proteolytic processing in the Nidovirales. *J. Gen. Virol.* 81 (4), 853–879. doi:10.1099/0022-1317-81-4-853
- Zumla, A., Chan, J. F., Azhar, E. I., Hui, D. S., and Yuen, K. Y. (2016). Coronaviruses - drug discovery and therapeutic options. *Nat. Rev. Drug Discov.* 15 (5), 327–347. doi:10.1038/nrd.2015.37

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Mengist, Dilnessa and Jin. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Pomegranate Peel Extract as an Inhibitor of SARS-CoV-2 Spike Binding to Human ACE2 Receptor (*in vitro*): A Promising Source of Novel Antiviral Drugs

Annalisa Tito<sup>1\*</sup>, Antonio Colantuono<sup>1</sup>, Luciano Pirone<sup>2</sup>, Emilia Pedone<sup>2</sup>, Daniela Intartaglia<sup>3</sup>, Giuliana Giamundo<sup>4</sup>, Ivan Conte<sup>3,5</sup>, Paola Vitaglione<sup>6</sup> and Fabio Apone<sup>1,7</sup>

<sup>1</sup> Arterra Bioscience SPA, Naples, Italy, <sup>2</sup> Institute of Biostructures and Bioimaging, National Research Council, Naples, Italy, <sup>3</sup> Telethon Institute of Genetics and Medicine, Pozzuoli, Italy, <sup>4</sup> Department of Neuroscience, Reproductive and Odontostomatological Sciences, University of Naples Federico II, Naples, Italy, <sup>5</sup> Department of Biology, University of Naples Federico II, Naples, Italy, <sup>6</sup> Department of Agricultural Science, University of Naples Federico II, Portici, Italy, <sup>7</sup> Vitalab Srl, Naples, Italy

## OPEN ACCESS

### Edited by:

Maria Luisa Mangoni,  
Sapienza University of Rome, Italy

### Reviewed by:

Alfonso Carotenuto,  
University of Naples Federico II, Italy  
Surajit Bhattacharjya,  
Nanyang Technological  
University, Singapore

### \*Correspondence:

Annalisa Tito  
annalisa@arterrabio.it

### Specialty section:

This article was submitted to  
Chemical Biology,  
a section of the journal  
Frontiers in Chemistry

Received: 05 December 2020

Accepted: 25 March 2021

Published: 28 April 2021

### Citation:

Tito A, Colantuono A, Pirone L,  
Pedone E, Intartaglia D, Giamundo G,  
Conte I, Vitaglione P and Apone F  
(2021) Pomegranate Peel Extract as  
an Inhibitor of SARS-CoV-2 Spike  
Binding to Human ACE2 Receptor (*in  
vitro*): A Promising Source of Novel  
Antiviral Drugs.  
Front. Chem. 9:638187.  
doi: 10.3389/fchem.2021.638187

Plant extracts are rich in bioactive compounds, such as polyphenols, sesquiterpenes, and triterpenes, which potentially have antiviral activities. As a consequence of the coronavirus disease 2019 pandemic, caused by the severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2) virus, thousands of scientists have been working tirelessly trying to understand the biology of this new virus and the disease pathophysiology, with the main goal of discovering effective preventive treatments and therapeutic agents. Plant-derived secondary metabolites may play key roles in preventing and counteracting the rapid spread of SARS-CoV-2 infections by inhibiting the activity of several viral proteins, in particular those involved in the virus entry into the host cells and its replication. Using *in vitro* approaches, we investigated the role of a pomegranate peel extract (PPE) in attenuating the interaction between the SARS-CoV-2 Spike glycoprotein and the human angiotensin-converting enzyme 2 receptor, and on the activity of the virus 3CL protease. Although further studies will be determinant to assess the efficacy of this extract *in vivo*, our results opened new promising opportunities to employ natural extracts for the development of effective and innovative therapies in the fight against SARS-CoV-2.

**Keywords:** pomegranate peels, SARS-CoV-2, ACE2, COVID-19, polyphenols, pomegranate (*Punica granatum* L.) peel extracts, polyphenols

## INTRODUCTION

Plants synthesize a large variety of secondary metabolites having a wide range of biological activities and vital roles for their development and survival (Isah, 2019). Most of those metabolites serve as the plant's defense chemicals against both biotic stresses (e.g., herbivore insects, parasitic nematodes, and microbial pathogens) and abiotic stress (e.g., low or high temperatures, deficient or excessive water, high salinity, heavy metals, and UV radiations) (Yang et al., 2018). For centuries, humans have used plant extracts for medicinal and beneficial health purposes, even though the

active compounds responsible for the extract efficacy were mostly unknown. There are thousands of examples of the use of plant-derived compounds as drugs, nutraceuticals, and cosmetic ingredients (Nasri et al., 2014; Atanasov et al., 2015; Barbulova et al., 2015). The active compounds within plant extracts are mainly secondary metabolites that can be classified into four main categories according to their different chemical properties and structures: terpenoids, polyphenols, nitrogen, and sulfur-containing compounds (Ahmed et al., 2017).

Polyphenols are the largest and the most widely distributed group of bioactive compounds in the plant kingdom. They have a distinctive structural skeleton consisting of one or more aromatic phenyl rings connected to hydroxyl groups and exhibiting a wide spectrum of health-related properties including antioxidant protection, anti-inflammatory, anti-allergic, anti-atherogenic, and anti-cancer (Gorzynik-Debicka et al., 2018; Serreli and Deiana, 2019). Moreover, several studies demonstrated the antiviral potential of some classes of polyphenols against Epstein–Barr virus (Yiu et al., 2010), enterovirus 71 (Zhang et al., 2015), herpes simplex virus (HSV) (Annunziata et al., 2018), influenza virus (Lin et al., 2015), and other viruses causing respiratory tract-related infections (Zang et al., 2011). The mechanisms underpinning the antiviral activity of polyphenols are varied (for a review, see Denaro et al., 2020) and include the inhibition of the virus entry because of their permanent attachment on the virion envelope (Lin et al., 2011) or the inhibition of the enzyme responsible for the virus replication (Nutan et al., 2013). The severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2) is a zoonotic pathogenic virus identified for the first time in December 2019 (Zhu et al., 2020), responsible for one of the most serious pandemics in human history, the coronavirus disease 2019 (COVID-19): so far, the number of COVID-19 cases have amounted to over 60 million people with more than 1.4 million deaths from all over the world (<https://covid19.who.int/>). SARS-CoV-2, like other coronaviruses, is an enveloped positive-sense single-stranded RNA virus exposing a highly glycosylated Spike (S) protein on its surface, which facilitates the viral entry into host cells. Entry depends on the binding of the surface unit S1 (a portion of the S protein) to a cellular receptor, facilitating viral attachment to the surface of target cells (Hu et al., 2020). Upon binding of the S protein to the host receptor angiotensin-converting enzyme 2 (ACE2), the virus uses the cellular serine protease TMPRSS2 for the priming of S protein itself (Hoffmann et al., 2020). The transcription of TMPRSS2 is promoted by androgen receptors, which could explain the predominance and the severity of pathological signs in COVID-19-affected men compared with women (Guan et al., 2020; Remuzzi and Remuzzi, 2020), the higher proportion of men's hospitalization (Espinosa et al., 2020) and their higher mortality rates (Onder et al., 2020).

Even though recently, alternative molecular mechanisms were hypothesized to explain the virus's entry into the cells (Pirone

et al., 2020; Tresoldi et al., 2020), the binding of SARS-CoV-2 S protein to human ACE2 remains the main route of the virus's access to the cells and more directly related to the subsequent levels of infectivity (Davidson et al., 2020). After the virus's entry, the RNA genome is released into the cytoplasm and translated into two polyproteins using the translational machinery of each host cell. The two polyproteins are cleaved into the virus proteins by the main protease M<sup>Pro</sup> (Anand et al., 2003), also referred to as 3CL<sup>Pro</sup>, and the papain-like protease PL<sup>Pro</sup> (Shin et al., 2020), while the RNA gets replicated by its own RNA-dependent RNA polymerase (Ahmad et al., 2020). Once the components are all assembled, matured, and packaged into new viral copies, the viruses can then exit the host cell via exocytosis and continue their infection cycles. Sars-CoV-2 mainly targets the respiratory system, intestine, cardiovascular tissues, brain, and kidneys because these organs have the highest expression of ACE2 (Zhang et al., 2020), resulting in symptoms such as fever, headache, dry cough, and dyspnea (Pascarella et al., 2020). At the moment, there are no generally proven effective therapy for COVID-19 but, thanks to joint efforts of scientific communities, three vaccines are now available worldwide (<https://ourworldindata.org/covid-vaccinations>). As reviewed by Dube et al. (2020), antivirals can be broadly categorized into two classes: the first includes those targeting viral proteins involved in the viral life cycle or virus structure, and the other those mostly targeting host proteins, which are important for viral infection or the host's immune response.

A large number of plant-derived compounds are under investigation for their potential therapeutic effects against SARS-CoV-2. Many reports based on molecular docking analysis suggested the potential capacity of polyphenols, such as curcumin, kaempferol, catechin, naringenin, quercetin (Khaerunnisa et al., 2020) or hesperidin, rutin, and diosmin (Adem et al., 2020) to inhibit the activity of SARS-CoV-2 main protease, and consequently, the virus replication. One study also suggested that the binding of two polyphenols, punicalagin (PC), and theaflavin, to the S protein, could be exploited as a strategy to inhibit the virus's entry into human cells (Bhatia et al., 2020).

Pomegranate (*Punica granatum* L.) fruits, extensively produced by Mediterranean countries, including Tunisia, Turkey, Egypt, Spain, Morocco, and Italy, are rich in polyphenols, such as ellagitannins (ETs), mainly including  $\alpha$  and  $\beta$  isomers of PC, gallic acid (GA), ellagic acid (EA), and its glycosylated derivatives, and anthocyanins (Reddy et al., 2007). Pomegranates are majorly processed by food industries to obtain juices or jams from the arils, while the peels, which constitute around 50% of the fresh fruit weight, are discarded. It has been reported that the peels had a higher content of dietary fiber and total polyphenols, as well as a stronger antioxidant capacity (AC) than the pulp fraction of the fruit itself, thus they could be a valuable source of extracts for cosmetic and nutraceutical applications (Akhtar et al., 2015). Evidence suggests that these compounds may have protective activity against degenerative chronic diseases, such as some types of cancer, type 2 diabetes, atherosclerosis, and cardiovascular diseases (Viuda-Martos et al., 2010; Landete, 2011). Furthermore, several studies on PPEs focused on their antibacterial and antiviral activity (Howell and D'Souza, 2013)

**Abbreviations:** PPE, pomegranate peel extract; S-protein, Spike glycoprotein; PC, punicalagin; EA, ellagic acid; GA, gallic acid; EAs, ellagic acid derivatives; ACE2, angiotensin-converting enzyme 2; SARS-CoV-2, severe acute respiratory syndrome coronavirus-2; COVID-19, coronavirus disease 2019.

as well as on the property to inhibit influenza (Moradi et al., 2019) and herpes virus replication (Houston et al., 2017). These observations indicated that PPEs may be successfully employed as antiviral agents against SARS-CoV2. Therefore, this work aimed to assess the potential of PPEs to counteract SARS-CoV2 infection. We found that a hydroalcoholic extract obtained from pomegranate peels and its main constituents were able to inhibit the binding between SARS-CoV-2 S glycoprotein and ACE2 *in vitro*, suggesting a potential of the extract in the prevention of SARS-CoV-2 entry into host cells. Moreover, PPE compounds inhibited the virus 3CL protease, indicating a potential use of the extract as a natural remedy to enhance protection against SARS-CoV-2.

## MATERIALS AND METHODS

### Preparation of PPE

Dried pomegranate peels were provided by Giovomel–Azienda Biologica (Aiello del Sabato, Italy), an Italian company that produces pomegranate juice. The preparation of the PPE was performed by adding 700 mL of a solution ethanol/water (70/30, v/v) to 150 g of dried peels, at 4°C, according to Malviya et al., 2014. The mixture was homogenized 3 min at 1,500 rpm and 2 min at 3,000 rpm using a Grindomix GM 300 knife mill (Retsch GmbH, Haan, Germany). The resulting suspension was left under stirring at 150 rpm for 2 h at 25°C, avoiding light exposure. The suspension was then centrifuged at 6,300 rpm for 10 min at 4°C. The supernatant was filtered through a filter paper (FILTER-LAB, qualitative filter paper, Barcelona, Spain) and concentrated under vacuum in a rotary evaporator (IKA RV8, IKA-Werke GmbH & Co, Staufen, Germany) set to 25°C. Finally, the pH of the concentrated extract was adjusted to 7.0 with 10N NaOH and then freeze-dried until obtaining a fine powder.

### High-Resolution Mass Spectrometry Analysis of PPE

Liquid chromatography–mass spectrometry data were acquired on an Accela U-HPLC system coupled to an Exactive Orbitrap mass spectrometer (Thermo Fisher Scientific, San Jose, CA, USA) equipped with a heated electrospray interface. The chromatographic separation was carried out according to Colantuono et al. (2017). Briefly, we used a Gemini C18-110Å column, 150 × 2.0 mm, 5 µm (Phenomenex, Torrance, CA, USA) heated to 30°C and the mobile phases consisted of 0.1% formic acid water (A) and 0.1% formic acid acetonitrile (B) with a flow rate of 200 µL/min. The dry extracts were dissolved in methanol/water (50:50, v/v) and 10 µL were injected into the column. MS data acquisition was performed in negative ionization modes, in the mass range of *m/z* 100–1,300. The resolving power was set to 50,000 full width at half-maximum (FWHM, *m/z* 200) resulting in a scan time of 1 s. The automatic gain control was used in balanced mode (1 × 10<sup>6</sup> ions); maximum injection time was 100 ms. The interface parameters were the following: spray voltage 3,500 kV, capillary voltage 50 V,

capillary temperature 275°C, sheath gas 30 arbitrary units, and auxiliary gas 15 arbitrary units.

Calibration curves were constructed in the linearity ranges of 1 to 50 µg/mL for PC and 0.1 to 5 µg/mL for EA and GA. Metabolite identification was performed using exact mass values up to the fifth decimal digit with mass tolerance ±5 ppm. **Table 1** reports the polyphenols identified in PPE and individual molecular formula, retention time, theoretical mass, experimental mass, and error. The amount of each compound in the extract was determined using PC, EA, and GA as reference standards for ETs, EA derivatives (EAs), and GA, respectively. Punicalin (α, β isomers), granatin B, Causarinin, Galloyl-HHDP-hexoside, pedunculagin I (bis-HHDP-hex), and pedunculagin II (Digalloyl-HHDP-hex) were expressed as equivalents of PC. EA hexoside, EA pentoside, EA deoxyhexoside were expressed as equivalents of EA. GA was quantified with the correspondent standard. Total polyphenols were calculated as the sum of all the compounds retrieved.

### Antioxidant Activity of PPE

The AC of PPE was measured by using the ABTS assay as reported by Re et al. (1999). Briefly, a stable stock solution of ABTS<sup>•+</sup> was produced by reacting a 7 mmol/L aqueous solution of ABTS with 2.45 mmol/L potassium persulfate (final concentration) and allowing the mixture to stand in the dark at 4°C for 16 h before use. The ABTS<sup>•+</sup> solution was diluted with ethanol to an absorbance of 0.700 ± 0.05 at 734 nm. Freeze-dried PPE was appropriately diluted in water and 0.1 mL of reconstituted extract was added to 1 mL of ABTS<sup>•+</sup> solution. The mixture was allowed to stand at room temperature for 2.5 min before the absorbance was recorded at 734 nm by using the multiplate reader Victor Nivo (Perkin Elmer, Woodbridge, ON, Canada). Results were expressed as µmol Trolox equivalents (TE)/g of powder.

### SARS-CoV-2 Spike Receptor-Binding Domain (Spike RBD)/ACE2 Binding Inhibitor Assay

The inhibition of the Spike-ACE2 interaction was measured using the SARS-CoV2 Inhibitor Screening Assay kit (Adipogen Corporation, San Diego, CA, USA; Cat. N° AG-44B-0007-KI01). According to the manufacturer's instructions, 100 µL per well of Spike RBD, produced as a recombinant protein in human cells (1 µg/ml), was used to coat a 96-well plate for 16 h at 4°C. The plate was then treated with the blocking buffer (2% albumine bovine serum (BSA) in phosphate buffered saline (PBS)) for 2 h at room temperature, washed in wash buffer (0.1% Tween<sup>®</sup> 20 in PBS), and incubated with the PPE or compounds for 1 h at 37°C in the Inhibitor Mix Solution (IMS), composed of 0.2% BSA and 0.05% Tween<sup>®</sup> 20 in PBS, containing biotin-conjugated-ACE2 (0.5 µg/mL). ACE2 was prepared as a recombinant protein in human cells. After incubation, HRP labeled-streptavidin diluted (1:200 dilution) was added to each well and incubated for 1 h at room temperature. The reaction was developed by adding 100 µL of tetramethylbenzidine (Neogen, Lansing, MI, USA) for 5 min at

**TABLE 1** | High-resolution mass spectrometry identification of the compounds in pomegranate peels extract (PPE) achieved by Orbitrap MS.

Compound	Molecular formula	Theoretical	experimental	Mass accuracy (ppm)	Retention time (min)
		[M-H] <sup>-</sup> m/z			
Punicalin	C <sub>34</sub> H <sub>22</sub> O <sub>22</sub>	781.053	781.05389	1.14	6.3–6.7
Punicalagin	C <sub>48</sub> H <sub>28</sub> O <sub>30</sub>	1083.05926	1083.05994	0.63	7.6–7.9
Pedunculagin I (bis-HHDP-hex)	C <sub>34</sub> H <sub>24</sub> O <sub>22</sub>	783.06865	783.06915	0.64	6.8–7.2–7.8–8.3
Pedunculagin II (digalloyl-HHDP-hex)	C <sub>34</sub> H <sub>26</sub> O <sub>22</sub>	785.0843	785.08502	0.92	8.03–8.6–9.1–9.4
Lagerstannin B	C <sub>41</sub> H <sub>26</sub> O <sub>27</sub>	949.05887			
Causarinin	C <sub>41</sub> H <sub>28</sub> O <sub>26</sub>	935.0796	935.08118	1.69	8.2
Galloyl-HHDP-hexoside	C <sub>27</sub> H <sub>22</sub> O <sub>18</sub>	633.07334	633.0741	1.2	8.6
Granatin B	C <sub>41</sub> H <sub>28</sub> O <sub>27</sub>	951.07452	951.07556	1.09	9.2
Ellagic acid hexoside	C <sub>20</sub> H <sub>16</sub> O <sub>13</sub>	463.05181	463.05225	0.95	8.7
Ellagic acid dihexoside	C <sub>26</sub> H <sub>26</sub> O <sub>18</sub>	625.10464			
Ellagic acid pentoside	C <sub>19</sub> H <sub>14</sub> O <sub>12</sub>	433.04125	433.04135	0.23	9.6
Ellagic acid deoxyhexoside	C <sub>20</sub> H <sub>16</sub> O <sub>12</sub>	447.0569	447.05701	0.25	9.7
Ellagic acid	C <sub>14</sub> H <sub>6</sub> O <sub>8</sub>	300.99899	300.99915	0.53	10.4
Gallic acid	C <sub>7</sub> H <sub>6</sub> O <sub>5</sub>	169.01425	169.01378	−2.78	6.4

RT and measured at 450 nm by the microplate reader Victor Nivo (Perkin Elmer, Woodbridge, ON, Canada).

## Microscale Thermophoresis (MST)

MST experiments were performed on a Monolith NT 115 system (Nano Temper Technologies, Munchen, Germany) and designed to evaluate the ability of the PPE to bind ACE2, S protein, and RBD (10108-H08H, 40589-V08B, 40592-V08B from Sino Biological, Wayne, PA, USA). The proteins used in the study were: ACE2 (NP\_068576.1) (Met1-Ser740), Spike FL (YP\_009724390.1) (Val16-Pro1296), and RBD Spike (YP\_009724390.1) (Arg319-Phe541); all three produced as recombinant in baculovirus-insect cells and carrying a polyhistidine tag at the C-terminus. Each protein (10 μM) was labeled with NT-647-NHS reactive dye (30 μM) (NanoTemper Technologies, GmbH, München, Germany), which reacts efficiently with the primary amines of the proteins to form a stable dye protein conjugate. PPE was used in the concentration range of 65–1.92 × 10<sup>−3</sup> μM in the experiment with ACE2, 32.5–9.92 × 10<sup>−4</sup> μM with Spike and 3.25–9.92 × 10<sup>−5</sup> μM with RBD Spike, respectively, preparing 16-point serial dilution (1:2) in PBS supplemented with tween 0.05%. These values corresponded to the quantity of PC, the most abundant extract polyphenol, as determined by chemical analysis. The MST was carried out using 100% LED and 20% IR-laser power at 37°C. The ligand in the experiments with Spike FL and RBD induced quenching of fluorescence and an SDS denaturation test (SD test) was performed to confirm the specificity of the interaction. An equation implemented by the software MO-S002 MO Affinity Analysis (Nano Temper Technologies, Munchen, Germany), provided by the manufacturer was used for fitting the normalized fluorescence values at different concentrations of the ligands (Mercurio et al., 2018; Bellia et al., 2019).

## Lentivirus Infection

Human kidney-2 cells (HK-2) were obtained from American Type Culture Collection (ATCC) and were cultured in Dulbecco's modified Eagle medium (DMEM) (EuroClone, Milano Italy) supplemented with 5% (v/v) FBS, 1% insulin-transferrin-sodium selenite media supplement (ITS) (Sigma-Aldrich-Merck KGaA, Darmstadt, Germany) and 1% penicillin-streptomycin. The cells were maintained at 37°C, 5% CO<sub>2</sub> in a humidified incubator according to the guidelines provided by the vendors, plated in 96-well plates (CellCarrier-96 ultra with lid, Perkin Elmer, Woodbridge, ON, Canada), at a density of 5 × 10<sup>3</sup> per well in 100 μL culture medium. After 24 h, the cells were incubated with either 0.04 mg/mL of PPE extract or water for 4 h. The cells were then infected with SARS-CoV-2 Spike-pseudotyped lentivirus (Firefly Luciferase SARS-CoV-2 lentiviral particles, GeneCopoeia, Inc., Rockville, MD, USA) and the control vesicular stomatitis virus G (VSVG) protein pseudotyped lentivirus (HLUC-Lv201 Firefly luciferase + eGFP lentifect, GeneCopoeia, Inc., Rockville, MD, USA) at a concentration of 4.9E + 9 GC/mL and 1.2E + 9 GC/mL, respectively. After 72 h, the cells were fixed in 4% paraformaldehyde and washed three times in PBS. Nuclei were counterstained with 4',6-diamidino-2-phenylindole, and after washing, the cells were imaged by the Operetta High Content Imaging System (Perkin Elmer, Woodbridge, ON, Canada), using a 20× magnification objective. Acquired images were analyzed by the software Columbus (Perkin Elmer, Woodbridge, ON, Canada), version 2.6.0. Image analysis consisted of identifying and counting viral-infected HK-2 cells based on 488-intensity fluorescence. The infection rate was calculated as the ratio between the number of infected cells and the number of total cells counted per well. The plot shows the percentage of 488-positive cells after pomegranate treatment compared with that in H<sub>2</sub>O-treated cells.

## Gene Expression Analysis on HK-2 Cells

Cells were plated in 24-well plates at a density of  $5 \times 10^4$  per well in a 500- $\mu$ L culture medium. After 24 h the cells were incubated with 0.04 mg/mL of PPE for 72 h and then collected for RNA extraction, performed by the GeneElute Mammalian total RNA purification kit (Sigma Aldrich, Merck KGaA, Darmstadt, Germany). The RNA was treated with deoxyribonuclease (DNase) I (Thermo Fisher Scientific, Dallas, TX, USA) at 37°C for 30 min. Reverse transcription was performed using the RevertAid™ First Strand cDNA Synthesis Kit (Thermo Fisher Scientific, Dallas, TX, USA). Semiquantitative RT-PCR was performed with the Quantum RNA™ kit (Thermo Fisher Scientific, Dallas, TX, USA) containing primers to amplify 18S ribosomal RNA (18S rRNA) along with competitors, that reduced the amplified 18S rRNA product within the range to be used as endogenous standard. The amplification reactions were made using specific oligonucleotides by the Mastercycler™ ProS (Eppendorf, Milan, Italy) with the following general scheme: 2 min at 94°C followed by 35 cycles of 94°C for the 30 s, 50°C for 30 s, and 72°C for 30 s, with a 10 min final extension at 72°C. The PCR products were loaded on 1.5% agarose gel, and the amplification bands were visualized and quantified with the Geliance 200 Imaging system (Perkin Elmer, Woodbridge, ON, Canada). The amplification band corresponding to the analyzed gene was normalized to the amplification band

corresponding to the 18S and reported as a percentage of untreated controls set as 100%. The used primer sequences for the amplifications were the following: ACE2 Fw ATG TCACTTTCTGCAGCC; ACE2 Rv GTTGAGCAGTGGCCTTACAT; TMPRSS2 Fw ATTGCCGGCACTTGTGTTCA; TMPRSS2 Rv ACAGTGTGCACCTCAAAGAC.

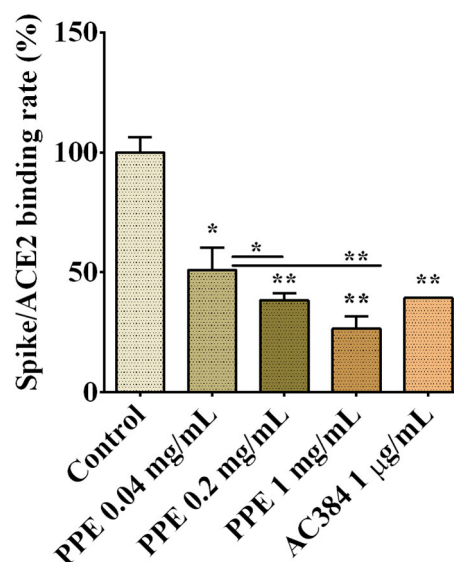
## 5alpha Reductase Activity

Hair follicle dermal papilla cells (HFDPC) were seeded in a 96-well plate at a density of  $8 \times 10^3$ , after 16 h they were stimulated with testosterone 600 nM and treated with pomegranate extract or finasteride 100 nM for 24 h. Another 96-well plate was coated with 100 ng of dihydrotestosterone (DHT)-conjugated BSA, the day after the plate was washed with PBS +0.05% Tween20 and

**TABLE 2 |** Total amount of ETs, EA derivatives, and GA in PPE.

Compounds	PPE (mg/g)
Punicalagin	182.31 $\pm$ 0.75
Punicalin	61.95 $\pm$ 2.34
Granatin B	61.04 $\pm$ 7.25
Causarinin	20.79 $\pm$ 2.52
Galloyl-HHDP-hexoside	45.4 $\pm$ 1.53
Lagerstannin B	<LOD
Pedunculagin I	50.25 $\pm$ 0.98
Pedunculagin II	28.04 $\pm$ 0.42
<b>Ellagitannins</b>	449.78 $\pm$ 8.31
Ellagic acid	10.71 $\pm$ 1.17
Ellagic acid hexoside	3 $\pm$ 0.13
Ellagic acid pentoside	1.88 $\pm$ 0.09
Ellagic acid deoxyhexoside	1.87 $\pm$ 0.11
Ellagic acid dihexoside	<LOD
<b>Ellagic acid derivatives</b>	17.45 $\pm$ 1.49
Galllic acid	0.98 $\pm$ 0.11
<b>Total*</b>	468.2 $\pm$ 9.69

The amount of each compound in the extract was determined using PC, EA, and GA as reference standards for ETs, EAs, and GA, respectively. Punicalin, granatin B, causarinin, galloyl-HHDP-hexoside, pedunculagin I (bis-HHDP-hex), pedunculagin II (digalloyl-HHDP-hex) were expressed as equivalents of PC. EA hexoside, EA pentoside, EA deoxyhexoside were expressed as equivalents of EA. GA was quantified with the correspondent standard. Total polyphenols were calculated as sum of all the compounds retrieved. The values are expressed as mg/g of dry powder (mean values  $\pm$  standard deviation). EA, ellagic acid derivatives; ET, ellagitannin; GA, gallic acid; PC, punicalagin; PPE, pomegranate peels extract. \*Expressed as sum of mg of punicalagin equiv. + mg of ellagic acid equiv. + mg of gallic acid equiv.; <LOD, lower than the limit of detection.



**FIGURE 1 |** Spike/ACE2 binding in the presence of PPE, used at three concentrations, compared with control and antibody inhibitor AC384. The results are the averages of three independent experiments, expressed as percentages respect to control arbitrarily set as 100%. The error bars represent SDs and the asterisks indicate statistically significant values (\**p*-value is between 0.01 and 0.05; \*\*0.001 and 0.01) according to *T*-test. ACE2, angiotensin-converting enzyme 2; PPE, pomegranate peel extract.

**TABLE 3 |** Spike/ACE2 binding (%) in the presence of punicalagin, ellagic acid, and gallic acid, at concentrations corresponding to those present in 0.04 mg/mL of PPE and equal to 7.29, 0.43, and 0.04  $\mu$ g/mL, respectively.

Sample	Binding (%)	SD	<i>p</i> -value
Spike/ACE2	100	$\pm$ 10	
Spike/ACE2 + PPE	51	$\pm$ 11	0.04
Spike/ACE2 + Punicalagin	36	$\pm$ 4	0.01
Spike/ACE2 + Gallic acid	100	$\pm$ 2	0.5
Spike/ACE2 + Ellagic acid	64	$\pm$ 10	0.03

The results are the averages of three independent experiments, expressed as percentage respect to control arbitrarily set as 100%. ACE2, angiotensin-converting enzyme 2; PPE, pomegranate peels extract.

incubated with a blocking solution containing PBS, Tween20, and 3% of BSA for 1 h. After three washes, the plate was loaded with 50  $\mu$ L of cell supernatants derived from cell treatments, plus 50  $\mu$ L of biotin-conjugated anti-DHT antibody (1:1,000 dilution in PBS + BSA 1%). After 2 h, the plate was washed three times, 5  $\mu$ g/mL of peroxidase streptavidin conjugated was added to the plate and incubated for 1 h at room temperature. After three washes, 0.5 mg/mL of OPD in 50 mM citrate buffer + 0.012%  $H_2O_2$  was added and the absorbance was measured at 490 nm by the microplate reader Victor Nivo (Perkin Elmer, Woodbridge, ON, Canada).

### 3CL Protease Activity Assay

To measure the activity of the viral 3CL protease in the presence of PPE extract we used the Untagged (SARS-CoV-2) Assay kit provided by BPS Bioscience (San Diego, CA, USA), according to the procedure described in the provider's instructions. Briefly, 15 ng of 3CL protease was incubated with the extract at the indicated concentrations or with 500  $\mu$ M of GC376, used as a positive control. After 30 min of incubation at room temperature, the enzymatic reaction was carried on for 24 h by the addition of 40  $\mu$ M 3CL protease substrate. The fluorescence was measured by the Victor Nivo Microplate reader (Perkin Elmer Woodbridge, ON, Canada) exciting at 360 nm and detecting at 460 nm.

### Statistical Analysis

All the measures were expressed as means  $\pm$  standard deviations (SD) of three independent experiments. A paired-samples *t*-test

was conducted using Microsoft Excel; a *p*-value lower than 0.05 was considered statistically significant.

## RESULTS AND DISCUSSION

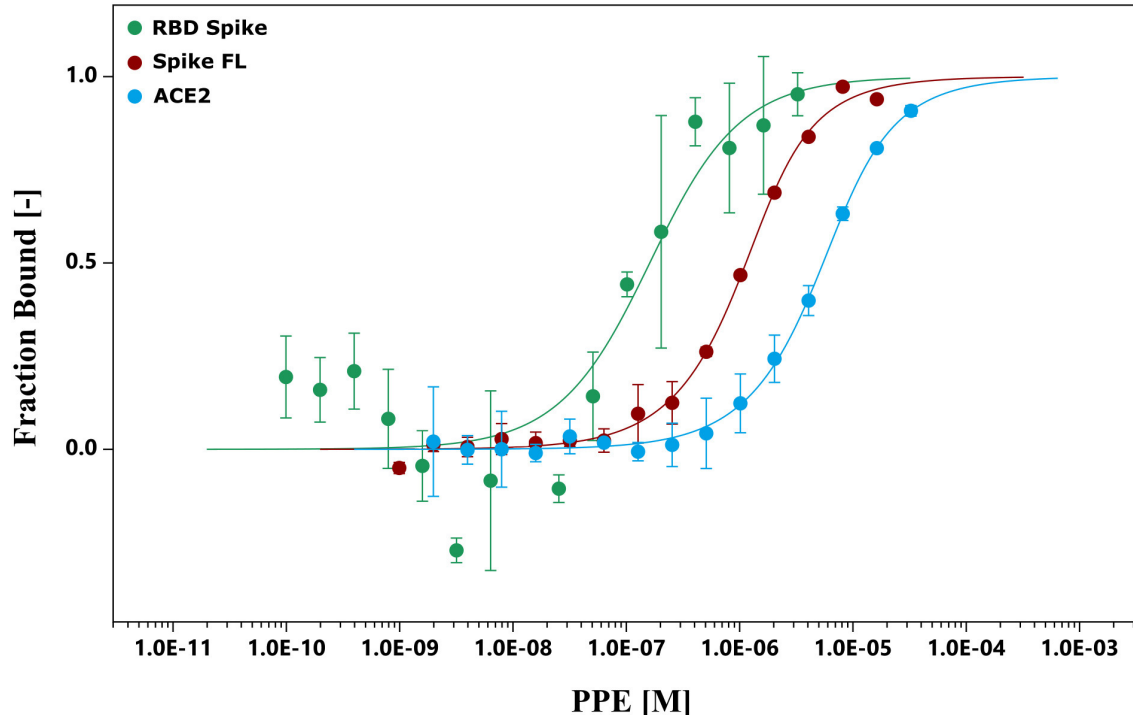
### Chemical Characterization of PPE

The concentration of polyphenols in PPE is reported in Table 2. ETs were the most abundant compounds. Specifically, PC represented 38.9% of all the polyphenols detected in the extract, followed by pedunculagin anomers and punicalin anomers representing 16.7 and 13.2% of total polyphenols, respectively. These results were in accordance with previous studies published by Lu et al. (2008) and Fischer et al. (2011). The sum of EAs and GA represented 3.9% of the total polyphenols in PPE.

Notably, the AC of PPE measured by the ABTS method was 3,590  $\mu$ mol TE/g of PPE. Our data showed that 1 g of freeze-dried PPE was obtained by 3.45 g of dried pomegranate peels. According to these data, the correspondent AC calculated for dried pomegranate peels was 1,041  $\mu$ mol TE/g of dried pomegranate peels, in line with the data showed by Marchi et al. (2015) (872–1,056  $\mu$ mol TE/g of dried peels) and by Fischer et al. (2011) (1,362  $\mu$ mol TE/g of dried peels and 2,887  $\mu$ mol TE/g of dried mesocarp).

### Effect of PPE on Spike/ACE2 Binding

To assess whether PPE had inhibitory activity on Spike/ACE2 binding, we used a SARS-CoV-2 inhibitor screening kit by Adipogen Corporation (San Diego, CA, USA). PPE, used



**FIGURE 2 |** MST. The binding curves were obtained incubating PPE with the RBD Spike, Spike FL protein, and ACE2. ACE2, angiotensin-converting enzyme 2; MST, microscale thermophoresis; PPE, pomegranate peel extract; RBD Spike, Spike receptor-binding domain; Spike FL, Spike full-length.

at three concentrations ranging from 0.04 mg/mL to 1 mg/mL, inhibited the interaction between Spike and ACE2 up to 74%, and this effect was dose-dependent (**Figure 1** and **Supplementary Figure 1**). As a positive control, we used AC384, a monoclonal antibody that inhibited the binding between Spike and ACE2 by specifically recognizing ACE2 itself.

To provide insights into which of the PPE polyphenols were relevant for that inhibition, the three most abundant components of PPE, that is PC, EA, and GA were individually tested at the same concentrations as present in 0.04 mg/mL PPE. The results in **Table 3** and **Supplementary Table 1** showed that PC had the most effect on the binding between Spike and ACE2 by exerting 49% inhibition, followed by EA with 36% inhibition, whereas GA did not have any effect.

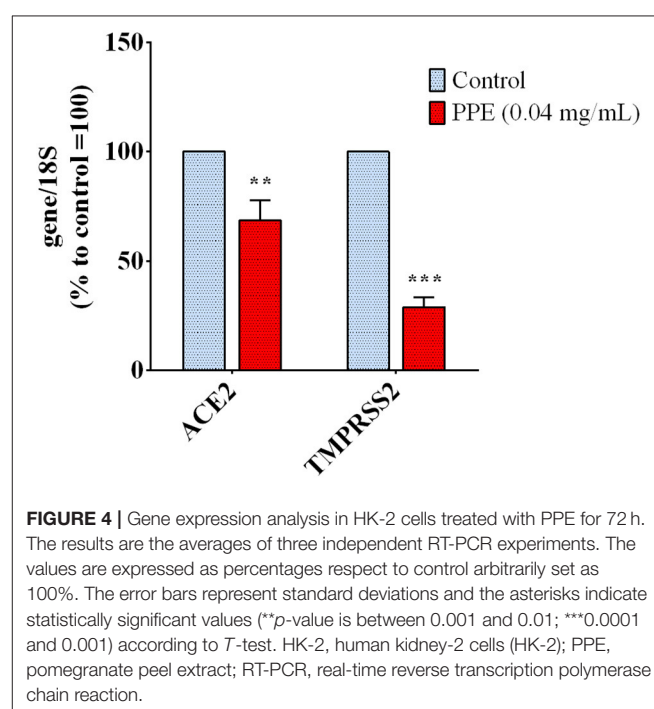
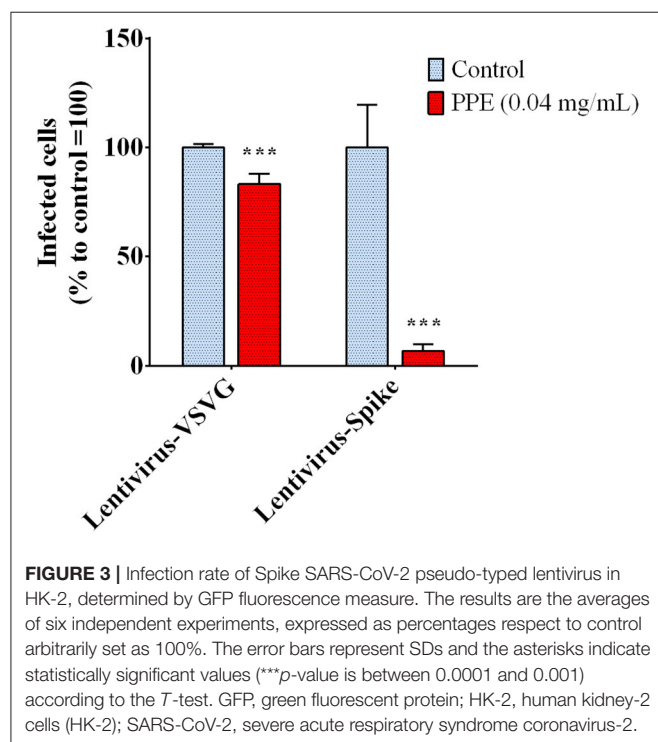
To further investigate the binding capacity of the pomegranate compounds, the chemical interactions between the extract and Spike, and between the extract and ACE2, were analyzed by MST experiments (**Figure 2** and **Supplementary Figures 2, 3**). The results showed that the PPE bound both the proteins (**Figure 2**), even though the interaction with Spike was 10-fold as strong as the one with ACE2. Moreover, we supposed that the binding of PPE compounds to Spike was mostly due to a high affinity toward the RBD of the protein, as the binding curve of PPE compounds plus Spike full-length was very similar to that of PPE compounds plus RBD.

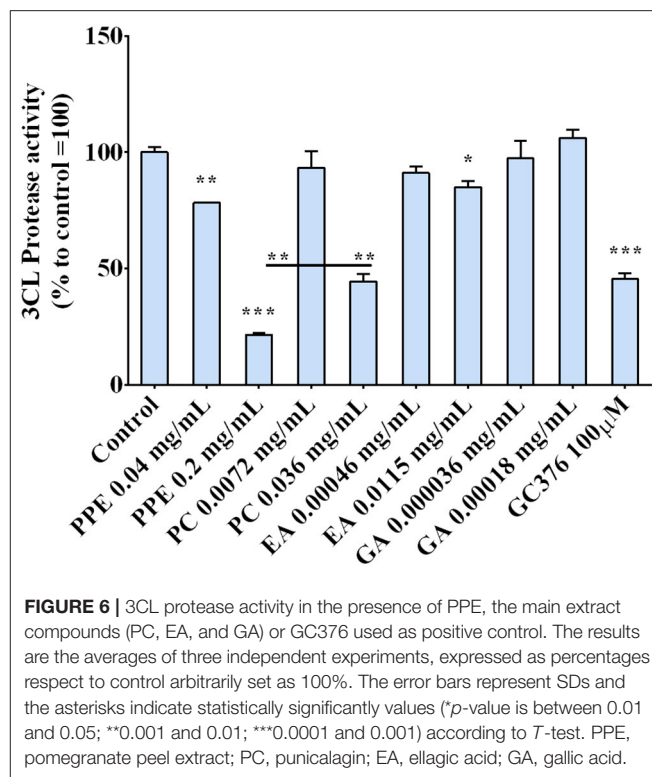
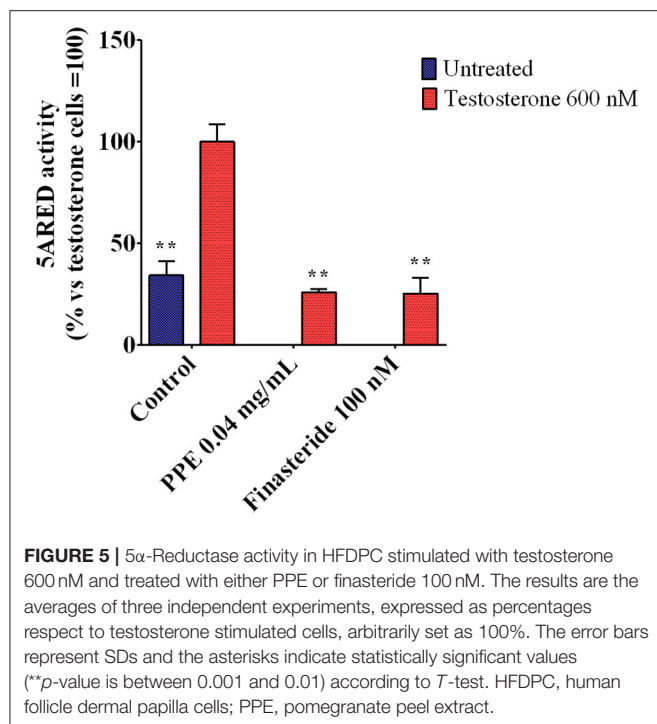
The biochemical data prompted us to investigate the capacity of PPE to effectively inhibit the interaction between Spike and ACE2 in a cellular model. To do that, we used a system based on a Spike-carrying Lentivirus, infecting human kidney-2 cells (HK-2), already known to express ACE2 (Koka et al., 2008). As a control, we used a lentivirus that carried the VSVG

protein instead of Spike, thus it entered the cells without a specific recognition of any receptor. Both viruses carried the green fluorescent protein (GFP) gene in their RNA genome, which was expressed and easily detected in the cells upon infection. PPE was used at the safe dose of 0.04 mg/mL, as determined by the cytotoxicity MTT assay (data reported in **Supplementary Material**). As shown in **Figure 3**, when the cells were infected by the lentivirus carrying the Spike protein in the presence of PPE, the percentage of GFP fluorescent cells (infected cells) was almost significantly abolished after 72 h. In contrast, when the cells were infected by the lentivirus carrying VSVG protein, the percentage of infected cells was reduced only by 18%, suggesting a specific inhibitory effect of PPE toward Spike/ACE2 binding.

To investigate whether PPE could regulate host genes involved in the virus uptake, we measured the expression level of ACE2 and TMPRSS2 genes in HK-2 cells treated with the extract for 72 h. As reported in **Figure 4**, the gene expression analysis showed that the treatment of HK-2 cells with the PPE at 0.04 mg/mL reduced the level of ACE2 and TMPRSS2 gene expression by 30 and 70%, respectively. This suggested that PPE, besides Spike/ACE2 binding inhibition, was able to downregulate the expression of two genes responsible for the virus access into the cells.

As the expression of TMPRSS2 was mainly regulated by androgens (Hong et al., 2008; Oyelowo et al., 2019), we analyzed whether PPE inhibited the 5 $\alpha$ -reductase activity, the primary enzyme involved in DHT synthesis. As shown in **Figure 5**, PPE at 0.04mg/mL reduced the activity of the 5 $\alpha$ -reductase by 65% in HFDPC, after stimulation by testosterone. This effect was similar to that obtained by finasteride, used as a positive control (Rattanachitthawat et al., 2019).





## Activity of PPE on SarsCov-2 Main Protease

The regulation of the 3CL protease, one of the main proteins involved in the virus replication, by the extract was investigated by incubating the enzyme with PPE and its main components, PC, EA, and GA. The results, reported in **Figure 6**, indicate that PPE at both concentrations (0.04 and 0.2 mg/ml), inhibited the activity of the 3CL protease up to 80% when used at 0.2 mg/ml. Among the compounds, PC was the most effective in inhibiting the enzymatic activity (by 50%), EA inhibited only by 10%, while GA did not have any effect, suggesting a synergic effect of the PPE polyphenols in inhibiting the protease activity.

## CONCLUSIONS

The activity of plants secondary metabolites against SARS-CoV-2 infection and replication has been extensively reviewed in the last months (da Silva Antonio et al., 2020; Sayed et al., 2020; Weng, 2020) and many studies, based on *in silico* approaches, suggested some of them as potential drug candidates for COVID-19 treatment (Majumder and Mandal, 2020; Singh et al., 2020). Both viral structural proteins, like Spike, and non-structural proteins, such as 3CLpro, PLpro, and RdRp, have been proposed as valuable targets for anti-SARS-CoV-2 therapeutic strategies. Through molecular-docking studies, Khalifa et al. (2020) found that some hydrolysable tannins, in particular pedunculagin, tercatin, and castalin, might serve as potential inhibitors of SARS-CoV-2 as they were able to specifically bind the 3CL protease catalytic site.

In parallel studies, Hariprasad et al. (2020) tested the virtual interaction between many plant secondary metabolites and four target proteins involved in COVID-19, the host protease TMPRSS2 and the three virus proteins, Spike, main protease, and RNA-dependent RNA polymerase, and predicted among the class of triterpenoids the most active compounds in blocking the Spike-binding site. Bhatia et al. (2020) also identified PC among dietary polyphenols as a potential inhibitor of Spike and other viral proteases. On the other side, human targets have been taken under consideration as well: ACE2 is certainly the most explored as it turned out to be the main “door lock” that the virus uses to get into the cells. However, ACE2 also has a pivotal role in many physio-pathological processes in human tissues, thus targeting this enzyme needs careful evaluation to ensure that the benefit–risk balance turns favorable (Lacroix et al., 2015; Mostafa-Hedeab, 2020).

In the present study, we found that the polyphenols contained in an ethanolic extract derived from pomegranate peels inhibited the interaction between Spike and ACE2, and reduced the activity of the viral 3CL protease *in vitro*, potentially suggesting the use of the extract as an adjuvant in the treatment against SARS-CoV-2 infections. Data showed that the most effective polyphenols in the extract were PC and EA possibly through a chemical interaction of the hydroxyl and galloyl groups in their molecules with amino acid residues at the active sites of the protein targets of SARS-CoV-2 or human cells, as supported by other studies based on molecular docking analysis (Surucic et al., 2021). The inhibitory effect on Spike/ACE2 binding was confirmed by experiments with a pseudotyped lentivirus, whose

entry into the human cells was dependent on Spike protein. Consistent with the *in-vitro* observations, our data showed that lentivirus infection was almost completely abolished by the polyphenol-containing PPE. This inhibition was also associated with the downregulation of the gene expression of both ACE2 and the protease TMPRSS2, the one involved in Spike priming. In particular, in this study 0.04 mg/ml of PPE was used for the *in vitro* experiments in HK-2 cell culture. This concentration appears higher than that used in some reports in which it was indicated that the CC50 of PPE was about 55.6  $\mu$ g/ml (Moradi et al., 2019). Anyway, in literature, different cytotoxicity data are reported in relation to different cell cultures used (Dana et al., 2015; Mastrogianni et al., 2019; Sorrenti et al., 2019; Keta et al., 2020). Moreover, we also provided evidence that PPE was able to inhibit the activity of the 3CL protease up to 80%, suggesting that PPE might have multiple biological roles in reducing the chance of virus to anchor the cells and get internalized.

In conclusion, inhibiting Spike/ACE2 binding still represents one of the most popular strategies to control SARS-CoV-2, and polyphenol-rich extracts represent promising candidates to reduce virus infection and replication thus being proposed as bioactive ingredients in pharmaceutical, nutraceutical, and/or cosmetic formulations. In agreement with our results, a recent report demonstrated that pomegranate juice was effective in reducing the infectious capacity of SARS-CoV-2 and influenza

virus in VeroE6 cells suggesting an antiviral activity of both viruses (Frank et al., 2020). The study here presented paves the way for longer and in depth-investigations on the activity of pomegranate peel polyphenols in preventing SARS-CoV-2 infection *in vivo* and it may also promote new ideas on how to reuse an agroindustry byproduct for valuable and healthy applications.

## DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found at: MetaboLights, <https://www.ebi.ac.uk/metabolights/>, MTBLS2309.

## AUTHOR CONTRIBUTIONS

AT, AC, LP, DI, GG, and PV have performed experimental work. AT, AC, IC, and FA contributed to conception and design of work. AT, AC, EP, and FA contribute to manuscript preparation and revision.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fchem.2021.638187/full#supplementary-material>

## REFERENCES

- Adem, S., Eyupoglu, V., Sarfraz, I., Rasul, A., and Ali, M. (2020). Identification of potent COVID-19 main protease (Mpro) inhibitors from natural polyphenols: an *in silico* strategy unveils a hope against CORONA. *Preprints* 2020:2020030333. doi: 10.20944/preprints202003.0333.v1
- Ahmad, J., Ikram, S., Ahmad, F., Rehman, I. U., and Mushtaq, M. (2020). SARS-CoV-2 RNA dependent RNA polymerase (RdRp) - a drug repurposing study. *Heliyon* 6:e04502. doi: 10.1016/j.heliyon.2020.e04502
- Ahmed, E., Arshad, M., Khan, M., Amjad, M., Sadaf, H., Rawalpindi, P., et al. (2017). Secondary metabolites and their multidimensional prospective in plant life. *J. Pharmacogn. Phytochem.* 6, 205–214.
- Akhtar, S., Ismail, T., Fraternali, D., and Sestili, P. (2015). Pomegranate peel and peel extracts: chemistry and food features. *Food Chem.* 174, 417–425. doi: 10.1016/j.foodchem.2014.11.035
- Anand, K., Ziebuhr, J., Wadhwani, P., Mesters, J. R., and Hilgenfeld, R. (2003). Coronavirus main proteinase (3CLpro) structure: basis for design of anti-SARS drugs. *Science* 300, 1763–1767. doi: 10.1126/science.1085658
- Annunziata, G., Maisto, M., Schisano, C., Ciampaglia, R., Narciso, V., Tenore, G. C., et al. (2018). Resveratrol as a novel anti-herpes simplex virus nutraceutical agent: an overview. *Viruses* 10:473. doi: 10.3390/v10090473
- Atanasov, A. G., Waltenberger, B., Pferschy-Wenzig, E. M., Linder, T., Wawrosch, C., Uhrin, P., et al. (2015). Discovery and resupply of pharmacologically active plant-derived natural products: a review. *Biotechnol. Adv.* 33, 1582–1614. doi: 10.1016/j.biotechadv.2015.08.001
- Barbulova, A., Colucci, G., and Apone, F. (2015). New trends in cosmetics: by-products of plant origin and their potential use as cosmetic active ingredients. *Cosmetics* 2, 82–92. doi: 10.3390/cosmetics2020082
- Bellia, F., Lanza, V., García-Viñuales, S., Ahmed, I. M. M., Pietropaolo, A., Iacobucci, C., et al. (2019). Ubiquitin binds the amyloid,  $\beta$  peptide and interferes with its clearance pathways. *Chem. Sci.* 10, 2732–2742. doi: 10.1039/C8SC03394C
- Bhatia, S., Giri, S., Lal, A. F., and Singh, S. (2020). Identification of potential inhibitors of dietary polyphenols for SARS-CoV-2 M protease: an *in silico* study. *Trop. Public Health* 1, 21–29.
- Colantuono, A., Vitaglione, P., Ferracane, R., Campanella, O. H., and Hamaker, B. R. (2017). Development and functional characterization of new antioxidant dietary fibers from pomegranate, olive and artichoke by-products. *Food Res. Int.* 101, 155–164. doi: 10.1016/j.foodres.2017.09.001
- da Silva Antonio, A., Wiedemann, L. S. M., and Veiga-Junior, V. F. (2020). Natural products' role against COVID-19. *RSC Adv.* 10, 23379–23393. doi: 10.1039/D0RA03774E
- Dana, N., Haghjooy, S. H., and Rafiei, L. (2015). Antiangiogenic and antiproliferative effects of black pomegranate peel extract on melanoma cell line. *Res. Pharm. Sci.* 10, 117–124.
- Davidson, A. M., Wysocki, J., and Battle, D. (2020). Interaction of SARS-CoV-2 and other coronavirus with ACE (Angiotensin-Converting Enzyme)-2 as their main receptor: therapeutic implications. *Hypertension* 76, 1339–1349. doi: 10.1161/HYPERTENSIONAHA.120.15256
- Denaro, M., Smeriglio, A., Barreca, D., De Francesco, C., Occhiuto, C., Milano, G., et al. (2020). Antiviral activity of plants and their isolated bioactive compounds: an update. *Phytother. Res.* 34, 742–768. doi: 10.1002/ptr.6575
- Dube, T., Ghosh, A., Mishra, J., Kompella, U. B., and Panda, J. J. (2020). Repurposed drugs, molecular vaccines, immune-modulators, and nanotherapeutics to treat and prevent COVID-19 associated with SARS-CoV-2, a deadly nanovector. *Adv. Ther.* 4:2000172. doi: 10.1002/adtp.20200172
- Espinosa, O. A., Zanetti, A., Antunes, E. F., Longhi, F. G., Matos, T. A., and Battaglini, P. F. (2020). Prevalence of comorbidities in patients and mortality cases affected by SARS-CoV2: a systematic review and meta-analysis. *Rev. Inst. Med. Trop. Sao Paulo* 62:e43. doi: 10.1590/s1678-9946202062043
- Fischer, U. A., Carle, R., and Kammerer, D. R. (2011). Identification and quantification of phenolic compounds from pomegranate (*Punica granatum* L.) peel, mesocarp, aril and differently produced juices by HPLC-DAD-ESI/MS(n). *Food Chem.* 127, 807–821. doi: 10.1016/j.foodchem.2010.12.156

- Frank, B., Conzelmann, C., Weil, T., Groß, R., Jungke, P., Egger, M., et al. (2020). Antiviral activity of plant juices and green tea against SARS-CoV-2 and influenza virus *in vitro*. *bioRxiv [Preprint]*. doi: 10.1101/2020.10.30.360545
- Gorzynik-Debicka, M., Przychodzen, P., Cappello, F., Kuban-Jankowska, A., Marino Gammazza, A., Knap, N., et al. (2018). Potential health benefits of olive oil and plant polyphenols. *Int. J. Mol. Sci.* 19:686. doi: 10.3390/ijms19030686
- Guan, W. J., Ni, Z. Y., Hu, Y., Liang, W. H., Ou, C. Q., He, J. X., et al. (2020). Clinical characteristics of coronavirus disease 2019 in China. *N. Engl. J. Med.* 382, 1708–1720. doi: 10.1056/NEJMoa2002032
- Hariprasad, P., Gowtham, H. G., Monu, D. O., Ajay, Y., Gourav, C., Vasantharaja, R., et al. (2020). Exploring structurally diverse plant secondary metabolites as a potential source of drug targeting different molecular mechanisms of Severe Acute Respiratory Syndrome Coronavirus-2 (SARS-CoV-2) pathogenesis: an *in silico* approach. doi: 10.21203/rs.3.rs-27313/v1. [Epub ahead of print].
- Hoffmann, M., Kleine-Weber, H., Schroeder, S., Krüger, N., Herrler, T., Erichsen, S., et al. (2020). SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. *Cell* 181, 271–280.e8. doi: 10.1016/j.cell.2020.02.052
- Hong, M. Y., Seeram, N. P., and Heber, D. (2008). Pomegranate polyphenols down-regulate expression of androgen-synthesizing genes in human prostate cancer cells overexpressing the androgen receptor. *J. Nutr. Biochem.* 19, 848–855. doi: 10.1016/j.jnutbio.2007.11.006
- Houston, D., Bugert, J. J., Denyer, S. P., and Heard, C. M. (2017). Correction: potentiated virucidal activity of pomegranate rind extract (PRE) and punicalagin against Herpes simplex virus (HSV) when co-administered with zinc (II) ions, and antiviral activity of PRE against HSV and aciclovir-resistant HSV. *PLoS ONE* 12:e0188609. doi: 10.1371/journal.pone.0188609
- Howell, A. B., and D'Souza, D. H. (2013). The pomegranate: effects on bacteria and viruses that influence human health. *Evid. Based Complement. Altern. Med.* 2013:606212. doi: 10.1155/2013/606212
- Hu, B., Guo, H., Zhou, P., and Shi, Z. L. (2020). Characteristics of SARS-CoV-2 and COVID-19. *Nat. Rev. Microbiol.* 19, 141–154. doi: 10.1038/s41579-020-00459-7
- Isah, T. (2019). Stress and defense responses in plant secondary metabolites production. *Biol. Res.* 52:39. doi: 10.1186/s40659-019-0246-3
- Keta, O., Deljanin, M., Petkovic, V., Zdunic, G., Jankovic, T., Zivkovic, J., et al. (2020). Pomegranate (*Punica granatum* L.) Peel Extract: potential cytotoxic agent against different cancer cell lines. *Rec. Nat. Prod.* 14, 326–339. doi: 10.25135/rnp.170.19.11.1477
- Khaerunnisa, S., Kurniawan, H., Awaluddin, R., Suhartati, S., and Soetjipto, S. (2020). Potential inhibitor of COVID-19 main protease (Mpro) from several medicinal plant compounds by molecular docking study. *Preprints* 2020:2020030226. doi: 10.20944/preprints202003.0226.v1
- Khalifa, I., Zhu, W., Mohammed, H., Dutta, K., and Li, C. (2020). Tannins inhibit SARS-CoV-2 through binding with catalytic dyad residues of 3CLpro: an *in silico* approach with 19 structural different hydrolysable tannins. *J. Food Biochem.* 44:e13432. doi: 10.1111/jfbc.13432
- Koka, V., Huang, X. R., Chung, A. C., Wang, W., Truong, L. D., and Lan, H. Y. (2008). Angiotensin II up-regulates angiotensin I-converting enzyme (ACE), but down-regulates ACE2 via the AT1-ERK/p38 MAP kinase pathway. *Am. J. Pathol.* 172, 1174–1183. doi: 10.2353/ajpath.2008.070762
- Lacroix, C., de Wouters, T., and Chassard, C. (2015). Integrated multi-scale strategies to investigate nutritional compounds and their effect on the gut microbiota. *Curr. Opin. Biotechnol.* 32, 149–155. doi: 10.1016/j.copbio.2014.12.009
- Landete, J. (2011). Ellagitannins, ellagic acid and their derived metabolites: a review about source, metabolism, functions and health. *Food Res. Int.* 44, 1150–1160. doi: 10.1016/j.foodres.2011.04.027
- Lin, C. J., Lin, H. J., Chen, T. H., Hsu, Y. A., Liu, C. S., Hwang, G. Y., et al. (2015). Correction: polygonum cuspidatum and its active components inhibit replication of the influenza virus through toll-like receptor 9-induced interferon beta expression. *PLoS ONE* 10:e0125288. doi: 10.1371/journal.pone.0125288
- Lin, L. T., Chen, T. Y., Chung, C. Y., Noyce, R. S., Grindley, T. B., McCormick, C., et al. (2011). Hydrolyzable tannins (chebulagic acid and punicalagin) target viral glycoprotein-glycosaminoglycan interactions to inhibit herpes simplex virus 1 entry and cell-to-cell spread. *J. Virol.* 85, 4386–4398. doi: 10.1128/JVI.01492-10
- Lu, J., Ding, K., and Yuan, Q. (2008). Determination of punicalagin isomers in pomegranate husk. *Chroma* 68, 303–306. doi: 10.1365/s10337-008-0699-y
- Majumder, R., and Mandal, M. (2020). Screening of plant-based natural compounds as a potential COVID-19 main protease inhibitor: an *in silico* docking and molecular dynamics simulation approach. *J. Biomol. Struct. Dyn.* 8, 1–16. doi: 10.1080/07391102.2020.1817787
- Malviya, S., Arvind, J. A., and Hettiarachchy, N. (2014). Antioxidant and antibacterial potential of pomegranate peel extracts. *J. Food Sci. Technol.* 51, 4132–4137. doi: 10.1007/s13197-013-0956-4
- Marchi, L. B., Monteiro, A. R., Mikcha, J., Santos, A., Chinelatto, M., Marques, D., et al. (2015). Evaluation of antioxidant and antimicrobial capacity of pomegranate peel extract (*Punica granatum* L.) Under different drying temperatures. *Chem. Eng. Transac.* 44, 121–126. doi: 10.3303/CET1544021
- Mastrogiovanni, F., Mukhopadhyay, A., Lacetera, N., Ryan, M. T., Romani, A., Bernini, R., et al. (2019). Anti-inflammatory effects of pomegranate peel extracts on *in vitro* human intestinal Caco-2 cells and *ex vivo* porcine colonic tissue explants. *Nutrients* 11:548. doi: 10.3390/nu11030548
- Mercurio, F. A., Pirone, L., Di Natale, C., Marasco, D., Pedone, E. M., and Leone, M. (2018). Sam domain-based stapled peptides: structural analysis and interaction studies with the Sam domains from the EphA2 receptor and the lipid phosphatase Ship2. *Bioorg. Chem.* 80, 602–610. doi: 10.1016/j.bioorg.2018.07.013
- Moradi, M. T., Karimi, A., Shahrani, M., Hashemi, L., and Ghaffari-Gosheh, M. S. (2019). Anti-influenza virus activity and phenolic content of pomegranate (*Punica granatum* L.) peel extract and fractions. *Avicenna J. Med. Biotechnol.* 11, 285–291.
- Mostafa-Hedeab, G. (2020). ACE2 as drug target of COVID-19 virus treatment, simplified updated review. *Rep. Biochem. Mol. Biol.* 9, 97–105. doi: 10.29252/rbmb.9.1.97
- Nasri, H., Baradaran, A., Shirzad, H., and Rafeian-Kopaei, M. (2014). New concepts in nutraceuticals as alternative for pharmaceuticals. *Int. J. Prev. Med.* 5, 1487–1499.
- Nutan, M. M., Goel, T., Das, T., Malik, S., Suri, S., Rawat, A. K. S., et al. (2013). Ellagic acid and gallic acid from *Lagerstroemia speciosa* L. inhibit HIV-1 infection through inhibition of HIV-1 protease and reverse transcriptase activity. *Indian J. Med. Res.* 137, 540–548.
- Onder, G., Rezza, G., and Brusaferro, S. (2020). Case-fatality rate and characteristics of patients dying in relation to COVID-19 in Italy. *JAMA* 323, 1775–1776. doi: 10.1001/jama.2020.4683
- Oyelowo, O., Okafor, C., Ajibare, A., Ayomide, B., Dada, K., Adelakun, R., et al. (2019). Fatty acids in some cooking oils as agents of hormonal manipulation in a rat model of benign prostate cancer. *Nigerian J. Physiol. Sci.* 34, 69–75.
- Pascarella, G., Strumia, A., Piliego, C., Bruno, F., Del Buono, R., Costa, F., et al. (2020). COVID-19 diagnosis and management: a comprehensive review. *J. Intern. Med.* 288, 192–206. doi: 10.1111/joim.13091
- Pirone, L., Del Gatto, A., Di Gaetano, S., Saviano, M., Capasso, D., Zaccaro, L., et al. (2020). A multi-targeting approach to fight SARS-CoV-2 attachment. *Front. Mol. Biosci.* 7:186. doi: 10.3389/fmolb.2020.00186
- Rattanachitthawat, N., Pinkhien, T., Opanasopit, P., Ngawhirunpat, T., and Chanvorachote, P. (2019). Finasteride enhances stem cell signals of human dermal papilla cells. *In Vivo* 33, 1209–1220. doi: 10.21873/invivo.11592
- Re, R., Pellegrini, N., Proteggente, A., Pannala, A., Yang, M., and Rice-Evans, C. (1999). Antioxidant activity applying an improved ABTS radical cation decolorization assay. *Free Radic. Biol. Med.* 26, 1231–1237. doi: 10.1016/S0891-5849(98)00315-3
- Reddy, M. K., Gupta, S. K., Jacob, M. R., Khan, S. I., and Ferreira, D. (2007). Antioxidant, antimalarial and antimicrobial activities of tannin-rich fractions, ellagitannins and phenolic acids from *Punica granatum* L. *Planta Med.* 73, 461–467. doi: 10.1055/s-2007-967167
- Remuzzi, A., and Remuzzi, G. (2020). COVID-19 and Italy: what next? *Lancet* 395, 1225–1228. doi: 10.1016/S0140-6736(20)30627-9
- Sayed, A. M., Khattab, A. R., AboulMagd, A. M., Hassan, H. M., Rateb, M. E., Zaid, H., et al. (2020). Nature as a treasure trove of potential anti-SARS-CoV drug leads: a structural/mechanistic rationale. *RSC Adv.* 10, 19790–19802. doi: 10.1039/D0RA04199H
- Serrelli, G., and Deiana, M. (2019). *In vivo* formed metabolites of polyphenols and their biological efficacy. *Food Funct.* 10, 6999–7021. doi: 10.1039/C9FO01733J

- Shin, D., Mukherjee, R., Grewe, D., Bojkova, D., Baek, K., Bhattacharya, A., et al. (2020). Papain-like protease regulates SARS-CoV-2 viral spread and innate immunity. *Nature* 587, 657–662. doi: 10.1038/s41586-020-2601-5
- Singh, Y. D., Jena, B., Ningthoujam, R., Panda, S., Priyadarsini, P., Panda, M. K., et al. (2020). Potential bioactive molecules from natural products to combat against coronavirus. *Adv. Tradit. Med.* doi: 10.1007/s13596-020-00496-w. [Epub ahead of print].
- Sorrenti, V., Randazzo, C. L., Caggia, C., Ballistreri, G., Romeo, F. V., Fabroni, S., et al. (2019). Beneficial effects of pomegranate peel extract and probiotics on pre-adipocyte differentiation. *Front. Microbiol.* 10:660. doi: 10.3389/fmicb.2019.00660
- Surucic, R., Tubic, B., Stojiljkovic, M. P., Djuric, D. M., Travar, M., Grabez, M., et al. (2021). Computational study of pomegranate peel extract polyphenols as potential inhibitors of SARS-CoV-2 virus internalization. *Mol. Cell. Biochem.* 476, 1179–1193 doi: 10.1007/s11010-020-03981-7
- Tresoldi, I., Sangiuolo, C. F., Manzari, V., and Modesti, A. (2020). SARS-COV-2 and infectivity: possible increase in infectivity associated to integrin motif expression. *J. Med. Virol.* 92, 1741–1742. doi: 10.1002/jmv.25831
- Viuda-Martos, M., Fernández-López, J., and Pérez-Álvarez, J. (2010). Pomegranate and its many functional components as related to human health: a review. *Compr. Rev. Food Sci. Food Saf.* 9, 635–654. doi: 10.1111/j.1541-4337.2010.00131.x
- Weng, J. K. (2020). Plant solutions for the covid-19 pandemic and beyond: historical reflections and future perspectives. *Mol. Plant* 13, 803–807. doi: 10.1016/j.molp.2020.05.014
- Yang, L., Wen, K. S., Ruan, X., Zhao, Y. X., Wei, F., and Wang, Q. (2018). Response of plant secondary metabolites to environmental factors. *Molecules* 23:762. doi: 10.3390/molecules23040762
- Yiu, C. Y., Chen, S. Y., Chang, L. K., Chiu, Y. F., and Lin, T. P. (2010). Inhibitory effects of resveratrol on the Epstein-Barr virus lytic cycle. *Molecules* 15, 7115–7124. doi: 10.3390/molecules15107115
- Zang, N., Xie, X., Deng, Y., Wu, S., Wang, L., Peng, C., et al. (2011). Resveratrol-mediated gamma interferon reduction prevents airway inflammation and airway hyperresponsiveness in respiratory syncytial virus-infected immunocompromised mice. *J. Virol.* 85, 13061–13068. doi: 10.1128/JVI.05869-11
- Zhang, L., Li, Y., Gu, Z., Wang, Y., Shi, M., Ji, Y., et al. (2015). Resveratrol inhibits enterovirus 71 replication and pro-inflammatory cytokine secretion in rhabdomyosarcoma cells through blocking IKKs/NF- $\kappa$ B signaling pathway. *PLoS ONE* 10:e0116879. doi: 10.1371/journal.pone.0116879
- Zhang, Y., Geng, X., Tan, Y., Li, Q., Xu, C., Xu, J., et al. (2020). New understanding of the damage of SARS-CoV-2 infection outside the respiratory system. *Biomed. Pharmacother.* 127:110195. doi: 10.1016/j.biopha.2020.110195
- Zhu, N., Zhang, D., Wang, W., Li, X., Yang, B., Song, J., et al. (2020). A novel coronavirus from patients with pneumonia in China, 2019. *N. Engl. J. Med.* 382, 727–733. doi: 10.1056/NEJMoa2001017

**Conflict of Interest:** FA is an Administration Board Member of Vitalab srl.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The reviewer AC declared a shared affiliation with several of the authors, GG, IC, and PV to the handling editor at time of review.

Copyright © 2021 Tito, Colantuono, Pirone, Pedone, Intartaglia, Giamundo, Conte, Vitaglione and Apone. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Pharmacophore Modelling-Based Drug Repurposing Approaches for SARS-CoV-2 Therapeutics

Shailima Rampogu and Keun Woo Lee\*

Department of Bio and Medical Big Data (BK21 Four Program), Research Institute of Natural Science (RINS), Gyeongsang National University (GNU), Jinju, South Korea

## OPEN ACCESS

### Edited by:

Chandrabose Selvaraj,  
Alagappa University, India

### Reviewed by:

Gopinath Krishnasamy,  
University of Turku, Finland  
Konda Reddy Karnati,  
Bowie State University, United States  
Padmanabhan Anbazhagan,  
Experimental Drug Development  
Centre (EDDC), Singapore

### \*Correspondence:

Keun Woo Lee  
kwlee@gnu.ac.kr

### Specialty section:

This article was submitted to  
Theoretical and Computational  
Chemistry,  
a section of the journal  
Frontiers in Chemistry

**Received:** 01 December 2020

**Accepted:** 15 January 2021

**Published:** 10 May 2021

### Citation:

Rampogu S and Lee KW (2021)  
Pharmacophore Modelling-Based  
Drug Repurposing Approaches for  
SARS-CoV-2 Therapeutics.  
Front. Chem. 9:636362.  
doi: 10.3389/fchem.2021.636362

The recent outbreak of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has caused a devastating effect globally with no effective treatment. The swift strategy to find effective treatment against coronavirus disease 2019 (COVID-19) is to repurpose the approved drugs. In this pursuit, an exhaustive computational method has been used on the DrugBank compounds targeting nsp16/nsp10 complex (PDB code: 6W4H). A structure-based pharmacophore model was generated, and the selected model was escalated to screen DrugBank database, resulting in three compounds. These compounds were subjected to molecular docking studies at the protein-binding pocket employing the CDocker module available with the Discovery Studio v18. In order to discover potential candidate compounds, the co-crystallized compound S-adenosyl methionine (SAM) was used as the reference compound. Additionally, the compounds remdesivir and hydroxychloroquine were employed for comparative docking. The results have shown that the three compounds have demonstrated a higher dock score than the reference compounds and were upgraded to molecular dynamics simulation (MDS) studies. The MDS results demonstrated that the three compounds, framycetin, kanamycin, and tobramycin, are promising candidate compounds. They have represented a stable binding mode at the targets binding pocket with an average protein backbone root mean square deviation below 0.3 nm. Additionally, they have prompted the hydrogen bonds during the entire simulations, inferring that the compounds have occupied the active site firmly. Taken together, our findings propose framycetin, kanamycin, and tobramycin as potent putative inhibitors for COVID-19 therapeutics.

**Keywords:** SARS-CoV-2, novel coronavirus, COVID-19, drug repurposing, pharmacophore modelling

## INTRODUCTION

Coronaviruses are the responsible agents that cause respiratory tract infections in humans (Rothan and Byrareddy, 2020). The recent outbreak has caused a serious challenge in finding effective therapeutics (Cao et al., 2020). SARS-CoV-2 has originated from Coronaviridae family and is known as novel coronavirus (2019-nCoV) (Pal et al., 2020; Zhou et al., 2020). Being the largest group of viruses, Coronaviruses (CoVs) has four families, of which the *Coronavirinae* a subfamily of *Coronaviridae* family can be divided into  $\alpha$ ,  $\beta$ ,  $\gamma$ , and  $\delta$  (Woo et al., 2010; Fehr and Perlman, 2015). These viruses have a club-shaped protrusion on their surface, resembling a “crown” referring

to “corona” in *Latin* (Li, 2016; Chorba, 2020). COVID-19 is the seventh type of human infecting coronavirus (Andersen et al., 2020).

The SARS-CoV-2 genome is made up of nonstructural polyprotein, open reading frame (ORF)1a/b and about 30 kb in size. It is further cleaved proteolytically to generate 15/16 proteins, 4 structural and 5 accessory proteins (Chan et al., 2020; Sinha et al., 2020; Wu et al., 2020). When the spike protein of SARS-CoV-2 (COVID-19) interacts with ACE-2, the infection initiates aided by the enzyme TMPRSS2 protease (Hoffmann et al., 2020; Mousavizadeh and Ghasemi, 2020). Once the initiation of the infection happens, ORF1a and ORF1ab are translated and subsequently cleaved proteolytically to form functional proteins and are majorly involved in viral replication (Mousavizadeh and Ghasemi, 2020; Krichel et al., (2020)).

In order to escape the degradation process and to secure efficient translation and additionally to escape from the host innate immune system recognition (Encinar and Menendez, 2020), the RNA molecule of SARS-CoV-2 (COVID-19) is capped at their 5' end. One of the nonstructural proteins, nsp16, encodes the 2'-O-methyltransferase (2'-O-MTase) and executes the RNA cap modification, which requires nsp10 for activation. This unique feature of SARS-CoV-2 is not discovered in host cell or other viruses. The binding of nsp10 to nsp16 happens through a  $\sim 930 \text{ \AA}^2$  activation surface in nsp10 (Decroly et al., 2011) correspondingly, favoring the nsp16 interaction to capped RNA substrate and methyl donor S-adenosyl-L-methionine (SAM), thus stabilizing the SAM-binding pocket and elongating the capped RNA-binding groove (Encinar and Menendez, 2020). A recent finding has documented three binding sites for the nsp16/nsp10 complex, the SAM-binding site, the interface between the nsp16-nsp10, and the RNA-binding groove (Encinar and Menendez, 2020). In the current investigation, we have targeted the SAM-binding site for exploiting new leads to combat SARS-CoV-2, adapting several computational approaches.

With no specific medication till date, it is essential to develop and/or design new therapeutics to counter this pandemic. A host of drugs are under investigation, and the search for new compounds is underway. Here, we targeted the nonstructural protein, the nsp16/nsp10 complex, to identify an effective prospective drug employing the *drug repurposing* approach using several computational studies. Additionally, the use of the already-known antiviral drugs offers several advantages due its known safety profiles (Das et al., 2020).

## MATERIALS AND METHODS

### Selection of the Target Structure

The protein for the present study is the nsp16/nsp10 complexed with the co-crystallized ligand SAM bearing the PDB code 6W4H (Rosas-Lemus et al., 2020). The protein was prepared by removing the water molecules and the heteroatoms. Subsequently, the protein was minimized enabling the “*Minimize and Refine Protein*” protocol available with the Discovery Studio v18 (DS). The chain A of the protein was used, and the active site residues were marked around the SAM ligand at 12 Å radius.

### Pharmacophore Generation and Virtual Screening Process

The receptor ligand pharmacophore model was generated exploiting the key interacting features present between the target and SAM. Correspondingly, the *Receptor-Ligand Pharmacophore Generation* module was used retaining all the other parameters as default (Meslamani et al., 2012). The best pharmacophore model was chosen based upon the selectivity score and the key residue interactions and was upgraded to screen the small molecules.

### Preparation of the Small Molecule Dataset

The FDA-approved small molecule dataset was downloaded from the DrugBank (Wishart et al., 2018) in the .sdf format and was subsequently imported onto the DS. The small molecules were initially checked for the presence of any duplicates and were thereafter minimized employing the “*Full Minimization*” protocol accessible with the DS.

### Binding Affinity Studies

Molecular docking was conducted to evaluate the binding affinities between the target protein and the small molecules. Herein, the CDOCKER (Wu et al., 2003) program available with the DS was employed. Each ligand was allowed to generate 50 conformations, and the best pose was selected based upon the highest dock score (-CDOCKER interaction energy) from the largest cluster and the key residue interactions (Rampogu et al., 2018). The compound SAM was used as the reference compound, while remdesivir and hydroxychloroquine were used for comparative docking.

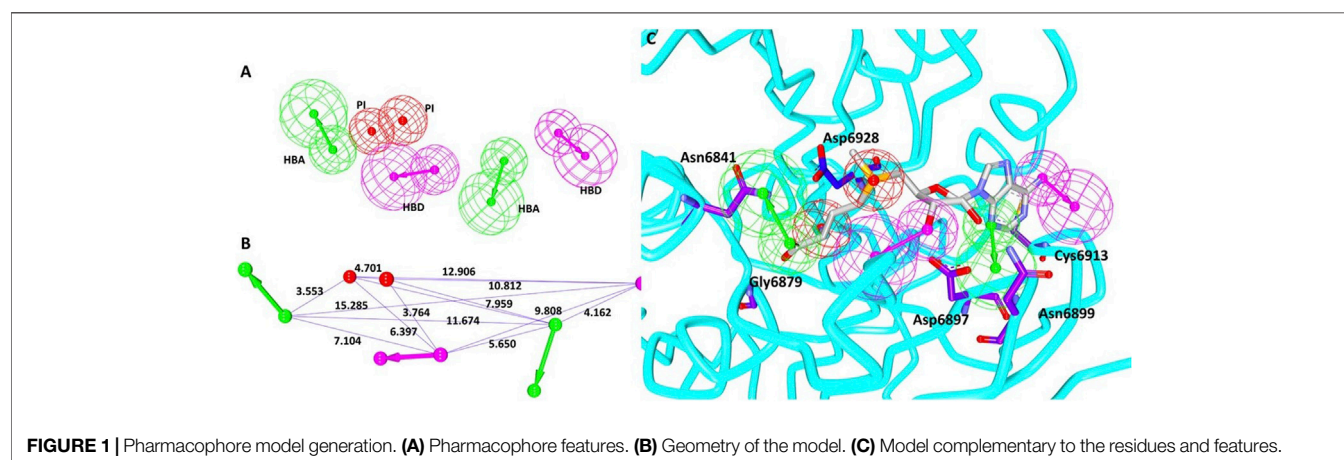
### Molecular Dynamics Simulation Studies

Molecular dynamics simulation (MDS) studies were undertaken to envisage on the behavior of the ligand at the binding pocket of the protein. The protein parameters were generated from the CHARMM 27 all-atom force field, and the ligand topologies were garnered from SwissParam (Zoete et al., 2011) employing the GROMACS v 2016.6 (Van Der Spoel et al., 2005). The systems were solvated in dodecahedron water box consisting of TIP3P water model and, subsequently, the counter ions were added. The systems were subsequently minimized followed by coupling of the protein and the ligand. This was proceeded by double equilibration method using the conserved number of particles (N), volume (V) and temperature (T) (NVT) and the constant number of particles (N), pressure (P), and temperature (T) (NPT) for 1 ns each. The NPT ensembles were escalated to MDS for 50 ns. All the analyses were carried out using visual molecular dynamics (VMD) (Humphrey et al., 1996) and DS. The results were evaluated according to the root mean square deviation (RMSD), root mean square fluctuations (RMSFs), radius of gyration (Rg), potential energy, number of hydrogen bonds, distance between the hydrogen bond interacting residues and ligand atoms, interaction energy, and the mode of ligand binding.

**TABLE 1** | Generation of different pharmacophore models and their characterization.

Pharmacophore model	Number of features	Feature set <sup>a</sup>	Selectivity score
Model 1	6	HBA, HBA, HBD, HBD, PI, PI	14.834
Model 2	6	HBA, HBA, HBD, HBD, PI, PI	14.834
Model 3	6	HBA, HBA, HBD, HBD, PI, PI	14.834
Model 4	6	HBA, HBA, HBD, HBD, PI, PI	14.834
Model 5	6	HBA, HBA, HBD, HBD, PI, PI	14.834
Model 6	6	HBA, HBA, HBD, HBD, PI, PI	14.834
Model 7	6	HBA, HBA, HBD, HBD, PI, PI	14.834
Model 8	6	HBA, HBA, HBD, HBD, PI, PI	14.834
Model 9	6	HBA, HBA, HBD, HBD, PI, PI	14.834
Model 10	6	HBA, HBA, HBD, HBD, PI, PI	14.834

<sup>a</sup>HBA refers to hydrogen bond acceptor, HBD refers to hydrogen bond donor, and PI indicates positive ionizable.



## RESULTS

### Pharmacophore Generation

Utilizing the interactions between the co-crystallized ligand SAM and the nsp16 target protein, the structure-based pharmacophore model was generated. The results have generated 10 pharmacophore models with the same features and selectivity score as shown in **Table 1**. When the receptor–ligand generation protocol was initiated, a total of 37 features in ligand were observed. However, 10 features have matched the receptor–ligand interactions: AAAAAADDPP [where A refers to hydrogen bond acceptor (HBA), D refers to hydrogen bond donor (HBD), and P indicates positive ionizable (PI)], out of which six features were incorporated in each model. For the scoring method, the *Rules* was used, which is an internal scoring method. This scoring function is based on a genetic function approximation (GFA model), which is a function of the feature set in the pharmacophore model and the feature–feature distances of different types of features. In order to select the ideal model, the pharmacophore that covers the entire structure of the SAM along with the key residue interactions was selected. Accordingly, model 3 was selected with the following features: 2 hydrogen bond acceptors (HBAs), 2 hydrogen bond donors (HBDs), and 2 positive ionizables (PIs), as shown in **Figure 1A** with its

geometry in **Figure 1B**. These features were complementary to the key residues such as Asn6841, Asp6897, Asn6899, Cys6913, and Asp6928, respectively, as illustrated in **Figure 1C**. We then evaluated if the model could retrieve potential inhibitors. Since there are not many compounds specific to nsp16/nsp10 complex, the pan-MTase inhibitor sinefungin was used. This compound is bound to SARS-CoV-2 nsp10–nsp16 methyltransferase complex bearing the PDB code 6YZ1 (Krafchikova et al., 2020). Upon enabling the *Ligand Pharmacophore Mapping* tool embedded with the DS, it was revealed that the compound has mapped perfectly with the model, **Supplementary Figure S1** suggesting that the model could be upgraded for retrieving new potential inhibitors.

### Pharmacophore Mapping to Discover New Indications for FDA-Approved Small Molecules

The prepared FDA-approved drugs were mapped with the pharmacophore model enabling the *Ligand Pharmacophore Mapping* tool, which resulted in three compounds, as shown in **Supplementary Figure S2**, inferring that these compounds possess the important features vital for inhibition. The resultant compounds were upgraded to binding affinity studies employing the CDocker program available with DS.

**TABLE 2** | Binding affinity scores according to the CDOCKER.

Compound name	DrugBank ID	-CDOCKER interaction energy (kcal/mol)
Framycetin	DB00452	68.80
Kanamycin	DB01172	62.18
Tobramycin	DB00684	58.40
S-Adenosylmethionine	Reference compound	64.26
Remdesivir		57.16
Hydroxychloroquine		49.13

## Binding Affinity Studies

The molecular docking was undertaken to predict the binding modes of the small molecules and further to estimate their binding affinities (Rampogu et al., 2020). Upon subjecting the obtained compounds to molecular docking mechanism along with the reference compound, three compounds were found to generate a higher to comparable dock score than the reference compound, as displayed in **Table 2**. Therefore, these compounds were upgraded to MDS studies to evaluate their stabilities.

## Molecular Dynamics Simulation Studies

Molecular dynamics simulation study was carried out to understand the behavior of small molecules at the binding pocket of the target. The results were delineated based upon the root mean square deviation (RMSD), radius of gyration (Rg), potential energy, root mean square fluctuation (RMSF), mode of ligand binding, number of hydrogen bond analysis, and the interaction energy.

## MDS Inferred Stability Analysis

### Root Mean Square Deviation

Protein stability with respect to its conformation can be measured *via* the deviations generated, if any, during the simulation period,

logically inferring that the smaller the deviation the greater the stability of the protein (Aier et al., 2016). Correspondingly, the RMSD plots were plotted for the protein backbone atoms to examine if there are any variations in the plot. The RMSD plot was found to be stable below 0.23 nm for framycetin, an average of 0.18 nm for tobramycin, and an average of 0.25 nm for kanamycin. These findings elucidate that the system is stable, demonstrating no major aberrations, as described in **Figure 2A**.

### Radius of Gyration

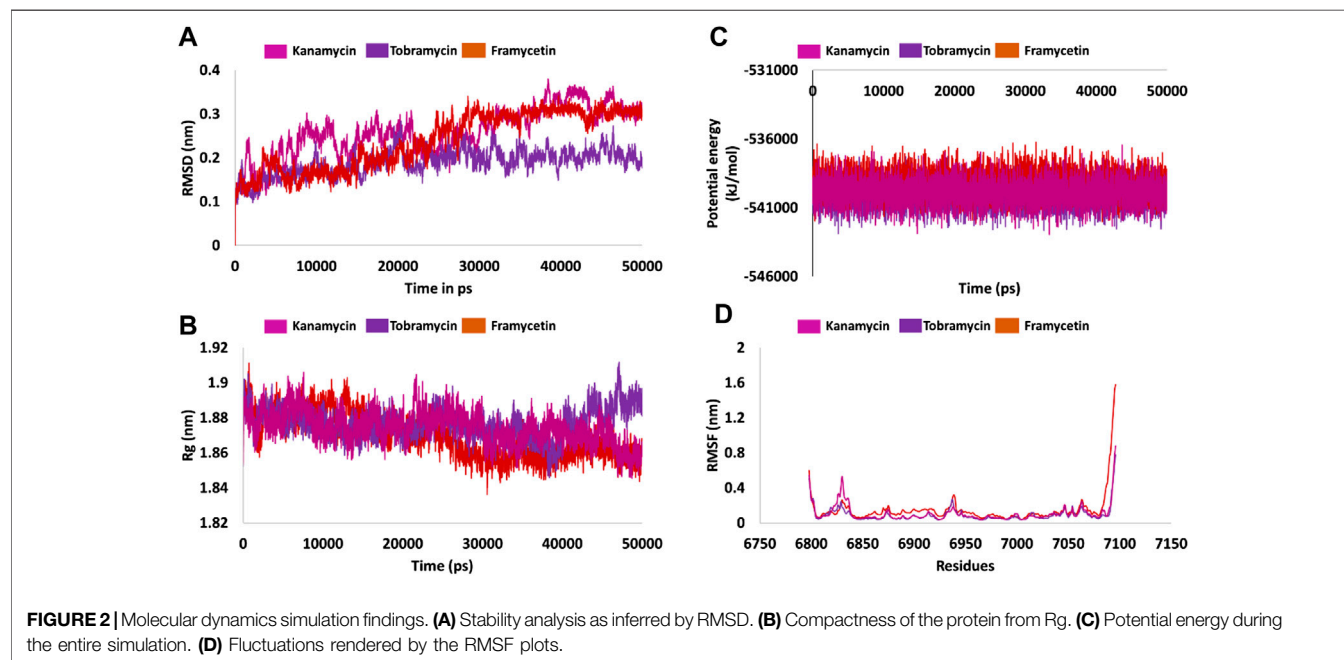
Radius of gyration serves as an indicator for protein structure compactness, further stating that the three-state folding of the proteins is relatively more compact than the proteins of the same size but in two-state folding (Lobanov et al., 2008). In the current study, we investigated the compactness of the protein (backbone) during the whole simulation. It was noted that the protein was highly stable ranging between 1.86 and 1.90 nm. Interestingly, the Rg of the three systems was observed to drop at 25 ns thereafter being stable. However, in case of tobramycin, the Rg has raised marginally at 40 ns and remained stable thereafter. Overall, the three systems have revealed a greater compactness as shown in **Figure 2B**.

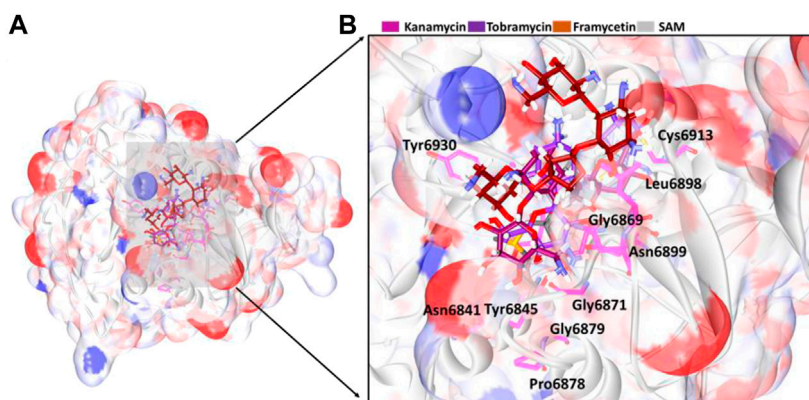
### Potential Energy

Another parameter, that is, the potential energy, was also calculated. The three systems have displayed a stable potential energy between  $-542000$  and  $-537000$  kJ/mol. The average potential energy was calculated as  $-539864$  kJ/mol for kanamycin,  $-539133$  kJ/mol for framycetin, and  $-540066$  kJ/mol for tobramycin. These results additionally strengthen the argument that the systems were stable with no major variations, as shown in **Figure 2C**.

### Root Mean Square Fluctuation

Root mean square fluctuation (RMSF) is defined as a measure of residue-specific flexibility (Dong et al., 2018). Here, the RMSF was





**FIGURE 3 |** Binding mode analysis of the compounds. **(A)** Accommodation of ligands at the binding pocket and **(B)** its enlarged view surrounded by the binding pocket residues.

conducted on the protein backbone of the three systems. It was observed that there were no fluctuations noticed in the backbone residues, as shown in **Figure 2D**. The binding pocket is made up of the residues Asn6841, Tyr6845, Gly6879, Gly6871, Pro6878, Gly6879, Asp6897, Leu6898, Asn6899, Asp6912, Cys6913, Asp6928, Met6929, and Tyr6930. Interestingly, no significant fluctuations were noticed with these residues as described in **Supplementary Figure S3A**. The residues between 6,820–6,830 and 6,932–6,942 were observed to demonstrate a marginal surge in the RMSF plots. However, these residues do not belong to the active site as shown in **Figure 2D** and **Supplementary Figure S3B**. The putative inhibitors were accommodated at the binding pocket by interacting with the key residues. The detailed computational analysis has unraveled their binding potential at the atomistic level. Furthermore, mechanistically, we speculate that binding with these key residues could illuminate the credibility of the compounds as new inhibitors.

## Binding Mode Analysis

From the stable RMSD, the last 5 ns structure was extracted and was superimposed against the X-ray crystal structure. The results have revealed that the compounds have occupied the same binding pocket as that of the co-crystallized compound SAM, as shown in **Figure 3**. The intermolecular interactions have shown that the compound framycetin has formed four hydrogen bond interactions with residues Asn6899, Tyr6930, and Asp6931. The residue Asn6899 has generated two hydrogen bonds as illustrated in **Figure 4A**. Additionally, the residues, Leu6898, Met6929, and Lys6933 have formed the van der Waals interactions firmly holding the compound at the binding pocket as illustrated in **Supplementary Figure S5**.

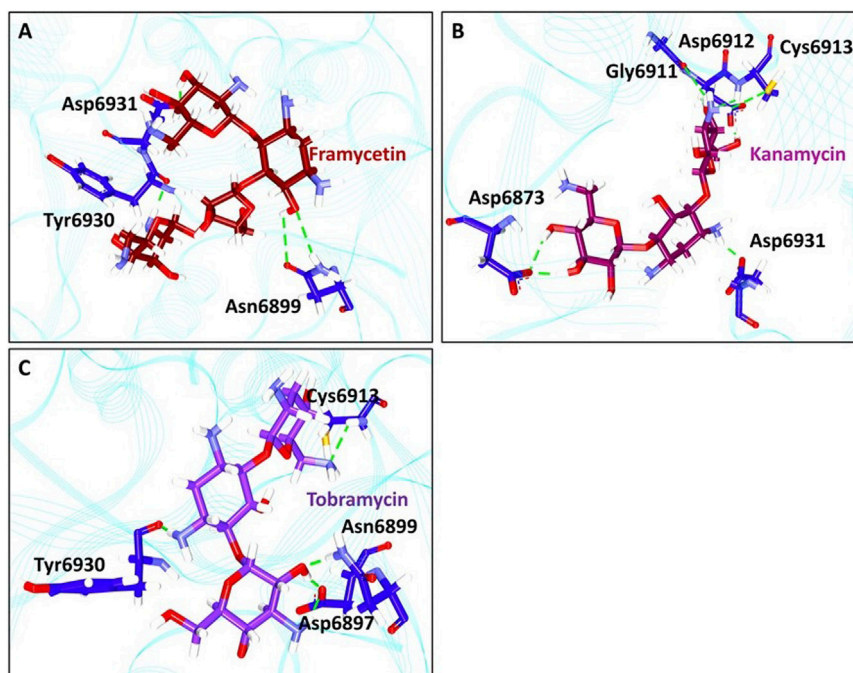
The compound kanamycin has occupied the similar binding pocket as that of the co-crystallized compound SAM, as shown in **Figure 3**. The intermolecular interactions have shown that the compound has formed hydrogen bond interactions with residues Asp6873, Gly6911, Asp6912, Cys6913, and Asp6931, respectively. The residues Asp6873 and Cys6913 have generated two hydrogen bonds as illustrated in **Figure 4B**. Additionally, the compounds Phe6868, Ser6896, Met6929, Gly6869, Asp6897, Ser6872, Gly6871, Pro6932, Lys6933, and Leu6898 have formed the van der Waals

interactions firmly holding the compound at the binding pocket, as illustrated in **Supplementary Figure S5**.

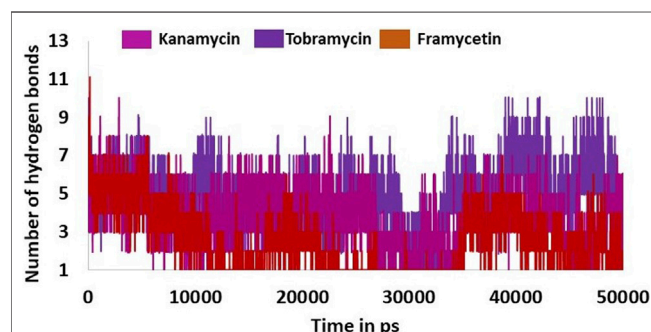
The compound tobramycin has occupied the binding pocket in the same fashion as that of the co-crystallized compound SAM, as shown in **Figure 3**. The intermolecular interactions have shown that the compound has formed hydrogen bond with residues Asn6899, Asp6897, Cys6913, and Tyr6930, respectively. The residue Asp6897 has generated two hydrogen bonds as illustrated in **Figure 4C**. Additionally, the compounds Ala6914, Asp6912, Tyr6950, Phe6947, Lys6933, Asp6931, Pro6932, Asp6873, Ser6872, Gly6871, and Gly6911 have formed the van der Waals interactions accommodating the compound at the binding pocket, as illustrated in **Supplementary Figure S5**.

## Hydrogen Bond Interactions and Distance Between the Interacting Residues and Atoms

The hydrogen bond interactions were monitored during the 50 ns simulation run. It was noted that the hydrogen bonds were formed during the entire simulations, stating that the compounds were present within the active site of the protein. The average hydrogen bonds were computed to be 2.5, 3.8, and 4.8 for framycetin, kanamycin, and tobramycin, respectively, as displayed in **Figure 5**. Additionally, we meticulously investigated the distance between the interacting residues that are making the hydrogen bonds and the ligand atoms. The compound framycetin forms four hydrogen bonds noticed to be interacting throughout the simulation run with an acceptable average distance below 0.3 nm, as shown in **Figure 6**. The seven hydrogen bonds demonstrated by the compound kanamycin have displayed an acceptable average distance of 0.3 nm. A distance beyond 0.3 nm was noticed with Asp6912\_OD2:H49, Cys6913\_SH:H48 during the initial simulation run, thereafter demonstrating an acceptable bond length. The interaction, Asp6931\_OD2:H52 showed longer distance that ranged from 0.3 nm to 0.9 nm until 30 ns, after which the distance has settled at an acceptable bond length projecting an average distance of 0.36 nm as described in **Figure 7**. The compound tobramycin has



**FIGURE 4 |** Intermolecular hydrogen bond interactions. **(A)** Molecular interactions between framycetin and the key residues. **(B)** Hydrogen bond interactions between kanamycin and the protein residues. **(C)** Intermolecular hydrogen bond interactions between tobramycin and the protein residues.



**FIGURE 5 |** Hydrogen bond interactions between the protein and ligand during the whole simulations.

shown relatively high degree of variation in the distance measure, while maintaining the average distance below 0.3 nm. The interactions with residue Cys6913 and Asn6899 were highly rigid, illuminating their strong affinity toward the ligand, as described in **Figure 8**. Notably, the residue Asp has displayed high degree of variation, whilst maintaining the distance below 0.3 nm, as described in **Figure 8**.

## Interaction Energy Analysis

Furthermore, to quantify the strength of the interaction between the protein and the small molecule, the interaction energies were computed and read according to the terms Coul-SR:protein-lig and LJ-SR:protein-lig. For framycetin, the Coul-SR: protein-lig interaction energy existed between  $-400$  and  $-100$  kJ/mol with an average of  $-178.28$  kJ/mol. The LJ-SR:protein-lig interaction energy

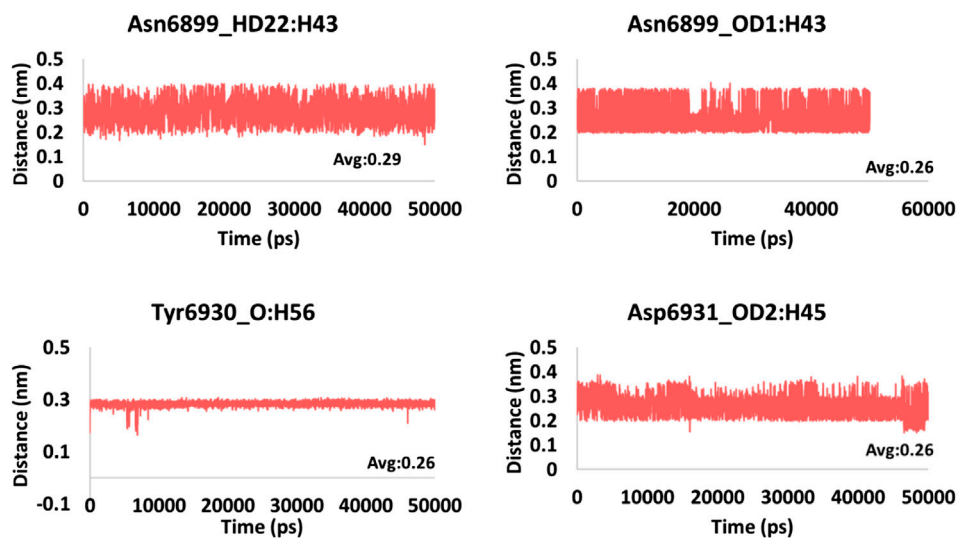
was  $-200$  to  $-50$  kJ/mol with an average of  $-121.35$  kJ/mol. For kanamycin, the Coul-SR:protein-lig interaction energy existed between  $-400$  and  $-100$  kJ/mol with an average of  $-167.46$  kJ/mol. The LJ-SR:protein-lig interaction energy was  $-100$  to  $-200$  kJ/mol with an average of  $-127.43$  kJ/mol. For tobramycin, the Coul-SR:protein-lig interaction energy existed between  $-300$  and  $-100$  kJ/mol with an average of  $-158.01$  kJ/mol. The LJ-SR: protein-lig interaction energy was  $-200$  to  $-100$  kJ/mol with an average of  $-121.15$  kJ/mol, as illustrated in **Figures 9A,B**.

## DISCUSSION

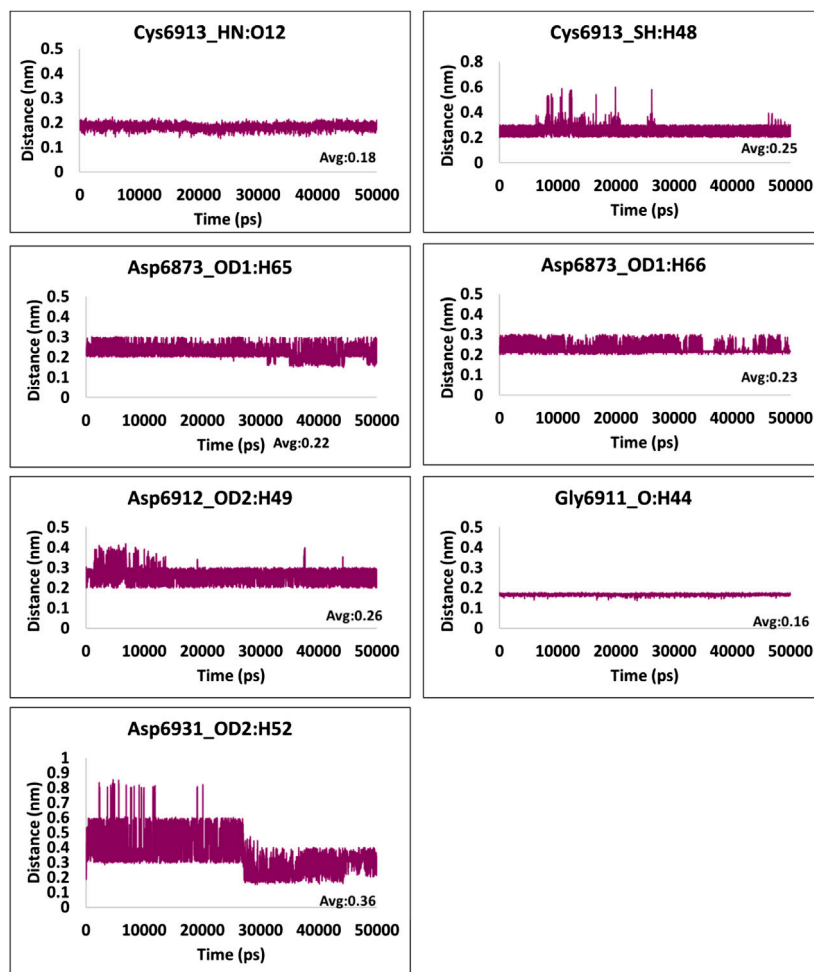
COVID-19 is a global pandemic (Osier et al., 2020) that the world has been witnessing with no effective medication yet. The World Health Organization (WHO) has stated that COVID-19 is “public enemy number 1” and potentially more powerful than terrorism (Yi et al., 2020). To find promising therapeutics, in the current research, we have adapted computational drug repurposing approaches.

Drug repurposing (drug repositioning, reprofiling, or re-tasking) is a method for discovering new uses for approved or investigational drugs different from the original use, and has several advantages such as reduced risk of failure, reduced time frame for drug development, and less investment (Pushpakom et al., 2018).

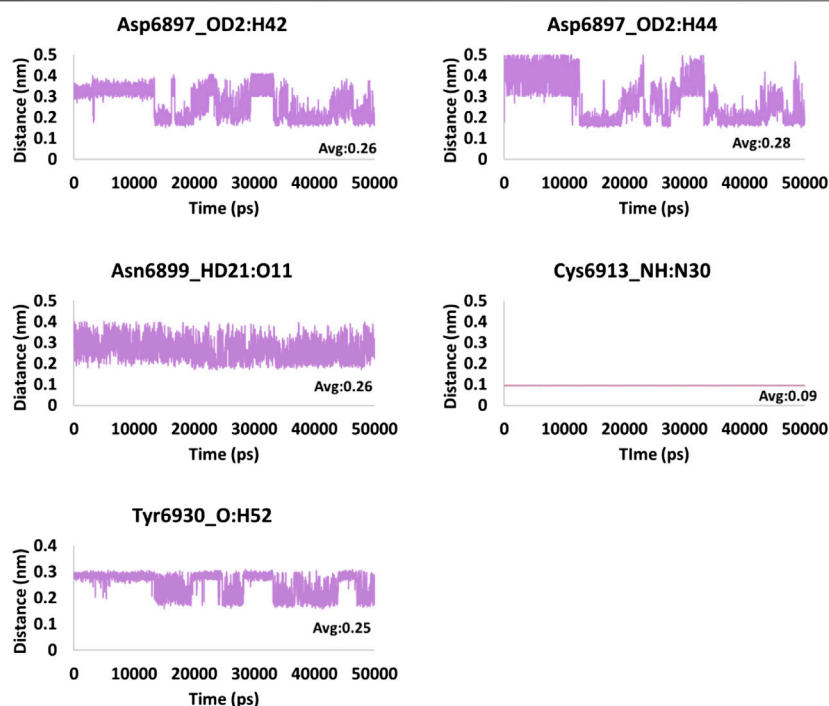
A few reports exist targeting the SAM binding site. Encinar et al. identified twelve drugs that have occupied the SAM binding site with a high affinity (Encinar and Menendez, 2020). In another study, Tazikeh-Lemeski et al. reported the drugs raltegravir and maraviroc, to be effective against nsp 16 (Tazikeh-Lemeski et al., 2020). Bilal et al. discovered



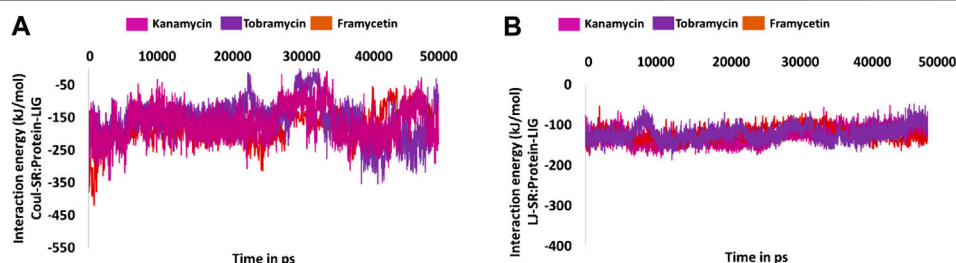
**FIGURE 6 |** Hydrogen bond distance of the interacting residues (atoms) and the ligand atoms of framycetin.



**FIGURE 7 |** Hydrogen bond distance of the interacting residues (atoms) and the ligand atoms of kanamycin.



**FIGURE 8 |** Hydrogen bond distance of the interacting residues (atoms) and the ligand atoms of tobramycin.



**FIGURE 9 |** Molecular dynamics simulation-guided interaction energy during the whole simulation. **(A)** Interaction energy from Coul-SR:Protein-lig and **(B)** interaction energy from LJ-SR:Protein-lig.

naphthyridine and quinoline derivatives as possible nsp16-nsp10 inhibitors (Aldahham et al., 2020). However, these computational investigations have not reported our compounds.

Herein, we have used pharmacophore-based molecular docking and molecular dynamics simulation methods to discover potential candidates for SARS-CoV-2, targeting the 6W4H. Upon performing the structural alignment using the protocol *Align Structures* of SARS-CoV (3R24) and SARS-CoV-2 (6W4H), both the sequences have shown a sequence identity of 86.8% and sequence similarity of 89.4%, as shown in **Supplementary Figure S4B**. Additionally, upon superimposition of the two structures, it was revealed that the SAM binding site was similar, inferring that the two structures share high similarity, as shown in **Supplementary Figure S4A**.

Our investigation has retrieved framycetin, kanamycin, and tobramycin as potential candidate drugs as they have demonstrated higher dock score than the reference compound. In order to understand their affinities, in case to avoid false positives and to elucidate the behavior of these compounds at the active site of the target, the MDS analysis was undertaken. The results have shown that the three compounds have projected acceptable results, whilst maintaining the key residue interactions with the target.

The antibiotic framycetin is generally employed to treat bacterial eye infections (Wishart et al., 2006; Wishart et al., 2018). The compound framycetin has formed hydrogen bonds with residues Asn6899, Tyr6930, and Asp6931. The residue Asn6899 has generated a hydrogen bond with remdesivir, while demonstrating a van der Waals interaction with

hydroxychloroquine. The residues Tyr6930 and Asp6931 have prompted a van der Waals interaction with the reference compounds. The compound framycetin was also found to be a potential inhibitor against SARS-CoV-2 M<sup>Pro</sup> (Rampogu and Lee, 2021). Kanamycin is an aminoglycoside antibiotic and acts by binding to the bacterial 30S ribosomal subunit, leading to misread t-RNA, thereby leaving the bacterium unable to synthesize proteins vital to its growth (Wishart et al., 2006; Wishart et al., 2018). The compound kanamycin has established hydrogen bond interactions with residues Asp6873, Gly6911, Asp6912, Cys6913, and Asp6931, respectively. The residue Asp6873 has formed van der Waals interaction with remdesivir and hydroxychloroquine. The residue Gly6911 has generated a van der Waals interaction with the docked pose of SAM and the cocrystallized structure. The residue Asp6912 has generated hydrogen bond interaction with the docked pose of SAM and the co-crystallized structure. The residue Cys6913 has formed van der Waals interaction with remdesivir and  $\pi$ -alkyl interaction with hydroxychloroquine. Interestingly, this residue has interacted with the hydrogen bond interaction in both the docked pose of SAM and the co-crystallized structure. The residue Asp6931 has formed van der Waals interaction with remdesivir and the docked pose of SAM. The compound tobramycin is a broad-spectrum aminoglycoside, antibiotic produced by *Streptomyces tenebrarius* (Wishart et al., 2006; Wishart et al., 2018). It is effective against Gram-negative bacteria, especially the *pseudomonas* species. The compound tobramycin has formed five hydrogen bonds. The residue Asp6897 has formed hydrogen bond with tobramycin as was noticed with all the reference compounds. The residue Asn6899 has formed hydrogen bond interaction with tobramycin, remdesivir, and co-crystallized ligand, while it formed van der Waals interactions with remdesivir and the docked pose of SAM. The residue Cys6913 formed hydrogen bond interactions with tobramycin, docked pose of SAM, and the cocrystallized structure. It has generated a van der Waals interaction with remdesivir. The residue Tyr6930 represented a hydrogen bond interaction with tobramycin and van der Waals interaction with hydroxychloroquine. The residue has demonstrated a carbon-hydrogen bond with remdesivir and docked pose of SAM.

Furthermore, our investigations have shown that the three compounds have been accommodated within the binding pocket during the entire simulation run without any significant variations, as predicted by RMSF profiles in **Supplementary Figure S3**.

Additionally, the distance plots of the hydrogen bonds determine the strength of the bonds inferring that the compounds are held strongly at the active site of the target protein. Furthermore, they were clamped by several residues at the binding pocket locking them at the active site throughout the simulation via van der Waals interaction, as shown in **Supplementary Figure S5**. Since these are the key residues to

bring out the biological processes as noticed with the cocrystallized ligand, we speculate that our newly identified compounds could serve as effective inhibitors. Mechanistically, interacting with these residues is essential to bring about the desirable result. These compounds have also demonstrated the pharmacophore features inferring the key features as that of the cocrystallized ligand. Taken together, we propose these compounds as potential leads targeting nsp16 protein to combat SARS-CoV-2.

## CONCLUSION

SARS-CoV-2 causing COVID-19 is the recent pandemic the world is fighting currently with no effective therapeutics yet. In pursuit of finding effective drugs to this disease, we have performed exhaustive pharmacophore-based drug repurposing approach to discover candidate compounds from DrugBank. Our results have retrieved three potential candidates, which have shown promising computational results. These compounds have displayed the pharmacophore features as possessed by the cocrystallized compound SAM complemented by the key residue interactions and stable MDS results. Together, we propose three compounds, framycetin, kanamycin, and tobramycin, as promising therapeutics for SARS-CoV-2 infections.

## DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**; further inquiries can be directed to the corresponding author.

## AUTHOR CONTRIBUTIONS

SR and KWL conceived the idea. SR wrote the manuscript. SR and KWL approved and validated the manuscript.

## FUNDING

This research was supported by the Bio & Medical Technology Development Program of the National Research Foundation (NRF) and funded by the Korean government (MSIT) (no. NRF-2018M3A9A7057263).

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fchem.2021.636362/full#supplementary-material>

## REFERENCES

- Aier, I., Varadwaj, P. K., and Raj, U. (2016). Structural insights into conformational stability of both wild-type and mutant EZH2 receptor. *Sci. Rep.* 6, 34984. doi:10.1038/srep34984
- Aldahham, B. J. M., Al-Khafaji, K., Saleh, M. Y., Abdelhakem, A. M., Alanazi, A. M., and Islam, M. A. (2020). Identification of naphthyridine and quinoline derivatives as potential Nsp16-Nsp10 inhibitors: a pharmacoinformatics study. *J. Biomol. Struct. Dyn.*, 1–8. doi:10.1080/07391102.2020.1851305
- Andersen, K. G., Rambaut, A., Lipkin, W. I., Holmes, E. C., and Garry, R. F. (2020). The proximal origin of SARS-CoV-2. *Nat. Med.* 26, 450. doi:10.1038/s41591-020-0820-9
- Cao, B., Wang, Y., Wen, D., Liu, W., Wang, J., Fan, G., et al. (2020). A trial of lopinavir–ritonavir in adults hospitalized with Severe covid-19. *N. Engl. J. Med.* 382, 1787–1799. doi:10.1056/nejmoa2001282
- Chan, J. F.-W., Kok, K.-H., Zhu, Z., Chu, H., To, K. K.-W., Yuan, S., et al. (2020). Genomic characterization of the 2019 novel human-pathogenic coronavirus isolated from a patient with atypical pneumonia after visiting Wuhan. *Emerg. microbes AND infect.* 9, 221–236. doi:10.1080/22221751.2020.1719902
- Chorba, T., and Perlman, S. (2020). The concept of the crown and its potential role in the downfall of coronavirus. *Emerg. Infect. Dis.* 1282, doi:10.3201/eid2609.ac2609
- Das, S., Sarmah, S., Lyndem, S., and Singha Roy, A. (2020). An investigation into the identification of potential inhibitors of SARS-CoV-2 main protease using molecular docking study. *J. Biomol. Struct. Dyn.*, 1–11. doi:10.1080/07391102.2020.1763201
- Decroly, E., Debarnot, C., Ferron, F., Bouvet, M., Coutard, B., Imbert, I., et al. (2011). Crystal structure and functional analysis of the SARS-coronavirus RNA cap 2'-O-methyltransferase nsp10/nsp16 complex. *PLoS Pathog.* 7, e1002059. doi:10.1371/journal.ppat.1002059
- Dong, Y.-w., Liao, M.-l., Meng, X.-l., and Somero, G. N. (2018). Structural flexibility and protein adaptation to temperature: molecular dynamics analysis of malate dehydrogenases of marine molluscs. *Proc. Natl. Acad. Sci. USA* 115, 1274. doi:10.1073/pnas.1718910115
- Encinar, J. A., and Menendez, J. A. (2020). Potential drugs targeting early innate immune evasion of SARS-coronavirus 2 via 2'-O-methylation of viral RNA. *Viruses* 12, 525. doi:10.3390/v12050525
- Fehr, A. R., and Perlman, S. (2015). Coronaviruses: an overview of their replication and pathogenesis. *Methods Mol. Biol.* 1282, 1–23. doi:10.1007/978-1-4939-2438-7\_1
- Hoffmann, M., Kleine-Weber, H., Schroeder, S., Krüger, N., Herrler, T., Erichsen, S., et al. (2020). SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. *Cell* 181, 271. doi:10.1016/j.cell.2020.02.052
- Humphrey, W., Dalke, A., and Schulten, K. (1996). VMD: visual molecular dynamics. *J. Mol. Graph.* 14, 33. doi:10.1016/0263-7855(96)00018-5
- Krafčikova, P., Silhan, J., Nencka, R., and Boura, E. (2020). Structural analysis of the SARS-CoV-2 methyltransferase complex involved in RNA cap creation bound to sinefungin. *Nat. Commun.* 11, 3717. doi:10.1038/s41467-020-17495-9
- Krichel, B., Falke, S., Hilgenfeld, R., Redecke, L., and Uetrecht, C. (2020). Processing of the SARS-CoV pp1a/ab nsp7-10 region. *Biochem. J.* 477, 1009–1019. doi:10.1042/BCJ20200029
- Li, F. (2016). Structure, function, and evolution of coronavirus spike proteins. *Annu. Rev. Virol.* doi:10.1146/annurev-virology-110615-042301
- Lobanov, M. Y., Bogatyreva, N. S., and Galzitskaya, O. V. (2008). Radius of gyration as an indicator of protein structure compactness. *Mol. Biol.* 42, 623. doi:10.1134/S0026893308040195
- Meslamani, J., Li, J., Sutter, J., Stevens, A., Bertrand, H. O., and Rognan, D. (2012). Protein-ligand-based pharmacophores: generation and utility assessment in computational ligand profiling. *J. Chem. Inf. Model.* 52, 943. doi:10.1021/ci300083r
- Mousavizadeh, L., and Ghasemi, S. (2020). Genotype and phenotype of COVID-19: their roles in pathogenesis. *J. Microbiol. Immunol. Infect.*, 30082. doi:10.1016/j.jmii.2020.03.022
- Osier, F., Ting, J. P. Y., Fraser, J., Lambrecht, B. N., Romano, M., Gazzinelli, R. T., et al. (2020). The global response to the COVID-19 pandemic: how have immunology societies contributed? *Nat. Rev. Immunol.* 20, 594. doi:10.1038/s41577-020-00428-4
- Pal, M., Berhanu, G., Desalegn, C., and Kandi, V. (2020). Severe Acute respiratory Syndrome coronavirus-2 (SARS-CoV-2): an update. *Cureus* 12, e7423. doi:10.7759/cureus.7423
- Pushpakom, S., Iorio, F., Eyers, P. A., Escott, K. J., Hopper, S., Wells, A., et al. (2018). Drug repurposing: progress, challenges and recommendations. *Nat. Rev. Drug Discov.* 18, 41. doi:10.1038/nrd.2018.168
- Rampogu, S., Kim, S. M., Son, M., Baek, A., Park, C., Lee, G., et al. (2020). A computational approach with biological evaluation: combinatorial treatment of curcumin and exemestane synergistically regulates ddx3 expression in cancer cell lines. *Biomolecules* 10, 857. doi:10.3390/biom10060857
- Rampogu, S., and Lee, K. W. (2021). Old drugs for new purpose—fast pace therapeutic identification for SARS-CoV-2 infections by pharmacophore guided drug repositioning approach. *Bull. Korean Chem. Soc.* 42, 212–226. doi:10.1002/bkcs.12171
- Rampogu, S., Park, C., Son, M., Baek, A., Zeb, A., Lee, G., et al. (2019). Modulation of aromatase by natural compounds—a pharmacophore guided molecular modelling simulations. *South Afr. J. Bot.* 120, 230. doi:10.1016/j.sajb.2018.06.019
- Rosas-Lemus, M., Minasov, G., Shuvalova, L., Inniss, N. L., Kiryukhina, O., Wiersum, G., et al. (2020). The crystal structure of nsp10-nsp16 heterodimer from SARS-CoV-2 in complex with S-adenosylmethionine. *bioRxiv* 4, 047498. doi:10.1101/2020.04.17.047498
- Rothan, H. A., and Byrareddy, S. N. (2020). The epidemiology and pathogenesis of coronavirus disease (COVID-19) outbreak. *J. Autoimmun.* 109, 102433. doi:10.1016/j.jaut.2020.102433
- Sinha, S. K., Shakya, A., Prasad, S. K., Singh, S., Gurav, N. S., Prasad, R. S., et al. (2020). An in-silico evaluation of different Saikosaponins for their potency against SARS-CoV-2 using NSP15 and fusion spike glycoprotein as targets. *J. Biomol. Struct. Dyn.* 1, 1–12. doi:10.1080/07391102.2020.1762741
- Tazikheh-Lemeski, E., Moradi, S., Raoufi, R., Shahlaei, M., Janlou, M. A. M., and Zolghadri, S. (2020). Targeting SARS-COV-2 non-structural protein 16: a virtual drug repurposing study. *J. Biomol. Struct. Dyn.*, 1–14. doi:10.1080/07391102.2020.1779133
- Van Der Spoel, D., Lindahl, E., Hess, B., Groenhof, G., Mark, A. E., and Berendsen, H. J. (2005). GROMACS: fast, flexible, and free. *J. Comput. Chem.* 26, 1701. doi:10.1002/jcc.20291
- Wishart, D. S., Knox, C., Guo, A. C., Shrivastava, S., Hassanali, M., Stothard, P., et al. (2006). DrugBank: a comprehensive resource for in silico drug discovery and exploration. *Nucleic Acids Res.* 34, D668. doi:10.1093/nar/gkj067
- Wishart, D. S., Feunang, Y. D., Guo, A. C., Lo, E. J., Marcu, A., Grant, J. R., et al. (2018). DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Res.* doi:10.1093/nar/gkx1037
- Woo, P. C. Y., Huang, Y., Lau, S. K. P., and Yuen, K. Y. (2010). Coronavirus genomics and bioinformatics analysis. *Viruses* 2, 1804–1820. doi:10.3390/v2081803
- Wu, A., Peng, Y., Huang, B., Ding, X., Wang, X., Niu, P., et al. (2020). Genome composition and divergence of the novel coronavirus (2019-nCoV) originating in China. *Cell Host Microbe* 27, 325–328. doi:10.1016/j.chom.2020.02.001
- Wu, G., Robertson, D. H., Brooks, C. L., and Vieth, M. (2003). Detailed analysis of grid-based molecular docking: a case study of CDOCKER-A CHARMM-based MD docking algorithm. *J. Comput. Chem.* 24, 1549. doi:10.1002/jcc.10306
- Yi, Y., Lagniton, P. N. P., Ye, S., Li, E., and Xu, R.-H. (2020). COVID-19: what has been learned and to be learned about the novel coronavirus disease. *Int. J. Biol. Sci.* 16, 1753. doi:10.7150/ijbs.45134
- Zhou, P., Yang, X.-L., Wang, X.-G., Wang, B., Zhang, L., Zhang, W., et al. (2020). A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* 579, 270. doi:10.1038/s41586-020-2012-7
- Zoete, V., Cuendet, M. A., Grosdidier, A., and Michielin, O. (2011). SwissParam: a fast force field generation tool for small organic molecules. *J. Comput. Chem.* 32, 2359. doi:10.1002/jcc.21816

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Rampogu and Lee. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

# Advantages of publishing in Frontiers



## OPEN ACCESS

Articles are free to read  
for greatest visibility  
and readership



## FAST PUBLICATION

Around 90 days  
from submission  
to decision



## HIGH QUALITY PEER-REVIEW

Rigorous, collaborative,  
and constructive  
peer-review



## TRANSPARENT PEER-REVIEW

Editors and reviewers  
acknowledged by name  
on published articles

## Frontiers

Avenue du Tribunal-Fédéral 34  
1005 Lausanne | Switzerland

Visit us: [www.frontiersin.org](http://www.frontiersin.org)

Contact us: [frontiersin.org/about/contact](http://frontiersin.org/about/contact)



## REPRODUCIBILITY OF RESEARCH

Support open data  
and methods to enhance  
research reproducibility



## DIGITAL PUBLISHING

Articles designed  
for optimal readership  
across devices



## FOLLOW US

@frontiersin



## IMPACT METRICS

Advanced article metrics  
track visibility across  
digital media



## EXTENSIVE PROMOTION

Marketing  
and promotion  
of impactful research



## LOOP RESEARCH NETWORK

Our network  
increases your  
article's readership