

COMPUTER-AIDED DRUG DESIGN: DRUG DISCOVERY, COMPUTATIONAL MODELLING, AND ARTIFICIAL INTELLIGENCE

EDITED BY: Fei Ye, Zhongjie Liang and Ylvain Broussy
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 978-2-88976-784-7

DOI 10.3389/978-2-88976-784-7

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

COMPUTER-AIDED DRUG DESIGN: DRUG DISCOVERY, COMPUTATIONAL MODELLING, AND ARTIFICIAL INTELLIGENCE

Topic Editors:

Fei Ye, Zhejiang Sci-Tech University, China

Zhongjie Liang, Soochow University, China

Ylvain Broussy, University Of Paris, France

Citation: Ye, F., Liang, Z., Broussy, Y., eds. (2022). Computer-Aided Drug Design: Drug Discovery, Computational Modelling, and Artificial Intelligence. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-88976-784-7

Table of Contents

- 05 Editorial: Computer-aided Drug Design: Drug Discovery, Computational Modelling, and Artificial Intelligence**
Fei Ye, Min Lin, Jia Jin and Sylvain Broussy
- 07 Discovery of New α -Glucosidase Inhibitors: Structure-Based Virtual Screening and Biological Evaluation**
Shan-Kui Liu, Haifang Hao, Yuan Bian, Yong-Xi Ge, Shengyuan Lu, Hong-Xu Xie, Kai-Ming Wang, Hongrui Tao, Chao Yuan, Juan Zhang, Jie Zhang, Cheng-Shi Jiang and Kongkai Zhu
- 16 Exploring the Interaction between Vancomycin/Teicoplanin and Receptor Binding Domain (RBD) of SARS-CoV-2**
Wen-Tao Tao, Qian Yu, Ya-Li Li, Mei Ge, Yi-Lei Zhao and Ting Shi
- 26 Revisiting Activity of Some Nocodazole Analogues as a Potential Anticancer Drugs Using Molecular Docking and DFT Calculations**
Muhammad Khattab and Ahmed A. Al-Karmalawy
- 36 Applications of Virtual Screening in Bioprospecting: Facts, Shifts, and Perspectives to Explore the Chemo-Structural Diversity of Natural Products**
Kauê Santana, Lidiane Diniz do Nascimento, Anderson Lima e Lima, Vinícius Damasceno, Claudio Nahum, Rodolpho C. Braga and Jerônimo Lameira
- 59 Molecular Docking and Dynamics Simulation Revealed the Potential Inhibitory Activity of ACEIs Against SARS-CoV-2 Targeting the hACE2 Receptor**
Ahmed A. Al-Karmalawy, Mohammed A. Dahab, Ahmed M. Metwaly, Sameh S. Elhady, Eslam B. Elkaeed, Ibrahim H. Eissa and Khaled M. Darwish
- 80 Using Network Pharmacology and Molecular Docking to Explore the Mechanism of Shan Ci Gu (*Cremastra appendiculata*) Against Non-Small Cell Lung Cancer**
Yan Wang, Yunwu Zhang, Yujia Wang, Xinyao Shu, Chaorui Lu, Shiliang Shao, Xingting Liu, Cheng Yang, Jingsong Luo and Quanyu Du
- 94 Molecular Dynamics Simulations Reveal Interactions of an IgG1 Antibody With Selected Fc Receptors**
Sebastjan Kralj, Milan Hodošček, Barbara Podobnik, Tanja Kunej, Urban Bren, Dušanka Janežič and Janez Konc
- 105 Computational Insights Into the Inhibition Mechanism of Proanthocyanidin B2 on Tau Hexapeptide (PHF6) Oligomer**
Qin Li, Chunmei Xiong, Hongli Liu, Huizhen Ge, Xiaojun Yao and Huanxiang Liu
- 114 A Machine Learning Model to Predict Drug Transfer Across the Human Placenta Barrier**
Juan I. Di Filippo, Mariela Bollini and Claudio N. Cavasotto

- 125 Identification of Potential Binding Sites of Sialic Acids on the RBD Domain of SARS-CoV-2 Spike Protein**
Bingqian Li, Lin Wang, Huan Ge, Xianglei Zhang, Penxuan Ren, Yu Guo, Wuyan Chen, Jie Li, Wei Zhu, Wenzhang Chen, Lili Zhu and Fang Bai
- 134 Potential Inhibitors of Fascin From A Database of Marine Natural Products: A Virtual Screening and Molecular Dynamics Study**
Lirui Lin, Kai Lin, Xiaodong Wu, Jia Liu, Yinwei Cheng, Li-Yan Xu, En-Min Li and Geng Dong
- 150 Synthesis, Computational Pharmacokinetics Report, Conceptual DFT-Based Calculations and Anti-Acetylcholinesterase Activity of Hydroxyapatite Nanoparticles Derived From Acorus Calamus Plant Extract**
Sushma Pradeep, Anisha S. Jain, Chandan Dharmashekara, Shashanka K. Prasad, Nagaraju Akshatha, R. Pruthvish, Raghavendra G Amachawadi, Chandrashekar Srinivasa, Asad Syed, Abdallah M. Elgorban, Abdulaziz A. Al Kheraif, Joaquín Ortega-Castro, Juan Frau, Norma Flores-Holguín, Chandan Shivamallu, Shiva Prasad Kollur and Daniel Glossman-Mitnik
- 168 Virtual Screening on Marine Natural Products for Discovering TMPRSS2 Inhibitors**
Mehdi Mahmudpour, Iraj Nabipour, Mohsen Keshavarz and Maryam Farrokhnia
- 182 Virtual Screening of Drug-Like Compounds as Potential Inhibitors of the Dengue Virus NS5 Protein**
Leidy L. García-Ariza, Cristian Rocha-Roa, Leonardo Padilla-Sanabria and Jhon C. Castaño-Osorio



OPEN ACCESS

EDITED AND REVIEWED BY
Thomas S. Hofer,
University of Innsbruck, Austria

*CORRESPONDENCE

Fei Ye,
yefeizst@zstu.edu.cn
Sylvain Broussy,
sylvain.broussy@u-paris.fr

SPECIALTY SECTION

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

RECEIVED 14 June 2022

ACCEPTED 29 June 2022

PUBLISHED 18 July 2022

CITATION

Ye F, Lin M, Jin J and Broussy S (2022), Editorial: Computer-aided drug design: Drug discovery, computational modelling, and artificial intelligence. *Front. Chem.* 10:968687. doi: 10.3389/fchem.2022.968687

COPYRIGHT

© 2022 Ye, Lin, Jin and Broussy. 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.

Editorial: Computer-aided drug design: Drug discovery, computational modelling, and artificial intelligence

Fei Ye^{1,2*}, Min Lin¹, Jia Jin¹ and Sylvain Broussy^{3*}

¹College of Life Sciences and Medicine, Zhejiang Sci-Tech University, Hangzhou, China, ²School of Pharmaceutical Sciences, Zhejiang Chinese Medical University, Hangzhou, China, ³University of Paris, Pharmacy, Paris, France

KEYWORDS

computer-aided drug design, lead compound discovery, hit optimization, allosteric regulation, conformational dynamics, artificial intelligence

Editorial on the Research Topic

Computer-aided drug design: Drug discovery, computational modelling, and artificial intelligence

Owing to the rapid improvement of computational methodologies and high-performance computational resources, computer-aided drug design (CADD) has been validated as an efficient and powerful strategy in almost every stage of drug discovery and development.

Generally, CADD can be divided into structure-based drug design (SBDD) and ligand-based drug design (LBDD). Due to the rapid development of crystallography and homology modeling, structure-based virtual screening has emerged as a useful technique to identify potential hits during early stage of drug discovery. LBDD strategies based on available information of known bioactive molecules, such as QSAR (Quantitative Structure–Activity Relationship) analysis, scaffold hopping, pharmacophore modeling, are also widely used for hit optimization and activity prediction. In addition, computational techniques like quantum chemistry calculation, molecular dynamics (MD) simulations and elastic network models can be used to study protein catalytic mechanism, conformational transition and allosteric regulations at an atomic level of detail, which provide useful information for mechanism-based drug design. Recently, with the development of machine learning theory and the accumulation of pharmacological data, artificial intelligence (AI), a powerful data mining technology, has been widely used in various fields of drug design, including virtual screening, *de novo* drug design, QSAR analysis, as well as *in silico* evaluation of absorption, distribution, metabolism, excretion and toxicity (ADME/T) properties.

In this Research Topic, we have invited some scientists worldwide to contribute original research and review articles which could enhance our understanding of some of the above issues. Several studies utilized multiple computational approaches, such as

molecular docking, DFT calculations, molecular dynamics (MD) simulations, ADME/T prediction, as well as biological evaluations to identify novel compounds against a series of important targets, like Tubulin (Khattab and Al-Karmalawy), Dengue Virus NS5 protein (García-Ariza et al.), α -Glucosidase (Liu et al.), hACE2 receptor of SARS-CoV-2 (Al-Karmalawy et al.), Fascin (Lin et al.), TMPRSS2 (Mahmudpour et al.) and Alzheimer's disease targets (Pradeep et al.). Santana et al. discussed the development of computational approaches to explore the chemo-structural diversity of natural products. CADD methods are also widely used for exploring interactions between ligand and receptor, as well as inhibition mechanisms of active compounds. Tao et al. performed molecular docking and MD simulations to study the interaction between RBD and two glycopeptide antibiotics (Vancomycin and Teicoplanin). Wang et al. used network pharmacology and molecular docking to explore the mechanism of Shan Ci Gu (*Cremastra appendiculata*) against non-small cell lung cancer. In addition, Di Filippo et al. proposed a machine learning model to predict drug transfer across the human placenta barrier, which could be used as a filter for chemical libraries in virtual screening campaigns.

In summary, the above works presented in this special Research Topic illustrate the applications of CADD approaches and highlight the importance of developing new methods. At last, as the Guest Editors of this Research Topic, we would like to thank all the authors for their contributed articles and all the referees for their comments on the manuscripts. We hope that the readers will find this Research Topic interesting and useful for their research. Finally, we

appreciate the editorial staff of Frontiers in Chemistry for their work in publishing this Research Topic.

Author contributions

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

Funding

This work is financially supported by the National Natural Science Foundation of China (81803339).

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.



Discovery of New α -Glucosidase Inhibitors: Structure-Based Virtual Screening and Biological Evaluation

Shan-Kui Liu¹, Haifang Hao¹, Yuan Bian¹, Yong-Xi Ge¹, Shengyuan Lu¹, Hong-Xu Xie¹, Kai-Ming Wang¹, Hongrui Tao², Chao Yuan³, Juan Zhang^{1*}, Jie Zhang⁴, Cheng-Shi Jiang^{1*} and Kongkai Zhu^{1,5*}

¹School of Biological Science and Technology, University of Jinan, Jinan, China, ²Drug Discovery and Design Center, State Key Laboratory of Drug Research, Shanghai Institute of Materia Medica, Chinese Academy of Sciences, Shanghai, China, ³Zoucheng Administration for Market Regulation, Zoucheng, China, ⁴Lunan Pharmaceutical Group Corporation, Linyi, China, ⁵Shandong Key Laboratory of Biochemical Analysis, College of Chemistry and Molecular Engineering, Qingdao University of Science and Technology, Qingdao, China

OPEN ACCESS

Edited by:

Zhongjie Liang,
Soochow University, China

Reviewed by:

Xuben Hou,
Shandong University, China
Xiao Jun Yao,
Macau University of Science and
Technology, Macau

*Correspondence:

Juan Zhang
bio_zhangj@ujn.edu.cn
Cheng-Shi Jiang
bio_jiangcs@ujn.edu.cn
Kongkai Zhu
bio_zhukk@ujn.edu.cn

Specialty section:

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

Received: 09 December 2020

Accepted: 20 January 2021

Published: 08 March 2021

Citation:

Liu S-K, Hao H, Bian Y, Ge Y-X, Lu S, Xie H-X, Wang K-M, Tao H, Yuan C, Zhang J, Zhang J, Jiang C-S and Zhu K (2021) Discovery of New α -Glucosidase Inhibitors: Structure-Based Virtual Screening and Biological Evaluation. *Front. Chem.* 9:639279. doi: 10.3389/fchem.2021.639279

α -Glycosidase inhibitors could inhibit the digestion of carbohydrates into glucose and promote glucose conversion, which have been used for the treatment of type 2 diabetes. In the present study, 52 candidates of α -glycosidase inhibitors were selected from commercial Specs compound library based on molecular docking-based virtual screening. Four different scaffold compounds (7, 22, 37, and 44) were identified as α -glycosidase inhibitors with IC_{50} values ranging from 9.99 to 35.19 μ M. All these four compounds exerted better inhibitory activities than the positive control (1-deoxynojirimycin, IC_{50} = 52.02 μ M). The fluorescence quenching study and kinetic analysis revealed that all these compounds directly bind to α -glycosidase and belonged to the noncompetitive α -glycosidase inhibitors. Then, the binding modes of these four compounds were carefully investigated. Significantly, these four compounds showed nontoxicity (IC_{50} > 100 μ M) toward the human normal hepatocyte cell line (LO2), which indicated the potential of developing into novel candidates for type 2 diabetes treatment.

Keywords: α -glycosidase, virtual screening, cytotoxicity, type 2 diabetes, molecular docking

INTRODUCTION

Diabetes is a metabolic disorder that causes high blood sugar and could directly increase the risk of other deadly diseases, such as cancer, stroke, and cardiovascular diseases (Cohen and Goedert, 2004; Zeng et al., 2019). According to the statistics of the World Health Organization (WHO), about 422 million people suffered from diabetes in 2014 around the world, and its prevalence is projected to be 642 million by 2040 (Reusch and Manson, 2017; World Health Organization, 2020). The ineffective use of insulin could result in the type 2 diabetes and accounts for more than 90% of diabetes cases (Proença et al., 2017).

Controlling blood glucose levels is thought to be the main strategy for treating diabetes and reducing diabetes complications (Ye et al., 2019). α -Glucosidase is a key carbohydrate hydrolase that regulates blood glucose by specifically hydrolyzing 1,4- α -glucopyranosidic bond to produce α -glucose (Kazmi et al., 2018). Early studies have shown that the inhibition of α -glucosidase activity could retard the absorption of glucose and decrease the postprandial blood glucose levels

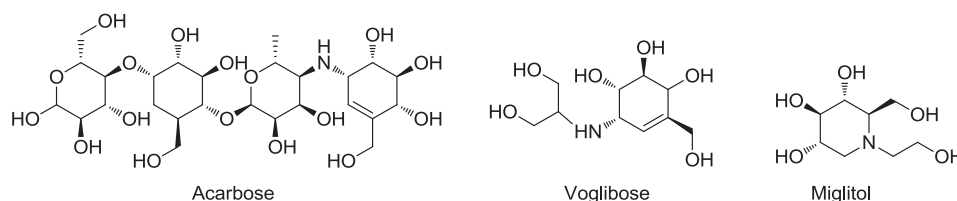


FIGURE 1 | Clinically Approved α -glucosidase inhibitors.

(Park et al., 2008; Kim et al., 2019). Therefore, α -glucosidase has been taken as a key target for treating diabetes, and the inhibitors of α -glucosidase can be developed into effective therapeutic drugs to treat this disease (Li et al., 2010). α -Glucosidase inhibitors such as acarbose, miglitol, and voglibose (shown in **Figure 1**) are the most well-known ones (Joshi et al., 2015). Acarbose, the first approved drug in α -glucosidase inhibitor category, was used to delay the release of glucose from polysaccharides by binding with α -glucosidase. Voglibose was used to discontinue the uptake and hydrolysis of saccharides by selectively inhibiting α -glucosidase vs. pancreatic α -amylase and lactase. Miglitol, the first pseudo-monosaccharide α -glucosidase inhibitor, was approved to reduce postprandial glucose (Hossain et al., 2020). However, some unexpected adverse effects (for instance, flatulence, diarrhea, and stomachache) limited their clinical application. Based on this background, numerous efforts have been carried out to discover new α -glucosidase inhibitors from diverse sources, such as natural products and chemical synthetic compounds (Chen et al., 2017; Liu and Ma, 2017; Abbas et al., 2019; Dhameja and Gupta, 2019).

Virtual screening has been proven to be a very effective tool capable of providing drug hits or leads with structural diversity and makes drug discovery faster and more efficient (Kitchen et al., 2004; Kontoyianni, 2017). In this study, molecular docking-based virtual screening on Specs database was conducted to identify α -glucosidase inhibitors with new chemotypes. After testing the purchased 52 compounds that were obtained by docking screening, four compounds, namely, 7, 22, 37, and 44 with different scaffolds, were disclosed as new α -glycosidase inhibitors. Kinetic analysis of these compounds revealed that they inhibited α -glycosidase activity in a noncompetitive type. Then, the binding modes of these compounds with α -glycosidase were investigated, and the results indicated that all of these compounds could be well located in the acarbose-binding site and displayed very similar binding poses. Moreover, the cytotoxicity of these compounds toward the human normal hepatocyte cell line (LO2) was evaluated. The present results provided new α -glycosidase inhibitors serving as hit compounds for developing novel medications used in the treatment of type 2 diabetes.

METHODS AND MATERIALS

Molecular Docking-Based Virtual Screening

The protein coordinates in the α -glucosidase crystal complex structure (PDB code 3W37) were prepared by the Protein

Preparation Wizard panel inserted in the Maestro with the default settings. Residues within 15 Å centered on acarbose were defined as compound-binding sites in which the docking grid was generated by the Receptor Grid Generation panel. The default settings were adopted for the cutoff, neutralization, etc. The docked compounds in Specs database were prepared with LigPrep panel. Then, the prepared compounds were docked to the aforementioned docking grid with extra precision (XP) mode. “Clustering Molecules” protocol inserted in Pipeline Pilot 7.5 was employed to achieve the cluster analysis. The top ranked compounds assessed by XP GScore were clustered into 30 clusters. To increase the diversity of selected compounds, at least one candidate was selected in each cluster. In addition, we gave priority to the compounds with simple structure and/or small molecular weight.

α -Glucosidase Inhibitory Assay

The α -glucosidase inhibitory evaluation of the purchased 52 compounds was performed according to the previously described protocol (Tang et al., 2014; Ye et al., 2019). α -Glucosidase (Sigma, G5003) derived from baker's yeast, and pNPG (Sigma, N1377) and the substrate were both purchased from Sigma-Aldrich. 1-Deoxynojirimycin was used as the positive control. The tested compounds and 1-deoxynojirimycin were dissolved in DMSO, the α -glucosidase and the substrate pNPG were both dissolved in phosphate buffer (pH = 6.8). The compounds and α -glucosidase were preincubated in phosphate buffer (37°C, 15 min). Then, 25 μ L substrate buffer was added to the system to start the reaction, and the incubation was continued at 37°C for 15 min. Finally, the reaction was terminated by the addition of 50 μ L 0.2 M reaction termination solution. The optical density (OD) was measured at an absorbance wavelength of 405 nm using a microplate reader (Tecan, Switzerland). The IC₅₀ values were estimated with six different concentrations, and each sample was measured three times in parallel experiments.

Fluorescence Quenching Experiment

According to the previously reported method (Aguilar-Moncayo et al., 2010), all fluorescence spectra were measured on a fluorescence spectrophotometer (Agilent Cary Eclipse) equipped with a 10.0-mm quartz cell and a thermostat bath. In the fluorescence spectrophotometer, α -glucosidase (1 U/ml) was pretreated with certain concentrations of inhibitors for 30 min at 37°C. 100 μ L of the above solution

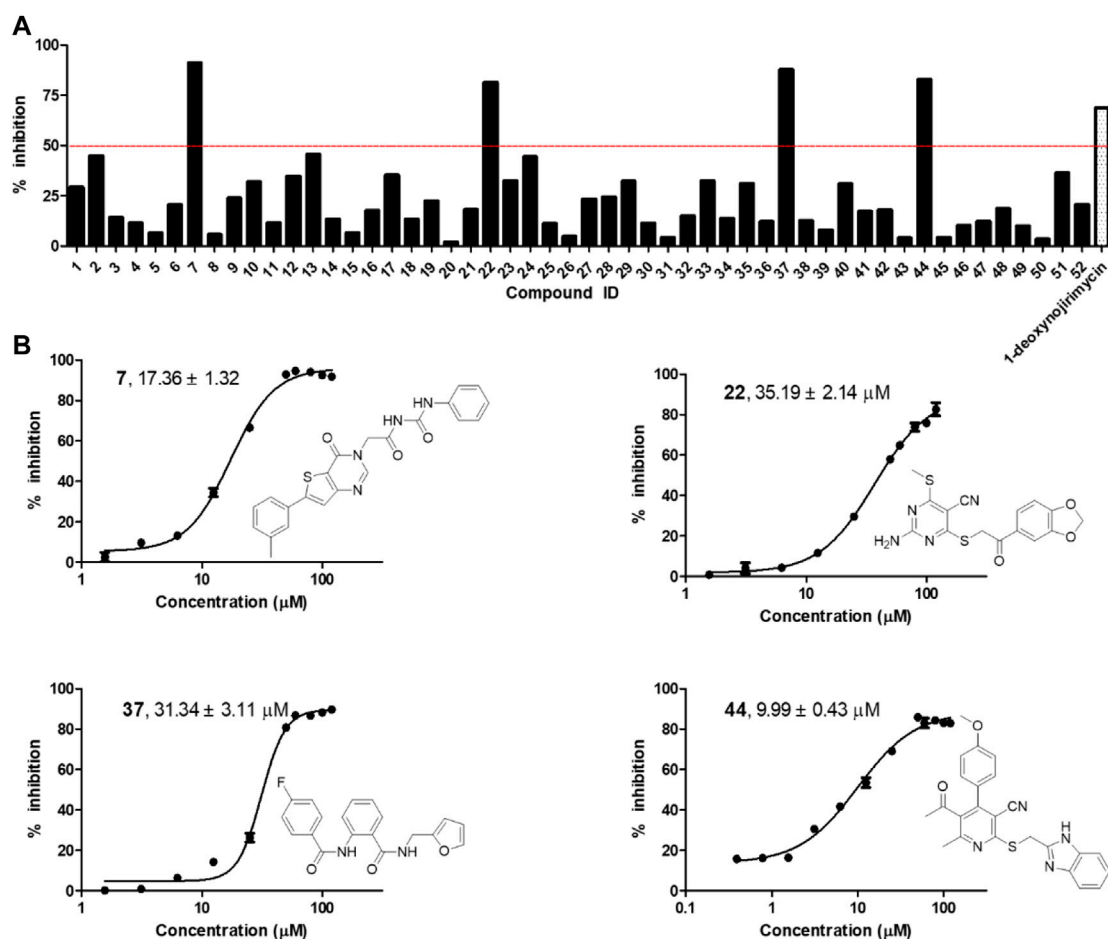


FIGURE 2 | (A) α -glucosidase inhibitory activity of the 52 selected candidates at 100 μ M; **(B)** The chemical structures and IC_{50} curves of compounds 7, 22, 37, and 44. IC_{50} data are shown as mean $\pm$ SD of three independent experiments.

(pH 6.8) was added accurately to the quartz cell. The blank was used for buffer spectrum values. The fluorescence emission spectra were measured at 37°C. The excitation wavelength was 290 nm, and the emission spectrum was recorded from 320 to 500 nm.

Kinetic Assay

The inhibition type of the inhibitors against α -glucosidase activities was evaluated based on a described method (Hou et al., 2009). Increasing concentrations of substrates pNPG were used in the absence or presence of tested compounds at four different concentrations around the IC_{50} values. The inhibitory kinetics of the investigated compounds on α -glucosidase was analyzed using the Lineweaver–Burk plot of the substrate concentration and velocity.

Cell Viability Assay

The LO2 cell line was cultured in a proper medium supplemented with 10% fetal bovine serum in a humidified atmosphere of 5% CO_2 at 37°C. Cell suspensions were plated in 96-well plates at a

density of 2×10^4 cells/cm³. Compounds were solubilized in DMSO at six different concentrations. After incubation for 24 h, the cells were treated with various concentrations of tested substances for 48 h and then incubated with 100 μ L of MTT at 37°C for 2 h. The formazan dye product was measured by the absorbance at 490 nm on a Tecan Spark multimode microplate reader (Switzerland).

RESULTS AND DISCUSSIONS

Fifty-Two Candidates of α -Glucosidase Inhibitor Were Selected From the Molecular Docking–Based Virtual Screening Result

As the crystal structure of α -glucosidase–acarbose complex has been determined (PDB code 3W37) (Tagami et al., 2013), molecular docking–based virtual screening could be performed. Specs database that contains 200,000 compounds was chosen as the screening database. The redock result of

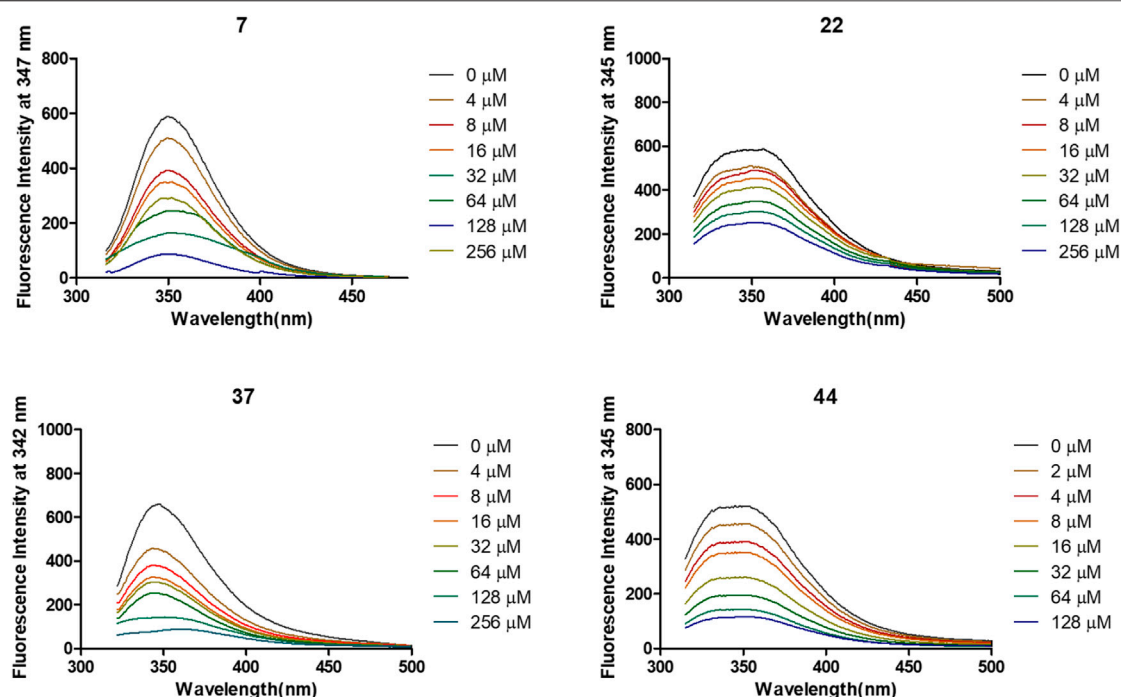


FIGURE 3 | Variation of fluorescence emission spectra of α -glucosidase (1 U/ml) in the presence of compounds 7, 22, 37, and 44 with increasing concentration for 30 min at 37°C.

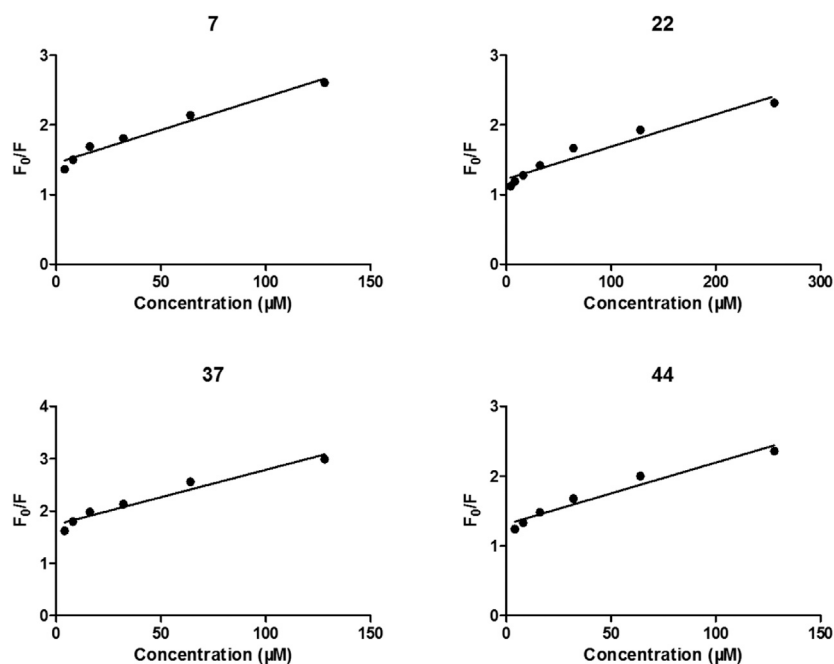


FIGURE 4 | Stern-Volmer plots for the fluorescence quenching of α -glucosidase by compounds 7, 22, 37, and 44.

acarbose (**Supplementary Figure S1**) declared that GLIDE program (Halgren et al., 2004) inserted in the Schrödinger program suite could well reproduced the binding mode of

acarbose in the crystal structure. The top 300 molecules ranked by the docking score were selected for the following cluster analysis. Finally, 52 compounds were retained and

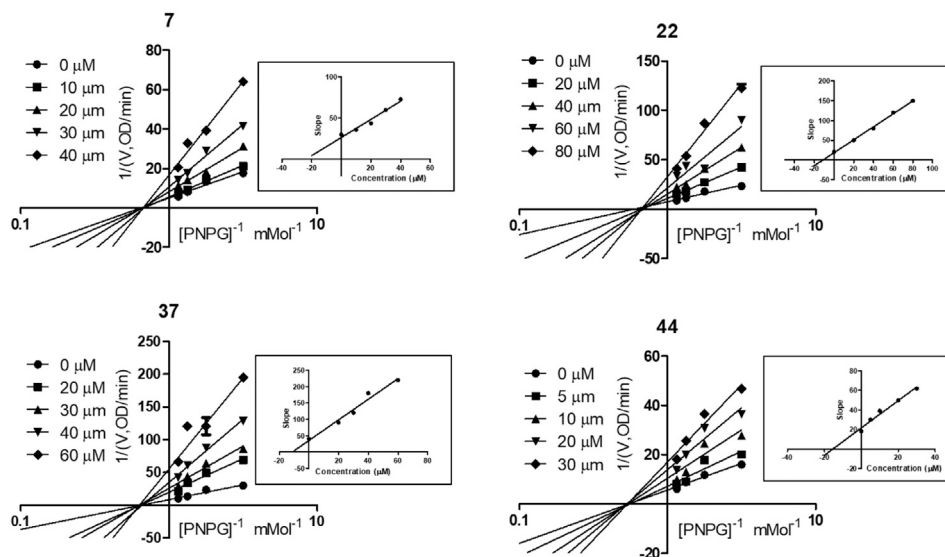


FIGURE 5 | Kinetic assay on α -glucosidase inhibition by compounds 7, 22, 37, and 44, respectively. Lineweaver-Burk reciprocal plots of initial velocity and increasing substrate (PNPG) concentration with secondary plot of slopes vs. the concentration of compounds.

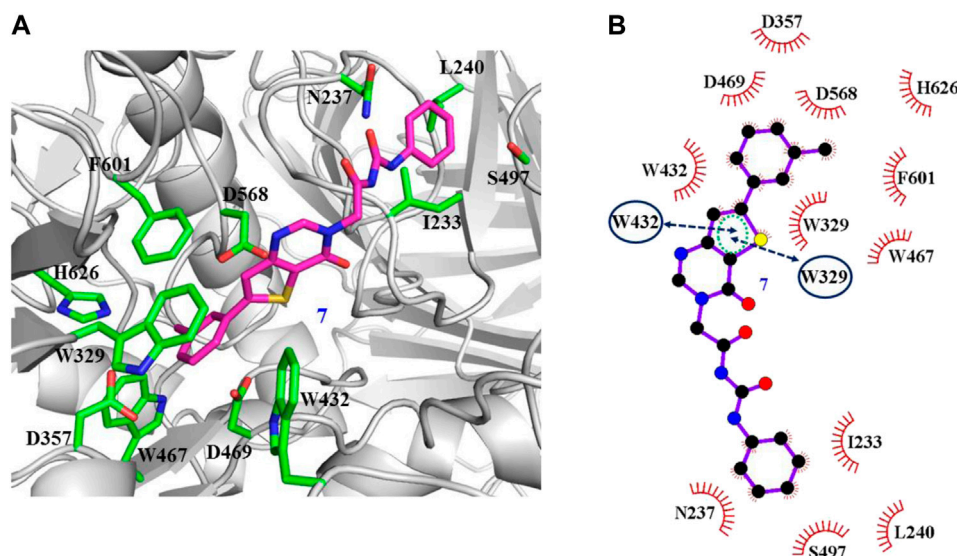


FIGURE 6 | Docking pose of compound 7 bound to the acarbose binding site in α -glucosidase. **(A)** The three-dimensional interacting modes between 7 and α -glucosidase. α -Glucosidase, 7 and the interacting residues were shown as cartoon, sticks (carbon atoms colored in magenta), sticks (carbon atom colored in green), respectively. **(B)** Schematic representation displayed the hydrophobic interactions (shown as starbursts) and Pi-Pi interactions (shown as oval) of 7 with α -Glucosidase.

purchased from the Specs database supplier for further α -glucosidase enzymatic inhibition activity evaluation.

***In vitro* Inhibition Test Against α -Glucosidase Identified Four Active Compounds 7, 22, 37, and 44**

The selected 52 candidates were initially evaluated for their inhibitory ratios against α -glucosidase at 100 μ M with 1-

deoxynojirimycin as positive reference. The α -glucosidase enzymatic inhibition bioassay results indicated that four compounds, namely, 7, 22, 37, and 44 with representing totally different scaffolds, exhibited an inhibition ratio above 50% at 100 μ M (**Figure 2A**). Then, the IC_{50} values of these four compounds were further determined. As shown in **Figure 2B**, compounds 7, 22, 37, and 44 displayed IC_{50} values of 17.36 ± 1.32 , 35.19 ± 2.14 , 31.34 ± 3.11 , and 9.99 ± 0.43 μ M, respectively. All of them showed better activity than the

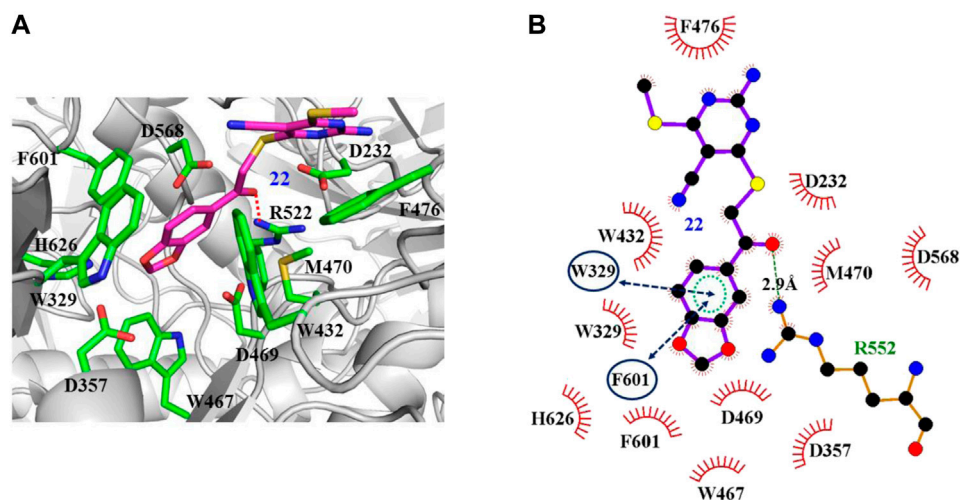


FIGURE 7 | Docking pose of compound 22 bound to the acarbose binding site in α -glucosidase. **(A)** The three-dimensional interacting modes between 22 and α -glucosidase. α -Glucosidase, 22 and the interacting residues were shown as cartoon, sticks (carbon atoms colored in magenta), sticks (carbon atom colored in green), respectively. **(B)** Schematic representation displayed the hydrophobic interactions (shown as starbursts), Pi-Pi interactions (shown as oval), and H-bond interactions (denoted by dotted green lines) of 22 with α -Glucosidase.

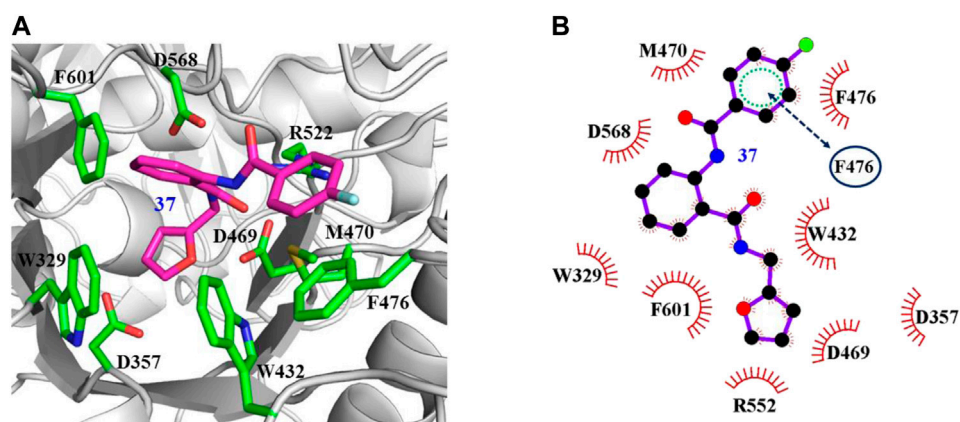


FIGURE 8 | Docking pose of compound 37 bound to the acarbose binding site in α -glucosidase. **(A)** The three-dimensional interacting modes between 37 and α -glucosidase. α -Glucosidase, 37 and the interacting residues were shown as cartoon, sticks (carbon atoms colored in magenta), sticks (carbon atom colored in green), respectively. **(B)** Schematic representation displayed the hydrophobic interactions (shown as starbursts) of 37 with α -Glucosidase.

positive reference control 1-deoxynojirimycin (IC_{50} , $52.02 \pm 3.78 \mu M$), and compound 44 exhibited the most potent activity.

Fluorescence Quenching Assay Confirmed the Binding of 7, 22, 37, and 44 to α -Glucosidase

The interactions of 7, 22, 37, and 44 with α -glucosidase were explored through the fluorescence quenching experiments. As displayed in Figure 3, the variations of the intrinsic fluorescence emission of α -glucosidase ($2 \mu M$) in the presence of increasing concentration of molecules 7, 22, 37, and 44, respectively, were recorded at $37^\circ C$ with the wavelength range from 320 to 500 nm.

The intrinsic fluorescence emission peak at 345 nm was observed after being excited at 290 nm.

After treated by compounds 7, 22, 37, and 44 with increasing concentration (Figure 4), the fluorescence intensities of enzyme in all tested systems were gradually quenched in a type of concentration-dependent manner. Thus, these results confirmed the binding of these inhibitors to α -glucosidase.

Kinetic Study on α -Glucosidase Inhibition Declared the Noncompetitive Manner of These Four Compounds

To explore the mechanism of the interaction modes of compounds 7, 22, 37, and 44 with the enzyme, kinetic assay

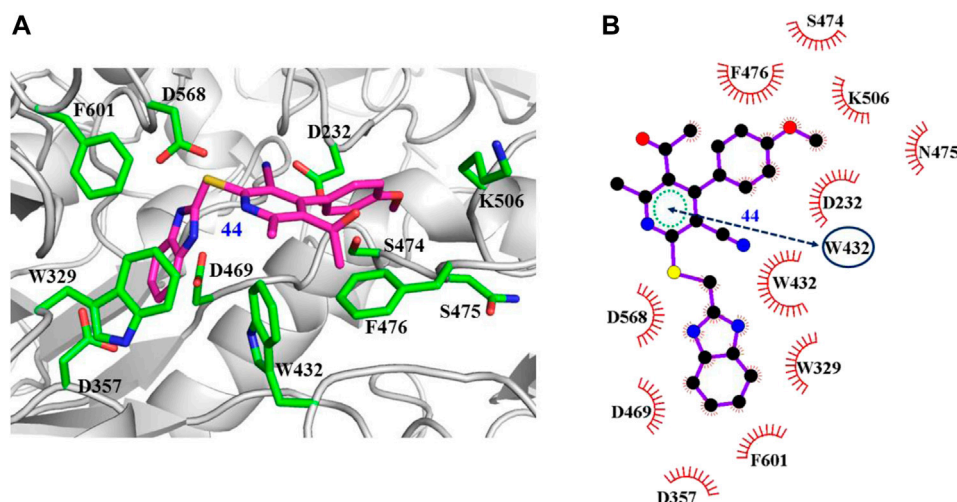


FIGURE 9 | Docking pose of compound 44 bound to the acarbose binding site in α -glycosidase. **(A)** The three-dimensional interacting modes between 44 and α -glycosidase. Compound 44 and the interacting residues were shown as cartoon, sticks (carbon atoms colored in magenta), sticks (carbon atom colored in green), respectively. **(B)** Schematic representation displayed the hydrophobic interactions (shown as starbursts) and Pi-Pi interactions (shown as oval) of 44 with α -Glycosidase.

was conducted to study their inhibition types using Lineweaver–Burk plot analysis (Wang et al., 2004; Sun et al., 2018). The results shown in **Figure 5** indicate that compounds 7, 22, 37, and 44 were noncompetitive α -glycosidase inhibitors, with estimated K_i values of 24.18, 11.34, 11.27, and 15.39 μ M, respectively.

Molecular Docking Simulation Revealed the Binding Mode of These Four Compounds

The interaction mechanisms of compounds 7, 22, 37, and 44 with α -glycosidase were carefully analyzed with the molecular docking results, as shown in **Figures 6–9**. All these four compounds could well bind to the allosteric sites away from the active site (Asp214, Glu276, and Asp349) (Ye et al., 2019) in α -glycosidase and formed hydrophobic interactions with nearby residues. These results were consistent with the noncompetitive property. The docking scores of these four hits were -3.811, -2.825, -3.627, and -6.283. Specifically, inhibitor 7 established hydrophobic interactions with residues D357, D469, W432, N237, S497, L240, I233, W329, W467, F601, H626, and D568 and formed Pi–Pi stacking with residues W329 and W432. Compound 22 formed hydrophobic interactions with residues W329, F476, D357, D469, W432, D232, M470, W467, F601, H626, and D568 and formed Pi–Pi stacking with residues W329 and F601. Besides, 22 established H bond interaction with R552 residue. Compound 37 formed hydrophobic interaction with residues M470, W329, F476, D357, D469, W432, F601, D568, and R552 and formed Pi–Pi stacking with residue F476. Additionally, compound 44 formed hydrophobic interaction with residues S474, W329, F476, D357, D469, W432, N475, D568, D232, F601, and K506 and formed Pi–Pi stacking with residue W432. From these data, we could find that residues

W432, W329, F601, D357, D469, and D568 were the key residues contributing interaction with all of the four compounds.

In vitro Cytotoxicity

Since most of the drugs are metabolized in the liver, there is a great focus on the hepatic safety of new medicines. Thus, the cytotoxicity of inhibitors 7, 22, 37, and 44 was evaluated in human normal hepatocyte (LO2) cells using the MTT method (Ge et al., 2020). The results disclosed that all of these compounds had IC_{50} values more than 100 μ M toward LO2 cells, suggesting they are nontoxic toward liver cells. Thus, further structural optimization and biological evaluation for 7, 22, 37, and 44 deserved further investigation.

CONCLUSION

In this study, four novel α -glycosidase inhibitors 7, 22, 37, and 44 with distinct structural features were identified through virtual screening and *in vitro* evaluation. Among them, compound 44 had the best α -glycosidase inhibitory activity with IC_{50} and K_i values of 9.99 ± 0.43 and 15.39 μ M, respectively. The fluorescence quenching experiment indicated all these compounds could directly bind to α -glycosidase, and the kinetic study revealed a noncompetitive α -glycosidase inhibitory mechanism of these compounds toward α -glycosidase. In addition, binding mode analysis provided the detailed binding mechanism of these four α -glycosidase inhibitors, which made further structural optimization feasible. Moreover, the *in vitro* cytotoxicity bioassay demonstrated these α -glycosidase inhibitors were nontoxic toward LO2 cells. Based on these results, these compounds can serve as promising hit compounds for further bioactivity optimization and anti-type 2 diabetes study.

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 authors.

AUTHOR CONTRIBUTIONS

All authors made significant contributions to the preparation of the manuscript and approved it before submission. KZ, C-SJ, and JuZ developed the paper design and finalized the manuscript. S-KL wrote the manuscript; HH and YB performed molecular docking study; Y-XG, SL, and H-XX helped to collect the data and complete the figure; K-MW and HT edited the references and revised the style of the manuscript, CY and JiZ revised the manuscript.

REFERENCES

- Abbas, G., Al-Harrasi, A., Hussain, H., Hamaed, A., and Supuran, C. T. (2019). The management of diabetes mellitus-imperative role of natural products against dipeptidyl peptidase-4, α -glucosidase and sodium-dependent glucose Co-transporter 2 (SGLT2). *Bioorg. Chem.* 86, 305–315. doi:10.1016/j.bioorg.2019.02.009
- Aguilar-Moncayo, M., García-Moreno, M. I., Stütz, A. E., García, Fernández, J. M., Wrodnigg, T. M., and Ortiz-Mellet, C. (2010). Fluorescent-tagged sp²-iminosugars with potent β -glucosidase inhibitory activity. *Bioorg. Med. Chem.* 18 (21), 7439–7445. doi:10.1016/j.bmc.2010.09.003
- Chen, H., Xiong, L., and Wang, W. J. (2017). Reviews on α -glucosidase inhibitor from plant secondary metabolites. *Zhongguo Zhong Yao Za Zhi* 42 (15), 2915–2924. doi:10.19540/j.cnki.cjcmm.20170714.021
- Cohen, P., and Goedert, M. (2004). GSK3 inhibitors: development and therapeutic potential. *Nat. Rev. Drug Discov.* 3 (6), 479–487. doi:10.1038/nrd1415
- Dhameja, M., and Gupta, P. (2019). Synthetic heterocyclic candidates as promising α -glucosidase inhibitors: an overview. *Eur. J. Med. Chem.* 176, 343–377. doi:10.1016/j.ejmech.2019.04.025
- Ge, Y. X., Cheng, Z. Q., Zhou, L., Xie, H. X., Wang, Y. Y., Zhu, K., et al. (2020). Synthesis and biological evaluation of quinoline/cinnamic acid hybrids as amyloid-beta aggregation inhibitors. *Monatsh Chem.* 151, 845–852. doi:10.1007/s00706-020-02609-2
- 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 (7), 1750–1759. doi:10.1021/jm030644s
- Hossain, U., Das, A. K., Ghosh, S., and Sil, P. C. (2020). An overview on the role of bioactive α -glucosidase inhibitors in ameliorating diabetic complications. *Food Chem. Toxicol.* 145, 111738. doi:10.1016/j.fct.2020.111738
- Hou, W., Li, Y., Zhang, Q., Wei, X., Peng, A., Chen, L., et al. (2009). Triterpene acids isolated from Lagerstroemia speciosa leaves as α -glucosidase inhibitors. *Phytother. Res.* 23 (5), 614–618. doi:10.1002/ptr.2661
- Joshi, S. R., Standl, E., Tong, N., Shah, P., Kalra, S., and Rathod, R. (2015). Therapeutic potential of α -glucosidase inhibitors in type 2 diabetes mellitus: an evidence-based review. *Expert Opin. Pharmacother.* 16 (13), 1959–1981. doi:10.1517/14656566.2015.1070827
- Kazmi, M., Zaib, S., Ibrar, A., Amjad, S. T., Shafique, Z., Mehsum, S., et al. (2018). A new entry into the portfolio of α -glucosidase inhibitors as potent therapeutics for type 2 diabetes: Design, bioevaluation and one-pot multi-component

FUNDING

This research work was financially supported by the Major Science and Technology Innovation Project of Shandong Province (No. 2019JZZY011116), National Natural Science Foundation of China (Nos. 21672082 and 81803438), Natural Science Foundation of Shandong Province (Nos. ZR2019YQ31 and ZR201910300056), Key Technology Research and Development Program of Shandong (No. 2019GSF108043), and Science and Technology Project of University of Jinan (No. XKY2004).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fchem.2021.639279/full#supplementary-material>.

- synthesis of diamine-bridged coumarinyl oxadiazole conjugates. *Bioorg. Chem.* 77, 190–202. doi:10.1016/j.bioorg.2017.12.022
- Kim, J. H., Kim, H. Y., and Jin, C. H. (2019). Mechanistic investigation of anthocyanidin derivatives as α -glucosidase inhibitors. *Bioorg. Chem.* 87, 803–809. doi:10.1016/j.bioorg.2019.01.033
- Kitchen, D. B., Decornez, H., Furr, J. R., and Bajorath, J. (2004). Docking and scoring in virtual screening for drug discovery: methods and applications. *Nat. Rev. Drug Discov.* 3 (11), 935–949. doi:10.1038/nrd1549
- Kontoyianni, M. (2017). Docking and virtual screening in drug discovery. *Methods Mol. Biol.* 1647, 255–266. doi:10.1007/978-1-4939-7201-2_18
- Li, D. Q., Qian, Z. M., and Li, S. P. (2010). Inhibition of three selected beverage extracts on α -glucosidase and rapid identification of their active compounds using HPLC-DAD-MS/MS and biochemical detection. *J. Agric. Food Chem.* 58 (11), 6608–6613. doi:10.1021/jf100853c
- Liu, Z., and Ma, S. (2017). Recent advances in synthetic α -glucosidase inhibitors. *ChemMedChem* 12 (11), 819–829. doi:10.1002/cmdc.201700216
- Park, H., Hwang, K. Y., Oh, K. H., Kim, Y. H., Lee, J. Y., and Kim, K. (2008). Discovery of novel α -glucosidase inhibitors based on the virtual screening with the homology-modeled protein structure. *Bioorg. Med. Chem.* 16 (1), 284–292. doi:10.1016/j.bmc.2007.09.036
- Proença, C., Freitas, M., Ribeiro, D., Oliveira, E. F. T., Sousa, J. L. C., Tomé, S. M., et al. (2017). α -Glucosidase inhibition by flavonoids: an *in vitro* and *in silico* structure-activity relationship study. *J. Enzyme Inhib. Med. Chem.* 32 (1), 1216–1228. doi:10.1080/14756366.2017.1368503
- Reusch, J. E., and Manson, J. E. (2017). Management of type 2 diabetes in 2017: getting to goal. *JAMA* 317(10), 1015–1016. doi:10.1001/jama.2017.0241
- Sun, H., Song, X., Tao, Y., Li, M., Yang, K., Zheng, H., et al. (2018). Synthesis & α -glucosidase inhibitory & glucose consumption-promoting activities of flavonoid-coumarin hybrids. *Future Med. Chem.* 10 (9), 1055–1066. doi:10.4155/fmc-2017-0293
- Tagami, T., Yamashita, K., Okuyama, M., Mori, H., Yao, M., and Kimura, A. (2013). Molecular basis for the recognition of long-chain substrates by plant α -glucosidase. *J. Biol. Chem.* 288 (26), 19296–19303. doi:10.1074/jbc.M113.465211
- Tang, C., Zhu, L., Chen, Y., Qin, R., Mei, Z., Xu, J., et al. (2014). Synthesis and biological evaluation of oleanolic acid derivative-chalcone conjugates as α -glucosidase inhibitors. *RSC Adv.* 4, 10862–10874. doi:10.1039/c3ra46492j
- Wang, Y., Ma, L., Pang, C., Huang, M., Huang, Z., and Gu, L. (2004). Synergistic inhibition of genistein and d-glucose on α -glucosidase. *Bioorg. Med. Chem. Lett.* 14 (11), 2947–2950. doi:10.1016/j.bmcl.2004.03.035
- World Health Organization. (2020). Diabetes. Available at: <https://www.who.int/news-room/fact-sheets/detail/diabetes> [Accessed August 27, 2020].

- Ye, G. J., Lan, T., Huang, Z. X., Cheng, X. N., Cai, C. Y., Ding, S. M., et al. (2019). Design and synthesis of novel xanthone-triazole derivatives as potential antidiabetic agents: α -Glucosidase inhibition and glucose uptake promotion. *Eur. J. Med. Chem.* 177, 362–373. doi:10.1016/j.ejmech.2019.05.045
- Zeng, Z., Yin, X., Wang, X., Yang, W., Liu, X., and Hong, Y. (2019). Synthesis of water soluble pentacyclic dihydroxyterpene carboxylic acid derivatives coupled amino acids and their inhibition activities on α -glucosidase. *Bioorg. Chem.* 86, 277–287. doi:10.1016/j.bioorg.2019.02.001

Conflict of Interest: Author JIZ was employed by the company Lunan Pharmaceutical Group Corporation.

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 © 2021 Liu, Hao, Bian, Ge, Lu, Xie, Wang, Tao, Yuan, Zhang, Zhang, Jiang and Zhu. 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.



Exploring the Interaction between Vancomycin/Teicoplanin and Receptor Binding Domain (RBD) of SARS-CoV-2

Wen-Tao Tao¹, Qian Yu¹, Ya-Li Li², Mei Ge², Yi-Lei Zhao¹ and Ting Shi^{1*}

¹State Key Laboratory of Microbial Metabolism, Joint International Research Laboratory of Metabolic and Developmental Sciences, School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai, China, ²Shanghai Laiyi Center for Biopharmaceutical R&D, Shanghai, China

OPEN ACCESS

Edited by:

Fei Ye,
Zhejiang Sci-Tech University, China

Reviewed by:

Fei Xia,
East China Normal University, China
Min Huang,
Northwestern Polytechnical
University, China

*Correspondence:

Ting Shi
tshi@sjtu.edu.cn

Specialty section:

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

Received: 10 December 2020

Accepted: 31 December 2020

Published: 19 January 2021

Citation:

Tao W-T, Yu Q, Li Y-L, Ge M, Zhao Y-L
and Shi T (2021) Exploring the
Interaction between Vancomycin/
Teicoplanin and Receptor Binding
Domain (RBD) of SARS-CoV-2.
Front. Chem. 8:639918.
doi: 10.3389/fchem.2020.639918

The recent pandemic caused by SARS-CoV-2 has spread to over 100 countries, infected more than 47 million people and resulted in more than 1.2 million deaths worldwide until October. It is well known that, the SARS-CoV-2 starts an infection by binding its Receptor Binding Domain (RBD) of spike protein to human Angiotensin converting enzyme 2 (ACE2) receptor, and strenuous efforts had been made to prevent the infection. However, no successful drugs or vaccines have appeared. Herein, molecular docking and molecular simulations were carried out to study the interaction between RBD and two glycopeptide antibiotics (Vancomycin and Teicoplanin). Key residues in binding pocket were highlighted and the binding free energies were calculated. Our results suggested that Vancomycin and Teicoplanin, as natural and accepted antibiotics, could block the interaction between RBD of spike protein and human ACE2 receptor, which might be developed to potential drugs against the SARS-CoV-2.

Keywords: SARS-CoV-2, RBD, ACE2, molecular docking, MD simulation

INTRODUCTION

Since the end of 2019, the epidemic disease caused by Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) has spread in more than 100 countries, causing over 47 million infections with more than 1.2 million deaths, which severely threatens the global public health, the economic development, and the social stability (Jiang et al., 2020). SARS-CoV-2 is a kind of novel enveloped positive-stranded RNA viruses. The SARS-CoV-2 genome can encode four major structural proteins (**Figure 1**): the spike (S) protein, the membrane (M) protein, the envelope (E) protein and the nucleocapsid (N) protein (Dutta et al. 2020). Among them, the S protein has attracted much attention, because it is a critical determinant of viral host range and tissue tropism, as well as a major inducer of host immune response (Walls et al. 2020).

The SARS-CoV-2 S protein consists of three segments: a large ectodomain (ED), a single-pass transmembrane anchor (TM) and a short intracellular (IC) tail (**Figure 1C**). The ED is a clove-shaped trimer, comprising three receptor-binding subunits (S1) heads and a trimeric membrane-fusion subunits (S2) stalk. The C-terminal domain of S1 (S1-CTD), also named as receptor binding domain (RBD), is responsible to recognize some protein receptors like ACE2, APN, and DPP4. Furthermore, the SARS-CoV-2 RBD contains a core structure and a receptor-binding motif (RBM) as two subdomains. The core structure is formed by a group of five-stranded antiparallel β -sheets and RBM contains another two short β -sheets and long loop areas (**Figure 1D**). The RBM presents a

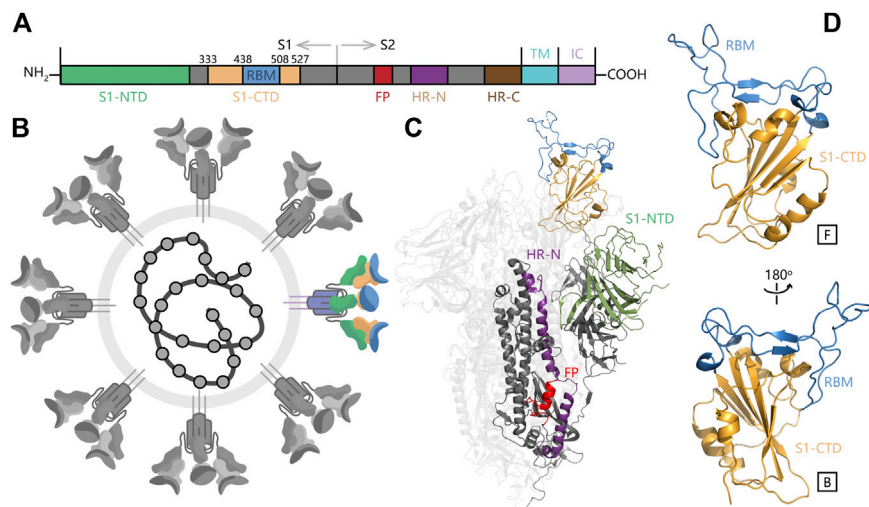


FIGURE 1 | The overall structure of SARS-CoV-2 S protein and RBM domain. **(A)** Important monomers in SARS-CoV-2 S protein. **(B)** Schematic of SARS-CoV-2 and S protein in color. **(C)** The overall structure of SARS-CoV-2 S protein **(D)** The structure of S1-NTD (RBD) in SARS-CoV-2 S protein.

bowl-like concave surface to accommodate the conformation of N-terminal α -helix in ACE2 for better binding (Li, 2016).

During the invasion process of SARS-CoV-2, S1 binds to the receptor on the host cell surface firstly and S2 fuses with the host cell membrane, allowing viral genomes to enter into the host cell (Walls et al., 2017). After that, viral genomes can encode N protein to promote RNA replication intracellularly. S protein is involved in receptor binding and membranes fusion that are indispensable steps in the coronavirus infectious cycle, so S protein has been considered to be the primary drug target for recent pandemic caused by SARS-CoV-2.

Although antibiotics become the effective therapy method of microbial infections after the introduction of sulfonamides and the discovery of penicillin in 1928, the overuse of them has resulted in the subsequent emergence of antibiotic-resistant bacteria, which dramatically reduces the therapeutic effect (Hutchings et al., 2019). Under this serious circumstance, antimicrobial peptides (AMP) has been quickly used as an alternative to antibiotics to treat bacterial and viral infections because of its natural source and high efficacy. The mechanism of AMPs against pathogens includes membrane permeabilization, membrane destabilization, inhibition of macromolecular synthesis, intracellular translocation of the peptide and inhibition of DNA/RNA/protein synthesis (Yeaman and Yount, 2003). A previous study has also shown that the peptide compounds Mo-CBP3-PepII and PepKAA can interact with the S1 and S2 domains of the S protein through molecular docking respectively (Souza et al., 2020). Peptides have been reported as the potential chemical compound to prevent SARS-CoV-2 from infecting the cell for the first time, which inspires us to study the role of glycopeptides in inhibiting the invasion of SARS-CoV-2 further.

Vancomycin (V) is a tricyclic glycopeptide antibiotic originally derived from the organism *Streptococcus Orientalis*.

As the first member of Vancomycin family, Vancomycin was soon approved by the Food and Drug Administration in 1978 as a therapy of penicillin-resistant *Staphylococcus aureus* infection (Levine, 2006). After introduced into the clinical use, it was soon replaced by the β -lactam antibiotics including methicillin, ethoxypenicillin and cloxacillin due to their higher efficiency and lower cytotoxicity. Since 2000, Vancomycin has returned as the clinical therapy of MRSA after the cytotoxic problem was solved. Nowadays, Vancomycin is still used to treat severe infections that cannot respond to other antibiotics (Pais et al., 2020).

Teicoplanin (T), another member of the Vancomycin family glycopeptide antibiotics, was first discovered in 1975 (Boger, 2001). The antibacterial spectrum and antibacterial activity of Teicoplanin were similar to those of Vancomycin (Figure 2), which is attributed to the same backbone structure. Compared with the structure of Vancomycin, the additional fat acid side chain in Teicoplanin is beneficial for penetrating cells more easily (Costa et al., 2011).

Considering the glycopeptide structure and the pre-existing therapeutic effect for pneumonia (Pais et al., 2020), Vancomycin and Teicoplanin were chosen as the potential compounds against SARS-CoV2. Their feasibility was investigated by exploring the interaction with SARS-CoV2-RBD and antiviral experiments in this study. We constructed the systems of SARS-CoV2-RBD with Vancomycin (RBD-V) and Teicoplanin (RBD-T) by molecular docking experiments. For each system, the most stable conformations were selected from 200 structures to perform 3×50 ns molecular dynamics (MD) simulations. Then the binding energy were calculated via MMPBSA method. The conformational characteristics were analyzed during MD simulations and key residues that play a critical role in different binding modes were highlighted. The strong and stable interaction between SARS-CoV2-RBD and Vancomycin/Teicoplanin may be beneficial to prevent receptor binding and membranes fusion process of SARS-CoV-2 infectious cycle.

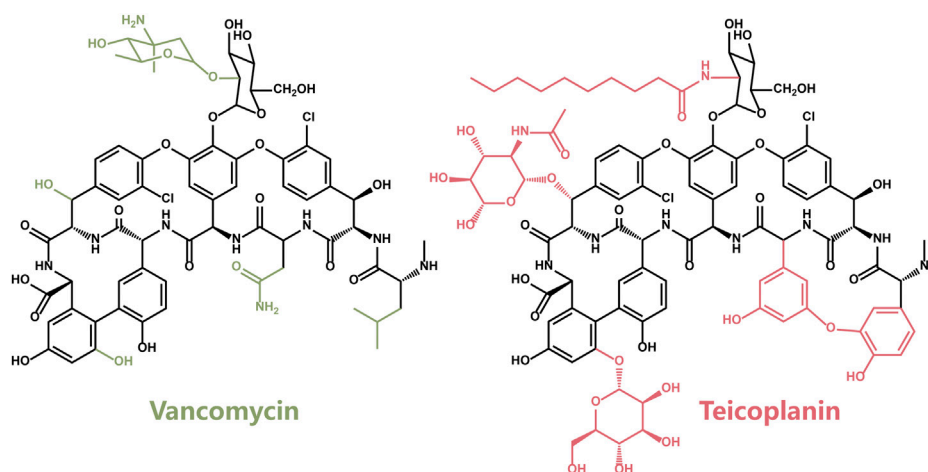


FIGURE 2 | The chemical structure of Vancomycin (V) and Teicoplanin (T).

METHODS

System Preparation

The crystal structure of SARS-CoV2-RBD was derived from the Protein Data Bank database (PDB ID: 6M0J) (Lan et al., 2020). We used H++ website (<http://biophysics.cs.vt.edu/>) to compute the pKa values of ionizable residues in RBD and determine their protonation states at pH = 7.0 (Anandakrishnan et al., 2012). The chemical structures of two glycopeptides, Vancomycin and Teicoplanin, were obtained from PubChem. Their three-dimensional structures were optimized by Gaussian16 (Steen et al., 2015) at m062x (Zhao and Truhlar, 2008)-D3 (Grimme et al., 2011)/6-31G(d)(Krishnan et al., 1980) level. Then the electrostatic surface potential (ESP) charges were calculated for force field preparation. Afterward, a two-step restrained electrostatic potential (RESP) (Wang et al., 2000) model was applied to determine the charges distribution on the substrates.

Docking

In this work, the AutoDock4.2 software (Goodsell et al., 1996) was utilized to build SARS-CoV-2 RBD-V complex and RBD-T complex. According to the binding position of RBD and ACE2, the map of 126,126,126 grid points in the point interval of 0.375 Å was set and calculated. The Lamarckian Genetic Algorithm (LGA) (Fuhrmann et al., 2010) was adopted to search for stable complexes. The number of runs and maximum energy evaluations were fixed at 200 and 2,500,000. Other parameters were set as default values. Finally, the results were ranked by docking energy. The top five for each system with the low binding energy were extracted to carry out classical MD simulations later.

Classical Molecular Dynamics Simulation

MD simulations were performed on RBD-V and RBD-T complexes using AMBER16 accelerated by GPU (Gotz et al., 2012; Le Grand et al., 2013; Salomon-Ferrer et al., 2013) under ff14SB force field (Maier et al., 2015). The complexes were solvated in an octahedral box of TIP3P (Mark and Nilsson, 2001) water molecules with the thickness of the external water layer exceeding 10 Å, totally 10,190 water molecules in RBD-V system and 10,120 water molecules in RBD-T system. To achieve the charge neutralization, two chloride ions were added in both systems.

Both solvated systems were firstly subjected to 10,000 steps of minimization followed by heating and equilibration cycles. In the heating cycle, the systems were gradually heated from 0 to 300 K through 25,000 iterations. After equilibration for 50 ps in the NPT ensemble, three 50 ns molecular dynamics simulation (300 K, 1 atm) were conducted for each pose of each system with different random seeds. The Particle Mesh Ewald (PME) method was employed to account for long-range electrostatic interactions, and the SHAKE algorithm in its matrix form was used to fix bonds and angles involving hydrogen atoms (Ryckaert et al. 1977). The cutoff of Van der Waals interactions was set to 10.0 Å. Each system included top five docking results and three parallel trajectories for each mode, 15 × 50 ns trajectories in total (Table 1). Then trajectories analyses were carried out using Cpptraj (Roe and Cheatham, 2013) in Ambertools18.

Trajectories Analysis

Root mean square deviation (RMSD) was widely used to measure the variation between two structures. To judge the change of protein structures, the coordinate difference of alpha carbon, carbon atom of carbonyl groups and nitrogen atom in the system

TABLE 1 | List of MD runs performed.

	RBD-V					RBD-T				
Mode	V1	V2	V3	V4	V5	T1	T2	T3	T4	T5
Time (ns)	3 × 50	3 × 50	3 × 50	3 × 50	3 × 50	3 × 50	3 × 50	3 × 50	3 × 50	3 × 50

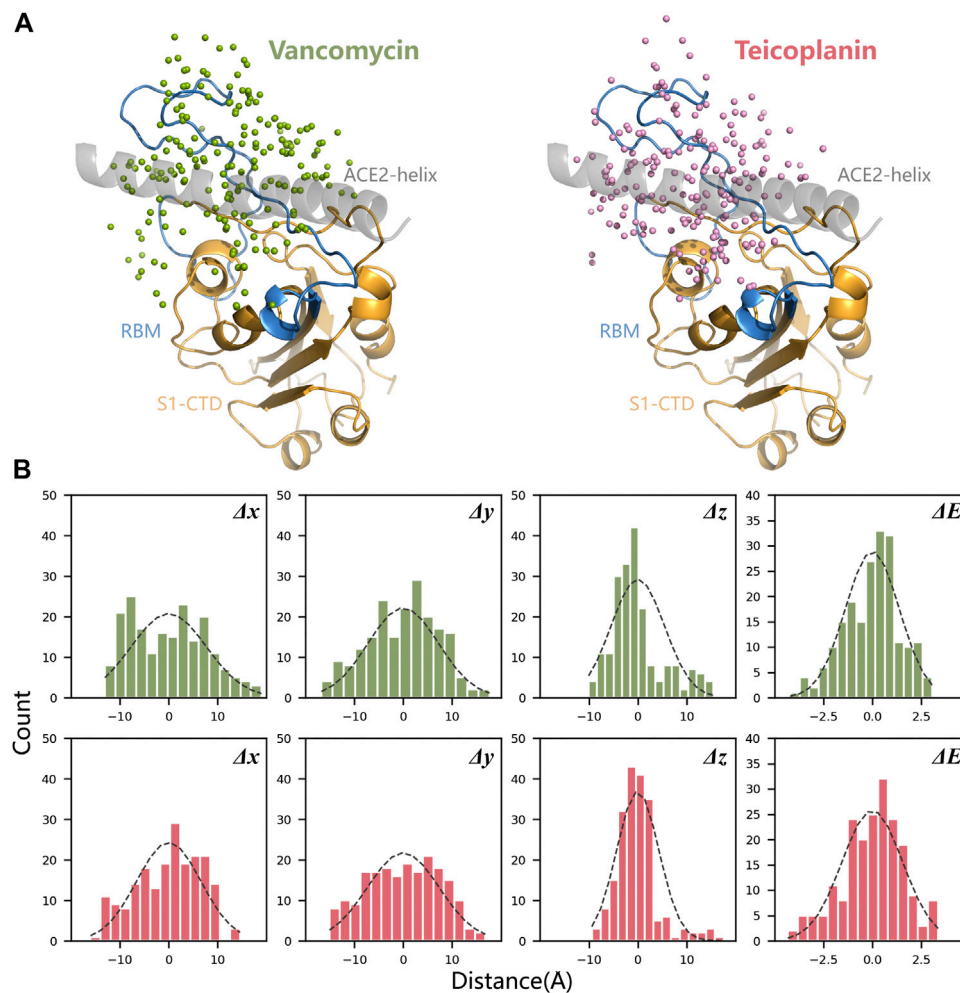


FIGURE 3 | Spatial position analysis of RBD-V and RBD-T complexes. **(A)** Centroids distribution of V and T in all docking modes were marked with green dots and pink dots respectively. N-terminal α -helix of ACE2 was displayed in grey, the core area of RBD (S1-CTD) was displayed in orange and RBM was displayed in blue. **(B)** The mean and variance of each three-dimensional coordinate vector (Δx , Δy , and Δz) and docking energy score (ΔE) were calculated and represented by histograms.

relative to the initial structure was calculated and averaged during the MD simulation. In the Eq. 1, d_i was the distance between the original and the present coordinates of atom i , and N was the atom number of the system to calculate the distance.

$$\text{RMSD} = \sqrt{\frac{\sum d_i^2}{N_{\text{atoms}}}} \quad (1)$$

Root mean square fluctuation (RMSF) represented the average variation of single alpha atom over time. In the Eq. 2, x_i means the coordinate of atom i and N means the number of frames to be calculated.

$$\text{RMSF}_i = \sqrt{\frac{1}{N_{\text{Frames}}} \sum_{j=1}^{N_{\text{Frames}}} [x_i(t_j) - \bar{x}_i]^2} \quad (2)$$

Binding Free Energy Calculation

Molecular Mechanics Poisson-Boltzmann Surface Area (MMPBSA) and Molecular Mechanics Generalized Born

Surface Area (MMGBSA) were two efficient methods to calculate the binding ability between ligands and receptors (Genheden and Ryde, 2015). The total binding energy was calculated on the basis of the difference of complex free energy and sum of the free energy of receptor plus ligand. The module anti-MMPBSA.py in AMBER18 was used to create three topology files (complex, receptor and ligand) for binding energy calculation (Miller et al., 2012). Each term on the right-hand side of Eq. 3 was estimated according to Eq. 4.

$$\Delta G_{\text{bind,solvated}} = \Delta G_{\text{complex,solvated}} - [\Delta G_{\text{receptor,solvated}} + \Delta G_{\text{ligand,solvated}}] \quad (3)$$

$$\Delta G_{\text{solvated}} \cong \frac{1}{N} \sum_i E_{i,\text{gas}} + \frac{1}{N} \sum_i \Delta G_{i,\text{solvation}} - \frac{T}{N} \sum_i S_{i,\text{solvate}} \quad (4)$$

In this study, one 50 ns MD simulation trajectory of each mode of each system was used to calculate the binding free energy. Each trajectory consists of 25,000 frames.

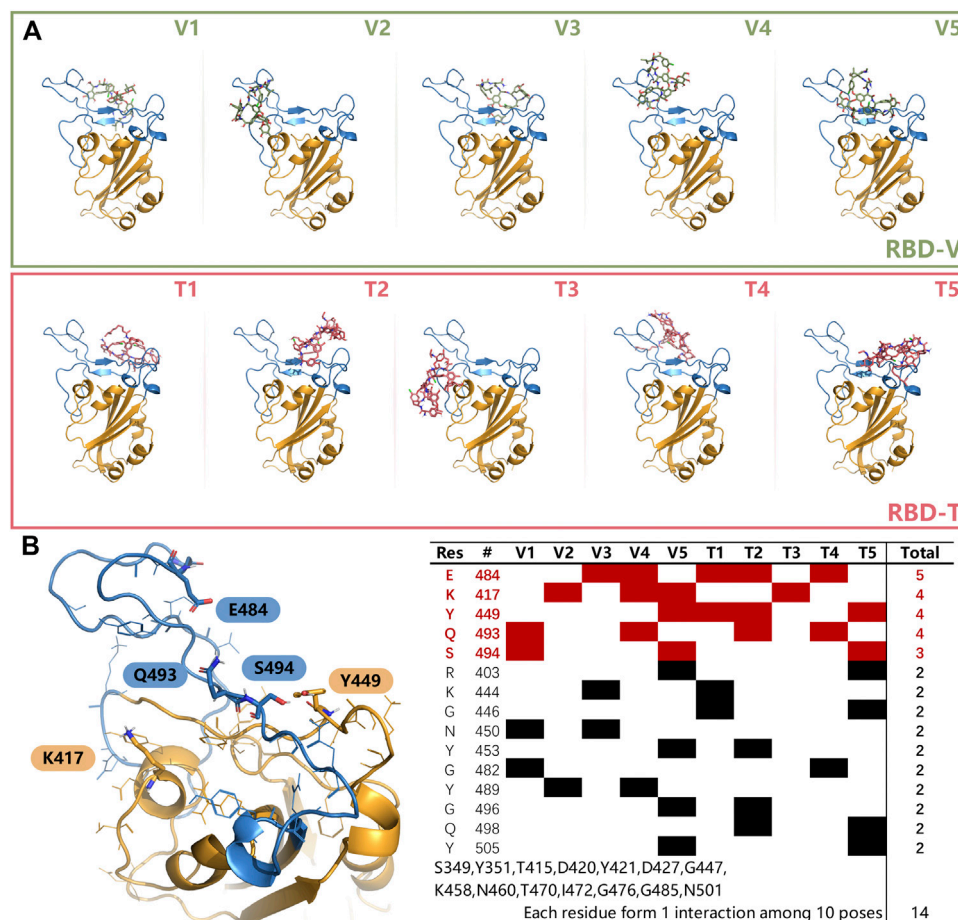


FIGURE 4 | (A) The top 5 modes obtained from docking experiments of RBD-V and RBD-T system. **(B)** Residues that could form interactions more than twice in top five modes with V and T were listed in the table. The residues, appeared once, were merged together on the bottom. Key residues K417, Y449, E484, Q493, and S494 were shown in sticks in left. The color of RBD is as the same as **Figure 3**.

RESULTS

Docking Results of SARS-CoV-2 RBD and Vancomycin/Teicoplanin

According to the binding mode of SARS-CoV-2 RBD and human ACE2, a docking box was set at the top of RBD. After 200 times molecular docking in each system, docking modes were ranked by binding energy in **Supplementary Figure S1**. The sampling of the docking structures was sufficient according to the distribution of binding energy and normal distribution of each three-dimensional coordinate vector (**Figure 3**). The flexible binding positions of T and V to ACE2 implied that they could efficiently block the interactions of RBD with human ACE2 receptor by occupying their interface.

Next, top five structures with the low energy were collected for each system and shown in **Figure 4A**. It can be seen that the receptor-binding motif (RBM) can easily accommodated V and T because of its bowl-like concave surface, except for T3, which seems escaped from the bowl. After counting up all polar interaction between ligands and RBD in those docking modes,

five important residues were found and shown in **Figure 4B**. The hydrogen bonds formed by K417, Y449, E484, Q493 and S494 and existed in more than three docking modes. These residues primarily concentrated on RBM region and had been reported widely to play a key role in RBM binding. Also, other residues listed in **Figure 4B** might be involved in the recognition process between two ligands and RBD. These results suggested that our docking results are reliable.

Structural Analysis of Receptor-Binding Motif

To study the structural characteristic of RBD-V and RBD-T complexes, the top five structures mentioned above were utilized to perform MD simulations in each system (**Supplementary Figure S2**). The RMSD of the RBD-V system was shown in **Supplementary Figure S3**. In both systems, the RMSD values with slight fluctuations (mean is 1.63 ± 0.31 Å in RBD-V system and 1.55 ± 0.25 Å in RBD-T system) suggested that the initial structures of RBD-V system and RBD-T system

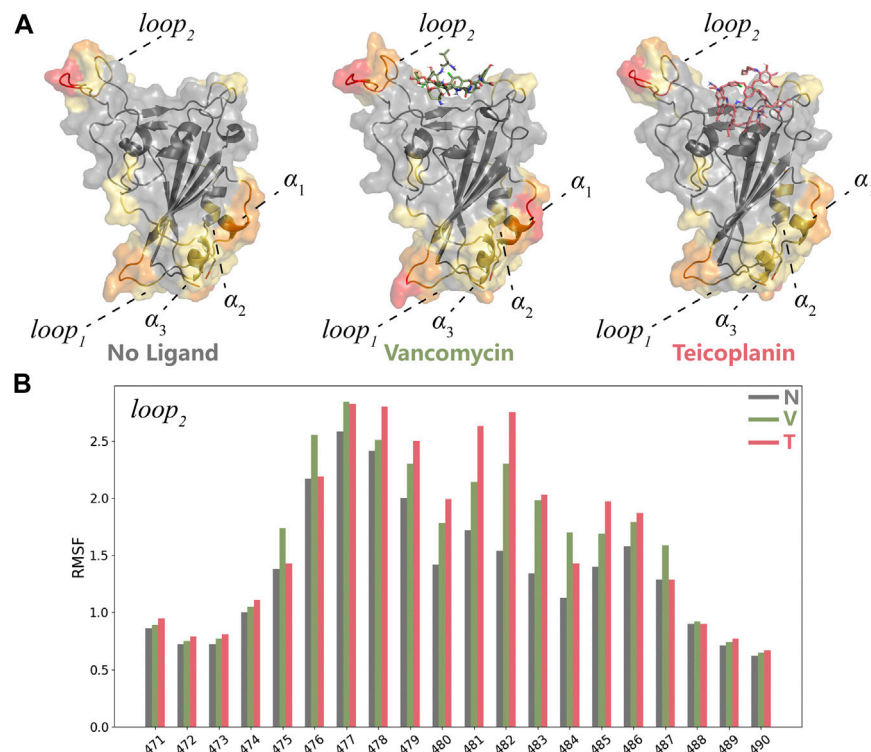


FIGURE 5 | MD simulation analyses of RBD-V and RBD-T (A) The representative conformations of the system without ligand, RBD-V system and RBD-T system according to the cluster analysis. The structures were colored by RMSF value. (B) The histogram of RMSF distribution. The RMSF value of each residue in the system without ligand (N), RBD-V (V) system and RBD-T (T) system was colored by grey, green and red respectively.

obtained by docking experiments were suitable and the binding states reached their respective equilibria during 50 ns MD simulations.

The structural fluctuations of RBD in two systems were depicted in **Supplementary Figure S3**. Based on the RMSF analysis, residues located on the top and the bottom regions of RBD including $loop_1$, $loop_2$, and α_1 were highlighted (**Figure 5A**). According to the structure of the whole S protein (**Figure 1C**), $loop_1$ and α_2 played important roles in the interaction of RBD with S1-NTD. The loss of S1-NTD during MD simulations might lead to the instability of these areas. Compared with the structure of RBD protein without ligands, the same RMSF variations in two systems meant the structure flexibilities of $loop_1$ and α_1 -3 were not caused by the existence of ligands. And $loop_2$ exhibited more remarkable structural flexibilities in RBD-V and RBD-T systems (**Figure 5B**), which suggested the existence of ligands could influence the conformational change of RBD protein. Besides $loop_2$, V445, G446 and T500 residues at RBM also showed flexible structure movements in both RBD-V and RBD-T systems. Furthermore, two systems exhibited the similar structural movement tendency, suggesting RBD consisted of a stable core structure and the flexible RBM region responsible for binding with V and T.

In a word, V and T could bind with RBD steadily and influence the key residues in RBM involved in the recognition of ACE2.

Therefore, more attention would be paid to RBM in the subsequent studies.

Linear Regression Analysis between MMPBSA Binding Free Energy and Interactions

Considering that RBM was a wide binding surface, which was much different from the traditional binding site. Also, V and T were large amphiphilic ligands. The binding modes of RBD-V and RBD-T systems were not only a single stable configuration, so conformers obtained by MD simulations every 1 ns were collected to study their structural characteristics. Hydrogen bond (HB) and hydrophobic (HP) interactions were studied. The number of the residues that interacted with V and T was counted for each system (**Supplementary Table S3**) and the mean was shown in **Figure 6A**. Compared with the chemical structure of V, T had more glycosylated modifications and a saturated nine-membered carbon chain which could increase the possibility of forming HB and HP interactions with RBD. During 50 ns MD simulations, T could form HP interaction with 3.25 residues in average, while 2.95 residues in RBD-V system. Similarly, more glycosylated modifications and polar functional group were beneficial for T to form nearly 0.5 HB interactions more than V. In a conclusion, T could bind with RBD

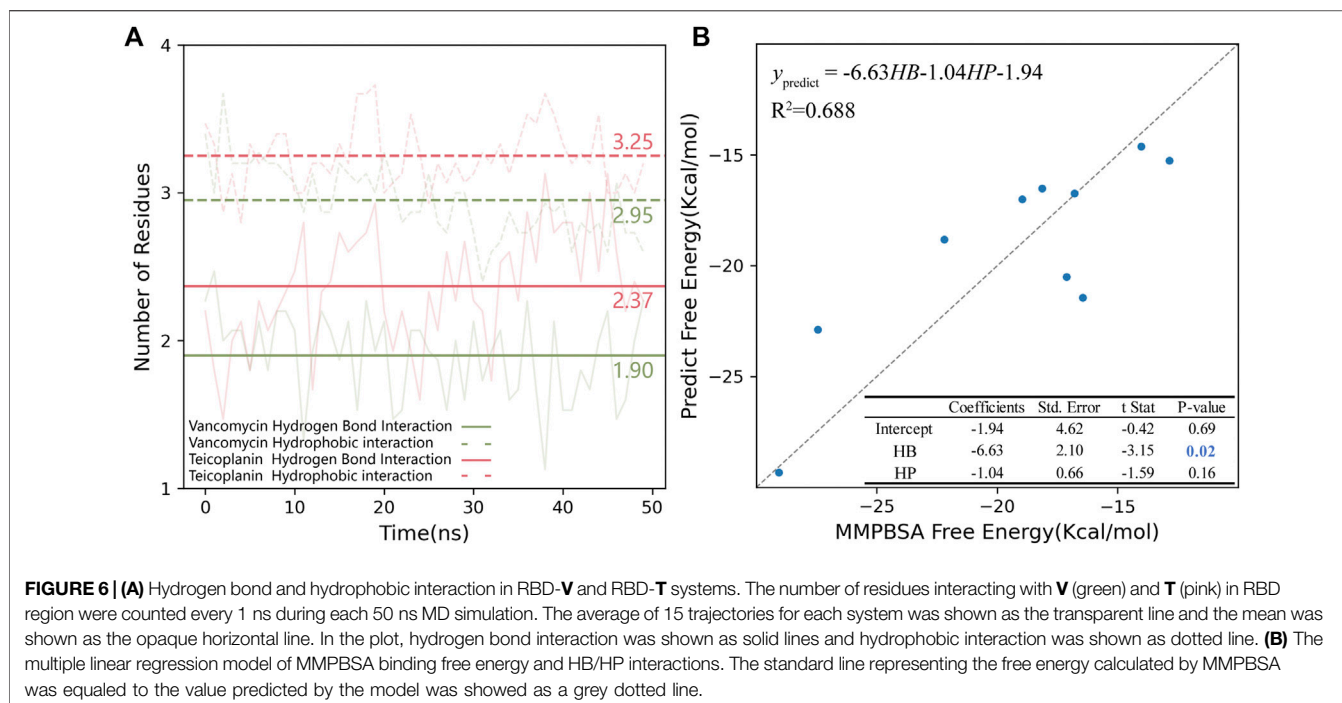


TABLE 2 | Binding free energy calculated via MMPBSA.

System	State	MMPBSA (kcal/mol)			Ave.(Sta)	Ave.(Sys)
		MD1	MD2	MD3		
RBD-V	V1	-19.4	-21.3	-10.7	-17.1	-17.9
	V2	-25.8	-19.7	-11.5	-18.0	
	V3	-21.5	-20.4	-12.6	-18.1	
	V4	-13.7	-15.5	-9.4	-12.9	
	V5	-13.5	-29.3	-23.8	-22.2	
RBD-T	T1	-30.5	-33.6	-24.7	-29.6	-20.9
	T2	-25.9	-37.0	-19.4	-27.4	
	T3	-9.8	-27.4	-12.2	-16.5	
	T4	-13.1	-14.4	-22.9	-16.8	
	T5	-10.9	-13.8	-17.4	-14.0	

more closely than **V**. What's more, five residues in average could interact with these glycopeptide ligands for each system, indicating the existence of effective binding.

The binding energy was calculated via MMPBSA method on the top five complexes and the results were all shown in **Table 2**. Comparing the average binding energy of two systems, we found RBD-**T** system was lower than RBD-**V** by 3.0 kcal/mol, which validated the aforementioned conclusion that **T** could bind with RBD more closely than **V**. Then a line regression analysis was performed to investigate the correlation between MMPBSA binding free energies and HB as well as HP interactions (**Figure 6B**). The squared correlation coefficient (R^2), also known as the coefficient of determination, was 0.688 that means the positive relationship between the binding free energy and the interactions. The coefficient between the free energy and the number of residues forming HB interaction was equaled to -6.63 with the p -value below 0.05. The binding

energy contributed by HB interaction lay within the reasonable range (Emamian et al., 2019). Although the P -value of the coefficient between the free energy and the number of residues forming HP interaction was larger than 0.05, which was possibly attributed to the obscure definition of HP interaction. Besides, the tendency that the binding free energy was lower with more HP interaction was in accordance with our knowledge.

Diverse binding modes led to different binding energies. The regression analysis above showed that lower energy values were caused by better conformations with more HB and HP interactions. The average binding energies of RBD-**V5** complex and RBD-**T1** complex are the lowest in each system because of more HB and HP interactions, so the key residues of the two modes were explored subsequently.

Key Residues Found through Dynamic Hydrogen Bond Analysis and MMPBSA Decomposition

Basing on the results mentioned above, HB interaction showed the positive correlation with the binding free energy. Therefore, we counted up all HBs in RBD-**V** and RBD-**T** systems and the survival time was used to define the strength of these HBs to search important residues.

The HB whose survival time was longer than 1% MD simulation time was listed as dots in **Figure 7**. As we can see, most dots were concentrated at RBM. The HB distribution in different modes was complicated and dissimilar, but Y449, E484, S494, and Y501 interacted with glycopeptide ligands in most modes. Combined with the results of MMPBSA decomposition (**Supplementary Table S3**), the key residues were found out.

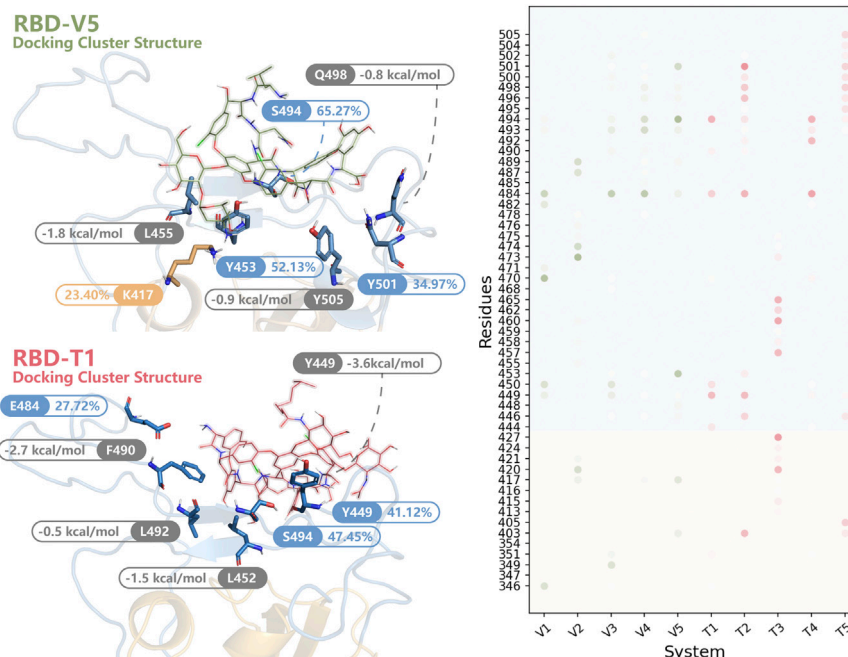


FIGURE 7 | The key residues in RBD-V and RBD-T systems. In the structures of RBD-V5 and RBD-T1, residues forming HBs for more than 20% MD simulation time (blue and orange) were highlighted. Besides, residues listed in the top 10 of the MMPBSA decomposition (grey) were shown. All residues forming HBs in more than 1% modes were summarized in a dot matrix where the transparency was proportional to the survival time. The dots in RBD-V system were colored in green and the dots of RBD-T system were colored in pink.

In RBD-V5 complex, K417, Y453, S494 and Y501 stabilized V by long-time HB interaction. These HBs formed by K417, Y453, S494, and Y501 maintained 23.40, 52.13, 65.27, and 34.97% of MD simulation time respectively. Meanwhile, L455, Q498, and Y505 could be helpful to lower the binding free energy by interacting with V.

Except S494, the interaction mode in RBD-T1 was significantly different from that of RBD-V5 complex. The HBs formed by Y449, E484 and S494 could retain 41.12, 27.72, and 47.45% of MD simulation time respectively. Furthermore, L452 donated -1.5 kcal/mol, F490 donated -2.7 kcal/mol and L492 donated -0.5 kcal/mol to stabilize the binding mode between T and RBD.

The importance also showed through alanine scanning base on the wild type trajectories (Liu et al., 2018). The difference of the MMGBSA energies were shown in **Supplementary Table S4**. Except L492, the mutation of the rest residues showed energy difference more than 0.5 kcal/mol in at least one system, indicating their important roles in stabilizing the ligands with RBD.

Aforementioned K417, Y449, Y453, L455, Q498, Y501, and Y505 had been reported to play a crucial part in the binding with ACE2 (Hussain et al., 2020; Lan et al., 2020; Veeramachaneni et al., 2020). The remaining key residues we found including L452, E484, F490, L492, and S494 were adjacent to these important residues associated with ACE2. The binding site of V and T was similar to that of ACE2, so V and T were thought to be able to prevent the infection of SARS-CoV2 by competitive binding.

The Antiviral Potential of Vancomycin and Teicoplanin

Besides all the computational analyses mentioned above, antiviral experiments were performed to confirm our proposal. We used pseudovirus to simulate the SARS-CoV-2 under 50 μ M Vancomycin or Teicoplanin. The antiviral potential of V and T were tested through an antiviral sever. The results revealed that V and T respectively with 83 and 87% being antiviral against SARS-CoV-2. (**Supplementary Table S2**).

CONCLUSION

In conclusion, two glycopeptides, Vancomycin and Teicoplanin, were investigated as potential antiviral molecules by analyzing the interactions with SARS-CoV-2 RBD of spike protein. Firstly, 200 times docking experiments were used to choose the proper binding modes and K417, Y449, E484, Q493, and S494 were found out as key residues to stabilize the ligands with RBD. After classical MD simulations of top five structures with low energy, the RMSD analysis showed the binding states reached their equilibriums and the RMSF analysis showed both ligands could cause structural change of loop2 by binding with RBD. Afterwards, the number of hydrogen bonds and hydrophobic interactions was counted as independent variable and the binding free energy was calculated as dependent variable. The linear regression analysis between binding free energy and

interactions suggested that hydrogen bonds played important roles in maintaining the stability of the conformation. Also compared two systems, More glycosylated modifications and the fat acid side chain in T increases the possibility of forming more hydrogen bonds and hydrophobic interactions with RBD. Furthermore, dynamic hydrogen bond analysis and MMPBSA decomposition helped us to find out more important residues. With the help of alanine scanning analysis of those important residues, we more convinced that they played an important role of enhancing the combination of two ligands and RBD. Part of those key residues were also found to bind with ACE2 reported in other studies (Hussain et al., 2020; Lan et al., 2020; Veeramachaneni et al., 2020), such as K417, Y449, Y453, L455, Q498, Y501 and Y505. Although the remaining key residues such as L452, E484, F490, L492, and S494 did not directly forming interactions, they were adjacent to these important residues associated with ACE2. Based on those results mentioned above, we implied that Vancomycin and Teicoplanin could be the potential competitive inhibitor to block the binding of SARS-CoV-2 RBD and ACE2, and further prevent the SARS-CoV-2 infectious cycle.

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.

REFERENCES

- 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
- Boger, D. L. (2001). Vancomycin, teicoplanin, and ramoplanin: synthetic and mechanistic studies. *Med. Res. Rev.* 21, 356–381. doi:10.1002/med.1014
- Costa, F., Carvalho, I. F., Montelaro, R. C., Gomes, P., and Martins, M. C. (2011). Covalent immobilization of antimicrobial peptides (AMPs) onto biomaterial surfaces. *Acta Biomater.* 7, 1431–1440. doi:10.1016/j.actbio.2010.11.005
- Dutta, N. K., Mazumdar, K., and Gordy, J. T. (2020). The nucleocapsid protein of SARS-CoV-2: a target for vaccine development. *J. Virol.* 94, e00647–20. doi:10.1128/JVI.00647-20
- Emamian, S., Lu, T., Kruse, H., and Emamian, H. (2019). Exploring nature and predicting strength of hydrogen bonds: a correlation analysis between atoms-in-molecules descriptors, binding energies, and energy components of symmetry-adapted perturbation theory. *J. Comput. Chem.* 40 (40), 2868–2881. doi:10.1002/jcc.26068
- Fuhrmann, J., Rurainski, A., Lenhof, H. P., and Neumann, D. (2010). A new Lamarckian genetic algorithm for flexible ligand-receptor docking. *J. Comput. Chem.* 31, 1911–1918. doi:10.1002/jcc.21478
- Genheden, S., and Ryde, U. (2015). The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. *Expet. Opin. Drug Discov.* 10, 449–461. doi:10.1517/17460441.2015.1032936
- Goodsell, D. S., Morris, G. M., and Olson, A. J. (1996). Automated docking of flexible ligands: applications of AutoDock. *J. Mol. Recogn.* 9, 1–5. doi:10.1002/(SICI)1099-1352(199601)9:1<1::AID-JMR241>3.0.CO;2-6
- 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
- Grimme, S., Ehrlich, S., and Goerigk, L. (2011). Effect of the damping function in dispersion corrected density functional theory. *J. Comput. Chem.* 32, 1456–1465. doi:10.1002/jcc.21759
- Hussain, M., Jabeen, N., Raza, F., Shabbir, S., Baig, A. A., Amanullah, A., et al. (2020). Structural variations in human ACE2 may influence its binding with SARS-CoV-2 spike protein. *J. Med. Virol.* 92, 1580–1586. doi:10.1002/jmv.25832
- Hutchings, M. I., Truman, A. W., and Wilkinson, B. (2019). Antibiotics: past, present and future. *Curr. Opin. Microbiol.* 51, 72–80. doi:10.1016/j.mib.2019.10.008
- Jiang, S., Hillyer, C., and Du, L. (2020). Neutralizing antibodies against SARS-CoV-2 and other human coronaviruses: (trends in immunology 41, 355–359; 2020). *Trends Immunol.* 41, 545–359. doi:10.1016/j.it.2020.04.008
- Krishnan, R., Binkley, J. S., Seeger, R., and Pople, J. A. (1980). Self-consistent molecular orbital methods. XX. A basis set for correlated wave functions. *J. Chem. Phys.* 72, 650–654. doi:10.1063/1.438955
- 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 Grand, S., Götz, A. W., and Walker, R. C. (2013). SPFP: speed without compromise-A mixed precision model for GPU accelerated molecular dynamics simulations. *Comput. Phys. Commun.* 184, 374–380. doi:10.1016/j.cpc.2012.09.022
- Levine, D. P. (2006). Vancomycin: a history. *Clin. Infect Dis.* (42 Suppl. 1), S5–S12. doi:10.1086/491709
- Li, F. (2016). Structure, function, and evolution of coronavirus spike proteins. *Annu. Rev. Virol.* 3 (3), 237–261. doi:10.1146/annurev-virology-110615-042301
- Liu, X., Peng, L., Zhou, Y., Zhang, Y., and Zhang, J. Z. H. (2018). Computational alanine scanning with interaction entropy for protein-ligand binding free energies. *J. Chem. Theory Comput.* Mar. 13 (14), 1772–1780. doi:10.1021/acs.jctc.7b01295

AUTHOR CONTRIBUTIONS

Conceptualization, TS, Y-LZ, and MG; formal analysis, W-TT; experiment, Y-LL and MG; funding acquisition, TS, Y-LZ, and MG; investigation, W-TT; methodology, TS, Y-LZ, and MG; writing-original draft, W-TT, QY, and TS; writing-review and editing, W-TT and TS.

FUNDING

This research was supported by the National Key R&D Program of China (No. 2019YFA0905400), the National Natural Science Foundation of China (Nos. 32070041, 31770070, and 31970041) and the Natural Science Foundation of Shanghai (No. 19ZR1427300 and 20S11900900).

ACKNOWLEDGMENTS

The authors thank SJTU-ZH2018ZDA26 for financial supports and SJTU-HPC computing facility award for computational hours.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fchem.2020.639918/full#supplementary-material>.

- 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. Theor. Comput.* 11 (11), 3696–3713. doi:10.1021/acs.jctc.5b00255
- Mark, P., and Nilsson, L. (2001). Structure and dynamics of the TIP3P, SPC, and SPC/E water models at 298 K. *J. Phys. Chem. B* 105, 24a. doi:10.1021/jp003020w
- 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. Theor. Comput.* 8 (8), 3314–3321. doi:10.1021/ct300418h
- Pais, G. M., Liu, J., Zepcan, S., Avedissian, S. N., Rhodes, N. J., Downes, K. J., et al. (2020). Vancomycin-Induced kidney injury: animal models of toxicodynamics, mechanisms of injury, human translation, and potential strategies for prevention. *Pharmacotherapy* 40, 438–454. doi:10.1002/phar.2388
- Roe, D. R., and Cheatham, T. E., 3rd (2013). PTRAJ and CPPTRAJ: software for processing and analysis of molecular dynamics trajectory Data. *J. Chem. Theor. Comput.* 9 (9), 3084–3095. doi:10.1021/ct400341p
- 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
- 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. Theor. Comput.* 9 (9), 3878–3888. doi:10.1021/ct400314y
- Souza, P. F. N., Lopes, F. E. S., Amaral, J. L., Freitas, C. D. T., and Oliveira, J. T. A. (2020). A molecular docking study revealed that synthetic peptides induced conformational changes in the structure of SARS-CoV-2 spike glycoprotein, disrupting the interaction with human ACE2 receptor. *Int. J. Biol. Macromol.* 164, 66–76. doi:10.1016/j.ijbiomac.2020.07.174
- Steen, G. W., Fuchs, E. C., Wexler, A. D., and Offerhaus, H. L. (2015). Identification and quantification of 16 inorganic ions in water by Gaussian curve fitting of near-infrared difference absorbance spectra. *Appl. Optic.* 54, 5937–5942. doi:10.1364/ao.54.005937
- Veeramachaneni, G. K., Thunuguntla, V. B. S. C., Bobbillaipati, J., and Bondili, J. S. (2020). Structural and simulation analysis of hotspot residues interactions of SARS-CoV 2 with human ACE2 receptor. *J. Biomol. Struct. Dyn.* 4, 1–11. doi:10.1080/07391102.2020.1773318
- 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–e6. doi:10.1016/j.cell.2020.02.058
- Walls, A. C., Tortorici, M. A., Snijder, J., Xiong, X., Bosch, B. J., Rey, F. A., et al. (2017). Tectonic conformational changes of a coronavirus spike glycoprotein promote membrane fusion. *Proc. Natl. Acad. Sci. U.S.A.* 114, 11157–11162. doi:10.1073/pnas.1708727114
- Wang, J., Cieplak, P., and Kollman, P. A. (2000). How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules?. *J. Comput. Chem.* 21, 1049–1074. doi:10.1002/1096-987x(200009)21:12<1049::aid-jcc3>3.0.co;2-f
- Yeaman, M. R., and Yount, N. Y. (2003). Mechanisms of antimicrobial peptide action and resistance. *Pharmacol. Rev.* 55, 27–55. doi:10.1124/pr.55.1.2
- Zhao, Y., and Truhlar, D. G. (2008). The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: two new functionals and systematic testing of four M06-class functionals and 12 other functionals. *Theor. Chem. Account.* 120, 215–241. doi:10.1007/s00214-007-0310-x

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 Tao, Yu, Li, Ge, Zhao and Shi. 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.



Revisiting Activity of Some Nocodazole Analogues as a Potential Anticancer Drugs Using Molecular Docking and DFT Calculations

Muhammad Khattab^{1*} and Ahmed A. Al-Karmalawy^{2*}

¹Department of Chemistry of Natural and Microbial Products, Division of Pharmaceutical and Drug Industries, National Research Centre, Cairo, Egypt, ²Department of Pharmaceutical Medicinal Chemistry, Faculty of Pharmacy, Horus University-Egypt, New Damietta, Egypt

OPEN ACCESS

Edited by:

Zhongjie Liang,
Center for Systems Biology, Soochow
University, China

Reviewed by:

Daniel Glossman-Mitnik,
Advanced Materials Research Center
(CIMAV), Mexico
Geng Dong,
College of Medicine, Shantou
University, China

*Correspondence:

Muhammad Khattab
mkhattab@swin.edu.au
Ahmed A. Al-Karmalawy
akarmalawy@horus.edu.eg

Specialty section:

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

Received: 11 November 2020

Accepted: 02 February 2021

Published: 24 March 2021

Citation:

Khattab M and Al-Karmalawy AA
(2021) Revisiting Activity of Some
Nocodazole Analogues as a Potential
Anticancer Drugs Using Molecular
Docking and DFT Calculations.
Front. Chem. 9:628398.
doi: 10.3389/fchem.2021.628398

Although potential anticancer activities of benzimidazole-based anthelmintic drugs have been approved by preclinical and clinical studies, modes of binding interactions have not been reported so far. Therefore, in this study, we aimed to propose binding interactions of some benzimidazole-based anthelmintics with one of the most important cancer targets (Tubulin protein). Studied drugs were selected based on their structural similarity with the cocrystallized ligand (Nocodazole) with tubulin protein. Quantum mechanics calculations were also employed for characterization of electronic configuration of studied drugs at the atomic and molecular level. Order of binding affinities of tested benzimidazole drugs toward colchicine binding site on tubulin protein is as follows: Flubendazole > Oxfendazole > Nocodazole > Mebendazole > Albendazole > Oxibendazole > Fenbendazole > Ciclobendazole > Thiabendazole > Bendazole. By analyzing binding mode and hydrogen bond length between the nine studied benzimidazole drugs and colchicine binding site, Flubendazole was found to bind more efficiently with tubulin protein than other benzimidazole derivatives. The quantum mechanics studies showed that the electron density of HOMO of Flubendazole and Mebendazole together with their MEP map are quite similar to that of Nocodazole which is also consistent with the calculated binding affinities. Our study has ramifications for considering the repurposing of Flubendazole as a promising anticancer candidate.

Keywords: anthelmintic, anticancer, tubulin inhibitors, molecular docking, DFT calculations, drug repurposing

INTRODUCTION

Microtubules play a key role in the invasion and metastatic spread of tumor cells, depending on its crucial roles in mitosis, signaling, trafficking, proliferation, and migration of eukaryotic cells (Honore, Pasquier and Braguer, 2005). Drugs targeting microtubular proteins constitute a major and promising anticancer drug category exhibiting both antimetabolic and antiangiogenic properties, besides inhibiting tumor progression in cancer and endothelial cells (D Katsetos and Draber, 2012). Colchicine binding site (CBS) is one of the five important identified binding sites on tubulin protein with the longest history of research as an anticancer target (Wang and Zhang, 2016).

The success rate for new anticancer drugs from Phase I trial to commercial use by FDA approval is estimated to be around 6.7% from 2003 to 2011, taking about 8.3 years as an average (Hay et al., 2014;

Takebe, Imai and Ono, 2018). The overall numbers of cancer deaths massively grow, rendering it the leading cause of death across all age groups by 2020 (Weir et al., 2016). Therefore, there is a great global need for rapidly approved and effective anticancer drug candidates.

Computational drug repurposing, a new area of drug repurposing, has been intensively developed due to breakthrough advances in fields of molecular, genomic and phenotypic data of pharmacological compounds (Park, 2019). Drug repurposing is an accelerated tool for drug development by seeking new indications for already approved drugs rather than discovering *de novo* drug compounds and constitutes nowadays 30% of the newly marked drugs in the United States (Pushpakom et al., 2018; Park, 2019). Many successful repurposed drugs have been introduced by FDA as in case of Aspirin, used for the treatment of stroke and/or myocardial infarction, Topiramate, used for the treatment of obesity, and Mifepristone, used for the treatment of hyperglycaemia in Cushing's syndrome (Polamreddy and Gattu, 2019).

Benzimidazole nucleus is a pharmacophore in lots of bioactive heterocyclic compounds with a wide range of biological and clinical applications. Moreover, benzimidazole derivatives constitute the isosteric structures of many naturally occurring nucleotides, which allows interacting easily with the living system biopolymers (Narasimhan et al., 2012). As a result, it constitutes a large group of drugs exhibiting various therapeutic activities such as anthelmintic, antiviral, antihypertensive, antioxidant, and anticancer (Yadav and Ganguly, 2015). Benzimidazole anthelmintic drugs are widely used due to their low toxicity and high efficiency against a wide range of helminth species (Köhler, 2001). Their mechanism of action is based on specific binding to tubulin subunit of microtubular protein, which results in a disruption of its structure and function (Lacey, 1990).

Based on the aforementioned data, benzimidazole derivatives became an attractive target for drug repurposing trials. The cytotoxic studies on benzimidazole-based drugs revealed their potential activity as colchicine binding site inhibitors (CBSIs) (Wang and Zhang, 2016; Al-Karmalawy and Khatab, 2020). Despite none of the benzimidazole-based drugs have been granted food and drug administration (FDA) approval for targeting the colchicine binding site on tubulin till now, the preclinical and clinical studies revealed the CBSIs are less susceptible to drug resistance development rendering them a potential target for cancer treatment (McLoughlin and O'Boyle, 2020).

Benzimidazole anthelmintic drugs having over 40 years of safe use as over the counter medications (Geary et al., 2010). They are meeting many characteristic features to be desirable for repurposing such as well-known safety profiles, well-described pharmacokinetic studies, and low prices (Biodegradabilní and Konjugátů, 2015). *In vitro* and *in vivo* studies revealed the potential of some members of benzimidazole anthelmintic drugs to suppress tumor progress through inhibition of multiple biological targets such as tubulin polymerization and angiogenesis (Mukhopadhyay et al., 2002; Králová et al., 2013; Raghunath and Viswanathan, 2014; Guerini et al., 2019).

Nocodazole (NZO) is a benzimidazole-based experimental drug targeting both protein kinases and microtubules. It is used as a lead compound for the discovery of novel CBSIs

(Geary et al., 2010). In this manuscript, we propose modes of binding interactions between nine benzimidazole-based drugs in comparison with the reference cocrystallized NZO drug (Figure 1) at colchicine binding site of tubulin protein. The first seven members of the nine elected benzimidazole anthelmintic drugs (1–7) exhibit structural similarity (benzimidazole core and carbamate moiety) with NZO. Studies reported here is a continuation to our previous work on NZO and a benzimidazole-based anthelmintic drug (Mebendazole) (Al-Karmalawy and Khatab, 2020; Khatab, 2020).

METHODOLOGY

Docking Studies

Tested Compounds Optimization

Docking studies using the Molecular Operating Environment (MOE) software package (Inc., 2016) were performed to evaluate the activities and binding modes of nine benzimidazole anthelmintic drugs compared to NZO complexed with CBS (Wang and Zhang, 2016). Mebendazole (1), Albendazole (2), Ciclobendazole (3), Fenbendazole (4), Flubendazole (5), Oxibendazole (6), Oxfendazole (7), Thiabendazole (8) and Bendazole (9) were downloaded from the PubChem website (<https://pubchem.ncbi.nlm.nih.gov/>). The nine tested compounds underwent energy minimization after examining the structure and the formal charges on atoms using a 2D depiction model. The partial charges were also automatically calculated. Structure of the co-crystallized NZO obtained from two subunits of the target tubulin protein (B and D subunits) and the nine tested benzimidazole drugs were imported together in the same database and saved in the form of MDB file to be docked in two separate processes for each subunit pocket, B and D, respectively.

Target Tubulin Active Site Optimization

The X-ray structure for tubulin protein (PDB: 5CA1) complexed with the native NZO was obtained from the Protein Data Bank (<http://www.rcsb.org/>). All steps for the preparation of the target protein for docking calculations were done. The addition of hydrogen atoms with their standard 3D geometry to the system, automatic correction to check for any errors in the connections and types of atoms, and fixation of the potential of the receptor were also performed. The selection of the same active site of co-crystallized inhibitor in the protein structure was done by using Site Finder where dummy atoms were created at the same binding site of the pocket.

Docking of Molecules into Colchicine Binding Site of Tubulin

Docking of the database composed of the nine benzimidazole drugs together with NZO was performed. The following methodology was applied: the prepared protein active site file was loaded, and then the general and template docking processes using dock tools were initiated. Dummy atoms act as the docking site, triangle matcher is the placement methodology, London dG is the scoring methodology, rigid receptor represents the

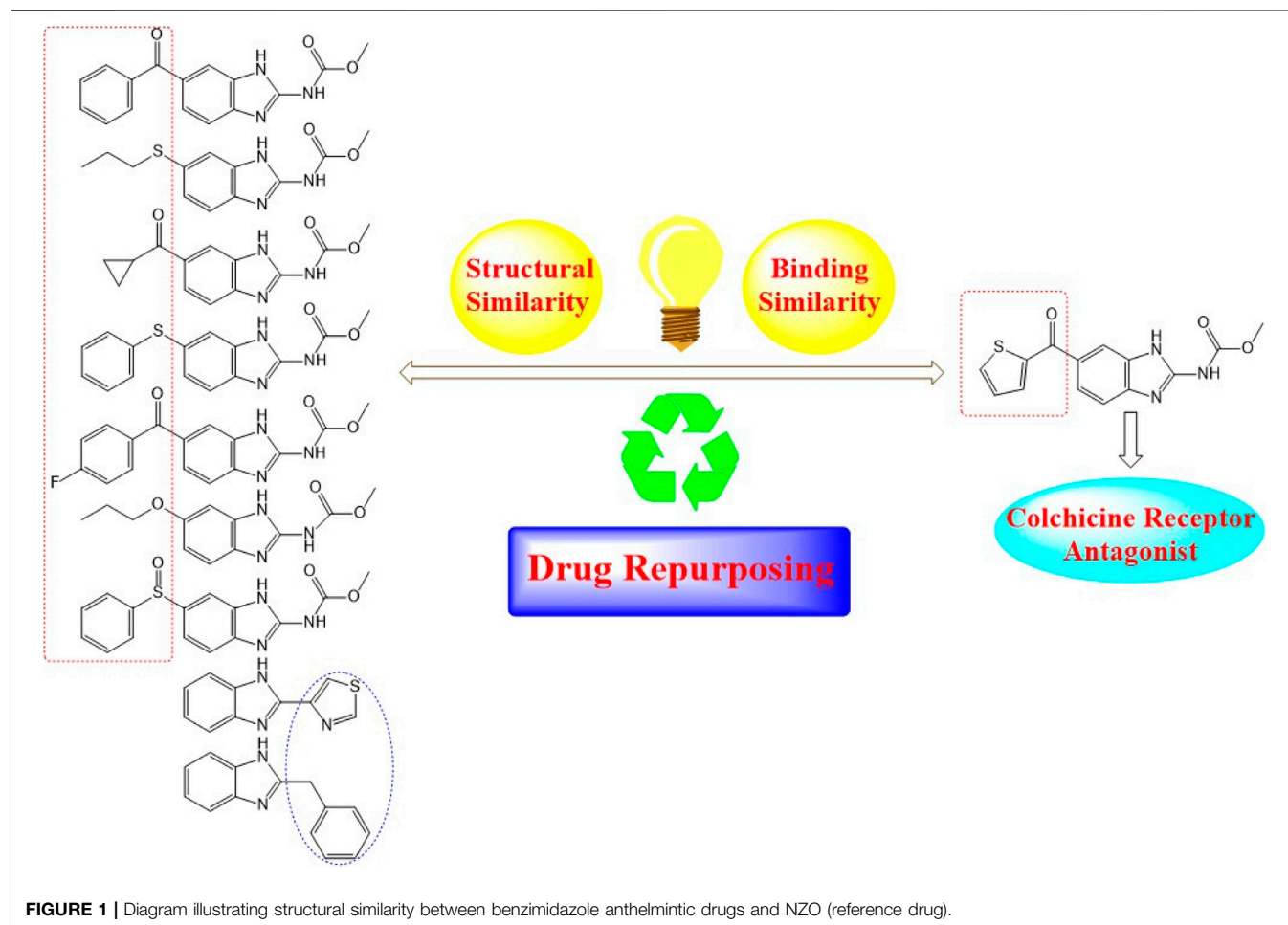
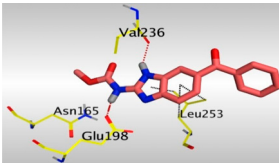
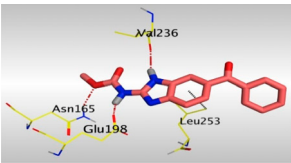
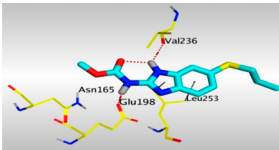
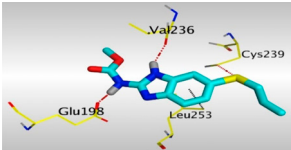
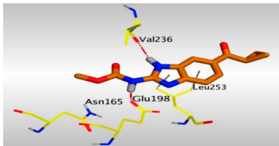
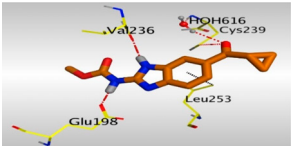
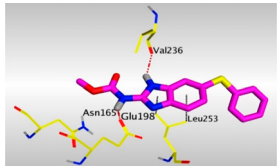
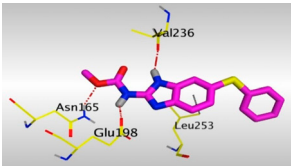
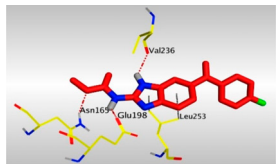
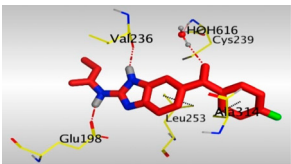
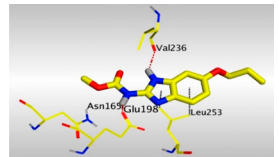
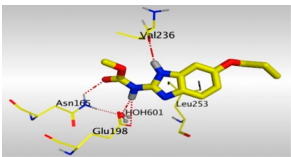
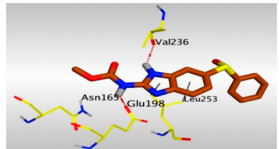
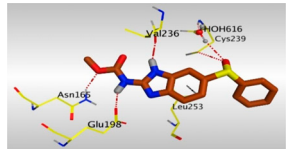


TABLE 1 | The 3D view of binding interactions between tested benzimidazole drugs and NZO-binding pocket within Tubulin subunit B and D (PDB: 5CA1) beside the solved NZO complex (Native) and the docked complex (Docked). Red dashed lines refer to hydrogen bonds, while the black ones denote hydrophobic interactions.

Drug	Protein (5CA1) subunit B	Protein (5CA1) subunit D
NZO (native)		
NZO (docked)		

(Continued on following page)

TABLE 1 | (Continued) The 3D view of binding interactions between tested benzimidazole drugs and NZO-binding pocket within Tubulin subunit B and D (PDB: 5CA1) beside the solved NZO complex (Native) and the docked complex (Docked). Red dashed lines refer to hydrogen bonds, while the black ones denote hydrophobic interactions.

Drug	Protein (5CA1) subunit B	Protein (5CA1) subunit D
Mebendazole		
Albendazole		
Ciclobendazole		
Fenbendazole		
Flubendazole		
Oxibendazole		
Oxfendazole		

(Continued on following page)

TABLE 1 | (Continued) The 3D view of binding interactions between tested benzimidazole drugs and NZO-binding pocket within Tubulin subunit B and D (PDB: 5CA1) beside the solved NZO complex (Native) and the docked complex (Docked). Red dashed lines refer to hydrogen bonds, while the black ones denote hydrophobic interactions.

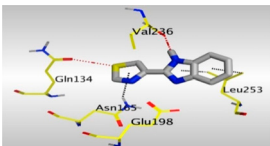
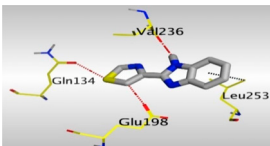
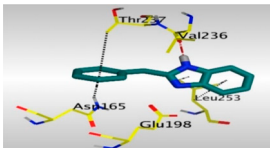
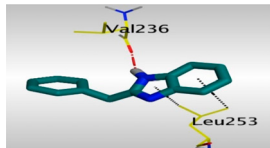
Drug	Protein (5CA1) subunit B	Protein (5CA1) subunit D
Thiabendazole		
Bendazole		

TABLE 2 | Calculated parameters obtained from docking of different benzimidazole drugs and NZO in the binding pockets within subunit B and D of Tubulin protein.

Phys. Prop.	S (subunit B)	S (subunit D)	rmsd_refine	E_conf	E_refine
NZO (docked)	-7.34	—	0.95	0.64	-26.68
	—	-7.04	2.07	-7.66	-27.13
Mebendazole	-7.36	—	1.54	-21.83	-37.08
	—	-7.28	1.40	-20.16	-37.04
Albendazole	-7.23	—	0.97	-53.87	-36.45
	—	-6.86	1.06	-40.74	-30.12
Ciclobendazole	-6.94	—	0.90	62.46	-35.58
	—	-6.78	1.40	63.16	-32.88
Fenbendazole	-7.09	—	1.71	-47.42	-37.47
	—	-7.39	1.22	-46.84	-37.31
Flubendazole	-7.36	—	1.19	-18.75	-36.26
	—	-7.25	1.38	-3.47	-34.07
Oxibendazole	-7.36	—	1.12	-52.05	-38.65
	—	-6.86	1.82	-40.80	-29.96
Oxfendazole	-7.31	—	2.47	-28.07	-38.11
	—	-7.55	2.03	-27.11	-38.19
Thiabendazole	-4.59	—	1.81	43.97	-8.75
	—	-4.68	0.23	39.50	-11.37
Bendazole	-5.44	—	1.12	39.08	-12.60
	—	-5.03	1.60	43.80	-6.76

S, the score of placement of a compound into the binding pocket of protein; rmsd_refine, the root-mean-squared-deviation (RMSD) between the heavy atoms of the predicted pose (after refinement) and those of the crystal structure (before refinement); E_conf, conformer energy in kcal/mol; E_refine, the score of refinement step of ligand conformer.

refinement methodology, and GBVI/WSA dG is the scoring methodology for selection of the best 10 poses from 100 poses for each compound. After completion of docking processes, the obtained poses were studied, and the best ones showing the best interactions of ligand–colchicine binding site of tubulin were selected and studied accordingly.

Molecular Mechanics Calculations

Density functional theory (DFT) based Becke's three-parameters Lee-Yang-Parr hybrid functional (B3LYP) (Becke, 1993; Becke, 2001) was employed in the calculations. The geometry of all compounds was initially optimized using the B3LYP/3-21G model and reoptimized using the B3LYP/6-31G model and B3LYP/6-311+G* model. No imaginary frequencies were

obtained from the optimized structures demonstrating that the corresponding geometries are true local minima. Conductor-like polarizable continuum model (CPCM) (Cossi et al., 2003) and dielectric constant of solvent water ($\epsilon = 78.35$) were used to approximately describe the polarity of colchicine binding pocket. All simulations were performed using GAUSSIAN 09 Revision C.01 (Frisch et al., 2016) on Swinburne supercomputing facilities.

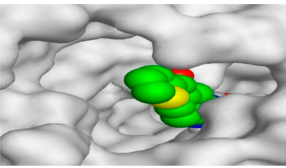
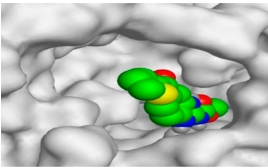
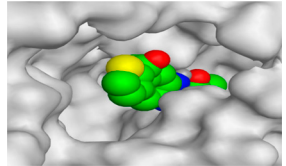
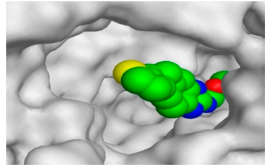
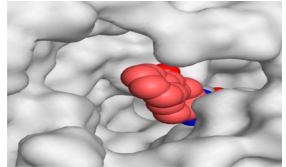
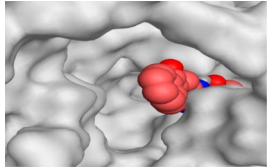
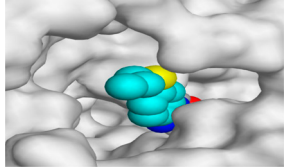
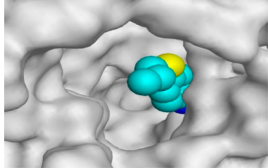
RESULTS AND DISCUSSION

Crystallographic data of co-crystallized NZO complexed with tubulin protein (PDB: 5CA1) revealed the existence of two binding sites for NZO inside the protein, subunits B and D,

TABLE 3 | Calculated hydrogen bond length (in Å) between different benzimidazole drugs and NZO and the crucial amino acids at binding sites in subunits B and D of Tubulin protein.

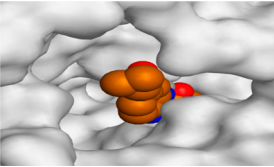
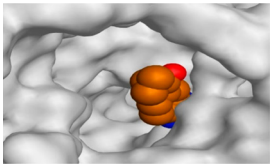
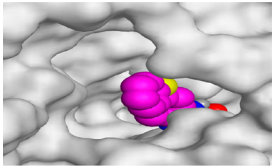
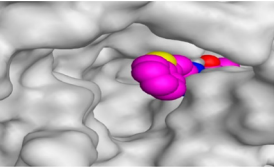
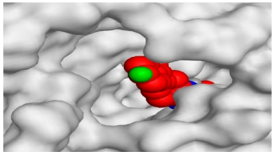
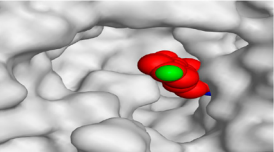
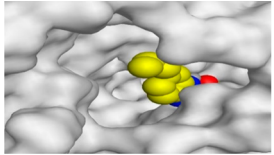
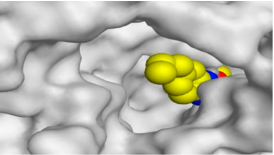
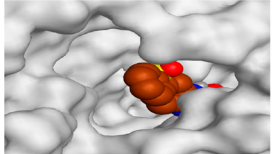
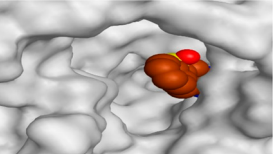
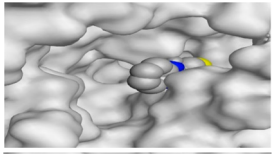
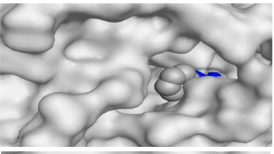
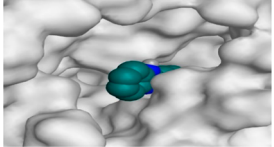
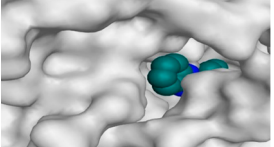
Phys. Prop.	Subunit	O-carbamate	NH-carbamate	N (1)H-benzimidazole	Cys239
					(Direct, indirect)
NZO (native)	B	3.19	2.70	2.94	—
	D	—	2.72	2.92	3.23, 2.70
NZO (docked)	B	3.03	2.95	2.87	—
	D	2.93	2.76	2.80	—
Mebendazole	B	—	2.89	2.91	—
	D	2.99	3.03	2.83	—
Albendazole	B	—	2.85	2.78	—
	D	—	2.71	2.81	3.40, —
Ciclobendazole	B	—	2.99	2.97	—
	D	—	2.92	2.81	3.23, 3.20
Fenbendazole	B	—	2.96	2.86	—
	D	2.95	3.03	2.81	—
Flubendazole	B	3.03	2.93	2.92	—
	D	—	2.79	2.84	—, 3.14
Oxibendazole	B	—	2.89	2.82	—
	D	2.87	3.53	3.00	—
Oxfendazole	B	—	2.97	2.87	—
	D	2.95	3.04	2.81	3.30, 3.38
Thiabendazole	B	—	—	2.74	—
	D	—	—	2.66	—
Bendazole	B	—	—	2.72	—
	D	—	—	2.78	—

TABLE 4 | The 3D positioning of different forms tested benzimidazole drugs inside the deep binding pocket of NZO within Tubulin subunit B and D (PDB: 5CA1) alongside with the solved NZO complex (Native) and the docked complex (Docked).

Drug	Protein (5CA1) subunit B		Protein (5CA1) subunit D	
NZO (native)				
				
NZO (docked)				
				
Mebendazole				
				
Albendazole				

(Continued on following page)

TABLE 4 | (Continued) The 3D positioning of different forms tested benzimidazole drugs inside the deep binding pocket of NZO within Tubulin subunit B and D (PDB: 5CA1) alongside with the solved NZO complex (Native) and the docked complex (Docked).

Drug	Protein (5CA1) subunit B	Protein (5CA1) subunit D
Ciclobendazole		
Fenbendazole		
Flubendazole		
Oxibendazole		
Oxfendazole		
Thiabendazole		
Bendazole		

respectively. It was obvious that subunit D is larger than subunit B in size. By analyzing the binding modes (3D graphic views) of NZO inside the two subunits (**Table 1**), it was found that NZO exhibits almost the same binding interactions within protein subunits B and D, where three H-bonds were observed between Glu198 and NH-carbamate (one H-bond) and NH(3)-benzimidazole (two H-bonds). Another H-bond was found between Val236 and the protonated NH(1)-benzimidazole. Besides, NZO docked in subunit B was able to form H-bond between CO of carbamate moiety and Asn165. Moreover, a hydrophobic interaction was found between Ala314

and the thiophene ring. In the case of subunit D, two additional H-bonds were observed between NZO and Cys239, which were a direct H-bond with the oxygen atom of the carbonyl group and an indirect H-bond with the carbonyl oxygen through the bridging water molecule (H₂O616). Hydrophobic interactions were also observed between Leu253 and the benzene and imidazole moieties. The 2D graphic views of binding interactions between studied benzimidazole ligands and NZO with the target protein are deposited in **Supplementary Material**.

By studying the binding site in both B and D subunits, it was concluded that the crucial amino acids for binding interactions

TABLE 5 | The charge density of the highest occupied molecular orbital (HOMO), the lowest unoccupied molecular orbital (LUMO), and the molecular electrostatic potential (MEP) map of studied compounds. Red and blue color codes represent the most electronegative and electropositive density, respectively. B3LYP/6-311+G* level of theory was employed to compute molecular orbital energies.

	HOMO	LUMO	MEP
NZO (docked)			
Mebendazole			
Albendazole			
Ciclobendazole			
Fenbendazole			
Flubendazole			
Oxibendazole			
Oxfendazole			
Thiabendazole			
Bendazole			

between CBSIs and the two subunits are Glu198 and Val236, besides Asn165 for B subunit and Cys239 for D subunit. The refined NZO was also docked among other studied benzimidazole drugs into subunits B and D to validate the molecular dynamic model used in conducting the current study. For the B subunit, the docked NZO showed almost a fingerprint binding mode as the original co-crystallized one. Interestingly, loss of binding interactions between the oxygen atom of its carbonyl group and Cys239 was also noticed. Furthermore, as expected, the binding interaction between N (3)-benzimidazole and Glu198 was lost in two subunits (B and D) as a result of using the neutral form of refined NZO in contrary to the protonated form of native NZO. The other nine benzimidazole drugs were refined and docked into NZO binding pocket subunits (B and D), respectively. Results revealed that all of the tested drugs, except for Thiabendazole (8) and Bendazole (9) members, showed very similar binding modes to the native NZO due to the great

structural similarity. Regarding B subunit, Flubendazole (5) showed a nearly similar fingerprint binding mode compared to the native and docked forms of NZO. Concerning D subunit, surprisingly, Albendazole (2), Ciclobendazole (3), Flubendazole (5), and Oxfendazole (7) showed the binding interaction with Cys239 similar to the native NZO which was lost in the docked NZO itself. Especially Ciclobendazole (3) and Oxfendazole (7) showed a fingerprint binding interaction with Cys239 with the same direct and indirect pathways. It was concluded that Flubendazole (5) showed the highest similarity in binding modes compared to the native NZO in both B and D subunits of protein pocket. Furthermore, Thiabendazole (8) and Bendazole (9) lost some of the binding interactions compared to the native NZO due to the differences in the structure of side chains of their benzimidazole moieties but maintained the binding interactions with both Val236 and Asn165. By reviewing molecular dynamics scoring and other parameters

listed in **Table 2**, all of the tested benzimidazole drugs showed comparable high absolute values for MD scoring and energies (except for Thiabendazole (8) and Bendazole (9) members), and the lowest rmsd_refine values compared to the docked NZO. Mebendazole (1) and Flubendazole (5) showed better scores for binding of its poses inside B and D subunits calculated at -7.36 , -7.36 , and -7.28 , -7.25 respectively compared to the docked poses of NZO at -7.34 and -7.04 . Again, Flubendazole (5) is the most promising drug to be repurposed as a colchicine binding site inhibitor compared to the docked NZO and other tested drugs. Moreover, the order of binding affinities for our tested benzimidazole drugs with target Tubulin protein depending on the docking and the calculated physical properties is as follows: Flubendazole (5) > Oxfendazole (7) > NZO (docked) > Mebendazole (1) > Albendazole (2) > Oxibendazole (6) > Fenbendazole (4) > Ciclobendazole (3) > Thiabendazole (8) > Bendazole (9).

Another confirmatory tool for validating the binding interactions between the tested ligands and the target protein is through measuring the lengths of H-bonds involved. **Table 3** shows H-bonds between our tested benzimidazole drugs compared to the native and docked forms of NZO with the similar amino acids involved in binding interactions. As predicted before, Flubendazole (5) was the best drug exhibiting similar binding mode to the native and docked forms of NZO by forming alone, at B subunit, an H-bond with Asn136 through the oxygen atom of its carbamate moiety calculated at $3.03 \text{ \AA}$ comparable to $3.19 \text{ \AA}$ and $3.03 \text{ \AA}$ formed by the native and docked forms of NZO, respectively. Additionally, it was one of the four drugs reported to interact with Cys239, at D subunit, superior to the docked NZO form by forming an H-bond with the bridging water molecule ($\text{H}_2\text{O616}$) calculated at $3.14 \text{ \AA}$ which is comparable to the same bond formed at $2.70 \text{ \AA}$ by the native NZO.

By representing the 3D filling positions of our tested compounds inside the deep protein pockets of two subunits B and D compared to the co-crystallized inhibitor, we observed a very close similarity between all in each subunit confirming the same binding mode and structural similarity between them as shown in **Table 4**.

Moreover, the large size of the two pockets of subunits B and D, especially D subunit, gives us an idea about the possible drug modifications especially at the side chain of thiophene ring of NZO to obtain larger compounds that can occupy the binding pockets more efficiently maintaining the same essential binding interactions with the crucial amino acids and at the same time forming extra binding ones for better inhibition.

It is well known that outermost electrons are those involved in the binding interaction between a ligand and target protein, we therefore, computed the electron density of the highest occupied molecular orbital (HOMO), the lowest unoccupied molecular orbital (LUMO), and the molecular electrostatic potential (MEP) map. Figures of the electron density distribution of HOMO, LUMO, and MEP are depicted in **Table 5**.

The electron density of HOMO in NZO was found to be localized mainly on benzimidazole moiety, while the electron density of LUMO was solely localized on the thiophene side chain. The electron density of HOMO and LUMO of all other studied drugs

are delocalized except for Mebendazole (1), Fenbendazole (4), and Flubendazole (5) in the case of HOMO and Oxibendazole (6) in case of LUMO. The electron density of HOMO of Mebendazole (1) and Flubendazole (5) is quite similar to that of NZO which is consistent with the calculated binding affinities [Flubendazole (5) > Oxfendazole (7) > NZO (docked) > Mebendazole (1)]. Despite of this consistency, the binding affinity does not solely depend on the energies and the electronic density distribution of HOMO and LUMO, but also on other determinants such as hydrogen bonding, electrostatic interactions, hydrophobic and Van der Waals forces and presence of clusters of water. We also noted that the electrons of HOMO are localized mainly on the benzene moiety and the aromatic side chain of Fenbendazole (4) in a different pattern to that observed with NZO.

The MEP maps of all studied compounds revealed that one of benzimidazole nitrogen acts as an electron donating site capable of forming H-bond with amino acids in colchicine binding sites. Whereas the other benzimidazolyl nitrogen is electrophilic moiety acting as H-donor group. Interestingly, the MEP map of Mebendazole (1), Albendazole (2) and Flubendazole (5) are quite similar which is also consistent with the calculated binding affinities, where Flubendazole (5) > Oxfendazole (7) > NZO (docked) > Mebendazole (1) > Albendazole (2) > Oxibendazole (6) > Fenbendazole (4) > Ciclobendazole (3) > Thiabendazole (8) > Bendazole (9). Lack of 5 (6) substitution on the benzimidazole pharmacophoric group leads to a significant change in the MEP of Thiabendazole and Bendazole. Weaker binding affinities of these two drugs can be ascribed to the change in electronic configuration due to the absence of 5 (6) substitution.

CONCLUSION

Among the nine tested members of benzimidazole anthelmintic drugs, Flubendazole (5) exhibits the most similar binding interactions, scoring, electron density distribution, and electrostatic map to that reported for NZO in tubulin protein (PDB: 5CA1). This suggests that Flubendazole (5) would be the most active member exerting antitumor activity mainly at CBS. At the same time, more structural modifications are required for FDA approved benzimidazole anthelmintic drugs and the original NZO inhibitor to obtain larger compounds with better fitting and binding modes inside the two large pockets of subunits B and D respectively.

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 authors.

AUTHOR CONTRIBUTIONS

AA-K was responsible for conducting some calculations besides drafting manuscript. MK was responsible for conceptualization, performing some calculations, and revising manuscript draft.

ACKNOWLEDGMENTS

The authors acknowledge Swinburne University of Technology for providing them with the use of supercomputing facilities.

REFERENCES

- Al-Karmalawy, A. A., and Khattab, M. (2020). Molecular modelling of Mebendazole polymorphs as a potential colchicine binding site inhibitor. *New J. Chem.* 44 (33), 13990–13996. doi:10.1039/d0nj02844d
- Becke, A. D. (1993). A new mixing of Hartree-Fock and local density-functional theories. *J. Chem. Phys.* 98, 1372. doi:10.1063/1.464304
- Becke, A. D. (2001). Density-functional thermochemistry III. the role of exact exchange. *J. Chem. Phys.* 98, 5648. doi:10.1063/1.464913
- Biodegradabilní, U., and Konjugát, P. (2015). Biodegradable high molecular weight polymeric conjugates for efficient delivery of cytostatic drugs into solid tumors. Available at: <http://hdl.handle.net/20.500.11956/84067> (Accessed September 9, 2015).
- Cossi, M., Rega, N., Scalmani, G., and Barone, V. (2003). Energies, structures, and electronic properties of molecules in solution with the C-PCM solvation model. *J. Comput. Chem.* 24, 669. doi:10.1002/jcc.10189
- D Katsetos, C., and Dráber, P. (2012). Tubulins as therapeutic targets in cancer: from bench to bedside. *Curr. Pharm. Des.* 18 (19), 2778–2792. doi:10.2174/138161212800626193
- Geary, T. G., Woo, K., McCarthy, J. S., Mackenzie, C. D., Horton, J., Prichard, R. K., et al. (2010). Unresolved issues in anthelmintic pharmacology for helminthiases of humans. *Int. J. Parasitol.* 40 (1), 1–13. doi:10.1016/j.ijpara.2009.11.001
- Guerini, A. E., Triggiani, L., Maddalo, M., Bonù, M. L., Frassine, F., Baiguini, A., et al. (2019). Mebendazole as a candidate for drug repurposing in oncology: an extensive review of current literature. *Cancers* 11 (9), 1284–1322. doi:10.3390/cancers11091284
- Hay, M., Thomas, D. W., Craighead, J. L., Economides, C., and Rosenthal, J. (2014). Clinical development success rates for investigational drugs. *Nat. Biotechnol.* 32 (1), 40–51. doi:10.1038/nbt.2786
- Honore, S., Pasquier, E., and Braguer, D. (2005). Understanding microtubule dynamics for improved cancer therapy. *Cell. Mol. Life Sci.* 62 (24), 3039–3056. doi:10.1007/s00018-005-5330-x
- Inc., C. C. G. (2016). *Molecular operating environment (MOE)*. Montreal, QC: Chemical Computing Group Inc.
- Khattab, M. (2020). Theoretical study of the geometric and electronic characterization of carbendazim-based drug (Nocodazole). *Heliyon* 6, e04055. doi:10.1016/j.heliyon.2020.e04055
- Köhler, P. (2001). The biochemical basis of anthelmintic action and resistance. *Int. J. Parasitol.* 31 (4), 336–345. doi:10.1016/S0020-7519(01)00131-X
- Králová, V., Hanušová, V., Staňková, P., Knoppová, K., Čáňová, K., and Skálová, L. (2013). Antiproliferative effect of benzimidazole anthelmintics albendazole, ricobendazole, and flubendazole in intestinal cancer cell lines. *Anti-Cancer Drugs* 24 (9), 911–919. doi:10.1097/CAD.0b013e3283648c69
- Lacey, E. (1990). Mode of action of benzimidazoles. *Parasitol. Today* 6 (4), 112–115. doi:10.1016/0169-4758(90)90227-U
- McLoughlin, E. C., and O'Boyle, N. M. (2020). Colchicine-binding site inhibitors from chemistry to clinic: a review. *Pharmaceuticals* 13 (1), 8–43. doi:10.3390/ph13010008

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fchem.2021.628398/full#supplementary-material>.

- Frisch, M. J., Trucks, G. W., Schlegel, H. B., Scuseria, G. E., Robb, M. A., Cheeseman, J. R., et al. (2016). Gaussian16 {R}evision {C}.01. Available at: <https://gaussian.com/citation/>
- Mukhopadhyay, T., Sasaki, J., Ramesh, R., and Roth, J. A. (2002). Mebendazole elicits a potent antitumor effect on human cancer cell lines both *in vitro* and *in vivo*. *Clin. Cancer Res.* 8 (9), 2963–2969.
- Narasimhan, B., Sharma, D., and Kumar, P. (2012). Benzimidazole: a medically important heterocyclic moiety. *Med. Chem. Res.* 21 (3), 269–283. doi:10.1007/s00044-010-9533-9
- Park, K. (2019). A review of computational drug repurposing. *Transl. Clin. Pharmacol.* 27 (2), 59–63. doi:10.12793/tcp.2019.27.2.59
- Polamreddy, P., and Gattu, N. (2019). The drug repurposing landscape from 2012 to 2017: evolution, challenges, and possible solutions. *Drug Discov. Today* 24 (3), 789–795. doi:10.1016/j.drudis.2018.11.022
- 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 (1), 41–58. doi:10.1038/nrd.2018.168
- Raghunath, M., and Viswanathan, C. L. (2014). Benzimidazole-2-carbamic acid as a privileged scaffold for antifungal, anthelmintic and antitumor activity a review. *Int. J. Pharm. Pharm. Sci.* 6 (5), 17–25. doi:10.18579/jpcrk/2014/13/4/78372
- Takebe, T., Imai, R., and Ono, S. (2018). The current status of drug discovery and development as originated in United States academia: the influence of industrial and academic collaboration on drug discovery and development. *Clin. Translational Sci.* 11 (6), 597–606. doi:10.1111/cts.12577
- Wang, Y., Zhang, H., Gigant, B., Yu, Y., Wu, Y., Chen, X., et al. (2016). Structures of a diverse set of colchicine binding site inhibitors in complex with tubulin provide a rationale for drug discovery. *Febs J.* 283, 102–111. doi:10.1111/febs.13555
- Weir, H. K., Anderson, R. N., Coleman King, S. M., Soman, A., Thompson, T. D., Hong, Y., et al. (2016). Heart disease and cancer deaths—trends and projections in the United States, 1969–2020. *Prev. Chronic Dis.* 13, E157. doi:10.5888/pcd13.160211
- Yadav, G., and Ganguly, S. (2015). Structure activity relationship (SAR) study of benzimidazole scaffold for different biological activities: a mini-review. *Eur. J. Med. Chem.* 97 (1), 419–443. doi:10.1016/j.ejmech.2014.11.053

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 Khattab and Al-Karmalawy. 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.



Applications of Virtual Screening in Bioprospecting: Facts, Shifts, and Perspectives to Explore the Chemo-Structural Diversity of Natural Products

OPEN ACCESS

Edited by:

Fei Ye,

Zhejiang Sci-Tech University, China

Reviewed by:

José L. Medina-Franco,

National Autonomous University of Mexico, Mexico

José Pedro Cerón-Carrasco,

Catholic University San Antonio de Murcia, Spain

*Correspondence:

Kauê Santana

kaue.costa@ufopa.edu.br

†ORCID:

Kauê Santana

orcid.org/0000-0002-2735-8016

Lidiane Diniz do Nascimento

orcid.org/0000-0003-1370-4472

Anderson Lima e Lima

orcid.org/0000-0002-8451-9912

Vinícius Damasceno

orcid.org/0000-0003-2263-2124

Rodolpho C. Braga

orcid.org/0000-0003-3814-3464

Jerônimo Lameira

orcid.org/0000-0001-7270-1517

Specialty section:

This article was submitted to Theoretical and Computational Chemistry,

a section of the journal Frontiers in Chemistry

Received: 01 February 2021

Accepted: 25 February 2021

Published: 29 April 2021

Citation:

Santana K, do Nascimento LD, Lima e Lima A, Damasceno V, Nahum C, Braga RC and Lameira J (2021) Applications of Virtual Screening in Bioprospecting: Facts, Shifts, and Perspectives to Explore the Chemo-Structural Diversity of Natural Products. *Front. Chem.* 9:662688. doi: 10.3389/fchem.2021.662688

Kauê Santana^{1*}, Lidiane Diniz do Nascimento^{2†}, Anderson Lima e Lima^{3†}, Vinícius Damasceno^{3†}, Claudio Nahum³, Rodolpho C. Braga^{4†} and Jerônimo Lameira^{5†}

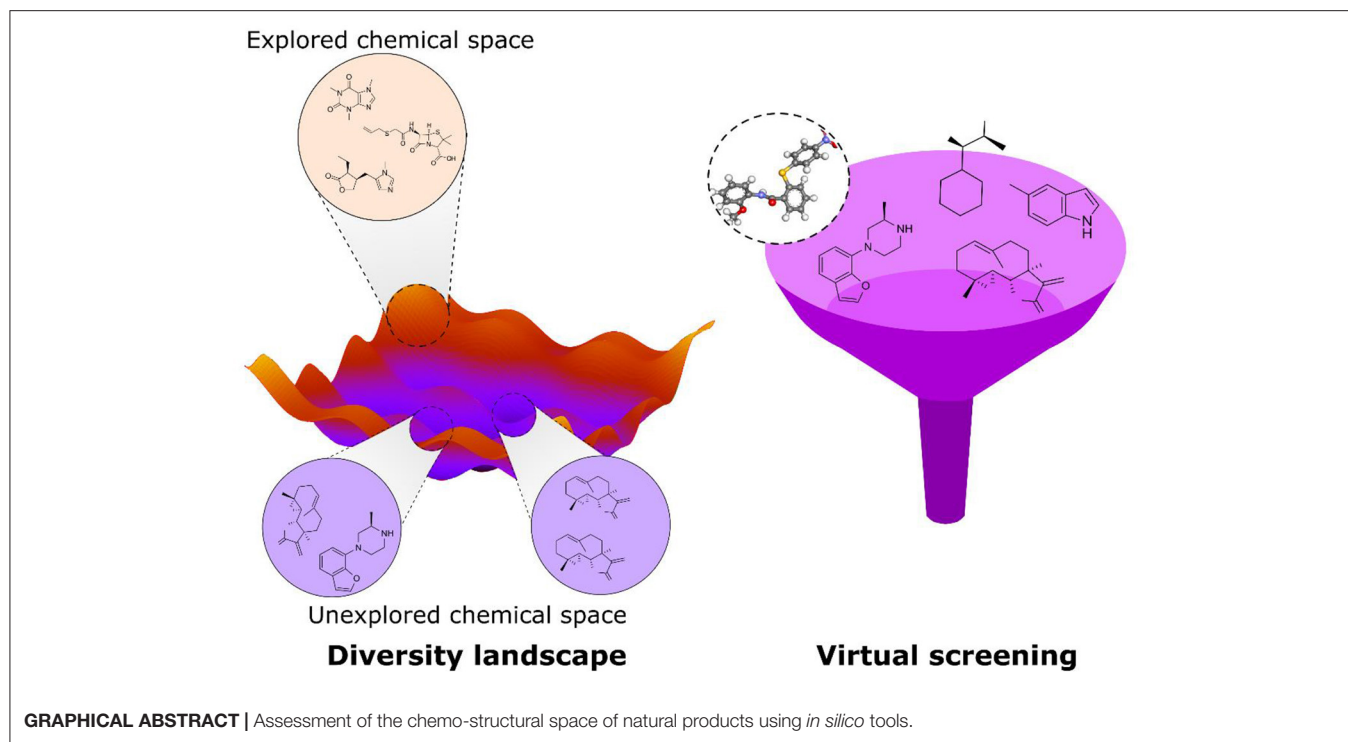
¹ Instituto de Biodiversidade, Universidade Federal do Oeste do Pará, Santarém, Brazil, ² Laboratório Adolpho Ducke, Coordenação de Botânica, Museu Paraense Emílio Goeldi, Belém, Brazil, ³ Instituto de Ciências Exatas e Naturais, Universidade Federal do Pará, Belém, Brazil, ⁴ InsilicAll Ltda, São Paulo, Brazil, ⁵ Instituto de Ciências Biológicas, Universidade Federal do Pará, Belém, Brazil

Natural products are continually explored in the development of new bioactive compounds with industrial applications, attracting the attention of scientific research efforts due to their pharmacophore-like structures, pharmacokinetic properties, and unique chemical space. The systematic search for natural sources to obtain valuable molecules to develop products with commercial value and industrial purposes remains the most challenging task in bioprospecting. Virtual screening strategies have innovated the discovery of novel bioactive molecules assessing *in silico* large compound libraries, favoring the analysis of their chemical space, pharmacodynamics, and their pharmacokinetic properties, thus leading to the reduction of financial efforts, infrastructure, and time involved in the process of discovering new chemical entities. Herein, we discuss the computational approaches and methods developed to explore the chemo-structural diversity of natural products, focusing on the main paradigms involved in the discovery and screening of bioactive compounds from natural sources, placing particular emphasis on artificial intelligence, cheminformatics methods, and big data analyses.

Keywords: machine learning, big data, natural products, bioprospecting, cheminformatics, virtual screening, drug discovery, chemical data

NATURAL PRODUCTS AS SOURCES OF NOVEL BIOACTIVE COMPOUNDS AND THE PARADIGMS OF THEIR EXPLORATION

The high structural and physicochemical diversity of natural products makes them a valuable source to discover and develop new bioactive compounds with different pharmaceutical, cosmetic, biotechnological, agrochemical, and food applications (Rayan et al., 2017). Success histories of natural product-based drugs have been reported in the pharmaceutical industry and include pilocarpine, quinine, morphine, and artemisinin (Newman and Cragg, 2016; Zhang L. et al., 2020). Natural products represent relevant importance in the discovery and development of new bioinspired bioactive compounds, and more than 50% of the developed drugs approved by the



United States Food and Drug Administration (USFDA, 1981–2019) are derived or bioinspired from compounds obtained from natural sources (Newman and Cragg, 2020). Natural products are chemically complex and differ from synthetic compounds in different aspects; as an example, these structures contain a high percentage of oxygen as well as a larger fraction of sp³-hybridized atoms and chiral centers (Lee and Schneider, 2001; Feher and Schmidt, 2003; Rodrigues et al., 2016), and their chemical space is highly diverse, containing different structural scaffolds, when compared with synthetic compound libraries (Chen et al., 2018). Due to their unique features, their structures can provide an innovative solution for the design and synthesis of new bioactive compounds (Kumar et al., 2017; Silva et al., 2019; Bradley et al., 2020; Morais et al., 2020).

The systematic search for natural sources to obtain valuable compounds to develop products with commercial value and industrial purposes remains the most challenging task in bioprospecting (Skirycz et al., 2016; Roumpeka et al., 2017; Cubillos et al., 2019). The traditional approach to discover new bioactive compounds from natural sources includes sequential steps that are obtained from the biological material using ethnological knowledge, extraction, fractionation/isolation, chemical characterization, and, finally, the execution of the biological assays of the isolated or fractionated natural products (Zhang L. et al., 2020). Subsequent analyses include the lead compound optimization using chemical synthesis to perform structural modifications in order to improve their pharmacodynamic and pharmacokinetic properties and to increase their biological activities (Huffman and Shenvi, 2019). In contrast, bioprospecting strategies that use computational tools

have been reported as efficient, low-cost, low-labor, and low-time approaches when compared to experimental methods that use solely *in vitro* and *in vivo* assays (Li and Vederas, 2009; Wingert and Camacho, 2018; Trujillo-Correa et al., 2019).

Despite natural products being continually explored in drug development programs, attracting the attention of scientific research efforts due to their pharmacophore-like structures, pharmacokinetic properties, and unique chemical space, the big pharma industry has focused on cutting-edge technologies that combine high-throughput screening and combinatorial chemistry methods to obtain and evaluate synthetic compound libraries (Henninot et al., 2018; Batool et al., 2019). This decision is, in part, a consequence of the complex structures of natural products that impose limitations in synthetic routes and due to the time-consuming and laborious process involved in the isolation of a single chemical constituent, which often requires a significant amount of reagents and adequate infrastructure, obtaining low yields of purified target compounds (Huffman and Shenvi, 2019). Based on these limitations, the isolation and the characterization of compounds from natural sources have been indicated only for those with potential applications and desirable biological activities (Olivon et al., 2017). However, it has been suggested that the reduced new chemical entities found by the pharmaceutical industry that reach the final market could be due to the strategic decision to prioritize combinatorial synthetic libraries instead of natural product-based libraries (Over et al., 2013; Rodrigues, 2017). Currently, we are witnessing a resurgence of natural products in the development and research of novel bioactive compounds; besides, some structural scaffolds obtained from different classes of natural products,

such as alkaloids, phenylpropanoids, polyketides, and terpenoids, have served as an inspiration to design new drug candidates (Thomford et al., 2018; Davison and Brimble, 2019; Galúcio et al., 2019; Li et al., 2019). Natural products remain inspiring the development of new drugs, cosmetics, and other bioactive compounds for human use (Newman and Cragg, 2020; Atanasov et al., 2021).

Recently, metabolomics and metabolic profiling approaches have explored novel taxonomic groups from the unique environment, providing opportunities for finding novel natural bioactive compounds, and some examples include bacteria (Kleigrewe et al., 2015; Gosse et al., 2019), cnidaria (Santacruz et al., 2020), marine sponge (Abdelhameed et al., 2020), insects (Klupczynska et al., 2020), and fungi (Oppong-Danquah et al., 2018). Special attention has been given to novel chemical entities that originated from marine environments due to their diverse and unique drug-like scaffolds (Shang et al., 2018) and physicochemical properties (Jagannathan, 2019) when compared with natural products of terrestrial origin, which make them a valuable source for exploration by the pharmaceutical and biotechnological industries. Advances in the experimental methods applied in metabolomic approaches coupled with computational methods have been useful to identifying new natural products with plausible biological activities as well as to understanding their molecular mechanisms of action (Atanasov et al., 2021).

Currently, artificial intelligence algorithms (Wolfe et al., 2018; Lima et al., 2020; Stokes et al., 2020) and omics-based technologies (Floros et al., 2016; Huang et al., 2017; Jones and Bunnage, 2017; Merwin et al., 2020) have emerged as approaches to characterize and select interesting chemo-structures with appropriate physicochemical properties and biological activities as well as to prioritize the isolation of natural compounds from biological sources (Chen et al., 2018; Wolfender et al., 2019), which open up new opportunities to explore their industrial applications. Combined with other *in silico* analyses, artificial intelligence and cheminformatics methods can screen a high diversity of chemo-structures isolated from natural sources or deposited in public databases (Chen and Kirchmair, 2020), analyzing their bioactivity, pharmacodynamics, and their pharmacokinetic properties, thus reducing the financial efforts involved in research programs that aim to find new chemical agents (Chen et al., 2018; Al Sharie et al., 2020; Medina-Franco and Saldívar-González, 2020).

In this review, we discuss the computational approaches and methods applied to explore the chemo-structural diversity of natural products, giving particular attention to the main paradigms involved in the discovery and screening of bioactive natural compounds with different industrial applications (e.g., herbicides, insecticides, etc.) that are beyond the discovery of new drugs. Here, we emphasize computational strategies that use artificial intelligence, cheminformatics, and big data analyses that have been developed in the last years. We also explore the limitations and biases of these methods and demonstrate practical applications to evaluate the chemical entities obtained from natural sources aiming at bioprospecting.

COMPUTATIONAL APPROACHES APPLIED IN THE VIRTUAL SCREENING OF BIOACTIVE COMPOUNDS

Virtual screening methods have innovated the discovery of new compounds with specific bioactivity, assessing *in silico* large structural libraries against a bioreceptor or biological system, thus favoring the reduction of financial efforts, infrastructure, and the time involved in the process of discovering new chemo-structures (Macalino et al., 2015). These methods apply sequential and hierarchical steps that aim at filtering and selecting compounds with desirable physicochemical, pharmacokinetic, and pharmacodynamic properties while discarding those that do not fit the desirable characteristics. A virtual screening workflow comprises two main computational tasks (**Figure 1A**): (1) the first one is the library preparation, which includes, among other computational tasks, obtaining the structures of the compounds, file conversion to readable formats, such as SMILES (simplified molecular-input line entry system), SDF (structure data file), and MOL2 (MDL Molfile) (Saldívar-González et al., 2020), conformer generation, and the correction of stereochemical and valence errors (Ropp et al., 2019); (2) the second one corresponds to the application of computational techniques to filter the desirable compounds (Gimeno et al., 2019). The final step corresponds to experimental validation using *in vitro* and *in vivo* assays, which include enzymatic inhibition assays and/or cell line inhibition (Spyrakakis et al., 2019; Ye et al., 2019).

Different computational methods have been developed over the years and implemented in virtual screening strategies (Tomar et al., 2018), applying knowledge of artificial intelligence (Gupta et al., 2013; Yang et al., 2018; Schaduengrat et al., 2019; Shoombuatong et al., 2019; Kong et al., 2020), molecular modeling (Semighini et al., 2011; Rampogu et al., 2018; Da Costa et al., 2019; Jin et al., 2020; Mascarenhas et al., 2020), statistics, and probability (Pire et al., 2015; Daina and Zoete, 2016; Blanco et al., 2018; Madzhidov et al., 2020; Cai et al., 2021). These methods, when combined with experimental approaches, increase the success to finding novel bioactive compounds (Kumar and Zhang, 2015; Coimbra et al., 2020; Gorgulla et al., 2020; Stokes et al., 2020). Two computational approaches are related to the virtual screening of compounds: (1) the ligand-based virtual screening (LBVS) and (2) structure-based virtual screening (SBVS) approaches (**Figure 1B**). Both computational approaches have been combined in virtual screening strategies that aim to identify novel bioactive compounds against a specific molecular target or a biological system (Da Costa et al., 2019; Galúcio et al., 2019; Wang et al., 2020).

The LBVS approach depends solely on the analyses of the intrinsic characteristics of the compound structure, such as the electronic, topological, physicochemical, and structural properties that are related to its molecular activity using, as a starting point, a set of compounds with experimentally proven biological activity (Hamza et al., 2012; Berenger et al., 2017; Garcia-Hernandez et al., 2019). Computational methods applied in the LBVS approach include structural-, three-dimensional

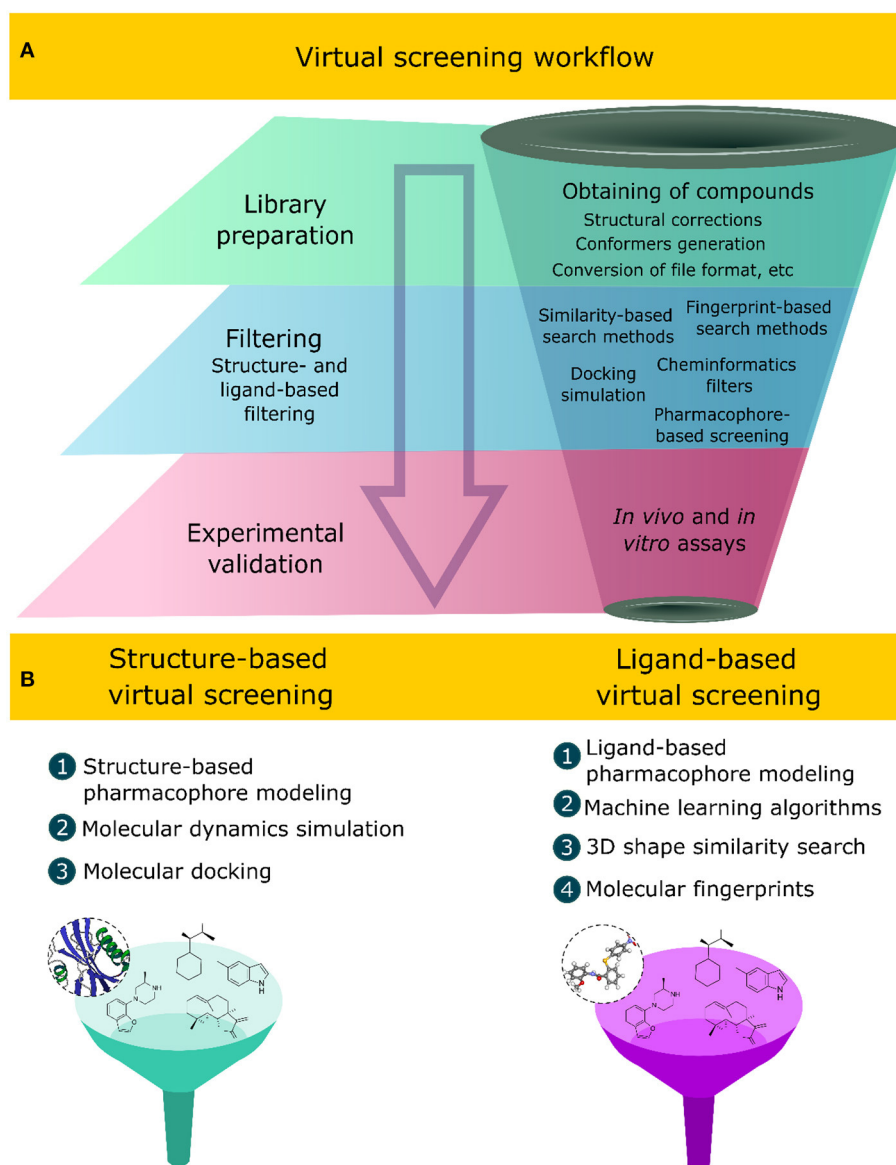


FIGURE 1 | (A) Sequential steps applied in virtual screening workflows to select bioactive natural products. **(B)** Ligand- and structure-based virtual screening approaches and some of their associated computational methods.

(3D) shape-, and fingerprint-based similarity search methods, cheminformatics filters, machine learning algorithms, ligand-based pharmacophore modeling, and quantitative structure–activity relationship (QSAR) methods (Yan et al., 2016; Tahir et al., 2020). In contrast, the SBVS approach uses, as a starting point, information related to the molecular recognition of the ligand in the bioreceptor structure to design and discover new bioactive compounds. This information includes bioreceptor conformation, the ligand-binding affinity, intermolecular interactions, molecular surface charge, and the composition of the residue of the binding site (Gonczarek et al., 2018; Guedes et al., 2018; Yasuo and Sekijima, 2019;

Maia E. H. B. et al., 2020). These methods require the elucidated 3D structure of the receptor and, preferably, in complex with the bioactive compound. The 3D structure informs the structural conformation and molecular binding site of the bioactive ligands. Among the computational methods applied in the SBVS approach, we can cite molecular docking, molecular dynamics simulation, and structure-based pharmacophore modeling (Wang et al., 2020). Currently, virtual screening methods are an integral part of the design and discovery process of new bioactive compounds, and their applications have become popular in the academia and industry (Kar and Roy, 2013).

COMPUTATIONAL METHODS APPLIED IN VIRTUAL SCREENING APPROACHES

Cheminformatics Filters (Molecular Filters)

The prediction of the pharmacokinetics and drug-likeness properties of chemical entities represents an important task for the discovery of structures with interesting biological activity (Mignani et al., 2018). In essence, drug-likeness represents a measure of the overall similarity of the analyzed compounds to a chemical space occupied by known drugs (Mignani et al., 2018; Jia et al., 2020).

The prediction of the chemical properties of compounds usually involves the application of a set of simple empirical chemical rules (Gfeller et al., 2014; Lagorce et al., 2015; Daina and Zoete, 2016). Over the years, different cheminformatics filters (also known as molecular filters) have been developed as useful tools to screen structures that have desirable pharmacokinetic and pharmacodynamic properties, low toxicity, and/or low promiscuity/reactivity in inhibition assays, thus guiding the decision-making process in the discovery of new chemical entities with pharmaceutical, cosmetic, agrochemical, and biotechnological interest (Huggins et al., 2011). The most commonly used filters are intended to remove from structural libraries the compounds with low cell membrane permeability or distribution. Among the well-known cheminformatics filters, we can cite those developed by Lipinski (Lipinski et al., 1997), Veber (Veber et al., 2002), and Jeffrey (Jeffrey and Summerfield, 2010). Some structural properties evaluated by these molecular filters predict some pharmacodynamic properties, such as compound promiscuity, i.e., their non-selectivity against a molecular target (Walters and Namchuk, 2003; Lovering, 2013). Some filters are based on the selection of a range of physicochemical and structural properties that are representative of specific pharmacokinetics (e.g., gastrointestinal absorption or penetration into the blood–brain barrier) and pharmacodynamic properties (e.g., specificity or promiscuity to a macromolecular target). These properties are selected using a statistical cutoff (e.g., 90th percentile limit) for each molecular descriptor that is representative to explain the interesting feature of the analyzed compounds (Daina and Zoete, 2016).

Since the first report of the chemical rules elected by Lipinski et al. (1997)—also known as the rule of five (RO5) and Pfizer rules—different chemical extensions to these chemical properties have been developed over the years to better define the “drug-like” features and bioavailability of compounds (Doak et al., 2014). More recently, hybrid methods that combine some counting schemes similar to Lipinski’s rules with a set of functional groups identified as reactive, toxic, and problematic moieties have also been developed to eliminate promiscuous structures from the high-throughput screening assays (Walters and Murcko, 2002; Bruns and Watson, 2012). Filters have also been developed to screen fragment-based chemical libraries (rule of three, RO3) (Jhoti et al., 2013). Similar to filters developed for drugs, molecular filters have also been developed to select herbicide-, fungicide-, and insecticide-likeness due to their applications in the agrochemical industry (Tice, 2001; Avram et al., 2014).

Despite these molecular filters having been widely applied in virtual screening approaches to select natural products from large chemo-structural libraries (Thireou et al., 2018; Da Costa et al., 2019; Galúcio et al., 2019), caution must be taken to avoid removal of the chemo-structures with appropriate bioavailability (Shultz, 2019). Most natural products break some chemical rules applied in molecular filtering; furthermore, some chemical classes of compounds, such as peptides and polyketides (e.g., macrolides), are located beyond the chemical limits determined by the rule of five (beyond the rule of five, bRO5) (Doak et al., 2014; Naylor et al., 2017; Rossi Sebastiano et al., 2018). Contrasting to the drug-likeness, the natural product-likeness concept has been developed to measure the overall molecular diversity of the natural product space, and it has been used as a selection criteria to screen substructures for the prioritization of combinatorial synthesis, aiming at novelty and the easy design of building blocks (Ertl et al., 2008; Jayaseelan et al., 2012). Currently, there are a great variety of cheminformatics programs that calculate these chemical properties that compose the cheminformatics filters, including the open-source programs Osiris DataWarrior [operating system (OS) compatibility: Linux/MS-Windows/Mac OS] (Sander et al., 2015) and RDKit (OS compatibility: Linux/MS-Windows/macOS) (Lovrić et al., 2019), and some commercial solutions, such as Instant JChem (OS compatibility: Linux/MS-Windows/macOS) (Instant JChem 21.4.0, 2021). Similar to these applications, the FAF-Drugs4 web server also predicts some chemical properties to screen structures from large compound libraries using some in-house cheminformatics filters, such as the Drug-Like Soft and Lead-Like Soft that predict compound similarity to drugs and leads, respectively (Miteva et al., 2006). Some databases also offer online tools to evaluate the drug-likeness and natural product-likeness (Sorokina and Steinbeck, 2019; Jia et al., 2020). **Table 1** exhibits an overview of the main molecular filters applied to screen natural products from chemical libraries.

Molecular Fingerprint-Based Methods

Similarity search methods applied in the screening of natural products are based on the premise that molecules with similar structures have similar biological activities (Cereto-Massagué et al., 2015). These methods have been applied to evaluate natural compound similarities, their bioactivity (Muegge and Mukherjee, 2016), and potential molecular targets (Huang et al., 2018).

Molecular fingerprint-based methods use representations of chemical structures to allow the quantitative assessment of the pairwise similarity of compounds with computationally efficient calculations (Riniker and Landrum, 2013; Bajusz et al., 2015). Molecular fingerprints are binary representations (bits) of a chemical structure in which 1 (present) denotes the existence of a certain molecular feature and 0 (absent) denotes inexistence (Rácz et al., 2018). **Figure 2A** shows a schematic view of the binary representation of a molecular fingerprint of a compound structure. Molecular fingerprints can vary greatly concerning the applied molecular descriptors, and some of them are based solely on the chemical structure, such as

TABLE 1 | Structural and physicochemical properties present in some cheminformatics filters applied in virtual screening.

	MW (Da)	PSA (Å ²)	HBA	HBD	cLogP/cLogD	RTB	NAR	Formal charge	References
Lipinski's rule (RO5)	≤500	–	0–10	0–5	≤5	–	–	–	Lipinski et al., 1997
Ghose's rule	160–480	–	–	–	–0.4 to +5.6	–	20–70	–	Ghose et al., 1999
Oprea's drug-like rule	–	–	2–9	0–2	–	2–8	–	–	Oprea, 2000
Walters	200–500	≤120	0–10	0–5	–	0–8	–	–	Walters and Murcko, 2002
Veber's rule	–	≤140	–	–	–	0–10	–	–	Veber et al., 2002
REOS	200–500	–	–	0–5	–5.0 to 5.0	0–8	–	–2 to +2	Walters and Namchuk, 2003
Beyond rule of five (bRO5)	≤1,000	<250	<15	≤6	–2 to 10	≤20	–	–	Doak et al., 2014
Congreve's rule (RO3)	<300	–	≤6	≤3	≤3	–	–	–	Congreve et al., 2003
Herbicide-likeness	150–500	–	2–12	<3	≤3.5	<12	–	–	Tice, 2001
Insecticide-likeness	150–500	–	1–18	≤2	0–5	<12	–	–	Tice, 2001
Hao's rule (pesticide-likeness)	≤435	–	≤6	≤2	≤6	≤9	≤17	–	Hao et al., 2011

MW, molecular weight; PSA, polar surface area; HBD, hydrogen bond donor; HBA, hydrogen bond acceptor; RTB, rotatable bonds; NAR, number of aromatic rings.

topological distances and the presence/absence of functional groups (Cereto-Massagué et al., 2015). However, some molecular fingerprints use information from pharmacophore models, allowing the comparison of the ligand poses (pharmacophore fingerprints) (Wood et al., 2012). Some molecular fingerprints, such as SMILES fingerprint (SMIfp) (Schwartz et al., 2013), and structural interaction fingerprint (SIFt) (Deng et al., 2004), evaluate structural features related to intermolecular interactions, such as hydrophobic contacts, polar interactions, and hydrogen bond acceptors and donors (interaction fingerprints) (Desaphy et al., 2013). Considering that natural products are chemically complex and structurally different from the synthetic libraries, the analyses of their structures using molecular fingerprints can provide insights, evidencing some structural similarities (see example in **Figure 2B**) (Gu et al., 2013; Tao et al., 2015; Floros et al., 2016; Galúcio et al., 2019; Chávez-Hernández et al., 2020).

Molecular fingerprints offer a cost-efficient computational calculation to be implemented with other computational approaches. Molecular fingerprints have been widely applied in the representation of chemical space networks to evaluate the structural similarities of natural products (see example in **Figure 2C**) (Zhang et al., 2015) as well as in hierarchical clustering methods (**Figure 2D**) (Sánchez-Cruz and Medina-Franco, 2018). In chemical network representations, the nodes (vertices) represent the analyzed compounds and edges of the pairwise fingerprint similarity relationships calculated by a structural metric. The edge drawn between a pair of nodes uses a satisfying threshold criterion for the structural similarity value (e.g., a cutoff = 0.7) between the analyzed compounds (Maggiora and Bajorath, 2014; Kunimoto and Bajorath, 2018). The investigation of the chemical space of natural products is an intelligent way to identify some classes of compounds, their bioactivity, and the structural scaffolds present in known active compounds (Opassi et al., 2018). Due to the high diversity of the derived structures of natural products containing modified functional groups; different strategies have been applied to investigate their chemical space, which include the modeling of hypothetical structural

modification (Skinnider et al., 2017) and the application of less restrictive similarity-based cutoffs (Pavadai et al., 2017).

Recently, machine learning algorithms using MACCS keys and Morgan molecular fingerprints have been used to differentiate natural products from synthetic molecules. The authors also used similarity maps to classify natural product substructures according to their similarity to natural or synthetic compounds (Chen et al., 2019). Galúcio et al. (2019) used fingerprint-based similarity to find correspondences between natural products and FDA-approved anticancer drugs, and the authors identified an interesting correspondence (see **Figure 2B**) between the bisdethiobis(methylthio)gliotoxin obtained from bacterial strain and the FDA-approved anticancer drug mitomycin.

Several programs and web servers have been developed to compute molecular fingerprints, and among them, we can cite ChemDes (web server) (Dong et al., 2015), ChemoPy (open-source Python package) (Cao et al., 2013), PaDEL (open-source Java program) (Yap, 2011), and jCompoundMapper (open-source Java program) (Hinselmann et al., 2011).

Similarity and Distance Metrics

Structural similarity is a key concept in the discovery of new bioactive compounds from natural sources due to the assumption that similar compounds perform similar molecular activities. Different similarity and distance metrics have been applied to compare molecular fingerprints (Bajusz et al., 2015); some of them are available in cheminformatics tools, such as Konstanz Information Miner (KNIME) (Berthold et al., 2009), PyBel (O'Boyle et al., 2008), the Chemistry Development Kit (CDK) (Willighagen et al., 2017), and RDKit (Lovrić et al., 2019). Similarity metrics could use two-dimensional (2D) or 3D similarities of compounds, but studies have demonstrated that the 2D similarity coefficient neglects some important structural/functional features in the identification of the target compound (Gohlke et al., 2015; Kim et al., 2016).

Several similarities and distance metrics have been applied to compare the pairwise similarities of molecules and their

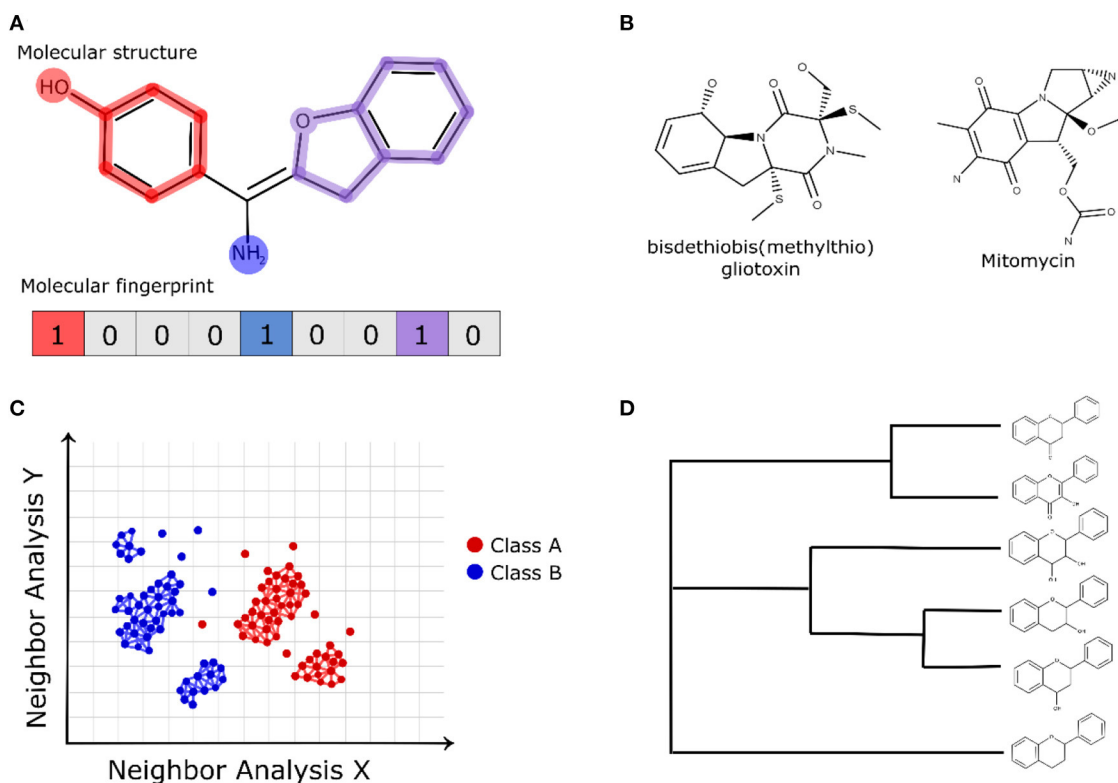


FIGURE 2 | (A) Schematic representation of *bits* applied in the molecular fingerprints of chemical structures. **(B)** Fingerprint-based similarity of the natural compound bisdethiobis (methylthio)gliotoxin and the FDA-approved anticancer drug mitomycin (Galúcio et al., 2019). **(C)** Schematic view of the chemical space network and **(D)** hierarchical clustering that apply fingerprint-based descriptors to analyze natural compounds.

TABLE 2 | Structural similarity and distance metrics applied in virtual screening.

Similarity and distance metrics	Equations for dichotomous variables
Cosine coefficient	$S_{A,B} = c/[ab]^{1/2}$
Dice coefficient	$S_{A,B} = 2c/[a + b]$
Tanimoto coefficient	$S_{A,B} = c/[a + b - c]$
Tversky coefficient	$S_{A,B} = c/[\alpha a + \beta b - c]$
Soergel distance	$D_{A,B} = 1 - \frac{c}{a+b-c}$
Manhattan distance	$D_{A,B} = a + b - 2c$
Euclidean distance	$D_{A,B} = [a + b - 2c]^{1/2}$

substructures (Bajusz et al., 2015; O'Hagan and Kell, 2016; Rácz et al., 2018). **Table 2** exhibits the main similarity coefficients and their dichotomous equations applied to compare molecular fingerprints, where *a* correspond to *on* bits (presence) in structure A, *b* is the number of the *on* bits in structure B, while *c* corresponds to bits that are *on* in both molecular structures. Differently from other similarity metrics, Tversky is an asymmetric coefficient that has two user-defined parameters, α and β . If α is set to 1 and β is set to 0, the Tversky coefficient will measure the substructural similarity between two molecules, where a Tversky value equal to 1 indicates that a given structural

moiety is a substructure of the compared compound (Senger, 2009).

Tanimoto has been the most used similarity coefficient in fingerprint-based similarity in virtual screening strategies, and its results have been described, in some cases, as equivalent to other similarity metrics applied to compare two molecules, such as Soergel, Dice, and Cosine, while the similarity measures derived from Euclidean and Manhattan distances have been described as unsatisfactory (Bajusz et al., 2015; Rácz et al., 2018). However, the Tversky coefficient has been indicated to compare moieties of natural products or non-symmetrical scaffolds seeking to identify drug-like similarities (O'Hagan and Kell, 2016). Tanimoto and Tversky coefficient values range from 0 to 1, where values close to 1 correspond to a high similarity between the two analyzed molecules and values close to 0 represent a low similarity (Senger, 2009; Bajusz et al., 2015).

Ligand-Based and Structure-Based Pharmacophore Modeling

A pharmacophore model consists of a set of chemical groups with a specific 3D arrangement that are involved in biological activity against a specific molecular target (Schaller et al., 2020). The functional characteristics present in a pharmacophore model include hydrogen bond acceptors, hydrogen bond donors, hydrophobic groups, positive or negative ionizable groups,

and coordination with metal ions (Vuorinen and Schuster, 2015; Schaller et al., 2020). The binding sites of ligands have physicochemical and spatial restrictions that impose limitations to the non-specific interaction of certain molecules, such as the physicochemical properties of the amino acid residue composition, the volume, and the shape of the cavity. These spatial restrictions dictate the binding mode of the ligands, thus allowing different molecules, even with different structures, to act against a specific bioreceptor due to the presence of the same pharmacophore model (Vuorinen and Schuster, 2015).

Pharmacophore modeling has been extensively applied in virtual screening, lead compound optimization strategies, and *de novo* drug design strategies (Akram et al., 2017; Azminah et al., 2019; Da Costa et al., 2019; El Kerdawy et al., 2019; Jade et al., 2020). Two computational approaches are distinguished in pharmacophore modeling: (1) ligand-based and (2) structure-based approaches. To predict the pharmacophore model, the ligand-based methods use 3D alignment to obtain the chemical information (e.g., shape, functional groups, etc.), shared by a set of active compounds, and select the functional groups that are relevant for the interaction of the ligand with the macromolecular target (Pal et al., 2019). In contrast, the structure-based approach uses the spatial information of the ligand complexed with the molecular target (e.g., ligand poses, conformations, etc.); thus, this approach is applied only in the presence of experimentally elucidated structures of the molecular targets (e.g., by X-ray crystallography) complexed with an active ligand (Jiang et al., 2020).

The ligand-based pharmacophore-based virtual screening comprises different stages: (1) selection of the active compounds validated experimentally; (2) generation of the 3D conformation of the ligands, followed by their structural alignment; (3) identification of the structural characteristics and functional groups involved in molecular recognition; (4) generation and validation of the pharmacophore model using a compound library as a testing dataset; and (5) screening of the natural product library (Figure 3).

In ligand-based pharmacophore modeling, the pharmacophore model is generated using a 3D alignment of the conformers of a set of bioactive compounds (training dataset). Then, active (true-positive compounds or hits) and inactive compounds (false-positive compounds or decoys) are used as a testing dataset to validate the pharmacophore model (Shahin et al., 2016; Pal et al., 2019). It is important to note that, despite the choice of strict pharmacophore models leading to the selection of compounds with better activities against the molecular target, it also could reduce the structural diversity of the analyzed natural products. In contrast, the choice of less restrictive models could retrieve a larger number of false-positive compounds (Schaller et al., 2020).

Pharmacophore modeling methods could be divided into two scoring function methods to predict the fitness of the analyzed compounds to the predicted pharmacophore models: the root of the mean square deviation (RMSD)-based and the overlay-based scoring function (Sanders et al., 2012). In RMSD-based methods, the distances between the functional groups of the compounds to the center of pharmacophore features are used to assess the fitness of the compounds concerning the predicted pharmacophore model. In contrast, the overlay-based methods use the *radii* of the functional groups and/or atoms to estimate the functional similarity of the structures with the pharmacophore model (Vuorinen and Schuster, 2015). Pharmacophore-based methods that apply RMSD-based scoring functions are better at predicting the ligand poses than the overlay-based scoring functions (Sanders et al., 2012). Nevertheless, the ratio of correctly predicted poses vs. incorrectly predicted poses is better obtained using overlay-based scoring functions (Sanders et al., 2012). Regarding structure-based pharmacophore modeling, the use of experimental structures to build the models must prioritize some structural features obtained from both methods; as an example, it has been demonstrated that a higher flexibility obtained in structures elucidated by nuclear magnetic resonance (NMR) spectroscopy helps to focus the models on the most essential interactions with the receptor due to the presence of structural

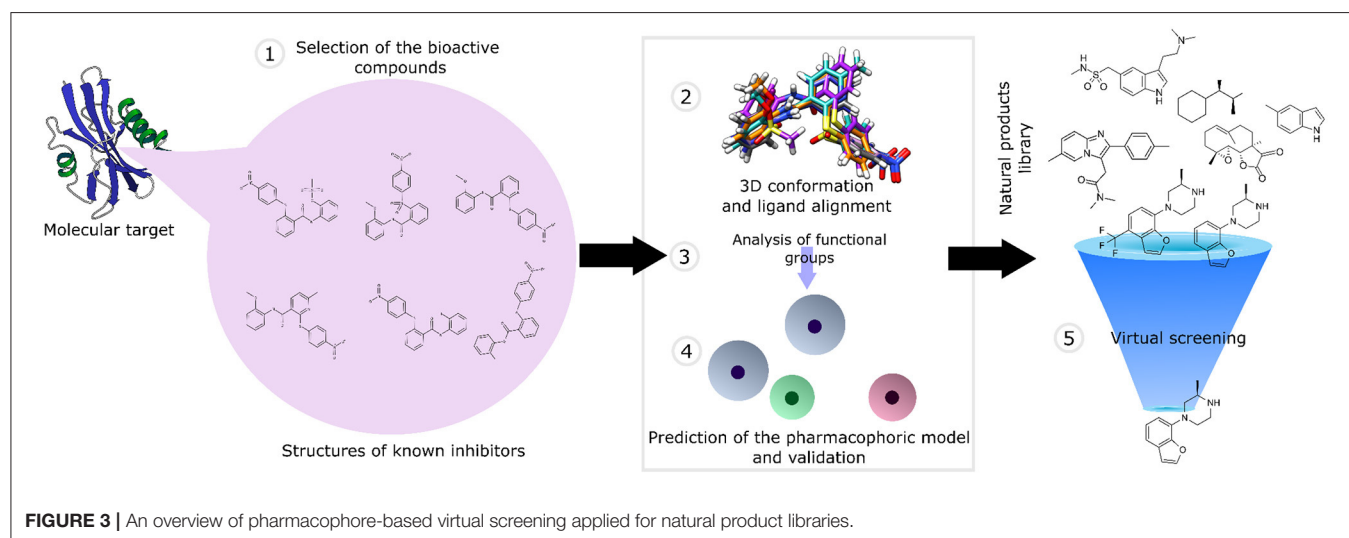


FIGURE 3 | An overview of pharmacophore-based virtual screening applied for natural product libraries.

flexibility of the complexes evidenced by the method. On the other hand, models obtained by X-ray crystallography had more pharmacophore elements compared to those obtained by NMR spectroscopy (Ghanakota and Carlson, 2017).

Pharmacophoric screening has been applied to screen compounds with cosmetic purposes using essential oils (Santana et al., 2018; Da Costa et al., 2019). Essential oils contain diverse classes of volatile and low-molecular-weight compounds with a broad spectrum of biological activities (Do Nascimento et al., 2020), and due to their reported repellent activities against mosquitos, these compounds have been investigated in virtual screening strategies (Santana et al., 2018; Thireou et al., 2018). Recently, a study performed an *in silico* analysis of 1,633 compounds from the essential oils of 71 botanical families by combining a structural similarity-based search method (ligand-based virtual screening) with a pharmacophore-based virtual screening (structure-based strategy). The authors used, as a reference, the structure of *N,N*-diethyl-*meta*-toluamide (DEET) complexed to the odorant-binding protein of *Anopheles gambiae*, and they found seven natural volatile compounds with potential repellent activity against mosquitos, such as *p*-cymen-8-yl, thymol acetate, carvacryl acetate, thymyl isovalerate, and *p*-anisyl hexanoate (Da Costa et al., 2019).

Currently, different programs generate pharmacophore models, differing in the algorithm applied to evaluate the conformational ligand flexibility as well as to perform the structural alignment. Some commercial programs applied to pharmacophore prediction include LigandScout (Wolber and Langer, 2005) and Molecular Operating Environment (MOE) (Molecular Operating Environment, 2019). Both programs apply ligand- and structure-based pharmacophore modeling and are compatible with the most used operating systems. Some open-source programs that use ligand-based pharmacophore prediction include Pharmer (<https://sourceforge.net/projects/pharmer/>) (Koes and Camacho, 2011) and Align-it (previously named Pharaoh; OS compatibility: OS X) (Taminau et al., 2008). Free-access web servers have also been developed to screen compounds using the structure-based pharmacophore approaches, such as Pharmit (<http://pharmit.csb.pitt.edu/>) (Sunseri and Koes, 2016) and PharmMapper (<http://www.lilab-ecust.cn/pharmmapper/>) (Liu et al., 2010).

3D Shape-Similarity Search Methods

The molecular shape acquired by a ligand is crucial to defining its affinity and selectivity against the protein binding site (Kortagere et al., 2009). Based on this assumption, the 3D shape-similarity search methods assume the premise that two compounds could be recognized by the same bioreceptor and then modulate their activity (Koes and Camacho, 2014; Kumar and Zhang, 2018). Shape-similarity methods can screen vast compound libraries against a reference ligand with known bioactivity (Ai et al., 2014; Koes and Camacho, 2014).

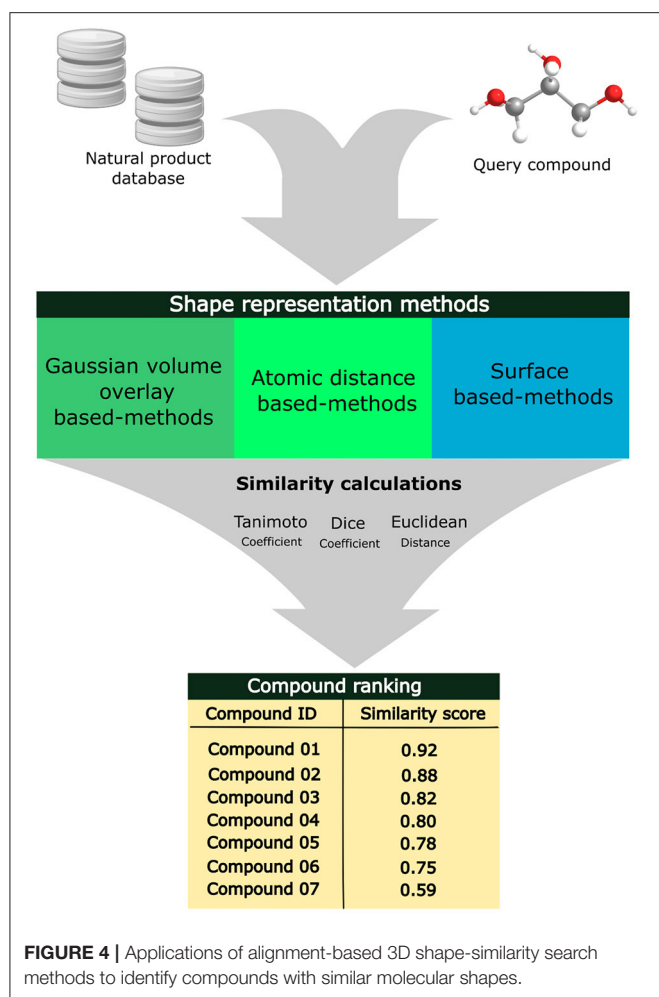
These methods are subdivided into two categories: (1) alignment-free methods that are usually computationally faster because they do not require overlapping the molecules or evaluating properties related to the surface (Seddon et al., 2019) and (2) alignment-based methods that are computationally costly since these methods superimpose molecular shapes and analyze

surface properties, such as polarity and hydrophobicity (Fontaine et al., 2007; Kumar and Zhang, 2018). Different methods have been used in the representation of the 3D molecular shape of the ligands, such as Gaussian overlay-based methods (Cai et al., 2013), atomic distance-based methods (Ballester et al., 2009; Ballester, 2011; Bonanno and Ebejer, 2020), and surface-based methods (Karaboga et al., 2013; Cleves et al., 2019). The recognized molecular shapes are transformed into the 3D molecular fingerprints that are then compared using similarities or distance indexes, such as Tanimoto, Dice, and Tversky coefficients (Shin et al., 2015). Due to the complex structure of natural products, the identification of their molecular targets has been challenging even using computational tools; however, the 3D shape-based similarity search methods have emerged as an efficient strategy to predict the macromolecular targets of these compounds (Shin et al., 2015; Chen et al., 2020). Web servers that apply shape-similarity search methods include the SHAFTS (Liu et al., 2011) and USR-VS (Li et al., 2016). Some installable open-source programs include Shape-it (OS compatibility: Linux) (Grant et al., 1996), gWEGA (Yan et al., 2014), and OptiPharm (Puertas-Martín et al., 2019). Some commercial solutions include Shape TK (OS compatibility: Linux/MS-Windows/macOS) (Software O Scientific, 2008).

Shape-based similarity methods have been used in virtual screening workflows alone or combined with different computational techniques (Pavadai et al., 2017; Thireou et al., 2018). Pavadai et al. applied shape-based and fingerprint-based similarity search against natural product libraries to find new steroid-like natural products as antiparasitic agents using, as a search key, fusidic acid. The hit compounds were filtered based on the predicted partition coefficient, $\log P$, and the authors identified nine new compounds that inhibited parasite growth with IC_{50} values of $<20 \mu M$ (Pavadai et al., 2017). **Figure 4** exhibits an overview of the 3D shape-similarity search methods applied to identify compounds in chemical libraries with similar molecular shapes despite their different structures.

Machine Learning Algorithms

Machine learning (ML) is the computational practice of using intelligent algorithms to learn and make decisions in order to solve problems related to an amount of data. Artificial Intelligence has made important progress toward the acceleration of research and development of novel bioactive natural compounds with industrial applications. This approach has been widely applied in different steps related to the virtual screening strategies, for example to predict some pharmacokinetic properties (Wei et al., 2017; Qiang et al., 2018) [e.g., penetration of compounds into the blood-brain barrier (Zhang et al., 2017; Dai et al., 2021) and cell membrane (Wei et al., 2017; Wolfe et al., 2018)], compounds' side effects (Dimitri and Lió, 2017), their toxicity (Mayr et al., 2016; Pu et al., 2019; Zheng et al., 2020), molecular targets (Wang et al., 2013; Jeon et al., 2014), and their bioactivity (Li and Huang, 2012; Schaduangrat et al., 2019; Shoombuatong et al., 2019) [e.g., anti-tuberculosis (Gomes et al., 2017; Maia S. M. et al., 2020), anticancer (Charoenkwan et al., 2021), and insecticidal activities (Soares Rodrigues et al., 2021)] as well as to identify the pan-assay interference compounds (PAINS), i.e., highly reactive and



promiscuous molecules that are often false positives in high-throughput screening assays (Jasial et al., 2018). In some cases, the ML algorithms have been reported with superior efficiency and, thus, are more suitable to predict hit compounds from chemical libraries than are the traditional QSAR methods (Tsou et al., 2020).

ML algorithms are trained using a large number of data that are used as a benchmark to accomplish a particular computational problem (Vamathevan et al., 2019). The main aim of an ML framework in virtual screening strategies is to generalize the results obtained from the training dataset to better evaluate the test dataset and, then, make the decision (Sieg et al., 2019; Vamathevan et al., 2019). ML algorithms applied in the LBVS approach aim to predict the bioactivity or pharmacodynamic/pharmacokinetic properties of molecules based on their similarity to known actives. Therefore, to evaluate the similarity of the molecules, these algorithms use, as datasets, molecular descriptors calculated from the compound structures (Li and Huang, 2012; Challa et al., 2020) using different molecular modeling and cheminformatics toolkits, such as RDKit (Lovrić et al., 2019) and CDK (Willighagen et al., 2017). Some chemo-structural and bioactivity information deposited in public databases, as well as experimental results, have also been used to

train these algorithms (Martínez-Treviño et al., 2020). Molecular descriptors applied to evaluate the similarity of molecules include the physicochemical [cLogP, topological polar surface area (tPSA), molecular weight, etc.] and structural properties (rotatable bonds, aromatic rings, etc.) (Lo et al., 2018), molecular fingerprints (Zhang et al., 2018), functional groups, molecular shape (Bonanno and Ebejer, 2020), and pharmacophores (Sato et al., 2010); in the case of proteins and peptides, some molecular descriptors include amino acid sequence composition (Wei et al., 2017; Manavalan et al., 2018; Qiang et al., 2018). The choice of the molecular representation and the type of molecular descriptor determine the efficiency and the interpretability of the final results obtained by the ML algorithms (David et al., 2020; Jiménez-Luna et al., 2020). In structure-based strategies, ML algorithms have been used in scoring the functions of molecular docking methods, seeking rank compound libraries based on their predicted affinity against a molecular target, and discriminating between hits and decoy compounds. To reach these results, the ML algorithms are trained using the binding affinities of active molecules against protein targets (Wójcikowski et al., 2017; Li et al., 2020). Different open-source programs have been applied to develop machine learning models [e.g., scikit-learn (Pedregosa et al., 2011) and SciPy (Virtanen et al., 2020), both Python modules] and pipelines [e.g., KNIME (Berthold et al., 2009), a data analytics platform].

ML algorithms are classified into supervised and unsupervised learning (Figure 5). Supervised ML algorithms require a retrospective validation using a dataset of active and inactive compounds to better select the methods that are suitable to differentiate the bioactive molecules (Sieg et al., 2019). Supervised learning techniques are divided into two subgroups: (1) regression analysis and (2) classifier methods. The first one includes decision trees, artificial neural networks, support vector machines, and random forest methods. In contrast, the unsupervised algorithms recognize patterns in the dataset of compounds without the presence of inactive ones, thus trying to organize the data in a logical form. These methods have been used for exploratory analyses using clustering data (Patel et al., 2020). Unsupervised algorithms include clustering methods, such as the hidden Markov model, hierarchical clustering, and *k*-means (Vamathevan et al., 2019).

Supervised ML algorithms have been widely applied to discover new bioactive natural products (Bilsland et al., 2015; Galúcio et al., 2019; Grisoni et al., 2019; Schaduengrat et al., 2019). Figure 6 exhibits a general overview of the computational steps involved in obtaining a validated supervised ML algorithm to predict the bioactivity of natural products. The first step to modeling a machine learning algorithm involves the preparation of a molecule dataset, i.e., obtaining the molecular structures/sequences that will be used in the algorithm using online databases, literature, or experimental data. This step also includes the correction of possible stereochemical and valence errors present in the molecular structures as well as the correction and conversions of the files to readable formats recognized by the cheminformatics programs. Then, the molecular properties are calculated using molecular modeling and cheminformatics toolboxes, extracted from online databases, or obtained from experimental results, then these descriptors are

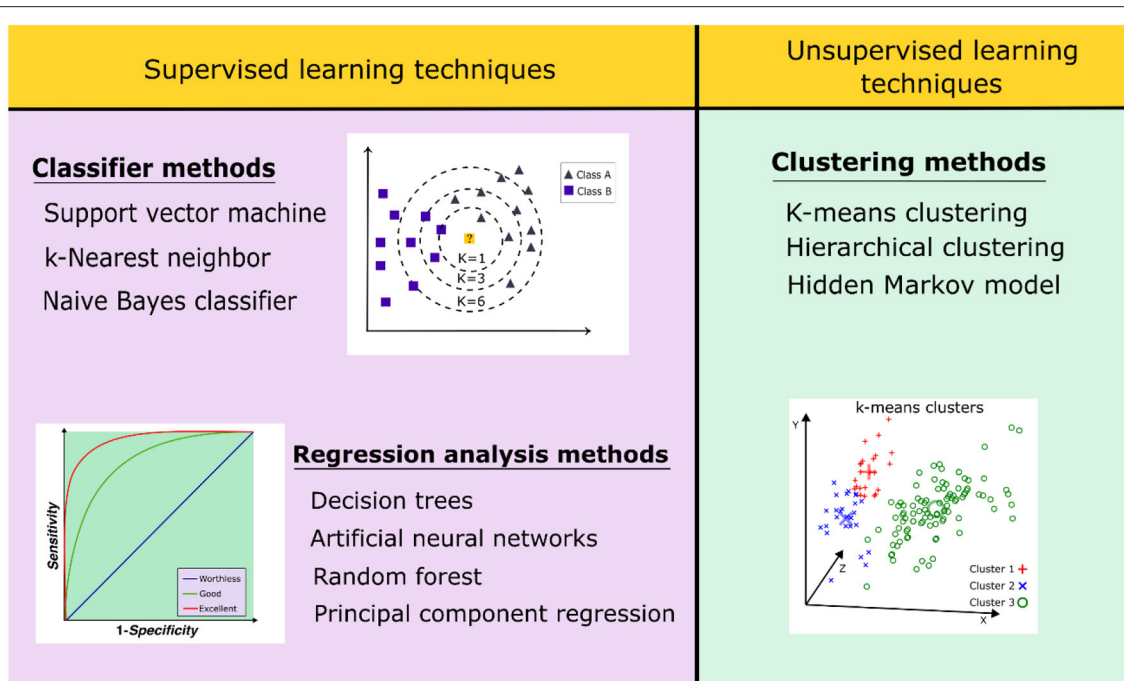


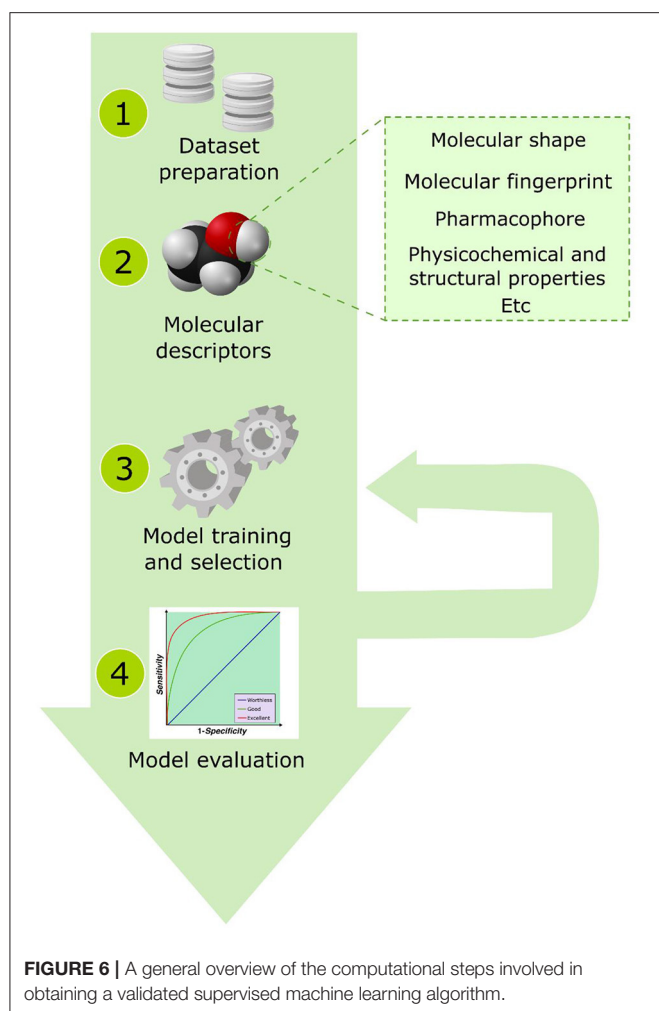
FIGURE 5 | Classification of supervised and unsupervised learning techniques applied in virtual screening strategies.

evaluated to compose the features of the ML model. Currently, different online databases have been developed with information regarding the structural and physicochemical properties of the molecular structure of natural products that could be used in the feature composition (Dunkel et al., 2006; Pilon et al., 2017; Pilón-Jiménez et al., 2019; Sorokina and Steinbeck, 2019). In this step, some statistical methods are applied to select the features, such as Kendall correlation, analysis of variance (ANOVA), and Spearman's test. Finally, the ML model is evaluated regarding its performance to discriminate the true and positive bioactive compounds. Several metrics have been applied to evaluate these models, such as the receiver operating characteristic (ROC) curve, enrichment factors, and mean squared error (R^2) applied for linear regression methods. We do not intend to extend the discussion about the application and the choice of the most adequate method to select the feature composition or to evaluate ML models; thus, we recommend the readers to consult previous reviews (Hossin and Sulaiman, 2015; Rácz et al., 2019). In the present sessions, we will discuss the functioning of some ML algorithms most applied in virtual screening strategies focusing on the *k*-nearest neighbor, decision tree, random forest, artificial, and neural network.

Decision tree algorithms are a supervised learning technique and their construction model is based on two steps: (1) selection of the features and (2) the building of the decision trees. This method is commonly represented by a tree, where the internal nodes represent the selected features (molecular descriptors), the branches represent the testing results of the molecule (decision criteria), and the leaf nodes represent the molecules (molecular structure) (Figure 7A). Compounds are classified based on the leaf nodes that are reached through a series of

algorithm decisions (branches). Decision tree (DT) models are constructed focusing on the selection of the best test conditions to expand the extremities of the tree. Some test metrics, such as the information-gain ratio and entropy, are applied to select the best test classification for the algorithm (Lavecchia, 2015). Decision trees have been applied in different virtual screenings of natural products to predict their bioactivity and drug-likeness (Pereira et al., 2015; Wang et al., 2019). Random forest is an ensemble learning technique considered an improvement of the decision tree algorithms to correct the overfitting in the training set (Svetnik et al., 2003). Random forest algorithms generate a model composed of several randomly sampled decision trees from the original dataset obtaining its random features. Random forest models have been applied in virtual screening pipelines to predict compound drug-likeness, bioactivity (Svetnik et al., 2003; Zoffmann et al., 2019), and the pharmacokinetic profile (Dong et al., 2018).

Artificial neural networks are the most studied learning techniques with widely diverse applications in the investigation of a compound's bioactivity (Lata et al., 2007; Liu et al., 2019, 2020; Stokes et al., 2020). Methods that apply neural networks mimic brain functioning and structure, building a model that reaches a decision based on previous experiences obtained from the training dataset (Jing et al., 2018). The architecture of an artificial neural network model comprises several units, named neurons which are connected to form a network arranged in different layers. Depending upon their position in the network, these layers are classified as output layers, input layers (external), and hidden layers (internal) (Zhang R. et al., 2020). A multilayer feed-forward neural network contains neurons connected only to those located in the following layers (Figure 7B), and this class is included in



radial basis function networks, multilayer perceptrons, and self-organizing maps (Kohonen maps) (Lavecchia, 2015). In contrast, the recurrent neural networks contain feedbacks between the layers, i.e., interconnections between neurons from the same and consecutive layers; thus, their outputs are determined by the previous outputs and the current inputs (Figure 7B), which form a “memory” during the learning process.

The k -nearest neighbor is instance-based learning and is one of the simplest and intuitive ML algorithms applied to classify and rank compounds based on the nearest training examples present in the chemical space (analyzed feature composition) (Kauffman and Jurs, 2001; Medina-Franco et al., 2005). The algorithm compares the molecular descriptors of the query molecule with k -neighbors that have the smallest distance (k -value), where the k -value corresponds to the number of closest neighbors (a positive integer) and classifies them by majority votes of their closest neighbors (Figure 7C). The number of neighbors is the most important parameter for the model, deciding its complexity. k -nearest neighbor is a classifier algorithm; thus, irrelevant features can lead to disturbances in the compound classification. It is indicated to first preprocess

the molecular descriptors to remove the irrelevant or the most correlated ones.

Despite the majority of the computational screening approaches using ML algorithms lacking experimental validations, we have some interesting successful studies that aimed to find and characterize novel natural products with experimentally validated biological activity (Rupp et al., 2010; Zhang et al., 2017; Nocedo-Mena et al., 2019; Patsilinakos et al., 2019; Lee et al., 2020; Liu et al., 2020). Recently, Reher et al. reported on the SMART 2.0, an NMR-based machine learning tool designed for the discovery and characterization of natural products. The tool was successfully applied to investigate the environmental extract of *Symplocos* sp., a filamentous marine cyanobacterium, leading to the isolation and identification of a new chimeric macrolide named symplocolide A. The molecular structure of this novel natural product was confirmed by 1D/2D NMR and tandem liquid chromatography mass spectrometry (LC-MS²) analysis (Reher et al., 2020). Similarly, Lee et al. applied SMART 2.0 to prioritize the isolation and characterization of sesquiterpene lactones from the *Eupatorium fortune* plant. The isolated natural compounds were experimentally tested against five cancer cell lines and exhibited cytotoxic activities (Lee et al., 2020).

ML algorithms have been successfully applied to predict the bioactivity of compounds. Recently, Nocedo-Mena et al. (2019) combined machine learning, perturbation theory, and information fusion techniques to investigate the antibacterial activity of terpenes from the *Cissus incisa* plant, and the authors found that phytol and α -amyrin showed minimum inhibitory concentrations equal to 100 μ g/ml against the carbapenem-resistant *Acinetobacter baumannii* and the vancomycin-resistant *Enterococcus faecium*. In another study, Liu et al. applied deep learning algorithms to find natural products with anti-osteoporosis activity. The selected hits successfully suppressed the osteoclastogenesis-related genes *Rank*, *Tracp*, *Ctsk*, and *Nfatc1* *in vitro* (Liu et al., 2020). Some studies have also reported experimental validations of ML models to predict pharmacokinetic properties. Zhang et al. used a hybrid ML algorithm using support vector machine, probabilistic neural network, naive Bayes classifier, and random forest models combined with *in vitro* assays to predict the blood–brain barrier penetration of natural compounds from the Traditional Chinese Medicine database (TCMD_B). The authors found an overall accuracy for experimental validation around 81% (Zhang et al., 2017).

BIASES AND LIMITATIONS OF VIRTUAL SCREENING METHODS

Virtual screening approaches have been predictive, useful, and cost-effective in identifying novel bioactive compounds when compared with the traditional methods applied solely. However, despite their well-known success, these methods have limitations and their models are prone to biases (Sieg et al., 2019; Slater and Kontoyianni, 2019). It has been demonstrated that the presence of stereochemical and valence errors in the chemical data libraries

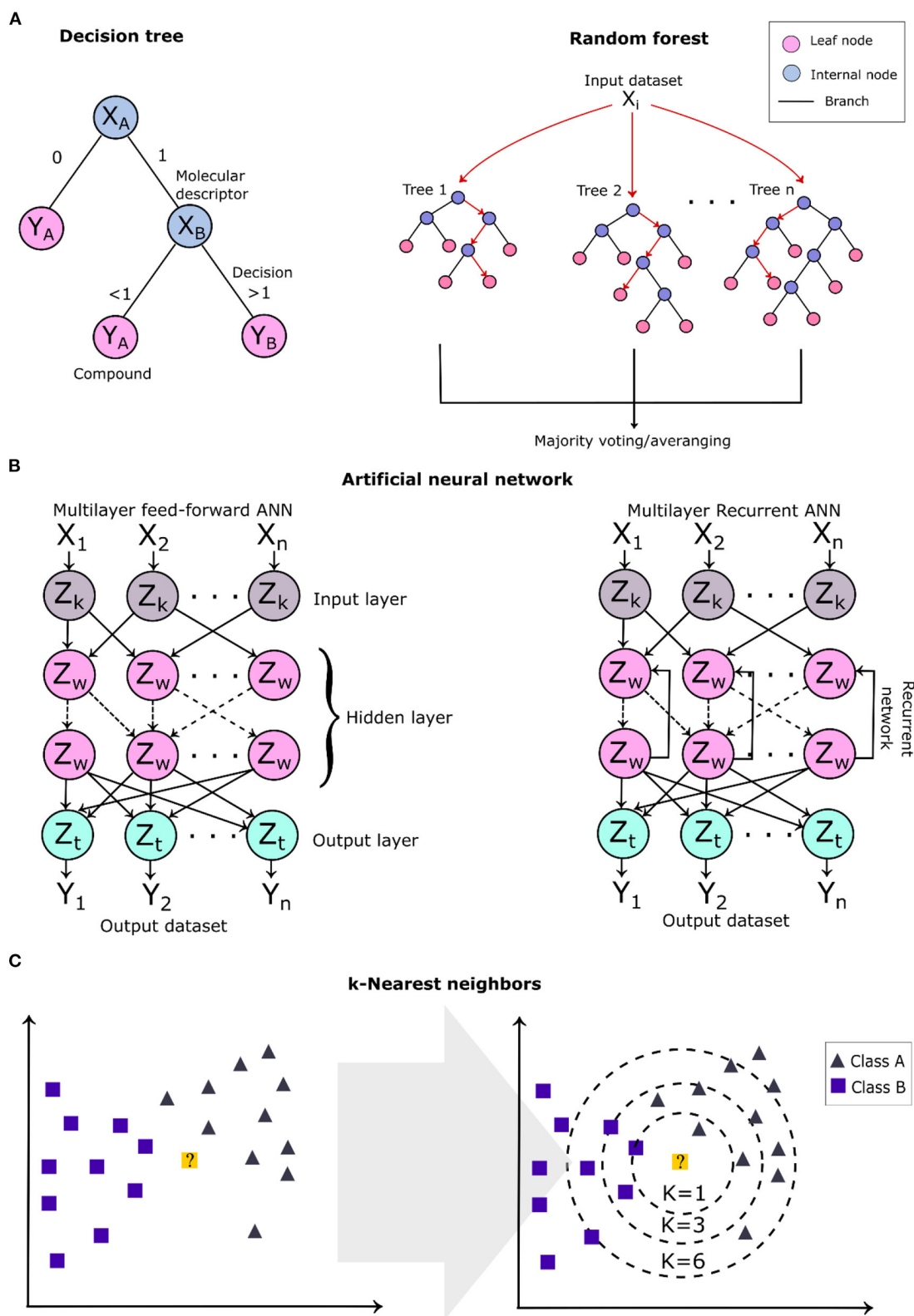


FIGURE 7 | Schematic overview of some of the machine learning algorithms applied in virtual screening. **(A)** Two-dimensional (2D) diagram of a single root tree of a decision tree algorithm and the general architecture of a random forest. **(B)** The architecture of a multilayer feed-forward and recursive artificial neural network. Z_w refers to neurons of the hidden layers (internal); Z_k and Z_t , to the neurons of the input and output layers, respectively. **(C)** k -Nearest neighbor algorithm showing the learning technique to classify a new data represented by the 2D yellow point, which is classified as belonging to class A (gray triangles).

could also induce investigators to choose unfeasible compounds (Williams and Ekins, 2011; Williams et al., 2012).

Biases, in essence, correspond to distortions from the true underlying relationship between the investigated objects. The investigation of the chemo-structural diversity of natural products and their bioactivity using similarity-based search methods is biased because it considers an assumption that the discovery of novel active compounds must consider the similarity of known active ones (Sieg et al., 2019). This assumption is susceptible to drive the decision-making process to erroneous directions and can reduce the structural diversity of new chemo-structures. Combining low time-consuming computational simulations and more realistic results also remains a challenge for some 3D similarity-based search algorithms, which, in general, require superimposing many conformation pairs of compounds from large chemical libraries, thus requiring high-performance computing (Yan et al., 2016).

Despite the chemical space being considered infinite, the pharmacological space of bioactive compounds of the “druggable human genome” is limited, and its exploration remains a difficult task even from a computational point of view (Opassi et al., 2018). This assumption has been proven to be true for other classes of bioactive compounds with industrial applications, such as pesticides and herbicides (Avram et al., 2014). Therefore, the exclusion of some compounds during the filtering process is comprehensive, but can also reduce the investigation of new chemical entities with specific bioactivity.

In pharmacophore-based virtual screening, the selection of inappropriate models, or very restricted ones, could eliminate an interesting structural diversity of natural compounds. However, the choice of less restrictive models could retrieve a larger number of false-positive compounds (Lans et al., 2020; Schaller et al., 2020). Based on these biases, a balanced choice between strict and loose criteria to select the pharmacophore model to filter natural products could be decided by prioritizing pharmacophore moieties better associated with a higher compound activity; thus, the information obtained from structure–activity analyses might be useful to decide on the most appropriate pharmacophore model to screen natural products (Qing et al., 2014). Regarding the limitation of ligand-based pharmacophore modeling methods, it has been reported that their dependence on structurally similar compounds reduces their application since compounds with high structural dissimilarities may not share the same binding mode (Schaller et al., 2020). Furthermore, few ligand-based methods consider the conformational flexibility of the macromolecular receptor in the determination of the pharmacophore model (Lans et al., 2020). In molecular docking, for example, the elimination of compounds with poor fitness could be biased due to the choice of wrong or inappropriate scoring functions, i.e., those that contain chemical information that contradicts the physical reality or that were not calibrated for the class of investigated molecules (Luo et al., 2017).

Supervised machine learning algorithms are also prone to biases, which can lead to a misleading interpretation of the final results obtained for chemical data libraries. It has been demonstrated that highly correlated training and testing datasets, i.e., containing chemical data too closely similar (e.g., same

molecular scaffold with a high frequency between the datasets), could limit the applicability of the machine learning model, reaching false accuracies in its predictiveness (Wallach and Heifets, 2018; Sieg et al., 2019). Therefore, low training errors are insufficient to justify the choice of a machine learning model since the satisfactory predictive performance could be due to redundancy between the training and testing datasets rather than accuracy (Wallach and Heifets, 2018). It has also been demonstrated that some biased machine learning models could be obtained using a training dataset composed of active molecules that are easily differentiated from inactive ones by coarse properties, such as cLogP, the number of HBA, and molecular weight (Ripphausen et al., 2011). Based on these biases of machine learning models, it is necessary to investigate whether chemical data benchmarks contain design flaws that might lead to optimistic performances that are distorted from the chemical reality. Some computational methods have been developed to avoid overfitting in chemical datasets. Wallach and Heifets (2018) developed the asymmetric validation embedding (AVE) bias using Python language to predict the performance across common benchmarks and standard machine learning algorithms, and Ripphausen et al. (2011) developed a public compound database, named REPROVIS-DB, that contains information from successful ligand-based virtual screening strategies including experimentally confirmed hits, reference compounds, screening databases, and selection criteria.

NATURAL PRODUCTS DATABASES APPLIED IN VIRTUAL SCREENING

The development of computational approaches for virtual screening has been incentivized by the presence of numerous biological and chemo-structural information of natural products deposited in public databases (Valli et al., 2013; Harvey et al., 2015; Pilon et al., 2017), as well as by the advances of computer processing and storage capacity (Walters, 2019). High scientific efforts to isolate and characterize natural products have increased the interest of the academia and industry to comprehensively organize this information using public databases to better explore these natural sources and also to contribute to our knowledge regarding their ethnobotanical information, biological activities, chemical structures, natural origin, and physicochemical properties. Herein, we do not intend to provide exhaustive information regarding these online databases with public access, but we will exhibit those with potential applications in virtual screening strategies of natural products.

Nuclei of Bioassays, Ecophysiology, and Biosynthesis of Natural Products Database (NuBBE_{DB})

NuBBE_{DB} (<https://nubbe.iq.unesp.br/portal/nubbe-search.html>) provides information regarding chemo-structures obtained from Brazilian biodiversity (Valli et al., 2013). Currently, the database contains more than 2,200 structures of natural compounds obtained from different Brazilian biomes (Pilon et al., 2017). NuBBE_{DB} contains the 3D structures of natural products in an

MOL2 file format, which is compatible with the most widely used molecular modeling and cheminformatics programs.

Comprehensive Marine Natural Products Database (CMNPD)

The Comprehensive Marine Natural Products Database (CMNPD) (<https://www.cmnpd.org/>) is a comprehensive and curated marine natural products database that contains more than 32,000 structures (accessed on January 06, 2020) with different physicochemical and pharmacokinetic properties. Besides, it includes information regarding their biological activity, natural origin, and the geographical distribution of source organisms (Lyu et al., 2020). The database also contains the complete molecule datasets freely available for download (<https://docs.cmnpd.org/downloads>).

Natural Product-Likeness Software Suite and Database (NaPLeS)

The natural product-likeness software suite NaPLeS (<https://naples.naturalproducts.net/>) is an MySQL database of natural products and an open-source web application that computes the natural product-likeness scores of large chemical libraries. Currently, the database contains 315,916 natural products from various public databases (Sorokina and Steinbeck, 2019).

Universal Natural Product Database (UNaProd)

The Universal Natural Product Database (UNaProd) (<http://jafarilab.com/unaprod/index.php>) is an online and public database of natural products used in Iranian traditional medicine. The database currently contains 2,696 compounds of botanical, animal, and mineral origins (accessed on January 06, 2020) (Naghizadeh et al., 2020).

Natural Product Activity and Species Source Database (NPASS)

The Natural Product Activity and Species Source Database (NPASS) (<http://bidd.group/NPASS/index.php>) provides biological activity results and information regarding the origin species of more than 35,032 natural products (accessed on January 06, 2020) (Zeng et al., 2018). The database also contains a structural compound library freely available for download in SDF and SMILES formats (<http://bidd.group/NPASS/downloadnpass.html>).

BIOFACQUIM

BIOFACQUIM (<https://biofacquim.herokuapp.com/>) is a free and public database of natural products isolated and characterized from Mexican biodiversity. Compounds from this database are also available in the ZINC database (Pilón-Jiménez et al., 2019). Currently, the database contains 423 natural compounds (accessed on January 08, 2020) which are identified by their respective names, accession codes, source organisms, in SMILE format, and references.

Natural Products Atlas

The Natural Products Atlas (<https://www.npatlas.org/joomla/>) is an open-access database of microbial natural products that contain 24,594 compound structures (accessed on January 07, 2020) and information related to their structure, IUPAC name, source organisms, and literature (van Santen et al., 2019). The database also contains information of other natural product databases, such as the Minimum Information about a Biosynthetic Gene Cluster (MIBiG) repository and the Global Natural Products Social Molecular Networking (GNPS) platform (van Santen et al., 2019).

African Natural Products Database (ANPDB)

The African Natural Products database (ANPDB) is a free database of natural products from different regions of the African continent (available at ANPDB|ANPDB (African-compounds.org) and contains ~4,500 structures (accessed on January 12, 2020). The available data content comprises sources covering the period from 1962 to 2019 (Ntie-Kang et al., 2017). The database also contains the 3D structures of natural products in SMILES and SDF formats available for non-commercial uses.

Natural Products for Cancer Regulation (NPCARE)

The Natural Products for Cancer Regulation (NPCARE) is a free online database (<http://silver.sejong.ac.kr/npcare/>) that provides more than 6,000 natural products and more than 2,000 extracts isolated from 1,952 different species including microorganisms, marine organisms, and plants, as well as information related to the action of these extracts and isolated natural compounds against the gene expression levels and cancer cell line inhibition (Choi et al., 2017). The database is an interesting source to discover potential anticancer compounds and to understand the anticancer molecular mechanisms underlying natural products.

StreptomeDB 3.0

StreptomeDB (<http://www.pharmbioinf.uni-freiburg.de/streptomedb>) is a free and online database used to explore natural products isolated or mutasynthesized from streptomycetes using an interactive phylogenetic analysis (Lucas et al., 2013; Moumbock et al., 2021). StreptomeDB 3.0 provides more than 6,500 natural products obtained from ~3,300 *Streptomyces* strains (Moumbock et al., 2021). These metabolites show interesting biological activities, such as antimicrobial, anticancer, and immunosuppressant properties. The compound structures are identified by their respective source organisms, references, biological role, and the routes of biosynthesis.

FINAL CONSIDERATIONS

Natural products offer an interesting structural scaffold, helping to find new chemical entities with several industrial applications, thus offering innovative solutions to solve old worldwide problems, such as bacterial resistance against antibiotics (Smith et al., 2018; Newman and Cragg, 2020). However, the complex and highly diverse structure and the peculiar chemical space

occupied by natural products have imposed pharmacokinetic and pharmacodynamic limitations, thus restricting their use for specific purposes by the pharmaceutical and cosmetic industries.

Several computational methods applied in virtual screening strategies have been developed over the years, thus increasing the rational explorations of natural sources aiming at the identification of specific bioactive compounds from large chemo-structural libraries. These computational strategies have also opened up new opportunities to discover new industrial applications of natural compounds justifying the financial and time efforts for their exploration. Natural products present a high structural diversity when compared with their synthetic counterparts, and their difference is, in part, due to the existing intricate biosynthetic pathways in living organisms that produce derived structures, containing modified functional groups, such as glycosylation and methylation. Based on these, the virtual screening strategies must investigate the chemical space of natural products, seeking to identify some classes of compounds with bioactivity or structural scaffolds present in known active molecules. Some of these screening strategies include applying less restrictive structural-based similarity cutoffs (Pavadai et al., 2017) and the modeling of hypothetically derived natural product structures (Skinnider et al., 2017). Regarding the application of molecular filters, some “bioactivity-likeness” criteria must be used with caution to avoid misleading screening or remotion of the important structural diversity of the compound libraries since the structural complexity of natural products situates them beyond the acceptable limits of some empirical rules determined by these filters.

Artificial intelligence algorithms employed in ligand-based approaches have demonstrated high success rates in finding interesting compounds with reduced computational time, and their combined uses with cheminformatics and molecular modeling methods have increased the efficiency of virtual

screening strategies, allowing us to explore the highly diverse chemo-structural landscapes of natural products.

Here, we hope to encourage the use of these computational tools by experimental groups, helping researchers to familiarize themselves with their concepts and capabilities as well as alert them of some of the common biases faced by investigators during the investigation of natural sources using computational tools, citing some possible solutions. Finally, we indicate that the automatic process represented by virtual screening must be oriented by human expert decision to avoid misinterpretation or false findings, and also to select compounds based on their desirable features, such as commercial availability, low cost, and synthetic feasibility.

AUTHOR CONTRIBUTIONS

KS, RB, and JL: conceptualization. KS, LN, AL, and VD: investigation. KS, LN, VD, and RB: writing—original draft preparation. KS, AL, CN, RB, and JL: writing—review and editing. CN, RB, and JL: supervision. All authors have read and agreed to the published version of the manuscript.

FUNDING

We would like to thank the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) and Pró-Reitoria de Pesquisa e Pós-Graduação (PROPESP/UFPA) for providing the financial support for the scientific research. KS was also grateful for the scholarship from the Brazilian funding agency Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES, grant number: 88882.466102/2019-01). The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

REFERENCES

- Abdelhameed, R. F. A., Habib, E. S., Eltahawy, N. A., Hassanean, H. A., Ibrahim, A. K., Mohammed, A. F., et al. (2020). New cytotoxic natural products from the red sea sponge *styliassa carteri*. *Mar Drugs* 18:241. doi: 10.3390/md18050241
- Ai, N., Welsh, W. J., Santhanam, U., Hu, H., and Lyga, J. (2014). Novel virtual screening approach for the discovery of human tyrosinase inhibitors. *PLoS ONE* 9:e112788. doi: 10.1371/journal.pone.0112788
- Akram, M., Waratchareyakul, W., Hauptenthal, J., Hartmann, R. W., and Schuster, D. (2017). Pharmacophore modeling and *in silico/in vitro* screening for human cytochrome P450 11B1 and cytochrome P450 11B2 inhibitors. *Front. Chem.* 5:104. doi: 10.3389/fchem.2017.00104
- Al Sharie, A. H., El-Elmat, T., Al Zu'bi YO, Aleshawi, A. J., and Medina-Franco, J. L. (2020). Chemical space and diversity of seaweed metabolite database (SWMD): a cheminformatics study. *J. Mol. Graph. Model.* 100:107702. doi: 10.1016/j.jmgm.2020.107702
- Atanasov, A. G., Zotchev, S. B., Dirsch, V. M., and Supuran, C. T. (2021). Natural products in drug discovery: advances and opportunities. *Nat. Rev. Drug Discov.* 20, 200–216. doi: 10.1038/s41573-020-00114-z
- Avram, S., Funar-Timofei, S., Borota, A., Chennamaneni, S. R., Manchala, A. K., and Muresan, S. (2014). Quantitative estimation of pesticide-likeness for agrochemical discovery. *J. Cheminform.* 6:42. doi: 10.1186/s13321-014-0042-6
- Azminah, A., Erlina, L., Radji, M., Mun'im A, Syahdi, R. R., and Yanuar, A. (2019). *In silico* and *in vitro* identification of candidate SIRT1 activators from Indonesian medicinal plants compounds database. *Comput. Biol. Chem.* 83:107096. doi: 10.1016/j.compbiolchem.2019.107096
- Bajusz, D., Rácz, A., and Héberger, K. (2015). Why is Tanimoto index an appropriate choice for fingerprint-based similarity calculations? *J. Cheminform.* 7:20. doi: 10.1186/s13321-015-0069-3
- Ballester, P. J. (2011). Ultrafast shape recognition: method and applications. *Future Med. Chem.* 3, 65–78. doi: 10.4155/fmc.10.280
- Ballester, P. J., Finn, P. W., and Richards, W. G. (2009). Ultrafast shape recognition: evaluating a new ligand-based virtual screening technology. *J. Mol. Graph. Model.* 27, 836–45. doi: 10.1016/j.jmgm.2009.01.001
- Batool, M., Ahmad, B., and Choi, S. (2019). A structure-based drug discovery paradigm. *Int. J. Mol. Sci.* 20:2783. doi: 10.3390/ijms20112783
- Berenger, F., Vu, O., and Meiler, J. (2017). Consensus queries in ligand-based virtual screening experiments. *J. Cheminform.* 9:60. doi: 10.1186/s13321-017-0248-5
- Berthold, M. R., Cebron, N., Dill, F., Gabriel, T. R., Kötter, T., Meinl, T., et al. (2009). KNIME—the Konstanz information miner. *ACM SIGKDD Explor. Newsl.* 11, 26–31. doi: 10.1145/1656274.1656280
- Bilsland, A. E., Pugliese, A., Liu, Y., Revie, J., Burns, S., McCormick, C., et al. (2015). Identification of a selective G1-phase benzimidazolone inhibitor by a

- senescence-targeted virtual screen using artificial neural networks. *Neoplasia (United States)* 17, 704–715. doi: 10.1016/j.neo.2015.08.009
- Blanco, J. L., Porto-Pazos, A. B., Pazos, A., and Fernandez-Lozano, C. (2018). Prediction of high anti-angiogenic activity peptides *in silico* using a generalized linear model and feature selection. *Sci. Rep.* 8:15688. doi: 10.1038/s41598-018-33911-z
- Bonanno, E., and Ebejer, J. P. (2020). Applying machine learning to ultrafast shape recognition in ligand-based virtual screening. *Front. Pharmacol.* 10:1675. doi: 10.3389/fphar.2019.01675
- Bradley, S. A., Zhang, J., and Jensen, M. K. (2020). Deploying microbial synthesis for halogenating and diversifying medicinal alkaloid scaffolds. *Front. Bioeng. Biotechnol.* 8:594126. doi: 10.3389/fbioe.2020.594126
- Bruns, R. F., and Watson, I. A. (2012). Rules for identifying potentially reactive or promiscuous compounds. *J. Med. Chem.* 55, 9763–72. doi: 10.1021/jm301008n
- Cai, C., Gong, J., Liu, X., Gao, D., and Li, H. (2013). SimG: an alignment based method for evaluating the similarity of small molecules and binding sites. *J. Chem. Inf. Model.* 53, 2103–2115. doi: 10.1021/ci400139j
- Cai, C., Wu, Q., Hong, H., He, L., Liu, Z., Gu, Y., et al. (2021). *In silico* identification of natural products from Traditional Chinese Medicine for cancer immunotherapy. *Sci. Rep.* 11:3332. doi: 10.1038/s41598-021-82857-2
- Cao, D. S., Xu, Q. S., Hu, Q. N., and Liang, Y. Z. (2013). ChemoPy: freely available python package for computational biology and chemoinformatics. *Bioinformatics* 29, 1092–1094. doi: 10.1093/bioinformatics/btt105
- Cereto-Massagué, A., Ojeda, M. J., Valls, C., Mulero, M., Garcia-Vallvé, S., and Pujadas, G. (2015). Molecular fingerprint similarity search in virtual screening. *Methods* 71, 58–63. doi: 10.1016/j.ymeth.2014.08.005
- Challa, A. P., Beam, A. L., Shen, M., Peryea, T., Lavieri, R. R., Lippmann, E. S., et al. (2020). Machine learning on drug-specific data to predict small molecule teratogenicity. *Reprod. Toxicol.* 95, 148–58. doi: 10.1016/j.reprotox.2020.05.004
- Charoenkwan, P., Chiangjong, W., Lee, V. S., Nantasenamat, C., Hasan, M. M., and Shoombuatong, W. (2021). Improved prediction and characterization of anticancer activities of peptides using a novel flexible scoring card method. *Sci. Rep.* 11:3017. doi: 10.1038/s41598-021-82513-9
- Chávez-Hernández, A. L., Sánchez-Cruz, N., and Medina-Franco, J. L. (2020). Fragment library of natural products and compound databases for drug discovery. *Biomolecules* 10:1518. doi: 10.3390/biom10111518
- Chen, Y., Garcia de Lomana, M., Friedrich, N. O., and Kirchmair, J. (2018). Characterization of the chemical space of known and readily obtainable natural products. *J. Chem. Inf. Model.* 58, 1518–1532. doi: 10.1021/acs.jcim.8b00302
- Chen, Y., and Kirchmair, J. (2020). Cheminformatics in natural product-based drug discovery. *Mol. Inform.* 39:2000171. doi: 10.1002/minf.202000171
- Chen, Y., Mathai, N., and Kirchmair, J. (2020). Scope of 3D shape-based approaches in predicting the macromolecular targets of structurally complex small molecules including natural products and macrocyclic ligands. *J. Chem. Inf. Model.* 60, 2858–75. doi: 10.1021/acs.jcim.0c00161
- Chen, Y., Stork, C., Hirte, S., and Kirchmair, J. (2019). NP-scout: machine learning approach for the quantification and visualization of the natural product-likeness of small molecules. *Biomolecules* 9:43. doi: 10.3390/biom9020043
- Choi, H., Cho, S. Y., Pak, H. J., Kim, Y., Choi, J. Y., Lee, Y. J., et al. (2017). NPCARE: database of natural products and fractional extracts for cancer regulation. *J. Cheminform.* 9:2. doi: 10.1186/s13321-016-0188-5
- Cleves, A. E., Johnson, S. R., and Jain, A. N. (2019). Electrostatic-field and surface-shape similarity for virtual screening and pose prediction. *J. Comput. Aided Mol. Des.* 33, 865–86. doi: 10.1007/s10822-019-00236-6
- Coimbra, J. R. M., Baptista, S. J., Dinis, T. C. P., Silva, M. M. C., Moreira, P. I., Santos, A. E., et al. (2020). Combining virtual screening protocol and *in vitro* evaluation towards the discovery of BACE1 inhibitors. *Biomolecules* 10:535. doi: 10.3390/biom10040535
- Congreve, M., Carr, R., Murray, C., and Jhoti, H. (2003). A 'Rule of Three' for fragment-based lead discovery? *Drug Discov. Today* 8, 876–877. doi: 10.1016/S1359-6446(03)02831-9
- Cubillos, F. A., Gibson, B., Grijalva-Vallejos, N., Krogerus, K., and Nikulin, J. (2019). Bioprospecting for brewers: exploiting natural diversity for naturally diverse beers. *Yeast* 36, 383–398. doi: 10.1002/yea.3380
- Da Costa, K. S., Galúcio, J. M., Da Costa, C. H. S., Santana, A. R., Dos Santos Carvalho, V., Do Nascimento, L. D., et al. (2019). Exploring the potentiality of natural products from essential oils as inhibitors of odorant-binding proteins: a structure- and ligand-based virtual screening approach to find novel mosquito repellents. *ACS Omega* 4, 22475–22486. doi: 10.1021/acsomega.9b03157
- Dai, R., Zhang, W., Tang, W., Wynendaele, E., Zhu, Q., Bin, Y., et al. (2021). BBPPred: sequence-based prediction of blood-brain barrier peptides with feature representation learning and logistic regression. *J. Chem. Inf. Model.* 61, 525–34. doi: 10.1021/acs.jcim.0c01115
- Daina, A., and Zoete, V. (2016). A BOILED-egg to predict gastrointestinal absorption and brain penetration of small molecules. *ChemMedChem* 11, 1117–1121. doi: 10.1002/cmdc.201600182
- David, L., Thakkar, A., Mercado, R., and Engkvist, O. (2020). Molecular representations in AI-driven drug discovery: a review and practical guide. *J. Cheminform.* 12:56. doi: 10.1186/s13321-020-00460-5
- Davison, E. K., and Brimble, M. A. (2019). Natural product derived privileged scaffolds in drug discovery. *Curr. Opin. Chem. Biol.* 52, 1–8. doi: 10.1016/j.cbpa.2018.12.007
- 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
- Desaphy, J., Raimbaud, E., Ducrot, P., and Rognan, D. (2013). Encoding protein–ligand interaction patterns in fingerprints and graphs. *J. Chem. Inf. Model.* 53, 623–637. doi: 10.1021/ci300566n
- Dimitri, G. M., and Lió, P. (2017). DrugClust: a machine learning approach for drugs side effects prediction. *Comput. Biol. Chem.* 68, 204–210. doi: 10.1016/j.compbiolchem.2017.03.008
- Do Nascimento, L. D., de Moraes, A. A. B., da Costa, K. S., Galúcio, J. M. P., Taube, P. S., Costa, C. M. L., et al. (2020). Bioactive natural compounds and antioxidant activity of essential oils from spice plants: new findings and potential applications. *Biomolecules* 10, 1–37. doi: 10.3390/biom10070988
- Doak, B. C., Over, B., Giordanetto, F., and Kihlberg, J. (2014). Oral druggable space beyond the rule of 5: insights from drugs and clinical candidates. *Chem. Biol.* 21, 1115–1142. doi: 10.1016/j.chembiol.2014.08.013
- Dong, J., Cao, D. S., Miao, H. Y., Liu, S., Deng, B. C., Yun, Y. H., et al. (2015). ChemDes: an integrated web-based platform for molecular descriptor and fingerprint computation. *J. Cheminform.* 7:60. doi: 10.1186/s13321-015-0109-z
- Dong, J., Wang, N. N., Yao, Z. J., Zhang, L., Cheng, Y., Ouyang, D., et al. (2018). Admetlab: a platform for systematic ADMET evaluation based on a comprehensively collected ADMET database. *J. Cheminform.* 10:29. doi: 10.1186/s13321-018-0283-x
- Dunkel, M., Fullbeck, M., Neumann, S., and Preissner, R. (2006). SuperNatural: a searchable database of available natural compounds. *Nucleic Acids Res.* 34, D678–D683. doi: 10.1093/nar/gkj132
- El Kerdawy, A. M., Osman, A. A., and Zaater, M. A. (2019). Receptor-based pharmacophore modeling, virtual screening, and molecular docking studies for the discovery of novel GSK-3 β inhibitors. *J. Mol. Model.* 25:171. doi: 10.1007/s00894-019-4032-5
- Ertl, P., Roggo, S., and Schuffenhauer, A. (2008). Natural product-likeness score and its application for prioritization of compound libraries. *J. Chem. Inf. Model.* 48, 68–74. doi: 10.1021/ci700286x
- Feher, M., and Schmidt, J. M. (2003). Property distributions: differences between drugs, natural products, and molecules from combinatorial chemistry. *J. Chem. Inf. Comput. Sci.* 43, 218–227. doi: 10.1021/ci0200467
- Floros, D. J., Jensen, P. R., Dorrestein, P. C., and Koyama, N. (2016). A metabolomics guided exploration of marine natural product chemical space. *Metabolomics* 12:145. doi: 10.1007/s11306-016-1087-5
- Fontaine, F., Bolton, E., Borodina, Y., and Bryant, S. H. (2007). Fast 3D shape screening of large chemical databases through alignment-recycling. *Chem. Cent. J.* 1:12. doi: 10.1186/1752-153X-1-12
- Galúcio, J. M., Monteiro, E. F., de Jesus, D. A., Costa, C. H., Siqueira, R. C., Santos, G. B. dos, et al. (2019). *In silico* identification of natural products with anticancer activity using a chemo-structural database of Brazilian biodiversity. *Comput. Biol. Chem.* 83:107102. doi: 10.1016/j.compbiolchem.2019.107102
- García-Hernández, C., Fernández, A., and Serratos, F. (2019). Ligand-based virtual screening using graph edit distance as molecular similarity measure. *J. Chem. Inf. Model.* 59, 1410–21. doi: 10.1021/acs.jcim.8b00820
- Gfeller, D., Grosdidier, A., Wirth, M., Daina, A., Michielin, O., and Zoete, V. (2014). SwissTargetPrediction: a web server for target prediction of bioactive small molecules. *Nucleic Acids Res.* 42, W32–W38. doi: 10.1093/nar/gku293

- Ghanakota, P., and Carlson, H. A. (2017). Comparing pharmacophore models derived from crystallography and NMR ensembles. *J. Comput. Aided Mol. Des.* 31, 979–993. doi: 10.1007/s10822-017-0077-7
- 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, 55–68. doi: 10.1021/cc9800071
- Gimeno, A., Ojeda-Montes, M., Tomás-Hernández, S., Cereto-Massagué, A., Beltrán-Debón, R., Mulero, M., et al. (2019). The light and dark sides of virtual screening: what is there to know? *Int. J. Mol. Sci.* 20:1375. doi: 10.3390/ijms20061375
- Gohlke, B. O., Overkamp, T., Richter, A., Richter, A., Daniel, P. T., Gillissen, B., et al. (2015). 2D and 3D similarity landscape analysis identifies PARP as a novel off-target for the drug Vatalanib. *BMC Bioinformatics* 16:308. doi: 10.1186/s12859-015-0730-x
- Gomes, M. N., Braga, R. C., Grzelak, E. M., Neves, B. J., Muratov, E., Ma, R., et al. (2017). QSAR-driven design, synthesis and discovery of potent chalcone derivatives with antitubercular activity. *Eur. J. Med. Chem.* 137, 126–138. doi: 10.1016/j.ejmech.2017.05.026
- Gonczarek, A., Tomczak, J. M., Zareba, S., Kaczmar, J., Dabrowski, P., and Walczak, M. J. (2018). Interaction prediction in structure-based virtual screening using deep learning. *Comput. Biol. Med.* 100, 253–258. doi: 10.1016/j.combiomed.2017.09.007
- Gorgulla, C., Boeszormentyi, A., Wang, Z. F., Fischer, P. D., Coote, P. W., Padmanabha Das, K. M., et al. (2020). An open-source drug discovery platform enables ultra-large virtual screens. *Nature* 580, 663–668. doi: 10.1038/s41586-020-2117-z
- Gosse, J. T., Ghosh, S., Sproule, A., Overy, D., Cheeptham, N., and Boddy, C. N. (2019). Whole genome sequencing and metabolomic study of cave *Streptomyces* isolates ICC1 and ICC4. *Front. Microbiol.* 10:1020. doi: 10.3389/fmicb.2019.01020
- Grant, J. A., Gallardo, M. A., and Pickup, B. T. (1996). A fast method of molecular shape comparison: a simple application of a Gaussian description of molecular shape. *J. Comput. Chem.* 17, 1653–1666. doi: 10.1002/(SICI)1096-987X(19961115)17:14<1653::AID-JCC7>3.0.CO;2-K
- Grisoni, F., Merk, D., Friedrich, L., and Schneider, G. (2019). Design of natural-product-inspired multitarget ligands by machine learning. *ChemMedChem* 14, 1129–1134. doi: 10.1002/cmdc.201900097
- Gu, J., Gui, Y., Chen, L., Yuan, G., Lu, H. Z. Z., and Xu, X. (2013). Use of natural products as chemical library for drug discovery and network pharmacology. *PLoS ONE* 8:e62839. doi: 10.1371/journal.pone.0062839
- Guedes, I. A., Pereira, F. S. S., and Dardenne, L. E. (2018). Empirical scoring functions for structure-based virtual screening: applications, critical aspects, and challenges. *Front. Pharmacol.* 9:1089. doi: 10.3389/fphar.2018.01089
- Gupta, S., Kapoor, P., Chaudhary, K., Gautam, A., Kumar, R., and Raghava, G. P. S. (2013). *In silico* approach for predicting toxicity of peptides and proteins. *PLoS ONE* 8:e73957. doi: 10.1371/journal.pone.0073957
- Hamza, A., Wei, N. N., and Zhan, C. G. (2012). Ligand-based virtual screening approach using a new scoring function. *J. Chem. Inf. Model.* 52, 963–974. doi: 10.1021/ci200617d
- Hao, G., Dong, Q., and Yang, G. (2011). A comparative study on the constitutive properties of marketed pesticides. *Mol. Inform.* 30, 614–622. doi: 10.1002/minf.201100020
- Harvey, A. L., Edrada-Ebel, R., and Quinn, R. J. (2015). The re-emergence of natural products for drug discovery in the genomics era. *Nat. Rev. Drug Discov.* 14, 111–129. doi: 10.1038/nrd4510
- Henninot, A., Collins, J. C., and Nuss, J. M. (2018). The current state of peptide drug discovery: back to the future? *J. Med. Chem.* 61, 1382–1414. doi: 10.1021/acs.jmedchem.7b00318
- Hinselmann, G., Rosenbaum, L., Jahn, A., Fechner, N., and Zell, A. (2011). jCompoundMapper: an open source Java library and command-line tool for chemical fingerprints. *J. Cheminform.* 3:3. doi: 10.1186/1758-2946-3-3
- Hossin, M., and Sulaiman, M. N. (2015). A review on evaluation metrics for data classification evaluations. *Int. J. Data Min. Knowl. Manag. Process.* 5, 1–11. doi: 10.5121/ijdkp.2015.5201
- Huang, C., Yang, Y., Chen, X., Wang, C., Li, Y., Zheng, C., et al. (2017). Large-scale cross-species chemogenomic platform proposes a new drug discovery strategy of veterinary drug from herbal medicines. *PLoS ONE* 12:e0184880. doi: 10.1371/journal.pone.0184880
- Huang, Y., You, Z., and Chen, X. (2018). A systematic prediction of drug-target interactions using molecular fingerprints and protein sequences. *Curr. Protein Pept. Sci.* 19, 468–478. doi: 10.2174/1389203718666161122103057
- Huffman, B. J., and Shenvi, R. A. (2019). Natural products in the “Marketplace”: interfacing synthesis and biology. *J. Am. Chem. Soc.* 141, 3332–3346. doi: 10.1021/jacs.8b11297
- Huggins, D. J., Venkitaraman, A. R., and Spring, D. R. (2011). Rational methods for the selection of diverse screening compounds. *ACS Chem. Biol.* 6, 208–217. doi: 10.1021/cb100420r
- Instant JChem 21.4.0 (2021). *ChemAxon*. Available online at: <http://www.chemaxon.com>
- Jade, D. D., Pandey, R., Kumar, R., and Gupta, D. (2020). Ligand-based pharmacophore modeling of TNF- α to design novel inhibitors using virtual screening and molecular dynamics. *J. Biomol. Struct. Dyn.* doi: 10.1080/07391102.2020.1831962. [Epub ahead of print].
- Jagannathan, R. (2019). Characterization of drug-like chemical space for cytotoxic marine metabolites using multivariate methods. *ACS Omega* 4, 5402–5411. doi: 10.1021/acsomega.8b01764
- Jasial, S., Gilberg, E., Blaschke, T., and Bajorath, J. (2018). Machine learning distinguishes with high accuracy between pan-assay interference compounds that are promiscuous or represent dark chemical matter. *J. Med. Chem.* 61, 10255–10264. doi: 10.1021/acs.jmedchem.8b01404
- Jayaseelan, K. V., Moreno, P., Truszkowski, A., Ertl, P., and Steinbeck, C. (2012). Natural product-likeness score revisited: an open-source, open-data implementation. *BMC Bioinformatics* 13:106. doi: 10.1186/1471-2105-13-106
- Jeffrey, P., and Summerfield, S. (2010). Assessment of the blood-brain barrier in CNS drug discovery. *Neurobiol. Dis.* 37, 33–37. doi: 10.1016/j.nbd.2009.07.033
- Jeon, J., Nim, S., Teyra, J., Datti, A., Wrana, J. L., Sidhu, S. S., et al. (2014). A systematic approach to identify novel cancer drug targets using machine learning, inhibitor design and high-throughput screening. *Genome Med.* 6:57. doi: 10.1186/s13073-014-0057-7
- Jhoti, H., Williams, G., Rees, D. C., and Murray, C. W. (2013). The “rule of three” for fragment-based drug discovery: where are we now? *Nat. Rev. Drug Discov.* 12:644. doi: 10.1038/nrd3926-c1
- Jia, C. Y., Li, J. Y., Hao, G. F., and Yang, G. F. (2020). A drug-likeness toolbox facilitates ADMET study in drug discovery. *Drug Discov. Today* 25, 248–258. doi: 10.1016/j.drudis.2019.10.014
- Jiang, S., Feher, M., Williams, C., Cole, B., and Shaw, D. E. (2020). AutoPH4: an automated method for generating pharmacophore models from protein binding pockets. *J. Chem. Inf. Model.* 60, 4326–4338. doi: 10.1021/acs.jcim.0c00121
- Jiménez-Luna, J., Grisoni, F., and Schneider, G. (2020). Drug discovery with explainable artificial intelligence. *Nat. Mach. Intell.* 2, 573–584. doi: 10.1038/s42256-020-00236-4
- Jin, Z., Wang, Y., Yu, X. F., Tan, Q. Q., Liang, S. S., Li, T., et al. (2020). Structure-based virtual screening of influenza virus RNA polymerase inhibitors from natural compounds: molecular dynamics simulation and MM-GBSA calculation. *Comput. Biol. Chem.* 85:107241. doi: 10.1016/j.compbiolchem.2020.107241
- Jing, Y., Bian, Y., Hu, Z., Wang, L., and Xie, X. Q. S. (2018). Deep learning for drug design: an artificial intelligence paradigm for drug discovery in the big data era. *AAPS J.* 20:58. doi: 10.1208/s12248-018-0210-0
- Jones, L. H., and Bunnage, M. E. (2017). Applications of chemogenomic library screening in drug discovery. *Nat. Rev. Drug Discov.* 16, 285–296. doi: 10.1038/nrd.2016.244
- Kar, S., and Roy, K. (2013). How far can virtual screening take us in drug discovery? *Expert Opin. Drug Discov.* 8, 245–261. doi: 10.1517/17460441.2013.761204
- Karaboga, A. S., Petronin, F., Marchetti, G., Souchet, M., and Maigret, B. (2013). Benchmarking of HPCC: a novel 3D molecular representation combining shape and pharmacophoric descriptors for efficient molecular similarity assessments. *J. Mol. Graph. Model.* 41, 20–30. doi: 10.1016/j.jmkgm.2013.01.003
- Kauffman, G. W., and Jurs, P. C. (2001). QSAR and k-nearest neighbor classification analysis of selective cyclooxygenase-2 inhibitors using topologically-based numerical descriptors. *J. Chem. Inf. Comput. Sci.* 41, 1553–1560. doi: 10.1021/ci010073h

- Kim, S., Bolton, E. E., and Bryant, S. H. (2016). Similar compounds versus similar conformers: complementarity between PubChem 2-D and 3-D neighboring sets. *J. Cheminform.* 8:62. doi: 10.1186/s13321-016-0163-1
- Kleigrew, K., Almaliti, J., Tian, I. Y., Kinnel, R. B., Korobeynikov, A., Monroe, E. A., et al. (2015). Combining mass spectrometric metabolic profiling with genomic analysis: a powerful approach for discovering natural products from cyanobacteria. *J. Nat. Prod.* 78, 1671–1682. doi: 10.1021/acs.jnatprod.5b00301
- Klupczynska, A., Plewa, S., Dereziński, P., Garrett, T. J., Rubio, V. Y., Kokot, Z. J., et al. (2020). Identification and quantification of honeybee venom constituents by multiplatform metabolomics. *Sci. Rep.* 10:21645. doi: 10.1038/s41598-020-78740-1
- Koes, D. R., and Camacho, C. J. (2011). Pharmer: efficient and exact pharmacophore search. *J. Chem. Inf. Model.* 51, 1307–1314. doi: 10.1021/ci200097m
- Koes, D. R., and Camacho, C. J. (2014). Shape-based virtual screening with volumetric aligned molecular shapes. *J. Comput. Chem.* 35, 1824–1834. doi: 10.1002/jcc.23690
- Kong, W., Wang, W., and An, J. (2020). Prediction of 5-hydroxytryptamine transporter inhibitors based on machine learning. *Comput. Biol. Chem.* 87:107303. doi: 10.1016/j.compbiolchem.2020.107303
- Kortagere, S., Krasowski, M. D., and Ekins, S. (2009). The importance of discerning shape in molecular pharmacology. *Trends Pharmacol. Sci.* 30, 138–147. doi: 10.1016/j.tips.2008.12.001
- Kumar, A., and Zhang, K. Y. J. (2015). Hierarchical virtual screening approaches in small molecule drug discovery. *Methods* 71, 26–37. doi: 10.1016/j.ymeth.2014.07.007
- Kumar, A., and Zhang, K. Y. J. (2018). Advances in the development of shape similarity methods and their application in drug discovery. *Front. Chem.* 6:315. doi: 10.3389/fchem.2018.00315
- Kumar, D., Sharma, P., Singh, H., Nepali, K., Gupta, G. K., Jain, S. K., et al. (2017). The value of pyrans as anticancer scaffolds in medicinal chemistry. *RSC Adv.* 7, 36977–36999. doi: 10.1039/C7RA05441F
- Kunimoto, R., and Bajorath, J. (2018). Combining similarity searching and network analysis for the identification of active compounds. *ACS Omega* 3, 3768–3777. doi: 10.1021/acsomega.8b00344
- Lagorce, D., Sperandio, O., Baell, J. B., Miteva, M. A., and Villoutreix, B. O. (2015). FAF-Drugs3: a web server for compound property calculation and chemical library design. *Nucleic Acids Res.* 43, W200–W207. doi: 10.1093/nar/gkv353
- Lans, I., Palacio-Rodríguez, K., Cavasotto, C. N., and Cossio, P. (2020). Flexi-pharma: a molecule-ranking strategy for virtual screening using pharmacophores from ligand-free conformational ensembles. *J. Comput. Aided Mol. Des.* 34, 1063–1077. doi: 10.1007/s10822-020-00329-7
- Lata, S., Sharma, B., and Raghava, G. (2007). Analysis and prediction of antibacterial peptides. *BMC Bioinformatics* 8:263. doi: 10.1186/1471-2105-8-263
- Lavecchia, A. (2015). Machine-learning approaches in drug discovery: methods and applications. *Drug Discov. Today* 20, 318–331. doi: 10.1016/j.drudis.2014.10.012
- Lee, J., Park, J., Kim, J., Jeong, B., Choi, S. Y., Jang, H. S., et al. (2020). Targeted isolation of cytotoxic sesquiterpene lactones from *Eupatorium fortunei* by the NMR annotation tool, SMART 2.0. *ACS Omega* 5, 23989–23995. doi: 10.1021/acsomega.0c03270
- Lee, M. L., and Schneider, G. (2001). Scaffold architecture and pharmacophoric properties of natural products and trade drugs: application in the design of natural product-based combinatorial libraries. *J. Comb. Chem.* 3, 284–289. doi: 10.1021/cc000097l
- Li, F., Wang, Y., Li, D., Chen, Y., and Dou, Q. P. (2019). Are we seeing a resurgence in the use of natural products for new drug discovery? *Expert Opin. Drug Discov.* 14, 417–420. doi: 10.1080/17460441.2019.1582639
- Li, G. H., and Huang, J. F. (2012). CDRUG: A web server for predicting anticancer activity of chemical compounds. *Bioinformatics* 28, 3334–3335. doi: 10.1093/bioinformatics/bts625
- Li, H., Leung, K. S., Wong, M. H., and Ballester, P. J. (2016). USR-VS: a web server for large-scale prospective virtual screening using ultrafast shape recognition techniques. *Nucleic Acids Res.* 44, W436–W441. doi: 10.1093/nar/gkw320
- Li, H., Sze, K. H., Lu, G., and Ballester, P. J. (2020). Machine-learning scoring functions for structure-based virtual screening. *Wiley Interdiscipl. Rev. Comput. Mol. Sci.* 11:e1478. doi: 10.1002/wcms.1478
- Li, J. W. H., and Vederas, J. C. (2009). Drug discovery and natural products: end of an era or an endless frontier? *Science* 325, 161–165. doi: 10.1126/science.1168243
- Lima, M. N. N., Borba, J. V. B., Cassiano, G. C., Mottin, M., Mendonça, S. S., Silva, A. C., et al. (2020). Artificial intelligence applied for the rapid identification of new antimalarial candidates with dual-stage activity. *ChemMedChem* 16, 1093–1103. doi: 10.1002/cmdc.202000685
- Lipinski, C. A., Lombardo, F., Dominy, B. W., Feeney, P. J., Lombardo, F., Dominy, B. W., et al. (1997). Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Deliv. Rev.* 23, 3–25. doi: 10.1016/S0169-409X(96)00423-1
- Liu, X., Jiang, H., and Li, H. (2011). SHAFTS: a hybrid approach for 3D molecular similarity calculation. 1. Method and assessment of virtual screening. *J. Chem. Inf. Model.* 51, 2372–2385. doi: 10.1021/ci200060s
- Liu, X., Ouyang, S., Yu, B., Liu, Y., Huang, K., Gong, J., et al. (2010). PharmMapper server: a web server for potential drug target identification using pharmacophore mapping approach. *Nucleic Acids Res.* 38, W609–W614. doi: 10.1093/nar/gkq300
- Liu, Z., Du, J., Fang, J., Yin, Y., Xu, G., and Xie, L. (2019). DeepScreening: a deep learning-based screening web server for accelerating drug discovery. *Database* 2019:baz104. doi: 10.1093/database/baz104
- Liu, Z., Huang, D., Zheng, S., Song, Y., Liu, B., Sun, J., et al. (2020). Deep learning enables discovery of highly potent anti-osteoporosis natural products. *Eur. J. Med. Chem.* 210:112982. doi: 10.1016/j.ejmech.2020.112982
- Lo, Y. C., Rensi, S. E., Torng, W., and Altman, R. B. (2018). Machine learning in chemoinformatics and drug discovery. *Drug Discov. Today* 23, 1538–1546. doi: 10.1016/j.drudis.2018.05.010
- Lovering, F. (2013). Escape from Flatland 2: complexity and promiscuity. *MedChemComm* 4:515. doi: 10.1039/c2md20347b
- Lovrić, M., Molero, J. M., and Kern, R. (2019). PySpark and RDKit: moving towards big data in cheminformatics. *Mol. Inform.* 38:1800082. doi: 10.1002/minf.201800082
- Lucas, X., Senger, C., Erxleben, A., Grüning, B. A., Döring, K., Mosch, J., et al. (2013). StreptomeDB: a resource for natural compounds isolated from *Streptomyces* species. *Nucleic Acids Res.* 41, D1130–D1136. doi: 10.1093/nar/gks1253
- Luo, Q., Zhao, L., Hu, J., Jin, H., Liu, Z., and Zhang, L. (2017). The scoring bias in reverse docking and the score normalization strategy to improve success rate of target fishing. *PLoS ONE* 12:e0171433. doi: 10.1371/journal.pone.0171433
- Lyu, C., Chen, T., Qiang, B., Liu, N., Wang, H., Zhang, L., et al. (2020). CMNPD: a comprehensive marine natural products database towards facilitating drug discovery from the ocean. *Nucleic Acids Res.* 49, D509–D515. doi: 10.1093/nar/gkaa763
- Macalino, S. J. Y., Gosu, V., Hong, S., and Choi, S. (2015). Role of computer-aided drug design in modern drug discovery. *Arch. Pharm. Res.* 38, 1686–1701. doi: 10.1007/s12272-015-0640-5
- Madzhidov, T. I., Rakhimbekova, A., Kutlushuna, A., and Polishchuk, P. (2020). Probabilistic approach for virtual screening based on multiple pharmacophores. *Molecules* 25:385. doi: 10.3390/molecules25020385
- Maggiora, G. M., and Bajorath, J. (2014). Chemical space networks: a powerful new paradigm for the description of chemical space. *J. Comput. Aided Mol. Des.* 28, 795–802. doi: 10.1007/s10822-014-9760-0
- Maia, E. H. B., Assis, L. C., de Oliveira, T. A., da Silva, A. M., and Taranto, A. G. (2020). Structure-based virtual screening: from classical to artificial intelligence. *Front. Chem.* 8:343. doi: 10.3389/fchem.2020.00343
- Maia, S. M., de Sousa, N. F., Rodrigues, G. C. S., Monteiro, A. F. M., Scotti, M. T., and Scotti, L. (2020). Lignans and neolignans anti-tuberculosis identified by QSAR and molecular modeling. *Comb. Chem. High Throughput Screen.* 23, 504–516. doi: 10.2174/1386207323666200226094940
- Manavalan, B., Subramaniam, S., Shin, T. H., Kim, M. O., and Lee, G. (2018). Machine-learning-based prediction of cell-penetrating peptides and their uptake efficiency with improved accuracy. *J. Proteome Res.* 17, 2715–2726. doi: 10.1021/acs.jproteome.8b00148
- Martínez-Treviño, S. H., Uc-Cetina, V., Fernández-Herrera, M. A., and Merino, G. (2020). Prediction of natural product classes using machine learning and ¹³C NMR spectroscopic data. *J. Chem. Inf. Model.* 60, 3376–3386. doi: 10.1021/acs.jcim.0c00293

- Mascarenhas, A. M. S., de Almeida, R. B. M., de Araujo Neto, M. F., Mendes, G. O., da Cruz, J. N., dos Santos, C. B. R., et al. (2020). Pharmacophore-based virtual screening and molecular docking to identify promising dual inhibitors of human acetylcholinesterase and butyrylcholinesterase. *J. Biomol. Struct. Dyn.* doi: 10.1080/07391102.2020.1796791. [Epub ahead of print].
- Mayr, A., Klambauer, G., Unterthiner, T., and Hochreiter, S. (2016). DeepTox: toxicity prediction using deep learning. *Front. Environ. Sci.* 3:80. doi: 10.3389/fenvs.2015.00080
- Medina-Franco, J. L., Golbraikh, A., Oloff, S., Castillo, R., and Tropsha, A. (2005). Quantitative structure-activity relationship analysis of pyridinone HIV-1 reverse transcriptase inhibitors using the k nearest neighbor method and QSAR-based database mining. *J. Comput. Aided Mol. Des.* 19, 229–242. doi: 10.1007/s10822-005-4789-8
- Medina-Franco, J. L., and Saldívar-González, F. I. (2020). Cheminformatics to characterize pharmacologically active natural products. *Biomolecules* 10:1566. doi: 10.3390/biom10111566
- Merwin, N. J., Mousa, W. K., Dejong, C. A., Skinnider, M. A., Cannon, M. J., Li, H., et al. (2020). DeepRiPP integrates multiomics data to automate discovery of novel ribosomally synthesized natural products. *Proc. Natl. Acad. Sci. U.S.A.* 117, 371–380. doi: 10.1073/pnas.1901493116
- Mignani, S., Rodrigues, J., Tomas, H., Jalal, R., Singh, P. P., Majoral, J. P., et al. (2018). Present drug-likeness filters in medicinal chemistry during the hit and lead optimization process: how far can they be simplified? *Drug Discov. Today* 23, 605–615. doi: 10.1016/j.drudis.2018.01.010
- Miteva, M. A., Violas, S., Montes, M., Gomez, D., Tuffery, P., and Villoutreix, B. O. (2006). FAF-Drugs: free ADME/tox filtering of compound collections. *Nucleic Acids Res.* 34, W738–W744. doi: 10.1093/nar/gkl065
- Molecular Operating Environment (2019). *Chemical Computing Group ULC*. Montreal, QC: Molecular Operating Environment, 202.
- Morais, T. R., Conserva, G. A. A., Varella, M. T., Costa-Silva, T. A., Thevenard, F., Ponci, V., et al. (2020). Improving the drug-likeness of inspiring natural products - evaluation of the antiparasitic activity against *Trypanosoma cruzi* through semi-synthetic and simplified analogues of licarín A. *Sci. Rep.* 10:5467. doi: 10.1038/s41598-020-62352-w
- Moumbock, A. F. A., Gao, M., Qaseem, A., Li, J., Kirchner, P. A., Ndingokhar, B., et al. (2021). StreptomeDB 3.0: an updated compendium of streptomycetes natural products. *Nucleic Acids Res.* 49, D600–D604. doi: 10.1093/nar/gkaa868
- Muegge, I., and Mukherjee, P. (2016). An overview of molecular fingerprint similarity search in virtual screening. *Expert Opin. Drug Discov.* 11, 137–148. doi: 10.1517/17460441.2016.1117070
- Naghizadeh, A., Hamzeheian, D., Akbari, S., Mohammadi, F., Otoufat, T., Asgari, S., et al. (2020). UNaProd: a universal natural product database for materia medica of Iranian traditional medicine. *Evid. Based Complement. Altern. Med.* 2020, 1–14. doi: 10.1155/2020/3690781
- Naylor, M. R., Bockus, A. T., Blanco, M. J., and Lokey, R. S. (2017). Cyclic peptide natural products chart the frontier of oral bioavailability in the pursuit of undruggable targets. *Curr. Opin. Chem. Biol.* 38, 141–147. doi: 10.1016/j.cbpa.2017.04.012
- 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
- Newman, D. J., and Cragg, G. M. (2020). Natural products as sources of new drugs over the nearly four decades from 01/1981 to 09/2019. *J. Nat. Prod.* 83, 770–803. doi: 10.1021/acs.jnatprod.9b01285
- Nocedo-Mena, D., Cornelio, C., Camacho-Corona M del R, Garza-González, E., Waksman de Torres, N., Arrasate, S., et al. (2019). Modeling antibacterial activity with machine learning and fusion of chemical structure information with microorganism metabolic networks. *J. Chem. Inf. Model.* 59, 1109–1120. doi: 10.1021/acs.jcim.9b00034
- Ntie-Kang, F., Telukunta, K. K., Döring, K., Simoben, C. V., Moumbock, A. F. A., Malange, Y. I., et al. (2017). NANPDB: a resource for natural products from northern African Sources. *J. Nat. Prod.* 80, 2067–2076. doi: 10.1021/acs.jnatprod.7b00283
- O'Boyle, N. M., Morley, C., and Hutchison, G. R. (2008). Pybel: a Python wrapper for the OpenBabel cheminformatics toolkit. *Chem. Cent. J.* 2:5. doi: 10.1186/1752-153X-2-5
- O'Hagan, S., and Kell, D. B. (2016). MetMaxStruct: a Tversky-similarity-based strategy for analysing the (sub)structural similarities of drugs and endogenous metabolites. *Front. Pharmacol.* 7:266. doi: 10.3389/fphar.2016.00266
- Olivon, F., Allard, P. M., Koval, A., Righi, D., Genta-Jouve, G., Neyts, J., et al. (2017). Bioactive natural products prioritization using massive multi-informational molecular networks. *ACS Chem. Biol.* 12, 2644–2651. doi: 10.1021/acschembio.7b00413
- Opassi, G., Gesù, A., and Massarotti, A. (2018). The hitchhiker's guide to the chemical-biological galaxy. *Drug Discov. Today* 23, 565–574. doi: 10.1016/j.drudis.2018.01.007
- Oppong-Danquah, E., Parrot, D., Blümel, M., Labes, A., and Tasdemir, D. (2018). Molecular networking-based metabolome and bioactivity analyses of marine-adapted fungi co-cultivated with phytopathogens. *Front. Microbiol.* 9:2072. doi: 10.3389/fmicb.2018.02072
- Oprea, T. I. (2000). Property distribution of drug-related chemical databases. *J. Comput. Aided Mol. Des.* 14, 251–264. doi: 10.1023/A:1008130001697
- Over, B., Wetzel, S., Grütter, C., Nakai, Y., Renner, S., Rauh, D., et al. (2013). Natural-product-derived fragments for fragment-based ligand discovery. *Nat. Chem.* 5, 21–28. doi: 10.1038/nchem.1506
- Pal, S., Kumar, V., Kundu, B., Bhattacharya, D., Preethy, N., Reddy, M. P., et al. (2019). Ligand-based pharmacophore modeling, virtual screening and molecular docking studies for discovery of potential topoisomerase I inhibitors. *Comput. Struct. Biotechnol. J.* 17, 291–310. doi: 10.1016/j.csbj.2019.02.006
- Patel, L., Shukla, T., Huang, X., Ussery, D. W., and Wang, S. (2020). Machine learning methods in drug discovery. *Molecules* 25:5277. doi: 10.3390/molecules25225277
- Patsilinos, A., Artini, M., Papa, R., Sabatino, M., Božović M, Garzoli, S., et al. (2019). Machine learning analyses on data including essential oil chemical composition and *in vitro* experimental antibiofilm activities against *Staphylococcus* species. *Molecules* 24:890. doi: 10.3390/molecules24050890
- Pavadai, E., Kaur, G., Wittlin, S., and Chibale, K. (2017). Identification of steroid-like natural products as antiplasmodial agents by 2D and 3D similarity-based virtual screening. *MedChemComm* 8, 1152–1157. doi: 10.1039/C7MD00063D
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2011). Scikit-learn: machine learning in Python. *J. Mach. Learn. Res.* 12, 2825–2830.
- Pereira, F., Latino, D., and Gaudêncio, S. (2015). QSAR-assisted virtual screening of lead-like molecules from marine and microbial natural sources for antitumor and antibiotic drug discovery. *Molecules* 20, 4848–4873. doi: 10.3390/molecules20034848
- Pilon, A. C., Valli, M., Dametto, A. C., Pinto, M. E. F., Freire, R. T., Castro-Gamboa, I., et al. (2017). NuBBE DB: an updated database to uncover chemical and biological information from Brazilian biodiversity. *Sci. Rep.* 7:7215. doi: 10.1038/s41598-017-07451-x
- Pilón-Jiménez, B., Saldívar-González, F., Díaz-Eufracio, B., and Medina-Franco, J. (2019). BIOFACQUIM: a Mexican compound database of natural products. *Biomolecules* 9:31. doi: 10.3390/biom9010031
- Pire, D. E. V., Blundell, T. L., and Ascher, D. B. (2015). pkCSM: predicting small-molecule pharmacokinetic and toxicity properties using graph-based signatures. *J. Med. Chem.* 58, 4066–4072. doi: 10.1021/acs.jmedchem.5b00104
- Pu, L., Naderi, M., Liu, T., Wu, H. C., Mukhopadhyay, S., and Brylinski, M. (2019). EToxPred: a machine learning-based approach to estimate the toxicity of drug candidates 11 Medical and Health Sciences 1115 Pharmacology and Pharmaceutical Sciences 03 Chemical Sciences 0305 Organic Chemistry 03 Chemical Sciences 0304 Medicinal and Biomolecular Chemistry. *BMC Pharmacol. Toxicol.* 20:2. doi: 10.1186/s40360-018-0282-6
- Puertas-Martín, S., Redondo, J. L., Ortigosa, P. M., and Pérez-Sánchez, H. (2019). OptiPharm: an evolutionary algorithm to compare shape similarity. *Sci. Rep.* 9:1398. doi: 10.1038/s41598-018-37908-6
- Qiang, X., Zhou, C., Ye, X., Du, P., Su, R., and Wei, L. (2018). CPPred-FL: a sequence-based predictor for large-scale identification of cell-penetrating peptides by feature representation learning. *Brief. Bioinform.* doi: 10.1093/bib/bby091. [Epub ahead of print].
- Qing, X., Lee, X. Y., De Raeymaekers, J., Tame, J. R., Zhang, K. Y., De Maeyer, M., et al. (2014). Pharmacophore modeling: advances, limitations, and current utility in drug discovery. *J. Receptor Ligand Channel Res.* 7, 81–92. doi: 10.2147/JRLCR.S46843
- Rácz, A., Bajusz, D., and Héberger, K. (2018). Life beyond the Tanimoto coefficient: similarity measures for interaction fingerprints. *J. Cheminform.* 10:48. doi: 10.1186/s13321-018-0302-y

- RÁCZ, A., Bajusz, D., and Héberger, K. (2019). Multi-level comparison of machine learning classifiers and their performance metrics. *Molecules* 24:2811. doi: 10.3390/molecules24152811
- Rampogu, S., Son, M., Baek, A., Park, C., Rana, R. M., Zeb, A., et al. (2018). Targeting natural compounds against HER2 kinase domain as potential anticancer drugs applying pharmacophore based molecular modelling approaches. *Comput. Biol. Chem.* 74, 327–338. doi: 10.1016/j.compbiolchem.2018.04.002
- Rayan, A., Raiyn, J., and Falah, M. (2017). Nature is the best source of anticancer drugs: indexing natural products for their anticancer bioactivity. *PLoS ONE* 12:e0187925. doi: 10.1371/journal.pone.0187925
- Reher, R., Kim, H. W., Zhang, C., Mao, H. H., Wang, M., Nothias, L. F., et al. (2020). A convolutional neural network-based approach for the rapid annotation of molecularly diverse natural products. *J. Am. Chem. Soc.* 142, 4114–4120. doi: 10.1021/jacs.9b13786
- Riniker, S., and Landrum, G. A. (2013). Similarity maps—a visualization strategy for molecular fingerprints and machine-learning methods. *J. Cheminform.* 5:43. doi: 10.1186/1758-2946-5-43
- Ripphausen, P., Wassermann, A. M., and Bajorath, J. (2011). REPROVIS-DB: a benchmark system for ligand-based virtual screening derived from reproducible prospective applications. *J. Chem. Inf. Model.* 51, 2467–2473. doi: 10.1021/ci200309j
- Rodrigues, T. (2017). Harnessing the potential of natural products in drug discovery from a cheminformatics vantage point. *Org. Biomol. Chem.* 15, 9275–9282. doi: 10.1039/C7OB02193C
- Rodrigues, T., Reker, D., Schneider, P., and Schneider, G. (2016). Counting on natural products for drug design. *Nat. Chem.* 8, 531–541. doi: 10.1038/nchem.2479
- Ropp, P. J., Spiegel, J. O., Walker, J. L., Green, H., Morales, G. A., Milliken, K. A., et al. (2019). Gypsum-DL: an open-source program for preparing small-molecule libraries for structure-based virtual screening. *J. Cheminform.* 11:34. doi: 10.1186/s13321-019-0358-3
- Rossi Sebastiano, M., Doak, B. C., Backlund, M., Poongavanam, V., Over, B., Ermondi, G., et al. (2018). Impact of dynamically exposed polarity on permeability and solubility of chameleonic drugs beyond the rule of 5. *J. Med. Chem.* 61, 4189–4202. doi: 10.1021/acs.jmedchem.8b00347
- Roumpeka, D. D., Wallace, R. J., Escalettes, F., Fotheringham, I., and Watson, M. (2017). A review of bioinformatics tools for bio-prospecting from metagenomic sequence data. *Front. Genet.* 8:23. doi: 10.3389/fgene.2017.00023
- Rupp, M., Schroeter, T., Steri, R., Zettl, H., Proschak, E., Hansen, K., et al. (2010). From machine learning to natural product derivatives that selectively activate transcription factor PPAR γ . *ChemMedChem* 5, 191–194. doi: 10.1002/cmdc.200900469
- Saldívar-González, F. I., Huerta-García, C. S., and Medina-Franco, J. L. (2020). Chemoinformatics-based enumeration of chemical libraries: a tutorial. *J. Cheminform.* 12:64. doi: 10.1186/s13321-020-00466-z
- Sánchez-Cruz, N., and Medina-Franco, J. L. (2018). Statistical-based database fingerprint: Chemical space representation of compound databases. *J. Cheminform.* 10:55. doi: 10.1186/s13321-018-0311-x
- Sander, T., Freyss, J., von Korff, M., Rufener, C., von Korff, M., and Rufener, C. (2015). DataWarrior: an open-source program for chemistry aware data visualization and analysis. *J. Chem. Inf. Model.* 55, 460–473. doi: 10.1021/ci500588j
- Sanders, M. P. A., Barbosa, A. J. M., Zarzycka, B., Nicolaes, G. A. F., Klomp, J. P. G., de Vlieg, J., et al. (2012). Comparative analysis of pharmacophore screening tools. *J. Chem. Inf. Model.* 52, 1607–1620. doi: 10.1021/ci2005274
- Santacruz, L., Hurtado, D. X., Doohan, R., Thomas, O. P., Puyana, M., and Tello, E. (2020). Metabolomic study of soft corals from the Colombian Caribbean: PSYCHE and 1H-NMR comparative analysis. *Sci. Rep.* 10:5417. doi: 10.1038/s41598-020-62413-0
- Santana, I. B., Leite, F. H. A., and Santos Junior, M. C. (2018). Identification of lutzomyia longipalpis odorant binding protein modulators by comparative modeling, hierarchical virtual screening, and molecular dynamics. *J. Chem.* 2018, 1–10. doi: 10.1155/2018/4173479
- Sato, T., Honma, T., and Yokoyama, S. (2010). Combining machine learning and pharmacophore-based interaction fingerprint for *in silico* screening. *J. Chem. Inf. Model.* 50, 170–185. doi: 10.1021/ci900382e
- Schaduangrat, N., Nantasenamat, C., Prachayasittikul, V., and Shoombuatong, W. (2019). ACPred: a computational tool for the prediction and analysis of anticancer peptides. *Molecules* 24:1973. doi: 10.3390/molecules24101973
- Schaller, D., Šribar, D., Noonan, T., Deng, L., Nguyen, T. N., Pach, S., et al. (2020). Next generation 3D pharmacophore modeling. *WIREs Comput. Mol. Sci.* 10:e1468. doi: 10.1002/wcms.1468
- Schwartz, J., Awale, M., and Reymond, J. L. (2013). SMIFp (SMILES fingerprint) chemical space for virtual screening and visualization of large databases of organic molecules. *J. Chem. Inf. Model.* 53, 1979–1989. doi: 10.1021/ci400206h
- Seddon, M. P., Cosgrove, D. A., Packer, M. J., and Gillet, V. J. (2019). Alignment-free molecular shape comparison using spectral geometry: the framework. *J. Chem. Inf. Model.* 59, 98–116. doi: 10.1021/acs.jcim.8b00676
- Semighini, E. P., Resende, J. A., De Andrade, P., Morais, P. A. B., Carvalho, I., Taft, C. A., et al. (2011). Using computer-aided drug design and medicinal chemistry strategies in the fight against diabetes. *J. Biomol. Struct. Dyn.* 28, 787–796. doi: 10.1080/07391102.2011.10508606
- Senger, S. (2009). Using tversky similarity searches for core hopping: finding the needles in the haystack. *J. Chem. Inf. Model.* 49, 1514–1524. doi: 10.1021/ci900092y
- Shahin, R., Swellmeen, L., Shaheen, O., Aboalhaja, N., and Habash, M. (2016). Identification of novel inhibitors for Pim-1 kinase using pharmacophore modeling based on a novel method for selecting pharmacophore generation subsets. *J. Comput. Aided Mol. Des.* 30, 39–68. doi: 10.1007/s10822-015-9887-7
- Shang, J., Hu, B., Wang, J., Zhu, F., Kang, Y., Li, D., et al. (2018). Cheminformatic insight into the differences between terrestrial and marine originated natural products. *J. Chem. Inf. Model.* 58, 1182–1193. doi: 10.1021/acs.jcim.8b00125
- Shin, W. H., Zhu, X., Bures, M., and Kihara, D. (2015). Three-dimensional compound comparison methods and their application in drug discovery. *Molecules* 20, 12841–12862. doi: 10.3390/molecules200712841
- Shoombuatong, W., Schaduangrat, N., Pratiwi, R., and Nantasenamat, C. (2019). THPep: a machine learning-based approach for predicting tumor homing peptides. *Comput. Biol. Chem.* 80, 441–451. doi: 10.1016/j.compbiolchem.2019.05.008
- Shultz, M. D. (2019). Two decades under the influence of the rule of five and the changing properties of approved oral drugs. *J. Med. Chem.* 62, 1701–1714. doi: 10.1021/acs.jmedchem.8b00686
- Sieg, J., Flachsenberg, F., and Rarey, M. (2019). In need of bias control: evaluating chemical data for machine learning in structure-based virtual screening. *J. Chem. Inf. Model.* 59, 947–961. doi: 10.1021/acs.jcim.8b00712
- Silva, G. N. S. da, Primon-Barros, M., Macedo, A. J., and Gnoatto, S. C. B. (2019). Triterpene derivatives as relevant scaffold for new antifilm drugs. *Biomolecules* 9:58. doi: 10.3390/biom9020058
- Skininder, M. A., Dejong, C. A., Franczak, B. C., McNicholas, P. D., and Magarvey, N. A. (2017). Comparative analysis of chemical similarity methods for modular natural products with a hypothetical structure enumeration algorithm. *J. Cheminform.* 9:46. doi: 10.1186/s13321-017-0234-y
- Skirycz, A., Kierszniowska, S., Méret, M., Willmitzer, L., and Tzotzos, G. (2016). Medicinal bioprospecting of the Amazon rainforest: a modern Eldorado? *Trends Biotechnol.* 34, 781–790. doi: 10.1016/j.tibtech.2016.03.006
- Slater, O., and Kontoyianni, M. (2019). The compromise of virtual screening and its impact on drug discovery. *Expert Opin. Drug Discov.* 14, 619–637. doi: 10.1080/17460441.2019.1604677
- Smith, P. A., Koehler, M. F. T., Girgis, H. S., Yan, D., Chen, Y., Chen, Y., et al. (2018). Optimized arylomycins are a new class of Gram-negative antibiotics. *Nature* 561, 189–194. doi: 10.1038/s41586-018-0483-6
- Soares Rodrigues, G. C., Maia M dos S, Silva Cavalcanti, A. B., Costa Barros, R. P., Scotti, L., Cespedes-Acuña, C. L., et al. (2021). Computer-assisted discovery of compounds with insecticidal activity against *Musca domestica* and *Mythimna separata*. *Food Chem. Toxicol.* 147:111899. doi: 10.1016/j.fct.2020.111899
- Software O Scientific (2008). *OpenEye Scientific Software*. Available online at: <http://www.eyesopen.com/products/applications/filter.html>
- Sorokina, M., and Steinbeck, C. (2019). NaPLeS: a natural products likeness scorer—web application and database. *J. Cheminform.* 11:55. doi: 10.1186/s13321-019-0378-z
- Spyrakakis, F., Bellio, P., Quotadamo, A., Linciano, P., Benedetti, P., D'Arrigo, G., et al. (2019). First virtual screening and experimental validation of inhibitors

- targeting GES-5 carbapenemase. *J. Comput. Aided Mol. Des.* 33, 295–305. doi: 10.1007/s10822-018-0182-2
- Stokes, J. M., Yang, K., Swanson, K., Jin, W., Cubillos-Ruiz, A., Donghia, N. M., et al. (2020). A deep learning approach to antibiotic discovery. *Cell* 180, 688–702.e13. doi: 10.1016/j.cell.2020.01.021
- Sunseri, J., and Koes, D. R. (2016). Pharmit: interactive exploration of chemical space. *Nucleic Acids Res.* 44, W442–W448. doi: 10.1093/nar/gkw287
- Svetnik, V., Liaw, A., Tong, C., Culberson, J. C., Sheridan, R. P., and Feuston, B. P. (2003). Random forest: a classification and regression tool for compound classification and QSAR modeling. *J. Chem. Inf. Comput. Sci.* 43, 1947–1958. doi: 10.1021/ci034160g
- Tahir, R. A., Hassan, F., Kareem, A., Iftikhar, U., and Sehgal, S. A. (2020). Ligand-based pharmacophore modeling and virtual screening to discover novel CYP1A1 inhibitors. *Curr. Top. Med. Chem.* 19, 2782–2794. doi: 10.2174/1568026619666191112104217
- Taminiau, J., Thijs, G., and De Winter, H. (2008). Pharao: pharmacophore alignment and optimization. *J. Mol. Graph. Model.* 27, 161–169. doi: 10.1016/j.jmgm.2008.04.003
- Tao, L., Zhu, F., Qin, C., Zhang, C., Chen, S., Zhang, P., et al. (2015). Clustered distribution of natural product leads of drugs in the chemical space as influenced by the privileged target-sites. *Sci. Rep.* 5:9325. doi: 10.1038/srep09325
- Thireou, T., Kythreoti, G., Tsitsanou, K. E., Koussis, K., Drakou, C. E., Kinnarsley, J., et al. (2018). Identification of novel bioinspired synthetic mosquito repellents by combined ligand-based screening and OBP-structure-based molecular docking. *Insect Biochem. Mol. Biol.* 98, 48–61. doi: 10.1016/j.ibmb.2018.05.001
- Thomford, N., Senthane, D., Rowe, A., Munro, D., Seele, P., Maroyi, A., et al. (2018). Natural products for drug discovery in the 21st century: innovations for novel drug discovery. *Int. J. Mol. Sci.* 19:1578. doi: 10.3390/ijms19061578
- Tice, C. M. (2001). Selecting the right compounds for screening: does Lipinski's Rule of 5 for pharmaceuticals apply to agrochemicals? *Pest Manag. Sci.* 57, 3–16. doi: 10.1002/1526-4998(200101)57:1<3::AID-PS269>3.0.CO;2-6
- Tomar, V., Mazumder, M., Chandra, R., Yang, J., and Sakharkar, M. K. (2018). "Small molecule drug design," in *Encyclopedia of Bioinformatics and Computational Biology: ABC of Bioinformatics*, Vol. 1–3, eds S. Ranganathan, M. Gribskov, K. Nakai, and C. B. Schönbach (Oxford: Academic Press), 741–760. doi: 10.1016/B978-0-12-809633-8.20157-X
- Trujillo-Correa, A. I., Quintero-Gil, D. C., Diaz-Castillo, F., Quiñones, W., Robledo, S. M., and Martínez-Gutiérrez, M. (2019). *In vitro* and *in silico* anti-dengue activity of compounds obtained from *Psidium guajava* through bioprospecting. *BMC Complement. Altern. Med.* 19:298. doi: 10.1186/s12906-019-2695-1
- Tsou, L. K., Yeh, S. H., Ueng, S. H., Chang, C. P., Song, J. S., Wu, M. H., et al. (2020). Comparative study between deep learning and QSAR classifications for TNBC inhibitors and novel GPCR agonist discovery. *Sci. Rep.* 10:16771. doi: 10.1038/s41598-020-73681-1
- Valli, M., dos Santos, R. N., Figueira, L. D., Nakajima, C. H., Castro-Gamboa, I., Andricopulo, A. D., et al. (2013). Development of a natural products database from the biodiversity of Brazil. *J. Nat. Prod.* 76, 439–444. doi: 10.1021/np3006875
- Vamathevan, J., Clark, D., Czodrowski, P., Dunham, I., Ferran, E., Lee, G., et al. (2019). Applications of machine learning in drug discovery and development. *Nat. Rev. Drug Discov.* 18, 463–477. doi: 10.1038/s41573-019-0024-5
- van Santen, J. A., Jacob, G., Singh, A. L., Aniebok, V., Balunas, M. J., Bunsko, D., et al. (2019). The natural products atlas: an open access knowledge base for microbial natural products discovery. *ACS Cent. Sci.* 5, 1824–1833. doi: 10.1021/acscentsci.9b00806
- Veber, D. F., Johnson, S. R., Cheng, H. Y. 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
- Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., et al. (2020). SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat. Methods* 17, 261–272. doi: 10.1038/s41592-019-0686-2
- Vuorinen, A., and Schuster, D. (2015). Methods for generating and applying pharmacophore models as virtual screening filters and for bioactivity profiling. *Methods* 71, 113–134. doi: 10.1016/j.ymeth.2014.10.013
- Wallach, I., and Heifets, A. (2018). Most ligand-based classification benchmarks reward memorization rather than generalization. *J. Chem. Inf. Model.* 58, 916–932. doi: 10.1021/acs.jcim.7b00403
- Walters, W. P. (2019). Virtual chemical libraries. *J. Med. Chem.* 62, 1116–1124. doi: 10.1021/acs.jmedchem.8b01048
- Walters, W. P., and Murcko, M. A. (2002). Prediction of 'drug-likeness.' *Adv. Drug Deliv. Rev.* 54, 255–271. doi: 10.1016/S0169-409X(02)00003-0
- Walters, W. P., and Namchuk, M. (2003). Designing screens: how to make your hits a hit. *Nat. Rev. Drug Discov.* 2, 259–266. doi: 10.1038/nrd1063
- Wang, L., Ma, C., Wipf, P., Liu, H., Su, W., and Xie, X. Q. (2013). TargetHunter: an *in silico* target identification tool for predicting therapeutic potential of small organic molecules based on chemogenomic database. *AAPS J.* 15, 395–406. doi: 10.1208/s12248-012-9449-z
- Wang, Y., Jafari, M., Tang, Y., and Tang, J. (2019). Predicting Meridian in Chinese traditional medicine using machine learning approaches. *PLoS Comput. Biol.* 15:e1007249. doi: 10.1371/journal.pcbi.1007249
- Wang, Z., Sun, H., Shen, C., Hu, X., Gao, J., Li, D., et al. (2020). Combined strategies in structure-based virtual screening. *Phys. Chem. Chem. Phys.* 22, 3149–3159. doi: 10.1039/C9CP06303J
- Wei, L., Xing, P., Su, R., Shi, G., Ma, Z. S., and Zou, Q. (2017). CPPred-RF: a sequence-based predictor for identifying cell-penetrating peptides and their uptake efficiency. *J. Proteome Res.* 16, 2044–2053. doi: 10.1021/acs.jproteome.7b00019
- Williams, A. J., and Ekins, S. (2011). A quality alert and call for improved curation of public chemistry databases. *Drug Discov. Today* 16, 747–750. doi: 10.1016/j.drudis.2011.07.007
- Williams, A. J., Ekins, S., and Tkachenko, V. (2012). Towards a gold standard: regarding quality in public domain chemistry databases and approaches to improving the situation. *Drug Discov. Today* 17, 685–701. doi: 10.1016/j.drudis.2012.02.013
- Willighagen, E. L., Mayfield, J. W., Alvarsson, J., Berg, A., Carlsson, L., Jeliakova, N., et al. (2017). The chemistry development kit (CDK) v2.0: atom typing, depiction, molecular formulas, and substructure searching. *J. Cheminform.* 9:33. doi: 10.1186/s13321-017-0220-4
- Wingert, B. M., and Camacho, C. J. (2018). Improving small molecule virtual screening strategies for the next generation of therapeutics. *Curr. Opin. Chem. Biol.* 44, 87–92. doi: 10.1016/j.cbpa.2018.06.006
- Wójcikowski, M., Ballester, P. J., and Siedlecki, P. (2017). Performance of machine-learning scoring functions in structure-based virtual screening. *Sci. Rep.* 7:46710. doi: 10.1038/srep46710
- Wolber, G., and Langer, T. (2005). LigandScout: 3-D pharmacophores derived from protein-bound ligands and their use as virtual screening filters. *J. Chem. Inf. Model.* 45, 160–169. doi: 10.1021/ci049885e
- Wolfe, J. M., Fadzen, C. M., Choo, Z. N., Holden, R. L., Yao, M., Hanson, G. J., et al. (2018). Machine learning to predict cell-penetrating peptides for antisense delivery. *ACS Cent. Sci.* 4, 512–520. doi: 10.1021/acscentsci.8b00098
- Wolfender, J. L., Litaudon, M., Touboul, D., and Queiroz, E. F. (2019). Innovative omics-based approaches for prioritisation and targeted isolation of natural products—new strategies for drug discovery. *Nat. Prod. Rep.* 36, 855–868. doi: 10.1039/C9NP00004F
- Wood, D. J., Vlieg J de, Wagener, M., and Ritschel, T. (2012). Pharmacophore fingerprint-based approach to binding site subpocket similarity and its application to bioisostere replacement. *J. Chem. Inf. Model.* 52, 2031–2043. doi: 10.1021/ci3000776
- Yan, X., Li, J., Gu, Q., and Xu, J. (2014). gWEGA: GPU-accelerated WEGA for molecular superposition and shape comparison. *J. Comput. Chem.* 35, 1122–1130. doi: 10.1002/jcc.23603
- Yan, X., Liao, C., Liu, Z., Hagler, T. A., Gu, Q., and Xu, J. (2016). Chemical structure similarity search for ligand-based virtual screening: methods and computational resources. *Curr. Drug Targets* 17, 1580–1585. doi: 10.2174/1389450116666151102095555
- Yang, H., Sun, L., Li, W., Liu, G., and Tang, Y. (2018). *In silico* prediction of chemical toxicity for drug design using machine learning methods and structural alerts. *Front. Chem.* 6:30. doi: 10.3389/fchem.2018.00030
- Yap, C. W. (2011). PaDEL-descriptor: an open source software to calculate molecular descriptors and fingerprints. *J. Comput. Chem.* 32, 1466–1474. doi: 10.1002/jcc.21707

- Yasuo, N., and Sekijima, M. (2019). Improved method of structure-based virtual screening via interaction-energy-based learning. *J. Chem. Inf. Model.* 59, 1050–1061. doi: 10.1021/acs.jcim.8b00673
- Ye, W. L., Zhang, L. X., Guan, Y. D., Xue, W. W., Chen, A. F., Cao, Q., et al. (2019). Virtual screening and experimental validation of eEF2K inhibitors by combining homology modeling, QSAR and molecular docking from FDA approved drugs. *New J. Chem.* 43, 19097–19106. doi: 10.1039/C9NJ02600B
- Zeng, X., Zhang, P., He, W., Qin, C., Chen, S., Tao, L., et al. (2018). NPASS: natural product activity and species source database for natural product research, discovery and tool development. *Nucleic Acids Res.* 46, D1217–D1222. doi: 10.1093/nar/gkx1026
- Zhang, B., Vogt, M., Maggiora, G. M., and Bajorath, J. (2015). Design of chemical space networks using a Tanimoto similarity variant based upon maximum common substructures. *J. Comput. Aided Mol. Des.* 29, 937–950. doi: 10.1007/s10822-015-9872-1
- Zhang, H., Liu, W., Liu, Z., Ju, Y., Xu, M., Zhang, Y., et al. (2018). Discovery of indoleamine 2,3-dioxygenase inhibitors using machine learning based virtual screening. *MedChemComm* 9, 937–945. doi: 10.1039/C7MD00642J
- Zhang, L., Song, J., Kong, L., Yuan, T., Li, W., Zhang, W., et al. (2020). The strategies and techniques of drug discovery from natural products. *Pharmacol. Ther.* 216:107686. doi: 10.1016/j.pharmthera.2020.107686
- Zhang, R., Li, X., Zhang, X., Qin, H., and Xiao, W. (2020). Machine learning approaches for elucidating the biological effects of natural products. *Nat. Prod. Rep.* 38, 346–361. doi: 10.1039/d0np00043d
- Zhang, X., Liu, T., Fan, X., and Ai, N. (2017). *In silico* modeling on ADME properties of natural products: classification models for blood-brain barrier permeability, its application to traditional Chinese medicine and *in vitro* experimental validation. *J. Mol. Graph. Model.* 75, 347–354. doi: 10.1016/j.jmgm.2017.05.021
- Zheng, S., Wang, Y., Liu, W., Chang, W., Liang, G., Xu, Y., et al. (2020). *In silico* prediction of hemolytic toxicity on the human erythrocytes for small molecules by machine-learning and genetic algorithm. *J. Med. Chem.* 63, 6499–6512. doi: 10.1021/acs.jmedchem.9b00853
- Zoffmann, S., Vercruysse, M., Benmansour, F., Maunz, A., Wolf, L., Blum Marti, R., et al. (2019). Machine learning-powered antibiotics phenotypic drug discovery. *Sci. Rep.* 9:5013. doi: 10.1038/s41598-019-39387-9

Conflict of Interest: RB was employed by company InsilicAll Ltda.

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 © 2021 Santana, do Nascimento, Lima e Lima, Damasceno, Nahum, Braga and Lameira. 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.



Molecular Docking and Dynamics Simulation Revealed the Potential Inhibitory Activity of ACEIs Against SARS-CoV-2 Targeting the hACE2 Receptor

Ahmed A. Al-Karmalawy¹, Mohammed A. Dahab^{2*}, Ahmed M. Metwaly³, Sameh S. Elhady⁴, Eslam B. Elkaeed^{5,6}, Ibrahim H. Eissa^{2*} and Khaled M. Darwish⁷

OPEN ACCESS

Edited by:

Lalith Perera,
National Institute of Environmental
Health Sciences (NIEHS),
United States

Reviewed by:

Asanga Bandara,
Pledge-Tx, United States
Imtaiyaz Hassan,
Jamia Millia Islamia, India

*Correspondence:

Mohammed A. Dahab
mohammeddahab@azhar.edu.eg
Ibrahim H. Eissa
ibrahimeissa@azhar.edu.eg

Specialty section:

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

Received: 30 January 2021

Accepted: 24 March 2021

Published: 04 May 2021

Citation:

Al-Karmalawy AA, Dahab MA,
Metwaly AM, Elhady SS, Elkaeed EB,
Eissa IH and Darwish KM (2021)
Molecular Docking and Dynamics
Simulation Revealed the Potential
Inhibitory Activity of ACEIs Against
SARS-CoV-2 Targeting the hACE2
Receptor. *Front. Chem.* 9:661230.
doi: 10.3389/fchem.2021.661230

¹ Department of Pharmaceutical Medicinal Chemistry, Faculty of Pharmacy, Horus University-Egypt, Damietta, Egypt,

² Pharmaceutical Medicinal Chemistry & Drug Design Department, Faculty of Pharmacy (Boys), Al-Azhar University, Cairo, Egypt, ³ Pharmacognosy Department, Faculty of Pharmacy (Boys), Al-Azhar University, Cairo, Egypt, ⁴ Department of Natural Products, Faculty of Pharmacy, King Abdulaziz University, Jeddah, Saudi Arabia, ⁵ Department of Pharmaceutical Sciences, College of Pharmacy, AlMaarefa University, Ad Diriyah, Saudi Arabia, ⁶ Department of Pharmaceutical Organic Chemistry, Faculty of Pharmacy (Boys), Al-Azhar University, Cairo, Egypt, ⁷ Department of Medicinal Chemistry, Faculty of Pharmacy, Suez Canal University, Ismailia, Egypt

The rapid and global spread of a new human coronavirus, Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) has produced an immediate urgency to discover promising targets for the treatment of COVID-19. Here, we consider drug repurposing as an attractive approach that can facilitate the drug discovery process by repurposing existing pharmaceuticals to treat illnesses other than their primary indications. We review current information concerning the global health issue of COVID-19 including promising approved drugs, e.g., human angiotensin-converting enzyme inhibitors (hACEIs). Besides, we describe computational approaches to be used in drug repurposing and highlight examples of *in-silico* studies of drug development efforts against SARS-CoV-2. Alacepril and lisinopril were found to interact with human angiotensin-converting enzyme 2 (hACE2), the host entranceway for SARS-CoV-2 spike protein, through exhibiting the most acceptable rmsd_refine values and the best binding affinity through forming a strong hydrogen bond with Asn90, which is assumed to be essential for the activity, as well as significant extra interactions with other receptor-binding residues. Furthermore, molecular dynamics (MD) simulations followed by calculation of the binding free energy were also carried out for the most promising two ligand-pocket complexes from docking studies (alacepril and lisinopril) to clarify some information on their thermodynamic and dynamic properties and confirm the docking results as well. These results we obtained probably provided an excellent lead candidate for the development of therapeutic drugs against COVID-19. Eventually, animal experiments and accurate clinical trials are needed to confirm the potential preventive and treatment effect of these compounds.

Keywords: COVID-19, molecular docking, molecular dynamics, ACEIs, hACE2

INTRODUCTION

In December 2019, rumors began to spread about the prevalence of a new unknown pneumonia-like illness in Wuhan, the capital of Hubei Province in China. Afterward, on February 11, 2020, the WHO reported a novel coronavirus as the causative agent of clusters of the new illness. Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) or COVID-19 was the name that the WHO designated for the disease caused by the novel coronavirus (Coronaviridae Study Group of the International Committee on Taxonomy of Viruses, 2020). Since the beginning of the outbreak, infections have expanded rapidly into multiple simultaneous epidemics worldwide. As of January 23, 2021, 99,071,240 confirmed COVID-19 cases and 2,124,086 COVID-19-related deaths have been reported across more than 221 countries (Culp, 2021).

The COVID-19 with influenza-like symptoms ranging from mild discomfort to severe lung injury and multi-organ failure, eventually leading to death (Rothe et al., 2020). Effective treatments for SARS-CoV-2 infection do not currently exist. Thus, it will be of great benefit to identify and repurpose already well-characterized compounds and approved drugs for use in combating COVID-19 (<https://www.who.int/emergencies/diseases/novel-coronavirus-2019>).

Drug repurposing or drug reprofiling is a promising field in drug discovery for identifying new therapeutic uses for already studied drugs (Khattab and Al-Karmalawy, 2021; Khattab et al., 2021). These drugs could be either currently approved and marketed for another use or withdrawn because of adverse effects (Ashburn and Thor, 2004). Available clinical trials at ClinicalTrials.gov (<https://clinicaltrials.gov/>) include the investigation of previously approved drugs for different indications, e.g., telmisartan and losartan. It offers a great opportunity to the traditional *de novo* drug discovery since the success rate of developing a new molecular entity is 2.01% only, and the number of approved drugs has been declining since the 1990's (Yeu et al., 2015). In the last decade, about one-third of the approvals correspond to drug repurposing, and repurposed drugs currently generate around 25% of the annual revenue for the pharmaceutical industry (Talevi and Bellera, 2020). As examples of the most common treatment, hydroxychloroquine, an antimalarial agent with anti-inflammatory and immunomodulatory activities, has shown inhibitory activity for SARS-CoV-2 similar to previous studies on SARS-CoV-1 (Sanders et al., 2020). It has been investigated for use by COVID-19 patients based on positive *in vitro* and limited clinical data. Also, azithromycin, a macrolide antibiotic, was found to raise the efficacy of hydroxychloroquine as a complementary therapy (Lover, 2020).

Computer-aided drug discovery is one of the most important approaches to investigate the activity of a drug through computational structure-based drug discovery. Different software tested the interaction between the tested compounds and the binding site through physics-based equations used to calculate their binding affinities (Sliwoski et al., 2014). SARS-CoV-2 proteins, particularly proteases and spike proteins (Prajapat et al., 2020), have been targeted in many docking

investigations hoping to understand the key amino acids essential for the interactions at the active site in SARS-CoV-2 (Calligari et al., 2020; Dahab et al., 2020; Khan et al., 2020; Kumar et al., 2020; Mohammad et al., 2020; Wu et al., 2020; Jairajpuri et al., 2021).

In general, various organ systems are believed to participate in COVID-19 due to the widespread expression of the primary SARS-CoV-2 entry receptor, human angiotensin-converting enzyme 2 (*hACE2*) (Groß et al., 2020). Angiotensinogen (AGT) as a key substrate of the Renin-Angiotensin System (RAS) is mainly synthesized by the liver and is cleaved by renin to form Ang I (proangiotensin). In the pulmonary circulation, Ang I is easily activated to *hACE2* (Wu et al., 2018). ACE is a zinc metallopeptidase ectoenzyme predominantly found in the lungs and was originally isolated in 1956 as (hypertension converting enzyme) (Skeggs et al., 1955). In 2000, genomic-based strategies led to the discovery of *hACE2*, a human ACE homolog. *hACE2* receptors which are the door through which the virus enters into cells and also the conductor of several pathophysiological reactions associated with the clinical features of the disease, with potential therapeutic implications (Donoghue et al., 2000).

Taking into account the characteristics of the mode of entry of this coronavirus to human cells through binding with *hACE2* and extensive scientific and clinical evidence information on the RAS, the hypothesis of the involvement of this system in the pathophysiology of COVID-19 was born (Gurwitz, 2020). The SARS-CoV-2 virus enters the airway and binds, utilizing the S (Spike) protein on its surface, to the membrane protein *hACE2* in type 2 alveolar cells. The S protein-*hACE2* complex is internalized by endocytosis and facilitates the entry of each virion into the cytoplasm (Wan et al., 2020).

hACE2 is involved in modulating blood pressure and establishing blood pressure homeostasis. Recently, a debatable question has risen, whether using antihypertensive medications will have a favorable impact on people infected with SARS-CoV-2 or a deleterious one, mainly since ACEIs and ARBs therapy can modulate the expression of *hACE2* protein (Vaduganathan et al., 2020).

We suppose that inhibition of the *hACE2* catalytic pocket by small molecules, e.g., ACEIs, could change the conformation of *hACE2* in such a way that it could block SARS-CoV-2 entry inside host cells through *hACE2* (Du et al., 2009).

Recently, a new promising success was reported: a group of scientists claimed that human recombinant soluble ACE2 (*hrsACE2*) can block the early stages of SARS-CoV-2 infections (Monteil et al., 2020). Moreover, telmisartan (ClinicalTrials.gov ID: NCT04355936) and losartan (ClinicalTrials.gov ID: NCT04312009) were proposed as alternative options for treating COVID-19 patients before the development of acute respiratory distress syndrome (ARDS) (Alnajjar et al., 2020; Gurwitz, 2020).

Interestingly, Zhang et al. found that among patients with hypertension hospitalized with COVID-19, inpatient treatment with ACEIs or angiotensin receptor blockers (ARBs) was associated with a lower risk of all-cause mortality compared with ACEI/ARB non-users (Zhang et al., 2020). Also, ACEIs proved to be particularly beneficial not only in controlling high blood pressure but also in reducing the incidence of stroke, by

downregulating tissue factor synthesis in monocytes (Dézsi, 2000; Napoleone et al., 2000).

For these reasons and in continuation to our previous works targeting SARS-CoV-2 (Alnajjar et al., 2020; Zaki et al., 2020; Al-Karmalawy et al., in press), the authors present a promising computational study including molecular docking and dynamics simulation for almost all FDA approved members of ACEIs (**Figure 1**) against the receptor-binding domain (RBD) of the spike protein of SARS-CoV-2 in complex with *hACE2* hoping to repurpose them effectively for the potential treatment of COVID-19 infection. However, we propose that ACEIs having the ability to block the *hrsACE2* receptor and so prevent the entrance of SARS-CoV-2 through its spike protein (**Figure 2**). Collectively, the main aim of the study is to investigate the potentiality of ACEIs, as promising small ligand molecules with drug-likeness properties, to accommodate the *N*-acetyl- β -glucosamine (NAG) specific binding site at the *hACE2* protein target. Accommodation of such pocket could permit distrusted glycan stability within such site being at proximity to the *hACE2*/SARS-CoV-2 Spike protein receptor-binding domain (RBD) interface. Accommodating this site by small molecules may impact the SARS-CoV-2 Spike protein owing to the reported findings of the glycan-mediated influence/interference with the *hACE2*/SARS-CoV-2 Spike protein association as well as spike epitopic recognition (Li et al., 2005; Banerjee et al., 2020; de Andrade et al., 2020; Devaux et al., 2020; Grant et al., 2020). Therefore, the affinity of ACEIs against the *hACE2*-NAG binding site was investigating through molecular docking and dynamics studies having the glycan NAG as a competitor binder and reference ligand.

MATERIALS AND METHODS

Both the molecular docking studies using MOE 2014.09 suite (Vilar et al., 2008) and molecular dynamics simulation using the GROMACS-2019 software package and CHARMM36 force field (da Silva et al., 2020) were applied in this study.

Molecular Docking Studies

To find a potential candidate for treating COVID-19, molecular docking studies were performed over 14 ACEIs on the binding pocket of the SARS-CoV-2 chimeric receptor-binding domain complexed with its receptor human *hACE2* (PDB IDs: 6VW1) (Shang et al., 2020). The chemical structures of drugs tested for docking study are depicted in **Figure 1**. The co-crystallized ligand *N*-Acetyl-*D*-Glucosamine (NAG) was used as a reference standard.

The tested compounds were sketched using ChemDraw 2014, imported into MOE, and subjected to 3D protonation and energy minimization up to 0.01 gradient. Then the co-crystallized ligand (NAG) and the tested compounds were imported into the same database and saved in the form of an MDB file to be used in the docking calculations with SARS-CoV-2 spike protein, 6VW1. The crystal structure was obtained from Protein Data Bank (<http://www.rcsb.org>) with good resolutions 2.68 Å (Shang et al., 2020). The crystal structures were prepared following the detailed procedure described earlier (Al-Karmalawy and Khattab, 2020;

Ghanem et al., 2020). They were imported into MOE and the structure preparation wizard of MOE was used to correct all the issues in protein structures. The hydrogen atoms were added to structures in their standard geometry, and all solvent molecules were removed from the structures then subjected to energy minimization. The final optimized structures were saved in the working directory. Triangle matcher and refinement methods were used for performing docking studies. Rigid receptor as refinement methodology and GBVI/WSA dG as the scoring methodology for selection of the best 20 poses from 100 different poses for each tested compound. The scoring methods were adjusted to their default values (Samra et al., 2021). After completion of docking processes, the obtained poses were studied and the best ones showing the best acceptable rmsd_refine values with the same binding mode of the native ligand were selected. Also, a program validation process was performed at first and confirmed by a low RMSD value (< 1 Å) as described before (Eliaa et al., 2020).

Molecular Dynamics Simulation

The best-docking scored models of the most promising leads, alacepril and lisinopril, in complex with *hACE2* protein were chosen as starting coordinates for 100 ns all-atom molecular dynamics simulation using a GROMACS-2019 software package (GNU, General Public License; <http://www.gromacs.org>) and CHARMM36 force field (da Silva et al., 2020). Each ligand-protein complex was solvated within a cubic box of the transferable intermolecular potential with a three-points (TIP3P) water model (100 × 100 × 100 Å) allowing a minimum of 10 Å marginal distance between protein and each side of the 3D box (Izadi et al., 2014). The CHARMM force field parameters for the investigated ligands were automatically generated using the CHARMM General Force Field (CGenFF) program (Vanommeslaeghe et al., 2009) (ParamChem project; <https://cgenff.umaryland.edu/>). Under periodic boundary conditions implementation, the protein residues were assigned for their standard ionization states at physiological conditions (pH 7.0), and the whole complexes were neutralized via sufficient numbers of K⁺ and Cl⁻ ions added via Monte-Carlo ion-placing method (Ross et al., 2019). The MD simulation was conducted over three stages and 1,000 kJ/mol.nm² force constant was used for restraining all heavy atoms and preserving original protein folding (Helal et al., 2020). The first stage involved initial optimization of each system geometry using 5,000 iterations (5 ps) with the steepest descent algorithm. The subsequent step involved system two-staged equilibration where the system was conditioned for 100,000 iterations (100 ps) at each stage. The first equilibration stage was proceeded under constant Number of particles, Volume, and Temperature (NVT) ensemble guided by the Berendsen temperature coupling method for regulating the temperature within the 3D box (Golo and Shaitan, 2002). Subsequently, the second equilibration stage was performed under a constant Number of particles, Pressure, and Temperature (NPT) ensemble at 1 atm and 303.15 K guided by using the Parrinello-Rahman barostat (Tuble et al., 2004).

Finally, the MD simulations were run for 100 ns under constant pressure (NPT ensemble) and long-range electrostatic

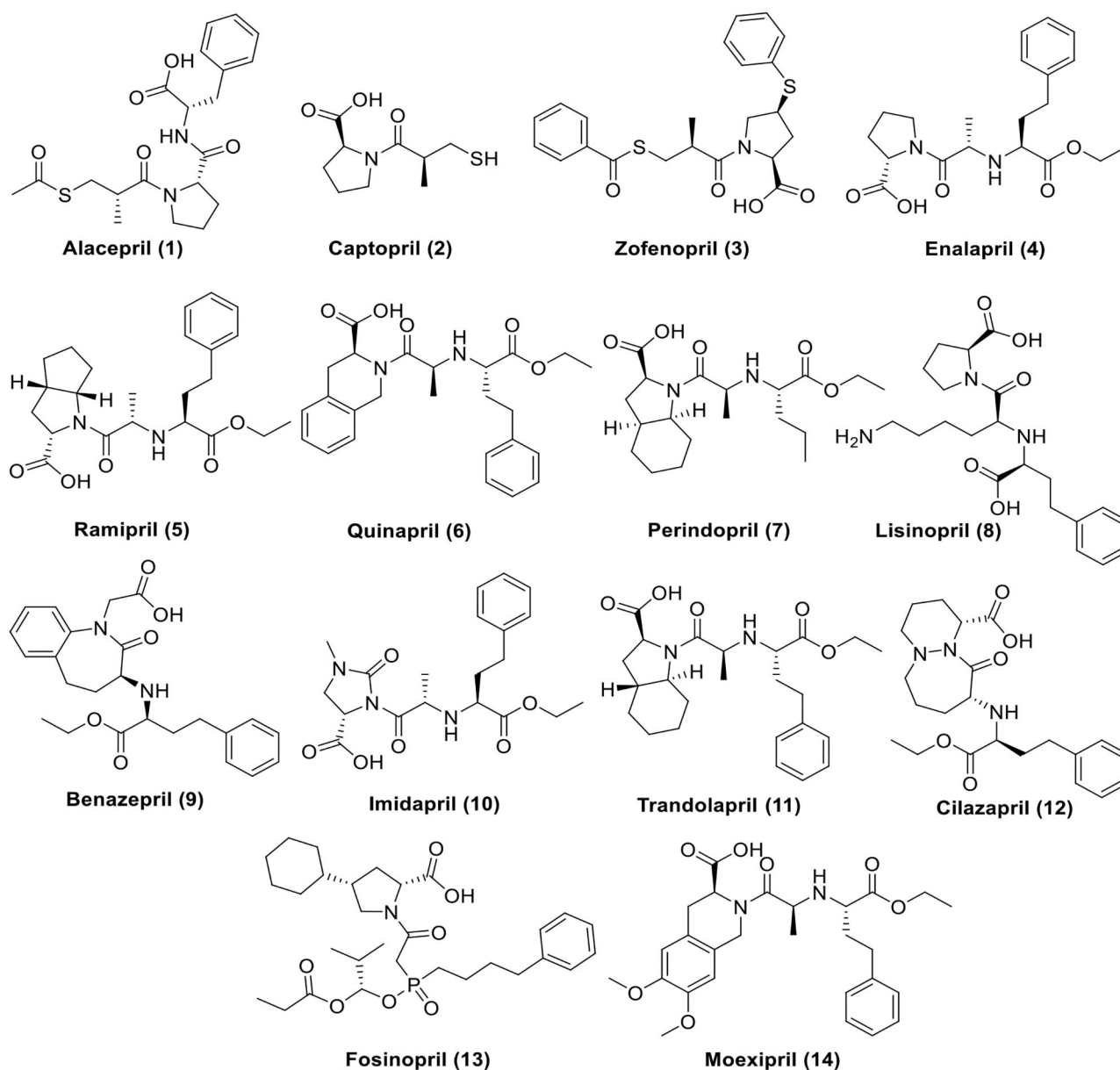
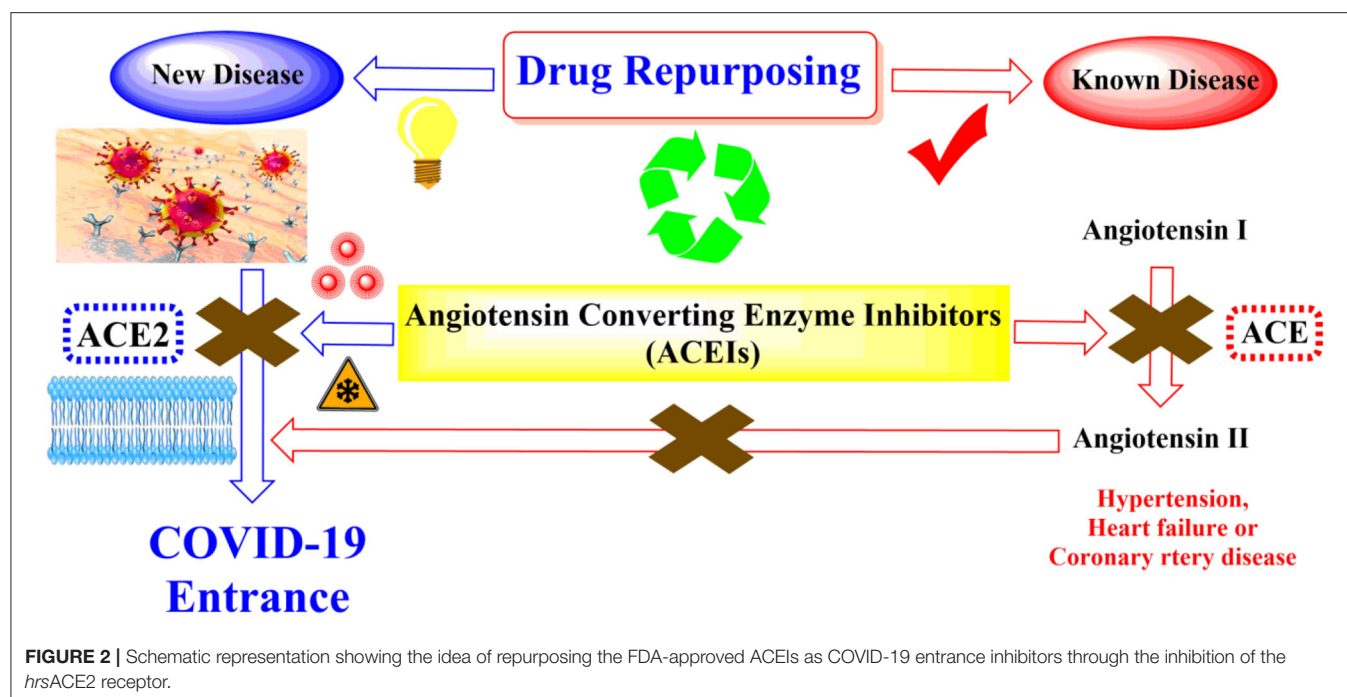


FIGURE 1 | Chemical structures of the tested ACEIs.

interactions were computed using Particle Mesh Ewald (PME) algorithm (Darden et al., 1998). Adopting such a highly accurate and rapid algorithm for treating long-range Coulomb interactions to achieve stable nanosecond trajectories within highly polar biomolecules like proteins. However, the implemented linear constraint LINCS method was used to constrain all covalent bond lengths, including hydrogens, allowing an integration time step size of 2 fs (Hess et al., 1997). The non-bounded interactions, Coulomb (electrostatic potential), and Lennard Jones (Pauli repulsion and hydrophobic/van der Waals attractions) interactions were truncated at 10 Å using the Verlet cut-off scheme (Páll and Hess,

2013). Throughout the MD simulation, the CHARMM36m all-atom force field was applied for both the ions and protein (Best et al., 2013). Computing comparative data, including RMSD and radius of gyration (R_g), was performed through analyzing the MD trajectories using the GROMACS built-in tools. Moreover, the Distance Calculation Tool, at Visual Molecular Dynamics 1.9.3 (VMD) package (the University of Illinois at Urbana-Champaign, USA), was utilized to calculate the change in the distance between the specified ligand/protein atoms over the whole simulation period (Humphrey et al., 1996). Such an approach permitted monitoring and investigating the possibility of interactions of ligands with the most important protein



residues. Finally, the binding-free energy between the ligand and protein was estimated via the GROMACS “*g_mmpbsa*” module (Kumari et al., 2014). The Pymol graphical software ver. 2.0.6 (Schrödinger™, NY, USA) was utilized for figure generation of ligand–protein conformational analysis (Delano, 2002).

RESULTS AND DISCUSSION

Molecular Docking Studies

Molecular docking simulations were performed in order to investigate the potentiality of small drug-like molecules, like ACEIs, to engage the *h*ACE2 glycosylated site and/or vicinal cavity in a way that would disrupt the glycosylation process of the *h*ACE2, leading to the modulation of *h*ACE2-RBD interactions. Actually, this crystallized *N*-glycan is covalently linked to the aimed nitrogen of the asparagine residue of the protein. Nevertheless, the approach of *N*-glycan and its existence within the pocket is highly guided by both Coulomb’s electrostatic interactions and Lenard-Johns van der Waal potential energy with different target residues comprising the *h*ACE2 pocket lining. In these regards, this *N*-glycan was considered as a reference ligand to investigate the ability of the investigated ACEIs to compete with it for engaging this glycosylated site and vicinal cavity. Throughout the adopted docking protocol, this *N*-glycan binder was fitted inside the binding pocket of SARS-CoV-2 spike protein showing one hydrogen bond with Asn90 (2.84 Å, binding score = −4.4, RMSD = 1.3), **Figure 3A**.

A molecular docking simulation of the target compounds and the native ligand into the spike protein active site was carried out. Many poses were obtained with better binding modes and interactions inside the receptor pocket. The poses with the most acceptable rmsd_refine values (related to the closeness of the

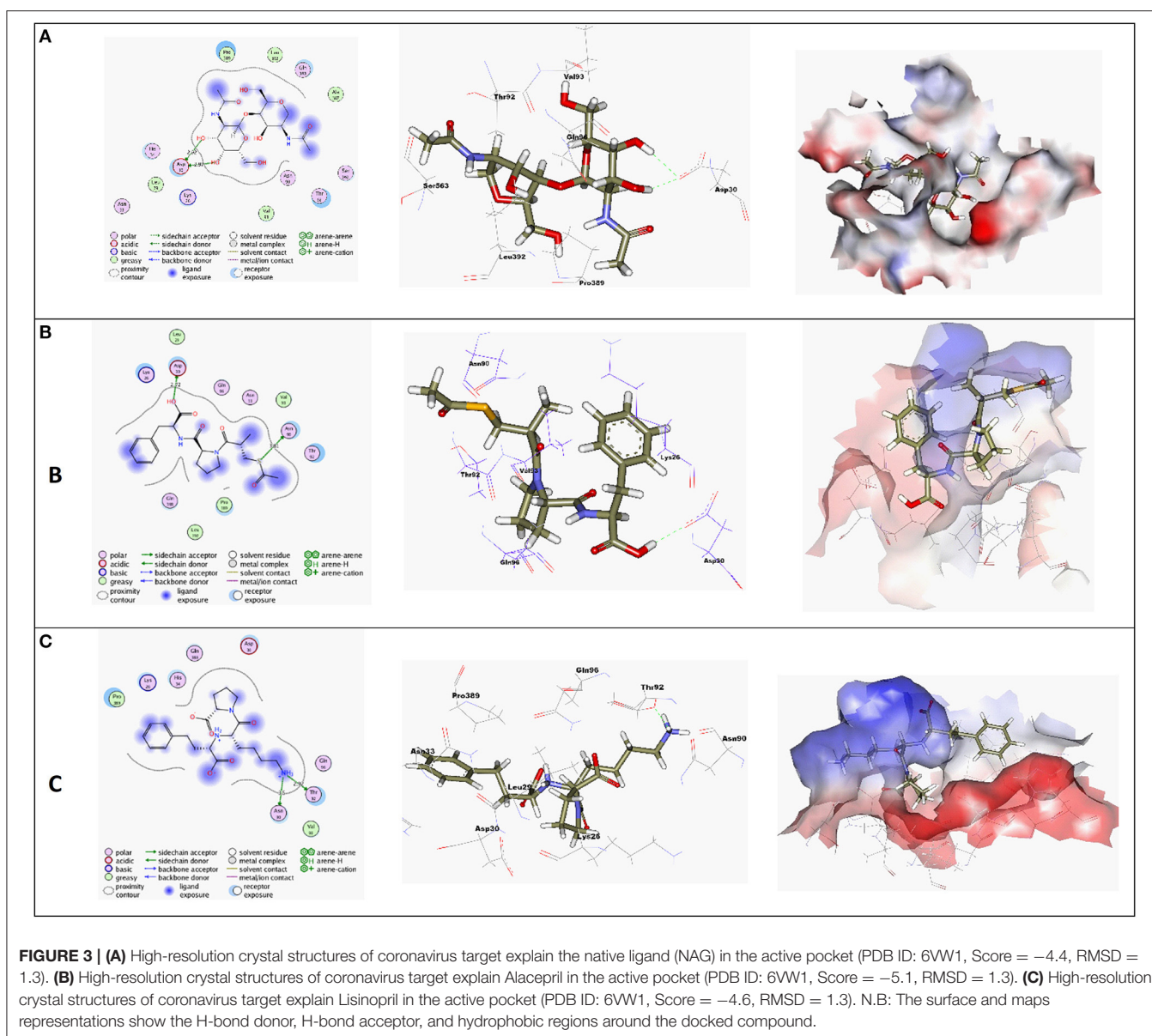
selected pose to the original ligand position inside the receptor pocket) and the same binding mode of the ligand were selected. Results of energies and different interactions with amino acids of the spike protein pocket are shown in **Table 1**. They got stabilized at the binding site of spike protein by variable several electrostatic bonds.

Most compounds showed acceptable RMSD values close to the NAG inhibitor, but only alacepril and lisinopril have the same binding mode of the NAG. For alacepril, binding interactions with 6VW1 (binding score = −5.1, RMSD = 1.3) are given in **Figure 3B**, two hydrogen bonds were recorded, one of them with Asn90 (3.81 Å), which is assumed to be essential for the activity. In addition, another hydrogen bond was observed with Asn30 (2.72 Å), whereas, in the case of lisinopril, binding interactions with 6VW1 (binding score = −4.7, RMSD = 1.3) are given in **Figure 3C**, and two hydrogen bonds also were recorded, one of them with Asn90 (3.50 Å), which is assumed to be essential for the activity. Furthermore, another hydrogen bond was observed with Asn30 (2.92 Å).

Finally, some ACEIs such as trandolapril, fosinopril, and moexipril have excellent binding scores (−5.60, −5.04, and −5.10, respectively), better than the native ligand NAG (−4.4), but, unfortunately, their binding modes are different. For trandolapril, two hydrogen bonds were observed with Asp30 and the third one with Gln95 (2.75, 2.75, and 2.91 Å). For fosinopril, one hydrogen bond was observed with Gln96 (4.36 Å). For moexipril, three hydrogen bonds were observed with Asp30 (4.25, 3.16, and 3.36 Å).

Molecular Dynamics Simulation

Considering it as an efficacious approach for validating the stability of the predicted docked ligand–*h*ACE2 complex, an



all-atom molecular dynamics (MD) simulation study was performed. Adopting such a study would also provide valuable information regarding the dynamic behavior of both the ligand and *hACE2* protein as well as evaluate the ligand's key binding interactions with important catalytic site residues (Karplus and Petsko, 1990). Therefore, the predicted ligand-protein complexes, for both alacepril and lisinopril, as well as the glycosylated *hACE2* protein were enrolled within 100 ns all-atom MD simulation.

Trajectory Analysis of Ligand-*hACE2* Complexes

The stability profile of both alacepril and lisinopril in complex with the human angiotensin-converting enzyme 2 (*hACE2*) was monitored using the GROMACS command line *gmx_rmsd* to estimate their respective RMSD values throughout the simulation runs. Generally, RMSD provides an inference regarding the

deviation extent for a group of atoms (protein, ligand, or even ligand-protein complex) to the respective initial reference structure (Schreiner et al., 2012). Thus, high RMSD values would be correlated to significant instability, being related to changes within the conformation of the investigated molecule. Moreover, ligands depicting high RMSD values, for their respective ligand-protein complex, would suggest inadequate ligand accommodation within the studied pocket across the adopted MD simulation time-frames (Liu et al., 2017).

Within the presented MD simulation, both investigated ligand-protein targets exhibited successful conversion following 20 ns of MD simulation start (**Figure 4A**). The obtained complex RMSD trajectories, in respect of their backbone, rises throughout the initial frames till the RMSDs level off at around 20 ns where the following trajectories proceeded around respective average values till the 70 ns of the MD simulation. It worth

TABLE 1 | Receptor interactions and binding energies of ACEIs drugs and NAG inhibitor into the spike protein of SARS-CoV-2.

No.	ACEIs	S ^a Kcal/mole	RMSD_Refine ^b	Amino acid bond	Distance Å
1	Alacepril	−5.10	1.3	Asn90/H-acceptor	3.81
				Asp30/H-acceptor	2.72
2	Captopril	−3.40	1.4	Asp30/H-acceptor	3.76
3	Zofenopril	−4.6	1.6	Pro389/arene-H	4.34
4	Enalapril	−4.8	1.5	Asp30/H-donor	2.94
				Asp30/H-donor	2.94
5	Ramipril	−4.6	1.7	Lys26/H-acceptor	4.29
				Lys26/H-acceptor	3.98
6	Quinapril	−4.60	1.7	Pro389/arene-H	4.52
				Gln96/H- acceptor	3.07
7	Perindopril	−4.2	1.7	Asp30/H-donor	3.31
				Asp30/H-donor	3.32
				Asp30/H- acceptor	3.31
				Asp30/H- acceptor	3.32
8	Lisinopril	−4.70	1.3	Asn90/H-acceptor	3.5
				Thr92/H-acceptor	2.92
9	Benazepril	−4.70	1.3	Lys25/H-donor	3.07
				Lys25/H-donor	3.07
10	Imidapril	−4.4	1.8	Asp30/H-donor	3.45
				Asp30/H-donor	3.45
11	Trandolapril	−5.60	1.2	Asp30/H-donor	2.75
				Asp30/H-donor	2.75
				Gln95/H-acceptor	2.91
12	Cilazapril	−4.5	1.6	Pro389/arene-H	4.49
				Asp30/H- donor	3.24
				Asp30/H- donor	3.24
				Asp30/H- donor	3.60
13	Fosinopril	−5.04	1.7	Gln96/H-acceptor	4.36
14	Moexipril	−5.10	1.7	Asp30/H- donor	4.25
				Asp30/H- donor	3.16
				Asp30/H- donor	3.36
15	NAG	−4.4	1.3	Asp30/H- donor	2.97
				Asp30/H- donor	2.92

^aS: the score of placement of a compound into the binding pocket of protein using London dG scoring function.

^bRMSD_Refine: the root-mean-squared-deviation (RMSD) between the heavy atoms of the predicted pose (after refinement) and those of the crystal structure (before refinement).

noting that the average RMSD values, throughout the plateau MD simulation interval (20–70 ns), were higher for lisinopril compared to alacepril (2.610 ± 0.20 Å vs. 3.786 ± 0.13 Å). The latter differential dynamic behavior confers a more stabilized and confinement accommodation for alacepril within the *h*ACE2 binding site throughout the plateau interval. However, both ligands converge around comparable RMSD values (~ 3.400 Å) where only the alacepril–protein trajectories were depicted steady till the end of the MD simulation at 100 ns. A second RMSD trajectory increase at the last 10 ns of the MD simulation was shown for lisinopril–protein complex tones, which further confirms a significant ligand shift out of the *h*ACE2 pocket. On the other hand, alacepril depicted a minimal increase within RMSD trajectories (from 2.316 to 3.110 Å) following the 70 ns suggesting a limited chance of the alacepril orientation within

the *h*ACE2 pocket rather than a dramatic escape out of the binding site. All latter findings confer maintained binding of alacepril within the *h*ACE2 binding site. Compared to lisinopril, the alacepril–protein complex depicted comparable RMSD tones to those of NAG-bound (glycosylated) protein along the 100 ns all-atom MD simulation run. All above findings suggest a more preferential binding for alacepril, over lisinopril, within the *h*ACE2 NAG-binding site.

Further investigation of ligand stability within the protein binding site was proceeded through monitoring the ligand RMSD tones (**Figure 4B**). Monitoring these trajectories would provide valuable information regarding the conformational/orientation of the simulated ligands in respective to their binding pocket. Following convergence, the bound NAG molecule showed the steadiest RMSD tones (8.970 ± 1.14 Å) across the entire 100

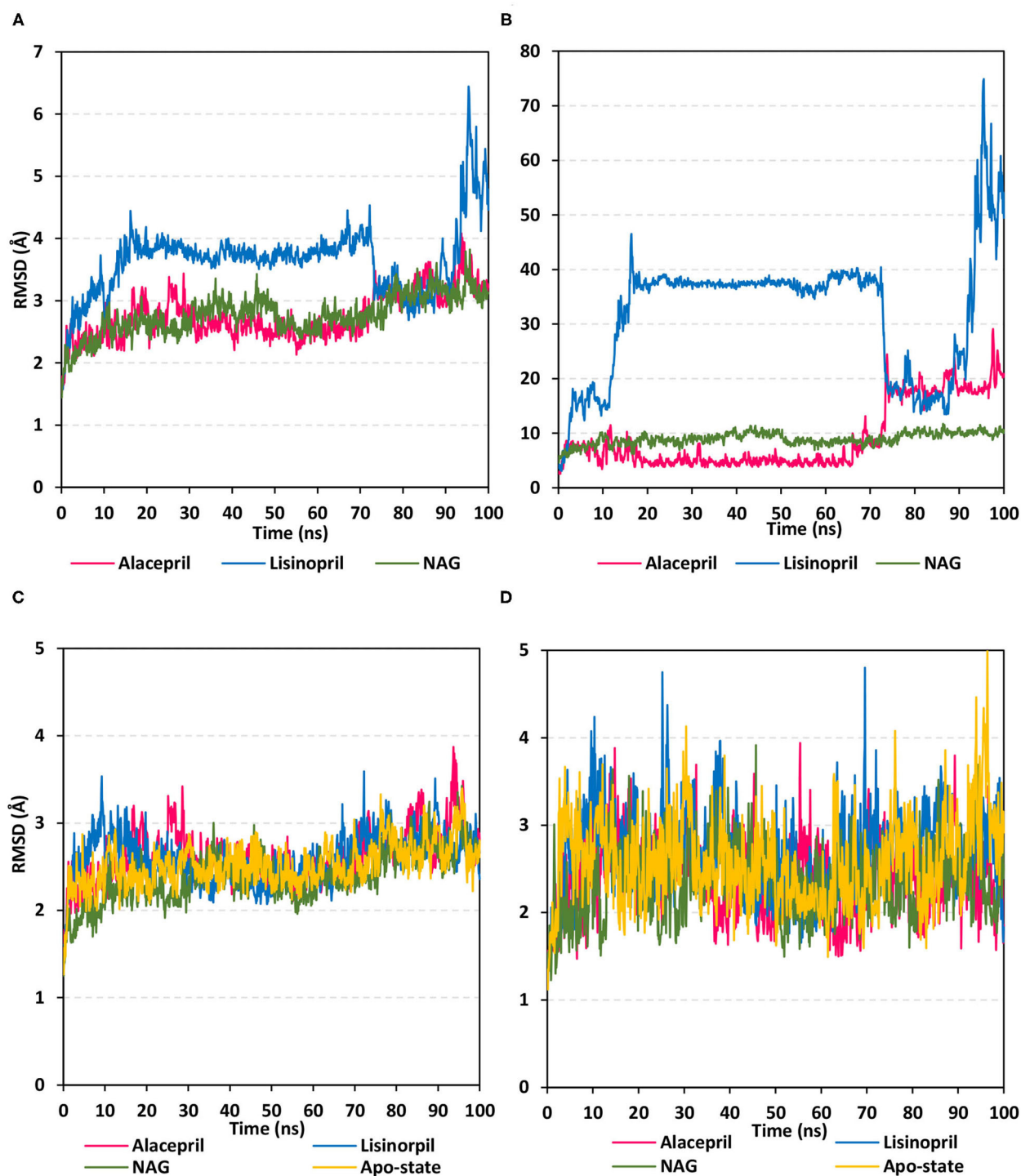
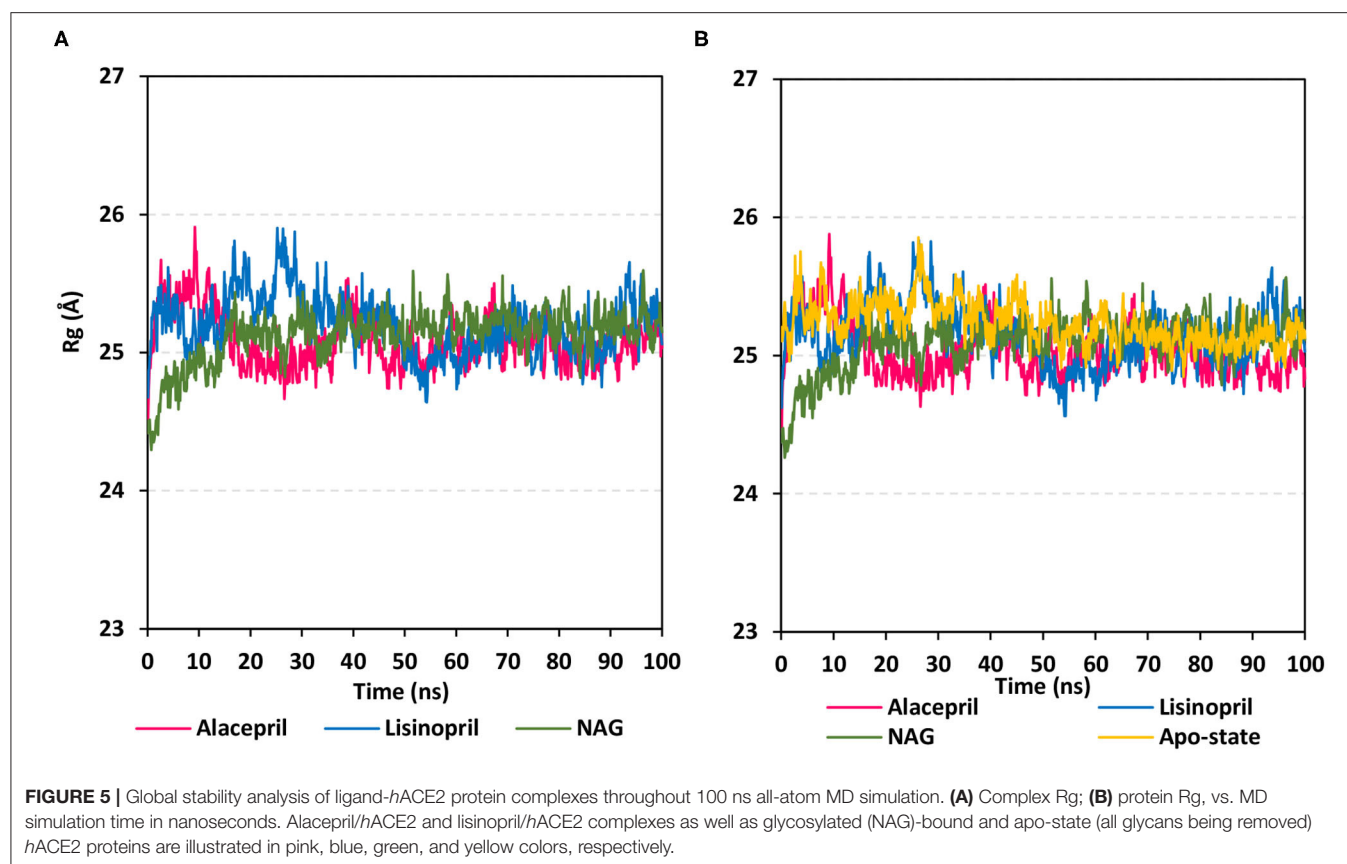


FIGURE 4 | Analysis of RMSD trajectories for the ligand-*hACE2* protein complexes throughout 100 ns all-atom MD simulation. **(A)** Complex RMSD; **(B)** ligand RMSD; **(C)** protein RMSD; **(D)** binding pocket residues RMSD, relative to backbone vs. MD simulation time in nanoseconds. Alacepril/*hACE2* and lisinopril/*hACE2* complexes as well as glycosylated (NAG)-bound and apo-state (all glycans being removed) *hACE2* proteins are illustrated in pink, blue, green, and yellow colors, respectively.

ns all-atom MD simulation. Nevertheless, alacepril depicted the lowest RMSD trajectories (4.962 ± 1.28 Å) around the 20–70 ns MD simulation run being at ~ 1.5 Å RMSD values below those of its respective ligand–protein complex. With

only limited fluctuations, the alacepril RMSD tones emphasize its preferential accommodation of the *hACE2* NAG-binding site as compared with lisinopril. The latter ligand depicted an extreme orientation/conformation shift relative to its initial



coordinates (37.542 ± 0.92 Å) following 20 ns and up to 70 ns.

Beyond the 70 ns MD simulation runs, both ACEIs ligands exhibited comparable trajectories around 75–90 ns with the highest fluctuations being assigned for lisinopril. Finally, another elevated lisinopril RMSD values (> 50 Å), near the end of the MD simulation timeframe, suggested that lisinopril has left the protein interaction side while being strayed at the solvent site. Further monitoring of the pocket residue RMSD trajectories, with the crystal structure, was informative regarding the differential ligand binding within the *hACE2* NAG-binding site (**Figure 4C**). As expected, the highest RMSD tones (2.777 ± 0.48 Å) were assigned to lisinopril-pocket residues with high fluctuations being depicted around 25 ns and 70 ns (4.750 Å and 4.800 Å, respectively). Notably, pocket residues showed lower RMSDs with both alacepril and NAG binding (2.394 ± 0.42 Å and 2.346 ± 0.41 Å, respectively), as compared to *hACE2* with all glycans being removed (apo-state; 2.570 ± 0.49 Å), particularly near the end of the MD simulation. The latter behaviors confer preferential ligand-pocket mutual stability relationship for alacepril and NAG across the MD simulation runs.

For excluding the presence of any artifacts within the adopted MD simulation runs, the *hACE2* protein RMSD trajectories were monitored both for the apo (unbounded) and glycosylated (NAG-bound) states as well as in complex with both investigated ligands, alacepril, and lisinopril. Interestingly, the RMSD tones

were comparable for the apo and complexed proteins since limited differential RMSD values were obtained across the 100 ns MD simulation window (**Figure 4D**). A little elevation of the protein RMSD tones, concerning their C-alpha atoms, was depicted at first frames of MD simulation and then an equilibrium plateau was achieved around an average RMSD of 2.558, 2.524, 2.661, and 2.611 Å, for apo, NAG, alacepril, and lisinopril-bound proteins, respectively. Such protein behavior is typical for optimum MD runs since all the applied constraints, before the simulation, were released and the protein starts to relax till reaching an equilibration state around which the RMSD revolves until reaching the MD simulation end. Showing comparable average RMSD values for apo *hACE2*, relative to those for NAG, alacepril, and lisinopril-bound proteins could exclude the presence of differential significant secondary structure rearrangement/folding within the three MD simulation runs. The latter findings further correlate the RMSD complex trajectory fluctuations to the ligand behaviors rather than that of respective proteins within the MD simulation runs. It worth noting that all protein RMSDs reached comparable values (~ 2.600 Å) at the end of the MD run which further validate the 100 ns MD simulation time frame being able to bring both the apo, glycosylated and complexed proteins at comparable equilibration/relaxed states. Moreover, the latter dynamic behaviors further ensure sufficient conditioning stages before the production of the MD simulation runs.

To gain more insight regarding the investigated complex stability, the radii of gyration (R_g) were monitored across the whole MD trajectories using the GROMACS “*gmx_gyr*” command script. This stability parameter accounts for global stability of either ligand or protein ternary structure, where R_g is the mass-weighted RMSD for a group of atoms relative to their common mass center (Likić et al., 2005). Therefore, sustained stability/compactness of the investigated molecule would be inferred through depicted low R_g values achieving a plateau around an average value. Within the furnished study, the obtained R_g tones confirm the preferential stability of the alacepril-*hACE2* complex as compared to those of lisinopril (Figure 5). Steadier R_g trajectories were obtained for the alacepril complex with lower maximum, average, and minimum values (Table 2), suggesting compactness and stability of the ligand within the protein active site. Comparable values were depicted for alacepril and glycosylated (NAG)-bound protein complexes. The latter complex R_g findings were highly correlated with those of respective proteins. Minimal fluctuations and low R_g standard deviations were observed with alacepril and NAG as compared to that of lisinopril (25.03 ± 0.17 Å and 25.08 ± 0.19 Å; vs. 25.20 ± 0.21 Å, respectively). Interestingly, lower R_g s was assigned for the alacepril-bound and glycosylated (NAG) *hACE2* proteins with the protein's apo-state (25.23 ± 0.15 Å) suggesting a more compacted secondary structure upon ligand binding as well as protein glycosylation. All obtained R_g findings showed high agreement with the previous RMSD analysis confirming preferential better stability of alacepril over lisinopril within the *hACE2* NAG-binding site.

Protein Flexibility and Root-Mean-Square Fluctuation of Target Residues

For gaining more insights regarding the stability of the complex binding site, the per residue root-mean-square fluctuation (Δ RMSF) profile was estimated for each ligand-bound protein relative to the *hACE2* apo-state. The individual backbone RMSF of each protein was estimated using the GROMACS “*gmx_rmsf*” command line. This flexibility validation criterion provides information regarding the contribution of protein individual residues within the ligand/protein complex structural fluctuations. RMSF estimates the time evolution of the average deviation for each residue from its reference position within the minimized starting structures (Benson and Daggett, 2013). Adopting a Δ RMSF cut-off value of 0.30 Å was relevant for estimating the significant change within structural movements, where residues with > 0.30 Δ RMSF values were considered of decreased mobility (de Souza et al., 2019).

Findings within Figure 6 showed expected terminal-free residue behavior with high negative Δ RMSF values since they are most likely to fluctuate at the highest deviations in comparison to core residues the thing that is typically depicted in well-behaved MD simulation. However, a different terminal-free residue pattern was assigned for each ligand. Lower RMSF negative values or even positive RMSF values were depicted for alacepril and NAG, respectively, for the C-terminal-free residues and vicinal residues. Since the *hACE2*-NAG pocket residues

are at proximity to the protein C-terminal side, such findings confer more stabilized alacepril and NAG-protein complexes as compared to lisinopril. At the N-terminal, lower negative RMSF values were assigned to lisinopril relative to alacepril and NAG, suggesting that N-terminal-free residues and vicinal residues might impact lisinopril-protein binding through MD simulation. As these latter residues are at > 30 Å distant from the reference *hACE2*-NAG binding site, they may be highly correlated to stabilization of lisinopril following the dramatic conformational/orientation shift beyond 20 ns and up to 70 ns of the MD simulation run.

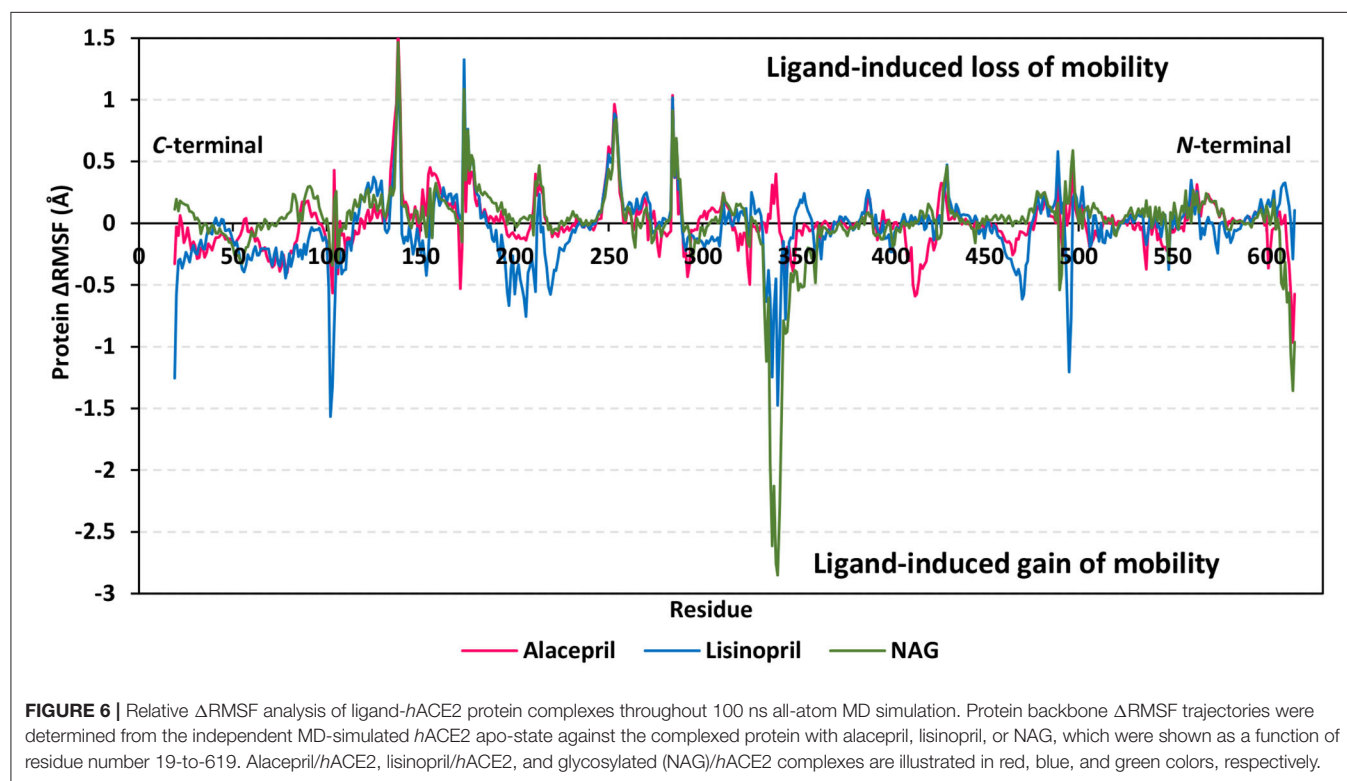
Concerning core protein residues, the three bounded ligands induced significant limited mobility (Δ RMSF > 0.3 Å) for *hACE2* residues at four distinct residue ranges including; range-I (134–140), range-II (173–178), range-III (248–256), and range-IV (284–286). The earlier two residue ranges-I and -II exhibited the greatest immobility with Δ RMSF values up to 1.55 Å and 0.91 Å, respectively. on the other hand, the other two less mobile residue ranges (-III and -IV) were at comparable Δ RMSF trajectories across the designated MD simulation window. Within the four top immobile residue ranges, the Δ RMSF trajectories for the three bound ligands were depicted as comparable. It worth noting that residues within the four residue ranges are at distances being > 29 Å from the bounded ligands the thing that can infer the impact of ligand binding site to induce stabilization of the protein secondary structures distant from the NAG-binding site.

Regarding residues with the highest fluctuations, there is a general trend of high negative RMSF values being assigned to the lisinopril-bound protein residues. Designated residue ranges (101–110, 195–220, and 462–473) exhibited high negative Δ RMSF values in particular for the protein in complex with lisinopril. Nevertheless, residues at these latter ranges showed limited flexibility regarding both alacepril and NAG-bound protein. Notably, one residue range (333–359) did not exhibit a similar pattern to the above highly mobile or immobile ranges, where residues of both lisinopril and NAG-bound protein were of great fluctuation/flexibility (maximum Δ RMSF –1.48 and –2.85 Å, respectively). On the contrary, positive Δ RMSF values (up to 0.40 Å) were assigned for the latter contradictory residue range up on alacepril binding suggesting the great impact of these residues on the alacepril-protein binding, which may be highly related to the suggested second conformation/orientation of alacepril following the 70 ns MD simulation run.

Further comparative analysis of the furnished Δ RMSF trajectories for the key residues lining the *hACE2*-NAG binding site permitted more insights regarding differential ligand-protein interactions. To the most interest, several pocket residues illustrated significant immobility with a Δ RMSF value of > 0.30 Å for alacepril-bound protein (Table 3). Pocket residues including Asn90, Leu91, Leu560, and Ser563 depicted the highest Δ RMSF values being the most positive for Leu91 suggesting the residue's key role in alacepril-pocket anchoring. Concerning the pocket residues of the NAG-bound protein, Asn90 and its vicinal residues (Leu91 and Thr92) depicted significant rigidity. This was not surprising

TABLE 2 | The Rg values for investigated ligand-*hACE2* complexes across the all-atom MD simulation.

Reference atom group	Alacepril- <i>hACE2</i> complex			Lisinopril- <i>hACE2</i> complex			Glycosylated (NAG) <i>hACE2</i>		
	Maximum (Å)	Average (Å)	Minimum (Å)	Maximum (Å)	Average (Å)	Minimum (Å)	Maximum (Å)	Average (Å)	Minimum (Å)
Complex	25.78	25.08 ± 0.09	24.49	25.90	25.20 ± 0.21	24.63	25.59	25.12 ± 0.19	24.29
Protein	25.75	25.03 ± 0.17	24.45	25.88	25.15 ± 0.21	24.56	25.57	25.08 ± 0.19	24.26



since crystallized NAG molecule is linked to *hACE2* at Asn90 within *hACE2* crystal structure. This observation ensures the stability of NAG as well as alacepril within the binding site along with the MD simulation frames. Moreover, the ability of alacepril to exhibit comparable immobility pattern or Asn90 and vicinal residues further emphasize the competitive capability of alacepril to replace NAG at its binding site. Moving toward the protein in complex with lisinopril, only Leu560, and Ser563 showed relevant rigidity with ΔRMSF values at the borderline (0.250 and 0.258 Å, respectively) being lower than those depicted with alacepril. It worth mentioning that several lisinopril-pocket residues, even those at the initial docking study, exhibited significant flexibility/fluctuations with ΔRMSF being of negative values (−0.035 to −0.264 Å). This finding can be correlated with the earlier suggestion that lisinopril has left the *hACE2*-NAG binding site exhibiting dramatic orientation/conformation shift. All above ΔRMSF analysis infer the inferior impact of lisinopril, as compared to alacepril and NAG, on the immobility/stability of the protein pocket residues. Therefore,

the ΔRMSF analysis is considered relevant as it came in great agreement with the above ΔRMSD and Rg findings suggesting the higher alacepril-*hACE2* complex stability relative to that of lisinopril.

Conformational Analysis Across Selected Trajectories

For gaining more insight regarding the newly adopted ligand-protein conformations by each ligand within the late MD simulation runs, the selected frames of each system were extracted and minimized to a gradient of 0.001 Kcal/mol/Å² using MOE software for further analysis of key changes. **Figure 7A** illustrates the comparative conformations of the alacepril-protein complex at 0, 70, and 100 ns. Interestingly, there is no significant orientation change for the ligand within the *hACE2* binding site between the time frames 0 and 65 ns. There was only a relevant shift toward the main chain of the Asp90 residue furnishing significant hydrogen bonding with its backbone amide. Such a shift caused a loss of the initial hydrogen bond with Asp30 and Gln96. Stabilization of alacepril within its new conformation/orientation was further mediated by several

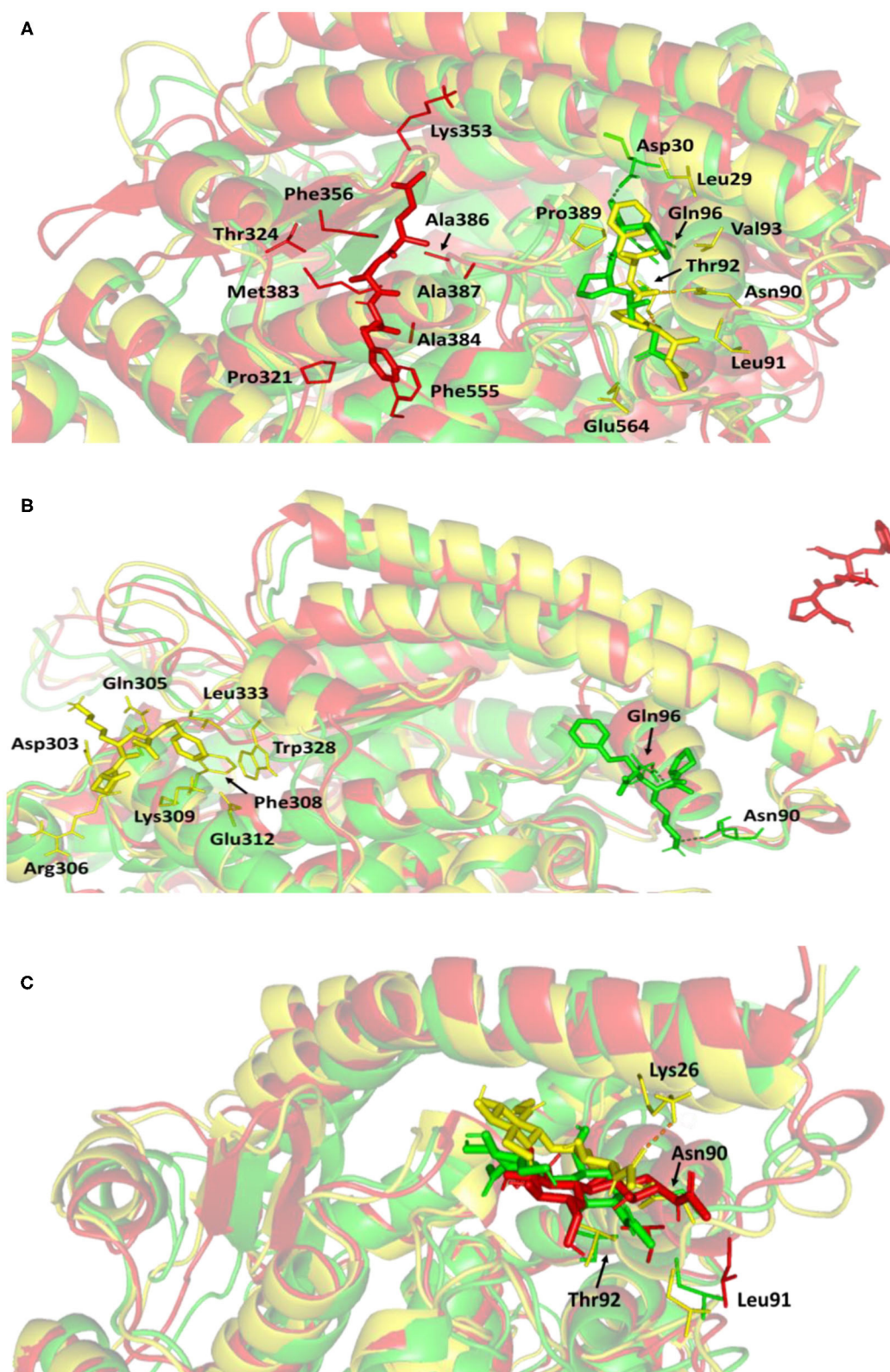


FIGURE 7 | Conformations of the ligand-protein complex at *hACE2* binding site through selected trajectories. **(A)** Alacepril; **(B)** lisinopril; **(C)** NAG. Protein is represented in green, yellow, and red cartoon 3D-representation corresponding to initial (0 ns), dynamic equilibrium (70 ns), and last (100 ns) extracted trajectories, respectively. The key binding residues (lines), ligands (sticks), and hydrophilic interactions (hydrogen bonding; dashed lines) are all presented in colors corresponding to their respective extracted trajectory.

TABLE 3 | Calculated Δ RMSF^a trajectories of ligand-*h*ACE2 proteins along with the MD simulation.

Residues of <i>h</i> ACE2-NAG binding site	Alacepril	Lisinopril	NAG
Ala25	−0.114	−0.236	0.148
Lys26	−0.037	−0.264	0.146
Asp30	0.239	−0.182	0.030
Lys31	0.286	−0.252	0.043
Asn90	0.381	−0.146	0.295
Leu91	0.404	−0.076	0.300
Thr92	0.053	−0.035	0.274
Val93	0.082	−0.06	0.204
Leu95	0.028	−0.053	0.221
Gln96	−0.006	−0.039	0.130
Ala387	0.091	0.198	0.070
Gln388	0.217	0.267	0.089
Pro389	0.199	0.191	0.054
Leu560	0.315	0.250	0.005
Ser563	0.314	0.258	0.241
Glu564	0.197	0.105	0.088

^aRelative difference root-mean-square fluctuation (Δ RMSF) was estimated for each ligand-bound protein relative to *h*ACE2 apo-state being without any glycan. Residues exhibiting significant immobility (Δ RMSF above 0.30 Å) are only written in bold and representative Δ RMSF value is highlighted.

non-polar residues, including Leu29, Lue91, Val93, Pro389, in addition to the C β of Glu564 side chain.

Concerning the ligand conformation at frame 100 ns, a more significant shift was depicted by alacepril toward a transient opened cleft at proximity to the SARS-CoV-2 spike-protein recognition domain-III. Such shift came in good agreement with the RMSD fluctuation following the 70 ns. Notably, the ligand was mainly maintained within this transit cleft through hydrophobic interaction with pocket lining residues. Being anchored at proximity to the protein's hydrophobic residues, Pro321, Phe356, Ala383, Ala386, Ala387, and Phe555, favored non-polar interactions were depicted with the ligand's terminal phenyl ring and pyrrolidine hydrophobic cage.

Validating the stability of alacepril within this transit cleft was achieved through extending the MD simulation. The last alacepril frame at 100 ns was extracted, minimized, and then proceeded within an extra 50 ns all-atom MD simulation adopting the same parameters at the initial 100 ns MD simulation run. Notably, alacepril showed great stability across the additional trajectories where the RMSD tones for the alacepril-*h*ACE2 complex and protein were maintained at low values (2.511 ± 0.33 Å and 2.482 ± 0.34 Å, respectively), following convergence (**Supplementary Figure 1**). Showing minimal fluctuations across the extended trajectories confirms the stability of alacepril at the transit cleft being still bounded with the pocket residues.

Concerning the lisinopril-*h*ACE2 complex, a more dramatic conformational and orientation shift was depicted for the ligand (**Figure 7B**). Throughout the dynamic equilibration shown from 20 to 70 ns, lisinopril was anchored at a distant pocket seated at ~ 25.00 Å from the initial *h*ACE2 binding site. These deviations

can be correlated to the high complex RMSD-C α fluctuations (**Figure 4A**) and the high maximum value of complex Rg (25.90 Å) compared to the alacepril-protein complex system. At this new distant pocket, relevant hydrophobic contacts between lisinopril and lining residues (Phe308, Trp328, and Leu 333) greatly mediated the ligand-protein complex stability. Interestingly, this distant pocket is near the *N*-terminal free residues and their vicinal residues. The binding of lisinopril within this distant pocket can explain the lower negative Δ RMSF trajectories of the *N*-terminal free residues, as compared to alacepril and NAG. Therefore, it is suggested that these residues impose a crucial role in stabilizing the lisinopril-protein complex within the 20–70 ns timeframe. Based on the furnished results, inferior stability within the *h*ACE2 binding site was assigned to lisinopril as compared to alacepril. The latter was further confirmed since lisinopril was found at the solvent side as being drifted away from the *h*ACE2 protein at the end of the MD simulation (100 ns).

Investigating the conformational changes for the glycosylated *h*ACE2 protein showed that NAG was retained within the binding pocket along with the whole MD simulation timeframe (**Figure 7C**). There is a quite comparable orientation for the NAG conformation at the 70 ns frame concerning its initial position at 0 ns time. Polar hydrogen bonding with the pocket hydrophilic residue, Lys26, was shown to provide extra stability for the NAG at the binding site. On the other hand, significant movement of NAG, as well as the pocket residues (Asn90, Leu91, and Thr92), was illustrated at the end of the MD simulation. Despite that, these particular pocket residues have exhibited relevant immobility with high positive Δ RMSF values (**Table 3**), a significant change in their respective position as depicted. This could raise the assumption that NAG is not fully occluding the binding site of interest the thing that could make it at least partially accessible across the designated MD simulation. Proving such a concept would provide relevant evidence that small druggable molecules, like alacepril, could manage to accommodate the *h*ACE2-NAG binding site of the glycosylated protein.

Extent of *h*ACE2 Binding Site Coverage by NAG

To speculate the possibility of small ligand inhibitors to accommodate the *h*ACE2-NAG binding site, an investigation of the extent of *h*ACE2 binding site coverage by NAG was within the glycosylated protein was proceeded. The GROMACS “*gmx sasa*” tool was used to compare the solvent-accessible-surface area (SASA) of the binding region in the absence and presence of glycan. Generally, SASA correlates for the molecular surface area being assessable to solvent molecules providing a quantitative measurement about the extent of protein/solvent interaction (Pirolli et al., 2014). The analysis was calculated for the atoms of lining residues comprising the *h*ACE2 binding site using spherical probes estimating the area exposed to the solvent. The % area of the binding site coverage was calculated as the percentage difference between the solvent-exposed area in the presence and absence of NAG. The solvent-sized probes (small radii, 1.4 Å) were applied to detect the binding site regions being within direct contact with the glycan. These small-sized probes are appropriate

for checking the accessibility of small drug-like molecules. However, larger probes (5–10 Å radius) are more correlated with more accurate SASA calculations for macromolecules including antibodies and protein-based molecules (Urbanowicz et al., 2019). Three different probe sizes (1.4 Å, 7.2 Å, and 10 Å radii) were utilized for investigating distinct types of binding site coverage (Grant et al., 2020).

Findings of the adapted SASA calculations illustrated insignificant binding site coverage by NAG ($1.318 \pm 5.79\%$) using the small probes (1.4 Å) (Figure 8). On the other hand, moderate surface occlusion was depicted on larger probes, 7.2 Å and 10.4 Å, where less than half of the binding site was covered by NAG ($5.502 \pm 6.40\%$ and $15.874 \pm 6.86\%$, respectively) throughout the MD simulation run. With several SASA trajectories having negative area coverage values, the simulated NAG molecule is considered to have a lower number of interactions with the binding site residues as well as non-complete coverage particularly with the 1.4 Å sized probes. Based on the above SASA findings, the binding site of interest has shown significant accessibility for small drug-like molecules as compared to peptidomimetic and antibody-related macromolecules during the simulation. Evaluation of the binding interactions for alacepril within the significant accessible hACE2-NAG binding site would identify the “hot spot” residues showing long-term hydrophilic interaction-related stabilization of the ligand within the binding site. Such information is highly relevant for understanding the evolution of ligand stability inside the protein pocket.

Binding Interaction Analysis

Investigating the hydrogen bond network interactions between the hACE2 residues and alacepril, over the 100 ns MD simulation, was considered crucial for understanding the observed conformational changes and stability of ligand–protein complexes. Using the VMD “Hydrogen bonds” tool, it was useful to explore the established ligand–protein hydrogen bond interactions and their relative frequencies (Humphrey et al., 1996). The cut-off values for hydrogen bond (Donor–H...Acceptor) distance and angle were assigned at 3.0 Å and 20°, respectively (de Souza et al., 2019; Albuquerque et al., 2020).

As expected, the hydrogen bond pairs between alacepril and either Asn90 or Gln96 were of the highest frequency, 55 and 37%, respectively, mediating the ligand–protein stabilization within the MD simulation interval 30–70 ns (Figure 9A). Following the 70 ns MD simulation frame, the latter polar interactions were lost as alacepril adopted the new shifted orientation/conformation at the transient opened cleft near the SARS-CoV-2 spike-protein recognition domain-III. On the other hand, the initial hydrogen bond pair Thr92:HG1-Alacepril:O4 was lost following the 10 ns of the MD simulation starts showing a minimal frequency of 4% (Figure 9B). This confers a limited contribution of Thr92 for the stabilization of the alacepril–hACE2 complex.

Surprisingly, the initial hydrogen bond interaction between alacepril and Asp30 was conserved up to 40 ns of the MD simulation. Despite limited fluctuations up to 8 Å hydrogen bond distances, the Asp30:OD1-Alacepril:H5 hydrogen bond pair was quite relevant particularly between the 57 ns and 65 ns MD simulation frames. Typically, Asp30 is reported as a key

polar residue for anchoring the SARS-CoV-2 spike glycoprotein on the receptor-binding domain of hACE2 through hydrogen bond interaction with Lys417 of the spike protein (Shang et al., 2020; Wang et al., 2020). Therefore, the depicted occurrence of hydrogen bonding between alacepril and Asp30 for more than 40 ns arose the promising role of alacepril to counter SARS-CoV-2/host entrance. It is suggested that polar anchoring of alacepril with any of the polar residues, involved at the S-protein-ACE2 connective interface, would probably impact both subdomains binding affinity (Hoffmann et al., 2020). Both suggested scenarios would halt the crucial stage of COVID-19 infection which is the virus–host membrane fusion and subsequent release of viral payload RNA into the host cytoplasm.

Binding-Free Energy Calculations

By illustrating the accessibility of the glycosylated site, we carried out an investigation of the differential binding affinity for the small molecules of interest and the N-glycan chain. Illustrating the potentiality of alacepril to compete with N-glycan for engaging the cavity near the glycan site would be beneficial to suggest an ability for disrupting the glycosylation process of the hACE2, leading to the modulation of hACE2-RBD interactions. Based on this, the following binding-free energy calculation was adopted to understand the nature of the alacepril–protein binding, explore the comparative alacepril/N-glycan-binding site affinity, and obtain more information concerning alacepril/residue contribution (Cavasotto, 2020). The MD-based Molecular Mechanics/Poisson Boltzmann Surface Area (MM/PBSA) approach was adopted for the designated binding-free energy calculations, using the “g_mmpbsa” tool on GROMACS. The approach accounts for more accurate ligand–protein affinity as compared to the most sophisticated flexible molecular docking technique (Kumari et al., 2014). Generally, MM/PBSA estimates binding-free energy as a contribution of several energy terms through these given Equations (Kumari et al., 2014):

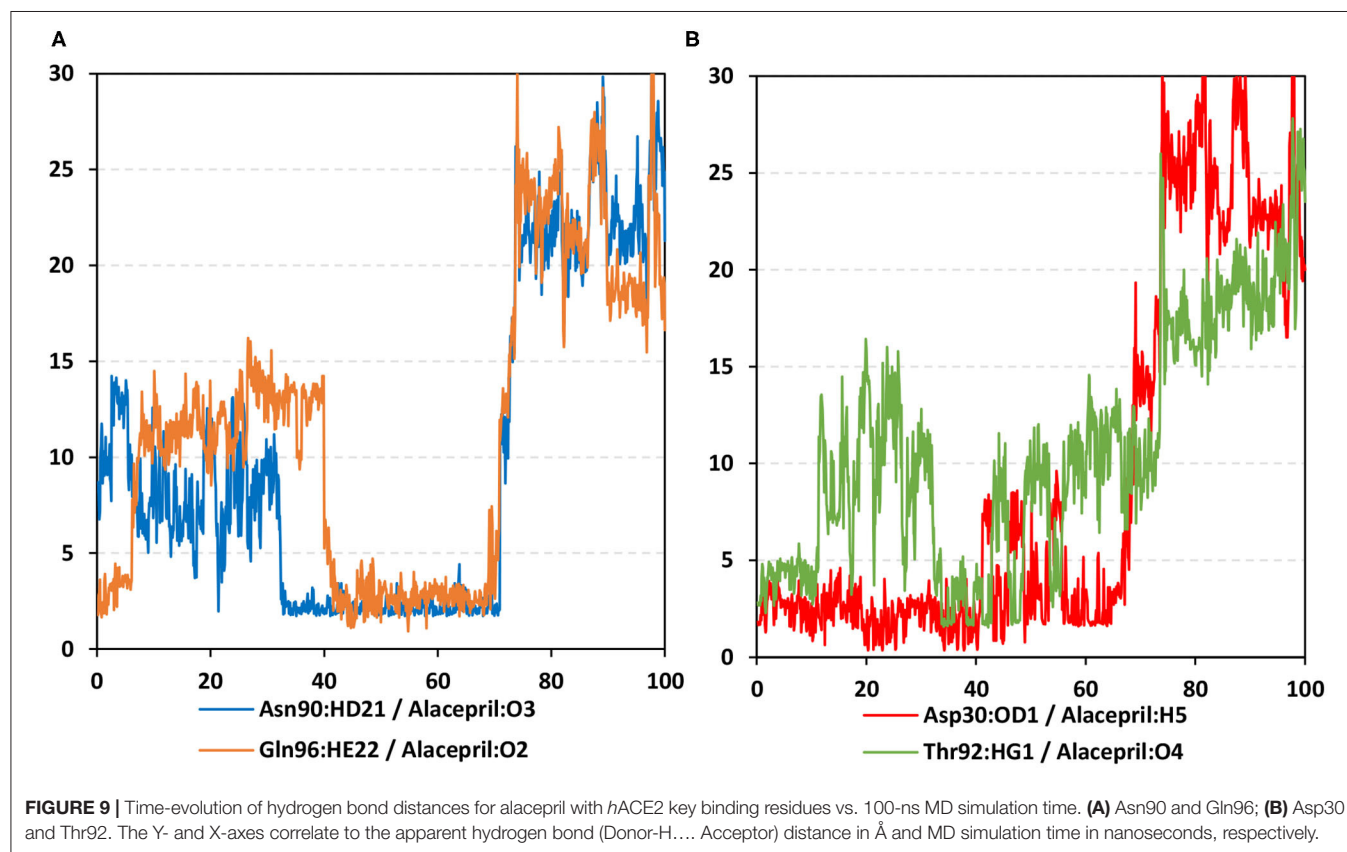
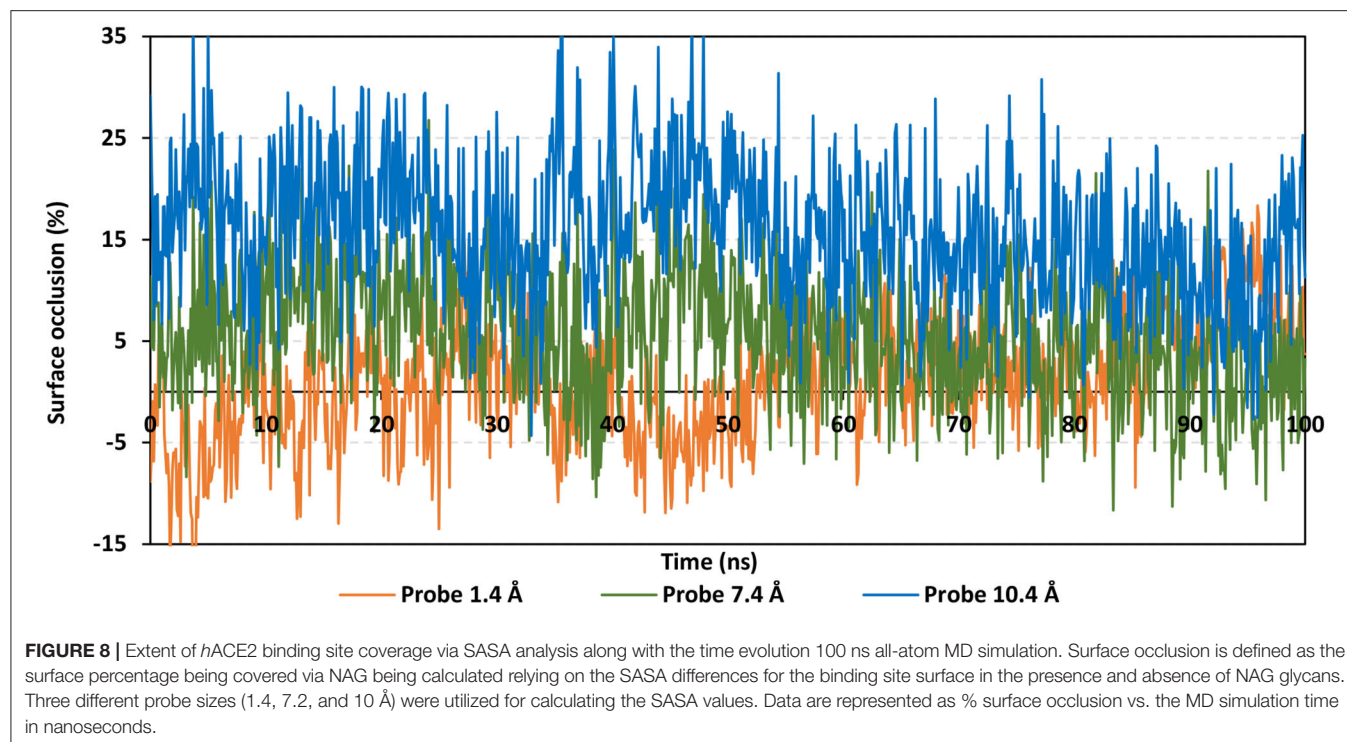
$$\Delta G_{\text{binding}} = G_{\text{complex}} - (G_{\text{ligand}} + G_{\text{protein}})$$

$$G_x = (E_{\text{MM}}) - TS + G_{\text{solvation}}$$

$$E_{\text{MM}} = E_{\text{bonded}} + (E_{\text{vdW}} + E_{\text{electrostatic}})$$

$$G_{\text{solvation}} = G_{\text{polar}} + \gamma \text{SASA} + b$$

where $\Delta G_{\text{binding}}$ is the binding-free energy correlating to ligand–protein binding where the higher negative energy values infer greater protein–ligand affinity. The energy terms G_{complex} , G_{protein} , and G_{ligand} are the total free energies of ligand–protein complex, isolated protein, and isolated ligand in the solvent, respectively. Vacuum MM potential energy (E_{MM}) together with the entropic contribution to free energy (TS) and free energy of solvation ($G_{\text{solvation}}$) provided the total free energy of protein, ligand, or ligand–protein complex (E_x). Terms T and S denote temperature and entropy, respectively, while E_{MM} was calculated based on molecular mechanics force-field parameters. Using the solvent-accessible surface area (SASA)–Non-polar Model, the $G_{\text{solvation}}$ energy term comprises polar and non-polar parts, where the latter was estimated via SASA and fitting constant (b). Finally, G_{polar} is to be solved from the Poisson–Boltzmann equation.



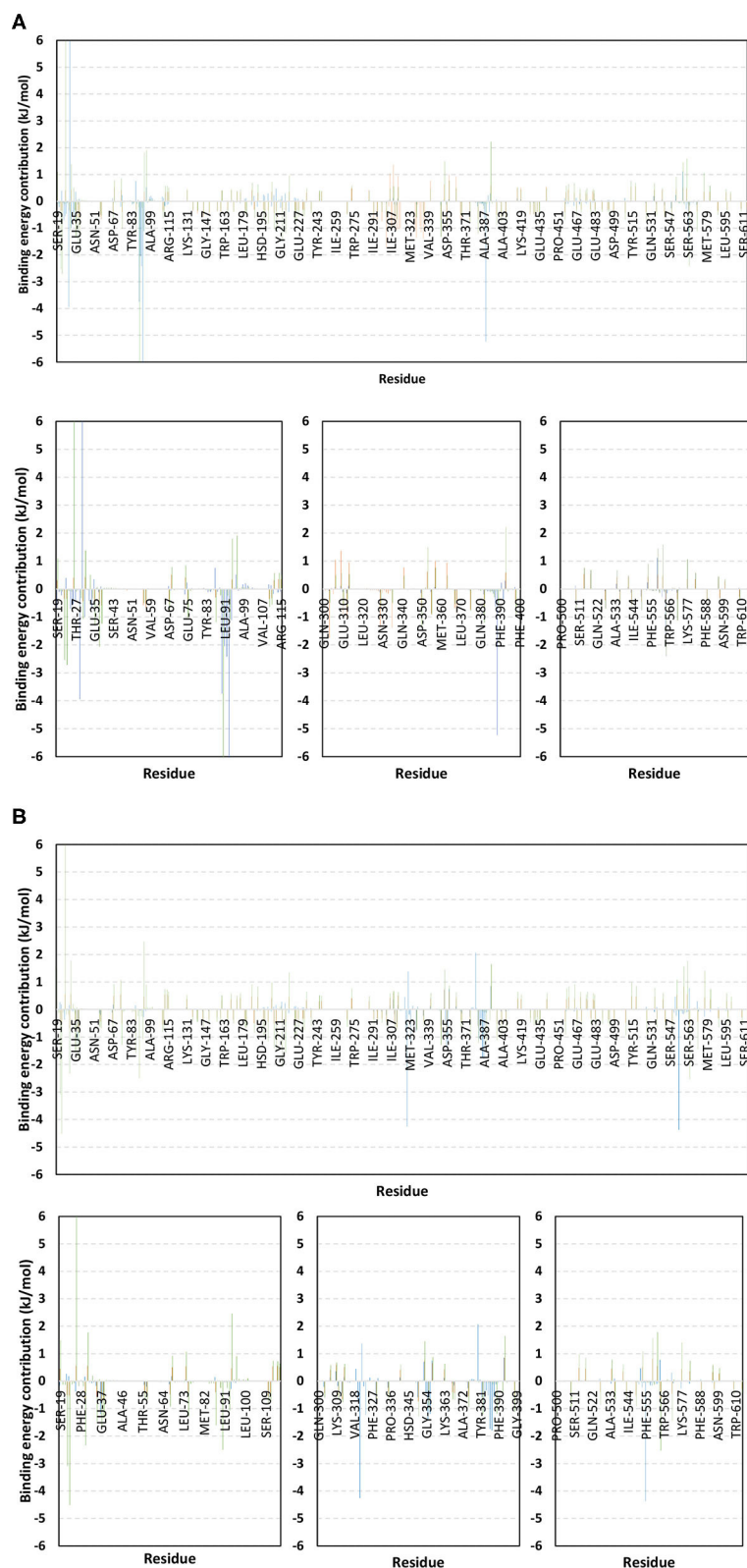


FIGURE 10 | Binding-free energy/residue decomposition illustrating the protein residue contribution at alacepril-*hACE2* protein complex $\Delta G_{\text{binding}}$ calculation. The residue-wise energy contributions across **(A)** 30–70 ns and **(B)** 80–100 ns MD simulation timeframes were represented in blue, brown, and green colored bars for alacepril, lisinopril, and NAG, respectively. Lower panels are expanded versions of three designated residue regions (19–115, 300–400, and 500–614) of the upper panels.

Typically, the binding-free energy should be estimated from the MD simulation trajectories depicting stabilized protein–ligand systems. Thus, the free energy calculation was adopted across the 30–70 ns and last 20 ns intervals where representative frames were extracted and saved to be enrolled within the calculation of each energy term. Adopting these specific time frames was rationalized by the above complex backbone RMSD analysis where equilibrated plateau tones were illustrated within the 30-to-70 ns and last 20 ns timeframe interval (**Figure 4**).

Interestingly, the $\Delta G_{\text{binding}}$ of the alacepril-*hACE2* complex was estimated at higher negative values around the 30–70 ns MD simulation interval as compared to that at 80–100 ns (-51.812 ± 17.494 kJ/mol vs. -37.898 ± 10.993 kJ/mol, respectively) (**Table 4**). A similar pattern was shown with lisinopril where its respective free-binding energy was lower across the last 20 ns MD simulation timeframes as compared to 30–70 ns ones. This less favored lisinopril–protein-binding energy came in good agreement with highly fluctuated RMSD and Rg tones near the end of the MD simulation. On the other hand, the $\Delta G_{\text{binding}}$ of the NAG-*hACE2* complex was of comparable values (-45.384 ± 47.279 and -48.729 ± 34.272 kJ/mol) across the two designated MD simulation time frames. The latter was expected since NAG depicted the steadiest complex RMSD trajectories along the whole MD simulation run.

Dissecting the furnished alacepril-*hACE2* $\Delta G_{\text{binding}}$ around both MD simulation intervals showed a preferential contribution of the hydrophobic van der Waal interactions as compared to that of the electrostatic energy term. However, the significant occupancy of the depicted hydrogen-bond interaction analysis around 30–70 ns MD simulation interval can suggest a somewhat balanced contribution between both energy terms. Moreover, the low electrostatic ($\Delta G_{\text{electrostatic}}$) contribution for 80–100 ns binding energy came in great agreement with the above bonding analysis findings where hydrogen bonding between alacepril and *hACE2* residues were limited as well as of minimal frequencies/occupancies. It worth mentioning that moderate $\Delta G_{\text{solvation}}$ energy term for alacepril at 30–70 ns interval (85.130 ± 25.313 kJ/mol) is considered favored for ligand–protein binding, as being balanced for the advent of the high electrostatic and Van der Waal energy contribution (-126.034 kJ/mol) afforded by alacepril scaffold. The latter compensated $\Delta G_{\text{solvation}}$ energy contribution further ensures the favored stability of the alacepril-*hACE2* complex since ligand binding is a solvent-substitution process.

For lisinopril, almost equal van de Waal/electrostatic energy contributions were assigned for the first MD simulation interval, whereas the electrostatic energy term depicted dominant free-binding energy contribution, nearly 3-fold higher than that of $\Delta G_{\text{Van der Waal}}$, within the last 20 ns. This came in great agreement with the above conformational analysis since the ligand showed an escape from the pocket side while becoming more solvent-exposed near the end of the MS simulation run. It worth mentioning that much higher $\Delta G_{\text{solvation}}$ values were depicted for the lisinopril-*hACE2* complex imposing a great penalty for the total free-binding energy calculation and ligand–protein binding. This could partially explain why lisinopril would exhibit dramatic conformational/orientation shift beyond 20 ns

as well as moving toward the solvent side while escaping the protein interface at the end of the MD simulation run.

Considering the NAG, the van der Waal energy term contribution was insignificant within the ligand–protein free-binding energy calculation depicting very low negative values across both designated MD simulation intervals. On the contrarily, the electrostatic energy term was of higher contributions across both MD intervals. This differential $\Delta G_{\text{Van der Waal}}/\Delta G_{\text{Electrostatic}}$ pattern could be reasoned for the chemical nature of NAG scaffold being rich in polar oxygen-based functionalities, which serve as excellent hydrogen bond donor/acceptor. The latter is expected to impose a higher energy penalty upon close contact with the hydrophobic residues lining the *hACE2*-pocket (Leu91, Val93, Ala387, and Leu560). Additionally, the polar sugar scaffold of NAG imposed high unfavored $\Delta G_{\text{solvation}}$, which negatively impacted the ligand–protein binding since such a process is a solvent-substitution approach.

For identifying the critical residues involved within the binding of ligands with *hACE2* protein, the residue-wise energy contribution to the obtained $\Delta G_{\text{binding}}$ was also estimated using *g_mmpbsa* (**Figure 10**) (Kumari et al., 2014). As a general observation, both alacepril and NAG depicted high residue-wise energy contribution near the C-terminal, particularly across the 30–70 ns MD simulation interval (**Figure 10A**). The latter is in great agreement with the previously discussed ΔRMSF analysis where the C-terminal free residues and their vicinal amino acids showed significant immobility with high positive values ($\Delta \text{RMSF} > 0.30$ Å). This further confirms the significant stabilized binding of alacepril within the *hACE2*-NAG pocket within this simulation interval. Comprehensive analysis of residue-wise energy contribution for alacepril across 30–70 ns showed significant contributions by Asn90 and Thr93, conferring their key role for stabilizing the alacepril–protein complex. Moreover, several key residues, which have participated in relevant within the initial docking analysis, showed significant contributions to the calculated $\Delta G_{\text{binding}}$. The high energy contribution by Asp30 came in great agreement with the previous hydrogen bonding analysis as the Asp30:OD1–Alacepril:H5 hydrogen bond pair was conserved for significant MD simulation frames. This high energy contribution further ensures the promising antiviral activity of alacepril in countering the SARS-CoV-2/host entrance through hampering the polar interaction role of Asp30 within *hACE2* spike-protein annealing and anchoring (Hoffmann et al., 2020). Other initial *hACE2* binding site residue and vicinal amino acids depicted significant contribution within alacepril–complex free-binding, including Leu29, Lue91, Val93, Pro389, and Glu312. These dominant non-polar energy contributions further confirm the superiority of $\Delta G_{\text{Van der Waal}}$ energy as compared $\Delta G_{\text{Velectrostatic}}$ energy term.

Moving toward the MD simulation 80–100 ns, residues of the transient opened cleft which is at proximity to the SARS-CoV-2 spike-protein recognition domain-III have depicted significant free-binding energy contributions (**Figure 10B**). The latter involves Glu37, Pro321, Asn322, Thr324, Asp355, Phe356, Gln380, Met383, Ala386, Ala387, and Phe555 residues. Notably, the highest energy contributions (-4.259 and -4.373 kJ/mol) were assigned for the aromatic hydrophobic residues (Pro321 and

TABLE 4 | Binding-free energies calculations ($\pm$ standard deviation; SD) for the investigated ligand-*hACE2* protein complexes.

Energy terms (kJ/mol $\pm$ SD)	30–70 ns			80–100 ns		
	Alacepril	Lisinopril	NAG	Alacepril	Lisinopril	NAG
$\Delta G_{\text{Van der Waal}}$	-101.954 ± 19.491	-121.265 ± 7.925	-3.784 ± 44.849	-76.329 ± 16.017	-30.148 ± 30.735	-7.143 ± 52.568
$\Delta G_{\text{Electrostatic}}$	-24.080 ± 21.066	-134.936 ± 45.623	-142.512 ± 26.408	-9.010 ± 12.093	-99.892 ± 188.971	-137.297 ± 28.240
$\Delta G_{\text{Solvation (Polar)}}$	85.130 ± 25.313	231.848 ± 0.866	109.767 ± 18.873	59.004 ± 21.113	111.265 ± 11.102	104.107 ± 43.007
$\Delta G_{\text{Solvation (SASA)}}$	-10.908 ± 1.736	-14.900 ± 0.929	-8.855 ± 2.190	-11.563 ± 2.503	-4.288 ± 4.118	-8.396 ± 2.045
$\Delta G_{\text{Binding}}$	-51.812 ± 17.494	-39.255 ± 49.430	-45.384 ± 47.279	-37.898 ± 10.993	-23.063 ± 21.715	-48.729 ± 34.272

Phe555, respectively), while moderate contributions (-1.074 to -1.798 kJ/mol) were assigned to Phe356, Met383, Ala386, and Ala387 residues. The latter confers dominance of hydrophobic interactions ($\Delta G_{\text{Van der Waal}}$) for stabilizing alacepril at *hACE2* transient opened pit.

Moving toward the glycosylated *hACE2*, a similar pattern of residue-wise energy contributions was depicted across the 30–70 ns and 80–100 ns intervals. Both pocket and vicinal residues showed significant contribution within the NAG $\Delta G_{\text{binding}}$ calculation. Across the 30–70 ns time frame, the highest energy contributions were assigned to Lys26 and Asn90 (6.184 and -6.867 kJ/mol, respectively), conferring their key role in NAG stabilization within the protein pocket. Other pocket/vicinal residues such as Glu22, Glu23, Asp30, Glu35, Glu37, Asp38, Glu87, Gln96, Asp213, Asp216, Arg393, Glu564, and Glu571, showed moderate energy contributions (-1.081 to -2.545 kJ/mol). Owing to the hydrophilic nature of these residues, an explanation of the dominant $\Delta G_{\text{electrostatic}}$ energy term contribution within the NAG free-binding energy calculation is to be rationalized. Regarding the last 20 ns interval, a general trend of increased residue-wise energy contributions was depicted for several residues, particularly those at proximity to C-terminal (Ser19, Glu22, Glu23, Lys26, Asp30, Glu35, Glu37, and Asp38). Nearly a 2-fold increase in the Lys26 energy contribution was depicted at 80–100 ns as compared to the 30–70 ns interval.

Other N-terminal pocket residues (Arg559, Lys562, Glu564, Glu571, and Lys577) showed a similar trend of increased residue-wise energy contributions. Contrarily, other pocket residues including Asp90, Val93, Gln96, and Arg393 showed lower energy contributions at the last 20 ns time with the highest descent for Asp90 being from -6.867 to -2.573 kJ/mol. Such differential pattern of residue-wise energy contribution shift came in great agreement with the previously described conformational analysis where a significant change in NAG respective position was depicted at the 100 ns frame. This further confirms the assumption that NAG is not fully occluding the binding site of interest the thing that could make it at least partially accessible across the designated MD simulation. Therefore, small druggable molecules, like alacepril, could manage to accommodate the *hACE2*-NAG binding site of the glycosylated protein effectively.

Considering the last investigated complex, the lisinopril residue-wise energy contribution showed minimal values for the C-terminal free residues and their vicinal amino acids. This came in adherence with the above ΔRMSF analysis confirming the escape of lisinopril from the initial *hACE2*-NAG binding site.

Significant energy contribution was depicted for Asp299, Asp303, Arg306, Ile307, Phe308, Lys309, Glu310, Glu312, Lys313, Phe314, Phe315, Trp328, Glu329, Met332, Leu333, Asp335, and Pro336. In worth noting that these latter residues comprise the distant pocket, or its vicinal residues, being accommodated by lisinopril throughout the previously described conformational analysis along with the 30–70 ns interval. Interestingly, the balanced hydrophilic/hydrophobic nature of these residues could explain the comparable contributions of $\Delta G_{\text{electrostatic}}$ and $\Delta G_{\text{Van der Waal}}$ energy terms within the lisinopril-protein binding throughout 30–70 ns. As expected, a significant decrease within the latter residue-wise energy contribution profile was observed across the 80–100 ns trajectories, since lisinopril showed instability and dramatic shift toward the solvent side.

CONCLUSION

A total of 14 ACEIs were subjected to virtual screening molecular docking against the spike protein of COVID-19. The tested drugs exhibited variable degrees of affinities toward the COVID-19 spike protein comparing to the native inhibitor. Alacepril and lisinopril were found to interact with COVID-19 spike protein by exhibiting the most acceptable rmsd_refine values and the best binding affinity through forming a strong hydrogen bond with Asn90, which is assumed to be essential for the activity, as well as significant extra interactions with other receptor-binding residues. Throughout the all-atom 100 ns MD simulation, alacepril depicted superior stability at the *hACE2* binding site for more than 70 ns, where the solvation energy was greatly compensated by the electrostatic and Van der Waal binding energies. SASA calculations for *hACE2* pocket in the presence and absence of glycan showed significant accessibility of the pocket for small drug-like molecules like alacepril. Moreover, alacepril mediated a stabilized favored hydrogen bond interaction with Asn30 which was conserved for significant MD simulation intervals. Depicting this favored hydrogen bond pair as well as the reported key role in *hACE2*/SARS-CoV-2 spike-protein association introduces the promising action of alacepril to counter COVID-19/host entrance and subsequent release of viral payload RNA into the host cytoplasm through hampering *hACE2* spike-protein annealing and anchoring. Based on the furnished evidence, these drugs are recommended to be tested clinically for proposed activity against COVID-19. They may be tested either alone or in combinations. Also, our results may give a clear spot about SAR required for the spike-protein targeting drug

to facilitate the future design and synthesis of new candidates against 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

AA-K: docking studies. SE and KD: molecular dynamic. IE: research plan. EE: data collection. AM: revision editing and references. MD: writing manuscript. All authors contributed to the article and approved the submitted version.

REFERENCES

- Albuquerque, S. O., Barros, T. G., Dias, L. R. S., Lima, C. H. D. S., Azevedo, P. H. R. A., Flores-Junior, L. A. P., et al. (2020). Biological evaluation and molecular modeling of peptidomimetic compounds as inhibitors for O-GlcNAc transferase (OGT). *Eur. J. Pharm. Sci.* 154:105510. doi: 10.1016/j.ejps.2020.105510
- Al-Karmalawy, A. A., Alnajjar, R., Dahab, M., Metwaly, A., and Eissa, I. (in press). Molecular docking and dynamics simulations reveal the potential of anti-HCV drugs to inhibit COVID-19 main protease. *Pharm Sci.* doi: 10.34172/PS.2021.3
- Al-Karmalawy, A. A., and Khattab, M. (2020). Molecular modelling of mebendazole polymorphs as a potential colchicine binding site inhibitor. *N. J. Chem.* 44, 13990–13996. doi: 10.1039/D0NJ02844D
- Alnajjar, R., Mostafa, A., Kandeil, A., and Al-Karmalawy, A. A. (2020). Molecular docking, molecular dynamics, and *in vitro* studies reveal the potential of angiotensin II receptor blockers to inhibit the COVID-19 main protease. *Heliyon* 6:e05641. doi: 10.1016/j.heliyon.2020.e05641
- Ashburn, T. T., and Thor, K. B. (2004). Drug repositioning: identifying and developing new uses for existing drugs. *Nat. Rev. Drug Disc.* 3, 673–683. doi: 10.1038/nrd1468
- Banerjee, A., Santra, D., and Maiti, S. (2020). Energetics and IC50 based epitope screening in SARS CoV - 2 (COVID 19) spike protein by immunoinformatic analysis implicating for a suitable vaccine development. *J. Trans. Med.* 18, 1–14. doi: 10.1186/s12967-020-02435-4
- Benson, N. C., and Daggett, V. (2013). A comparison of multiscale methods for the analysis of molecular dynamics simulations. *J. Phys. Chem. B* 116, 8722–8731. doi: 10.1021/jp302103t
- Best, R. B., Zhu, X., Shim, J., Lopes, P. E., Mittal, J., Feig, M., et al. (2013). Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone ϕ , ψ and sidechain χ_1 and χ_2 dihedral angles. *J. Chem. Theory Comput.* 8, 3257–3273. doi: 10.1021/ct300400x
- Calligari, P., Bobone, S., Ricci, G., and Bocedi, A. (2020). Molecular investigation of SARS-CoV-2 proteins and their interactions with antiviral drugs. *Viruses* 12:445. doi: 10.3390/v12040445
- Cavasotto, C. N. (2020). Binding free energy calculation using quantum mechanics aimed for drug lead optimization. *Quant. Mech. Drug Disc.* 2114, 257–268. doi: 10.1007/978-1-0716-0282-9_16
- 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
- Culp, W. C. (2021). Coronavirus disease 2019 situation update. *World Health Organ.* 14:e01218. doi: 10.1213/XAA.00000000000001218
- da Silva, T. U., Pougy, K. C., Albuquerque, M. G., da Silva Lima, C. H., and Machado, S. P. (2020). Development of parameters compatible with the CHARMM36 force field for [Fe4S4] 2+ clusters and molecular dynamics simulations of adenosine-5'-phosphosulfate reductase in GROMACS 2019. *J. Biomol. Struct. Dyn.* 13, 1–11. doi: 10.1080/07391102.2020.1847687

ACKNOWLEDGMENTS

The simulations in this work were performed at King Abdulaziz University's High-Performance Computing Center (Aziz Supercomputer) (<http://hpc.kau.edu.sa>), and the authors, therefore, acknowledge with thanks the center for technical support.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fchem.2021.661230/full#supplementary-material>

- Dahab, M. A., Hegazy, M. M., and Abbass, H. S. (2020). Hordatinas as a potential inhibitor of COVID - 19 main protease and RNA polymerase: an *in silico* approach. *Nat. Prod. Bioprospect.* 10, 453–462. doi: 10.1007/s13659-020-00275-9
- Darden, T., York, D., and Pedersen, L. (1998). Particle mesh Ewald: An N² log (N) method for Ewald sums in large systems. *J. Chem. Phys.* 98, 10089–10092. doi: 10.1063/1.464397
- de Andrade, J., Gonçalves, P. F. B., and Netz, P. A. (2020). Why does the novel coronavirus spike protein interact so strongly with the human ACE2? A thermodynamic answer. *ChemBioChem* 22, 865–875. doi: 10.1002/cbic.202000455
- de Souza, A. S., Pacheco, B. D. C., Pinheiro, S., Muri, E. M. F., Dias, L. R. S., Lima, C. H. S., et al. (2019). 3-Acyltetramic acids as a novel class of inhibitors for human kallikreins 5 and 7. *Bioorg. Med. Chem. Lett.* 29, 1094–11098. doi: 10.1016/j.bmcl.2019.02.031
- Delano, W. L. (2002). Pymol: an open-source molecular graphics tool. *CCP4 Newsl. Protein Crystallogr.* 40, 82–92.
- Devaux, C. A., Rolain, J. M., and Raoult, D. (2020). ScienceDirect ACE2 receptor polymorphism: Susceptibility to SARS-CoV-2, hypertension, multi-organ failure, and COVID-19 disease outcome. *J. Microbiol. Immunol. Infect.* 53, 425–435. doi: 10.1016/j.jmii.2020.04.015
- Dézi, L. (2000). Fibrinolytic actions of ACE inhibitors: a significant plus beyond antihypertensive therapeutic effects. *Elsevier Sci.* 47, 642–644. doi: 10.1016/S0008-6363(00)00163-2
- Donoghue, M., Hsieh, F., Baronas, E., Godbout, K., Gosselin, M., Stagliano, N., et al. (2000). A novel angiotensin-converting enzyme-related carboxypeptidase (ACE2) converts angiotensin I to angiotensin 1-9. *Circ. Res.* 87:E1–9. doi: 10.1161/01.RES.87.5.e1
- Du, L., He, Y., Zhou, Y., Liu, S., Zheng, B. J., and Jiang, S. (2009). The spike protein of SARS-CoV — a target for vaccine and therapeutic development. *Nat. Rev. Microbiol.* 7, 226–236. doi: 10.1038/nrmicro2090
- Eliaa, S. G., Al-Karmalawy, A. A., Saleh, R. M., and Elshal, M. F. (2020). Empagliflozin and doxorubicin synergistically inhibit the survival of triple-negative breast cancer cells via interfering with the mtor pathway and inhibition of calmodulin: *in vitro* and molecular docking studies. *ACS Pharmacol. Trans. Sci.* 3, 1330–1338. doi: 10.1021/acspstsci.0c00144
- Ghanem, A., Emara, H. A., Muawia, S., Abd El Maksoud, A. I., Al-Karmalawy, A. A., and Elshal, M. F. (2020). Tanshinone IIA synergistically enhances the antitumor activity of doxorubicin by interfering with the PI3K/AKT/mTOR pathway and inhibition of topoisomerase II: *in vitro* and molecular docking studies. *N. J. Chem.* 44, 17374–17381. doi: 10.1039/D0NJ04088F
- Golo, V., and Shaitan, K. (2002). Dynamic attractor for the Berendsen thermostat and the slow dynamics of biomacromolecules. *Biofizika* 47, 611–617.
- Grant, O. C., Montgomery, D., Ito, K., and Woods, R. J. (2020). Analysis of the SARS - CoV - 2 spike protein glycan shield reveals implications for immune recognition. *Sci. Rep.* 10, 1–11. doi: 10.1038/s41598-020-71748-7
- Groß, S., Jahn, C., Cushman, S., Bär, C., and Thum, T. (2020). SARS-CoV-2 receptor ACE2-dependent implications on the cardiovascular system: from

- basic science to clinical implications. *J. Mol. Cell. Cardiol.* 144, 47–53. doi: 10.1016/j.yjmcc.2020.04.031
- Gurwitz, D. (2020). Angiotensin receptor blockers as tentative SARS-CoV-2 therapeutics. *Drug Dev. Res.* 81, 537–540. doi: 10.1002/ddr.21656
- Helal, M. A., Shouman, S., Abdelwaly, A., Elmeharath, A. O., Essawy, M., Sayed, S. M., et al. (2020). Molecular basis of the potential interaction of SARS-CoV-2 spike protein to CD147 in COVID-19 associated-lymphopenia. *J. Biomol. Struct. Dyn.* 16, 1–11. doi: 10.1080/07391102.2020.1822208
- Hess, B., Bekker, H., Berendsen, H. J., and Fraaije, J. G. (1997). LINCS: a linear constraint solver for molecular simulations. *J. Comput. Chem.* 18, 1463–1472. doi: 10.1002/(SICI)1096-987X(199709)18:12<1463::AID-JCC4>3.0.CO;2-H
- 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
- 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
- Izadi, S., Anandakrishnan, R., and Onufriev, A. V. (2014). Building water models: a different approach. *J. Phys. Chem. Lett.* 5, 3863–3871. doi: 10.1021/jz501780a
- Jairajpuri, D. S., Hussain, A., Nasreen, K., Mohammad, T., Anjum, F., Tabish Rehman, M., et al. (2021). Identification of natural compounds as potent inhibitors of SARS-CoV-2 main protease using combined docking and molecular dynamics simulations. *Saudi J. Biol. Sci.* 28, 2423–2431. doi: 10.1016/j.sjbs.2021.01.040
- Karplus, M., and Petsko, G. A. (1990). Molecular dynamics simulations in biology. *Nature* 347, 631–639. doi: 10.1038/347631a0
- Khan, S., Fakhar, Z., Hussain, A., Ahmad, A., Jairajpuri, D. S., Alajmi, M. F., et al. (2020). Structure-based identification of potential SARS-CoV-2 main protease inhibitors. *J. Biomol. Struct. Dyn.* 19, 1–14. doi: 10.1080/07391102.2020.1848634
- Khattab, M., and Al-Karmalawy, A. A. (2021). Revisiting activity of some nocodazole analogues as a potential anticancer drugs using molecular docking and DFT calculations. *Front. Chem.* 9:92. doi: 10.3389/fchem.2021.628398
- Khattab, M., Khalifa, M., and Al-Karmalawy, A. A. (2021). Revisiting activity of some glucocorticoids as a potential inhibitor of SARS-CoV-2 main protease: theoretical study. *RSC Adv.* 11, 10027–10042. doi: 10.1039/D0RA10674G
- Kumar, D., Chauhan, G., Kalra, S., Kumar, B., and Gill, M. S. (2020). Bioorganic chemistry A perspective on potential target proteins of COVID-19: comparison with SARS-CoV for designing new small molecules. *Bioorg. Chem.* 104:104326. doi: 10.1016/j.bioorg.2020.104326
- Kumari, R., Kumar, R. O., Consortium, S. D. D., and Lynn, A. (2014). g_mmpbsa-A GROMACS tool for high-throughput MM-PBSA calculations. *J. Chem. Inf. Model.* 5, 1951–1962. doi: 10.1021/ci500020m
- Li, W., Zhang, C., Sui, J., Kuhn, J. H., Moore, M. J., Luo, S., et al. (2005). Receptor and viral determinants of SARS-coronavirus adaptation to human ACE2. *EMBO J.* 24, 1634–1643. doi: 10.1038/sj.emboj.7600640
- Likić V. A., Gooley, P. R., Speed, T. P., and Strehler, E. E. (2005). A statistical approach to the interpretation of molecular dynamics simulations of calmodulin equilibrium dynamics. *Protein Sci.* 14, 2955–2963. doi: 10.1110/ps.051681605
- Liu, K., Watanabe, E., and Kokubo, H. (2017). Exploring the stability of ligand binding modes to proteins by molecular dynamics simulations. *J. Comput. Aided Mol. Des.* 31, 201–211. doi: 10.1007/s10822-016-0005-2
- Lover, A. A. (2020). Quantifying treatment effects of hydroxychloroquine and azithromycin for COVID-19: a secondary analysis of an open label non-randomized clinical trial. *MedRxiv [Preprint]*. doi: 10.1101/2020.03.22.20040949
- Mohammad, T., Shamsi, A., Anwar, S., Umair, M., Hussain, A., Rehman, M. T., et al. (2020). Identification of high-affinity inhibitors of SARS-CoV-2 main protease : Towards the development of effective COVID-19 therapy. *Virus Res.* 288:198102. doi: 10.1016/j.virusres.2020.198102
- Monteil, V., Kwon, H., Prado, P., Hagelkrüys, A., Wimmer, R. A., Stahl, M., et al. (2020). Inhibition of SARS-CoV-2 infections in engineered human tissues using clinical-grade soluble human ACE2. *Cell* 181, 905–913.e7. doi: 10.1016/j.cell.2020.04.004
- Napoleone, E., Di Santo, A., Camera, M., Tremoli, E., and Lorenzet, R. (2000). Angiotensin-converting enzyme inhibitors downregulate tissue factor synthesis in monocytes. *Circ. Res.* 86, 139–143. doi: 10.1161/01.RES.86.2.139
- Páll, S., and Hess, B. (2013). A flexible algorithm for calculating pair interactions on SIMD architectures. *Comput. Phys. Commun.* 184, 2641–2650. doi: 10.1016/j.cpc.2013.06.003
- Pirolli, D., Sciandra, F., Bozzi, M., Giardina, B., Brancaccio, A., and De Rosa, M. C. (2014). Insights from molecular dynamics simulations: structural basis for the V567D mutation-induced instability of zebrafish alpha-dystroglycan and comparison with the murine model. *PLoS ONE* 9:e103866. doi: 10.1371/journal.pone.0103866
- Prajapat, M., Sarma, P., Shekhar, N., Avti, P., Sinha, S., Kaur, H., et al. (2020). Drug targets for corona virus: a systematic review. *Indian J. Pharmacol.* 52, 56–65. doi: 10.4103/ijp.IJP_115_20
- Ross, G. A., Rustenburg, A. S., Grinaway, P. B., Fass, J., and Chodera, J. D. (2019). Biomolecular simulations under realistic macroscopic salt conditions. *J. Phys. Chem. B* 122, 5466–5486. doi: 10.1021/acs.jpcc.7b11734
- Rothe, C., Schunk, M., Sothmann, P., Bretzel, G., Froeschl, G., and Wallrauch, C. (2020). Correspondence transmission of 2019-nCoV infection from an asymptomatic contact in Germany. *N. Engl. J. Med.* 382, 970–971. doi: 10.1056/NEJMc2001468
- Samra, R. M., Soliman, A. F., Zaki, A. A., Ashour, A., Al-Karmalawy, A. A., Hassan, M. A., et al. (2021). Bioassay-guided isolation of a new cytotoxic ceramide from *Cyperus rotundus* L. *South Afr. J. Botany* 139, 210–216. doi: 10.1016/j.sajb.2021.02.007
- Sanders, J. M., Monogue, M. L., Jodlowski, T. Z., and Cutrell, J. B. (2020). Pharmacologic treatments for coronavirus disease 2019 (COVID-19) a review. *JAMA* 323, 1824–1836. doi: 10.1001/jama.2020.6019
- Schreiner, W., Karch, R., Knapp, B., and Ilieva, N. (2012). Relaxation estimation of RMSD in molecular dynamics immunosimulations. *Comput. Math. Methods Med.* 2012:173521. doi: 10.1155/2012/173521
- Shang, J., Ye, G., Shi, K., Wan, Y., Aihara, H., and Li, F. (2020). Structural basis of receptor recognition by SARS-CoV-2. *Nature*. 581, 221–224. doi: 10.1038/s41586-020-2179-y
- Skeggs, L. T., Kahn, J. R., and Shumway, N. P. (1955). The preparation and function of the hypertension converting enzyme. *J. Exp. Med.* 103:295. doi: 10.1084/jem.103.3.295
- Slivski, G., Kothiwale, S., Meiler, J., and Lowe, E. W. (2014). Computational methods in drug discovery. *Pharmacol. Rev.* 66, 334–395. doi: 10.1124/pr.112.007336
- Talevi, A., and Bellera, C. L. (2020). Expert opinion on drug discovery challenges and opportunities with drug repurposing: finding strategies to find alternative uses of therapeutics. *Expert Opin. Drug Disc.* 15, 397–401. doi: 10.1080/17460441.2020.1704729
- Tubtle, S. C., Anwar, J., and Gale, J. D. (2004). An approach to developing a force field for molecular simulation of martensitic phase transitions between phases with subtle differences in energy and structure. *J. Am. Chem. Soc.* 126, 396–405. doi: 10.1021/ja0356131
- Urbanowicz, R. A., Wang, R., Schiel, J. E., Keck, Z. Y., Kerzic, M. C., Lau, P., et al. (2019). Antigenicity and immunogenicity of differentially glycosylated hepatitis C virus E2 envelope proteins expressed in mammalian and insect cells. *J. Virol.* 93:e01403–e01418. doi: 10.1128/JVI.01403-18
- Vaduganathan, M., Vardeny, O., Michel, T., McMurray, J. J., Pfeffer, M. A., and Solomon, S. D. (2020). Renin-angiotensin-aldosterone system inhibitors in patients with Covid-19. *N. Engl. J. Med.* 382, 1653–1659. doi: 10.1056/NEJMs2005760
- Vanommeslaeghe, K., Hatcher, E., Acharya, C., Kundu, S., Zhong, S., Shim, J., et al. (2009). CHARMM general force field: a force field for drug-like molecules compatible with the CHARMM All-atom additive biological force fields. *J. Comput. Chem.* 31, 671–690. doi: 10.1002/jcc.21367
- Vilar, S., Cozza, G., and Moro, S. (2008). Medicinal chemistry and the molecular operating environment (MOE): application of QSAR and molecular docking to drug discovery. *Curr. Top. Med. Chem.* 8, 1555–1572. doi: 10.2174/156802608786786624
- Wan, Y., Shang, J., Graham, R., Baric, R. S., and Li, F. (2020). Receptor recognition by the novel coronavirus from Wuhan: an analysis based on decade-long structural studies of SARS coronavirus. *J. Virol.* 94, 1–7. doi: 10.1128/JVI.00127-20

- Wang, Q., Zhang, Y., Wu, L., Niu, S., Song, C., Zhang, Z., et al. (2020). 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
- 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
- Wu, C.-H., Mohammadmoradi, S., Chen, J. Z., Sawada, H., Daugherty, A., and Lu, H. S. (2018). Renin-angiotensin system and cardiovascular functions. *Arterioscler. Thromb. Vasc. Biol.* 38, e108–e116. doi: 10.1161/ATVBAHA.118.311282
- Yeu, Y., Yoon, Y., and Park, S. (2015). Molecular BioSystems protein localization vector propagation: a method for improving the accuracy of drug repositioning. *Mol. BioSyst.* 11, 2096–2102. doi: 10.1039/C5MB00306G
- Zaki, A. A., Al-Karmalawy, A. A., El-Amier, Y. A., and Ashour, A. (2020). Molecular docking reveals the potential of cleome amblyocarpa isolated compounds to inhibit COVID-19 virus main protease. *N. J. Chem.* 44, 16752–16758. doi: 10.1039/D0NJ03611K
- Zhang, P., Zhu, L., Cai, J., Lei, F., Qin, J.-J., Xie, J., et al. (2020). Association of inpatient use of angiotensin-converting enzyme inhibitors and angiotensin II receptor blockers with mortality among patients with hypertension hospitalized with COVID-19. *Circ. Res.* 126, 1671–1681. doi: 10.1161/CIRCRESAHA.120.317134

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 Al-Karmalawy, Dahab, Metwaly, Elhady, Elkaeed, Eissa and Darwish. 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.



Using Network Pharmacology and Molecular Docking to Explore the Mechanism of Shan Ci Gu (*Cremastra appendiculata*) Against Non-Small Cell Lung Cancer

Yan Wang¹, Yunwu Zhang^{2*}, Yujia Wang¹, Xinyao Shu¹, Chaorui Lu¹, Shiliang Shao¹, Xingting Liu³, Cheng Yang⁴, Jingsong Luo³ and Quanyu Du^{1*}

¹Hospital of Chengdu University of Traditional Chinese Medicine, Chengdu, China, ²Department of Biochemistry and Molecular Biology, West China School of Basic Medical Sciences and Forensic Medicine, Sichuan University, Chengdu, China, ³Chengdu University of Traditional Chinese Medicine, Chengdu, China, ⁴Faculty of Geosciences and Environmental Engineering, Southwest Jiaotong University, Chengdu, China

OPEN ACCESS

Edited by:

Zhongjie Liang,
Soochow University, China

Reviewed by:

Guang Hu,
Soochow University, China
Andrei I. Khlebnikov,
Tomsk Polytechnic University, Russia

*Correspondence:

Yunwu Zhang
zhangyunwu@stu.scu.edu.cn
Quanyu Du
QuanyuDu@163.com

Specialty section:

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

Received: 19 March 2021

Accepted: 20 May 2021

Published: 09 June 2021

Citation:

Wang Y, Zhang Y, Wang Y, Shu X,
Lu C, Shao S, Liu X, Yang C, Luo J and
Du Q (2021) Using Network
Pharmacology and Molecular Docking
to Explore the Mechanism of Shan Ci
Gu (*Cremastra appendiculata*) Against
Non-Small Cell Lung Cancer.
Front. Chem. 9:682862.
doi: 10.3389/fchem.2021.682862

Background: In recent years, the incidence and mortality rates of non-small cell lung cancer (NSCLC) have increased significantly. Shan Ci Gu is commonly used as an anticancer drug in traditional Chinese medicine; however, its specific mechanism against NSCLC has not yet been elucidated. Here, the mechanism was clarified through network pharmacology and molecular docking.

Methods: The Traditional Chinese Medicine Systems Pharmacology database was searched for the active ingredients of Shan Ci Gu, and the relevant targets in the Swiss Target Prediction database were obtained according to the structure of the active ingredients. GeneCards were searched for NSCLC-related disease targets. We obtained the cross-target using VENN to obtain the core targets. The core targets were imported into the Search Tool for the Retrieval of Interacting Genes/Proteins database, and Cytoscape software was used to operate a mesh chart. R software was used to analyze the Gene Ontology biological processes (BPs) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment. The core targets and active compounds were molecularly docked through Auto-Dock Vina software to predict the detailed molecular mechanism of Shan Ci Gu for NSCLC treatment. We did a simple survival analysis with hub gene to assess the prognosis of NSCLC patients.

Results: Three compounds were screened to obtain 143 target genes and 1,226 targets related to NSCLC, of which 56 genes were related to NSCLC treatment. Shan Ci Gu treatment for NSCLC involved many BPs and acted on main targets including epidermal growth factor receptor (EGFR), ESR1, and SRC through signaling pathways including the

Abbreviations: TCMSP, Traditional Chinese Medicine Systems Pharmacology; OB, Oral bioavailability; DL, Drug likeness; NSCLC, Non-small cell lung cancer; PPI, Protein-protein interaction; STRING, Search Tool for the Retrieval of Interacting Genes/Proteins database; GO, Gene Ontology; BPs, Biological processes; MF, Molecular function; CC, Cellular component; KEGG, Kyoto Encyclopedia of Genes and Genomes; PDB, Protein Data Bank; SDF, Standard Delay Format; NED, Neuro-endocrine dedifferentiation.

endocrine resistance, EGFR tyrosine kinase inhibitor resistance, and ErbB signaling pathways. Shan Ci Gu might be beneficial for treating NSCLC by inhibiting cell proliferation and migration. Molecular docking revealed that the active compounds β -sitosterol, stigmasterol, and 2-methoxy-9,10-dihydrophenanthrene-4,5-diol had good affinity with the core target genes (EGFR, SRC, and ESR1). Core targets included EGFR, SRC, ESR1, ERBB2, MTOR, MCL1, matrix metalloproteinase 2 (MMP2), MMP9, KDR, and JAK2. Key KEGG pathways included endocrine resistance, EGFR tyrosine kinase inhibitor resistance, ErbB signaling, PI3K-Akt signaling, and Rap1 signaling pathways. These core targets and pathways have an inhibitory effect on the proliferation of NSCLC cells.

Conclusion: Shan Ci Gu can treat NSCLC through a multi-target, multi-pathway molecular mechanism and effectively improve NSCLC prognosis. This study could serve as a reference for further mechanistic research on wider application of Shan Ci Gu for NSCLC treatment.

Keywords: Shan Ci Gu, non-small cell lung cancer, network pharmacology, molecular docking, molecular mechanism

INTRODUCTION

Lung cancer is a malignant tumor with the highest incidence and mortality rates in China (Zhu et al., 2020), with cough, hemoptysis, chest pain, fever, and shortness of breath being the main clinical manifestations. Over the past 50 years, the incidence and mortality rates of lung cancer have increased significantly in many countries, and it now ranks first among the causes of death from malignant tumors in China's urban citizens. Non-small cell lung cancer (NSCLC) accounts for more than 80% of lung cancers, and the vast majority of patients are diagnosed at advanced inoperable stages. NSCLC remains the single most common malignancy of lung cancer, which has caused an increasing number of deaths in recent years (Ye et al., 2021). The main treatment methods for lung cancer include surgery, radiotherapy, chemotherapy, targeted therapy, and immunotherapy, among which concurrent radiotherapy and chemotherapy are the standard modes of treatment for NSCLC, but its five-year survival rate is only approximately 5%, and the side effects of radiotherapy and chemotherapy seriously reduce the quality of life (Cheng and Chen, 2020). In the last decade, NSCLC treatment has achieved certain efficacy through the use of targeted therapy, but all of the commonly used drugs have a single pathway and are subject to increasing drug resistance. Thus, the use of Chinese medicine in combination with the commonly used drugs and treatments can enhance the effectiveness of conventional treatment, reduce drug resistance, reduce adverse effects and toxicity, alleviate patient suffering, and improve the quality of life (Su et al., 2020).

Shan Ci Gu is the dried pseudostem of plants of the Orchidaceae family, and is sweet, slightly pungent, and cool in nature. In traditional Chinese Medicine theory, Shan Ci Gu belongs to the liver and spleen meridians. It is used to treat carbuncles and furuncles and to heal sores and phlegm ulcers, snake and insect bites, and traumatic wounds Guo and Wang, 2012. It has antibacterial, antihypertensive, gout, antitumor, and

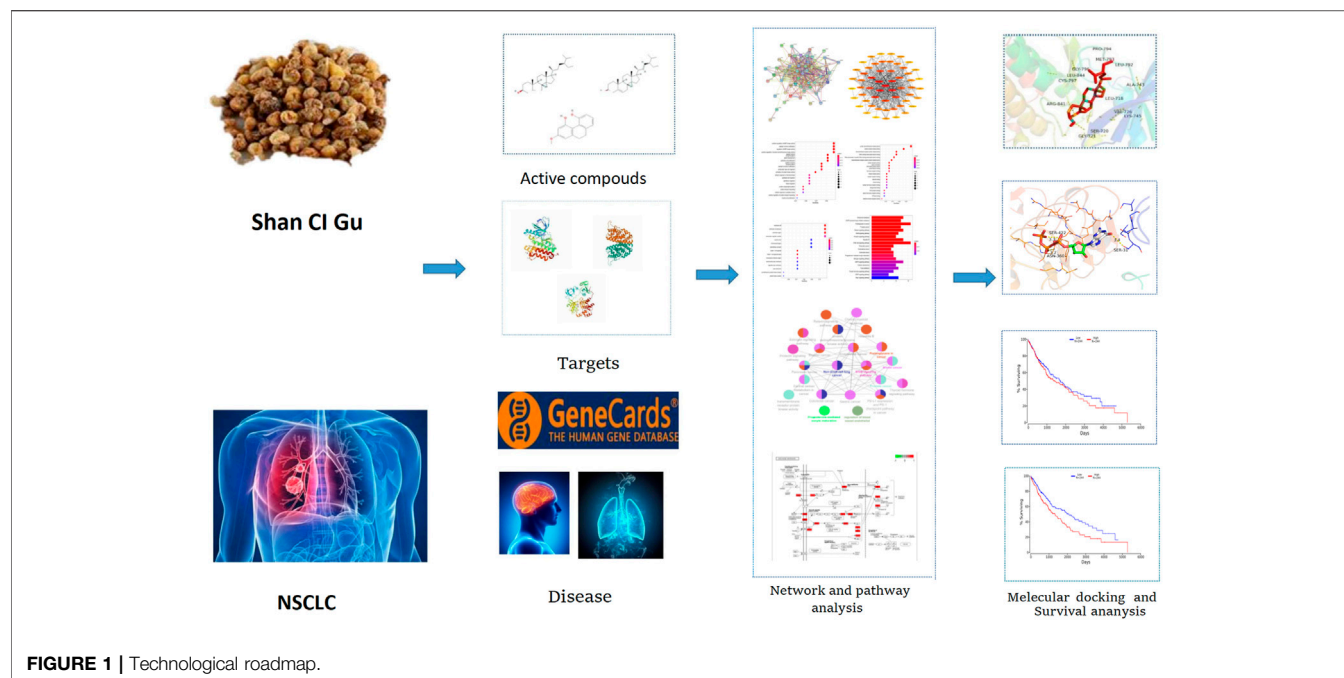
acetylcholine receptor M3-blocking effects, providing it high medicinal value (Zhang et al., 2019). Shan Ci Gu is commonly used as a clinical anti-tumor herbal medicine to treat a variety of cancers. The extracts of this herb can be used to treat Lewis lung cancer, liver cancer, and breast cancer (as well as human breast cancer MDA-MB-231 cells) (Si et al., 2020).

Shan Ci Gu also has a level of clinical efficacy against NSCLC and is effective in improving the quality of life of patients with advanced NSCLC, reducing the side effects of radiotherapy and chemotherapy, as well as enhancing the sensitivity to radiotherapy and chemotherapy (Shan et al., 2015). Its extracts can inhibit the growth of tumor cells directly or indirectly through cytotoxic effects and improve the body's immunity (Li et al., 2018). Network pharmacology is a new discipline based on the theory of systems biology, the network analysis of biological systems, and the selection of specific signal nodes (Nodes) for the design of multi-target drug molecules, which can predict the molecular mechanism of drug action in disease. Molecular docking is mainly used to study intermolecular interactions and predict the binding mode and relationship. Molecular docking is also used for drug and protein function prediction. By using network pharmacology and molecular docking, we aimed to explore the mechanism of action of Shan Ci Gu for the treatment of NSCLC to improve the condition of patients with NSCLC and reduce mortality rates. **Figure 1** shows our technology roadmap.

METHODS

Search and Collection of Active Ingredients of Shan Ci Gu

"Shan Ci Gu" was searched for in the Traditional Chinese Medicine Systems Pharmacology (TCMSP) database (<http://tcmspw.com/tcmsp.php>) to find multiple active compounds. Oral bioavailability (OB) represents the rate and degree of



absorption of traditional Chinese medicine in the human circulatory system, while drug-likeness (DL) is the similarity of compounds to known drugs. $OB \geq 30\%$ and $DL \geq 0.18$ were used to screen the active ingredients of Shan Ci Gu. The names of the compounds collected from the TCMSP database were entered into the PubChem database (<https://pubchem.ncbi.nlm.nih.gov/>) and the SMILE and 3D “Standard Delay Format” (SDF) structures of the corresponding compounds were downloaded for the prediction of target genes and molecular docking.

Acquisition and collection of targets for the active ingredients of Shan Ci Gu for the treatment of NSCLC.

The SMILE structure of the active compound was uploaded into the Swiss Target Prediction database (<http://www.swisstargetprediction.ch/>), the predicted target gene data were downloaded in CSV format, and all active compounds were filtered and integrated using Microsoft Excel software. Predicted targets of the components were imported into UniProt for normalization and then restricted to human species, and all retrieved target proteins were corrected to their official names. The anti-cancer targets of the main components of Shan Ci Gu were imported into Cytoscape (3.7.2) to generate a “component-target” network. The nodes in the network diagram are the chemical components and targets. The correlation between components and targets is represented by edges.

Through the GeneCards database (<https://www.gene-cards.org/>), targets related to NSCLC were searched by entering the keyword “Non-Small Cell Lung Cancer.” Then, we combined the components and targets of Shan Ci Gu with those of the drug. Predicted target mapping of the relevant targets with NSCLC targets was used to obtain the target of action of Shan Ci Gu for NSCLC. Venny (https://bioinfogp.cnb.csic.es/tools/venny_old/) was used to draw a Venn diagram of matsutake and NSCLC targets.

Protein-Protein Interaction Analysis and Core Target Screening

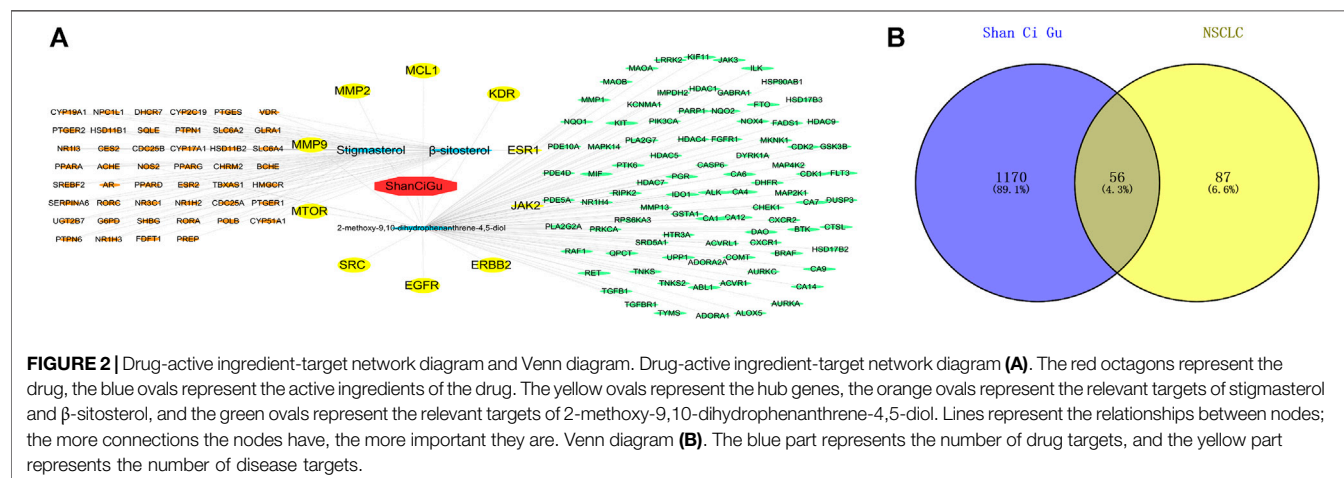
The intersection of Shan Ci Gu and NSCLC targets was uploaded to the online site of STRING version 10.5 (<https://string-db.org/>). The protein type was set to “*Homo sapiens*,” a high confidence level of 0.7 was selected, and the other parameters were set to default values. The protein interaction relationships were retrieved. Node1, Node2, and the combined score from the export file were imported into Cytoscape, and the interaction network was constructed. The node size reflected the degree value and the thickness of the edge reflected the combined core in the final PPI network diagram, and the core proteins with the top 10⁴ values were selected.

Gene Ontology and Kyoto Encyclopedia of Genes and Genomes Pathway Enrichment Analyses

GO analysis of the relevant obtained intersection target proteins was performed using R software to select the biological process (BP), cellular component (CC), and molecular function (MF). These data were plotted as bubble charts, and R software was used to construct a bar graph of the KEGG biological pathway results, collect the targets from the pathway, and upload the active components, pathways, and targets to Cytoscape 3.7.2 software to create a “component-target-pathway” map. A network diagram was constructed to visualize, and thus, explore the mechanisms of Shan Ci Gu related to the treatment of NSCLC in detail. Cluster ONE was then used to sift the candidate genes and retrieve them for further analysis. The KEGG function of the genes was further understood by using the Cytoscape plugin Clue GO, and the relevant KEGG pathways were selected for enrichment analysis. The intersecting genes were imported

TABLE 1 | Active ingredients of Shan Ci Gu.

MOL	English name	Pubchem CID	OB	DL
MOL000358	beta-sitosterol	222,284	36.91	0.75
MLL000449	Stigmasterol	5280794	43.83	0.76
MOL007991	2-methoxy-9,10-dihydrophenanthrene-4,5-diol	11506999	44.97	0.18



into R software, and the script was run to obtain a predictive map of the endocrine pathway mechanism.

Molecular Docking

Molecular docking is a method for drug design by exploring the interaction and recognition between receptors and ligands. It is a theoretical simulation method that focuses on the study of intermolecular interactions and the prediction of their binding patterns and affinities. In recent years, molecular docking methods have become an important technique in the field of computer-aided drug research (Chen et al., 2020). AutoDock Vina is an open-source molecular docking program designed by the Scripps Research Institute for the computation of semi-flexible molecular docking. AutoDock Vina uses a complex gradient algorithm and multi-threaded techniques to make more accurate and faster predictions than AutoDock4. Semi-flexible docking means that the conformation of the ligand molecule can be changed according to the receptor molecule and is flexible, while the receptor molecule does not change and is rigid. The SDF structure of the active ingredient was imported into Chem3D 18.0 for optimization. The main protein targets selected were passed through the PDB database (<https://www.rcsb.org/>). Therefore, we searched the PDB database for the 3D structures of the ten potential targets of Shan Ci Gu in the treatment of NSCLC and found 3D structures for ten of the targets (EGFR, SRC, ESR1, ERBB2, MTOR, MCL1, MMP2, MMP9, KDR, and JAK2). Then, the best protein crystal structure was selected [images with lower resolution (Å) with observable ligands and a relatively intact structure were more desirable] and downloaded from the PDB database. The PDB

files of the active compound and ligand molecules were imported into AutoDock Tools. We removed these target proteins' water molecules, added polar hydrogen, and built active pockets active pockets, which were saved as PDBQT format files for later use. By adjusting target protein X-Y-Z coordinates and grid size, optimizing the position of protein structure-binding sites for molecular docking. AutoDock Vina was run to dock the treated active compound to the target protein ten times, and the lowest binding energy for each docking was taken as the final result. The complexes were then observed and plotted using PyMOL.

Survival Analysis

We used OncoLnc (<http://www.oncolnc.org/>) to obtain OS (Overall Survival) and DFS (Disease-free survival) significance data for ten core genes (EGFR, SRC, ESR1, ERBB2, MTOR, MCL1, MMP2, MMP9, KDR, and JAK2) in all squamous lung cancers at TCGA. High (50%) and low (50%) cutoff values were used as expression thresholds to split high and low expression cohorts, and 488 samples were analyzed using log-rank test and Kaplan-Meier to obtain survival maps. $p < 0.05$ was considered a statistically significant difference.

RESULTS

Search and Collection of the Active Ingredients of Shan Ci Gu

Based on the screening conditions of the compounds, three active ingredients were collected from the TCMS database, namely, 2-

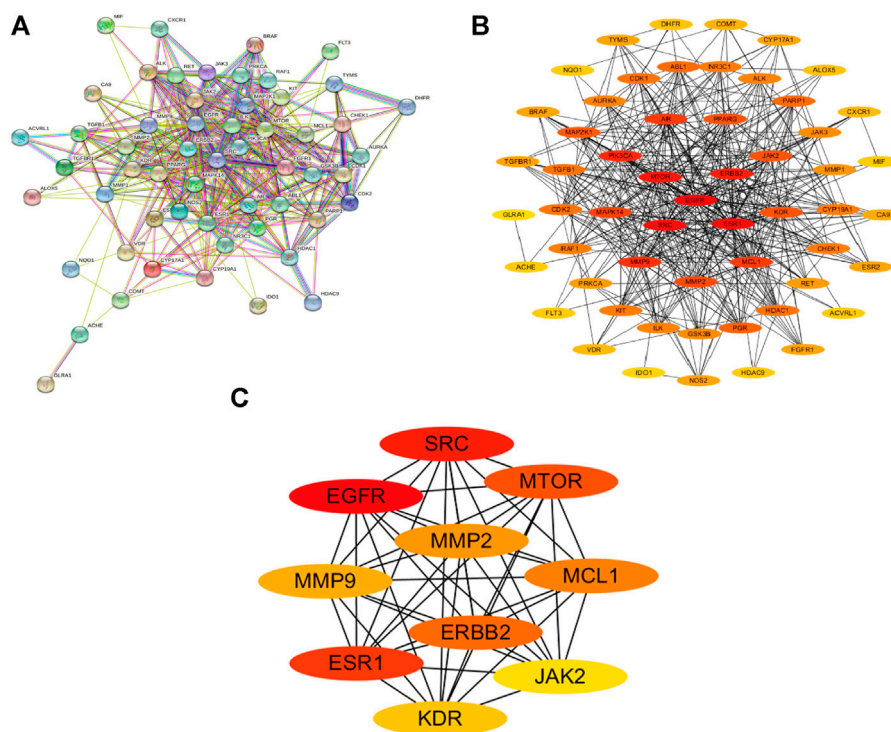


FIGURE 3 | Protein-protein interaction (PPI) analysis. **(A)** The PPI network was constructed using the plug-in targets from the Search Tool for the Retrieval of Interacting Genes/Proteins database, which were imported into Cytoscape, and the targets were the candidates used for non-small cell lung cancer treatment. **(B)** Proteins are represented by nodes (colors from red to yellow illustrate the extent to which the medical targets have combined with each other). Edges indicate protein-protein associations. **(C)** The top 10 targets (hub targets) in the PPI network ranked by maximal clique centrality using the cytoHubba plug-in.

methoxy-9,10-dihydrophenanthrene-4,5-diol, stigmasterol, and β -sitosterol. The details are shown in Table 1.

Acquisition of Active Targets for the Treatment of Non-Small Cell Lung Cancer From Shan Ci Gu

The Swiss Target Prediction database of predicted targets was compiled, resulting in 42 predicted targets for the active ingredient β -sitosterol and 45 predicted targets for the active ingredient stigmasterol, with an active ingredient of 2-methoxy-9,10-dihydrophenanthrene-4,5-diol. The number of predicted targets for dihydrophenanthrene-4,5-diol was 100, and the total number of targets was 143 by combining the duplicates. The “drug-active compound-target” network graph built in Cytoscape 3.7.2 reflects the correspondence of the compound targets, as shown in Figure 2A. The GeneCards database was searched with the keyword “Non-Small Cell Cancer,” and 1226 NSCLC targets were obtained based on the relevance score. The 56 intersecting genes of Shan Ci Gu and NSCLC were obtained, as shown in Figure 2B.

Protein-Protein Interaction Network Analysis

The 56 intersection targets of the predicted Shan Ci Gu and NSCLC were imported into the STRING database to select a *H.*

sapiens-generated PPI network map and obtain protein interaction relationships (Figure 3A). The intersection targets were imported into Cytoscape 3.7.2 to create a network diagram of potential target interactions (Figures 3B,C). Nodes represent proteins and edges represent relationships between proteins, resulting in a total of 56 nodes and 409 edges. The colors from yellow to red represent small to large degree values, respectively, and according to the degree values, the key nodes, which are epidermal growth factor receptor (EGFR), ESR1, SRC, ERBB2, MTOR, MCL1, matrix metalloproteinase 2 (MMP2), MMP9, KDR, and JAK2, were selected for interactions with NSCLC.

Gene Ontology Analysis and Kyoto Encyclopedia of Genes and Genomes Pathway Enrichment Analysis

To further explore possible mechanisms of the 56 candidate targets for the treatment of NSCLC, R software was used for GO enrichment analysis with the candidate target and KEGG pathway analysis of these targets for the treatment of NSCLC with Shan Ci Gu. The results showed that the number of BP terms was 1,385, CC was 35, and MF was 69. The top 15 BPs are shown in bubble charts (Figures 4A–C). KEGG pathway enrichment analysis was conducted using R software and involved 125 terms. The bar chart (Figure 4D) reflects the top 20 entries.

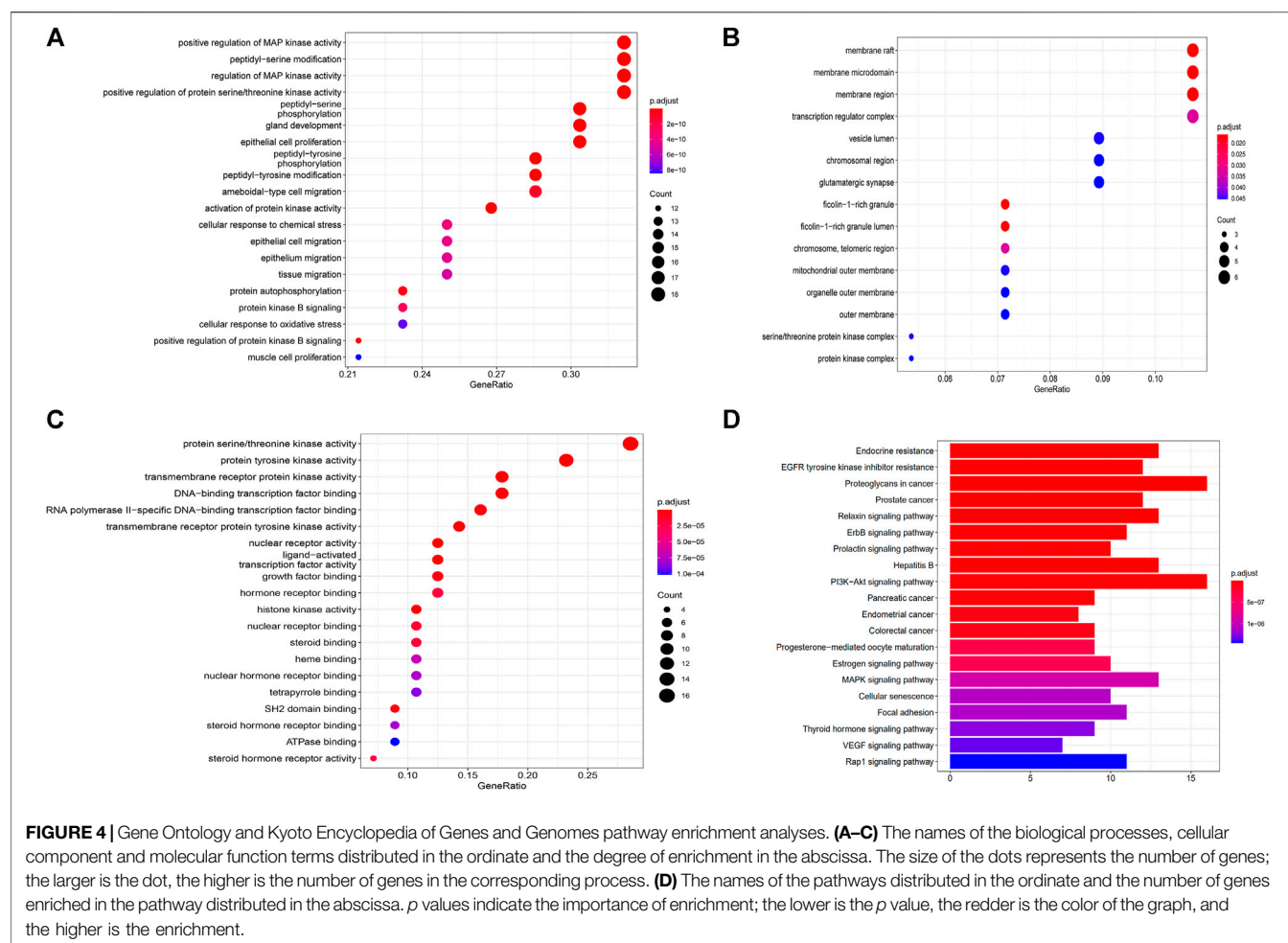
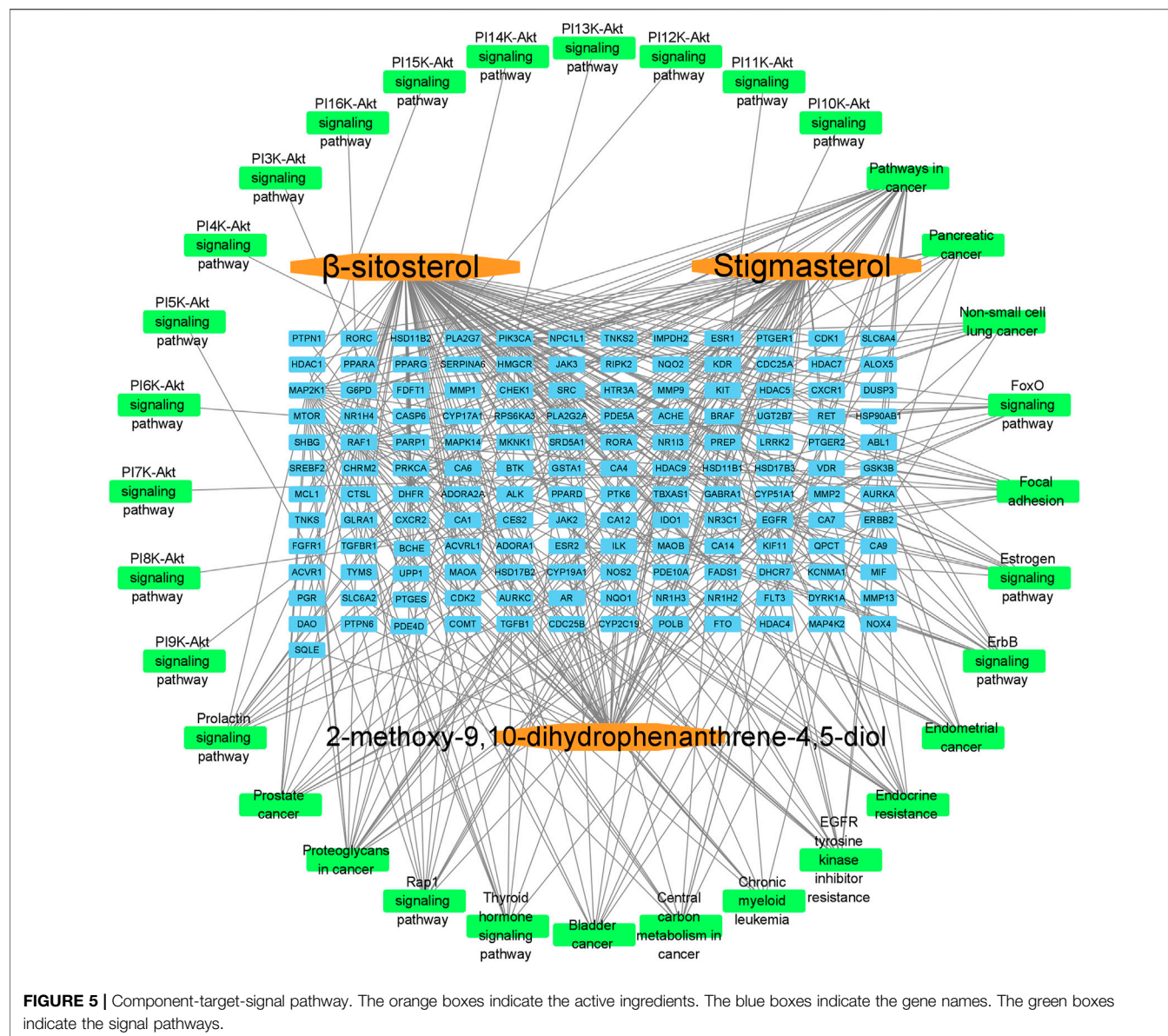


TABLE 2 | Kyoto Encyclopedia of Genes and Genomes pathway analysis for the treatment of non-small cell lung cancer, using R software.

Pathway ID	Paythway	pvalue	Number of target hits
hsa01522	Endocrine resistance	6.61E-14	13
hsa01521	EGFR tyrosine kinase inhibitor resistance	1.26E-13	12
hsa05205	Proteoglycans in cancer	2.76E-13	16
hsa05215	Prostate cancer	1.60E-12	12
hsa04926	Relaxin signaling pathway	2.48E-12	13
hsa05219	Bladder cancer	5.53E-12	9
hsa04012	ErbB signaling pathway	9.02E-12	11
hsa05224	Breast cancer	1.34E-11	13
hsa04917	Prolactin signaling pathway	3.15E-11	10
hsa05230	Central carbon metabolism in cancer	3.15E-11	10
hsa05223	Non-small cell lung cancer	4.22E-11	10
hsa05161	Hepatitis B	4.63E-11	13
hsa04151	PI3K-Akt signaling pathway	1.09E-09	16
hsa05212	Pancreatic cancer	1.87E-09	9
hsa05226	Gastric cancer	4.19E-09	11
hsa05213	Endometrial cancer	4.62E-09	8
hsa05210	Colorectal cancer	5.72E-09	9
hsa05225	Hepatocellular carcinoma	1.49E-08	11
hsa04914	Progesterone-mediated oocyte maturation	2.20E-08	9
hsa04915	Estrogen signaling pathway	2.75E-08	10



Detailed data are shown in **Table 2**. The active ingredients, candidate targets, and 20 pathways were imported into Cytoscape to create a “component-target-pathway” network diagram and visualize it (**Figure 5**). The predictive map of the mechanism of the endocrine pathway is shown in **Figure 6**. Relevant targets in the signaling pathway of Shan Ci Gu and endocrine resistance is shown in **Figure 7**.

Molecular Docking

Table 3 shows the results acquired from the molecular docking software (AutoDock Vina). Processed by PyMOL software, the docked complex and ‘Best-Docked Complex’ show images of the best docking of the receptor and ligand. According to **Table 3**, we can conclude that the binding energies of the three active ingredients to the top three ranked target gene (EGFR, SRC, and ESR1) transcriptional proteins were all less than -5 kcal/mol,

and most of the binding energies of the remaining core target proteins were also less than 0, indicating a high affinity between the compounds and the core target genes. The binding energies of MMP2 and MMP9 to the three components were all greater than 0, indicating that the two target proteins had low binding ability to the active components of Shanzi mushroom and cannot be used as active sites.

Survival Analysis

We divided the lung squamous carcinoma cases into high and low expression groups based on the expression levels of ten core genes and investigated the correlation between the expression of each of the ten core genes and the prognosis of lung squamous carcinoma patients mainly using TCGA. As shown in **Figure 8**, highly expressed genes were associated with poor prognosis. Among the ten hub genes analyzed for survival, ESR1 ($p = 0.00454$) and

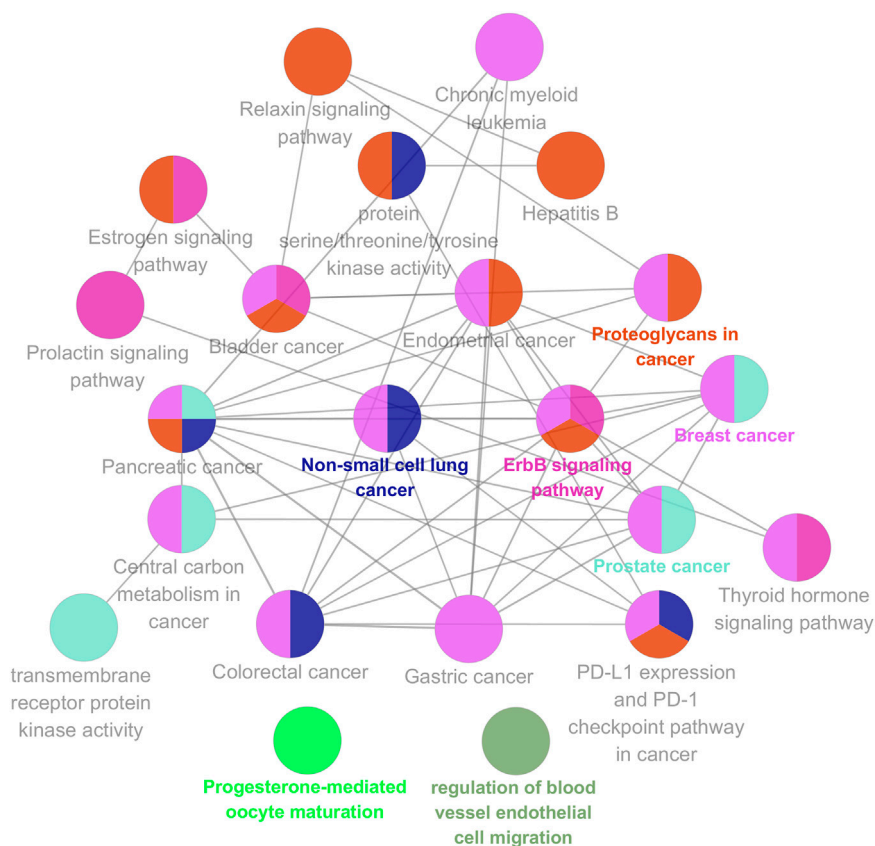


FIGURE 6 | KEGG analysis using the ClueGO plug-in. Kyoto Encyclopedia of Genes and Genomes pathway analyses of the potential targets of Shan Ci Gu against non-small cell lung cancer by the ClueGO plug-in. Each node is a representative enrichment pathway. The nodes indicate the number of genes shared between pathways. The color represents the enrichment classification of the node.

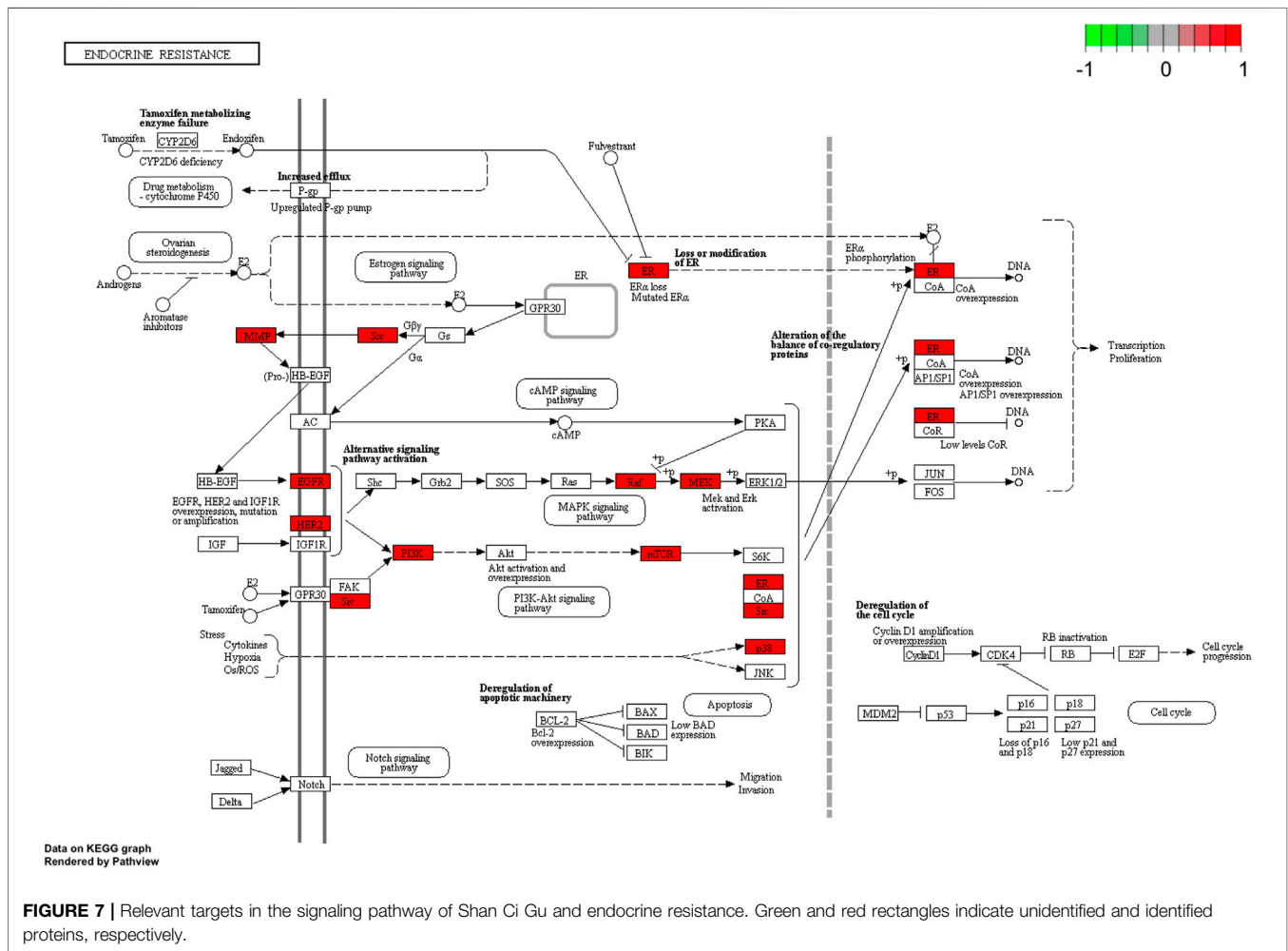
MMP2 ($p = 0.0388$) were associated with overall survival in patients with lung squamous carcinoma (**Figures 8C,G**), and overall survival analysis of other hub genes with high and low expression did not show statistical significance ($p < 0.05$).

DISCUSSION

As of 2021, lung cancer will continue to be the leading cause of cancer death for both men and women, accounting for 22% of all cancer deaths. To further improve the survival rate of patients with lung cancer, this study identified the active ingredients and the possible detailed molecular mechanism of Shan Ci Gu in the treatment of NSCLC. This study supports the wider application and further mechanistic surveys of Shan Ci Gu for the treatment of NSCLC.

First, we aimed to derive the feasible active ingredients and targets according to the OB and DL values of the components of TCMSP. Consequently, three active components of Shan Ci Gu were obtained, and only 143 candidate targets of these active components were retrieved through Swiss Target Prediction. We constructed an herb-ingredient-target network, which reflected the relationship between numerous components and

targets of the drug. In this study, three active components of Shan Ci Gu for the treatment of NSCLC were identified: β -sitosterol, stigmasterol, and 2-methoxy-9,10-dihydrophenanthrene-4,5-diol. Previous studies have mainly focused on anti-breast cancer, -colon cancer, and -prostate cancer effects of Shan Ci Gu. β -sitosterol, as a potential natural drug, can effectively prevent the occurrence and growth of many types of tumors. Zhou et al. (Zhou et al., 2016), in 2016, showed that β -sitosterol promotes the apoptosis of A549 cells in a dose-dependent manner in the range of 0–40 μ M. The higher is the concentration of β -sitosterol, the higher is the apoptosis rate of the A549 cells. 2-Methoxy-9,10-dihydrophenanthrene-4,5-diol is a type of phenanthrene component, which has anti-tumor, anti-bacterial, anti-spasmodic, anti-inflammatory, anti-allergic, and anti-platelet aggregation biological activities. Xue et al. (Xue et al., 2006), in 2006, carried out cytotoxicity experiments that showed that blestiarane C and other phenanthrene components have strong inhibitory effects on the proliferation of HepG2 cells of liver cancer, and blestiarane A has certain inhibitory effects on the proliferation of A549 cells of lung cancer *in vitro*. Therefore, β -sitosterol, stigmasterol, and 2-methoxy-9,10-dihydrophenanthrene-4,5-diol are good anti-tumor agents, and related studies show that they are of importance in the proliferation



and apoptosis of liver, colon, and breast cancer cells, while studies regarding NSCLC are scarce, and the mechanism of action of Shan Ci Gu remains unclear.

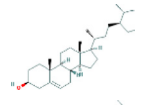
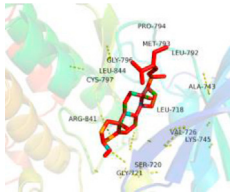
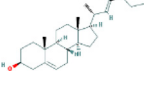
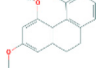
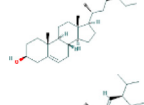
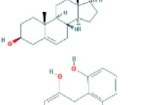
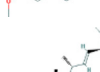
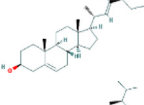
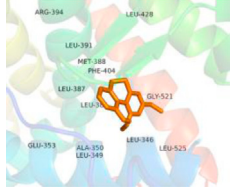
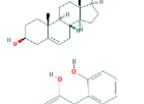
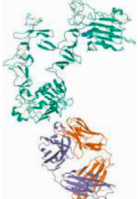
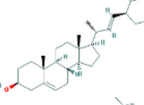
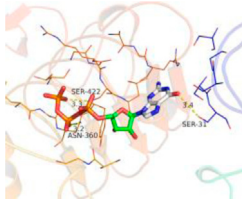
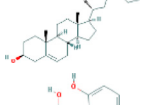
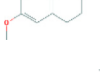
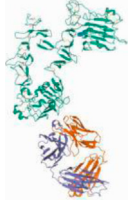
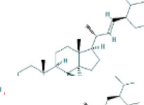
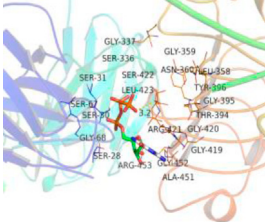
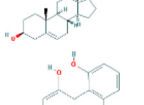
We retrieved 1,226 candidate NSCLC targets from the GeneCards database and 143 candidate targets of Shan Ci Gu through Swiss Target Prediction. We found 56 targets in common between the disease and Shan Ci Gu, which were considered potential targets for treating NSCLC. The degree of association between these genes is shown in **Figures 3A,B**. The top 10 core genes are shown in **Figure 3C**, including EGFR, ESR1, SRC, ERB2, MTOR, MCL1, MMP2, MMP9, KDR, and JAK2. These genes play a significant role in the proliferation, migration, and apoptosis of NSCLC cells. As EGFR and ERBB2 are human EGFRs, they have important impacts on the physiological processes of cell growth, proliferation, and differentiation in humans. Increased expression of EGFR is commonly observed in malignancies such as lung, breast, and pancreatic cancers, making this receptor a major target for the development of anti-tumor therapies (de Laveria et al., 2021). ESR1 and SRC are involved in endocrine regulation, and MMP2 and MMP9 regulate complex kinases, thus participating in cell proliferation and migration.

The target genes of Shan Ci Gu for the treatment of NSCLC were used to obtain an enrichment map of the GO and KEGG

pathway analyses using R software (**Figure 4**). **Figure 4** shows that the BP is mainly associated with the regulation of protein kinases such as MAP kinase and serine threonine kinase, and CC is mainly associated with various cell bodies such as the receptor complex, cell body, neuronal cell body, dendrite, and dendritic tree. The perinuclear region of cytoplasm MF is mainly associated with protein kinase activity. The main signaling pathways are the endocrine, EGFR-TKI resistance, ErbB, PI3K-Akt, and Rap1 signaling pathways. The diseases involved are prostate cancer and hepatitis B.

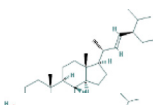
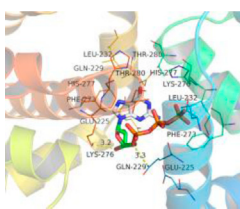
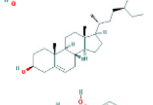
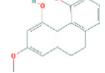
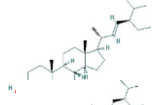
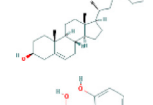
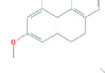
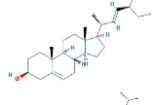
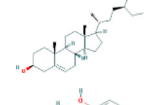
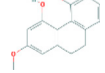
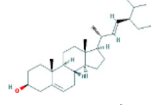
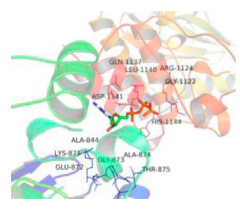
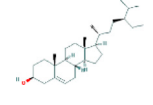
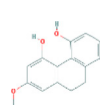
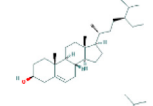
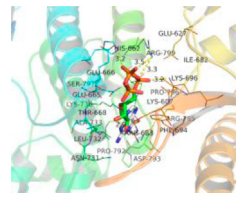
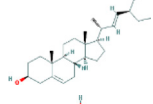
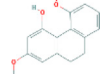
EGFR plays an important role in lung carcinogenesis, while its expression level is independent of EGFR positivity rate, lung cancer stage, lung cancer differentiation, the presence of lymph node metastasis, and so on. It should be used as an index of prognosis together with other cytokines. In human lung adenocarcinoma A549 cells, MMP secretes epidermal growth factor, activates the EGFR-ERK signaling pathway, and promotes the expression of claudin-2, thus promoting tumor colonization (ciardiello and Tortora, 2008). Among the EGFR-TKI resistance signaling pathways, KRAS is an important signaling pathway downstream of EGFR, and the mutated KRAS gene directly activates the MAPK signaling pathway without relying on the activation of upstream EGFR, leading to tumor proliferation and metastasis (Wang et al.,

TABLE 3 | Molecular docking results.

Target	Target structure	Compound	Compound 2D structure	Affinity (kcal/mol)	Best-docked complex
EGFR		Beta-sitosterol		-9.0	
		Stigmasterol		-9.1	
		2-Methoxy-9,10-dihydrophenanthrene-4,5-diol		-8.0	
SRC		Beta-sitosterol		-10.1	
		Stigmasterol		-10.1	
		2-Methoxy-9,10-dihydrophenanthrene-4,5-diol		-8.4	
ESR1		Beta-sitosterol		-7.3	
		Stigmasterol		-6.8	
		2-Methoxy-9,10-dihydrophenanthrene-4,5-diol		-8.4	
MTOR		Beta-sitosterol		-5.0	
		Stigmasterol		-5.0	
		2-Methoxy-9,10-dihydrophenanthrene-4,5-diol		-5.0	
ERBB2		Beta-sitosterol		-4.7	
		Stigmasterol		-5.1	
		2-Methoxy-9,10-dihydrophenanthrene-4,5-diol		-5.0	

(Continued on following page)

TABLE 3 | (Continued) Molecular docking results.

Target	Target structure	Compound	Compound 2D structure	Affinity (kcal/mol)	Best-docked complex
MCL1		Beta-sitosterol		-6.9	
		Stigmasterol		-6.8	
		2-Methoxy-9,10-dihydrophenanthrene-4,5-diol		-6.7	
MMP2		Beta-sitosterol		4.5	NA
		Stigmasterol		4.4	
		2-Methoxy-9,10-dihydrophenanthrene-4,5-diol		4.5	
MMP9		Beta-sitosterol		7.1	NA
		Stigmasterol		2.8	
		2-Methoxy-9,10-dihydrophenanthrene-4,5-diol		1.8	
KDR		Beta-sitosterol		-2.8	
		Stigmasterol		-2.7	
		2-Methoxy-9,10-dihydrophenanthrene-4,5-diol		-3.3	
JAK2		Beta-sitosterol		-1.8	
		Stigmasterol		-1.9	
		2-Methoxy-9,10-dihydrophenanthrene-4,5-diol		-1.7	

Note: The results of the docking data of MAPP2 and MAPP9 show that these two proteins cannot be binding sites; thus, NA means there is no optimal binding image.

2020). Inhibition of EGFR expression can be used to treat NSCLC by modulating the immune microenvironment. This is because EGFR can upregulate immune checkpoints such as PD-L1 and

IDO1, making NSCLC more resistant to drugs. In addition, ESR1 encodes an estrogen receptor involved in hormone binding, DNA binding, and transcription activation and participates in breast

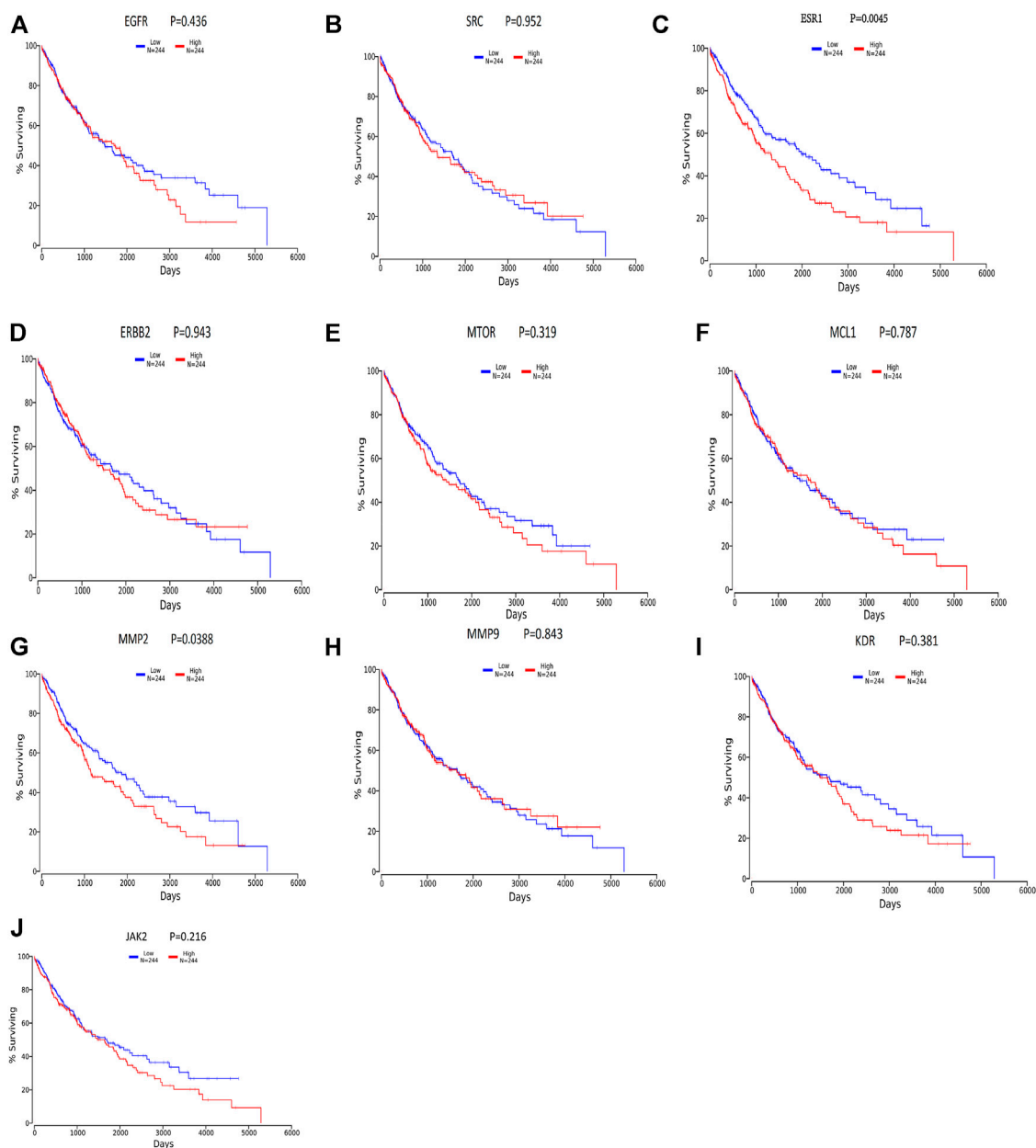


FIGURE 8 | Survival analysis for hub genes by oncolnc. Red lines represent sample groups with high gene expression, while green lines represent sample groups with low gene expression.

cancer, endometrial cancer, osteoporosis, and other pathological processes. ESR1 mRNA overexpression is associated with the prognosis of NSCLC (Teng et al., 2018). The existence of a ligand-independent ESR signaling pathway has been demonstrated, wherein ESR and its functional pathways are subject to multiple regulations such as those of the growth and differentiation of target cells, either through their own phosphorylation or through binding or interaction with different growth factors, co-regulators, oncogenes, or oncogenic proteins (Zhang et al., 2014). A study by Atmaca, in 2020, showed that assessment of ESR1 mRNA by qPCR is a feasible method to

examine ESR1 expression in NSCLC, and ESR1 expression determines the prognosis of metastatic NSCLC. ESR1 is a predictive biomarker that is of therapeutic importance in breast cancer, and this study indicates that it can play a similar role in lung cancer (Atmaca et al., 2014). The Src tyrosine kinase inhibitor can be selectively used for the molecular targeting of NSCLC with high activation of Src proteins (Zheng et al., 2011). A study by Yao, in 2020, further demonstrated that hesperidin effectively inhibits the proliferation of NSCLC cells (A549 and H460) by inhibiting the SRC3-mediated ubiquitination of IGF-1R-PI3K-AKT signaling to induce apoptosis and exert an inhibitory effect on the tumor

growth in NSCLC (Yao et al., 2020). Dong et al. found that quercetin inhibited the expression of Src, and subsequently, inhibited Fln14/NF-kappa B signaling, thereby suppressing the proliferation and metastasis of NSCLC. This suggests that Src expression promotes NSCLC progression and could be a target for the treatment of NSCLC (Dong et al., 2020). In the ErbB signaling pathway, the ErbB receptor family and its downstream pathways may regulate epithelial-mesenchymal transition, migration, and tumor invasion by regulating components of the extracellular matrix (ECM) (Kern et al., 1992). This mechanism occurs not only in NSCLC but also in other tumor growth processes such as those of breast, ovarian, and bladder cancer. EGFR is a member of the ErbB family and can bind to ECM components such as matrilins to promote tumor cell expansion (Appert-Collin et al., 2015). Yu et al., in 2020, found that TMPO-AS1 was upregulated in the cancerous tissues of NSCLC samples, which enhanced the expression of ERBB2, promoting the deterioration of NSCLC cells (Yu et al., 2020). The activation of mTOR causes an accelerated tumor cell cycle, shortened G1 phase duration, rapid cell proliferation, and increased secretion of oncoproteins, and this promotes tumor development. mTOR acts in a synergistic manner to inhibit tumor growth in mouse prostate and lung cancer models, and phosphorylated or activated mTOR is found in 74% of the NSCLCs, making it an additional target for NSCLC therapy (Marinov et al., 2007). In A549 and primary human NSCLC cells, GDC-0349 inhibits NSCLC cell growth, proliferation, cell cycle progression, migration, and invasion through the Akt-Akt-mTOR pathway, while inducing significant apoptotic activation (Yang et al., 2020). Analysis of a receiver operating characteristic curve (area under curve = 0.6785) showed that the expression of MCL-1 is an important critical value for predicting prognosis in 30.0% of the NSCLC tumor cell types. Curcumin inhibits the expression of radiation-induced EMT and E-cad by reducing the expression of MMP9, thereby inhibiting the migration and invasion of NSCLC (Deng et al., 2020). CCK8 and Transwell invasion assays have shown that A549 cells transfected with the miR-4448 inhibitor have higher proliferation and metastatic abilities. High expression of MMP2 and MMP9 in A549 cells transfected with the miR-4448 inhibitor has been confirmed by qRT-PCR and western blot (Xu et al., 2020). miR-142-3p overexpression inhibits the expression of NR2F6, MMP2, and MMP9 and improves caspase-3 expression, thereby inhibiting lung adenocarcinoma cell proliferation, migration, and invasion and enhancing apoptosis, demonstrating that miR-142-3p may be a new therapeutic target for lung adenocarcinoma treatment (Jin et al., 2019). In one study, compared to that in the normal lung cell line, miR-204 expression was found to be downregulated, while that of JAK2 was upregulated in four NSCLC cell lines (A549, H1299, H1650, and H358). These findings indicate that miR-204 functions as a tumor suppressor in NSCLC by acting on JAK2. Therefore, we can consider miR-204 as a biomarker for the diagnosis and

treatment of NSCLC (Wang et al., 2016). Among the endocrine signaling pathways, the core targets EGFR, ERBB2, ESR1, MTOR, MMP2, MMP9, and SRC are in endocrine signaling pathways. Neuroendocrine dedifferentiation (NED) is widely found in tumors of prostate, gastrointestinal tract, and lung cancers, among others. Chen (Chen et al., 2014), have found that 0–20% of NSCLC is associated with NED and inhibition of the Akt signaling pathway. Clinical manifestations, natural course, pathological changes, and treatment response are all characteristic of NED and have become a new field of lung cancer research (Dudzińska et al., 2020). Ma et al. (Ma and Zhang, 2015), in 2015, found that with a positive rate of 72.3%, the expression of Rap1b was significantly higher in NSCLC tissues compared to that in paraneoplastic tissues. Further studies have confirmed that Rap1b is closely related to tumor differentiation, supporting the conclusion that Rap1b may have an oncogene function in the development of NSCLC. The results of survival analysis showed that the survival rate of the low-expression group of ESR1 and MMP2 was higher than that of the high-expression group, which further confirmed that reducing the expression of ESR1 and MMP2 could improve the quality of life of NSCLC patients.

CONCLUSION

We speculate that Shan Ci Gu may play a role in inhibiting tumor cell proliferation by targeting several proteins such as EGFR, SRC, and ESR1 in NSCLC. The therapeutic effect of Shan Ci Gu involves a variety of BPs mainly involved in the inhibition of cell proliferation and endocrine effects, such as the endocrine, EGFR-TKI resistance, ErbB, and PI3K-Akt signaling pathways. In conclusion, Shan Ci Gu plays a role in the treatment of NSCLC through multiple targets and pathways. Therefore, the results of this study provide a basis for further research on the clinical application of Shan Ci Gu in NSCLC.

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

YW selected the topic, designed the research methodology, collected and sorted data, and wrote the first draft. YZ selected the topic and designed the research methodology. YW, XS, and CL analyzed the data. SS, XL, CY, and JL collected and sorted data. QD reviewed and finalized the manuscript. All authors have read and approved the final manuscript.

REFERENCES

- Appert-Collin, A., Hubert, P., Crémel, G., and Bennisroune, A. (2015). Role of ErbB Receptors in Cancer Cell Migration and Invasion. *Front. Pharmacol.* 6, 283. doi:10.3389/fphar.2015.00283
- Atmaca, A., Al-Batran, S.-E., Wirtz, R. M., Werner, D., Zirlik, S., Wiest, G., et al. (2014). The Validation of Estrogen Receptor 1 mRNA Expression as a Predictor of Outcome in Patients with Metastatic Non-small Cell Lung Cancer. *Int. J. Cancer* 134 (10), 2314–2321. doi:10.1002/ijc.28571
- Chen, G., Seukep, A. J., and Guo, M. (2020). Recent Advances in Molecular Docking for the Research and Discovery of Potential Marine Drugs. *Mar. Drugs* 18 (11), 545. doi:10.3390/md18110545
- Chen, Y., Nowak, I., Huang, J., Keng, P. C., Sun, H., Xu, H., et al. (2014). Erk/MAP Kinase Signaling Pathway and Neuroendocrine Differentiation of Non-small-cell Lung Cancer. *J. Thorac. Oncol.* 9 (1), 50–58. doi:10.1097/jto.000000000000034
- Cheng, W., and Chen, X. M. (2020). Research Progress on the Mechanism of Action of Chinese Medicine against Non-small Cell Lung Cancer. *Chin. J. Exp. Formulary* 26 (24), 227–234. (Chinese).
- ciardiello, F., and Tortora, G. (2008). EGFR Antagonists in Cancer Treatment. *N. Engl. J. Med.* 358 (11), 1160–1174. doi:10.1056/nejmra0707704
- de Lavera, I., Merklings, P. J., Oliva, J. M., Sayagués, M. J., Cotán, D., Sánchez-Alcázar, J. A., et al. (2021). EGFR-targeting Antitumor Therapy: Neuregulins or Antibodies? *Eur. J. Pharm. Sci.* 158, 105678. doi:10.1016/j.ejps.2020.105678
- Deng, X., Chen, C., Wu, F., Qiu, L., Ke, Q., Sun, R., et al. (2020). Curcumin Inhibits the Migration and Invasion of Non-small-cell Lung Cancer Cells through Radiation-Induced Suppression of Epithelial-Mesenchymal Transition and Soluble E-Cadherin Expression. *Technol. Cancer Res. Treat.* 19, 1533033820947485. doi:10.1177/1533033820947485
- Dong, Y., Yang, J., Yang, L., and Li, P. (2020). Quercetin Inhibits the Proliferation and Metastasis of Human Non-small Cell Lung Cancer Cell Line: The Key Role of Src-Mediated Fibroblast Growth Factor-Inducible 14 (Fn14)/ Nuclear Factor Kappa B (NF- κ B) Pathway. *Med. Sci. Monit.* 26, e920537. doi:10.12659/msm.920537
- Dudzińska, M., Szczyrek, M., Wojas-Krawczyk, K., Świrski, J., Chmielewska, I., and Zwolak, A. (2020). Endocrine Adverse Events of Nivolumab in Non-small Cell Lung Cancer Patients-Literature Review. *Cancers (Basel)*. 12 (8), 2314. doi:10.3390/cancers12082314
- Guo, J. H., and Wang, Z. H. (2012). Zhou Zhongying's Academic Thoughts on the Treatment of Lung Cancer from Cancer Toxicity. *J. Tradit. Chin. Med.* 39 (03), 410–411. (Chinese).
- Jin, C. e., Xiao, L., Zhou, Z., Zhu, Y., Tian, G., and Ren, S. (2019). MiR-142-3p Suppresses the Proliferation, Migration and Invasion through Inhibition of NR2F6 in Lung Adenocarcinoma. *Hum. Cel* 32 (4), 437–446. doi:10.1007/s13577-019-00258-0
- Kern, J. A., Robinson, R. A., Gazdar, A., Torney, L., and Weiner, D. B. (1992). Mechanisms of p155HER2Expression in Human Non-small Cell Lung Cancer Cell Lines. *Am. J. Respir. Cel Mol Biol.* 6 (4), 359–363. doi:10.1165/ajrcmb/6.4.359
- Li, X., Zhang, Y., Yang, P., and Sun, B.-X. (2018). The Effect of Xiaoyan Decoction on the Migration and Invasion of Lung Cancer A549 Cells and Lung Cancer Stem Cells. *Chin. Herbal Med.* 49 (4), 870–873. doi:10.7501/j.issn.0253-2670.2018.04.019
- Lu, J. (2008). *Professor Zhou Zhongying's Clinical Experience in Treating Lung Cancer and the Inheritance of Academic Ideas*. Nanjing University of Traditional Chinese Medicine, (Chinese).
- Ma, M. M., and Zhang, M. X. (2015). Expression and Significance of Rap1b in NSCLC. *Mod. Oncol. Med.* 23 (17), 2426–2428. (Chinese).
- Marinov, M., Fischer, B., and Arcaro, A. (2007). Targeting mTOR Signaling in Lung Cancer. *Crit. Rev. Oncology/Hematology* 63 (2), 172–182. doi:10.1016/j.critrevonc.2007.04.002
- Shan, S., Guo, M., and Zhu, D. (2015). Etc. Professor Zhou Zhongying, a Master of Traditional Chinese Medicine, Experience in Treating Lung Cancer. *Chin. J. Traditional Chin. Med.* 30 (12), 4332–4335. (Chinese).
- Si, H. R., Si, Y., Jiao, Y. F., Wang, H., Li, P. Y., and Zhong, S. X. (2020). Research Progress on the Chemical Composition and Pharmacological Effects of the Mountain Cichlid Mushroom. *J. Liaoning Univ. Chin. Med.* 22 (05), 151–155. (Chinese).
- Su, X.-L., Wang, J.-W., Che, H., Wang, C.-F., Jiang, H., Lei, X., et al. (2020). Clinical Application and Mechanism of Traditional Chinese Medicine in Treatment of Lung Cancer. *Chin. Med. J. (Engl)*. 133 (24), 2987–2997. doi:10.1097/cm9.0000000000001141
- Teng, Y., Ding, Y., Zhang, M., Chen, X., Wang, X., Yu, H., et al. (2018). Genome-wide Haplotype Association Study Identifies Risk Genes for Non-small Cell Lung Cancer. *J. Theor. Biol.* 456, 84–90. doi:10.1016/j.jtbi.2018.08.007
- Wang, P., Lv, H. Y., Zhou, D. M., and Zhang, E. N. (2016). miR-204 Suppresses Non-small-cell Lung Carcinoma (NSCLC) Invasion and Migration by Targeting JAK2. *Genet. Mol. Res.* 15 (2). doi:10.4238/gmr.15026415
- Wang, X., Jia, X., and Tong, D. (2020). Drug Resistance Mechanism and Countermeasures after Treatment with Osimertinib for NSCLC. *Basic Clin. Oncol.* 33 (04), 365–368. (Chinese).
- Xu, Z., Wang, Z., Sun, H., and Xin, H. (2020). Evaluation of Exosomal miRNA in Blood as a Potential Diagnostic Biomarker for Human Non-small Cell Lung Cancer. *Med. Sci. Monit.* 26, e924721. doi:10.12659/msm.924721
- Xue, Z., Li, S., Wang, S., Wang, Y., Yang, Y., Shi, J., et al. (2006). Mono-, Bi-, and Triphenanthrenes from the Tubers of *Cremastra appendiculata*. *J. Nat. Prod.* 69 (6), 907–913. doi:10.1021/np060087n
- Yang, H., Zhao, J., Zhao, M., Zhao, L., Zhou, L. N., Duan, Y., et al. (2020). GDC-0349 Inhibits Non-small Cell Lung Cancer Cell Growth. *Cell Death Dis.* 11 (11), 951. doi:10.1038/s41419-020-03146-w
- Yao, C., Su, L., Zhang, F., Zhu, X., Zhu, Y., Wei, L., et al. (2020). Thevebioside, the Active Ingredient of Traditional Chinese Medicine, Promotes Ubiquitin-Mediated SRC-3 Degradation to Induce NSCLC Cells Apoptosis. *Cancer Lett.* 493, 167–177. doi:10.1016/j.canlet.2020.08.011
- Ye, Z., Huang, Y., Ke, J., Zhu, X., Leng, S., and Luo, H. (2021). Breakthrough in Targeted Therapy for Non-small Cell Lung Cancer. *Biomed. Pharmacother.* 133, 111079. doi:10.1016/j.biopha.2020.111079
- Yu, X., Lin, Q., Liu, F., Yang, F., Mao, J., and Chen, X. (2020). LncRNA TMPO-AS1 Facilitates the Proliferation and Metastasis of NSCLC Cells by Up-Regulating ERBB2 via Sponging miR-204-3p. *Int. J. Immunopathol Pharmacol.* 34, 2058738420958947. doi:10.1177/2058738420958947
- Zhang, M. J., Liu, D., Liu, X. Z., Qu, B., Zhou, L., Liang, C. X., et al. (2019). Research Progress on the Chemical Composition of Mountain Cichlid Mushroom. *Agric. Development Equipment* (12), 35–36. (Chinese).
- Zhang, Q., Chen, C., Gong, S., and Huang, S. (2014). The Association between Genetic Variation of ESR1 and ESR2 Genes and Postpartum Depression in Women. *China Matern. Child Health Care* 29 (28), 4634–4637. (Chinese).
- Zheng, R., Qin, X., Li, W., and Kang, J. (2011). The Role of Src Protein in Lung Cancer Cell Proliferation and Infiltration. *Chin. J. Lung Cancer* 14 (04), 311–316.
- Zhou, L., Cai, X., Wang, L., Siyu, W., Lu, L., and Lian, X. (2016). Effect of β -sitosterol on the Proliferation and Apoptosis of Human Lung Cancer A549 Cells. *Int. J. Lab. Med.* 37 (07), 865–867. (Chinese).
- Zhu, C., Zhuang, W., Chen, L., Yang, W., and Ou, W.-B. (2020). Frontiers of ctDNA, Targeted Therapies, and Immunotherapy in Non-small-cell Lung Cancer. *Transl Lung Cancer Res.* 9 (1), 111–138. doi:10.21037/tlcr.2020.01.09

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 © 2021 Wang, Zhang, Wang, Shu, Lu, Shao, Liu, Yang, Luo and Du. 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.



Molecular Dynamics Simulations Reveal Interactions of an IgG1 Antibody With Selected Fc Receptors

Sebastjan Kralj^{1,2}, Milan Hodošek¹, Barbara Podobnik³, Tanja Kunej⁴, Urban Bren², Dušanka Janežič^{5*} and Janez Konc^{1,2*}

¹Theory Department, National Institute of Chemistry, Ljubljana, Slovenia, ²Laboratory of Physical Chemistry and Chemical Thermodynamics, Faculty of Chemistry and Chemical Engineering, University of Maribor, Maribor, Slovenia, ³Biologics Technical Development Mengeš, Technical Research and Development Novartis, Lek Pharmaceuticals d.d., Mengeš, Slovenia, ⁴Department of Animal Science, Biotechnical Faculty, University of Ljubljana, Ljubljana, Slovenia, ⁵Faculty of Mathematics, Natural Sciences and Information Technologies, University of Primorska, Koper, Slovenia

OPEN ACCESS

Edited by:

Fei Ye,
Zhejiang Sci-Tech University, China

Reviewed by:

Ruihan Zhang,
Yunnan University, China
Ting Shi,
Shanghai Jiao Tong University, China

*Correspondence:

Dušanka Janežič
dusanka.janezic@upr.si
Janez Konc
konc@cmm.ki.si

Specialty section:

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

Received: 06 May 2021

Accepted: 24 June 2021

Published: 02 July 2021

Citation:

Kralj S, Hodošek M, Podobnik B,
Kunej T, Bren U, Janežič D and Konc J
(2021) Molecular Dynamics
Simulations Reveal Interactions of an
IgG1 Antibody With Selected
Fc Receptors.
Front. Chem. 9:705931.
doi: 10.3389/fchem.2021.705931

In a survey of novel interactions between an IgG1 antibody and different Fc γ receptors (Fc γ R), molecular dynamics simulations were performed of interactions of monoclonal antibody involved complexes with Fc γ Rs. Free energy simulations were also performed of isolated wild-type and substituted Fc regions bound to Fc γ Rs with the aim of assessing their relative binding affinities. Two different free energy calculation methods, Molecular Mechanical/Generalized Born Molecular Volume (MM/GBMV) and Bennett Acceptance Ratio (BAR), were used to evaluate the known effector substitution G236A that is known to selectively increase antibody dependent cellular phagocytosis. The obtained results for the MM/GBMV binding affinity between different Fc γ Rs are in good agreement with previous experiments, and those obtained using the BAR method for the complete antibody and the Fc-Fc γ R simulations show increased affinity across all Fc γ Rs when binding to the substituted antibody. The Fc γ R1a, a key determinant of antibody agonistic efficacy, shows a 10-fold increase in binding affinity, which is also consistent with the published experimental results. Novel interactions between the Fab region of the antibody and the Fc γ Rs were discovered with this *in silico* approach, and provide insights into the antibody-Fc γ R binding mechanism and show promise for future improvements of therapeutic antibodies for preclinical studies of biological drugs.

Keywords: free energy calculation, homology modeling, molecular dynamics, fab-fc γ receptor interactions, monoclonal antibody, biological drugs

INTRODUCTION

As therapeutic agents, monoclonal antibodies possess key advantages over small-molecule drugs. These include target specificity, lower toxicity profiles, longer serum half-life and multiple cytotoxic modes of action. This versatility has led to a valuation predicted to be \$137–220 billion by the end of year 2020 for the antibody drug market. With this potential of antibodies, the pharmaceutical industry is searching for ways to improve existing therapies and cutting into the future market share (Grilo and Mantalaris, 2019). Currently, all FDA-approved therapeutic antibodies belong to the immunoglobulin isotype G (IgG) (Brezski and Georgiou, 2016), one of five isotypes of human antibodies or immunoglobulins (Franklin, 1975). The IgG antibody is a heterodimer consisting of two light chains and two heavy chains. The fragment antigen binding (Fab) domain is responsible for

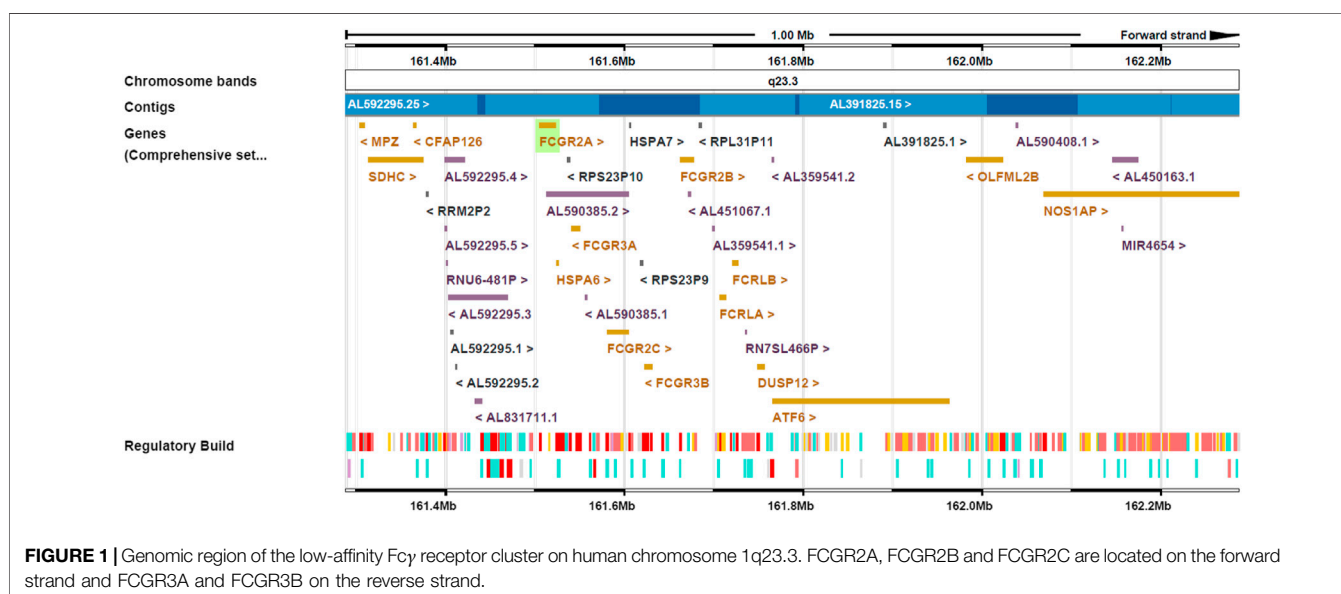
TABLE 1 | Characteristics of the FcγRs. Data extracted from the HGNC database (<https://www.genenames.org/>).

	Affinity	Function	Gene symbol	Gene name	Gene HGNC ID	Chromosomal location	Alias symbols
FcγRIa	High	Activating	<i>FCGR1A</i>	Fc fragment of IgG receptor Ia	3,613	1q21.2	CD64, CD64A
FcγRIIb	High	Activating	<i>FCGR1B</i>	Fc fragment of IgG receptor Ib	3,614	1p11.2	CD64b
FcγRIIa	Low	Activating	<i>FCGR2A</i>	Fc fragment of IgG receptor IIa	3,616	1q23.3	CD32, CD32A
FcγRIIb	Low	Inhibiting	<i>FCGR2B</i>	Fc fragment of IgG receptor IIb	3,618	1q23.3	CD32, CD32B
FcγRIIc	Low	Activating	<i>FCGR2C</i>	Fc fragment of IgG receptor IIc	15,626	1q23.3	hFcRII-C, D32C
FcγRIIIa	Low	Activating	<i>FCGR3A</i>	Fc fragment of IgG receptor IIIa	3,619	1q23.3	CD16, CD16a
FcγRIIIb	Low	Activating	<i>FCGR3B</i>	Fc fragment of IgG receptor IIIb	3,620	1q23.3	CD16, CD16b

specific antigen recognition, while the C-terminal part of both heavy chains forms the fragment crystallizable (Fc) domain. This domain is responsible for immune effector functions associated with antibodies.

The efficacy of many antibodies is associated with antibody dependent cellular cytotoxicity (ADCC), complement dependent cytotoxicity (CDC) and antibody dependent cellular phagocytosis (ADCP) to deplete target cells is mediated through interaction of the Fc region with the complementary component C1q or Fcγ receptors (FcγR). These are expressed in a broad spectrum of immune cells, and formation of an Fc/FcγR complex recruits these cells to sites of the bound antigen. The IgG antibodies predominantly elicit ADCC and ADCP by interacting with FcγRs. In humans, the FcγR protein family consists of FcγRI, FcγRII (subtypes a/b/c) and FcγRIII (subtypes a/b) (Raghavan and Bjorkman, 1996). All FcγRs bind the same region on the IgG Fc, with the FcγRI classified as a high affinity FcγRs and the FcγRII and FcγRIII as the low affinity FcγRs. The FcγRI, FcγRIIa/c and FcγRIIIa are activating receptors characterized by an intracellular immunoreceptor tyrosine-based activation motif, while the FcγRIIb is an inhibitory receptor characterized by an inhibition motif (Maenaka et al., 2001) (Table 1). The genomic region of the low-affinity Fcγ receptor cluster on human chromosome 1q23.3 is presented in Figure 1.

Antibodies possess multiple cytotoxic modes of action, but many have failed in clinical trials due to insufficient efficacy. This has led to attempts to increase their potency through enhancement of their ability to mediate cellular cytotoxicity functions such as ADCC and ADCP (Weiner and Carter, 2005). Further studies have found the Fc region to be essential for the therapeutic efficacy of antibodies that rely on ADCC or ADCP (Clynes et al., 2000; Arce Vargas et al., 2018; Lešnik et al., 2020). In order to achieve optimal therapeutic efficacy, specific sub-types of FcγR must engage with the Fc region. Increased binding to FcγRIIa or FcγRIIIa results in greater ADCP and ADCC activity, a desirable effect in many therapeutic antibodies, but an increase in FcγRIIb binding is desired for inhibitory antibodies. Achieving this is difficult as FcγRIIb and FcγRIIIa demonstrate ~92% homology in their extracellular domains despite the fact that they differ functionally. However, several successful engineered Fc variants with increased binding affinity to human FcγRIIIa have been reported (Lazar et al., 2006; Richards et al., 2008; Liu et al., 2014; Wang et al., 2018). These variants include the single mutants S239D and I332E, the double mutant S239D/I332E, and the triple mutant S239D/I332E/A330L (Liu et al., 2014). All of these variants have also been linked to enhanced ADCC activity (Lazar et al., 2006). On the spectrum of increasing FcγRIIa binding and enhancing



phagocytosis, the reported Fc variant G236A has selectively enhanced binding to FcγRIIa compared to FcγRIIb and mediates enhanced phagocytosis of antibody-coated target cells by macrophages (Richards et al., 2008; Wang et al., 2018).

We have examined the molecular dynamics of both the complete antibody structures and the isolated Fc region and have conducted binding free energy calculations to gain insight into their interactions with various FcγRs and to obtain directions which could lead to improved antibodies. We used homology modeling (Šali and Blundell, 1993) to obtain a new structure of a complete therapeutic IgG1 antibody, to which we examined the binding of FcγRIIa, FcγRIIb and FcγRIIIa using existing structures from the Protein Data Bank (PDB). Using the CHARMM biomolecular simulation program (Brooks et al., 2009) we performed 100 ns simulations for both the wild type (wt) antibody and the antibody with the G236A substitution in the lower hinge of the Fc region, a substitution known to increase the ADCP selectively (Richards et al., 2008). The calculated free energy values for the complete antibody simulations agree with the published experimental results (Richards et al., 2008), which show that FcγRIIIa enjoys a higher affinity than both FcγRIIa and FcγRIIb, with FcγRIIa having a higher affinity than FcγRIIb. However, the wild type and substituted antibody free energy values are not entirely in agreement with the published results for complete antibodies (Richards et al., 2008). The relatively higher agreement of the Fc-FcγR simulations with the experimental data is probably due to the negation of the effects of Fab-FcγR interactions seen in these simulations. With the more stringent conditions of the BAR free energy calculation method, a large increase in affinity is observed exclusively for FcγRIIa, as has been reported by Richards (Richards et al., 2008). The agreement of the calculated energies with the experimental data lends credibility to the *in silico* approach to future prospective evaluations of potential effector substitutions.

METHODS

Homology Modeling of the IgG1 Therapeutic Antibody and Preparation of Structures

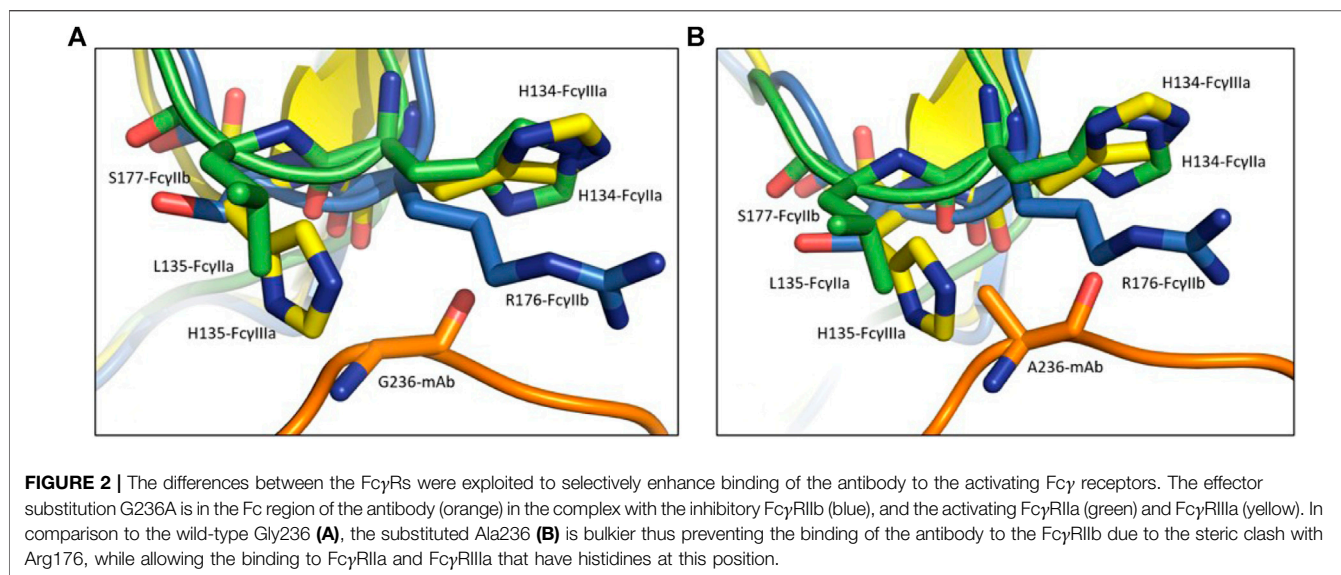
The sequence of the therapeutic IgG1 antibody in the FASTA format was obtained from the DrugBank database (<https://www.drugbank.ca>, accessed on date August 28, 2019). Structural templates for this sequence were found using the blastp algorithm (Altschul et al., 1990) with default settings. Only similar sequences found in the PDB were retained. The templates were selected based on sequence identity, query cover and resolution of the solved structure. For modeling of the heavy chain of a therapeutic antibody we used the structure of an intact human IgG1 (PDB ID:1HZH) (Saphire, 2001) which has 100% query cover and 84% sequence identity to the target sequence. 1HZH, a complete IgG1 antibody structure, enabled us to correctly spatially orient the Fab arms in our model. As monoclonal antibodies differ significantly in the Fab variable region which has both the light and heavy chains, an additional template was used to improve the quality of the model in this

region. The structure of ipilimumab bound to the human receptor CTLA-4 (PDB ID:5TRU) was selected based on the 100% query cover and 100% sequence identity to the light chain of the investigated therapeutic antibody (Ramagopal et al., 2017). High query cover and sequence identity are crucial for the final quality of the obtained homology model. Templates were aligned to their corresponding targets using the MUSCLE sequence alignment algorithm (Edgar, 2004). The alignment was checked for potential gaps or misaligned residues before the modeling. Protein models were constructed using the MODELLER software, builds a model of the protein by satisfying all spatial restraints (Šali and Blundell, 1993). The models obtained in this way were evaluated using the discrete optimized protein energy (DOPE) score. The lowest scoring model, with a DOPE value of −131,997 was further checked for quality using several homology model validating tools (see **Supplementary Material**) and was chosen for further work (Shen and Sali, 2006).

Models of the selected FcγRs complexed with the complete antibody structure were generated (**Figure 2**). The criteria for the selection of FcγR structures from the PDB were the resolution of the solved structure and the scope of the FcγR glycosylation profile, since glycosylation of both the Fc region and the FcγR play a crucial role in the binding mechanics of these two proteins (Hayes et al., 2014). The PDB structure of the FcγRIIIa bound to antibody Fc region (PDB ID:3SGK) contains the intact IgG1 Fc region bound to FcγRIIIa (Ferrara et al., 2011). The Fc region in the PDB structure 3SGK was superimposed on the modeled complete therapeutic IgG1 antibody to position the FcγRIIIa so that it was bound to our model. The redundant Fc region of 3SGK crystal structure was deleted to produce the final model of the antibody in a complex with FcγRIIIa. The other two models were then constructed by superimposing the structures of FcγRIIa (1FCG) and FcγRIIb (5OCC) onto the correctly positioned antibody (Hogarth et al., 1999; Sutton et al., 2018). This was possible because the selected FcγRs, FcγRIIa, FcγRIIb and FcγRIIIa bind to the same region of the IgG1 Fc region and in a similar conformation. Structural files for each of the FcγRs bound to the antibody were saved as PDB files to be used as inputs for molecular dynamics. For the isolated Fc regions, no homology modeling was performed. The PDB structure 3SGJ was chosen because of its extended hinge region which was not expected to interfere with the receptors during the simulation (Ferrara et al., 2011). The same three FcγRs used for the complete antibody simulations, were superimposed to the structure and saved as separate files.

Molecular Dynamics Simulations of IgG1-FcγR Complexes

We performed MD simulations of the complete IgG1 therapeutic antibody with selected FcγRs and simulations of the isolated Fc region with various bound FcγRs, using the latter to calculate the impact of substitution on binding free energy. Finally, we compared the *in silico* results obtained with experimental binding affinities. The use of MD simulations allows us to study the motion of our system through time. This is achieved



by numerically integrating Newton's second law of motion. The simulation inputs were prepared using the CHARMM-GUI web interface for the CHARMM biomolecular simulations program (Jo et al., 2008; Brooks et al., 2009). CHARMM-GUI Glycan Modeler was used to apply the most prevalent experimentally determined human glycosylation profile: {bDGal (14)bDGlNAc(1→2)aDMan (1→6) [bDGlNAc(1→2)-aDMan (1→3)] bDMan (1→4) bDGlNAc(1→4)[aLFuc (1→6)]bDGlNAc(1→)ASN-297)} for IgG1 antibodies based on experimental data (Sonneveld et al., 2018), as the homology model PDB structure was unglycosylated (Park et al., 2019). The protein structures were solvated using TIP3P water, and then neutralized using Na⁺ and Cl⁻ ions (0.1 M) to approximate physiological conditions. For the removal of steric clashes, that could be present after merging coordinate files of water molecules with proteins and, to optimize atomic coordinates of the complexes, 50 steps of steepest descent and 250 steps of adopted basis Newton-Raphson (ABNR) energy minimizations were performed. Both functions attempt to minimize the potential energy of the system, by slightly nudging the atomic coordinates of the protein followed by potential energy calculation and examination of the first derivatives to determine the direction of the gradient. Nudges of coordinates which result in lower potential energy i.e. moving towards a local minimum are saved, and the process is repeated with new coordinates until the specified step number.

This is followed by a short MD simulation during which the protein was equilibrated at 310.15 K using the HOOVER thermostat and the integration time-step set to 1 fs. The total length of equilibration molecular dynamics with NVT ensemble applied was 1 ns. Final molecular dynamics production runs were carried out using an NPT ensemble with periodic boundary conditions applied, the time-step set to 2 fs and the HOOVER thermostat set to 310.15 K. Van der

Waals interactions were cutoff between 10 and 12 Å using the force switch method (VFSWIt). Electrostatic potential used force shifting method (FSHift) with a cutoff of 12 Å. The particle-mesh Ewald summation (Darden et al., 1999) was used to calculate electrostatic interactions. Bonds to hydrogens were constrained using the SHAKE algorithm. This allows for a 1–2 fs integration step as otherwise unconstrained hydrogens, which have high frequency vibrating bonds lead to errors when integrating Newton's second law of motion. The force field, a simplified representation of reality allows us to derive the forces required for solving Newton's second law. The CHARMM36m force field was used for all simulations (Brooks et al., 2009; Guvench et al., 2011). For each of the selected FcγRs bound to either the complete antibody structure or the isolated Fc region production runs were generated using GPU acceleration with the final analysis performed on the last 100 ns for the complete antibodies and 200 ns for the isolated Fc region. In order to solve Newton's second law of motion velocities of atoms beside forces are required as well. Velocities are randomly generated at the start of the simulations. For this reason the first 20 ns of production runs were ignored to minimize the error arising from different initial velocities, as the additional time before sampling allows the protein to settle.

Calculation of the Binding Free Energy

The final result of MD simulations is the trajectory file that contains the information of how the protein moved in time. Beside visual cues that are offered from this file, we can calculate thermodynamic properties from it. In this paper two different approaches were used to calculate the relative binding free energies for the simulated complexes, the end-point Molecular Mechanical/Generalized Born molecular volume (MM/GBMV) method and the Bennett Acceptance Ratio (BAR) method (Bennett, 1976). The relative binding free energy was

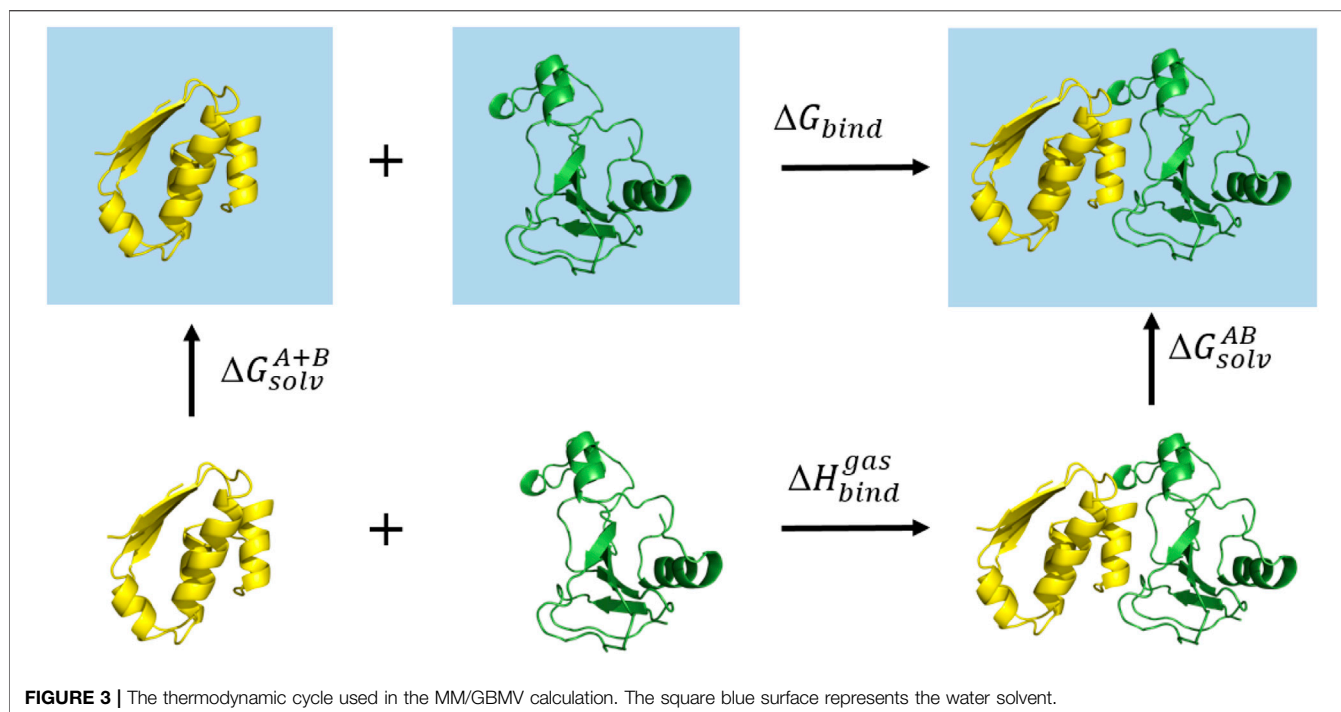


FIGURE 3 | The thermodynamic cycle used in the MM/GBMV calculation. The square blue surface represents the water solvent.

calculated, rather than the absolute binding free energy, since calculation of absolute binding free energies for biological events requires much longer simulations.

The MM/GBMV (Lee et al., 2002) method implemented in CHARMM decomposes the free energy of binding of the ligand, the protein and the complex, into contributions of different interactions (Figure 3) and can be expressed as follows:

$$\Delta G_{bind} = \Delta H - T \cdot \Delta S \approx \Delta E_{MM} + \Delta G_{sol} - T \cdot \Delta S \quad (1)$$

$$\Delta E_{MM} = \Delta E_{int} + \Delta E_{el-st} + \Delta E_{vdw} \quad (2)$$

$$\Delta G_{sol} = \Delta G_{generalized-Born} + \Delta G_{molecular\ volume} \quad (3)$$

In these equations, ΔE_{MM} represents the changes in the gas-phase molecular mechanics energy and includes changes in the internal energy (ΔE_{int}) (bond, angle and dihedral), electrostatic (ΔE_{el-st}) and the van der Waals energy (ΔE_{vdw}). The effect between the solute and the implicit solvent is described by ΔG_{sol} , which represents the sum of the polar and non-polar contribution to the desolvation free energy with the polar contribution ($\Delta G_{generalized-Born}$) calculated by the Generalized Born using Molecular Volume (GBMV) model implemented in CHARMM and the molecular volume contribution ($\Delta G_{molecular\ volume}$) estimated by the molecular volume calculation implemented within the GB module (Lee et al., 2002, 2003). The entropic contribution due to vibrational modes of the system to the binding free energy was neglected as the aim was to calculate relative binding free energies. Furthermore, the calculation of the entropic contribution of protein-ligand binding is only relevant in binding events where large conformational changes occur. Such conformational changes were not expected in this case, as we had very similar complexes (wild type and G236A), and

consequently we neglected the entropy term (Konc et al., 2013). For calculation of free energy using the MM/GBMV approach a single MD simulation of a protein-ligand complex was used, from which we obtained three separate trajectories of all components, the ligand, the apo protein and the protein-ligand complex (Lee and Olson, 2006). For the complete antibody simulations, all energy terms with the exception of the entropy terms were calculated for the final 100 ns (10,000 snapshots) of the production run for each MD trajectory. The energy terms for the isolated Fc region with the bound FcγR were calculated for 200 ns (20,000 snapshots).

The Bennett Acceptance Ratio (BAR) method is a rigorous method for calculating relative binding free energy but it offers greater accuracy (Bennett, 1976). Using the coupling parameter, λ to define intermediate states, it calculates the free energy difference between end-states A and B. This is useful as two similar systems, such as a wild type and substituted antibody can show very little overlap in phase space, making estimation of the free energy difficult when relying solely on the end states. In the production runs that were obtained, each snapshot of the simulation was submitted to an energy minimization of 100 steps of ABNR or until the specified tolerance (tolgrd) was 1.0 before calculating the energy. This was done to prevent steric clashes that might arise when inserting (mutating) the amino-acid into the non-mutated simulation snapshot. In the case of glycine to alanine the presence of the additional $-CH_3$ after mutation might cause steric clashes with the amino acids in the vicinity, giving false energy calculations. The non mutated simulation paths (0:0,1:1, Figure 4) were minimized as well to achieve consistency of energies. The energy data was filtered using an in-house python script, and the average $\Delta\Delta G$ differences for the simulations were obtained using the BAR script for

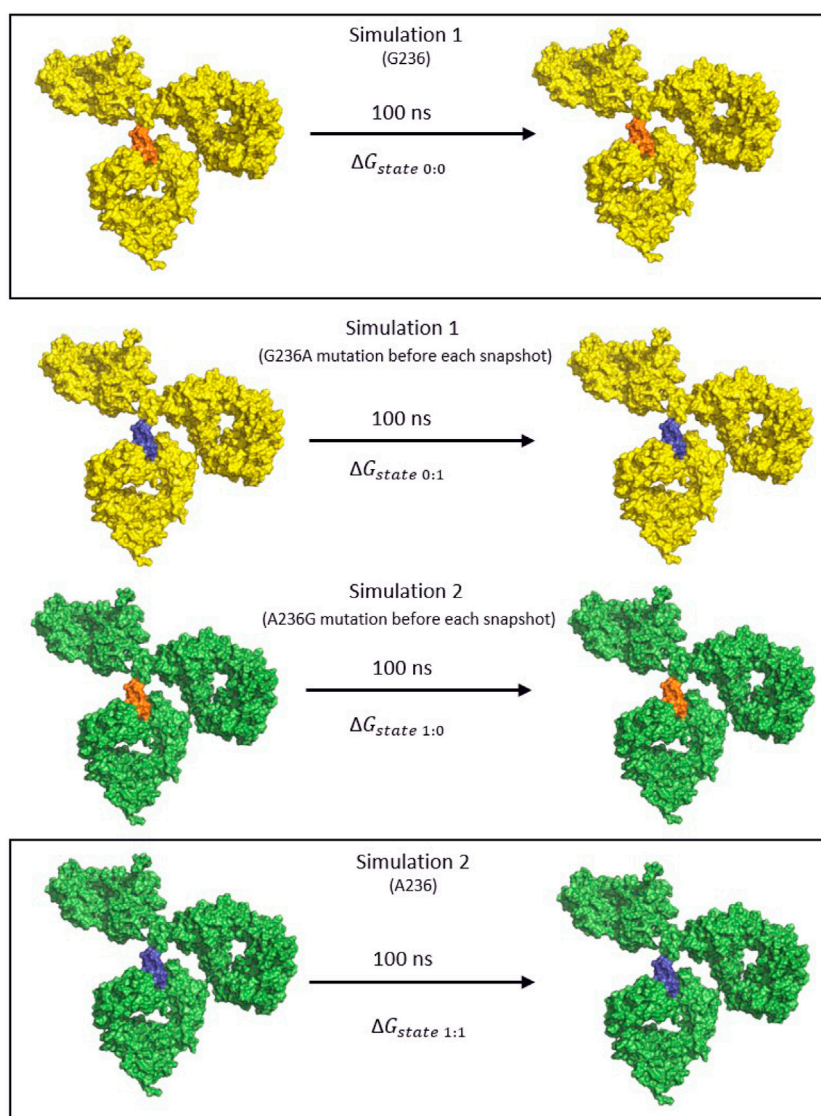


FIGURE 4 | The thermodynamic cycle for the calculation of the binding free energy using the BAR method. The IgG1 antibody is colored yellow for the first simulation and green for the second, the Gly236 variant is orange and the mutated Ala236 is blue. Two simulations were performed for each of the FcγR bound to the antibody. Subsequently, the intermediate trajectories 0→1 and 1→0 were obtained by mutating the original amino acid of the given simulation at every snapshot of the simulation before calculating the free energy terms.

CHARMM. No additional intermediate λ states were defined, as great overlap between the phase space of both simulations was achieved using only the end states (Figure 4).

RESULTS AND DISCUSSION

Free Energy Calculations of the Effector G236A Substitution

The effect of the G236A substitution on the binding affinity of the monoclonal antibody with different Fcγ receptors (FcγRs) was assessed by the free energy MM/GBMV and BAR methods using the simulations of the complete antibody as well as its isolated Fc

region. The calculated binding affinities are in agreement with the experiment by Richards et al. using the MM/GBMV method; in both simulations, FcγRIIIa exhibits the highest overall affinity of the three receptors, FcγRIIb the lowest, while FcγRIIa is intermediate (Figure 5).

It was also reported (Richards et al., 2008) that the G236A substitution in the Fc region of the antibody results in selective increase of the binding of FcγRIIa over that of FcγRIIb or FcγRIIIa. The complete antibody simulations do not confirm this (Figure 5A), as the calculated binding free energy of the wild type antibody in complex with FcγRIIa is -36.9 ± 20 kcal/mol, but increases to -25.2 ± 13 kcal/mol for the mutated antibody. This indicates that the G236A substitution decreases the binding

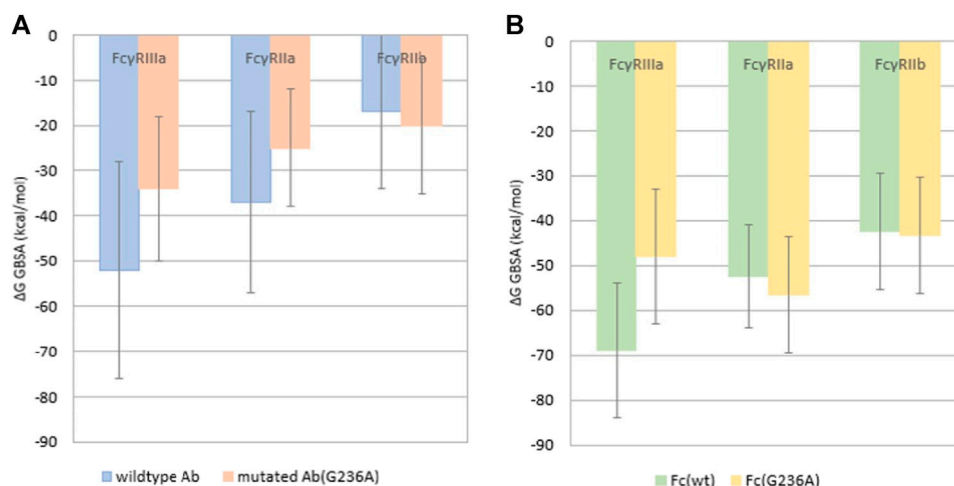


FIGURE 5 | Binding free energy calculated using the MM/GBMV method for the complex of **(A)** the complete antibody with FcγRs and **(B)** the isolated Fc region with FcγRs. The error bars represent standard deviations.

affinity. For the FcγRIIIa wild type antibody the binding affinity drops significantly, as the binding free energy for the wild type antibody is -52.1 ± 24 kcal/mol and for the substituted it is -33.8 ± 16 kcal/mol. When comparing the number of hydrogen bonds formed between the two antibodies (wild type and Ala236) bound to FcγRIIIa, the Ala236 antibody surprisingly forms more hydrogen bonds, on average 11.49 for the Ala236 antibody, and 4.85 for the wild type antibody. This difference could be due to different types of hydrogen bonds, with variable strengths. The binding affinity of the complete antibody towards FcγRIIb increases slightly from -17.4 ± 17 to -20 ± 15 kcal/mol, but the number of hydrogen bonds formed decreases from 4.79 (78%) in the wild type compared to 2.11 (32%) in the Ala236 antibody, suggesting a decrease in the binding affinity. When comparing the wild type antibody with the G236A substituted antibody we were unable to establish a clear correlation with the reference experimental values.

It was observed that the trajectories of the complete antibodies contained, in addition to the expected interactions between the Fc regions and the FcγRs, many interactions between the Fab regions and the FcγRs. It is thought that these Fab interactions may have disrupted the Fc-FcγRs interactions that were studied. This can be seen by comparison of **Figures 5A,B**, which shows that the binding affinity for the isolated Fc region (**Figure 5B**) is higher in all cases than the binding affinity of the complete antibody simulations (**Figure 5A**). This higher affinity for isolated Fc regions is thought to be due to the destabilizing effects of the Fab-FcγRs interactions that were seen with the complete antibody simulations, which may reduce the binding of the FcγRs with the complete antibody.

To examine the possible effects of the Fab region on the binding affinity we performed additional simulations using the isolated Fc regions in complexes with FcγRs (**Figure 5B**). Here, in accordance with existing data (Richards et al., 2008), the FcγRIIIa

exhibits a ~ 4 kcal/mol decrease in binding free energy, consequently an increase in binding affinity after introducing the G236A substitution, which drops from -53.2 ± 11 kcal/mol to -57.1 kcal/mol. The FcγRIIIa shows an increase in binding free energy, or a decrease in binding affinity after substitution, probably due to the steric clash of the His135 in FcγRIIIa with the Ala236 methyl group (**Figure 2**). The decrease in the binding affinity for FcγRIIIa is -21 kcal/mol, from -68.9 ± 15 kcal/mol for the wild type Fc region to -48 ± 15 kcal/mol for the Ala236 substituted Fc region. The FcγRIIb shows no significant increase in binding free energy, which goes from -42.4 ± 13 kcal/mol for the wild type Fc region to -43.3 ± 13 kcal/mol for the Ala236 substituted Fc region, indicating no effect of this substitution on binding affinity. These findings are in agreement with experimental results (Richards et al., 2008).

Results from the BAR method show that in fact the G236A substitution causes an increase of the binding affinity for all FcγRs (**Figure 6**). The effect of the Fab region on the FcγR binding is clearly visible as the simulation of the isolated Fc region shows a larger increase in binding affinity compared to the simulations of the complete antibody (*cf* **Figures 6A,B**). The value for the FcγRIIIa is in alignment with published data (Richards et al., 2008) as the binding free energy increase of -1.29 kcal/mol indicates an almost 10-fold increase in the binding affinity. The binding affinity increases along all of the receptors when the substitution is present, but the largest increase for the Fc-FcγR simulations is seen in the activation of FcγRIIIa, an important mediator of ADCP, compared to the inhibitory FcγRIIb and ADCC stimulating FcγRIIIa (**Figure 6B**). This indicates that introduction of this substitution strengthens the binding to FcγRIIIa, resulting in a higher activation of macrophages (ADCP) and a better therapeutic outcome, a result that has been observed experimentally (Richards et al., 2008).

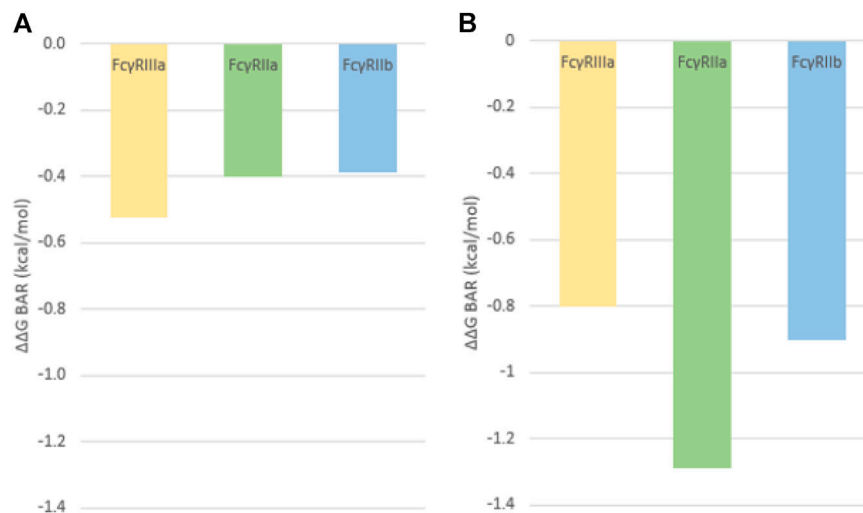


FIGURE 6 | Relative binding free energies calculated using the BAR method and representing the relative change in binding free energy after the induction of the G236A substitution for **(A)** the complete antibody simulation and **(B)** the isolated Fc region simulation.

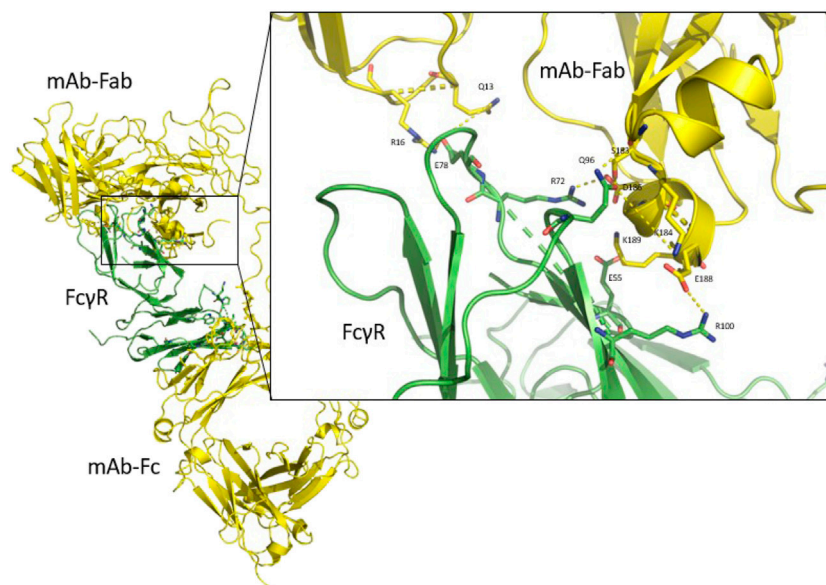


FIGURE 7 | The complete antibody structure (yellow) interacting with FcγR (green) with both the Fc region and the Fab region. Close-up of the interaction between the Fab region, which interacts predominantly with its α -helix (residues 183–190), and the FcγR. Hydrogen bond interactions are shown as yellow dashed lines.

Novel Interactions Between the FcγRs and the Fab Regions of the Therapeutic Antibody

Upon inspection of the obtained trajectories of all the FcγRIIa, FcγRIIIa and FcγRIIb FcγRs, we discovered novel interactions of the Fab region of the therapeutic antibody with the selected FcγRs (**Figure 7**). This was unexpected as it is generally

thought that interactions between an antibody and FcγRs occur exclusively in the Fc region of the antibody (Lu et al., 2015). The FcγR-Fab interactions observed in the simulations may be present *in vivo* and may play a previously unrecognized role in the binding of antibodies to FcγRs (Hogarth et al., 1999). To date, one study [see **Figure 6** in Ref (Yogo et al., 2019)] reports such interactions with the receptor FcγRIIIa, and similar interactions were observed using MD simulations

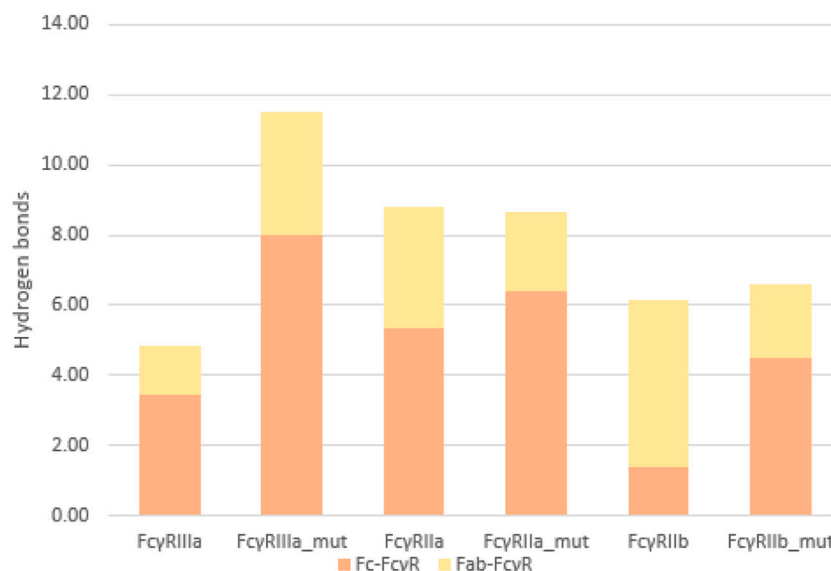


FIGURE 8 | Average number of hydrogen bonds formed between the FcγR and the antibody during simulation: comparison between the Fab-FcγR (yellow) interactions and the Fc-FcγR interactions (orange).

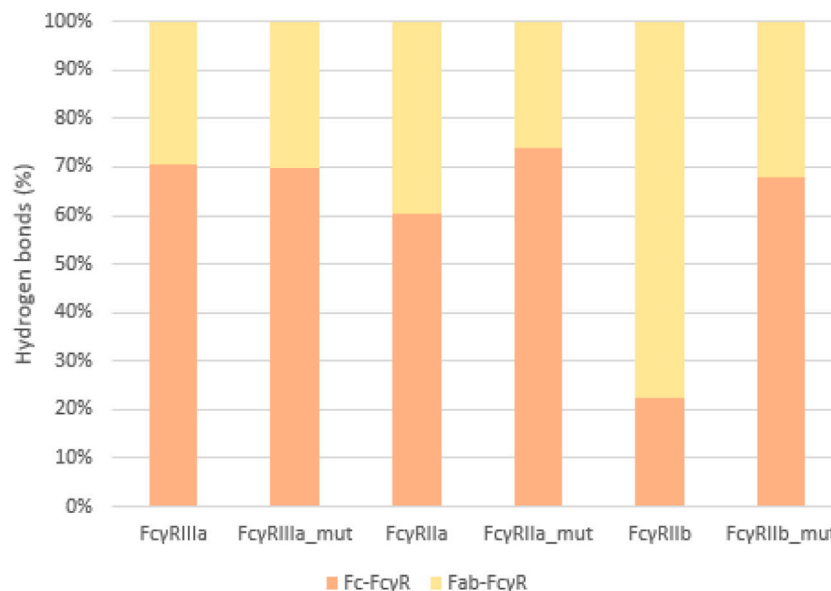


FIGURE 9 | Hydrogen bonds as percentage of all interactions. Comparison between Fc-FcγR (orange) and Fab-FcγR (yellow) interactions shows that hydrogen bonds formed between Fab region and FcγR represent at least 30% of the interactions, indicating their relevance.

with the FcγRI (Zhao et al., 2019). Our simulations show that antibodies can interact similarly with FcγRs through their Fab regions, confirming this previous report (Yogo et al., 2019).

Our simulations show that the CH1 domain of the antibody's Fab region forms the strongest interactions with the FcγR through an α -helix (residues 183–190) of the Fab light chain, which interacts with the upper region of the receptor (Figure 7, close-up). In addition, the loop (residues 12–17) of the heavy

chain CH1 is also seen to interact with the FcγR, as has been observed experimentally (Yogo et al., 2019).

To quantify the interactions between the antibody and FcγRs, we calculated the average number of hydrogen bonds formed during simulations between the individual regions (Fc or Fab) and the FcγRs (Figure 8). The significant percentage of hydrogen bonds between the Fab region and the FcγRs (Figure 9) indicates that the Fab region may indeed play an important role in binding

of the antibody to the FcγRs. Particularly, the wild type antibody in complex with FcγRIIb had predominantly the Fab region binding with the receptor, as in this complex, and the bonds between the Fab and the FcγR bonds amounted to 77% of all hydrogen bonds formed between the antibody and the FcγR. Across all the simulated complexes, the average proportion of the Fab-FcγR hydrogen bonds is 38% or 2.7 hydrogen bonds. These newly discovered hydrogen bond interactions reaffirm that the Fab region influences the binding of FcγRs, which has been reported in just one study to date (Yogo et al., 2019). Further, the comparison of the calculated binding free energies (Figure 5) and the hydrogen bond interactions (Figure 8) suggests that the interactions formed between the Fab region and FcγRs decrease the binding affinity of the Fc region for the receptors. This decrease in affinity however, is partially compensated for by the formation of new hydrogen bonds with the Fab region.

CONCLUSIONS

Using the free energy simulation methods MM/GBMV and BAR, we have shown that the known substitution (G236A) has a selective effect on the binding of the antibody with FcγRs. The substitution increases the binding with FcγRIIa, thereby increasing ADCP, and to a lesser extent it increases binding affinity for the inhibitory FcγRIIb and the ADCC-activating FcγRIIIa, consistent with experiments (Richards et al., 2008). Through simulations of the complete antibody, we found novel interactions between the Fab region of the antibody and the FcγRs, which were experimentally determined recently using high-speed atomic force microscopy (Yogo et al., 2019). Our

simulation results should be very valuable for future improvement of therapeutic monoclonal antibodies and could contribute to the development of new antibody therapeutic approaches.

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 authors.

AUTHOR CONTRIBUTIONS

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

ACKNOWLEDGMENTS

Funding from the Slovenian Research Agency under project numbers L7-8269, N1-0142, J1-1715, J1-9186 and N1-0209 is acknowledged.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fchem.2021.705931/full#supplementary-material>

REFERENCES

- Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic Local Alignment Search Tool. *J. Mol. Biol.* 215, 403–410. doi:10.1016/S0022-2836(05)80360-2
- Arce Vargas, F., Furness, A. J. S., Litchfield, K., Joshi, K., Rosenthal, R., Ghorani, E., et al. (2018). Fc Effector Function Contributes to the Activity of Human Anti-CTLA-4 Antibodies. *Cancer Cell* 33, 649–e4. doi:10.1016/j.ccell.2018.02.010
- Bennett, C. H. (1976). Efficient Estimation of Free Energy Differences from Monte Carlo Data. *J. Comput. Phys.* 22, 245–268. doi:10.1016/0021-9991(76)90078-4
- Brezski, R. J., and Georgiou, G. (2016). Immunoglobulin Isotype Knowledge and Application to Fc Engineering. *Curr. Opin. Immunol.* 40, 62–69. doi:10.1016/j.coi.2016.03.002
- Brooks, B. R., Brooks, C. L., Mackerell, A. D., Nilsson, L., Petrella, R. J., Roux, B., et al. (2009). CHARMM: The Biomolecular Simulation Program. *J. Comput. Chem.* 30, 1545–1614. doi:10.1002/jcc.21287
- Clynes, R. A., Towers, T. L., Presta, L. G., and Ravetch, J. V. (2000). Inhibitory Fc Receptors Modulate *In Vivo* Cytotoxicity against Tumor Targets. *Nat. Med.* 6, 443–446. doi:10.1038/74704
- Darden, T., Perera, L., Li, L., and Pedersen, L. (1999). New Tricks for Modelers from the Crystallography Toolkit: the Particle Mesh Ewald Algorithm and its Use in Nucleic Acid Simulations. *Structure* 7, R55–R60. doi:10.1016/S0969-2126(99)80033-1
- Edgar, R. C. (2004). MUSCLE: Multiple Sequence Alignment with High Accuracy and High Throughput. *Nucleic Acids Res.* 32, 1792–1797. doi:10.1093/nar/gkh340
- Ferrara, C., Grau, S., Jager, C., Sondermann, P., Brunker, P., Waldhauer, I., et al. (2011). Unique Carbohydrate-Carbohydrate Interactions Are Required for High Affinity Binding between Fc RIII and Antibodies Lacking Core Fucose. *Proc. Natl. Acad. Sci.* 108, 12669–12674. doi:10.1073/pnas.1108455108
- Franklin, E. C. (1975). Structure and Function of Immunoglobulins. *Acta Endocrinologica* 80, S77–S95. doi:10.1530/acta.0.080S077
- Grilo, A. L., and Mantalaris, A. (2019). The Increasingly Human and Profitable Monoclonal Antibody Market. *Trends Biotechnol.* 37, 9–16. doi:10.1016/j.tibtech.2018.05.014
- Guvench, O., Mallajosyula, S. S., Raman, E. P., Hatcher, E., Vanommeslaeghe, K., Foster, T. J., et al. (2011). CHARMM Additive All-Atom Force Field for Carbohydrate Derivatives and its Utility in Polysaccharide and Carbohydrate-Protein Modeling. *J. Chem. Theor. Comput.* 7, 3162–3180. doi:10.1021/ct200328p
- Hayes, J. M., Cosgrave, E. F. J., Struwe, W. B., Wormald, M., Davey, G. P., Jefferis, R., et al. (2014). “Glycosylation and Fc Receptors,” in *Fc Receptors Current Topics in Microbiology and Immunology*. Editors M. Daeron and F. Nimmerjahn (Cham: Springer International Publishing), 165–199. doi:10.1007/978-3-319-07911-0_8
- Hogarth, P. M., Maxwell, K. F., Powell, M. S., Hulett, M. D., Barton, P. A., McKenzie, I. F. C., et al. (1999). Crystal Structure of the Human Leukocyte Fc Receptor, Fc gammaRIIa. *Nat. Struct. Biol.* 6, 437–442. doi:10.1038/8241
- Jo, S., Kim, T., Iyer, V. G., and Im, W. (2008). CHARMM-GUI: A Web-Based Graphical User Interface for CHARMM. *J. Comput. Chem.* 29, 1859–1865. doi:10.1002/jcc.20945
- Konc, J., Hodošek, M., Ogrizek, M., Trykowska Konc, J., and Janežič, D. (2013). Structure-Based Function Prediction of Uncharacterized Protein Using Binding Sites Comparison. *Plos Comput. Biol.* 9, e1003341. doi:10.1371/journal.pcbi.1003341

- Lazar, G. A., Dang, W., Karki, S., Vafa, O., Peng, J. S., Hyun, L., et al. (2006). Engineered Antibody Fc Variants with Enhanced Effector Function. *Proc. Natl. Acad. Sci.* 103, 4005–4010. doi:10.1073/pnas.0508123103
- Lee, M. S., Feig, M., Salsbury, F. R., and Brooks, C. L. (2003). New Analytic Approximation to the Standard Molecular Volume Definition and its Application to Generalized Born Calculations. *J. Comput. Chem.* 24, 1348–1356. doi:10.1002/jcc.10272
- Lee, M. S., and Olson, M. A. (2006). Calculation of Absolute Protein-Ligand Binding Affinity Using Path and Endpoint Approaches. *Biophysical J.* 90, 864–877. doi:10.1529/biophysj.105.071589
- Lee, M. S., Salsbury, F. R., and Brooks, C. L. (2002). Novel Generalized Born Methods. *J. Chem. Phys.* 116, 10606–10614. doi:10.1063/1.1480013
- Lešnik, S., Hodošek, M., Podobnik, B., and Konc, J. (2020). Loop Grafting between Similar Local Environments for Fc-Silent Antibodies. *J. Chem. Inf. Model.* 60, 5475–5486. doi:10.1021/acs.jcim.9b01198
- Liu, Z., Gunasekaran, K., Wang, W., Razinkov, V., Sekirov, L., Leng, E., et al. (2014). Asymmetrical Fc Engineering Greatly Enhances Antibody-dependent Cellular Cytotoxicity (ADCC) Effector Function and Stability of the Modified Antibodies. *J. Biol. Chem.* 289, 3571–3590. doi:10.1074/jbc.M113.513366
- Lu, J., Chu, J., Zou, Z., Hamacher, N. B., Rixon, M. W., and Sun, P. D. (2015). Structure of FcγRI in Complex with Fc Reveals the Importance of Glycan Recognition for High-Affinity IgG Binding. *Proc. Natl. Acad. Sci. USA* 112, 833–838. doi:10.1073/pnas.1418812112
- Maenaka, K., van der Merwe, P. A., Stuart, D. I., Jones, E. Y., and Sondermann, P. (2001). The Human Low Affinity Fcγ Receptors IIa, IIb, and III Bind IgG with Fast Kinetics and Distinct Thermodynamic Properties. *J. Biol. Chem.* 276, 44898–44904. doi:10.1074/jbc.M106819200
- Park, S.-J., Lee, J., Qi, Y., Kern, N. R., Lee, H. S., Jo, S., et al. (2019). CHARMM-GUIGlycan Modeler for Modeling and Simulation of Carbohydrates and Glycoconjugates. *Glycobiology* 29, 320–331. doi:10.1093/glycob/cwz003
- Raghavan, M., and Bjorkman, P. J. (1996). Fc Receptors and Their Interactions with Immunoglobulins. *Annu. Rev. Cel Dev. Biol.* 12, 181–220. doi:10.1146/annurev.cellbio.12.1.181
- Ramagopal, U. A., Liu, W., Garrett-Thomson, S. C., Bonanno, J. B., Yan, Q., Srinivasan, M., et al. (2017). Structural Basis for Cancer Immunotherapy by the First-In-Class Checkpoint Inhibitor Ipilimumab. *Proc. Natl. Acad. Sci. USA* 114, E4223–E4232. doi:10.1073/pnas.1617941114
- Richards, J. O., Karki, S., Lazar, G. A., Chen, H., Dang, W., and Desjarlais, J. R. (2008). Optimization of Antibody Binding to FcγRIIa Enhances Macrophage Phagocytosis of Tumor Cells. *Mol. Cancer Ther.* 7, 2517–2527. doi:10.1158/1535-7163.MCT-08-0201
- Šali, 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
- Saphire, E. O. (2001). Crystal Structure of a Neutralizing Human IgG against HIV-1: A Template for Vaccine Design. *Science* 293, 1155–1159. doi:10.1126/science.1061692
- Shen, M.-V., and Sali, A. (2006). Statistical Potential for Assessment and Prediction of Protein Structures. *Protein Sci.* 15, 2507–2524. doi:10.1110/ps.062416606
- Sonneveld, M. E., Koeleman, C. A. M., Plomp, H. R., Wührer, M., van der Schoot, C. E., and Vidarsson, G. (2018). Fc-Glycosylation in Human IgG1 and IgG3 Is Similar for Both Total and Anti-red-blood Cell Anti-K Antibodies. *Front. Immunol.* 9, 129. doi:10.3389/fimmu.2018.00129
- Sutton, E. J., Bradshaw, R. T., Orr, C. M., Freundéus, B., Larsson, G., Teige, I., et al. (2018). Evaluating Anti-CD32b F(ab) Conformation Using Molecular Dynamics and Small-Angle X-Ray Scattering. *Biophysical J.* 115, 289–299. doi:10.1016/j.bpj.2018.03.040
- Wang, X., Mathieu, M., and Brezski, R. J. (2018). IgG Fc Engineering to Modulate Antibody Effector Functions. *Protein Cell* 9, 63–73. doi:10.1007/s13238-017-0473-8
- Weiner, L. M., and Carter, P. (2005). Tunable Antibodies. *Nat. Biotechnol.* 23, 556–557. doi:10.1038/nbt0505-556
- Yogo, R., Yamaguchi, Y., Watanabe, H., Yagi, H., Satoh, T., Nakanishi, M., et al. (2019). The Fab Portion of Immunoglobulin G Contributes to its Binding to Fcγ Receptor III. *Sci. Rep.* 9, 11957. doi:10.1038/s41598-019-48323-w
- Zhao, J., Nussinov, R., and Ma, B. (2019). Antigen Binding Allosterically Promotes Fc Receptor Recognition. *mAbs* 11, 58–74. doi:10.1080/19420862.2018.1522178

Conflict of Interest: BP was employed by the company Lek Pharmaceuticals d.d.

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 © 2021 Kralj, Hodošek, Podobnik, Kunej, Bren, Janežič and Konc. 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 Insights Into the Inhibition Mechanism of Proanthocyanidin B2 on Tau Hexapeptide (PHF6) Oligomer

Qin Li¹, Chunmei Xiong¹, Hongli Liu², Huizhen Ge³, Xiaojun Yao³ and Huanxiang Liu^{1*}

¹School of Pharmacy, Lanzhou University, Lanzhou, China, ²Jiangsu Key Laboratory of New Drug Research and Clinical Pharmacy, Xuzhou Medical University, Xuzhou, China, ³State Key Laboratory of Applied Organic Chemistry and Department of Chemistry, Lanzhou University, Lanzhou, China

OPEN ACCESS

Edited by:

Zhongjie Liang,
Soochow University, China

Reviewed by:

Guang Hu,
Soochow University, China
Tingjun Hou,
Zhejiang University, China

*Correspondence:

Huanxiang Liu
hxliu@lzu.edu.cn

Specialty section:

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

Received: 09 February 2021

Accepted: 14 April 2021

Published: 14 July 2021

Citation:

Li Q, Xiong C, Liu H, Ge H, Yao X and
Liu H (2021) Computational Insights
Into the Inhibition Mechanism of
Proanthocyanidin B2 on Tau
Hexapeptide (PHF6) Oligomer.
Front. Chem. 9:666043.
doi: 10.3389/fchem.2021.666043

The formation of amyloid fibrils from Tau is a key pathogenic feature of Alzheimer's disease (AD). To disturb the formation of Tau aggregates is considered as a promising therapeutic strategy for AD. Recently, a natural product proanthocyanidin B2 (PB2) was confirmed to not only inhibit Tau aggregation, but also disaggregate Tau fibrils. Herein, to explore the inhibition mechanism of PB2 against Tau fibril and to provide the useful information for drug design and discovery, all-atom molecular dynamics simulations were carried out for the ordered Tau hexapeptide PHF6 oligomer in the presence and absence of PB2. The obtained result shows that PB2 can transform PHF6 oligomer from the ordered β -sheet structure into disordered one. Moreover, the clustering analysis and binding free energy calculations identify that S3 site is the most potential binding site. At S3 site, by hydrophobic and hydrogen bond interactions, the residues V309, Y310 and K311 are essential for binding with PB2, especially K311. In a word, our study reveals the molecular mechanism of PB2 inhibiting PHF6 aggregation and it will provide some valuable information for the development of Tau aggregation inhibitors.

Keywords: PHF6, Tau aggregation, proanthocyanidin B2, molecular dynamics simulation, Alzheimer's disease

INTRODUCTION

Alzheimer's disease (AD), as the most common neurodegenerative disorder, is clinically distinguished by progressive declines in cognitive functions, causing severe dementia (Goedert and Spillantini, 2006). It is also the sixth leading cause of death in the United States. There are approximately 5.8 million people diagnosed with AD and the cost of care for these individuals in 2019 is about \$244 billion, causing an enormous psychological and economic stress on families, caregivers, and the health care system in the United States (Alzheimer's Association, 2020). AD is histopathologically identified by the presence of extracellular amyloid plaques composed of amyloid-beta ($A\beta$) and intracellular neurofibrillary tangles (NFTs) composed of hyperphosphorylated Tau proteins in paired helical (PHFs) or straight filaments (SFs) (Crowther, 1991; Bloom, 2014; Götz et al., 2019). Tau hyperphosphorylation triggers neurodegeneration due to Tau propagation and aggregation into NFTs (Spillantini and Goedert, 2013). Based on the widely accepted amyloid cascade hypothesis, the aggregation and spreading of Tau seems to be facilitated by aggregation of $A\beta$. Nevertheless, many compounds targeting $A\beta$ have failed to demonstrate efficacy in slowing disease progression during clinical trials (Aisen et al., 2006; Wilcock et al., 2008; Chiang and Koo,

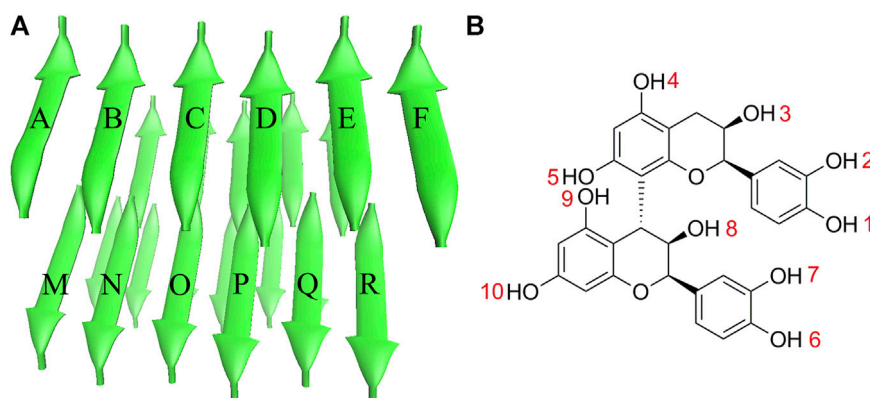


FIGURE 1 | The initial structures of (A) PHF6 oligomer and (B) PB2.

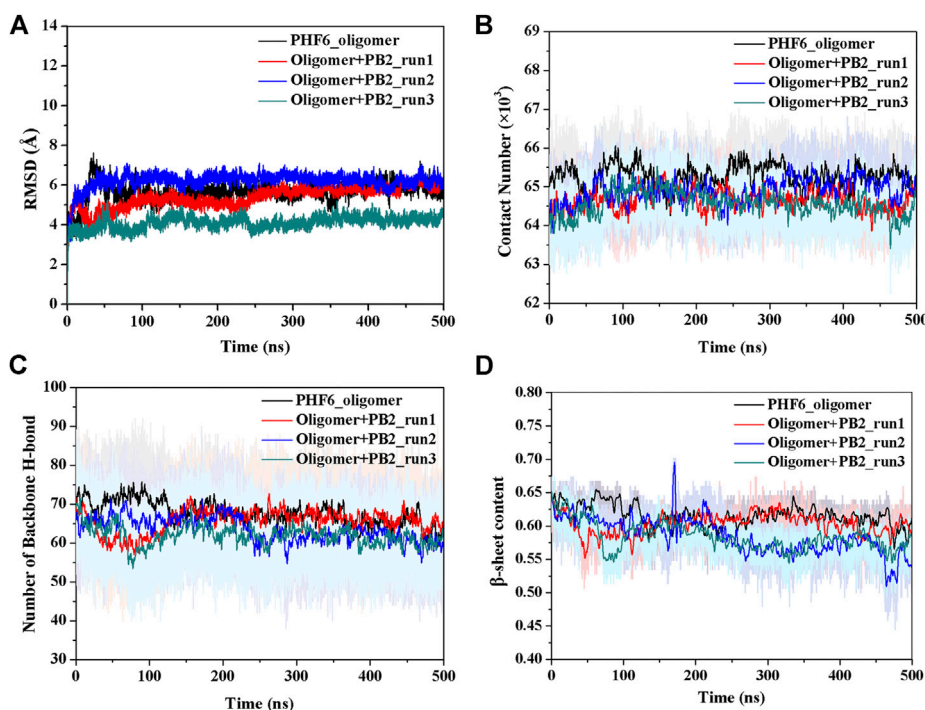


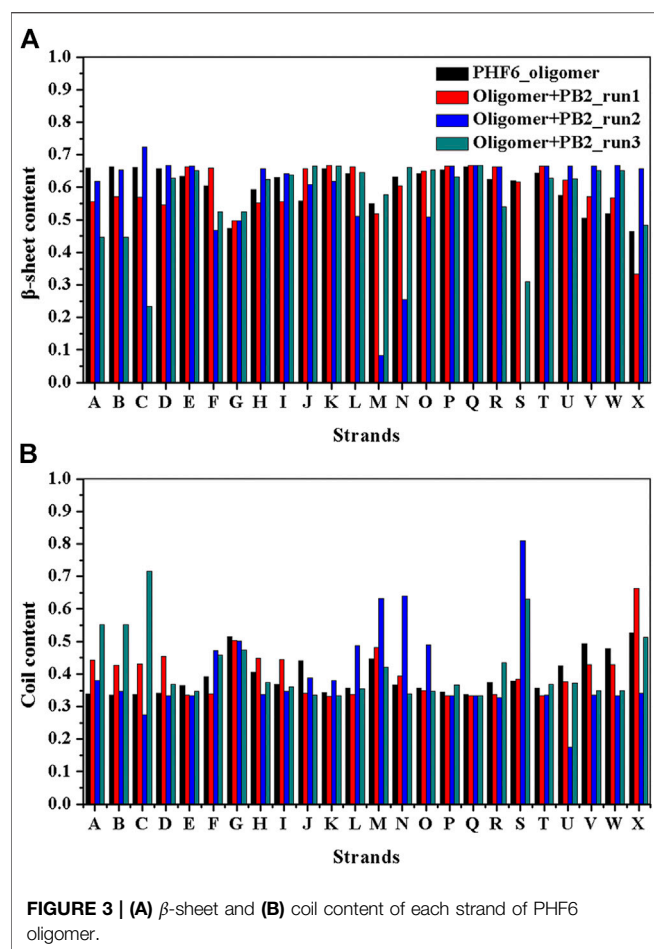
FIGURE 2 | The monitoring of structural characteristics of PHF6 oligomer calculated from each run. (A) The RMSD of protein backbone atoms. (B) The contact number between peptides. (C) The number of backbone H-bonds between peptides. (D) Time evolution of β -sheet content.

2014). Moreover, recent research has suggested that compared to A β pathology, Tau pathology that is described as the accumulation of Tau and the deposition of NFTs has better correlation with disease severity (Arriagada et al., 1992; Kametani and Hasegawa, 2018). In this context, considerable attention is now focused on targeting Tau as a therapeutic strategy for AD.

Tau is a microtubule-associated protein (MAP), abundantly expressed in the central nervous system. Under normal conditions, Tau acts as a cytoskeleton stabilizer through its interface with tubulin heterodimers (Kadavath et al., 2015).

But in the certain conditions, Tau may experience some abnormal post-translational modifications including hyperphosphorylation, acetylation, methylation, ubiquitination and cleavage. Abnormally phosphorylated Tau protein no longer binds to microtubules, but assembles into fibrils which are insoluble and toxic, leading to neuronal death (Grundke-Iqbal et al., 1986; Alonso et al., 2001; Šimić et al., 2016).

To develop new therapeutic agents in AD, several Tau-based therapeutic approaches are currently emerging, including Tau phosphorylation inhibitors (Mazanetz and Fischer, 2007), microtubule stabilizers (Zhang et al., 2005), Tau aggregation



inhibitors (Wischnik et al., 1996; Pickhardt et al., 2015), and immune therapy (Kontseikova et al., 2014). Among them, the most widely studied are Tau aggregation inhibitors. A screening from over 200,000 compounds finds that polyphenols, phenothiazines, anthraquinones and porphyrins are capable of inhibiting Tau fibril formation not only *in vitro* but also in cultured cells (Pickhardt et al., 2005; Taniguchi et al., 2005; Crowe et al., 2007), in addition, some compounds are under clinical trials (Gauthier et al., 2016). Attractively, natural polyphenolic compounds such as myricetin, curcumin, oleocanthal and EGCG have been found to have anti-amyloid effects that prevent amyloid aggregation and fibril formation (Taniguchi et al., 2005; Li et al., 2009; Wobst et al., 2015; Rane et al., 2017). Compared to synthetic compounds, natural polyphenols from food or herbal extracts usually exhibit higher availability, stability, convenience and lower side effects. Proanthocyanidins, the most abundant polyphenols present in human diets, are potentially effective in the prevention and treatment for AD due to their antioxidant and neuroprotective capacity (Zhao et al., 2019). Proanthocyanidin B2 (PB2), a major type of proanthocyanidins, has been reported to cross blood-brain-barrier and have potent inhibitory activity on Tau and A β aggregates for the treatment in AD. It is also shown that PB2 can not only inhibit Tau aggregation, but also disaggregate Tau fibrils

(Snow et al., 2019). However, the potential mechanism of PB2 exerting its effects is still unclear.

This study was to explore the inhibition mechanism of PB2 on Tau oligomer at atom level, where the used oligomer is formed of a hexapeptide motif ³⁰⁶VQIVYK³¹¹ (PHF6), the most important nucleation sequence in Tau aggregation. PHF6 has been reported to self-assemble to form the steric-zipper conformation composed of an ordered antiparallel-layered parallel β -sheet structures (von Bergen et al., 2000; Sawaya et al., 2007; Plumley and Dannenberg, 2010). Moreover, PHF6 is capable of forming fibrils *in vitro* similar to those formed by full-length Tau (Goux et al., 2004; Rojas Quijano et al., 2006). In this work, we started with a preformed PHF6 oligomer and then performed all-atom molecular dynamics (MD) simulations to explore the inhibition mechanism of PB2 on PHF6 oligomer. Contrast to the traditional experimental approaches, molecular dynamics simulation method can obtain more structural dynamics information of protein and clarify the significant effects of inhibitors of amyloid protein (Wang et al., 2015; Liu et al., 2018). It is also able to predict the detailed binding mode of the inhibitor and search for the key residues of PHF6 oligomer. The results will give some helpful guidance for the discovery and design of Tau aggregation inhibitors in the future.

COMPUTATIONAL METHODS

Preparation of Starting Structures

The stable PHF6 oligomer was used to investigate the disaggregation of PB2 on Tau fibril. The three-dimensional coordinates of PHF6 were gained from the Protein Data Bank (PDB ID: 2ON9) (Sawaya et al., 2007). Then PyMOL software version 1.3 (DeLano Scientific LLC) was applied to construct the three-dimensional structure of PHF6 oligomer, consisting of four sheets and six strands per sheet, made of a total of 24 PHF6 monomers. In order to neutralize the N- and C-terminals in peptide strands, ACE and NME residues were added to cap the N- and C-terminals. The obtained starting PHF6 oligomer structure as well as the structure of PB2 was given in **Figure 1**. Here, 24 strands were numbered as A-X sequentially. The applied molar ratio of PB2/PHF6 was about 1:5 in this model according to the experimental condition (Snow et al., 2019). Therefore, in each system, five PB2 molecules were randomly placed around PHF6 oligomer, and their minimum distance from the oligomer were at least 8 Å. Gaussian 09 software (Frisch et al., 2009) was used to optimize the structure of PB2 at the Hartree-Fock level with 6-31G* basis set. The partial atomic charges were derived using RESP fitting technique (Bayly et al., 1993). The GAFF force field (Wang et al., 2004) and the Amber ff99SB force field (Hornak et al., 2006) was applied to describe PB2 and the oligomer, respectively.

Details of Molecular Dynamics Simulations

All molecular dynamics simulations were performed using Amber18 software (Case et al., 2018). Each system prepared for simulation was placed in a cube periodic box filled with TIP3P water (Jorgensen et al., 1983) molecules, with more than

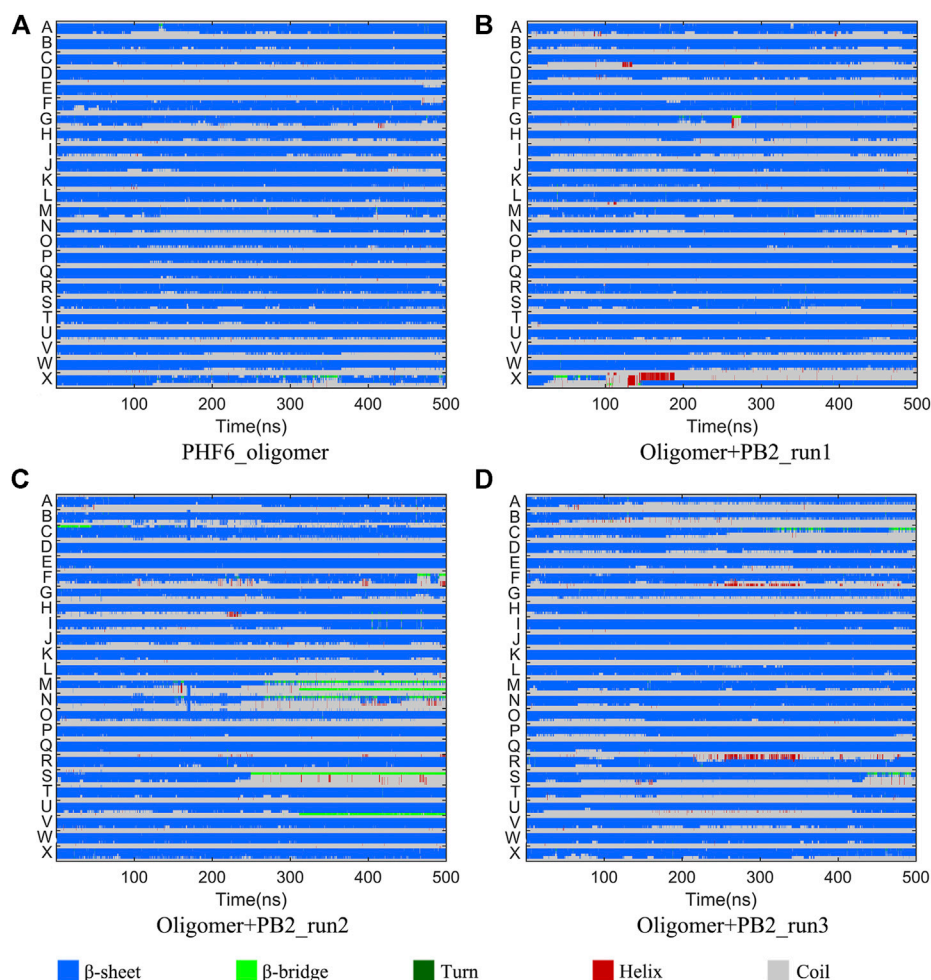


FIGURE 4 | Secondary structure changes of each strand in four systems.

12 Å distance around the oligomer. In order to maintain the electrical neutrality of the system, an appropriate amount of chloride ions were added to each system. Subsequently, the steepest descent method and the conjugate gradient method were used to optimize the system to eliminate unnatural collisions. Then, each system was heated from 0 to 300 K under the NVT ensemble. In the meantime, all oligomer atoms were constrained by a harmonic force of 5.0 kcal/(mol·Å²). And then the further five steps equilibrium process was performed in the NPT ensemble with decreased restraint force on the complexes from 5.0 to 0 kcal/(mol·Å²) to release all the restraints. Finally, 500 ns molecular dynamics simulations were performed without any restraints. The temperature was controlled by the Langevin thermostat. The SHAKE algorithm (Ryckaert et al., 1977) was used to limit the bond length concerning hydrogen atoms. The particle mesh Ewald (PME) method (Essmann et al., 1995) was used to calculate long-range electrostatic interactions. Totally, four separate trajectories which include three parallel runs for PHF6 oligomer with PB2 (oligomer + PB2) and one for PHF6 oligomer without PB2

(PHF6_oligomer), were performed to explore the disrupting mechanism of PB2 against PHF6 oligomer.

Molecular Dynamics Trajectory Analysis

All the trajectory analysis was performed in Amber and VMD programs (Humphrey et al., 1996). The contact between strands of PHF6 oligomer is considered to be formed when the distance between the pair of heavy atoms is less than 4.0 Å. The hydrogen bond is considered to be formed when the hydrogen-acceptor distance is less than 3.5 Å and the donor-hydrogen-acceptor angle should be larger than 120°. Principal components analysis (PCA) (Amadei et al., 1993) was applied to obtain the first two eigenvectors to draw the free energy landscape. Secondary structure tendency for every residue was calculated by employing the DSSP method (Kabsch and Sander, 1983). The K-means clustering algorithm (Feig et al., 2004) was applied to cluster the geometrically similar conformations. The molecular mechanics/generalized Born surface area (MM-GBSA) method (Hou et al., 2011) was used to calculate the binding free energy between the oligomer and PB2. By MM-GBSA approach, 5,000

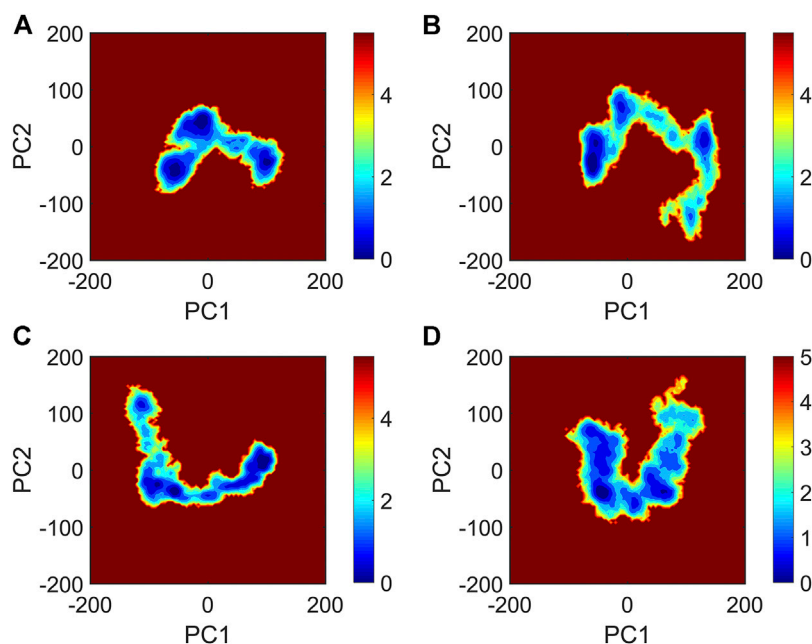


FIGURE 5 | The free energy landscape map of (A) PHF6_oligomer, (B) PHF6_oligomer + PB2_run1, (C) PHF6_oligomer + PB2_run2 and (D) PHF6_oligomer + PB2_run3.

snapshots from the last 100 ns were extracted to calculate the binding free energy between protein PHF6 oligomer and PB2. The calculated interaction energy was further decomposed to each residue to obtain the contribution of each residue to the binding energy.

RESULTS AND DISCUSSION

The Stability of Studied Systems

The convergence of four simulations was firstly examined to monitor if the simulations were up to the equilibration. First, the root mean square deviation (RMSD) of backbone atoms was calculated for all four trajectories. As shown in **Figure 2A**, the RMSD values of all systems fluctuated slightly after 250 ns, indicating all systems are up to the convergence of trajectories. To further explore the influence of inhibitor on the structure of PHF6 oligomer in each system, the total contact number between peptides was calculated (**Figure 2B**). For three oligomer systems with PB2, the contact number of oligomer decreased obviously compared with that of the system without PB2, suggesting that the oligomer becomes less stable in the present of PB2. The hydrogen bond (H-bond) interactions between peptides generally play an important part in the aggregation and the formation of oligomer (Zheng et al., 2006; Matthes et al., 2012; Zhou et al., 2016). The ordered PHF6 oligomer is stabilized by a complex network of inter-strand H-bond interactions. On this account, the H-bond number between peptides (**Figure 2C**) was calculated and the result shows that H-bond number of run2 and run3 of oligomer system with PB2 are obviously less than apo oligomer. The interrupted inter-strand hydrogen bonds in both run2 and run3 implies that the interaction of PB2 with oligomer will

interrupt the formed hydrogen bonds between peptides and result in the decrease of stability of PHF6 oligomer. While little difference of H-bond number in run1 can be explained by weak binding of PB2. By analyzing the contact number and H-bond number, it is evident that the stability of PHF6 oligomer is indeed reduced by PB2.

The Conformational Changes of PHF6 Oligomer

It is now well-accepted that the β -sheet-rich structure is the typical structural feature of the amyloid oligomer. Thus, in order to explore the conformational change of PHF6 oligomer, the β -sheet content during the simulation was calculated to study the influence of PB2 on the PHF6 oligomer. As can be seen from **Figure 3**, the β -sheet content of oligomer with PB2 is notably lower than apo oligomer. As results, the coil content of oligomer with PB2 increases. It is proved that β -sheet structures convert into coil structures (**Figure 3** and **Figure 4**). From **Figure 4**, the secondary structure analysis suggests that PB2 molecules change the secondary structures significantly, including strand X in run1, strand A, B and C in run3, and especially for strand L, M, N, O and S in run2 (**Figure 3** and **Figure 4**). This implies that PB2 molecules may bind most strongly to the oligomer in run2 trajectory, causing great parts (strand L, M, N, O and S) of the β -sheet structure of PHF6 oligomer to convert into disordered random coil and turn structures. Then, PCA analysis was applied to investigate the influence of PB2 on the general conformational space of PHF6 oligomer. We can see from **Figure 5** that PHF6 oligomers in complex with PB2 exhibit the larger conformational space and more disperse basins appear on the free energy landscape than PHF6 oligomer without PB2. These results

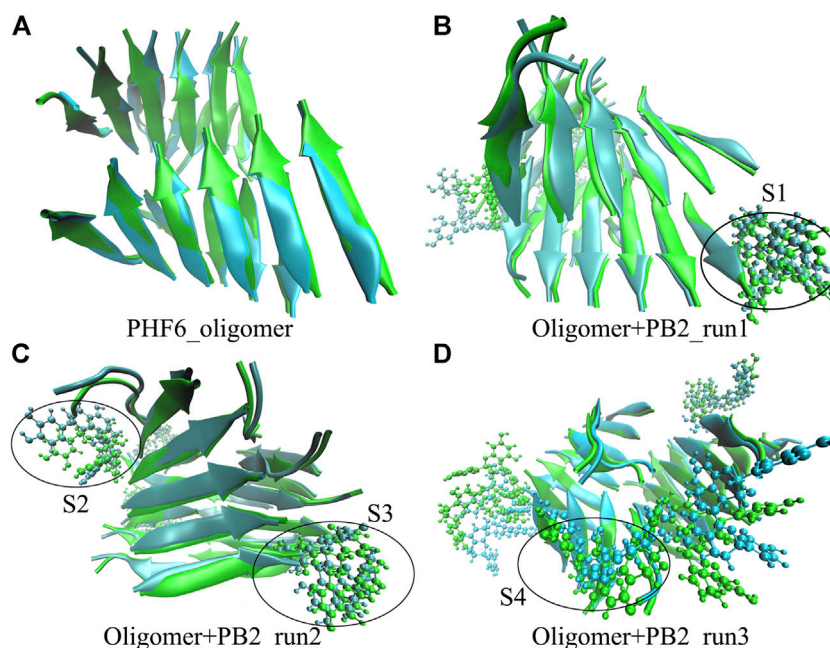


FIGURE 6 | The superposed representative structures of the first two classes of each run. Green and ice blue represent the first and second cluster, respectively. The possible binding site S1, S2, S3 and S4 are circled in black from each parallel run.

TABLE 1 | Binding free energy (kcal/mol) between PHF6 oligomer and PB2 at four possible binding sites.

Contribution	S1	S2	S3	S4
ΔE_{vdw}	-26.35 ± 3.75	-41.70 ± 6.58	-32.80 ± 3.99	-38.09 ± 3.92
ΔE_{ele}	-40.30 ± 16.88	-44.07 ± 12.72	-71.18 ± 9.15	-26.02 ± 10.38
ΔG_{GB}	55.36 ± 15.34	65.24 ± 12.61	83.63 ± 7.66	49.91 ± 9.22
ΔG_{np}	-3.41 ± 0.38	-4.78 ± 0.65	-4.78 ± 0.35	-5.46 ± 0.44
ΔE_{MM}	-66.65 ± 19.15	-85.77 ± 16.84	-103.98 ± 9.48	-64.11 ± 10.30
ΔG_{sol}	51.96 ± 15.15	60.45 ± 12.12	78.85 ± 7.55	44.44 ± 9.14
ΔG_{bind}	-14.69 ± 4.90	-25.32 ± 6.29	-25.13 ± 3.90	-19.67 ± 3.90

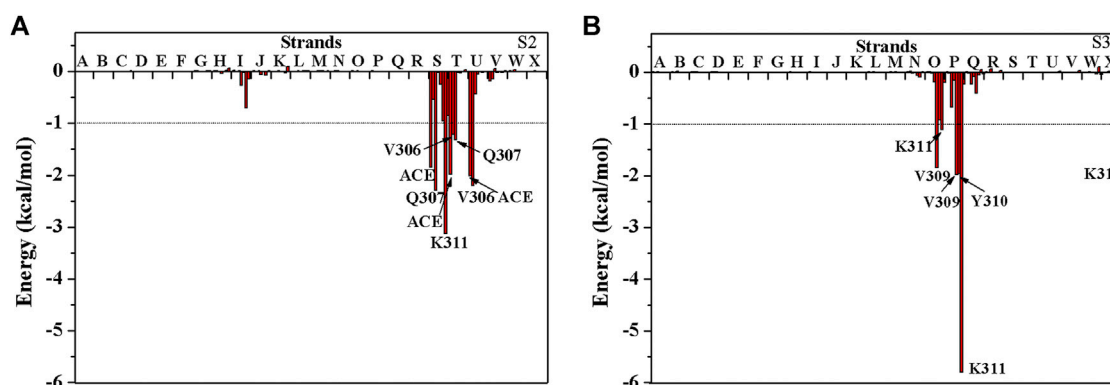


FIGURE 7 | The decomposition of binding free energy of PHF6 oligomer with PB2 at (A) S2 site and (B) S3 site. The residues with energy contribution larger than 1 kcal/mol are labeled.

TABLE 2 | The H-bond occupancy between PHF6 oligomer and PB2 (only gave the hydrogen bonds with occupancy larger than 15%).

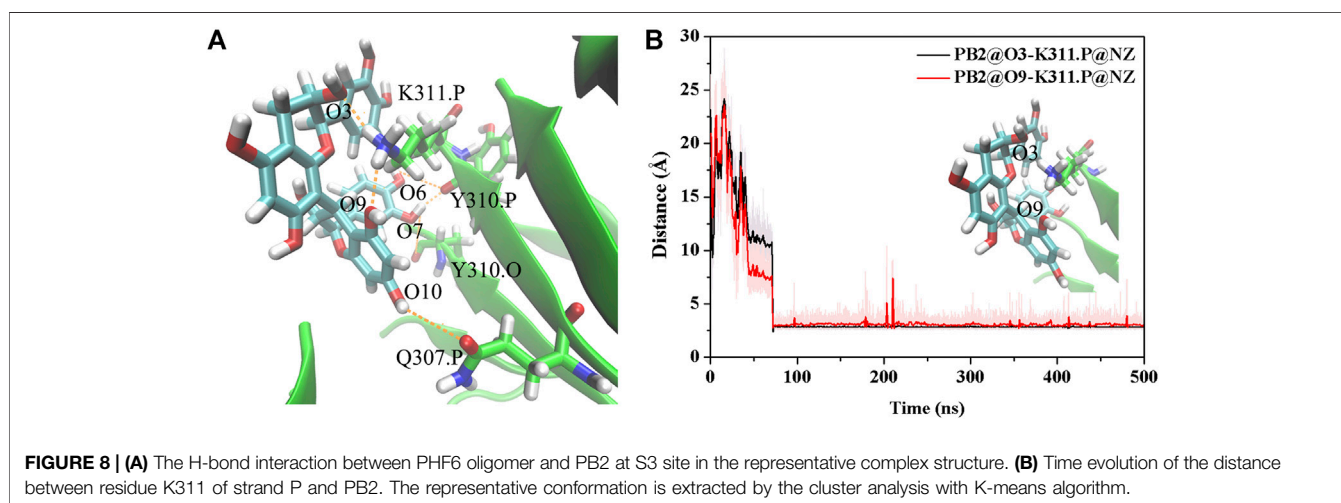
Acceptor	Donor	Occupancy (%)
Y310.P@O	PB2@HO6	74.23
PB2@O3	K311.P@HZ2NZ	32.44
Q307.P@OE1	PB2@HO10	31.78
PB2@O3	K311.P@HZ1NZ	31.04
PB2@O3	K311.P@HZ3NZ	31.04
PB2@O9	K311.P@HZ3NZ	29.71
PB2@O9	K311.P@HZ2NZ	28.28
PB2@O9	K311.P@HZ1NZ	28.24
Y310.P@O	PB2@HO7	22.62
Y310.O@O	PB2@HO7	17.77

reveal that the initial ordered structure of PHF6 oligomer is partly disrupted by PB2, which are consistent with the contact and H-bond analysis results.

The Identification of the Most Possible Binding Site of Proanthocyanidin B2 on PHF6 Oligomer

The above results show that PHF6 oligomer can be influenced by PB2 inhibitor, but it is still unknown how PB2 influences the structure of PHF6 oligomer. Further analysis was performed to uncover the detailed inhibition mechanism. First, clustering analysis was executed to obtain the representative conformations of trajectories with MMTSB toolset based on K-means algorithm. During this process, 4, 5 classes of conformations were generated for apo oligomer and three trajectories of oligomer with PB2, respectively. **Figure 6** shows the first two representative classes for these four trajectories. Compared with apo oligomer, the oligomers with PB2 exhibit more twisted and disordered structure. It is suggested that PB2 can disrupt the ordered β -sheet structure of PHF6 oligomer. Based on the representative conformations from clustering analysis, four possible binding sites S1, S2, S3 and S4 are identified as shown in **Figure 6**. Further, to examine the binding ability of PB2 molecules at different sites, the binding

free energy of each binding site was then calculated by using the MM-GBSA method. The binding free energy and detailed statistic results were listed in **Table 1**. From **Table 1**, the electrostatic interactions are the driving force and play an important role in binding of PB2 to oligomer. What's more, the van der Waals interactions also contribute a lot to the total binding free energy. By comparison, the ranking of binding free energy is -25.31 (S2) < -25.05 (S3) < -19.62 (S4) < -14.68 (S1) kcal/mol. Due to the lower binding free energy, S2 and S3 sites are considered as more possible binding sites of PB2 on oligomer. Then the key residues of S2 and S3 sites in interaction with PB2 were further analyzed. **Figure 7** shows that K311 residues contribute most at both S2 and S3 site. Unexpectedly, a few ACE residues also have an obvious contribution at S2 site (**Figure 7A**). But ACE terminal caps are not natural residues, and they were added to cap the N-terminals to avoid the abnormal electrostatic action between C-terminal and N-terminal when the oligomer structure was prepared. Considering that these residues make a great contribution to the binding, the practical interaction between PB2 and PHF6 oligomer at S2 site may not be as strong as the predicted binding free energy. Therefore, S3 site seems to be more reasonable than S2 site. **Figure 7B** shows that V309 residues of strand O and P, K311 residues of strand O, P and X as well as Y310 of strand P make favorable contributions to the binding of PB2 to oligomer at S3 site. According to the binding mode analysis, V309 residue exerts enormous functions on hydrophobic interactions with PB2. It is worth mentioning that the favorable contribution of K311 of strand P is mainly derived from the strong H-bond interactions with PB2. There are six H-bonds between O3/O9 of hydroxy group of PB2 and NZ of K311 in strand P of the oligomer (**Table 2**). Here, Y310.P represents residue Y310 of strand P for simplification. To show the dynamics changes of hydrogen bonds during MD simulation, we monitored the distances between O3/O9 of PB2 and NZ of K311 in strand P of the oligomer (**Figure 8B**). The distances are rapidly narrowed and keep stable around 3.0 Å from 70 ns, validating the strong H-bond interactions are formed in PB2 binding to PHF6 oligomer during MD simulation. It coincides well with some previous reports that the inhibitor can bind to lysine side chain located in the steric



zipper of PHF6 oligomer (Landau et al., 2011; Mohamed et al., 2013). As shown in **Table 2** and **Figure 8**, there are numerous critical H-bonds formed between hydroxy groups of PB2 and Q307, Y310 and K311 of strand P as well as Y310 of strand O, which may disturb initial inter-strand hydrogen bond network and the stability of the oligomer. This result also explains why the oligomer without PB2 keeps stable while the oligomer with PB2 becomes less stable. These detailed interaction results reveal that PB2 can stably bind to adjacent strands (strand O and P) of PHF6 oligomer with hydrophobic and hydrogen bond interactions at the S3 site. The binding process of PB2 to PHF oligomer along 500 ns simulation time is captured and shown in **Supplementary Movie S1**.

CONCLUSION

In this work, we simulated PHF6 oligomer in the absence and presence of PB2 to explore the molecular mechanism of disruption of PB2 on PHF6 oligomer. Through comparing and analyzing the change of contact and H-bond number, secondary structure and conformation space, we find that PB2 can indeed destabilize PHF6 oligomer. The results are in accordance with experimental observations (Snow et al., 2019). Then to identify the binding site of PB2 on the oligomer, cluster analysis was applied and four possible binding sites were recognized. Among them, S3 site is considered as the most possible one. Our results show that PB2 can stably bind to PHF6 oligomer with hydrophobic and H-bond interactions. Residues V309, Y310 and K311 are essential to the binding of PB2 to PHF6 oligomer, especially residues K311. There are many H-bonds formed between O3/O9 of hydroxy groups of PB2 and NZ of

K311 of the oligomer. These interactions can disrupt the inter-strand H-bonds and convert the ordered β -sheet structure into the disordered one, ultimately disaggregating the PHF6 oligomer. In general, PB2 is a promising Tau aggregation inhibitor and clarifying the molecular inhibition mechanism will help to develop more effective drugs to prevent Tau aggregation for AD.

DATA AVAILABILITY STATEMENT

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

AUTHOR CONTRIBUTIONS

HXL and QL designed the research. QL, CX, HLL, and HG carried out the research and analyzed the data. QL and HXL wrote the manuscript. All authors discussed the results and revised the manuscript.

FUNDING

This work was supported by the National Natural Science Foundation of China (Grant No.21973035).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fchem.2021.666043/full#supplementary-material>

REFERENCES

- Aisen, P. S., Saumier, D., Briand, R., Laurin, J., Gervais, F., Tremblay, P., et al. (2006). A Phase II Study Targeting Amyloid-Beta with 3APS in Mild-To-Moderate Alzheimer Disease. *Neurology* 67, 1757–1763. doi:10.1212/01.wnl.0000244346.08950.64
- Alonso, A., Zaidi, T., Novak, M., Grundke-Iqbal, I., and Iqbal, K. (2001). Hyperphosphorylation Induces Self-Assembly of Tau into Tangles of Paired Helical Filaments/straight Filaments. *Proc. Natl. Acad. Sci. U S A* 98, 6923–6928. doi:10.1073/pnas.121119298
- Alzheimer's Association (2020). Alzheimer's Disease Facts and Figures. *Alzheimers Dement* 16 (3), 391–460. doi:10.1002/alz.12068
- Amadei, A., Linssen, A. B., and Berendsen, H. J. (1993). Essential Dynamics of Proteins. *Proteins* 17, 412–425. doi:10.1002/prot.340170408
- Arriagada, P. V., Growdon, J. H., Hedley-Whyte, E. T., and Hyman, B. T. (1992). Neurofibrillary Tangles but Not Senile Plaques Parallel Duration and Severity of Alzheimer's Disease. *Neurology* 42, 631–639. doi:10.1212/wnl.42.3.631
- Bayly, C. I., Cieplak, P., Cornell, W. D., and Kollman, P. A. (1993). A Well-Behaved Electrostatic Potential Based Method Using Charge Restraints for Deriving Atomic Charges: the RESP Model. *J. Phys. Chem.* 97, 10269–10280. doi:10.1021/j100142a004
- Bloom, G. S. (2014). Amyloid- β and Tau: the Trigger and Bullet in Alzheimer Disease Pathogenesis. *JAMA Neurol.* 71, 505–508. doi:10.1001/jamaneurol.2013.5847
- Case, D. A., Ben-Shalom, I. Y., Brozell, S. R., Cerutti, D. S., Cheatham, T. E., Cruzeiro, V. W., et al. (2018). *Amber 18*. San Francisco, CA: University of California.
- Chiang, K., and Koo, E. H. (2014). Emerging Therapeutics for Alzheimer's Disease. *Annu. Rev. Pharmacol. Toxicol.* 54, 381–405. doi:10.1146/annurev-pharmtox-011613-135932
- Crowe, A., Ballatore, C., Hyde, E., Trojanowski, J. Q., and Lee, V. M. (2007). High Throughput Screening for Small Molecule Inhibitors of Heparin-Induced Tau Fibril Formation. *Biochem. Biophys. Res. Commun.* 358, 1–6. doi:10.1016/j.bbrc.2007.03.056
- Crowther, R. A. (1991). Straight and Paired Helical Filaments in Alzheimer Disease Have a Common Structural Unit. *Proc. Natl. Acad. Sci. U S A* 88, 2288–2292. doi:10.1073/pnas.88.6.2288
- 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
- Feig, M., Karanicolas, J., and Brooks, C. L. (2004). MMTSB Tool Set: Enhanced Sampling and Multiscale Modeling Methods for Applications in Structural Biology. *J. Mol. Graph. Model.* 22, 377–395. doi:10.1016/j.jmgl.2003.12.005
- Frisch, M. J., Trucks, G. W., Schlegel, H. B., Scuseria, G. E., Robb, M. A., Cheeseman, J. R., et al. (2009). *Gaussian 09*. Wallingford, CT: Inc.
- Gauthier, S., Feldman, H. H., Schneider, L. S., Wilcock, G. K., Frisoni, G. B., Hardlund, J. H., et al. (2016). Efficacy and Safety of Tau-Aggregation Inhibitor Therapy in Patients with Mild or Moderate Alzheimer's Disease: a Randomised, Controlled, Double-Blind, Parallel-Arm, Phase 3 Trial. *Lancet* 388, 2873–2884. doi:10.1016/s0140-6736(16)31275-2
- Goedert, M., and Spillantini, M. G. (2006). A Century of Alzheimer's Disease. *Science* 314, 777–781. doi:10.1126/science.1132814
- Götz, J., Halliday, G., and Nisbet, R. M. (2019). Molecular Pathogenesis of the Tauopathies. *Annu. Rev. Pathol.* 14, 239–261. doi:10.1146/annurev-pathmechdis-012418-012936
- Goux, W. J., Kopplin, L., Nguyen, A. D., Leak, K., Rutkowsky, M., Shanmuganandam, V. D., et al. (2004). The Formation of Straight and Twisted Filaments from Short Tau Peptides. *J. Biol. Chem.* 279, 26868–26875. doi:10.1074/jbc.M402379200

- Grundke-Iqbal, I., Iqbal, K., Tung, Y. C., Quinlan, M., Wisniewski, H. M., and Binder, L. I. (1986). Abnormal Phosphorylation of the Microtubule-Associated Protein Tau (Tau) in Alzheimer Cytoskeletal Pathology. *Proc. Natl. Acad. Sci. U S A.* 83, 4913–4917. doi:10.1073/pnas.83.13.4913
- Hornak, V., Abel, R., Okur, A., Strockbine, B., Roitberg, A., and Simmerling, C. (2006). Comparison of Multiple Amber Force Fields and Development of Improved Protein Backbone Parameters. *Proteins* 65, 712–725. doi:10.1002/prot.21123
- Hou, T., Wang, J., Li, Y., and Wang, W. (2011). Assessing the Performance of the MM/PBSA and MM/GBSA Methods. 1. The Accuracy of Binding Free Energy Calculations Based on Molecular Dynamics Simulations. *J. Chem. Inf. Model.* 51, 69–82. doi:10.1021/ci100275a
- Humphrey, W., Dalke, A., and Schulten, K. (1996). VMD: Visual Molecular Dynamics. *J. Mol. Graph.* 14 (33–38), 27–38. doi:10.1016/0263-7855(96)00018-5
- 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
- Kabsch, W., and Sander, C. (1983). Dictionary of Protein Secondary Structure: Pattern Recognition of Hydrogen-Bonded and Geometrical Features. *Biopolymers* 22, 2577–2637. doi:10.1002/bip.360221211
- Kadavath, H., Hofe, R. V., Biernat, J., Kumar, S., Tepper, K., Urlaub, H., et al. (2015). Tau Stabilizes Microtubules by Binding at the Interface between Tubulin Heterodimers. *Proc. Natl. Acad. Sci. U S A.* 112, 7501–7506. doi:10.1073/pnas.1504081112
- Kametani, F., and Hasegawa, M. (2018). Reconsideration of Amyloid Hypothesis and Tau Hypothesis in Alzheimer's Disease. *Front. Neurosci.* 12, 25. doi:10.3389/fnins.2018.00025
- Kontsekova, E., Zilka, N., Kovacech, B., Skrabana, R., and Novak, M. (2014). Identification of Structural Determinants on Tau Protein Essential for its Pathological Function: Novel Therapeutic Target for Tau Immunotherapy in Alzheimer's Disease. *Alzheimers Res. Ther.* 6, 45. doi:10.1186/alzrt277
- Landau, M., Sawaya, M. R., Faull, K. F., Laganowsky, A., Jiang, L., Sievers, S. A., et al. (2011). Towards a Pharmacophore for Amyloid. *Plos Biol.* 9, e1001080. doi:10.1371/journal.pbio.1001080
- Li, W., Sperry, J. B., Crowe, A., Trojanowski, J. Q., Smith, A. B., and Lee, V. M. (2009). Inhibition of Tau Fibrillization by Oleocanthol via Reaction with the Amino Groups of Tau. *J. Neurochem.* 110, 1339–1351. doi:10.1111/j.1471-4159.2009.06224.x
- Liu, X., Zhou, S., Shi, D., Bai, Q., Liu, H., and Yao, X. (2018). Influence of EGCG on α -synuclein (α S) Aggregation and Identification of Their Possible Binding Mode: A Computational Study Using Molecular Dynamics Simulation. *Chem. Biol. Drug Des.* 91, 162–171. doi:10.1111/cbdd.13067
- Matthes, D., Gapsys, V., and de Groot, B. L. (2012). Driving Forces and Structural Determinants of Steric Zipper Peptide Oligomer Formation Elucidated by Atomistic Simulations. *J. Mol. Biol.* 421, 390–416. doi:10.1016/j.jmb.2012.02.004
- Mazanetz, M. P., and Fischer, P. M. (2007). Untangling Tau Hyperphosphorylation in Drug Design for Neurodegenerative Diseases. *Nat. Rev. Drug Discov.* 6, 464–479. doi:10.1038/nrd2111
- Mohamed, T., Hoang, T., Jelokhani-Niaraki, M., and Rao, P. P. (2013). Tau-derived-hexapeptide 306VQIVYK311 Aggregation Inhibitors: Nitrocatechol Moiety as a Pharmacophore in Drug Design. *ACS Chem. Neurosci.* 4, 1559–1570. doi:10.1021/cn400151a
- Pickhardt, M., Neumann, T., Schwizer, D., Callaway, K., Vendruscolo, M., Schenk, D., et al. (2015). Identification of Small Molecule Inhibitors of Tau Aggregation by Targeting Monomeric Tau as a Potential Therapeutic Approach for Tauopathies. *Curr. Alzheimer Res.* 12, 814–828. doi:10.2174/156720501209151019104951
- Pickhardt, M., von Bergen, M., Gazova, Z., Hascher, A., Biernat, J., Mandelkow, E. M., et al. (2005). Screening for Inhibitors of Tau Polymerization. *Curr. Alzheimer Res.* 2, 219–226. doi:10.2174/1567205053585891
- Plumley, J. A., and Dannenberg, J. J. (2010). The Importance of Hydrogen Bonding between the Glutamine Side Chains to the Formation of Amyloid VQIVYK Parallel Beta-Sheets: an ONIOM DFT/AM1 Study. *J. Am. Chem. Soc.* 132, 1758–1759. doi:10.1021/ja909690a
- Rane, J. S., Bhaumik, P., and Panda, D. (2017). Curcumin Inhibits Tau Aggregation and Disintegrates Preformed Tau Filaments In Vitro. *J. Alzheimers Dis.* 60, 999–1014. doi:10.3233/jad-170351
- Rojas Quijano, F. A., Morrow, D., Wise, B. M., Brancia, F. L., and Goux, W. J. (2006). Prediction of Nucleating Sequences from Amyloidogenic Propensities of Tau-Related Peptides. *Biochemistry* 45, 4638–4652. doi:10.1021/bi052226q
- 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
- Sawaya, M. R., Sambashivan, S., Nelson, R., Ivanova, M. I., Sievers, S. A., Apostol, M. I., et al. (2007). Atomic Structures of Amyloid Cross-Beta Spines Reveal Varied Steric Zippers. *Nature* 447, 453–457. doi:10.1038/nature05695
- Šimić, G., Babić Leko, M., Wray, S., Harrington, C., Delalle, I., Jovanov-Milošević, N., et al. (2016). Tau Protein Hyperphosphorylation and Aggregation in Alzheimer's Disease and Other Tauopathies, and Possible Neuroprotective Strategies. *Biomolecules* 6, 6. doi:10.3390/biom6010006
- Snow, A. D., Castillo, G. M., Nguyen, B. P., Choi, P. Y., Cummings, J. A., Cam, J., et al. (2019). The Amazon Rain Forest Plant *Uncaria Tomentosa* (Cat's Claw) and its Specific Proanthocyanidin Constituents Are Potent Inhibitors and Reducers of Both Brain Plaques and Tangles. *Sci. Rep.* 9, 561. doi:10.1038/s41598-019-38645-0
- Spillantini, M. G., and Goedert, M. (2013). Tau Pathology and Neurodegeneration. *Lancet Neurol.* 12, 609–622. doi:10.1016/s1474-4422(13)70090-5
- Taniguchi, S., Suzuki, N., Masuda, M., Hisanaga, S., Iwatsubo, T., Goedert, M., et al. (2005). Inhibition of Heparin-Induced Tau Filament Formation by Phenothiazines, Polyphenols, and Porphyrins. *J. Biol. Chem.* 280, 7614–7623. doi:10.1074/jbc.M408714200
- von Bergen, M., Friedhoff, P., Biernat, J., Heberle, J., Mandelkow, E. M., and Mandelkow, E. (2000). Assembly of Tau Protein into Alzheimer Paired Helical Filaments Depends on a Local Sequence Motif ((306)VQIVYK(311)) Forming Beta Structure. *Proc. Natl. Acad. Sci. U S A.* 97, 5129–5134. doi:10.1073/pnas.97.10.5129
- 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
- Wang, Q., Zhou, S., Wei, W., Yao, X., Liu, H., and Hu, Z. (2015). Computational Insights into the Inhibition and Destabilization of Morin on the Oligomer of Full-Length Human Islet Amyloid Polypeptide. *Phys. Chem. Chem. Phys.* 17, 29103–29112. doi:10.1039/c5cp03991f
- Wilcock, G. K., Black, S. E., Hendrix, S. B., Zavitz, K. H., Swabb, E. A., and Laughlin, M. A. (2008). Efficacy and Safety of Tarenflurbil in Mild to Moderate Alzheimer's Disease: a Randomised Phase II Trial. *Lancet Neurol.* 7, 483–493. doi:10.1016/s1474-4422(08)70090-5
- Wischik, C. M., Edwards, P. C., Lai, R. Y., Roth, M., and Harrington, C. R. (1996). Selective Inhibition of Alzheimer Disease-like Tau Aggregation by Phenothiazines. *Proc. Natl. Acad. Sci. U S A.* 93 (20), 11213–11218. doi:10.1073/pnas.93.20.11213
- Wobst, H. J., Sharma, A., Diamond, M. I., Wanker, E. E., and Bieschke, J. (2015). The Green Tea Polyphenol (-)-epigallocatechin Gallate Prevents the Aggregation of Tau Protein into Toxic Oligomers at Substoichiometric Ratios. *FEBS Lett.* 589, 77–83. doi:10.1016/j.febslet.2014.11.026
- Zhang, B., Maiti, A., Shively, S., Lakhani, F., McDonald-Jones, G., Bruce, J., et al. (2005). Microtubule-binding Drugs Offset Tau Sequestration by Stabilizing Microtubules and Reversing Fast Axonal Transport Deficits in a Tauopathy Model. *Proc. Natl. Acad. Sci. U S A.* 102, 227–231. doi:10.1073/pnas.0406361102
- Zhao, S., Zhang, L., Yang, C., Li, Z., and Rong, S. (2019). Procyanidins and Alzheimer's Disease. *Mol. Neurobiol.* 56 (8), 5556–5567. doi:10.1007/s12035-019-1469-6
- Zheng, J., Ma, B., Tsai, C. J., and Nussinov, R. (2006). Structural Stability and Dynamics of an Amyloid-Forming Peptide GNNQQNY from the Yeast Prion Sup-35. *Biophys. J.* 91, 824–833. doi:10.1529/biophysj.106.083246
- Zhou, S., Shi, D., Liu, X., Liu, H., and Yao, X. (2016). Protective V127 Prion Variant Prevents Prion Disease by Interrupting the Formation of Dimer and Fibril from Molecular Dynamics Simulations. *Sci. Rep.* 6, 21804. doi:10.1038/srep21804

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 Li, Xiong, Liu, Ge, Yao and Liu. 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.



A Machine Learning Model to Predict Drug Transfer Across the Human Placenta Barrier

Juan I. Di Filippo^{1,2,3}, Mariela Bollini⁴ and Claudio N. Cavasotto^{1,2,3*}

¹Computational Drug Design and Biomedical Informatics Laboratory, Instituto de Investigaciones en Medicina Traslacional (IIMT), CONICET-Universidad Austral, Pilar, Argentina, ²Facultad de Ciencias Biomédicas and Facultad de Ingeniería, Universidad Austral, Pilar, Argentina, ³Austral Institute for Applied Artificial Intelligence, Universidad Austral, Pilar, Argentina, ⁴Centro de Investigaciones en BioNanociencias (CIBION), Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Buenos Aires, Argentina

OPEN ACCESS

Edited by:

Sylvain Broussy,
University of Paris, France

Reviewed by:

Yuguang Mu,
Nanyang Technological University,
Singapore
Tanmoy Chakraborty,
Sharda University, India

*Correspondence:

Claudio N. Cavasotto
cnc@cavasotto-lab.net
ccavasotto@austral.edu.ar

Specialty section:

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

Received: 26 May 2021

Accepted: 07 July 2021

Published: 20 July 2021

Citation:

Di Filippo JI, Bollini M and
Cavasotto CN (2021) A Machine
Learning Model to Predict Drug
Transfer Across the Human
Placenta Barrier.
Front. Chem. 9:714678.
doi: 10.3389/fchem.2021.714678

The development of computational models for assessing the transfer of chemicals across the placental membrane would be of the utmost importance in drug discovery campaigns, in order to develop safe therapeutic options. We have developed a low-dimensional machine learning model capable of classifying compounds according to whether they can cross or not the placental barrier. To this aim, we compiled a database of 248 compounds with experimental information about their placental transfer, characterizing each compound with a set of ~5.4 thousand descriptors, including physicochemical properties and structural features. We evaluated different machine learning classifiers and implemented a genetic algorithm, in a five cross validation scheme, to perform feature selection. The optimization was guided towards models displaying a low number of false positives (molecules that actually cross the placental barrier, but are predicted as not crossing it). A Linear Discriminant Analysis model trained with only four structural features resulted to be robust for this task, exhibiting only one false positive case across all testing folds. This model is expected to be useful in predicting placental drug transfer during pregnancy, and thus could be used as a filter for chemical libraries in virtual screening campaigns.

Keywords: placenta barrier permeability, machine learning, toxicology, clearance index, fetus:mother ratio

INTRODUCTION

Drug prescribing in pregnancy remains a complex and controversial issue for both pregnant women and clinicians (Leong et al., 2019). According to the Center for Disease Control and Prevention (CDC), 9 out of 10 women take at least one medication during pregnancy; and 70% of pregnant women take at least one prescribed medication (<https://www.cdc.gov/pregnancy/meds/treatingfortwo/index.html>). Over the past 30 years, the use of prescription drugs during the first quarter trimester of pregnancy has increased by more than 60%. This suggests that at the beginning of pregnancy, many women either present pre-chronic conditions (e.g., pre-gestational diabetes) or develop pregnancy-specific diseases (e.g., hyperemesis gravidarum, intrahepatic cholestasis of pregnancy, HELLP syndrome) which will require the administration of medications, including those which might cause fetal toxicity or teratogenesis (Eke et al., 2020). To guarantee drug safety during pregnancy, *in vitro* and *in vivo* experimental models were developed to study the transfer and metabolism of drugs across the human placental barrier. Since the placenta is the most

species-specific organ, human cell lines and tissue models are considered more appropriate than *in vivo* assays performed in rodent models for evaluating the transfer of chemicals across the human placental barrier (Giaginis et al., 2012). In this regard, the *ex vivo* human placental perfusion model, which preserves placental structural integrity, and mimics the maternal and fetal blood circulation, is more suitable (Gordon et al., 2016). Unfortunately, *in vitro* and *ex vivo* methods cannot directly predict *in vivo* outcomes, making the assessment of placental transfer difficult (Hutson et al., 2011). On the other hand, *in vivo* assays are more accurate in evaluating drug toxicity. *In vivo* data can be obtained by measuring drug concentrations in the umbilical cord blood and maternal blood at delivery (Freriksen et al., 2020). The fetal-maternal concentration ratio is a widely used indicator of placental permeability that has been applied to drug monitoring (Hutson et al., 2011). However, there is an obvious ethical barrier to develop *in vivo* studies to assess the risk of transfer of chemicals across the placental membrane from the mother to the fetus. In this scenario, there is an urgent need for an integrated approach incorporating all the range of methodologies (*in vitro*, *ex vivo*, *in silico* and *in vivo* studies) to accelerate the availability of pharmacology data in pregnant women to allow the safe and effective use of medication during this physiological state.

Several Quantitative Structure–Activity Relationship (QSAR) models have been published on this topic. Based on *ex vivo* human placental perfusion results, Giaginis et al. (2009) developed a model to predict placental transfer through the calculation of the Clearance Index (CI) values for a set of 88 compounds. Using this approach, Zhang et al. (2015) estimated the placental barrier permeability, also expressed as CI values, for a set of 65 compounds. Takaku et al. (2015) developed a QSAR model for predicting the *in vivo* fetal–maternal blood concentration ratio (F/M ratio) for a set of 55 compounds. Later, Wang et al., using the same chemical library of 55 compounds as Takaku et al. (2015), developed a QSAR model following the Organization for Economic Co-operation and Development (OECD) guidelines based on multiple linear regression adjustments for predicting *in vivo* log (F/M) values (Wang et al., 2020). These studies achieved a reasonable predictive potential (the correlation between measured and predicted values is acceptable); however, all of them were validated with few samples. Giaginis et al. used only nine compounds as a test set, Takaku et al. and Wang et al. used a test set of 14 compounds, and Zhang et al. utilized 19 compounds for the test set. Takaku et al. used three features for their QSAR model, and Wang et al. utilized two descriptors, which is a reasonable approach taking into account the number of samples in their set; however, Zhang et al. utilized 48 descriptors to construct their QSAR model.

In this study, we used available information on drug placental transfer to train machine learning (ML) algorithms in order to carry out the *in silico* prediction of whether a compound will cross the placental barrier or not. ML approaches have been consistently implemented in the last decade with different degrees of success in the drug discovery pipeline (Carpenter et al., 2018; Chen et al., 2018; Chan et al., 2019; Mak and

Pichika, 2019; Cavasotto and Di Filippo, 2021a); while a ML model would not necessarily provide a clearer understanding of why some drugs cross or do not cross the placental barrier, its importance lies on the direct use for practical purposes, namely, serving as a filter in a high throughput screening campaign of a chemical library. To this purpose, we compiled a database of 248 compounds, collecting for each compound its CI value, and/or F/M ratio, and/or assessment from the literature whether it crosses or not the placental barrier. Considering the variability of the experimental parameters collected between different laboratories (Hutson et al., 2011), we decided to label each compound in a binary fashion according to whether it crosses (C) or does not cross (NC) the placental membrane, using the above mentioned information and based on a proposed set of criteria (see Methods). We used molecular descriptors as inputs and the binary output (C/NC) to train the ML classifiers to predict whether a molecule will cross the placental barrier or not. After an extensive feature selection process and the evaluation of different models, we present in this work a robust LDA classifier trained with only four features that exhibits an excellent performance. Furthermore, the model exhibits a critical characteristic, namely, the amount of molecules that cross the placenta that are misclassified as not crossing is almost null.

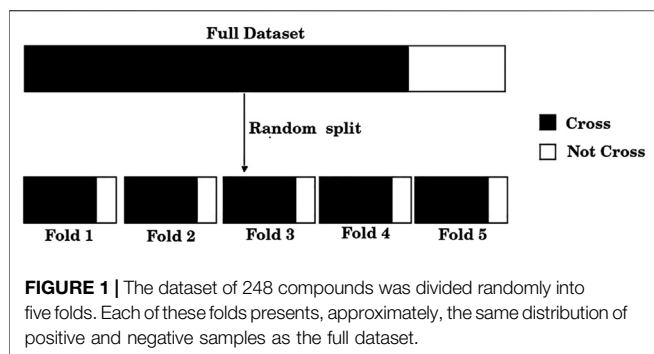
MATERIALS AND METHODS

Data Collection

We collected a dataset of 248 molecules with at least one of these pieces of information: CI, F/M ratio (F/M), evidence from the literature that the molecule crosses or not the placenta barrier (**Supplementary Table S1**). If $F/M \leq 0.15$, the molecule was labeled as NC; if $F/M \geq 0.3$, the molecule was labeled as C; to avoid dubious cases, molecules in the range $0.15 < F/M < 0.3$ were not included in the set. In cases where only the CI value was available, the molecule was labeled as C if $CI > 0.80$ (this threshold was chosen based on the fact that whenever both F/M and CI values were available, all molecules with $CI > 0.8$ have $F/M > 0.3$, i.e., they were labeled as C). If $F/M \geq 0.3$ and $CI < 0.8$ the molecule was labeled as C, since we privileged results from *in vivo* assays over those using the perfusion method. Several molecules lacked of F/M and CI values, but evidence was found in the literature to classify them as C or NC (cf. **Supplementary Table S1**). Using these criteria, the dataset contained 213 molecules (~86%) that cross the placental barrier, and 35 (~14%) that do not. Following the standard convention, we defined the larger class as the negative one.

Dataset Split

The standard training set/test set split is useful only for large size datasets, which is clearly not our case. If, for example, 20% of the dataset were used for testing, results would be reported only over 50 samples; furthermore, the results could be biased due to the unique random split of the training and test sets. Instead, we adopted a standard procedure when dealing with small datasets, a 5-fold cross-validation scheme. For this purpose, the dataset was split randomly into five folds, where each fold approximately



exhibits the C/NC distribution of the entire dataset, as shown in **Figure 1**. Unlike a single training set/test set split, this scheme allows the use of each of the samples both in the training set (four times) and in the test set (once).

Molecular Descriptors

Molecules were protonated at physiological pH using the ICM software (MolSoft, San Diego, CA, 2019) (Abagyan et al., 1994), in a similar fashion as in earlier works (Cavasotto and Aucar, 2020; Cavasotto and Di Filippo, 2021b), and then each molecule was visually inspected. To generate model inputs, molecules were described using a set of 5,379 features, which are summarized in **Table 1**. These were calculated with OpenBabel (O'Boyle et al., 2008; O'Boyle et al., 2011) and PaDEL (Yap, 2011), and included both physicochemical properties and substructure fingerprint counts. These fingerprint count features encompass electrotopological state indices (Hall and Kier, 1995), the presence of SMARTS patterns (Klekota and Roth, 2008), and the presence of chemical substructures.

Evaluation Metrics

A binary classifier predicts all the instances as either positive (1) or negative (0). Considering that these instances can be classified correctly or incorrectly, four types of outcomes can be distinguished: True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN). In general, classification algorithms predict the probability that an observation will belong to the positive class, i.e., will be 1. To make discrete predictions based on the probability provided by the classifier, that is to say, to have a binary outcome, it is necessary to define a threshold: Probabilities below this threshold are discretized as 0 and above the threshold as 1.

The Accuracy (A) is the percentage of accurate predictions, and is defined as

$$A = \frac{TP + TN}{TP + FP + TN + FN} \quad (1)$$

Precision (P), Recall (R), and the False Positive Rate (FPR) are defined as

$$P = \frac{TP}{TP + FP}$$

$$R = \frac{TP}{TP + FN} \quad (2)$$

$$FPR = \frac{FP}{FP + TN}$$

The F_β score, which is the weighted harmonic mean of P and R , is expressed as

$$F_\beta = (1 + \beta^2) \frac{P \times R}{\beta^2 \times P + R} \quad (3)$$

where β is a parameter that controls the balance to give more weight to P ($\beta < 1$) or R ($\beta > 1$).

Due to the imbalance of the dataset classes, it is evident that accuracy (**Eq. 1**) would not be a proper score for the classification task. Indeed, a classifier that predicts the negative class for all cases would have an accuracy of 86%. It has been shown that, for imbalanced sets, computing precision and recall (**Eq. 2**) gives a better insight about the classifier's performance than the Receiver Operating Characteristic curve, a common metric in classification tasks (Saito and Rehmsmeier, 2015). In this context, a low false positive rate is represented by a high precision score, while false negatives are addressed by the recall. In this work, we chose the F_β score (**Eq. 3**) using $\beta = 0.5$ to penalize the classifying of molecules that cross the barrier as not crossing, i.e., e, classification of negative samples as positive samples. Thus, we favor models that have a low number of false positives. A common metric for unbalanced classification problems is the Mathews Correlation Coefficient (MCC); since a recent study discourages its use in unbalanced sets (Zhu, 2020), we decided to use only the F_β score due to its direct implementation in penalizing false positives.

The Precision-Recall Curve (PRC) is constructed by plotting P in terms of R for different probability thresholds. The Average Precision (AP) is a scalar that summarizes the PRC, in the same manner as the area under the curve (AUC) of the receiver operating characteristic (ROC) curve. Strictly, the AP is the area under the PRC.

$$AP = \int_0^1 P(R) dR \quad (4)$$

TABLE 1 | Molecular features calculated with OpenBabel and PaDEL. Physicochemical properties include classical descriptors such as molecular weight, rotatable bonds, number of Hydrogen bond donors and acceptors, etc.

Source	Name	Descriptors	Number
OpenBabel	1D and 2D descriptors	Physicochemical properties	13
PaDEL	EState fingerprints	Electrotopological state indices	79
PaDEL	KlekotaRoth fingerprints	Presence of SMARTS patterns	307
PaDEL	Substructure fingerprints	Presence of chemical substructures	4,860

In this work, we approximated this integral by a sum over the precisions at every possible threshold value (n) multiplied by the change in R , according to

$$AP \approx \sum_n P_n (R_n - R_{n-1}) \quad (5)$$

where R_n and P_n are, the recall and precision values at the n th threshold value, respectively.

RESULTS AND DISCUSSION

The objective of this study was to provide a ML model capable of classifying compounds either as crossing or not crossing the placental barrier. To this aim, using a dataset of 248 compounds (see Methods), we trained and compared several ML models, searching for optimal low-dimensional sets of descriptors. Considering that the odds of classifying a molecule that crosses the placenta as not crossing must be reduced to a minimum, we chose $F_{1/2}$ as the metric to evaluate performance, thus favoring models that have a low number of FPs; while having a high false rate of predictions is always undesirable, it would be highly risky in this specific case. Due to the high features/samples ratio, we decided to keep the number of descriptors in the final models to a minimum.

Design of the Feature Selection Protocol

Considering the size of our dataset (248 samples), and the number of calculated features (~5.4 thousand descriptors), we performed a feature selection process to avoid over-fitting. Initially we considerably reduced the high dimensionality of the feature space by eliminating from the PaDEL set of descriptors variables that did not provide significant information, by eliminating features (specifically, fingerprint counts) that had less than three matches within the molecules of the dataset. This decision was principally based on the trade-off between the number of remaining descriptors (by removing features) and the information loss. After this process, the number of descriptors fell to 760. Needless to say, this procedure is independent of the class labels, and thus can be done before the cross validation split.

To reduce even further the set of 760 features, we used a genetic algorithm (GA) which essentially searches for sets of features with a high $F_{1/2}$ score over a given training set, as described below.

Genetic Algorithm

From a training dataset composed of a set of molecules with their corresponding descriptors, the GA generated a population of 1,000 individuals, where each individual was defined as a set of six randomly selected features; we also explored the use of individuals described with nine and 12 features, but did not find any improvement over the use of six features (see *Additional Genetic Algorithm Runs Using Identical Initial Conditions*). Then, each individual was used to train a ML classifier, and subsequently ranked in terms of the obtained $F_{1/2}$ score over

the training data. After having this initial population of 1,000 individuals ranked, the following iterative process was carried out: 1) the set of features with the best score of the population (the optimal individual) was assessed; 2) random sets from the top half of the population were selected in pairs and combined until 500 new sets were obtained; with two individuals, a new agent was generated by retrieving the first three features from one individual and the other three features from the other individual; 3) the $F_{1/2}$ score was calculated for each of these 500 generated sets and, independently of the results, these new individuals replaced the bottom half from the past population; 4) the new population of 1,000 features was re-ranked. This iterative process was carried out for 199 iterations, which allowed both the convergence of the method (the top ranked individuals were very similar) and the exploration of the feature space (as explained below).

Within this process, three operations were performed: 1) every time a new set of features was generated (by the combination of two other sets), it was assigned a probability of 0.2 of being mutated. If it was mutated, the new agent would change all its variables with those of the optimal individual, replacing two features with two random ones from the major set; in certain sense, this is a way to explore the “vicinity” of the best scored individual; 2) for each generation we replaced one third of the reproducible population (top half of the population) with new random agents; 3) finally, after 50 iterations, a new initial population of 1,000 individuals was generated and ranked, and the current population was replaced entirely except for the top 10 individuals. The last two operations were performed for the sake of augmenting the exploration of the feature space.

Coupling the Genetic Algorithm With the Cross Validation Scheme

Following Hastie et al. Hastie et al, (2009), we first split the data according to the cross validation scheme, and then used the feature selection method described above with the training data. Specifically, we proceeded as follows:

- 1) Divide the total amount of samples into 5 cross-validation folds ($k = 1, 5$) at random as shown in **Figure 1** and generate five partitions, where partition k corresponds to using fold k as a test set, and the remaining four folds as the training set.
- 2) For each partition, the GA is used to find a set of predictors that exhibits a high value of the $F_{1/2}$ score, calculated only on the training samples.
- 3) Assess shared features between the optimal sets found in each of the five partitions (common features), as illustrated in **Figure 2**.
- 4) Evaluate the performance of the set of common features over the corresponding test sets of each partition, as shown in **Figure 3**.

As is standard in the use of cross validation schemes, we report the average $F_{1/2}$ scores over the five training sets, and over the five test sets. For simplicity, the process depicted in **Figure 2** of finding a set of “Common features” and evaluating it as shown in **Figure 3** will be referred from now on as a “run”.

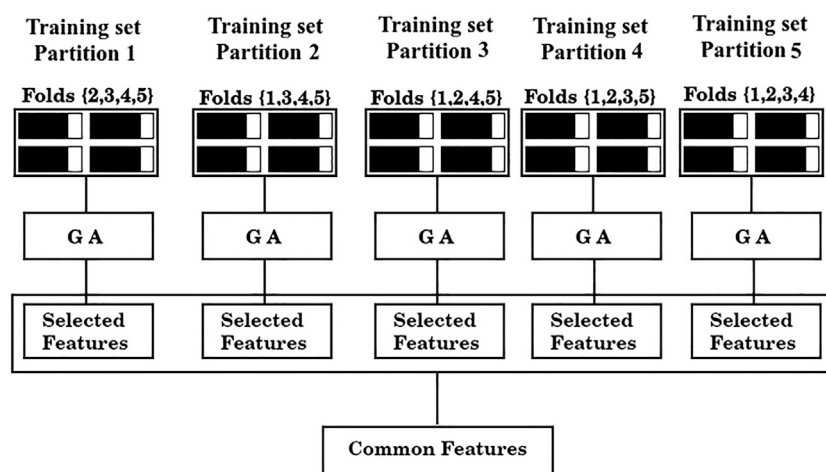


FIGURE 2 | Feature selection scheme. From left to right, training data from partitions 1-5 are fed to a GA. The GA yields a solution for each partition ("Selected features") and finally, the shared features between those solutions are collected ("Common features").

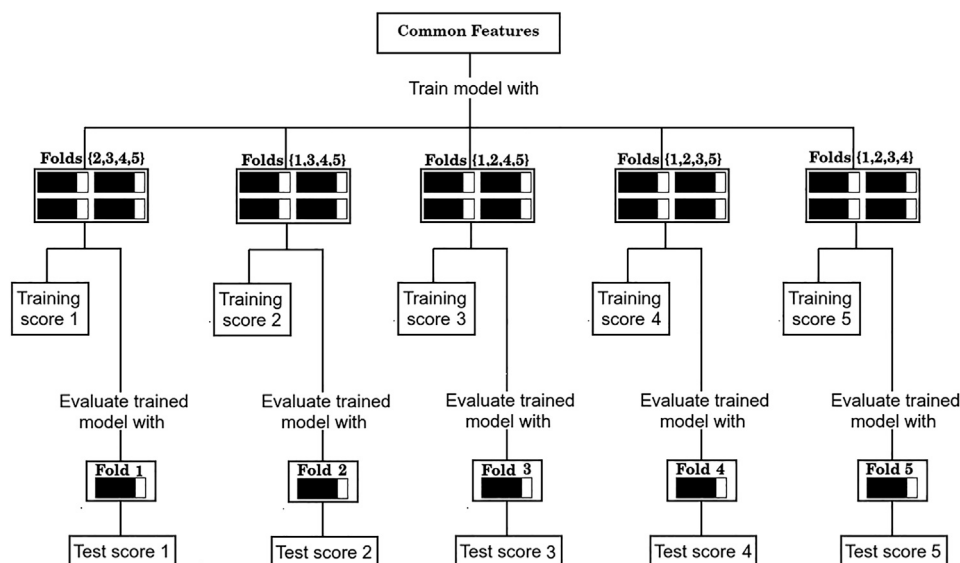


FIGURE 3 | Evaluation procedure of the common sets of features with the 5-fold cross validation scheme.

Selection of the Best Machine Learning Model

Common sets of descriptors were searched for four different ML algorithms: Linear Discriminant Analysis (LDA), Logistic Regression (LR), Random Forest (RF), and Support Vector Machines (SVM). For this task, we ran one GA per model, feeding each algorithm with the same initial population. Before displaying the results corresponding to the four methods, we will illustrate the feature selection protocol with the LDA. In **Table 2** we show the best sets of features found in each partition for the LDA model by running a single GA. As mentioned earlier, only the $F_{1/2}$ score over the training set is

reported at this stage. These sets of features correspond to the "Selected features" shown in **Figure 2**.

Across the five sets of features shown in **Table 2**, there are four repeated descriptors: KRFPC413 (2 times), KRFPC566 (4 times), KRFPC608 (4 times), and KRFPC4830 (2 times). Although the GA was fed with sets of six features, only these four repeated features constitute the set of "Common features" (cf. **Figure 2**). Using these four features we trained another LDA model (**Figure 3**). This model exhibited mean $F_{1/2}$ scores of 0.80 and 0.77 in the training and test sets, respectively (see **Table 3**). The average $F_{1/2}$ score of 0.77 over the test sets corresponded to average values of P and R of 0.93 and 0.51, respectively. This

TABLE 2 | Best set of features (“Selected features”, see **Figure 2**) obtained on each partition of the cross validation split based on the training $F_{1/2}$ score for the LDA model.

Partition	Feature 1	Feature 2	Feature 3	Feature 4	Feature 5	Feature 6	$F_{1/2}$ Train
1	KRFPC608	KRFPC841	KRFPC1773	KRFPC3224	KRFPC3591	KRFPC4830	0.87
2	KRFPC413	KRFPC566	KRFPC608	KRFPC1638	KRFPC3399	SubFPC19	0.90
3	KRFPC442	KRFPC557	KRFPC566	KRFPC3400	KRFPC3741	KRFPC3948	0.89
4	KRFPC413	KRFPC566	KRFPC608	KRFPC3139	KRFPC3737	KRFPC4006	0.87
5	KRFPC326	KRFPC566	KRFPC592	KRFPC608	KRFPC3730	KRFPC4830	0.83

TABLE 3 | Repeated features across different partitions for the first run of the GA (“Common features”, see **Figures 2, 3**) using different ML models. The frequency each feature is repeated within partitions is shown in parentheses. The $F_{1/2}$ Train and $F_{1/2}$ Test columns refer to the average score across the different training folds and test folds, respectively.

Model	Feature 1	Feature 2	Feature 3	Feature 4	Feature 5	Feature 6	$F_{1/2}$ train	$F_{1/2}$ test
RF	KRFPC476 (2)	KRFPC3707 (2)	KRFPC4556 (2)	SubFPC3 (2)	SubFPC301 (2)	MP (2)	1.0	-
LDA	KRFPC413 (2)	KRFPC566 (3)	KRFPC608 (4)	KRFPC4830 (2)	-	-	0.80	0.77
SVM	KRFPC1564 (2)	KRFPC3946 (2)	SubFPC169 (2)	-	-	-	0.72	-
LR	KRFPC608 (4)	ROTB (2)	-	-	-	-	0.59	0.54

represents a very good performance, and *a priori* indicates that it is plausible to select features in this manner.

The same process was carried out for the other 3 ML methods. The sets of common features found for each model, as well as the training and test $F_{1/2}$ scores are summarized in **Table 3**. RF and SVM models were prone to over-fitting, as they achieved a null averaged R over the test sets (non-defined $F_{1/2}$ score), and the LR model displayed a significantly poorer performance compared to LDA. We thus continued the analysis with only the LDA model. While different alternatives could be pursued to improve the performance of the other ML models, the aim of this study is to find a robust and accurate model exhibiting high performance.

Linear Discriminant Analysis Model Analysis

Despite of having promising results with the LDA model (**Table 3**), at this point it is not yet clear whether the found set of features is robust. Considering the random nature of the GA, we analyzed how the different parameters of the feature selection process could impact on the results. First, we performed five additional runs using the same initial conditions of the GA used for the LDA model shown in **Table 3**. Then, we focused on three main parameters of the initial conditions of the GA, namely, the number of features used to describe the individuals of the GA, the cross validation split, and the initial population fed to the GA, and performed additional runs maintaining two of the mentioned initial parameters fixed, while varying the third one. In the following results, “run” refers to the finding of a set of common features” and evaluating it (see **Figures 2, 3**).

Additional Genetic Algorithm Runs Using Identical Initial Conditions

Using the same initial population and cross validation split as in the first ML model selection (**Table 3**), we performed five additional runs of the GA for the LDA model, obtaining

another five sets of common features. Results are summarized in **Supplementary Table S2**. In four of the five runs, KRFPC566 and KRFPC608 belonged to the set of common features and, remarkably, KRFPC3948 was repeated in the five sets. This indicates, that the KRFPC566 and KRFPC608 features, which were found in the first LDA model (**Table 3**), are retrieved despite of the inherent randomness of the GA, and that the first obtained solution missed an apparently important feature, KRFPC3948.

Extending the Size of the Genetic Algorithm Individuals

We performed five runs (see **Figures 2, 3**) with sets of nine features, and five runs with sets of 12 features. Results are shown in **Supplementary Tables S3, S4**.

Every set of common features exhibited a low performance in comparison to the LDA model using six features in the GA (**Table 3**). Over the training data, the highest $F_{1/2}$ score was of 0.60. In the test data, we found one common set for which the model’s performance was of 0.50 (run 3, **Supplementary Table S3**), and in the rest of the runs, the corresponding models achieved null recall values. This shows that using nine or 12 features in the GA shows no advantage on the performance of the LDA model.

Genetic Algorithm Runs Changing the Cross Validation Split

We performed fifteen more runs using the same initial population fed to the GA, but changing the cross validation split three times—five runs per cross validation split. Results are summarized in **Supplementary Tables S5–S7**. In the first split (**Supplementary Table S5**) the KRFPC566 feature was found in the common set of features in four of the five runs, which further supports the hypothesis of this descriptor being a key feature. The same applies to the KRFPC3948 descriptor, which was found in three of the five common sets. Two additional features were found: the KRFPC435 descriptor, repeated in two of the five

common sets, and the KRFPC4830 descriptor, found in three of the five common sets. Remarkably, one of the common sets found consisted of these four features and obtained an average training $F_{1/2}$ score of 0.81 and an average test $F_{1/2}$ score of 0.78, matching the top performance of the first LDA model (Table 3). Both these sets of common features that display top performances (at least up to this point) share two features, KRFPC566 and KRFPC4830, which indicates that KRFPC4830 may also be a key descriptor.

Although in the second cross validation split (Supplementary Table S6) the observed top performance was of 0.54 in the test set, an already encountered descriptor, the KRFPC608 feature, was found repeated in three of the five runs. In the last cross validation split (Supplementary Table S7), the top $F_{1/2}$ score achieved in the test set was of 0.64. The KRFPC435 descriptor was found again in these sets of runs - repeated in three of the five runs-, and also the KRFPC3392 descriptor, found in two of the five runs.

Genetic Algorithm Runs Changing the Initial Population

We also performed five extra runs in which the cross validation split was maintained (the same as in the initial run), but changing the initial population fed to the GA. Strictly speaking, this was performed with three different initial populations, totalizing fifteen extra runs. Results are shown in Supplementary Tables S8-S10. The KRFPC566 descriptor was found to be repeated in eight of the fifteen runs, thus clearly indicating that this feature is indeed important to achieve a high $F_{1/2}$ score with the LDA model. For the first change in the initial population, i.e., the first five runs, the KRFPC3948 descriptor was found in four of the five common sets. Although it was not found repeated in the remaining ten runs, it must be taken into consideration that this descriptor had already been found previously in a high performance set (Supplementary Table S5). The KRFPC435 descriptor shows a similar behavior, which was found earlier along with the KRFPC3948 descriptor (Supplementary Table S5): the results of the second change in the initial population (Supplementary Table S9) show that the KRFPC435 descriptor is repeated in two of five common sets. Other descriptors were also found repeated within the common sets, but at this point we considered them as irrelevant since they did not show up in any of the previous results, specifically, the KRFPC3899 and the KRFPC669 descriptors. Similar to the change in the cross validation split, where one particular change led to a top performing model (Supplementary Table S5), and the two other led to models with a poor performance (Supplementary Tables S6, S7), the same happens with the change in the initial population: The performances shown in Supplementary Tables S9-S10 are low in comparison to previous results. Nonetheless, a particular run shown in Supplementary Table S8 presents the best performance so far. This set of features included the KRFPC435, KRFPC566 and KRFPC3948 descriptors, along with KRFPC3399 and KRFPC3899. The first three descriptors were already included in a high performance model (Run 1 from Supplementary Table S5). Presumably, the last two descriptors are only complementary features (to the first three) related to the change of the initial population.

Final Linear Discriminant Analysis Model

From the previous results (Table 3; Supplementary Table S4-S10), we show in Table 4 the three sets with the best performances. These correspond to the LDA run from Table 3, run 1 from Supplementary Table S5, and run 1 from Supplementary Table S8.

To compare these sets of features, we assessed, feature by feature, in which of the previous runs (Table 3; Supplementary Table S4-S10), each descriptor was present. For each table, we distinguish three cases: 1) the feature was not present in any of the runs of the corresponding table; 2) the feature was present only in one run; 3) the feature was present in more than one run. To quantify the appearance of features across different runs we assigned a partial score to each of the cases described before, being 0 for 1), $\frac{1}{2}$ for 2) and 1 for 3). Taking into account that in Table 3 there is only one LDA run, the sum of partial scores ranges from 0 to 7.5. The feature importance was defined, for each feature, as the sum of partial scores normalized by 7.5, so that the ranking goes between 0 and 1. This information is summarized in Table 5, which allows the visualization of which features are repeated even when the initial conditions of the selection process were changed consistently, like KRFPC566, and which features appear to be dependent on a particular condition of the same process, such as KRFPC608, which appears only in Table 3; Supplementary Table S4, corresponding to the exact same conditions. The comparison between the sets of features presented in Table 4 in terms of the feature importance of each descriptor supports the fact that the set of features composed by KRFPC435, KRFPC566, KRFPC3948 and KRFPC4830 descriptors (Table 4, highlighted in bold) is the most robust. Intuitively, the selected features were the ones which were found more often in the different runs in which the initial conditions of the optimization process were changed. Given a compound, these four features describe the number of times a specific SMARTS pattern is repeated along the molecular structure. The SMARTS associated with each descriptor are shown in Table 6.

Analysis of Misclassified Molecules Within the Final Model

It is important to bear in mind that the scores of the final model shown in Table 4 (highlighted in bold) were achieved over a particular cross validation split. To ensure that the score achieved with these features was not highly dependent on that particular split, we generated 100 different splits and evaluated the model's scores on each one (see Figure 4).

As can be seen, the mean scores are close to the achieved values in the initial split. For this reason, we present the full performance over the test set on each fold (Table 7) using that initial cross validation split. TP, FP, FN and TN are also computed to show exactly how the LDA model is classifying the compounds.

Remarkably, there is only one negative sample misclassified, thus achieving the most important objective sought for this classifier. A great balance is observed between the total number of TPs (17) and FNs (18), which in conjunction with the correct classification of the negative class, gives an overall

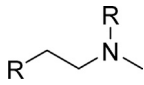
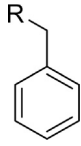
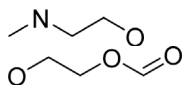
TABLE 4 | Best set of common features found in the complete set of runs.

	Feature 1	Feature 2	Feature 3	Feature 4	Feature 5	F _{1/2} Train	F _{1/2} Test
LDA (Table 3)	KRFPC435	KRFPC566	KRFPC3948	KRFPC4830	-	0.81	0.78
Run 1 (Supplementary Table S5)	KRFPC413	KRFPC566	KRFPC608	KRFPC4830	-	0.80	0.77
Run 1 (Supplementary Table S6)	KRFPC435	KRFPC566	KRFPC3948	KRFPC3399	KRFPC3899	0.80	0.78

TABLE 5 | Analysis of the repeated features over different runs. The column headers display each of the features that appear in Table 4. The rows contain information on whether these features were present or not in each of the performed runs: XX indicates that the feature was repeated across common sets of features, and X indicates the presence of the feature in only one common set.

	KRFPC435	KRFPC566	KRFPC3948	KRFPC4830	KRFPC413	KRFPC608	KRFPC3399	KRFPC3899
Initial run (Table 3)	-	X	-	X	X	X	-	-
Extra five runs (Supplementary Table S2)	-	XX	XX	X	-	XX	-	-
Change in cross validation split (Supplementary Table S5)	XX	XX	XX	X	-	-	-	X
Change in cross validation split (Supplementary Table S6)	-	X	-	-	-	XX	-	-
Change in cross validation split (Supplementary Table S7)	XX	-	-	X	-	X	-	-
Change initial population (Supplementary Table S8)	X	XX	XX	-	-	-	X	XX
Change initial population (Supplementary Table S9)	XX	XX	X	X	X	-	-	-
Change initial population (Supplementary Table S10)	X	XX	-	X	-	-	-	-
Feature importance	0.53	0.8	0.47	0.4	0.13	0.4	0.07	0.2

TABLE 6 | SMARTS patterns associated with set of descriptors of the final LDA model. R represents any atom other than Hydrogen.

KlekotahRoth Fingerprint Count	SMARTS	Molecular Structure
KRFPC435	[#6]-[#7](-[#1])-[#6]-[#6]-[#1]	
KRFPC566	[!#1]-[#6]-[#6]-1 = [#6]-[#6] = [#6]-[#6] = [#6]-1	
KRFPC3948	[#6]-[#7]-[#6]-[#6]-[#8]	
KRFPC4830	[#8]-[#6]-[#6]-[#8]-[#6] = [#8]	

excellent performance. To use this model prospectively, given a new set of molecules, the final model would have to be trained with our entire dataset of 248 compounds. For the new compounds, we would only calculate the KRFPC435, KRFPC566, KRFPC3948 and KRFPC4830 descriptors, and placental transfer would be predicted by inputting the new set of molecules to the ML model.

False Positive Case

The only FP observed in the test sets corresponds to Tubocuraine (CID = 6,000), which belongs to fold 2. As a matter of fact, when

this compound is used to train the LDA model, and this trained model is used to make predictions over the corresponding training set (partitions 1, 3, 4, and 5), this compound is also miss-classified, so this is the only compound belonging to the negative class that is misclassified both in the training and test sets.

Compound 6,000 is described with the following descriptors: KRFPC435 = 4, KRFPC566 = 0, KRFPC3948 = 0 and KRFPC4830 = 0. Similar molecules from the database in terms of these four features, i.e., compounds with KRFPC435 > 0 and the rest of the descriptors equal to zero, are listed in

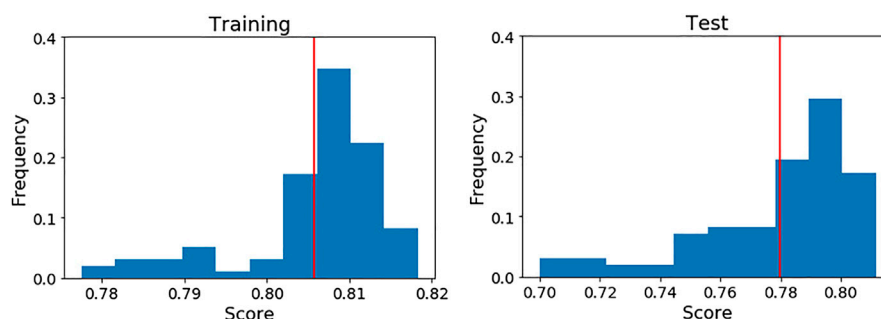


FIGURE 4 | Performance of the final LDA model over 100 different five cross validation splits. The red line indicates the mean score on each case. Left: Training data. Right: Test data.

TABLE 7 | Results using the best set of features (KRFPC435, KRFPC566, KRFPC3948 and KRFPC4830) on each partition over the corresponding test sets.

Partition	F _{1/2} Test	Precision	Recall	AP	TP	FP	FN	TN
1	0.79	1.0	0.43	0.54	3	0	4	43
2	0.74	0.8	0.57	0.63	4	1	3	42
3	0.66	1.0	0.29	0.40	2	0	5	43
4	0.79	1.0	0.43	0.73	3	0	4	42
5	0.93	1.0	0.71	0.76	5	0	2	42
Average	0.78	0.96	0.49	0.61	-	-	-	-

Table 8, together with their actual placental transfer class (C or NC).

Taking into account that compound 6,000 is the only FP in the training and test sets, and considering that the rest of the compounds that cross the placenta shown in **Table 8** are correctly classified whether they were in the training or the test set, it is reasonable to suppose that in the case of having null values in the KRFPC566, KRFPC3948 and KRFPC4830 descriptors, classes are distinguished based on a threshold in the KRFPC435 value: compounds with $KRFPC435 \leq 4$ cross the placenta, while compounds with a $KRFPC435 > 5$ do not cross it.

To assess which threshold our LDA model—trained with the 248 compounds has learned, we inputted several artificial samples with KRFPC435 values ranging from one to nine and the rest of

the descriptors with values equal to zero. We confirmed that compounds are classified as not crossing the placental barrier with $KRFPC435 \geq 4$, which explains why compound 6,000 is misclassified.

False Negative Cases

From **Table 7** 18 FNs were identified in the test sets. Inspecting the representation of the database in terms of the optimal descriptors, we found that the majority of the compounds that cross the placenta were described by null values in the four descriptors (161 compounds), or had only $KRFPC3948 > 0$ (31 compounds). Of the 18 FNs, we found 17 compounds that had one of the representations described before (corresponding to compounds crossing the placenta): 12 compounds had all the four values equal to zero and five compounds had only $KRFPC3948 > 0$. The remaining FN corresponds to compound 441243. Surprisingly, there is another compound (CID = 5362440) with the same representation ($KRFPC435 = 0$, $KRFPC566 = 1$, $KRFPC4830 = 0$ and $KRFPC3948 = 3$) that does not cross the placenta and which is not misclassified. As these two compounds belong to different folds, and effectively checking that there is no compound that crosses the placenta with this exact representation, we assume that the misclassification of compound 441243 is directly related to the cross validation split. Unlike the other 17 FNs, miss-classifications like compound 441243 could be avoided in prospective applications (by the use of both 441243 and 5362440 compounds in the training set).

TABLE 8 | Compounds from the database similar to Tubocurarine (CID = 6,000, in bold) in terms of the four descriptors of the final model. The “Cross” column contains the actual placental transfer class (C or NC).

CID	Name	KRFPC435	KRFPC566	KRFPC3948	KRFPC4830	Cross
47,320	Atracurium Besilate	8	0	0	0	NC
21,233	Dimethyl-Tubocurarine	6	0	0	0	NC
6,000	Tubocurarine	4	0	0	0	C
5,750	Pethidine (Meperidine)	2	0	0	0	C
4,062	Mepivacaine	1	0	0	0	C
43,708	Cefotiam	1	0	0	0	C
89,594	Nicotine	1	0	0	0	C
5288826	Morphine	1	0	0	0	C
2,801	Clomipramine	1	0	0	0	C

It is clear that the majority of FNs arise due to there being compounds belonging to different classes (C-NC) with the same representation. In fact, the positive samples that were correctly classified in the test sets (Table 7) presented clear distinctions in their representations with respect to negative samples. This indicates that to reduce the amount of FNs, at least one more feature should be incorporated. As simple as this may sound, directly incorporating new descriptors to this particular set of features would introduce a bias into the solution [because the relationship between descriptors and classes (C-NC) is already known for the entire dataset], and would finally result in an overfitted model. Although this is a limitation of our method, taking into account that the amount of TPs is at an acceptable level, and that the main goal of having low amounts of false positives was fulfilled by the use of the $F_{1/2}$ score, we consider performing further GA searches or modifying any of the feature selection protocol parameters unnecessary.

CONCLUSION

The study of chemical transfer across the placental membrane from the mother to the fetus is of the utmost importance due to its importance to drug safety, especially in a time when drug prescription during pregnancy is common. Taking into account that *in vivo* data cannot be obtained for ethical reasons, the main difficulty arises from the fact that *in vitro* and *ex vivo* methods cannot directly predict *in vivo* outcomes. In this scenario, the use of *in silico* approaches to complement *ex vivo* and *in vitro* models constitutes an interesting strategy to tackle this challenge.

Although QSAR models have been developed, the datasets used for developing these models were rather small (<100 compounds), and the models validated only on small test sets (<20 compounds). In this study, a database of 248 compounds was compiled, and although this still remains a small dataset, to our knowledge it is the largest reported so far. Also, unlike those studies, which predicted either the CI or the F/M ratio, we treated the placental transfer as a binary classification problem (cross/not cross) rather than as a regression task for a continuous variable.

The results shown in this work support the use of our feature selection protocol, which involves the implementation of a GA that maximizes the $F_{1/2}$ score in conjunction with a five cross validation scheme. The final LDA model displayed key characteristics that are desirable for a ML classifier in this context: 1) it relies on a set of only four features to discriminate between classes; 2) it correctly classifies the majority of both

classes; 3) most importantly, the number of molecules that cross the placenta predicted by the LDA model as not crossing was very low.

One limitation of our ML model is that it was trained with a low amount of data ($N \sim 250$). Strictly speaking, this limitation is not intrinsic to the model itself, but related to our knowledge of placental transfer itself, since there is scarce reliable information publicly available.

As we highlighted before, despite having a low amount of positive (non-crossing) samples, the fact of having only one false positive along the test sets is remarkable. Considering also that a significant number of molecules within the positive class was correctly classified in the test sets (approximately, half of the corresponding positive samples), this supports the incorporation of a ML predictor of placental membrane crossing in a drug discovery campaign.

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

CC conceived and designed the research. CC and MB collected data. JD wrote the code and evaluated its performance. CC, MB, and JD performed analysis, interpretation, and wrote the paper. All authors approved the manuscript for publication.

ACKNOWLEDGMENTS

CC thanks Molsoft LLC (San Diego, CA) for providing an academic license for the ICM program. The Authors thank C. Cortada for helping in the early stages of this work in collecting compound data. Discussions with Julian Maggini, Ventura Simonovich and Waldo Belloso are greatly appreciated.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fchem.2021.714678/full#supplementary-material>

REFERENCES

- Abagyan, R., Totrov, M., and Kuznetsov, D. (1994). ICM?A New Method for Protein Modeling and Design: Applications to Docking and Structure Prediction from the Distorted Native Conformation. *J. Comput. Chem.* 15, 488–506. doi:10.1002/jcc.540150503
- Carpenter, K. A., Cohen, D. S., Jarrell, J. T., and Huang, X. (2018). Deep Learning and Virtual Drug Screening. *Future Med. Chem.* 10, 2557–2567. doi:10.4155/fmc-2018-0314
- Cavasotto, C. N., and Aucar, M. G. (2020). High-Throughput Docking Using Quantum Mechanical Scoring. *Front. Chem.* 8, 246. doi:10.3389/fchem.2020.00246
- Cavasotto, C. N., and Di Filippo, J. I. (2021b). In Silico Drug Repurposing for COVID-19: Targeting SARS-CoV-2 Proteins through Docking and Consensus Ranking. *Mol. Inform.* 40, e2000115. doi:10.1002/minf.202000115
- Cavasotto, C. N., and Di Filippo, J. I. (2021a). Artificial Intelligence in the Early Stages of Drug Discovery. *Arch. Biochem. Biophys.* 698, 108730. doi:10.1016/j.abb.2020.108730

- Chan, H. C. S., Shan, H., Dahoun, T., Vogel, H., and Yuan, S. (2019). Advancing Drug Discovery via Artificial Intelligence. *Trends Pharmacol. Sci.* 40, 592–604. doi:10.1016/j.tips.2019.06.004
- Chen, H., Engkvist, O., Wang, Y., Olivecrona, M., and Blaschke, T. (2018). The Rise of Deep Learning in Drug Discovery. *Drug Discov. Today* 23, 1241–1250. doi:10.1016/j.drudis.2018.01.039
- Eke, A. C., Olagunju, A., Best, B. M., Mirochnick, M., Momper, J. D., Abrams, E., et al. (2020). Innovative Approaches for Pharmacology Studies in Pregnant and Lactating Women: A Viewpoint and Lessons from HIV. *Clin. Pharmacokinet.* 59, 1185–1194. doi:10.1007/s40262-020-00915-w
- Freriksen, J. J. M., Schalkwijk, S., Colbers, A. P., Abduljalil, K., Russel, F. G. M., Burger, D. M., et al. (2020). Assessment of Maternal and Fetal Dolutegravir Exposure by Integrating Ex Vivo Placental Perfusion Data and Physiologically-Based Pharmacokinetic Modeling. *Clin. Pharmacol. Ther.* 107, 1352–1361. doi:10.1002/cpt.1748
- Giaginis, C., Theocharis, S., and Tsantili-Kakoulidou, A. (2012). Current Toxicological Aspects on Drug and Chemical Transport and Metabolism across the Human Placental Barrier. *Expert Opin. Drug Metab. Toxicol.* 8, 1263–1275. doi:10.1517/17425255.2012.699041
- Giaginis, C., Zira, A., Theocharis, S., and Tsantili-Kakoulidou, A. (2009). Application of Quantitative Structure-activity Relationships for Modeling Drug and Chemical Transport across the Human Placenta Barrier: a Multivariate Data Analysis Approach. *J. Appl. Toxicol.* 29, 724–733. doi:10.1002/jat.1466
- Gordon, Z., Glaubach, L., Elad, D., Zaretsky, U., and Jaffa, A. J. (2016). Ex Vivo Human Placental Perfusion Model for Analysis of Fetal Circulation in the Chorionic Plate. *J. Ultrasound Med.* 35, 553–560. doi:10.7863/ultra.15.04080
- 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. Inf. Comput. Sci.* 35, 1039–1045. doi:10.1021/ci00028a014
- Hastie, T., Tibshirani, R., and Friedman, J. H. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. New York, NY: Springer. doi:10.1007/978-0-387-84858-7
- Hutson, J. R., Garcia-Bournissen, F., Davis, A., and Koren, G. (2011). The Human Placental Perfusion Model: A Systematic Review and Development of a Model to Predict In Vivo Transfer of Therapeutic Drugs. *Clin. Pharmacol. Ther.* 90, 67–76. doi:10.1038/clpt.2011.66
- Klekota, J., and Roth, F. P. (2008). Chemical Substructures that Enrich for Biological Activity. *Bioinformatics* 24, 2518–2525. doi:10.1093/bioinformatics/btn479
- Leong, C., Chateau, D., Dahl, M., Falk, J., Katz, A., Bugden, S., et al. (2019). Prescription Medication Use during Pregnancies that Resulted in Births and Abortions (2001–2013): A Retrospective Population-Based Study in a Canadian Population. *PLoS One* 14, e0211319. doi:10.1371/journal.pone.0211319
- Mak, K.-K., and Pichika, M. R. (2019). Artificial Intelligence in Drug Development: Present Status and Future Prospects. *Drug Discov. Today* 24, 773–780. doi:10.1016/j.drudis.2018.11.014
- O'Boyle, N., Banck, M., James, C., Morley, C., Vandermeersch, T., and Hutchison, G. (2011). Open Babel: An Open Chemical Toolbox. *J. Cheminf.* 3, 33. doi:10.1186/1758-2946-3-33
- O'Boyle, N. M., Morley, C., and Hutchison, G. R. (2008). Pybel: a Python Wrapper for the OpenBabel Cheminformatics Toolkit. *Chem. Cent. J.* 2, 5. doi:10.1186/1752-153x-2-5
- Saito, T., and Rehmsmeier, M. (2015). The Precision-Recall Plot Is More Informative Than the ROC Plot when Evaluating Binary Classifiers on Imbalanced Datasets. *PLoS One* 10, e0118432. doi:10.1371/journal.pone.0118432
- Takaku, T., Nagahori, H., Sogame, Y., and Takagi, T. (2015). Quantitative Structure-Activity Relationship Model for the Fetal-Maternal Blood Concentration Ratio of Chemicals in Humans. *Biol. Pharm. Bull.* 38, 930–934. doi:10.1248/bpb.b14-00883
- Wang, C. C., Lin, P., Chou, C. Y., Wang, S. S., and Tung, C. W. (2020). Prediction of Human Fetal-Maternal Blood Concentration Ratio of Chemicals. *PeerJ* 8, e9562. doi:10.7717/peerj.9562
- Yap, C. W. (2011). PaDEL-descriptor: An Open Source Software to Calculate Molecular Descriptors and Fingerprints. *J. Comput. Chem.* 32, 1466–1474. doi:10.1002/jcc.21707
- Zhang, Y.-H., Xia, Z.-N., Yan, L., and Liu, S.-S. (2015). Prediction of Placental Barrier Permeability: a Model Based on Partial Least Squares Variable Selection Procedure. *Molecules* 20, 8270–8286. doi:10.3390/molecules20058270
- Zhu, Q. (2020). On the Performance of Matthews Correlation Coefficient (MCC) for Imbalanced Dataset. *Pattern Recognition Lett.* 136, 71–80. doi:10.1016/j.patrec.2020.03.030

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 Di Filippo, Bollini and Cavaotto. 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 Potential Binding Sites of Sialic Acids on the RBD Domain of SARS-CoV-2 Spike Protein

Bingqian Li^{1,2†}, Lin Wang^{1†}, Huan Ge^{3†}, Xianglei Zhang¹, Penxuan Ren¹, Yu Guo⁴, Wuyan Chen⁵, Jie Li⁵, Wei Zhu¹, Wenzhang Chen¹, Lili Zhu^{3*} and Fang Bai^{1*}

¹Shanghai Institute for Advanced Immunochemical Studies and School of Life Science and Technology, ShanghaiTech University, Shanghai, China, ²Department of Chemistry, Imperial College London, London, United Kingdom, ³State Key Laboratory of Bioreactor Engineering, Shanghai Key Laboratory of New Drug Design, School of Pharmacy, East China University of Science and Technology, Shanghai, China, ⁴State Key Laboratory of Medicinal Chemical Biology and College of Pharmacy, Nankai University, Tianjin, China, ⁵National Center for Protein Science Shanghai, Shanghai, China

OPEN ACCESS

Edited by:

Fei Ye,
Zhejiang Sci-Tech University, China

Reviewed by:

Zhijian Xu,
Shanghai Institute of Materia Medica,
Chinese Academy of Sciences, China
Feixiong Cheng,
Cleveland Clinic, United States
Min-Yeh Tsai,
Tamkang University, Taiwan

*Correspondence:

Lili Zhu
zhulili@ecust.edu.cn
Fang Bai
baifang@shanghaitech.edu.cn

[†]These authors have contributed
equally to this work

Specialty section:

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

Received: 28 January 2021

Accepted: 25 February 2021

Published: 22 July 2021

Citation:

Li B, Wang L, Ge H, Zhang X, Ren P,
Guo Y, Chen W, Li J, Zhu W, Chen W,
Zhu L and Bai F (2021) Identification of
Potential Binding Sites of Sialic Acids
on the RBD Domain of SARS-CoV-2
Spike Protein.
Front. Chem. 9:659764.
doi: 10.3389/fchem.2021.659764

COVID-19, caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), is still an emergent pandemic for humans. The virus infection is achieved by penetrating its spike protein to host cells via binding with ACE2. Moreover, recent studies show that SARS-CoV-2 may have multiple receptors that need to be further revealed. SARS-CoV-2 shares similar sequences of the spike protein with the Middle East Respiratory Syndrome Coronavirus (MERS-CoV), which can invade host cells by binding to either DPP4 or sialic acids. Sialic acids can be linked to the terminal of glycoproteins and gangliosides are used as one of the receptors of many types of viruses. Therefore, it is very interesting to determine whether sialic acid is a potential receptor of SARS-CoV-2. To address this question, we took N-Acetylneuraminic acid (Neu5Ac), a type of predominant sialic acid found in human cells, as the molecular probe to computationally search the surface of the spike protein to locate the potential binding sites of Neu5Ac. SPR analysis and mass spectrum analysis confirmed the interaction between Neu5Ac and spike protein. This study shows that sialic acids can moderately interact with the spike protein of SARS-CoV-2 by binding between the two RBDs of the spike protein, indicating it could be a potential secondary or auxiliary receptor of SARS-CoV-2.

Keywords: SARS-CoV-2, receptor, sialic acid, RBD domain, spike protein

INTRODUCTION

The new coronavirus disease 2019 (COVID-19) is caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). COVID-19 has caused a worldwide health emergency with parallel effects on the economy. Over a hundred million cases were reported by February 20, 2021, with thousands of deaths every day (World Health Organization, 2021). The molecular mechanisms of SARS-CoV-2 infection are still not clear and urgently needed to be explored. To date, several medical agents, including small molecular agents and vaccines are in the process of clinical trials (Liu et al., 2020).

SARS-CoV-2 belongs to the beta-coronavirus family which contains Human beta-Coronavirus (HCoV-OC43), Human beta-Coronavirus (HCoV-HKU1), Severe Acute Respiratory Syndrome Coronavirus (SARS-CoV), and the Middle East Respiratory Syndrome Coronavirus (MERS-CoV) (Hu et al., 2015; Hulswit et al., 2019). It shares similarity in sequence with SARS-CoV and MERS-

CoV, being with identity of 79 and 50%, respectively, (Lu et al., 2020; Petrosillo et al., 2020). Compared with MERS-CoV and SARS-CoV, the SARS-CoV-2 virus has a relatively low mortality rate (around 2.3%) (9.5% for SARS-CoV and 34.4% for MERS-CoV), but a significantly higher rate of transmission (Dömling and Gao, 2020; Petrosillo et al., 2020).

Coronavirus enters the host cell mainly by binding to the host cell receptor. Both SARS-CoV and SARS-CoV-2 share the same human cell receptor, angiotensin-converting enzyme 2 (hACE2), while MERS-CoV enters human cells by binding to dipeptidyl-peptidase (DPP4) (Wan et al., 2020). Cell surface protease TMPRSS2 and lysosomal cathepsins activate the SARS-CoV-2 and may cleave the spike protein at two distinct sites. This presence of pre-activation enables SARS-CoV-2 to be less dependent on target cell activation. Studies have also shown a higher binding affinity to hACE2 for SARS-CoV-2 than for SARS-CoV (Hoffmann et al., 2020). Two other potential host receptors for SARS-CoV-2 entry, kringle containing transmembrane protein 1 (KREMEN1) and asialoglycoprotein receptor 1 (ASGR1), were recently discovered (Gu et al., 2020).

Sialic acid is a generic term for a family of derivatives of neuraminic acid, an acid sugar with a nine-carbon backbone. It is generally found in the terminal position on a variety of glycoconjugates, which cover the surfaces of many different cell types, playing important cellular functions, including mediating the attachment, and entry of types of viruses, such as influenza viruses, orthomyxoviruses, infectious salmon anemia virus, as well as coronavirus (Matrosovich et al., 2015). HCoV-OC43 and HCoV-HKU1 can interact with 9-O-acetyl-sialic acid to infect the host cell (Hulswit et al., 2019; Tortorici et al., 2019). Different from HCoV-OC43, MERS-CoV also shows a stronger preference interaction with α 2,3-linked sialosides other than α 2,6-linked sialosides (Park et al., 2019). One recent study reported the identification of binding between SARS-CoV-2 and sialic acids (N-acetyl neuraminic acid) by using a new lateral flow detection system. (Baker et al., 2020) This indicates that sialic acids may be a candidate receptor, and their binding molecular mechanisms with spike protein need to be further studied.

SARS-CoV-2 is formed as an enveloped structure that contains RNA genome, spike (S) protein, nucleocapsid (N) protein, membrane (M) protein, and envelop (E) protein. The homo-trimeric S protein contains two subunits, S1 and S2, covering the cleavage sites at R685 and S686 (Hu et al., 2015; Woo et al., 2020). The N-terminal S1 subunit mainly comprises the N-terminal domain (NTD) and receptor-binding domain (RBD), which is responsible for hACE2 binding. However, KREMEN1 and ASGR1 bind to both NTD and RBD (Gu et al., 2020). The C-terminal S2 subunit is mainly made up of heptad repeats 1 and 2 (HR1 and HR2), as well as the transmembrane domain (TM), which specializes in membrane fusion while entering the cell (Dömling and Gao, 2020; Woo, et al., 2020).

The spike glycoprotein of SARS-CoV-2 is usually in a “down” conformational state to escape from the immune response. When it approaches a target cell receptor, RBD shifts its position to bind with a human cell receptor, which turns the protein into an “up”

conformational state (Shang et al., 2020). The types of conformational structures of S protein and compositions were fully discovered using cryo-EM, with 31% S protein in the “down” conformational state, 55% in the state with one RBD “up”, and 14% in the state with two RBDs “up” (Cai et al., 2020; Ke et al., 2020). In reality, S protein is largely shielded by glycans, which are utilized for thwarting immune response from the host. N-glycans at N165 and N234 play a critical role in the process of the state changes of RBD (Casalino et al., 2020). Previous studies have predicted 22 N-glycosylation and 4 O-glycosylation sites on the surface of S protein (Woo et al., 2020). 17 of 22 N-glycosylated and 2 O-glycosylated sites were observed using the cryo-EM technique. (Shajahan et al., 2020; Woo et al., 2020).

Computational techniques have already been widely used in drug discovery. Although experimental technologies provide straightforward observation in studies, they are normally time-consuming and laborious. Moreover, recent techniques focus on studying biological molecular mechanisms by using molecular dynamics (MD), which improve the understanding of reaction mechanisms and protein dynamic behavior (Karplus and McCammon, 2002). For example, Arantes' group used MD simulations to explore strategies for developing vaccines of SARS-CoV-2 (Arantes et al., 2020). Deganutti's group focused on identifying druggable binding sites on the SARS-CoV-2 spike protein by using supervised molecular dynamics. (Deganutti et al., 2021). Chauhan's team outlined some key aspects in molecular structure that may affect inhibition performance in organic corrosion inhibitors using molecular dynamics techniques (Chauhan et al., 2021).

Yadav's group tested FDA-approved drugs on several new SARS-CoV-2 proteases using molecular docking techniques (Yadav et al., 2020).

The present study designed a comprehensive framework by combining multiple computational modeling methods with experimental technologies, aiming to determine whether and how sialic acids bind with the spike protein of SARS-CoV-2. Several studies have shown glycosylation can alter the thermodynamic stability and folding as well as conformations of proteins, resulting in an increase in protein free energies (Shental-Bechor and Levy, 2008; Gavrillov et al., 2015). Hence, this work also studies whether the binding of sialic acids may also be affected by the existence of glycan ligands on the surface of the spike protein.

RESULTS AND DISCUSSION

Identification of Possible Binding Sites of Sialic Acids

To explore the potential binding sites of sialic acids on the surface of the spike protein, in which both the conformational change and the glycosylation states were considered, a series of ligandable binding site identification simulations were performed on the four different modeled protein structures of the spike protein (illustrated in **Supplementary Figure S1**). Both RBD “down” and one RBD “up” conformations, as well as the glycosylation, were considered. As a result, four protein structures were constructed based on the different RBD conformational states and

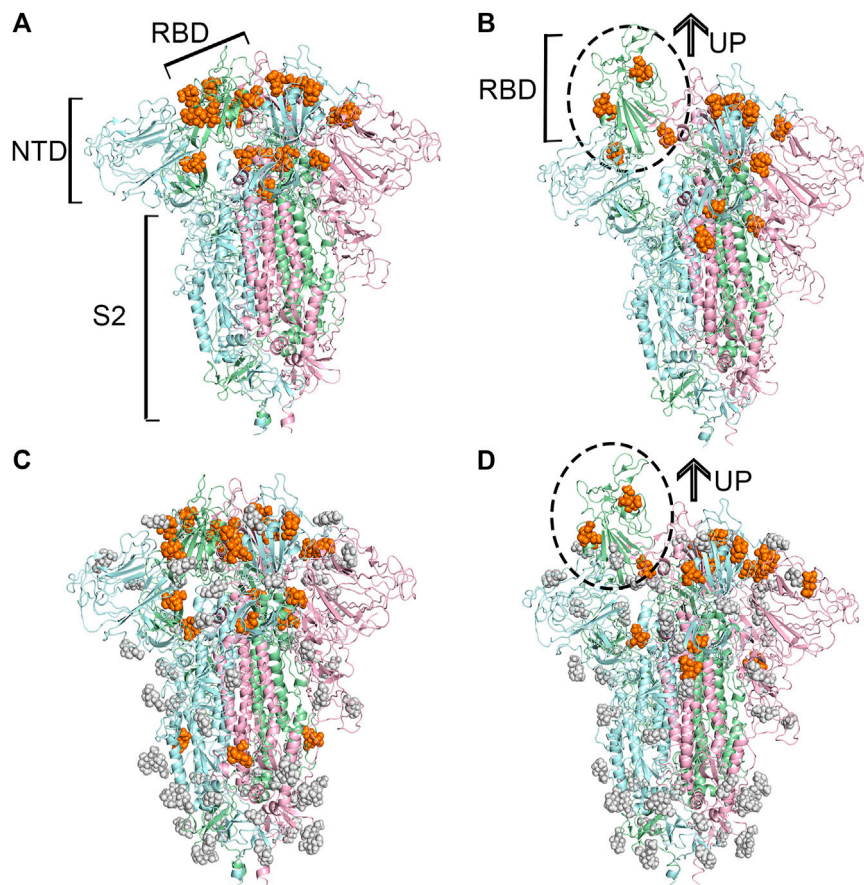


FIGURE 1 | Predicted potential bindings of sialic acids on the surface of the spike protein. Potential ligandable binding sites were identified on the four constructed models of the trimeric spike protein by using FTMap (Ngan et al., 2012) and SiteMap (Halgren, 2009). The different colors of the cartoon models in each figure represent different chains of the protein: chain A is shown in pale-green, chain B is in pale-blue, and chain C is in light-pink. The protein structures which contain gray sphere balls represent glycosylated state S trimer (the gray sphere models are the glycosylation), and the orange balls represent sialic acid molecules. **(A)** 21 and **(B)** 15 sialic acid molecules were observed to bind to the different places of the surface of unglycosylated spike protein in “down” and “up” conformational states, respectively, **(C)** 23 potential sites for sialic acid binding were identified on the surface of glycosylated spike protein in “down” conformational state, and 17 were found **(D)** in the “up” conformational state.

glycosylated states. A detailed description of the protein structure modeling process has been given in the methods section. Firstly, twenty-one potential binding sites mainly locating on the S1 domain were identified by using SiteMap (Halgren, 2009), which is a computational druggable binding site characterizing method for proteins. To double-check the prediction results, another binding site identification method, which was a fragment-based druggable “hot spot” searching method developed with a different algorithm (the Fourier domain correlation algorithm), named as FTMap, was also used to search again on the surface of these four structures. The spike protein is trimer and some positions identified should be symmetrically located. However, some of such binding sites were only captured once or twice in our calculations. In this way, we artificially corrected the results by adding the missing ones. Finally, forty candidate positions were obtained. Interestingly, the obtained active sites were shown at similar positions with the results from SiteMap. These candidate binding sites should be evaluated further by using other techniques.

Based on these identified candidate binding sites, several rounds of molecular docking simulations were performed to verify whether sialic acids could interact with or not. Theoretically, the sialic acid may be extended with oligosaccharides to decorate glycoproteins and gangliosides at the host cell surface (Schauer and Kamerling, 2018). Hence, the binding sites of sialic acids could be relatively exposed to solvent, in other words, on the surface of a protein. Given this, the candidate binding sites located on the surface of proteins were extensively explored. Because of this, we used N-Acetylneuraminic acid (Neu5Ac), a type of predominant sialic acid, as a small molecular probe to detect the potential anchor site of glycogen chains on the surface of the S protein. The spike protein is a pivotal trimeric structure, therefore, once the candidate binding sites were discovered on one chain of the protein, the additional binding sites symmetrically located on the other chains would be manually added to our candidate list if they were not observed accidentally. Finally, for the protein in the “up” conformational state, 15 potential binding sites on the

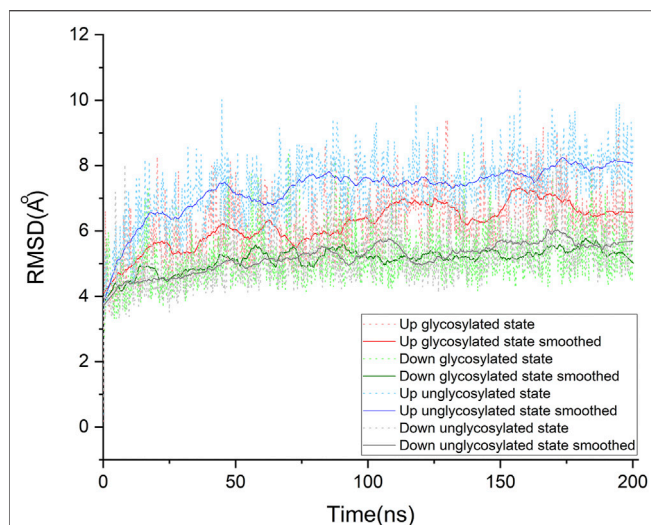


FIGURE 2 | Ca root-mean-square deviation (RMSD) of the molecular dynamics simulations for our four systems. Each system was studied extensively by running three times of molecular dynamic simulations. For each protein system, only one of the simulations was taken out to make this plot representative. The raw data of RMSDs are shown in dot lines, and the fluctuations of RMSD are smoothed by using the Savitzky-Golay method in OriginPro, version 2020 (OriginLab Corporation, Northampton, MA, and United States), with the polynomial order as 1 and polynomial order as 50. The details for other trajectories are shown in supporting information (See Supporting Information, **Supplementary Figure S2**). All systems show an RMSD variation around 3–4 Å, which is contributed by the large conformational motion of flexible loops of the spike protein. Compared with these “down” conformational states, these “up” conformational states show more obvious fluctuations.

unglycosylated spike protein, and 17 on the glycosylated spike protein were observed, respectively. Compared with the “up” state, 21 and 23 potential binding sites were found to be able to accommodate sialic acids in the unglycosylated and glycosylated “down” state of the protein. Through the comprehensive analysis of the locations of all the candidate binding sites, a total of 40 unique candidate binding sites were collected from those four protein models and numbered from 1 to 40. In general, most of the predicted binding sites of sialic acid molecules were found to locate in the RBD of the protein, as shown in **Figure 1**. Based on the above docking simulations and artificial correlation, these four modeled protein structures were modeled as multiple sialic acid-bound complex structures.

Determining the Most Likely Binding Site for the Sialic Acid

To find the most likely binding site for the sialic acid, where a sialic acid molecule could stably bind, a series of molecular dynamics simulations were carried out to monitor the stability of the bindings of sialic acids against the candidate binding sites, obtained from the molecular docking simulations mentioned above. Theoretically, the weaker bound sialic acids would dissociate faster. For each protein structure, three repeated MD simulations were performed. These modeled four sialic

acid-protein complex structures were subjected to molecular dynamics simulations for 200 ns and generated twelve independent trajectories. As shown in **Figure 1**. The Ca-RMSD of each trajectory shows relative fluctuations of the proteins in a range of about 3–4 Å (**Figure 2**; **Supplementary Figure S2**). These curves which show obvious fluctuations of the conformational change of the protein along the trajectories are mainly contributed by the larger numbers of flexible loops of the protein. The “down” states of S protein generally show relative small-scale fluctuations than “up” states, suggesting that later structures could be less stable. During the simulations, some sialic acid molecules docked to the protein surface fly away after 20 ns of simulations, whereas some are stably staying in their positions after 200 ns. Therefore, the most probable (strongly bound) sites for the sialic acid can be distinguished from others. To compare the strength of those binding sites, all unique potential pockets were numbered from 1 to 40 by simply aligning all four structures. The strength of interactions of every ligand (sialic acid) in each frame of the trajectories was analyzed and plotted in heatmaps (See Supporting Information, **Supplementary Figures S3–S6**). Moreover, the depth of color suggests the number of molecular interactions, i.e., hydrogen bonds, hydrophobic contacts, and ionic bonds, etc., between ligand and protein in each frame. Interestingly, we found some sialic acids that flew away after 20 ns and re-bounded to the protein at different positions from their initial binding sites, which then left again after several nanoseconds. However, this action did not show up in any repeats for a particular sialic acid and therefore is random interactions. From an overall perspective, the glycosylation may be beneficial to strengthening the binding of sialic acids to the spike protein, as the interactions observed in the glycosylated spike protein are significantly more than unglycosylated proteins (as shown in **Supplementary Figures S3–S6**). By comparing the heatmaps, sialic acids at positions SA_6, SA_7, and SA_24 are appeared to show stable interactions within the 200 ns trajectories for glycosylated “RBD” down conformational state (**Supplementary Figure S3**). The positions of SA_7 and SA_24 can also be observed to be stable within the 200 ns trajectories for unglycosylated “RBD” down conformational state (**Supplementary Figure S4**). Sialic acid is strongly bound to the position of SA_6 in glycosylated “RBD” up conformational state (**Supplementary Figure S5**). By superposing the last frames from the trajectories for these four protein structures (**Figure 3**), we found that the positions of SA_6, SA_7, and SA_24 are similar and symmetrically located between every two adjacent RBDs from different chains.

The predicted binding modes of sialic acid molecules in the positions of SA_6, SA_7, and SA_24 for glycosylated “RBD” down conformational state are shown in **Figure 4**. In general, the residue of D405 cooperated with its neighboring residue of R403 or R408, participating in all sialic acid interactions in the positions of SA_6, SA_7, and SA_24. On the other hand, sialic acids form a salt bridge with K378 of an adjacent chain, whereas in the position of SA_24, the predicted binding orientation is slightly different from the other two positions, i.e., interacting with the residue of S375 but not K378. Three sialic acid molecules symmetrically bind around the residue of D405 in each chain

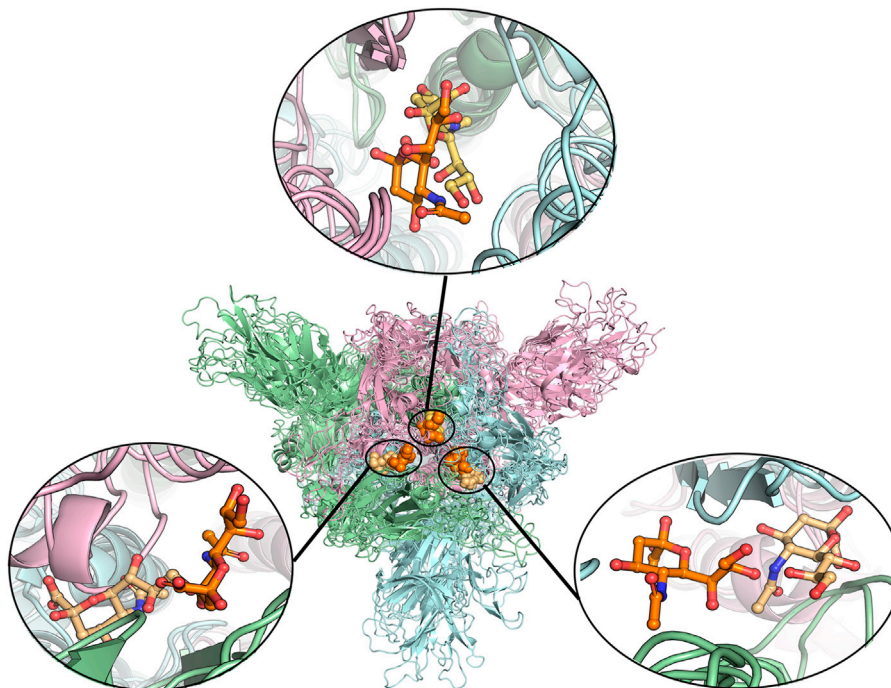


FIGURE 3 | Alignment of different protein structures to identify the overlapped sialic acid-binding positions, that is, SA₆, SA₇, and SA₂₄. The figure shows the alignment of these final stable complex structures of sialic acid with spike protein, generated by the MD simulations starting from those four different protein structures. The sialic acid-binding positions, i.e., SA₆, SA₇, and SA₂₄ were found to be conserved in “down” conformational states of the spike protein and to be observable in “up” conformational states. Position SA₆ is between chain B (pale-blue) and chain C (light-pink), position SA₇ is between chain A (pale-green) and chain B and position SA₂₄ is between chain A and chain C. The stable bound sialic acids are shown in different colors. Three sialic acids in the glycosylated “down” conformational state are shown in orange color. Two sialic acids in the unglycosylated “down” conformational state are shown in light orange color. One sialic acid bound on the glycosylated “up” conformational state is shown in yellow-orange color. The detailed interaction modes for the bound modes are shown in **Figure 4** and **Supplementary Figures S7, S8**. No stable sialic acid appears in these three positions for the glycated “up” state.

with forming couples of molecular interactions, such as the hydrogen bonds with G504, G404, and K417. Moreover, the observed binding mode of sialic acid molecules against these three positions in the different conformational or glycosylation states are shown in **Supplementary Figures S7, S8**. Both “down” state systems show sialic acids stably bind between two adjacent RBDs in different chains, whereas in the “up” state system, there is only one sialic acid in the same position because one “up” chain of RBD could distort the binding sites (**Supplementary Figure S7**). Alternatively, this can be explained by the shifting position of one RBD. As the RBD shifting upwards, it moves further from the other two RBDs, causing loss of stable interactions. The positions of sialic acids in each trajectory are slightly shifted from the origin docking position but still in the same region. These findings indicate that sialic acids could bind to the RBD domain of the spike protein of SARS-CoV-2, but not the N-terminal domain of the S1 domain (NTD) that the binding sites of the sialic acid locate on the MERS’s or other viruses’ spike protein. Therefore, experimental validations were performed against Neu5Ac and the RBD of the spike protein.

On the other hand, as shown in **Supplementary Figures S3–S6**, SA₂₈ shows an obvious preference for binding with sialic acid. The binding mode has been illustrated in

Supplementary Figure S9. This position is embedded inside of the protein. Despite its strong interaction with sialic acid, it should not be the binding site for glycans.

One recent study reported conformational accessibility and binding strength of the S protein to its receptor of ACE2. In these reported simulations, five potential ligand-binding pockets were identified to expose and correlate with the conformational shifts of S protein (Peng et al., 2020). The authors also screened the compound database to identify potential ligands and reported one polyhydroxy (Quercetin) compound that is somehow like the sialic acid. This makes us curious whether this pocket is the site of the sialic acid. By carefully comparing, pocket four was found in the report to be close to our predicted site, but not fully overlapping. This indicates that the binding of the sialic site in this position may be involved in some relationship with the conformational change of spike protein, but we do not know how and why at this stage. It is a very interesting topic that needs to be explored further in the future.

Mass Spectrometry Analysis

To validate the binding of sialic acid to spike protein. Mass spectrometry analysis was firstly performed to qualitatively determine the bindings. According to our computational

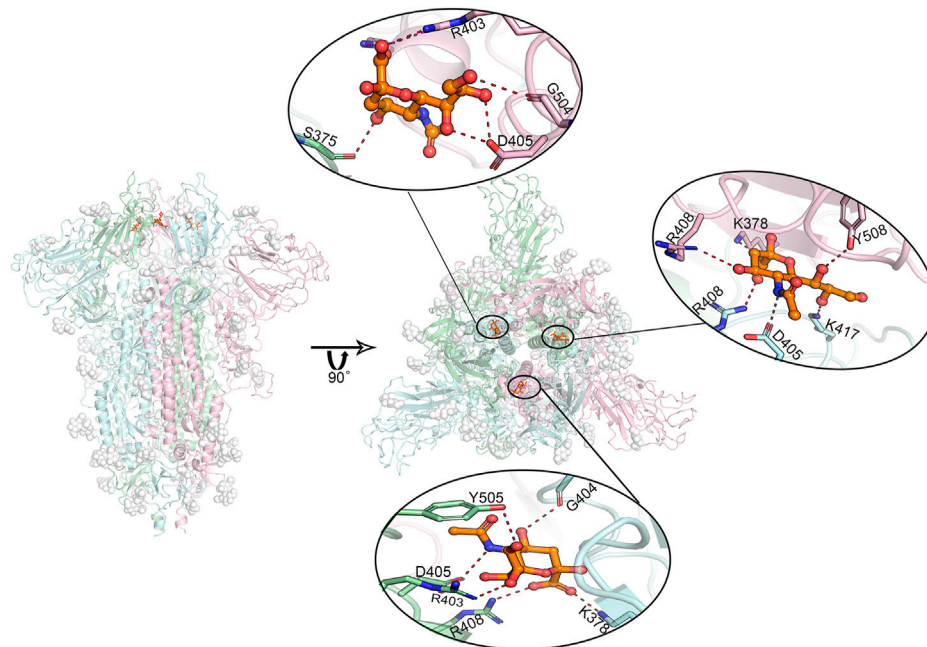


FIGURE 4 | Predicted most likely binding modes of sialic acids with glycosylated spike protein at positions SA_6, SA_7, and SA_24. Molecular dynamics simulations identified three stable interacting sites, locating between every two adjacent RBD domains. The figure in the middle shows the relative positions of these three bound sialic acids. The enlarged binding areas show the detailed molecular interactions at each binding site. Position SA_6 is between chain B (pale-cyan) and chain C (light-pink). Sialic acid can form hydrogen bonds with the residues of R403, G504 as well as D405 from chain B, and the residue of S375 from chain C. Position SA_7 locates between chain A (pale-green) and chain B, where the sialic acid molecule is observed to specifically interact with the residues of R403, D405, R408 and Y505 of chain A, and the residues of K378 as well as G404 of chain B in the form of salt bridges or hydrogen bonds. Position SA_24 lies between chain A and chain C, where the sialic acid binds to the residues of D405, R408 as K417 of chain B, and the residues of K378, R408, and Y508 of chain C. Overall, the residue D405 in each of these three chains shows a significant role in sialic acid binding. The interactions to K378/S375 on their adjacent chains may play a role in further stabilizing sialic acid molecules. Dash lines represent hydrogen or ionic bonds. White spheres are the glycans that are artificially modified on the protein.

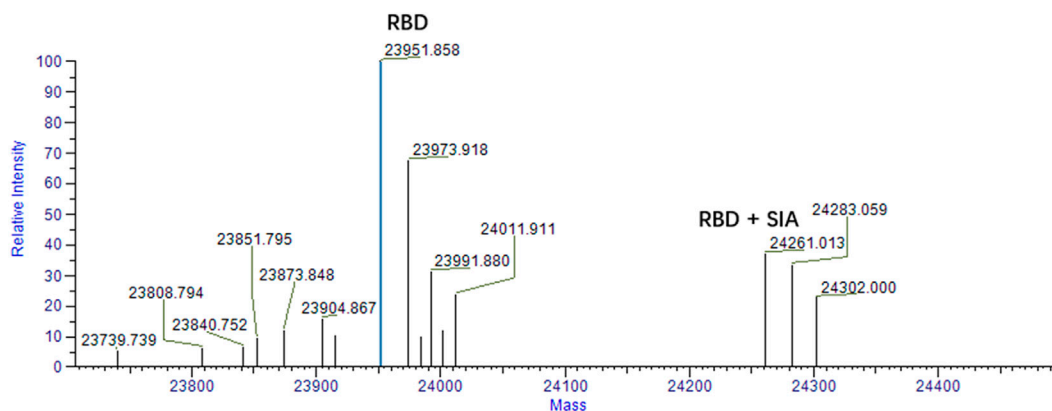


FIGURE 5 | Mass spectrum analysis of RBD and sialic acid binding complex. The peak at 23,951 represents the RBD domain. The ligand-protein binding complex is shown by the peak at 24,261 m/z. The condition for analysis is RBD: SIA = 1:50.

prediction, sialic acids may bind the site which is around the residue of D405 and belongs to the RBD domain. Therefore, mass spectrometry analysis was carried out between sialic acids and the RBD domain of spike protein. The experiment was carried out in a protein-ligand ratio of 1:50. As shown in **Figure 5**, the presence of a peak at 24,261 suggests the presence of a ligand-protein

complex. The ratio of peaks at 23,951 and 24,261 is approximately 2.5:1.

Surface Plasmon Resonance Analysis

To further confirm our findings, we analyzed the binding affinity between the RBD domain and sialic acid by using Surface

Plasmon Resonance (SPR). SPR is a biophysical method that can quantify molecular-molecular binding interactions. It allows “label-free” detection in real time and has been widely used to monitor interface processes (Shepherd et al., 2014). The experimental result shows that sialic acid effectively binds to the RBD domain with a rapid dissociation rate ($k_{\text{off}} = 0.0127$ 1/s), and the binding is concentration-dependent, as illustrated in **Figure 6**. The experiment shows a binding affinity of $K_D = 27.26$ μM . From previous computational simulations, sialic acid is most likely bound between two adjacent RBD domains, while the SPR only detects the interaction between ligand and one RBD (1:1 model). In other words, the real binding affinity should be stronger than the observed 27.26 μM .

Further Discussion

SARS-CoV-2 is experiencing rapid evolution. A number of mutations have been observed and most of them have occurred on the spike protein (CNCB, 2021; CoVariants, 2021). At present, the most important reported mutations are D614G and N501Y, which have been found can increase the binding affinities between the spike protein and ACE2. The two mutations are away from the identified bindings sites of sialic acid, and may not impact their bindings (Leung et al., 2021; Weissman et al., 2021). We also compared all reported single point mutations of S protein of SARS-CoV-2 (**Supplementary Table S1**), and mutations that happened outside the proposed binding sites for sialic acids, therefore, are less likely to affect the binding of sialic acids.

This paper proposes a new binding site for sialic acids on the RBD domain of Spike proteins. Drug repurposing can be done on this pocket through artificial intelligence (Zhou et al., 2020). Apart from Yadav's work, last year, Martin's group reported that Toremifene, an FDA-approved drug, could work on SARS-CoV-

2 S protein and NSP14 (Martin and Cheng, 2020). This provides a new scientific orientation for further studies.

Moreover, the allosteric binding sites on the S protein of SARS-CoV-2 have been discovered and reported by designing a comprehensive framework, combining computational methods and experimental validation (Paola et al., 2020). The allosteric sites, being different from the conventional active sites, can allosterically alter the conformation of the proteins and regulate the functions. Therefore, it is important to identify the potential allosteric binding sites of sialic acids and probe the corresponding allosteric molecular mechanisms, to better understand the functions of sialic acids in triggering the virus invasion.

METHODS

Constructing the Spike Protein Structures

Full-length spike protein structure models were built based on experimentally obtained protein structures, the PDB codes of these proteins are 6VXX for “down” and 6VSB for “up” conformational states. The missing fragments of the sequence were added by comparing different spike protein PDB models. Gaps between loops were filled by referring to full-length sequence of S protein by using Maestro (Zhu et al., 2014). A part of the incomplete RBD in all three chains was replaced by a modeled fully-sequenced model (modification based on the structure with the PDB code of 6M17). Based on the built-up reference models, two spike protein trimer models, one “up” and one “down” conformational state, was built using the Maestro Homology modeling method (Cappel et al., 2016). The glycosylation in the above PDBs were kept, and an additional one missing O-glycosylation at N801 was added manually according to the literature (Woo et al., 2020). The other two models without glycans were built by removing the glycans from previously built structures (**Supplementary Figure S1**).

Detecting Potential Druggable Binding Sites for Sialic Acids on Spike Protein

Based on the above built four protein structures, SiteMap (Halgren, 2007; Halgren, 2009) and FTMap (Ngan et al., 2012) were used as two individual methods, which gave complementary results, identifying active sites on the surface of the spike protein structure. The SiteMap is a server of Schrödinger which predicts possible binding sites by scanning through the protein surface (Halgren, 2007). FTMap scans the entire protein by placing lots of probes in the funnel. More detailed descriptions of FTMap have already been published (Brenke et al., 2009; Ngan et al., 2012; Kozakov et al., 2015). Both methods were used by setting the parameters as default. Compared with SiteMap, the binding sites found by FTMap were relatively embedded into the protein. Overall, approximately 21 candidate sites were detected on the surface of the trimer. Then, sialic acid molecules were placed onto the trimer structure using molecular docking simulations by Glide (Tubert-Brohman et al., 2013). Two potential binding sites on “down” glycosylated state were added artificially because there were similar binding sites had been observed on the symmetric chain.

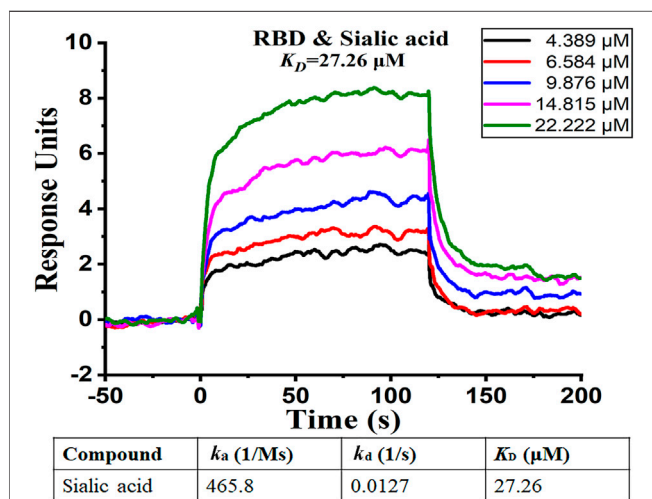


FIGURE 6 | SPR analysis of binding affinity between sialic acid and RBD domain. The figure shows different levels of response with different sialic acid concentrations. The curve suggests the presence of binding between sialic acid and the RBD domain.

Molecular Dynamics Simulations

All simulation systems were built and minimized using Desmond (Bowers et al., 2006), TIP3P (Gillan et al., 2016) as water model, neutralized by Na cation and Cl anion. Simulations for “down” state without glycans initially measured 45,770,357 Å³, NaCl at a concentration of 0.15 M, for a total of about 440,000 atoms. Simulation for “down” state with glycans initially measured 4,804,764 Å³, NaCl at a concentration of 0.15 M, for a total of about 442,000 atoms. Simulation for “up” state without glycans initially measured 4,770,069 Å³, NaCl at a concentration of 0.15 M, for a total of about 447,000 atoms. Simulation for “down” state with glycans initially measured 4,902,837 Å³, NaCl at a concentration of 0.15 M, for a total of about 460,000 atoms. The systems were modeled in the OPLS_2005 force field (Shivakumar et al., 2010). A molecular dynamics simulation was carried out using Desmond (Bowers et al., 2006). The systems were pre-production run for 50 ns. After that, each trajectory was set for a longer simulation as long as 200 ns, an ensemble at 310 K (37°C), and 1 bar. Every system was then repeated three times with the same conditions but various initial velocities. Trajectories were analyzed using a simulation interaction analysis module in Maestro (Bowers et al., 2006).

Mass Spectrometry Analysis

Proteins were dissolved in 25 ammonium acetate at a concentration of 10 μM, the drugs dissolved DMSO were diluted by 25 ammonium acetate to 100 μM. Then proteins were incubated with an equal volume of the drugs.

The above-mixed solutions were then injected into Orbitrap Fusion MS (Thermo Scientific) through direct injection. The MS was operated in intact protein mode. Data were analyzed with BioPharma Finder (Thermo Scientific) software (Marcoux et al., 2015).

Surface Plasmon Resonance Analysis

We carried out surface plasmon resonance (SPR) experiments using BIAcore T200 to evaluate the kinetic parameters of sialic acid binding to RBD. The purified RBD (residues 319–591), which was diluted in sodium acetate solution (pH 4.5) with a final concentration of 50 μg/ml, was immobilized covalently on a CM5 sensor chip. The final immobilization level was 4,430.3 resonance units (RU). The running buffer was PBS, 0.005% (vol/vol) surfactant P20, pH 7.4, and 1% DMSO. Sialic acid was diluted using the running

buffer from the top concentration. The measurements were performed at a flow rate of 30 μL/min. For each binding cycle, the analyte was injected for 120 s and the dissociation time was 180 s. Data were analyzed using BIAevaluation 1.1 software.

DATA AVAILABILITY STATEMENT

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

AUTHOR CONTRIBUTIONS

FB and LZ designed the research. BL, LW, PR, and FB performed computational simulations and analysis. HG, XZ, and LZ performed experimental binding assays. YG, WC, and JL prepared and provided the recombinant protein samples for validation. WZ and WZ performed mass spectrometry experiments and analysis. All the authors were involved in writing the articles.

FUNDING

This work was supported by the Science and Technology Commission of Shanghai Municipality grants (Grant IDs: 20431900102, 20431900100, 2043190012, and 20430780300); the startup package from ShanghaiTech University; the National Natural Science Foundation of China (NSFC) (Grant ID: 82003654t) and the National Science and Technology Major Project “Key New Drug Creation and Manufacturing Program”, China (Grant ID: 2018ZX09711002); and Shanghai Science and Technology Development Funds (Grant ID: 20QA1406400).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fchem.2021.659764/full#supplementary-material>

REFERENCES

- Arantes, P. R., Saha, A., and Palermo, G. (2020). Fighting COVID-19 Using Molecular Dynamics Simulations. *ACS Cent. Sci.* 6, 1654–1656. doi:10.1021/acscentsci.0c01236
- Baker, A. N., Richards, S.-J., Guy, C. S., Congdon, T. R., Hasan, M., Zwetsloot, A. J., et al. (2020). The SARS-COV-2 Spike Protein Binds Sialic Acids and Enables Rapid Detection in a Lateral Flow Point of Care Diagnostic Device. *ACS Cent. Sci.* 6, 2046–2052. doi:10.1021/acscentsci.0c00855
- Bowers, K. J., Chow, D. E., Xu, H., Dror, R. O., Eastwood, M. P., Gregersen, B. A., et al. (2006). Scalable Algorithms for Molecular Dynamics Simulations on Commodity Clusters. *IEEE* 06, 43. doi:10.1109/SC.2006.54
- Brenke, R., Kozakov, D., Chuang, G.-Y., Beglov, D., Hall, D., Landon, M. R., et al. (2009). Fragment-based Identification of Druggable ‘hot Spots’ of Proteins Using Fourier Domain Correlation Techniques. *Bioinf.* 25, 621–627. doi:10.1093/bioinformatics/btp036
- Cai, Y., Zhang, J., Xiao, T., Peng, H., Sterling, S. M., Walsh, R. M., Jr, et al. (2020). Distinct Conformational States of SARS-CoV-2 Spike Protein. *Science* 369, 1586–1592. doi:10.1126/science.abd4251
- Cappel, D., Hall, M. L., Lenselink, E. B., Beuming, T., Qi, J., Bradner, J., et al. (2016). Relative Binding Free Energy Calculations Applied to Protein Homology Models. *J. Chem. Inf. Model.* 56, 2388–2400. doi:10.1021/acs.jcim.6b00362
- Casalino, L., Gaieb, Z., Goldsmith, J. A., Hjorth, C. K., Dommer, A. C., Harbison, A. M., et al. (2020). Beyond Shielding: The Roles of Glycans in the SARS-CoV-2 Spike Protein. *ACS Cent. Sci.* 6, 1722–1734. doi:10.1021/acscentsci.0c01056
- Chauhan, D. S., Verma, C., and Quraishi, M. A. (2021). Molecular Structural Aspects of Organic Corrosion Inhibitors: Experimental and Computational Insights. *J. Mol. Struct.* 1227, 129374. doi:10.1016/j.molstruc.2020.129374

- CNCB (2021). *Mutation Sites in the Spike Protein*. Available at: <https://bigd.big.ac.cn/ncov/protein?lang=en> (Accessed February 19, 2021).
- CoVariants (2021). *Variants*. Available at: <https://covariants.org/variants> (Accessed February 19, 2021).
- Deganutti, G., Prisch, F., and Reynolds, C. A. (2021). Supervised Molecular Dynamics for Exploring the Druggability of the SARS-CoV-2 Spike Protein. *J. Comput. Aided. Mol. Des.* 35, 195–207. doi:10.1007/s10822-020-00356-4
- Dömling, A., and Gao, L. (2020). Chemistry and Biology of SARS-CoV-2. *Chem.* 6, 1283–1295. doi:10.1016/j.chempr.2020.04.023
- Gavrilov, Y., Shental-Bechor, D., Greenblatt, H. M., and Levy, Y. (2015). Glycosylation May Reduce Protein Thermodynamic Stability by Inducing a Conformational Distortion. *J. Phys. Chem. Lett.* 6, 3572–3577. doi:10.1021/acs.jpclett.5b01588
- Gillan, M. J., Alfè, D., and Michaelides, A. (2016). Perspective: How Good Is DFT for Water?. *J. Chem. Phys.* 144, 130901. doi:10.1063/1.4944633
- Gu, Y., Cao, J., Zhang, X., Gao, H., Wang, Y., Wang, J., et al. (2020). Interaction Network of SARS-CoV-2 with Host Receptome through Spike Protein. *BioRxiv* [Epub ahead of print]. doi:10.1101/2020.09.09.287508
- Halgren, T. (2007). New Method for Fast and Accurate Binding-Site Identification and Analysis. *Chem. Biol. Drug Des.* 69, 146–148. doi:10.1111/j.1747-0285.2007.00483.x
- Halgren, T. A. (2009). Identifying and Characterizing Binding Sites and Assessing Druggability. *J. Chem. Inf. Model.* 49, 377–389. doi:10.1021/ci800324m
- 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
- Hu, B., Ge, X., Wang, L.-F., and Shi, Z. (2015). Bat Origin of Human Coronaviruses. *Virology* 12, 221. doi:10.1186/s12985-015-0422-1
- Hulswit, R. J. G., Lang, Y., Bakkers, M. J. G., Li, W., Li, Z., Schouten, A., et al. (2019). Human Coronaviruses OC43 and HKU1 Bind to 9-O-Acetylated Sialic Acids via a Conserved Receptor-Binding Site in Spike Protein Domain A. *Proc. Natl. Acad. Sci. U.S.A.* 116, 2681–2690. doi:10.1073/pnas.1809667116
- Karplus, M., and McCammon, J. A. (2002). Molecular Dynamics Simulations of Biomolecules. *Nat. Struct. Biol.* 9, 646–652. doi:10.1038/nsb0902-646
- Ke, Z., Oton, J., Qu, K., Cortese, M., Zila, V., McKeane, L., et al. (2020). Structures and Distributions of SARS-CoV-2 Spike Proteins on Intact Virions. *Nature* 588, 498–502. doi:10.1038/s41586-020-2665-2
- Kozakov, D., Grove, L. E., Hall, D. R., Bohnuud, T., Mottarella, S. E., Luo, L., et al. (2015). The FTMap Family of Web Servers for Determining and Characterizing Ligand-Binding Hot Spots of Proteins. *Nat. Protoc.* 10, 733–755. doi:10.1038/nprot.2015.043
- Leung, K., Shum, M. H., Leung, G. M., Lam, T. T., and Wu, J. T. (2021). Early Transmissibility Assessment of the N501Y Mutant Strains of SARS-CoV-2 in the United Kingdom, October to November 2020. *Euro. Surveill.* 26 (1), 15–21. doi:10.2807/1560-7917.ES.2020.26.1.2002106
- Liu, Y., Wang, K., Massoud, T. F., and Paulmurugan, R. (2020). SARS-CoV-2 Vaccine Development: An Overview and Perspectives. *ACS Pharmacol. Transl. Sci.* 3, 844–858. doi:10.1021/acspstci.0c00109
- 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, 565–574. doi:10.1016/s0140-6736(20)30251-8
- Marcoux, J., Champion, T., Colas, O., Wagner-Rousset, E., Corvaia, N., Van Dorsselaer, A., et al. (2015). Native Mass Spectrometry and Ion Mobility Characterization of Trastuzumab Emtansine, a Lysine-Linked Antibody Drug Conjugate. *Protein Sci.* 24, 1210–1223. doi:10.1002/pro.2666
- Martin, W. R., and Cheng, F. (2020). Repurposing of FDA-Approved Toremifene to Treat COVID-19 by Blocking the Spike Glycoprotein and NSP14 of SARS-CoV-2. *J. Proteome Res.* 19, 4670–4677. doi:10.1021/acs.jproteome.0c00397
- Matrosovich, M., Herrler, G., and Klenk, H. D. (2013). Sialic Acid Receptors of Viruses. *Top. Curr. Chem.* 367, 1–28. doi:10.1007/128_2013_466
- Ngan, C. H., Bohnuud, T., Mottarella, S. E., Beglov, D., Villar, E. A., Hall, D. R., et al. (2012). FTMAP: Extended Protein Mapping with User-Selected Probe Molecules. *Nucleic Acids Res.* 40, W271–W275. doi:10.1093/nar/gks441
- Di Paola, L., Hadi-Aljanvand, H., Song, X., Hu, G., and Giuliani, A. (2020). The Discovery of a Putative Allosteric Site in the SARS-CoV-2 Spike Protein Using an Integrated Structural/Dynamic Approach. *J. Proteome Res.* 19, 4576–4586. doi:10.1021/acs.jproteome.0c00273
- Park, Y.-J., Walls, A. C., Wang, Z., Sauer, M. M., Li, W., Tortorici, M. A., et al. (2019). Structures of MERS-CoV Spike Glycoprotein in Complex with Sialoside Attachment Receptors. *Nat. Struct. Mol. Biol.* 26, 1151–1157. doi:10.1038/s41594-019-0334-7
- Peng, C., Zhu, Z., Shi, Y., Wang, X., Mu, K., Yang, Y., et al. (2020). Computational Insights into the Conformational Accessibility and Binding Strength of SARS-CoV-2 Spike Protein to Human Angiotensin-Converting Enzyme 2. *J. Phys. Chem. Lett.* 11, 10482–10488. doi:10.1021/acs.jpclett.0c02958
- Petrosillo, N., Viceconte, G., Viceconte, G., 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
- Schauer, R., and Kamerling, J. P. (2018). Exploration of the Sialic Acid World. *Adv. Carbohydr. Chem. Biochem.* 75, 1–213. doi:10.1016/bs.accb.2018.09.001
- Shajahan, A., Supekar, N. T., Gleinich, A. S., and Azadi, P. (2020). Deducing the N- and O-Glycosylation Profile of the Spike Protein of Novel Coronavirus SARS-CoV-2. *Glycobiology* 30, 981–988. doi:10.1093/glycob/cwaa042
- 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
- Shental-Bechor, D., and Levy, Y. (2008). Effect of Glycosylation on Protein Folding: A Close Look at Thermodynamic Stabilization. *Proc. Natl. Acad. Sci.* 105 (24), 8256–8261. doi:10.1073/pnas.0801340105
- Shepherd, C. A., Hopkins, A. L., and Navratilova, I. (2014). Fragment Screening by SPR and Advanced Application to GPCRs. *Prog. Biophys. Mol. Biol.* 116, 113–123. doi:10.1016/j.pbiomolbio.2014.09.008
- 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. Theor. Comput.* 6, 1509–1519. doi:10.1021/ct900587b
- Tortorici, M. A., Walls, A. C., Lang, Y., Wang, C., Li, Z., Koerhuis, D., et al. (2019). Structural Basis for Human Coronavirus Attachment to Sialic Acid Receptors. *Nat. Struct. Mol. Biol.* 26, 481–489. doi:10.1038/s41594-019-0233-y
- Tubert-Brohman, I., Sherman, W., Repasky, M., and Beuming, T. (2013). Improved Docking of Polypeptides with Glide. *J. Chem. Inf. Model.* 53, 1689–1699. doi:10.1021/ci400128m
- Wan, Y., Shang, J., Graham, R., Baric, R. S., and Li, F. (2020). Receptor Recognition by the Novel Coronavirus from Wuhan: an Analysis Based on Decade-Long Structural Studies of SARS Coronavirus. *J. Virol.* 94, e00127–20. doi:10.1128/JVI.00127-20
- Weissman, D., Alameh, M.-G., de Silva, T., Collini, P., Hornsby, H., Brown, R., et al. (2021). D614G Spike Mutation Increases SARS CoV-2 Susceptibility to Neutralization. *Cell Host & Microbe* 29 (1), 23–31. doi:10.1016/j.chom.2020.11.012
- Woo, H., Park, S.-J., Choi, Y. K., Park, T., Tanveer, M., Cao, Y., et al. (2020). Developing a Fully Glycosylated Full-Length SARS-CoV-2 Spike Protein Model in a Viral Membrane. *J. Phys. Chem. B* 124, 7128–7137. doi:10.1021/acs.jpcb.0c04553
- World Health Organization (2021). WHO Coronavirus Disease (COVID-19) Dashboard. Available at: <https://covid19.who.int/> (Accessed February, 2021).
- Yadav, R., Parihar, R., Dhiman, U., Dhamija, P., Upadhyay, S., and Imran, M. (2020). Docking of FDA Approved Drugs Targeting NSP-16, N-Protein and Main Protease of SARS-CoV-2 as Dual Inhibitors. *Biointerface Res. Appl. Chem.* 11 (3), 9848–9861. doi:10.33263/BRIAC113.98489861
- Zhou, Y., Wang, F., Tang, J., Nussinov, R., and Cheng, F. (2020). Artificial Intelligence in COVID-19 Drug Repurposing. *Lancet Digit. Health* 2 (12), 667–676. doi:10.1016/S2589-7500(20)30192-8
- Zhu, K., Day, T., Warshaviak, D., Murrett, C., Friesner, R., and Pearlman, D. (2014). Antibody Structure Determination Using a Combination of Homology Modeling, Energy-Based Refinement, and Loop Prediction. *Proteins* 82, 1646–1655. doi:10.1002/prot.24551

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 Li, Wang, Ge, Zhang, Ren, Guo, Chen, Li, Zhu, Chen, Zhu and Bai. 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.



Potential Inhibitors of Fascin From A Database of Marine Natural Products: A Virtual Screening and Molecular Dynamics Study

Lirui Lin^{1,2}, Kai Lin^{1,2}, Xiaodong Wu¹, Jia Liu¹, Yinwei Cheng^{3,4}, Li-Yan Xu^{3,4*}, En-Min Li^{1,3*} and Geng Dong^{1,2*}

¹Department of Biochemistry and Molecular Biology, Shantou University Medical College, Shantou, China, ²Medical Informatics Research Center, Shantou University Medical College, Shantou, China, ³Key Laboratory of Molecular Biology in High Cancer Incidence Coastal Area of Guangdong Higher Education Institutes, Shantou University Medical College, Shantou, China, ⁴Cancer Research Center, Shantou University Medical College, Shantou, China

OPEN ACCESS

Edited by:

Zhongjie Liang,
Soochow University, China

Reviewed by:

José Pedro Cerón-Carrasco,
Catholic University San Antonio de
Murcia, Spain
Fei Ye,
Northeast Agricultural University,
China

*Correspondence:

Li-Yan Xu
lyxu@stu.edu.cn
En-Min Li
nmli@stu.edu.cn
Geng Dong
gdong@stu.edu.cn

Specialty section:

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

Received: 03 June 2021

Accepted: 17 September 2021

Published: 07 October 2021

Citation:

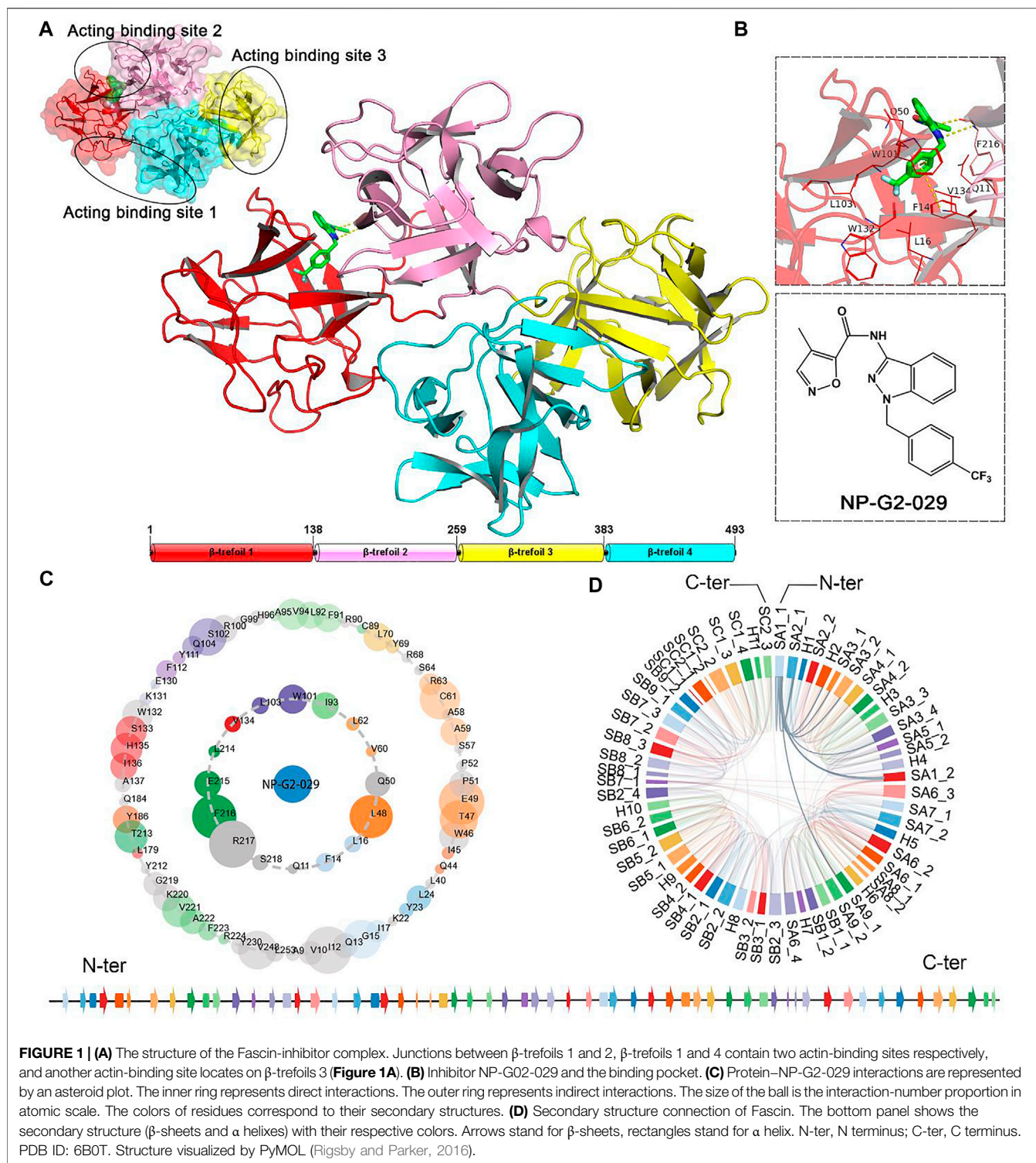
Lin L, Lin K, Wu X, Liu J, Cheng Y,
Xu L-Y, Li E-M and Dong G (2021)
Potential Inhibitors of Fascin From A
Database of Marine Natural Products:
A Virtual Screening and Molecular
Dynamics Study.
Front. Chem. 9:719949.
doi: 10.3389/fchem.2021.719949

Marine nature products are unique compounds that are produced by the marine environment including plants, animals, and microorganisms. The wide diversity of marine natural products have great potential and are versatile in terms of drug discovery. In this paper, we use state-of-the-art computational methods to discover inhibitors from marine natural products to block the function of Fascin, an overexpressed protein in various cancers. First, virtual screening (pharmacophore model and molecular docking) was carried out based on a marine natural products database (12015 molecules) and provided eighteen molecules that could potentially inhibit the function of Fascin. Next, molecular mechanics generalized Born surface area (MM/GBSA) calculations were conducted and indicated that four molecules have higher binding affinities than the inhibitor NP-G2-029, which was validated experimentally. ADMET analyses of pharmacokinetics demonstrated that one of the four molecules does not match the criterion. Finally, ligand Gaussian accelerated molecular dynamics (LiGaMD) simulations were carried out to validate the three inhibitors binding to Fascin stably. In addition, dynamic interactions between protein and ligands were analyzed systematically. Our study will accelerate the development of the cancer drugs targeting Fascin.

Keywords: marine nature product, fascin, virtual screening, docking, molecular dynamics

INTRODUCTION

With a deeper understanding of the particularity of the marine environment and the diversity of marine biology, researchers have developed many applications based on aquatic and marine resources (Carroll et al., 2021). Extreme conditions in the ocean in terms of temperature, salinity, pressure, and illumination promote marine organisms to evolve and create a unique system with different processes of absorption and metabolism (Montaser and Luesch, 2011). In the metabolism of marine organisms, enormous and innovative marine natural products (MNPs) are produced, and those products can be exploited to develop new functional materials and drugs (Barbosa and Roque, 2019). In recent years, many new compounds have been discovered from marine life, which have also benefited from the rapid development of technology (Hu et al., 2011; Hu et al., 2015; Greco and Cinquegrani, 2016; Ruiz et al., 2016; Blunt et al., 2017; Bilal et al., 2018; Blunt



et al., 2018). To exploit the data of MNPs for the treatment of diseases conveniently, some databases of MNPs are built for drug screening and other research on ocean resources (Haroun et al., 2019).

One of the applications of MNPs is drug discovery, e.g., drugs for cancer treatment, as tumor metastasis is the main

cause of cancer-related deaths (Chen et al., 2010). Cell invasion and migration are essential features of tumor cells and actin cytoskeleton reconstruction triggers the switch of protrusive tissue, e.g., filopodia, lamellipodia, and lamellipodia (Machesky and Li, 2010). Fascin is one of the actin-binding proteins and it is overexpressed in various

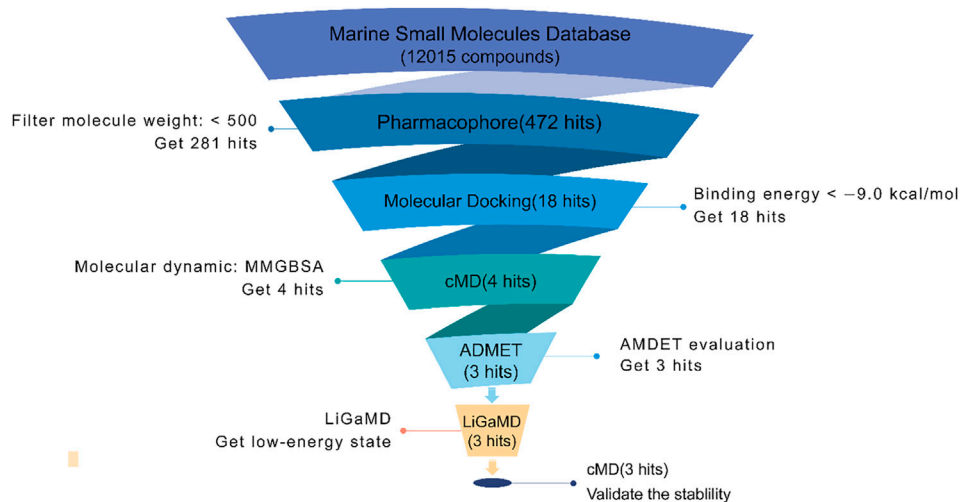


FIGURE 2 | The workflow of inhibitors screening in this study. The pharmacophore model used LigandScout software; docking used the AutoDock Vina module of LigandScout (Nguyen et al., 2020).

types of cancer. Fascin plays a key role in the formation of filopodia, which leads to increased cell movability in multiple transformed cells (Conesa-Zamora et al., 2013; Tan et al., 2013). Some studies have indicated that Fascin can be used as a diagnostic marker and therapeutic target for aggressive tumors (Tan et al., 2013; Rodrigues et al., 2017).

Fascin was first found as a cross-linking protein in sea urchin (Kane, 1975) and later identified in *Drosophila*, *Xenopus* (Holthuis, Schoonderwoert, and Martens, 1994), mice (Edwards et al., 1995), and human beings (Duh et al., 1994; Yamashiro-Matsumura and Matsumura, 1985). Fascin is one of the components of actin bundles, with 55 k Da and four β -trefoil domains (**Figures 1A,B**) (Yamashiro-Matsumura and Matsumura, 1985). There are six pairs of two-stranded β -hairpins in each β -trefoil domain with 3-fold symmetry (Murzin, Lesk, and Chothia, 1992; Ponting and Russell, 2000). These four β -trefoils of Fascin form a quadrilateral-like shape and each β -trefoil located in the catercorner (Yamashiro-Matsumura and Matsumura, 1985). Fascin is a monomeric protein and functions by bundling actin filament at its monomeric state. Previous studies have suggested that Fascin has three individual surfaces for its bundling activity to actin, i.e., binding site 1, 2, and 3 (**Figure 1A**) (Yang et al., 2013). The junction between β -trefoils 1 and 2 of Fascin is suggested to be essential for its actin-binding activity, which is termed actin-binding site 2. (**Figure 1A**) (Ono et al., 1997).

To block actin-Fascin interaction and inhibit filament assembly, several small molecule inhibitors have been developed from chemical libraries for biochemical and pathological research (Chen et al., 2010; Huang et al., 2015; Huang et al., 2018). However, the inhibitor exploration for Fascin is still under development, due to the limitation of current inhibitors on efficiency and specificity. NP-G2-029 and NP-G2-044 are two inhibitors targeting Fascin, which show a

strong effect, weakening the migration ability of human breast cancer cells (Han et al., 2016; Huang et al., 2018). The IC_{50} values of NP-G2-029 and NP-G2-044 are 0.19 and 0.07 μ m in the F-actin-bundling assay. The crystal structure of the Fascin-NP-G2-029 complex was solved by Huang et al. (2018). Six hydrophobic residues surround the benzene ring of NP-G2-029, i.e., Glu11, Phe14, Leu16, Gln50, Trp101, Leu103, Trp132, Val134, and Phe216 (**Figure 1B**), and the benzene ring also forms edge-to-face pi-pi stacking with Phe14 and Trp101. Two hydrogen-bond interactions are formed between the backbone of Phe216 and the pyrazole and amide groups of NP-G2-029.

The second structure connections of Fascin (**Figures 1C,D**) show the residues in the binding pocket of NP-G2-029, and the connections of β -sheets and helices in Fascin. It can be seen from **Figure 1D** that interactions between secondary structures are complex, indicating that the correlations between those structures are strong. The bottom panel shows the secondary structure (β -sheets and α helices) with their respective colors (Conducted by Protein Contacts Atlas server) (Kayikci et al., 2018).

In recent years, computer-aided drug discovery (CADD) methods are extensively used for new drug discovery. The pharmacophore model is a ligand-based method to screen lead compounds (Gupta et al., 2019; Wang et al., 2019; Fu et al., 2020; Liu et al., 2020; Liu et al., 2020). It is a rapid and powerful method for the first screening from a large chemical library. The pharmacophore model is often used in combination with structure-based methods, e.g., molecular docking (Saikia and Bordoloi, 2019). Molecular docking programs can be used to predict the bound poses of ligands and to rank them with scoring functions. (Huang and Zou, 2010; Lopez-Vallejo et al., 2011; Garcia-Sosa and Maran, 2021). With CADD approaches, the cost of drug research and development can be reduced markedly (Xiang et al., 2012). These approaches can provide a

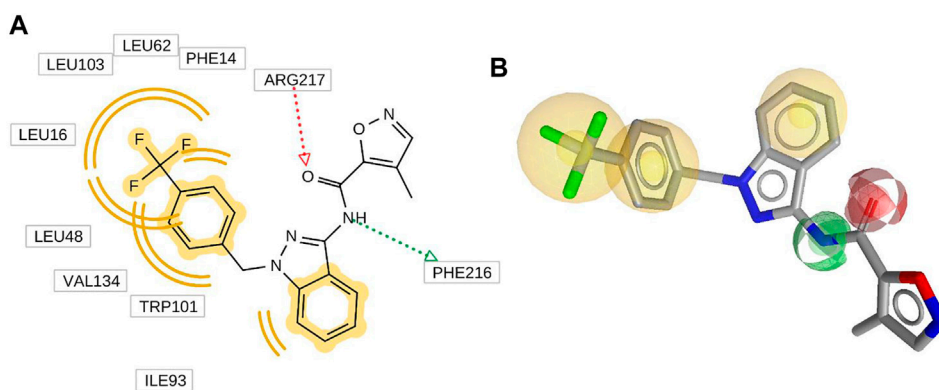


FIGURE 3 | 2D **(A)** and 3D **(B)** inhibitor NP-G2-029 with its abstract pharmacophore model generated by LigandScout. Hydrogen bond acceptor (red arrow), hydrogen bond donor (green arrow), hydrophobic interaction, aromatic ring feature interaction (yellow sphere).

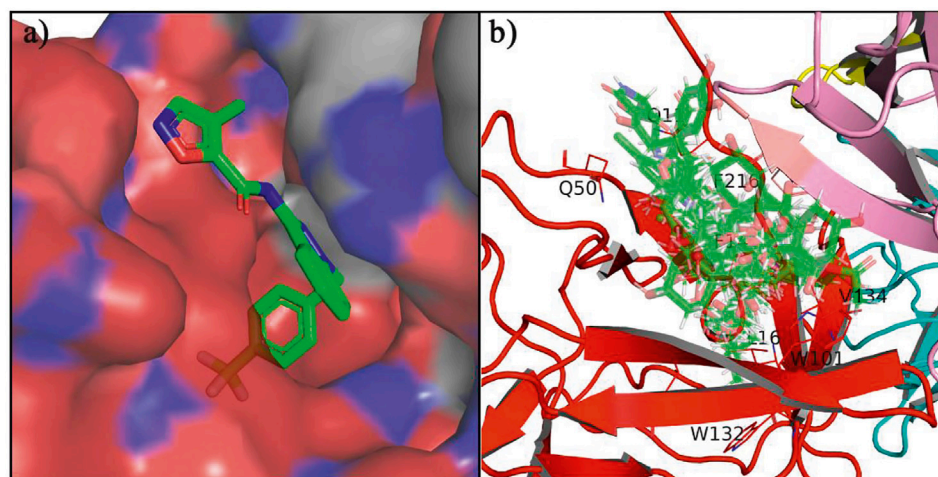


FIGURE 4 | **(A)** The binding pocket of Fascin with inhibitor NP-G2-029; **(B)** The molecular docking results for 18 small-molecules.

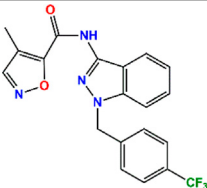
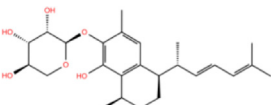
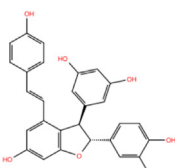
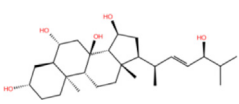
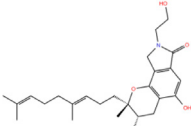
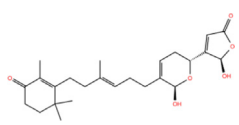
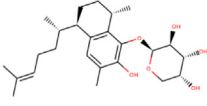
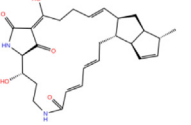
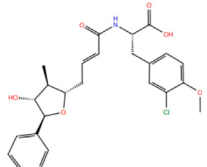
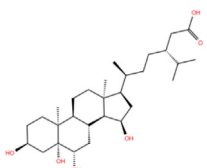
comprehensive insight into biomolecule mechanisms and improve the effectiveness of the drug development process (Macalino et al., 2015).

It is noteworthy that molecule docking results still need further evaluation and analyses (Rastelli and Pinzi, 2019), and molecular dynamics (MD) simulation is an often-used method to improve the accuracy of molecular docking. Meanwhile, the dynamic properties of proteins can be investigated in depth by MD simulations, which can provide detailed information on the process of ligand binding at an atomic level and this information is significant for drug discovery (De Vivo et al., 2016). Molecular mechanics generalized born surface area (MM/GBSA) is an efficient method for binding free energy calculation, which is used to assess docking poses, determine structural stability and predict binding affinities (Ylilauri and Pentikainen, 2013; Wang et al., 2019). On the other hand, the free energy landscape can be calculated to explore the intermediate states and global minimum of

biomolecule (Buckley et al., 2017). However, conformation transition overcoming energy barrier usually needs a millisecond time scale or even longer, depending on the height of the barrier (Miao, Feher, and McCammon, 2015). To overcome this challenge, many enhanced sampling methods have been developed (Bernardi, Melo, and Schulten, 2015). In addition, small molecules have various conformations because of their flexibility in solvent, and the dynamics of small molecules are significant for the induced-fit process (Francis et al., 2019). Thus, exploring the binding state of the inhibitor is important for drug design.

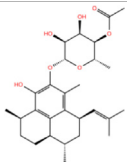
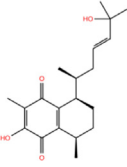
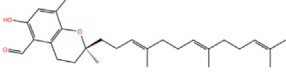
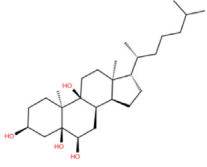
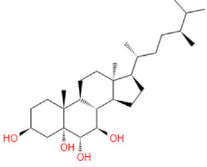
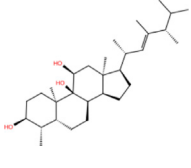
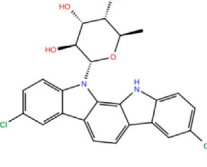
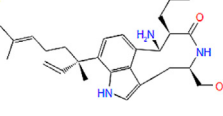
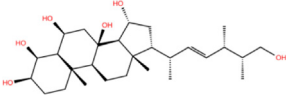
In this study, we use several CADD methods to screen small molecules from an MNP library, as indicated by the workflow in **Figure 2**. First, based on the marine natural products database (12,015 molecules), virtual screening using the pharmacophore model and molecular docking were carried out to discover potential inhibitors of Fascin. Then, the top 18 compounds were selected for further MD simulations, and the binding affinity of each inhibitor was calculated. ADMET

TABLE 1 | Data collection of potential inhibitors for Fascin by molecular docking. Unit: kcal/mol.

No./Compound	Library ID	Structure	Molecular weight	Binding affinity
01 _a NP-G2-029 C ₂₀ H ₁₅ F ₃ N ₄ O ₂	No data		400.36	−10.80
02 C ₂₅ H ₃₆ O ₆	ZINC000238749885		432.56	−10.50
03 C ₂₈ H ₂₂ O ₇	ZINC000014693073		470.48	−9.90
04 C ₂₇ H ₄₆ O ₅	ZINC000044387599		450.66	−9.70
05 C ₂₅ H ₃₅ NO ₅	ZINC000014714664		429.56	−9.60
06 C ₂₅ H ₃₄ O ₆	ZINC000238761262		430.54	−9.50
07 C ₂₅ H ₃₈ O ₆	ZINC000040915743		434.57	−9.40
08 C ₂₇ H ₃₄ N ₂ O ₅	ZINC000042851223		466.58	−9.40
09 C ₂₅ H ₂₅ NO ₆ Cl	No data		470.93	−9.30
10 C ₂₉ H ₅₀ O ₆	ZINC000255214715		494.71	−9.30

(Continued on following page)

TABLE 1 | (Continued) Data collection of potential inhibitors for Fascin by molecular docking. Unit: kcal/mol.

No./Compound	Library ID	Structure	Molecular weight	Binding affinity
11 C ₂₈ H ₄₀ O ₇	ZINC000042888842		488.62	-9.20
12 C ₂₀ H ₂₈ O ₄	ZINC000005890667		332.44	-9.10
13 C ₂₈ H ₄₀ O ₃	ZINC000014767734		424.62	-9.10
14 C ₂₇ H ₄₈ O ₄	ZINC000137671675		436.68	-9.10
15 C ₂₈ H ₅₀ O ₄	ZINC000137547990		450.70	-9.10
16 C ₃₀ H ₅₂ O ₄	No data		476.74	-9.10
17 C ₂₄ H ₂₀ Cl ₂ N ₂ O ₄	ZINC000085599962		471.34	-9.10
18 C ₂₇ H ₃₉ N ₃ O ₂	No data		437.63	-9.00
19 C ₂₈ H ₄₈ O ₆	ZINC000044387005		480.69	-9.00

^aActive controlled indicator.

predictions were also performed to study pharmacokinetic properties. Furthermore, Ligand Gaussian accelerated Molecular Dynamics (LiGaMD) were carried out on the three potential inhibitors to study low-energy states (Miao,

Bhattacharai, and Wang, 2020). To validate the low-energy states in LiGaMD, we performed an extended conventional MD. Finally, we analyzed the binding pockets of Fascin with different potential inhibitors.

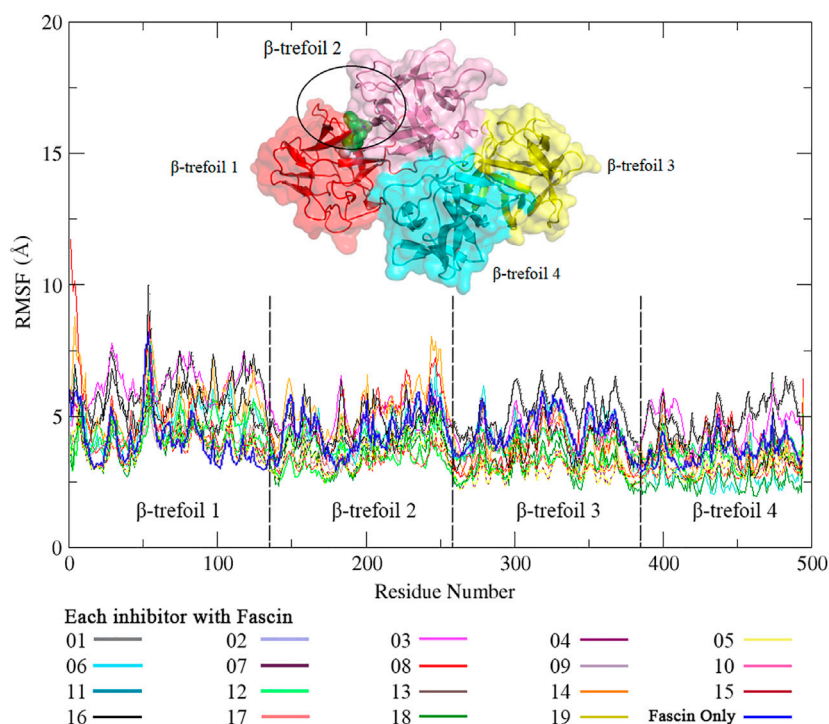


FIGURE 5 | The RMSF of residues in complex with NP-G2-029 and 18 inhibitors in conventional MD simulations. Different color lines stand for the different inhibitors.

TABLE 2 | Binding affinity for each inhibitor by MM/GBSA (Unit: kcal/mol. Potential inhibitors those meet the criterion are highlighted as bold values.)

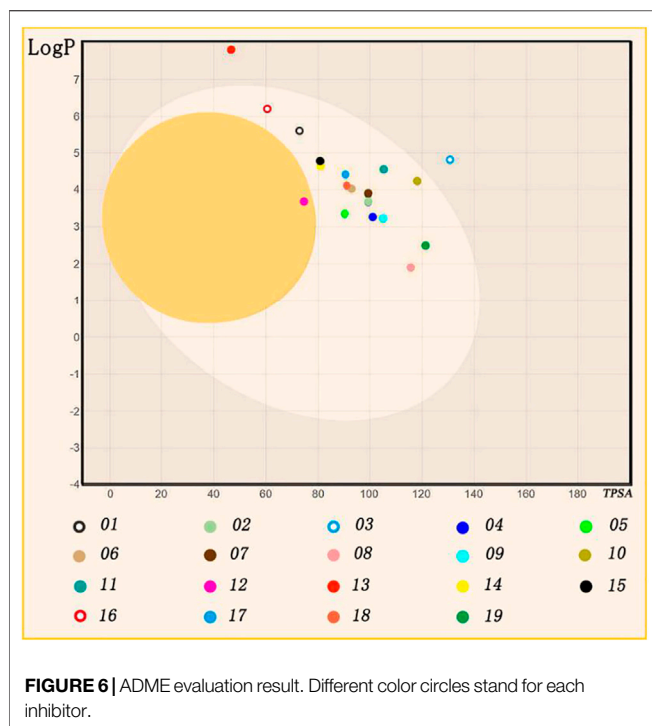
Receptor	Ligand No.	ΔE_{vdw}	ΔE_{ele}	ΔE_{polar}	$\Delta E_{non-polar}$	ΔG_{bind}	SD
Fascin	01(NP-G02-029)	-50.42	-20.92	35.29	-4.93	-40.97	0.28
Fascin	02	-43.71	-19.74	34.51	-4.98	-33.91	3.54
Fascin	03	-44.63	-28.99	51.13	-5.18	-27.66	3.07
Fascin	04	-47.44	-15.72	33.96	-5.04	-34.23	3.42
Fascin	05	-47.86	-11.60	27.43	-5.38	-37.41	2.51
Fascin	06	-45.80	-11.71	31.96	-5.03	-30.58	3.79
Fascin	07	-52.78	-10.03	27.01	-5.35	-41.14	2.53
Fascin	08	-42.05	-29.61	42.83	-4.60	-33.41	2.66
Fascin	09	-45.42	-16.57	36.71	-5.24	-30.52	2.62
Fascin	10	-40.84	-17.14	32.68	-4.80	-30.10	2.65
Fascin	11	-43.02	-19.21	33.22	-5.21	-34.22	3.07
Fascin	12	-38.20	-6.21	21.80	-4.28	-26.89	1.79
Fascin	13	-58.00	-15.99	32.75	-5.97	-47.20	3.94
Fascin	14	-47.53	-8.89	25.55	-5.01	-35.88	2.09
Fascin	15	-49.60	-18.33	32.55	-5.26	-40.64	0.69
Fascin	16	-46.01	-7.04	21.67	-5.06	-36.44	2.42
Fascin	17	-42.97	-32.28	45.93	-4.74	-34.06	2.59
Fascin	18	-56.03	-12.02	29.80	-5.55	-43.79	2.56
Fascin	19	-47.41	-9.60	31.15	-5.18	-31.04	2.86

MATERIALS AND METHODS

Data Preparation

In terms of the target protein, the crystal structure of Fascin was obtained from an online protein database (<https://www.rcsb.org>), PDB: 6B0T, 2.80 Å resolution (Huang et al., 2018).

The crystal structure was solved with its inhibitor NP-G2-029, which was set as an active controlled sample in our study. In addition, NP-G2-044, another effective inhibitor, was used as an active control (Han et al., 2016; Huang et al., 2018). On the other hand, inhibitors NP-G2-112 and NP-G2-113 were used as an inactive control since they do not have



any effect on Fascin (Han et al., 2016). Because no complex structures were solved for the NP-G2-044, -112, and -113, complex structures were prepared by molecular docking.

For the ligand database, Marine Natural Products Library (Marvin annotated) series (<http://docking.umh.es>) was used (Bugni et al., 2008; Encinar et al., 2015; Galiano et al., 2016). OMEGA was used for generating the conformations of all compounds (Hawkins et al., 2010).

Pharmacophore Model

Ligand-based pharmacophore modeling is one of the widely used methods in CADD (Leach et al., 2010). In this work, the pharmacophore model was built by LigandScout V_{4.4.5} (Salam, Nuti, and Sherman, 2009; Dixon et al., 2006; Maia et al., 2020). Directed hydrogen-bond interactions, hydrophobic interactions, charge interactions, and steric exclusions were detected directly. In this work, the HypoGen algorithm was used to produce the model, which contains three hydrophobic, one hydrogen-bond donor, and one hydrogen-bond acceptor (Figure 3) (Koes and Camacho, 2011). All features are added as 3D objects. It can be seen from Figure 3A that there are three hydrophobic models for this inhibitor, so the hydrophobic effect is the main pharmacophore feature. In addition, two hydrogen-bond interactions were formed between the inhibitor and Fascin, and inhibitor acted as hydrogen acceptor and donor, respectively.

Molecular Docking

Molecular docking is a structure based virtual screening method for drug discovery (Liu et al., 2020). It explores small ligand binding to biomacromolecule by searching the possible degrees of freedom of the whole system and finding the global energy

minimum. The binding sites of the ligand are evaluated by different score functions. It is widely used for lead screening and optimization (Saikia and Bordoloi, 2019). In this work, the AutoDock Vina module of LigandScout was used for docking (Roy, Srinivasan, and Skolnick, 2015; Nguyen et al., 2020). The scoring function of Vina includes a finite repulsion term, Gaussian steric interaction terms, Piecewise linear hydrophobic, hydrogen-bond interaction terms, etc. (Gaillard, 2018). All docking calculations were performed with default values in LigandScout.

Molecular Dynamics Simulation

In this paper, we performed molecular dynamic simulations on Fascin with 19 inhibitors. The coordinate of Fascin for all systems was taken from the 2.80 Å crystal structure of the Fascin-NP-G02-029 complex (PDB code: 6B0T) (Huang et al., 2018). Missing residues of structure (fragment 1–6) were modeled by using Chimera (Pettersen et al., 2004). TIP3P water models were used for solvating all systems (Jorgensen et al., 1983) in an octahedral box with a minimum distance of 12 Å from protein structures to box boundary (Gillan, Alfe, and Michaelides, 2016). Each His residue protonation state was identified by the pKa value from PROPKA (Olsson et al., 2011). All of the His residues were protonated at NE2 atoms, except His96, His108, His154, His198, His310 which are assumed to be doubly protonated, and His135 is protonated at ND1 atoms.

For all ligands, AM1-BCC atomic charges were calculated by the antechamber program (Wang and Kollman, 2001) (Jakalian et al., 2000; Jakalian, Jack, and Bayly, 2002). The general AMBER force field (GAFF) and Amber ff14sb force field were used for inhibitors and Fascin, respectively (Wang et al., 2004; Maier et al., 2015). In addition, an optimal amount of counterions was added to generate a neutral system.

The conventional MD simulation of each Fascin-inhibitor system was performed by using Amber 20 (Belfon et al., 2020). Langevin dynamics (Wu and Brooks, 2003) were performed at a constant temperature of 300 K. Collision frequency was set to 2.0 ps⁻¹. For NPT ensemble, pressure was kept at 1 atm (Berendsen et al., 1984). Particle mesh Ewald summation was used to handle the long-range electrostatics (Darden, York, and Pedersen, 1993).

For all simulations, we first ran a 5000-step minimization. Then, 20 ps NVT and 20 ps NPT pre-equilibration were carried out with restraints for heavy atoms of the protein. To further equilibrate the system, we ran a 1 ns NPT simulation without any restraints. Finally, 20 ns NPT production simulations were performed and coordinates were printed every 1 ps. For each inhibitor with Fascin, we performed five replicates of production calculations. For all systems, root mean square deviation (RMSD), root mean square fluctuation (RMSF), and radius of gyration (Rg) were calculated by using the cpptraj module in AMBER 20 (Belfon et al., 2020).

Ligand Gaussian Accelerated Molecular Dynamics

Conformation transition of protein usually happens in a millisecond time scale due to the high energy barrier between

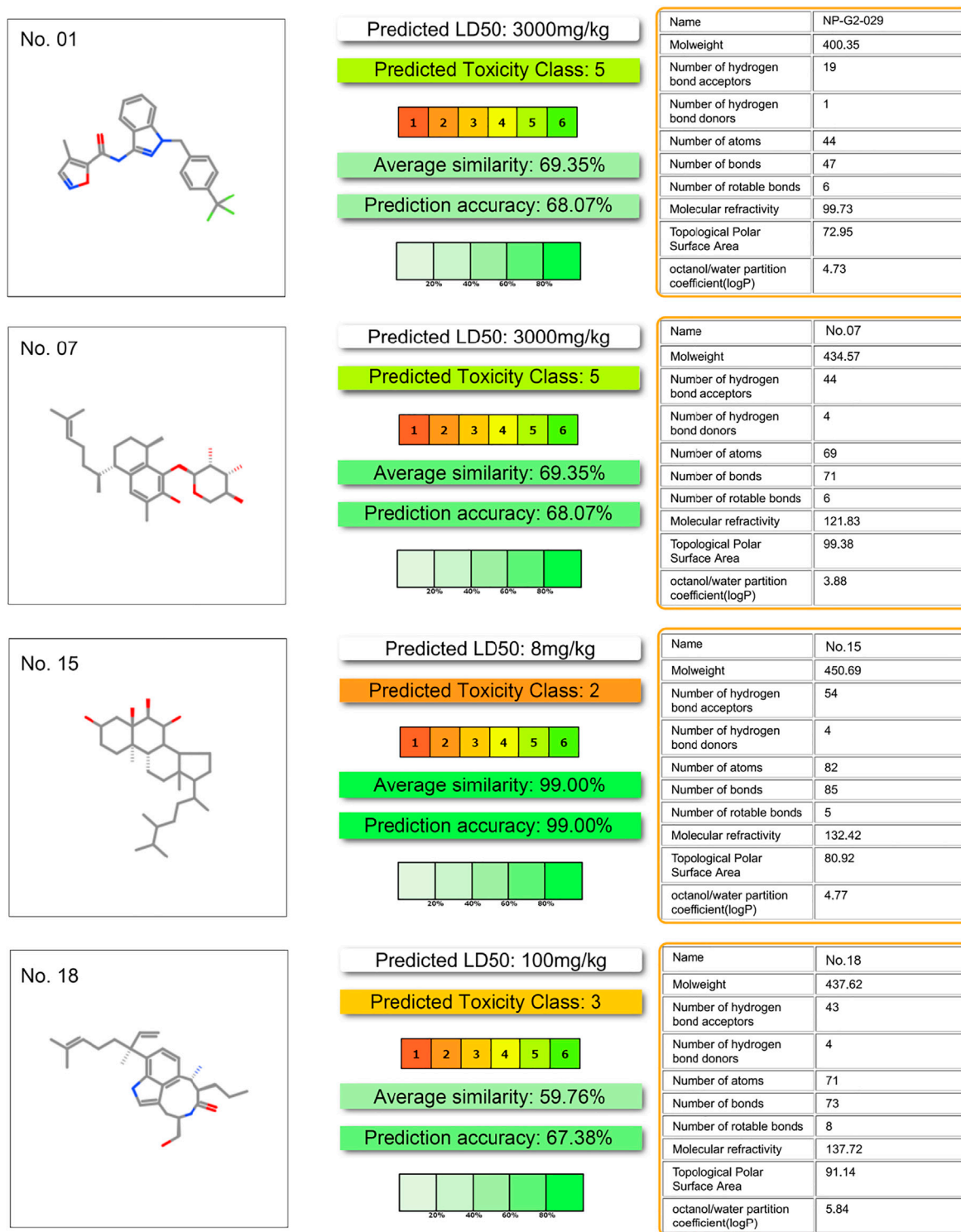


FIGURE 7 | Toxicity evaluation result. Evaluation processed by ProTox-II server (<https://tox-new.charite.de/>) (Banerjee et al., 2018).

different states. Thus, it is hard to capture the most stable state of protein from the whole potential surface. To investigate the conformational changes of Fascin with different inhibitors effectively, we used Li-GaMD (Miao, Feher, and McCammon,

2015; Miao, Bhattarai, and Wang, 2020) for simulation, which is developed from the enhanced sampling method GaMD (Miao, Feher, and McCammon, 2015), LiGaMD can accelerate simulations of the receptor with ligand between binding and

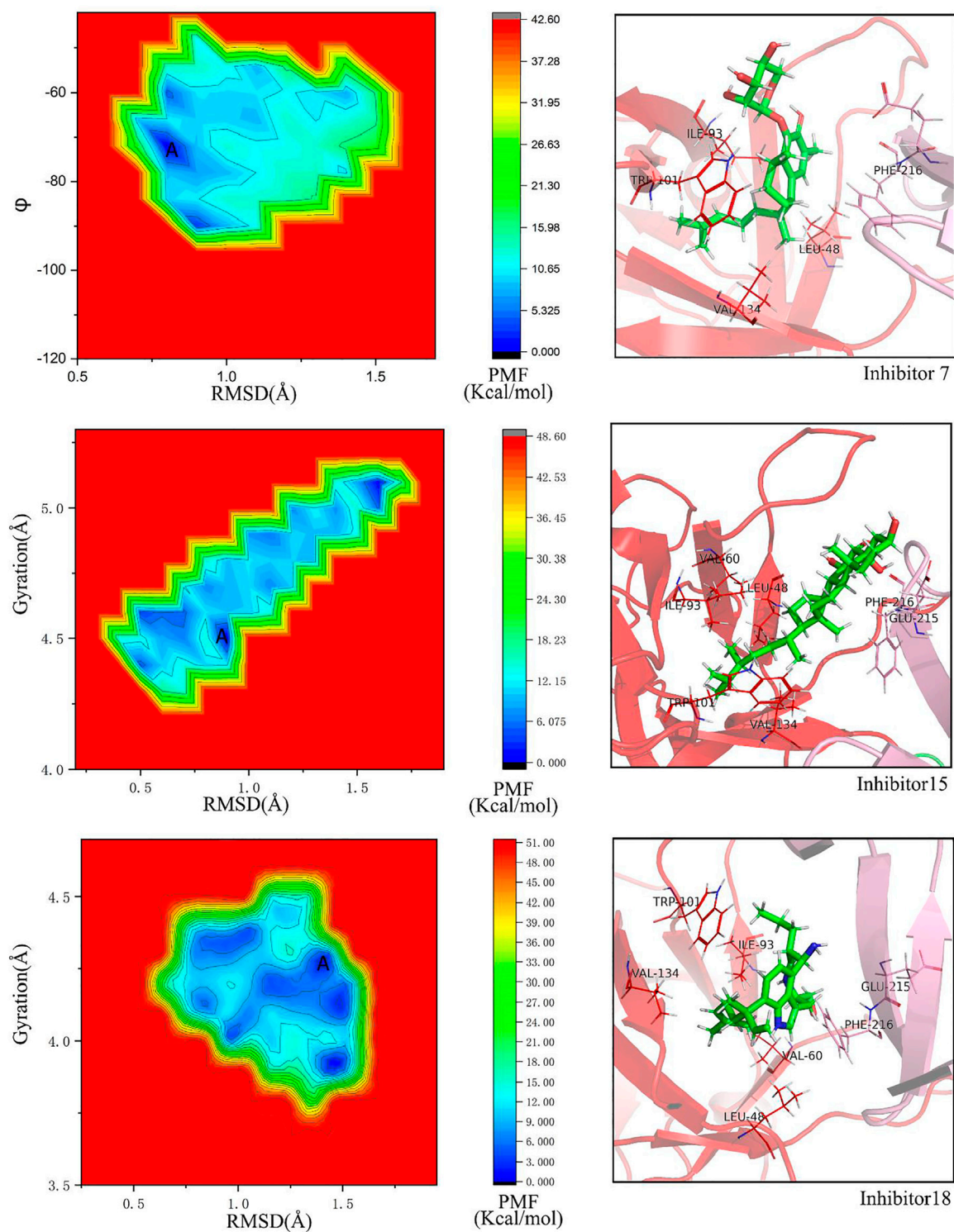


FIGURE 8 | Free energy landscapes and low-energy conformational states of Fascin with inhibitors, whose was modeled with the GAFF force field by using LiGaMD simulation: **(A)** PMF profile of inhibitor No. 07, collective variables (CVs) are backbone dihedrals (ϕ) and RMSD of inhibitor No.07. **(B)** PMF profiles of inhibitor No. 15, CVs are the radius of gyration and RMSD. **(C)** PMF profiles of inhibitor No. 18, CVs are the same with inhibitor No.15.

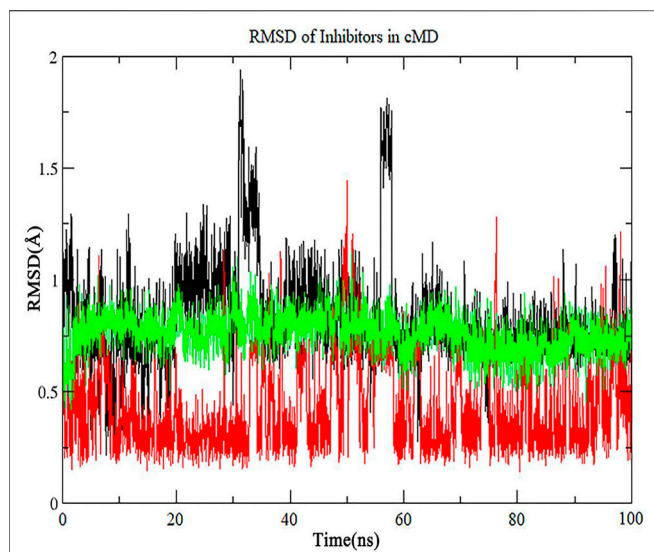


FIGURE 9 | The RMSD of inhibitor No. 07, 15, 18 in extended conventional MD simulation. The line with black, red, green color is inhibitor No. 07, 15, 18 respectively.

unbinding, explore protein conformational transition efficiently.

In a system comprising ligand L , protein P , and environment E , the system comprises N atoms with their coordinates $r \equiv \{\vec{r}_1, \dots, \vec{r}_N\}$ and momenta $p \equiv \{\vec{p}_1, \dots, \vec{p}_N\}$. The system Hamiltonian can be expressed as:

$$H(r, p) = K(p) + V(r) \quad (1)$$

where $K(p)$ and $V(r)$ are the systems kinetic and total potential energies, respectively. Then, the potential energy could be decomposed into the following terms:

$$\begin{aligned} V(r) = & V_{P,b}(r_P) + V_{L,b}(r_L) + V_{E,b}(r_E) \\ & + V_{PP,nb}(r_P) + V_{LL,nb}(r_{PL}) + V_{EE,nb}(r_{PL}) \\ & + V_{PL,nb}(r_{PL}) + V_{PE,nb}(r_{PE}) + V_{LE,nb}(r_{LE}) \end{aligned} \quad (2)$$

where $V_{P,b}$, $V_{L,b}$, and $V_{E,b}$ are the bonded potential energies in protein P , ligand L , and environment E , respectively. $V_{PP,nb}$, $V_{LL,nb}$, and $V_{EE,nb}$ are the nonbonded potential energies. $V_{PL,nb}$, $V_{PE,nb}$, and $V_{LE,nb}$ are the nonbonded interaction energies. Based on classical force fields (Duan et al., 2003; Vanommeslaeghe and MacKerell, 2015), the non-bonded potential energies are usually presented as:

$$V_{nb} = V_{elec} + V_{vdw} \quad (3)$$

Presumably, ligand binding mainly involves the nonbonded interaction energies of the ligand, $V_{L,nb}(r) = V_{LL,nb}(r_L) + V_{PL,nb}(r_{PL}) + V_{LE,nb}(r_{LE})$. LiGaMD adds a boost potential selectively to the ligand non-bonded potential energy according to the GaMD algorithm:

$$\Delta V_{L,nb}(r) = \begin{cases} \frac{1}{2} k_{L,nb} (E_{L,nb} - V_{L,nb}(r))^2, & V_{L,nb}(r) < E_{L,nb} \\ 0, & V_{L,nb}(r) \geq E_{L,nb} \end{cases} \quad (4)$$

where $E_{L,nb}$ is the threshold energy for applying boost potential and $k_{L,nb}$ is the harmonic constant. These parameters in LiGaMD are derived similarly as in the GaMD algorithm (Miao, Feher, and McCammon, 2015).

In addition to optional boosting non-bonded potential energy term of ligand, a second boost potential can be added on protein to explore protein conformational changes. The second boost potential is calculated using the total system potential energy as:

$$\Delta V_D(r) = \begin{cases} \frac{1}{2} k_D (E_D - V_D(r))^2, & V_D(r) < E_D \\ 0, & V_D(r) \geq E_D \end{cases} \quad (5)$$

where V_D is the total potential energy without the nonbonded potential energy of ligand, E_D is the threshold energy for applying the second boost potential and k_D is the harmonic constant. In this study, we applied dual-boost LiGaMD and total boost potential $\Delta V(r) = \Delta V_{L,nb}(r) + \Delta V_D(r) = \Delta V_{L,nb}(r) + \Delta V_D(r)$. For the analysis of the results, we used the PyReweighting program to calculate the free energy surface with different collective variables (Miao et al., 2014).

Binding Affinity Calculation With MM/GBSA

In order to calculate the binding free energies for different potential inhibitors, we used molecular mechanics MM/GBSA methods. It is an end-point based free energy calculation method,

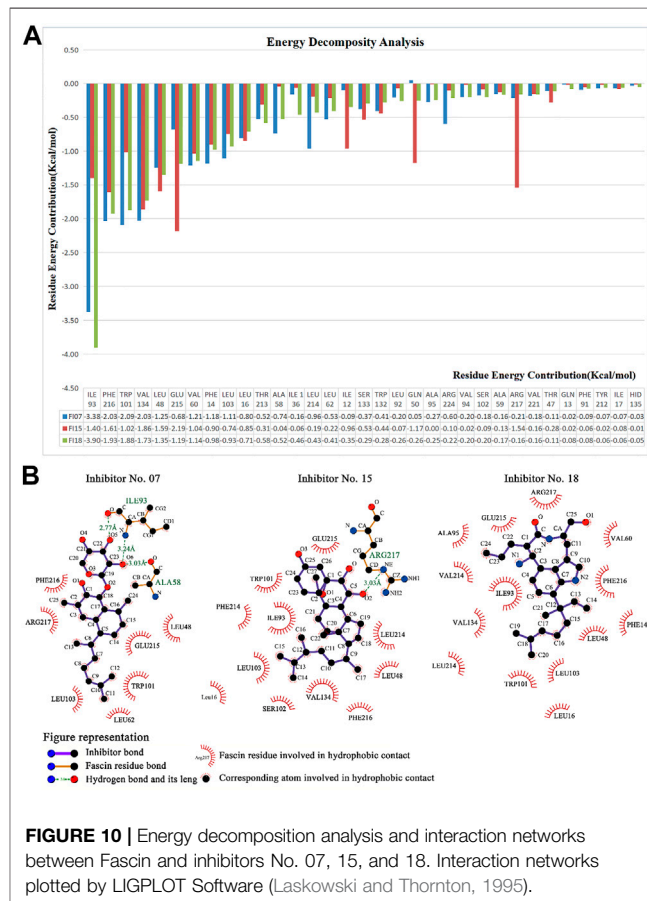


FIGURE 10 | Energy decomposition analysis and interaction networks between Fascin and inhibitors No. 07, 15, and 18. Interaction networks plotted by LIGPLOT Software (Laskowski and Thornton, 1995).

i.e., the binding free energy is calculated by

$$\Delta G_{\text{bind}} = G_{\text{RL}} - G_{\text{R}} - G_{\text{L}} \quad (6)$$

where G_{L} , G_{R} , and G_{L} represent the free energy of the complex and the receptor and ligand, respectively. Each free energy is calculated with

$$G = (E_{\text{bond}}) + (E_{\text{ele}}) + (E_{\text{vdW}}) + (E_{\text{pol}}) + (E_{\text{np}}) - TS \quad (7)$$

where E_{bond} is the energy of covalent interactions, E_{ele} is the electrostatic potential, E_{vdW} is the energy of van der Waals interactions, and G_{pol} and G_{np} are the polar and nonpolar contributions. The conformational entropy contribution ($-TS$) is estimated by normal-mode analysis (Srinivasan et al., 1998), but it is usually neglected from consideration due to its high computational cost and low prediction accuracy (Hou and Yu, 2007; Sun et al., 2018). In this work, MMPBSA.py (Miller et al., 2012) module in Amber20 (Belfon et al., 2020) was used to calculate the MM/GBSA for each system based on the last 2,500 frames extracted from the 20 ns conventional MD trajectory.

Pharmacokinetics Evaluation

ADMET evaluation is a comprehensive study of drug absorption, distribution, metabolism, excretion, and toxicity properties (Acuna, Hopper, and Yoder, 2020). Evaluation of ADMET properties at the early stage of drug development can significantly improve the success rate of drug discovery. It is used to efficiently and accurately calculate the physicochemical properties, toxicity information, and pharmacokinetic properties of candidate drug molecules, provide the basis for prediction and improve the interpretability of structure and drug credibility (Wenzel, Matter, and Schmidt, 2019). The small molecule hits were predicted by ADMET based on the Swissadme server (<http://www.swissadme.ch/>) (Daina, Michielin, and Zoete, 2017), and LogP and TPSA were pointed out as the main reference indexes of the results.

Log P refers to the equilibrium distribution of the undissociated molecules in the oil and water phases, which is an important indicator in the passage of compounds through biofilm. TPSA refers to the topological polar surface area. TPSA <60 indicates that it has good membrane permeability and is completely absorbed. 60 < TPSA <140 indicates the molecular permeability decreases with the increase of polar surface area. TPSA >140 indicates poor permeability of the molecule. Lipid solubility is an important parameter of small molecules in pharmaceutical chemistry. Log P is the logarithm of the oil-water partition coefficient P of the compound, which refers to the equilibrium of the distribution of the undissociated molecules in the oil phase and water phase. When oral drugs are permeated by passive diffusion, logP in the 0–3 range is the best. High logP compounds have poor water solubility and low logP compounds have poor lipid permeability.

RESULTS AND DISCUSSION

Pharmacophore Model

In this work, the pharmacophore model with three hydrophobic, one hydrogen-bond donor, and one hydrogen-bond acceptor

(Figure 3) was used for virtual screening. First, the MNPs database (14,064 compounds) was processed by Openbabel V_{2.4.1} for 3D structure generation, hydrogenation, and charge processing operation (O'Boyle et al., 2011), and 12,015 compounds were generated. Then, the pharmacophore virtual screening was performed on the 12,015 compounds (Figure 3). In total, 472 compounds with high fitness were found. Finally, compounds that have a molecular weight larger than 500 were removed with the *Filter* module, which provided 281 results for further study.

Molecular Docking

In the crystal structure, inhibitor NP-G2-029 resides in the pocket formed by the residues located in the junction of domains 1 and 2 (Figure 4A). The surface volume of the active site inherent is 1130 Å³ calculated by the Proteins Plus server (Schoning-Stierand et al., 2020). To rank the 281 ligands from the screening based on the Pharmacophore model, molecular docking was performed with the AutoDock Vina module of the Ligandscout program (Wolber and Langer, 2005). The top 18 ligands with binding energy ≤ −9 kcal/mol were selected as the potential inhibitors for further calculations. The molecular structures and molecular binding energy are shown in Table 1 (For convenience, we have also provided the ZINC ID for those MNPs).

As is shown in Figure 4B, all the 18 small molecules that are embedded in the binding pocket are approximately as same as NP-G2-029. The residues of Fascin involved in the binding pocket are mainly Leu48, Glu49, Gln50, Ile93, Trp101, Val 103, Glu215, Phe216, and Arg217. All the detailed interactions between ligands and proteins are shown in Supplementary Figure S1.

Conventional Molecular Dynamics for Fascin-Inhibitor Complex

To find a better inhibitor than NP-G2-029, conventional MD dynamics were carried for the Fascin with 18 inhibitors from AutoDock Vina. For each system, we run 20 ns × 5 replicates. RMSDs for all systems indicate that all simulations are converged (Supplementary Figure S2). It can be seen from RMSF data (Figure 5) that the binding sites of inhibitor in Fascin-inhibitor complexes are more dynamic with a range of 3–10 Å in β-trefoil 1, whereas other regions are relatively rigid, compared to Fascin without inhibitor (Blue line in Figure 5). These findings are consistent with a study by Huang et al. (2018). Overall, the inhibitors affect the RMSF of Fascin significantly.

Binding Free Energy by MM/GBSA

To improve the accuracy of ranking in molecular docking, we calculated the binding affinities of inhibitors in each complex using MM/GBSA with conventional MD trajectories. Binding free energy results were obtained based on the five replicate (20 ns × 5 replicas) simulations (Table 2). For inhibitor NP-G2-029 with Fascin, the calculated binding affinity is −41 kcal/mol, indicating the two objectives are favorable for binding, which is consistent

with experimental data that NP-G2-029 inhibits Fascin (Huang et al., 2018). For the other active inhibitor NP-G2-044, the binding affinity is -42 kcal/mol (Supplementary Table S2). For the inactive inhibitors NP-G2-112 and NP-G2-113, they are -38 and -35 kcal/mol, respectively. Thus, -41 kcal/mol was used as a threshold value, i.e., ligands with binding affinity larger than -40 kcal/mol are thought of as potential inhibitors. It can be found from Table 2, No. 07, 13, 15, 18 inhibitors have qualified binding affinities to Fascin with binding free energies of -41 , -47 , -41 , -44 kcal/mol, respectively.

Pharmacokinetics Evaluation

Pharmacokinetics prediction was performed for the 19 compounds (including NP-G2-029) on the ADMETlab server (Dong et al., 2018), which is based on a comprehensive database that includes 288,967 entries (Ferreira and Andricopulo, 2019). There are four function modules for drug-likeness analysis, ADME endpoint prediction, systematic evaluation, and similarity searching, these results give an overall understanding of compounds and can check the rapid screening process.

For NP-G2-029 and the 18 small molecules, we perform ADMET assessments, which include Lipid solubility (Dong et al., 2018; Ferreira and Andricopulo, 2019). Lipid solubility is an important parameter for small molecules in pharmaceutical chemistry (Williams et al., 2013). When oral drugs are permeated by passive diffusion, the $\log P$ 0–3 range is the best. High $\log P$ compounds have poor water solubility, low $\log P$ compounds have poor lipid permeability. TPSA <60 denotes good membrane permeability and is completely absorbed. $60 < \text{TPSA} < 140$ denotes that the molecular permeability decreases with the increase of polar surface area. TPSA >140 denotes the poor permeability of the molecule.

ADME results in Figure 6 show that TPSA of No. 02, 04, 05, 06, 07, 09, 10, 11, 12, 14, 15, 17, 18, 19 are in range of 60–120, and whose $\log P$ are in a range of 3–5. Notable, No.13 inhibitor is out of 60–120 TPSA and 3–5 $\log P$, therefore, we exclude No.13 inhibitor for the next assessment.

Toxicity predictions are performed on the potential inhibitor No. 07, 15, 18, and the NP-G2-029. Data in Figure 7 shows that the toxicity score of NP-G2-029 is 5, the toxicity scores of inhibitor No. 07, 15, 18 are 5, 2, 3 respectively, whose are lower than NP-G2-029, signify that the potentials inhibitor are less toxic than NP-G2-029.

Ligand GAMD Simulation

To confirm whether the docking pose of compounds 07, 15, and 18 are stable in the pocket of Fascin, we performed LiGAMD simulations. The boost potential added in LiGAMD simulations is according to Gaussian distribution, accurate reweighting and recovery of the original biomolecular free energy landscapes can be achieved by using cumulant expansion to the second order.

For No. 07 inhibitor (Figure 8A), 2D PMF with backbone dihedrals (φ) and RMSD is calculated by reweighting 100 ns LiGAMD simulations. One low-energy state (labeled as A) can be found from the potential surface. The binding pocket in this state includes Ile93, Trp101, Val134, Phe216, Leu48, Val60, Phe14,

Leu103, Leu16, and they contribute the binding free energy much according to the energy decomposition in MM/GBSA (Figure 10).

For No. 15 inhibitor (Figure 8B) 2D PMF with the radius of gyration of protein and RMSD. The state with the lowest energy is shown in Figure 8B. For this inhibitor, the binding pocket is slightly modulated due to ligand binding. The pocket consists of Glu215, Val134, Phe216, Arg217, Leu48, Ile93, Val60, Trp101, Phe14, and Leu16.

For the No.18 inhibitor (Figure 8C), the 2D PMF was plotted with the same collective variables as inhibitor No. 15. Again, the binding pocket modulates slightly, including Ile93, Phe216, Trp101, Val134, Leu48, Glu215, Val60, Phe14, and Leu103.

In addition, for binding poses of the ligands, AutoDock Vina gives almost the same pose as LiGAMD simulations in this study.

Extended Conventional MD From Low-Energy States

To confirm that whether the three inhibitors reside in the binding pocket of Fascin in LiGAMD, we ran 100 ns conventional MD simulations that start from the structures at A position in Figure 8. As is shown in Figure 9, RMSD values for the three potential inhibitors are mainly lower than 1 Å, indicating that all of them stay at the binding position. On the other hand, our results indicate that the docking structures can be trusted for this system.

Finally, we run the binding affinity and binding energy decomposition analysis for the three systems (Figure 10A). The binding affinities of inhibitor No. 07, 15, 18 are -42 , -45 , and -41 kcal/mol, respectively. Our results indicate the three compounds can bind to Fascin as well as NP-G2-029. The crucial residues contributed to the binding affinity of the inhibitor and are mainly in the binding pocket, as shown in Figure 10B. It can be seen from the interaction network that the hydrophobic interactions have a large contribution for binding, i.e., 7 hydrophobic interactions for ligand No. 07, 11 hydrophobic interactions for ligand No. 15, and 24 hydrophobic interactions for ligand No. 18. In addition, 2 hydrogen bonds are formed for ligand No. 07 and one hydrogen bond is formed for ligand No. 15.

CONCLUSION

Fascin is overexpressed in many cancers, e.g., esophageal cancer. In this paper, we performed CADD methods to predict the potential inhibitors for Fascin from a library of marine natural products including 14,064 compounds, viz. pharmacophore model, molecular docking, molecular dynamics, MM/GBSA, and predictions of absorption, distribution, metabolism, excretion and toxicity properties (AMDET).

First, we built the pharmacophore model for the inhibitor NP-G2-029, which was confirmed experimentally (Huang et al., 2018). With the pharmacophore model, we achieved 472 results. In addition, compounds that have a molecular weight larger than 500 were kicked out, which gives 281 hits for further study. Next, molecular docking was carried out to rank all the 281 hits. The top 18 inhibitors with binding affinity larger than 9 kcal/mol were selected for further study.

To accurately assess the binding affinity, MM/GBSA calculations are performed for the 19 compounds (including NP-G02-029). Four compounds (No. 07, 13, 15, and 18) were found to have larger affinities to Fascin than NP-G02-029 and were deemed potential inhibitors.

To predict the AMDET, we used the web server ADMETLab server (Dong et al., 2018) and ProTox-II server (Banerjee et al., 2018). AMDET results indicate that compound No.13 does not satisfy the criteria. Thus, three compounds (No. 07, 15, and 18) potentially inhibit the function of Fascin.

Finally, we ran LiGaMD and other conventional MD simulations to confirm whether the three potential inhibitors reside in the binding site or not. Our results demonstrate that all of them stay at the binding site stably.

Thus, we predict three potential inhibitors for Fascin from marine natural products in this investigation. These inhibitors could have higher binding affinities than the one (NP-G02-029) found in the previous study (Huang et al., 2018), and block the function of Fascin. All the computational methods used in this study could accelerate drug discovery dramatically.

DATA AVAILABILITY STATEMENT

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

REFERENCES

- Acuna, V. V., Hopper, R. M., and Yoder, R. J. (2020). Computer-Aided Drug Design for the Organic Chemistry Laboratory Using Accessible Molecular Modeling Tools. *J. Chem. Educ.* 97, 760–763. doi:10.1021/acs.jchemed.9b00592
- Banerjee, P., Eckert, A. O., Schrey, A. K., and Preissner, R. (2018). ProTox-II: a webserver for the prediction of toxicity of chemicals. *Nucleic Acids Res.* 46, W257–W263. doi:10.1093/nar/gky318
- Barbosa, A. J. M., and Roque, A. C. A. (2019). Free Marine Natural Products Databases for Biotechnology and Bioengineering. *Biotechnol. J.* 14, e1800607. doi:10.1002/biot.201800607
- Belfon, D. A., K., Ben-Shalom, I. Y., Brozell, S. R., Cerutti, D. S., Cheatham, T. E., III, Darden, T. A., et al. (2020). *AMBER 2020*. San Francisco: University of California.
- Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., DiNola, A., and Haak, J. R. (1984). Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* 81, 3684–3690. doi:10.1063/1.448118
- Bernardi, R. C., Melo, M. C. R., and Schulten, K. (2015). Enhanced sampling techniques in molecular dynamics simulations of biological systems. *Biochim. Biophys. Acta (Bba) - Gen. Subjects* 1850, 872–877. doi:10.1016/j.bbagen.2014.10.019
- Bilal, M., Rasheed, T., Sosa-Hernández, J. E., Raza, A., Nabeel, F., and Iqbal, H. M. N. (2018). Biosorption: An Interplay between Marine Algae and Potentially Toxic Elements-A Review. *Mar. Drugs*, 16. doi:10.3390/md16020065
- Blunt, J. W., Carroll, A. R., Copp, B. R., Davis, R. A., Keyzers, R. A., and Prinsep, M. R. (2018). Marine natural products. *Nat. Prod. Rep.* 35, 8–53. doi:10.1039/c7np00052a
- Blunt, J. W., Copp, B. R., Keyzers, R. A., Munro, M. H. G., and Prinsep, M. R. (2017). Marine natural products. *Nat. Prod. Rep.* 34, 235–294. doi:10.1039/c6np00124f
- Buckley, C. L., Kim, C. S., McGregor, S., and Seth, A. K. (2017). The free energy principle for action and perception: A mathematical review. *J. Math. Psychol.* 81, 55–79. doi:10.1016/j.jmp.2017.09.004

AUTHOR CONTRIBUTIONS

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

FUNDING

This investigation has been supported by grants from the National Natural Science Foundation of China (81872372, 21907063, and 81902469), the Li Ka Shing Foundation (project LD0101), and the 2020 Li Ka Shing Foundation Cross-Disciplinary Research Grant (2020LKSFG07B).

ACKNOWLEDGMENTS

We acknowledge the Big Data Platform of Shantou University Medical College for providing resources for computations and storage.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fchem.2021.719949/full#supplementary-material>

- Bugni, T. S., Richards, B., Bhoite, L., Cimbor, D., Harper, M. K., and Ireland, C. M. (2008). Marine natural product libraries for high-throughput screening and rapid drug discovery. *J. Nat. Prod.* 71, 1095–1098. doi:10.1021/np800184g
- Carroll, A. R., Copp, B. R., Davis, R. A., Keyzers, R. A., and Prinsep, M. R. (2021). Marine natural products. *Nat. Prod. Rep.* 38, 362–413. doi:10.1039/d0np00089b
- Chen, L., Yang, S., Jakoncic, J., Zhang, J. J., and Huang, X.-Y. (2010). Migrastatin analogues target fascin to block tumour metastasis. *Nature* 464, 1062–1066. doi:10.1038/nature08978
- Conesa-Zamora, P., García-Solano, J., García-García, F., Turpin, M. d. C., Trujillo-Santos, J., Torres-Moreno, D., et al. (2013). Expression profiling shows differential molecular pathways and provides potential new diagnostic biomarkers for colorectal serrated adenocarcinoma. *Int. J. Cancer* 132, 297–307. doi:10.1002/ijc.27674
- 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. doi:10.1038/srep42717
- 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
- De Vivo, M., Masetti, M., Bottegoni, G., and Cavalli, A. (2016). Role of Molecular Dynamics and Related Methods in Drug Discovery. *J. Med. Chem.* 59, 4035–4061. doi:10.1021/acs.jmedchem.5b01684
- Dixon, S. L., Smondyrev, A. M., Knoll, E. H., Rao, S. N., Shaw, D. E., and Friesner, R. A. (2006). PHASE: a new engine for pharmacophore perception, 3D QSAR model development, and 3D database screening: 1. Methodology and preliminary results. *J. Comput. Aided Mol. Des.* 20, 647–671. doi:10.1007/s10822-006-9087-6
- Dong, J., Wang, N.-N., Yao, Z.-J., Zhang, L., Cheng, Y., Ouyang, D., et al. (2018). ADMETLab: a platform for systematic ADMET evaluation based on a comprehensively collected ADMET database. *J. Cheminform.* 10, 29. doi:10.1186/s13321-018-0283-x
- Duan, Y., Wu, C., Chowdhury, S., Lee, M. C., Xiong, G., Zhang, W., et al. (2003). A point-charge force field for molecular mechanics simulations of proteins based

- on condensed-phase quantum mechanical calculations. *J. Comput. Chem.* 24, 1999–2012. doi:10.1002/jcc.10349
- Duh, F.-M., Latif, F., Weng, Y., Geil, L., Modi, W., Stackhouse, T., et al. (1994). cDNA cloning and expression of the human homolog of the sea urchin fascin and *Drosophila* singed genes which encodes an actin-bundling protein. *DNA Cel Biol.* 13, 821–827. doi:10.1089/dna.1994.13.821
- Edwards, R. A., Herrera-Sosa, H., Otto, J., and Bryan, J. (1995). Cloning and expression of a murine fascin homolog from mouse brain. *J. Biol. Chem.* 270, 10764–10770. doi:10.1074/jbc.270.18.10764
- Encinar, J. A., Fernández-Ballester, G. J., Galiano-Ibarra, V., and Micol-Molina, V. (2015). In silico approach for the discovery of new PPAR γ modulators among plant-derived polyphenols. *Dddt* 9, 5877–5895. doi:10.2147/dddt.s93449
- Ferreira, L. L. G., and Andricopulo, A. D. (2019). ADMET modeling approaches in drug discovery. *Drug Discov. Today* 24, 1157–1165. doi:10.1016/j.drudis.2019.03.015
- Francis, S., Croft, D., Schüttelkopf, A. W., Parry, C., Pugliese, A., Cameron, K., et al. (2019). Structure-based design, synthesis and biological evaluation of a novel series of isoquinoline and pyrazolo[4,3-c]pyridine inhibitors of fascin 1 as potential anti-metastatic agents. *Bioorg. Med. Chem. Lett.* 29, 1023–1029. doi:10.1016/j.bmcl.2019.01.035
- Fu, Y., Ye, T., Liu, Y. X., Wang, J., and Ye, F. (2020). Based on the Virtual Screening of Multiple Pharmacophores, Docking and Molecular Dynamics Simulation Approaches toward the Discovery of Novel HPPD Inhibitors. *Int. J. Mol. Sci.* 21, 21. doi:10.3390/ijms21155546
- Gaillard, T. (2018). Evaluation of AutoDock and AutoDock Vina on the CASF-2013 Benchmark. *J. Chem. Inf. Model.* 58, 1697–1706. doi:10.1021/acs.jcim.8b00312
- Galiano, V., Garcia-Valtanen, P., Micol, V., and Encinar, J. A. (2016). Looking for inhibitors of the dengue virus NS5 RNA-dependent RNA-polymerase using a molecular docking approach. *Dddt* Vol. 10, 3163–3181. doi:10.2147/dddt.s117369
- García-Sosa, A. T., and Maran, U. (2021). Combined Naïve Bayesian, Chemical Fingerprints and Molecular Docking Classifiers to Model and Predict Androgen Receptor Binding Data for Environmentally- and Health-Sensitive Substances. *Int. J. Mol. Sci.* 22, 22. doi:10.3390/ijms22136695
- Gillan, M. J., Alfè, D., and Michaelides, A. (2016). Perspective: How good is DFT for water. *J. Chem. Phys.* 144, 130901. doi:10.1063/1.4944633
- Greco, G. R., and Cinquegrani, M. (2016). Firms Plunge into the Sea. Marine Biotechnology Industry, a First Investigation. *Front. Mar. Sci.*, 2. doi:10.3389/fmars.2015.00124
- Gupta, C. L., Babu Khan, M., Ampasala, D. R., Akhtar, S., Dwivedi, U. N., and Bajpai, P. (2019). Pharmacophore-based virtual screening approach for identification of potent natural modulatory compounds of human Toll-like receptor 7. *J. Biomol. Struct. Dyn.* 37, 4721–4736. doi:10.1080/07391102.2018.1559098
- Han, S., Huang, J., Liu, B., Xing, B., Bordeleau, F., Reinhart-King, C. A., et al. (2016). Improving fascin inhibitors to block tumor cell migration and metastasis. *Mol. Oncol.* 10, 966–980. doi:10.1016/j.molonc.2016.03.006
- Haroun, R., Gil-Rodríguez, M. C., Neto, A. I., Machín-Sánchez, M., and Viera-Rodríguez, M. A. (2019). A review of current uses and potential biotechnological applications of seaweeds from the Macaronesian region (Central-East Atlantic Ocean). *J. Appl. Phycol.* 31, 3777–3790. doi:10.1007/s10811-019-01889-4
- Hawkins, P. C. D., Skillman, A. G., Warren, G. L., Ellingson, B. A., and Stahl, M. T. (2010). Conformer generation with OMEGA: algorithm and validation using high quality structures from the Protein Databank and Cambridge Structural Database. *J. Chem. Inf. Model.* 50, 572–584. doi:10.1021/ci100031x
- Holthuis, J. C. M., Schoonderwoert, V. T. G., and Martens, G. J. M. (1994). A vertebrate homolog of the actin-bundling protein fascin. *Biochim. Biophys. Acta (Bba) - Gene Struct. Expr.* 1219, 184–188. doi:10.1016/0167-4781(94)90267-4
- Hou, T., and Yu, R. (2007). Molecular dynamics and free energy studies on the wild-type and double mutant HIV-1 protease complexed with amprenavir and two amprenavir-related inhibitors: mechanism for binding and drug resistance. *J. Med. Chem.* 50, 1177–1188. doi:10.1021/jm0609162
- Hu, G.-P., Yuan, J., Sun, L., She, Z.-G., Wu, J.-H., Lan, X.-J., et al. (2011). Statistical research on marine natural products based on data obtained between 1985 and 2008. *Mar. Drugs* 9, 514–525. doi:10.3390/md9040514
- Hu, Y., Chen, J., Hu, G., Yu, J., Zhu, X., Lin, Y., et al. (2015). Statistical research on the bioactivity of new marine natural products discovered during the 28 years from 1985 to 2012. *Mar. Drugs* 13, 202–221. doi:10.3390/md13010202
- Huang, F.-K., Han, S., Xing, B., Huang, J., Liu, B., Bordeleau, F., et al. (2015). Targeted inhibition of fascin function blocks tumour invasion and metastatic colonization. *Nat. Commun.* 6, 7465. doi:10.1038/ncomms8465
- Huang, J., Dey, R., Wang, Y., Jakoncic, J., Kurinov, I., and Huang, X.-Y. (2018). Structural Insights into the Induced-fit Inhibition of Fascin by a Small-Molecule Inhibitor. *J. Mol. Biol.* 430, 1324–1335. doi:10.1016/j.jmb.2018.03.009
- Huang, S.-Y., and Zou, X. (2010). Advances and challenges in protein-ligand docking. *Ijms* 11, 3016–3034. doi:10.3390/ijms11083016
- 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
- Jakalian, A., Jack, D. B., and Bayly, C. I. (2002). Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation. *J. Comput. Chem.* 23, 1623–1641. doi:10.1002/jcc.10128
- 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
- Kane, R. E. (1975). Preparation and purification of polymerized actin from sea urchin egg extracts. *J. Cel Biol.* 66, 305–315. doi:10.1083/jcb.66.2.305
- Kaykci, M., Venkatakrishnan, A. J., Scott-Brown, J., Ravarani, C. N. J., Flock, T., and Babu, M. M. (2018). Visualization and analysis of non-covalent contacts using the Protein Contacts Atlas. *Nat. Struct. Mol. Biol.* 25, 185–194. doi:10.1038/s41594-017-0019-z
- Koes, D. R., and Camacho, C. J. (2011). Pharmer: efficient and exact pharmacophore search. *J. Chem. Inf. Model.* 51, 1307–1314. doi:10.1021/ci200097m
- 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
- Leach, A. R., Gillet, V. J., Lewis, R. A., and Taylor, R. (2010). Three-dimensional pharmacophore methods in drug discovery. *J. Med. Chem.* 53, 539–558. doi:10.1021/jm900817u
- Liu, Y.-X., Gao, S., Ye, T., Li, J.-Z., Ye, F., and Fu, Y. (2020). Combined 3D-quantitative structure-activity relationships and topomer technology-based molecular design of human 4-hydroxyphenylpyruvate dioxygenase inhibitors. *Future Med. Chem.* 12, 795–811. doi:10.4155/fmc-2019-0349
- Liu, Y.-X., Zhao, L.-X., Ye, T., Gao, S., Li, J.-Z., Ye, F., et al. (2020). Identification of key residues determining the binding specificity of human 4-hydroxyphenylpyruvate dioxygenase. *Eur. J. Pharm. Sci.* 154, 105504. doi:10.1016/j.ejps.2020.105504
- Lopez-Vallejo, F., Caulfield, T., Martinez-Mayorga, K., A. Giulianotti, M., Nefzi, A., A. Houghten, R., et al. (2011). Integrating virtual screening and combinatorial chemistry for accelerated drug discovery. *Cchts* 14, 475–487. doi:10.2174/138620711795767866
- Macalino, S. J. Y., Gosu, V., Hong, S., and Choi, S. (2015). Role of computer-aided drug design in modern drug discovery. *Arch. Pharm. Res.* 38, 1686–1701. doi:10.1007/s12272-015-0640-5
- Machesky, L. M., and Li, A. (2010). Fascin. *Communicative Integr. Biol.* 3, 263–270. doi:10.4161/cib.3.3.11556
- Maia, E. H. B., Medaglia, L. R., da Silva, A. M., and Taranto, A. G. (2020). Molecular Architect: A User-Friendly Workflow for Virtual Screening. *ACS Omega* 5, 6628–6640. doi:10.1021/acsomega.9b04403
- 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. Theor. Comput.* 11, 3696–3713. doi:10.1021/acs.jctc.5b00255
- Miao, Y., Bhattarai, A., and Wang, J. (2020). Ligand Gaussian Accelerated Molecular Dynamics (LiGaMD): Characterization of Ligand Binding Thermodynamics and Kinetics. *J. Chem. Theor. Comput.* 16, 5526–5547. doi:10.1021/acs.jctc.0c00395
- Miao, Y., Feher, V. A., and McCammon, J. A. (2015). Gaussian Accelerated Molecular Dynamics: Unconstrained Enhanced Sampling and Free Energy Calculation. *J. Chem. Theor. Comput.* 11, 3584–3595. doi:10.1021/acs.jctc.5b00436

- Miao, Y., Sinko, W., Pierce, L., Bucher, D., Walker, R. C., and McCammon, J. A. (2014). Improved Reweighting of Accelerated Molecular Dynamics Simulations for Free Energy Calculation. *J. Chem. Theor. Comput.* 10, 2677–2689. doi:10.1021/ct500090q
- 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. Theor. Comput.* 8, 3314–3321. doi:10.1021/ct300418h
- Montaser, R., and Luesch, H. (2011). Marine natural products: a new wave of drugs. *Future Med. Chem.* 3, 1475–1489. doi:10.4155/fmc.11.118
- Murzin, A. G., Lesk, A. M., and Chothia, C. (1992). β -Trefoil fold. *J. Mol. Biol.* 223, 531–543. doi:10.1016/0022-2836(92)90668-a
- Nguyen, N. T., Nguyen, T. H., Pham, T. N. H., Huy, N. T., Bay, M. V., Pham, M. Q., et al. (2020). Autodock Vina Adopts More Accurate Binding Poses but Autodock4 Forms Better Binding Affinity. *J. Chem. Inf. Model.* 60, 204–211. doi:10.1021/acs.jcim.9b00778
- 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
- Olsson, M. H. M., Søndergaard, C. R., Rostkowski, M., and Jensen, J. H. (2011). PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pKa Predictions. *J. Chem. Theor. Comput.* 7, 525–537. doi:10.1021/ct100578z
- Ono, S., Yamakita, Y., Yamashiro, S., Matsudaira, P. T., Gnarr, J. R., Obinata, T., et al. (1997). Identification of an actin binding region and a protein kinase C phosphorylation site on human fascin. *J. Biol. Chem.* 272, 2527–2533. doi:10.1074/jbc.272.4.2527
- 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
- Ponting, C. P., and Russell, R. B. (2000). Identification of distant homologues of fibroblast growth factors suggests a common ancestor for all β -trefoil proteins 1 Edited by J. Thornton. *J. Mol. Biol.* 302, 1041–1047. doi:10.1006/jmbi.2000.4087
- Rastelli, G., and Pinzi, L. (2019). Refinement and Rescoring of Virtual Screening Results. *Front. Chem.* 7, 498. doi:10.3389/fchem.2019.00498
- Rigsby, R. E., and Parker, A. B. (2016). Using the PyMOL application to reinforce visual understanding of protein structure. *Biochem. Mol. Biol. Educ.* 44, 433–437. doi:10.1002/bmb.20966
- Rodrigues, P. C., Sawazaki-Calone, I., Ervolino de Oliveira, C., Soares Macedo, C. C., Dourado, M. R., Cervigne, N. K., et al. (2017). Fascin promotes migration and invasion and is a prognostic marker for oral squamous cell carcinoma. *Oncotarget* 8, 74736–74754. doi:10.18632/oncotarget.20360
- Roy, A., Srinivasan, B., and Skolnick, J. (2015). PoLi: A Virtual Screening Pipeline Based on Template Pocket and Ligand Similarity. *J. Chem. Inf. Model.* 55, 1757–1770. doi:10.1021/acs.jcim.5b00232
- Ruiz, J., Olivieri, G., de Vree, J., Bosma, R., Willems, P., Reith, J. H., et al. (2016). Towards industrial products from microalgae. *Energy Environ. Sci.* 9, 3036–3043. doi:10.1039/c6ee01493c
- Saikia, S., and Bordoloi, M. (2019). Molecular Docking: Challenges, Advances and its Use in Drug Discovery Perspective. *Cdt* 20, 501–521. doi:10.2174/1389450119666181022153016
- Salam, N. K., Nuti, R., and Sherman, W. (2009). Novel method for generating structure-based pharmacophores using energetic analysis. *J. Chem. Inf. Model.* 49, 2356–2368. doi:10.1021/ci900212v
- Schöning-Stierand, K., Diedrich, K., Fährrolfs, R., Flachsenberg, F., Meyder, A., Nittinger, E., et al. (2020). ProteinsPlus: interactive analysis of protein-ligand binding interfaces. *Nucleic Acids Res.* 48, W48–W53. doi:10.1093/nar/gkaa235
- Srinivasan, J., Cheatham, T. E., Cieplak, P., Kollman, P. A., and Case, D. A. (1998). Continuum Solvent Studies of the Stability of DNA, RNA, and Phosphoramidate–DNA Helices. *J. Am. Chem. Soc.* 120, 37. doi:10.1021/ja981844+
- Sun, H., Duan, L., Chen, F., Liu, H., Wang, Z., Pan, P., et al. (2018). Assessing the performance of MM/PBSA and MM/GBSA methods. 7. Entropy effects on the performance of end-point binding free energy calculation approaches. *Phys. Chem. Chem. Phys.* 20, 14450–14460. doi:10.1039/c7cp07623a
- Tan, V. Y., Lewis, S. J., Adams, J. C., and Martin, R. M. (2013). Association of fascin-1 with mortality, disease progression and metastasis in carcinomas: a systematic review and meta-analysis. *BMC Med.* 11, 52. doi:10.1186/1741-7015-11-52
- Vanommeslaeghe, K., and MacKerell, A. D., Jr. (2015). CHARMM additive and polarizable force fields for biophysics and computer-aided drug design. *Biochim. Biophys. Acta (Bba) - Gen. Subjects* 1850, 861–871. doi:10.1016/j.bbagen.2014.08.004
- Wang, E., Sun, H., Wang, J., Wang, Z., Liu, H., Zhang, J. Z. H., et al. (2019). End-Point Binding Free Energy Calculation with MM/PBSA and MM/GBSA: Strategies and Applications in Drug Design. *Chem. Rev.* 119, 9478–9508. doi:10.1021/acs.chemrev.9b00055
- Wang, J., and Kollman, P. A. (2001). Automatic parameterization of force field by systematic search and genetic algorithms. *J. Comput. Chem.* 22, 1219–1228. doi:10.1002/jcc.1079
- 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
- Wenzel, J., Matter, H., and Schmidt, F. (2019). Predictive Multitask Deep Neural Network Models for ADME-Tox Properties: Learning from Large Data Sets. *J. Chem. Inf. Model.* 59, 1253–1268. doi:10.1021/acs.jcim.8b00785
- Williams, H. D., Trevaskis, N. L., Charman, S. A., Shanker, R. M., Charman, W. N., Pouton, C. W., et al. (2013). Strategies to address low drug solubility in discovery and development. *Pharmacol. Rev.* 65, 315–499. doi:10.1124/pr.112.005660
- Wolber, G., and Langer, T. (2005). LigandScout: 3-D pharmacophores derived from protein-bound ligands and their use as virtual screening filters. *J. Chem. Inf. Model.* 45, 160–169. doi:10.1021/ci049885e
- Wu, X., and Brooks, B. R. (2003). Self-guided Langevin dynamics simulation method. *Chem. Phys. Lett.* 381, 512–518. doi:10.1016/j.cplett.2003.10.013
- Xiang, M., Cao, Y., Fan, W., Chen, L., and Mo, Y. (2012). Computer-aided drug design: lead discovery and optimization. *Cchts* 15, 328–337. doi:10.2174/138620712799361825
- Yamashiro-Matsumura, S., and Matsumura, F. (1985). Purification and characterization of an F-actin-bundling 55-kilodalton protein from HeLa cells. *J. Biol. Chem.* 260, 5087–5097. doi:10.1016/s0021-9258(18)89183-9
- Yang, S., Huang, F.-K., Huang, J., Chen, S., Jakoncic, J., Leo-Macias, A., et al. (2013). Molecular mechanism of fascin function in filopodial formation. *J. Biol. Chem.* 288, 274–284. doi:10.1074/jbc.m112.427971
- Ylilauri, M., and Pentikäinen, O. T. (2013). MMGBSA as a tool to understand the binding affinities of filamin-peptide interactions. *J. Chem. Inf. Model.* 53, 2626–2633. doi:10.1021/ci4002475

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 Lin, Lin, Wu, Liu, Cheng, Xu, Li and Dong. 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.



Synthesis, Computational Pharmacokinetics Report, Conceptual DFT-Based Calculations and Anti-Acetylcholinesterase Activity of Hydroxyapatite Nanoparticles Derived From *Acorus Calamus* Plant Extract

OPEN ACCESS

Edited by:

Zhongjie Liang,
Soochow University, China

Reviewed by:

Muthumuni Managa,
University of South Africa, South Africa
Philani Mashazi,
Rhodes University, South Africa

*Correspondence:

Chandan Shivamallu
chandans@jssuni.edu.in
Shiva Prasad Kollur
shivachemist@gmail.com
Daniel Glossman-Mitnik
daniel.glossman@cimav.edu.mx

Specialty section:

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

Received: 15 July 2021

Accepted: 17 September 2021

Published: 07 October 2021

Citation:

Pradeep S, Jain AS,
Dharmashekara C, Prasad SK,
Akshatha N, Pruthvish R,
Amachawadi RG, Srinivasa C, Syed A,
Elgorban AM, Al Kheraif AA,
Ortega-Castro J, Frau J,
Flores-Holguín N, Shivamallu C,
Kollur SP and Glossman-Mitnik D
(2021) Synthesis, Computational
Pharmacokinetics Report, Conceptual
DFT-Based Calculations and Anti-
Acetylcholinesterase Activity of
Hydroxyapatite Nanoparticles Derived
From *Acorus Calamus* Plant Extract.
Front. Chem. 9:741037.
doi: 10.3389/fchem.2021.741037

Sushma Pradeep¹, Anisha S. Jain¹, Chandan Dharmashekara¹, Shashanka K. Prasad¹, Nagaraju Akshatha², R. Pruthvish³, Raghavendra G Amachawadi⁴, Chandrashekar Srinivasa⁵, Asad Syed⁶, Abdallah M. Elgorban⁶, Abdulaziz A. Al Kheraif⁷, Joaquín Ortega-Castro⁸, Juan Frau⁸, Norma Flores-Holguín⁹, Chandan Shivamallu^{1*}, Shiva Prasad Kollur^{10*} and Daniel Glossman-Mitnik^{9*}

¹Department of Biotechnology and Bioinformatics, School of Life Sciences, JSS Academy of Higher Education and Research, Mysuru, India, ²Department of Physics, Marimallappa PU College, Mysuru, India, ³Department of Biotechnology, Acharya Institute of Technology, Bengaluru, India, ⁴Department of Clinical Sciences, College of Veterinary Medicine, Kansas State University, Manhattan, KS, United States, ⁵Department of Studies in Biotechnology, Davangere University, Shivagangothri, Davangere, India, ⁶Department of Botany and Microbiology, College of Science, King Saud University, Riyadh, Saudi Arabia, ⁷Dental Health Department, College of Applied Medical Sciences, King Saud University, Riyadh, Saudi Arabia, ⁸Departament de Química, Universitat de les Illes Balears, Palma de Mallorca, Spain, ⁹Laboratorio Virtual NANOCOSMOS, Departamento de Medio Ambiente y Energía, Centro de Investigación en Materiales Avanzados, Chihuahua, México, ¹⁰Department of Sciences, Amrita School of Arts and Sciences, Amrita Vishwa Vidyapeetham, Mysuru Campus, Mysuru, India

Over the years, Alzheimer's disease (AD) treatments have been a major focus, culminating in the identification of promising therapeutic targets. A herbal therapy approach has been required by the demand of AD stage-dependent optimal settings. Present study describes the evaluation of anti-acetylcholinesterase (AChE) activity of hydroxyapatite nanoparticles derived from an *Acorus calamus* rhizome extract (AC-HAp NPs). The structure and morphology of as-prepared (AC-HAp NPs) was confirmed using powder X-ray diffractometer (XRD), scanning electron microscopy (SEM), transmission electron microscopy (TEM) and high-resolution transmission electron microscopy (HR-TEM). The crystalline nature of as-prepared AC-HAp NPs was evident from XRD pattern. The SEM analysis suggested the spherical nature of the synthesized material with an average diameter between 30 and 50 nm. Further, the TEM and HR-TEM images revealed the shape and size of as-prepared (AC-HAp NPs). The interplanar distance between two lattice fringes was found to be 0.342 nm, which further supported the crystalline nature of the material synthesized. The anti-acetylcholinesterase activity of AC-HAp NPs was greater as compared to that of pure HAp NPs. The mechanistic evaluation of such an activity carried out using in silico studies suggested that the anti-acetylcholinesterase activity of phytoconstituents derived from *Acorus calamus* rhizome extract was mediated by BDNF, APOE4, PKC- γ , BACE1 and γ -secretase proteins. The global and local descriptors, which are the underpinnings of Conceptual Density Functional Theory

(CDFT), have been predicted through the MN12SX/Def2TZVP/H₂O model chemistry to help in the comprehension of the chemical reactivity properties of the five ligands considered in this study. With the further objective of analyzing their bioactivity, the CDFT studies are complemented with the estimation of some useful computed pharmacokinetics indices, their predicted biological targets, and the ADMET parameters related to the bioavailability of the five ligands are also reported.

Keywords: Alzheimer's disease, neuropathology, neurofibrillary tangles, amyloid plaques, molecular docking, computational pharmacokinetics, Conceptual DFT

1 INTRODUCTION

The most frequent form of dementia is Alzheimer's disease. It affects millions of individuals around the world, and the number is rapidly increasing. Alzheimer's disease has been found to impact socially and financially the lives of those affected (Wang et al. 2016). According to the amyloid hypothesis, misfolding of the extracellular protein collected in senile plaques and intracellular deposition of misfolded tau protein in neurofibrillary tangles induce memory loss and disorientation, as well as personality and cognitive decline over time (Förstl and Kurz 1999). Over 24 million people worldwide are estimated to have dementia, with Alzheimer's disease accounting for the bulk of cases (Mayeux and Stern 2012). As a result, research into Alzheimer's disease, which is a huge public health concern, must be prioritized. The treatments available are aimed at alleviating the symptoms of Alzheimer's disease, implying the necessity to have a better understanding of disease pathophysiology to find/develop treatments that can lessen symptoms or repair harm already done (Solfrizzi et al. 2011). The most important component in focusing therapy efforts is integrating both pharmaceutical and psychosocial support systems towards early diagnosis and studying the disease further.

The absence of treatment for Alzheimer's disease and dementia has become a major public health concern. Alzheimer's disease is a neurological disease that worsens with time Heneka et al. (2015). The therapeutic drugs used to treat Alzheimer's disease must either cure or slow the illness. There are few first-line medications/drugs available to treat AD and these works as acetylcholinesterase inhibitors and have FDA approval in the United States Wilson et al. (2011). However, because none of the treatments are designed to boost neural functioning, none of them can entirely heal the disease or improve the patient's cognitive or memory skills. As a result, there is a need for an alternative supply of medication to treat Alzheimer's disease Blennow et al. (2006). The amyloid hypothesis, presented in 1991, claimed that AD was caused by the accumulation of A β proteins. The APP (amyloid β -protein precursor) gene, which produces A β protein, is found on the 21st chromosome, and people with Down's syndrome (trisomy 21) have an extra copy of this gene, resulting in the earliest reported symptom(s) of Alzheimer's disease at the age of 40 Waring and Rosenberg (2008). APOE4 (Apolipoprotein E4) has long been thought to be a key risk factor for Alzheimer's disease since it aids in the breakdown of A β proteins. However, some isoforms of APOE4

are ineffective, resulting in amyloid buildup in the brain Selkoe (1999). Also, the enzyme acetylcholinesterase is involved in cholinergic neurotransmission. It degrades acetylcholine stopping the neurotransmission process. The assay of AChE activity can be used to confirm the efficacy of various test substances as herbal extracts or herbal extracts derived nanoparticles in terms of treatment Kim (2018).

Ayurveda is an ancient medicinal system that employs a variety of herbs and plants to effectively treat a wide range of diseases Strittmatter et al. (1993). Herbal treatments from plants contain a blend of phytochemicals with varying pharmacobiological relevance and can treat a variety of disorders. Plants, in fact, have long been a primary source of medications in a variety of therapeutic traditions Mahley et al. (2006). The plant's pharmacological activity, such as anti-amyloidogenic, anti-inflammatory, antioxidants and anti-cholinesterase properties, are due to phytochemical components such as polyphenols, alkaloids, triterpenes, tannins, lignins, sterols and flavonoids Francis et al. (1999). *Acorus calamus*, a member of the *Acoraceae* family, is native to India. In Ayurveda, this plant is revered for its revitalizing effects on the neurological system, brain, and digestive systems. Alkaloids, volatile oil, steroids, tannins, sesquiterpenes, polyphenols, saponin, lignin, mucilage, monoterpenes, flavonoids and glycosides chemicals are among the phytoconstituents found in Ayurveda Martorana et al. (2010). *A. calamus* has anti-microbial, anti-ulcer, antidiabetic, insecticidal, neuroprotective, anti-allergic, anti-inflammatory, cardioprotective, pesticidal, anti-cancer and anti-oxidant activities. Ferreira-Vieira et al. (2016).

Nanomedicines (NMs) have a number of unique qualities that allow them to deliver anti-AD therapies to specific brain locations Spuch et al. (2012). NMs benefit from smaller dimensions and enhanced biocompatibility, making therapeutic chemicals easier to move into the brain Fakhoury et al. (2015). NMs that are small (about 100–10,000 times smaller than a human cell) can easily interact with proteins and chemicals on the cell surface and inside the cell. The essential core structures of NP-functionalized NMs ensure drug encapsulation or conjugation, as well as protection and sustained blood circulation Leszek et al. (2017). NMs can also target cells or even an intracellular compartment such as A β in cells, allowing the drug to be delivered at a predetermined dosage straight to the diseased spot Kim et al. (2012). NMs can reduce the dose and frequency of treatment, resulting in better patient compliance Altinoglu and Adali (2020). Nanomedicines have potential advantages over other conventional ways of drug

delivery to the brain to cure AD, such as favorability to the brain, greater stability, biocompatibility and biodegradability, protection from enzymatic degradation, increased half-life, improved bioavailability, and controlled release, despite some clinical issues Knop et al. (2010).

Hydroxyapatite ($\text{Ca}_{10}(\text{PO}_4)_6(\text{OH})_2$) is composed of 70% apatite calcium phosphate and remaining 30% of natural materials Gopi et al. (2013). As a result, it is frequently employed in biomedical applications including fillers for bone deformities, scaffolds for tissue engineering, coatings on metallic implants to increase biocompatibility, and drug/protein delivery carriers Youness et al. (2017). On the other hand, green synthesis of nanoparticles using plant leaf extracts has opened a new era in research.

A quite often method of predicting a small molecule's orientation when it is bound to a target molecule to create a stable complex is the *in silico* molecular docking approach Davies (1999). Predicting the strength of association or binding affinity between two molecules requires knowledge of the preferred orientation. The study of how two or more molecular structures fit together is known as molecular docking. As a result, molecular docking can be used to forecast the strength that will be created between the molecules Polvikoski et al. (1995). The binding behavioral studies have aided in understanding the fundamental biological processes that help in rational drug discovery techniques Lacor et al. (2007). Rational drug discovery (RDD) study was to find an inhibitor that binds and stops the action of some toxic proteins produced in the human body. RDD allows researchers to forecast how tiny molecules like ligands bind in the receptor target site. One of the most often used strategies in structure-based drug design is molecular docking Nikolaev et al. (2009). The present study investigates the *in silico* and AChE inhibitory activity of pure HApNPs and Ac-HApNPs prepared using aqueous extract of rhizome of *Acorus calamus* against AD proteins.

2 MATERIALS AND METHODS

All the chemicals and reagents were procured from Loba chemicals (Bangalore, India). Demineralized water was collected from an ELGA RO system and was used throughout the experiments (Elga Veolia, Lane End, United Kingdom). The crystalline phases were recorded on Bruker X-ray diffractometer with a scan range of 20–70° at a 2°/min scan rate using Cu K α (1.5406 Å) radiation (Bruker, Karlsruhe, Germany). The morphology and elemental composition were studied using Scanning electron microscopy (SEM) and Energy dispersive X-ray (EDX) mapping, respectively, which were recorded on a Zeiss microscope (Carl Zeiss, White Plains, NY, United States). Transmission electron microscopy (TEM) images and Selected Area Electron Diffraction (SAED) patterns were recorded on a JEOL 2100F FEG apparatus operating at 200 kV after casting a drop of sample material for dispersion in ethanol over a Cu grid (JEOL, Akishima, Tokyo, Japan).

2.1 Plant Material Collection

The matured rhizomes of the *Acorus calamus* plant grown were collected in the region around Mysuru, Karnataka, India. The

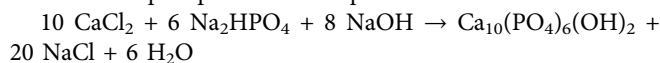
rhizomes collected were washed with single distilled water and 0.5% sodium hypochlorite solution and lastly with double distilled water to remove microscopic entities and other dust particles and later the rhizomes were shade dried for 45 days at room temperature ($28 \pm 5^\circ\text{C}$) Turner et al. (2003). The dried materials were then crushed using a blender and made into fine powder.

2.2 Aqueous Rhizome Extract Preparation

The powdered sample was extracted using the Soxhlet apparatus. Around 60 g of sample was added into thimble for extraction using water as the solvent for 8 h (24 cycles). The obtained extracts were air-dried and stored at 4°C. Further, it was subjected to qualitative and quantitative phytochemical analysis to quantify the presence of various phytochemicals present in the rhizome extract Mudher and Lovestone (2002). Further, the prepared extract was sent for GC-MS (gas chromatography-mass spectrometry) analysis to identify the important phytochemical constituents and functional groups Goedert et al. (1991).

2.3 Preparation of Hydroxyapatite Nanoparticles (AC-HAP NPs)

1 M CaCl_2 and 0.6 M Na_2HPO_4 were prepared using leaf extract as the solvent and separately raised to pH 10.0 using 0.8 M NaOH solution for the synthesis of HAP nanospheres. The CaCl_2 solution was then aggressively agitated at room temperature with a magnetic stirrer, and then Na_2HPO_4 solution was added drop by drop to generate a gelatinous precipitate. The formation of precipitation of HAP is described as follows:



The precipitate formed was centrifuged to eliminate byproducts before being dried in a hot air oven at 130°C for 6 h and resulting in a dry cake which was crushed to form powder Iqbal et al. (2005). In addition, for comparison, HAP without rhizome extract was made and termed control (pure HAP).

2.4 Anti-Acetylcholinesterase Inhibition Assay

Ellman's method was slightly modified to measure AChE inhibition. Briefly, 150 μL of 0.1 M sodium phosphate buffer (pH 8.0), 10 μL of test chemical solution, and 20 μL of AChE enzyme solution (0.1 units/mL) were combined and incubated at 25°C for 15 min. After that, 10 μL of DTNB (10 mM) (5,5-dithio-bis-(2-nitrobenzoic acid)) was added, and the reaction was started by adding substrate (10 μL of ATCI (acetylthiocholine iodide), 14 mM solution). The formation of the colored product, 5-thio-2-nitrobenzoate anion generated by the reaction of DTNB and thiocholine, which is released by the ATCI's hydrolysis, can be used to determine the enzyme's hydrolysis. After 10 min, the colored product was detected at 410 nm wavelength. Tacrine was utilized as a positive control Kametani and Hasegawa (2018). Inhibition (%) was estimated using the following equation:

$$\text{Inhibition Activity (\%)} = \left[1 - \frac{\text{Absorbance of sample}}{\text{Absorbance of control}} \right] \times 100$$

2.5 In Silico Bioinformatics Studies

2.5.1 Lead Optimization

GC-MS analysis is one of the fast, best and accurate technique used to detect various compounds that includes organic acids, long chain hydrocarbons, alcohols, steroids, amino acids, nitro compounds, alkaloids and esters Pradeep et al. (2020). The GC-MS analysis of aqueous extract of rhizomes of the *Acorus calamus* detected the presence of 110 compounds among which a library of 20 molecules were created based on the review of literature Singh et al. (2010). Subsequently, all 20 molecules were analyzed for their bioactivity through *in silico* molecular docking studies.

The 2D chemical structure of all the 20 molecules were sketched using ChemSketch software Tian et al. (2010). These files were further converted to 3D structures (pdb format) using OpenBabel GUI2.4.1 software Obulesu and Rao (2011). Before carrying out the molecular docking studies, the geometry of all the structures was cleaned using ArgusLab program Jaiswal et al. (2011).

2.5.2 Protein Preparation

Proteins/enzymes that are mainly involved in amyloid hypothesis of AD are BNDF (brain-derived neurotrophic factor), APOE4, PKC- γ (protein kinase c), BACE1 and γ -secretase. By altering synaptic plasticity, BDNF plays a fundamental role in cognition, learning, and memory formation, making it a critical molecule in dementia and neurodegenerative illnesses Prasad et al. (2021). The biggest genetic risk factor for Alzheimer's disease is APOE4. It is important for the metabolism of lipids such as cholesterol and for the repair of neuronal injury in the brain Kollur et al. (2021). PKC isoforms have crucial functions as tau kinases in addition to their role in memory formation. PKC- γ is involved in the maintenance of synaptic plasticity Prasad et al. (2020c). BACE1 (β -secretase 1) catalyses the amyloid precursor protein's initial cleavage to produce A β proteins. As a result, inhibiting BACE1 activity could prevent one of the earliest pathogenic events in Alzheimer's disease Ankegowda et al. (2020). γ -Secretase is a protease complex that cuts the transmembrane domain of the APP to create the amyloid β -protein (A β), an aggregation-prone product that builds up in the brain of Alzheimer's patients Desikan et al. (2009). The above mentioned protein play very important role in memory and cognition functions thus, all the 5 enzymes were selected for the *in silico* inhibition studies to dock the screened phytocompounds against them.

The three dimensional structures of BNDF, APOE4, PKC- γ , BACE1 and γ -secretase with their respective PDB IDs such as 1B8M, 1GS9, 3PFQ, 4L7G and 5A63, required for the *in silico* studies were obtained from Protein Data Bank (PDB) [https://www.rcsb.org/], a protein structural database Tiraboschi et al. (2004). Before beginning with the docking analysis, all of the protein structures were refined and energy-optimized Mendez (2006). The cleanup of the proteins was accomplished by finishing incomplete residues with hydrogen atoms. External

ligands and non-essential ions were removed from the protein structure Waldemar et al. (2007).

2.5.3 Protein Structure Validation

Using the PROCHECK module of the PDBSum server [https://servicesn.mbi.ucla.edu/PROCHECK/], the stereochemical stability of the predicted models was further verified using various protein quality-based parameters such as percentage of residues lying in favored and allowed regions, number of glycine and proline residues, and orientation of dihedral angles including phi (ϕ) and psi (ψ), as well as backbone conformation Schroeter et al. (2009).

2.5.4 Binding Site Prediction

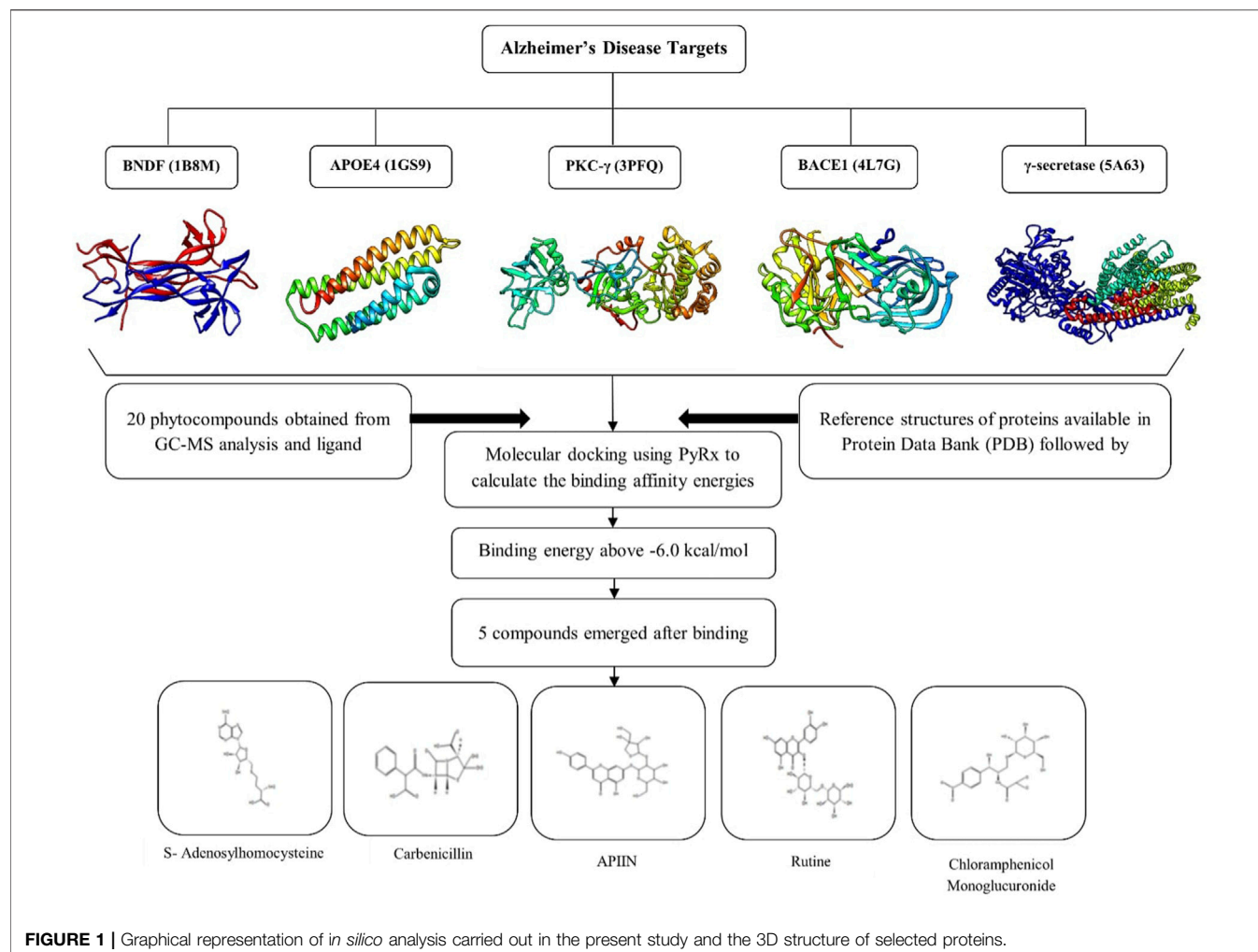
Residues in the protein interacting with the ligand is termed as a binding site of that protein. This binding site was predicted using the CASTp 3.0 server (http://sts.bioe.uic.edu/castp/index.html?4jii) which stands for Computed Atlas of Surface Topography of proteins. Surface pockets and interior cavities are identified and measured by CASTp Jain et al. (2021). The modeled protein is used to predict ligand binding sites, and the server identifies the amino acids that are relevant for binding interactions.

2.5.5 Molecular Docking Studies

MD is a technique for studying the molecular behavior of target proteins when they bind. It is a tool widely utilized in drug development. PyRx 0.8 [https://pyrx.sourceforge.io/], a virtual screening tool was used to accomplish molecular docking Prasad et al. (2020a). A genetic algorithm is an effective approach for searching the docked conformer's space globally. It also allows for the existence of a population of solutions, which can evolve through processes like 'breeding' and 'mutation' Prasad et al. (2020b). Poor solutions are extinguished, while good ones are passed down to future generations. In a few tens of generations, such algorithms may usually obtain an excellent answer Uppar et al. (2021). The MD results were analyzed for their bonded and non-bonded interactions using Discovery Studio 3.1 (Accelrys, San Diego, United States) visualization software Avinash et al. (2021). The whole process is depicted in **Figure 1**

2.6 Computational Pharmacokinetics Analysis

It is critical to learn about the pharmacokinetics, or the fate of a molecule in the body, during the creation of a novel therapeutic medicine. Individual indices known as Absorption, Distribution, Metabolism, Excretion, and Toxicity (ADMET) factors are typically used to do this. As an alternative to employing experimental approaches to determine these parameters, computer models are commonly used. Chemicalize, a software developed by ChemAxon [http://www.chemaxon.com], and the internet available SwissADME program were used to estimate some ADME parameters in this study Daina et al. (2017). Additional information about the Pharmacokinetics parameters and the ADMET properties were obtained by resorting to pkCSM Pires et al. (2015), a software for the prediction of small-molecule pharmacokinetic properties using SMILES [https://biosig.unimelb.edu.au/pkcsml/] (accessed, June 2021). Molinspiration



Chemoinformatics' freely available Molinspiration software [https://www.molinspiration.com/] (accessed, June 2021) was used to conduct similarity searches in the chemical space of compounds with molecular structures comparable to those being researched and to predict bioactivity ratings for a variety of pharmacological targets. A Webtool named SwissTargetPrediction for efficient prediction of protein targets of small molecules was used for the determination of the potential bioactivity of the five ligands considered in this study Daina et al. (2019). The associated website allows the estimation of the most probable macromolecular targets of a small molecule, assumed as bioactive.

2.7 Conceptual DFT Studies

The molecular energy, electronic density, and orbital energies of a particular system, including the Highest Occupied Molecular Orbital (HOMO) and the Lowest Unoccupied Molecular Orbital (LUMO) were determined using the Kohn-Sham (KS) approach Lewars (2003); Young (2001); Jensen (2007); Cramer (2004) while making use of the Conceptual DFT (CDFT) methodology Parr and Yang (1989); Chermette (1999); Geerlings et al. (2003, 2020); Toro-Labbé (2007); Chattaraj (2009); Chakraborty and Chattaraj (2021). The conformers of the compounds studied in this work were determined

using MarvinView 17.15 from ChemAxon [http://www.chemaxon.com] by using the entire MMFF94 force field to perform Molecular Mechanics calculations Halgren (1996a,b, 1999); Halgren and Nachbar (1996); Halgren (1996c). This was followed by a geometry optimization and frequency calculation by means of the Density Functional Tight Binding (DFTBA) methodology Frisch et al. (2016). This last step was required for the verification of the absence of imaginary frequencies as a check for the stability of the optimized structures as being a minimum in the energy landscape. The electronic properties and the chemical reactivity descriptors of the studied molecules involved the use of MN12SX/Def2TZVP/H₂O model chemistry Peverati and Truhlar (2012); Weigend and Ahlrichs (2005); Weigend (2006) on the optimized molecular structures due to its ability in the verification of the 'Koopmans in DFT' (KID) protocol Flores-Holguín et al. (2019b); Flores-Holguín et al. (2019d); Frau and Glossman-Mitnik (2018a); Frau and Glossman-Mitnik (2018b); Frau and Glossman-Mitnik (2018c); Frau and Glossman-Mitnik (2018d); Frau and Glossman-Mitnik (2018e); Frau and Glossman-Mitnik (2018f); Flores-Holguín et al. (2019a); Frau et al. (2019); Flores-Holguín et al. (2019c); Flores-Holguín et al. (2020c); Flores-Holguín et al. (2020a); Flores-Holguín et al. (2020b); Flores-Holguín et al. (2021) using Gaussian 16 Frisch et al. (2016) and the SMD model for

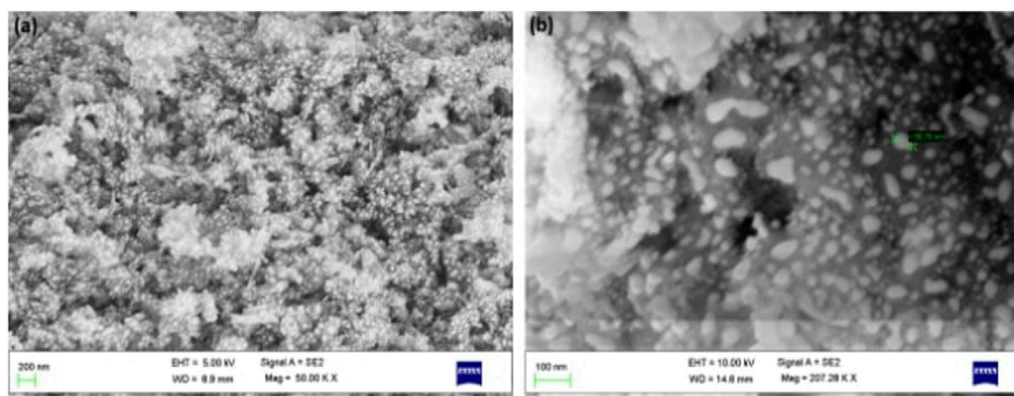


FIGURE 2 | SEM images of *A. calamus* rhizome extract derived HAp NPs.

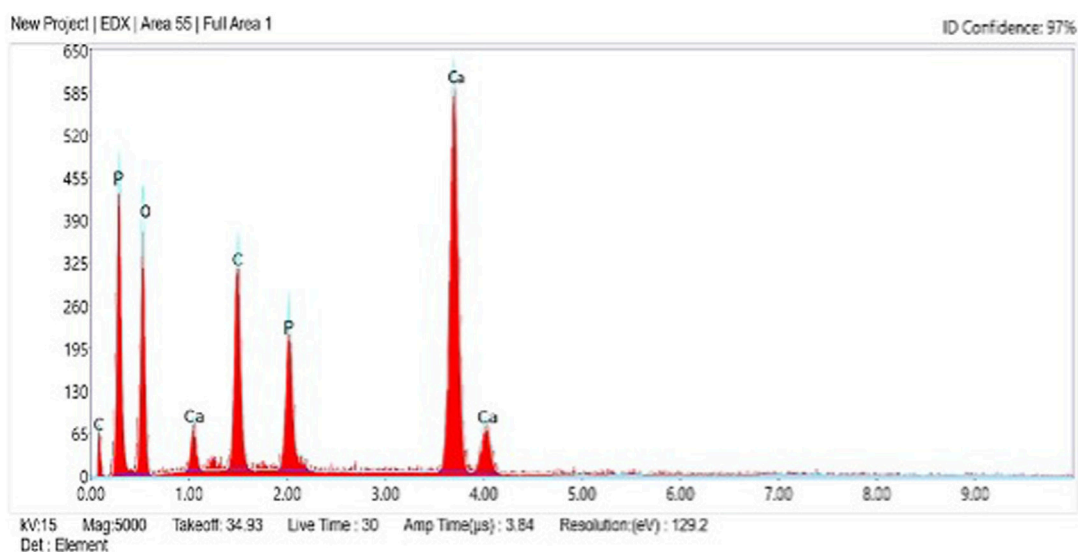


FIGURE 3 | EDX spectra showing the elements present in as-prepared HAp NPs.

the simulation of the solvent Marenich et al. (2009). This model chemistry considers the MN12SX screened-exchange density functional Peverati and Truhlar (2012) together with the Def2TZVP basis set Weigend and Ahlrichs (2005); Weigend (2006) and in all cases the charge of the molecules is equal to zero while the radical anion and cation have been considered in the doublet spin state.

3 RESULTS

3.1 SEM Analysis

The surface morphology of as-prepared AC-HAp NPs showed spherical shaped particles which are highly agglomerated. The average particles size ranged between 30 and 50 nm (Figure 2). Further, EDAX analysis was carried out to explore the composition of the AC-HAp NPs. Figure 3 depicts the EDX

spectra of as-obtained AC-HAp NPs showing the characteristic peaks of Ca, P and O with the atomic and weight percentages of the elemental particles providing the mean relative calcium to phosphate ratios, and was found to be 1.68, which is quite close to the Ca/P ratio of the human bone.

3.2 X-ray Diffraction Analysis

The crystalline phases of the as-prepared AC-HAp NPs was determined using XRD diffraction pattern (Figure 4). The position of observed diffraction peaks are in good agreement with the JCPDS (89-6438). The peaks observed at $2\theta = 27.8^\circ, 30.1^\circ, 33.3^\circ, 35.1^\circ, 36.2^\circ, 45.8^\circ, 49.8^\circ$, and 60.2° corresponds to the (hkl): (002), (210), (211), (112), (212), (400), (222), and (323), matching exactly with the hexagonal system with primitive lattice. Furthermore, the average particle size of the as-prepared AC-HAp NPs was said to 36 nm, which was calculated (using FWHM) by Scherrer' formulam $D = k \lambda / \beta \cos\theta$ Ankegowda et al. (2020).

3.3 TEM Analysis

The structure and morphology of as-prepared AC-HAp NPs was determined by Transmission electron microscopy (TEM). The TEM images as shown in **Figure 5**, reveals that the average sizes were between 30 and 40 nm. The spherical nature of AC-HAp NPs is evident from the TEM image. Moreover, the aggregate blocks with porous structure of the material can be seen from TEM image. Further, the HR-TEM image showed that the inter-planer spacing between two lattice fringes was 0.342 nm (**Figure 5B**), corresponding to (102) lattice plane of AC-HAp NPs, and same has been confirmed by SAED pattern, which shows the crystalline structure of as-prepared AC-HAp NPs (**Figure 5C**).

3.4 AChE Inhibition

By inhibiting AChE of the cholinergic synapse, AChE inhibitors enhance acetylcholine levels by inhibiting AChE of the cholinergic synapse thus enhancing the function and relieving the symptoms of neurological illnesses, including Alzheimer's disease. In addition to alkaloid-derived chemicals as the most well-known natural AChE inhibitors, plant-derived extract are

also a major source of AChE inhibitors. HAp NPs from the rhizome of *A. calamus* has been demonstrated to inhibit AChE. As a result, AC-HAp NPs of the *A. calamus* rhizome were found to have AChE inhibitory action.

Surprisingly, the anti-AChE activity was evidently greater with AC-HAp NPs when compared to positive control Tacrine and pure HAp NPs with IC₅₀ value of about 22.39 µg/ml (**Figure 6**).

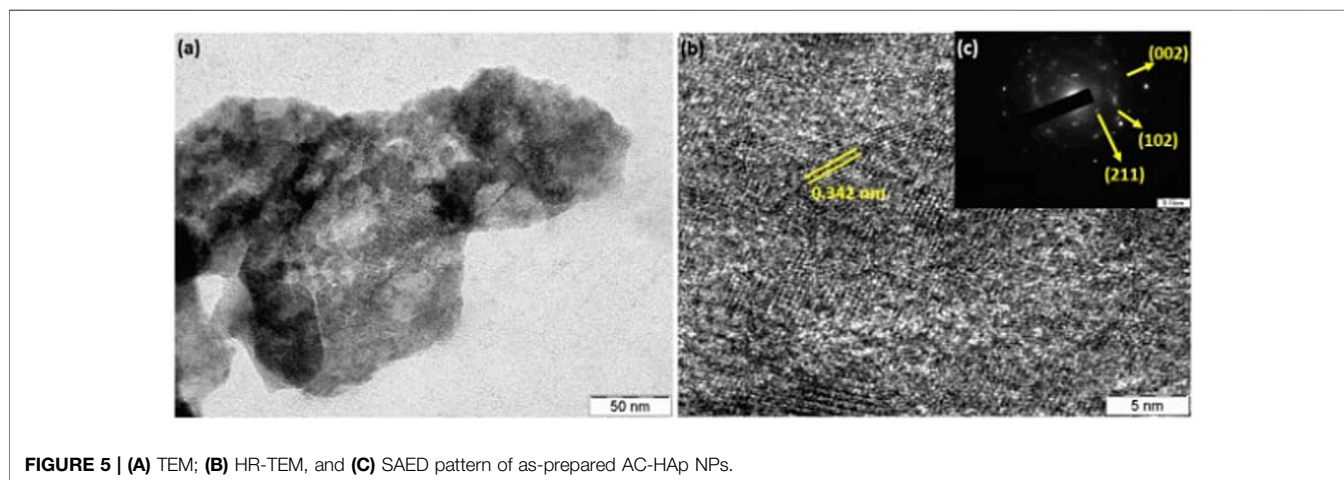
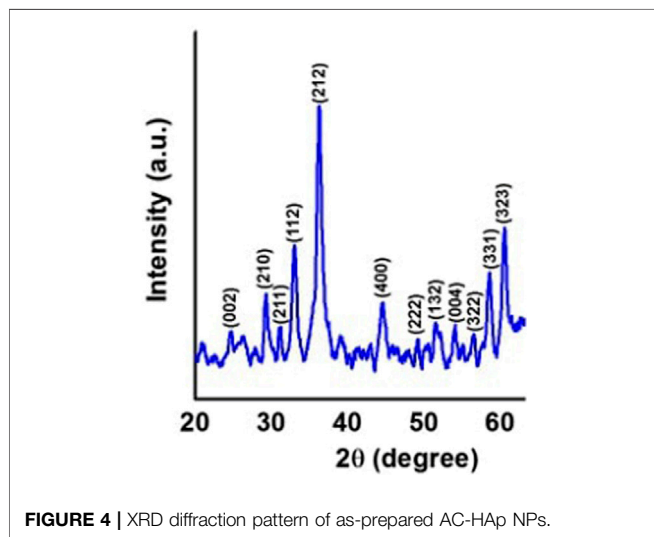
3.5 MD Interactions

Interaction affinity describes the strength of protein-ligand binding. The binding affinity is determined by the strength of the attractive force between the protein and the ligand. The best molecularly docked poses were analyzed and visualized. The docking procedure was validated before the ligands were screened. The optimum ligand-protein complex orientations were investigated. The docking score was used to identify the excellent docking conformation. The binding affinity of a specific protein-ligand complex with a known 3D structure is computed using the binding energy score. Van der Waals interactions, hydrogen bonding and hydrophobic effects are all included in the binding energy (**Table 1**). The 3D and 2D interactions between all the protein-Apiin complexes were analyzed and its images were taken using Discovery Studio 3.1 visualization software from **Figures 7–11**.

The MD studies revealed that only 5 molecules out of 20 molecules has the ability to bind to the active site with the selected targets by forming greater binding affinity and least binding energy against the targets. Also, all the 25 protein-ligand complexes were capable of forming a very good amount of bonded and non-bonded interaction between them. Therefore, all the five phytoconstituents obtained from the GC-MS analysis of aqueous extracts of *A. calamus* were able to form a great interaction with all the 5 selected targets and thus showed the *in silico* inhibition activity against AD.

3.6 Computational Pharmacokinetics Report

The Bioactivity Scores, that is a measure of the ability of the molecules to behave or interact with different receptors, for the five ligands are presented in **Table 2**



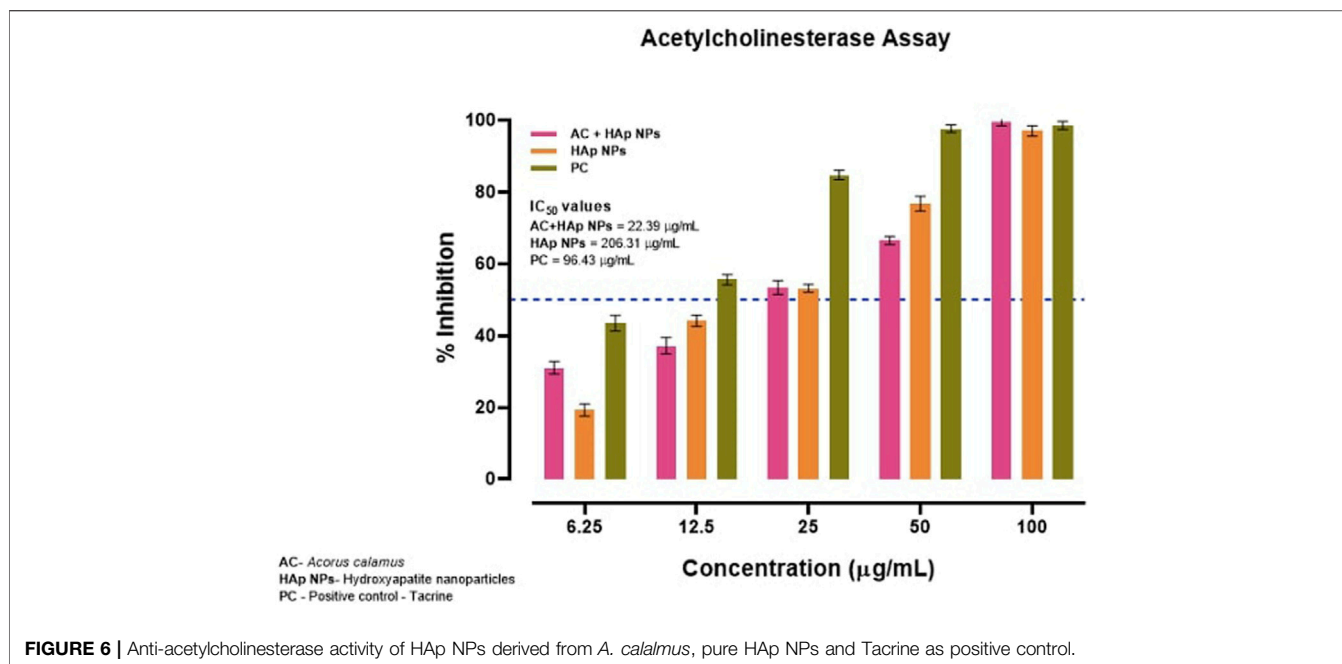


FIGURE 6 | Anti-acetylcholinesterase activity of HAp NPs derived from *A. calamus*, pure HAp NPs and Tacrine as positive control.

A chemical with a bioactivity score more than 0 is predicted to have significant biological activities, while values between -0.50 and 0.00 are moderately active. The molecular system is considered inactive if the bioactivity score is less than -0.50 . The findings clearly show that the drug complexes' physiological activities may be mediated by many pathways, including interactions with GPCR ligands, protease inhibitors, and other enzymes. The results from **Table 2** indicate that S-Adenosylhomocysteine will mostly act as a GPCR ligand, an enzyme inhibitor and a protease inhibitor. For the case of Carbenicillin, the main interaction are going to be as a protease and enzyme inhibitor. Lastly, by considering the Apiin, Rutine and Chloramphenicol monoglucoronide ligands, they may be regarded as enzyme inhibitors, and with the exception of Rutine, also as protease inhibitors.

An ADMET study is the assessment of pharmacokinetics of a drug which stands for Absorption, Distribution, Metabolism, Excretion and Toxicity. The prediction of the fate of a drug and the effects caused by a drug inside the body, such as how much drug is absorbed if administered orally and how much is absorbed in the gastrointestinal tract, is an indispensable part of drug discovery. In a similar way, if the absorption is poor, its distribution and metabolism would be affected, which can lead to causing neurotoxicity and nephrotoxicity.

The computed ADMET properties of the five studied ligands are presented in **Table 3**.

A chemical can reach a tissue if it is injected into the bloodstream. Before being taken up by target cells, a drug is usually given through mucous surfaces such as the digestive tract, i.e. intestinal absorption. Drug absorption is limited following oral delivery due to poor substance solubility, intestinal transit time, gastric emptying time, difficulty permeating the intestinal wall, and chemical instability in the stomach. Absorption is

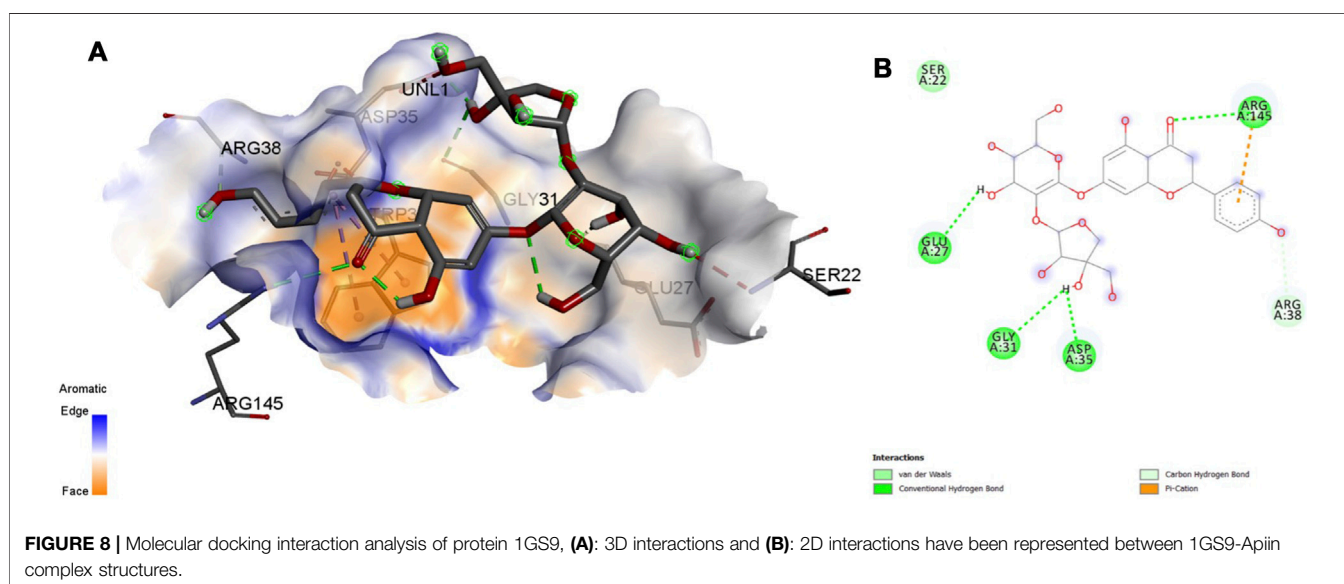
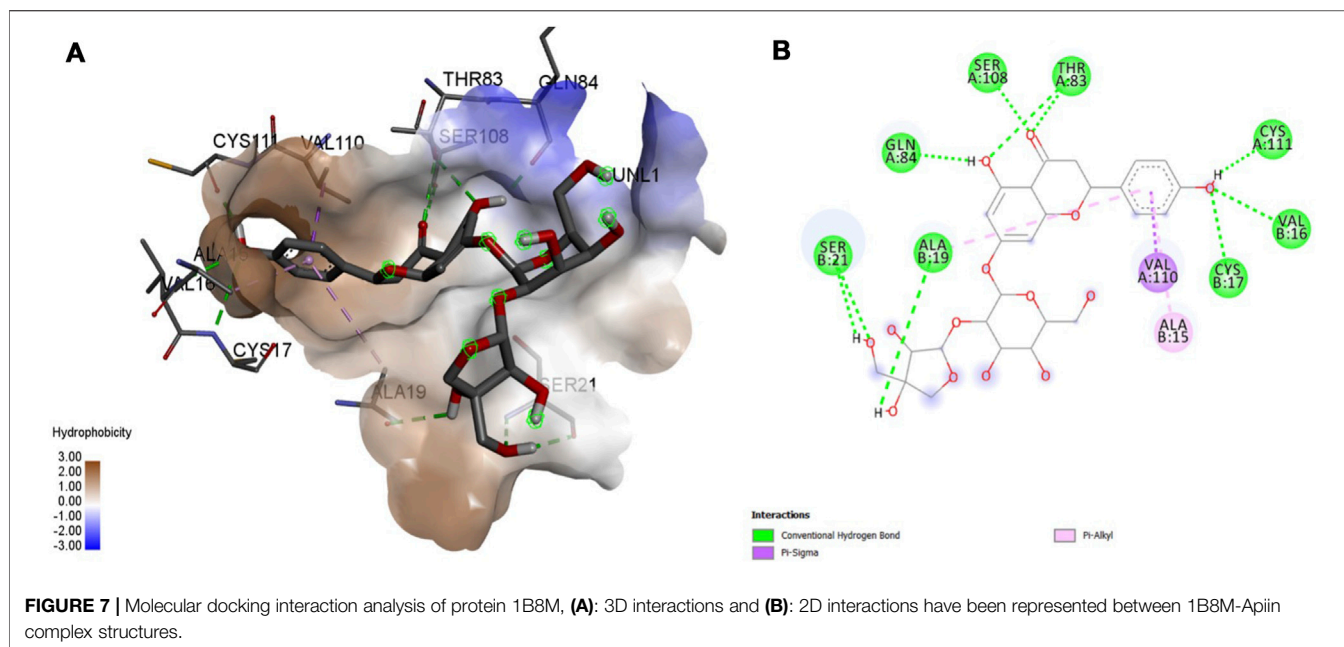
important because it affects the bioavailability of a chemical. For medications with low absorption, oral delivery, such as inhalation or intravenously, is less desirable Jujjavarapu et al. (2019); Pires et al. (2015). For projected values >0.90 , a substance is deemed to have a high Caco-2 permeability across the human intestinal mucosa, being Apiin with a value of 0.737 the only drug that could be considered in this regard. In most cases, the gut is the principal location of medication absorption from an orally delivered solution. Intestinal Absorption forecasts the percentage of a substance that will be absorbed through the human intestine, with less than 30% being considered poorly absorbed. Again, Apiin is the only drug that satisfies this requirement, according to **Table 3**. The model forecast whether or not a particular substance will be a P-glycoprotein substrate. This is verified for all the molecules considered in this study. Modulation of P-glycoprotein-mediated transport has significant pharmacokinetic implications for P-glycoprotein substrates, which might have therapeutic benefits or create contraindications. As a result, this study indicates that none of the molecules will inhibit P-glycoprotein I and II, with the exception of S-Adenosylhomocysteine which will be an inhibitor of P-glycoprotein I. Furthermore, it may be predicted whether a certain substance will be skin permeable. If a chemical has a $\log K_p > -2.5$, it is regarded to have low skin permeability, meaning that all five drug may be useful in the development of transdermal medication administration Pires et al. (2015). The total dose of a drug requires a certain volume to be uniformly distributed in blood plasma known as VDss. The drug will be more distributed in the tissue rather than in the plasma for higher VDss. From **Table 3**, low values of VDss are found for the five drugs. The efficacy of a given drug may be affected by the degree to which it binds proteins within the blood. The Fraction Unbound predicts the fraction that will be unbound in plasma

TABLE 1 | Protein-ligand complex binding energy and interaction details.

Protein ID	Ligand name	Binding affinity (Kcal/mol)	Number of hydrogen bonds	Residues forming hydrogen bonds
1B8M	S-Adenosylhomocysteine	-6.0	7	THR-83, GLN-84, SER-21, ALA-19, CYS-119
	Carbenicillin	-7.4	8	THR-59, THR-83, GLN-84, TYR-86, SER-108, THR-117
	Apiin	-7.2	10	SER-21, ALA-19, GLN-84, SER-108, THR-83, CYS-111 VAL-16, CYS-17
	Rutine	-6.9	11	SER-17, CYS-13, TYR, 54, THR-56, GLN-94, LYS-93 CYS-13, ASP-14
	Chloramphenicol Monoglucoronide	-7.0	10	GLU-9, SER-11, ASP-14, SER-15, LYS-93, TYR-96
1GS9	S-Adenosylhomocysteine	-5.7	6	GLY-23, GLU-27, ASP-36, ASP-153, GLN-156
	Carbenicillin	-6.8	3	Trp-34, ASP-35, GLN-156
	Apiin	-6.0	6	GLU-27, ARG-145, GLN-156
	Rutine	-6.6	4	VAL-103, SER-104, ARG-108, THR-188, HIS-222 GLU-245, ILE-246
	Chloramphenicol Monoglucoronide	-6.9	3	GLU-27, ARG-145, GLN-156
3PFQ	S-Adenosylhomocysteine	-6.4	9	VAL-103, SER-104, ARG-108, THR-188, HIS-222 GLU-245, ILE-246
	Carbenicillin	-8.6	7	VAL-103, SER-104, ARG-108, PRO-244, GLU-245 ARG-652
	Apiin	-9.0	9	SER-104, THR-188, LYS-239, PRO-244, GLU-245 ILE-246, ARG-652
	Rutine	-9.5	8	SER-67, SER-102, VAL-103, ARG-108, THR-188, LYS-654 LYS-654, ARG-657, ARG-659
	Chloramphenicol Monoglucoronide	-8.6	9	VAL-103, SER-104, ARG-108, GLU-184, HIS-222, ILE-246 ARG-652
5A63	S-Adenosylhomocysteine	-7.7	7	LEU-348, TYR-422, VAL-423, GLY-426, ASP-427, ASP-470
	Carbenicillin	-7.2	2	TYR-422, MET-473
	Apiin	-9.0	8	TYR-123, GLN-128, HIS-140, TYR-422, ASN-424, SER-476
	Rutine	-7.8	8	LEU-348, GLY-351, GLY-354, GLY-426, ASP-427, ASP-470
	Chloramphenicol Monoglucoronide	-7.4	5	TYR-123, GLN-128, GLY-129, LYS-141, SER-476
4L7G	S-Adenosylhomocysteine	-7.8	5	SER-35, ASN-37, TYR-71, ILE-126
	Carbenicillin	-7.4	5	GLY-11, TYR-71, THR-232
	Apiin	-7.4	4	GLY-11, SER-35, SER-36, ASN-37, TYR-71, ILE-126 TYR-198, THR-232
	Rutine	-9.2	8	ASP-32, SER-35, ASN-37, ALA-39, TYR-198, LYS-224 THR-231, ARG-235
	Chloramphenicol Monoglucoronide	-10.7	10	GLY-11, SER-35, TYR-71, THR-72, GLN-73, GLY-230 THR-232

resulting in the values shown in **Table 3**. A drug's ability to cross into the brain is a significant descriptor because it will be able to contribute to the reduction of toxicities and side effects, and is evaluated through the Blood-Brain Permeability parameter. For a given potential therapeutic drug, a logBBB > -0.3 value is estimated to readily cross the blood-brain barrier while molecules with logBBB > -1 will be badly distributed to the brain. The CNS Permeability is another measurement having low values which indicates that these drugs cannot penetrate the Central Nervous System (CNS) Pires et al. (2015). Cytochrome P450 is an important detoxification enzyme in the body, mostly present in the liver, since it oxidizes xenobiotics to enhance excretion Pires et al. (2015). **Table 3** shows that none of the studied molecules will be inhibitors or substrates of any P450 cytochrome isoform. Drug clearance happens as a combination of renal and hepatic clearance, and is associated with bioavailability; consequently, it is important for determining dosing rates. The AMES toxicity test utilizes microbes in order to ascertain a compound's mutagenesis potential. A positive test shows that

the substance is mutagenic; therefore, it could result in cancer. The predictions are negative for all the molecules with the exception of Rutine. The main causes of acquiring long QT syndrome are the blocking of the potassium channels encoded by hERG (the human Ether-a-go-go-Related Gene), which leads to fatal ventricular arrhythmia. The predictions indicate that none of the molecules will be an hERG inhibitor, but Apiin and Rutine will be hERG II inhibitors. The lethal dosage value (LD50) can be assessed in terms of the ORAT (Oral Rat Acute Toxicity) and the ORCT (Oral Rat Chronic Toxicity) parameters. Drug-induced liver injury is a major safety concern for drug development and a significant cause of drug attrition. Thus, Hepatotoxicity is related to the disruption of the normal liver function and the predictions for Apiin, Rutine and Chloramphenicol Monoglucoronide are negative. Skin Sensitization is predicted negative in all cases. *T. Pyriformis* is a protozoa bacteria whose toxicity is frequently applied as a toxic endpoint. A forecasted value > -0.5 for a given compound is considered toxic Pires et al. (2015).

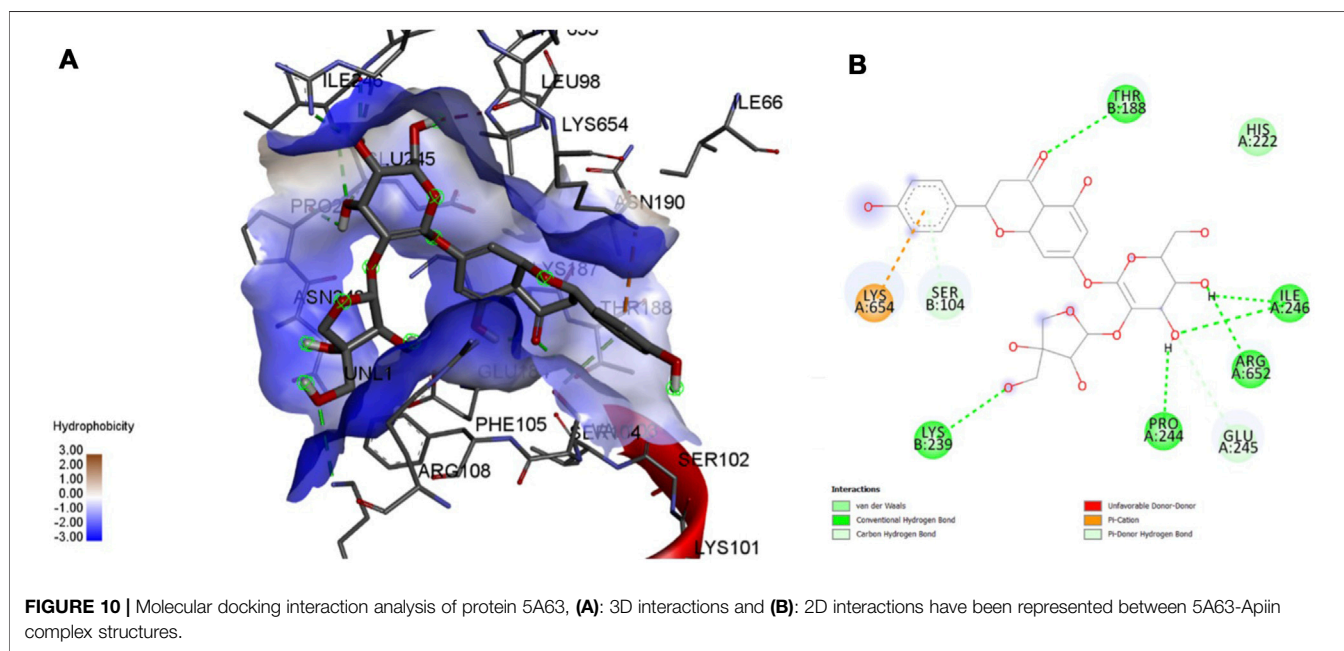
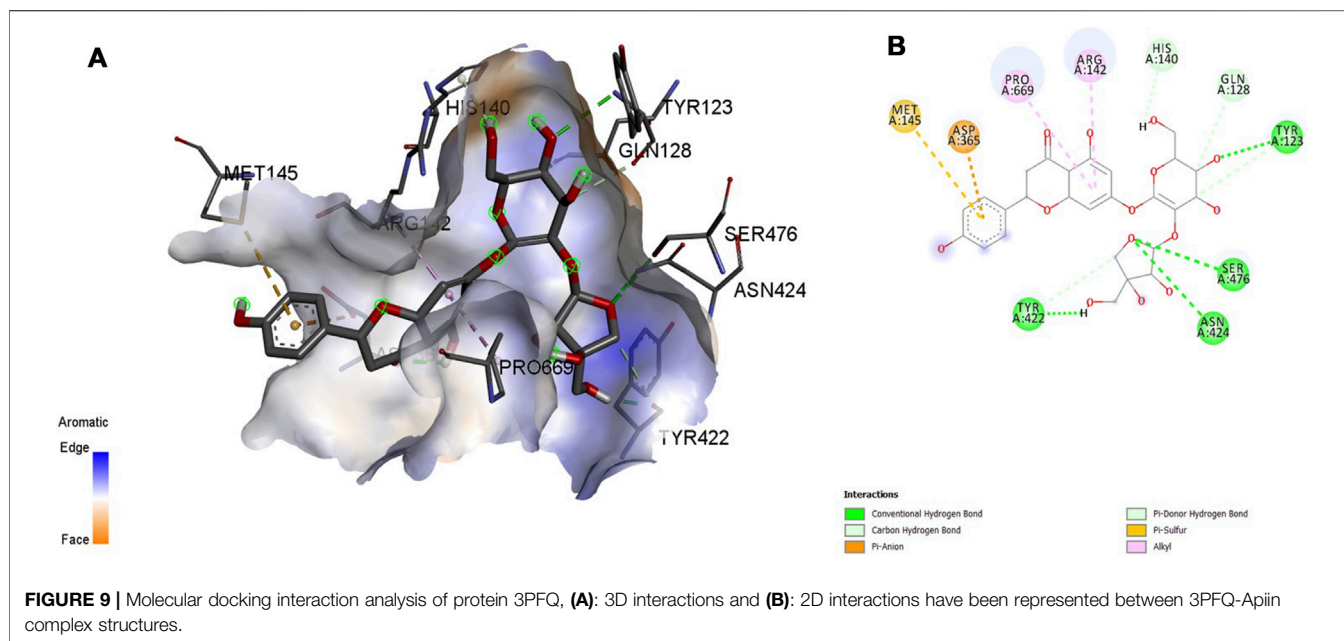


3.7 Conceptual DFT Studies

The calculated global reactivity descriptors: Electronegativity (χ), Hardness (η), Electrophilicity (ω) (all in eV), Softness (S), Nucleophilicity (N), Electrodonating Power (ω^-), Electroaccepting Power (ω^+) and Net Electrophilicity ($\Delta\omega^\pm$) Parr and Yang (1989); Chermette (1999); Geerlings et al. (2003, 2020); Toro-Labbé (2007); Chattaraj (2009); Chakraborty and Chattaraj (2021), estimated following the methodology presented in the 2.7 subsection together with the in-house developed CDFT software tool are displayed in **Table 4**.

As the global hardness η can be regarded as a direct measure of the deformation of the electron density and of the chemical

reactivity being related to the HOMO-LUMO gap, it can be seen that Carbenicillin will be the less reactive ligand being the others very similar in their reactivity. The electrodonating ability ω^- is more important than its electroaccepting power ω^+ for all the ligands because of their molecular structures. However, after a comparison of the values of ω^- and ω^+ for each molecule, it can be concluded that the reactivity of the Chloramphenicol monoglucuronide will be very different from the other ligands. The electrophilicity ω index encompasses the equilibrium between an electrophile's tendency to acquire extra electron density and a molecule's resistance to exchanging electron density with the environment Domingo et al. (2016). By



studying the electrophilicities of a series of reagents involved in Diels-Alder reactions Domingo et al. (2002); Domingo and Sáez (2009); Pérez et al. (2003), an electrophilicity ω scale for the classification of organic molecules as strong, moderate or marginal electrophiles was proposed being $\omega > 1.5$ eV for the first case, $0.8 < \omega < 1.5$ eV for the second case and $\omega < 0.8$ eV for the last case Domingo et al. (2002); Domingo and Sáez (2009); Pérez et al. (2003). By inspection of **Table 4**, it can be said that with the exception of Carbenicillin all the ligands may be regarded as strong electrophiles.

Besides the global reactivity descriptors, their local counterparts have been developed to get an idea of the differences in chemical reactivity between the atoms within the molecule. Among these local reactivity descriptors are the Fukui functions Parr and Yang (1989); Chermette (1999); Geerlings et al. (2003) and the Dual Descriptor Toro-Labbé (2007); Morell et al. (2005, 2006); Martínez-Araya (2012a); Martínez-Araya (2012b); Martínez-Araya (2015), which have been defined as: Nucleophilic Fukui Function (NFF) = $f^+(\mathbf{r}) = \rho_{N+1}(\mathbf{r}) - \rho_N(\mathbf{r})$, Electrophilic Fukui Function (EFF) = $f^-(\mathbf{r}) = \rho_N(\mathbf{r}) - \rho_{N-1}(\mathbf{r})$, and

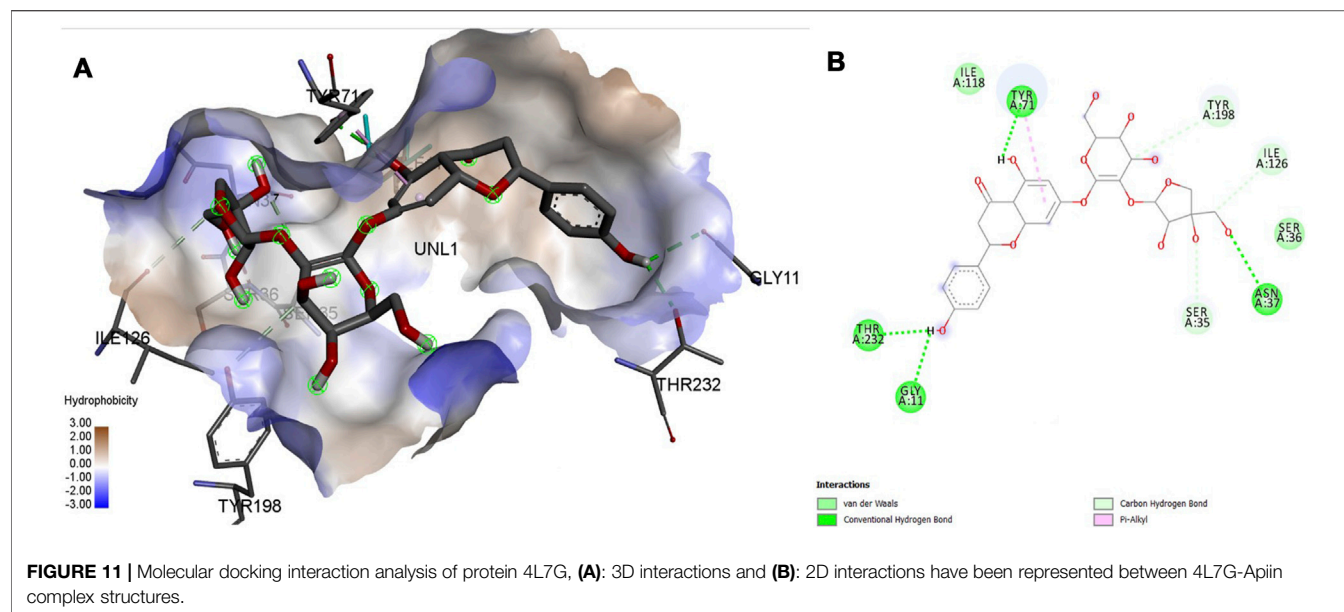


TABLE 2 | Bioactivity scores of the studied molecules calculated on the basis of the GPCR ligand, ion channel modulator, nuclear receptor ligand, kinase inhibitor, protease inhibitor, and enzyme inhibitor interactions.

Property	S-adenosyl homocysteine	Carbenicillin	Apiin	Rutine	Chloramphenicol monoglucuronide
GPCR Ligand	1.04	0.05	0.18	-0.05	0.08
Ion Channel Modulator	0.44	-0.40	-0.17	-0.52	-0.06
Nuclear Receptor Ligand	0.47	-0.75	0.09	-0.14	-0.28
Kinase Inhibitor	-1.18	-0.37	0.18	-0.23	-0.12
Protease Inhibitor	0.51	0.85	0.17	-0.07	0.08
Enzyme Inhibitor	1.23	0.30	0.42	0.12	0.28

Dual Descriptor (DD) = $\Delta f(\mathbf{r}) = (\partial f(\mathbf{r})/\partial N)_{v(\mathbf{r})}$, relating the electronic densities of the neutral, positive and negative species.

The NFF, $f^+(\mathbf{r})$, is associated with the sites within a molecular system which are prone to nucleophilic attacks while the EFF, $f^-(\mathbf{r})$, describes those sites that are more susceptible to electrophilic attacks. Although the NFF and the EFF have been used successfully for the identification of reactive sites, the Dual Descriptor $\Delta f(\mathbf{r})$ or DD, has been shown to describe unambiguously nucleophilic and electrophilic sites within a molecule Martínez-Araya (2015). Graphical representations of the DD for the five studied ligands is displayed in Figure 12 showing the zones where DD > 0 and DD < 0:

Although there is some overlap between the different regions within the ligands, these graphical representations allow to clearly distinguish the regions within the molecules where the Dual Descriptor will be greater or smaller than zero, implying the differences in their chemical reactivities.

4 DISCUSSION

Alzheimer's disease (AD) is a chronic neurodegenerative disorder characterized by the progressive impairment of

memory, cognition and behavior that usually exhibits a slow onset before worsening over time and ultimately leading to death. The causes of AD are poorly understood, although several etiological factors, such as genetic abnormalities, history of head injuries, environmental factors, general lifestyles, depression or hypertension, deposition of extracellular β -amyloid protein ($A\beta$) and microtubule associated tau protein in the brain, and cholinergic dysfunction have all been implicated in AD. At present, there are no drugs available that are capable of curing Alzheimer's disease or any of the other common types of dementia, but two conceptual approaches for the treatment of AD have been developed. Currently, only three cholinesterase inhibitors such as donepezil, galantamine and rivastigmine are the Food and Drug Administration (FDA) approved drugs to treat AD. Unfortunately, they only work for a short period of time, primarily in the early stages of the illness, to help patients delay the loss of cognitive functions as much as possible.

In this study, the HAp nanoparticles were selected as the drug delivery system with precise targeting. In general, nanoparticles are divided into two types: inorganic (metallic, metal oxide, and ceramic particles) and organic (organic, metal oxide, and

TABLE 3 | ADMET properties of the five studied ligands.

Property	S-adenosyl homocystein	Carbenicillin	Apiin	Rutine	Chloramphenicol monoglucoronide
Absorption					
Caco-2 Permeability	-0.506	0.377	0.737	-0.763	-0.868
Intestinal Absorption	27.464	23.953	29.350	28.135	0.000
Skin Permeability	-2.735	-2.735	-2.735	-2.735	-2.735
P-glycoprotein Substrate	Yes	Yes	Yes	Yes	Yes
P-glycoprotein I Inhibitor	Yes	No	No	No	No
P-glycoprotein II Inhibitor	No	No	No	No	No
Distribution					
VDss	-0.575	-1.804	-0.108	0.013	-2.144
Fraction Unbound	0.559	0.427	0.175	0.292	0.427
BBB Permeability	-1.630	-1.051	-1.523	-2.080	-5.744
CNS Permeability	-4.090	-3.572	-5.144	-5.744	-4.556
Metabolism					
CYP2D6 Substrate	No	No	No	No	No
CYP3A4 Substrate	No	No	No	No	No
CYP1A2 Inhibitor	No	No	No	No	No
CYP2C19 Inhibitor	No	No	No	No	No
CYP2C9 Inhibitor	No	No	No	No	No
CYP2D6 Inhibitor	No	No	No	No	No
CYP3A4 Inhibitor	No	No	No	No	No
Excretion					
Total Clearance	0.721	0.081	0.117	-0.200	0.373
Renal OCT2 Substrate	No	No	No	No	No
Toxicity					
AMES Toxicity	No	No	No	Yes	No
Maximum Tolerated Dose	0.514	1.717	0.516	0.427	0.759
hERG I Inhibitor	No	No	No	No	No
hERG II Inhibitor	No	No	Yes	Yes	No
Oral Rat Acute Toxicity	2.403	1.856	2.417	2.445	2.439
Oral Rat Chronic Toxicity	2.771	2.956	5.414	5.414	5.158
Hepatotoxicity	Yes	Yes	No	No	No
Skin Sensitisation	No	No	No	No	No
<i>T. Pyriformis</i> Toxicity	0.285	0.285	0.285	0.2.85	0.2.85

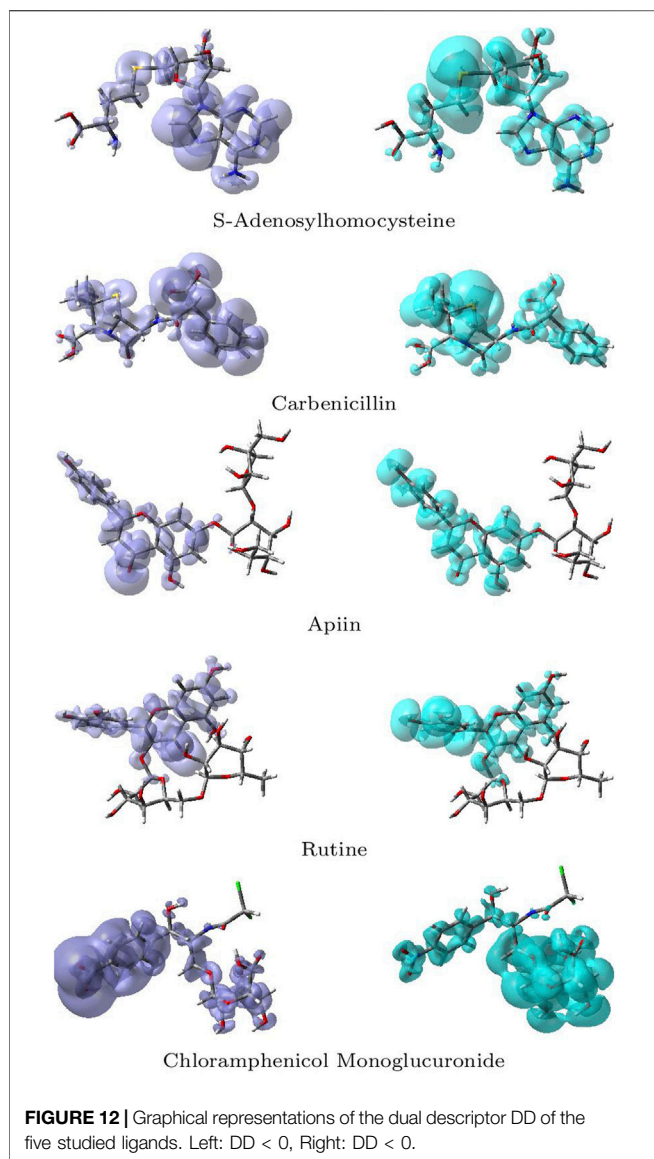
TABLE 4 | Global Reactivity Descriptors of the Five Studied Ligands: Electronegativity (χ), Hardness (η), Electrophilicity (ω) (all in eV), Softness (S) (in eV⁻¹), Nucleophilicity (N), Electrodonating Power (ω^-), Electroaccepting Power (ω^+) and Net Electrophilicity ($\Delta\omega^+$) (also in eV).

Molecule	χ	H	ω	S	N	ω^-	ω^+	$\Delta\omega^+$
S-Adenosylhomocysteine	3.8897	4.2687	1.7722	0.2343	2.7684	5.7561	1.8664	7.6225
Carbenicillin	3.8718	5.6238	1.3328	0.1778	2.1088	4.9530	1.0812	6.0341
Apiin	4.1168	4.6227	1.8332	0.2163	2.3643	6.0136	1.8968	7.9105
Rutine	4.1346	4.1184	2.0754	0.2428	2.5986	6.4756	2.3410	8.8166
Chloramphenicol Monoglucoronide	5.2081	4.2387	3.1996	0.2359	2.8369	9.2682	4.0601	13.3283

ceramic particles) (lipidic and polymeric particles). The metallic NPs have some limitations over organic NPs because of the presence of metals, but few metallic NPs like gold, selenium and cerium NPs are reported to exhibit significant anti-AD properties. Recently in a study, the solid lipid NPs was shown to have significant inhibitory effects against amyloid aggregation Sathya et al. (2020). The selenium based NPs were seen to reduce the ROS level in the brain which is a key strategy to relieve AD because of the presence of many trace elements such as sodium selenite (VI), sodium selenite (IV) and selenium

selenite (II) Fernandes and Gandin (2015). In the AD mouse model, the cerium NPs coupled with triphenylphosphonium (TPP) was seen to localize in the mitochondria to prevent the neuronal death Kwon et al. (2016). In another study, the gold NPs (AuNPs) showed significant results in reducing the symptoms of AD by modulating the mitochondrial functions dos Santos Tramontin et al. (2019).

When compared to the above mentioned nanoparticles AC-HAp NPs are not only bioactive but also non-toxic and non-immunogenic and do not contain any toxic elements Yasukawa



et al. (1994). The AC-HAP NPs exhibit improved densification and better bioactivity than pure HAP NPs. None of the above discussed NPs were checked for their AChE inhibitory activity but in our present study, *A. calamus* rhizome with revitalizing neurological properties has been used to produce the HAP NPs and their AChE inhibitory activity was evaluated and has shown a promising AChE inhibitory action.

Inhibition of AChE, the key enzyme in the breakdown of acetylcholine, is considered one of the treatment strategies against Alzheimer's disease. Plants have been traditionally used to enhance cognitive function and to alleviate other symptoms associated nowadays with Alzheimer's disease. The AC-HAP NPs drastically enhanced the AChE inhibitory action even at very low concentrations when compared to pure HAP NPs. The IC₅₀ values of 206.31 and 22.39 µg/ml were recorded with HAP NPs and AC-HAP NPs (Figure 6). Tacrine was used as a positive control which

showed IC₅₀ of 96.43 µg/ml. Surprisingly, significant AChE inhibition activity results was observed in *A. calamus* mediated HAP NPs, suggesting that the preparation of HAP NPs from *A. calamus* rhizome extract enhanced the AChE inhibition. Similar observations were made by Uddin et al. (2021) where the extracts of *Blumea lacera*, *Cyclea barbata*, *Smilax guianensis* and *Byttneria Pilosa* inhibited AChE with an IC₅₀ values of 150 ± 11, 176 ± 14, 205 ± 31, and 221 ± 2 µg/ml respectively Uddin et al. (2021). They also proved that the plant extracts selected in their study showed a promising effect in inhibiting AChE activity, wherein the present study has shown improvement in biological activities in the medicinal plants when in combination with nanoparticle.

In the Molecular Docking studies, the strength of protein-ligand complex binding is well known as binding affinity. The affinity determines if the ligand binds to the target. Further, amongst the 20 phytochemicals screened, 5 compounds exhibited highest binding affinity and lowest binding energy values for the selected target proteins of AD, such as 1B8M, 1GS9, 3PFQ, 4L7G, and 5A63. The binding energy for all the target proteins was in the range of -5.7 to -10.7 kcal/mol with the formation of at least 6 to 11 hydrogen bonds. Depending upon the obtained binding energy, bonded and non-bonded interactions between the targets and 5 ligands (S-Adenosylhomocysteine, Carbenicillin, Apiin, Rutine and Chloramphenicol Monoglucuronide) the present study concludes that *A. calamus* phytochemicals have an effective anti-neurodegenerative activity.

5 CONCLUSION

A. calamus rhizome extract was used to successfully produce hydroxyapatite nanoparticles (AC-HAP NPs). The formation of nanoparticles was confirmed by SEM, EDX, XRD, TEM, HR-TEM and SAED techniques. The formation of AC-HAP NPs with high crystallinity and well defined forms was demonstrated by XRD, SEM, and TEM analysis. The goal of this study was to find phytochemicals that can bind to the critical targets of amyloid hypothesis of AD using a computational approach and also to check the AChE inhibition activity of the synthesized AC-HAP NPs. The findings of the present study shows that as-prepared AC-HAP NPs can inhibit AChE, which was compared with pure AC-HAP NPs. *In silico* molecular docking approach revealed that most of the compounds derived from *A. calamus* rhizome extract have the ability to bind to the selected targets, according to the binding scores and analysis of the interactions of the compounds. Further, *in vivo* studies to evaluate substances like S-Adenosylhomocysteine, Carbenicillin, Apiin, Rutine and Chloramphenicol Monoglucuronide would lead to therapeutically effective molecules for treating a variety of chronic pain problems. It is also suggested that multipurpose NPs with multitherapeutic capabilities can be used. Given the present medications' major targets of tau proteins, neuroinflammation, and Aβ proteins, there is an urgent need to create drugs with novel targets that

can not only treat the symptoms but also prevent the disease from progressing at an early stage, resulting in a better life.

With the additional goal of analyzing their bioactivities, the predicted biological targets and the ADMET parameters related to the bioavailability and computational pharmacokinetics of the five ligands have been reported. The chemical reactivities of these five ligands have been thoroughly investigated through the optimization of their structures using the DFTBA methodology and the estimation of their electronic properties using the MN12SX/Def2TZVP/H₂O model chemistry, which has already been used in previous research for the study of potentially therapeutic molecules, proving its suitability for this type of calculation and supporting this and previous research on this important subject.

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 authors.

AUTHOR CONTRIBUTIONS

SP: Research; AJ: Research; CD: Research; SKP: Research; NA: Research; RP: Research; RA: Research; CS: Research; AS: Research; AE: Research; AA: Research; JO-C: Research; JF: Research; NF-H: Research; CS: Research and Writing the

article; SPK: Research and Writing the article; DG-M: Research and Writing the article.

FUNDING

This work was funded by Researchers Supporting Project No. RSP-2021/31, King Saud University, Riyadh, Saudi Arabia.

ACKNOWLEDGMENTS

Authors thank the Director and Head, JSS Academy of Higher Education and Research (JSSAHER), Mysuru, India for the support and infrastructure facilities. SPK is grateful to the Director, Amrita Vishwa Vidyapeetham, Mysuru campus for infrastructure support. The authors extend their appreciation to the Researchers Supporting Project Number RSP-2021/31, King Saud University, Riyadh, Saudi Arabia. NFH and DGM are researchers of CIMAV and CONACYT and want to thank both institutions for partial support.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fchem.2021.741037/full#supplementary-material>

REFERENCES

- Altinoglu, G., and Adali, T. (2020). Alzheimer's Disease Targeted Nano-Based Drug Delivery Systems. *Curr. Drug Targets* 21, 628–646. doi:10.2174/1389450120666191118123151
- Ankegowda, V. M., Kollur, S. P., Prasad, S. K., Pradeep, S., Dhramashekara, C., Jain, A. S., et al. (2020). Phyto-Mediated Synthesis of Silver Nanoparticles Using Terminalia Chebula Fruit Extract and Evaluation of its Cytotoxic and Antimicrobial Potential. *Molecules* 25, 5042. doi:10.3390/molecules25215042
- Avinash, K. O., Pradeep, S., Shivamallu, C., Gopenath, T. S., Kumar, M. N. K., Kanthesh, B. M., et al. (2021). In Silico Screened Flavanoids of Glycyrrhiza Glabra Inhibit CPLA2 and SPLA2 in LPS Stimulated Macrophages. *Bull. Environ. Pharmacol. Life Sci.* 10, 24.
- Blennow, K., de Leon, M. J., and Zetterberg, H. (2006). Alzheimer's Disease. *Lancet* 368, 387–403. doi:10.1016/s0140-6736(06)69113-7
- Chakraborty, D., and Chattaraj, P. K. (2021). Conceptual Density Functional Theory Based Electronic Structure Principles. *Chem. Sci.* 12, 6264–6279. doi:10.1039/d0sc07017c
- Chattaraj, P. K. (2009). *Chemical Reactivity Theory - A Density Functional View*. Boca Raton, FL: CRC Press. Taylor & Francis Group.
- Chermette, H. (1999). Chemical Reactivity Indexes in Density Functional Theory. *J. Comput. Chem.* 20, 129–154. doi:10.1002/(sici)1096-987x(19990115)20:1<129:aid-jcc13>3.0.co;2-a
- Cramer, C. (2004). *Essentials of Computational Chemistry - Theories and Models*. 2nd edn. Chichester, England: John Wiley & Sons.
- 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. doi:10.1038/srep42717
- Daina, A., Michielin, O., and Zoete, V. (2019). SwissTargetPrediction: Updated Data and New Features for Efficient Prediction of Protein Targets of Small Molecules. *Nucleic Acids Res.* 47, W357–W364. doi:10.1093/nar/gkz382
- Davies, P. (1999). Challenging the Cholinergic Hypothesis in Alzheimer Disease. *JAMA* 281, 1433. doi:10.1001/jama.281.15.1433
- Desikan, R. S., Cabral, H. J., Hess, C. P., Dillon, W. P., Glastonbury, C. M., Weiner, M. W., et al. (2009). Automated MRI Measures Identify Individuals with Mild Cognitive Impairment and Alzheimer's Disease. *Brain* 132, 2048–2057. doi:10.1093/brain/awp123
- Domingo, L. R., Aurell, M. J., Pérez, P., and Contreras, R. (2002). Quantitative Characterization of the Global Electrophilicity Power of Common diene/Dienophile Pairs in Diels-Alder Reactions. *Tetrahedron* 58, 4417–4423. doi:10.1016/s0040-4020(02)00410-6
- Domingo, L., Ríos-Gutiérrez, M., and Pérez, P. (2016). Applications of the Conceptual Density Functional Theory Indices to Organic Chemistry Reactivity. *Molecules* 21, 748. doi:10.3390/molecules21060748
- Domingo, L. R., and Sáez, J. A. (2009). Understanding the Mechanism of Polar Diels-Alder Reactions. *Org. Biomol. Chem.* 7, 3576–3583. doi:10.1039/b909611f
- dos Santos Tramontin, N., da Silva, S., Arruda, R., Ugioni, K. S., Canteiro, P. B., de Bem Silveira, G., et al. (2019). Gold Nanoparticles Treatment Reverses Brain Damage in Alzheimer's Disease Model. *Mol. Neurobiol.* 57, 926–936. doi:10.1007/s12035-019-01780-w
- Fakhoury, M., Takechi, R., and Al-Salami, H. (2015). Drug Permeation across the Blood-Brain Barrier: Applications of Nanotechnology. *BJMMR* 6, 547–556. doi:10.9734/bjmmr/2015/15493
- Fernandes, A. P., and Gandin, V. (2015). Selenium Compounds as Therapeutic Agents in Cancer. *Biochim. Biophys. Acta Gen. Subjects* 1850, 1642–1660. doi:10.1016/j.bbagen.2014.10.008
- Flores-Holguín, N., Frau, J., and Glossman-Mitnik, D. (2020c). A Fast and Simple Evaluation of the Chemical Reactivity Properties of the Pristinamycin Family of Antimicrobial Peptides. *Chem. Phys. Lett.* 739, 137021. doi:10.1016/j.cplett.2019.137021

- Flores-Holguín, N., Frau, J., and Glossman-Mitnik, D. (2019a). Chemical-Reactivity Properties, Drug Likeness, and Bioactivity Scores of Seragamides A-F Anticancer Marine Peptides: Conceptual Density Functional Theory Viewpoint. *Computation* 7, 52. doi:10.3390/computation7030052
- Flores-Holguín, N., Frau, J., and Glossman-Mitnik, D. (2019b). Computational Peptidology Assisted by Conceptual Density Functional Theory for the Study of Five New Antifungal Tripeptides. *ACS Omega* 4, 12555–12560. doi:10.1021/acsomega.9b01463
- Flores-Holguín, N., Frau, J., and Glossman-Mitnik, D. (2019c). Computational Prediction of Bioactivity Scores and Chemical Reactivity Properties of the Parasin I Therapeutic Peptide of Marine Origin through the Calculation of Global and Local Conceptual DFT Descriptors. *Theor. Chem. Acc.* 138, 1–9. doi:10.1007/s00214-019-2469-3
- Flores-Holguín, N., Frau, J., and Glossman-Mitnik, D. (2019d). “Conceptual DFT as a Helpful Chemoinformatics Tool for the Study of the Clavanin Family of Antimicrobial Marine Peptides,” in *Density Functional Theory Calculations* (Rijetia: IntechOpen), 1–11.
- Flores-Holguín, N., Frau, J., and Glossman-Mitnik, D. (2021). “Conceptual DFT as a Helpful Chemoinformatics Tool for the Study of the Clavanin Family of Antimicrobial Marine Peptides,” in *Density Functional Theory*. Editors S. R. De Lazaro, L. H. Da Silveira Lacerda, and R. A. Pontes Ribeiro (London, UK: IntechOpen), 57–67. doi:10.5772/intechopen.88657
- Flores-Holguín, N., Frau, J., and Glossman-Mitnik, D. (2020a). Conceptual DFT-Based Computational Peptidology of Marine Natural Compounds: Discodermins A-H. *Molecules* 25, 4158. doi:10.3390/molecules25184158
- Flores-Holguín, N., Frau, J., and Glossman-Mitnik, D. (2020b). Virtual Screening of Marine Natural Compounds by Means of Chemoinformatics and CDFT-Based Computational Peptidology. *Mar. Drugs* 18, 478. doi:10.3390/md18090478
- Förstl, H., and Kurz, A. (1999). Clinical Features of Alzheimer’s Disease. *Eur. Arch. Psychiatry Clin. Neurosci.* 249, 288–290. doi:10.1007/s004060050101
- Francis, P. T., Palmer, A. M., Snape, M., and Wilcock, G. K. (1999). The Cholinergic Hypothesis of Alzheimer’s Disease: a Review of Progress. *J. Neurol. Neurosurg. Psychiatry* 66, 137–147. doi:10.1136/jnnp.66.2.137
- Frau, J., Flores-Holguín, N., and Glossman-Mitnik, D. (2019). Chemical Reactivity Theory and Empirical Bioactivity Scores as Computational Peptidology Alternative Tools for the Study of Two Anticancer Peptides of Marine Origin. *Molecules* 24, 1115. doi:10.3390/molecules24061115
- Frau, J., and Glossman-Mitnik, D. (2018a). Blue M2: An Intermediate Melanoidin Studied via Conceptual DFT. *J. Mol. Model.* 24, 1–13. doi:10.1007/s00894-018-3673-0
- Frau, J., and Glossman-Mitnik, D. (2018b). Chemical Reactivity Theory Applied to the Calculation of the Local Reactivity Descriptors of a Colored Maillard Reaction Product. *Chem. Sci. Int. J.* 22, 1–14. doi:10.9734/csji/2018/41452
- Frau, J., and Glossman-Mitnik, D. (2018c). Computational Study of the Chemical Reactivity of the Blue-M1 Intermediate Melanoidin. *Comput. Theor. Chem.* 1134, 22–29. doi:10.1016/j.comptc.2018.04.018
- Frau, J., and Glossman-Mitnik, D. (2018e). Conceptual DFT Study of the Local Chemical Reactivity of the Dilylsdipyrrolones A and B Intermediate Melanoidins. *Theor. Chem. Acc.* 137, 1210. doi:10.1007/s00214-018-2244-x
- Frau, J., and Glossman-Mitnik, D. (2018d). Local Molecular Reactivity of the Colored Dansylglycine in Water and Dioxane Studied through Conceptual DFT. *J. Chem.* 2018, 1–7. doi:10.1155/2018/3172412
- Frau, J., and Glossman-Mitnik, D. (2018f). Molecular Reactivity and Absorption Properties of Melanoidin Blue-G1 through Conceptual DFT. *Molecules* 23, 559–615. doi:10.3390/molecules23030559
- Frisch, M. J., Trucks, G. W., Schlegel, H. B., Scuseria, G. E., Robb, M. A., Cheeseman, J. R., et al. (2016). *Gaussian 16 Revision C.01*. [Dataset]. Wallingford CT: Gaussian Inc.
- Geerlings, P., Chamorro, E., Chattaraj, P. K., De Proft, F., Gázquez, J. L., Liu, S., et al. (2020). Conceptual Density Functional Theory: Status, Prospects, Issues. *Theor. Chem. Acc.* 139, 36. doi:10.1007/s00214-020-2546-7
- Geerlings, P., De Proft, F., and Langenaeker, W. (2003). Conceptual Density Functional Theory. *Chem. Rev.* 103, 1793–1874. doi:10.1021/cr990029p
- Goedert, M., Spillantini, M. G., and Crowther, R. A. (1991). Tau Proteins and Neurofibrillary Degeneration. *Brain Pathol.* 1, 279–286. doi:10.1111/j.1750-3639.1991.tb00671.x
- Gopi, D., Bhuvaneshwari, N., Indira, J., Kanimozhi, K., and Kavitha, L. (2013). A Novel Green Template Assisted Synthesis of Hydroxyapatite Nanorods and Their Spectral Characterization. *Spectrochimica Acta A: Mol. Biomol. Spectrosc.* 107, 196–202. doi:10.1016/j.saa.2013.01.052
- Halgren, T. A. (1996a). Merck Molecular Force Field. I. Basis, Form, Scope, Parameterization, and Performance of MMFF94. *J. Comput. Chem.* 17, 490–519. doi:10.1002/(sici)1096-987x(199604)17:5/6<490:aid-jcc1>3.0.co;2-p
- Halgren, T. A. (1996b). Merck Molecular Force Field. II. MMFF94 van der Waals and Electrostatic Parameters for Intermolecular Interactions. *J. Comput. Chem.* 17, 520–552. doi:10.1002/(sici)1096-987x(199604)17:5/6<520:aid-jcc2>3.0.co;2-w
- Halgren, T. A. (1996c). Merck Molecular Force Field. V. Extension of MMFF94 Using Experimental Data, Additional Computational Data, and Empirical Rules. *J. Comput. Chem.* 17, 616–641. doi:10.1002/(sici)1096-987x(199604)17:5/6<616:aid-jcc5>3.0.co;2-x
- Halgren, T. A. (1999). MMFF VI. MMFF94s Option for Energy Minimization Studies. *J. Comput. Chem.* 20, 720–729. doi:10.1002/(sici)1096-987x(199905)20:7<720:aid-jcc7>3.0.co;2-x
- Halgren, T. A., and Nachbar, R. B. (1996). Merck Molecular Force Field. IV. Conformational Energies and Geometries for MMFF94. *J. Comput. Chem.* 17, 587–615. doi:10.1002/(sici)1096-987x(199604)17:5/6<587:aid-jcc4>3.0.co;2-q
- Heneka, M. T., Carson, M. J., Khoury, J. E., Landreth, G. E., Brosseron, F., Feinstein, D. L., et al. (2015). Neuroinflammation in Alzheimer’s Disease. *Lancet Neurol.* 14, 388–405. doi:10.1016/s1474-4422(15)70016-5
- H. Ferreira-Vieira, T., M. Guimaraes, I., R. Silva, F., and M. Ribeiro, F. (2016). Alzheimer’s Disease: Targeting the Cholinergic System. *Curr. Neuropharmacol.* 14, 101–115. doi:10.2174/1570159x13666150716165726
- Iqbal, K., Alonso Adel, C., Chen, S., Chohan, M. O., El-Akkad, E., Gong, C.-X., et al. (2005). Tau Pathology in Alzheimer Disease and Other Tauopathies. *Biochim. Biophys. Acta Mol. Basis Dis.* 1739, 198–210. doi:10.1016/j.bbdis.2004.09.008
- Jain, A. S., Sushma, P., Dharmashekar, C., Beelagi, M. S., Prasad, S. K., Shivamallu, C., et al. (2021). In Silico Evaluation of Flavonoids as Effective Antiviral Agents on the Spike Glycoprotein of SARS-CoV-2. *Saudi J. Biol. Sci.* 28, 1040–1051. doi:10.1016/j.sjbs.2020.11.049
- Jaiswal, P., Kumar, P., Singh, V. K., and Singh, D. K. (2011). Areca Catechu L.: A Valuable Herbal Medicine against Different Health Problems. *Res. J. Med. Plant* 5, 145–152. doi:10.3923/rjmp.2011.145.152
- Jensen, F. (2007). *Introduction to Computational Chemistry*. 2nd edn. Chichester, England: John Wiley & Sons.
- Jujavarapu, S. E., Dhagat, S., and Yadav, M. (2019). *Computer-Aided Design of Antimicrobial Lipopeptides as Prospective Drug Candidates*. Boca Raton: CRC Press LLC.
- Kametani, F., and Hasegawa, M. (2018). Reconsideration of Amyloid Hypothesis and Tau Hypothesis in Alzheimer’s Disease. *Front. Neurosci.* 12, 25. doi:10.3389/fnins.2018.00025
- Kim, J.-h., Hong, C.-O., Koo, Y.-c., Choi, H.-D., and Lee, K.-W. (2012). Anti-Glycation Effect of Gold Nanoparticles on Collagen. *Biol. Pharm. Bull.* 35, 260–264. doi:10.1248/bpb.35.260
- Kim, J. H. (2018). Genetics of Alzheimer’s Disease. *Dement Neurocognitive Disord.* 17, 131. doi:10.12779/dnd.2018.17.4.131
- Knop, K., Hoogenboom, R., Fischer, D., and Schubert, U. S. (2010). Poly(ethylene Glycol) in Drug Delivery: Pros and Cons as Well as Potential Alternatives. *Angew. Chem. Int. Edition* 49, 6288–6308. doi:10.1002/anie.200902672
- Kollur, S. P., Prasad, S. K., Pradeep, S., Veerapur, R., Patil, S. S., Amachawadi, R. G., et al. (2021). Luteolin-Fabricated ZnO Nanostructures Showed PLK-1 Mediated Anti-breast Cancer Activity. *Biomolecules* 11, 385. doi:10.3390/biom11030385
- Kwon, H. J., Cha, M.-Y., Kim, D., Kim, D. K., Soh, M., Shin, K., et al. (2016). Mitochondria-Targeting Ceria Nanoparticles as Antioxidants for Alzheimer’s Disease. *ACS Nano* 10, 2860–2870. doi:10.1021/acsnano.5b08045
- Lacor, P. N., Buniel, M. C., Furlow, P. W., Sanz Clemente, A., Velasco, P. T., Wood, M., et al. (2007). A Oligomer-Induced Aberrations in Synapse Composition, Shape, and Density Provide a Molecular Basis for Loss of Connectivity in Alzheimer’s Disease. *J. Neurosci.* 27, 796–807. doi:10.1523/jneurosci.3501-06.2007

- Leszek, J., Md Ashraf, G., Tse, W. H., Zhang, J., Gasiorowski, K., Avila-Rodriguez, M. F., et al. (2017). Nanotechnology for Alzheimer Disease. *Curr. Alzheimer Res.* 14, 1182–1189. doi:10.2174/1567205014666170203125008
- Lewars, E. (2003). *Computational Chemistry - Introduction to the Theory and Applications of Molecular and Quantum Mechanics*. Dordrecht: Kluwer Academic Publishers.
- Mahley, R. W., Weisgraber, K. H., and Huang, Y. (2006). Apolipoprotein E4: A Causative Factor and Therapeutic Target in Neuropathology, Including Alzheimer's Disease. *Proc. Natl. Acad. Sci.* 103, 5644–5651. doi:10.1073/pnas.0600549103
- Marenich, A. V., Cramer, C. J., and Truhlar, D. G. (2009). Universal Solvation Model Based on Solute Electron Density and on a Continuum Model of the Solvent Defined by the Bulk Dielectric Constant and Atomic Surface Tensions. *J. Phys. Chem. B* 113, 6378–6396. doi:10.1021/jp810292n
- Martínez-Araya, J. I. (2012a). Explaining Reaction Mechanisms Using the Dual Descriptor: A Complementary Tool to the Molecular Electrostatic Potential. *J. Mol. Model.* 19, 2715–2722. doi:10.1007/s00894-012-1520-2
- Martínez-Araya, J. I. (2012b). Revisiting Caffate's Capabilities as a Complexation Agent to Silver Cation in Mining Processes by Means of the Dual Descriptor-A Conceptual DFT Approach. *J. Mol. Model.* 18, 4299–4307. doi:10.1007/s00894-012-1405-4
- Martínez-Araya, J. I. (2015). Why Is the Dual Descriptor a More Accurate Local Reactivity Descriptor Than Fukui Functions? *J. Math. Chem.* 53, 451–465. doi:10.1007/s10910-014-0437-7
- Martorana, A., Esposito, Z., and Koch, G. (2010). Beyond the Cholinergic Hypothesis: Do Current Drugs Work in Alzheimer's Disease? *CNS Neurosci. Ther.* 16, 235–245. doi:10.1111/j.1755-5949.2010.00175.x
- Mayeux, R., and Stern, Y. (2012). Epidemiology of Alzheimer Disease. *Cold Spring Harbor Perspect. Med.* 2, a006239. doi:10.1101/cshperspect.a006239
- Mendez, M. F. (2006). The Accurate Diagnosis of Early-Onset Dementia. *Int. J. Psychiatry Med.* 36, 401–412. doi:10.2190/q6j4-r143-p630-kw41
- Morell, C., Grand, A., and Toro-Labbé, A. (2005). New Dual Descriptor for Chemical Reactivity. *J. Phys. Chem. A* 109, 205–212. doi:10.1021/jp046577a
- Morell, C., Grand, A., and Toro-Labbé, A. (2006). Theoretical Support for Using the $\Delta f(r)$ Descriptor. *Chem. Phys. Lett.* 425, 342–346. doi:10.1016/j.cplett.2006.05.003
- Mudher, A., and Lovestone, S. (2002). Alzheimer's Disease - Do Tauists and Baptists Finally Shake Hands?. *Trends Neurosciences* 25, 22–26. doi:10.1016/s0166-2236(00)02031-2
- Nikolaev, A., McLaughlin, T., O'Leary, D. D. M., and Tessier-Lavigne, M. (2009). APP Binds DR6 to Trigger Axon Pruning and Neuron Death via Distinct Caspases. *Nature* 457, 981–989. doi:10.1038/nature07767
- Obulesu, M., and Rao, D. M. (2011). Effect of Plant Extracts on Alzheimer's Disease: An Insight into Therapeutic Avenues. *J. Neurosci. Rural Pract.* 02, 056–061. doi:10.4103/0976-3147.80102
- Parr, R., and Yang, W. (1989). *Density-Functional Theory of Atoms and Molecules*. New York: Oxford University Press.
- Pérez, P., Domingo, L. R., José Aurell, M., and Contreras, R. (2003). Quantitative Characterization of the Global Electrophilicity Pattern of Some Reagents Involved in 1,3-Dipolar Cycloaddition Reactions. *Tetrahedron* 59, 3117–3125. doi:10.1016/s0040-4020(03)00374-0
- Peverati, R., and Truhlar, D. G. (2012). Screened-Exchange Density Functionals with Broad Accuracy for Chemistry and Solid-State Physics. *Phys. Chem. Chem. Phys.* 14, 16187–16191. doi:10.1039/c2cp42576a
- Pires, D. E. V., Blundell, T. L., and Ascher, D. B. (2015). pkCSM: Predicting Small-Molecule Pharmacokinetic and Toxicity Properties Using Graph-Based Signatures. *J. Med. Chem.* 58, 4066–4072. doi:10.1021/acs.jmedchem.5b00104
- Polvikoski, T., Sulkava, R., Haltia, M., Kainulainen, K., Vuorio, A., Verkkoniemi, A., et al. (1995). Apolipoprotein E, Dementia, and Cortical Deposition of β -Amyloid Protein. *N. Engl. J. Med.* 333, 1242–1248. doi:10.1056/nejm199511093331902
- Pradeep, S., Jain, A. S., Dharmashekara, C., Prasad, S. K., Kollur, S. P., Syed, A., et al. (2020). Alzheimer's Disease and Herbal Combination Therapy: A Comprehensive Review. *J. Alzheimers Dis. Rep.* 4, 417–429. doi:10.3233/adr-200228
- Prasad, A., Shruthi, G., Sushma, P., Jain, A. S., Chandan, D., Prasad, M. N. N., et al. (2020a). *Helicobacter pylori* Infection: A Bioinformatic Approach. *Int. J. Pharm. Sci. Res.* 17, 5469–5483. doi:10.13040/IJPSR.0975-8232.11(11).5469-83
- Prasad, K. S., Pillai, R. R., Ghimire, M. P., Ray, R., Richter, M., Shivamallu, C., et al. (2020b). Indole Moiety Induced Biological Potency in Pseudo-peptides Derived from 2-Amino-2-(1H-Indole-2-Yl) Based Acetamides: Chemical Synthesis, *In Vitro* Anticancer Activity and Theoretical Studies. *J. Mol. Struct.* 1217, 128445. doi:10.1016/j.molstruc.2020.128445
- Prasad, K. S., Pillai, R. R., Shivamallu, C., Prasad, S. K., Jain, A. S., Pradeep, S., et al. (2020c). Tumoricidal Potential of Novel Amino-1,10-Phenanthroline Derived Imine Ligands: Chemical Preparation, Structure, and Biological Investigations. *Molecules* 25, 2865. doi:10.3390/molecules25122865
- Prasad, S. K., Pradeep, S., Shimavallu, C., Kollur, S. P., Syed, A., Marraiki, N., et al. (2021). Evaluation of Annona Muricata Acetogenins as Potential Anti-SARS-CoV-2 Agents through Computational Approaches. *Front. Chem.* 8, 624716. doi:10.3389/fchem.2020.624716
- Sathya, S., Shanmuganathan, B., and Devi, K. P. (2020). Deciphering the Anti-apoptotic Potential of α -bisabolol Loaded Solid Lipid Nanoparticles against A β Induced Neurotoxicity in Neuro-2a Cells. *Colloids Surf. B: Biointerfaces* 190, 110948. doi:10.1016/j.colsurfb.2020.110948
- Schroeter, M. L., Stein, T., Maslowski, N., and Neumann, J. (2009). Neural Correlates of Alzheimer's Disease and Mild Cognitive Impairment: A Systematic and Quantitative Meta-Analysis Involving 1351 Patients. *NeuroImage* 47, 1196–1206. doi:10.1016/j.neuroimage.2009.05.037
- Selkoe, D. J. (1999). Translating Cell Biology into Therapeutic Advances in Alzheimer's Disease. *Nature* 399, A23–A31. doi:10.1038/399a023
- Singh, A. K., Gupta, A., Mishra, A. K., Gupta, V., Bansal, P., and Kumar, S. (2010). Medicinal Plant for Curing Alzheimer's Disease. *Int. J. Pharm. Biol. Arch.* 1, 108–114.
- Solfrizzi, V., Panza, F., Frisardi, V., Seripa, D., Logroscino, G., Imbimbo, B. P., et al. (2011). Diet and Alzheimer's Disease Risk Factors or Prevention: the Current Evidence. *Expert Rev. Neurother.* 11, 677–708. doi:10.1586/ern.11.56
- Spuch, C., Saida, O., and Navarro, C. (2012). Advances in the Treatment of Neurodegenerative Disorders Employing Nanoparticles. *Recent Pat. Drug Deliv. Formul.* 6, 2–18. doi:10.2174/187221112799219125
- Strittmatter, W. J., Saunders, A. M., Schmechel, D., Pericak-Vance, M., Enghild, J., Salvesen, G. S., et al. (1993). Apolipoprotein E: High-Avidity Binding to Beta-Amyloid and Increased Frequency of Type 4 Allele in Late-Onset Familial Alzheimer Disease. *Proc. Natl. Acad. Sci.* 90, 1977–1981. doi:10.1073/pnas.90.5.1977
- Tian, J., Shi, J., Zhang, X., and Wang, Y. (2010). Herbal Therapy: a New Pathway for the Treatment of Alzheimer's Disease. *Alzheimer's Res. Ther.* 2, 30. doi:10.1186/alzrt54
- Tiraboschi, P., Hansen, L. A., Thal, L. J., and Corey-Bloom, J. (2004). The Importance of Neuritic Plaques and Tangles to the Development and Evolution of AD. *Neurology* 62, 1984–1989. doi:10.1212/01.wnl.0000129697.01779.0a
- Toro-Labbé, A. (2007). *Theoretical Aspects of Chemical Reactivity*. Amsterdam: Elsevier Science.
- Turner, P. R., O'Connor, K., Tate, W. P., and Abraham, W. C. (2003). Roles of Amyloid Precursor Protein and its Fragments in Regulating Neural Activity, Plasticity and Memory. *Prog. Neurobiol.* 70, 1–32. doi:10.1016/s0301-0082(03)00089-3
- Uddin, M. J., Russo, D., Rahman, M. M., Uddin, S. B., Halim, M. A., Zidorn, C., et al. (2021). Anticholinesterase Activity of Eight Medicinal Plant Species: *In Vitro* and *In Silico* Studies in the Search for Therapeutic Agents against Alzheimer's Disease. *Evidence-Based Complement. Altern. Med.* 2021, 1–14. doi:10.1155/2021/9995614
- Uppar, V., Chandrashekarappa, S., Shivamallu, C., P, S., Kollur, S. P., Ortega-Castro, J., et al. (2021). Investigation of Antifungal Properties of Synthetic Dimethyl-4-Bromo-1-(Substituted Benzoyl) Pyrrolo[1,2-A] Quinoline-2,3-Dicarboxylates Analogues: Molecular Docking Studies and Conceptual DFT-Based Chemical Reactivity Descriptors and Pharmacokinetics Evaluation. *Molecules* 26, 2722. doi:10.3390/molecules26092722
- Waldemar, G., Dubois, B., Emre, M., Georges, J., McKeith, I. G., Rossor, M., et al. (2007). Recommendations for the Diagnosis and Management of Alzheimer's Disease and Other Disorders Associated with Dementia: EFNS Guideline. *Eur. J. Neurol.* 14, e1–e26. doi:10.1111/j.1468-1331.2006.01605.x
- Wang, H., Naghavi, M., Allen, C., Barber, R. M., Bhutta, Z. A., Carter, A., et al. (2016). Global, Regional, and National Life Expectancy, All-Cause Mortality,

- and Cause-specific Mortality for 249 Causes of Death, 1980-2015: A Systematic Analysis for the Global Burden of Disease Study 2015. *Lancet* 388, 1459–1544. doi:10.1016/S0140-6736(16)31012-1
- Waring, S. C., and Rosenberg, R. N. (2008). Genome-Wide Association Studies in Alzheimer Disease. *Arch. Neurol.* 65, 329–334. doi:10.1001/archneur.65.3.329
- Weigend, F. (2006). Accurate Coulomb-Fitting Basis Sets for H to Rn. *Phys. Chem. Chem. Phys.* 8, 1057–1065. doi:10.1039/b515623h
- Weigend, F., and Ahlrichs, R. (2005). Balanced Basis Sets of Split Valence, Triple Zeta Valence and Quadruple Zeta Valence Quality for H to Rn: Design and Assessment of Accuracy. *Phys. Chem. Chem. Phys.* 7, 3297–3305. doi:10.1039/b508541a
- Wilson, R. S., Barral, S., Lee, J. H., Leurgans, S. E., Foroud, T. M., Sweet, R. A., et al. (2011). Heritability of Different Forms of Memory in the Late Onset Alzheimer's Disease Family Study. *J. Alzheimers Dis.* 23, 249–255. doi:10.3233/jad-2010-101515
- Yasukawa, A., Takase, H., Kandori, K., and Ishikawa, T. (1994). Preparation of Calcium Hydroxyapatite Using Amides. *Polyhedron* 13, 3071–3078. doi:10.1016/s0277-5387(00)83673-6
- Youness, R. A., Taha, M. A., Elhaes, H., and Ibrahim, M. (2017). Molecular Modeling, FTIR Spectral Characterization and Mechanical Properties of Carbonated-Hydroxyapatite Prepared by Mechanochemical Synthesis. *Mater. Chem. Phys.* 190, 209–218. doi:10.1016/j.matchemphys.2017.01.004
- Young, D. (2001). *Computational Chemistry - A Practical Guide for Applying Techniques to Real-World Problems*. New York: John Wiley & Sons.
- 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 Pradeep, Jain, Dharmashekara, Prasad, Akshatha, Pruthvish, Amachawadi, Srinivasa, Syed, Elgorban, Al Kheraif, Ortega-Castro, Frau, Flores-Holguín, Shivamallu, Kollur and Glossman-Mitnik. 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.



Virtual Screening on Marine Natural Products for Discovering TMPRSS2 Inhibitors

Mehdi Mahmudpour^{1,2}, Iraj Nabipour¹, Mohsen Keshavarz² and Maryam Farrokhnia^{1*}

¹The Persian Gulf Marine Biotechnology Research Center, The Persian Gulf Biomedical Sciences Research Institute, Bushehr University of Medical Sciences, Bushehr, Iran, ²The Persian Gulf Tropical Medicine Research Center, The Persian Gulf Biomedical Sciences Research Institute, Bushehr University of Medical Sciences, Bushehr, Iran

OPEN ACCESS

Edited by:

Sylvain Broussy,
University of Paris, France

Reviewed by:

Andrei I. Khlebnikov,
Tomsk Polytechnic University, Russia
Luca Rastrelli,
University of Salerno, Italy

*Correspondence:

Maryam Farrokhnia
farrokhnia.m@gmail.com

Specialty section:

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

Received: 09 June 2021

Accepted: 17 September 2021

Published: 12 October 2021

Citation:

Mahmudpour M, Nabipour I,
Keshavarz M and Farrokhnia M (2021)
Virtual Screening on Marine Natural
Products for Discovering
TMPRSS2 Inhibitors.
Front. Chem. 9:722633.
doi: 10.3389/fchem.2021.722633

Although SARS-CoV-2 entry to cells strictly depends on angiotensin-converting enzyme 2 (ACE2), the virus also needs transmembrane serine protease 2 (TMPRSS2) for its spike protein priming. It has been shown that the entrance of SARS-CoV-2 through ACE2 can be blocked by cellular TMPRSS2 blockers. The main aim of this study was to find potential inhibitor(s) of TMPRSS2 through virtual screening against a homology model of TMPRSS2 using the library of marine natural products (MNPs). The homology modeling technique for generating a three-dimensional structure of TMPRSS2 was applied. Molecular docking, MM-GBSA and absorption, distribution, metabolism, excretion (ADME) evaluations were performed to investigate the inhibitory activity of marine natural products (MNPs) against TMPRSS2 and their pharmacokinetic properties. Camostat and nafamostat mesylate were used as the standard inhibitory molecules. Seven MNPs were able to inhibit TMPRSS2 better than the standard compounds. MNP 10 with CAS number 107503-09-3, called Watasenia β -D- Preluciferyl glucopyrasoiuronic acid, was found to be the best inhibitor of TMPRSS2 with acceptable pharmacokinetic properties. Herein, for the first time, a new marine natural product was introduced with potent inhibitory effects against TMPRSS2. MNP 10 exhibited favorable drug-like pharmacokinetic properties and it promises a novel TMPRSS2 blocker to combat SARS-CoV-2.

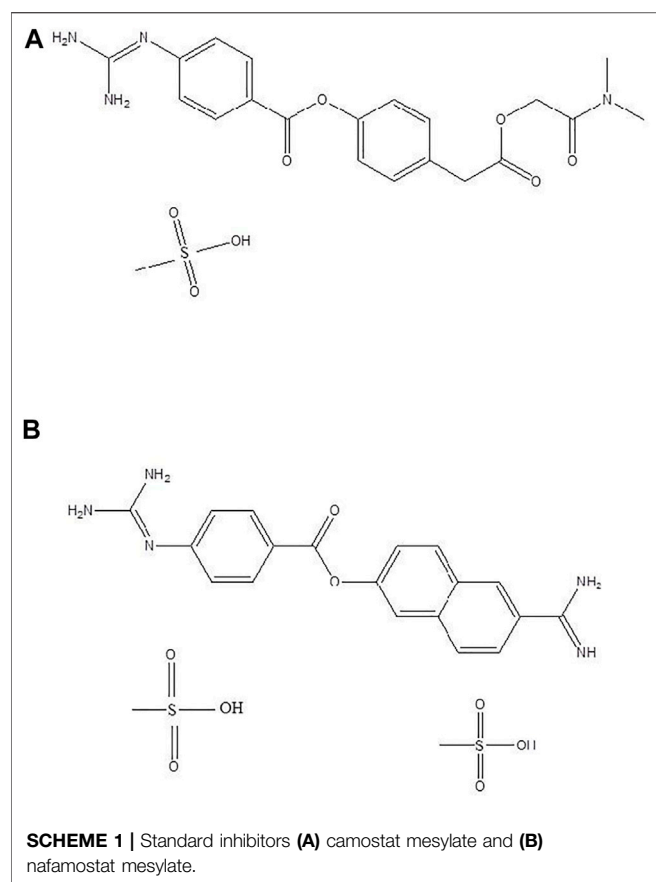
Keywords: SARS-CoV-2, COVID-19, tmprss2, molecular modeling, molecular docking

HIGHLIGHTS

- 1) Marine natural products (MNPs) are a valuable source for anti-SARS-CoV-2 drugs.
- 2) MNP 10 is a potent TMPRSS2 inhibitor to combat SARS-CoV-2.
- 3) MNP 10 has favorable drug-like pharmacokinetic characteristics.

INTRODUCTION

The devastating pandemic caused by SARS-CoV-2 (Astuti, 2020; Guan et al., 2020; Wang et al., 2020; Zhu et al., 2020) that first broke out in Wuhan, China in late 2019, has become the most important global health and socioeconomic issue. Although there is a worldwide effort to develop an effective vaccine against SARS-CoV-2 using both established and new vaccine production technologies (WHO, 2021), no one can yet claim what kind of therapy can be absolutely efficient for the treatment



of or protection against COVID-19. Based on the involved pathophysiologic pathways, different kinds of therapeutic modalities have been conducted in numerous clinical trials with conflicting results.

SARS-CoV-2 is a spherical shaped virus with a diameter of about 60–140 nm with some pleomorphism belonging to Coronaviridae family. The enveloped virus has distinctive spikes (Zhu et al., 2020). Its genome is around 29.8 kilobase with a single-stranded positive-sense RNA (Lu et al., 2020a; Chan et al., 2020; Zhou et al., 2020) encoding 12 putative structural and non-structural proteins; of which spike(S), envelope (E), membrane (M) and nucleocapsid (N) proteins are structural. The S protein complex has two subunits, the S1 subunit contains a single peptide, a receptor-binding domain (RBD) which mediates attachment of virion to host cell surface receptors, and an N-terminal domain (NTD). The S2 subunit mediates fusion between the viral and host cellular membranes which facilitates virus genome entry into the host cell (Gui et al., 2017; Kirchdoerfer et al., 2018; Song et al., 2018; Wan et al., 2020).

Although it has been discovered that the SARS-CoV-2 entry to cells strictly depends on Angiotensin-Converting Enzyme 2 (ACE2) (Li et al., 2003; Li and De Clercq, 2020), it has been shown in several studies that SARS-CoV-2 also needs transmembrane serine protease 2 (TMPRSS2) for S protein

priming (Hoffmann et al., 2020a; Hoffmann et al., 2020b; Stopsack et al., 2020). Hoffmann et al. showed that the entrance of SARS-CoV-2 through ACE2 can be blocked using cellular TMPRSS2 blockers. The SARS-CoV-2 spike protein contains several arginine residues with a high cleavability action at the communication point in the S1/S2 cleavage site. It has been suggested that this zoonotic-origin cleavage site sequence is required for SARS-CoV-2 entrance into human cells. These findings are consistent with previous observations from several clinically relevant viruses such as MERS, other Coronaviruses, and Influenza A virus (Kim et al., 2006; Matsuyama et al., 2010; Glowacka et al., 2011; Shulla et al., 2011; Kawase et al., 2012; Gierer et al., 2013; Zhou et al., 2015; Shen et al., 2017; Iwata-Yoshikawa et al., 2019; Kleine-Weber et al., 2019). Moreover, Heurich et al. showed that TMPRSS2 and other potentially related proteases cleave the ACE2 and SARS-S protein leading to the SARS-CoV entry and fusion of the virus S protein with the host cell membrane, respectively (Heurich et al., 2014). Therefore blocking S protein priming by specific serine protease might have the potential to control SARS-CoV-2 infection. There are some therapeutic agents like bromehehexine (Lucas et al., 2014), camostat mesylate (Shirato et al., 2013), and nafamostat mesylate (Yamamoto et al., 2016) that have been elucidated to be a good inhibitor of TMPRSS2.

Regarding TMPRSS2 inhibitors as potent candidates for anti SARS-CoV-2 infection, it may be promising to investigate natural resources to discover novel components with anti-TMPRSS2 activities. The extreme and unusual environment of the ocean has extraordinary organisms with astonishing properties, which can reveal new horizons for treatment in modern medicine including marine-derived secondary metabolites with evident anti-inflammatory, antitumor, antimicrobial, antiviral, antimalarial, and antioxidant activities (Molinski et al., 2009; Gogineni et al., 2015; Blunt et al., 2018; Riccio et al., 2020; Yi et al., 2020; Carroll et al., 2021). Based on the extraordinary self-defense capacities of marine organisms and the occurrence in them of some deadly viral infections, these organisms might be regarded as a source of novel antiviral agents, which may be able to combat a SARS-CoV-2 infection. Hence, marine-derived natural compounds should be considered in our efforts to overcome the challenges of COVID-19 treatments.

A promising method to investigate viral entry and proliferation is to apply a computer-aided active site directed inhibition study. To the best of our knowledge, there are no studies that have screened the marine natural products (MNPs) libraries specifically to find a blocker of TMPRSS2. However, several studies have screened the libraries of other natural products, particularly plant-derived compounds, to discover the potential inhibitors of TMPRSS2 (Chikhale et al., 2020; Da Silva Antonio et al., 2020; Idris et al., 2020; Rahman et al., 2020; Singh et al., 2020; Vivek-Ananth et al., 2020; Hu et al., 2021).

This study aimed to find potential inhibitor(s) of TMPRSS2 through virtual screening against a homology model of TMPRSS2 using the library of marine natural products (MNPs).

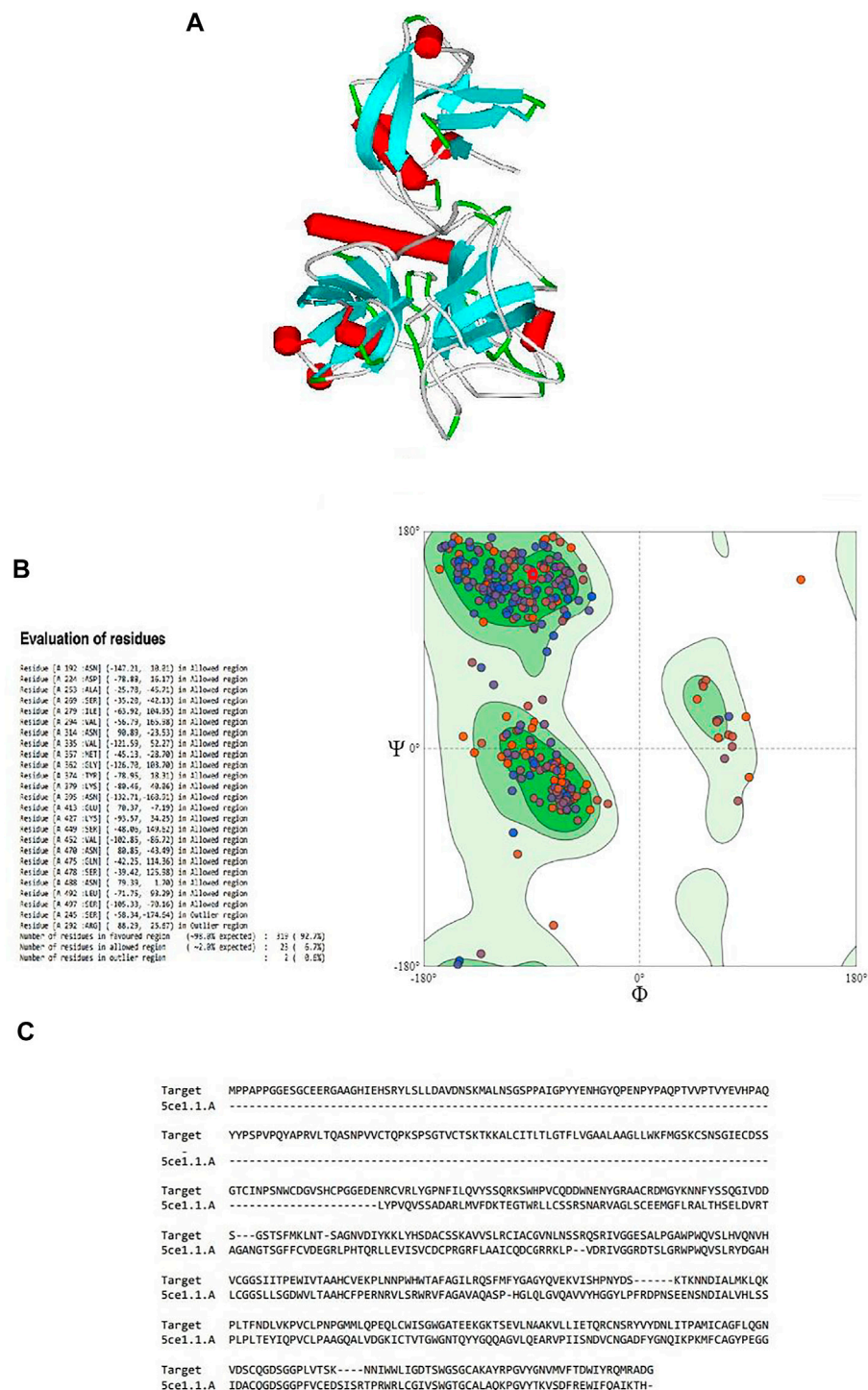


FIGURE 1 | (A) Three-dimensional structure of the modeled serine protease transmembrane protease serine 2 (TMPRSS2), **(B)** Ramachandran plot validation of the modeled 3D structure, **(C)** alignment of the target serine protease TMPRSS2 and the template serine protease hepsin (PDB ID: 5CE1.A.).

MATERIALS AND METHODS

Homology Modeling

The essential step in the study of the structural and functional aspects of any protein is to have its suitable crystal structure.

Unfortunately, the three-dimensional (3D) structure of TMPRSS2 had not been found at the time of the current study. Hence, in this case, the only option was to generate a 3D coordinate of TMPRSS2 by comparative prediction approach. Herein, the online server SWISS-MODEL (Guex et al., 2009;

TABLE 1 | Homology modeling results and validation for predicting 3D structure of TMPRSS2.

Server/tool	Parameter	Score
Swiss-model	QMEAN z score	-1.47
	GMQE	0.48
	Seq Similarity	51%
	Coverage	71%
	Seq identity	33.82%
RAMPAGE(Ramachandran plot)	favoured region	92.7%
	Allowed region	6.7%
	Outlier region	2%
MolProbity	Molprobability Score	1.89
	Ramachandran Favoured	92.20%
	Ramachandran Outliers	1.16%
	Rotamer Outliers	1.35%
	C-Beta Deviations	7%

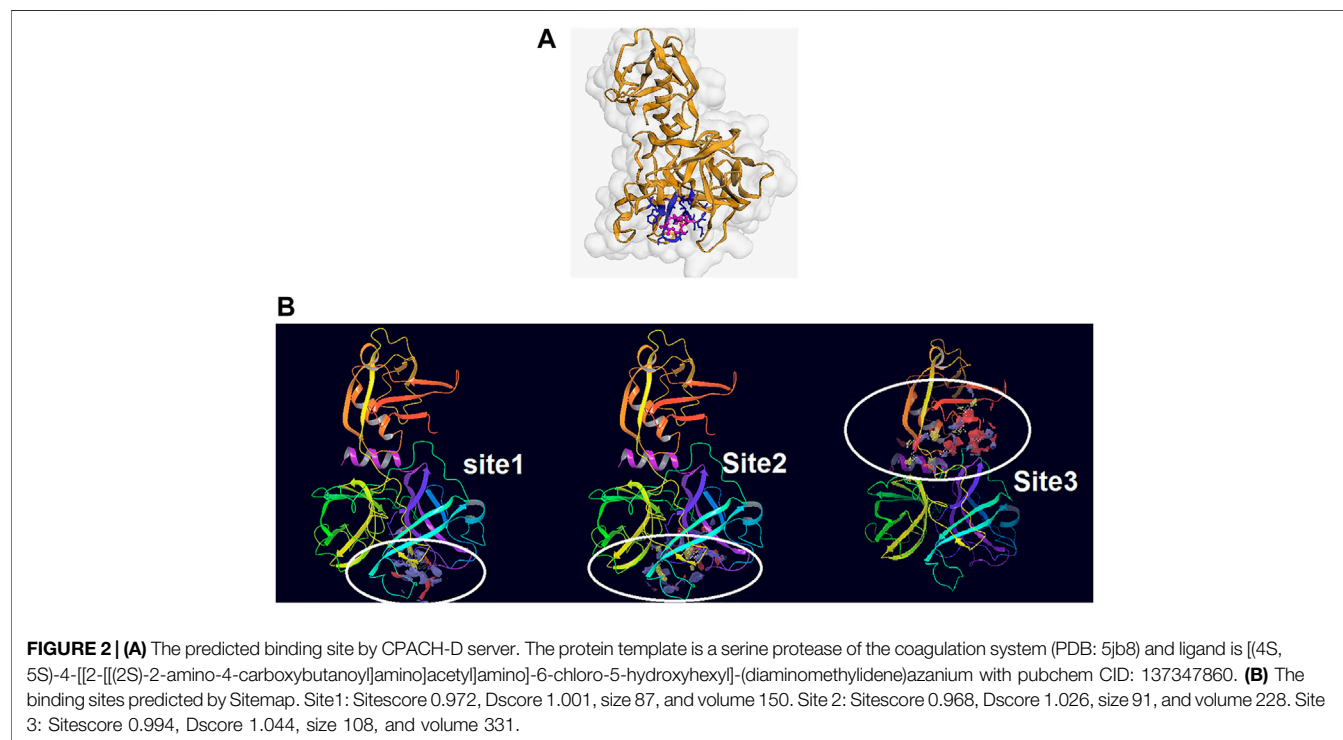
Bertoni et al., 2017; Waterhouse et al., 2018) (<https://swissmodel.expasy.org/>) was used to build the 3D structure of TMPRSS2. The amino acid sequence of human transmembrane protease serine 2, from the Universal Protein Resource “UniProtKB” (accession no: O15393) isoform-2,492 amino acids long (<https://www.uniprot.org/uniprot/O15393>), was selected for homology modeling that was performed by a template-based method. Then, the RAMPAGE server was used to validate the 3D modeled structure (<http://mordred.bioc.cam.ac.uk/~%7B%7Drapper/rampage.php>).

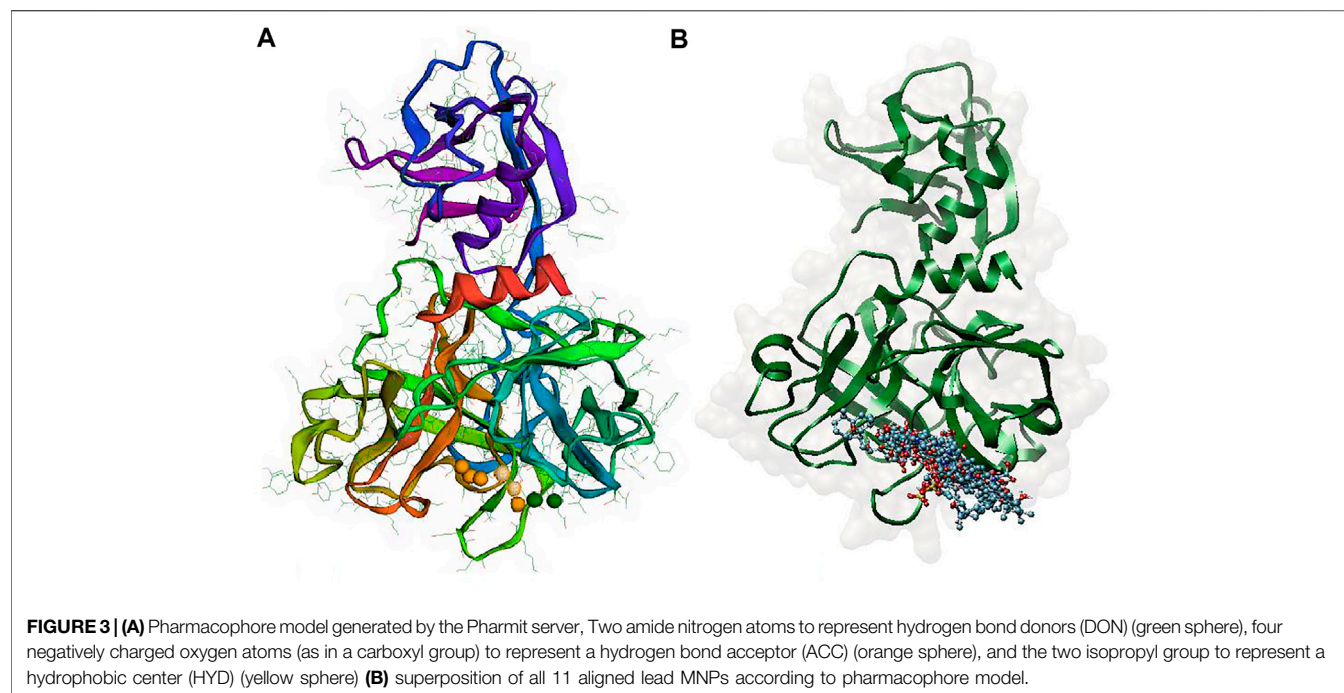
Pharmacophore-Based Virtual Screening

The pharmacophore model was created using the Pharmit server (<http://pharmit.csb.pitt.edu/>). First, pharmacophore features were automatically extracted from the co-crystallized inhibitor of Serine protease Hepsin (PDB: 5CE1.

TABLE 2 | Predicted binding site of homology model of TMPRSS2.

Predicted binding site	C-score	E-Value	Involved residues	Docking energy of representative ligand-template complex	Predicted binding residues
CDD-search	—	7.39e-100	ILE293-GLN524	—	HIS333 to SER478 Catalytic site: HIS333, ASP382 and SER478 Substrate binding site: ASP472, SER497 and GLY499
COACH-D	0.99	—	ILE293-GLN524	-6.0	HIS333, LYS379, ASP472-GLY476, SER478, THR496-CYS502 and GLY509





A), and camostat mesylate was also used as another ligand for the pharmacophore modeling to obtain a more realistic model for homology structure of TMPRSS2 as input receptor. Hence, the pharmacophore model used for TMPRSS2 virtual screening was generated based on the predicted binding interactions of TMPRSS2 with these two inhibitors. The Pharmit parameters for 3D-pharmacophore research were changed according to these pharmacophore parameters. Then the MNP library, which contains 164,952 conformers from 14,064 molecules, was searched on this model. The hit compounds with an RMSD ≥ 4 Å and minimized affinity ≥ -6 were discarded. The remaining poses were minimized using functions of Pharmit. Finally, the pharmacophore-based minimized entries were further docked with the target protein to identify the lead compounds with the best docking scores.

Active Site Identification and Preparation of TMPRSS2 for Docking

The conserved domain (CD) search was done on the Fasta sequence of the homology model of TMPRSS2 using the NCBI's conserved domain database (CDD/SPARCLE: <https://www.ncbi.nlm.nih.gov/cdd/>) (Lu et al., 2020b). Then, it was analyzed and its cleavage, active, and substrate binding site residues were predicted. The COACH-D server (<https://yanglab.nankai.edu.cn/COACH-D/>) (Yang et al., 2013; Wu et al., 2018) was also applied to predict TMPRSS2 putative ligand-binding sites. Then, in a model-template active site comparative study, the active and substrate binding residues of the (5CE1.A) template were similarly obtained.

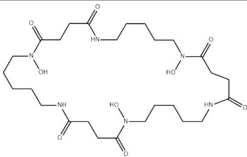
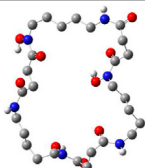
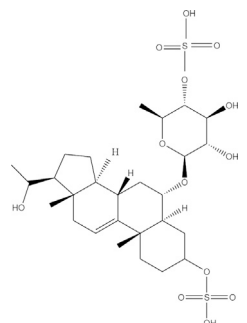
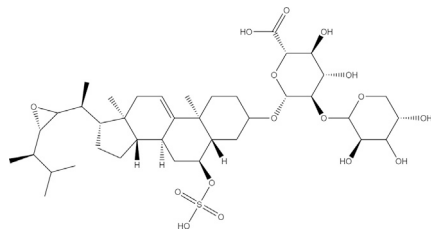
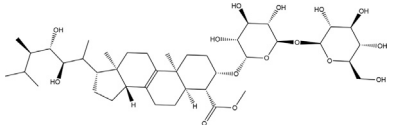
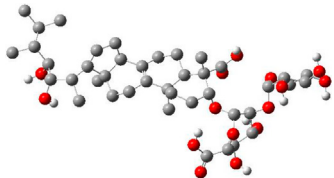
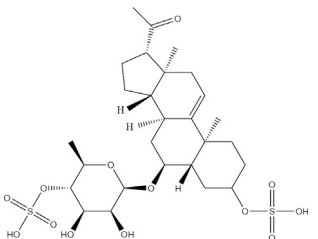
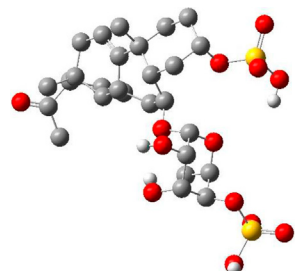
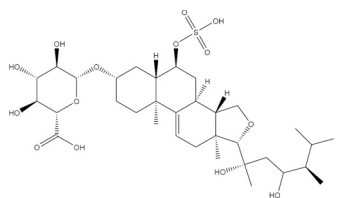
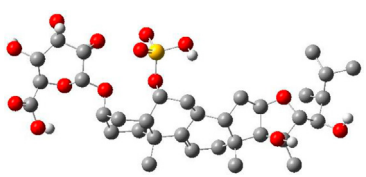
Finally, the homology structure of TMPRSS2 was imported into Maestro Protein Wizard and sitemap analysis was done by SiteMap (Halgren, 2007; Halgren, 2009; Schrödinger, 2015b).

This software was used with default settings, in which the top five possible binding sites by a minimum of 15 points were identified while cropping site maps set at 4 Å from the nearest site point. SiteMap used a more restrictive definition of hydrophobicity by standard grid. A SiteScore value above 0.80 is indicative of high druggability and promising drug-binding sites and is used in conjunction with Dscores, which serve as a measure of hydrophobicity. Dscore or druggability score penalizes increasing hydrophilicity and is thus used as a druggability measure for a pocket. In general, Dscore < 0.83 is considered as “undruggable,” 0.83–0.98 as “difficult to drug” and > 0.98 as “druggable” (Halgren, 2007; Halgren, 2009; Vidler et al., 2012).

Molecular Docking Ligand Preparation

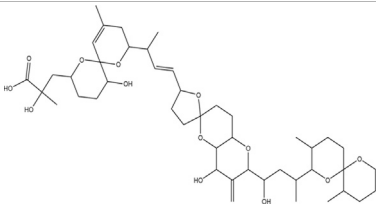
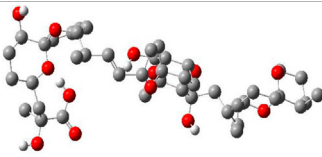
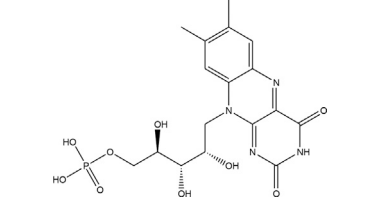
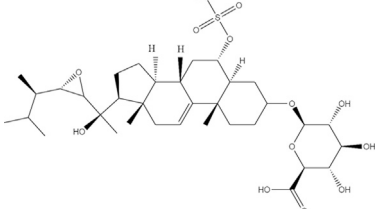
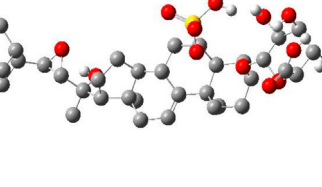
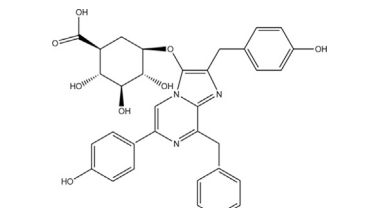
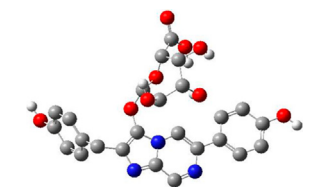
One conformation was generated per compound which was followed by geometry optimization with PM3 (Stewart, 1991), a semi-empirical method using Hyperchem release 7 for windows (HyperCube Inc., 2002). The geometry-optimized structures were retrieved in MOL type for further analysis with the LigPrep application which has been implemented in the Schrödinger 2015-2 suite of software (Schrödinger, 2015a). The ionization state was specified at pH = 7.00 ± 2.0 using Epik (Epik, 2015) based on Hammett and Taft methodologies (Sastri et al., 2013). The desalt option was the same as the program default. All 32 possible conformations were produced for each compound at pH 7.00 in the OPLS3 (Harder et al., 2016). The obtained ligands were then used in the docking calculations.

TABLE 3 | CAS number, 2D and 3D presentations of investigated marine natural products.

#	CAS#	2D structure	3D structure
1	26605-16-3		
2	125127-57-3		
3	174286-21-6		
4	454470-88-3		
5	143572-73-0		
6	170894-36-7		
7	81720-10-7		

(Continued on following page)

TABLE 3 | (Continued) CAS number, 2D and 3D presentations of investigated marine natural products.

#	CAS#	2D structure	3D structure
8	6184-17-4		
9	174286-17-0		
10	107503-09-3		
11	59985-26-1		

Generation of the Grid

As adequate generation of the grid is a key step in the prediction of a ligand binding to a receptor, the 3D boundary for ligand binding was produced by Glide, version 10.2 of Mastero, Schrödinger (Glide, 2015). First, the protein preparation wizard was used with the following settings: 1) The original hydrogens removal and subsequent addition of hydrogens. 2) The atomic charges and bond orders were assigned. 3) The N and C termini were capped. 4) The disulfide bonds were generated between sulfur atoms (within 3.2 Å). 5) Epik was applied to

generate possible protonation states at neutral pH. 6) The H-bonds were assigned, optimized by PROPKA (Olsson et al., 2011; Søndergaard et al., 2011) at pH 7.0, and then the structure was minimized with the OPLS3 force field. Glide was used to generate the grid on the catalytic domain of the receptor. The grid box size was set to 32*32*32 Å.

Molecular Interaction and Docking Studies

After grid generation, ligand docking was done according to the protocols in Glide version 10.2. The homology model of TMPRSS2

was used as the receptor, and the different internally produced conformations by the software were passed across some filters such as Euler angles, grid-based force field evaluation, and energy minimization by Monte Carlo. Finally, docking score is an important parameter for evaluating the conformations, and in this study, the output of standard precision (SP) docking was put forward in extra precision (XP) docking. The docked compounds were ranked based on their docking scores.

Pose Rescoring With Molecular Mechanics Generalized Born Surface Area

The binding energies of all docking poses were calculated using the molecular mechanics generalized Born surface area (MM-GBSA) approach implemented in the Prime program in the Schrödinger software suite (Prime, 2015). This approach employs a single minimized protein-ligand structure, and so is used as an efficient approach to rapidly refine and rescore docking results. A variable dielectric solvent model VSGB 2.0 (Li et al., 2011) was used, this solvent model contains several empirical corrections for modeling the directionality of hydrogen bond and π -stacking interactions. MM-GBSA has been shown to give good binding free energies for a wide range of protein-ligand complexes (Mulakala and Viswanadhan, 2013). It is also widely used to evaluate docking poses, to determine the stability of ligand-target complex for predicting binding affinity in drug design (Genheden and Ryde, 2011; Wang et al., 2019).

Absorption, Distribution, Metabolism, and Excretion and Drug-likeness Analysis

It is known that nearly 40% of drug candidates fail in clinical trials because of poor absorption, distribution, metabolism, and excretion (ADME). Hence, it is very crucial to recognize these problematic candidates at an early stage to avoid wasted time and resources. Accurate ADME prediction is based on full 3D molecular structures. Qikprop offers a set of several predictors including central nervous system (CNS) penetration, predicted apparent Caco-2 cell permeability across the gut-blood barrier in nm/sec (QPPCaco), apparent MDCK cell permeability (QPPMDCK), human oral absorption, Lipinski's rule of five, and predicted maximum transdermal transport rate (JM). Another option of Qikprop (Qikprop, 2015) is to rank compounds based on how drug-like they are. In the current study, the ADME, drug-likeness, and medicinal chemistry parameters of these 11 compounds were predicted by QikProp (Qikprop, 2015).

RESULTS AND DISCUSSION

Homology Modeling and Evaluation of Model Quality

Since the crystal structure of TMPRSS2 was unavailable at the time of this study, the 3D structure of TMPRSS2 was predicted using the online server SWISS-MODEL, as shown in **Figure 1A**. Its global model quality estimate (GMQE) score was 0.48, this score estimates the quality of the expected output model by a

particular template. Its QMEAN Z-score was -1.47, with sequence coverage of 71%, sequence identity of 33.82%, and sequence similarity of 50% in comparison with the template (PDB ID: 5CE1.A). It has been shown that when the sequence similarity with the template is more than 30%, the obtained model can be considered reliable and suitable for further study (Xiang, 2006). Benkert et al. (2011) showed that QMEAN Z-score is an estimation of the degree of native likeness of the model, and a value close to 0 (and not lower than -4) could be an acceptable agreement criteria for the experimental structure of similar size **Table 1**.

By validating the obtained results of SWISS-MODEL and cross-checking in RAMPAGE, it was observed that there were 319 (92.7%) residues were in the favored region, 23 (6.7%) residues in the allowed region, and 2 (0.6%) residues in the outlier region. Moreover, there were no steric clashes or deviations in bond length or bond angle compared to the protein structure report (**Supplementary Figure S1** and **Supplementary Figure S2**). These results indicated that the obtained model may have the correct geometry. The 3D arrangement of the model is shown in **Figure 1B**. The alignment of the template (PDB ID: 5CE1.A) and the target protein is shown in **Figure 1C**. The summary of obtained results is presented in **Table 1**.

Catalytic Site of the Homology Model of TMPRSS2

According to a conserved domain database (CDD) search, the protein classification of TMPRSS2 was Trypsin-like serine proteases with E-value 7.39e-100, and the residues of ILE293-GLN524 were involved in the characteristic domain of this protein. Based on the CDD algorithm, six amino acid residues have been identified as especially important in the active site of TMPRSS2, and one residue for its cleavage site. The active site included residue HIS333 to SER478 where (HIS333, ASP382, and SER478) were the three important amino acids at the catalytic site whereas (ASP472, SER497, and GLY499) residues were found to be the substrate binding site (**Table 2**). Moreover, the COACH-D results supported those findings. According to the COACH-D best prediction results, the protein template was a serine protease of the coagulation system (PDB ID: 5JB8) with a confidence score of 0.99.

The corresponding docking energy of the template for its representative ligand (with pubchem CID: 137347860) was -6.0 kcal/mol. HIS333, LYS379, ASP472-GLY476, SER478, THR496-CYS502, and GLY509 were identified as the predicted binding residues, of which the CDD search showed HIS333, ASP472, SER478, SER497, and GLY499, as the same catalytic domain residues.

From the five binding sites which were calculated by sitemap, two sites had a Dscore above 1, and their locations were completely in the agreement with the CDD search and COACH-D results (Site 1: Sitescore 0.972, Dscore 1.001, size 87, and volume 150. Site 2: Sitescore 0.968, Dscore 1.026, size 91, and volume 228). The Dscore of site 3 was 1.044 with Sitescore 0.994 but its position is not in agreement with the catalytic site of Trypsin-like serine proteases active domain which is located in the base of its S1 pocket where it contains Asp472, and therefore is predicted to cleave after lysine or arginine residues (Hedstrom, 2002; Wilson et al., 2005; Blay and Pei, 2019). The predicted binding sites by all

TABLE 4 | Glide Docking score, Glide energy, Glide emodel, and estimated free energy of binding for the best poses of investigated compounds in kcal/mol.

Ligand	Docking score	Glide energy	Glide emodel	ΔG_{bind}
1	-3.99	-40.89	-49.54	-36.06
2	-5.82	-47.34	-62.26	-64.33
3	-7.04	-51.70	-72.78	-73.48
4	-7.10	-47.50	-67.96	-67.30
5	-3.68	-36.62	-48.24	-54.64
6	-7.09	-42.77	-54.31	-62.99
7	-7.02	-48.18	-37.76	-53.89
8	-6.51	-47.50	-55.17	-76.00
9	-5.42	-45.15	-59.48	-67.82
10	-8.16	-59.21	-71.34	-76.00
11	-7.12	-49.37	-63.35	-67.68
Camostat	-4.52	-42.47	-53.19	-57.91
Nafamostat	-3.73	-38.27	-45.75	-47.49

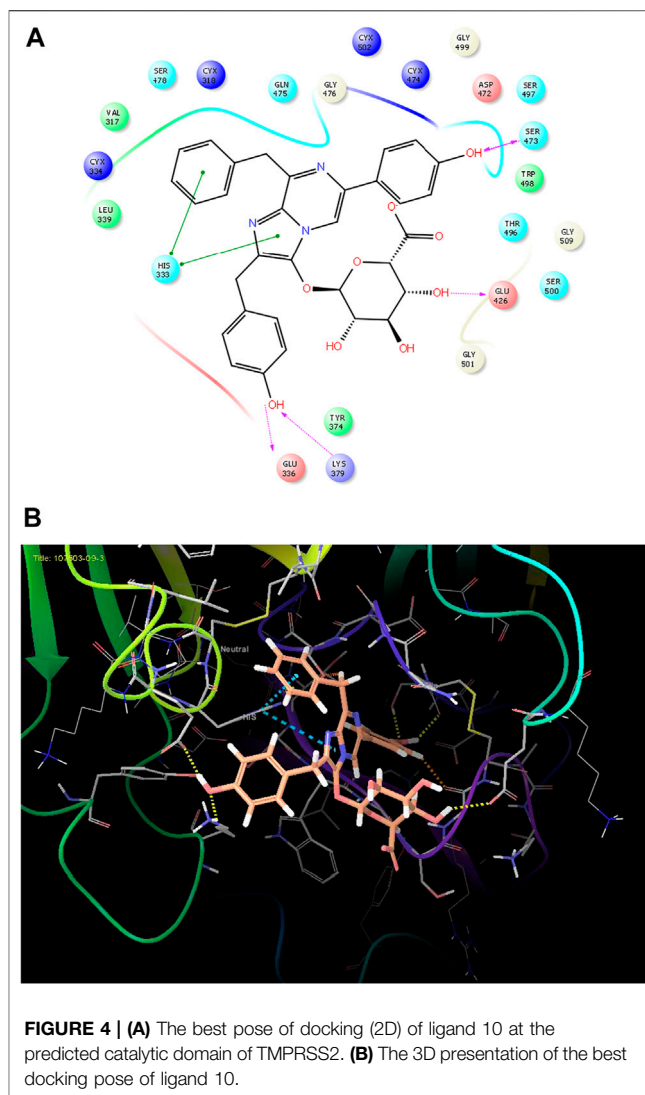
applied approaches are in agreement with the theoretical study by Idris et al. (2020). Two other predicted binding sites had a Dscore lower than 0.8, which indicates that they are not druggable. Hence, domains 1 and 2 were chosen for the docking study. **Figures 2A,B** illustrate the predicted binding site of TMPRSS2 from COACH-D and SiteMap analysis.

Pharmacophore Model

Since a pharmacophore states the crucial features of interactions, such as the spatial arrangement of each interaction in the close contact of ligand and the target, its accurate setup is very important in binding site pharmacophore modeling. In this study, the 3D structure of TMPRSS2 homology was used to set pharmacophore using the Pharmit server; this server provides both pharmacophore and molecular shape search options and the results are ranked by the energy. The generated pharmacophore features were selected according to the co-crystal inhibitor of the template and camostat mesylate. In this modeling, the binding-site derived pharmacophore models include three subgroups of ligand binding sites: i) two amide nitrogen atoms were added to represent hydrogen bond donors (DON), ii) four negatively charged oxygen atoms (as in a carboxyl group) were added to represent a hydrogen bond acceptor (ACC), and iii) the two isopropyl group were added to represent a hydrophobic center (HYD) (**Figure 3A**). According to the generated pharmacophore model, a vast library of MNP (14,064 molecules, 164,952 conformers) was filtered. A total of 25,000 hits that met the criteria were minimized, resulting in 114 conformers. A total of 11 structures were retained by using one conformer for each molecule, with an RMSD lower than 4 Å and a binding score lower than -6 (**Figure 3B** and **Table 3**).

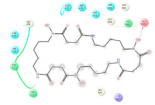
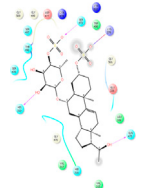
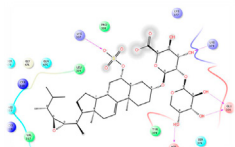
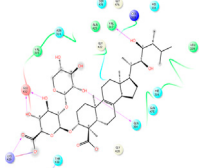
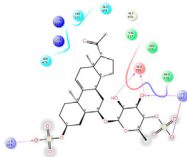
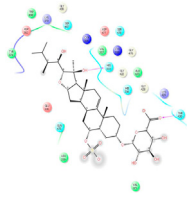
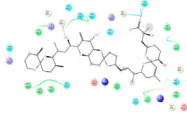
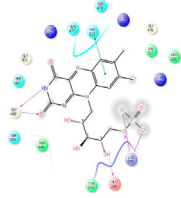
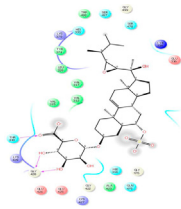
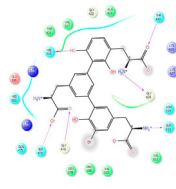
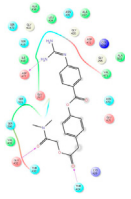
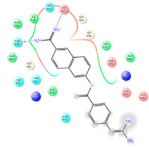
Molecular Docking and Molecular Mechanics Generalized Born Surface Area Studies

The top 11 selected MNPs of the 114 structures which are shown in **Table 3** were separately docked into the catalytic site of TMPRSS2. The results of docking of these structures, as well as camostat and nafamostat mesylate, as two standard inhibitors of TMPRSS2 (**Scheme 1**), are presented in **Table 4**.



The molecular docking analysis revealed that all the studied compounds had comparable or lower docking scores than those of the standard inhibitors. The highest docking Glide score is -8.16 in compound 10, whereas these scores were -4.52 and -3.73 in camostat and nafamostat mesylate respectively. Also, the docking scores in compounds 3, 4, 6, 7, and 11 are lower than -7. Thus, these MNPs could be considered as the most potent inhibitors for TMPRSS2. Moreover, the highest Glide energy value was -59.21 for compound 10, and its Glide emodel value is -71.34, which are almost the highest values. The Glide energies were -42.47 and -38.27 for camostat mesylate and nafamostat mesylate respectively, and Glide emodel values were -53.19 and -45.75 respectively. Therefore, from the observed theoretical superiority of these 11 MNPs compared to the standard inhibitors (camostat and nafamostat mesylate), these compounds may be encouraging for further studies. Interestingly, among these selected compounds, compound 10 had the most promising results. Compound 10 (CAS number 107503-09-3; Watasenia Preluciferyl β-D-glucopyrasoiuronic acid) is a bioluminescent substance that was derived from the

TABLE 5 | 2D presentation of best Glide docking pose at the catalytic domain of TMPRSS2 for investigated marine natural product.

MNP	2D presentation of best docking pose	MNP	2D presentation of best docking pose
1		2	
3		4	
5		6	
7		8	
9		11	
Camostate		Nafamostat	

liver of myctophina fish, *Diaphus elucens* (Inoue et al., 1987; Blunt and Munro, 2007).

As it is presented in **Table 5**, the 2D template of best poses demonstrates the types of contacts formed between the ligands and target with cutoff 4.00 Å. Remarkably, close contact/interactions within the catalytic domain were detected for all MNPs, however, all the important residues of catalytic domains significantly contributed in interactions with compound 10, such

as HIS333 which was involved in π - π stacking interaction, and SER497 and GLY499 which were the important residues of substrate binding, in the close vicinity of phenol and iduronic acid moieties (**Figure 4A**). In addition, SER473, GLU426, LYS379, and GLU336 participated in intermolecular H-bond interaction with the active domain (**Figure 4B**). Taken together, compounds 3, 4, 10, and 11 had stronger interactions with TMPRSS2 than the standard ligands based on both their

TABLE 6 | ADME parameters and drug similarity of investigated marine natural products.

MNP	S	MW	QLogP o/w	QLogS	QLogK	QPPCaco	QLogBB	PHOA	RO5	JRO3	The most similar drug (%)
1	11	600.71	-3.39	2.00	-3.19	0	-5.32	0	3	1	Trientine (50)
2	1	640.76	1.53	-4.29	-0.83	0	-3.40	7	2	1	Sulfamazone (65.2)
3	11	818.97	2.00	-5.16	-0.84	0	-4.83	0	3	2	Dirithromycin (60.45)
4	11	798.96	1.78	-3.96	-0.91	0	-3.88	0	3	2	Monoxerutin (64.25)
5	1	638.74	1.00	-3.17	-1.01	0	-3.00	6	2	1	Sulfamazone (69.7)
6	5	718.85	2.16	-6.24	-0.52	0	-3.80	0	3	2	Azithromycin (65.70)
7	11	819.04	7.21	-10.30	1.40	36	-2.68	59	3	2	Rifaximin (56.25)
8	3	456.35	-1.04	-2.04	-1.47	0	-3.68	0	2	1	Methotrexate (79.50)
9	2	702.85	3.34	-4.86	-0.26	1	-3.12	9	3	2	Cefotiam (66)
10	1	599.60	3.22	-4.24	-0.13	9	-2.45	23	3	2	Amprenavir (68.5)
11	9	618.44	-2.24	-0.86	-1.25	0	-4.93	0	3	2	Lymecycline (67.23)

S (STARS) = Number of property/descriptor values falling outside the 95% range of similar values for known drugs. Recommended value 0–5. MW = Molecular weight Recommended values 130.0–725.0. QLogPo/w = Predicted octanol/water partition coefficient. Recommended values -2.0–6.5. QLogK = hsa Serum Protein Binding. Recommended values -1.5–1.5. QLogS = Predicted aqueous solubility, log S. Recommended values -6.5–0.5. QPPCaco = Predicted apparent Caco-2 cell permeability in nm/sec. Recommended values < 25 poor, >500 great. QLogBB = Predicted brain/blood partition coefficient. Recommended values -3.0–1.2. PHOA = Predicted human oral absorption on 0–100% scale. Recommended values > 80% is high <25% is poor. RO5 = Rule of Five, The rules are: mol_MW < 500, QLogPo/w < 5, donorHB ≤5, and acptHB ≤10. Maximum is 4. JRO3 = Jorgensen Rule of 3 Violations. Maximum is 3.

Glide and XP-pose emodel energies and were significantly involved in active domain contacts/interactions. On the other hand, the energy values of compounds 2 and 9 were also considerable and close to the references. Molecule 3, *Downeyoside I*, firstly isolated from starfish *Henricia Downeyae* by Plagiano; molecule 2, *Forbeside E*, a sulfated sterol glycoside from starfish *Asterias forbesi*; molecule 4, *Ulososide E*, which has been derived from the sponge *Ulosa sp*; and molecule 9, *Downeyoside E*, sulfated steroid glycoside isolated from *Henricia Downeyae* (Blunt and Munro, 2007).

In addition, the MM-GBSA calculations which estimated the values of ΔG_{bind} are reported in **Table 4**. According to these results, the two standard inhibitors had values of -57.91 and -47.49 kcal/mol while eight molecules out of the 11 studied MNPs had more negative ΔG_{bind} than the standard inhibitors. The highest binding free energies were for compounds 8 and 10 (-76.00 kcal/mol) which is much higher than the standard inhibitors, and the values of compounds 2, 3, 4, 6, 9, and 11 were more negative than the standard inhibitors. After the rescoring of the 11 MNPs, compound 10 remained as a potent inhibitor of TMPRSS2, however other ligands might be also interesting for further studies. Although these compounds belong to various sources, starfish-derived products sound to be prominent. For example, 2, 3, and 9 have been isolated from starfish, whereas 7 has been isolated from the marine sponge *Merriamum oxeato* (Blunt and Munro, 2007). In this study, the marine natural products showed good measurable binding affinities for the TMPRSS2 residues. In other words, these binding affinities are indicative of the ligand's contribution to ligand-target interactions and their sensible flexibility for this target. Based on XP Glide docking score, compounds 10 and 11 have strong interactions with the enzyme but rescoring by MM-GBSA suggests that other compounds such as 2, 3, 4, 8, and 9 are also important. Nevertheless, our proposed lead compound is compound 10, which has concurrently both high docking scores and a comparable ΔG_{bind} to standard inhibitors.

Absorption, Distribution, Metabolism, and Excretion and Drug-likeness Analysis

In this study, the ADME <https://www.sciencedirect.com/topics/pharmacology-toxicology-and-pharmaceutical-science/admeproperties> of 11 MNPs were analyzed using the QikProp tool. This analysis represents the physicochemical properties of chemical compounds along with their biological functions. The resulting physicochemical and biological properties are molecular formula, molecular weight, volume, SASA, acceptor H-bond, donor H-bond groups, the number of ring atoms, QLogPw (-2–6.5), the percentage of human oral absorption, and CNS effects. However, it has been suggested that Lipinski's rule of five (Zhang and Wilkinson, 2007) is not a strict criterion for natural compounds (Lipinski, 2003) and it has been revealed that natural compounds mostly do not follow Lipinski's rule and they tend to keep their low hydrophobicity as well as their potential of donating the intermolecular H-bonds (Ganesan, 2008).

In general, Lipinski's rule of five (Lipinski et al., 1997) is applied for predicting the drug-likeness with the following criteria: molecular mass less than 500 Da, up to 5 hydrogen bond donors, no more than 10 hydrogen bond acceptors, and an octanol-water partition coefficient (logPo/w) no higher than 5. The rule states that a molecule or an inhibitor can be orally absorbed/active if two or more of these thresholds are not violated. However, Jorgensen's rule of three may also be used to evaluate the bioavailability of each marine natural product by estimating its solubility, permeability, and liver first-pass metabolism through the following rules: predicted aqueous solubility (logSwat) higher than -5.7 (with S in mol/dm³), predicted apparent Caco-2 cell rate permeability (BIPcaco-2) high than 22 nm/s, and number of primary metabolites up to 7 (Di and Kerns, 2015). In addition, the predicted qualitative human oral absorption (2 = medium and 3 = high) and the predicted skin permeability (logKp values between -8.0–1.0) are considered. Finally, the ADME-compliance score drug-likeness parameter (#star) was

used to evaluate the pharmacokinetic of the studied compounds, including 25 different properties within the acceptable range of 95% of the known drugs. Herein, compounds 2, 5, 9, and 10 had the fewest violations when ADME-compliance score drug-likeness parameter (#stars) was considered less than 2 (Table 6), however, the recommended value is 0–5, and compounds 6 and 8 were also considered in the acceptable range of #stars. It is concluded that these MNPs may be proper candidate drugs for TMPRSS2 inhibition.

Again, amongst the latter compounds, compound 10 showed a higher human oral absorption (23), as well as median aggregation to plasma proteins (Qlog k has serum protein binding: −0.13) and predicted aqueous solubility values (QPlog S: 3.22). Moreover, by considering the number of “stars” and the violations from the Lipsink and Jorgensen rules, the obtained results for compound 10 indicated a high degree of reliability to be a drug candidate.

To find chemical similarity to the known drug molecules, the QikProp (Qikprop, 2015) software database identified five similar drug molecules for each entry according to its predicted descriptors. In this study, the results of two of the compounds were very promising. Accordingly, Azithromycin with a similarity of 65.70% was suggested for compound 6. It is interesting to know that in the early phase of COVID-19, azithromycin could reduce the need for hospitalization or duration of clinical recovery (Echeverría-Esnal et al., 2020; Million et al., 2020; Molina et al., 2020). There is also an opinion supporting the potential effectiveness of Azithromycin in SARS-CoV-2 infection, as well as its antiviral activity and immunomodulatory effects (Bleyzac et al., 2020). Interestingly, Amprenavir (Shen et al., 2010), as an antiretroviral protease inhibitor for HIV infection has been identified with 68.5% similarity to compound 10.

In conclusion, marine natural product 10, with high similarity to a known antiretroviral protease, might be considered as a potent inhibitor of TMPRSS2. However, molecular docking and MM-GBSA studies demonstrated that compounds 3, 4, 6, 7, 8, 9, and

11 were also able to inhibit TMPRSS2 as well. In a comparison study, these compounds showed better results than the standard TMPRSS2 inhibitors (camostat and nafamostat mesylate). Further computational, experimental, and clinical investigations are warranted to reveal their anti-SARS-CoV-2 activities.

DATA AVAILABILITY STATEMENT

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

AUTHOR CONTRIBUTIONS

MF conceived and designed the format of the manuscript. All calculations have been done by MF. All the authors, MM, IN, and MK reviewed the manuscript and all of them contributed to the critical reading and discussion of the manuscript. All authors have read and agreed to the published version of the manuscript.

FUNDING

The authors wish to express their thanks for the partially financial support of Bushehr University of medical sciences.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fchem.2021.722633/full#supplementary-material>

REFERENCES

- Astuti, I., and Ysrafil, Y. (2020). Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2): An Overview of Viral Structure and Host Response. *Diabetes Metab. Syndr. Clin. Res. Rev.* 14, 407–412. doi:10.1016/j.dsx.2020.04.020
- 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
- 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, 10480–10495. doi:10.1038/s41598-017-09654-8
- Blay, V., and Pei, D. (2019). Serine Proteases: How Did Chemists Tease Out Their Catalytic Mechanism. *ChemTexts* 5, 1–7. doi:10.1007/s40828-019-0093-4
- Bleyzac, N., Goutelle, S., Bourguignon, L., and Tod, M. (2020). Azithromycin for COVID-19: More Than Just an Antimicrobial. *Clin. Drug Investig.* 40 (8), 683–686. doi:10.1007/s40261-020-00933-3
- Blunt, J. W., Carroll, A. R., Copp, B. R., Davis, R. A., Keyzers, R. A., and Prinsep, M. R. (2018). Marine Natural Products. *Nat. Prod. Rep.* 35, 8–53. doi:10.1039/c7np00052a
- Blunt, J. W., and Munro, M. H. (2007). *Dictionary of marine Natural Products with CD-ROM*. London, United Kingdom: Chapman and Hall/CRC Press.
- Carroll, A. R., Copp, B. R., Davis, R. A., Keyzers, R. A., and Prinsep, M. R. (2021). Marine Natural Products. *Nat. prod. Rep.* 38 (2), 362–413. doi:10.1039/D0NP00089B
- 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. *Emerging Microbes & Infections* 9, 221–236. doi:10.1080/22221751.2020.1719902
- Chikhale, R. V., Gupta, V. K., Eldesoky, G. E., Wabaidur, S. M., Patil, S. A., and Islam, M. A. (2020). Identification of Potential Anti-TMPRSS2 Natural Products through Homology Modelling, Virtual Screening and Molecular Dynamics Simulation Studies. *J. Biomol. Struct. Dyn.* 39, 6660–6675. doi:10.1080/07391102.2020.1798813
- 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
- Di, L., and Kerns, E. H. (2015). *Drug-Like Properties: Concepts, Structure Design and Methods from ADME to Toxicity Optimization*. Cambridge: Academic Press.
- Echeverría-Esnal, D., Martín-Ontiyuelo, C., Navarrete-Rouco, M. E., De-Antonio Cuscó, M., Ferrández, O., Horcajada, J. P., et al. (2020). Azithromycin in the Treatment of COVID-19: A Review. *Expert Rev. Anti Infect. Ther.* 19 (2), 147–163. doi:10.1080/14787210.2020.1813024
- Epik (2015). version 3. 2 ed. New York, NY: Schrödinger.

- Ganesan, A. (2008). The Impact of Natural Products upon Modern Drug Discovery. *Curr. Opin. Chem. Biol.* 12, 306–317. doi:10.1016/j.cbpa.2008.03.016
- Genheden, S., and Ryde, U. (2011). Comparison of the Efficiency of the LIE and MM/GBSA Methods to Calculate Ligand-Binding Energies. *J. Chem. Theor. Comput.* 7, 3768–3778. doi:10.1021/ct200163c
- Gierer, S., Bertram, S., Kauf, F., Wrensch, F., Heurich, A., Krämer-Kühl, A., et al. (2013). The Spike Protein of the Emerging Betacoronavirus EMC Uses a Novel Coronavirus Receptor for Entry, Can Be Activated by TMPRSS2, and Is Targeted by Neutralizing Antibodies. *J. Virol.* 87, 5502–5511. doi:10.1128/jvi.00128-13
- Glide (2015). version 6. 7 ed. New York, NY: Schrödinger.
- Glowacka, I., Bertram, S., Müller, M. A., Allen, P., Soilleux, E., Pfefferle, S., et al. (2011). Evidence that TMPRSS2 Activates the Severe Acute Respiratory Syndrome Coronavirus Spike Protein for Membrane Fusion and Reduces Viral Control by the Humoral Immune Response. *J. Virol.* 85, 4122–4134. doi:10.1128/jvi.02232-10
- Gogineni, V., Schinazi, R. F., and Hamann, M. T. (2015). Role of Marine Natural Products in the Genesis of Antiviral Agents. *Chem. Rev.* 115, 9655–9706. doi:10.1021/cr4006318
- 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
- Guex, N., Peitsch, M. C., and Schwede, T. (2009). Automated Comparative Protein Structure Modeling with SWISS-MODEL and Swiss-PdbViewer: A Historical Perspective. *Electrophoresis* 30, S162–S173. doi:10.1002/elps.200900140
- 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
- Halgren, T. A. (2009). Identifying and Characterizing Binding Sites and Assessing Druggability. *J. Chem. Inf. Model.* 49, 377–389. doi:10.1021/ci800324m
- Halgren, T. (2007). New Method for Fast and Accurate Binding-Site Identification and Analysis. *Chem. Biol. Drug Des.* 69, 146–148. doi:10.1111/j.1747-0285.2007.00483.x
- 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. Theor. Comput.* 12, 281–296. doi:10.1021/acs.jctc.5b00864
- Hedstrom, L. (2002). Serine Protease Mechanism and Specificity. *Chem. Rev.* 102, 4501–4524. doi:10.1021/cr000033x
- Heurich, A., Hofmann-Winkler, H., Gierer, S., Liepold, T., Jahn, O., and Pöhlmann, S. (2014). TMPRSS2 and ADAM17 Cleave ACE2 Differentially and Only Proteolysis by TMPRSS2 Augments Entry Driven by the Severe Acute Respiratory Syndrome Coronavirus Spike Protein. *J. Virol.* 88, 1293–1307. doi:10.1128/jvi.02202-13
- Hoffmann, M., Kleine-Weber, H., Schroeder, S., Krüger, N., Herrler, T., Erichsen, S., et al. (2020a). SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. *cell* 181, 271–280. e278. doi:10.1016/j.cell.2020.02.052
- Hoffmann, M., Schroeder, S., Kleine-Weber, H., Müller, M. A., Drosten, C., and Pöhlmann, S. (2020b). Nafamostat Mesylate Blocks Activation of SARS-CoV-2: New Treatment Option for COVID-19. *Antimicrob. Agents Chemother.* 64 (6), e00754–20. doi:10.1128/AAC.00754-20
- Hu, X., Shrimp, J. H., Guo, H., Xu, M., Chen, C. Z., Zhu, W., et al. (2021). Discovery of TMPRSS2 Inhibitors from Virtual Screening as a Potential Treatment of COVID-19. *ACS Pharmacol. translational Sci.* 4 (3), 1124–1135. doi:10.1021/acspstci.0c00221
- Hypercube Inc. (2002). *Hyperchem 7.0*. Toronto, C.: Hypercube Inc.
- Idris, M. O., Yekeen, A. A., Alakanse, O. S., and Durojaye, O. A. (2020). Computer-Aided Screening for Potential TMPRSS2 Inhibitors: a Combination of Pharmacophore Modeling, Molecular Docking and Molecular Dynamics Simulation Approaches. *J. Biomol. Struct. Dyn.* 39 (15), 5638–5656. doi:10.1080/07391102.2020.1792346
- Inoue, S., Okada, K., Tanino, H., and Kakoi, H. (1987). Chemical Studies on Myctophina Fish Bioluminescence. *Chem. Lett.* 16, 417–418. doi:10.1246/cl.1987.417
- Iwata-Yoshikawa, N., Okamura, T., Shimizu, Y., Hasegawa, H., Takeda, M., and Nagata, N. (2019). TMPRSS2 Contributes to Virus Spread and Immunopathology in the Airways of Murine Models after Coronavirus Infection. *J. Virol.* 93, e01815–18. doi:10.1128/JVI.01815-18
- Kawase, M., Shirato, K., Van Der Hoek, L., Taguchi, F., and Matsuyama, S. (2012). Simultaneous Treatment of Human Bronchial Epithelial Cells with Serine and Cysteine Protease Inhibitors Prevents Severe Acute Respiratory Syndrome Coronavirus Entry. *J. Virol.* 86, 6537–6545. doi:10.1128/jvi.00094-12
- Kim, T. S., Heinlein, C., Hackman, R. C., and Nelson, P. S. (2006). Phenotypic Analysis of Mice Lacking the Tmprss2 -Encoded Protease. *Mol. Cell. Biol.* 26, 965–975. doi:10.1128/mcb.26.3.965-975.2006
- Kirchdoerfer, R. N., Wang, N., Pallesen, J., Wrapp, D., Turner, H. L., Cottrell, C. A., et al. (2018). Stabilized Coronavirus Spikes Are Resistant to Conformational Changes Induced by Receptor Recognition or Proteolysis. *Sci. Rep.* 8, 15701–15711. doi:10.1038/s41598-018-34171-7
- Kleine-Weber, H., Elzayat, M. T., Wang, L., Graham, B. S., Müller, M. A., Drosten, C., et al. (2019). Mutations in the Spike Protein of Middle East Respiratory Syndrome Coronavirus Transmitted in Korea Increase Resistance to Antibody-Mediated Neutralization. *J. Virol.* 93, e01381–18. doi:10.1128/JVI.01381-18
- 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, 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
- Li, W., Moore, M. J., Vasilieva, 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
- Lipinski, C. A. (2003). Chris Lipinski Discusses Life and Chemistry after the Rule of Five. *Drug Discov. Today* 8, 12–16. doi:10.1016/s1359-6446(02)02556-4
- Lipinski, C. A., Lombardo, F., Dominy, B. W., and Feeney, P. J. (1997). Experimental and Computational Approaches to Estimate Solubility and Permeability in Drug Discovery and Development Settings. *Adv. Drug Deliv. Rev.* 23, 3–25. doi:10.1016/s0169-409x(96)00423-1
- Lu, R., Zhao, X., Li, J., Niu, P., Yang, B., Wu, H., et al. (2020a). Genomic Characterisation and Epidemiology of 2019 Novel Coronavirus: Implications for Virus Origins and Receptor Binding. *The lancet* 395, 565–574. doi:10.1016/s0140-6736(20)30251-8
- Lu, S., Wang, J., Chitsaz, F., Derbyshire, M. K., Geer, R. C., Gonzales, N. R., et al. (2020b). CDD/SPARCLE: The Conserved Domain Database in 2020. *Nucleic Acids Res.* 48, D265–D268. doi:10.1093/nar/gkz991
- Lucas, J. M., Heinlein, C., Kim, T., Hernandez, S. A., Malik, M. S., True, L. D., et al. (2014). The Androgen-Regulated Protease TMPRSS2 Activates a Proteolytic cascade Involving Components of the Tumor Microenvironment and Promotes Prostate Cancer Metastasis. *Cancer Discov.* 4, 1310–1325. doi:10.1158/2159-8290.cd-13-1010
- 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
- Matsuyama, S., Nagata, N., Shirato, K., Kawase, M., Takeda, M., and Taguchi, F. (2010). Efficient Activation of the Severe Acute Respiratory Syndrome Coronavirus Spike Protein by the Transmembrane Protease TMPRSS2. *J. Virol.* 84, 12658–12664. doi:10.1128/jvi.01542-10
- Million, M., Lagier, J.-C., Gautret, P., Colson, P., Fournier, P.-E., Amrane, S., et al. (2020). Early Treatment of COVID-19 Patients with Hydroxychloroquine and Azithromycin: A Retrospective Analysis of 1061 Cases in Marseille, France. *Trav. Med. Infect. Dis.* 35, 101738. doi:10.1016/j.tmaid.2020.101738
- Molina, J. M., Delaugerre, C., Le Goff, J., Mela-Lima, B., Ponscarne, D., Goldwirth, L., et al. (2020). No Evidence of Rapid Antiviral Clearance or Clinical Benefit with the Combination of Hydroxychloroquine and Azithromycin in Patients with Severe COVID-19 Infection. *Médecine et Maladies Infectieuses* 50, 384. doi:10.1016/j.medmal.2020.03.006
- Molinski, T. F., Dalisay, D. S., Lievens, S. L., and Saludes, J. P. (2009). Drug Development from Marine Natural Products. *Nat. Rev. Drug Discov.* 8, 69–85. doi:10.1038/nrd2487
- Mulakala, C., and Viswanadhan, V. N. (2013). Could MM-GBSA Be Accurate Enough for Calculation of Absolute Protein/Ligand Binding Free Energies. *J. Mol. Graphics Model.* 46, 41–51. doi:10.1016/j.jmgm.2013.09.005
- Olsson, M. H. M., Søndergaard, C. R., Rostkowski, M., and Jensen, J. H. (2011). PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pKa Predictions. *J. Chem. Theor. Comput.* 7, 525–537. doi:10.1021/ct100578z
- Prime (2015). version 4.0. New York, NY: Schrödinger, LLC.
- Qikprop (2015). version 4.4. New York, NY: Schrödinger, LLC.
- Rahman, N., Basharat, Z., Yousuf, M., Castaldo, G., Rastrelli, L., and Khan, H. (2020). Virtual Screening of Natural Products against Type II Transmembrane

- Serine Protease (TMPRSS2), the Priming Agent of Coronavirus 2 (SARS-CoV-2). *Molecules* 25, 2271. doi:10.3390/molecules25102271
- Riccio, G., Ruocco, N., Mutalipassi, M., Costantini, M., Zupo, V., Coppola, D., et al. (2020). Ten-Year Research Update Review: Antiviral Activities from Marine Organisms. *Biomolecules* 10, 1007. doi:10.3390/biom10071007
- Schrödinger, L. (2015a). *LigPrep, Version 3.1*. New York, NY: Schrödinger.
- Schrödinger, L. (2015b). *LigPrep, Version 3.5*. New York, NY: Schrödinger, LLC.
- Shen, C.-H., Wang, Y.-F., Kovalevsky, A. Y., Harrison, R. W., and Weber, I. T. (2010). Amprenavir Complexes with HIV-1 Protease and its Drug-Resistant Mutants Altering Hydrophobic Clusters. *FEBS J.* 277, 3699–3714. doi:10.1111/j.1742-4658.2010.07771.x
- Shen, L. W., Mao, H. J., Wu, Y. L., Tanaka, Y., and Zhang, W. (2017). TMPRSS2: A Potential Target for Treatment of Influenza Virus and Coronavirus Infections. *Biochimie* 142, 1–10. doi:10.1016/j.biochi.2017.07.016
- Shirato, K., Kawase, M., and Matsuyama, S. (2013). Middle East Respiratory Syndrome Coronavirus Infection Mediated by the Transmembrane Serine Protease TMPRSS2. *J. Virol.* 87, 12552–12561. doi:10.1128/jvi.01890-13
- Shulla, A., Heald-Sargent, T., Subramanya, G., Zhao, J., Perlman, S., and Gallagher, T. (2011). A Transmembrane Serine Protease Is Linked to the Severe Acute Respiratory Syndrome Coronavirus Receptor and Activates Virus Entry. *J. Virol.* 85, 873–882. doi:10.1128/jvi.02062-10
- Singh, R., Gautam, A., Chandel, S., Ghosh, A., Dey, D., Roy, S., et al. (2020). Protease Inhibitory Effect of Natural Polyphenolic Compounds on SARS-CoV-2: An In Silico Study. *Molecules* 25, 4604. doi:10.3390/molecules25204604
- Søndergaard, C. R., Olsson, M. H., Rostkowski, M., and Jensen, J. H. (2011). Improved Treatment of Ligands and Coupling Effects in Empirical Calculation and Rationalization of pKa Values. *J. Chem. Theor. Comput.* 7, 2284–2295. doi:10.1021/ct200133y
- 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
- Stewart, J. J. P. (1991). Optimization of Parameters for Semiempirical Methods. III Extension of PM3 to Be, Mg, Zn, Ga, Ge, As, Se, Cd, In, Sn, Sb, Te, Hg, Tl, Pb, and Bi. *J. Comput. Chem.* 12, 320–341. doi:10.1002/jcc.540120306
- Stopsack, K. H., Mucci, L. A., Antonarakis, E. S., Nelson, P. S., and Kantoff, P. W. (2020). TMPRSS2 and COVID-19: Serendipity or Opportunity for Intervention. *Cancer Discov.* 10, 779–782. doi:10.1158/2159-8290.cd-20-0451
- Vidlar, L. R., Brown, N., Knapp, S., and Hoelder, S. (2012). Druggability Analysis and Structural Classification of Bromodomain Acetyl-Lysine Binding Sites. *J. Med. Chem.* 55, 7346–7359. doi:10.1021/jm300346w
- Vivek-Ananth, R. P., Rana, A., Rajan, N., Biswal, H. S., and Samal, A. (2020). Silico Identification of Potential Natural Product Inhibitors of Human Proteases Key to SARS-CoV-2 Infection. *Molecules* 25, 3822. doi:10.3390/molecules25173822
- Wan, Y., Shang, J., Graham, R., Baric, R. S., and Li, F. (2020). Receptor Recognition by the Novel Coronavirus from Wuhan: An Analysis Based on Decade-Long Structural Studies of SARS Coronavirus. *J. Virol.* 94, e00127–20. doi:10.1128/JVI.00127-20
- Wang, C., Horby, P. W., Hayden, F. G., and Gao, G. F. (2020). A Novel Coronavirus Outbreak of Global Health Concern. *The Lancet* 395, 470–473. doi:10.1016/s0140-6736(20)30185-9
- Wang, E., Sun, H., Wang, J., Wang, Z., Liu, H., Zhang, J. Z. H., et al. (2019). End-Point Binding Free Energy Calculation with MM/PBSA and MM/GBSA: Strategies and Applications in Drug Design. *Chem. Rev.* 119, 9478–9508. doi:10.1021/acs.chemrev.9b00055
- 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
- WHO (2021). Draft Landscape and Tracker of COVID-19. candidate vaccines-21 February [Online]. Available: <https://www.who.int/publications/m/item/draft-landscape-of-covid-19-candidate-vaccines> (Accessed October 1, 2021).
- Wilson, S., Greer, B., Hooper, J., Zijlstra, A., Walker, B., Quigley, J., et al. (2005). The Membrane-Anchored Serine Protease, TMPRSS2, Activates PAR-2 in Prostate Cancer Cells. *Biochem. J.* 388, 967–972. doi:10.1042/bj20041066
- Wu, Q., Peng, Z., Zhang, Y., and Yang, J. (2018). COACH-D: Improved Protein-Ligand Binding Sites Prediction with Refined Ligand-Binding Poses through Molecular Docking. *Nucleic Acids Res.* 46, W438–W442. doi:10.1093/nar/gky439
- Xiang, X. (2006). Advances in Homology Protein Structure Modeling. *Cpps* 7, 217–227. doi:10.2174/13892030677452312
- Yamamoto, M., Matsuyama, S., Li, X., Takeda, M., Kawaguchi, Y., Inoue, J.-I., 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
- Yang, J., Roy, A., and Zhang, Y. (2013). Protein-ligand Binding Site Recognition Using Complementary Binding-Specific Substructure Comparison and Sequence Profile Alignment. *Bioinformatics* 29, 2588–2595. doi:10.1093/bioinformatics/btt447
- Yi, M., Lin, S., Zhang, B., Jin, H., and Ding, L. (2020). Antiviral Potential of Natural Products from marine Microbes. *Eur. J. Med. Chem.* 207, 112790. doi:10.1016/j.ejmech.2020.112790
- Zhang, M.-Q., and Wilkinson, B. (2007). Drug Discovery Beyond the 'Rule-Of-Five'. *Curr. Opin. Biotechnol.* 18, 478–488. doi:10.1016/j.copbio.2007.10.005
- 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
- Zhou, Y., Vedantham, P., Lu, K., Agudelo, J., Carrion, R., Jr, Nunneley, J. W., et al. (2015). Protease Inhibitors Targeting Coronavirus and Filovirus Entry. *Antivir. Res.* 116, 76–84. doi:10.1016/j.antiviral.2015.01.011
- Zhu, N., Zhang, D., Wang, W., Li, X., Yang, B., Song, J., et al. (2020). A Novel Coronavirus from Patients with Pneumonia in China. *New Engl. J. Med.* 382 (8), 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.

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 Mahmudpour, Nabipour, Keshavarz and Farrokhnia. 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.



Virtual Screening of Drug-Like Compounds as Potential Inhibitors of the Dengue Virus NS5 Protein

Leidy L. García-Ariza^{1*}, Cristian Rocha-Roa^{2,3}, Leonardo Padilla-Sanabria¹ and Jhon C. Castaño-Osorio¹

¹Grupo de Inmunología Molecular, Centro de Investigaciones Biomédicas, Universidad del Quindío, Armenia, Colombia, ²Grupo de Parasitología Molecular, Centro de Investigaciones Biomédicas, Universidad del Quindío, Armenia, Colombia, ³Biophysics of Tropical Diseases, Max Planck Tandem Group, Universidad de Antioquia, Medellín, Colombia

OPEN ACCESS

Edited by:

Sylvain Broussy,
University of Paris, France

Reviewed by:

Yanmin Zhang,
China Pharmaceutical University,
China

Sajjad Gharaghani,
University of Tehran, Iran
Mahmoud A. Al-Sha'er,
Zarqa Private University, Jordan

*Correspondence:

Leidy L. García-Ariza
llgarcia@uniquindio.edu.co

Specialty section:

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

Received: 03 December 2020

Accepted: 10 January 2022

Published: 10 February 2022

Citation:

García-Ariza LL, Rocha-Roa C,
Padilla-Sanabria L and
Castaño-Osorio JC (2022) Virtual
Screening of Drug-Like Compounds as
Potential Inhibitors of the Dengue Virus
NS5 Protein.
Front. Chem. 10:637266.
doi: 10.3389/fchem.2022.637266

Dengue virus (DENV) is the causative agent of dengue fever. Annually, there are about 400 million new cases of dengue worldwide, and so far there is no specific treatment against this disease. The NS5 protein is the largest and most conserved viral protein among flaviviruses and is considered a therapeutic target of great interest. This study aims to search drug-like compounds for possible inhibitors of the NS5 protein in the four serotypes of DENV. Using a virtual screening from a ~642,759-compound database, we suggest 18 compounds with NS5 binding and highlight the best compound per region, in the methyltransferase and RNA-dependent RNA polymerase domains. These compounds interact mainly with the amino acids of the catalytic sites and/or are involved in processes of protein activity. The identified compounds presented physicochemical and pharmacological properties of interest for their use as possible drugs; furthermore, we found that some of these compounds do not affect cell viability in Huh-7; therefore, we suggest evaluating these compounds *in vitro* as candidates in future research.

Keywords: dengue virus, NS5 protein, drug-like compounds, molecular docking, virtual screening

INTRODUCTION

Dengue virus (DENV) is a member of the genus *Flavivirus* belonging to the family *Flaviviridae* (Zou et al., 2011; Lim et al., 2013a). DENV is the causative agent of the viral disease known as dengue fever, which is transmitted through the bite of mosquito species *Aedes aegypti* and *Aedes albopictus* (Regato et al., 2008; Basavannacharya and Vasudevan., 2014). This disease mainly affects people who live in tropical and subtropical countries, with approximately 400 million new cases worldwide annually, and can lead to febrile illness and flu-like symptoms or can progress to the more severe dengue hemorrhagic fever or dengue shock syndrome (Klema et al., 2016). However, to date, there is no specific treatment that can inhibit the replication of DENV.

This virus has a non-segmented, single-stranded, positive-sense RNA genome of approximately 11 kb (Basavannacharya and Vasudevan., 2014), which codes for 10 proteins: 3 structural virion components (C, PRM, and E proteins) and 7 nonstructural proteins (NS1, NS2A, NS2B, NS3, NS4A, NS4B, and NS5) (Lim et al., 2013b; Basavannacharya and Vasudevan., 2014; Galiano et al., 2016). To date, four serotypes of DENV have been reported (DENV1 to DENV4) (Niyomrattanakit et al., 2015; Lai et al., 2017; Potisopon et al., 2017). Nevertheless, in the last years, the presence of a fifth serotype with a sylvatic cycle was reported (Mustafa et al., 2014).

The NS5 protein is the largest (Troost and Smit, 2020) and most highly conserved viral protein encoded by the flavivirus genome (Zou et al., 2011; Bollati et al., 2010; García et al., 2017; Wu, J et al., 2020; Bhatnagar et al., 2021). In particular, this protein shows approximately 67–82% amino acid sequence identity among the four dengue serotypes. NS5 consists of two domains, namely methyltransferase (MTase) domain and RNA-dependent RNA polymerase (RdRp) domain (Klema et al., 2016). These domains are linked through a short sequence of poorly conserved amino acids (Lim et al., 2015). The MTase and RdRp domains have enzymatic activity, and both are essential for the viral replication cycle (Bhattacharya et al., 2008).

The N-terminal region of the protein comprises the MTase domain (with a length of approximately 270 amino acids) (Galiano et al., 2016), which functions as a double methyltransferase that can methylate the 5'-end of the viral RNA genome at the N-7 position of the guanosine cap (N-7 MTase) as well as the 2'-OH position of the ribose of the first nucleotide (2'O MTase) (Lim et al., 2013b). Depending on the serotype and the experimental system, the MTase domain, within the context of NS5, can positively influence the polymerase activity. Specifically, the MTase domain of DENV2 can stimulate RNA loading within the adjacent polymerase domain and improve its stable catalytic state during *de novo* specific initiation and the elongation reaction (Potisopon et al., 2017).

On the other hand, the C-terminal region of NS5 harbors RdRp (with a length of approximately 630 amino acids) (Galiano et al., 2016), which plays a vital role in the viral life cycle through replication. After viral entry and the translation of proteins from its genome, the polymerase domain performs the *de novo* synthesis of RNA (first generating RNA of negative polarity from RNA of positive polarity) (Lim et al., 2015). Like all polymerases, the structure of the RdRp of flaviviruses resembles a right hand with the characteristic subdomain fingers (amino acids 273–315, 416–496, and 543–600), palm (amino acids 497–542 and 601–705), and thumb (amino acid 706–900) (Zou et al., 2011; Najera, 2013; Galiano et al., 2016). In addition, the RdRp domain is unique to RNA viruses, and it is absent in human cells; for this reason, the DENV NS5 protein is an attractive target in the search for antiviral compounds (Malet et al., 2008; De Burghgraeve et al., 2013; Meguellati et al., 2014; Alhossary et al., 2018; Mirza et al., 2019). The crystal structure of the RdRp catalytic domain of DENV was reported by Yap et al., allowing the exploration of regions in its structure that could be of interest for the design of anti-dengue compounds (Yap et al., 2007).

To date, several inhibitors of MTase and RdRp activities have been identified by large-scale *in vitro* screening (Bhatnagar et al., 2021). For instance, sinefungin (a SAM analog with a broad antiviral spectrum) has shown affinity six times greater than SAM for its binding site in the MTase domain (Lim et al., 2015). Ribavirin, a synthetic analog of guanosine, has been shown to inhibit dengue and hepatitis C virus replication (Chang et al., 2011; Tomlinson and Watowich, 2011); however, the use of ribavirin is limited by its oral toxicity, and its aerosol presentation diminishes its efficacy for clinical uses (Fusco and Chung, 2014).

Also, the activity of the RdRp enzyme of DENV is inhibited allosterically by blocking the RNA tunnel using N-sulfonylanthranilic acid derivatives, which are considered desirable for the development of antiviral compounds (Yin et al., 2009b; Niyomrattanakit et al., 2010). Likewise, the activity of this enzyme is inhibited by the action of beta-d-2'-ethenyl-7-deaza-adenosine triphosphate (2'E-7D-ATP) through competition with the natural nucleotide. This nucleoside analog, initially developed for hepatitis C (HCV), showed anti-dengue activity in cell culture and significantly reduced viremia in mouse models with DENV. However, the catalytic efficiency of incorporation of this molecule is 10 times lower than that of ATP (Latour et al., 2010; De Burghgraeve et al., 2013; García et al., 2017). Two non-nucleoside inhibitors, retinamide and ivermectin, were identified in binding assays as compounds that can block DENV NS5 (Lim et al., 2015). Ivermectin is reported as an inhibitor of the α/β importin and therefore of the NS5 polymerase since it is required for its activity. There are reports that a previous treatment with ivermectin inhibits dengue virus infection in Vero cells; in addition, a pretreatment with this compound strongly inhibits the nuclear localization of NS5 during infection with DENV1 and DENV2 in BHK-21 or Huh-7 cells (Fusco and Chung, 2014).

Despite many efforts in the search for antiviral compounds against DENV, success has been limited (Diosa-Toro et al., 2018); consequently, it is necessary to look for new alternatives with low or no toxicity and powerful anti-dengue activity. For this, compounds with similar properties to drugs already approved for use in humans can be of great help. Also, the application of computational tools, such as virtual screening, predictors of physical-chemical characteristics, and molecular dynamics (MD) simulations, can contribute to the design and improvement of new drugs. In recent years, most inhibitors have been first selected *via in silico* or high-throughput screening, which was followed by the evaluation of their antiviral activities *via in vitro* or in-cell based assays (Tian et al., 2018). Here, we explore seven binding sites in the NS5 protein of the four DENV serotypes and perform virtual screening strategies to select compounds that can be tested against DENV. Finally, we suggest a short list of 18 compounds that could be considered as candidates for *in vitro* evaluation.

MATERIALS AND METHODS

Structural Modeling of DENV1–4 NS5 Proteins

The structural models of the NS5 proteins of DENV1, DENV2, and DENV4 serotypes were constructed by homology modeling on the SWISS-MODEL web server (Waterhouse et al., 2018), using the full crystal structure of NS5 DENV3 (PDB accession code: 5JJR) (Lim et al., 2016) as a template. We used the consensus amino acid sequences of NS5 for each serotype obtained from the Virus Variation database (NCBI) (Hatcher et al., 2017). For the NS5 protein of DENV3, the non-crystallized regions were modeled but retained the rest of the crystallized structure.

The sequence alignment was performed by the UniProt web server (<https://www.uniprot.org/>) (Uniprot Consortium, 2021). The stereochemical quality of the constructed models was assessed by analyzing the Ramachandran plot (Hollingsworth and Karplus, 2010), which allows to evaluate the ϕ (Φ) and ψ (Ψ) angles of each amino acid. These plots were obtained from the MolProbity web server (Williams et al., 2018). Additionally, we calculated the Z-scores, obtained from the ProSA web server (Wiederstein and Sippl, 2007). The root mean square deviation (RMSD) between the structures of the serotypes was calculated with the MatchMaker module available in Chimera v1.11.2 software (Pettersen et al., 2004).

Virtual Screening of Drug-Like Compounds and Molecular Docking Calculations

The virtual screenings were performed using the supercomputer of the Texas Advanced Computing Center (TACC) (<https://drugdiscovery.tacc.utexas.edu>) (Viswanathan et al., 2014) linked to the ZINC database of compounds for virtual screening (<https://zinc.docking.org/>). AutoDock Tools v1.5.6 (Morris et al., 2009) was used for preparing the receptors (NS5 protein) and ligands (compounds used as controls). For proteins, we removed the water molecules, and co-crystallized ligands, polar hydrogens, and Kollman charges were added. For ligands, the polar hydrogens, Gasteiger charges, and rotatable bonds were added. The compounds in the ZINC Lrg database (~642,759) were not prepared as the TACC portal, where this library is available, has the compounds ready for molecular docking.

The molecular docking calculations were performed with AutoDock Vina software (Trott and Olson, 2010) in several important regions for functionality of NS5 in the four serotypes of DENV. These regions have been studied in other research studies (Zou et al., 2011; Galiano et al., 2016; Lim et al., 2015; Malet et al., 2008; Yap et al., 2007; Niyomrattanakit et al., 2010; Dong et al., 2008; Zhou et al., 2007; Zhao et al., 2015a), and they have been recognized as interesting sites for antiviral drug development. For the RdRp domain, we evaluated cavities A and B (Malet et al., 2008; Zhou et al., 2007), the RNA tunnel (Galiano et al., 2016; Yap et al., 2007; Niyomrattanakit et al., 2010), and the GDD motif (Galiano et al., 2016). Likewise, we assessed the KDKE tetrad (Dong et al., 2008; Zhou et al., 2007; Zhao et al., 2015a; Zhao et al., 2015c) and the SAM- and GTP-binding site (Lim et al., 2015; Dong et al., 2008) for the MTase domain. Overall, we explored seven regions in the two domains. The regions with a crystallized ligand, such as GTP-binding site, SAM-binding site, and GDD motif, were validated using redocking. All the models were aligned to be able to use the same boxes in the four serotypes, and the boxes were configured with a dimension of 24 Å³. The compounds were obtained from the library ZINC Lrg, available as a tab inside TACC options (<https://drugdiscovery.tacc.utexas.edu>). This library contains ~642,759 commercially available drug-like compounds. The visualizations of the 2D interactions and the generation of the protein–ligand complexes were performed with Chimera v1.11.2 software (Pettersen et al., 2004).

Selection of the Compounds

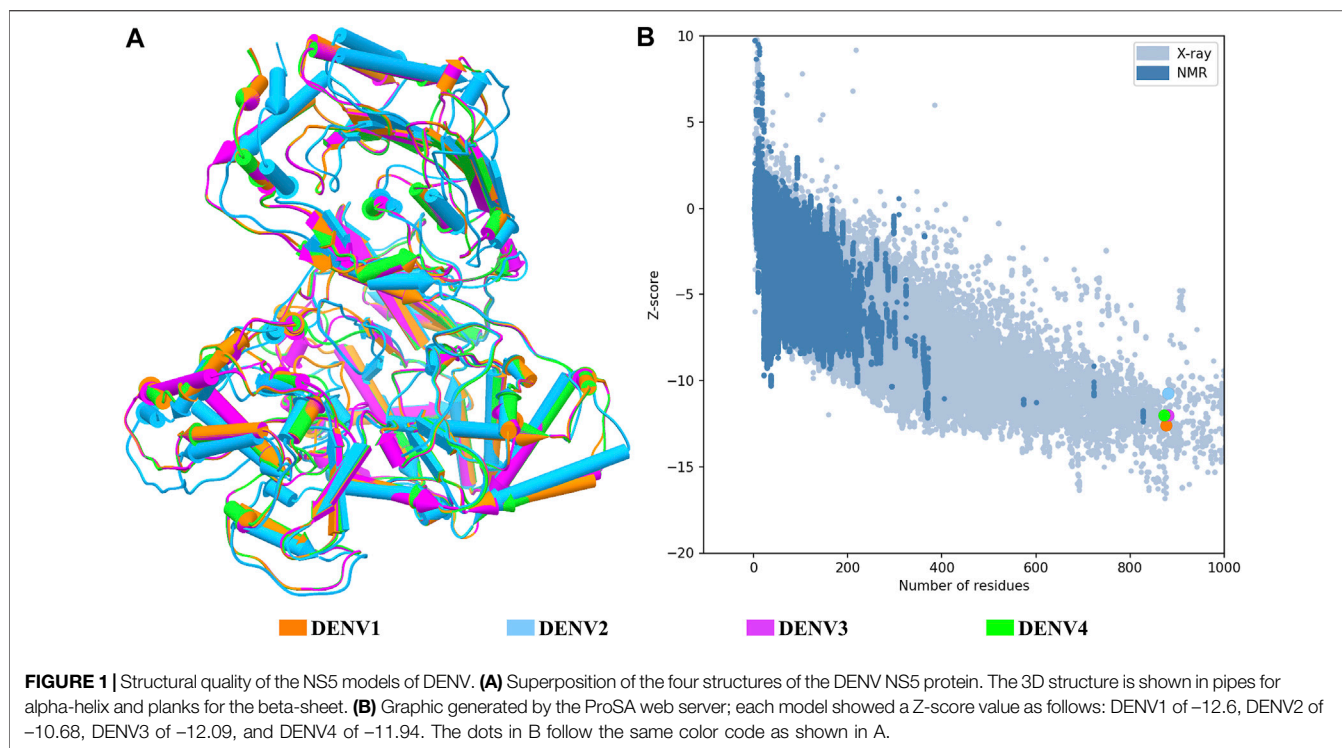
Initially, we selected the candidate compounds based on two filters: 1) multi-domain binding compounds (compounds that were docked to both the MTase and RdRp domains) and 2) single-domain binding compounds or those that were docked only to one domain. These analyses were performed using a classification script in the R package (R Foundation for Statistical Computing, s.f.). We applied a strict parameter to select compounds docked to the same region of the NS5 protein in the four DENV serotypes, which should theoretically increase their spectrum of activity and antiviral potential. The selected compounds were also subjected to other filters, including compliance to Lipinski's rules, solubility, gastrointestinal absorption, and prediction of toxicological risks (Figure 1).

Compliance to Lipinski's Rules, Solubility, Gastrointestinal Absorption, and Toxicological Risk Predictions

The prediction of Lipinski's rules, solubility, gastrointestinal absorption, and toxicological risk was performed using the SwissADME web server (Daina et al., 2017). For compliance with Lipinski's rules, this tool provides a qualification of *Yes* or *No*, accompanied by the number of rules violated. We discarded all the compounds that presented at least one violation. For solubility prediction, we used three predictors in SwissADME, which yields a qualification of insoluble, poorly, moderately, soluble, very, and highly. We decided to assign a score for each prediction as follows: insoluble and poorly, a value of 1; moderately and soluble, a value of 2; and very and highly, a value of 3. We discarded all the compounds that presented a value ≤ 5 . In order to estimate the gastrointestinal absorption of each query compound, we accepted only compounds with a high predicted gastrointestinal absorption. Finally, we used DataWarrior software (Sander et al., 2015), ProToxII (Banerjee et al., 2018), and CarcinoPred-EL (Zhang et al., 2017) web servers to predict several types of toxicological risks, including possible mutagenic, tumorigenic, reproductive, irritant, hepatotoxic, immunotoxic, cytotoxic, and carcinogenic risks. We discarded the compounds that presented three or more toxicological risks.

Molecular Dynamics Simulations and Binding Free-Energy Calculations

Molecular dynamics simulations were performed using GROMACS 2019 software (Berendsen et al., 1995). The complexes formed by the four DENV serotypes and the best selected compound per each binding site were taken as initial coordinates and were simulated. A total of 20 simulations were carried out. For the protein, the amber ff99SB-ILDN force field was used (Lindorff-Larsen et al., 2010); the ligands were parameterized using the general AMBER force field (Wang et al., 2004) and ACPYPE web server (based on ANTECHAMBER, <https://www.bio2byte.be/acpype/>), by which the parameters of the ligands to work in GROMACS are obtained (Sousa da Silva and Vranken, 2012). Each protein–ligand complex was solvated using the TIP3P water model, and its



charges were neutralized using $\text{Na}^+ \text{Cl}^-$ ions, and an excess of ions were added to reach a concentration of 0.15 M of NaCl. Each system was subjected to an energy minimization stage of 50,000 steps, followed by an NVT equilibration for 250 ps using a temperature of 310 K. Then, a series of equilibrations were performed under NPT conditions using a pressure of 1 bar, with decreasing restrictions on the heavy atoms of the protein and the ligand for 250 ps each (1,000, 100, 10, and 1 kJ/mol*nm²). In general, all systems were subjected to one energy minimization stage and five equilibration stages. The systems were subjected to a production stage for a total of 40 ns, in which a V-rescale thermostat and a Parrinello–Rahman barostat were implemented. A time-step of 2 fs was used. Once the simulations were finished, the RMSD, root mean square fluctuation (RMSF), radius of gyration (Rg), and contacts plots of each protein–ligand complex were obtained through the *gmx_rms*, *gmx_rmsf*, *gmx_gyr*, and *gmx_mindist* modules contained in the GROMACS package, respectively. Then, we calculated the binding free-energy (ΔG_{bind}) using the molecular mechanics/Poisson–Boltzmann surface area (MM/PBSA) method, available in the tool *gmx_MMPBSA* (Miller et al., 2012; Valdés et al., 2021). The ΔG_{bind} between a protein and ligand can be calculated as seen in the following equations (Hou et al., 2011):

$$\Delta G_{\text{bind}} = \Delta H - T\Delta S \approx \Delta E_{\text{MM}} + \Delta G_{\text{solvation}} - T\Delta S; \quad (1)$$

$$\Delta E_{\text{MM}} = \Delta E_{\text{internal}} + \Delta E_{\text{electrostatic}} + \Delta E_{\text{van der Waals}}; \quad (2)$$

$$\Delta G_{\text{solvation}} = \Delta G_{\text{PB/GB}} + \Delta G_{\text{SA}}, \quad (3)$$

where ΔE_{MM} represents the contribution of MM energy, and it can be obtained from the force field implemented in the MD

simulations. $\Delta G_{\text{solvation}}$ represents the solvation energy given by the sum of the polar contribution ($\Delta G_{\text{PB/GB}}$) and nonpolar contribution (ΔG_{SA}). $\Delta G_{\text{PB/GB}}$ can be obtained using the Poisson–Boltzmann or generalized-Born models, while ΔG_{SA} can be calculated by the solvent accessible surface area (SASA) (Hou et al., 2011; Genheden and Ryde., 2015). The term $T\Delta S$ can be added to refine the predictions. For ΔG_{bind} calculations, frames from the 10 ns to the end were taken for each 30 ps, for a total of 1,000 frames per each trajectory.

Pharmacophore Modeling

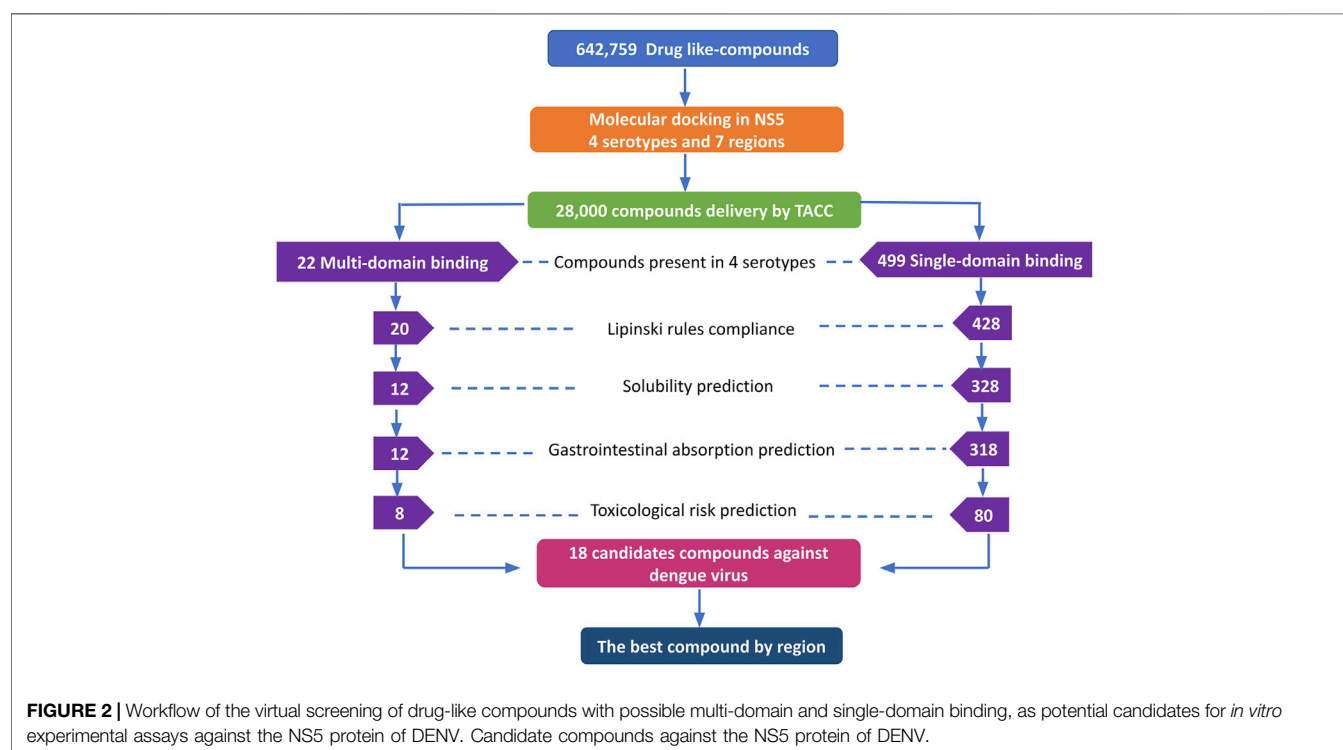
Once the list of the best compounds that passed all the previous filters was obtained, we used the binding poses on the four serotypes of the compounds with the best binding-free energy scores, and we generated a model of the pharmacophore resultant of a possible inhibitor for each binding site separately. For that, we use the PharmaGist web server (<http://bioinfo3d.cs.tau.ac.il/PharmaGist/>) (Schneidman-Duhovny et al., 2008). This allows a rational design of molecules, by identifying modifications that can improve the affinity of each compound for its binding site.

Cytotoxicity Assay in Huh-7 Cells

The cytotoxic effect of acquired compounds was evaluated on the Huh-7 cell line. A total of 2×10^4 cells were subcultured per well in 96-well plates and incubated for 24 h at 37°C and 5% CO₂. Once the cells reached 80% confluence, each of the compounds was added in serial concentrations from 0.7 to 50 μM and incubated again under the previously described conditions for 24 h. The medium was then replaced, and 15 μL of the staining solution with tetrazolium salt (MTT) was added (CellTiter 96 Non-Radioactive Cell Proliferation Assay kit, Reference G4001,

TABLE 1 | Percentages of sequence identity and structural RMSD (Å) between the NS5 protein of the four DENV serotypes.

NS5 serotype	DENV1		DENV2		DENV3		DENV4	
	Sequence identity	RMSD	Sequence identity	RMSD	Sequence identity	RMSD	Sequence identity	RMSD
DENV1	100	0	80.07	0.90	81.71	0.15	75.59	0.20
DENV2	80.07	0.90	100	0	79.42	0.91	74.38	0.89
DENV3	81.71	0.15	79.42	0.91	100	0	77.08	0.21
DENV4	75.59	0.20	74.38	0.89	77.08	0.21	100	0



Promega) and incubated again for 4 h under the previously mentioned conditions. After incubation, 100 μ L of solubilization solution was added and incubated again for an additional 1 h. The contents of the wells were mixed until obtaining uniformity in the coloration, and finally the absorbance of each well was read at 570 nm. As a control in the test, 0.3% dimethylsulfoxide (DMSO) was used, which corresponds to the vehicle control (solution in which the compounds are dissolved at a concentration of 50 μ M), and the cellular control was included, which corresponds to untreated Huh-7 cells. For data analysis, the absorbance of the medium was subtracted, and the percentage of cell viability was calculated by applying the following formula:

$$\% \text{ Cell Viability} = \frac{\text{Sample Abs}}{\text{Control Abs}} \times 100,$$

where

- Sample Abs corresponds to the absorbance value of each well with cells and treatment.
- Control Abs corresponds to the value of the absorbance of the wells with cells without treatment.

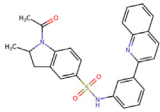
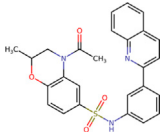
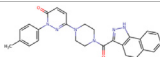
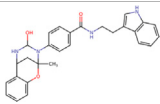
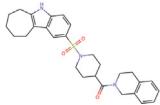
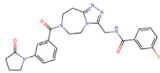
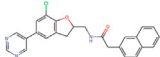
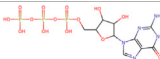
The data were evaluated by the Kruskal–Wallis test, with a comparison by Dunn's test. All data were analyzed in GraphPad Prism 6.0 software. A *p* value <0.05 was considered statistically significant.

RESULTS

Structural Modeling of DENV1–4 NS5 Proteins

The structural models of the NS5 proteins of DENV serotypes 1 and 4 were obtained by homology modeling. The models were generated from consensus sequences obtained with BioEdit using the sequences reported in the Virus Variation database (Hatcher et al., 2017) for each serotype. We found 2097 sequences of DENV1, 1559 of DENV2, and 370 of DENV4. The NS5 of DENV3 showed a high percentage of structural and sequence identity with the other serotypes of DENV as can be seen in Table 1. The sequence alignment is shown in Supplementary Figure S1, in which some relevant regions are highlighted.

TABLE 2 | Best compounds obtained for the GTP-binding site of the four DENV serotypes. The best compound is highlighted in bold.

Compound	Molecular structure	Docking score (kcal/mol)			
		DENV1	DENV2	DENV3	DENV4
1md ZINC15827835		-9.3	-9.3	-9.5	-8.9
2md ZINC15827831		-9.1	-9.1	-9.5	-8.7
3md ZINC15730200		-9.0	-8.8	-8.9	-8.8
4md ZINC9835726		-8.8	-8.6	-9.1	-8.8
5md ZINC9268549		-9.1	-8.8	-9.2	-9.3
1sd ZINC12122852		-9.2	-9.2	-9.4	-9.0
2sd ZINC11971657		-9.0	-9.0	-9.2	-9.0
GTP ^a		-7.9	-7.7	-7.9	-7.8

^aCrystal control: natural substrate of the MTase domain crystallized with the PDB accession code: 4V0R (Zhao et al., 2015a).

Once all the full-length models were obtained, the general quality was analyzed through the estimation of GMQE (Global Model Quality Estimate) and global QMEANDisCo (Qualitative Model Energy Analysis) provided in the SWISS-MODEL web server (overall measure of model quality between 0 and 1), the stereochemical quality by Ramachandran plot, and the structural quality based on the Z-score value, for each model. Based on the quality estimates, the following GMQE and QMEANDisCo values were found for the NS5 models of each serotype: GMQE 0.83 for DENV1 and DENV2, 0.85 for DENV3, and 0.82 for DENV4, and QMEANDisCo 0.82, 0.81, 0.83, and 0.81 ± 0.05 for DENV1, DENV2, DENV3, and DENV4, respectively. The values obtained are close to 1, which indicates favorable quality for these models. On the other hand, the Ramachandran plots (not shown) for these structures suggested the following percentages for the amino acid residues that are located in the favored regions: 91.6% for DENV1, 96.1% for DENV2, 98.3% for DENV3, and 97.6% for DENV4.

Figure 1A shows the four models after a structural alignment. In general, the models showed low RMSD values between them, which supports the idea of designing compounds with activity on the four dengue serotypes. Additionally, **Figure 1B** shows the spectrum provided by the ProSA web server that locates all the Z-scores for the structures that have been resolved by X-ray and NMR and shows the obtained values for the four models of the NS5 protein of DENV.

Virtual Screening and Compounds Selection

In this study, we used the DrugDiscovery@TACC web portal of the Texas Advanced Computing Center (<https://portal.tacc.utexas.edu>) and screened the largest library (Lgr), which contained approximately 642,759 compounds with structural characteristics similar to drugs. The regions with a crystallized

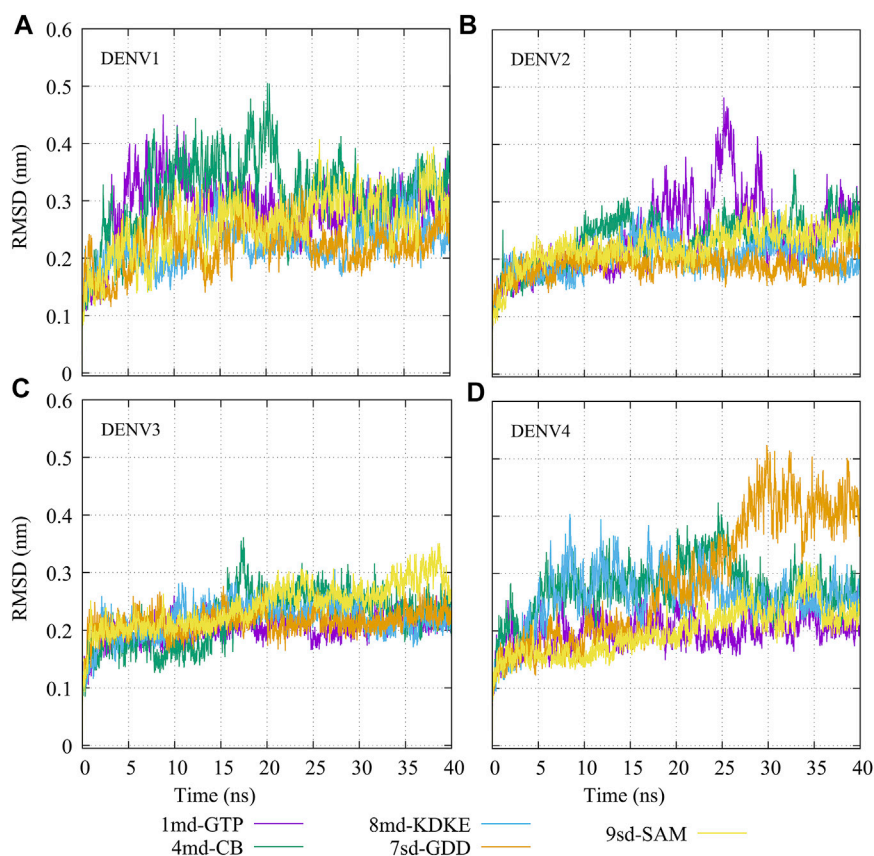


FIGURE 3 | RMSD as a function of the time for the protein–ligand complexes formed by the best compound of each binding site. **(A)** DENV1, **(B)** DENV2, **(C)** DENV3, and **(D)** DENV4.

ligand, such as GTP-binding site, SAM-binding site, and GDD motif, were validated using re-docking (**Supplementary Figure S2**). The similarity between these structures was evidenced, with a RMSD value between the experimental and the crystallographic pose of 1.545 Å, 1.707 Å, and 0.884 Å for GTP-binding site, SAM-binding site, and GDD motif, respectively.

The DrugDiscovery@TACC web portal provided a list of the first 1,000 compounds for each virtual screening run, ranked by their binding-free energy score calculated with AutoDock Vina software. In total, we studied seven regions in the NS5 protein (3 in the MTase domain and 4 in the RdRp domain) of the four DENV serotypes, retrieving a total of 28,000 compounds that were filtered according to our selection criteria. A simple R script was designed to perform the first filter that consisted in: 1) selecting all the compounds that were present in the four serotypes and docked to a region of the MTase domain as well as to a region of the RdRp domain and 2) selecting all the compounds that were present in all four DENV serotypes, but this time, only in one domain of the NS5 protein (MTase or RdRp). This selection resulted in 22 compounds with possible multi-domain binding and 499 compounds with possible single-domain binding in the NS5 protein of DENV (**Figure 2**).

Once we obtained the compounds for each strategy, we performed each of our filters. For prediction of compliance or

violation of Lipinski's rules, resulted in 20 compounds for strategy 1 and 428 compounds for strategy 2. After evaluating our scores for solubility prediction, strategy 1 was reduced to 12 compounds and strategy 2 to 328 compounds. All the compounds of strategy 1 presented high gastrointestinal absorption, which did not reduce the list, while for strategy 2 the list was reduced to 318 compounds. Toxicological risks prediction resulted in the removal of 4 compounds from strategy 1 according to our exclusion criteria, resulting in a list of 8 compounds. Meanwhile, strategy 2 was reduced to 80 compounds that did not present any predicted toxicological risks. Finally, based on the binding score from molecular docking, we postulate the best compounds from both strategies, as potential candidates to be inhibitors of DENV NS5 protein *in vitro* assays, resulting 8 compounds with possible multi-domain binding (strategy 1) and the top 10 compounds with possible single-domain binding (strategy 2).

The 8 compounds resulting from strategy 1 (multi-domain) showed mixed binding between four regions of the NS5 protein of the four DENV serotypes, such as the GTP-binding site, cavity B, KDKE tetrad, and GDD motif. In particular, two regions were located in the MTase domain (KDKE and GTP) and two in the RdRp domain (CB and GDD). For strategy 2 (single-domain), we obtained 80 compounds that passed the filters of the structural

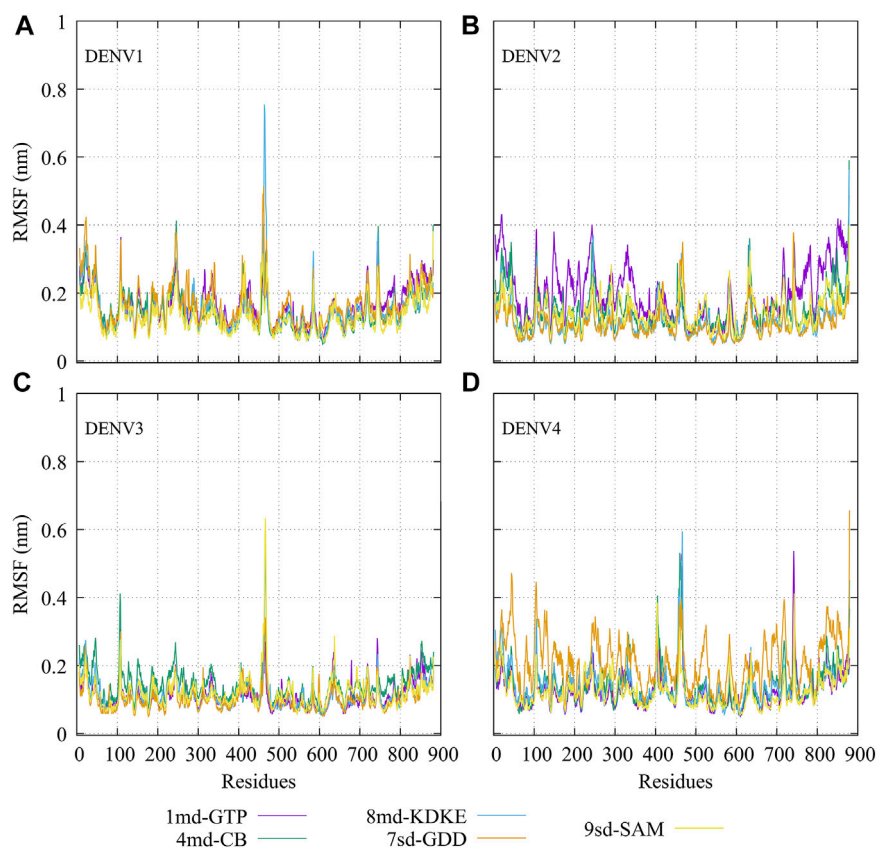


FIGURE 4 | RMSF as a function of numbers of residues for the protein–ligand complexes formed by the best compound of each binding site in (A) DENV1, (B) DENV2, (C) DENV3, and (D) DENV4.

predictions. These compounds showed dockings in regions such as the GTP-binding site, cavity B, KDKE tetrad, GDD motif, and SAM-binding site. The single-domain criteria allowed obtaining a much larger list than the list obtained with strategy 1. However, we only took the best 10 compounds (2 per region) for the subsequent analyses. Accordingly, we postulate a short list of 18 compounds, 8 with double binding sites and 10 compounds with single binding sites in the NS5 protein of the four DENV serotypes. The best compounds for the GTP-binding site are shown in **Table 2**, with *1md* (ID ZINC15827835) being the compound with the best affinity results.

Molecular Dynamics Simulations and Binding-Free Energy Calculations

The molecular dynamics simulations aimed to sample the protein–ligand complexes formed by the best compound at each binding site for 40 ns. Then, with these obtained trajectories and by means of the MM/PBSA method, we calculated a new binding-free energy, which in principle is more computationally robust than the score obtained in the molecular docking. This was in order to analyze on which serotype the interaction of each compound would be stronger. In addition, we obtained the RMSD, RMSF, Rg, and contact map graphs for each simulated complex. In

Figure 3, the RMSD values for the NS5 protein of the four DENV serotypes are shown. In brief, the RMSD measures the structural deviation of the protein along the time regarding its initial conformation. For complexes with DENV1 (**Figure 3A**), an approximate RMSD range of 0.2–0.4 nm was obtained. Compound *4md* (bound in cavity B) presented greater fluctuations at the beginning of the simulation than the other compounds, reaching values very close to 0.5 nm. For the complexes with DENV2 (**Figure 3B**), the approximate range of RMSD was smaller than that of the DENV1 complexes, being between 0.15 and 0.3 nm. Compound *1md* (bound at the GTP-binding site) shows greater fluctuations in the interval from 17 to 30 ns, reaching values less than 0.5 nm, for the rest of the time, its behavior was similar to the other DENV2 complexes. For complexes with DENV3 (**Figure 3C**), the approximate range of RMSD was narrower, being between 0.2 and 0.3 nm. For the complexes with DENV4 (**Figure 3D**), the approximate range of RMSD was 0.2–0.5 nm. All complexes except the one formed by *7sd* (bound in the GDD motif region) exhibited a plateau between 0.3 and 0.4 nm after ~27 ns. In general, the structural changes for all serotypes were less than or equal to 0.5 nm.

The RMSF plots for simulated protein–ligand complexes are shown in **Figure 4**. In brief, the RMSF measures the average deviation of the protein residues, that is, it measures the

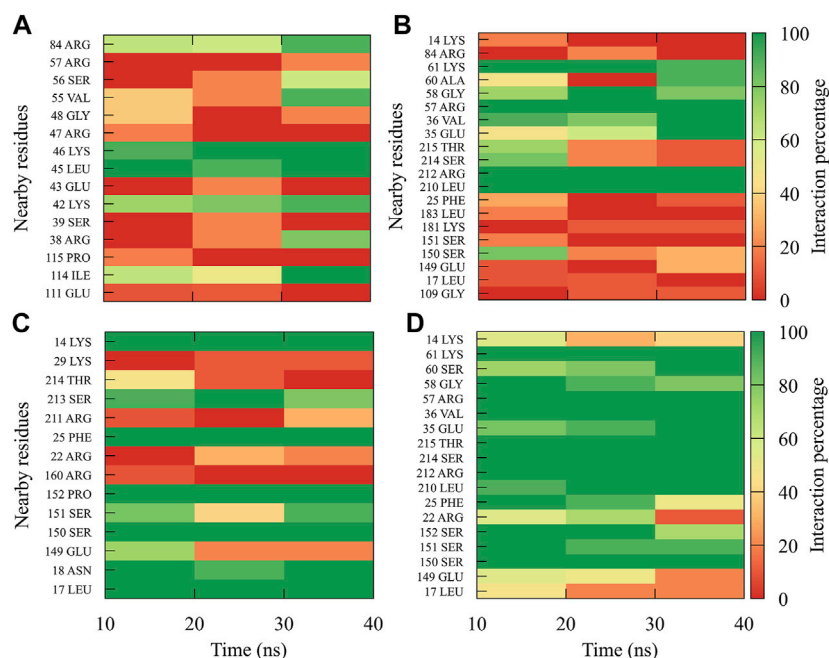


FIGURE 5 | Contact frequency maps for the compound *1md*, bound to the GTP-binding site of the (A) DENV1, (B) DENV2, (C) DENV3, and (D) DENV4. Red represents that the contact frequency was 0%, while green represents that the contact frequency was 100%.

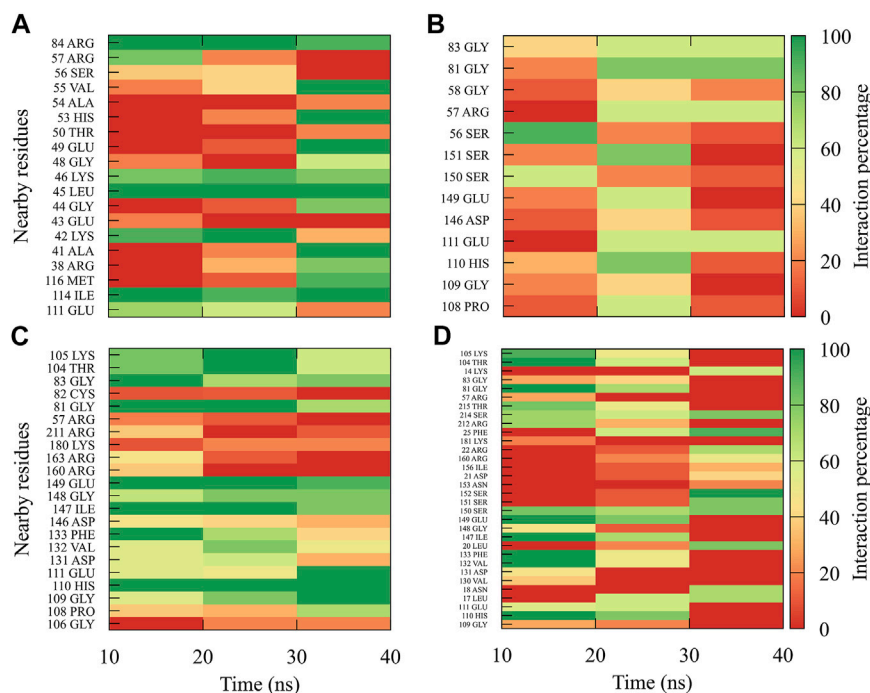


FIGURE 6 | Contact frequency maps for the compound *8md*, bound to the region of the KDKE tetrad of the (A) DENV1, (B) DENV2, (C) DENV3, and (D) DENV4. Red represents that the contact frequency was 0%, while green represents that the contact frequency was 100%.

fluctuation in their position. Thus, high values in the RMSF indicate portions in the protein with greater flexibility (Martínez, 2015). For the RMSF of the complexes with DENV1 (Figure 4A),

it was found that *8md* increased the flexibility of a short fragment (from residue 461–465) reaching a value of ~0.75 nm. For the complexes formed by DENV2 (Figure 4B), it was observed that

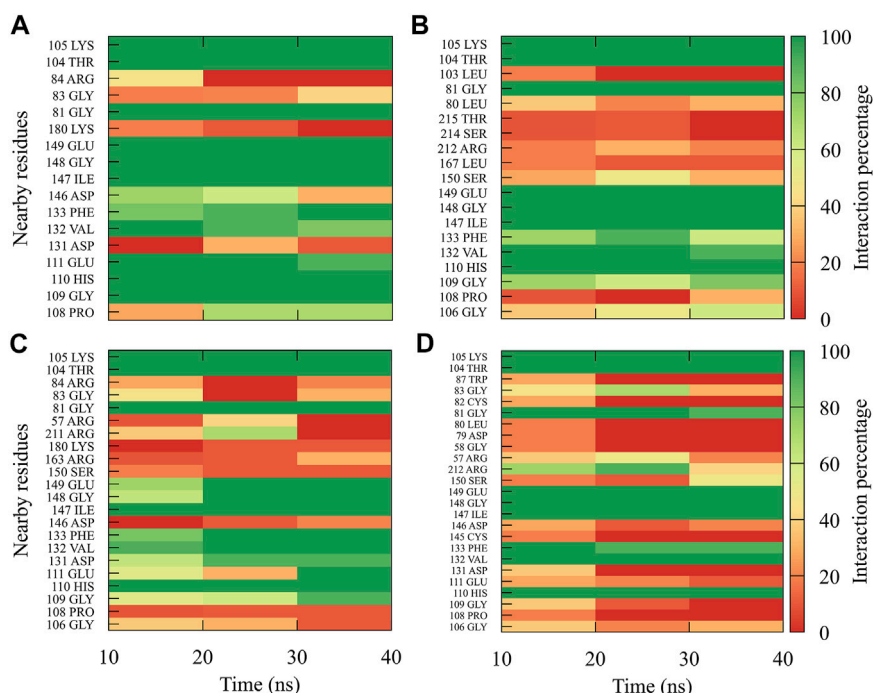


FIGURE 7 | Contact frequency maps for the compound 9sd, bound to the SAM binding-site of the (A) DENV1, (B) DENV2, (C) DENV3, and (D) DENV4. Red represents that the contact frequency was 0%, while green represents that the contact frequency was 100%.

1md caused a slight increase in the flexibility of the protein with respect to the flexibility of the other complexes, in fragments that go from the N-terminal to residue ~400 and from residue ~700 to the C-terminal. For the complexes formed by DENV3 (Figure 4C), compounds 8md and 9sd increased the flexibility of the protein in a short fragment from residue 462 to 466, reaching values close to ~0.63 nm. For the complexes formed by DENV4 (Figure 4D), 7sd presented an increase in flexibility in the protein with respect to the other complexes.

In order to dynamically track the interactions between the protein and the ligand, we calculated the average frequencies of interaction. Here, the interactions were all those contacts between the protein and the ligand at a distance less than or equal to 0.35 nm; in this way, only the strongest or closest interactions are considered. To do it dynamically, we divide the simulation time into blocks of 10 ns, obtaining 3 blocks of time (since the first 10 ns was omitted for equilibration). If a residue has an interaction percentage of 100% with the ligand, the green color will be assigned, which means that this interaction was conserved for all 10 ns of the time block.

In general, it can be observed that the interactions between compound 1md and the residues of the GTP-binding site of the four DENV serotypes (Figure 5) were both in a higher percentage and in greater quantity with DENV4 and DENV3 serotypes, and they were both in a lower percentage and in a lower quantity in DENV2 and DENV1 serotypes. 1md conserved interactions with residues such as Lys14, Leu17, Asn18, Phe25, Ser151, Leu210 (DENV2 and DENV4), and Ser213, which have been classified as important for the

stabilization of the natural substrate GTP (Geiss et al., 2009). To indicate the position of the other amino acids involved, the sequence of NS5 DENV3 was taken as reference, according to the alignment (Supplementary Figure S1).

The interactions between compound 8md and the residues in the region of the KDKE tetrad (Figure 6) suggest that the highest number of contacts with the highest frequency occurred in the DENV3 and DENV2 serotypes, while those with the lowest frequencies were DENV1 and DENV4 serotypes. 8md showed interactions with some amino acids of the KDKE tetrad and close ones, such as Asp146, Ile147, Lys180, and Thr215 (DENV4), which are considered important residues for the interaction of the natural substrate SAM (Zhou et al., 2007; Lim et al., 2011). To indicate the position of the other amino acids involved, the sequence of NS5 DENV3 was taken as reference, according to the alignment (Supplementary Figure S1).

For compound 9sd, bound to the SAM-binding site in the MTase domain, the highest frequencies of interactions were for the DENV1 and DENV2 serotypes (Figures 7A, B); therefore, the lowest frequencies of interactions were for the DENV3 and DENV4 serotypes (Figures 7C, D). Compound 9sd exhibited interactions with described amino acids important for the stabilization of SAM, such as Lys61, Arg84, Lys105, His110, Glu111, Asp131, Val132, Asp146, Ile147, Gly148, Lys180, and Glu216 (Zhou et al., 2007; Dong et al., 2008; Zhao et al., 2015a; Zhao et al., 2015c). To indicate the position of the amino acids involved, the sequence of NS5 DENV3 was taken as reference, according to the alignment (Supplementary Figure S1).

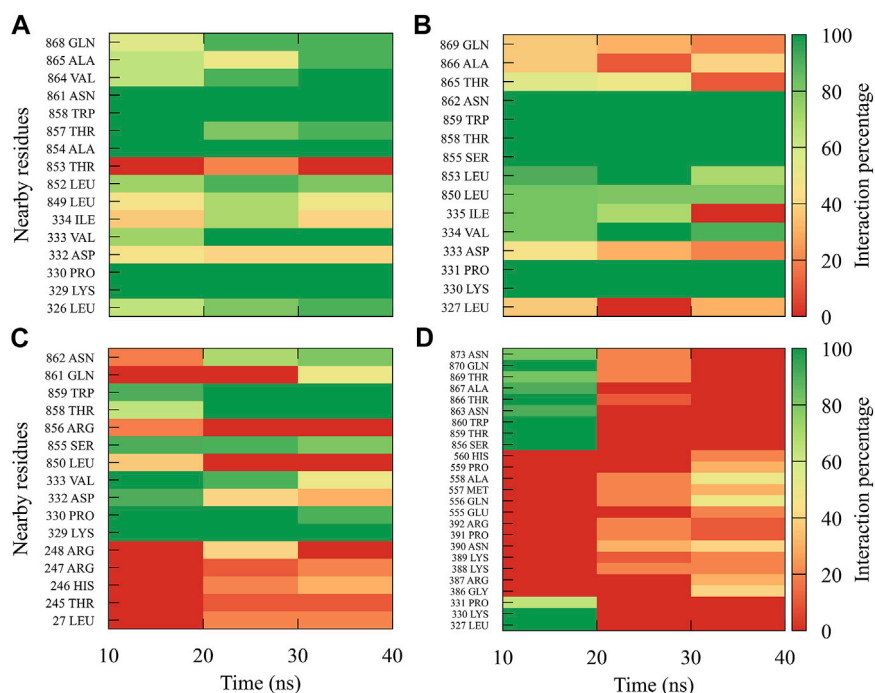


FIGURE 8 | Contact frequency maps for the compound *4md*, bound to the cavity B of the (A) DENV1, (B) DENV2, (C) DENV3, and (D) DENV4. Red represents that the contact frequency was 0%, while green represents that the contact frequency was 100%.

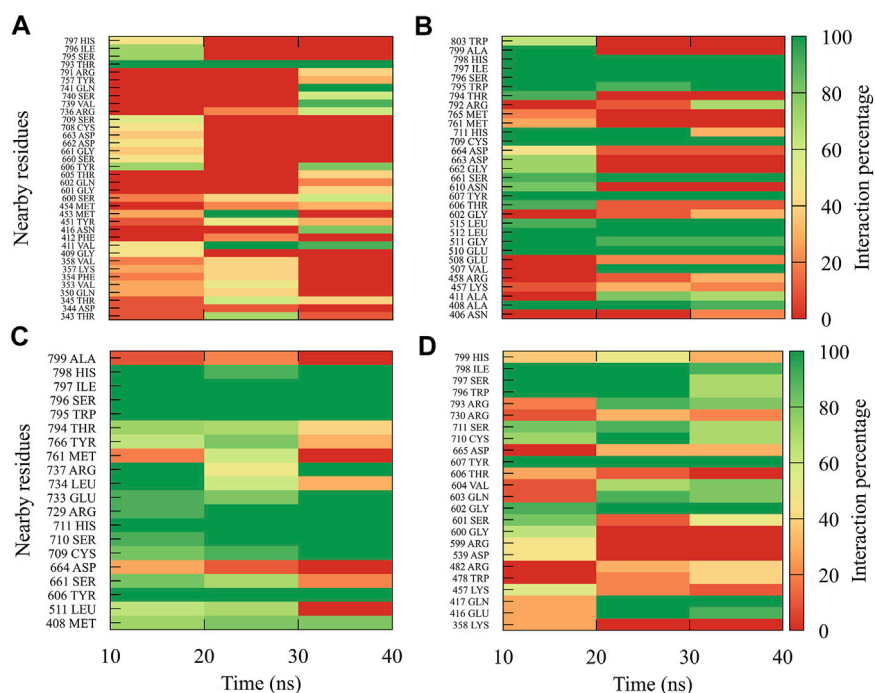


FIGURE 9 | Contact frequency maps for the compound *7sd*, bound to the region of the GDD motif in the RdRp domain of the (A) DENV1, (B) DENV2, (C) DENV3, and (D) DENV4. Red represents that the contact frequency was 0%, while green represents that the contact frequency was 100%.

TABLE 3 | Best compounds obtained for the KDKE tetrad site of the four DENV serotypes. The best compound is highlighted in bold.

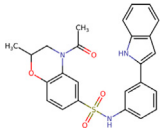
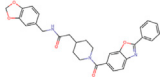
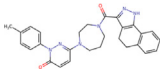
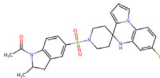
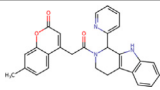
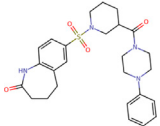
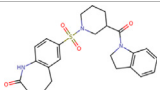
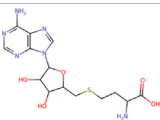
Compound	Molecular structure	Docking score (kcal/mol)			
		DENV1	DENV2	DENV3	DENV4
6md ZINC33106115		-9.7	-11.1	-9.9	-11.2
7md ZINC15869753		-9.8	-11.0	-9.8	-11.0
8md ZINC15730188		-10.4	-11.5	-10.8	-11.7
5sd ZINC21152560		-10.0	-11.1	-10.2	-11.0
6sd ZINC12307816		-9.8	-11.0	-10.0	-11.1

TABLE 4 | Best compounds obtained for the SAM-binding site of the four dengue serotypes. The best compound is highlighted in bold.

Compound	Molecular structure	Docking score (kcal/mol)			
		DENV1	DENV2	DENV3	DENV4
9sd ZINC20943220		-9.7	-11.3	-9.7	-11.2
10sd ZINC20943169		-9.7	-11.1	-9.6	-11.1
SAH ^a		-7.8	-7.9	-8.0	-7.9

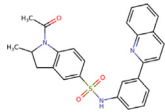
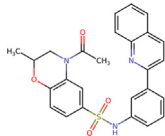
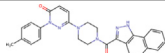
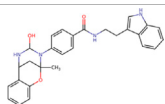
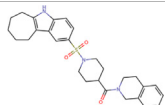
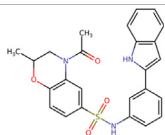
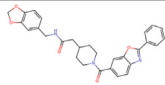
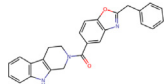
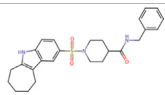
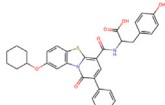
^aCrystal control: natural substrate of the MTase domain crystallized with the PDB accession code: 5JJR (Lim et al., 2016).

For compound *4md*, bound to cavity B of the RdRp domain, the highest frequencies and numbers of contacts were for the DENV1 and DENV2 serotypes (**Figures 8A, B**), while the lowest frequencies and numbers of contacts were for the DENV3 and DENV4 serotypes (**Figures 8C, D**). *4md* presented interactions with amino acids from cavity B, such as Leu327, Lys329, Pro330, Asp332, Thr858, Trp859, Asn862, Ile863, Ala866, and Gln869, which have been described as interacting with other inhibitors

designed for this cavity (Zou et al., 2011; Kaptein et al., 2018; Cannalire et al., 2020). To indicate the position of the amino acids involved, the sequence of NS5 DENV3 was taken as reference, according to the alignment (**Supplementary Figure S1**).

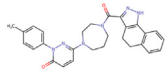
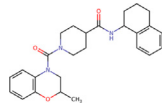
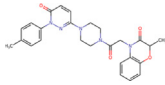
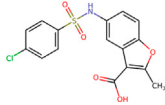
For compound *7sd*, bound to the GDD motif region of the RdRp domain, the highest frequencies and numbers of interactions were with DENV3 and DENV2 serotypes (**Figures 9B, C**), and the lowest frequencies of interactions were for

TABLE 5 | Best compounds obtained for cavity B of the four dengue serotypes. The best compound is highlighted in bold.

Compound	Molecular structure	Docking score (kcal/mol)			
		DENV1	DENV2	DENV3	DENV4
1md ZINC15827835		-8.9	-8.8	-8.8	-9.0
2md ZINC15827831		-9.1	-9.0	-9.0	-9.0
3md ZINC15730200		-8.6	-8.8	-8.9	-8.5
4md ZINC9835726		-9.6	-9.5	-9.5	-9.2
5md ZINC9268549		-8.6	-8.7	-8.7	-8.8
6md ZINC33106115		-8.8	-8.8	-8.8	-8.7
7md ZINC15869753		-8.8	-8.4	-8.5	-8.6
3sd ZINC12705528		-9.0	-9.1	-9.0	-9.1
4sd ZINC09268580		-9.4	-9.0	-9.1	-8.9
PBTZ1 ^a		-7.9	-7.6	-7.5	-7.6

^aCrystal control: compound with reported anti-dengue activity and with possible binding mode in cavity B (Cannalire et al., 2020).

TABLE 6 | Best compounds obtained for the GDD motif of the four dengue serotypes. The best compound is highlighted in bold.

Compound	Molecular structure	Docking score (kcal/mol)			
		DENV1	DENV2	DENV3	DENV4
8md ZINC15730188		-9.9	-10.4	-10.3	-9.8
7sd ZINC09405884		-10.3	-10.6	-9.8	-10.2
8sd ZINC21887378		-10.1	-10.2	-9.7	-10.4
NITD107 ^a		-8.4	-9.2	-8.0	-8.5

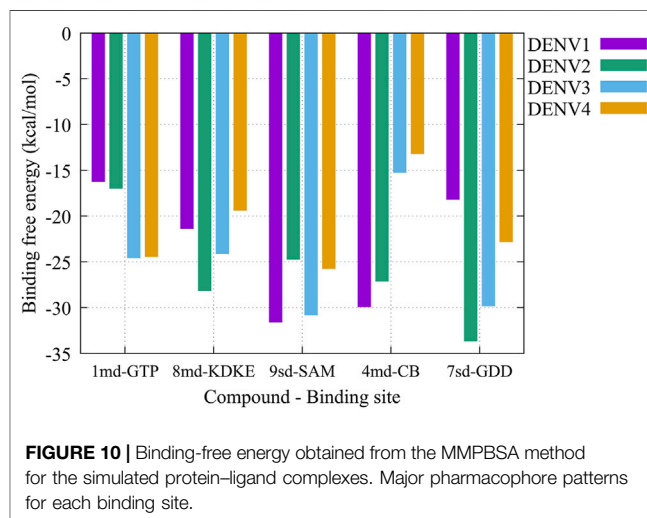
^aCrystal control: inhibitor of the RdRp domain of DENV3 crystallized with the PDB accession code: 3VWS (Noble et al., 2013).

DENV1 and DENV4 serotypes (**Figures 9A, D**). Also, *7sd* conserved interactions with some amino acids that have been involved in the stabilization of the inhibitor NITD107 crystallized in the GDD motif region and as important for the polymerase activity of the RdRp domain (Noble et al., 2013; Klema et al., 2016), such as Val411, Phe412, Val603, Thr605, Tyr606, Asp663, Asp664, Trp795, and Ile797. To indicate the position of the amino acids involved, the sequence of NS5 DENV3 was taken as reference, according to the alignment (**Supplementary Figure S1**).

On the other hand, to analyze possible preferences of the compounds for any of the serotypes, we calculated the binding-free energy from molecular dynamics simulations (**Figure 10**). Our results suggest that compound *1md* prefers the DENV3 and DENV4 serotypes over the other two serotypes, with DENV3 and DENV4 being the serotypes with the highest ΔG_{bind} values. For compound *8md*, the two best ΔG_{bind} values were for the DENV2 and DENV3 serotypes. For compound *9sd*, the best ΔG_{bind} values were for DENV1 and DENV3 serotypes. For compound *4md*, the best ΔG_{bind} values were for DENV1 and DENV2. Finally, for compound *7sd*, the best values were for the DENV2 and DENV3 serotypes. Thus, in general, we can conclude that the serotypes with the best ΔG_{bind} values were the DENV2 and DENV3 serotypes.

The pharmacophore approximation models were presented as an additional result after selecting the best compounds for each region. The idea behind this analysis was to take advantage of the predicted poses of the best compounds and obtain a pattern of chemical characteristics (**Supplementary Figure S3**) that can suggest structural modifications of the ligands that can be used to design new molecules in future research (**Figure 11**).

The prediction of the pharmacophore patterns in the GTP-binding site was predicted using compounds *1md*, *2md*, *5md*,



and *1sd* (**Table 2**), among which *1md* (ID ZINC15827835) being the compound with the best affinity value. In general, four pharmacophore patterns are observed (**Supplementary Figure S3A**) to consider in compounds with affinity for the GTP-binding site: (I) a highly hydrophobic region, (II) a highly polar region, (III) an aromatic ring with polar and hydrophobic radicals, and (IV) an aromatic region that could be represented by a two-ring fragment, where both can have polar atoms. Based on this, we postulate 3 sites in *1md*, in which 2 polar-type and 1 hydrophobic modifications can be explored (**Figure 11A**).

The best compounds for the KDKE tetrad are shown in **Table 3**, where *8md* (ID ZINC15730188) being the compound with the best affinity results. The pharmacophore patterns at this

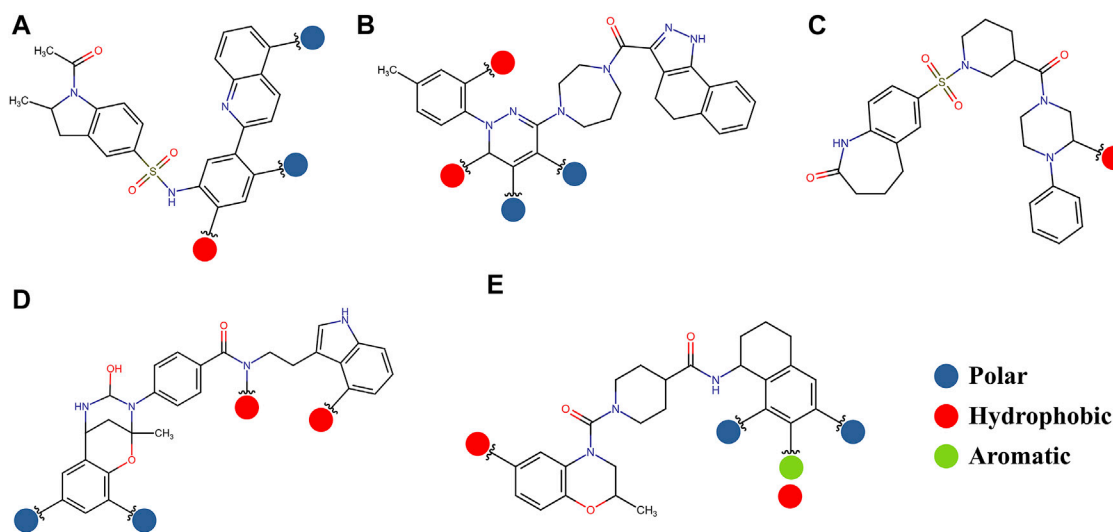


FIGURE 11 | Possible structural modifications for the compound *1md* bound to the GTP-binding site (**A**), compound *8md* bound to the region of the KDKE tetrad (**B**), compound *9sd* bound to the SAM-binding site (**C**), compound *4md* bound to the cavity B (**D**), and compound *7sd* bound to the region of the GDD motif (**E**). The polar, hydrophobic, and aromatic patterns are shown as blue, red, and green spheres, respectively. These spheres show the position on each compound where they can be located.

binding site were predicted using compounds *6md*, *8md*, *5sd*, and *6sd* (compounds with the best affinity values). We highlight four pharmacophore patterns for compounds with possible binding in the region of the KDKE tetrad (**Supplementary Figure S3B**): 1) an aromatic fragment with hydrophobic substituents, 2) an aromatic region with hydrophobic and polar radicals, 3) a highly polar fragment, 4) and a voluminous fragment that can be described by a three-ring fragment, an aromatic one with heteroatoms, followed by a hydrophobic one, and finally a fully aromatic one. Based on this, we postulate four sites in *8md*, in which 2 polar-type and 2 hydrophobic modifications can be explored (**Figure 11B**).

Among our results, only two compounds, *9sd* (ID ZINC20943220) and *10sd* (ID ZINC20943169), were in the SAM-binding site (**Table 4**), both presented affinities of similar values, although *9sd* has the best results. The both compounds were used for pharmacophore prediction because *9sd* and *10sd* have very similar chemical structures and the pharmacophore pattern mostly matches for both compounds. Region I, which is described by the identical part between both molecules, is conserved. Region II suggests that a more elongated aromatic group would be more representative, coinciding with the fact that it is a phenylpiperazine fragment for *9sd* and it is an indole ring for *10sd*, which is a smaller fragment (**Supplementary Figure S3C**). Based on this, we postulate only one site in *9sd*, in which hydrophobic-type modifications can be explored (**Figure 11C**).

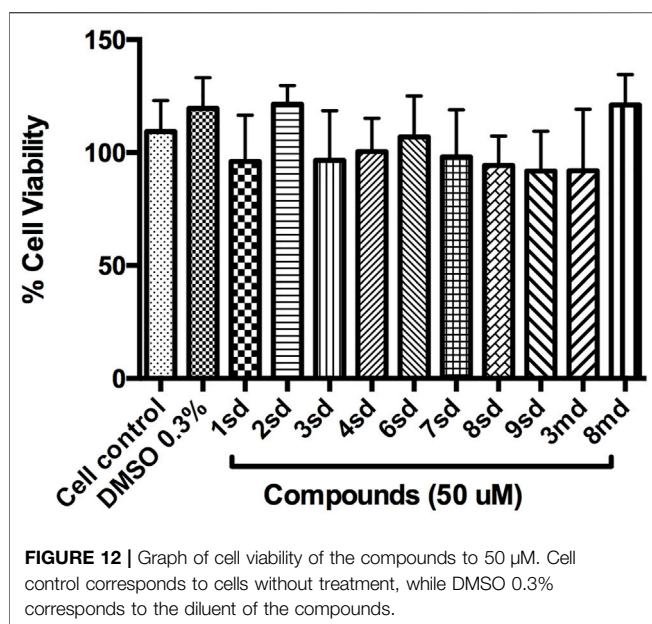
The best compounds for cavity B are shown in **Table 5**, where *4md* (ID ZINC9835726) being the compound with the best binding-free energy score. Then, in order to describe the chemical environment for inhibitors with binding in cavity B, we calculated a possible pharmacophore using compounds *2md*,

4md, *3sd*, and *4sd*. In general, five chemical patterns can be described to consider in the structure of the future ligand with binding in cavity B: 1) an aromatic region with polar substituents, 2) a polar region with small hydrophobic groups, 3) an aromatic fragment as a linker, 4) a highly polar region with a hydrophobic character, and 5) a voluminous hydrophobic region (**Supplementary Figure S3D**). Based on this, we postulate four sites in *4md*, in which 2 polar-type and 2 hydrophobic modifications can be explored (**Figure 11D**).

For the case of the GDD motif site, only 3 compounds were obtained (**Table 6**), of which *7ds* was the one that presented the best affinity scores. Therefore, the 3 compounds, *8md*, *7sd*, and *8sd*, were used for the prediction of pharmacophores. For this case, three regions are described: 1) a bulky region, which can be described by fragments of two rings, one aromatic ring with heteroatoms, and the other with one hydrophobic ring; 2) a region with a hydrophobic ring and an aromatic ring with substituents polar, thus forming a fragment of two rings; and finally 3) a region that suggests that aromatic-type modifications with hydrophobic substituents can be considered on the aromatic ring of region II (**Supplementary Figure S3E**). Based on this, we postulate four sites in *7sd*, in which 2 modifications of the polar type and 1 modification of the hydrophobic type were explored, and for this case we found that an aromatic fragment with hydrophobic substitutions could be included, for example, a benzene ring with hydrophobic groups in the para position (**Figure 11E**).

Cytotoxicity Assay in Huh-7 Cells

Initially, the cytotoxic effect of 10 of the 18 candidate compounds identified in this study was evaluated on the Huh-7 cell line. Cell



viability was close to 100% in the presence of the evaluated compounds (Figure 12), without statistically significant differences with respect to the control, indicating that they present CC_{50} greater than 50 μ M in Huh-7 cells.

DISCUSSION

In recent years, many efforts have been made to search for effective antiviral compounds as a treatment against DENV. Among the identified compounds, only a few have been further evaluated in preclinical or clinical trials (Troost and Smit, 2020). The difficulty that has arisen in finding effective treatments is related to the toxicity of some compounds evaluated, as has been reported in other investigations (Lim et al., 2013b; Tay et al., 2013; Cailliet-Saguy et al., 2014; Fusco and Chung, 2014; Lim et al., 2015; García et al., 2017), which is why it is important to continue searching for drugs. Computational tools have a large impact in drug discovery because of its fast and promising results (Halim et al., 2017). In the discovery, design, and development of new compounds with putative biological activity, it is common to find research papers that begin with listings or libraries of hundreds of thousands of compounds, which are subsequently reduced to lists of just a few compounds; for this, methodologies based on virtual screening have been proved to be very useful (Jenwitheesuk et al., 2008; Westermaier et al., 2015; Seyedi et al., 2016).

Target identification and validation is the first key stage in the drug-discovery pipeline (Li et al., 2008). In the present study, the DENV NS5 protein was selected as a target. NS5 represents a promising antiviral target (Tripathi and Shrivastava, 2018) to design specific inhibitors with low toxicity (El Sahili and Lescar, 2017). We obtained the three-dimensional structures for all dengue virus serotypes, based on PDB accession code: 5JJR. The identity of sequences between NS5 was consistent with that of the sequence identities reported in other studies (García et al., 2017; Sampath and

Padmanabhan, 2009). This crystal structure was an adequate template for the first stage of this study, finding values close to 1 in the measurement of the quality of each model according to GMQE and QMEANDisCo and valid stereochemical quality, confirmed with the Ramachandran plot. The structural alignment showed small RMSD values, being the lowest between DENV1–DENV4 and the biggest between DENV2–DENV3. The difference between these global topologies is related to the differences about the sequence (Saw et al., 2015). It has been shown that NS5 adopts multiple conformations owing to its flexible linker and that DENV4 NS5 is more compact and less flexible compared with NS5 from DENV1 to DENV3 (Saw et al., 2015; Subramanian et al., 2016). A ten-residue linker in the NS5 protein is important in communication between the MTase and RdRp domain (Zhao et al., 2015b; Saw et al., 2015). According to the alignment in this region, the higher sequence identity was 60% between NS5 linker DENV1–DENV3 and less between NS5 linker DENV1–DENV2, DENV1–DENV4, and DENV2–DENV4, with 30%, indicating variation in different amino acid in all serotypes, found only conserved E267 and E269 according to sequence of DENV3 (Supplementary Figure S1). Together with the high percentages of identity between sequences and the high conservation of the global topology of the protein, it is possible to think of drugs that can act on the NS5 protein in the four DENV serotypes. The Z-score values for each NS5 structure (Figure 1B) indicate that the built structures are comparable to other proteins resolved by X-Ray and NMR, indicating an acceptable overall quality, because Z-scores outside a range characteristic for native proteins indicate erroneous structures (Wiederstein and Sippl, 2007). Also, it is observed that the most distant model, with respect to NS5 DENV3, is DENV2; however, all the models obtained included most of the amino acids of NS5, allowing the models to be reliable and comparable. Additionally, the Ramachandran plot results for each NS5 allowed to continue with the virtual screening, suggesting that the structures of each model displayed a valid structural quality, according to torsion angles (Hollingsworth and Karplus, 2010).

After molecular docking in the DrugDiscovery@TACC web portal, all compounds were analyzed through physicochemical predictions and ADMET. One of the most important/common parameters in drug development is the Lipinski's rule, which plays an important role because it reveals that if the selected compounds possess the properties of possible drugs and they may be used in the future as drug candidates (Ahmad et al., 2020). As with the development of drug discovery, it was realized that it is important to filter and optimize the ADMET properties for drugs at an early stage, which has been accepted and widely used to reduce the attrition rate in drug research and development (Wu, F et al., 2020). So, starting with a total of 22 compounds with possible multi-domain binding and 499 compounds with possible single-domain binding in the NS5 protein of DENV, we came up with a list of 18 compounds with varied chemical structure (Tables 2–6). For the binding sites in which we have controls (GTP-binding site, SAM-binding site, cavity B, and GDD motif), the molecular docking results suggest that all our compounds have an affinity for the binding site greater than their respective control. In the case of GTP- and SAM-

binding sites, the results suggest a competitive-type interaction, presenting much higher affinities than natural substrates. For the case of cavity B, we used one compound reported by Cannalire et al., called PBTZ1. It presented a mean effective concentration (EC_{50}) to inhibit DENV2 replication of $2.1 \pm 0.22 \mu\text{M}$ (Cannalire et al., 2020). For the binding site of the GDD motif, we use the crystallized compound by Noble et al., in the RdRp domain of DENV3 (PDB accession code: 3VWS) called NITD107 as control. It inhibits the RdRp activity of DENV4 with an IC_{50} value of $113 \mu\text{M}$ and inhibits DENV2 replication with an IC_{50} value of $100 \mu\text{M}$ (Noble et al., 2013). Thus, in general, our molecular docking results suggest that all our compounds docked on cavity B and GDD motif region would have a better affinity than the PBTZ1 and NITD107 compounds, which in turn suggests that they, in theory, could have a better *in vitro* effect.

Subsequently, making use of the best compound for each binding site, we carried out a new calculation of the binding-free energy but this time from molecular dynamics simulations. First, our simulations show that, generally speaking, the backbone of the protein ranges roughly between RMSD values of 0.2–0.4 nm. However, for some time periods and specific serotypes, RMSD values of ~ 0.5 nm were reached. For example, for compound *7sd* in serotype DENV4. Here, it is necessary to mention that since the simulation time was only 40 ns, new simulations with a much more extensive sampling are necessary to delve into the behavior of each of the proteins once they interact with these compounds, although it has been reported that even shorter simulation times may be suitable for performing junction-free energy calculations using MMPBSA (Xu et al., 2013; Genheden and Ryde, 2015; Wang et al., 2019). In addition, other additional analysis could be useful to deepen the selectivity of these compounds (Al-Sha'er and Taha, 2010; Al-Sha'er and Taha, 2021). Recently Wu, J et al., reported the crystallographic structure of the DENV2 NS5 protein in two distant conformations (Wu, J et al., 2020): the structure with PDB accession code: 6KR2 adopts a conformation similar to the homologous protein in the Zika virus, and the structure with PDB accession code: 6KR3 adopts the same conformation as the crystallized structure for serotype DENV3 with PDB accession code: 5JJR. The RMSD value between the 6KR2 and 6KR3 crystals is ~ 0.57 nm, while the RMSD value between the 6KR2 and 5JJR crystals is ~ 0.91 nm (values calculated with the Chimera program). Zhao et al., performed molecular dynamics simulations to study the role of linker amino acid mutations in the DENV3 NS5 protein and obtained an RMSD for the wild-type protein that converges around 0.5 nm after 20 ns (Zhao et al., 2015b). So, if we consider that the DENV NS5 protein can modify its structure reaching RMSD values lower than 1 nm, it is feasible to think that our RMSD values could be considered natural fluctuations of the protein. In agreements, from the Rg plots (**Supplementary Figure S4**), it could be observed that the compaction or stiffness of all the serotypes with all the compounds oscillates in a range of 3–3.2 nm, being the complexes of serotype 3 those that presented minor changes in their values of the Rg in time with respect to the other serotypes.

The best compounds conserved interactions with important amino acids within each of the studied binding sites (**Figures 5–9**). The DENV2 and DENV3 serotypes were the ones with the highest frequencies and the highest number of contacts with the compounds. In addition, it can be observed that the DENV4 serotype tends, in a general way, to present interactions with very low frequencies with respect to the other serotypes. From now on, it could be hypothesized which will be the serotypes that present the best free binding energies with the compounds. In this order of ideas, our contact frequency results have a correlation with the results of the ΔG_{bind} calculation (**Figure 10**), since the DENV4 serotype presented the lowest frequency of interactions and also the lowest values were obtained for ΔG_{bind} . In this way, more economical computational analyses such as the frequency of interactions from classical molecular dynamics simulations could be used to filter compounds in this type of methodologies. Also, using the best compounds and analyzing their pharmacophore patterns, we postulate specific sites on the structure of the best molecule of each binding site (**Figure 11**) to be able to carry out modifications that can lead to increase the affinity for its binding site in the DENV NS5 protein.

According to our results, we found compounds with the possibility of binding to the two enzymatic domains of NS5 protein, which is interesting considering that both functions (MTase and RdRp) have been investigated as antiviral targets. Targets with multiple binding sites (prerequisites or allosteric) are of increasing importance in the drug design (Hetényi and Bálint, 2020). The exploration of multiple binding sites is of great importance in pharmacology (Hetényi and Bálint, 2020; Yuan et al., 2020) and has been a strategy considered in various studies (Hammoudeh et al., 2009; Ludlow et al., 2015). Most studies have focused on the function of NS5 as RdRp (Troost and Smit, 2020) because this activity is absent in the host cell (Galiano et al., 2016), and this is promising to design specific inhibitors with low toxicity (El Sahili and Lescar, 2017); and within RdRp, cavity B has been considered a site to be explored for the drug design (Malet et al., 2008; Alhossary et al., 2018); however, the MTase domain is also reported as an attractive strategy in the anti-flavivirus drug design (Santos et al., 2020), and regions such as GTP and SAM pockets are obvious targets for antiviral development since they have both been shown to bind to low-molecular-weight ligands (Lim et al., 2015).

The idea of looking for drugs for all serotypes must consider the differences between them because although any serotype is equally able to cause dengue, serotype differences have been postulated to lead to differences in pathogenesis such as the case for DENV2 which has been related with severe dengue (Trujillo-Correa et al., 2019); however, DENV inhibitors that protect toward two or three serotypes should not be neglected for further testing. Thus, serotype-specific treatment may help to treat serotype confirmed DENV patients, and when more antiviral drugs are available it may be possible to achieve a pan-protective effect *via* drug combinational therapy (Troost and Smit, 2020).

On the other hand, the *in silico* assays have served as a starting point for the identification of potential compounds as inhibitors;

many studies report this strategy as a good means for the discovery and development of drugs for different diseases. Some example for this is the use of these strategies to search, from libraries, for natural compounds that inhibit RdRp of DENV (Galiano et al., 2016) and also the identification of phytochemical compounds reported for various flaviviruses, on non-structural proteins of DENV (Tahir ul Qamar and Muner, 2019).

On the other hand, with respect to *in vitro* assays, 10 of the 18 candidate compounds were acquired and subsequently evaluated on Huh-7 cells, in order to determine their effect on cell viability. The Huh-7 cell line is derived from hepatocytes, which represent a target cell during natural DENV infection. These cells have been used in previous studies of dengue–host cell interaction (Pando-Robles et al., 2014). CC₅₀ is defined as the concentration that causes a 50% reduction or inhibition of cell viability, or that causes 50% cytotoxicity (Dewi et al., 2019). The results indicated that none of the compounds were cytotoxic at the highest concentration evaluated, which implies a CC₅₀ for each of them above 50 μ M, during the first 24 h of exposure (Figure 12). The evaluation of all concentrations reflected that cell viability remained above 80% for most of the compounds, which validates the evaluation of these compounds in antiviral tests on this cellular model.

In this study, we propose 18 compounds, identified in *in silico*, as possible candidates for the *in vitro* evaluation of the antiviral effect in the four serotypes of dengue virus. The compounds identified in this work bind to different regions of the DENV NS5 protein, such as cavities that are allosteric sites and active sites in the MTase and RdRp domains. However, these compounds are not nucleotide analogs, yet they interact on important sites of the NS5 protein and may be promising compounds to be evaluated by *in vitro* assays.

Future Outlook

According to our results, our interest is to evaluate the identified compounds in this study (18 compounds) *in vitro*. It is worth mentioning that all the compounds presented here (18) were selected under a rational identification criterion, for which the classification of “best compounds” according to the *in silico* analyses was supported by the binding-free energies that they presented in the selected regions of the NS5 protein; however, all the identified compounds correspond to compounds with desirable physicochemical and pharmacological properties to have been considered as candidates to be evaluated *in vitro*.

The experimental validation of the identified compounds has been carried out with prior identification of cellular toxicity on Huh-7 cells as an ideal model of dengue infection. Subsequently, it is intended to evaluate the effect on DENV by means of plaque-forming units (PFUs), on the synthesis of viral proteins and RNA and on the activity of the viral NS5 protein.

So far, we have evaluated the cytotoxicity of some of these compounds on Huh-7 cells, as 1sd, 2sd, 3sd, 4sd, 6sd, 7sd, 8sd, 9sd, 3md, and 8md (compounds that were possible to purchase from the seller). According to preliminary results, the compounds do not generate cytotoxic effect in lower

concentrations to 50 μ M, finding cell viability greater than 80% for most of these, which validates the evaluation of these compounds in other *in vitro* assays. A preliminary analysis of the effect of these compounds against Huh-7 cells infected with DENV2, showed that for compounds 8md and 9sd, there is a statistically significant difference (p -value <0.01) between the percentage of plaque-forming units (PFUs) of DENV compared to the control of DMSO 0.3%, indicating that was a reduction by treatment. On the other hand, other compounds identified in this study, such as compounds 6sd and 3md reduce Log PFUs, with a behavior similar to that presented by compound NITD-008, a nucleotide analog that it potently inhibits replication in DENV (Yin et al., 2009a; Lim et al., 2013b) and that was used as an experimental control; furthermore, the compound 3md has been found to reduce the percentage of infected cells and the number of viral copies. The *in vitro* antiviral assays are in process.

CONCLUSION

We performed a computational screening of several drug-like compounds with potential effects against the NS5 protein of the dengue virus. We report two virtual screening strategies focused on the search for compounds with binding to multiple sites within the same protein, called multi-domain compounds, and compounds of classical inhibition or those that bind to a single site within the NS5 protein, called single-domain compounds. As inclusion and exclusion criteria, we developed a series of filters that allowed us to recognize possible structural risks in the selected compounds to minimize complications in future experimental trials against DENV. Starting from a list of ~642,759 commercially available drug-like compounds for each strategy, we found 8 compounds with multi-domain binding and 80 with single-domain binding. Then, we identified the best compound for each region and analyzed their interaction with the four serotypes. According to our results, we highlight a short list of 18 compounds as the most promissory for future research. Additionally, we suggest that contact frequency analysis can be useful when filtering compounds from molecular dynamics simulations, being computationally cheaper than a calculation of free binding energies, and in the evaluation *in vitro* of 10 of these compounds we found that they are not cytotoxic in the Huh-7 cell line below 50 μ M. We are aware that experimental validation, more extensive simulations, and robust thermodynamic studies can be useful in order to validate our hypothesis and to expand the search for compounds with antiviral activity.

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

LG-A, LP-S, and JC-O contributed to conceptualization; LG-A and CR-R contributed to methodology; LG-A and CR-R contributed to software; LG-A, CR-R, LP-S, and JC-O contributed to validation; LG-A and CR-R contributed to formal analysis; LG-A contributed to investigation; LG-A, LP-S, and JC-O contributed to resources; LG-A and CR-R contributed to data curation; LG-A and CR-R contributed to writing—original draft preparation; LG-A, CR-R, LP-S, and JC-O contributed to writing—review and editing; LG-A, CR-R, LP-S, and JC-O contributed to visualization; LP-S and JC-O contributed to supervision; LG-A, LP-S, and JC-O contributed to project administration; LG-A, LP-S, and JC-O contributed to funding acquisition. All authors read and approved the final manuscript.

FUNDING

This research was funded by the Ministerio de Ciencia, Tecnología e Innovación (Minciencias)- Colombia, convocatory 757 Doctorados Nacionales 2016, contract

REFERENCES

- Ahmad, N., Rehman, A. U., Badshah, S. L., Ullah, A., Mohammad, A., and Khan, K. (2020). Molecular Dynamics Simulation of Zika Virus NS5 RNA Dependent RNA Polymerase with Selected Novel Non-nucleoside Inhibitors. *J. Mol. Struct.* 1203, 127428. doi:10.1016/j.molstruc.2019.127428
- Al-Sha'er, M. A., and Taha, M. O. (2021). Elaboration of Novel TTK1 Inhibitory Leads via QSAR-Guided Selection of Crystallographic Pharmacophores Followed by *In Vitro* Assay. *Cad* 17 (4), 511–522. doi:10.2174/1573409916666200611122736
- Al-Sha'er, M. A., and Taha, M. O. (2010). Discovery of Novel CDK1 Inhibitors by Combining Pharmacophore Modeling, QSAR Analysis and *In Silico* Screening Followed by *In Vitro* Bioassay. *Eur. J. Med. Chem.* 45 (9), 4316–4330. doi:10.1016/j.ejmech.2010.06.034
- Alhossary, A., Awuni, Y., Kwok, C. K., and Mu, Y. (2018). Proposing Drug Fragments for Dengue Virus NS5 Protein. *J. Bioinform. Comput. Biol.* 16, 1840017. doi:10.1142/S0219720018400176
- Banerjee, P., Eckert, A. O., Schrey, A. K., and Preissner, R. (2018). ProTox-II: a Webserver for the Prediction of Toxicity of Chemicals. *Nucleic Acids Res.* 46, W257–W263. doi:10.1093/nar/gky318
- Basavannacharya, C., and Vasudevan, S. G. (2014). Suramin Inhibits Helicase Activity of NS3 Protein of Dengue Virus in a Fluorescence-Based High Throughput Assay Format. *Biochem. Biophys. Res. Commun.* 453, 539–544. doi:10.1016/j.bbrc.2014.09.113
- Berendsen, H. J. C., Van Der Spoel, D., and Van Drunen, R. (1995). GROMACS: A Message-Passing Parallel Molecular Dynamics Implementation. *Comp. Phys. Commun.* 91 (1–3), 43–56. doi:10.1016/0010-4655(95)00042-e
- Bhatnagar, P., Sreekanth, G. P., Murali-Krishna, K., Chandele, A., and Sitaraman, R. (2021). Dengue Virus Non-structural Protein 5 as a Versatile, Multi-Functional Effector in Host-Pathogen Interactions. *Front. Cel. Infect. Microbiol.* 11, 574067. doi:10.3389/fcimb.2021.574067
- Bhattacharya, D., Hoover, S., Falk, S. P., Weisblum, B., Vestling, M., and Striker, R. (2008). Phosphorylation of Yellow Fever Virus NS5 Alters Methyltransferase Activity. *Virology* 380, 276–284. doi:10.1016/j.virol.2008.07.013
- Bollati, M., Alvarez, K., Assenberg, R., Baronti, C., Canard, B., Cook, S., et al. (2010). Structure and Functionality in Flavivirus NS-Proteins: Perspectives for Drug Design. *Antiviral Res.* 87, 125–148. doi:10.1016/j.antiviral.2009.11.009

number FP44842-269-2017, and Universidad del Quindío, Colombia.

ACKNOWLEDGMENTS

The authors acknowledge Grupo de Inmunología Molecular (GYMOL), Centro de Investigaciones Biomédicas, Universidad del Quindío; Ministerio de Ciencia, Tecnología e Innovación (Minciencias)- Colombia, and Mariano García Blanco's laboratory at the Department of Biochemistry and Molecular Biology, University of Texas Medical Branch at Galveston, United States, for supporting *in vitro* tests. It is also appreciated that CR-R has been supported by Minciencias, University of Antioquia and Ruta N, Colombia, and the Max Planck Society, Germany

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fchem.2022.637266/full#supplementary-material>

- Caillet-Saguy, C., Lim, S. P., Shi, P.-Y., Lescar, J., and Bressanelli, S. (2014). Polymerases of Hepatitis C Viruses and Flaviviruses: Structural and Mechanistic Insights and Drug Development. *Antiviral Res.* 105, 8–16. doi:10.1016/j.antiviral.2014.02.006
- Cannalire, R., Ki Chan, K. W., Burali, M. S., Gwee, C. P., Wang, S., Astolfi, A., et al. (2020). Pyridobenzothiazolones Exert Potent Anti-dengue Activity by Hampering Multiple Functions of NS5 Polymerase. *ACS Med. Chem. Lett.* 11, 773–782. doi:10.1021/acsmchemlett.9b00619
- Chang, J., Schul, W., Butters, T. D., Yip, A., Liu, B., Goh, A., et al. (2011). Combination of α -glucosidase Inhibitor and Ribavirin for the Treatment of Dengue Virus Infection *In Vitro* and *In Vivo*. *Antiviral Res.* 89, 26–34. doi:10.1016/j.antiviral.2010.11.002
- Chen, D., Oezguen, N., Urvil, P., Ferguson, C., Dann, S. M., and Savidge, T. C. (2016). Regulation of Protein-Ligand Binding Affinity by Hydrogen Bond Pairing. *Sci. Adv.* 2 (2), e1501240. doi:10.1126/sciadv.1501240
- Chung Rt, F. D., and Chung, R. T. (2014). Review of Current Dengue Treatment and Therapeutics in Development. *J. Bioanal. Biomed.* s8, 1–15. doi:10.4172/1948-593X.S8-002
- Crooks, G. E., Hon, G., Chandonia, J.-M., and Brenner, S. E. (2004). WebLogo: A Sequence Logo Generator: Figure 1. *Genome Res.* 14, 1188–1190. doi:10.1101/gr.849004
- 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. doi:10.1038/srep42717
- Diosa-Toro, M., Troost, B., van de Pol, D., Heberle, A. M., Urcuqui-Inchima, S., and Thedieck, K. (2018). Tomatidine, a novel antiviral compound towards dengue virus. *Antiviral Res.* 161, 90–99. doi:10.1016/j.antiviral.2018.11.011
- De Burghgraeve, T., Selisko, B., Kaptein, S., Chatelain, G., Leyssen, P., Debing, Y., et al. (2013). 3',5'-Di-O-trityluridine Inhibits *In Vitro* Flavivirus Replication. *Antiviral Res.* 98, 242–247. doi:10.1016/j.antiviral.2013.01.011
- Dewi, B. E., Angelina, M., Nuwwaaidya, F., Desti, H., and Sudiro, T. M. (2019). Antiviral Activity of Isobutyl Gallate to Dengue Virus Serotype 2 *In Vitro*. *IOP Conf. Ser. Earth Environ. Sci.* 251, 012018. doi:10.1088/1755-1315/251/1/012018
- Dong, H., Zhang, B., and Shi, P.-Y. (2008). Flavivirus Methyltransferase: a Novel Antiviral Target. *Antiviral Res.* 80, 1–10. doi:10.1016/j.antiviral.2008.05.003
- El Sahili, A., and Lescar, J. (2017). Dengue Virus Non-structural Protein 5. *Viruses* 9, 91. doi:10.3390/v9040091

- Galiano, V., García-Valtanan, P., Micol, V., and Encinar, J. A. (2016). Looking for Inhibitors of the Dengue Virus NS5 RNA-dependent RNA-Polymerase Using a Molecular Docking Approach. *Dddt* Vol. 10, 3163–3181. doi:10.2147/DDDT.S117369
- García, L. L., Padilla, L., and Castaño, J. C. (2017). Inhibitors Compounds of the Flavivirus Replication Process. *Virol. J.* 14, 95. doi:10.1186/s12985-017-0761-1
- Geiss, B. J., Thompson, A. A., Andrews, A. J., Sons, R. L., Gari, H. H., Keenan, S. M., et al. (2009). Analysis of Flavivirus NS5 Methyltransferase Cap Binding. *J. Mol. Biol.* 385, 1643–1654. doi:10.1016/j.jmb.2008.11.058
- 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
- Halim, S. A., Khan, S., Khan, A., Wadood, A., Mabood, F., Hussain, J., et al. (2017). Targeting Dengue Virus NS-3 Helicase by Ligand Based Pharmacophore Modeling and Structure Based Virtual Screening. *Front. Chem.* 5, 88. doi:10.3389/fchem.2017.00088
- Hammoudeh, D. I., Follis, A. V., Prochownik, E. V., and Metallo, S. J. (2009). Multiple Independent Binding Sites for Small-Molecule Inhibitors on the Oncoprotein C-Myc. *J. Am. Chem. Soc.* 131 (21), 7390–7401. doi:10.1021/ja900616b
- Hatcher, E. L., Zhdanov, S. A., Bao, Y., Blinkova, O., Nawrocki, E. P., Ostapchuck, Y., et al. (2017). Virus Variation Resource - Improved Response to Emergent Viral Outbreaks. *Nucleic Acids Res.* 45, D482–D490. doi:10.1093/nar/gkw1065
- Hetényi, C., and Bálint, M. (2020). Systematic Exploration of Binding Modes of Ligands on Drug Targets. *Methods Mol. Biol.* 2112, 107–121. doi:10.1007/978-1-0716-0270-6_8
- Hollingsworth, S. A., and Karplus, P. A. (2010). A Fresh Look at the Ramachandran Plot and the Occurrence of Standard Structures in Proteins. *Biomol. Concepts* 1, 271–283. doi:10.1515/BMC.2010.022
- Hou, T., Wang, J., Li, Y., and Wang, W. (2011). Assessing the Performance of the MM/PBSA and MM/GBSA Methods. 1. The Accuracy of Binding Free Energy Calculations Based on Molecular Dynamics Simulations. *J. Chem. Inf. Model.* 51, 69–82. doi:10.1021/ci100275a
- Jenwithesuk, E., Horst, J., Rivas, K., Vanvoorhis, W., and Samudrala, R. (2008). Novel Paradigms for Drug Discovery: Computational Multitarget Screening. *Trends Pharmacol. Sci.* 29, 62–71. doi:10.1016/j.tips.2007.11.007
- Kaptein, S. J. F., Vincetti, P., Crespan, E., Rivera, J. I. A., Costantino, G., Maga, G., et al. (2018). Identification of Broad-Spectrum Dengue/Zika Virus Replication Inhibitors by Functionalization of Quinoline and 2,6-Diaminopurine Scaffolds. *ChemMedChem* 13, 1371–1376. doi:10.1002/cmde.201800178
- Klema, V. J., Ye, M., Hindupur, A., Teramoto, T., Gottipati, K., Padmanabhan, R., et al. (2016). Dengue Virus Nonstructural Protein 5 (NS5) Assembles into a Dimer with a Unique Methyltransferase and Polymerase Interface. *Plos Pathog.* 12 (2), e1005451. doi:10.1371/journal.ppat.1005451
- Lai, J.-H., Lin, Y.-L., and Hsieh, S.-L. (2017). Pharmacological Intervention for Dengue Virus Infection. *Biochem. Pharmacol.* 129, 14–25. doi:10.1016/j.bcp.2017.01.005
- Latour, D. R., Jekle, A., Javanbakht, H., Henningsen, R., Gee, P., Lee, I., et al. (2010). Biochemical Characterization of the Inhibition of the Dengue Virus RNA Polymerase by Beta-D-2'-Ethynyl-7-Deaza-Adenosine Triphosphate. *Antiviral Res.* 87, 213–222. doi:10.1016/j.antiviral.2010.05.003
- Li, Honglin., Zheng, Mingyue., Luo, Xiaomin., Zhu, Weiliang., and Jiang, Hualiang. (2008). Drug Discovery and Development: Computational Approaches. *Wiley Encyclopedia Chem. Biol.* 1–9. doi:10.1002/9780470048672.wecb
- Lim, S. P., Koh, J. H. K., Seh, C. C., Liew, C. W., Davidson, A. D., Chua, L. S., et al. (2013a). A crystal Structure of the Dengue Virus Non-structural Protein 5 (NS5) Polymerase Delineates Interdomain Amino Acid Residues that Enhance its Thermostability and De Novo Initiation Activities. *J. Biol. Chem.* 288, 31105–31114. doi:10.1074/jbc.M113.508606
- Lim, S. P., Noble, C. G., Seh, C. C., Soh, T. S., El Sahili, A., Chan, G. K. Y., et al. (2016). Potent Allosteric Dengue Virus NS5 Polymerase Inhibitors: Mechanism of Action and Resistance Profiling. *Plos Pathog.* 12, e1005737. doi:10.1371/journal.ppat.1005737
- Lim, S. P., Noble, C. G., and Shi, P.-Y. (2015). The Dengue Virus NS5 Protein as a Target for Drug Discovery. *Antiviral Res.* 119, 57–67. doi:10.1016/j.antiviral.2015.04.010
- Lim, S. P., Wang, Q.-Y., Noble, C. G., Chen, Y.-L., Dong, H., Zou, B., et al. (2013b). Ten Years of Dengue Drug Discovery: Progress and Prospects. *Antiviral Res.* 100, 500–519. doi:10.1016/j.antiviral.2013.09.013
- Lim, S. V., Rahman, M. B. A., and Tejo, B. A. (2011). Structure-based and Ligand-Based Virtual Screening of Novel Methyltransferase Inhibitors of the Dengue Virus. *BMC Bioinformatics* 12 (Suppl. 13), S24. doi:10.1186/1471-2105-12-S13-S24
- Lindorff-Larsen, K., Piana, S., Palmo, K., Maragakis, P., Klepeis, J. L., Dror, R. O., et al. (2010). Improved Side-Chain Torsion Potentials for the Amber ff99SB Protein Force Field. *Proteins* 78, 1950–1958. doi:10.1002/prot.22711
- Ludlow, R. F., Verdonk, M. L., Saini, H. K., Tickle, I. J., and Jhoti, H. (2015). Detection of Secondary Binding Sites in Proteins Using Fragment Screening. *Proc. Natl. Acad. Sci. USA* 112 (52), 15910–15915. doi:10.1073/pnas.1518946112
- Malet, H., Massé, N., Selisko, B., Romette, J.-L., Alvarez, K., Guillemot, J. C., et al. (2008). The Flavivirus Polymerase as a Target for Drug Discovery. *Antiviral Res.* 80, 23–35. doi:10.1016/j.antiviral.2008.06.007
- Martínez, L. (2015). Automatic Identification of mobile and Rigid Substructures in Molecular Dynamics Simulations and Fractional Structural Fluctuation Analysis. *PLoS One* 10 (3), e0119264. doi:10.1371/journal.pone.0119264
- Meguellati, A., Ahmed-Belkacem, A., Yi, W., Haudecoeur, R., Crouillère, M., Brillet, R., et al. (2014). B-ring Modified Aurones as Promising Allosteric Inhibitors of Hepatitis C Virus RNA-dependent RNA Polymerase. *Eur. J. Med. Chem.* 80, 579–592. doi:10.1016/j.ejmech.2014.04.005
- 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. Theor. Comput.* 8, 3314–3321. doi:10.1021/ct300418h
- Mirza, M. U., Vanmeert, M., Froeyen, M., Ali, A., Rafique, S., and Idrees, M. (2019). In Silico structural Elucidation of RNA-dependent RNA Polymerase towards the Identification of Potential Crimean-Congo Hemorrhagic Fever Virus Inhibitors. *Sci. Rep.* 9, 6809. doi:10.1038/s41598-019-43129-2
- 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
- Mustafa, M. S., Rasotgi, V., Jain, S., and Gupta, V. (2015). Discovery of Fifth Serotype of Dengue Virus (DENV-5): A New Public Health Dilemma in Dengue Control. *Med. J. Armed Forces India* 71, 67–70. doi:10.1016/j.mjafi.2014.09.011
- Najera, I. (2013). Resistance to HCV Nucleoside Analogue Inhibitors of Hepatitis C Virus RNA-dependent RNA Polymerase. *Curr. Opin. Virol.* 3, 508–513. doi:10.1016/j.coviro.2013.08.011
- Niyomrattanakit, P., Chen, Y.-L., Dong, H., Yin, Z., Qing, M., Glickman, J. F., et al. (2010). Inhibition of Dengue Virus Polymerase by Blocking of the RNA Tunnel. *J. Virol.* 84, 5678–5686. doi:10.1128/JVI.02451-09
- Niyomrattanakit, P., Wan, K. F., Chung, K. Y., Abas, S. N., Seh, C. C., Dong, H., et al. (2015). Stabilization of Dengue Virus Polymerase in De Novo Initiation Assay Provides Advantages for Compound Screening. *Antiviral Res.* 119, 36–46. doi:10.1016/j.antiviral.2015.04.007
- Noble, C. G., Lim, S. P., Chen, Y.-L., Liew, C. W., Yap, L., Lescar, J., et al. (2013). Conformational Flexibility of the Dengue Virus RNA-dependent RNA Polymerase Revealed by a Complex with an Inhibitor. *J. Virol.* 87, 5291–5295. doi:10.1128/JVI.00045-13
- Pando-Robles, V., Osés-Prieto, J. A., Rodríguez-Gandarilla, M., Meneses-Romero, E., Burlingame, A. L., and Batista, C. V. F. (2014). Quantitative Proteomic Analysis of Huh-7 Cells Infected with Dengue Virus by Label-free LC-MS. *J. Proteomics* 111, 16–29. doi:10.1016/j.jprot.2014.06.029
- 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
- Potisopon, S., Ferron, F., Fattorini, V., Selisko, B., and Canard, B. (2017). Substrate Selectivity of Dengue and Zika Virus NS5 Polymerase towards 2'-modified

- Nucleotide Analogues. *Antiviral Res.* 140, 25–36. doi:10.1016/j.antiviral.2016.12.021
- Regato, M., Recarey, R., Moratorio, G., de Mora, D., Garcia-Aguirre, L., González, M., et al. (2008). Phylogenetic Analysis of the NS5 Gene of Dengue Viruses Isolated in Ecuador. *Virus Res.* 132, 197–200. doi:10.1016/j.virusres.2007.10.012
- Sampath, A., and Padmanabhan, R. (2009). Molecular Targets for Flavivirus Drug Discovery. *Antiviral Res.* 81, 6–15. doi:10.1016/j.antiviral.2008.08.004
- Sander, T., Freyss, J., Von Korff, M., and Rufener, C. (2015). DataWarrior: an Open-Source Program for Chemistry Aware Data Visualization and Analysis. *J. Chem. Inf. Model.* 55, 460–473. doi:10.1021/ci500588j
- Santos, F. R. S., Lima, W. G., Maia, E. H. B., Assis, L. C., Davyt, D., Taranto, A. G., et al. (2020). Identification of a Potential Zika Virus Inhibitor Targeting NS5 Methyltransferase Using Virtual Screening and Molecular Dynamics Simulations. *J. Chem. Inf. Model.* 60, 562–568. doi:10.1021/acs.jcim.9b00809
- Saw, W. G., Tria, G., Grüber, A., Subramanian Manimekalai, M. S., Zhao, Y., Chandramohan, A., et al. (2015). Structural Insight and Flexible Features of NS5 Proteins from All Four Serotypes of Dengue Virus in Solution. *Acta Cryst. D Biol. Crystallogr.* 71, 2309–2327. doi:10.1107/s1399004715017721
- Schneidman-Duhovny, D., Dror, O., Inbar, Y., Nussinov, R., and Wolfson, H. J. (2008). PharmaGist: a Webserver for Ligand-Based Pharmacophore Detection. *Nucleic Acids Res.* 36 (Web Server issue), W223–W228. doi:10.1093/nar/gkn187
- Seyedi, S. S., Shukri, M., Hassandarvish, P., Oo, A., Shankar, E. M., Abubakar, S., et al. (2016). Computational Approach towards Exploring Potential Antichikungunya Activity of Selected Flavonoids. *Sci. Rep.* 6, 24027. doi:10.1038/srep24027
- Sousa da Silva, A. W., and Vranken, W. F. (2012). ACPYPE - AnteChamber PYthon Parser interfAcE. *BMC Res. Notes* 5, 367. doi:10.1186/1756-0500-5-367
- Subramanian Manimekalai, M. S., Saw, W. G., Pan, A., Grüber, A., and Grüber, G. (2016). Identification of the Critical Linker Residues Conferring Differences in the Compactness of NS5 from Dengue Virus serotype 4 and NS5 from Dengue Virus serotypes 1–3. *Acta Cryst. Sect D Struct. Biol.* 72 (Pt 6), 795–807. doi:10.1107/S2059798316006665
- Tahir ul Qamar, M., Maryam, A., Muneer, I., Xing, F., Ashfaq, U. A., Khan, F. A., et al. (2019). Computational Screening of Medicinal Plant Phytochemicals to Discover Potent Pan-Serotype Inhibitors against Dengue Virus. *Sci. Rep.* 9, 1433. doi:10.1038/s41598-018-38450-1
- Tay, M. Y. F., Fraser, J. E., Chan, W. K. K., Moreland, N. J., Rathore, A. P., Wang, C., et al. (2013). Nuclear Localization of Dengue Virus (DENV) 1–4 Non-structural Protein 5; protection against All 4 DENV Serotypes by the Inhibitor Ivermectin. *Antiviral Res.* 99, 301–306. doi:10.1016/j.antiviral.2013.06.002
- Tian, Y.-S., Zhou, Y., Takagi, T., Kameoka, M., and Kawashita, N. (2018). Dengue Virus and its Inhibitors: A Brief Review. *Chem. Pharm. Bull.* 66, 191–206. doi:10.1248/cpb.c17-00794
- Tomlinson, S. M., and Watowich, S. J. (2011). Anthracene-based Inhibitors of Dengue Virus NS2B-NS3 Protease. *Antiviral Res.* 89, 127–135. doi:10.1016/j.antiviral.2010.12.006
- Tripathi, N. K., and Shrivastava, A. (2018). Recent Developments in Recombinant Protein-Based Dengue Vaccines. *Front. Immunol.* 9, 1919. doi:10.3389/fimmu.2018.01919
- Troost, B., and Smit, J. M. (2020). Recent Advances in Antiviral Drug Development towards Dengue Virus. *Curr. Opin. Virol.* 43, 9–21. doi:10.1016/j.coviro.2020.07.009
- Trott, O., and Olson, A. J. (2009). AutoDock Vina: Improving the Speed and Accuracy of Docking with a New Scoring Function, Efficient Optimization, and Multithreading. *J. Comput. Chem.* 31, NA. doi:10.1002/jcc.21334
- Trujillo-Correa, A. I., Quintero-Gil, D. C., Diaz-Castillo, F., Quiñones, W., Robledo, S. M., and Martínez-Gutierrez, M. (2019). In vitro and in silico anti-dengue activity of compounds obtained from Psidium guajava through bioprospecting. *BMC Complement Altern Med* 19 (1), 298. doi:10.1186/s12906-019-2695-1
- UniProt Consortium (2021). UniProt: the Universal Protein Knowledgebase in 2021. *Nucleic Acids Res.* 49 (D1), D480–D489. doi:10.1093/nar/gkaa1100
- Valdés, M. S., Valdes, M. E., Valiente, P., and Moreno, E. (2021). gmx_MMPBSA: A New Tool to Perform End-State Free Energy Calculations with GROMACS. *J. Chem. Theor. Comput.* 17 (10), 6281–6291. doi:10.1021/acs.jctc.1c00645
- Viswanathan, U., Tomlinson, S. M., Fonner, J. M., Mock, S. A., and Watowich, S. J. (2014). Identification of a Novel Inhibitor of Dengue Virus Protease through Use of a Virtual Screening Drug Discovery Web portal. *J. Chem. Inf. Model.* 54, 2816–2825. doi:10.1021/ci500531r
- Wang, E., Sun, H., Wang, J., Wang, Z., Liu, H., Zhang, J. Z. H., et al. (2019). End-Point Binding Free Energy Calculation with MM/PBSA and MM/GBSA: Strategies and Applications in Drug Design. *Chem. Rev.* doi:10.1021/acs.chemrev.9b000510.1021/acs.chemrev.9b00055
- 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
- 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
- Westermaier, Y., Barril, X., and Scapozza, L. (2015). Virtual Screening: an *In Silico* Tool for Interlacing the Chemical Universe with the Proteome. *Methods* 71, 44–57. doi:10.1016/j.jymeth.2014.08.001
- 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
- Williams, C. J., Headd, J. J., Moriarty, N. W., Prisant, M. G., Videau, L. L., Deis, L. N., et al. (2018). MolProbity: More and Better Reference Data for Improved All-Atom Structure Validation. *Protein Sci.* 27, 293–315. doi:10.1002/pro.3330
- Wu, F., Zhou, Y., Li, L., Shen, X., Chen, G., Wang, X., et al. (2020). Computational Approaches in Preclinical Studies on Drug Discovery and Development. *Front. Chem.* 8. doi:10.3389/fchem.2020.00726
- Wu, J., Ye, H.-Q., Zhang, Q.-Y., Lu, G., Zhang, B., and Gong, P. (2020). A Conformation-Based Intra-molecular Initiation Factor Identified in the Flavivirus RNA-dependent RNA Polymerase. *Plos Pathog.* 16, e1008484. doi:10.1371/journal.ppat.1008484
- Xu, L., Sun, H., Li, Y., Wang, J., and Hou, T. (2013). Assessing the Performance of MM/PBSA and MM/GBSA Methods. 3. The Impact of Force Fields and Ligand Charge Models. *J. Phys. Chem. B* 117 (28), 8408–8421. doi:10.1021/jp404160y
- Yap, T. L., Xu, T., Chen, Y.-L., Malet, H., Egloff, M.-P., Canard, B., et al. (2007). Crystal Structure of the Dengue Virus RNA-dependent RNA Polymerase Catalytic Domain at 1.85-angstrom Resolution. *J. Virol.* 81, 4753–4765. doi:10.1128/JVI.02283-06
- Yin, Z., Chen, Y.-L., Kondreddi, R. R., Chan, W. L., Wang, G., Ng, R. H., et al. (2009b). N-sulfonylanthranilic Acid Derivatives as Allosteric Inhibitors of Dengue Viral RNA-dependent RNA Polymerase. *J. Med. Chem.* 52, 7934–7937. doi:10.1021/jm901044z
- Yin, Z., Chen, Y.-L., Schul, W., Wang, Q.-Y., Gu, F., Duraiswamy, J., et al. (2009a). An Adenosine Nucleoside Inhibitor of Dengue Virus. *Proc. Natl. Acad. Sci.* 106, 20435–20439. doi:10.1073/pnas.0907010106
- Yuan, Y., Pei, J., and Lai, L. (2020). LigBuilder V3: A Multi-Target *De Novo* Drug Design Approach. *Front. Chem.* 8, 142. doi:10.3389/fchem.2020.00142
- Zhang, L., Ai, H., Chen, W., Yin, Z., Hu, H., Zhu, J., et al. (2017). CarcinoPred-EL: Novel Models for Predicting the Carcinogenicity of Chemicals Using Molecular Fingerprints and Ensemble Learning Methods. *Sci. Rep.* 7, 2118. doi:10.1038/s41598-017-02365-0
- Zhao, Y., Soh, T. S., Chan, K. W. K., Fung, S. S. Y., Swaminathan, K., Lim, S. P., et al. (2015b). Flexibility of NS5 Methyltransferase-Polymerase Linker Region Is Essential for Dengue Virus Replication. *J. Virol.* 89, 10717–10721. doi:10.1128/JVI.01239-15
- Zhao, Y., Soh, T. S., Lim, S. P., Chung, K. Y., Swaminathan, K., Vasudevan, S. G., et al. (2015c). Molecular Basis for Specific Viral RNA Recognition and 2'-O-Ribose Methylation by the Dengue Virus Nonstructural Protein 5 (NS5). *Proc. Natl. Acad. Sci. USA* 112, 14834–14839. doi:10.1073/pnas.1514978112

- Zhao, Y., Soh, T. S., Zheng, J., Chan, K. W. K., Phoo, W. W., Lee, C. C., et al. (2015a). A crystal Structure of the Dengue Virus NS5 Protein Reveals a Novel Inter-domain Interface Essential for Protein Flexibility and Virus Replication. *Plos Pathog.* 11, e1004682. doi:10.1371/journal.ppat.1004682
- Zhou, Y., Ray, D., Zhao, Y., Dong, H., Ren, S., Li, Z., et al. (2007). Structure and Function of Flavivirus NS5 Methyltransferase. *J. Virol.* 81, 3891–3903. doi:10.1128/JVI.02704-06
- Zou, G., Chen, Y.-L., Dong, H., Lim, C. C., Yap, L. J., Yau, Y. H., et al. (2011). Functional Analysis of Two Cavities in Flavivirus NS5 Polymerase. *J. Biol. Chem.* 286, 14362–14372. doi:10.1074/jbc.M110.214189

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 © 2022 García-Ariza, Rocha-Roa, Padilla-Sanabria and Castaño-Osorio. 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

Contact us: 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