

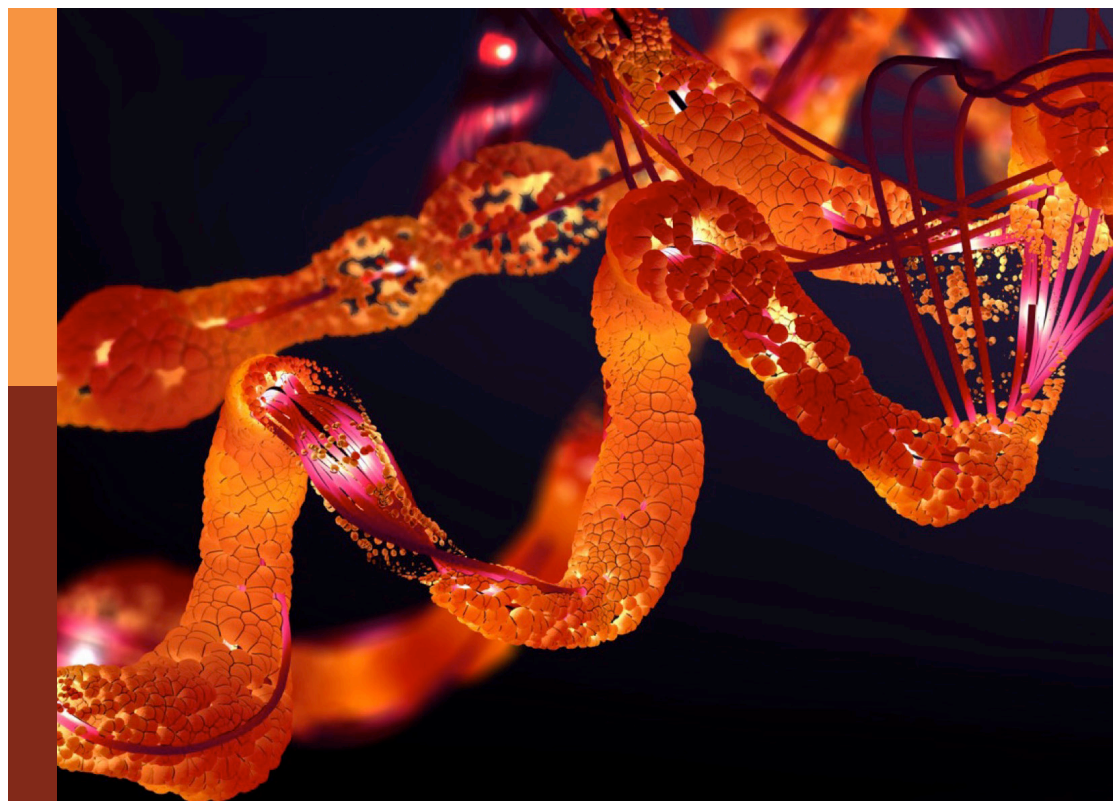
Molecular level atomistic and structural insights on biological macromolecules, inhibition, and dynamics studies

Edited by

Chandrabose Selvaraj, Emilia Pedone, Jung-Kul Lee
and Sanjeev Kumar Singh

Published in

Frontiers in Molecular Biosciences
Frontiers in Physiology



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 978-2-8325-4627-7
DOI 10.3389/978-2-8325-4627-7

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

Molecular level atomistic and structural insights on biological macromolecules, inhibition, and dynamics studies

Topic editors

Chandrabose Selvaraj — Saveetha Medical College & Hospital, India

Emilia Pedone — Institute of Biostructure and Bioimaging, Department of Biomedical Sciences, National Research Council (CNR), Italy

Jung-Kul Lee — Konkuk University, Republic of Korea

Sanjeev Kumar Singh — Alagappa University, India

Citation

Selvaraj, C., Pedone, E., Lee, J.-K., Singh, S. K., eds. (2024). *Molecular level atomistic and structural insights on biological macromolecules, inhibition, and dynamics studies*. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-8325-4627-7

Table of contents

- 05 Editorial: Molecular level atomistic and structural insights on biological macromolecules, inhibition, and dynamics studies
Chandrabose Selvaraj, Emilia Pedone, Jung-Kul Lee and Sanjeev Kumar Singh
- 08 Exploring the mechanism of action of dapansutride in the treatment of gouty arthritis based on molecular docking and molecular dynamics
Jun-Feng Cao, Xingyu Yang, Li Xiong, Mei Wu, Shengyan Chen, Hengxiang Xu, Yunli Gong, Lixin Zhang, Qilan Zhang and Xiao Zhang
- 25 Investigation of bioactive compounds from *Bacillus* sp. against protein homologs CDC42 of *Colletotrichum gloeosporioides* causing anthracnose disease in cassava by using molecular docking and dynamics studies
Narendra Kumar Papathoti, Kishore Mendam, Bala Hanumath Sriram Kanduri, Wannaporn Thepbandit, Rungthip Sangpueak, Chanon Saengchan, Nguyen Huy Hoang, Vineela Sai Megavath, Madhuri Kurakula, Toan Le Thanh and Natthiya Buensanteai
- 41 Quantum-based machine learning and AI models to generate force field parameters for drug-like small molecules
Sathish Kumar Mudedla, Abdennour Braka and Sangwook Wu
- 59 Structural characterization and conformational dynamics of alpha-1 antitrypsin pathogenic variants causing alpha-1-antitrypsin deficiency
Noor Ahmad Shaik, Najla Bint Saud Al-Saud, Thamer Abdulhamid Aljuhani, Kaiser Jamil, Huda Alnuman, Deema Aljeaid, Nasreen Sultana, Ashraf AbdulRahman El-Harouni, Zuhier Ahmed Awan, Ramu Elango and Babajan Banaganapalli
- 74 PCW-A1001, AI-assisted *de novo* design approach to design a selective inhibitor for FLT-3(D835Y) in acute myeloid leukemia
Seong Hun Jang, Dakshinamurthy Sivakumar, Sathish Kumar Mudedla, Jaehan Choi, Sungmin Lee, Minjun Jeon, Suneel Kumar Bvs, Jinha Hwang, Minsung Kang, Eun Gyeong Shin, Kyu Myung Lee, Kwan-Young Jung, Jae-Sung Kim and Sangwook Wu
- 88 Ab-initio evaluation of acid influence on chemical stability of hydrophilic diglycolamides
Jakub Luštinec, Tomáš Koubský and Ladislav Kalvoda
- 98 Gene expression, molecular docking, and molecular dynamics studies to identify potential antifungal compounds targeting virulence proteins/genes VelB and THR as possible drug targets against *Curvularia lunata*
Himanshu Kamboj, Lovely Gupta, Pawan Kumar, Pooja Sen, Abhishek Sengupta and Pooja Vijayaraghavan

- 116 **Microsecond-long simulation reveals the molecular mechanism for the dual inhibition of falcipain-2 and falcipain-3 by antimalarial lead compounds**
Ammar Usman Danazumi and Emmanuel Oluwadare Balogun
- 125 **A role of salt bridges in mediating drug potency: A lesson from the N-myristoyltransferase inhibitors**
Danislav S. Spassov, Mariyana Atanasova and Irini Doytchinova
- 140 **Targeting potential receptor molecules in non-small cell lung cancer (NSCLC) using in silico approaches**
C. Kirubhanand, J. Merciline Leonora, S. Anitha, R. Sangeetha, K. T. Nachammai, K. Langeswaran and S. Gowtham Kumar
- 156 **Functional metagenomics uncovers nitrile-hydrolysing enzymes in a coal metagenome**
Arunmozhi Bharathi Achudhan, Priya Kannan and Lilly M. Saleena
- 166 **Protein structural insights into a rare PCSK9 gain-of-function variant (R496W) causing familial hypercholesterolemia in a Saudi family: whole exome sequencing and computational analysis**
Noor Ahmad Shaik, Najla Al-Shehri, Mohammad Athar, Ahmed Awan, Mariam Khalili, Hadiyah Bassam Al Mahadi, Gehan Hejazy, Omar I. Saadah, Sameer Eida Al-Harhi, Ramu Elango, Babajan Banaganapalli, Eman Alefishat and Zuhier Awan



OPEN ACCESS

EDITED AND REVIEWED BY

Ioan Andricioaei,
University of California, Irvine, United States

*CORRESPONDENCE

Chandrabose Selvaraj,
✉ selva@csrdd.org,
✉ boselab2w@gmail.com

RECEIVED 27 December 2023

ACCEPTED 19 February 2024

PUBLISHED 07 March 2024

CITATION

Selvaraj C, Pedone E, Lee J-K and Singh SK (2024), Editorial: Molecular level atomistic and structural insights on biological macromolecules, inhibition, and dynamics studies. *Front. Mol. Biosci.* 11:1362215. doi: 10.3389/fmolb.2024.1362215

COPYRIGHT

© 2024 Selvaraj, Pedone, Lee and Singh. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](#). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Editorial: Molecular level atomistic and structural insights on biological macromolecules, inhibition, and dynamics studies

Chandrabose Selvaraj^{1*}, Emilia Pedone², Jung-Kul Lee³ and Sanjeev Kumar Singh⁴

¹Computational and Structural Research in Drug Discovery Lab (CSRDD), Center for Global Health Research, Saveetha Medical College, Saveetha Institute of Medical and Technical Sciences, Chennai, Tamil Nadu, India, ²Institute of Biostructures and Bioimaging, CNR, Naples, Italy, ³Department of Chemical Engineering, Konkuk University, Seoul, Republic of Korea, ⁴Computer Aided Drug Design and Molecular Modelling Lab, Department of Bioinformatics, Science Block, Alagappa University, Karaikudi, Tamil Nadu, India

KEYWORDS

molecular modeling, molecular dynamics, quantum mechanics, atomic insights, drug targets

Editorial on the Research Topic

Molecular level atomistic and structural insights on biological macromolecules, inhibition, and dynamics studies

Introduction

Atoms are the fundamental components of matter, and when they come together, they create molecules. These molecules can then join to create intricate biological structures. Having a deep understanding of how molecules behave at the atomic level has had a significant influence on the pharmaceutical, biotechnology, and chemical sectors (De Vivo et al. 2016). In various scientific disciplines such as chemistry, physics, materials science, and biology, it is essential to thoroughly examine and comprehend the behaviour, structure, and interactions of atoms and molecules (Selvaraj et al. 2023). In particular, researchers are uncovering novel enzyme structures using different experimental and computational techniques. These methods provide a detailed understanding of how enzymes function at the atomic level, their mechanisms, their roles in reactions, and how they can be inhibited (Carvalho et al. 2014). The atom-level illustrations primarily emphasize enzyme kinetics, inhibition, and the analysis of mutations and conformational changes using quantum mechanical and molecular dynamics techniques (Liu et al. 2018). By uncovering the atomic details of the macromolecule, we can gain insights that will aid in the identification of new agonists or antagonists. This, in turn, could lead to the development of potential drug candidates for the treatment of different diseases (Yu and MacKerell 2017). In order to develop a new inhibitor that specifically targets a particular protein, it is essential to thoroughly understand how the active site of the target protein interacts with

potential inhibitors. The main goal in designing a new inhibitor is to fully comprehend the molecular interactions between the inhibitor and the target, improve these interactions to ensure a strong binding and specificity, and rigorously test the effectiveness and safety of the inhibitor (Li and Kang 2020).

The following articles in this Research Topic align with the theme of offering insights at the molecular level to identify drug candidates that can bind to the desired drug targets. This is achieved through various computational methods such as modeling calculations, Quantum Mechanics, and molecular dynamics, which demonstrate the wide range of calculations and predictions. Cao et al. conducted a study to investigate how dapansutride works on NLRP3 and other protein targets in gouty arthritis. They used bioinformatics analysis and a computer simulation framework. The analysis at the molecular and atomic level, using techniques like molecular docking and molecular dynamics simulations, showed that dapansutride may not only directly inhibit NLRP3 to reduce the inflammatory response and pyroptosis, but also hinder the movement and activation of inflammatory cells by regulating IL1B, IL6, IL17A, IL18, MMP3, CXCL8, and TNF. Mudedla et al. have applied Quantum-based machine learning and AI models to generate force field parameters for drug-like small molecules. They have applied density functional theory (DFT) calculation for 31,770 small molecules that covered the chemical space of drug-like molecules. They also developed the neural network model for assigning atom types, phase angles, and periodicities. They found that an AI-generated force field was influential in the fast and accurate generation of partial charges and other force field parameters for small drug-like molecules. Papathoti et al. have used the molecular docking and simulation methods for investigating the bioactive compounds extracted from the *Bacillus sp* that target the protein homologs CDC42 of *Colletotrichum gloeosporioides* causing anthracnose disease in cassava. Five potent compounds from *B. megaterium* were used to target the protein. The interaction of β -sitosterol and phenylacetic acid with the critical residue of CDC42 demonstrated that ligands may inhibit growth-related functional proteins. They have also constructed the protein-protein interactions network, and from that, they have revealed that targeting the CDC42 protein could impart MAPK (Mitogen-activated protein kinases) signaling pathway. Shaik et al. have come up with a new computational biology dimension to interpret the genotype-protein phenotype relationship between SERPINA1 pathogenic variants with its structural plasticity and functional behaviour with NE ligand molecule contributing to the Alpha-1-antitrypsin deficiency. The molecular docking approach findings have demonstrated that the most missense variants negatively impact the affinity of NE (Neutrophil Elastase) and A1AT binding in a molecular complex, lowering A1AT functionality and contributing to its deficiency. Kamboj et al. have applied Gene expression analysis, molecular docking, and molecular dynamics studies to identify the strong antifungal compounds that show specificity with VelB and THR drug targets to inhibit *Curvularia lunata*. Luštinec et al. have performed the Ab-initio evaluation for evaluating the acid influence on the chemical stability of hydrophilic diglycolamides. Their results show strong theoretical findings on including an

acid influence on the diglycolamides chemical structure, treated in the frame of the density functional theory. Spassov et al. have used the molecular dynamics simulation methods for protonated and non-protonated forms of the inhibitors and suggested that the salt bridge has an unexpected role in stabilizing the NMT protein conformation and that this may be a significant factor in mediating its effects on NMT inhibitors potency. Danazumi et al. conducted microsecond-level MD simulations to comprehend the role of quinolinyl oxamide derivative (QOD) and an indole carboxamide derivative (ICD) as antimalarial lead drugs with dual inhibition of falcipain-2 and falcipain-3. Jang et al. have come up with the AI-assisted *de novo* design approach to identify a potent and selective inhibitor for the FLT3/FLT-3 (D835Y) mutant. They have optimized the compound PCW-1001 and generated the 10,416 analogues using the LSTM approach. Achudhan et al. identified the novel nitrilases compounds from a coal metagenome using the *in silico* mining methods. The binding scores produced by the novel nitrilase were approximately similar to those of the other prokaryotic nitrilase crystal structures, with a deviation of ± 0.5 . Kirubhanand et al. analyzed the bioactive nature of lochnericine against Non-Small Cell Lung Cancer (NSCLC) using various computational approaches such as quantum chemical calculations, molecular docking, and molecular dynamic simulation. Also, they confirmed the molecule's potential bioactivity based on the band gap energy value associated with bioactive compounds through Frontier Molecular Orbital (FMO). Shaik et al. provide comprehensive computational and structural insights into the genotype-protein phenotype correlation of the PCSK9 (Proprotein convertase subtilisin/kexin type 9) pathogenic variant with a PCSK9 inhibitor monoclonal antibody.

In general, the authors of these articles have used Artificial Intelligence and molecular modeling approaches to bring insightful information on atomistic mechanisms and explore functions of the biological macromolecule using atom-level calculations (Huggins et al. 2012; Selvaraj et al. 2022). Some studies have performed extensive molecular dynamics simulations like microsecond level molecular dynamics simulations and accurate Quantum Mechanical Calculations for understanding the atomic role in molecular mechanisms (Sakkiah et al. 2021). The conclusions of the major articles are based on theoretical approaches from the software and publicly available information, with very little confirmation in laboratory conditions. In the future, the added advantage of experimental findings supporting these theoretical findings is required to confirm these findings.

Author contributions

CS: Funding acquisition, Project administration, Supervision, Validation, Writing—original draft, Writing—review and editing. EP: Data curation, Formal Analysis, Funding acquisition, Resources, Visualization, Writing—review and editing. J-KL: Formal Analysis, Funding acquisition, Investigation, Resources, Validation, Visualization, Writing—review and editing. SS: Data curation,

Formal Analysis, Funding acquisition, Methodology, Resources, Visualization, Writing–review and editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This research was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) and funded by the Ministry of Science, ICT, and Future Planning (NRF-2022M3A9I3082366).

Acknowledgments

The author CS thankfully acknowledges the Saveetha Institute of Medical and Technical Sciences (SIMATS), Chennai for providing the infrastructure facility.

References

- Carvalho, A. T., Barrozo, A., Doron, D., Kilshtain, A. V., Major, D. T., and Kamerlin, S. C. (2014). Challenges in computational studies of enzyme structure, function and dynamics. *J. Mol. Graph Model* 54, 62–79. doi:10.1016/j.jmgm.2014.09.003
- 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 (9), 4035–4061. doi:10.1021/acs.jmedchem.5b01684
- Huggins, D. J., Sherman, W., and Tidor, B. (2012). Rational approaches to improving selectivity in drug design. *J. Med. Chem.* 55 (4), 1424–1444. doi:10.1021/jm2010332
- Li, Q., and Kang, C. (2020). Mechanisms of action for small molecules revealed by structural biology in drug discovery. *Int. J. Mol. Sci.* 21 (15), 5262. doi:10.3390/ijms21155262
- Liu, X., Shi, D., Zhou, S., Liu, H., Liu, H., and Yao, X. (2018). Molecular dynamics simulations and novel drug discovery. *Expert Opin. Drug Discov.* 13 (1), 23–37. doi:10.1080/17460441.2018.1403419
- Sakkiah, S., Selvaraj, C., Guo, W., Liu, J., Ge, W., Patterson, T. A., et al. (2021). Elucidation of agonist and antagonist dynamic binding patterns in ER- α by integration of molecular docking, molecular dynamics simulations and quantum mechanical calculations. *Int. J. Mol. Sci.* 22 (17), 9371. doi:10.3390/ijms22179371
- Selvaraj, C., Chandra, I., and Singh, S. K. (2022). Artificial intelligence and machine learning approaches for drug design: challenges and opportunities for the pharmaceutical industries. *Mol. Divers* 26 (3), 1893–1913. doi:10.1007/s11030-021-10326-z
- Selvaraj, C., Panwar, U., Ramalingam, K. R., Vijayakumar, R., and Singh, S. K. (2023). Exploring the macromolecules for secretory pathway in cancer disease. *Adv. Protein Chem. Struct. Biol.* 133, 55–83. doi:10.1016/bs.apcsb.2022.10.003
- Yu, W., and MacKerell, A. D., Jr. (2017). Computer-aided drug design methods. *Methods Mol. Biol.* 1520, 85–106. doi:10.1007/978-1-4939-6634-9_5

Conflict of interest

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

The author(s) declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

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.



OPEN ACCESS

EDITED BY

Chandrabose Selvaraj,
Department of Bioinformatics, Alagappa
University, India

REVIEWED BY

Anca Margineanu,
Helmholtz Association of German
Research Centers (HZ), Germany
Bhuvaneshwari Sampath,
Anna University, India

*CORRESPONDENCE

Lixin Zhang,
534259627@qq.com
Qilan Zhang,
345539297@qq.com
Xiao Zhang,
954073462@qq.com

[†]These authors have contributed equally
to this work and share first authorship

SPECIALTY SECTION

This article was submitted to Biophysics,
a section of the journal
Frontiers in Physiology

RECEIVED 10 July 2022

ACCEPTED 12 August 2022

PUBLISHED 29 August 2022

CITATION

Cao J-F, Xingyu Yang, Li Xiong, Wu M,
Chen S, Xu H, Gong Y, Zhang L, Zhang Q
and Zhang X (2022), Exploring the
mechanism of action of dapansutril in
the treatment of gouty arthritis based on
molecular docking and
molecular dynamics.
Front. Physiol. 13:990469.
doi: 10.3389/fphys.2022.990469

COPYRIGHT

© 2022 Cao, Xingyu Yang, Li Xiong, Wu,
Chen, Xu, Gong, Zhang, Zhang and
Zhang. 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 mechanism of action of dapansutril in the treatment of gouty arthritis based on molecular docking and molecular dynamics

Jun-Feng Cao^{1†}, Xingyu Yang^{1†}, Li Xiong¹, Mei Wu¹,
Shengyan Chen¹, Hengxiang Xu¹, Yunli Gong², Lixin Zhang^{3*},
Qilan Zhang^{4*} and Xiao Zhang^{4*}

¹Clinical Medicine, Chengdu Medical College, Chengdu, China, ²Laboratory Medicine, Chengdu Medical College, Chengdu, China, ³Yunnan Academy of Forestry Sciences, Kunming, Yunnan, China, ⁴Chengdu Medical College of Basic Medical Sciences, Chengdu, China

Purpose: Dapansutril is an orally active β -sulfonyl nitrile compound that selectively inhibits the NLRP3 inflammasome. Clinical studies have shown that dapansutril is active *in vivo* and limits the severity of endotoxin-induced inflammation and joint arthritis. However, there is currently a lack of more in-depth research on the effect of dapansutril on protein targets such as NLRP3 in gouty arthritis. Therefore, we used molecular docking and molecular dynamics to explore the mechanism of dapansutril on NLRP3 and other related protein targets.

Methods: We use bioinformatics to screen active pharmaceutical ingredients and potential disease targets. The disease-core gene target-drug network was established and molecular docking was used for verification. Molecular dynamics simulations were utilized to verify and analyze the binding stability of small molecule drugs to target proteins. The supercomputer platform was used to measure and analyze the binding free energy, the number of hydrogen bonds, the stability of the protein target at the residue level, the radius of gyration and the solvent accessible surface area.

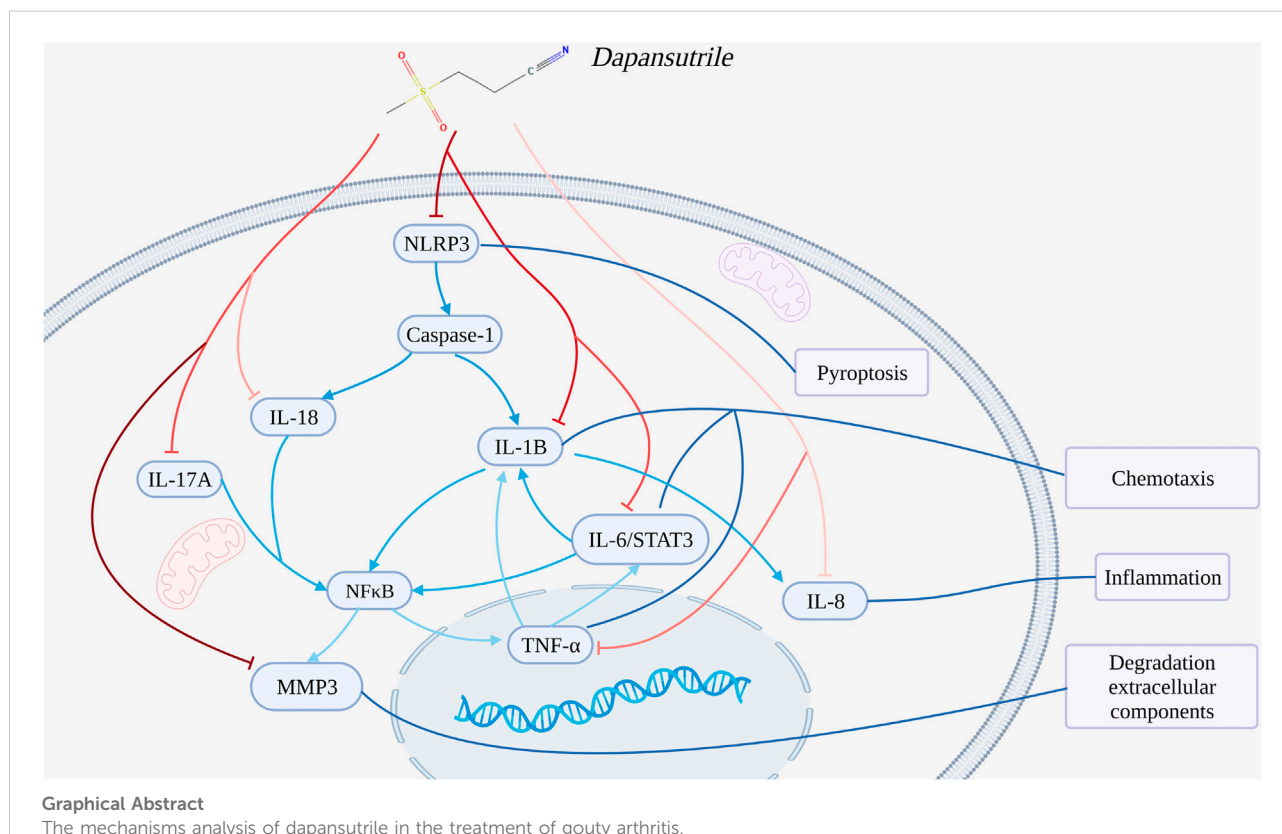
Results: The protein interaction network screened out the core protein targets (such as: NLRP3, TNF, IL1B) of gouty arthritis. Gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis revealed that gouty arthritis mainly played a vital role by the signaling pathways of inflammation and immune response. Molecular docking showed that dapansutril play a role in treating gouty arthritis by acting on the related protein targets such as NLRP3, IL1B, IL6, etc. Molecular dynamics was used to prove and analyze the binding stability of active ingredients and protein targets, the simulation results found that dapansutril forms a very stable complex with IL1B.

Conclusion: We used bioinformatics analysis and computer simulation system to comprehensively explore the mechanism of dapansutril acting on NLRP3 and other protein targets in gouty arthritis. This study found that

dapansutrine may not only directly inhibit NLRP3 to reduce the inflammatory response and pyroptosis, but also hinder the chemotaxis and activation of inflammatory cells by regulating IL1B, IL6, IL17A, IL18, MMP3, CXCL8, and TNF. Therefore, dapansutrine treats gouty arthritis by attenuating inflammatory response, inflammatory cell chemotaxis and extracellular matrix degradation by acting on multiple targets.

KEYWORDS

gouty arthritis, dapansutrine, NLRP3, molecular docking, molecular dynamics



Introduction

Gouty arthritis is considered one of the most common forms of inflammatory arthritis, it is a metabolic rheumatic disease (Roddy and Choi, 2014; Kim et al., 2016; Xie et al., 2017). Gouty arthritis is a form of arthritis caused by deposits of uric acid crystals called monosodium urate (MSU) crystals. The disease is usually secondary to chronic hyperuricemia, and the lesions are located in the joints and bursae (Klück et al., 2020; Wang and Wang, 2020). MSU crystal-induced gouty arthritis can occur in joints, periarticular tissues and kidneys (Wu et al., 2015). Acute gouty arthritis (AGA) is usually characterized by joint redness, swelling, warmth, and pain. As the disease progresses, gouty arthritis eventually leads to deformity of the diseased joint and severe limitation of joint movement (Yu et al., 2022). The principal

clinical treatment goals for gouty arthritis are stopping acute attacks, preventing recurrences and complications. The American College of Rheumatology (ACR) guidelines for the treatment of gout published in 2012 recommend the use of non-steroidal anti-inflammatory drugs (NSAIDs) or oral colchicine for the treatment of acute gout attacks (Khanna et al., 2012). It has been reported that NSAIDs are harmful to the gastrointestinal tract, liver and kidney, central nervous system, etc. (Bindu et al., 2020). Colchicine is an anti-inflammatory drug, which is widely used to treat acute gouty arthritis. However, colchicine can cause gastrointestinal discomfort, liver and kidney damage and multi-organ dysfunction when taken in high doses (Blackham et al., 2007). At the same time, it is under a narrow therapeutic index with no significant difference between non-toxic, toxic and lethal doses, which may lead to toxicity in patients (Finkelstein et al., 2010).

Therefore, there is an imperative for a safe and effective drug for the treatment of gouty arthritis.

Dapansutril (OLT1177) is a beta-sulfonitrile compound. It can inhibit the NLRP3 inflammasome and reverses the metabolic cost of inflammation (Marchetti et al., 2018a). The drug was originally formulated as a candidate for the topical treatment of degenerative arthritis, and an oral form was subsequently developed. As with topical gels, oral capsules have been shown to be safe and well tolerated in humans (Marchetti et al., 2018a; Marchetti et al., 2018b). Dapansutril has been shown to be safe in humans, and Dapansutril was the first NLRP3 inhibitor to complete two human proof-of-concept studies, one for acute gouty arthritis flares (Phase 2a) and one for stable systolic heart failure (NYHA II-III) (Phase 1b) (Aliaga et al., 2021). Dapansutril has been shown to specifically block NLRP3 inflammasome formation and prevent caspase-1 activation and IL1B maturation and release (Yang et al., 2019; Lonnemann et al., 2020). NLRP3 inflammasome formation also induces pyroptosis (Jorgensen and Miao, 2015). Interestingly, dapansutril reduced neutrophil infiltration and joint swelling and inhibited the secretion of pro-inflammatory factors IL1B, IL18, and IL6 in a mouse model of yeast glycan- and urate-induced arthritis (Marchetti et al., 2018b; Zhao and Zhao, 2020). At the same time, activation of the NLRP3 inflammasome induces the maturation of IL1B and IL18, both of which are effective targets for the treatment of acute and chronic inflammatory diseases (Dinarello et al., 2012).

However, there is currently a lack of more in-depth and systematic research on dapansutril in the treatment of gouty arthritis. Molecular dynamics can comprehensively and systematically simulate the interaction and binding stability between small molecule monomers and protein targets with the help of powerful computing power.

Molecular dynamics (MD) use large computer clusters as the carrier to obtain data such as the microstructure, physicochemical properties, and performance characterization parameters of small molecule drugs and proteins through calculation (Santos et al., 2019). It is a supplement and in-depth exploration of traditional biomedical disciplines based on experiments. Through the data obtained by calculation, the mechanism behind the experiment is analyzed from the micro, meso and macro scales of multi-level research. Molecular dynamics analyzes the behavioral laws of molecular motion by solving the potential functions and motion equations of intermolecular interactions, it simulates the dynamic evolution process of the system, and it provides microscopic quantities (such as the coordinates and velocity of molecules) and macroscopic observable quantities (such as: the relationship between the temperature, pressure, heat capacity of the system, etc.) (Sivakumar et al., 2020), so as to study the equilibrium

properties and mechanical properties of composite systems. Therefore, molecular dynamics can be systematically and comprehensively analyze the stability and affinity of dapansutril and gouty arthritis related protein targets.

Since the mechanism of action of dapansutril in the treatment of gouty arthritis is still unclear, this study used bioinformatics to screen core targets between dapansutril and gouty arthritis, and we used gene ontology (GO), Kyoto Encyclopedia of Genes and Genomes (KEGG) and protein-protein interaction (PPI) to analyze target genes and explore their mechanisms of action and potential pathways. We used molecular system motions to simulate the results of computing interactions from the cellular level to the chemical group level. Molecular docking was used to determine the affinity of monomeric compounds to protein targets, and molecular dynamics were used to model the stability of bound complexes. Research on the mechanism of action of dapansutril in the treatment of gouty arthritis will promote the related research and clinical application of the drug.

Materials and methods

Core gene targets screening and protein-protein interaction network building

"Gouty arthritis" was used to be the key word to obtain the disease gene targets through GeneCards database. The STRING database was used to analyze the protein-protein interaction (PPI) of dapansutril in the treatment of gouty arthritis (Szkarczyk et al., 2019). In order to further clarify the interaction between potential protein targets, all potential therapeutic protein targets of dapansutril on gouty arthritis were imported into Cytoscape 3.7.1 to analyze (Shannon et al., 2003), we defined the protein type as "*Homo sapiens*," and obtained relevant information on protein interactions by STRING database. Finally, the network topology parameters were analyzed by Cytoscape 3.7.1, and the core protein targets were screened out according to the criterion that the node degree value and the betweenness center value were greater than the average value.

The gene target enrichment analysis

The main biological processes and signaling pathways of dapansutril on gouty arthritis were analyzed through DAVID database. The interaction gene targets were used in DAVID database for gene ontology (GO) functional annotation and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis. We obtained molecular function (MF), cellular component (CC) and biological process (BP) of the

gene targets through GO enrichment. The disease related targets obtained from screening were input into the DAVID database by entering the list of target gene names and selecting the species as “*Homo sapiens*” (Huang et al., 2009). In this study, KEGG pathway enrichment analysis was performed on the relevant signaling pathways involved in the target, and gene target screening was performed under the condition of $p < 0.05$ (Cao et al., 2022).

Network diagram of “disease-core target gene-drug”

This study used the Cytoscape 3.7.1 network map software to construct a disease-core target gene-drug network and conduct topological analysis. The core gene targets can be screened based on the node degree value greater than 2 times the median (Cao et al., 2022).

Molecular docking and validation of the docking protocol

Molecular docking was used to study the molecular affinity of dapansutride with gouty arthritis protein targets. In this study, AutoDock Vina 1.1.2 software was used for molecular docking work. Before docking, PyMol 2.5 was used to process all receptor proteins, including removal of water molecules, salt ions and small molecules (Kim et al., 2016). ADFRsuite 1.0 was used to convert all processed small molecules and receptor proteins into the PDBQT format necessary for docking with AutoDock Vina 1.1.2. The output highest scoring docked conformation was considered to be the binding conformation for subsequent molecular dynamics simulations (Kim et al., 2016). The study used the original crystal ligand of the protein target as a positive reference, and we analyzed and compared the binding posture of the original crystal ligand and protein, the chemical bond length and the chemical bond angle by re-docking the original crystal ligand and protein. Finally, the consistency of the binding mode could indicate the correctness of the molecular docking protocol (Cao et al., 2022).

Molecule dynamics

In this study, the stability of the conformational complex of the dapansutride and gouty arthritis proteins was further verified by molecular dynamics simulations. Molecular dynamics (MD) simulation is a comprehensive set of molecular simulation methods combining physics, mathematics and chemistry. This method mainly relies on Newtonian mechanics to simulate the motion of molecular systems, we calculate macroscopic properties such as

thermodynamic quantities of a system by taking samples from an ensemble of different states of a molecular system (Burley et al., 2017). In this study, the small molecules and protein complexes obtained from the molecular docking results were used as the initial structures, and AMBER 18 software was used to conduct all-atom molecular dynamics simulations (Maier et al., 2015). The charge of the small molecule was calculated in advance by the antechamber module and the Hartree-Fock (HF) SCF/6-31G* of the gaussian 09 software before the simulation (Harrach and Drossel, 2014). Finally, the simulated topology and parameter files were exported. After the initial addition of hydrogen atoms to each system, the system used a steepest descent algorithm for vacuum minimization (Wang et al., 2006). The solvent was then added and the system ions were balanced with counter ions (Na⁺/Cl[−]).

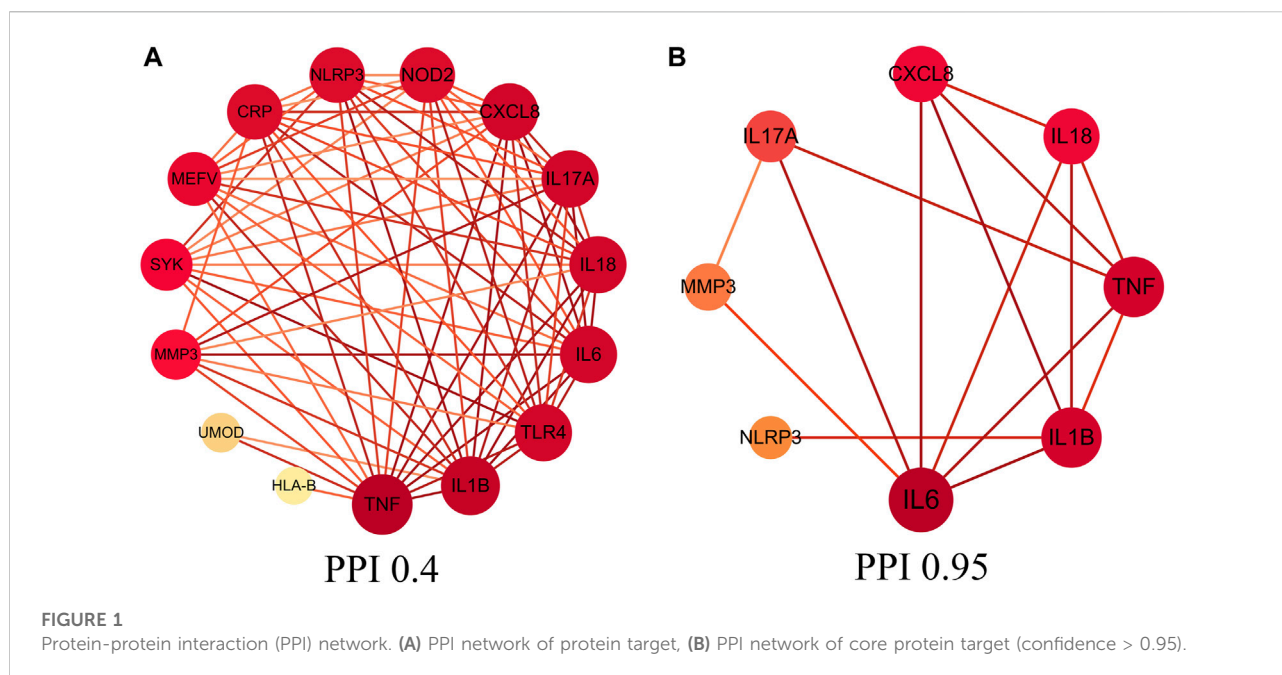
The proteins were all energy minimized using the steepest descent method and the conjugate gradient method. Subsequently, combined NVT and NPT (1,000 ps, 2 fs dt) and MD tests (100 ns, 2 fs dt) were performed at 298 K temperature and 1 bar pressure using a jump-integration algorithm. The coordinates and energy of the system are saved every 10 ps. Finally, 50 ns production simulations were carried out for each system under periodic boundary conditions. For all simulations, the van der Waals force (vdw) cutoff and short-range electrostatic interactions were set to 10 Å. The Particle-Mesh-Ewald (PME) method was used to evaluate long-range electrostatic interactions. Molecular dynamics simulation trajectories include protein-ligand complex root mean square deviation (RMSD), root mean square fluctuation (RMSF), radius of gyration and solvent accessible surface area (SASA).

MMGBSA binding free energy calculation

In this study, the binding free energy of the compound was investigated by MM-PBSA method, and its conformational stability was studied in detail. We calculated the binding free energies between proteins and ligands in all systems using the MM/GBSA method (Hou et al., 2011). The molecular dynamics trajectory of 50 ns was used for calculation, and the specific formula was as follows:

$$\begin{aligned}\Delta G_{\text{bind}} &= \Delta G_{\text{complex}} - (\Delta G_{\text{receptor}} + \Delta G_{\text{ligand}}) \\ &= \Delta E_{\text{internal}} + \Delta E_{\text{VDW}} + \Delta E_{\text{elec}} + \Delta G_{\text{LGB}} + \Delta G_{\text{SA}}\end{aligned}$$

In the formula, bond energy (E_{bond}), angular energy (E_{angle}), torsion energy (E_{torsion}), ΔG_{GA} and ΔG_{GB} are collectively called solvation free energy. The non-polar solvation free energy (ΔG_{GA}) was calculated based on the product of surface tension (γ) and solvent accessible surface area (SA), $\Delta G_{\text{GA}} = 0.0072 \times \text{SASA}$ (Cao et al., 2022).



Results

Core target screening and protein-protein interaction network diagram

A total of 220 gouty arthritis gene targets was screened through GeneCards database. We obtained core gene targets through relevance score, relevance score ≥ 20 which was considered as core gene target, the study analyzed 18 core gene targets through STRING database to construct the PPI network interaction map of target proteins of dapansutra for gouty arthritis, shown in Figure 1A. Eight core genes (such as: NLRP3, IL1B, CXCL8, etc.) were obtained by improving the confidence score (confidence degree > 0.95), and the study used the eight core gene targets to construct the core PPI network, shown in Figure 1B.

Gene ontology and Kyoto Encyclopedia of Genes and Genomes enrichment analysis

The 18 gene targets were imported into the DAVID database for enrichment analysis. Under the condition of $p < 0.05$, the GO enrichment analysis yielded 137 GO entries, including 126 BP entries, 6 CC entries, and 5 MF entries. According to the number of targets contained, the top 5 BP, CC, and MF compressions were screened. The results showed that biological processes were highly correlated with inflammation and cytokine regulation, mainly involving the inflammatory response, positive regulation of interleukin-1 beta production, and cellular response to lipopolysaccharide. Among cell components, extracellular space, extracellular region and cell surface account for a relatively large

amount. In molecular functions, cytokine activity, peptidoglycan binding and protein binding were relatively high, shown in Figures 2A–F. KEGG pathway analysis yielded 48 pathways with $p < 0.05$. According to the number of targets contained, the top 15 pathways were screened. The results showed that the enriched pathways involved multiple pathways related to inflammation and immune response, mainly rheumatoid arthritis, NOD-like receptor signaling pathway, IL17 signaling pathway and other signaling pathways, shown in Figures 2G,H.

Disease-core gene target-drug network

The disease-core gene target-drug network was constructed to show the main signal pathway and biological process of dapansutritile in the treatment of gouty arthritis, shown in Figure 3.

Molecular docking

The eight core gene targets were selected for molecular docking. The stability of receptor-ligand binding depends on the binding energy. The lower the binding energy of the complex, the more stable the receptor-ligand binding conformation. The small molecule dapansutritile interacted with various proteins mainly through hydrogen bonding and hydrophobic interactions. In addition, we were surprised that the cyano group of the small molecule dapansutritile was the main cyano group acceptor, which had hydrogen bonds with various proteins, while the sulfone group did not play the role of hydrogen bond acceptors. Moreover, we observed

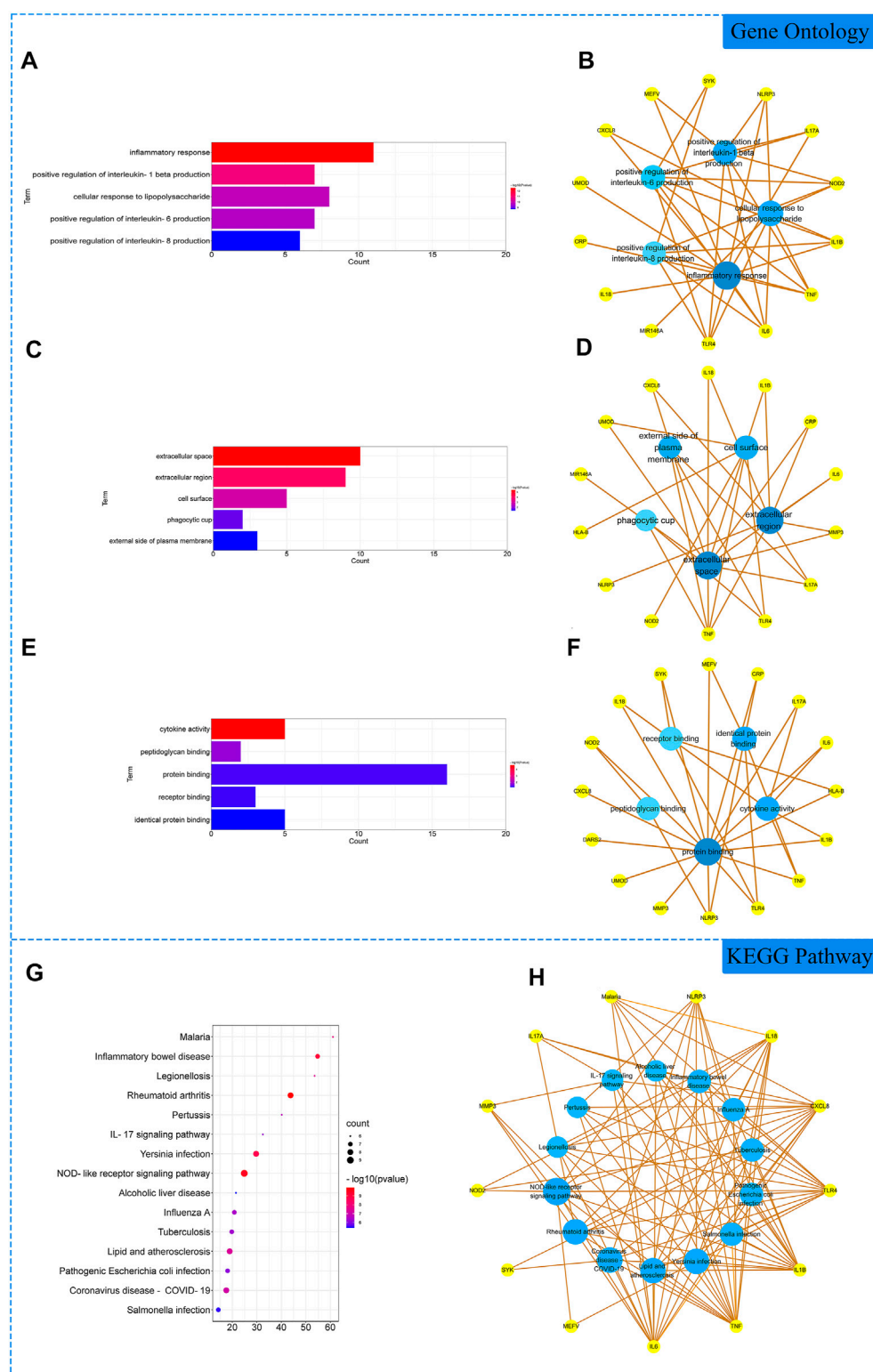
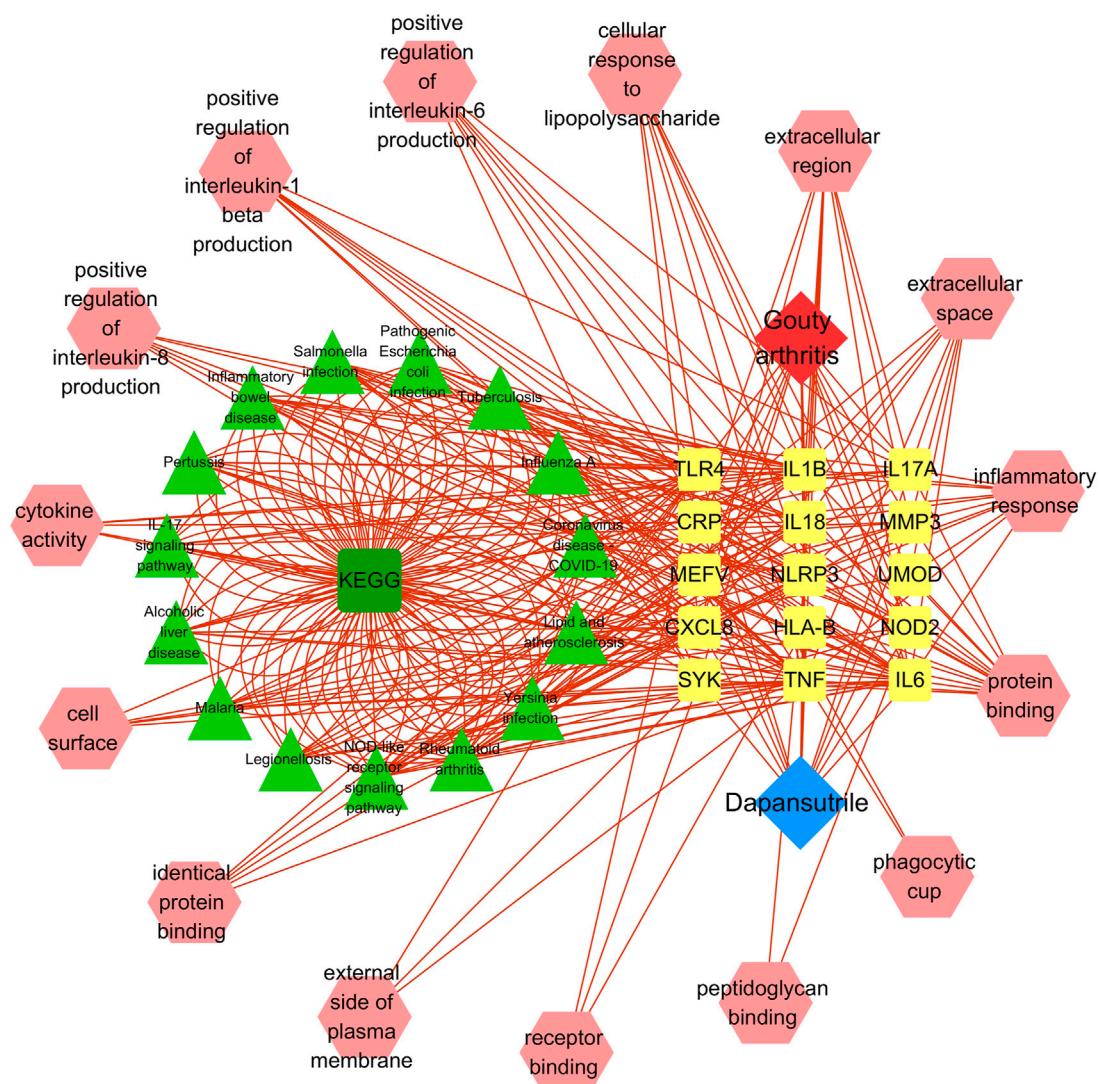


FIGURE 2

Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG). Analysis of related genes. (A) The top 5 terms in biological processes (BP) were greatly enriched. (B) The subnetwork displayed the top 5 BP terms and related genes. (C) The top 5 terms in cellular components (CC) were greatly enriched. (D) The subnetwork displayed the top 5 CC terms and related genes. (E) The top 5 terms in molecular function (MF) were greatly enriched. (F) The subnetwork displayed the top 5 MF terms and related genes. (G) The top 15 KEGG pathways were showed. (H) The subnetworks displayed the top 15 KEGG pathways.



RMSD, we could determine the stability of these complexes from strong to weak in order of Dapansutritile/IL18, Dapansutritile/MMP3, Dapansutritile/TNF, Dapansutritile/IL17A, Dapansutritile/IL1B, Dapansutritile/IL6, Dapansutritile/NLRP3, and Dapansutritile/CXCL8. However, RMSD results for all complexes indicated that small molecules could bind to proteins and maintain a relatively stable state. The results are shown in **Figure 5**.

Combined free energy calculation results

Based on the trajectory of molecular dynamics simulation, this study calculated the binding energy by using the MM-GBSA method, which can more accurately reflect the binding mode of

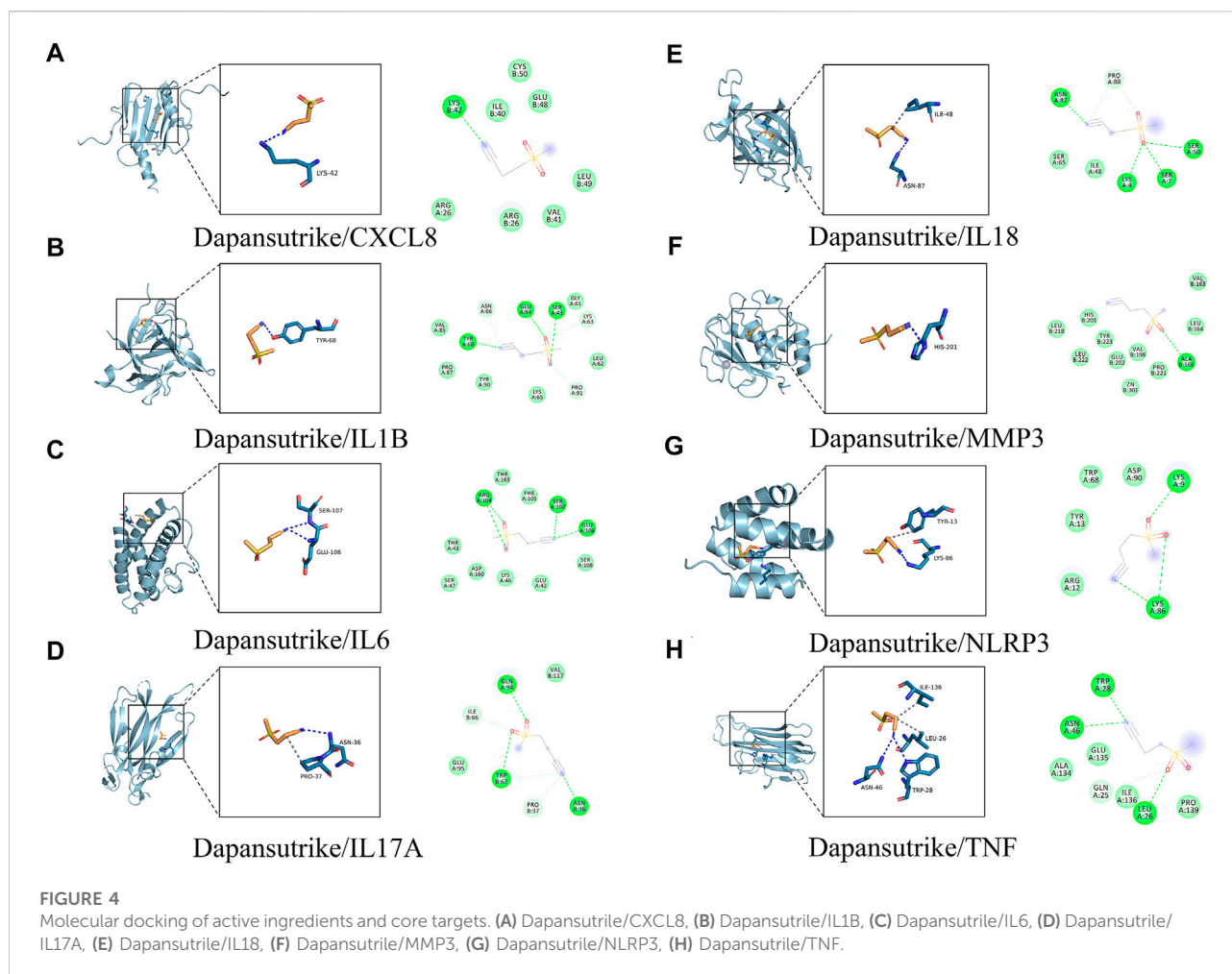


TABLE 1 The results of the amino acid residues of the complexes.

Complex	Van Der Waals	Conventional Hydrogen Bond	Carbon Hydrogen Bond
Dapansutrile/CXCL8	CYS-50, GLU-48, ILE-40, LEU-49, VAL-41, ARG-26	LYS-42	
Dapansutrile/IL1B	VAL-85, PRO-87, TYR -90, LYS-65, LEU-62, GLY-61	TYR-68, GLU-64, SER-43	SER-43, LYS-63, ASN-66, PRO-91
Dapansutrile/IL6	THR-163, PHE-105, SER-108, GLU-42, LYS-46, ASP-160, THR-43, SER-47	ARG-104, SER-107, GLU-106	
Dapansutrile/IL17A	VAL-117, GLU-95	GLN-94, TRP-67, ASN-36	TRP-67, PRO-37, ILE-66
Dapansutrile/IL18	SER-65, ILE-48	ASN-87, SER-50, SER-7, LYS-4	PRO-88
Dapansutrile/MMP3	VAL-163, LEU-164, PRO-221, VAL-198, ZN-301, GLU-202, TYR-233, HIS-201	ALA-165	
Dapansutrile/NLRP3	ASP-90, TRP-68, TYR-13, ARG-12	LYS-9, LYS-86	
Dapansutrile/TNF	ALA-134, GLU-135, ILE-136, PRO-139	TRP-28, ASN-46, LEU-26	GLN-25

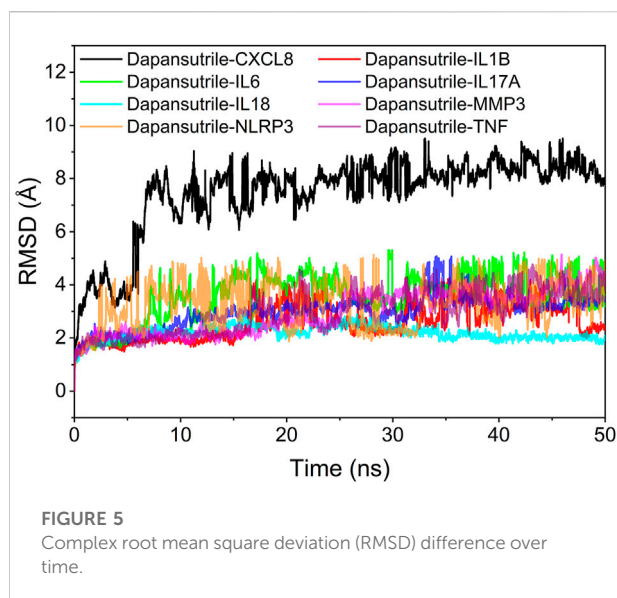


FIGURE 5
Complex root mean square deviation (RMSD) difference over time.

small molecules and target proteins. The simulation results suggested that the binding energies of Dapansutrile/CXCL8, Dapansutrile/IL1B, Dapansutrile/IL6, Dapansutrile/IL17A, Dapansutrile/IL18, Dapansutrile/MMP3, Dapansutrile/NLRP3, Dapansutrile/TNF were -2.79 ± 0.48 kcal/mol, respectively, -6.75 ± 0.52 kcal/mol, -5.67 ± 0.40 kcal/mol, -6.59 ± 0.41 kcal/mol, -3.85 ± 0.64 kcal/mol, -45.88 ± 0.85 kcal/mol, -7.06 ± 0.35 kcal/mol, -4.54 ± 0.35 kcal/mol. The value of the operation result indicates the affinity of the molecule to bind to the target protein, a lower value indicates a stronger binding affinity. The results showed that the small molecule and the corresponding proteins have strong binding affinity. The Dapansutrile/MMP3 binding was significant, the value was -45.88 ± 0.85 kcal/mol. The binding energies of these complexes are mainly contributed by van der Waals energy and electrostatic energy. The experimental results are shown in Table 2.

TABLE 2 Binding free energies and energy components predicted by MM/GBSA (kcal/mol).

System name	ΔE_{vdw}	ΔE_{elec}	ΔG_{GB}	ΔG_{SA}	ΔG_{bind}
Dapansutrile/CXCL8	-10.11 ± 0.63	-10.57 ± 1.41	19.83 ± 1.54	-1.93 ± 0.11	-2.79 ± 0.48
Dapansutrile/IL1B	-17.32 ± 0.48	-19.06 ± 1.36	32.38 ± 0.96	-2.74 ± 0.03	-6.75 ± 0.52
Dapansutrile/IL6	-17.18 ± 0.68	-15.22 ± 1.44	29.73 ± 1.88	-2.98 ± 0.10	-5.67 ± 0.40
Dapansutrile/IL17A	-17.26 ± 0.49	-19.15 ± 1.20	32.50 ± 1.32	-2.68 ± 0.06	-6.59 ± 0.41
Dapansutrile/IL18	-10.72 ± 0.76	-17.01 ± 1.81	26.24 ± 1.58	-2.36 ± 0.08	-3.85 ± 0.64
Dapansutrile/MMP3	-15.44 ± 0.65	-65.59 ± 0.98	38.69 ± 0.95	-3.53 ± 0.03	-45.88 ± 0.85
Dapansutrile/NLRP3	-6.04 ± 0.70	-7.13 ± 2.28	9.31 ± 2.62	-3.19 ± 0.11	-7.06 ± 0.35
Dapansutrile/TNF	-10.89 ± 0.41	-9.73 ± 1.06	18.10 ± 1.15	-2.00 ± 0.05	-4.54 ± 0.35

ΔE_{vdw} , van der Waals energy; ΔE_{elec} , electrostatic energy; ΔG_{GB} , electrostatic contribution to solvation; ΔG_{SA} , non-polar contribution to solvation; ΔG_{bind} , binding free energy.

Hydrogen bond analysis

Hydrogen bonding is one of the strongest non-covalent bonding interactions, the larger the number of hydrogen bonds indicates the better binding of the complex. The experimental results showed that the optimal hydrogen bond size and density for small molecules and proteins were dapansutril/IL18 and dapansutril/MMP3, the number of hydrogen bonds was stable at about three throughout the process. This was followed closely by Dapansutrile/IL17A. The formation of hydrogen bonds in the rest of the complexes was relatively sparse. The results are shown in Figure 6.

The stability of the target protein at the residue level

In this study, the vibrations of each residue after binding of small molecules and proteins are explored as root mean square fluctuations (RMSF). The RMSF can reflect the flexibility of the protein during molecular dynamics simulations. The binding of small molecule drugs to proteins reduces the flexibility of the proteins, which results in stabilization of the proteins and thus their efficacy. The results showed that most of the proteins had low RMSF except for the two ends, indicating that the protein core structure has good rigidity. Notably, the overall RMSF of IL1B, NLRP3 and TNF bound to small molecules was less than 2.5 \AA , indicating that these proteins are more rigid when bound to small molecules. The results are shown in Figure 7.

Analysis of the radius of gyration

The radius of gyration (RoG) can reflect the degree of compactness of the complex. The results reflected the variation of RoG over time for the six complexes during the molecular dynamics simulation. The experimental results showed the degree of denseness of the complexes from largest to smallest: Dapansutrile/NLRP3, Dapansutrile/IL1B, Dapansutrile/IL18, Dapansutrile/MMP3, Dapansutrile/TNF, Dapansutrile/IL6, Dapansutrile/IL17A,

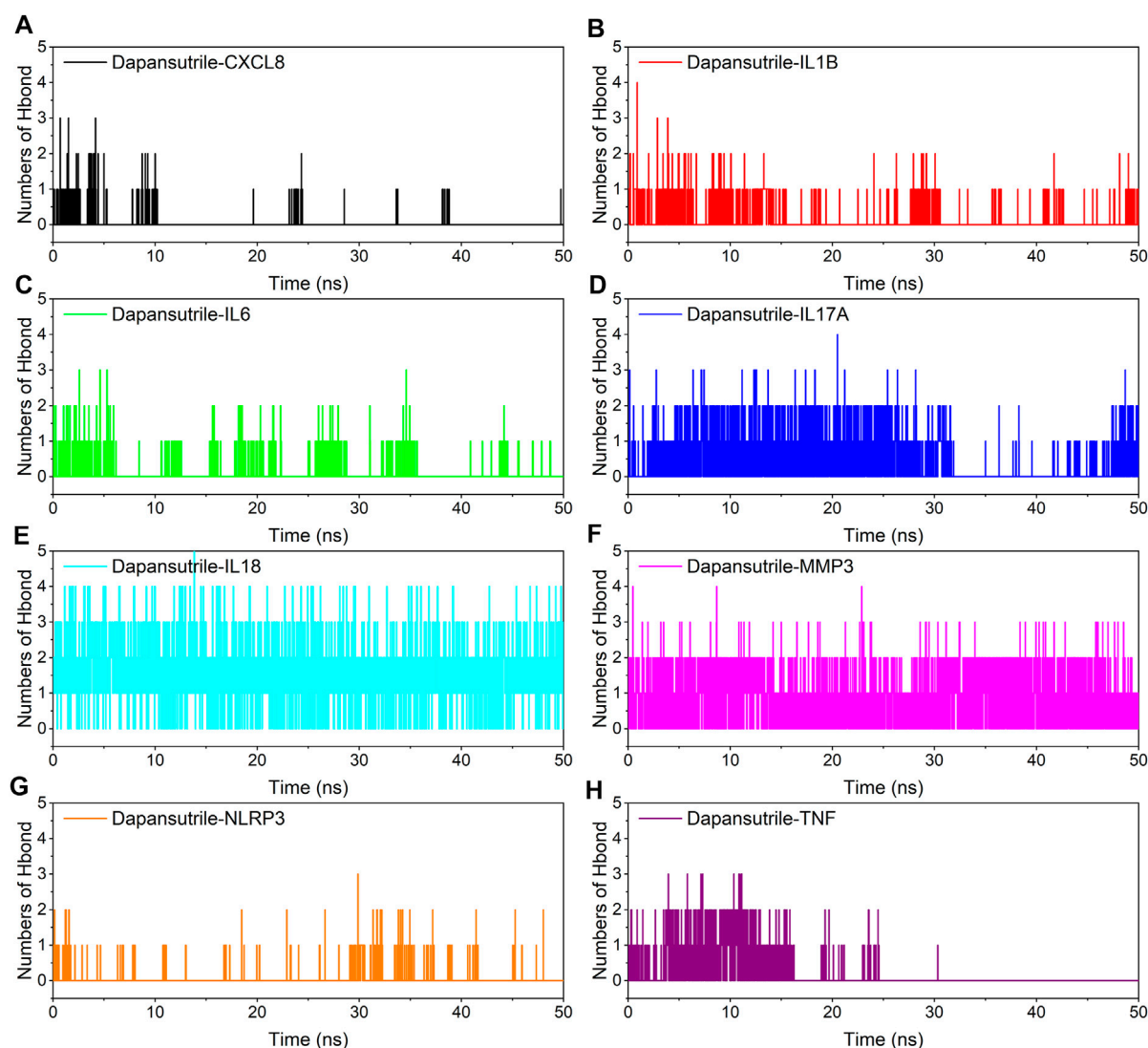


FIGURE 6
Changes in the number of hydrogen bonds between small molecule ligands and protein receptors in complex system simulations (A) Dapansutrine/CXCL8, (B) Dapansutrine/IL1B, (C) Dapansutrine/IL6, (D) Dapansutrine/IL17A, (E) Dapansutrine/IL18, (F) Dapansutrine/MMP3, (G) Dapansutrine/NLRP3, (H) Dapansutrine/TNF.

and Dapansutrine/CXCL8. The simulation results indicated that dapansutrine/NLRP3, dapansutrine/IL1B, dapansutrine/IL18, and dapansutrine/MMP3 have strong binding potential. The results of RoG experiments were consistent with those of RMSD. The results are shown in [Figure 8](#).

Analysis of solvent accessible surface area

The solvent accessible surface area (SASA) indicates the area where the complex can come into contact with the

aqueous solution. The contact area of the complex indicates the size of the interaction between the complex and the aqueous solution. In addition, the fluctuation of SASA responds to the exposure of the protein surface and the changes occurring in the buried area. The fluctuation analysis of SASA suggests that the fluctuations of Dapansutrine/NLRP3, Dapansutrine/TNF, Dapansutrine/IL1B, Dapansutrine/IL18, and Dapansutrine/MMP3 were small. This result implied the close interaction within the complex, which was the basis for the formation of stable binding of the complex. The results are shown in [Figure 9](#).

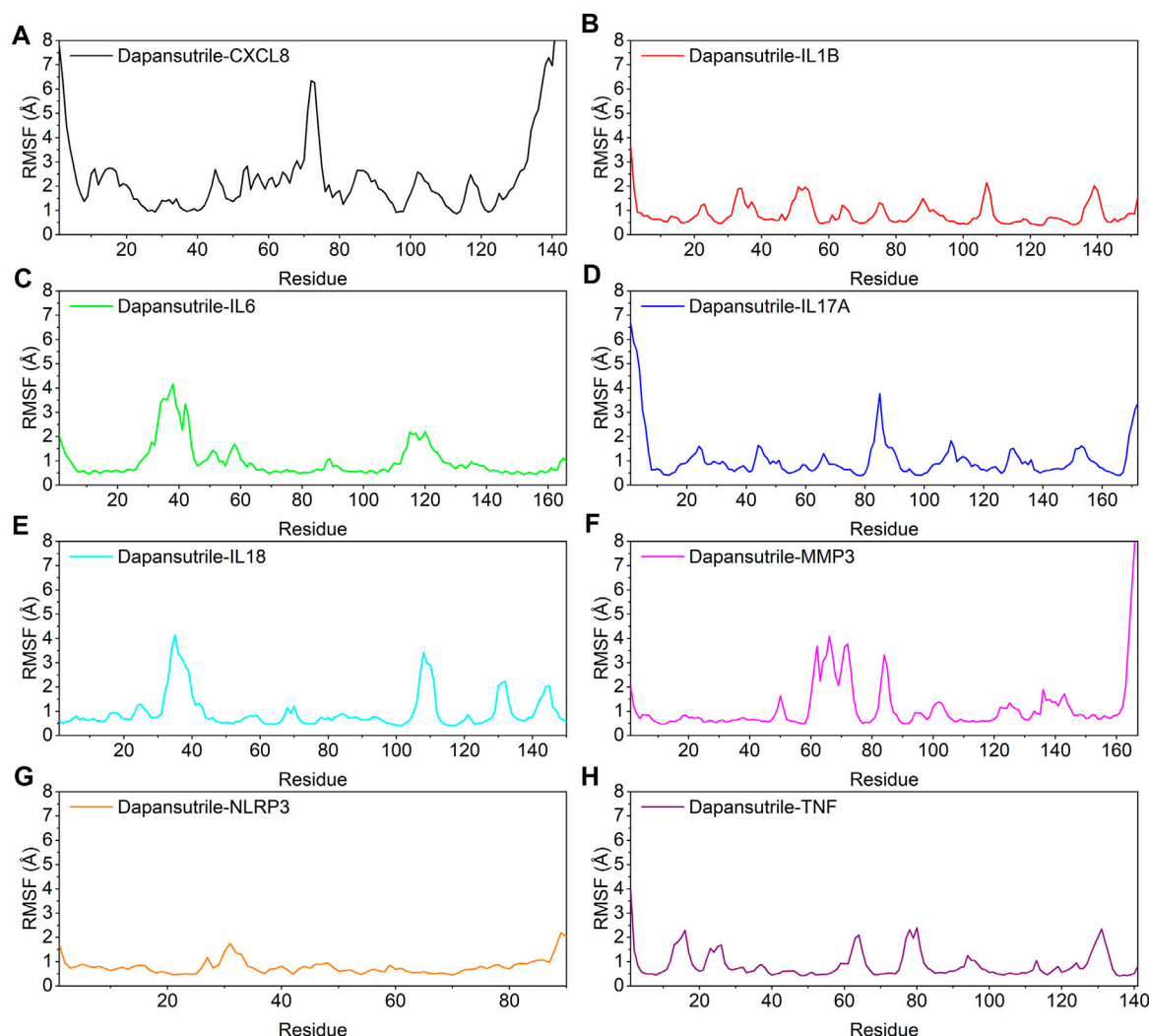


FIGURE 7

Changes in the stability of protein targets at the residue level (A) Dapansutride/CXCL8, (B) Dapansutride/IL1B, (C) Dapansutride/IL6, (D) Dapansutride/IL17A, (E) Dapansutride/IL18, (F) Dapansutride/MMP3, (G) Dapansutride/NLRP3, (H) Dapansutride/TNF.

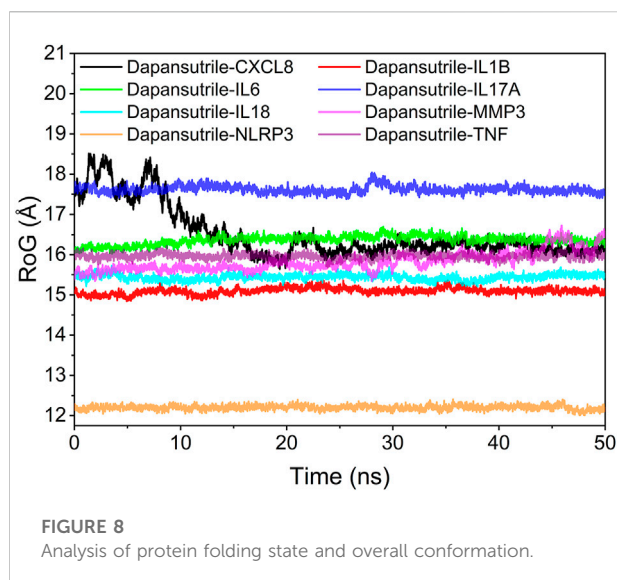
Discussion

This study explored the pharmacological mechanism of dapansutride in the treatment of gouty arthritis by molecular docking and molecular dynamics simulation based on molecular system movement. This study found that dapansutride may not only directly inhibit NLRP3 to reduce the inflammatory response and pyroptosis, but also hinder the chemotaxis and activation of inflammatory cells by regulating IL1B, IL6, IL17A, IL18, MMP3, CXCL8, and TNF. Firstly, dapansutride may attenuate inflammatory responses and reduce pyroptosis by directly inhibiting the NLRP3 inflammasome and hindering the activation of downstream inflammatory inflammation. Secondly, dapansutride may impede the activation of IL1B,

CXCL8, and TNF reducing the chemotaxis and activation of inflammatory cells. Finally, dapansutride may reduce the expression of MMP3 by regulating IL6, IL18, and IL17A, thereby degrading the extracellular matrix to treat gouty arthritis. Therefore, these results demonstrate that dapansutride treats gouty arthritis by inhibiting the inflammatory response from multiple targets.

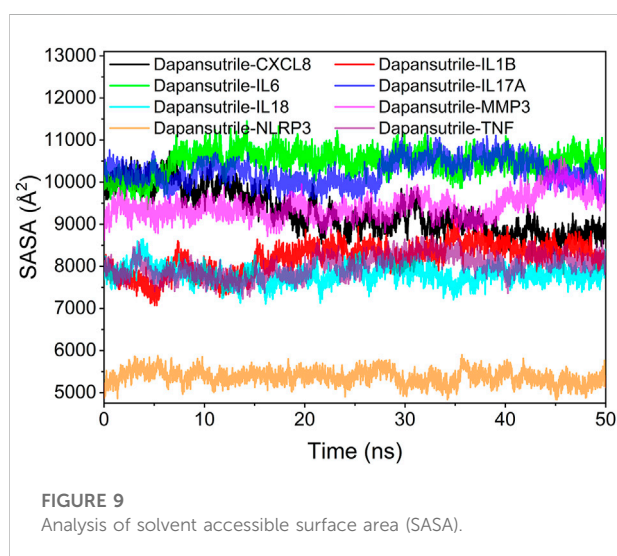
Analysis of molecular docking and molecular dynamics

Molecular docking can recognize each other through the spatial matching of drug small molecule dapansutride and protein



macromolecules *in vivo*, and predict their interaction and binding mode and affinity. Therefore, the results of molecular docking can be used to explore the mechanism of action of dapansutritile on gouty arthritis.

Firstly, molecular docking experiments indicated that dapansutritile had strong affinity for the protein targets NLRP3 and MMP3. The binding of dapansutritile in IL1B, CXCL8 and TNF were relatively stable, molecular docking showed that the binding of Dapansutritile/IL1B and Dapansutritile/IL18 is mainly maintained by hydrogen bonding and hydrophobic interaction. Dapansutritile/CXCL8 and Dapansutritile/MMP3 were mainly through hydrophobic interaction. Dapansutritile combined with IL6, IL18, and IL17A can form stable complexes, but there were



some abnormal fluctuations, which may be due to the influence of the number and angle of binding bonds.

Secondly, based on the trajectory of the molecular dynamics simulation, we calculated the binding energy using the MMGBSA method, which could more accurately reflect the binding mode of small molecules and target proteins. In the molecular dynamics simulation, the RMSDs of Dapansutritile/NLRP3 converged gradually in the first 5 ns of the simulation and preserved stable fluctuations in subsequent simulation. The RMSF results suggested that the overall RMSF of IL1B and TNF was less than 2.5 Å when they were bound to small molecules. These proteins are more rigid when bound to small molecules. The radius of gyration suggested that Dapansutritile/IL18 and Dapansutritile/MMP3 have stable fluctuations in size, which means they have high binding potential. Hydrogen bonding is also an important basis for the formation of stable binding between small molecule drugs and protein targets. The results of this study showed that the number of hydrogen bonds in Dapansutritile/IL18 and Dapansutritile/MMP3 were the best in terms of size and density. The number of their hydrogen bonds was stable throughout at about three. And all eight complexes can form stable hydrogen bonds. Therefore, the results of hydrogen bonding analysis showed that all the eight complexes had the basis for forming stable bonds.

Finally, the binding free energy consists mainly of non-bonding interactions (such as: Van der Waals interactions, electrostatic interactions and hydrogen bonding interactions). Because only non-bonding interactions are generally present in the actual drug small molecule (ligand) and protein (receptor) complex systems. The drug molecule binds reversibly to the proteins through non-bonding interactions, and this binding is more favorable for its own metabolism and excretion. Therefore, the results of free energy of binding were the comprehensive evaluation of the binding stability of drug and protein targets in this study. We analyzed our binding free energy consequences and other experimental results to derive a comprehensive ranking of the complex binding stability from strong to weak Dapansutritile/MMP3, Dapansutritile/NLRP3, Dapansutritile/IL1B, Dapansutritile/IL17A, Dapansutritile/IL6, Dapansutritile/TNF, Dapansutritile/IL18, Dapansutritile/CXCL8.

Dapansutritile may attenuate the inflammatory response in the treatment of gouty arthritis by inhibiting the NACHT, LRR, and PYD domains-containing protein 3 inflammasome

Dapansutritile may inhibit the activation of downstream inflammatory signaling pathways and reduce cell death by directly inhibiting the NLRP3 inflammasome.

Bioinformatics analysis suggested that dapansutritile can attenuate inflammatory responses and reduce pyroptosis by directly inhibiting the NLRP3 inflammasome and hindering the activation of downstream inflammatory inflammation. As a key component of inflammatory activation, NLRP3 plays a crucial role

in innate immunity and inflammation. NLRP3 has a regulatory role in inflammation, immune response and cellular scorching as an upstream activator of NF- κ B signaling. Analysis of protein interaction network PPI suggested that NLRP3 was closely related to inflammatory responses targets. KEGG signaling pathway analysis showed that NLRP3 was involved in protein metabolism and NOD-like receptor signaling pathway. GO analysis suggested that NLRP3 was involved in peptidoglycan binding.

The NLRP3 inflammasome is a complex containing the NLRP3 protein, the adaptor protein apoptosis-associated speck-like protein (ASC) and procaspase-1 (Inoue and Shinohara, 2013a, 2013b; Abderrazak et al., 2015; Song and Li, 2018). The interaction among the three proteins tightly regulates the function of the inflammasome to ensure immune activity only when appropriate (Shao et al., 2015). NLRP3 can trigger caspase-1 self-activation (Schroder and Tschopp, 2010). In the presence of immune activators [such as: Pathogen-associated molecular pattern molecules (PAMPs), danger-associated molecular patterns (DAMPs), other exogenous invaders, or environmental stress], NLRP3 opens and allows interaction between NLRP3 and the pyrin domain (PYD) in ASC. Subsequently, the caspase recruitment domain (CARD) of ASC in turn recruits the CARD domain on procaspase-1 for binding, resulting in the generation of the NLRP3 inflammasome (Martinon et al., 2006; Willingham et al., 2009; Schroder and Tschopp, 2010; Shao et al., 2015), NLRP3 inflammasome formation also triggers pyroptosis (Jorgensen and Miao, 2015). Gouty arthritis is driven by macrophage uptake of deposited sodium urate crystals and subsequent activation of the NLRP3 inflammasome (Martinon et al., 2006; Cavalcanti et al., 2016). However, dapansutrine is an orally active β -sulfonitrile molecule that inhibits NLRP3 inflammasome activation (Marchetti et al., 2018a; Sánchez-Fernández et al., 2019; Wohlford et al., 2020). Research showed that the concentrations of dapansutrine were found to inhibit NLRP3-ASC and NLRP3-caspase-1 interaction *in vitro* at 1 μ M or less. Interestingly, in LPS-stimulated human blood-derived macrophages, dapansutrine reduced IL1B levels by 60% and IL18 by 70%, *in vitro* at concentrations 100-fold lower than plasma concentrations safely achieved in humans (Marchetti et al., 2018a; Wohlford et al., 2020). Alba Sánchez-Fernández et al. (2019) found that prophylactic oral administration of dapansutrine resulted in a significant (2- to 3-fold) reduction in the protein levels of IL1B and IL18 as well as IL6 and TNF α in the spinal cord of EAE mice.

Bertinaria discussed whether the bond between dapansutrine and NLRP3 was covalent or non-covalent (Bertinaria et al., 2018). In fact, small molecules can have multiple binding sites with protein ligands at the same time. In our simulation studies, we found that the cyano and sulfone groups of dapansutrine could form hydrogen bonds with the protein NLRP3. Moreover, the drug usually binds to the protein target and acts through non-covalent bonds. Therefore, it is more reasonable to conclude that dapansutrine inhibits NLRP3 by forming a stable bond with NLRP3 through non-covalent bonds. At

the same time, our results showed that dapansutrine may not only act on NLRP3, but also block the downstream signaling pathway of NLRP3 by IL1B, IL18, IL6, thus reducing the inflammatory response. It is possible that our simulation results can better explain the strong NLRP3 inhibitory and anti-inflammatory effects of dapansutrine.

Therefore, dapansutrine may attenuate inflammation and reduces pyroptosis by inhibiting the NLRP3 inflammasome to treat gouty arthritis.

Dapansutrine in the treatment of gouty arthritis by inhibiting the occurrence of inflammation and chemotaxis

Dapansutrine may reduce the inflammatory response and the chemotaxis of inflammatory cells by blocking the activation of IL1B, CXCL8, and TNF for the treatment of gouty arthritis.

Dapansutrine may impede the activation of IL1B, CXCL8, and TNF reducing the chemotaxis and activation of inflammatory cells. IL1B is an important mediator of inflammatory response, and it is involved in various cellular activities such as cell proliferation, differentiation and apoptosis. IL1B is also involved in the pathogenesis of osteoarthritis. KEGG signaling pathway analysis suggested that IL1B regulated rheumatoid arthritis and glucocorticoids. GO analysis indicated that IL1B was associated with protein domain-specific binding. CXCL8 (also known as IL8) is a chemokine that attracts neutrophils, basophils and T cells. It is not only involved in neutrophil activation and chemotaxis, but also has a role in systemic inflammatory response syndrome (SIRS). KEGG signaling pathway analysis included cellular senescence and MIF-mediated glucocorticoid regulation. GO analysis included chemokine activity and interleukin 8 receptor binding. Protein interaction network analysis indicated that IL1B, CXCL8, and TNF were jointly engaged in the chemotaxis and activation of inflammatory cells. TNF is a multifunctional pro-inflammatory cytokine. This cytokine is mainly secreted by macrophages, it causes fever by direct action or through IL1B secretion, and has been implicated in the induction of cachexia. KEGG signaling pathway analysis suggested that TNF regulated inflammatory response and inflammatory bowel disease. GO analysis indicated that TNF affected cytokine activity.

Tengesdal et al. (2021) found that dapansutrine reduced pSTAT3 (Y705) by 82% and IL6 expression by 53%. IL6 binds to its receptor complex in IL6R/gp130 to activate downstream Janus kinases (JAKs), which subsequently activate signal converters and activators of transcription 3 (STAT3) through phosphorylation of tyrosine 705 (Mauer et al., 2015). König et al. (2021) found that activation of TNF was dependent on IL6 signaling, and TNF also limited the action of IL1B. Furthermore, activation of IL6 transsignal must be “downstream” of TNF signaling (König et al., 2021). Meanwhile, IL1B is highly expressed NF- κ B activator in triple negative breast cancer (TNBC) (Ignacio et al., 2019). NF- κ B can

increase the expression of TNF- α and IL6 (Huang et al., 2008; Liu et al., 2013). Therefore, IL1B and TNF- α can regulate each other. And You et al. (2021) demonstrated that IL1B enhanced the expression level of CXCL8 in TNBC cells.

Therefore, dapansutril may reduce the chemotaxis and activation of inflammatory cells for the treatment of gouty arthritis by blocking the activation of IL1B, CXCL8, and TNF.

Dapansutril in the treatment of gouty arthritis by degrading extracellular matrix

Dapansutril may regulate extracellular matrix degradation by reducing MMP3 expression through IL6, IL18, and IL17A in the treatment of gouty arthritis.

Dapansutril may reduce the expression of MMP3 by regulating IL6, IL18, and IL17A, thereby degrading the extracellular matrix to deal with gouty arthritis. IL6 is mainly produced at sites of acute and chronic inflammation, and it has been implicated in various inflammation-related disease states, including diabetes and systemic rheumatoid arthritis. Protein interaction network analysis showed that IL6, IL18, IL17, and MMP3 were closely connected with each other. KEGG signaling pathway analysis showed that IL6 was involved in dendritic cell developmental lineage pathways and cellular senescence. GO analysis included signaling receptor binding. IL17A is involved in inducing the production of inflammatory molecules, chemokines, antimicrobial peptides and remodeling proteins. IL17A plays a key role in inducing innate immune defenses, and IL17A stimulates non-hematopoietic cells and promotes the production of chemokines, which attract bone marrow cells to sites of inflammation. KEGG signaling pathway analysis included MIF-mediated glucocorticoid regulation and IL17 family signaling pathways. GO analysis included cytokine activity. IL18 is associated with tissue and organ damage and plays an important role in potentially fatal diseases characterized by cytokine storm. KEGG signaling pathway included MIF-mediated glucocorticoid regulation and IL-1 family signaling pathway. MMP3 is involved in the breakdown of extracellular matrix in arthritis and metastasis. MMP3 is regarded as involved in wound repair, atherosclerosis progression and tumor initiation. KEGG analysis included Gastrin-CREB signaling through PKC and MAPK. GO analysis included calcium binding and metalloproteinase activity.

Studies have shown that dapansutril can inhibit the IL6/STAT3 axis to inhibit breast cancer metastasis (Siersbæk et al., 2020; Tengesdal et al., 2021). Tantilertanant et al. (2019) found that circulating tension-upregulated IL6 increased MMP3 expression in human periodontal ligament cells. The IL6 amplifier (IL6 Amp) is an amplification mechanism, the

synergistic interaction of STAT3 with nuclear factor- κ B (NF- κ B) produces IL6 and various other cytokines and chemokines (Hirano, 2010; Murakami et al., 2013; Atsumi et al., 2014). Moreover, study found that IL6 could activate NF- κ B through the IL6-STAT3 axis (Hirano, 2021). Wang et al. (2019) found that IL18 promoted MMP secretion in human periodontal ligament fibroblasts by activating NF- κ B signaling. Koenders et al. (2005) found that IL17A promoted gastric cancer invasiveness through NF- κ B-mediated expression of MMP2 and MMP9. Therefore, IL17A and IL18 regulate MMP3 expression through NF- κ B. At the same time, the proinflammatory cytokines IL1B and TNF- α produced after activation of the NLRP3 inflammasome can also promote the expression of MMP3 (Burrage et al., 2006; You et al., 2021).

Therefore, Dapansutril may decrease the expression of MMP3 by regulating IL6, IL18, and IL17A, thereby degrading the extracellular matrix for the treatment of gouty arthritis.

Analyze the value of dapansutril in the treatment of gouty arthritis

Gouty arthritis is a metabolic rheumatic disease caused by disorders of purine metabolism and reduced synthesis or excretion of uric acid. Therefore, the treatment of gouty arthritis mainly revolves around correcting the metabolic abnormalities and reducing the inflammatory response of the body. Current studies have shown that dapansutril alleviates the clinical symptoms of gouty arthritis by suppressing the inflammatory response, and that dapansutril has good therapeutic effects and few side effects.

The results of our study further validated the important role of dapansutril in reducing the inflammatory response. And we believed that dapansutril could reduce the inflammatory response and body damage not only through NLRP3 but also through other protein targets (such as: MMP3, IL1B, and IL18). Moreover, the further development of related drugs should focus on and learn from some of the characteristics of dapansutril. Firstly, dapansutril has a very simple structure, and its molecular weight is around one hundred. The simple molecular structure allows dapansutril to have satisfactory transmembrane ability, the excellent transmembrane ability provides the basis for dapansutril to inhibit the intracellular inflammatory response. Secondly, dapansutril has good absorption, which can effectively improve drug utilization and reduce drug dosage, thus reducing drug side effects. Based on the good absorption properties, dapansutril is currently being developed for oral administration. Finally, the clear and reliable biosafety proof is the biggest advantage of dapansutril. The development and improvement of any related drug must be based on the principle that there is no or minimal biological toxicity.

The summary of the mechanisms analysis of dapansutril in the treatment of gouty arthritis is shown in Graphical Abstract.

Conclusion

This study explored the pharmacological mechanism of dapansutride in the treatment of gouty arthritis by molecular docking and molecular dynamics simulation based on molecular system movement. This study found that Dapansutride may not only directly inhibit NLRP3 to reduce the inflammatory response and pyroptosis, but also hinder the chemotaxis and activation of inflammatory cells by regulating IL1B, IL6, IL17A, IL18, MMP3, CXCL8, and TNF.

Therefore, Dapansutride treats gouty arthritis by attenuating inflammatory response, inflammatory cell chemotaxis and extracellular matrix degradation by acting on multiple targets.

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 authors.

Author contributions

J-FC, LX, XZ, and QZ contributed to the conception of the study; J-FC, XY, SC, MW, HX, and YG contributed

significantly to analysis and manuscript preparation; J-FC, XY, LZ, LX, and HX performed the data analyses and wrote the manuscript; XZ, QZ, J-FC, and LZ helped perform the analysis with constructive discussions.

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.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fphys.2022.990469/full#supplementary-material>

References

- Abderrazak, A., Syrovets, T., Couchie, D., El Hadri, K., Friguet, B., Simmet, T., et al. (2015). NLRP3 inflammasome: From a danger signal sensor to a regulatory node of oxidative stress and inflammatory diseases. *Redox Biol.* 4, 296–307. doi:10.1016/j.redox.2015.01.008
- Aliaga, J., Bonaventura, A., Mezzaroma, E., Dhakal, Y., Mauro, A. G., Abbate, A., et al. (2021). Preservation of contractile reserve and diastolic function by inhibiting the NLRP3 inflammasome with OLT1177(®) (dapansutride) in a mouse model of severe ischemic cardiomyopathy due to non-reperfusion anterior wall myocardial infarction. *Molecules* 26 (12), 3534. doi:10.3390/molecules26123534
- Atsumi, T., Singh, R., Sabharwal, L., Bando, H., Meng, J., Arima, Y., et al. (2014). Inflammation amplifier, a new paradigm in cancer biology. *Cancer Res.* 74 (1), 8–14. doi:10.1158/0008-5472.CAN-13-2322
- Bertinaria, M., Gastaldi, S., Marini, E., and Giorgis, M. (2018). Development of covalent NLRP3 inflammasome inhibitors: Chemistry and biological activity. *Arch. Biochem. Biophys.* 670, 116–139. doi:10.1016/j.abb.2018.11.013
- Bindu, S., Mazumder, S., and Bandyopadhyay, U. (2020). Non-steroidal anti-inflammatory drugs (NSAIDs) and organ damage: A current perspective. *Biochem. Pharmacol.* 180, 114147. doi:10.1016/j.bcp.2020.114147
- Blackham, R. E., Little, M., Baker, S., Augustson, B. M., and Macquillan, G. C. (2007). Unsuspected colchicine overdose in a female patient presenting as an acute abdomen. *Anaesth. Intensive Care* 35 (3), 437–439. doi:10.1177/0310057X0703500322
- Burley, S. K., Berman, H. M., Kleywegt, G. J., Markley, J. L., Nakamura, H., and Velankar, S. (2017). Protein data bank (PDB): The single global macromolecular structure archive. *Methods Mol. Biol.* 1607, 627–641. doi:10.1007/978-1-4939-7000-1_26
- Burrage, P. S., Mix, K. S., and Brinkerhoff, C. E. (2006). Matrix metalloproteinases: Role in arthritis. *Front. Biosci.* 11, 529–543. doi:10.2741/1817
- Cao, J., Li, L., Xiong, L., Wang, C., Chen, Y., and Zhang, X. (2022). Research on the mechanism of berberine in the treatment of COVID-19 pneumonia pulmonary fibrosis using network pharmacology and molecular docking. *Phytomed. Plus* 2 (2), 100252. doi:10.1016/j.phyplu.2022.100252
- Cavalcanti, N. G., Marques, C. D., Lins, E. L. T. U., Pereira, M. C., Rêgo, M. J., Duarte, A. L., et al. (2016). Cytokine profile in gout: Inflammation driven by IL6 and IL18? *Immunol. Invest.* 45 (5), 383–395. doi:10.3109/08820139.2016.1153651
- Dinarelli, C. A., Simon, A., and Van Der Meer, J. W. (2012). Treating inflammation by blocking interleukin-1 in a broad spectrum of diseases. *Nat. Rev. Drug Discov.* 11 (8), 633–652. doi:10.1038/nrd3800
- Finkelstein, Y., Aks, S. E., Hutson, J. R., Juurlink, D. N., Nguyen, P., Dubnov-Raz, G., et al. (2010). Colchicine poisoning: The dark side of an ancient drug. *Clin. Toxicol.* 48 (5), 407–414. doi:10.3109/15563650.2010.495348
- Harrach, M. F., and Drossel, B. (2014). Structure and dynamics of TIP3P, TIP4P, and TIP5P water near smooth and atomistic walls of different hydroaffinity. *J. Chem. Phys.* 140 (17), 174501. doi:10.1063/1.4872239
- Hirano, T. (2021). IL6 in inflammation, autoimmunity and cancer. *Int. Immunol.* 33 (3), 127–148. doi:10.1093/intimm/dxaa078
- Hirano, T. (2010). Interleukin 6 in autoimmune and inflammatory diseases: A personal memoir. *Proc. Jpn. Acad. Ser. B Phys. Biol. Sci.* 86 (7), 717–730. doi:10.2183/pjab.86.717
- 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 (1), 69–82. doi:10.1021/ci100275a
- Huang, D., Yang, C. Z., Yao, L., Wang, Y., Liao, Y. H., and Huang, K. (2008). Activation and overexpression of PARP-1 in circulating mononuclear cells promote TNF-alpha and IL6 expression in patients with unstable angina. *Arch. Med. Res.* 39 (8), 775–784. doi:10.1016/j.arcmed.2008.09.003

- Huang, Da W., Sherman, B. T., and Lempicki, R. A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* 4 (1), 44–57. doi:10.1038/nprot.2008.211
- Ignacio, R. M. C., Gibbs, C. R., Kim, S., Lee, E. S., Adunyah, S. E., and Son, D. S. (2019). Serum amyloid A predisposes inflammatory tumor microenvironment in triple negative breast cancer. *Oncotarget* 10 (4), 511–526. doi:10.18632/oncotarget.26566
- Inoue, M., and Shinohara, M. L. (2013a). NLRP3 inflammasome and MS/EAE. *Autoimmune Dis.* 2013, 859145. doi:10.1155/2013/859145
- Inoue, M., and Shinohara, M. L. (2013b). The role of interferon- β in the treatment of multiple sclerosis and experimental autoimmune encephalomyelitis - in the perspective of inflammasomes. *Immunology* 139 (1), 11–18. doi:10.1111/imm.12081
- Jorgensen, I., and Miao, E. A. (2015). Pyroptotic cell death defends against intracellular pathogens. *Immunol. Rev.* 265 (1), 130–142. doi:10.1111/imr.12287
- Khanna, D., Khanna, P. P., Fitzgerald, J. D., Singh, M. K., Bae, S., Neogi, T., et al. (2012). 2012 American College of Rheumatology guidelines for management of gout. Part 2: Therapy and antiinflammatory prophylaxis of acute gouty arthritis. *Arthritis Care Res.* 64 (10), 1447–1461. doi:10.1002/acr.21773
- Kim, S., Thiessen, P. A., Bolton, E. E., Chen, J., Fu, G., Gindulyte, A., et al. (2016). PubChem substance and compound databases. *Nucleic Acids Res.* 44 (D1), D1202–D1213. doi:10.1093/nar/gkv951
- Klück, V., Jansen, T., Janssen, M., Comarniceanu, A., Efdé, M., Tengesdal, I. W., et al. (2020). Dapansutril, an oral selective NLRP3 inflammasome inhibitor, for treatment of gout flares: An open-label, dose-adaptive, proof-of-concept, phase 2a trial. *Lancet Rheumatol.* 2 (5), e270–e280. doi:10.1016/s2665-9913(20)30065-5
- Koenders, M. I., Kolls, J. K., Oppers-Walgreen, B., Van Den Bersselaar, L., Joosten, L. A., Schurr, J. R., et al. (2005). Interleukin-17 receptor deficiency results in impaired synovial expression of interleukin-1 and matrix metalloproteinases 3, 9, and 13 and prevents cartilage destruction during chronic reactivated streptococcal cell wall-induced arthritis. *Arthritis Rheum.* 52 (10), 3239–3247. doi:10.1002/art.21342
- König, C., Vazquez, E., Eß, S., Ebbinghaus, M., Vorpahl, B., Ebersberger, A., et al. (2021). Spinal interleukin-1 β induces mechanical spinal hyperexcitability in rats: Interactions and redundancies with TNF and IL6. *J. Neurochem.* 158 (4), 898–911. doi:10.1111/jnc.15438
- Liu, Z., Guan, Y., Sun, X., Shi, L., Liang, R., Lv, X., et al. (2013). HSV-1 activates NF-kappaB in mouse astrocytes and increases TNF-alpha and IL6 expression via Toll-like receptor 3. *Neurol. Res.* 35 (7), 755–762. doi:10.1179/016164113X13703372991516
- Lonnemann, N., Hosseini, S., Marchetti, C., Skouras, D. B., Stefanoni, D., D'aleandro, A., et al. (2020). The NLRP3 inflammasome inhibitor OLT1177 rescues cognitive impairment in a mouse model of Alzheimer's disease. *Proc. Natl. Acad. Sci. U. S. A.* 117 (50), 32145–32154. doi:10.1073/pnas.2009680117
- Maier, J. A., Martinez, C., Kasavajhala, K., Wickstrom, L., Hauser, K. E., and Simmerling, C. (2015). ff14SB: Improving the accuracy of protein side chain and backbone parameters from ff99SB. *J. Chem. Theory Comput.* 11 (8), 3696–3713. doi:10.1021/acs.jctc.5b00255
- Marchetti, C., Swartzwelter, B., Gamboni, F., Neff, C. P., Richter, K., Azam, T., et al. (2018a). OLT1177, a β -sulfonyl nitrile compound, safe in humans, inhibits the NLRP3 inflammasome and reverses the metabolic cost of inflammation. *Proc. Natl. Acad. Sci. U. S. A.* 115 (7), E1530–E1539. doi:10.1073/pnas.1716095115
- Marchetti, C., Swartzwelter, B., Koenders, M. I., Azam, T., Tengesdal, I. W., Powers, N., et al. (2018b). NLRP3 inflammasome inhibitor OLT1177 suppresses joint inflammation in murine models of acute arthritis. *Arthritis Res. Ther.* 20 (1), 169. doi:10.1186/s13075-018-1664-2
- Martinon, F., Pétrilli, V., Mayor, A., Tardivel, A., and Tschopp, J. (2006). Gout-associated uric acid crystals activate the NALP3 inflammasome. *Nature* 440 (7081), 237–241. doi:10.1038/nature04516
- Mauer, J., Denson, J. L., and Brünig, J. C. (2015). Versatile functions for IL6 in metabolism and cancer. *Trends Immunol.* 36 (2), 92–101. doi:10.1016/j.it.2014.12.008
- Murakami, M., Harada, M., Kamimura, D., Ogura, H., Okuyama, Y., Kumai, N., et al. (2013). Disease-association analysis of an inflammation-related feedback loop. *Cell. Rep.* 3 (3), 946–959. doi:10.1016/j.celrep.2013.01.028
- Roddy, E., and Choi, H. K. (2014). Epidemiology of gout. *Rheum. Dis. Clin. North Am.* 40 (2), 155–175. doi:10.1016/j.rdc.2014.01.001
- Sánchez-Fernández, A., Skouras, D. B., Dinarello, C. A., and López-Vales, R. (2019). OLT1177 (dapansutril), a selective NLRP3 inflammasome inhibitor, ameliorates experimental autoimmune encephalomyelitis pathogenesis. *Front. Immunol.* 10, 2578. doi:10.3389/fimmu.2019.02578
- Santos, L. H. S., Ferreira, R. S., and Caffarena, E. R. (2019). Integrating molecular docking and molecular dynamics simulations. *Methods Mol. Biol.* 2053, 13–34. doi:10.1007/978-1-4939-9752-7_2
- Schroder, K., and Tschopp, J. (2010). The inflammasomes. *Cell.* 140 (6), 821–832. doi:10.1016/j.cell.2010.01.040
- Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., et al. (2003). Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13 (11), 2498–2504. doi:10.1101/gr.1239303
- Shao, B. Z., Xu, Z. Q., Han, B. Z., Su, D. F., and Liu, C. (2015). NLRP3 inflammasome and its inhibitors: A review. *Front. Pharmacol.* 6, 262. doi:10.3389/fphar.2015.00262
- Siersbæk, R., Scabia, V., Nagarajan, S., Chernukhin, I., Papachristou, E. K., Broome, R., et al. (2020). IL6/STAT3 signaling hijacks estrogen receptor α enhancers to drive breast cancer metastasis. *Cancer Cell.* 38 (3), 412–423. e419. doi:10.1016/j.ccell.2020.06.007
- Sivakumar, K. C., Haixiao, J., Naman, C. B., and Sajeevan, T. P. (2020). Prospects of multitarget drug designing strategies by linking molecular docking and molecular dynamics to explore the protein-ligand recognition process. *Drug Dev. Res.* 81 (6), 685–699. doi:10.1002/ddr.21673
- Song, N., and Li, T. (2018). Regulation of NLRP3 inflammasome by phosphorylation. *Front. Immunol.* 9, 2305. doi:10.3389/fimmu.2018.02305
- Szklarczyk, D., Gable, A. L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., et al. (2019). STRING v11: Protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* 47 (D1), D607–D613. doi:10.1093/nar/gky1131
- Tantilertanant, Y., Niyompanich, J., Everts, V., Supaphol, P., Pavasant, P., and Sanchavanakit, N. (2019). Cyclic tensile force-upregulated IL6 increases MMP3 expression by human periodontal ligament cells. *Arch. Oral Biol.* 107, 104495. doi:10.1016/j.archoralbio.2019.104495
- Tengesdal, I. W., Dinarello, A., Powers, N. E., Burchill, M. A., Joosten, L. A. B., Marchetti, C., et al. (2021). Tumor NLRP3-derived IL-1 β drives the IL6/STAT3 Axis resulting in sustained MDSC-mediated immunosuppression. *Front. Immunol.* 12, 661323. doi:10.3389/fimmu.2021.661323
- Wang, F., Guan, M., Wei, L., and Yan, H. (2019). IL-18 promotes the secretion of matrix metalloproteinases in human periodontal ligament fibroblasts by activating NF- κ B signaling. *Mol. Med. Rep.* 19 (1), 703–710. doi:10.3892/mmr.2018.9697
- Wang, J., Wang, W., Kollman, P. A., and Case, D. A. (2006). Automatic atom type and bond type perception in molecular mechanical calculations. *J. Mol. Graph. Model.* 25 (2), 247–260. doi:10.1016/j.jmgm.2005.12.005
- Wang, X., and Wang, Y. G. (2020). Progress in treatment of gout using Chinese and western medicine. *Chin. J. Integr. Med.* 26 (1), 8–13. doi:10.1007/s11655-019-3058-y
- Willingham, S. B., Allen, I. C., Bergstralh, D. T., Brickey, W. J., Huang, M. T., Taxman, D. J., et al. (2009). NLRP3 (NALP3, Cryopyrin) facilitates *in vivo* caspase-1 activation, necrosis, and HMGB1 release via inflammasome-dependent and -independent pathways. *J. Immunol.* 183 (3), 2008–2015. doi:10.4049/jimmunol.0900138
- Wohlford, G. F., Van Tassell, B. W., Billingsley, H. E., Kadariya, D., Canada, J. M., Carbone, S., et al. (2020). Phase 1B, randomized, double-blinded, dose escalation, single-center, repeat dose safety and pharmacodynamics study of the oral NLRP3 inhibitor dapansutril in subjects with NYHA II-III systolic heart failure. *J. Cardiovasc. Pharmacol.* 77 (1), 49–60. doi:10.1097/FJC.0000000000000931
- Wu, X. H., Wang, C. Z., Wang, S. Q., Mi, C., He, Y., Zhang, J., et al. (2015). Anti-hyperuricemia effects of allopurinol are improved by Smilax riparia, a traditional Chinese herbal medicine. *J. Ethnopharmacol.* 162, 362–368. doi:10.1016/j.jep.2015.01.012
- Xie, Z., Wu, H., Jing, X., Li, X., Li, Y., Han, Y., et al. (2017). Hypouricemic and arthritis relapse-reducing effects of compound tufuling oral-liquid in intercritical and chronic gout: A double-blind, placebo-controlled, multicenter randomized trial. *Med. Baltim.* 96 (11), e6315. doi:10.1097/MD.00000000000006315
- Yang, Y., Wang, H., Kouadir, M., Song, H., and Shi, F. (2019). Recent advances in the mechanisms of NLRP3 inflammasome activation and its inhibitors. *Cell. Death Dis.* 10 (2), 128. doi:10.1038/s41419-019-1413-8
- You, D., Jeong, Y., Yoon, S. Y., Kim, S. W., Nam, S. J., et al. (2021). Celastrol attenuates the inflammatory response by inhibiting IL-1 β expression in triple-negative breast cancer cells. *Oncol. Rep.* 45 (6), 89. doi:10.3892/or.2021.8040
- Yu, J., Li, L., Liu, J., and Chen, Z. (2022). Influence of intervention treatment by "heat-clearing and diuresis-promoting" prescription on NALP3, an inflammatory factor in acute gouty arthritis. *J. Orthop. Surg. Res.* 17 (1), 162. doi:10.1186/s13018-022-03046-z
- Zhao, C., and Zhao, W. (2020). NLRP3 inflammasome-A key player in antiviral responses. *Front. Immunol.* 11, 211. doi:10.3389/fimmu.2020.00211

Glossary

ACR The American College of Rheumatology

AGA Acute gouty arthritis

ASC Apoptosis-associated speck-like protein

BP Biological process

CARD The caspase recruitment domain

CC Cellular component

CXCL8 C-X-C motif ligand 8

DAMPs Damage-associated molecular patterns

GO Gene ontology

HF Hartree-Fock

IL17A Interleukin 17A

IL18 Interleukin 18

IL1B Interleukin 1 beta

IL6 Interleukin 6

KEGG Kyoto Encyclopedia of Genes and Genomes

LRR Leucine-rich repeat

MAPK Mitogen-activated protein kinases

MD Molecular dynamics

MF Molecular function

MIF Macrophage migration inhibitory factor

MMP2 Matrix metalloproteinase 2

MMP3 Matrix metalloproteinase 3

MMP9 Matrix metalloproteinase 9

MM-PBSA Molecular Mechanics-Poisson Boltzmann Surface Area

MSU Monosodium urate

NACHT Nucleotide-binding and oligomerization

NF- κ B Nuclear Factor Kappa B

NLRP3 The NACHT, LRR, and PYD domains-containing protein 3

NLRs NOD-like receptors

NSAIDs Non-steroidal anti-inflammatory drugs

NYHA New York Heart Association

PAMPs Pathogen-associated molecular pattern molecules

PKC Protein Kinase C

PME The Particle-Mesh-Ewald

PPI Protein-protein interaction

PRRs Pattern recognition receptors

PYD Pyrin domain

RMS Root mean square

RMSD Root mean square deviation

RMSF Root mean square fluctuation

SASA Solvent accessible surface area

SIRS Systemic inflammatory response syndrome

STAT3 Signal transducer and activator of transcription 3

TNBC Triple negative breast cancer

TNF Tumour necrosis factor alpha

Vdw Van der Waals force



OPEN ACCESS

EDITED BY
Chandrabose Selvaraj,
Alagappa University, India

REVIEWED BY
Ramar Vanajothi,
Bharathidasan University, India
Teng Fu,
Kangwon National University, South
Korea
Althaf Shaik,
King Saud University, Saudi Arabia

*CORRESPONDENCE
Natthiya Buensanteai,
natthiya@sut.ac.th

SPECIALTY SECTION
This article was submitted to Biophysics,
a section of the journal
Frontiers in Molecular Biosciences

RECEIVED 03 August 2022
ACCEPTED 30 August 2022
PUBLISHED 23 September 2022

CITATION
Papathoti NK, Mendam K,
Sriram Kanduri BH, Thepbandit W,
Sangpueak R, Saengchan C, Hoang NH,
Megavath VS, Kurakula M, Le Thanh T
and Buensanteai N (2022), Investigation
of bioactive compounds from *Bacillus*
sp. against protein homologs CDC42 of
Colletotrichum gloeosporioides
causing anthracnose disease in cassava
by using molecular docking and
dynamics studies.
Front. Mol. Biosci. 9:1010603.
doi: 10.3389/fmolb.2022.1010603

COPYRIGHT
© 2022 Papathoti, Mendam, Sriram
Kanduri, Thepbandit, Sangpueak,
Saengchan, Hoang, Megavath, Kurakula,
Le Thanh and Buensanteai. 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.

Investigation of bioactive compounds from *Bacillus* sp. against protein homologs CDC42 of *Colletotrichum gloeosporioides* causing anthracnose disease in cassava by using molecular docking and dynamics studies

Narendra Kumar Papathoti¹, Kishore Mendam²,
Bala Hanumath Sriram Kanduri³, Wannaporn Thepbandit¹,
Rungthip Sangpueak¹, Chanon Saengchan¹,
Nguyen Huy Hoang¹, Vineela Sai Megavath⁴, Madhuri Kurakula⁴,
Toan Le Thanh⁵ and Natthiya Buensanteai^{1*}

¹School of Crop Production Technology, Suranaree University of Technology, Nakhon Ratchasima, Thailand, ²Department of Zoology, Dr. B.R. Ambedkar Open University, Hyderabad, Telangana, India, ³R&D Division, Sri Yuva Biotech Pvt Ltd., Hyderabad, Telangana, India, ⁴Department of Biotechnology, Mahatma Gandhi University, Nalgonda, Telangana, India, ⁵Department of Plant Protection, Can Tho University, Can Tho City, Viet Nam

Manihot esculenta, commonly called cassava, is an economically valuable crop and important staple food, grown in tropical and subtropical regions of the world. Demand for cassava in the food and fuel industry is growing worldwide. However, anthracnose disease caused by *Colletotrichum gloeosporioides* severely affects cassava yield and production. The bioactive molecules from *Bacillus* are widely used to control fungal diseases in several plants. Therefore, in this study, bioactive compounds (erucamide, behenic acid, palmitic acid, phenylacetic acid, and β -sitosterol) from *Bacillus megaterium* were assessed against CDC42, a key protein for virulence, from *C. gloeosporioides*. Structure of the CDC42 protein was generated through the comparative homology modeling method. The binding site of the ligands and the stability of the complex were analyzed through docking and molecular dynamics simulation studies, respectively. Furthermore, a protein interaction network was envisaged through the STRING database, followed by enrichment analysis in the WebGestalt tool. From the enrichment analysis, it is apparent that bioactive from *B. megaterium* chiefly targets the MAP kinase pathway that is essential for filamentous growth and virulence. Further exploration through experimental studies could be advantageous for cassava improvement as well as to combat against *C. gloeosporioides* pathogen.

KEYWORDS

Colletotrichum gloeosporioides, Cdc42, *Manihot esculenta*, docking, molecular dynamics simulation

Introduction

Plants are an important source for humans, animals, birds, and other living organisms. Plants protect themselves from a variety of biotic and abiotic stresses (Gong et al., 2020). Biotic stress occurs due to bacterial and fungal pathogens (Peck and Mittler, 2020). One of the important plants on which more than 800 million people worldwide are depending for major food sources is *Manihot esculenta* Crantz (Atwijukire et al., 2019). The crop, commonly known as cassava, is enriched with several nutrients such as starch, carotenoids, vitamins, and minerals. Cassava is consumed as the primary food source mainly in the regions of tropical and sub-tropical countries. Later, due to increased industrial importance such as the production of animal feed, biomedicine, and cosmetics, the production of cassava has been highly increased (Li et al., 2017). In addition, the raw materials from *M. esculenta* were used for biopolymer, starch, and bioethanol production (El-Sharkawy 2004). Apart, the by-products from the cassava industries are rich with organic residues essential for the production of value-based products (Ayling et al., 2012). Hence, the crop with human value and with the immense industrial application was cultivated by both low-scale and high-scale cultivators. However, the plant is restricted at an economical level due to several factors such as the presence of cyanogen compounds (Balyejusa Kizito et al., 2007), a low level of proteins, and infectious disease.

Anthraxnose disease (AD) damages the healthy planting materials of cassava, leading to low yield and total economic loss for the planters. AD occurs in cassava due to the fungal pathogen *Colletotrichum gloeosporioides* f. sp. *manihotis* (Machado et al., 2020). The fungal strain infects the shoot tips of the healthy plants; develops cancerous growth on the stem and leaves. AD is notorious to cause shoot tip-die-back disease because the pathogen infects the stem region, weakens the parts and leads to major destruction during strong wind and rain (Pinweha et al., 2015). The primary interaction of the pathogen with the cassava plant was established by producing an infection cell known as appressorium. The melanized cell surrounding the aspersorium supports the internal solute concentration and rigidity of the cells (Wang et al., 2021). After the interaction with the host, the pathogen develops infection vesicles and primary hyphae. Later, the fungi develop secondary hyphae structures that spread the infection and kill the plants. Generally, after the successful infection into the host, the fungi adapt to the biotrophic mode of nutrition for their survival (Li et al., 2021). The pathogen produces lesions on leaves, stems, and other parts of the plant. Sequentially, switches to the necrotrophic mode of nutrition in which the pathogen

absorbs nutrition from the dead cells of the infected region. This nutrition adaptation by the pathogen is known as the hemibiotrophic mode of infection (Jacobs et al., 2019). Thus, it is very challenging to impair the growth and spread of infection by *C. gloeosporioides*. This pathogen also infects humans but knowledge about the type of disease and mode of infection is not clear so far.

To prevent fungal infection, chemical fungicides were widely used to control the disease (Ons et al., 2020). The use of several fungicides has resulted in impacts on human health and environmental issues. Hence, as an alternative approach to overcome AD-mediated damage in the cassava, different novel approaches were promoted for the development of fungal resistant-cassava crops (Koehorst-van Putten et al., 2012). The genetic engineering approach was one of the conventional strategies known to be the most economical, safe, and effective method to generate anthracnose disease-resistance cassava plants (Hormhuan et al., 2020). The use of a conventional breeding strategy with cassava crops leads to high heterozygosity, low fertility, delayed flowering, and prolonged vegetative stage. Hence, the approach of *Agrobacterium*-mediated transformation is considered to improve the acquired resistance in plants. One of the important plant-pathogen resistance genes, transferred into cassava has been reported to show improved resistance against a wide variety of plant pathogens. Thus, alternative strategies were required to incredulous the current scenario of anthracnose disease in cassava plants. The cell division cycle (CDC42) protein present in the fungi performs the function of the molecular switch by regulating signal transduction pathways and cytoskeleton-mediated cellular process. The protein belongs to the Rho-family of the GTP-binding protein, which plays a pivotal role in the transduction of polarity signals for morphogenetic development (Wang et al., 2018). The CDC42 protein also plays an important role in cell differentiation and appressoria development. CDC42 protein reported with plants is highly diverse, however, the protein is conserved in other eukaryotic species. The protein CDC42 from different fungal species (*Magnaporthe grisea*, *Claviceps purpurea*, and *Ustilago maydis*) has a key role in plant-pathogen interaction (Oeser et al., 2017; Zheng et al., 2009). Thus, the deletion of CDC42 from pathogens has significantly reduced the virulence mechanism during infection. Therefore, in the present study, CDC42 of *C. gloeosporioides* was selected as the therapeutic target to screen for inhibitors against the protein. Also, the detailed investigation using a protein-protein interaction network will pave the way to study the characteristic properties of CDC42 involved in the different biological processes of host-pathogen interaction.

Materials and methods

Generation of homology model and structure validation

The three-dimensional structure of Cell division control protein 42 homologs (CDC42) from *C. gloeosporioides* was determined through the comparative homology modeling method. The structure of CDC42 was built through the SWISS-MODEL server (<https://swissmodel.expasy.org/>). The accuracy of the model was assessed by QMEAN4 score analysis (Benkert et al., 2011). Later, energy minimization was performed using the steepest descent algorithm using GROMACS (Van Der Spoel et al., 2005). The structure of the predicted model was assessed through the structure validation tool SAVES v6.0 program - VERIFY 3D (Agarwal et al., 2021; Hasan et al., 2021), ERRAT (Adewole and Ishola, 2021), WHAT CHECK (Sekhar Pagadala et al., 2009), and PROCHECK analysis (Laskowski et al., 1996). The geometry and stereochemistry of the modeled structure were analyzed through the Ramachandran plot analysis method (Agarwal et al., 2021). In addition, structural validation of the generated model was performed through ProSA score analysis (Wiederstein and Sippl, 2007). Then, the overall quality score of the homology model was compared with the score of the template structure.

Binding site prediction

Prediction of druggable cavities is a crucial step for structure-based drug designing. The active site as predicted for CDC42 model protein using sitemap. The prediction of the active site reveals the shape, size and chemical interaction of the ligands with the receptor protein.

Molecular docking

Five monomeric bioactive compounds identified from the ethyl acetate extract of *Bacillus megaterium* erucamide, behenic acid, palmitic acid, phenylacetic acid, and β -sitosterol, were examined against CDC42 protein of *C. gloeosporioides*. The structure of the bioactive compounds was obtained from the PubChem database. The selected ligands were prepared and the three-dimensional (3D) coordinates were generated. For molecular docking, the proteins used for the study were prepared using protein preparation wizard. The proteins were subjected for H-bond optimization. The entire protein structure were relaxed using Uff force field. Energy minimization for protein and ligand was performed before docking using default parameters. Autodock tools were utilized for the addition of hydrogen, Kollman charges, and solvation parameters (Azam and Abbasi, 2013). Molecular docking was performed through the Autodock Vina program (Trott and Olson, 2010). The grid size of 3 Å for the coordinates X, Y,

and Z centered at X: 12.20; Y: 5.95; Z: 7.22 with a grid spacing of 0.375 Å was used for the docking program. The pose with the lowest binding energy was selected as the best conformation. The modeled structures were visualized through the BIOVIA Discovery Studio visualizer (Studio, 2015). Molecular mechanics of combined generalized born surface area and surface area continuum solvation (MM/PBSA and MM/GBSA) methods were performed for studying the effectiveness of interaction. The calculation for average binding free energy ΔG_{bind} was represented for estimating the free energy of ligands binding to the macromolecules. During molecular dynamics simulation of the receptor-ligand complex, the molecular mechanics is applied with empirical scoring and perturbation methods for predicting the accuracy during their simulation run. The formula for average binding free energy ΔG_{bind} was calculated as; $\Delta G_{\text{bind}} = \Delta E_{\text{MM}} + \Delta G_{\text{Solv}} + \Delta G_{\text{SA}}$.

ΔE_{MM} : denotes minimized energies of protein and ligand.

ΔG_{Solv} : solvation-free energy.

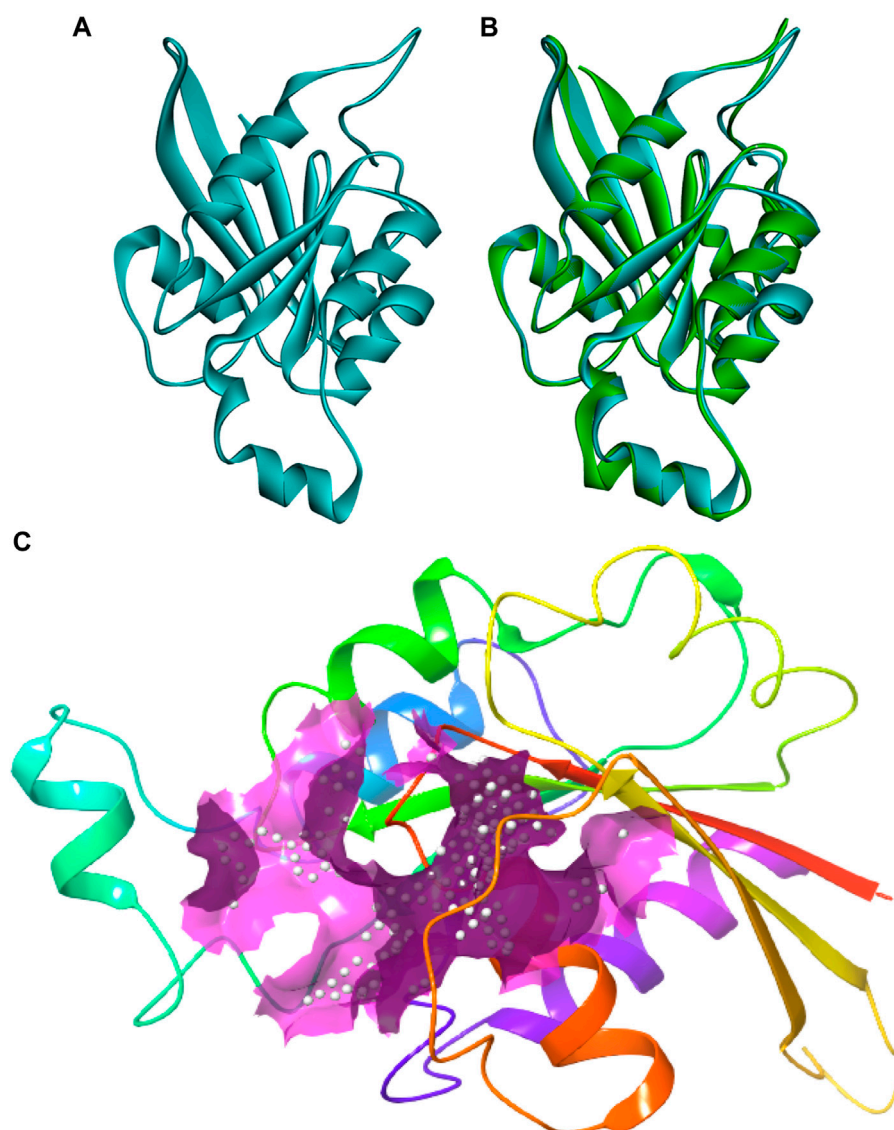
ΔG_{SA} : surface area energy.

ADMET properties of the ligands

SMILE structure of the lead molecules used for the present study were downloaded from Pubchem database. The pharmacokinetic properties of molecules were predicted using ADMETSAR2.0. The properties such as acute oral toxicity, BBB, fish aquatic toxicity and carcinogenicity of the molecules were analysis.

Molecular dynamics simulation

For each protein-ligand system, their pose with the lowest binding energy was assessed. The system was minimized and equilibrated under the number of particles, volume, and temperature (NVT) and the number of particles, pressure, and temperature (NPT) conditions. The molecular dynamics simulation was performed for 50 ns in DESMOND with GPU support. The Optimized Potential for Liquid Simulations (OPLS) force field was used. The system was solvated in a dodecahedron box using a simple point charge (SPC) model with a periodic boundary condition. The system was neutralized by adding sodium chloride ions. Energy minimization was performed through the steepest descent algorithm. Harmonic position restraints were applied during the NVT ensemble simulation. The molecular dynamics production runs were carried out at a 2 fs time step. Temperature and pressure were controlled by setting the Langevin dynamics and Berendsen barostat at 300 K and 1 bar, respectively. Standard periodic boundary conditions and cut-off distance (1 nm) were updated. The particle-mesh Ewald (PME) method was used to assess the interactions. The bonds were constrained with a linear constraint solver (LINCS) and the water molecules were constrained with SETTLE (Hess 2008; Tripathi et al., 2022). Molecular dynamics simulation was

**FIGURE 1**

Three-dimensional structure of CDC42 predicted through homology modeling **(B)** The structure of CDC42 protein superimposed with the template (2NGR) structure. **(C)** Binding site prediction representing the active site region (pink).

evaluated using root mean square deviation (RMSD) and hydrogen bond analysis.

STRING analysis

The interacting proteins of CDC42 were predicted through the STRING database (<http://string-db.org>) and the network is built by providing the CDC42 protein sequence in the input box. The search was performed against *C. gloeosporioides*. The confidence score was set to high (0.7). The interactions were based on the experiments, co-expression, databases, gene fusion,

neighborhood, and co-occurrence. The maximum number of interactions was set to no more than ten in the first and second shells.

Identification of clusters from the protein-protein interaction network

Clustering of interactions from large protein-protein interaction networks is essential to define the molecular complexes and topological modules. It is difficult to comprehend and interpret the network properties as such in

TABLE 1 Docking score of different ligands from *B. megaterium* with CDC42 protein.

S. No	PubChem ID	Compound name	Dock score (kJ mol ⁻¹)	ΔGbind (Kcal/mol)
1	222284	beta-Sitosterol	-10	-42.35
2	999	Phenylacetic acid	-10	-39.26
3	985	palmitic acid	-9.4	-41.51
4	8,215	Behenic acid	-9.2	-37.90
5	5365371	Eurcamide	-9.2	-39.49

TABLE 2 GO Slim summary is based on Entrez gene IDs.

S. No	Gene symbol	Gene name	Entrez gene
1	BMH1	14-3-3 family protein BMH1	856924
2	BUD6	Bud6p	851029
3	CLA4	serine/threonine protein kinase CLA4	855418
4	SHO1	osmosensor SHO1	856854
5	SPA2	Spa2p	850639
6	STE50	Ste50p	850325

large protein-protein interaction networks. Therefore, clustering of networks is significant in unraveling the pure network properties as well as finding the network connections in the dense regions. The network obtained from the STRING database is a network based interaction evidence for data support. The obtained network was reconstructed in Cytoscape 3.8.0. The constructed network was evaluated further using Molecular Complex Detection (MCODE) plug-in to visualize the central network. The cut off parameters were set as MCODE score >3 and node number >4. The subclusters generated were further visualized and group to study the closeness and degree of interaction in their group.

Gene ontology (GO) analysis and protein interaction analysis

The gene ontology (GO) analysis was performed in which the functional annotation was achieved through DAVID (database for annotation, visualization, and integrated discovery) database. GOView, a web-based WebGestalt (WEB-based GENE Set Analysis Toolkit) application, is used to visualize and compare the interactional relationship in the network (Zhang et al., 2005; Zhang et al., 2004). Furthermore, the central gene sets were annotated and the hierarchical associations were defined. The protein SHO1 involved in the MAPK signaling pathway was modeled using modeler. SHO1 from yeast was used as the template (2vkn. Pdb) with 61.02% with target sequence. The

model of SHO1 and CDC42 was loaded into Patchdock server and protein-protein docking was performed (Yousafi et al., 2021). The protein complex was analysed and results were represented.

Results

Generation of homology model and structure validation

The protein sequence of the CDC42 homolog of *C. gloeosporioides* (O94103) was retrieved from the NCBI database (<https://www.ncbi.nlm.nih.gov/protein/O94103>). The sequence contains 190 amino acids and belongs to the small GTPase superfamily, the Rho family. The sequence was predicted to contain three nucleotides (GTP) binding regions: 12–19 (GDGAVGKT), 59–63 (DTAGQ), and 117–120 (TERG). Based on sequence homology, the molecular function involves GTPase activity and the biological process involves cell cycle and cell division. The sequence contains a propeptide region from amino acid 188 to 190 (LVL), which is predicted to be cleaved during protein maturation or activation. The detail for the propeptide region was revealed through the prediction evidence sequence similarity search tool ECO: 0000250 mentioned in the NCBI protein sequence database. The structure of this CDC42 protein was predicted through comparative homology modeling. It showed 70.83% sequence

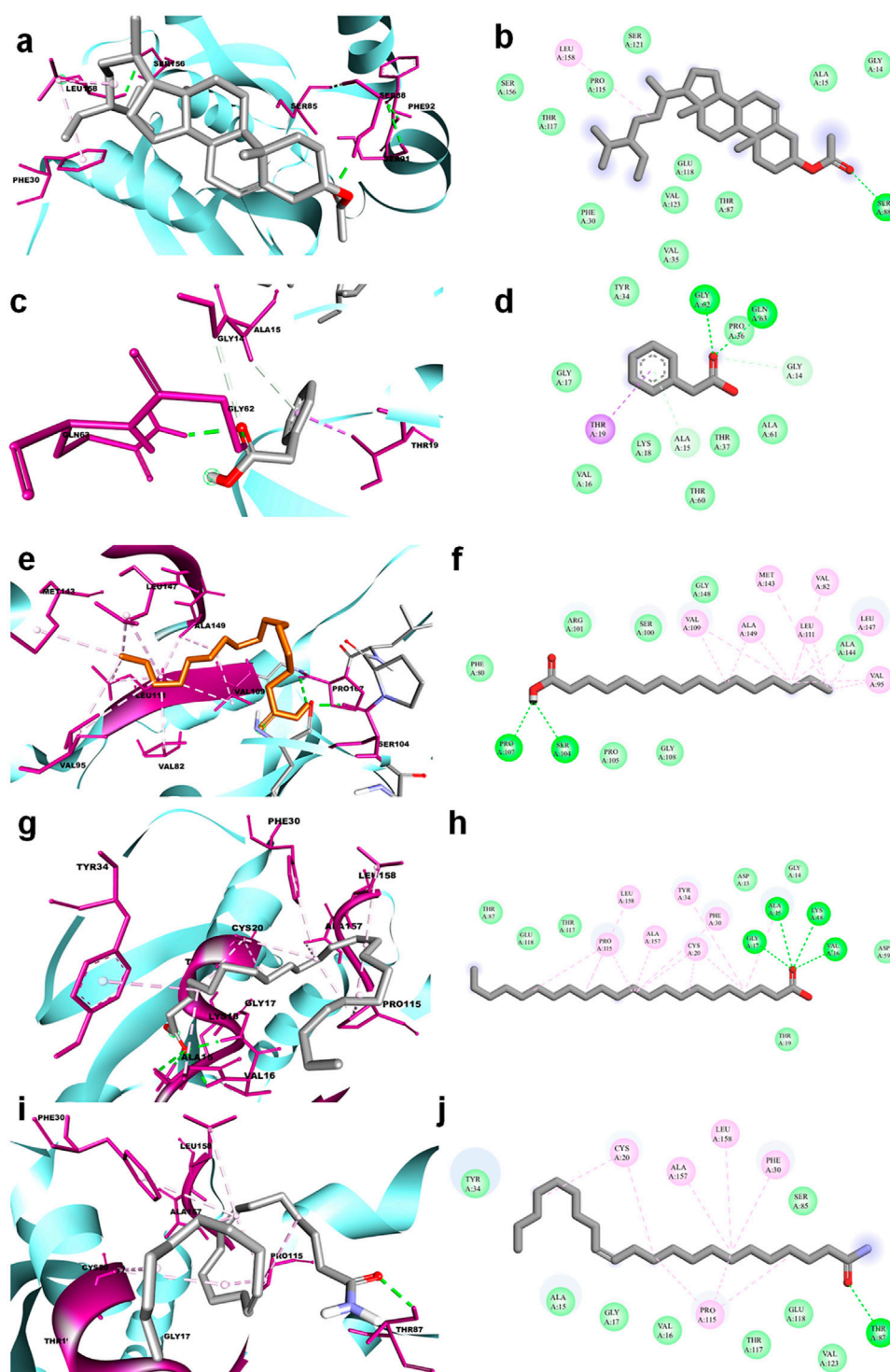


FIGURE 2

3D and 2D representation of molecular docking of CDC42 with ligands from *B. megaterium*. The ligands and their interaction are shown with the line diagram. The color code green color represents the hydrogen bond. Purple color represents pi-sigma interaction. Light pink color represents, pi-alkyl and alkyl interaction. (A,B) = beta sitosterol, (C,D) = Phenylacetic acid, (E,F) = palmitic acid, (G,H) = Behenic acid and (I,J) = Eurcamide.

TABLE 3 Predicted ADME physio-chemical properties of the docked compounds, all the tables cited correctly.

Compound name	Human intestinal absorption	BBB	Acute oral toxicity (log (1/(mol/kg)))	Fish aquatic toxicity	Carcinogenicity (binary)
Beta sitosterol	0.9930	0.9247	1.989	0.9917	0.9714
Phenylacetic acid	0.9490	0.9659	1.697	0.4220	0.7286
palmitic acid	0.8417	0.9725	1.16	0.9178	0.6571
Behenic acid	0.8417	0.9725	0.6378	0.9178	0.6571
Eurcamide	0.9186	0.9969	0.6537	0.7699	0.6429

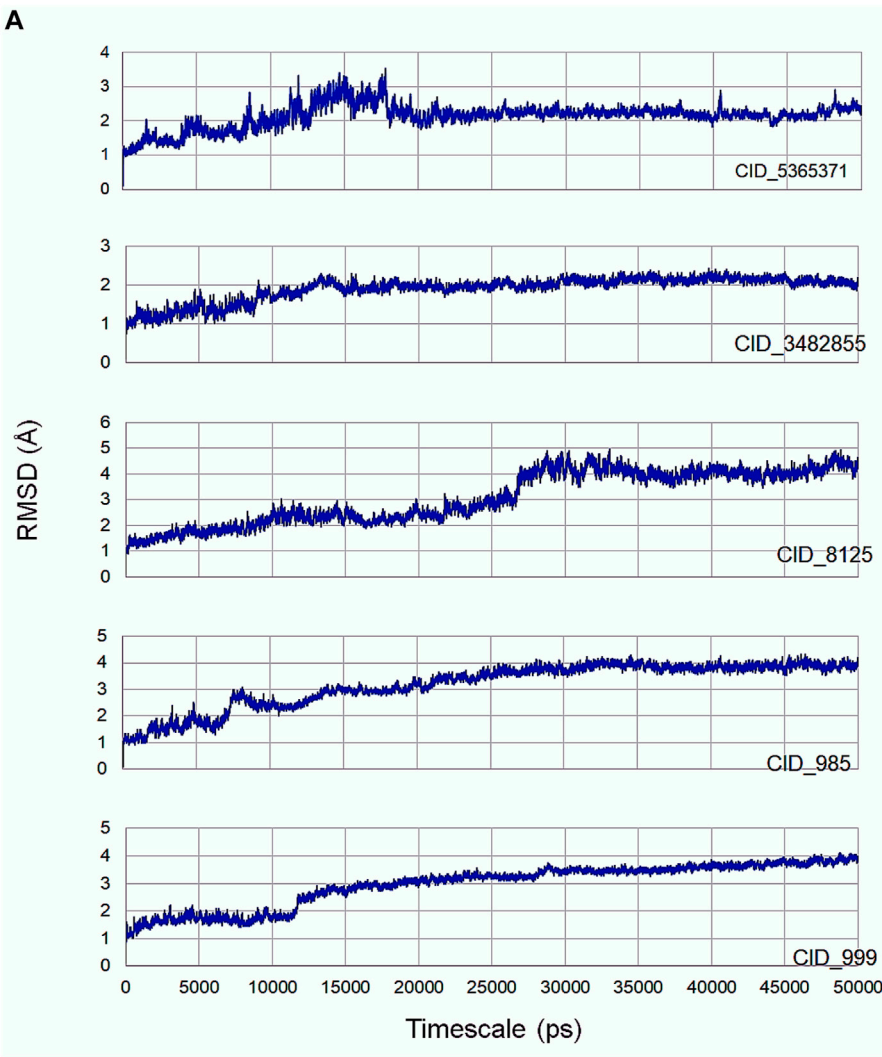


FIGURE 3 (continued).

similarity with 98% query coverage with human cell division control protein 42 homolog. The modeled structure was superimposed with the template structure and it is shown in

Figure 1. The overall quality factor obtained during ERRAT analysis is 77.5281. In Verify3D, around 80.65% of the amino acid residues have scored ≥ 0.2 in the 3D/1D profile.

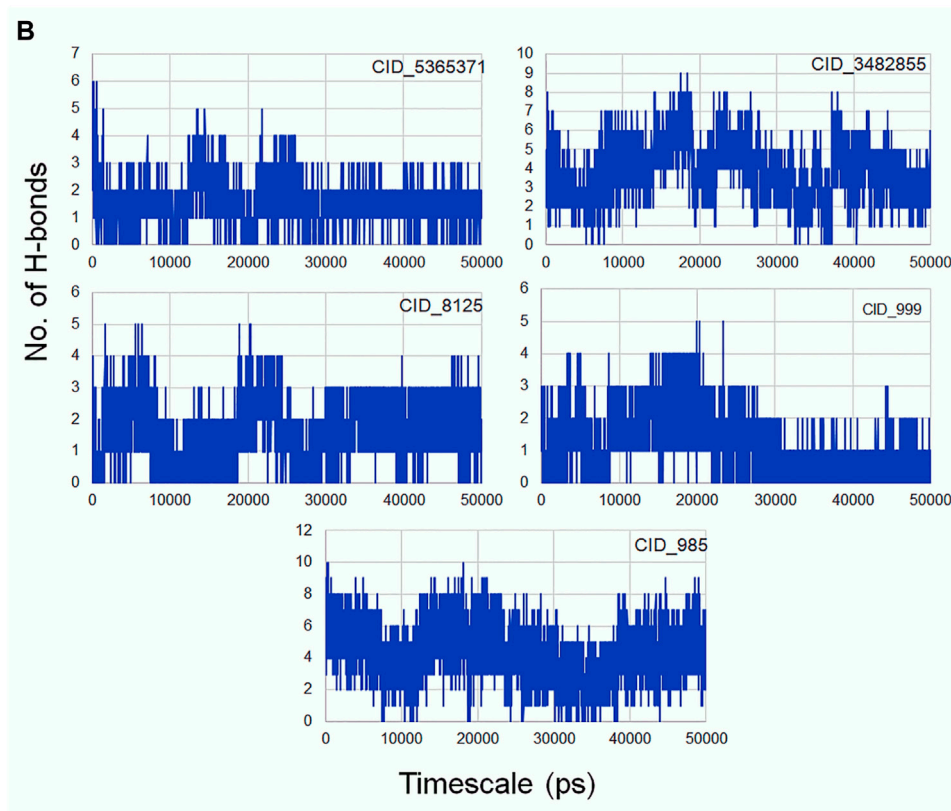


FIGURE 3

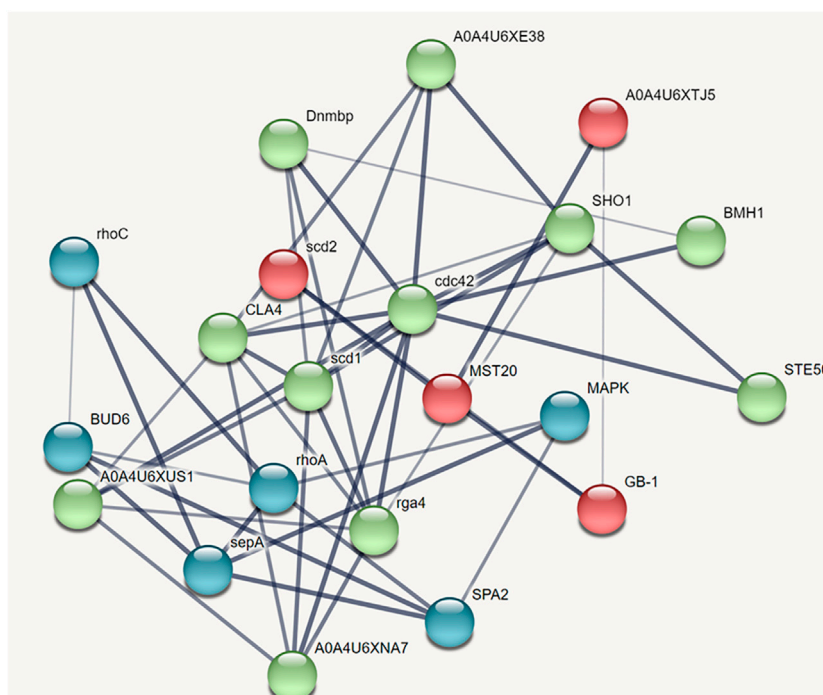
(A) Molecular dynamics simulation of CDC42 with ligands (β -sitosterol, phenylacetic acid, palmitic acid, behenic acid, and erucamide); (B) H-bond interaction of CDC42 and ligand molecules.

Ramachandran's Z-score was found to be -2.077 in the WHAT CHECK analysis. In Ramachandran plot analysis 88.7% amino acid residues were found in the most allowed regions. Around 10.7% residues were present in additionally allowed regions and 0.6% amino acid residues were found in generously allowed regions. The overall quality analyzed through ProSA Z-score displayed a -6.62 value for target CDC42 homology while the template displayed a Z-score value of -7.59 . This Z-score comparison between the target and the template suggests the resemblance in the geometry of the conformations between the target and template. Structural validation was shown in (Supplementary Figure S1). Altogether the structural verification suggested the consistency of the generated model.

Molecular docking

The molecular docking results for active compounds identified from *B. megaterium* were shown in Table 1. Inconsistent with previous reports from Xie et al. (2021), the results from the present study also showed β -sitosterol and

phenylacetic acid as the top hits in molecular docking. Through binding site prediction, it was observed that Leu158, Ser121, Thr117, Glu118, Ser88, Thr87, Ala15, Glu18, and Gly14 are the active residues of CDC42. Active site region is distributed with polar (Ser and Thr), hydrophobic residues (Leu and Ala) and negatively charged (Glu) residues. The presence of Ser residues in the active region are responsible for the interaction with the lead molecules. Presence of single Ser residues are responsible for enzymatic reaction. The binding site region consist of two Ser residues responsible for interaction of the lead molecules. Both the ligand, β -sitosterol and phenylacetic acid, presented the highest dock score of -10 kJ mol^{-1} . The next top hit obtained was palmitic acid (-9.4 kJ mol^{-1}) followed by behenic acid (-9.2 kJ mol^{-1}) and erucamide (-9.2 kJ mol^{-1}). The representative 2D and 3D images were presented in Figure 2. The CDC42 homolog protein with phenylacetic acid displayed van der Waals interactions with Gly17, Val16, Lys18, Thr37, Thr60, Ala61, Pro36, Val35, Tyr34; conventional hydrogen bond interactions with Gly62, Gln63; carbon-hydrogen bond and pi-donor hydrogen bond interactions with Ala15, Gly14; and pi-sigma bond interaction with Thr19. The CDC42 protein with β -

**FIGURE 4**

STRING network analysis displaying protein-protein interactions. Color nodes represents query proteins and first shell of interactions. Red color node represents cluster 1, green color represents cluster 2 and blue color represents cluster 3.

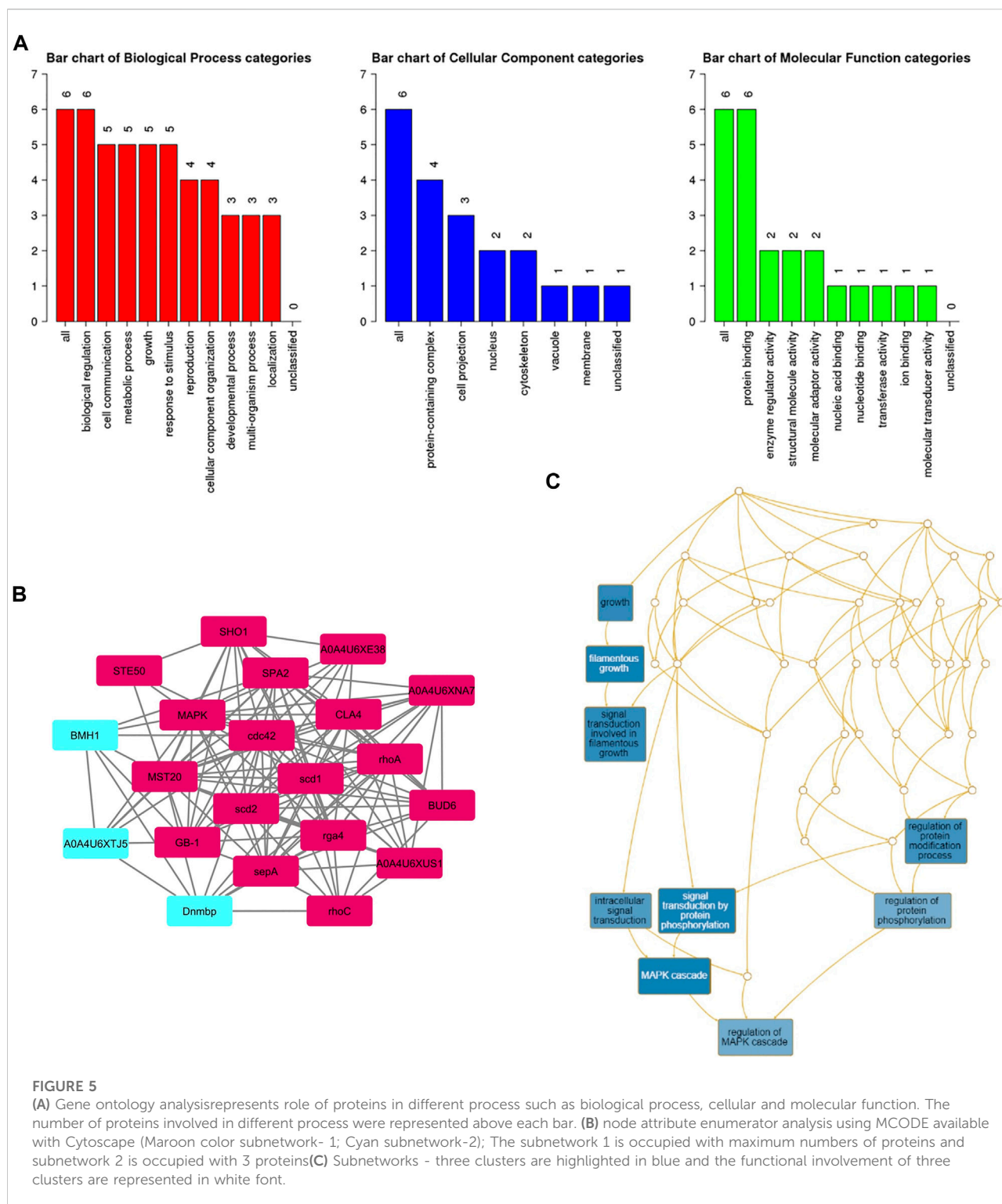
sitosterol displayed van der Waals interactions with Pro115, Ser121, Ser156, Thr117, Phe30, Val123, Glu118, Thr87, Ala15, and Gly14; conventional hydrogen bond interaction with Ser88; and alkyl bond interaction with Leu158. Palmitic acid displayed van der Waals interactions with Phe80, Pro105, Gly108, Arg101, Ser100, Gly148, Ala144; conventional hydrogen bond interactions with Ser104, Pro107; and alkyl bond interactions with Val109, Ala149, Met143, Leu111, Val82, Leu147, Val95. The CDC42 homolog with behenic acid displayed van der Waals interactions with Thr87, Glu118, Thr117, Asp13, Gly14, Asp59, Thr19; conventional hydrogen bond interactions with Ala15, Val16, Gly17, Lys18; alkyl and pi-alkyl bond interactions with Pro115, Ala157, Leu158, Tyr34, Phe30, Cys20. Erucamide showed van der Waals interactions with Tyr34, Ala15, Gly17, Val16, Thr117, Glu118, Val123, Ser85; (Table 2) conventional hydrogen bond interaction with Thr87; alkyl and pi-alkyl bond interactions with Cys20, Phe30, Pro115, Ala157, Leu158. From the free energy calculation, it was observed that β -sitosterol obtained the highest binding energy of -42.35 (Kcal/mol) compared to the other molecules used for docking. Palmitic acid with an energy of -41.51 (Kcal/mol) was observed as second highest compound showing highest binding affinity. The other compounds such as phenylacetic acid, Erucamide, and Behenic acid was observed with -39.26, -39.49 and -37.90 was observed with binding energy respectively.

ADMET

All the compounds were predicted positive for intestinal absorption and blood brain barrier. Also, from the predicted results it was observed that the compounds were non AMES toxic and non-carcinogenic. Hence the predicted compounds were determined non-toxic and can be used extensively for further studies. Also, β -sitosterol was previously predicted as FDA approved drug with no side-effects (Babu and Jayaraman, 2020). Based on the pharmacokinetic properties, the molecules were predicted to be lead molecules (Table 3).

Molecular dynamics simulation

Molecular dynamics simulation is an efficacious method for validating the stability of the ligands (β -sitosterol, phenylacetic acid, palmitic acid, behenic acid, and erucamide) docked into the binding pocket of CDC42. For this, all-atom molecular dynamics (MD) simulation study was applied which is regarded as a valuable approach to study the dynamic behavior of the ligands and proteins along with their key interacting residues. Thus, the obtained protein-ligand complexes through molecular docking were enrolled for 50 ns of all-atom MD simulation. MD simulation results revealed the protein-ligands exhibited



successful conversion of the initial start of run from 15 ns (Figure 3). The trajectories analysis of the MD run has shown the rise of initial frames at an average of 15 ns. However, the RMSD level of the trajectories proceeded with the average values

with minimal fluctuation until 20 ns. The ligand β -sitosterol showing interaction with CDC42 during MD simulation has exhibited an initial rise of the frames from 10 to 15 ns. The standard plateau throughout the MD simulation interval was

TABLE 4 KEGG enrichment analysis for enriched gene set.

GeneSet	Description	Size	Overlap	Expect	Enrichment ratio	pValue	FDR	Gene symbol
sce04011	MAPK signaling pathway	114	5	0.109986	45.46053	1.12E-08	4.02E-05	BMH1; CLA4; SHO1; SPA2; STE50
GO: 0030447	filamentous growth	135	5	0.130246	38.38889	2.64E-08	4.02E-05	BMH1; BUD6; SHO1; SPA2; STE50
GO: 0000165	MAPK cascade	42	4	0.040521	98.71429	2.67E-08	4.02E-05	CLA4; SHO1; SPA2; STE50
GO: 0023014	signal transduction by protein phosphorylation	47	4	0.045345	88.21277	4.25E-08	4.80E-05	CLA4; SHO1; SPA2; STE50
GO: 0040007	growth	177	5	0.170767	29.27966	1.04E-07	9.36E-05	BMH1; BUD6; SHO1; SPA2; STE50
GO: 0001402	signal transduction involved in filamentous growth	13	3	0.012542	239.1923	1.42E-07	1.07E-04	BMH1; SHO1; STE50
GO: 0031399	regulation of protein modification process	212	5	0.204534	24.44575	2.56E-07	1.66E-04	BMH1; CLA4; SHO1; SPA2; STE50
GO: 0035556	intracellular signal transduction	252	5	0.243126	20.56548	6.10E-07	3.45E-04	BMH1; CLA4; SHO1; SPA2; STE50
GO: 0043408	regulation of MAPK cascade	32	3	0.030873	97.17188	2.45E-06	0.00123	CLA4; SHO1; SPA2
GO: 0001932	regulation of protein phosphorylation	133	4	0.128316	31.17293	2.90E-06	0.001312	CLA4; SHO1; SPA2; STE50

observed from 20 to 50 ns. The average RMSD for β -sitosterol was observed as 2.10 ± 0.20 Å. This dynamic behavior confers a more stabilized accommodation of β -sitosterol into the binding pocket of the CDC42 throughout the MD simulation. The average RMSD values for phenylacetic acid, throughout the plateau MD simulation interval (12–28 ns) was 3.1 ± 0.50 Å. However, the plateau showed a rise in level after 30 ns and remained stable until 50 ns with an average RMSD of 3.5 ± 0.30 Å. Palmitic acid showed a higher shift in trajectory frames with an RMSD value rise from 3.1 ± 0.50 Å to 4.3 ± 0.10 Å after 25 ns. However, both the ligands phenylacetic acid and palmitic acid have converged around the comparable trajectory frames with an average RMSD value of approximately above 3.5 Å. This dynamic difference between palmitic and phenylacetic acid has shown that the ligands might have deviated from the original interaction compared to the β -sitosterol ligand complex. The other ligands such as behenic acid and eurcamide have shown differential dynamic behavior, which confers the ligands shift from the binding pocket. Eurcamide initial rise in the trajectory level at 15–20 ns was about 3.1 ± 0.10 Å. Later, the plateau was depicted as steady until the end of the simulation around 50 ns. Similarly, behenic acid has shown a rise in the level of the plateau after 28 ns, which confirms the significant ligand, shifts out of the CDC2 pocket, and remained stable till the end of the MD run. Thus, the overall analysis of the MD simulation run suggests the ligand β -sitosterol has stable conformation into the binding pocket of CDC42. Comparatively, phenylacetic acid and palmitic acid have also been found to be stable. Eurcamide and behenic acid ligands were observed to alter their position in

their binding pocket of CDC42. The hydrogen bond formation plays a significant role during molecular interaction between the ligand and protein (CDC42). The term hydrogen bond donor and acceptor during hydrogen bonding indicate hydrogen atom from the donor and the acceptor with lone pair of electrons. From the MD simulation run, it was observed that β -sitosterol and palmitic acid shared a maximum of eight hydrogen atoms throughout the run. Phenylacetic acid has shared a maximum of four hydrogen bonds throughout the simulation run. Overall, the number of hydrogen bond donors, as well as acceptors, are within the range for β -sitosterol and palmitic acid (Figure 3B).

STRING analysis

The protein-protein interaction network predicted through STRING analysis is shown in Figure 4. The network comprises 21 nodes; 133 edges; 12.7 average degree nodes; 0.812 average local clustering coefficient; 31 expected edges with protein-protein interaction enrichment values less than $1.0e^{-16}$. In Figure 4, color nodes represent the query protein and first shell of interactors while the white nodes represent the second shell of interactors. The empty nodes indicate proteins with an unknown 3D structure. The network edges represent the confidence mode in which the thickness of the line indicates the strength of the data support. From Figure 4, it is clear that there is no 3D structure available for the first and second shell interactors. Hence, further studies are required to understand the complex mechanism and detailed functions of the

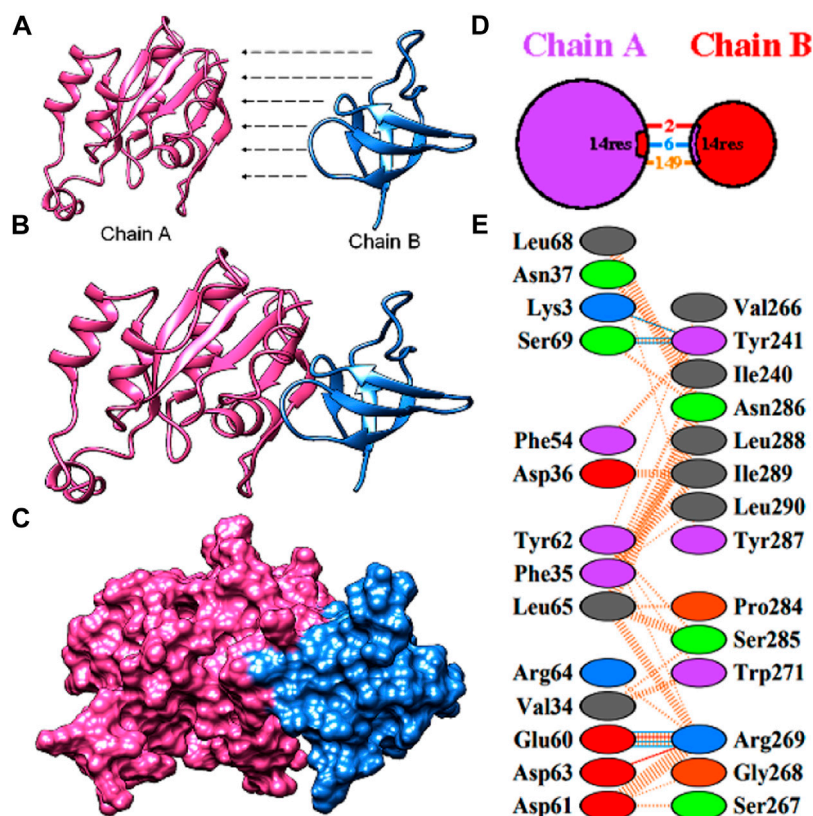


FIGURE 6

Protein-protein docking of the CDC42 with SHO1 from *C. gloeosporioides* (A) model proteins of CDC42 (pink) and SHO1 (blue). (B) and (C) interaction of the model proteins (D) pictorial representation of the interaction model and number of interactions (E) residues involved in interaction.

CDC42 homolog in establishing pathogenicity and diseases in plants.

Clusters identification through MCODE analysis and GO classification

A subnetwork was constructed and the result was visualized through Cytoscape. CDC42 has shown interaction with SHO1, STE50, SPEA2, A0A4U6XE38, MAPK, CLA4, A0A4U6XNA7, MST20, SCD2, SCD1, RHOA, GB-1, SEPA, RHOC, A0A4U6XUS1, BUD6, RGA4, BMH1, A0A4U6XTJ5, and DNMBP. Further functional enrichment and Gene ontology analysis performed through the WEB-based GENESeT AnaLysis Toolkit depicting three fundamental categories were presented in Figure 5A. The three main categories are biological process, cellular components, and molecular function. MCODE provides a real-time cluster assessment quality. The node attribute enumerator provides a numerical summary of node attribute values as shown in Figure 5B. Node attribute that is available for the loaded network is shown in box-1 which

contains 15 nodes and 94 edges. The members of the clusters are represented in red. Exploration of clusters is shown in box-2 which contains 3 nodes and 3 edges. The members of the clusters are represented in red. The node scoring the highest value in the cluster is referred to as the seed. It is the node from which the cluster was derived, and it is represented in squares, and other cluster members are represented in circles. Edges indicating the interactions are represented in blue while the edge directionality is represented by arrows. New sub-clusters formed from the main cluster is shown in Figures 5C,D. The GO Slim summary is based on 6 unique Entrez gene IDs including BMH1, CLA4, SHO1, SPA2, STE50, and CDC42. Among 6 unique Entrez gene IDs, 6 IDs are annotated to the selected functional categories and also in the reference list, which are used for the enrichment analysis. All the genes mentioned above are predicted to play an important role in the MAP kinase pathway. The enrichment analysis revealed that the gene is mainly involved in filamentous growth, signal transduction (by protein phosphorylation), and MAPK cascade. Altogether, the KEGG enrichment analysis revealed the association of the MAPK signaling pathway. The enriched gene set for the MAPK signaling pathway was found to

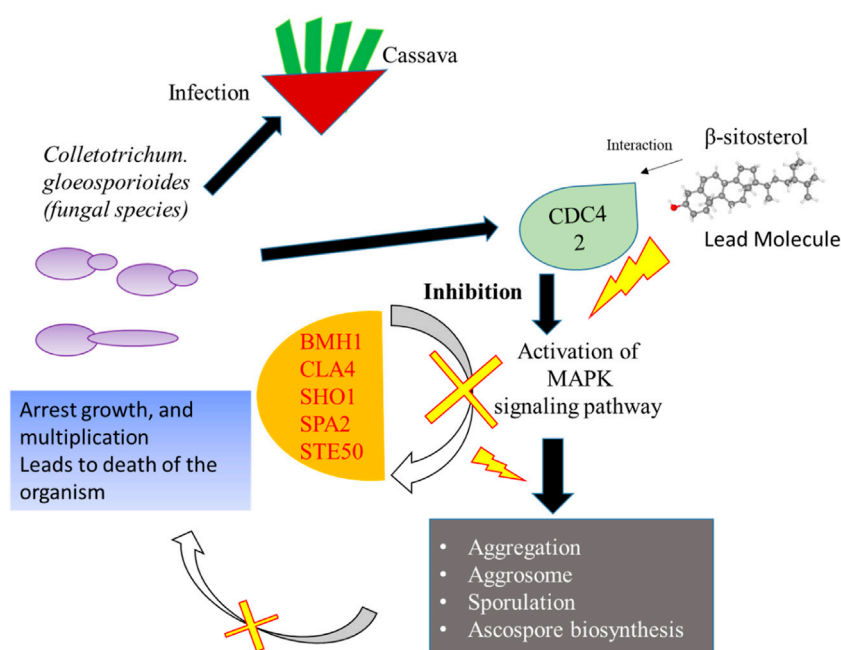


FIGURE 7

Details of the overall mechanism representing CDC42 from *C. gloeosporioides* involved infection and inhibition of the signaling pathway using lead molecules.

have a p -value of 1.1214×10^{-8} ; an enrichment ratio of 45.46. The false discovery rate for the network was predicted as 0.00004 for a gene set size of 114 (Table 4). The protein sequence of SHO1 of *C. gloeosporioides* with the accession id A0A1B2LQ50 was obtained from Uniprot database. The sequence alignment of target and template sequence (yeast SHO1) with the sequence coverage of 61.02% was used for modelling the protein. Interaction analysis of CDC42 and SHO protein has shown 14 residues from each protein have good binding affinity. H-bond (6), non-bonded 149 and 2 salt bridges were established between the drug target (CDC42) and SHO1 (Figure 6).

Discussion

Anthrax disease occurrence in cassava can lead to total economic loss for the cultivators by damaging the total harvest into the rotting waste. Anthracnose disease occurs in plants due to fungal species of the genus *Colletotrichum*. The species such as *C. fructicola*, *C. gloeosporioides*, *C. tropicale*, *C. theobromicola*, *C. siamense*, *C. brevisporum*, and *C. plurivorum* are the most common group of plant pathogens that are responsible for diseases on many plant species. Infected plants with fungal strains develop dark patches and lesions on stems, leaves, or any parts of the plant. The lesions occurring on the infected region (leaves, stem) appear to be the gelatinous masses of

spores. The fungi during infection come in close contact with the adherence of the spores. The germination starts after several hours with favorable conditions such as temperature. During the suitable temperature, the fungi germinate the conidia and produce the germ tubes. Fungi develop appressorium and arrest the elongation of the germ tube. The penetration of appressorium promotes turgor pressure and fungi colonize the plant tissues, which appear like a canker.

The role of genes in the penetration and development of infection has not been revealed so far in cassava. However, mitogen activator (PMK1, MPS1), ATPase (PDE1), Tetraspanins (PLS1), and fungal effector genes were reported as important genes for infection in rice blast fungus. CDC42, an important protein essential for cell division and cell cycle from cassava, was investigated as the molecular target for the present study. From the sequence analysis, CDC42 was revealed with 190 amino acids and belongs to the small GTPase family. In fungi, the presence of the small GTPase is essential for both beneficial and pathogenic interaction with the plant system. The small GTPases of cassava are characterized as Rho family, essential for the formation of virulence, a fusion of pathogen with plant cell, and production of reactive oxygen species (ROS). Thus, the protein with three motifs is structurally essential for the GTPase activity. In absence of the three-dimensional structure, the protein was modeled based on sequence similarity. The CDC42 protein from *C. gloeosporioides* has shown 70.83% homology with CDC42 of humans with query

coverage of 98%. The modeled structure was validated and used for further docking studies. Generally, *Bacillus* sp. is considered a promising source for bioactive secondary metabolites. Therefore, in the present study molecular docking was carried out with the bioactive compounds (erucamide, behenic acid, palmitic acid, phenylacetic acid, and β -sitosterol) of *B. megaterium*. There are several pieces of evidence for the compounds identified from *B. megaterium* as lead molecules. Erucamide identified from radish leaf has shown preventive effects against memory deficits related to Alzheimer's disease by modulation of cholinergic function. *In vivo* experiments have shown that erucamide has biological activity such as stimulation of angiogenesis, augmentation of neovascularization in regenerating skeletal muscle, and anti-depressive effects (Kim et al., 2018). Similarly, behenic acid-based nano micelles were prepared with dextran as the combination to deliver itraconazole as a drug. The nano micelles were used as anti-leishmaniasis for targeting the parasite (Shahriyar et al., 2021). The saturated fatty acid (C:16), the palmitic acid, selected as the lead molecule in the present study has been deeply investigated previously (Lee et al., 2009) as an antiviral agent. Palmitic acid specifically binds to the CD4 and prevents the entry of the HIV-1 virus. Moreover, β -sitosterol has been reported from plants and is known for anticancer effects against several cancers such as breast and ovary, prostate, lung, stomach, and liver. In addition, the compound can significantly inhibit several pathways, cell signaling, apoptosis, angiogenesis, and inflammation (Bin Sayeed and Ameen, 2015). Phenylacetic acid and its derivatives were extensively used in the preparation of drugs that can be used for several ailments. Diclofenac is used as a medication for the treatment of pain and inflammation (Gan 2010). Apart, the previous report indicated that the purified components possess significant antimicrobial activity against plant pathogens such as *A. tumefaciens* (T-37), *Erwinia carotovora* (EC-1), and *Ralstonia solanacearum* (RS-2) (Xie et al., 2021). Among the five components investigated against plant pathogens, β -sitosterol, behenic acid, and phenylacetic acid displayed significant antimicrobial activity. β -sitosterol showed a very low minimum inhibitory concentration (15.6 μ g/ml) against *R. solanacearum* RS-2 (Xie et al., 2021). Thus, the five compounds of choice used for the present study have already been investigated for several alignments. Mostly, these compounds were reported from different plant species; however, in the present study, the compounds identified from *B. megaterium* were used for investigation. The compounds were docked into the binding pocket of the CDC42 of *C. gloeosporioides*. β -sitosterol and phenylacetic acid showed the highest dock score of -10 kJ mol $^{-1}$. The next top hit obtained was palmitic acid (-9.4 kJ mol $^{-1}$) followed by behenic acid (-9.2 kJ mol $^{-1}$) and erucamide (-9.2 kJ mol $^{-1}$). From the binding energy analysis, it was observed that β -sitosterol obtained the highest binding affinity of -41.51 (Kcal/mol).

To study the stability during their interaction, molecular simulation was performed and the results showed that the compounds were stable throughout the simulation. For a

comprehensive analysis of the docked protein-ligand complex, a molecular dynamics simulation was carried out. It is the most powerful technique to study the conformational changes taking place at the atom level. Therefore, molecular dynamics simulation was performed for some time of 50 ns for all atom-docked protein-ligand complexes. The result showed β -sitosterol with stable conformation compared to the other docked complexes. The results are evidence that the docked protein- β -sitosterol complexes are highly stable for the entire period of 20–50 ns. Furthermore, RMSD plot analysis showed slight modification in the position, indicating the stable association and interactions between the ligand and the protein molecule. Also, the ligand has maintained a maximum of eight hydrogen bond interactions throughout the MD run. Thus, the stability of the ligand with CDC42 shows the ligands can be extended further as a biological agent to treat pathogenesis against *C. gloeosporioides*. Additionally, studies have shown that β -sitosterol has already been used for the treatment of various diseases due to its potent properties such as antinociceptive, anxiolytic and sedative effect, anticancer, antimicrobial, immunomodulatory, hepatoprotective, and wound healing effects. The chemical has already been approved by FDA and is a safe nutraceutical with no deleterious effects (Babu and Jayaraman, 2020).

Network-based approaches provide a deep insight to understand the biological process during the pathogenesis of *C. gloeosporioides*. The interacting partner revealed through the PPI network will pave the way to investigate the cellular activity, protein localization, and complex biological function of the protein. Besides, 20 genes have shown interaction with CDC42 and from the MCODE statistical analysis, two clusters were identified one with 15 nodes and 94 edges and the second cluster with 3 nodes and 3 edges. Furthermore, the functional enrichment analysis has revealed the BMH1, CLA4, SHO1, SPA2, and STE50 as the important genes involved in the MAPK signaling pathway. The protein BMH1 has shown to play important role in aggregation, and arrangement to form aggresomes. Additionally, BMH1 is involved in spore formation, sporulation, and ascospore biosynthesis. The CLA4 is very essential for imparting *Cladosporium* resistance in the organism. SHO1 protein, the osmosensor present in the plasma membrane of *C. gloeosporioides* activates the high osmolality glycerol (HOG) of the MAPK signaling pathway in response to high osmolality. SHO1 protein is found in bud and bud neck region of the fungal pathogen. Protein docking interaction reveals SHO1 and CDC42 has established binding affinity. Hence it is envisaged that inhibition of CDC42 can significantly prevent the signalling and inhibit the growth and development of the fungal pathogen. SPA2 perhaps a cytoskeletal protein is involved in pheromone-induced morphogenesis, budding, invasive filamentation growth, regulation of hyphal growth, cellular shape, and reproduction of *C. gloeosporioides*. STE50 protein has shown to play a significant role in signal transduction during filamentous growth, osmosensory signaling MAPK cascade thereby arrest the growth during conjugation.

Thus, the identified interacting partners of CDC42 are involved in the MAPK signaling pathway essential for growth, and virulence regulation in *C. gloeosporioides*. Therefore, from the present study, it is revealed that targeting CDC42 can impart the interaction network, prevent filament production, and arrest the reproduction in *C. gloeosporioides* (Figure 7).

Conclusion

Infections caused by *C. gloeosporioides* in cassava are very serious to impair, leading to economic damage to the cultivators. To date, there are no clear details about the type of infection and mode of transmission, and pathogenesis of the fungal pathogen. Hence, in the present study, CDC42 protein involved in cell division and cell cycle of the pathogen was selected as the target. Modeling of the protein revealed the key residues playing the functional role of the protein. The protein was targeted with five active compounds from *B. megaterium*. Interaction of β -sitosterol and phenylacetic acid with the key residue of CDC42 revealed that ligands could have a potential role in the inhibition of functional proteins that are involved in growth. Further PPI network constructed for CDC42 revealed that targeting the protein could impart MAPK signaling pathway. Additionally, targeting the interacting partner could also prevent the growth, filamentation, and hyphal growth which is essential for virulence regulation. However, further experimental insight can pave a way for preventing *C. gloeosporioides* mediated infection in cassava.

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

NP and NB, designed the experiment. NP, KM, and BK implemented and analyzed the data. WT, RS, CS, and NH, analysis the results, discussion, VM, TL, and MK, prepared the manuscript. NP, TL, and NB revised the manuscript.

References

- Adewole, K. E., and Ishola, A. A. (2021). A computational approach to investigate the HDAC6 and HDAC10 binding propensity of psidium guajava-derived compounds as potential anticancer agents. *Curr. Drug Discov. Technol.* 18, 423–436. doi:10.2174/1568009620666200502013657
- Agrawal, S., Jana, U. K., and Kango, N. (2021). Heterologous expression and molecular modelling of L-asparaginase from *Bacillus subtilis* ETMC-2. *Int. J. Biol. Macromol.* 1 (192), 28–37. doi:10.1016/j.ijbiomac.2021.09.186

Acknowledgments

The authors are thankful to Suranaree University of Technology, Thailand for providing financial support. This research received research funding support from the SUT research and development fund (Project code: IRD 3-302-65-12-20). The authors are also thankful to Dr. B.R. Ambedkar Open University (Hyderabad), Sri Yuva Biotech Pvt. Ltd (Hyderabad), and Mahatma Gandhi University (Nalgonda), India for their help and support in this work.

Conflict of interest

BK was employed by the company Sri Yuva Biotech Pvt Ltd.

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.

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.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmolb.2022.1010603/full#supplementary-material>

SUPPLEMENTARY FIGURE 1

Structural validation through Saves and ProSA program (A) Verify3D (B) Ramachandran plot for model obtained from swiss model server and modeller (C) ProSA score for the template, 2NGR (D) ProSA score for target, CDC42.

SUPPLEMENTARY FIGURE 2

superimposition of model from Swissmodel (blue) and modeler (pink).

- Atwijukire, E., Hawumba, J. F., Baguma, Y., Wembabazi, E., Esuma, W., Kawuki, R. S., et al. (2019). Starch quality traits of improved provitamin A cassava (*Manihot esculenta* Crantz). *Heliyon* 5, e01215. doi:10.1016/j.heliyon.2019.e01215

- Ayling, S., Ferguson, M., Rounsley, S., and Kulakow, P. (2012). Information resources for cassava research and breeding. *Trop. Plant Biol.* 5, 140–151. doi:10.1007/s12042-012-9093-x

- Azam, S. S., and Abbasi, S. W. (2013). Molecular docking studies for the identification of novel melatoninergic inhibitors for acetylserotonin-O-methyltransferase using different docking routines. *Theor. Biol. Med. Model.* 10, 63. doi:10.1186/1742-4682-10-63
- Babu, S., and Jayaraman, S. (2020). An update on β -sitosterol: A potential herbal nutraceutical for diabetic management. *Biomed. Pharmacother.* 131, 110702. doi:10.1016/j.biopha.2020.110702
- Balyejusa Kizito, E., Rönnberg-Wästljung, A. C., Egwang, T., Gullberg, U., Fregene, M., and Westerbergh, A. (2007). Quantitative trait loci controlling cyanogenic glucoside and dry matter content in cassava (*Manihot esculenta* Crantz) roots. *Hereditas* 144, 129–136. doi:10.1111/j.2007.0018-0661.01975.x
- 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
- Bin Sayeed, M. S., and Ameen, S. S. (2015). Beta-sitosterol: A promising but orphan nutraceutical to fight against cancer. *Nutr. Cancer* 67, 1214–1220. doi:10.1080/01635581.2015.1087042
- El-Sharkawy, M. A. (2004). Cassava biology and physiology. *Plant Mol. Biol.* 56, 481–501. doi:10.1007/s11103-005-2270-7
- Gan, T. J. (2010). Diclofenac: An update on its mechanism of action and safety profile. *Curr. Med. Res. Opin.* 26, 1715–1731. doi:10.1185/03007995.2010.486301
- Gong, Z., Xiong, L., Shi, H., Yang, S., Herrera-Estrella, L. R., Xu, G., et al. (2020). Plant abiotic stress response and nutrient use efficiency. *Sci. China. Life Sci.* 63, 635–674. doi:10.1007/s11427-020-1683-x
- Hasan, R., Rony, M. N. H., and Ahmed, R. (2021). *In silico* characterization and structural modeling of bacterial metalloprotease of family M4. *Plant Mol. Biol.* 2 (191), 25. doi:10.1186/s43141-020-00105-y
- Hess, B. (2008). P-LINCS: A parallel linear constraint solver for molecular simulation. *J. Chem. Theory Comput.* 4, 116–122. doi:10.1021/ct700200b
- Hormhuan, P., Viboonjun, U., Sojikul, P., and Narangajavana, J. (2020). Enhancing of anthracnose disease resistance indicates a potential role of antimicrobial peptide genes in cassava. *Genetica* 148, 135–148. doi:10.1007/s10709-020-00097-0 <https://swissmodel.expasy.org/> https://string-db.org/cgi/input?sessionId=b7c9uFOprzGmandinput_page_active_form=single_sequence
- Jacobs, R. L., Adhikari, T. B., Pattison, J., Yencho, G. C., Fernandez, G. E., and Louws, F. J. (2019). Inheritance of resistance to *Colletotrichum gloeosporioides* and *C. acutatum* in strawberry. *Phytopathology* 109, 428–435. doi:10.1094/PHYTO-08-18-0283-R
- Kim, C. R., Kim, H. S., Choi, S. J., Kim, J. K., Gim, M. C., Kim, Y. J., et al. (2018). Erucamide from radish leaves has an inhibitory effect against acetylcholinesterase and prevents memory deficit induced by trimethyltin. *J. Med. Food* 21, 769–776. doi:10.1089/jmf.2017.4117
- Koehorst-van Putten, H. J., Sudarmonowati, E., Herman, M., Pereira-Bertram, I. J., Wolters, A. M., Meima, H., et al. (2012). Field testing and exploitation of genetically modified cassava with low-amylose or amylose-free starch in Indonesia. *Transgenic Res.* 21, 39–50. doi:10.1007/s11248-011-9507-9
- Laskowski, R. A., Rullmann, J. A., MacArthur, M. W., Kaptein, R., and Thornton, J. M. (1996). AQUA and PROCHECK-NMR: Programs for checking the quality of protein structures solved by NMR. *J. Biomol. NMR* 8, 477–486. doi:10.1007/BF00228148
- Lee, D. Y., Lin, X., Paskaleva, E. E., Liu, Y., Puttamadappa, S. S., Thornber, C., et al. (2009). Palmitic acid is a novel CD4 fusion inhibitor that blocks HIV entry and infection. *AIDS Res. Hum. Retroviruses* 25, 1231–1241. doi:10.1089/aid.2009.0019
- Li, M., Liu, J., and Zhou, G. (2021). Histopathological and ultrastructural observations of *Camellia oleifera* infected with *Colletotrichum fructicola*. *Australas. Plant Pathol.* 50, 523–531. doi:10.1007/s13313-021-00811-2
- Li, S., Cui, Y., Zhou, Y., Luo, Z., Liu, J., and Zhao, M. (2017). The industrial applications of cassava: Current status, opportunities and prospects. *J. Sci. Food Agric.* 97, 2282–2290. doi:10.1002/jsfa.8287
- Machado, S., Silva Veloso, J., Camara, M., Campos, F. S., Sarmiento, R. A., Giongo, M. V., et al. (2020). First report of *Colletotrichum chrysophilum* causing cassava anthracnose in Brazil. *Plant. Dis.* 8, 1–8. doi:10.1094/PDIS-09-20-1925-PDN10.1094/PDIS-09-20-1925-PDN
- Miyamoto, S., and Kollman, P. A. (1992). Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models. *J. Comput. Chem.* 13, 952–962. doi:10.1002/jcc.540130805
- Oeser, B., Kind, S., Schurack, S., Schmutzer, T., Tudzynski, P., and Hinsch, J. (2017). Cross-talk of the biotrophic pathogen *Claviceps purpurea* and its host *Secale cereale*. *BMC Genomics* 18, 273. doi:10.1186/s12864-017-3619-4
- Ons, L., Bylemans, D., Thevissen, K., and Cammue, B. (2020). Combining biocontrol agents with chemical fungicides for integrated plant fungal disease control. *Microorganisms* 8, 1930. doi:10.3390/microorganisms8121930
- Peck, S., and Mittler, R. (2020). Plant signaling in biotic and abiotic stress. *J. Exp. Bot.* 71, 1649–1651. doi:10.1093/jxb/eraa051
- Pinweha, N., Asvarak, T., Viboonjun, U., and Narangajavana, J. (2015). Involvement of miR160/miR393 and their targets in cassava responses to anthracnose disease. *J. Plant Physiol.* 174, 26–35. doi:10.1016/j.jplph.2014.09.006
- Sekhar Pagadala, N., Arha, M., Reddy, P. S., Kumar, R., Sirisha, V. L., Prashant, S., et al. (2009). Phylogenetic analysis, homology modelling, molecular dynamics and docking studies of caffeoyl-CoA-O-methyl transferase (CCoAOMT 1 and 2) isoforms isolated from subabul (*Leucaena leucocephala*). *J. Mol. Model.* 15, 203–221. doi:10.1007/s00894-008-0395-8
- Shahriyar, S., Taymouri, S., Saberi, S., Asadi, P., and Tabbakhian, M. (2021). Preparation and characterization of itraconazole loaded nanomicelles based on dextran-behenic acid for cutaneous leishmaniasis treatment. *Drug Dev. Ind. Pharm.* 47, 416–428. doi:10.1080/03639045.2021.1890112
- Studio, D. (2015). *4.1 visualizer*. San Diego, USA: Accelrys Software Inc.
- Tripathi, N., Danger, R., Chesneau, M., Brouard, S., and Laurent, A. D. (2022). Structural insights into the catalytic mechanism of granzyme B upon substrate and inhibitor binding. *J. Mol. Graph. Model.* 114, 108167. doi:10.1016/j.jmgm.2022.108167
- Trott, O., and Olson, A. J. (2010). AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* 31, 455–461. doi:10.1002/jcc.21334
- Van Der Spoel, D., Lindahl, E., Hess, B., Groenhof, G., Mark, A. E., and Berendsen, H. J. (2005). Gromacs: Fast, flexible, and free. *J. Comput. Chem.* 26, 1701–1718. doi:10.1002/jcc.20291
- Wang, X., Lu, D., and Tian, C. (2021). Analysis of melanin biosynthesis in the plant pathogenic fungus *Colletotrichum gloeosporioides*. *Fungal Biol.* 125, 679–692. doi:10.1016/j.funbio.2021.04.004
- Wang, X., Xu, X., Liang, Y., Wang, Y., and Tian, C. (2018). A CDC42 homolog in *Colletotrichum gloeosporioides* regulates morphological development and is required for ROS-mediated plant infection. *Curr. Genet.* 64, 1153–1169. doi:10.1007/s00294-018-0833-9
- 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
- Xie, Y., Peng, Q., Ji, Y., Xie, A., Yang, L., Mu, S., et al. (2021). Isolation and identification of antibacterial bioactive compounds from *Bacillus megaterium* L2. *Front. Microbiol.* 12, 645484. doi:10.3389/fmicb.2021.645484
- Yousafi, Q., Amin, H., Bibi, S., Rafi, R., Khan, M. S., Ali, H., et al. (2021). Subtractive proteomics and immuno-informatics approaches for multi-peptide vaccine prediction against *Klebsiella oxytoca* and validation through in silico expression. *Int. J. Pept. Res. Ther.* 27, 4 2685–2701. doi:10.1007/s10989-021-10283-z
- Zheng, W., Zhao, Z., Chen, J., Liu, W., Ke, H., Zhou, J., et al. (2009). A CDC42 ortholog is required for penetration and virulence of *Magnaporthe grisea*. *Fungal Genet. Biol.* 46, 450–460. doi:10.1016/j.fgb.2009.03.005



OPEN ACCESS

EDITED BY
Chandrabose Selvaraj,
Alagappa University, India

REVIEWED BY
Sunil Kumar Tripathi,
Georgia State University, United States
Anupam Nath Jha,
Tezpur University, India

*CORRESPONDENCE
Sangwook Wu,
s.wu@pharmcadd.com

SPECIALTY SECTION
This article was submitted to Biophysics,
a section of the journal
Frontiers in Molecular Biosciences

RECEIVED 25 July 2022
ACCEPTED 15 September 2022
PUBLISHED 11 October 2022

CITATION
Mudedla SK, Braka A and Wu S (2022),
Quantum-based machine learning and
AI models to generate force field
parameters for drug-like
small molecules.
Front. Mol. Biosci. 9:1002535.
doi: 10.3389/fmolb.2022.1002535

COPYRIGHT
© 2022 Mudedla, Braka and Wu. 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.

Quantum-based machine learning and AI models to generate force field parameters for drug-like small molecules

Sathish Kumar Mudedla¹, Abdenmour Braka¹ and
Sangwook Wu^{1,2*}

¹R&D Center, PharmCADD, Busan, South Korea, ²Department of Physics, Pukyong National University, Busan, South Korea

Force fields for drug-like small molecules play an essential role in molecular dynamics simulations and binding free energy calculations. In particular, the accurate generation of partial charges on small molecules is critical to understanding the interactions between proteins and drug-like molecules. However, it is a time-consuming process. Thus, we generated a force field for small molecules and employed a machine learning (ML) model to rapidly predict partial charges on molecules in less than a minute of time. We performed density functional theory (DFT) calculation for 31770 small molecules that covered the chemical space of drug-like molecules. The partial charges for the atoms in a molecule were predicted using an ML model trained on DFT-based atomic charges. The predicted values were comparable to the charges obtained from DFT calculations. The ML model showed high accuracy in the prediction of atomic charges for external test data sets. We also developed neural network (NN) models to assign atom types, phase angles and periodicities. All the models performed with high accuracy on test data sets. Our code calculated all the descriptors that were needed for the prediction of force field parameters and produced topologies for small molecules by combining results from ML and NN models. To assess the accuracy of the predicted force field parameters, we calculated solvation free energies for small molecules, and the results were in close agreement with experimental free energies. The AI-generated force field was effective in the fast and accurate generation of partial charges and other force field parameters for small drug-like molecules.

KEYWORDS

partial charge prediction, AI force field, atomtype prediction, protein ligand binding, free energy, atomic charges

1 Introduction

Molecular dynamics (MD) simulations play an important role in rational drug design, (Marco De, et al., 2016), which is useful in the analysis of dynamical interactions between a target protein and drug molecules (Allinger, 1977; Lifson, et al., 1979; Burkert and Allinger, 1982; Brooks, et al., 1983; Jorgensen and Tirado-Rives, 1988; Allinger, et al., 1989; Clark, et al., 1989; Mayo, et al., 1990; Momany and Rone, 1992; Rappé et al., 1992; Hwang, et al., 1994; Cornell, et al., 1995; Halgren, 1996; Wang, et al., 2000; Wang, et al., 2005). MD simulations are less accurate than first principles approaches. However, a well-parameterized force field can be used to produce results comparable to those of quantum mechanical (QM) calculations (Weiner, et al., 1984). The dynamical properties of proteins, and DNA and RNA molecules can be described by performing MD simulations using well-established traditional force fields such as AMBER, CHARMM, GROMOS and OPLS-AA (Jorgensen, et al., 1996; Mackerell, et al., 1998; Ponder and Case, 2003; Oostenbrink, et al., 2004; Zgarbova, et al., 2011; Bergonzo and Cheatham, 2015; Maier, et al., 2015; Vanommeslaeghe and MacKerell, 2015; Galindo-Murillo, et al., 2016; Tian, et al., 2019). The generation of parameters for fundamental units for biological macromolecules is sufficient to describe the properties of proteins, DNA and RNA. However, the force field for small organic molecules should cover a large chemical space because each drug-like molecule contains different chemical fragments.

In general, a force field consists of bonded and nonbonded parameters (Jorgensen, et al., 1996; Mackerell, et al., 1998; Ponder and Case, 2003; Oostenbrink, et al., 2004; Zgarbova, et al., 2011; Bergonzo and Cheatham, 2015; Maier, et al., 2015; Vanommeslaeghe and MacKerell, 2015; Galindo-Murillo, et al., 2016; Tian, et al., 2019). Nonbonded parameters are van der Waals and electrostatic atomic charges. In molecular simulations, electrostatics are calculated using atom-centered point charges with the aid of a simple Coulombic model. The electrostatic energy component is the dominant one in nonbonded interactions such as ligand binding to a receptor, therefore, the generation of qualitative atomic charges plays a key role in studying the binding of ligands to receptors using simulations (Honig and Nicholls, 1995). An atomic charge should include the influence of the corresponding atom and its bonded atoms. Additionally, the point charge must account for the electronic effects from nearby electron-donating or electron-withdrawing functional groups and formal charges in the molecule (Jakalian, et al., 2002). Hence, charge models should take into account all these effects.

To generate electrostatics for a molecule, it is necessary to perform QM calculations. Several software packages, such as antechamber (Wang, et al., 2006) and CGenff (Vanommeslaeghe, et al., 2010) generate force field parameters for small organic molecules using quantum mechanical calculations at different levels. Charge methods, including AM1-BCC, CGenFF, CM1A,

CM3P and CM5, are used in conjunction with AMBER, CHARMM and OPLS force fields to generate force field parameters for drug-like molecules (Storer, et al., 1995; Jakalian, et al., 2000; Jakalian, et al., 2002; Thompson, et al., 2003; Marenich, et al., 2012). The charge methods CM1A (Storer, et al., 1995), CM3P (Thompson, et al., 2003) and AM1-BCC (Jakalian, et al., 2000; Jakalian, et al., 2002) and produce atomic charges by applying different empirical corrections to charges derived from semiempirical quantum methods such as AM1 and PM3. CM5 produces charges using Hirshfeld population analysis with the aid of density functional theory (DFT) methods (Marenich, et al., 2012). To consider the polarization effect by the environment, these methods increase the magnitude of charges by using scaling factors such as 1.14 for CM1A3 and 1.20 for CM5 (Udier-Blagovic, et al., 2004; Vilseck, et al., 2014). AM1-BCC utilizes bond-based incremental corrections to the charges obtained by Mulliken population analysis (Jakalian, et al., 2000). Bond charge corrections are parametrized by fitting to HF/6-31G* ESP of molecules in the training set (Jakalian, et al., 2000). These models have both pros and cons. For instance, AM1-BCC successfully describes electrostatics for nonpolar molecules such as saturated and aromatic hydrocarbons. However, it fails in the case of polar molecules such as pyridines, alkyl amines, alkyl and aryl halides, sulfides, and nitriles (Jakalian, et al., 2000; Jakalian, et al., 2002). The DFT-derived CM5 model suffers from a lack of a fixed scale factor to account for internal electron delocalization and external polarization effects (Marenich, et al., 2012). Recently, the 1.14*CM1A charge model with localized bond charge corrections showed high accuracy in reproducing experimental solvation free energies and heat of vaporization and densities with relatively small errors (Dodda, et al., 2017). In addition to AM1-BCC charge method, antechamber produces RESP charges using the ESP charges from user provided QM calculations for the molecule. CGenff program initially estimates ESP charges from the optimization calculations at MP2/6-31G* level which is computationally expensive to perform. Then it further optimizes the charges based on the QM data for the molecule which is interacting with water molecules in various orientations. Thus, the popular Antechamber and CGenff methods use ESP charges from different levels of theory and then introduces corrections to further improve the quality of charges. Despite the success in charge models, it is necessary to develop charge models which are optimized for efficiency and accuracy for small molecules to the accurate estimation of electrostatics in MD simulations.

Machine learning algorithms have been successfully applied to the generation of new scaffolds of small drug-like molecules (Lavecchia, 2015; Lipinski, et al., 2019; Patel, et al., 2020; Carracedo-Reboredo Jose et al., 2021), toxicity prediction (Wu and Wang, 2018), and omics pattern recognition (Stanke and Morgenstern, 2005). Machine learning algorithms have also been applied to predict partial charges and forces on atoms of small molecules in the field of quantum chemistry (Roman and

Dominik, 2019; Pattnaik, et al., 2020). The calculated force on the atom in a molecule is used to perform *ab initio* MD simulations. The contribution of electrostatic interactions is prominent in force field-based MD simulations (Jorgensen, 2005). The atomic charges of molecules alter the interaction with water thus sensitive to condensed phase properties including free energies of hydration and heats of vaporization (Jorgensen and Tirado-Rives, 2005). The accurate estimation of electrostatic interactions between proteins and ligands is important in calculating binding free energies, which are useful for screening small molecules in computer-aided drug design (Jorgensen 2009). Despite the progress in the polarizable force fields, the point charge models are still essential owing to their low computational cost and accuracy (Swope, et al., 2010). Hence, in this study, we have developed machine learning and DFT charge-based artificial intelligence (AI) models to predict atomic charges and to generate force fields for small molecules in less than a minute of time.

2 Computational methods

2.1 Force field parameters

The potential energy is the sum of the nonbonded (van der Waals and electrostatic) and bonded (bonds, angles and dihedrals) interactions in a molecule. The general functional form of potential energy in force fields is as follows in eqn. 1. (Jorgensen, et al., 1996; Mackerell, et al., 1998; Ponder and Case, 2003; Oostenbrink, et al., 2004; Zgarbova, et al., 2011; Bergonzo and Cheatham, 2015; Maier, et al., 2015; Vanommeslaeghe and MacKerell, 2015; Galindo-Murillo, et al., 2016; Tian, et al., 2019).

$$V = \sum_{\text{bonds}} K_b (r - r_0)^2 + \sum_{\text{angles}} K_\theta (\theta - \theta_0)^2 + \sum_{\text{dihedrals}} K_\phi [1 + \cos(n\phi - \gamma)] + \sum_{i,j \text{ pairs}} \left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^6} \right) + \sum_{i,j \text{ pairs}} \frac{q_i q_j}{\epsilon r_{ij}} \quad (1)$$

where K_b = force constant of bond, K_θ = force constant of angle, K_ϕ = force constant of dihedral angle, r = bond length, r_0 = equilibrium bond length, θ_0 = equilibrium angle, θ = angle, ϕ = dihedral angle, ϕ_0 = equilibrium dihedral angle, q_i , q_j = partial charges, A_{ij} , B_{ij} = well depth and r_{ij} = distance.

All the above-mentioned force field parameters are necessary to calculate the potential energy in MD simulations. In this study, we aimed to generate all these force field parameters except van der Waals potentials for drug molecules using machine learning tools. The existing van der Waals parameters for the atom types of organic molecules were developed with great care by matching the densities and enthalpies of vaporization (Cornell, et al., 1995; Jorgensen, et al., 1996). The van der Waals parameters also developed using QM methodologies and they were refined by fitting experimental properties including heat of vaporization, molecular volume and hydration free energy (Rupakheti et al.,

2018) and it needs an extensive of work to achieve. Also, the small changes in van der Waals potentials cause significant changes in the properties of molecules in the solution (Rupakheti et al., 2018; Boulanger et al., 2021). Therefore, in this study, we have not focused on the development of new van der Waals parameters using machine learning algorithms.

2.2 Generation of the training data set

To generate a training data set for machine learning, we collected 100,000 small molecules to represent the entire druggable chemical space of small molecules from the ChEMBL-2.5 database (Davies, et al., 2015) after careful removal of salts, ions and small fragments. We considered to perform quantum mechanical calculations for all 100,000 molecules and the collected data would be used for machine learning training. However, it needs 2 years of time to complete all these calculations with our existing computational resources. Thus, the calculations were subjected to three batches. The 31,770 molecules used in this work represent the first batch. To select molecules of this batch, we divided the 100,000 molecules into 10 parts based on their index. Then we selected the third part of each 10,000 molecules by random choice function on the index. To check the trainability of this batch, we have verified by principal component analysis (PCA) that the projection of this batch covers the chemical space of 100,000 molecules. The calculations for the other two batches are in progress. Figure 1 clearly shows that the selected 31770 molecules covered the entire chemical space of 100,000 molecules. This shows that the selected molecules can cover the whole chemical space.

2.3 Density functional theory calculations

The 2D structures in Simulation Description Format (SDF) were converted to 3D format using OpenBabel (O'Boyle, et al., 2011) software, and hydrogen atoms were added to all molecules. The 3D geometries of the collected small molecules were optimized using DFT at the B3LYP/6-31G** level of theory with the Gaussian16 package (Frisch, et al., 2016). The optimized geometries were subjected to frequency calculations to confirm that structures were stable on a potential energy surface (PES) at the same level of theory. The frequencies showed that there were no imaginary values, implying that the geometries corresponded to stationary points on the PES. Atomic charges are not observable in experiments or in quantum chemical calculations. Several methods have been suggested to estimate atomic charges. Here, we calculated electrostatic potential (ESP) charges for all atoms in a molecule using the Merz-Kollman method (Chandra Singh and Kollman, 1984) at the B3LYP/6-31G** level. The DFT functional is good in accuracy and

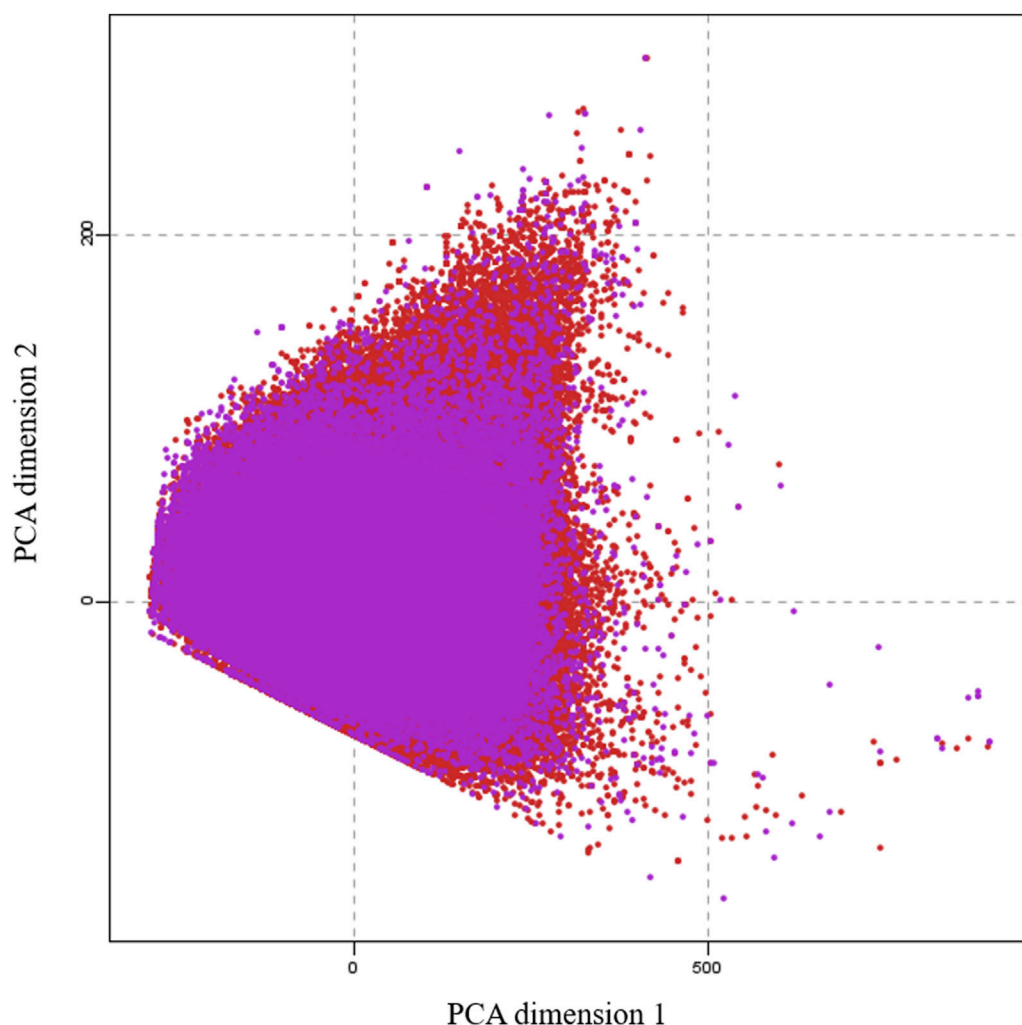


FIGURE 1

A principal component analysis (PCA) plot, showing the comparison of the chemical space defined by our dataset (purple) and the chemical space represented by CHEMBL25 databases (red).

predicting ground state properties of molecules compared AM1 method. DFT methods are computationally expensive than AM1 whereas cheaper than MP2 method to perform calculations on large number of molecules.

2.4 Machine learning and deep learning

The local environment of an atom in a molecule was described using atomic features. Bonding and neighbor atom information for the atoms in a molecule were extracted with the help of molecular graphs implemented in the MolMod package (Verstraelen, 2019). From the optimized geometries of ESP charges for atoms, bond lengths, bond angles and dihedral angles values were extracted for each molecule in the data set.

The local environment around an atom in a molecule strongly influences its atomic charge. Therefore, to train the atomic charge for an atom in a molecule, the atomic features such as atomic number, electronegativity, atomic size, valence, hybridization, aromatic nature, chiral, axial, hydrogen donor or acceptor are first extracted for each atom in a molecule.

Next, we included the features of bonding (first shell around the reference atom) information for each atom in a molecule. The local bonded atom information, such as neighboring atoms, number of bonds, bond orders and bond lengths for each atom in a molecule, was extracted from the optimized geometries. The properties of the atoms in the first shell were included using features such as aromaticity, hybridization and the presence or absence of rings, fused rings, and double bonds obtained from structures. We also added information about the

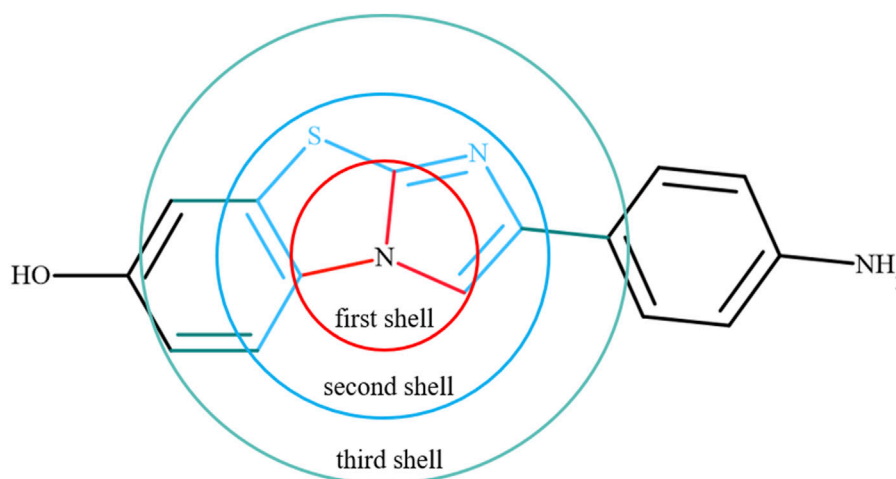


FIGURE 2

An example to show the chemical environment around a reference atom using the first, second and third shells. Red indicates the first shell, sky blue is the second shell, and Aztec blue represents the third shell around the reference atom.

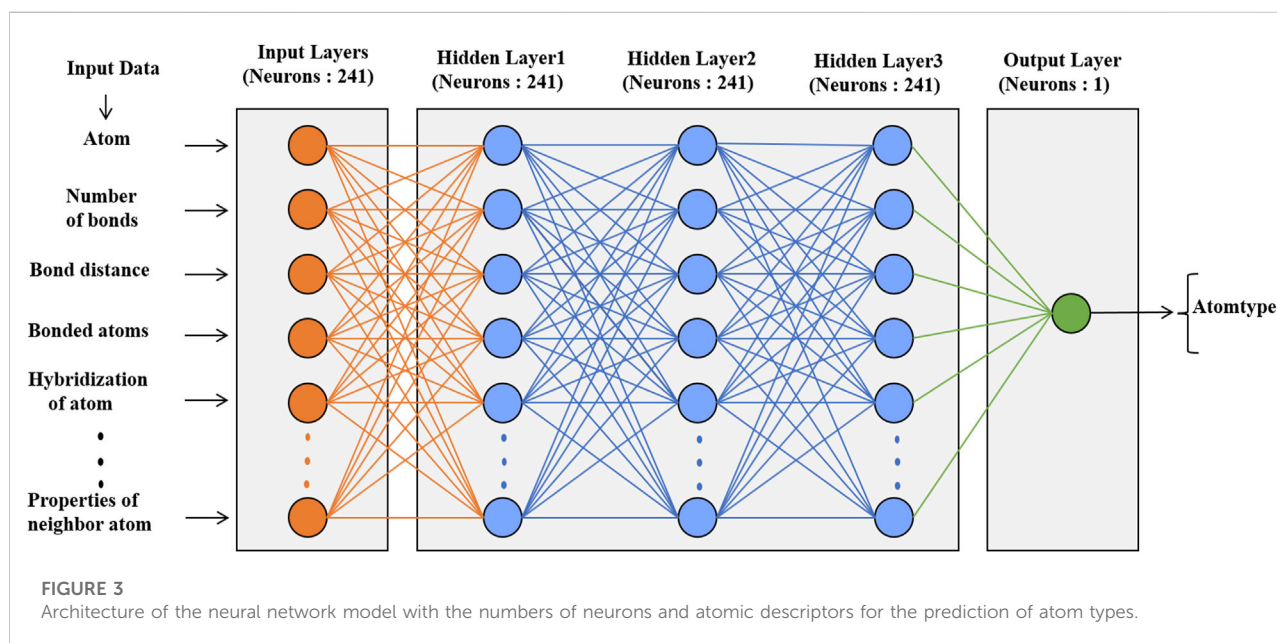


FIGURE 3

Architecture of the neural network model with the numbers of neurons and atomic descriptors for the prediction of atom types.

atoms present in the second and third shells around the reference atom in a molecule. Overall, the chemical environment was described around one atom in the molecule using the properties of the reference atom and atoms in the first, second and third shells. A schematic of the chemical environment around a reference atom is shown in Figure 2. The information was collected for 1.53 million atoms from 31770 molecules. Accessing such information was not straightforward, and it is not readily available in packages at

present. For this purpose, in-house scripts were used to extract all this information.

We applied classification and regression algorithms to train the derived data for small molecules selected from ChEMBL-2.5. A neural network classifier model was used for training to classify the data. Random forest and neural network regressors were employed to predict numbers for unknown data. We used the Python-based scikit-learn package to construct, train and validate the classification and regressor models (Pedregosa, et al., 2011).

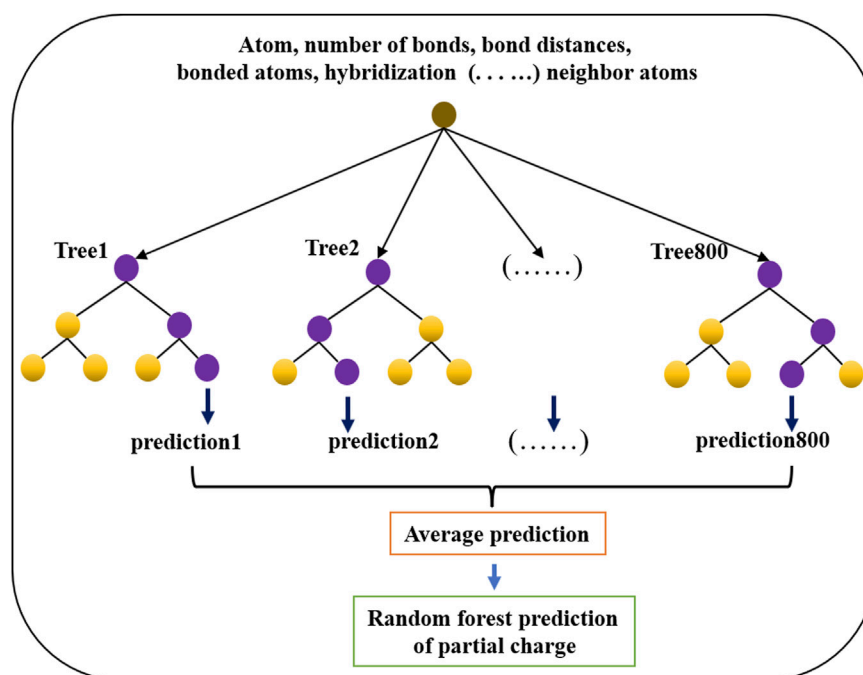


FIGURE 4

Architecture of the random forest regression model used to predict partial charges for atoms in drug-like small molecules.

2.4.1 Deep learning

The neural network classification model in the scikit-learn package was used for atom types, phase angles and periodicity classification. The architecture of these models is shown in Figure 3; Supplementary Figure S1. The data set had 31,770 molecules resulting in 1.53 million atoms and 4.8 million torsional terms for training atom types and phase angles and periodicities, respectively. The models were trained with a learning rate of 0.001, which controlled the step size in updating the weights, and a default batch size. The default log-loss was used as a loss function. Relu was used as the activation function for the hidden layers, and Adam (Diederik and Jimmy, 2015) a stochastic gradient-based optimizer, was used to update the weights. Similar parameters were used in the prediction of partial charges with the neural network regressor except for the loss function. Mean square error (MSE) was used as the loss function and to validate the model.

2.4.2 Machine learning

The random forest regressor estimator fits trees on various subsamples of a data set and uses averaging to improve the prediction. The random forest regressor model (shown in Figure 4) was used to train and predict the partial charges of atoms in molecules. The model was constructed with 800 trees, and the maximum depth was 100 for each tree. Mean square error was used to validate the regression model. All other parameters were used as default values in the scikit package.

In all models, 80% of the data were used for training, and the remaining 20% were used for validation. The parameters of the random forest regressor model were determined by employing k-fold cross validation with $k = 5$. The mean square error (MSE) was calculated for the predictions in each fold and then averaged.

2.5 Molecular dynamics simulations

Free energy calculation methods are generally implemented using the so-called lamination strategy or multistage sampling along a suitably defined chemical coordinate, λ , whereby the system is simulated in an appropriate number, n , of intermediate states corresponding to values of λ between 0 and 1. In this study, small molecules (33 compounds) were selected for solvation free energy calculations. Small molecules were solvated in a cubic box using the TIP3P water model (Jorgensen, et al., 1983). These systems were subjected to energy minimization using the steepest descent method and subsequently equilibrated for 1 ns at 298 K and 1 bar pressure. Velocity rescaling and Parrinello–Rahman algorithms were used to control temperature and pressure in the NPT ensemble (Parrinello and Rahman, 1981; Nose and Klein, 1983; Bussi, et al., 2007). Furthermore, equilibrated solvated structures were simulated for a production run of 1 ns in the NPT ensemble using a 2 fs time step (Berendsen, et al., 1995; Lindahl, et al., 2001; Hess, et al., 2008). The particle mesh Ewald method was used to calculate the electrostatic interactions with

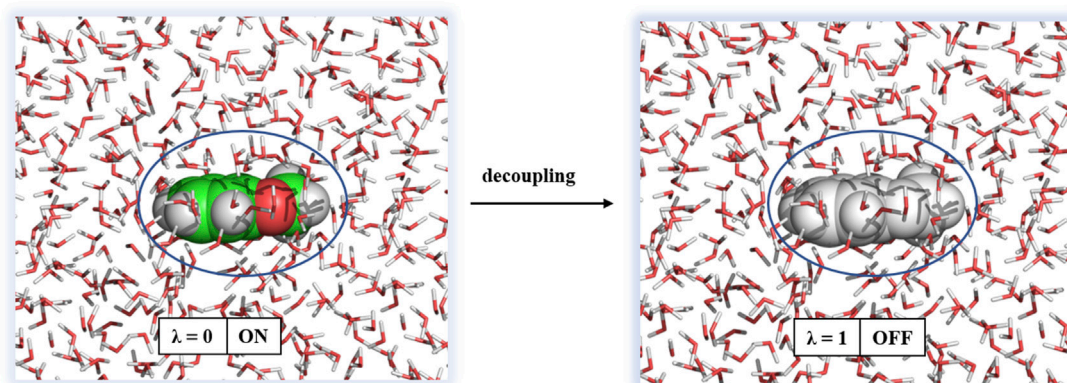


FIGURE 5

Decoupling of a ligand from solvation. Water molecules are shown in sticks and ligand as spheres.

an interpolation order of 4 and a grid spacing of 1.6 Å (Essmann, et al., 1995). Bonds between hydrogen and heavy atoms were constrained at equilibrium bond lengths using the LINCS algorithm (Hess, et al., 1997). All simulations were performed using the GROMACS-2020 package.

All solvation free energy calculations were performed by decoupling the ligand from the solvent environment. The initial conformation of the ligand in solvent was taken from the final snapshot of the 1 ns simulation. Decoupling of the ligand from solution was performed by turning off Coulombic interactions and subsequently van der Waals interactions. The approach of solvation free energy calculation is shown in Figure 5.

The Coulombic interactions were turned off by changing λ from 0 to 1 with a step size of $\Delta\lambda = 0.25$, and the van der Waals interactions were unperturbed. Then, the van der Waals interactions were turned off with nonuniformly distributed values of λ (0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.65, 0.7, 0.75, 0.8, 0.85, 0.9, 0.95, and 1.0). Therefore, a total of 20 windows, each 1 ns, were employed for decoupling of the ligand from solution. The free energy difference between two end states was calculated using the Bennett Acceptance Ratio (BAR) method (Bennett, 1976) and the following equation:

$$\left\langle \frac{1}{1 + \exp\{\beta(\Delta U_{ij} - \Delta G)\}} \right\rangle_i = \left\langle \frac{1}{1 + \exp\{\beta(-\Delta U_{ij} + \Delta G)\}} \right\rangle_j \quad (2)$$

where β is the reciprocal of the thermodynamic temperature, ΔG is the free energy difference between states i and j , and $\Delta U_{ij} = U_j - U_i$ is the potential energy difference.

2.6 Protein–ligand simulations

The crystal structure of the protein kinase, covid19 (main protease) and factor-IX with cocrystal ligand were taken from the protein data bank (PDB id: 4XUF, 7L10 and 5TNT). Protein structures were prepared by correcting the bond orders, adding missing hydrogens and optimizing H-bonding with protonation states of residues at pH 7.0 using protein preparation wizard (Sastry et al., 2013). The complex was solvated in a cubic box with a TIP3P water model. The total charge of the proteins was neutralized by inclusion of Na^+ and Cl^- ions. The AMBER99SB-ILDN force field was used for the proteins. The force field parameters for the cocrystal ligands were generated using generalized amber force fields (GAFF) and machine learning force field for the comparison. All solvated the protein and ligand complexes were subjected to energy minimization using steepest decent method. Temperature and pressure controls were imposed using the V-rescale and Parrinello-Rahman algorithms with 298 K and 1 bar, respectively (Parrinello and Rahman, 1981; Nose and Klein, 1983; Bussi, et al., 2007). The simulations were carried out with a time step of 2 fs for 1 ns to equilibrate the systems in the NPT ensemble. The production run was performed for 250 ns for each complex using a time step of 2 fs in NPT ensemble. The interpolation order of 4 and a grid spacing of 1.6 Å were used in the electrostatic calculations using particle mesh Ewald method (Essmann, et al., 1995). LINCS algorithm have used to constrain the bonds of hydrogens with heavy atoms (Hess, et al., 1997).

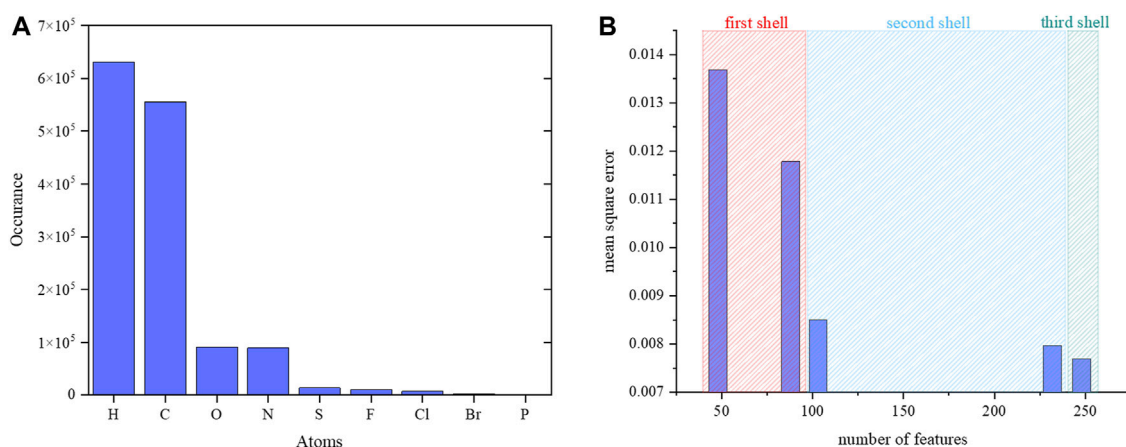


FIGURE 6

(A) Number of samples per element in the data set. Numbers for phosphorous atoms are not visible in the graph as very few phosphorus atoms were present in the data. (B) The calculated MSE vs. number of atomic features for a reference atom.

3 Results and discussion

3.1 Prediction of partial charges

The number of samples per element presented in the data set is shown in Figure 6. Each atom has its local chemical environments and reference partial charge in the data. The calculated MSE in Supplementary Table S1 shows that the random forest regressor is slightly better than the neural network regressor. Therefore, a random forest regression model was adopted for further validation. The MSE was optimized by increasing the number of descriptors for each atom in the data set. The descriptors were atoms and their properties in the first, second and third shells around a reference atom. The MSE with respect to the number of descriptors is shown in Figure 6. The addition of the chemical environment reduced the MSE value for the random forest regression model. The performance of the random forest regression model was best when all atoms and their properties were included in the three shells. In previous study also, it was shown that random forest regression produces reliable results compared to other machine learning algorithms. Previous study randomly collected the data for 10000 and 7,000 molecules from ATB (automated topology builder) and PRODRG servers. ATB applies symmetry-based averaging of atomic charges based on the ESP charges from B3LYP/6-31G* calculations for small molecules with the number of atoms less than 40 otherwise it carried out semiempirical calculations to generate the charges. Whereas in this study, we have performed calculations at B3LYP/6-31G* for all the molecules which are having more than 40 number atoms in addition smaller size molecules. The considered data for the training of random forest regression model has the molecules with atoms range from 10 to 120. The

number of data points and features are used in the training of our charge model (241 features) is higher than the previous study (61 features). The features include the bond orders, bond lengths, hybridizations and electronegativities for neighbor atoms to provide the chemical environment around an atom whereas the previous study does not consider them.

In Figure 7, the predicted charges were fitted to reference charges for elements including carbon, hydrogen, nitrogen and oxygen. The same plots for other elements, such as sulfur, fluorine, chlorine, bromine and phosphorous, are given in Supplementary Figure S2 of the Supplementary Information. Notably, the majority of the predicted charges were similar to the reference values. The calculated coefficient of determination (R^2) and MSE values for carbon, hydrogen, oxygen, nitrogen, sulfur, fluorine, chlorine, bromine and phosphorous are presented in Supplementary Table S2. The calculated coefficient of determination (R^2) values for carbon, hydrogen, oxygen, nitrogen, sulphur, fluorine, chlorine, bromine and phosphorous are 0.871, 0.847, 0.852, 0.880, 0.977, 0.632, 0.805, 0.714 and 0.664, respectively. MSE values are 0.0148, 0.001, 0.002, 0.013, 0.004, 0.0003, 0.004, 0.001 and 0.027 for carbon, hydrogen, oxygen, nitrogen, sulphur, fluorine, chlorine, bromine and phosphorous, respectively. The prediction accuracy was less for fluorine, bromine and phosphorous than for other elements. This may have been due to a smaller number of samples in the data. The prediction accuracy for atoms such as C, H, O, N, S, P, F, Cl and Br is low when compared to previous study. Because the number of data points and data for each atom is different and it increases variance in the atomic charges thus makes difficulty in the prediction. It is difficult to compare the charges from our charge model with other methods because the atomic charges for a molecule using QM calculations are often sensitive to functional and fitting method which are used to generate ESP

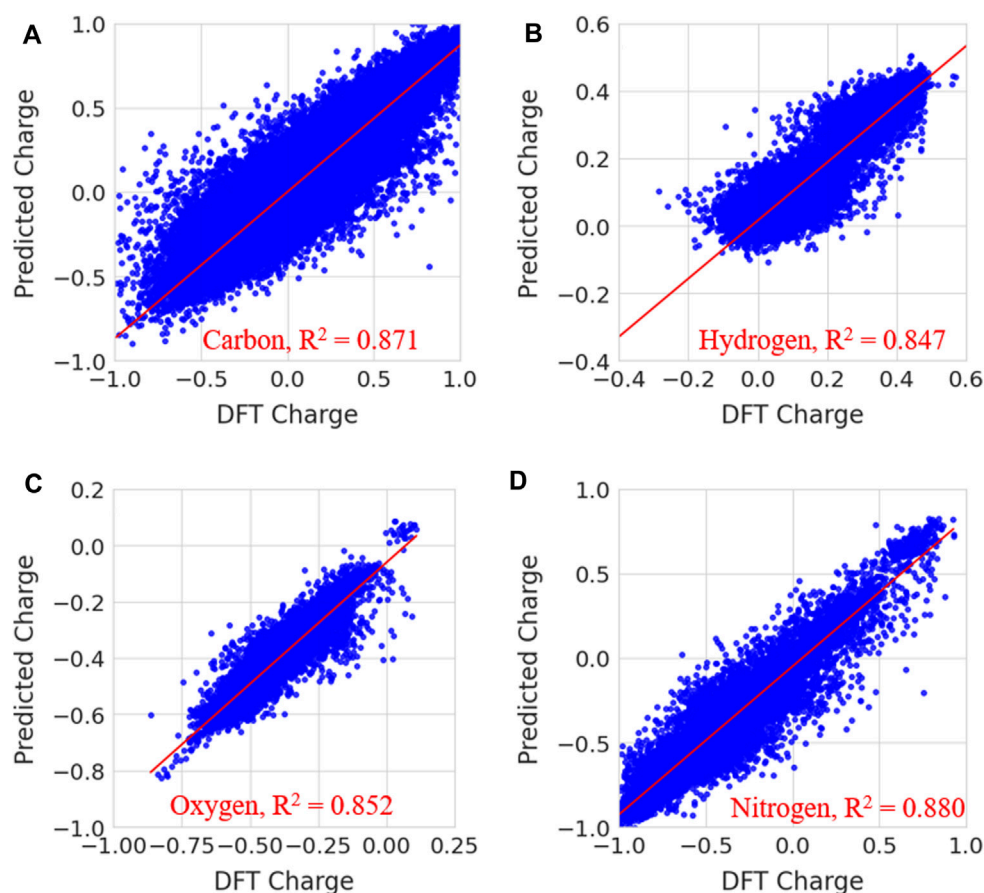


FIGURE 7

Prediction of partial charges for (A) carbon (B) hydrogen, (C) oxygen and (D) nitrogen atoms in the test data set using a random forest regression model.

charges. The charges from different fitting methods are not same for a molecule. However, we have provided the comparison of our charge predictions with ESP charges of ATB, QM and AM1-BCC methods for one molecule in [Supplementary Table S3](#). It can be clearly noted that charges in all these methods are not same. The quantity of atomic charges is different in each method whereas the sign (+ or -) is same in the case of all atoms. In order to understand the atomic charges produced from random regression model, the calculated solvation free energies for molecules using different charge methods are compared with experimental values in the validation section.

To validate, the performance of the trained random forest regression model was tested on two external test sets. Test set-1 consisted of 100 molecules that were randomly selected from the drug-induced liver injury database. This database consists of FDA approved drugs that are shown to be toxic to the liver. Test set-2 considered 33 molecules that had experimental solvation free energies in the literature. We have tested the charge model on two different datasets (i) first dataset is having molecules

which contains atom numbers range from 20 to 87 (ii) second dataset contains molecules with atom number range from 9 to 24. The small molecules in test set-2 consisted of various electron-donating and electron-withdrawing functional groups. Eight small molecules from test set-2 are shown in [Figure 8](#). The predicted charges are plotted against DFT charges for both test sets and displayed in [Figure 9](#). The R^2 and MSE values reveal that the prediction accuracy for the test sets was high. The predicted values for a few molecules are compared with ESP charges obtained from DFT calculations in [Table 1](#) and [Supplementary Table S4](#). [Table 1](#) shows that the predicted values were close to the DFT charges. The random forest regressor gave the correct sign (+ or -) and values similar to the ESP charges. It was evident that the random forest regressor model produced can work for small molecules as well as for large size molecules.

However, it is noted that machine learning charge model can assigns wrong sign (+ or -) for aliphatic carbon atoms compared to DFT ESP charges. For example, we considered the charges for

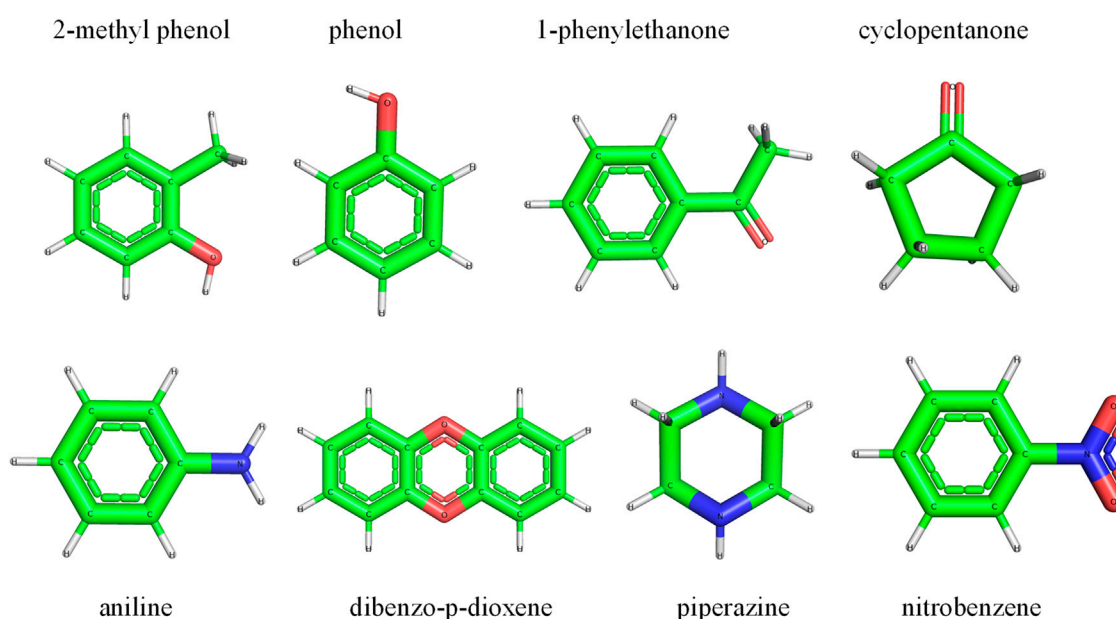


FIGURE 8

A few small molecules were selected from test set-2 for the validation of the random forest regression model.

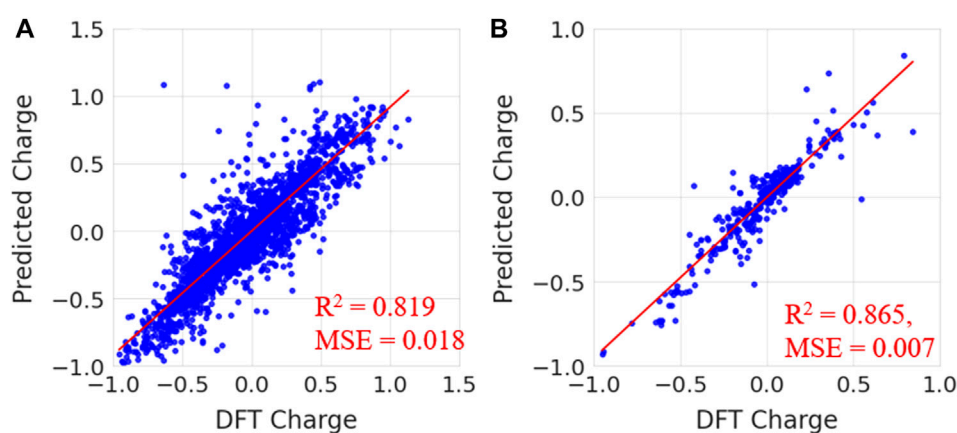


FIGURE 9

Predicted charges vs. DFT charges for (A) test set-1 from the drug-induced liver injury database and (B) test set-2 for known small molecules.

aliphatic molecule (1-Octanol) which are generated using random forest model and DFT calculation [Supplementary Table S5](#). It can be seen that C3, C4, C5, C6 and C7 have positive atomic charge in random forest prediction. Whereas the atoms C3, C5 and C6 are negative and C4 and C7 are having positive partial charge in the case of ESP from DFT calculations. Because all C3, C4, C5, C6 and C7 are sharing similar kind of

bonding environment thus random forest predicts positive charges for them. The prediction of charges can be improved by adding a greater number of diverse aliphatic molecules into the data set to reproduce the ESP of DFT.

Existing small molecule force field generate programs such as antechamber, CGenFF, ATB and PRODRG produces charges based on ESP. Antechamber program uses ESP charges from

TABLE 1 Comparison of predicted partial charges from the random forest regression model with DFT charges.

Cyclopentanone			Aniline		
Atom	Random forest	DFT	Atom	Random	DFT
O	-0.56910	-0.49271	N	-0.84978	-0.78174
C	-0.06156	-0.04073	C	0.18780	0.35203
C	-0.05696	-0.04086	C	-0.29226	-0.25455
C	-0.15291	-0.19790	C	-0.25365	-0.25456
C	-0.23938	-0.19782	C	-0.06735	-0.10162
C	0.59172	0.54435	C	-0.08070	-0.10157
H	0.02416	0.03785	C	-0.08311	-0.15603
H	0.04566	0.02639	H	0.16222	0.14008
H	0.04022	0.03786	H	0.14058	0.14009
H	0.05218	0.02643	H	0.13362	0.11626
H	0.08876	0.06753	H	0.13362	0.11625
H	0.08969	0.08105	H	0.13362	0.11343
H	0.07499	0.06753	H	0.36227	0.33594
H	0.07256	0.08101	H	0.37313	0.33595

quantum calculation and produces restrained electrostatic potential (RESP) using least square fitting method. We have used antechamber to generate RESP charges for small dataset of 2,700 molecules to train using random forest regression model. The obtained charge model shows the MSE of 0.027 on the test set. We have predicted the RESP charges for testset-1 and testset-2 datasets. The calculated coefficient of determinations are 0.71 and 0.61 [Supplementary Figure S3](#). The model shows promising result though the trained on a smaller number of atoms. The model can be improved a lot by incorporating a greater number of atoms into training set. We will develop a RESP charge prediction model using greater number of molecules in the near future.

3.2 Atom type prediction

Categorizing the atoms in a molecule into atom types is useful to assign the force field parameters. Antechamber programs were used to generate atom type data for atoms in the molecules. The trained neural network classifier model performed well with 98% accuracy on the test data set. The model identified the atom types based on the provided local chemical environment around a reference atom. The calculated confusion matrix produced precision, recall, F1-score and accuracy for each atom-type prediction. The model could identify only the following atom types: C, H, O, N, S, P, F, Cl, and Br. The prediction accuracy for each atom type is given in [Table 2](#). In [Table 2](#), from c to cy, from h1 to hx, from n to nq, from o to os, from p3 to py, from s to sy, f, cl and br are atom

types for C, H, N, O, P, S, F, Cl, and Br, respectively. The definition for each atom type is similar to the generalized amber force field (GAFF). The model clearly identifies all types of H, O, F, Cl, and Br atoms with 100% accuracy. Additionally, the sulfur and phosphorus atom type prediction accuracy was 100%. The most common aliphatic, cyclic and aromatic atom types of carbon (c, c1, c2, c3 and ca) and nitrogen (n, n1, n2, n3, na and nb) were predicted with accuracy above 95%. The predictions were the least accurate for atom types cf and nf. However, the model assigns incorrect atom type in the case of sp² carbons such as cc, cd, ce, cp and cf only with another sp² carbon type (cc, cd, ce, cp and cf). All sp² carbons (nitrogen's) have the same van der Waals parameters in GAFF. The same is true in the case of sp³ carbons and nitrogens. In our force field, atom type prediction is used to assign van der Waals parameters for the atoms in a molecule. Therefore, even the incorrect prediction of atom types would not effect the force field parameters.

To assess the accuracy, the model was used to predict the atom types for a few small molecules, as shown in [Figure 8](#). The predicted atom types were compared with antechamber-produced atom types; the results are presented in [Table 3](#) and [Supplementary Table S6](#). [Table 3](#) and [Supplementary Table S6](#) show that the atom types predicted by the neural network model were in good agreement with the predictions of the antechamber program. The neural network classifier accurately identified the atom types and their chemical environments. This ensured that the model successfully assigned atom types for small drug-like molecules.

3.3 Prediction of phase shift angles for dihedral terms

The phase shift angle is involved in the dihedral energy term, and it is important to calculate the energy contribution from the dihedral energy term to the total potential energy. Each dihedral term had a specific phase angle value and was restricted to the range between 0° and 180°. The 4.8 million dihedral terms in 31770 molecules were extracted along with their phase angles. Atomic descriptors were generated for the atoms involved in each dihedral angle. The calculated dihedral angle values were also included to train the neural network classifier to predict phase angles of 0° and 180°. The trained model classified the test data set as 0° and 180° with 94% accuracy. The predicted values were well correlated with the parameters generated by the antechamber program. The incorrect prediction of phase shift angle for dihedral angle can produces the unwanted angle rotations or restrictions thus causes changes in the conformation of ligand compared to GAFF. The phase shift angle is important to retain the planarity of aromatic ring and conjugated groups in the molecules. Our phase angle model predicts accurately for these kinds of molecules and retained the planarity of molecules. However, phase shift angle model (accuracy 94%) has to be improved further to avoid the unfavorable conformational

TABLE 2 Accuracy of the prediction of atom types using a neural network model.

Atomtype	Precision	Recall	f1-score	Atomtype	Precision	Recall	f1-score
br	1	1	1	n	1	0.99	0.99
c	1	1	1	n1	0.99	0.99	0.99
c1	0.99	0.97	0.98	n2	0.99	0.96	0.97
c2	0.96	0.98	0.97	n3	1	0.98	0.99
c3	1	1	1	n4	1	0.67	0.8
ca	0.96	0.99	0.98	na	0.99	0.98	0.98
cc	0.82	0.66	0.73	nb	0.92	0.97	0.95
cd	0.72	0.67	0.69	nc	0.75	0.44	0.56
ce	0.72	0.8	0.76	nd	0.74	0.87	0.8
cf	0.58	0.45	0.51	ne	0.66	0.8	0.72
cg	0.7	0.93	0.8	nf	0.17	0.05	0.08
ch	0.43	0.1	0.16	nh	0.96	0.99	0.98
cl	1	1	1	nj	1	1	1
cp	0.57	0.79	0.66	nm	1	1	1
cq	0	0	0	no	1	1	1
cv	0.5	0.5	0.5	np	1	1	1
cx	1	1	1	nq	1	1	1
cy	1	1	1	o	1	1	1
f	1	1	1	oh	1	1	1
h1	1	1	1	op	1	0.83	0.91
h2	1	0.93	0.96	os	1	1	1
h3	1	1	1	p5	1	1	1
h4	1	1	1	py	1	1	1
h5	1	1	1	s	1	1	1
ha	1	1	1	s4	1	1	1
hc	1	1	1	s6	1	1	1
hn	1	1	1	sh	1	1	1
ho	1	1	1	ss	1	1	1
hs	0.97	1	0.98	sx	1	1	1
hx	1	0.25	0.4	sy	1	1	1

changes in the molecules by increasing the number of data points and feature incorporation in the training dataset.

Neural network model training was conducted with the same atom features for the inclusion of phase angles for the prediction of periodicity for dihedral terms. The model performed the prediction with 93% accuracy. The accuracy for classification of terms with periodicity 2 and 3 was 96% and 95%, respectively. The predicted periodicities were retained the structures of aromatic and other types of molecules. The predicted phase angles and periodicities were compared with the antechamber-generated values; they are shown in [Supplementary Table S7](#).

3.4 Generation of topology for a molecule

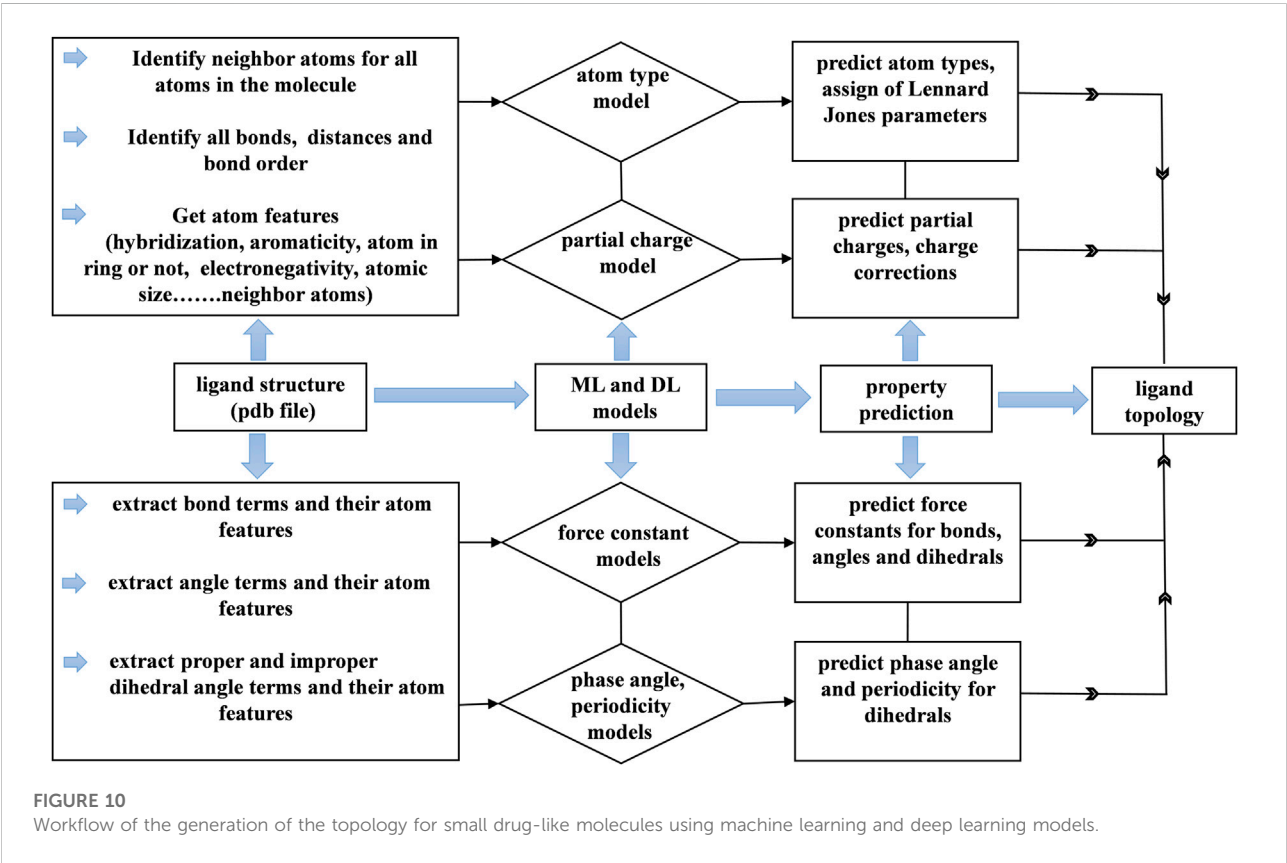
The concept of using AI algorithms was to generate parameters and topology for small molecules that generally

did not have parameters in conventional force fields. Few commercial and noncommercial software packages, such as ATB ([Stroet, et al., 2018](#)), antechamber, CGenFF and PRODG ([Schüttelkopf and Van Aalten, 2004](#)), are available to generate force field parameters for small molecules. We generated topologies for small drug-like molecules using machine learning models to predict atom types, DFT-based partial charges, phase angles, periodicity and force constants for bonds, angles and dihedrals. The work flow is shown in [Figure 10](#). In this study, flow, data collected from a molecule were used to perform predictions by employing machine learning and deep learning models. The collected information and predicted data were used to generate topologies in the format of used in most MD simulation programs, such as GROMACS and NAMD.

Topology generation started with the prediction of atom types for a given molecule. The small molecule force fields like

TABLE 3 Comparison of the atom types predicted by the neural network model and antechamber program.

Cyclopentanone			Aniline		
Atom	NN model atom type	Antechamber atom type	Atom	NN model atom type	Antechamber atom type
O	o	o	N	nh	nh
C	c3	c3	C	ca	ca
C	c3	c3	C	ca	ca
C	c3	c3	C	ca	ca
C	c3	c3	C	ca	ca
C	c	c	C	ca	ca
H	hc	hc	C	ca	ca
H	hc	hc	H	ha	ha
H	hc	hc	H	ha	ha
H	hc	hc	H	ha	ha
H	hc	hc	H	ha	ha
H	hc	hc	H	ha	ha
H	hc	hc	H	ha	ha
H	hc	hc	H	hn	hn
H	hc	hc	H	hn	hn



GAFF have limited number of atom types. Each atom type has corresponding Lennard Jones parameters. These parameters assignment depends on the atom type. The atom type prediction was done by our model with 98% accuracy with respect to GAFF. Based on atom types, Lennard Jones parameters were assigned accordingly. Lennard Jones parameters were taken from the GAFF force field. This gives the correct assignment of Lennard Jones parameters to the atoms in a molecule. Next, the partial charge model predicted atomic charges for all the atoms in a molecule. The sum of the predicted atomic charges was not equal to the formal charge of the molecule. Therefore, charge correction was applied in such a way that the sum of the predicted charges was subtracted from the formal charge of the molecule, and the difference was distributed among all the atoms to make the total charge of the molecule equal to the sum of the predicted atomic charges. Furthermore, a list of the bonded atoms and bond lengths was calculated, and the bond force constants were predicted with the aid of a trained model. Here, bond lengths from the structure were used as equilibrium distances for bonds. Subsequently, the angles and dihedral terms were added to the topology in the respective sections. Then, the nonbonded 1, 4 pairs for the molecule were generated by taking the first and fourth atoms in dihedral angle terms.

Next, we generated improper dihedral angle terms for the topology file. No tool was used to identify the improper dihedral angles in small molecules other than current force field generation programs. In general, improper angles are intended to maintain the planarity of aromatic and conjugated molecules. Aromatic and conjugated molecules are predominantly involved with carbon atoms. Three atoms are bonded to carbon atom that is involved in a double bond. We generated a list of improper dihedral angles based on the number of atoms bonded to carbon atoms and with the extraction of their neighboring atoms. Eventually, we generated force field parameters for drug-like molecules within a minute of CPU time. The correct assignment of partial charges and van der Waals parameters to the atoms enables the molecules to interact with environment such as water and protein through nonbonded interactions. The atomic features are the important in order to understand the chemical environment which effects partial charges, atom type and phase angle predictions. The user has to provide proper molecule structure by adding all hydrogens to heavy atoms otherwise user may end up with assigning of incorrect parameters which can collapses molecule structure.

3.5 Validation of the force field

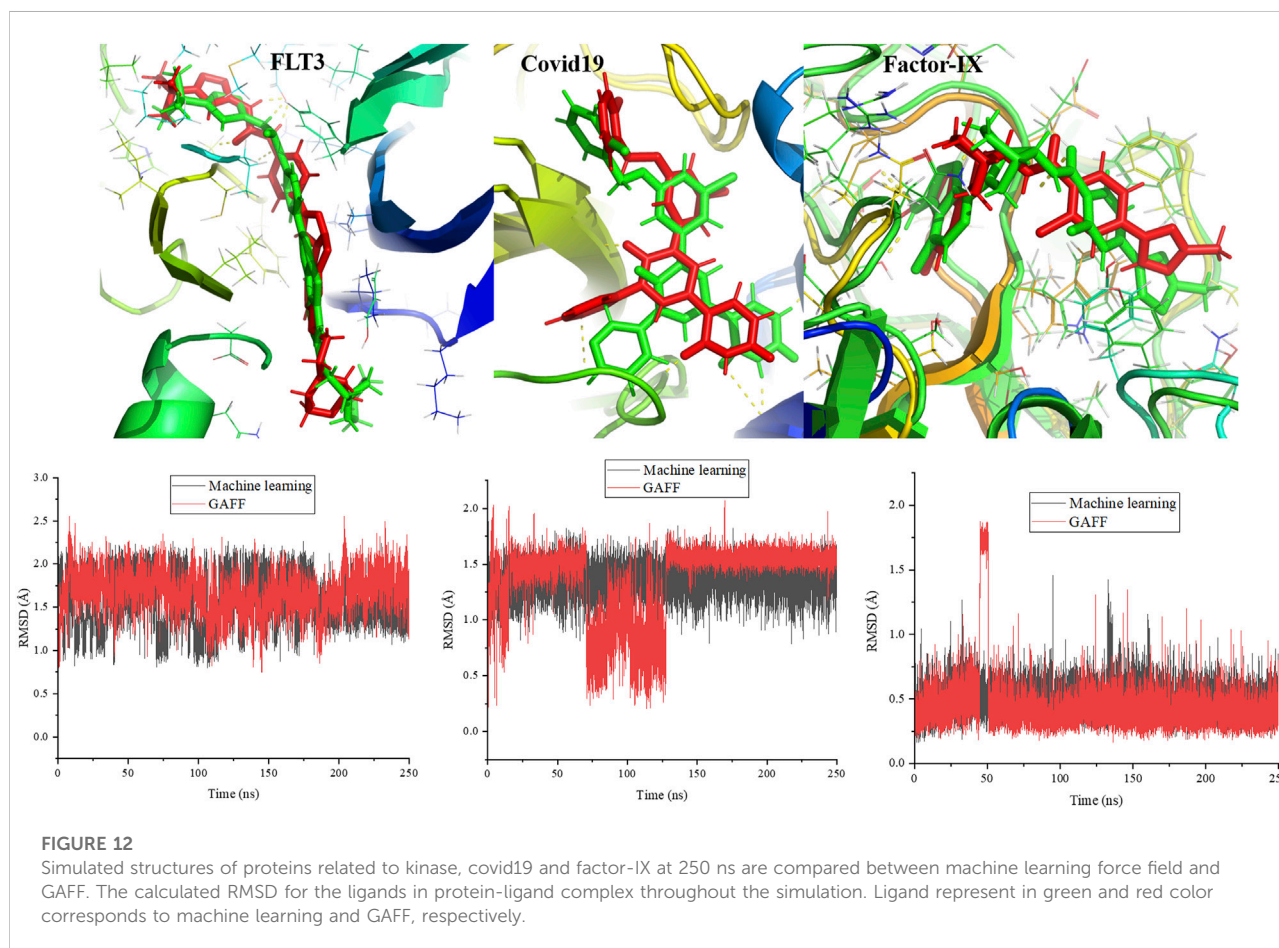
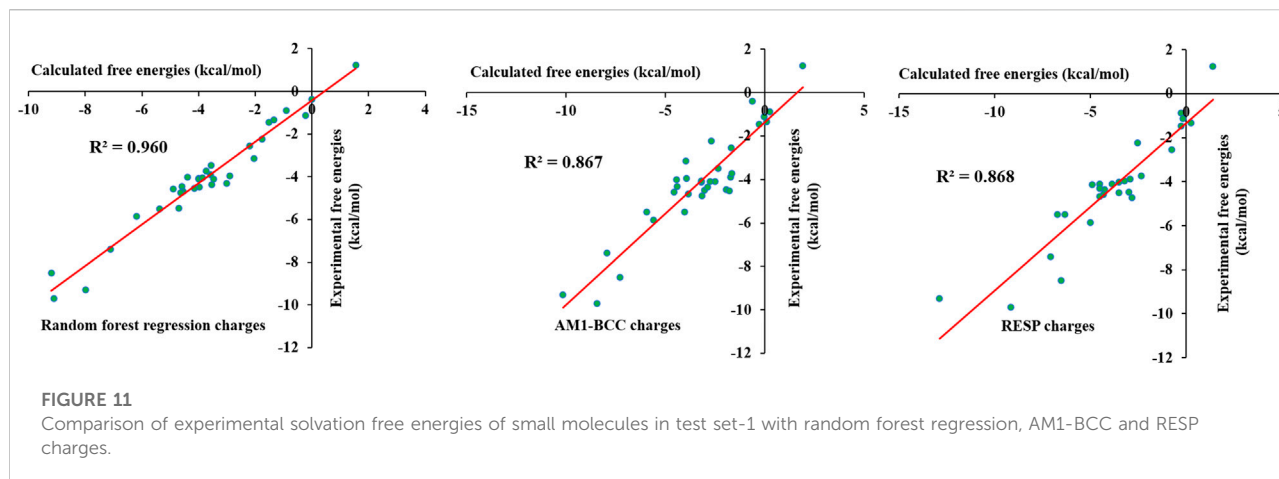
3.5.1 Solvation free energies

To verify the predicted partial charges and other force field parameters, solvation free energies were calculated for 33 selected small molecules using the λ -coupling method. This method is

reliable and accurate in the calculation of solvation free energies and has been used to calculate protein–ligand absolute binding free energies. The selected 33 molecules contained various functional groups, including alcohol, thiol, amide, amine, aldehyde, ketone, nitro, nitrile, and methyl groups and halogens. Aliphatic chains, aromatic rings and cyclic rings were also present in the chosen molecules. The calculated free energies were compared with the experimental free energies; the results are shown in [Supplementary Table S8](#). The calculated values were in close agreement with the experimental free energies. The calculated values were within 2 kcal/mol error from experimental free energies except for several molecules. To obtain reasonable free energy values, we introduced charge corrections to the atoms involved in specific bonds. This was done based on previous studies where localized bond charge corrections were added to improve the solvation free energies of small molecules (Dodda, et al., 2017). Localized bond charge corrections for few bonds were taken from the literature (Dodda, et al., 2017), and others were calculated based on a trial and error approach. The charge corrections for specified bonds are shown in [Supplementary Table S9](#). Charge corrections were performed for aliphatic, cyclic and aromatic bonds. The introduction of charge corrections significantly improved the free energy values, which were similar to the experimental numbers. The calculated values are shown in [Supplementary Table S8](#), and they reveal that the calculated values were similar to the experimental values. It can be seen that though the incorrect assignment of atomic charges for carbon atoms in 1-Octanol produces solvation free energy close to experimental value. [Figure 11](#) shows that the R^2 value reached 0.960. Thus, the corrected charges accurately described the interaction of molecules with the water environment. We have also compared the calculated solvation free energies from AI force field, AM1-BCC/GAFF and RESP/GAFF with experimental values. The calculated solvation free energies for AM1-BCC/GAFF and RESP/GAFF were taken from the literature (Shivakumar, et al., 2009) and given in [Supplementary Table S8](#). The calculated coefficient of determination for AI force field, AM1-BCC/GAFF and RESP/GAFF are 0.960, 0.867 and 0.868, respectively. The results shows that AI force field outperforms the other methods in reproducing the experimental values. However, further AI force field has to be tested on large number of molecules and compare with experimental values. Overall, the machine learning force field successfully reproduced the experimental free energies, revealing that the force field was accurate and reliable.

3.5.2 Protein–ligand interactions

To validate the force field parameters generated by the machine learning force field, MD simulations were performed for protein and ligand complexes and then compared with the results of simulations with the GAFF. The complexes were stable throughout the simulations, and the final snapshots at 250 ns are shown in [Figure 12](#). The ligand was composed of



aromatic and nonaromatic rings. There were no distortions in the ligand structure, and it was stable in the pocket. The surrounding interacting residues for the ligand were the same in the cases of the machine learning and amber force fields. However, the atoms involved in hydrogen bond formation were different in the final snapshots from both force fields.

Additionally, the ligand conformation was slightly different in the case of the machine learning force field compared to the GAFF (Figure 12). The calculated root mean square deviations (RMSDs) for the ligand throughout the simulations are presented in the Figure 12. Notably, structural changes in the ligand were not significant in either force field. The

average RMSDs of the ligand with respect to the starting conformation were 1.57 and 1.67 Å for the machine learning and GAFF force fields, respectively. In addition to protein kinase, we have performed simulations of 250 ns for the proteins such as covid 19 (main protease) (pdb id:7L10) and factor-IX (pdb id: 5TNT). In 5TNT, the ligand binds at the surface of protein, however, it is stable at the binding site throughout the simulation. We compared the snapshots of AI force field and GAFF and it is shown in [Figure 12](#). The structure of ligand at the binding is not same in both force field, however the difference is marginal. The ligands are stable at binding site through interactions with the residues of protein. The plots show that there is no significant structural changes in the ligand with respect to RMSD values.

The stability of complexes derived from the interaction energy was calculated and shown in the supporting information. The interaction energy was clearly less for the machine learning force field than the GAFF. Furthermore, the electrostatic and van der Waals energy contributions to the total interaction energy were calculated, and the results showed that electrostatic interactions were responsible for the difference in the interaction energies. The average electrostatic interaction energies between the protein and ligand were -17.4 and -30.5 kcal/mol for the machine learning and GAFF force fields, respectively. The electrostatic interaction energy was different due to variations in atomic charges between the machine learning force field and GAFF. The difference clearly shows that the machine learning force field should be improved to minimize the differences in the energies and conformations of the ligand compared to those obtained using the GAFF. We expect to study ways to improve the force field by including more data in the training data set to maximize interactions between proteins and ligands and enhance the prediction of phase angles.

4 Conclusion

A force field for small drug-like molecules was generated using machine learning and deep learning techniques. The random forest regression based charge model generates quality atomic charges comparable to DFT based ESP charges which are suitable for molecular dynamics simulations. In addition to the charge model, we developed AI-based models to predict atom types, force constants, phase angles and periodicities for dihedral terms. The classifications of atom types, phase angles and periodicity were achieved successfully with accuracies of 98, 94 and 93%, respectively. The AI models could be able to predict charges and atom types with high accuracy based on the provided atomic chemical environment through features around a reference atom. Using

all these models, we developed a module in the *pharmulator*[™] platform that generated topology files for small molecules in GROMACS and NAMD formats to perform molecular dynamics simulations. The code generates quality atomic charges and other compatible force field parameters within a minute of time. The generated force field parameters for small molecules reproduces the experimental solvation free energies with coefficient of determination value of 0.96. The calculated free energies are better reproduced than AM1-BCC and RESP charges. Further, the calculated structural changes in ligand molecules at protein binding sites are comparable with GAFF results. Overall, the results clearly revealed that the force field generated by machine and deep learning techniques was accurate and reliable for use in molecular dynamics simulations of small molecules as well as for complexes of proteins and ligands. The machine learning charge model differs from AM1-BCC and CGenff methods in terms of methodology and level of theory used to generate atomic charges. This method could optimize the efficiency and accuracy of calculations to produce reasonable ESP charges. Also, DFT calculations to obtain ESP charges were included at additional computational cost, which increased with the size of the molecule. Therefore, the rapid prediction of accurate ESP partial charges, within a minute of time and without quantum mechanical calculations, would be very helpful in the drug discovery process.

However, AI based force field models may have certain limitations that it assigns incorrect sign (+ or -) for aliphatic carbon atoms compared to ESP charges from DFT. In some cases, the prediction phase angle for dihedral angle can be incorrect that may introduce flexibility or rigidity in the molecules. These limitations can be overcome by adding large number of diverse aliphatic molecules into training data set. The increasing of data points and number of features for training of phase angle model would improve the accuracy to overcome the limitations.

Generation of force field parameters for ligand molecules is useful to perform molecular dynamics simulations to analyze the interactions and to estimate binding free energy in implicit and explicit water environment. The estimation of atomic partial charges of the small molecules to calculate molecular interaction fields (MIFs) is an important process in field-based quantitative structure-activity relationship (QSAR) ([Mittal et al., 2009](#); [Gadhe et al., 2011](#)). The predicted DFT based charges could also be useful to incorporate in docking calculations to perform virtual screening ([Cho et al., 2005](#)).

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors upon email request.

Author contributions

SM, performed and analyzed the AI training, DFT calculations and MD simulations. SM, written the entire code and executed. AB, provided the curated data set and helped in code writing. SW and SM, conceived and designed the experiments and wrote the paper.

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.

References

- Allinger, N. L. (1977). Conformational analysis. 130. MM2. A hydrocarbon force field utilizing V1 and V2 torsional terms. *J. Am. Chem. Soc.* 99 (25), 8127–8134. doi:10.1021/ja00467a001
- Allinger, N. L., Yuh, Y. H., and Lii, J.-H. (1989). Molecular mechanics. The MM3 force field for hydrocarbons. 1. *J. Am. Chem. Soc.* 111 (23), 8551–8566. doi:10.1021/ja00205a001
- Bennett, C. H. (1976). Efficient estimation of free energy differences from Monte Carlo data. *J. Comput. Phys.* 22 (2), 245–268. doi:10.1016/0021-9991(76)90078-4
- Berendsen, H. J. C., van der Spoel, D., and van Drunen, R. (1995). Gromacs: A message-passing parallel molecular dynamics implementation. *Comput. Phys. Commun.* 91 (1), 43–56. doi:10.1016/0010-4655(95)00042-E
- Bergonzo, C., and Cheatham, T. E., III (2015). Improved force field parameters lead to a better description of RNA structure. *J. Chem. Theory Comput.* 11 (9), 3969–3972. doi:10.1021/acs.jctc.5b00444
- Boulanger, E., Huang, L., Rupakheti, C., MacKerell, A. D., Jr, and Roux, B. (2021). Optimized Lennard-Jones parameters for drug-like small molecules. *J. Chem. Theory Comput.* 14 (6), 3121–3131. doi:10.1021/acs.jctc.8b00172
- Brooks, B. R., Brucoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S., and Karplus, M. J. (1983). CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.* 4 (2), 187–217. doi:10.1002/jcc.540040211
- Burkert, U., and Allinger, N. L. (1982). *Molecular mechanics*. Washington, D. C.: American Chemical Society. doi:10.1002/jcc.540040420
- Bussi, G., Donadio, D., and Parrinello, M. (2007). Canonical sampling through velocity rescaling. *J. Chem. Phys.* 126 (1), 014101–014107. doi:10.1063/1.2408420
- Carracedo-Reboredo Jose, P., Nereida, L. B., Francisco, R. F., Francisco, C., Adrian, J. N., Victor, C., et al. (2021). A review on machine learning approaches and trends in drug discovery. *Comput. Struct. Biotechnol. J.* 19, 4538–4558. doi:10.1016/j.csbj.2021.08.011
- Chandra Singh, U., and Kollman, P. A. (1984). An approach to computing electrostatic charges for molecules. *J. Comput. Chem.* 5 (2), 129–145. doi:10.1002/jcc.540050204
- Cho, A. E., Guallar, V., Berne, B. J., and Friesner, R. (2005). Importance of accurate charges in molecular docking: Quantum mechanical/molecular mechanical (QM/MM) approach. *J. Comput. Chem.* 26 (9), 915–931. doi:10.1002/jcc.20222
- Clark, M., III Cramer, R. D., and van Opdenbosch, N. (1989). Validation of the general purpose tripos 5.2 force field. *J. Comput. Chem.* 10 (8), 982–1012. doi:10.1002/jcc.540100804
- Cornell, W. D., Cieplak, P., Bayly, C. I., Gould, I. R., Merz, K. M., Jr., Ferguson, D. M., et al. (1995). A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.* 117 (19), 5179–5197. doi:10.1021/ja00124a002
- Davies, M., Nowotka, M., Papadatos, G., Dedman, N., Gaulton, A., Atkinson, F., et al. (2015). ChEMBL web services: Streamlining access to drug discovery data and utilities. *Nucleic Acids Res.* 43, W612–W620. doi:10.1093/nar/gkv352
- Diederik, P. K., and Jimmy, B. (2015). Adam: A method for stochastic optimization. doi:10.48550/ARXIV.1412.6980
- Dodda, L. S., Vilseck, J. Z., Tirado-Rives, J., and Jorgensen, W. L. (2017). 1.14*CM1A-LBCC: Localized bond-charge corrected CM1A charges for condensed-phase simulations. *J. Phys. Chem. B* 121 (15), 3864–3870. doi:10.1021/acs.jpcc.7b00272
- 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 (19), 8577–8593. doi:10.1063/1.470117
- 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*. Wallingford CT: Gaussian, Inc.
- Gadhe, C. G., Madhavan, T., Kothandan, G., Lee, T. B., Lee, K., and Cho, S. J. (2011). Various partial charge schemes on 3D-QSAR models for P-gp inhibiting adamantyl derivatives. *Bull. Korean Chem. Soc.* 32 (5), 1604–1612. doi:10.5012/bkcs.2011.32.5.1604
- Galindo-Murillo, R., Robertson, J. C., Zgarbovic, M., Sponer, J., Otyepka, M., Jureska, P., et al. (2016). Assessing the current state of amber force field modifications for DNA. *J. Chem. Theory Comput.* 12 (8), 4114–4127. doi:10.1021/acs.jctc.6b00186
- Halgren, T. A. (1996). Merck molecular force field. I. Basis, form, scope, parameterization, and performance of MMFF94. *J. Comput. Chem.* 17 (5), 490–519. doi:10.1002/(sici)1096-987x(199604)17:5<490::aid-jcc1>3.0.co;2-p(199604)17
- Hess, B., Bekker, H., Bendersen, H. J. C., and Fraaije, J. G. E. M. (1997). Lincs: A linear constraint solver for molecular simulations. *J. Comput. Chem.* 18 (12), 1463–1472. doi:10.1002/(sici)1096-987x(199709)18:12<1463::aid-jcc4>3.0.co;2-h
- Hess, B., Kutzner, C., van der Spoel, D., and Lindahl, E. (2008). Gromacs 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theory Comput.* 4 (3), 435–447. doi:10.1021/ct700301q
- Honig, B., and Nicholls, A. (1995). Classical electrostatics in biology and chemistry. *Science* 268 (5214), 1144–1149. doi:10.1126/science.7761829
- Hwang, M. J., Stockfish, T. P., and Halgler, A. T. (1994). Derivation of class II force fields. 2. Derivation and characterization of a class II force field, CFF93, for the alkyl functional group and alkane molecules. *J. Am. Chem. Soc.* 116 (6), 2515–2525. doi:10.1021/ja00085a036
- 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 (2), 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 (16), 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 (2), 926–935. doi:10.1063/1.445869
- Jorgensen, W. L. (2009). Efficient drug lead discovery and optimization. *Acc. Chem. Res.* 42 (6), 724–733. doi:10.1021/ar800236t

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.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmolb.2022.1002535/full#supplementary-material>

- Jorgensen, W. L., Maxwell, D. S., and Tirado-Rives, J. (1996). Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *J. Am. Chem. Soc.* 118 (45), 11225–11236. doi:10.1021/ja9621760
- Jorgensen, W. L., and Tirado-Rives, J. (2005). Molecular modeling of organic and biomolecular systems using BOSS and MCPRO. *J. Comput. Chem.* 26 (16), 1689–1700. doi:10.1002/jcc.20297
- Jorgensen, W. L., and Tirado-Rives, J. (1988). The OPLS [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin. *J. Am. Chem. Soc.* 110 (6), 1657–1666. doi:10.1021/ja00214a001
- 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
- Lifson, S., Hagler, A. T., and Dauber, P. (1979). Consistent force field studies of intermolecular forces in hydrogen-bonded crystals. 1. Carboxylic acids, amides, and the C:O...H...O...H- hydrogen bonds. *J. Am. Chem. Soc.* 101 (18), 5111–5121. doi:10.1021/ja00512a001
- Lindahl, E., Hess, B., and van der Spoel, D. (2001). Gromacs 3.0: A package for molecular simulation and trajectory analysis. *J. Mol. Model.* 7, 306–317. doi:10.1007/s008940100045
- Lipinski, C. F., Maltarollo, V. G., Oliveira, P. R., da Silva, A. B. F., and Honorio, K. M. (2019). Advances and perspectives in applying deep learning for drug design and discovery. *Front. Robot. AI* 6, 108. doi:10.3389/frobt.2019.00108
- MacKerell, A. D., Jr., Bashford, D., Bellott, M., Dunbrack, R. L., Evanseck, J. D., Field, M. J., et al. (1998). All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* 102 (18), 3586–3616. doi:10.1021/jp973084f
- Maier, J. A., Martinez, C., Kasavajhala, K., Wickstrom, L., Hauser, K. E., and Simmerling, C. (2015). ff14SB: Improving the accuracy of protein side chain and backbone parameters from ff99SB. *J. Chem. Theory Comput.* 11 (8), 3696–3713. doi:10.1021/acs.jctc.5b00255
- Marco De, V., Matteo, M., Giovanni, B., and Andrea, C. (2016). Role of molecular dynamics and related methods in drug discovery. *J. Med. Chem.* 59 (4), 4035–4061. doi:10.1021/acs.jmedchem.5b01684
- Marenich, A. V., Jerome, S. V., Cramer, C. J., and Truhlar, D. G. (2012). Charge model 5: An extension of Hirshfeld population analysis for the accurate description of molecular interactions in gaseous and condensed phases. *J. Chem. Theory Comput.* 8 (2), 527–541. doi:10.1021/ct200866d
- Mayo, S. L., Olafson, B. D., and Goddard, W. A., III (1990). DREIDING A generic force field for molecular simulations. *J. Phys. Chem.* 94 (26), 8897–8909. doi:10.1021/j100389a010
- Mittal, R. R., Harris, L., McKinnon, R. A., and Sorch, M. J. (2009). Partial charge calculation method affects CoMFA QSAR prediction accuracy. *J. Chem. Inf. Model.* 49 (3), 704–709. doi:10.1021/ci800390m
- Momany, F. M., and Rone, R. (1992). Validation of the general purpose QUANTA 3.2/CHARMM force field. *J. Comput. Chem.* 13 (7), 888–900. doi:10.1002/jcc.540130714
- Nose, S., and Klein, M. L. (1983). Constant pressure molecular dynamics for molecular systems. *Mol. Phys.* 50 (5), 1055–1076. doi:10.1080/00268978300102851
- 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.* 33 (3). doi:10.1186/1758-2946-3-33
- Oostenbrink, C., Villa, A., Mark, A. E., and Van Gunsteren, W. F. (2004). A biomolecular force field based on the free enthalpy of hydration and solvation: The GROMOS force-field parameter sets 53A5 and 53A6. *J. Comput. Chem.* 25 (13), 1656–1676. doi:10.1002/jcc.20090
- Parrinello, M., and Rahman, A. (1981). Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* 52 (12), 7182–7190. doi:10.1063/1.328693
- 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
- Pattanaik, P., Raghunathan, S., Kalluri, T., Bhimalapuram, P., Jawahar, C. V., and Priyakumar, U. D. (2020). Machine learning for accurate force calculations in molecular dynamics simulations. *J. Phys. Chem. A* 124 (34), 6954–6967. doi:10.1021/acs.jpca.0c03926
- 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. doi:10.48550/arXiv.1201.0490
- Ponder, J. W., and Case, D. A. (2003). Force fields for protein simulations. *Adv. Protein Chem.* 66, 27–85. doi:10.1016/s0065-3233(03)66002-x
- Rappé, A. K., Casewit, C. J., Colwell, K. S., Goddard, W. A., III, and Skiff, W. M. (1992). UFF, a full periodic table force field for molecular mechanics and molecular dynamics simulations. *J. Am. Chem. Soc.* 114 (25), 10024–10035. doi:10.1021/ja00051a040
- Roman, M., and Dominik, H. (2019). ContraDRG: Automatic partial charge prediction by machine learning. *Front. Genet.* 10, 990. doi:10.3389/fgene.2019.00990
- Rupakheti, C., MacKerell, A. D., Jr., and Roux, B. (2018). Global optimization of the Lennard-Jones parameters for the drude polarizable force field. *J. Chem. Theory Comput.* 17 (11), 7085–7095. doi:10.1021/acs.jctc.1c00664
- Sastry, G. M., Adzhigirey, M., Day, T., Annabhimoju, R., and Sherman, W. (2013). Protein and ligand preparation: Parameters, protocols, and influence on virtual screening enrichments. *J. Comput. Aided. Mol. Des.* 27, 221–234. doi:10.1007/s10822-013-9644-8
- Schüttelkopf, A. W., and Van Aalten, D. M. F. (2004). ProDRG: A tool for high-throughput crystallography of protein–ligand complexes. *Acta Crystallogr. D. Biol. Crystallogr.* 60 (8), 1355–1363. doi:10.1107/S0907444904011679
- Shivakumar, D., Deng, Y., and Roux, B. (2009). Computations of absolute solvation free energies of small molecules using explicit and implicit solvent model. *J. Chem. Theory Comput.* 5 (4), 919–930. doi:10.1021/ct800445x
- Stanke, M., and Morgenstern, B. (2005). Augustus: A web server for gene prediction in eukaryotes that allows user-defined constraints. *Nucleic Acids Res.* 33, 465–467. doi:10.1093/nar/gki458
- Storer, J., Giesen, D., Cramer, C., and Truhlar, D. (1995). Class IV charge models: A new semiempirical approach in quantum chemistry. *J. Comput. Aided. Mol. Des.* 9 (1), 87–110. doi:10.1007/BF00117280
- Stroet, M., Caron, B., Visscher, K. M., Geerke, D. P., Malde, A. K., and Mark, A. E. (2018). Automated topology builder version 3.0: Prediction of solvation free enthalpies in water and hexane. *J. Chem. Theory Comput.* 14 (11), 5834–5845. doi:10.1021/acs.jctc.8b00768
- Swope, W. C., Horn, H. W., and Rice, J. E. (2010). Accounting for polarization cost when using fixed charge force fields. II. Method and application for computing effect of polarization cost on free energy of hydration. *J. Phys. Chem. B* 114 (26), 8631–8645. doi:10.1021/jp911701h
- Thompson, J. D., Cramer, C. J., and Truhlar, D. G. (2003). Parameterization of charge model 3 for AM1, PM3, BLYP, and B3LYP. *J. Comput. Chem.* 24 (11), 1291–1304. doi:10.1002/jcc.10244
- Tian, C., Kasavajhala, K., Belfon, K. A. A., Raguette, L., Huang, H., Migues, J., et al. (2019). ff19SB: Amino-Acid-Specific protein backbone parameters trained against quantum mechanics energy surfaces in solution. *J. Chem. Theory Comput.* 16 (1), 528–552. doi:10.1021/acs.jctc.9b00591
- Udier-Blagovic, M., Morales De Tirado, P., Pearlman, S. A., and Jorgensen, W. L. (2004). Accuracy of free energies of hydration using CM1 and CM3 atomic charges. *J. Comput. Chem.* 25 (11), 1322–1332. doi:10.1002/jcc.20059
- Vanommeslaeghe, K., Hatcher, E., Acharya, C., Kundu, S., Zhong, S., Shim, J., et al. (2010). 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 (4), 671–690. doi:10.1002/jcc.21367
- Vanommeslaeghe, K., and MacKerell, A. D., Jr. (2015). CHARMM additive and polarizable force fields for biophysics and computer-aided drug design. *Biochim. Biophys. Acta* 1850 (5), 861–871. doi:10.1016/j.bbagen.2014.08.004
- Verstraelen, T. (2019). Molmod software library. Available at: <http://molmod.ugent.be/software>. (Accessed January 27, 2021).
- Vilseck, J. Z., Tirado-Rives, J., and Jorgensen, W. L. (2014). Evaluation of CM5 charges for condensed-phase modeling. *J. Chem. Theory Comput.* 10 (7), 2802–2812. doi:10.1021/ct500016d
- 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 (12), 1049–1074. doi:10.1002/1096-987x(200009)21:12<1049::aid-jcc3>3.0.co;2-f
- Wang, J., Wang, W., Kollman, P. A., and Case, D. A. (2006). Automatic atom type and bond type perception in molecular mechanical calculations. *J. Mol. Graph. Model.* 25 (2), 247–260. doi:10.1016/j.jmgm.2005.12.005
- Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A., and Case, D. A. (2005). Development and testing of a general amber force field. *J. Comput. Chem.* 25 (9), 1157–1174. doi:10.1002/jcc.20035
- Weiner, S. J., Kollman, P. A., Case, D. A., Singh, U. C., Ghio, C., Alagona, G., et al. (1984). A new force field for molecular mechanical simulation of nucleic acids and proteins. *J. Am. Chem. Soc.* 106 (3), 765–784. doi:10.1021/ja00315a051
- Wu, Y., and Wang, G. (2018). Machine learning based toxicity prediction: From chemical structural description to transcriptome analysis. *Int. J. Mol. Sci.* 19 (8), 2358. doi:10.3390/ijms19082358
- Zgarbova, M., Otyepka, M., Sponer, J., Mladek, A., Banas, P., Cheatham, T. E., et al. (2011). Refinement of the Cornell et al. Nucleic Acids Force Field Based on Reference Quantum Chemical Calculations of Glycosidic Torsion Profiles. *J. Chem. Theory Comput.* 7 (9), 2886–2902. doi:10.1021/ct200162x



OPEN ACCESS

EDITED BY
Chandrabose Selvaraj,
Alagappa University, India

REVIEWED BY
Mattia Laffranchi,
Sapienza University of Rome, Italy
George Priya Doss C.,
VIT University, India
Sher Zaman Safi,
Mahsa University, Malaysia

*CORRESPONDENCE
Babajan Banaganapalli,
bbabajan@kau.edu.sa

SPECIALTY SECTION
This article was submitted to Biophysics,
a section of the journal
Frontiers in Molecular Biosciences

RECEIVED 22 September 2022
ACCEPTED 02 November 2022
PUBLISHED 24 November 2022

CITATION
Shaik NA, Saud Al-Saud NB,
Abdulhamid Aljuhani T, Jamil K,
Alnuman H, Aljeaid D, Sultana N,
El-Harouni AA, Awan ZA, Elango R and
Banaganapalli B (2022), Structural
characterization and conformational
dynamics of alpha-1 antitrypsin
pathogenic variants causing alpha-1-
antitrypsin deficiency.
Front. Mol. Biosci. 9:1051511.
doi: 10.3389/fmolb.2022.1051511

COPYRIGHT
© 2022 Shaik, Saud Al-Saud,
Abdulhamid Aljuhani, Jamil, Alnuman,
Aljeaid, Sultana, El-Harouni, Awan,
Elango and Banaganapalli. This is an
open-access article distributed under
the terms of the [Creative Commons
Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use,
distribution or reproduction in other
forums is permitted, provided the
original author(s) and the copyright
owner(s) are credited and that the
original publication in this journal is
cited, in accordance with accepted
academic practice. No use, distribution
or reproduction is permitted which does
not comply with these terms.

Structural characterization and conformational dynamics of alpha-1 antitrypsin pathogenic variants causing alpha-1-antitrypsin deficiency

Noor Ahmad Shaik^{1,2}, Najla Bint Saud Al-Saud³,
Thamer Abdulhamid Aljuhani¹, Kaiser Jamil⁴, Huda Alnuman²,
Deema Aljeaid¹, Nasreen Sultana⁵,
Ashraf AbdulRahman El-Harouni¹, Zuhier Ahmed Awan^{6,7},
Ramu Elango^{1,2} and Babajan Banaganapalli^{1,2*}

¹Department of Genetic Medicine, Faculty of Medicine, King Abdulaziz University, Jeddah, Saudi Arabia, ²Princess Al-Jawhara Al-Brahim Center of Excellence in Research of Hereditary Disorders, King Abdulaziz University, Jeddah, Saudi Arabia, ³Department of Biological Sciences, Faculty of Science, King Abdulaziz University, Jeddah, Saudi Arabia, ⁴Department of Genetics, Bhagwan Mahavir Medical Research Centre, Hyderabad, India, ⁵Department of Biotechnology, Acharya Nagarjuna University, Guntur, India, ⁶Department of Clinical Biochemistry, Faculty of Medicine, King Abdulaziz University, Jeddah, Saudi Arabia, ⁷Department of Genetics, Al Borg Medical Laboratories, Jeddah, Saudi Arabia

Background: Alpha-1 antitrypsin deficiency (A1ATD) is a progressive lung disease caused by inherited pathogenic variants in the *SERPINA1* gene. However, their actual role in maintenance of structural and functional characteristics of the corresponding α -1 anti-trypsin (A1AT) protein is not well characterized.

Methods: The A1ATD causative *SERPINA1* missense variants were initially collected from variant databases, and they were filtered based on their pathogenicity potential. Then, the tertiary protein models were constructed and the impact of individual variants on secondary structure, stability, protein-protein interactions, and molecular dynamic (MD) features of the A1AT protein was studied using diverse computational methods.

Results: We identified that A1ATD linked *SERPINA1* missense variants like F76S, S77F, L278P, E288V, G216C, and H358R are highly deleterious as per the consensual prediction scores of SIFT, PolyPhen, FATHMM, M-CAP and REVEL computational methods. All these variants were predicted to alter free energy dynamics and destabilize the A1AT protein. These variants were seen to cause minor structural drifts at residue level (RMSD = <2Å) of the protein. Interestingly, S77F and L278P variants subtly alter the size of secondary structural elements like beta pleated sheets and loops. The residue level fluctuations at 100 ns simulation confirm the highly damaging structural consequences of all the six missense variants on the conformation dynamics of the A1AT protein. Moreover, these variants were also predicted to cause functional deformities by negatively impacting the binding energy of A1AT protein with NE ligand molecule.

Conclusion: This study adds a new computational biology dimension to interpret the genotype-protein phenotype relationship between *SERPINA1* pathogenic variants with its structural plasticity and functional behavior with NE ligand molecule contributing to the Alpha-1-antitrypsin deficiency. Our results support that A1ATD complications correlates with the conformational flexibility and its propensity of A1AT protein polymerization when misfolded.

KEYWORDS

SERPINA1 gene, alpha-1-antitrypsin, serpinopathies, molecular dynamics, AATD

Introduction

A1ATD is a rare autosomal recessive disease in which low levels of circulating alpha-1 antitrypsin enzyme in the plasma promote degenerative or destructive changes in the lung (Ferrarotti et al., 2018). A1AT protein is a serine protease inhibitor produced in the liver. This enzyme binds to different enzymes, including neutrophil elastase, that cleaves them and gets cleaved by them in a suicidal fashion (Visser et al., 1988). A proportion of A1ATD patients develop liver cirrhosis, which may be caused by aggregates of alpha-1-antitrypsin proteins (Köhnlein and Welte, 2008). While it is often undiagnosed (Quinn et al., 2020), it causes emphysema, which can be exacerbated by tobacco smoke (Tejwani and Stoller, 2021). It affects 1 in 2,500 people of European ancestry (Greulich et al., 2017).

The A1ATD is caused by different pathogenic variants in the *SERPINA1* gene, which is located on the long arm (q) of chromosome 14 at 32.1 and consists of 7 exons. It encodes a 394 amino acid long polypeptide, which acts as a molecular mouse trap that binds and blocks the function of a variety of proteases. The amino acid substitutions in A1AT may also alter the cellular function by forming polymer aggregates of the protein, causing liver and lung damage (Duvoix et al., 2014). Many pathogenic alleles in *SERPINA1* gene, also designated as protease inhibitor (PI), have already been described in A1ATD patients (Foil, 2021). The misfolding and aggregation of this serpin family member is thermodynamically favorable since the protein's native conformation is transient and built to be cleaved to reach stability (Cho et al., 2005).

The effects of missense variants on A1AT's structure and function are not yet well resolved. Classical *in-vivo* and *in-vitro* approaches to study the molecular characterization of pathogenic variants are time and cost-consuming. Alternative "*in silico*" approaches, owing to their sensitivity, specificity, and accuracy, act as pre-screens for laboratory studies (Shaik et al., 2020a; Shaik et al., 2020b; Awan et al., 2021). In this regard, a growing number of computational methods can effectively predict variant pathogenicity and stability, visualize their structures, map the conserved domains, compare their secondary structures with the wildtype protein, and simulate their ability to bind with a substrate. Therefore, it is aimed to utilize these computational

biology methods to study clinical missense variants to expand the knowledge on the nature of structural defects and conformational dynamics affecting the A1AT's function.

Materials and methods

SERPINA1 variant data collection and curation

The molecular details of the *SERPINA1* gene, including the nucleotide sequence, chromosome position, transcript, and the corresponding amino acid sequence, were obtained from the National Center for Bioinformatics (NCBI) (www.ncbi.nlm.nih.gov) and Ensembl databases (www.ensembl.org). The A1ATD causative variants were collected from the DisGenNET platform (<https://www.disgenet.org>), which is a webserver that contains disease-associated variants gathered from scientific literature, genome-wide association studies (GWAS) catalogues, and animal models (Piñero et al., 2021). The reported variants are downloaded in a list in Microsoft Excel format. Then, only the missense variants associated with the A1ATD phenotype were selected for further analysis, after sorting and filtering by Microsoft Excel 2019. Table 1 shows the molecular details of the selected *SERPINA1* missense variants.

Variant pathogenicity predictions and conservation analysis

The selected missense variants were uploaded into the Ensembl (www.ensembl.org) variant pathogenicity predictor (VEP) to assess their pathogenic potential. This webserver allows using multiple tools that measure whether a variant can be considered deleterious or not based on different features like whether it is located in a evolutionarily conserved sequence across species, or whether it causes structural and stability differences, etc. (Adzhubei et al., 2010). The data can be entered as a variant ID, VCF file, or nomenclature notation of HGVS coordinates. The output can be in text or html format. In this study, six tools, including Combined Annotation Dependent Depletion (CADD), Scale-invariant Feature Transform (SIFT)

TABLE 1 Molecular details of alpha-1-antitrypsin causative SERPINA1 missense variants.

#	dbSNP ID	Clinvar ID	Chromosomal position	c.DNA position	Exon	Amino acid change	Codon change	MAF
1	rs1555369172	VCV000444040.1	14:94383011-94383011	c.274T>C	2/5	F76S	tTc/tCc	—
2	rs55819880	VCV000017992.2	14:94383008-94383008	c.277C>T	2/5	S77F	tCc/tTc	—
3	rs756773408	VCV000444044.1	14:94382592-94382592	c.693G>T	2/5	G216C	Ggc/Tgc	—
4	rs1566753480	VCV000626306.1	14:94380955-94380955	c.880T>C	3/5	L278P	cTg/cCg	—
5	rs17580	VCV000626305.1	14:94380925-94380925	c.910A>T	3/5	E288V	gAa/gTa	0.0196
6	rs1555367891	RCV000512630.1	14:94378633-94378633	c.1120A>G	5/5	H358R	cAt/cGt	—

(Ng and Henikoff, 2003), Polymorphism Phenotyping (PolyPhen) (Adzhubei et al., 2013), Mendelian Clinically Applicable Pathogenicity (M-CAP) (Jagadeesh et al., 2016), and Functional Analysis through Hidden Markov Models (FATHMM) (Rogers et al., 2018), REVEL (rare exome variant ensemble learner) were used to evaluate the pathogenicity of variants (Ioannidis et al., 2016). SIFT predicts pathogenicity based on alteration in conserved regions of the nucleotide sequence (Shaik et al., 2020b; Shaik et al., 2021; Alharthi et al., 2022; Bima et al., 2022). PolyPhen predicts the variant effects based on the nucleotide sequence and changes in protein structure. CADD predicts the effects based on the integration of several parameters, including sequence context, evolutionary constraints on the genome, and epigenetic calculations. The M-CAP combines the predictions of PolyPhen, SIFT, and CADD together. FATHMM predicts the consequences based on combining sequence conservation within Hidden Markov models (HMMs) to depict the alignment of homologous sequences and conserved protein domains. REVEL is an ensemble technique for estimating the pathogenicity of missense variants utilizing the following methods: MutPred, FATHMM, VEST, PolyPhen, SIFT, PROVEAN, MutationAssessor, MutationTaster, LRT, GERP, SiPhy, phyloP, and phastCons. REVEL was trained exclusively on rare pathogenic and neutral missense variants (Ioannidis et al., 2016). To examine the amino acid sequence conservation pattern of A1AT gene across related species (8 primates), we have performed multiple sequence alignment using Clustal Omegat (<https://www.ebi.ac.uk/Tools/msa/clustalo/>).

3D structure mapping and superimposition

The PDB database contains the full-length x-ray crystallographic structure of the natural human native A1AT protein at 1.8 Å (PDB ID: 3NE4 chain A). This structure was used as a template to construct A1AT protein variants using Modeller-homology model software. The full-length amino acid sequence of A1AT retrieved from the NCBI database (accession number CAJ15161.1) was used to provide input to the Modeller tool to

construct tertiary structural models of A1AT variants. The Modeller is a readily available online tool that relies on protein NMR to meet spatial restrictions, using sets of geometrical requirements to establish atomic positions in protein models by generating probability density functions. This approach matches the input sequence with target amino acid sequences and the structure of the template protein. Using the steepest descent force field method in the GROMACS software, energy of three-dimensional mutant A1AT structure was optimized with steepest descent energy minimization method. All 3D models, including the wildtype and mutated structures, were viewed, and analyzed through Pymol2 software.

Structural deviation and secondary structure analysis of A1AT variants

The structural deviation of mutated proteins is a reliable metric to determine how an amino acid change affects the overall structure of the protein. The structural deviation of proteins is measured in form of RMSD values, which were computed with the Pymol2 software. The smaller the RMSD value, higher similarity in both structures is predicted. These RMSD values were estimated by superimposing each mutated model with the corresponding wildtype structure. To perform the secondary structure analysis, each mutated amino acid sequence was created manually in text format *via* manual amino acid substitution. Then, along with the wildtype, all mutated amino acid sequences were entered into the NetsurfP 2.0 web tool (<https://services.healthtech.dtu.dk/service.php?NetSurfP-2.0>) to generate secondary structure representations. These secondary structures were then analyzed to see if variant induced changes occurred at the secondary structural element like α -helix, β -pleated sheet or loop.

Stability analysis of A1AT variants

The missense variants were analyzed for their ability to increase or decrease the stability of the protein by using the DUET server (<http://biosig.unimelb.edu.au/duet>), which predicts

stability scores in $\Delta\Delta G$ in kcal/mol. This server uses variant cutoff scanning matrix (mCSM) and site-directed mutator (SDM) methods for estimation of a DUET score, which is calculated based on integrated scores of both the aforementioned methods (Pires et al., 2014). The input to this server can be the four-letter code of the wildtype protein, or the PDB structure, together with the amino acid change and the affected chain letter code. If the value is positive, the variant is making the structure more stable, and if it is negative, it is making the structure less stable.

Molecular dynamics simulation

The MD simulation of the wild and mutant protein models was done using the Desmond MD tool (Bowers et al., 2006; Release, 2017). The simulation systems were first prepared by applying the SPC/E solvation model to an orthorhombic box with a boundary distance of 10 Å. The system for all the models was neutralized by adding 10 Na⁺ and 0.15 M Na⁺, Cl⁻ ions using the OPLS3e forcefield. Energy minimization of the prepared solvent systems was minimized using the steepest decent method at 5,000 steps. Further, the minimized systems were equilibrated in constant temperature and volume (canonical or NVT) and constant temperature and pressure (NPT) ensembles using a “relax model system” before the simulation. In the initial steps, the energy minimized systems are simulated in the NVT ensemble with Brownian dynamics at 10K temperature for 100 ps and 12 ps with restraint on solute heavy atoms. In NPT ensemble systems, no restraints were on heavy atoms at 10 K and 300 K temperatures for 12 ps and 24 ps, respectively. The fully equilibrated systems were finally subjected to 100 ns unrestrained MD simulations in an NPT ensemble with 1.01325 pKa (pressure) and 300 K temperature. The 25 ns trajectories were recorded during the simulation period for post MD analysis.

A1AT—NE protein computational binding assay

The binding affinity between A1AT (both native and mutant forms) with NE (neutrophil elastase) proteins was analyzed using the ClusPro molecular docking online tool (<https://cluspro.org>). The input for this tool is the protein and ligand files in PDB format. The other options were set as “default.” This tool utilizes the PIPER algorithm where the center of mass of the receptor remains fixed, and the ligand molecule is rotated in a variety of positions to determine the best fit (Kozakov et al., 2013; Kozakov et al., 2017; Vajda et al., 2017; Desta et al., 2020). The resulting models were compared and only the best scores, depicted as 0 for each model, were selected for downloading and visual simulation in Pymol2. Moreover, the lowest negative energy score outputs

were recorded along with center scores for the analysis of binding energy variation.

Results

Predicting missense deleterious variants and evolutionary conservation analysis

The deleterious effects of all 6 (F76S, S77F, L278P, E288V, G216C, and H358R) missense variants were measured with various “*in silico*” prediction tools; SIFT, PolyPhen, CADD, FATHMM M-CAP, and REVEL in the Variant Effect Predictor (VEP) web server from Ensembl. One of the variants (L278P) was predicted to be pathogenic by all tools except CADD with a C-score of 29.1, which was close to the cutoff value of 30 (Table 2). The pattern of amino acid sequence conservation implies that all A1AT variants were mapped to the evolutionarily conserved region (F76S, S77F, L278P, E288V, G216C, and H358R) (Supplementary Figure S1). A1AT in humans has a relatively close phylogenetic link with bonobos and chimpanzees, but it is distinct from mouse lemurs, based on the phylogenetic relationships among 12 species of primates that are closely related to one another.

3D modeling and stereochemical quality assessment of wildtype protein

The crystal structure of A1AT is in metastable native fold form and consists of three sheets, nine helices, and a reactive center loop held at the apex of the protein. The amino acids 357 to 359 allowed the RCL region extend β -stand conformation (stressed external loop) and stabilize the structure by forming slat bridges between P5 Glu and Arg 196, 223, and 281. Upon 3D modeling of the wildtype protein in Pymol2, the resultant structure was subjected to energy minimization to remove bad physical configurations. This was achieved using the Modeller tool. The energy minimization output was shown to be -17874.756 kJ/mol, which shows the successful removal of unwanted bonding patterns from the 3D model (Figure 1A). SAVES v.6.0 was used in the analysis of protein structure stereochemical quality assessment through a Ramachandran plot (Figure 1B), which revealed that a small number of amino acids have their ϕ (phi), ψ (psi) angles in the non-core areas of the protein. The percentage of amino acid deposits in the center and non-center areas of the protein is 90.9%–8.8%, respectively. The hydrogen bond estimation (DSSP) algorithm also revealed the output for hydrogen bonding as a graphical representation (Figure 1C), which demonstrates the good quality of the built protein model. The average DSSP score falls between 0 and 1, indicating a good quality protein structure.

TABLE 2 The pathogenicity prediction output of different computational tools of the SERPINA1 missense variants.

#	Amino acid variant	SIFT	PolyPhen	CADD	FATHMM	M-CAP	REVEL
1	F76S	0.00	1.000	32.0	0.86815	0.75948	0.98304
2	S77F	0.00	1.000	31.0	0.95561	0.82026	0.99571
3	G216C	0.00	1.000	35.0	0.91589	0.83675	0.9706
4	L278P	0.00	0.999	29.1	0.90962	0.88091	0.96096
5	E288V	0.00	0.996	32.0	0.89496	—	0.8882
6	H358R	0.00	0.990	26.8	0.92158	0.80576	0.94428

SIFT: < 0.01 = damaging, PolyPhen: > 0.5 = damaging, CADD: > 30 damaging, FATHMM >0.5 = damaging, and M-CAP > 0.5 = damaging; REVEL >0.5 = damaging.

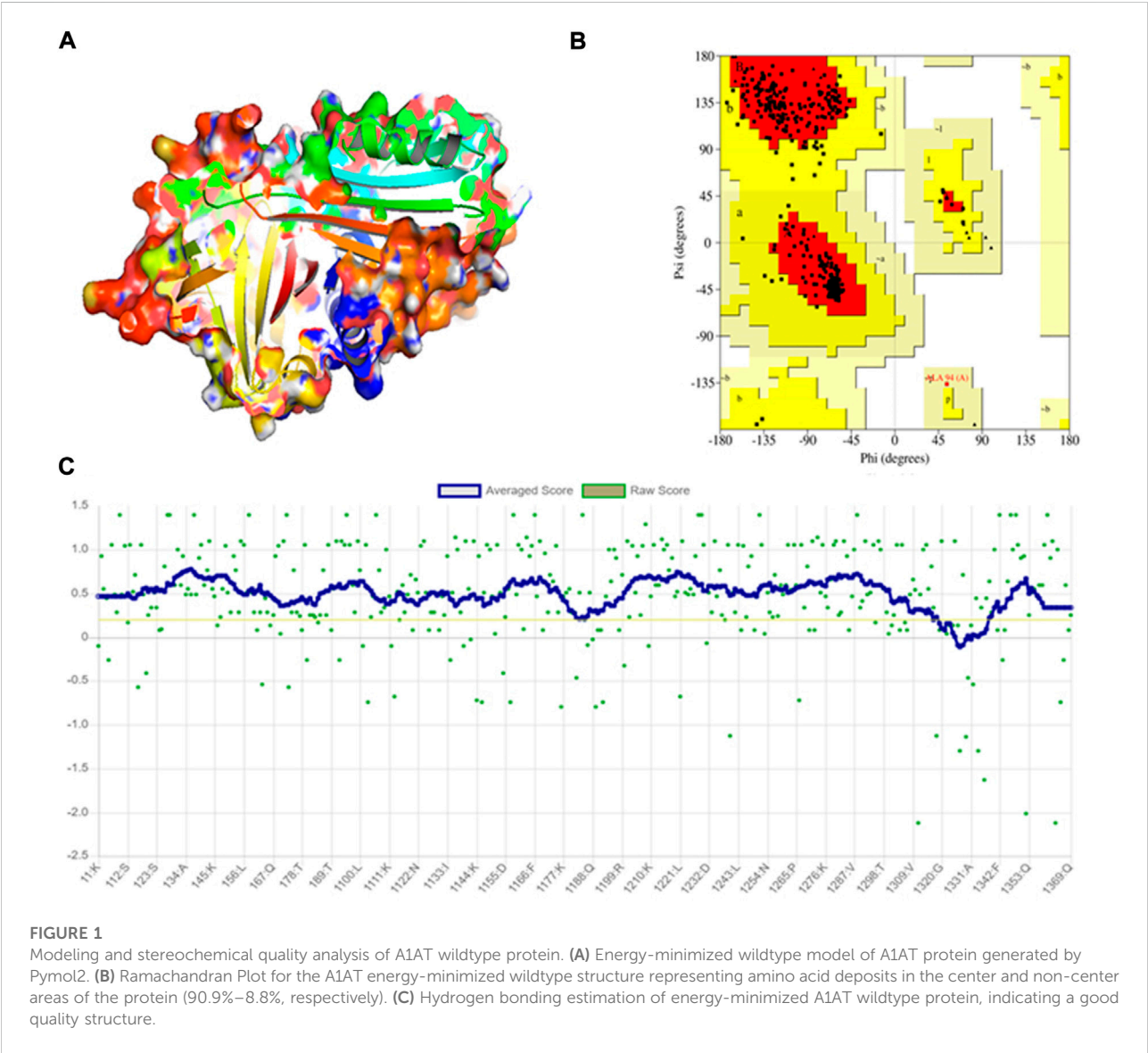
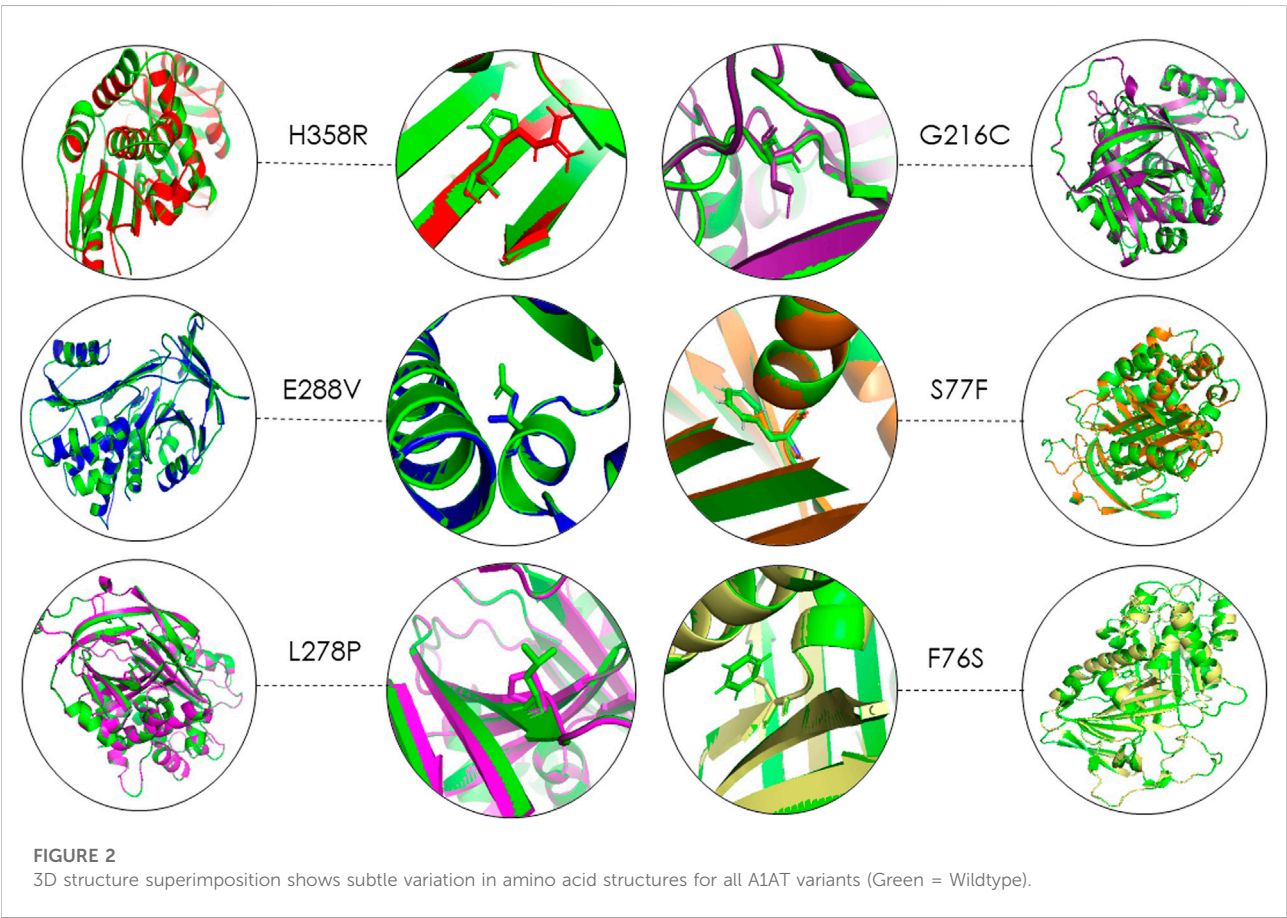


TABLE 3 3D structural deviation of mutated A1AT protein structures versus wildtype represented in the form of RMSD values.

#	dbSNP ID	Amino acid variant	RMSD values (Å)	
			Whole structure	Amino acid
1	rs1555369172	F76S	0.038	2.4
2	rs55819880	S77F	0.028	2.9
3	rs756773408	G216C	0.029	3.1
4	rs1566753480	L278P	0.039	2.5
5	rs17580	E288V	0.049	2.8
6	rs1555367891	H358R	0.069	2.9



Structural deviation analysis of A1AT variants

The stereochemical analysis of the energy minimized A1AT mutant protein models (F76S, S77F, L278P, E288V, G216C, and H358R) showed that approximately 99.7% of the amino acids fall in the allowed region and only 0.03% in the disallowed region (Figure 1B). Moreover, it displayed good overall structural

quality through the Procheck tool (SAVES v6.0 package) (Figure 1C).

Using 3D structure imposition between the folded wildtype and mutant A1AT proteins, the C-atom structural coordinates were estimated in the form of RMSD scores by rotating them in three-dimensional space. No significant structural differences between the wildtype and all six mutant models of A1AT protein was observed at whole protein structure level as their

TABLE 4 Thermodynamic stability analysis of SERPINA1 missense variants.

#	dbSNP ID	Nucleotide substitution	Exon	Amino acid variant	mCSM ($\Delta\Delta G$)	SDM ($\Delta\Delta G$)	DUET ($\Delta\Delta G$)
1	rs1555369172	c.274T>C	2/5	F76S	-2.9 kcal/mol (Destabilizing)	-3.57 kcal/mol (Destabilizing)	-3.199 kcal/mol (Destabilizing)
2	rs55819880	c.277C>T	2/5	S77F	-0.584 kcal/mol (Destabilizing)	0.74 kcal/mol (Stabilizing)	-0.419 kcal/mol (Destabilizing)
3	rs756773408	c.693G>T	2/5	G216C	-0.415 kcal/mol (Destabilizing)	-1.12 kcal/mol (Destabilizing)	-0.434 kcal/mol (Destabilizing)
4	rs1566753480	c.880T>C	3/5	L278P	-1.571 kcal/mol (Destabilizing)	-2.58 kcal/mol (Destabilizing)	-1.93 kcal/mol (Destabilizing)
5	rs17580	c.910A>T	3/5	E288V	0.909 kcal/mol (Stabilizing)	0.55 kcal/mol (Stabilizing)	1.333 kcal/mol (Stabilizing)
6	rs1555367891	c.1120A>G	5/5	H358R	-1.74 kcal/mol (Destabilizing)	-1.69 kcal/mol (Destabilizing)	-1.842 kcal/mol (Destabilizing)

RMSD scores fell less than 0.2 Å (Table 3). The RMSD values for F76S, S77F, L278P, E288V, G216C, and H358R are 0.038 Å, 0.028 Å, 0.029 Å, 0.039 Å, 0.049 Å, and 0.069 Å, respectively. Overall, the superimposition analysis of A1AT protein demonstrated that all six missense variants cause subtle structural changes in A1AT at the whole protein level (Figure 2). However, all mutant models show >2Å structural deviations at the residue level.

A1AT stability and secondary structure analysis

The function of the disease candidate protein will be affected by missense variants that negatively affect thermodynamic stability. Stability changes of A1AT mutated structures were analyzed by different tools to measure energy changes. The mCSM predicted F76S (-2.9 kcal/mol), S77F (-0.584 kcal/mol), L278P (-1.571 kcal/mol), G216C (-0.415 kcal/mol), and H358R (-1.74 kcal/mol) variants (5/6; 83.3%) are destabilizing the protein due to their negative free energy ($\Delta\Delta G$) values. SDM also predicted F76S (-3.57 kcal/mol), G216C (-1.12 kcal/mol), L278P (-2.58 kcal/mol), and H358R (-1.69 kcal/mol) variants as destabilizing owing to their free energy values.

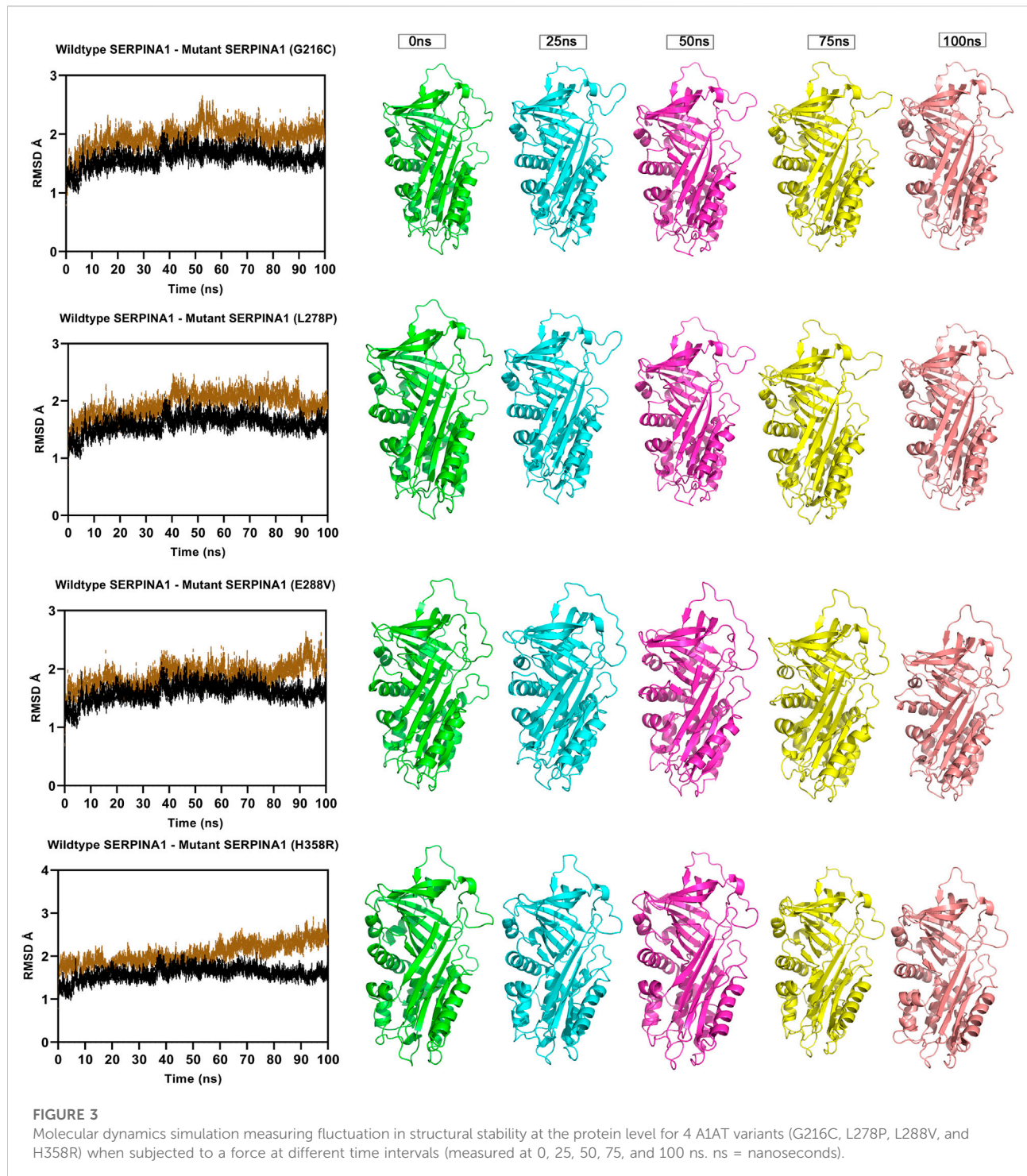
The DUET tool combines the output of both the mCSM and SDM tools to generate a consensual prediction. DUET webserver supported that 5/6 variants (83.3%) are destabilizing the A1AT structure because of their negative free energy values, i.e., F76S (-3.199 kcal/mol), S77F (-0.419 kcal/mol), G216C (-0.434 kcal/mol), L278P (-1.93 kcal/mol), and H358R (-1.842 kcal/mol). The destabilization data predicted that most of these variants are pathogenic as they disrupt the folding of the A1AT protein. Because of positive free energy values, E288V is, predicted by all the three tools (mCSM = 0.909 kcal/mol, SDM = 0.55 kcal/mol, and DUET = 1.333 kcal/mol), to further stabilize the A1AT protein structure (Table 4).

The secondary structure analysis provides information on how the mutant amino acid residue affects the substructures (alpha helices, loops, and beta pleated sheets) in the protein. The output of the secondary structure analysis predicted 2/6 (33%) of variants show alterations in the secondary structure of A1AT. This includes L278P variant, which shortens the beta pleated sheet (Figure 3) and the S77F variant that increases the protein length as this variant is located at the junction with a loop region (Figure 3).

MD simulation analysis

The Root Mean Square Deviation (RMSD), Root Mean Square Fluctuations (RMSF), and Secondary Structure Elements (SSE) were analyzed in all the six alpha-1 anti-trypsin models at 100 ns. Figure 4 represents the C-alpha RMSDs of wildtype and mutant alpha-1 anti-trypsin proteins for 100 ns. The RMSD analysis shows that four mutant structures (G216C, L278C, E288V, and H358R) significantly fluctuated compared to the wild-type alpha-1 anti-trypsin model. The wildtype model RMSD started at 1.4 Å at 0 ns and reached equilibrium at 50 ns, maintaining the RMSD range between 1.5 Å and 1.8 Å. Whereas the RMSD of the four mutant models increases steadily up to 10 ns, after 50 ns, it abruptly decreases from 0.2 Å to 0.42 Å and then increases to 1.8 Å at 70 ns, before fluctuating at 1.6 Å at the end of the simulation. In the H358R mutant model, a clear deviation was observed in the domain region, whereas in the other models, the helix 12 region showed more deviation compared to other secondary structural elements in the protein models. The mutations G216C, L278C, E288V, and H358R are localized near the RCL region of A1AT. After 50 ns, the β -sheets were opened (at RCL), allowing an increase in the stability of the motifs in the protein. We also discovered that the RCL region is stressed to a relaxing state during the simulation period.

To understand the flexible nature of mutated proteins, we performed RMSF analysis at 100 ns. The RMSF value over the entire simulation of six mutant models (F76S, S77F, G216C,



L28P, E288V, and H358R) showed significant fluctuations. When compared to the wildtype model, the mutant models S77F, G216C, L278P, and 288V exhibit more fluctuations ($>0.2 \text{ \AA}$), whereas the mutant models of F76S and H358R exhibit more rigidity ($>0.1 \text{ \AA}$). The higher RMSF values of mutated models of S77F, G216C, L28P, and E288V support the calculated RMSD values. The secondary

structure analysis was performed on mutated models of the alpha-1 anti-trypsin protein. Of the 6 mutated models, two models show a significant alteration in secondary structural elements. The mutated models S77F and G216C show $>1\%$ alterations in the secondary structural elements, whereas the remaining four models show lesser alterations in their β -strands (Figure 5).

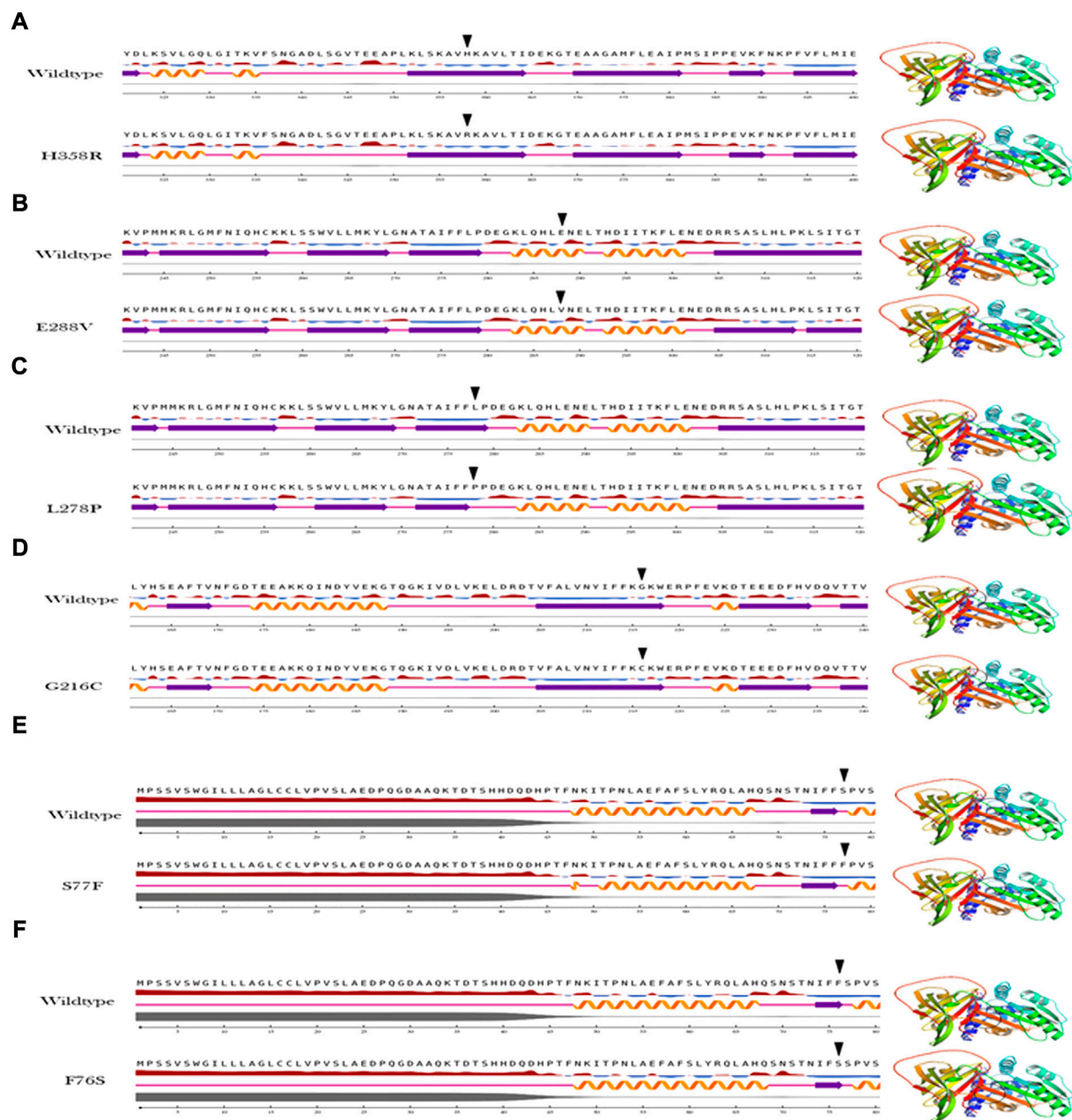


FIGURE 4
Secondary structure analysis output of the A1AT missense variants (A–F) in comparison with wildtype structure.

Molecular docking of A1AT-NE complex

ClusPro docking generated the best docking complexes of A1AT (receptor) and NE (ligand) with similar polarity and orientation based on their high-resolution models fitted to the electron microscopy density volumes above. The ClusPro software calculated the best docking pose based on highly populated clusters of low-energy models. Further, the best fit

pose for each receptor-ligand complex was identified by the PIPER algorithm based on electrostatic and van der Waals scores (Figure 6). A1AT-NE mutated complexes have shown significant alterations in lowest central energy score (>9.1 kcal/mol), compared to the wildtype-NE complex (Table 5). Therefore, major differences in binding configuration for all the six variants are predicted by this analysis. In the wildtype A1AT protein complex, the protein-

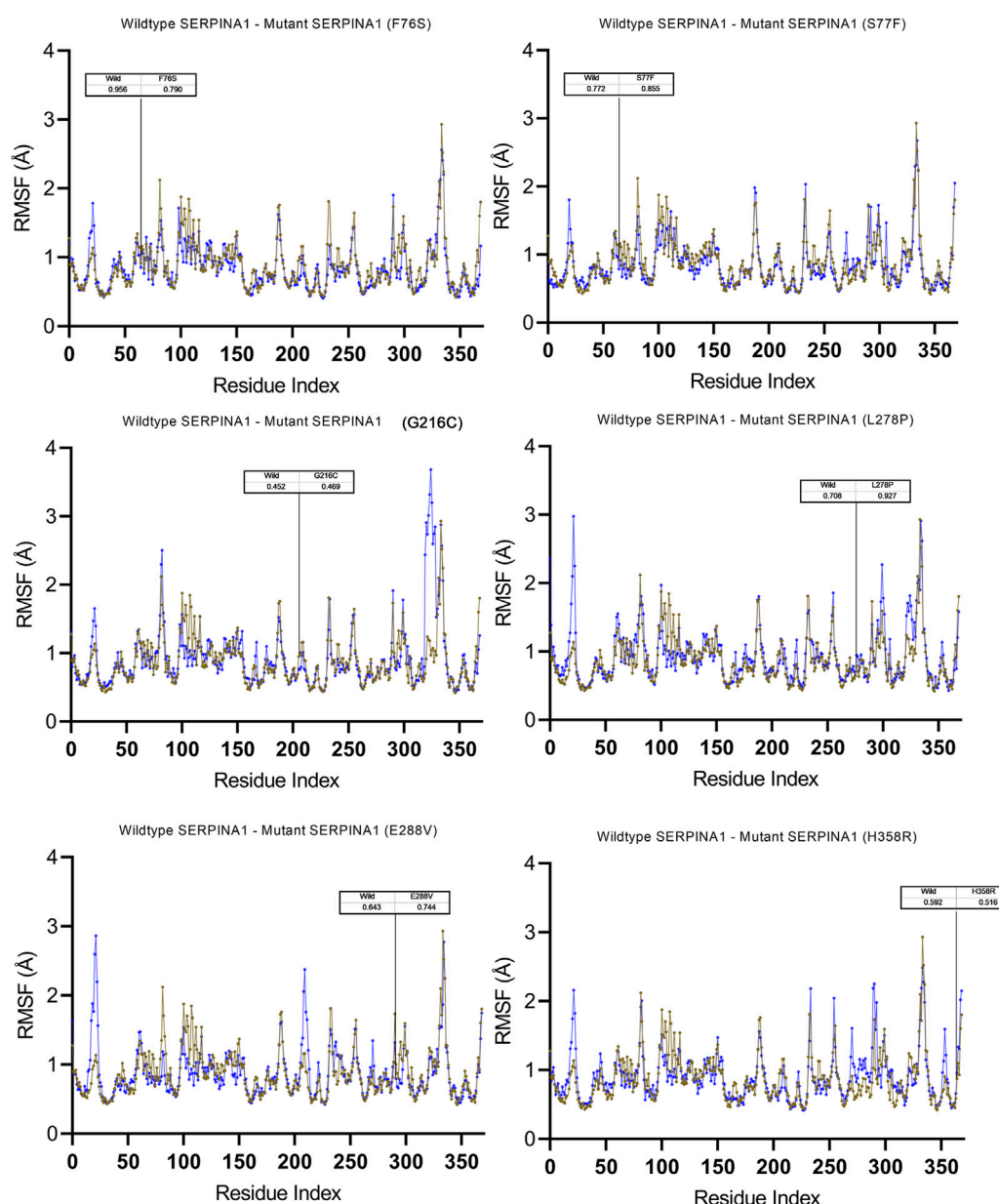


FIGURE 5

Molecular dynamics simulation measuring fluctuation in structural stability at the amino acid level for 6 A1AT variants (F76S, S77F, G216C, L278P, E288V, and H358R) after being subjected to a force.

protein interaction had the lowest free energy score of -914.9 kcal/mol, whereas the lowest free energy between F76S-NE complex is -941.2 kcal/mol, in S77F-NE is -944.6 kcal/mol, in G216C-NE is -941.5 kcal/mol, in L278P-NE is -945.4 kcal/mol, in E288V-NE is -941.1 kcal/mol, and in H358R-NE is -931.6 kcal/mol. The A1AT protein normally functions as a molecular mouse trap by having high affinity to its substrate and eliminates its target molecule in a suicidal fashion (Berclaz and Trapnell, 2006). Under A1AT

mutant conditions, differences in binding affinities with NE were evident. The NE interacts at RCL loop in both wild and mutant state. However, the mutation at AA348 created an expansion of B-sheet and allowed perturbation of the Helix in the A1AT structure. This will allow the formation of protein-inhibitor covalent like complex. Other mutations are not having any direct effect on the NE- A1AT inhibition. As a result of these findings, it is assumed that all six mutant models will tightly bind to their substrate,

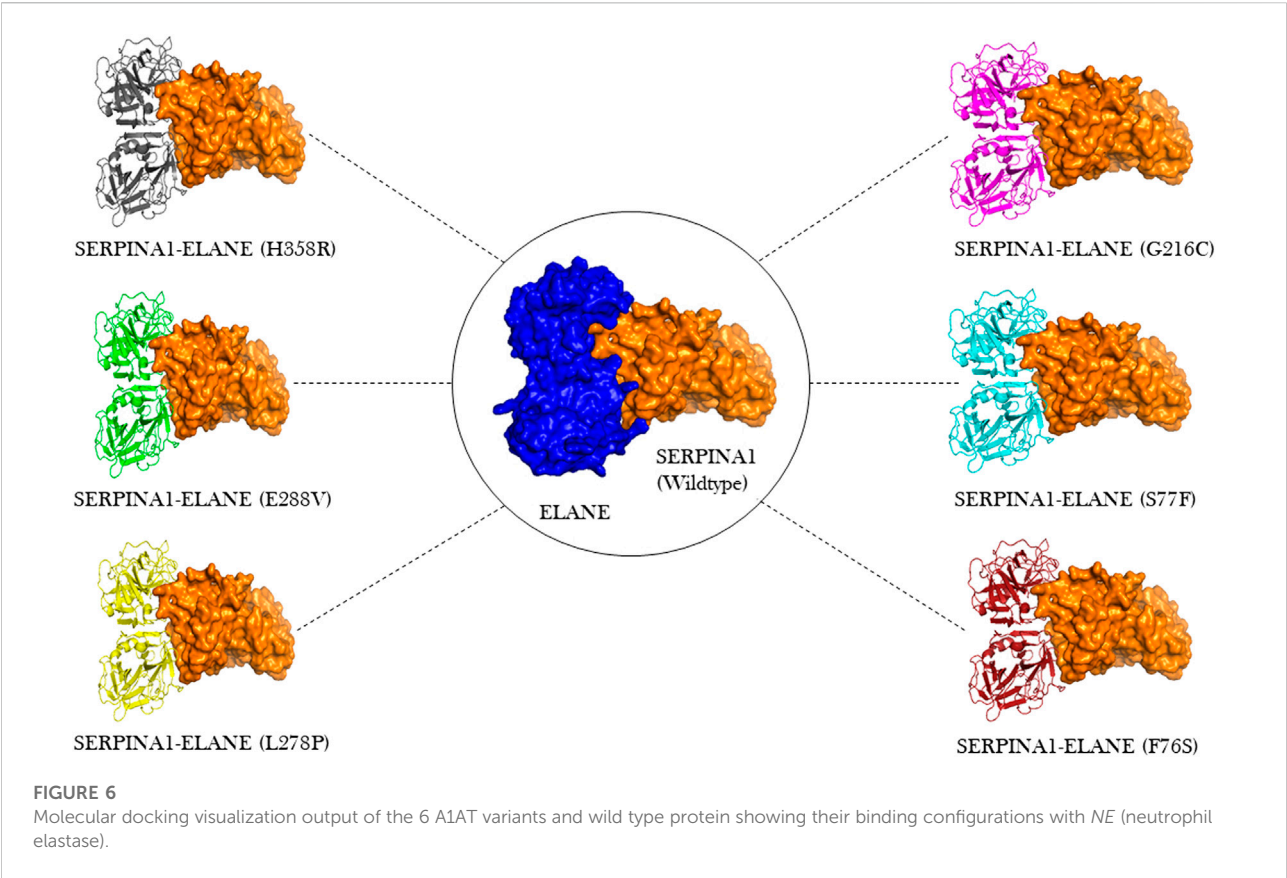


TABLE 5 Cluspro predictions of A1AT-NE molecular complex binding energy scores.

#	Receptor-ligand	Cluspro weighted scores	
		Center score (K.cal/mol)	Lowest energy (K.cal/mol)
1	A1AT-NE (Wildtype)	-798.7	-914.9
2	A1AT -NE (F76S)	-793.0	-941.2
3	A1AT -NE (S77F)	-798.5	-944.6
4	A1AT -NE (G216C)	-798.1	-941.5
5	A1AT -NE (L278P)	-794.8	-945.4
6	A1AT-NE (E288V)	-792.4	-941.1
7	A1AT-NE (H358R)	-775.0	-931.6

altering the way the alpha-1 antitrypsin protein functions inside the cell. The A1AT and protease molecular complex are recognized by hepatic receptors, which rapidly clear it from the blood circulation. A1AT has been shown to have a variety of different immunomodulatory actions in addition to its function as the main protease inhibitor, including an anti-inflammatory effect and the modulation of T- and B-lymphocyte functions.

Discussion

A1ATD manifests clinically with emphysema in the lungs around the fourth to fifth decade of life, with a proportion of patients developing liver cirrhosis later due misfolded A1AT protein aggregates accumulating in hepatocytes (Kelly et al., 2010). The *SERPINA1* gene is located on chromosome 14q32.1 and has three untranslated exons (IA, IB and IC) and

four coding exons (II–V). The first three exons regulate gene expression through three alternative transcription initiation sites: exons IA or IB in macrophages, exon IC in hepatocytes (Duvoix et al., 2014; Greulich et al., 2017; Tejwani and Stoller, 2021). Pathogenic variants in the *SERPINA1* gene underlie alpha-1-antitrypsin deficiency (A1ATD), which causes reduced protein levels. Many pathogenic *SERPINA1* variants associated with A1ATD have been reported in medical literature (Foil, 2021). Biological characterization of each genetic variant is impractical owing to their high number and time-consuming laboratory methods. In recent years, different researchers have shown the successful application of different computational methods like SIFT, Polyphen-2, M-CAP, and FATHMM in screening clinically pathogenic variants (Kelly et al., 2010; Hunt and Tuder, 2012; Lage, 2014). Computational biology-based pathogenicity prediction methods employ different support vector machine-based algorithms to identify deleterious variants from non-deleterious ones (Berclaz and Trapnell, 2006).

In the current study, computational predictions of SIFT, Polyphen, M-CAP, FATHMM, CADD and REVEL methods have confirmed all the six variants as pathogenic. However, computational predictions are often variable, based on the fact that each tool is trained on a unique variant data set (Thirumal Kumar et al., 2022; Thirumal Kumar et al., 2021; Abdel-Motal et al., 2018; Agrahari et al., 2019; Selvakumar et al., 2019).

This study found that two of the selected variants (F76S and S77F) fell in close range to the previously reported Trento variant of A1AT (E75V) (Miranda et al., 2017). The patient had a severe case of A1ATD with pathogenic intracellular polymer formation. The E75V variant affects the hydrogen bonding of the glutamic acid sidechain to the backbone of helix I which causes destabilization to the post helix I loop. This geometry has been shown to play a conserved role in preventing polymerization (Lomas et al., 1992). As a result of the expected geometry disturbance, F76S and S77F are expected to produce a polymerization phenotype similar to the Trento variant in affected individuals.

The evolutionary conservation analysis has identified that the H358R variant is in the Serpin conserved domain (344aa–369aa). This variant is reported in ClinVar as likely pathogenic. It is assumed that this variant might introduce a functional change in the protein conserved domain, but the secondary structure analysis for this variant did not show any clear alterations in secondary structural elements like α helices and β pleated sheets. Recent research has shown that missense variants TYK2 (Pro1104Ala), *IL6R* (Asp358Ala), and *PTPN22* (Trp620Arg) are unaffected by secondary structural element (SSE) changes in their respective proteins. The other variants were reported on ClinVar as follows: F76S as pathogenic, S77F as pathogenic, G216C as likely pathogenic, L278P as pathogenic, and E288V as pathogenic. The pathogenicity prediction tools provide a qualitative support for damaging or not damaging effect but do not show additional details on the structural changes caused

by the variants. *SERPINA1* missense alleles are associated with a significant reduction in A1AT serum levels because of the incorrect folding of the protein, poor stability, or degradation.

A1AT is a 52-kDa plasma glycoprotein with 394 amino acids. Its expression inversely correlates with expression of its binding partner NE [28]. The damaging effects of missense variants are better understood when their impact is studied at 3D structure level [29]. The superimposition of the investigated 3D protein structures onto the folded wildtype model is demonstrated to be a useful method for estimating root-mean-square deviation (RMSD), the average distance between backbone atoms of the superimposed variant and native protein structures [29]. In this investigation, we did not find major structural differences in the whole structure of A1AT in mutated state. However, all 6 missense variants cause subtle structural changes at the residue level. Amino acid substitutions often result in quantitative structural alterations that are accompanied by changes in fundamental physicochemical characteristics such as size, charge, side chains, molecular weight, and hydrophobicity. All of these modifications could affect the amino acid residues' chemical bonding properties (hydrogen bonds, ionic bonds, and Vander wall interactions), which are necessary for keeping the A1AT protein molecule in its secondary (α helices, β sheets, and coils), tertiary (3-dimensional form), and quaternary (protein subunit arrangement) structural conformations.

Thermodynamic stability can provide information about the nature of the A1AT's altered function. The greater the negative free energy, the more destabilizing the variant. In this combination, the DUET algorithm predicted two variants with highest negative free energy for H358R (−1.842 K. Cal/Mol) and F76S (−3.199 K. Cal/Mol) to be as destabilizing. The H358R falls into the Serpin conserved domain and the F76S has changed from a large hydrophobic residue (phenylalanine) to a small hydrophilic residue (serine). This might push the protein region towards the external environment, altering the overall thermodynamic and structural orientation.

To get a deeper insight into how variants influence the stability of the A1AT protein structure on the molecular level, molecular dynamics simulations were performed for the wild type and mutant models. The output trajectory of the simulation at multiple intervals (0, 25, 50, 75, and 100 ns) was subjected to two analyses: RMSD and RMSF. The RMSF value over the whole simulation (100 ns) of six mutant models (F76S, S77F, G216C, L28P, E288V, and H358R) showed significant fluctuations. However, when subjected to a force, they showed a significant structural behavior change. The wildtype model had more fluctuations ($>0.2 \text{ \AA}$) than the mutant models S77F, G216C, L278P, and E288V, whereas the F76S and H358R variants were more rigid in nature ($>0.1 \text{ \AA}$). This finding demonstrates that missense variants could somehow introduce mild structural changes that could affect the behavior of the protein in its environment. The increase in RMSF values of mutated models

of S77F, G216C, L28P, and E288V variants validated the calculated RMSD values. This suggests that these four variants must have some structural changes affecting the residue flexibility and behavior. Of the six mutated models, two models (S77F and G216C) showed a significant alteration in secondary structural elements. The S77F was already predicted in the previous secondary structure analysis as having an increased beta pleated sheet that is increased in length, giving more support to this finding. The secondary structure change can be attributed to the subtle changes giving rise to behavioral changes seen in MD simulations. However, we cannot assume it is the exact cause of this altered behavior.

Protein-protein interaction analysis is an important approach in understanding the variant impact on structural features of disease candidate proteins of genetic diseases (Lage, 2014). The reduction or increase in the binding affinity of a protein induced by residue alterations can make proteins lose their function and cause disease. The A1AT protein is the primary protease inhibitor in the human body, acting on a variety of targets including trypsin, collagenase, macrophage cathepsin, tissue kallikrein, factor IX and other molecules (Berclaz and Trapnell, 2006). However its main function is to keep the neutrophil elastase in balance during inflammation or infection through its inhibitory action (Hunt and Tuder, 2012). In this analysis, molecular docking has yielded a similar range of lowest energy (−931.6 K. Cal/Mol to −944.6 K. Cal/Mol), which predicts a significant difference versus the wildtype protein (−914.9 K. Cal/Mol). This results in a strong binding to NE (−16.7 K. Cal/Mol to 30.5 K. Cal/Mol deviation range), most likely causing a functional alteration. The binding configuration was chosen based on the best score, which could not have been compared to an available crystal structure. So, the actual amount was only an estimation.

Conclusion

This study concludes that computational methods like SIFT, PolyPhen, FATHMM, M-CAP and REVEL tools are very helpful in prioritizing *SERPINA1* loss-of-function pathogenic variants. These tools have lot of promise in screening Alpha-1-antitrypsin deficiency causative variants from next-generation sequencing data. It is important to note that laboratory experimental methods are required for definitive answers to the questions asked in this analysis. This present analysis highlighted general structural abnormalities caused by the reported missense variants of *SERPINA1*. The structural and stability prediction methods used in this study have shown how loss-of-function pathogenic variants could induce structural drifts, free energy value fluctuations, and alter the conformational dynamics of the A1AT protein molecule.

The *SERPINA1* mutations result in an unstable intermediate structure that is responsible for the β sheet-A opening, which can accept the RCL of another A1AT molecule to form a loop-sheet dimer. The latter can then be extended to form longer chains of loop-sheet polymers. These models are based on the “classic” loop-sheet model in which serpin polymers are formed by the intermolecular linkage of the reactive loop of one molecule with the β -sheet A of another. The findings from molecular docking have demonstrated how most missense variants negatively impact the affinity of NE and A1AT binding in a molecular complex, lowering A1AT functionality and contributing to its deficiency. Taken together, our computational approach offers an extra layer to study the deleterious potential of *SERPINA1* genetic variants from the structure and function context. Our findings recommend implementing computational variant assessment as a pre-invitro phase in improving the genomic medicine for A1ATD patients carrying *SERPINA1* pathogenic variants.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/Supplementary Material.

Author contributions

Conceptualization, NAS, NBS, and BB; Methodology, NAS, BB, and TA; Software, BB; Formal Analysis, NAS, TA, and BB; Investigation, NAS, TA, KJ, and BB; Resources, BB and NAS; Writing—Original Draft Preparation, NAS, NBS, TA, HA-N, NS, RE, and BB; Writing—Review and Editing, NAS, TA, DA, AE-H, ZA, RE, and BB; Visualization, TA and BB; Supervision, NAS, RE, and BB; Project Administration, NAS; Funding Acquisition, NAS.

Funding

This Project was funded by Deanship of Scientific Research (DSR), King Abdulaziz University, Jeddah, Saudi Arabia under grant no. (KEP-6-140-38).

Acknowledgments

The authors therefore acknowledge, with thanks, DSR for technical and financial support.

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.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmolb.2022.1051511/full#supplementary-material>

References

- Abdel-Motal, U. M., Abdelalim, E. M., Ponnuraja, C., Iken, K., Jahromi, M., Doss, G. P., et al. (2018). Prevalence of nephropathy in type 1 diabetes in the arab world: A systematic review and meta-analysis. *Diabetes. Metab. Res. Rev.* 34, e3026. doi:10.1002/dmrr.3026
- Adzhubei, I. A., Schmidt, S., Peshkin, L., Ramensky, V. E., Gerasimova, A., Bork, P., et al. (2010). A method and server for predicting damaging missense mutations. *Nat. Methods* 7, 248–249. doi:10.1038/nmeth0410-248
- Adzhubei, I., Jordan, D. M., and Sunyaev, S. R. (2013). Predicting functional effect of human missense mutations using PolyPhen-2. *Curr. Protoc. Hum. Genet.* 7, Unit7.20. doi:10.1002/0471142905.hg0720s76
- Agrahari, A. K., Doss, G. P. C., Siva, R., Magesh, R., and Zayed, H. (2019). Molecular insights of the G2019S substitution in LRRK2 kinase domain associated with Parkinson's disease: A molecular dynamics simulation approach. *J. Theor. Biol.* 469, 163–171. doi:10.1016/j.jtbi.2019.03.003
- Alharthi, A. M., Banaganapalli, B., Hassan, S. M., Rashidi, O., Al-Shehri, B. A., Alarifan, M. A., et al. (2022). Complex inheritance of rare missense variants in PAK2, TAP2, and PLCL1 genes in a consanguineous arab family with multiple autoimmune diseases including celiac disease. *Front. Pediatr.* 10, 895298. doi:10.3389/fped.2022.895298
- Awan, Z. A., Bahattab, R., Kutbi, H. I., Noor, A. O. J., Al-Nasser, M. S., Shaik, N. A., et al. (2021). Structural and molecular interaction studies on familial hypercholesterolemia causative PCSK9 functional domain mutations reveals binding affinity alterations with LDLR. *Int. J. Pept. Res. Ther.* 27, 719–733. doi:10.1007/s10989-020-10121-8
- Berclaz, P.-Y., and Trapnell, B. C. (2006). "Chapter 52 - rare childhood lung disorders: α 1-Antitrypsin deficiency, pulmonary alveolar proteinosis, and pulmonary alveolar microlithiasis," in *Kendig's disorders of the respiratory tract in children*. Editors V. Chernick, T. F. Boat, R. W. Wilmott, and A. Bush. Seventh Edition (Philadelphia: W.B. Saunders), 747–761.
- Bima, A. I. H., Elsamanoudy, A. Z., Albaqami, W. F., Khan, Z., Parambath, S. V., Al-Rayes, N., et al. (2022). Integrative system biology and mathematical modeling of genetic networks identifies shared biomarkers for obesity and diabetes. *Math. Biosci. Eng.* 19, 2310–2329. doi:10.3934/mbe.2022107
- Bowers, K., Chow, E., Xu, H., Dror, R., Eastwood, M., Gregersen, B., et al. (2006). In Proceedings of the 2006 ACM/IEEE Conference on Supercomputing, New York, 11–17 Nov. 2006.
- Cho, Y. L., Chae, Y. K., Jung, C. H., Kim, M. J., Na, Y. R., Kim, Y. H., et al. (2005). The native metastability and misfolding of serine protease inhibitors. *Protein Pept. Lett.* 12, 477–481. doi:10.2174/0929866054395365
- Desta, I. T., Porter, K. A., Xia, B., Kozakov, D., and Vajda, S. (2020). Performance and its limits in rigid body protein-protein docking. *Structure* 28, 1071–1081. doi:10.1016/j.str.2020.06.006
- Duvoix, A., Roussel, B. D., and Lomas, D. A. (2014). Molecular pathogenesis of alpha-1-antitrypsin deficiency. *Rev. Mal. Respir.* 31, 992–1002. doi:10.1016/j.rmr.2014.03.015
- Ferrarotti, I., Ottaviani, S., De Silvestri, A., and Corsico, A. G. (2018). Update on α 1-antitrypsin deficiency. *Breathe (Sheff)* 14, e17–e24. doi:10.1183/20734735.015018
- Foil, K. E. (2021). Variants of SERPINA1 and the increasing complexity of testing for alpha-1 antitrypsin deficiency. *Ther. Adv. Chronic Dis.* 12, 20406223211015954. doi:10.1177/20406223211015954
- Greulich, T., Nell, C., Hohmann, D., Grebe, M., Janciauskiene, S., Koczulla, A. R., et al. (2017). The prevalence of diagnosed α 1-antitrypsin deficiency and its comorbidities: results from a large population-based database. *Eur. Respir. J.* 49, 1600154. doi:10.1183/13993003.00154-2016
- Hunt, J. M., and Tudor, R. (2012). Alpha 1 anti-trypsin: one protein, many functions. *Curr. Mol. Med.* 12, 827–835. doi:10.2174/156652412801318755
- Ioannidis, N. M., Rothstein, J. H., Pejaver, V., Middha, S., McDonnell, S. K., Baheti, S., et al. (2016). REVEL: An ensemble method for predicting the pathogenicity of rare missense variants. *Am. J. Hum. Genet.* 99, 877–885. doi:10.1016/j.ajhg.2016.08.016
- Jagadeesh, K. A., Wenger, A. M., Berger, M. J., Guturu, H., Stenson, P. D., Cooper, D. N., et al. (2016). M-CAP eliminates a majority of variants of uncertain significance in clinical exomes at high sensitivity. *Nat. Genet.* 48, 1581–1586. doi:10.1038/ng.3703
- Kelly, E., Greene, C. M., Carroll, T. P., McElvaney, N. G., and O'Neill, S. J. (2010). Alpha-1 antitrypsin deficiency. *Respir. Med.* 104, 763–772. doi:10.1016/j.rmed.2010.01.016
- Köhnlein, T., and Welte, T. (2008). Alpha-1 antitrypsin deficiency: pathogenesis, clinical presentation, diagnosis, and treatment. *Am. J. Med.* 121, 3–9. doi:10.1016/j.amjmed.2007.07.025
- Kozakov, D., Beglov, D., Bohnuud, T., Mottarella, S. E., Xia, B., Hall, D. R., et al. (2013). How good is automated protein docking? *Proteins* 81, 2159–2166. doi:10.1002/prot.24403
- Kozakov, D., Hall, D. R., Xia, B., Porter, K. A., Padhorny, D., Yueh, C., et al. (2017). The ClusPro web server for protein-protein docking. *Nat. Protoc.* 12, 255–278. doi:10.1038/nprot.2016.169
- Lage, K. (2014). Protein-protein interactions and genetic diseases: The interactome. *Biochim. Biophys. Acta* 1842, 1971–1980. doi:10.1016/j.bbdis.2014.05.028
- Lomas, D. A., Evans, D. L., Finch, J. T., and Carrell, R. W. (1992). The mechanism of Z alpha 1-antitrypsin accumulation in the liver. *Nature* 357, 605–607. doi:10.1038/357605a0
- Miranda, E., Ferrarotti, I., Berardelli, R., Laffranchi, M., Cerea, M., Gangemi, F., et al. (2017). The pathological Trento variant of alpha-1-antitrypsin (E75V) shows nonclassical behaviour during polymerization. *FEBS J.* 284, 2110–2126. doi:10.1111/febs.14111
- Ng, P. C., and Henikoff, S. (2003). SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Res.* 31, 3812–3814. doi:10.1093/nar/gkg509
- Piñero, J., Saüch, J., Sanz, F., and Furlong, L. I. (2021). The DisGeNET cytoscape app: Exploring and visualizing disease genomics data. *Comput. Struct. Biotechnol. J.* 19, 2960–2967. doi:10.1016/j.csbj.2021.05.015
- Pires, D. E. V., Ascher, D. B., and Blundell, T. L. (2014). DUET: a server for predicting effects of mutations on protein stability using an integrated computational approach. *Nucleic Acids Res.* 42, W314–W319. doi:10.1093/nar/gku411
- Quinn, M., Ellis, P., Pye, A., and Turner, A. M. (2020). Obstacles to early diagnosis and treatment of alpha-1 antitrypsin deficiency: Current perspectives. *Ther. Clin. Risk Manag.* 16, 1243–1255. doi:10.2147/TCRM.S234377
- Release, S. (2017). 3: *Desmond molecular dynamics system, DE Shaw research*. New York, NY: Schrödinger. Maestro-Desmond Interoperability Tools.

- Rogers, M. F., Shihab, H. A., Mort, M., Cooper, D. N., Gaunt, T. R., and Campbell, C. (2018). FATHMM-XF: accurate prediction of pathogenic point mutations via extended features. *Bioinformatics* 34, 511–513. doi:10.1093/bioinformatics/btx536
- Selvakumar, J. N., Chandrasekaran, S. D., Doss, G. P. C., and Kumar, T. D. (2019). Inhibition of the ATPase domain of human topoisomerase IIa on HepG2 cells by 1, 2-benzenedicarboxylic acid, mono (2-ethylhexyl) ester: Molecular docking and dynamics simulations. *Curr. Cancer Drug Targets* 19, 495–503. doi:10.2174/1568009619666181127122230
- Shaik, N. A., Al-Qahtani, F., Nasser, K., Jamil, K., Alrayes, N. M., Elango, R., et al. (2020a). Molecular insights into the coding region mutations of low-density lipoprotein receptor adaptor protein 1 (LDLRAP1) linked to familial hypercholesterolemia. *J. Gene Med.* 22, e3176. doi:10.1002/jgm.3176
- Shaik, N. A., Bokhari, H. A., Masoodi, T. A., Shetty, P. J., Ajabnoor, G. M. A., Elango, R., et al. (2020b). Molecular modelling and dynamics of CA2 missense mutations causative to carbonic anhydrase 2 deficiency syndrome. *J. Biomol. Struct. Dyn.* 38, 4067–4080. doi:10.1080/07391102.2019.1671899
- Shaik, N. A., Nasser, K. K., Alruwaili, M. M., Alallasi, S. R., Elango, R., and Banaganapalli, B. (2021). Molecular modelling and dynamic simulations of sequestosome 1 (SQSTM1) missense mutations linked to Paget disease of bone. *J. Biomol. Struct. Dyn.* 39, 2873–2884. doi:10.1080/07391102.2020.1758212
- Tejwani, V., and Stoller, J. K. (2021). The spectrum of clinical sequelae associated with alpha-1 antitrypsin deficiency. *Ther. Adv. Chronic Dis.* 12, 2040622321995691. doi:10.1177/2040622321995691
- Thirumal Kumar, D., Jain, N., Udhaya Kumar, S., George Priya Doss, C., and Zayed, H. (2021). Identification of potential inhibitors against pathogenic missense mutations of PMM2 using a structure-based virtual screening approach. *J. Biomol. Struct. Dyn.* 39, 171–187. doi:10.1080/07391102.2019.1708797
- Thirumal Kumar, D., Shaikh, N., Udhaya Kumar, S., and George Priya Doss, C. (2022). Computational and structural investigation of palmitoyl-protein thioesterase 1 (PPT1) protein causing neuronal ceroid lipofuscinoses (NCL). *Adv. Protein Chem. Struct. Biol.* 132, 89–109. doi:10.1016/bs.apcsb.2022.07.002
- Vajda, S., Yueh, C., Beglov, D., Bohnuud, T., Mottarella, S. E., Xia, B., et al. (2017). New additions to the ClusPro server motivated by CAPRI. *Proteins* 85, 435–444. doi:10.1002/prot.25219
- Vissers, M. C., George, P. M., Bathurst, I. C., Brennan, S. O., and Winterbourn, C. C. (1988). Cleavage and inactivation of alpha 1-antitrypsin by metalloproteinases released from neutrophils. *J. Clin. Invest.* 82, 706–711. doi:10.1172/JCI113651



OPEN ACCESS

EDITED BY
Chandrabose Selvaraj,
Alagappa University, India

REVIEWED BY
Young Ho Seo,
Keimyung College, South Korea
Arul Murugan Natarajan,
Indraprastha Institute of Information
Technology Delhi, India

*CORRESPONDENCE
Kwan-Young Jung,
krjeong@krikt.re.kr
Jae-Sung Kim,
jaesung@kiram.re.kr
Sangwook Wu,
s.wu@pharmcadd.com

SPECIALTY SECTION
This article was submitted to Biophysics,
a section of the journal
Frontiers in Molecular Biosciences

RECEIVED 17 October 2022
ACCEPTED 16 November 2022
PUBLISHED 25 November 2022

CITATION
Jang SH, Sivakumar D, Mudedla SK,
Choi J, Lee S, Jeon M, Bvs SK, Hwang J,
Kang M, Shin EG, Lee KM, Jung K-Y,
Kim J-S and Wu S (2022), PCW-A1001,
AI-assisted *de novo* design approach to
design a selective inhibitor for FLT-
3(D835Y) in acute myeloid leukemia.
Front. Mol. Biosci. 9:1072028.
doi: 10.3389/fmolb.2022.1072028

COPYRIGHT
© 2022 Jang, Sivakumar, Mudedla, Choi,
Lee, Jeon, Bvs, Hwang, Kang, Shin, Lee,
Jung, Kim and Wu. 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.

PCW-A1001, AI-assisted *de novo* design approach to design a selective inhibitor for FLT-3(D835Y) in acute myeloid leukemia

Seong Hun Jang¹, Dakshinamurthy Sivakumar¹,
Sathish Kumar Mudedla¹, Jaehan Choi¹, Sungmin Lee¹,
Minjun Jeon¹, Suneel Kumar Bvs¹, Jinha Hwang¹,
Minsung Kang², Eun Gyeong Shin^{3,4}, Kyu Myung Lee³,
Kwan-Young Jung^{3,4*}, Jae-Sung Kim^{2*} and Sangwook Wu^{1,5*}

¹R&D Center, PharmCADD, Busan, South Korea, ²Division of Radiation Biomedical Research, Korea Institute of Radiological and Medical Sciences, Seoul, South Korea, ³Therapeutics & Biotechnology Division, Korea Research Institute of Chemical Technology, Daejeon, South Korea, ⁴Department of Medicinal Chemistry and Pharmacology, University of Science & Technology, Daejeon, South Korea, ⁵Department of Physics, Pukyong National University, Busan, South Korea

Treating acute myeloid leukemia (AML) by targeting FMS-like tyrosine kinase 3 (FLT-3) is considered an effective treatment strategy. By using AI-assisted hit optimization, we discovered a novel and highly selective compound with desired drug-like properties with which to target the FLT-3 (D835Y) mutant. In the current study, we applied an AI-assisted *de novo* design approach to identify a novel inhibitor of FLT-3 (D835Y). A recurrent neural network containing long short-term memory cells (LSTM) was implemented to generate potential candidates related to our in-house hit compound (PCW-1001). Approximately 10,416 hits were generated from 20 epochs, and the generated hits were further filtered using various toxicity and synthetic feasibility filters. Based on the docking and free energy ranking, the top compound was selected for synthesis and screening. Of these three compounds, PCW-A1001 proved to be highly selective for the FLT-3 (D835Y) mutant, with an IC₅₀ of 764 nM, whereas the IC₅₀ of FLT-3 WT was 2.54 μM.

KEYWORDS

FLT-3, AML, kinase inhibitors, *de novo* drug design, LSTM, AI

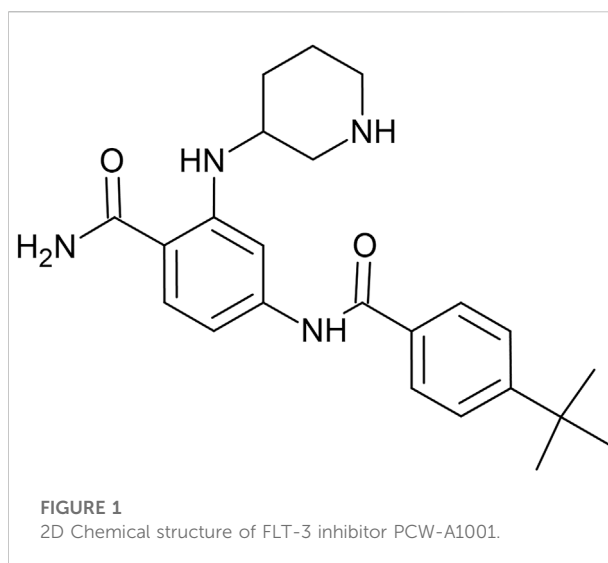
Introduction

Overexpression or mutation of some signaling proteins leads to cancer development (Kazi and Rönnstrand, 2019). Among the most mutated extracellular signaling mediators in cancer are the receptor tyrosine kinases (RTKs) (McDonnell et al., 2015). Among the five known types of RTKs, FMS-like tyrosine kinase (Kindler et al., 2010) (FLT-3) belongs to type III. It plays an essential role in regulating early hematopoiesis because it is selectively expressed on CD34+ hematopoietic stem cells and immature hematopoietic progenitors (Rosnet et al., 1996; Kindler et al., 2010). It is also expressed in the liver, spleen, lymph nodes, thymus, placenta, gonads, and brain (Del Zotto et al., 2001; Stirewalt and Radich, 2003; Brown and Small, 2004). Our work is focused on FLT-3, a gene that is highly mutated in acute myeloid leukemia (AML) (Grafone et al., 2012).

Interestingly, thousands of mutations (mostly insertions) have been reported for FLT-3. Many FLT-3 point mutations are commonly found in AML, and the activation loop residue D835, which stabilizes the inactive conformation is the predominant site of mutations (Yamamoto et al., 2001; Liang et al., 2003; Smith et al., 2012). Overexpression and frequent FLT-3 mutations are associated with poor prognoses and AML pathogenicity and activate downstream signaling molecules, which leads to stimulation and survival of cancerous cells (Zhang and Broxmeyer, 1999; Hayakawa et al., 2000; Lin et al., 2012). Treating AML patients by targeting FLT-3 and its mutants with small molecules is considered a promising strategy (Assouline et al., 2012; Leung et al., 2013; Gebru and Wang, 2020; Ambinder and Levis, 2021).

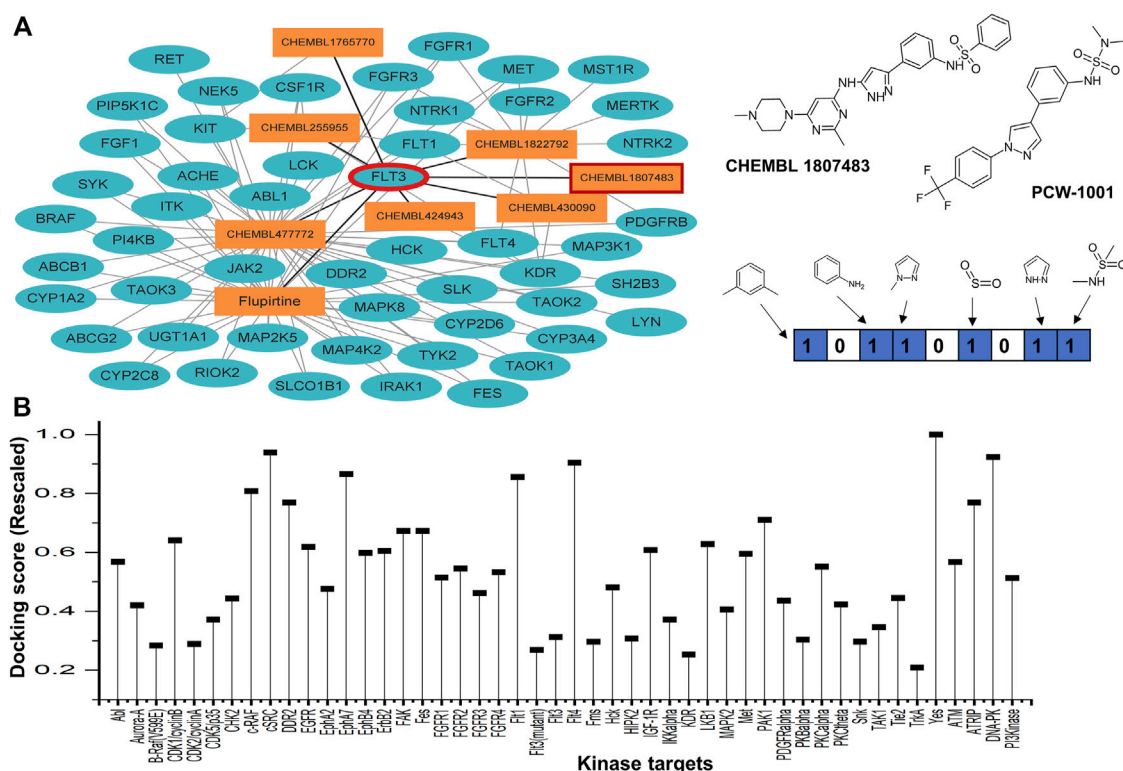
Since approval by the Food and Drug Administration (FDA) of the first tyrosine kinase inhibitor imatinib (Savage and Antman, 2002) two decades ago, several drugs targeting FLT-3 have entered clinical trials. Nevertheless, only midostaurin and gilteritinib have been approved by the FDA (Scholl et al., 2020). FLT-3 inhibitors are classified as type I or type II based on their binding with the protein. Type I inhibitors such as sunitinib (Schittenhelm et al., 2006), midostaurin (Stone et al., 2004), lestaurtinib (Smith et al., 2004), crenolanib (Heinrich et al., 2012), and gilteritinib (Grunwald and Levis, 2013) bind with the active state (DFG-in) of FLT-3, whereas type II inhibitors such as sorafenib (Auclair et al., 2007), ponatinib (O'Hare et al., 2009), and quizartinib (Zarrinkar et al., 2009) bind only with the inactive (DFG-out) FLT-3 conformation (Scholl et al., 2020). Studies have shown that type I inhibitors are more promising for use in AML treatment, as they target the predominant mutated kinase (Wodicka et al., 2010; Smith et al., 2012). There has been tremendous interest in developing FLT-3 inhibitors using classic computer-aided drug design approaches (Chang Hsu et al., 2014; Ke et al., 2015). In this study, we focus on developing a more rational approach for preparation of FLT-3 type-I inhibitors.

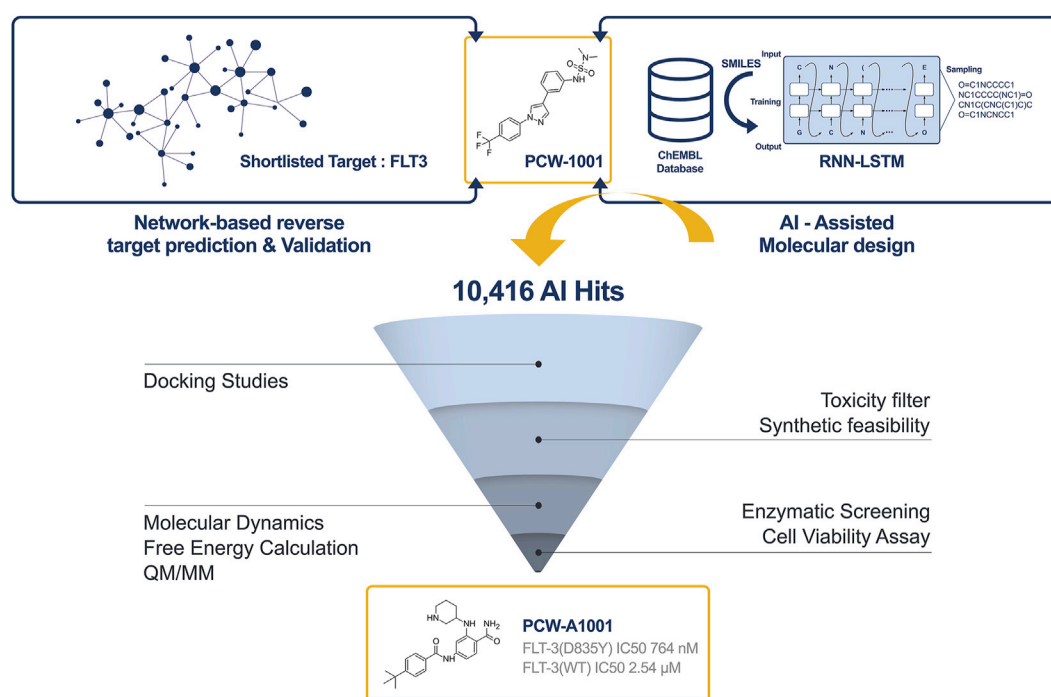
Recent breakthroughs show the significance of artificial intelligence (AI) in drug discovery, and AI reduces costs and



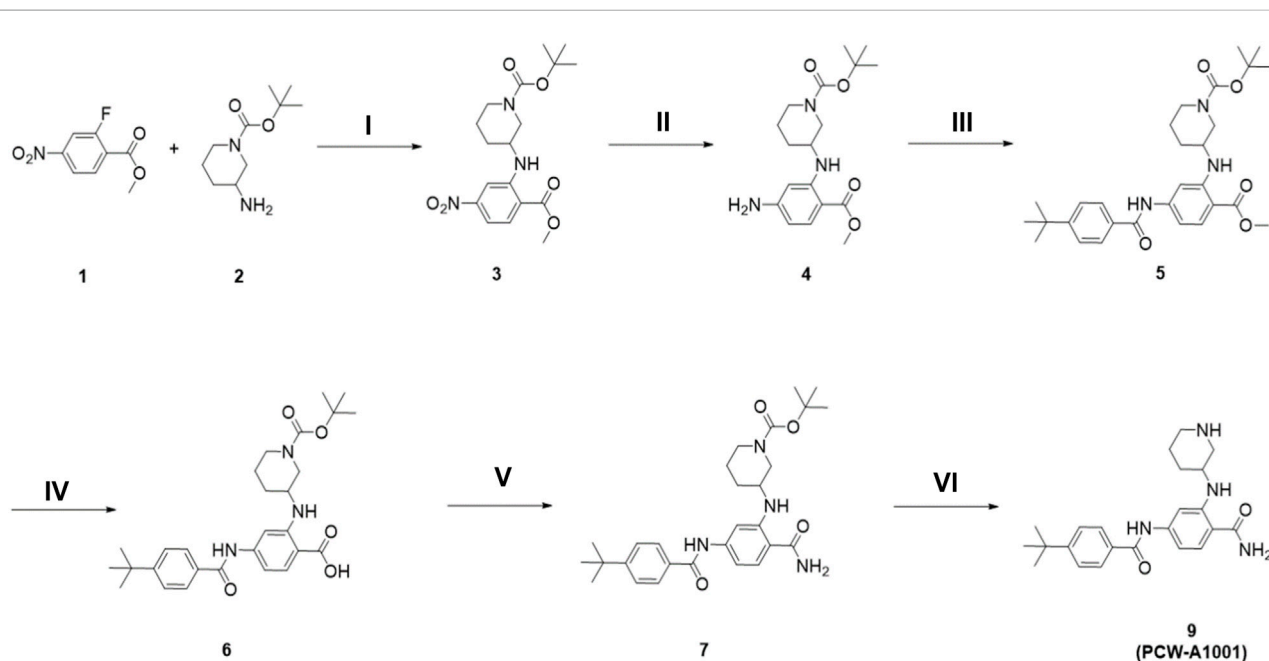
increases the speed of the drug discovery pipeline (Mak and Pichika, 2019). One of the main bottlenecks of traditional *de novo* drug design methods is the complicated synthetic routes; reported AI methods suggest synthetically feasible molecules or synthetic pathways that can help chemists (Corey and Wipke, 1969; Hessler and Baringhaus, 2018). Using AI, identification of a DDR1 kinase inhibitor was completed in just 60 days, including synthesis and experimental validation (Zhavoronkov et al., 2019). Excientia prepared the first AI-designed drug (DSP-1181) to treat obsessive-compulsive disorder (OCD), which subsequently entered clinical trials (Luo et al., 2022). They also discovered the AI-designed molecule EXS-21546 for immuno-oncology, which entered clinical trials in 8 months. Insilico medicine (www.insilico.com) used its AI program to develop a novel inhibitor (ISM001-055) for antifibrotic targets, and it reached clinical trials in 9 months. Recently, they have announced a preclinical candidate for the main protease of SARS-CoV-2, which was discovered with their novel AI platform, Chemistry42.

Network-based approaches are widely used to infer relationships between diseases and drugs (Guney et al., 2016) and are more focused on predicting novel protein targets and new uses of known drugs (Berger and Iyengar, 2009; Wu et al., 2013). In the current study, we used our reverse network theory approach developed in-house to identify a potential therapeutic target for PCW-1001. Based on the network theory and docking results, FLT3 was considered a potential target. Further biological screening studies showed that PCW-1001 exhibited an inhibitory IC_{50} of 13.6 μ M against FLT-3 WT and 1.83 μ M against the FLT-3 (D835Y) mutant (Kang et al., 2022). An AI-assisted *de novo* design approach was applied to identify a potent and selective inhibitor for the FLT3/FLT-3 (D835Y) mutant. This parent compound (PCW-1001) was considered for further optimization, and more than 10,416 analogs were generated using the LSTM approach. These hits were further evaluated for synthetic



**FIGURE 3**

Overall pipeline of AI based drug discovery approach to identify PCW-A1001 from PCW-1001. Step-1: The target protein for the PCW-1001 was identified as FLT3 from the network analysis. Step-2: AI-assisted drug design using the RNN-LSTM method. Step-3: Generated compounds were evaluated using various lead-like identification filters. Step-4: Binding mode analysis (Docking, Molecular dynamics, QM/MM, Free energy calculation) of the filtered molecules. Step-5: Synthesis and characterization. Step-6: *In vitro* Assay for the enzymatic activity and cell viability.

**SCHEME 1**

The synthesis of PCW-A1001. (i) K₂CO₃, DMF, 70°C, 12 h; (ii) H₂(g), Pd/C, EtOAc, r.t., 10 min; (iii) 4-tert-butylbenzoic acid, EDCI, DMAP, DCM, r.t., 48 h; (iv) LiOH·H₂O, THF/MeOH/H₂O, r.t., 12 h; (v) NH₄Cl, HBTU, DIPEA, DMF, r.t., 12 h; (vi) 20% TFA in MeOH, r.t., 3 h.

for *de novo* drug design (Merk et al., 2018) to generate potential hit candidates around PCW-1001 (Figure 3). We fine-tuned the model by using the transfer learning approach to optimize *de novo* generation of FLT-3 active compounds (Kang et al., 2022). We sampled 10,416 SMILES (Simplified Molecular Input Line Entry System) strings from 20 epochs from the resulting fine-tuned model. AI-generated hits were further evaluated using MOSES (Polykovskiy et al., 2020) for novelty, validity, diversity, scaffold similarity, and uniqueness. Benchmarking analyses indicated that 98.8% of the hits were valid, 85% of the hits were unique, and 90% of the hits were novel. The chemical space of AI-generated hits falls within the range of FLT3 known actives and PCW-1001 (Supplementary Figure S1). Furthermore, violin plot analysis also suggested that the distribution of molecular weights and LogP of AI-generated hits were within the range of known FLT3 actives (Supplementary Figure S2). Overall, AI-generated hits fell within the chemical space of known actives, and MOSES analysis suggested that AI-generated hits were diverse and novel compared with known FLT3 actives and PCW-1001.

Based on the binding analysis and interaction pattern studies, we identified 1750 compounds out of 10,416 hits as suitable for further studies. Following this preliminary evaluation, we assessed the resulting structures for toxicity endpoints *via* our Pharmulator™. We generated 9 toxicity models (different endpoints) using available literature data, and validated models were deployed in Pharmulator™ to assess the hit moieties quickly. Of 1750 hits, only 190 compounds passed synthetic feasibility, novelty, drug-like, all toxicity, and PAINS filters and were further subjected to a binding pose analysis and free energy calculations. We selected the top compound for synthesis and *in vitro* screening.

Synthesis and structural characterization of the *de novo* compound PCW-A1001

The synthetic route to PCW-A1001 is summarized in Scheme 1. Methyl 2-fluoro-4-nitrobenzoate **1** and tert-butyl 3-aminopiperidine-1-carboxylate **2** were reacted in the presence of K₂CO₃ to obtain **3** through nucleophilic aromatic substitution. The nitro group of resulting compound **3** was converted to an amino functional group *via* hydrogenation. Amine Compound **4** was coupled with 4-(tert-butyl)benzoic acid in the presence of EDCI and a catalytic amount of DMAP and then hydrolyzed using LiOH·H₂O to produce intermediate **6**. The acid functional group was efficiently converted to an amide with the HBTU coupling reagent. The Boc protecting group of the secondary amine in the piperidine ring was removed to obtain the desired compound PCW-A1001. The step-by-step synthesis and structural characterizations are shown in the Supplementary Material.

Structural interaction and stability analysis of PCW-A1001 with wild-type FLT-3 and the FLT-3 (D835Y) mutant

The top predicted binding mode of PCW-A1001 with FLT-3 WT showed two key hydrogen bonding interactions with Cys694 and Cys695 and a π - π interaction with Phe830. The docking complex of PCW-A1001 with FLT-3 WT was considered for molecular dynamics (MD) simulation for 100 ns. The MD simulation results also showed that the compound binding interactions observed in the initial docked complex were retained in PCW-A1001. The compound bound perfectly in the ATP binding site by forming hydrogen bonds with the two cysteine residues (Cys694 and Cys695) located in the hinge region. As observed in several inhibitor-kinase complexes, hydrogen bonding interactions with the inhibitor are essential for kinase inhibitory activity (Banks et al., 1979; Ke et al., 2015). The carbonyl moiety of the benzamide group formed a hydrogen bond with the NH group of the Cys694 residue (Figure 4).

The NH moiety of the next benzamide group in the compound formed a hydrogen bond with the backbone carbonyl group of the Cys695 residue. Phe830 in the DFG loop in wild-type (WT) FLT-3 formed π - π interactions with compound PCW-A1001. The binding free energy of PCW-A1001 was -13.4 kcal/mol with FLT-3 WT but -14.8 kcal/mol with the FLT-3 (D835Y) mutant, whereas those of the precursor compound PCW-1001 were -7.2 kcal/mol and -8.07 kcal/mol for FLT-3(WT) and FLT-3 (D835Y), respectively (Table 1).

In the FLT-3 (D835Y) mutant, the NH group of Cys694 formed a hydrogen bond with the CO moiety of the benzamide group in PCW-A1001, as seen with WT FLT-3. Cys695 also maintained its hydrogen bonding interactions, as in the WT; additionally, Lys614 and Tyr693 interacted with the protein. The C α -RMSD of the WT FLT-3 and the mutant complex showed that the complex was stable throughout the simulation (Figure 4).

Based on the binding study of PCW-A1001 and its precursor compound PCW-1001 against the panel of kinase enzymes, selectivity was achieved by PCW-A1001 for FLT-3 (D835Y). The selectivity scores of PCW-A1001 and PCW-1001 were calculated from the dock score (rescaled) of the selected kinase panel of enzymes as 0.33 and 0.46, respectively, for the Flt-3 (D835Y) mutant (Supplementary Figure S3).

QM/MM analysis of PCW-A1001

QM/MM optimization was used to validate the interactions between PCW-A1001 and the FLT-3 (D835Y) mutant in the MD-determined complex to study the electronic and structural properties of the ligand and selected atoms of the protein (Figure 5). The electrostatic and van der Waals interactions

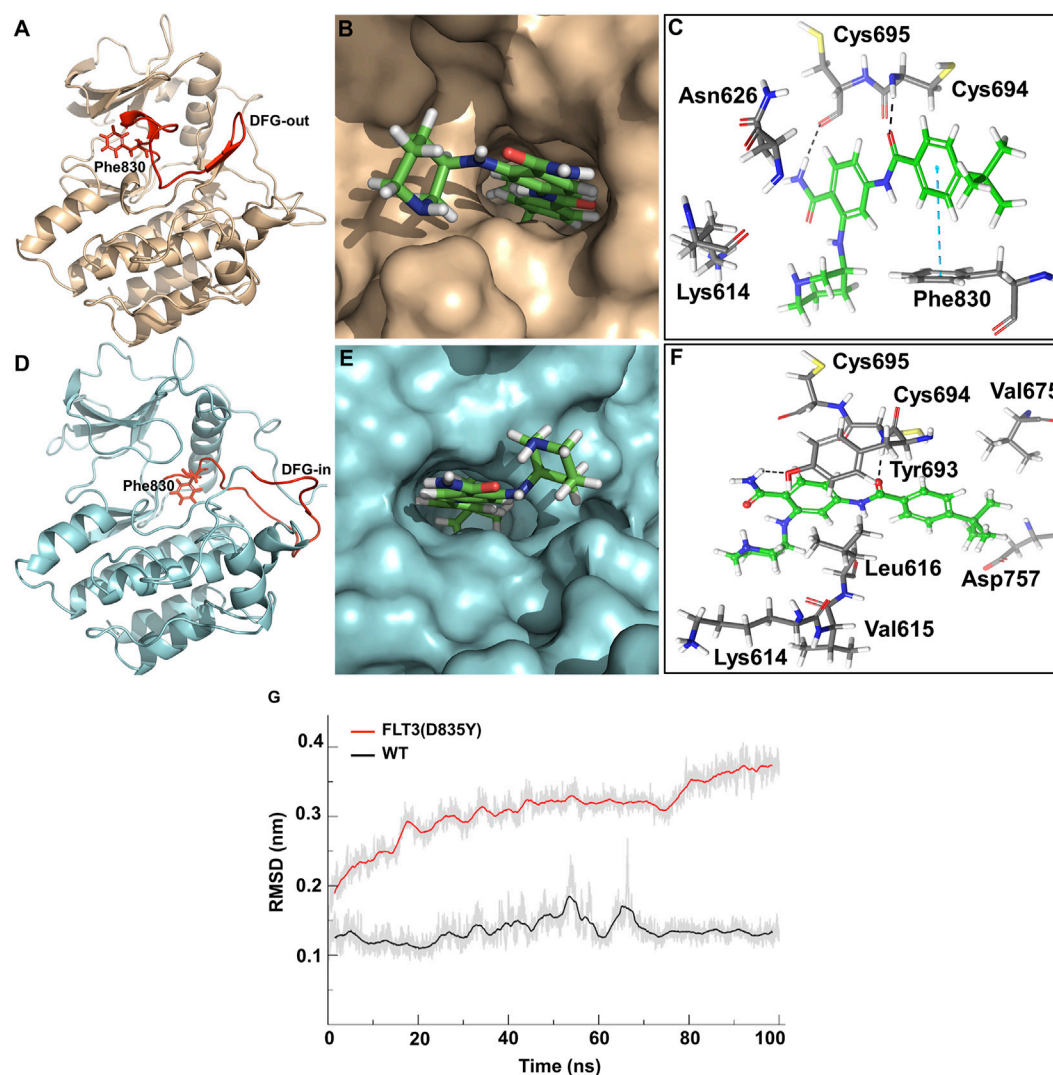


FIGURE 4

Binding analysis of PCW-A1001. (A) DFG-out conformation of FLT-3 wild type (B) Binding mode of PCW-A1001 with FLT-3 wild type (C) Crucial interactions observed in the FLT-3 wild-type (sticks). (D) DFG-in conformation of FLT-3 (D835Y) (E) Binding mode of PCW-A1001 with FLT-3 (D835Y), (F) Crucial interactions observed in the FLT-3 (D835Y) (sticks). Hydrogen bond interactions are represented in black dashed lines and pi-pi interactions in the blue dashed lines. (G) RMSD (Ca) plot of FLT-3 wild type and D835Y mutant.

TABLE 1 Binding free energy calculation for PCW-A1001 with WT and mutant FLT-3.

Compound	FLT-3 (WT)	FLT-3 (D835Y) mutant
PCW-A1001	-13.4 kcal/mol	-14.8 kcal/mol
PCW-1001	-7.2 kcal/mol	-8.07 kcal/mol

were intact in the protein-ligand complex, as with the MD structure. The ligand was stabilized at the binding site through hydrogen bonding and -CH- π and -NH- π interactions. The backbone -NH group from

Gly697 interacted with the phenyl ring of the ligand. In addition to hydrogen bonding interactions, -CH- π interactions were dominant in the complex formed between the ligand and protein. The Leu746, Phe691, Val624, and Leu616 residues were involved in the -CH- π interaction, as shown in Figure 5. The same pattern was also observed in the case of WT protein binding with the ligand, except for the -CH- π interactions with Ala642. The calculated interaction energies for the ligand and protein were -46.91 and -34.04 kcal/mol for the mutant and WT, respectively. These binding affinities were in good agreement with the free energy calculations for an explicit water environment.

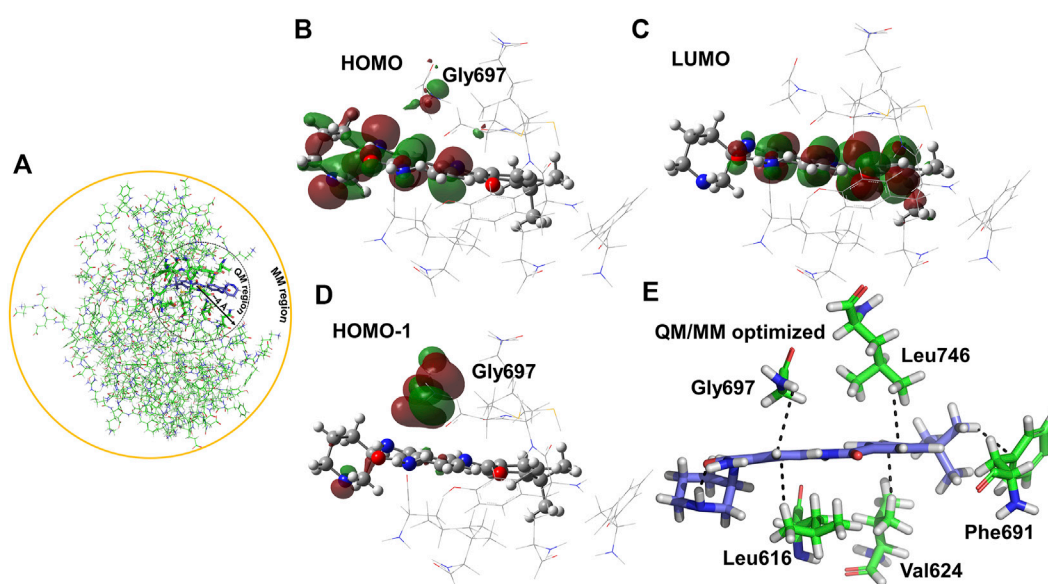


FIGURE 5

QM optimization of PCW-A1001 with FLT-3(D835Y). (A) Schematic representation of the QM and MM optimization region selected in the protein-ligand complex of the FLT-3-D835Y mutant. (B,D) HOMO/HOMO-1 surface of FLT-3 (D835Y) and (C) LUMO surface in FLT-3 (D835Y) in complex with PCW-A1001. Ligand PCW-A1001 shows in ball and sticks and the residues in the QM region shown in lines. (E) Hydrophobic interactions observed in the QM/MM optimized complex.

The electron distribution was determined from the calculated orbital energies. The highest occupied molecular orbital (HOMO), HOMO-1, and lowest unoccupied molecular orbital (LUMO) were computed and are displayed in Figure 6. The HOMO surface was predominantly localized in the hydrogen bonding interaction regions of the ligand, whereas the LUMO surface was distributed evenly across the ligand (Figure 6). Gly697, which was involved in the -NH- π interaction with the ligand, contributed less to the HOMO, whereas HOMO-1 was highly localized on Gly697. The piperidine ring in the ligand formed an intramolecular hydrogen bond and stabilized the ligand orbitals. Thus, the HOMO was localized on and near the piperidine ring. Furthermore, atomic charges were calculated with natural population analysis. The sum of the atomic charges on the ligand was found to be 0.02 au. No significant charge transfer from the ligand to the protein was observed.

We also analyzed the Frontier orbitals of PCW-A1001, as shown in Figure 6. Frontier orbitals direct the mode of interaction between drugs and proteins. The HOMO and LUMO contribute to the chemical stability of the molecule. If the energy gap is zero or negligible, the molecule is highly reactive. PCW-A1001 was stable and showed an energy gap of 3.7 eV. The HOMO was localized on the phenyl ring, and the LUMO was distributed across two phenyl rings of PCW-A1001. This indicated that intramolecular charge transfer might enhance the stability of PCW-A1001. The molecular electrostatic potential illustrates the charge distribution of a molecule. This explains how one molecule can interact with another. The electrostatic potential helps determine the electrophilic

and nucleophilic sites involved in hydrogen bond formation. The calculated electrostatic potential surface is shown in Figure 6. The positive and negative potentials are indicated by blue and red colors, respectively. Atoms in the positive potential region act as electron acceptors, whereas atoms with a negative potential behave as electron donors during hydrogen bond formation with FLT-3. The aromatic phenyl rings involved in -CH- π interactions were found between the positive and negative potentials. The results show that the charge distribution over PCW-A1001 was favorable for interacting with the binding pocket of FLT-3.

Inhibition of MV4-11 and acute myeloid leukemia cell lines by PCW-A1001

MV4-11 cells and FLT-3-mutated AML cells (Quentmeier et al., 2003) were used to examine the anticancer activity of PCW-A1001. It inhibited the proliferation of MV4-11 cells, with an IC_{50} of 1.98 μ M, showing that PCW-A1001 has potent anticancer activity in AML cells (Figure 7).

In vitro kinase activity of PCW-A1001

Next, we performed an *in vitro* kinase assay to evaluate the inhibitory activity of PCW-A1001 for WT FLT-3 and the D835Y-mutant kinase. Interestingly, our data indicated that PCW-

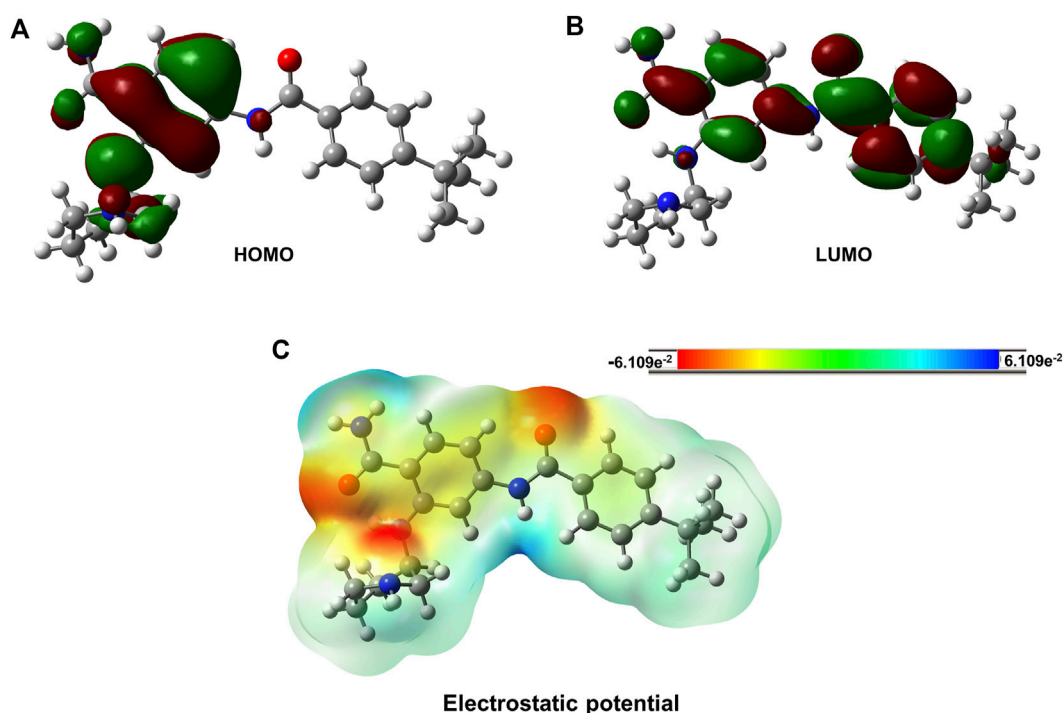


FIGURE 6

Isosurfaces of Frontier orbitals HOMO (Highest Occupied Molecular Orbital) and LUMO (Lowest Unoccupied Molecular Orbital). The charge distribution over PCW-A1001 molecule. (A) Isosurface of HOMO, (B) Isosurface of LUMO and (C) Electrostatic potential of the PCW-A1001. The red color indicates the negative charge and green color indicates the positive charge for the PCW-A1001. The electrostatic potential values were distributed from -6.109×10^{-2} to 6.109×10^{-2} .

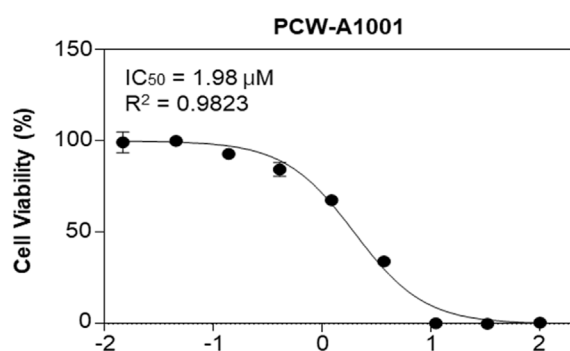


FIGURE 7

PCW-A1001 inhibits the viability of AML cell line. MV4-11 cells were treated with PCW-A1001 for 72 h, and cell viability was analyzed. The data are presented as the mean \pm standard error mean. The dose response curve was generated using OriginPro 2021 (OriginLab, United States).

A1001 inhibited the mutant kinase more effectively than the WT (Figure 8A). The IC_{50} determined from the kinase assay was 764 nM for the FLT-3 D835Y mutant, which was only one-third the IC_{50} for WT ($2.54 \mu M$) (Figures 8B,C). Thus, our data

provide proof-of-concept evidence for the AI-assisted *de novo* drug design approach.

Materials and methods

Reverse target prediction—a network-based approach

The input structures were used as SMILES structures and converted into fingerprints. We used six kinds of fingerprints: RDKit, MACCSkeys, AtomPair, Torsion, Morgan, and Morgan with Features (Landrum, 2016). A total of six fingerprints and six similarities were calculated and compared with the precalculated scores for the ligands in our protein–ligand interaction network database. A total of 10,647 compounds were used for the analysis among 262,327 compounds in the ChEMBL database. The list of compounds used with SMILES is included in the (Supplementary Table S2). In detail, the similarities between the input compound and all drugs in the network were calculated. Here, we used six similarity measures: Tanimoto, Dice, Sokal, Cosine, Kulczynski, and McConnaughey. We selected drugs with high similarity scores (sum of similarity values). Using protein–drug relationships in the

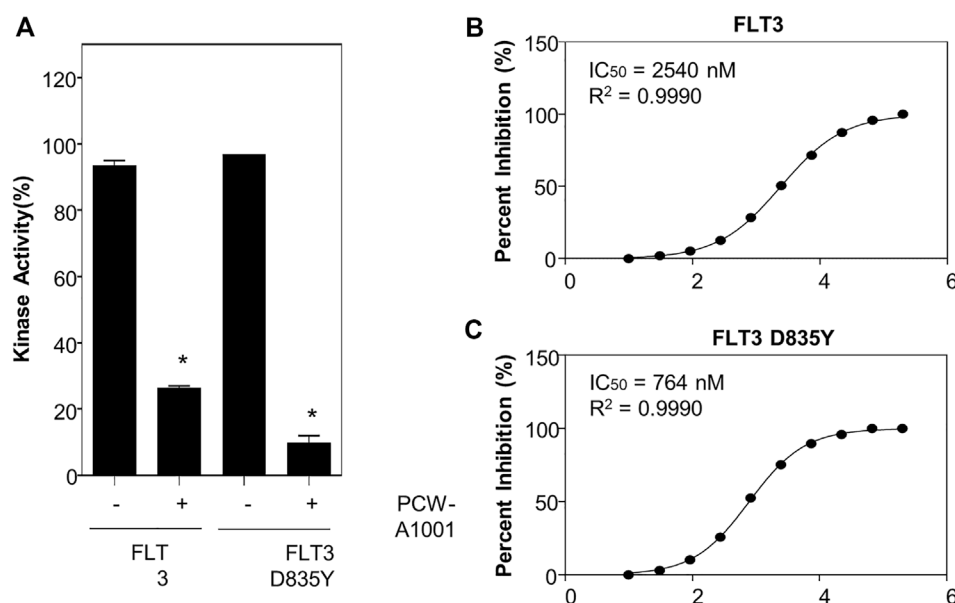


FIGURE 8

PCW-A1001 inhibits the kinase activity of FLT-3 WT and D835Y mutant. (A) PCW-A1001 inhibited FLT-3 WT and D835Y mutant kinase at 10 μ M (A). IC₅₀ of PCW-A1001 for FLT-3 WT (B) and D835Y mutation (C) were analyzed by ThermoFisher. The data are presented as the mean \pm standard error mean. * $p < 0.01$.

network, the similarity score of a drug was assigned to all proteins connected to the drug. Then, we obtained a list of proteins with similarity scores for each fingerprint. Based on the similarity score, the list was sorted in descending order, and a rank value, which is a descending value from the maximum rank, was assigned. The rank score was calculated using the following formula:

$$\text{rank score} = \frac{\text{rank value}}{\sum \text{rank values}} \quad (1)$$

The proteins obtained a total rank score R , which was the sum of the rank score of each fingerprint. The rank score was further modified as \tilde{R} by applying the following formula and assigned to each protein. The modified total rank score showed the potential of the target protein. The network model was validated using eight known Bruton's tyrosine kinase (BTK) inhibitors, the model prediction rank was given (Supplementary Figure S4), and the details of the targets along with their inhibitors are given (Supplementary Table S1).

$$\tilde{R} = \frac{1}{R_{\max}} \sum_{\text{Finger print}} \text{rank score} \quad (2)$$

AI-assisted *de novo* design

A recurrent neural network (RNN) is a type of neural network that is widely used for natural language processing

(NLP) tasks from simple language processing to complex cheminformatics problems. RNNs have been successfully applied for protein structure and function predictions from sequences (Liu, 2017; Zhang et al., 2018), property predictions, fragment-based hit generation (Awale et al., 2019), and hit identification (Segler et al., 2018; Erikawa et al., 2021).

For *de novo* drug design, we successfully applied generative recurrent neural networks (RNN) containing long short-term memory (LSTM) cells (Merk et al., 2018). The model considers the SMILES strings for molecular representation and learns the patterns and their probabilities from pretraining for use in generation of the SMILES structures. We fine-tuned the generated structures (SMILES) for specific molecular targets or chemical series by employing transfer learning. The generative LSTM approach has proven helpful in low-data drug discovery, hit expansion, molecular design (fragment-based), and lead optimization (Gupta et al., 2018; Segler et al., 2018; Erikawa et al., 2021).

All deep learning models were applied using TensorFlow (v2.1, www.TensorFlow.org) in Python (v3.7, www.python.org). We have used RDkit (www.rdkit.org) for most cheminformatics activities (property calculations, SMILES string validity calculations, molecular fingerprint calculations, and molecular clustering calculations). A detailed analysis of the generated SMILES strings was performed using the Jupyter notebook (www.anaconda.org).

RNNs were used as autoencoders, and the deep learning model employed for this study was initially pretrained to capture the structures of 438,552 bioactive small molecules (after carefully excluding the FLT-3 actives) retrieved from ChEMBL25 (KD, Ki, EC50, IC50 < 1 μ M) and represent them as simplified molecular-input line-entry system (SMILES) strings (Weininger, 1988). Using this pretrained model, we fine-tuned the model by transfer learning to bias *de novo* molecule generation toward the desired bioactivities of the templates (Merk et al., 2018). This fine-tuning step was employed to train the model for designing functional mimetics. Generated hits were further evaluated and benchmarked using Molecular sets (MOSES) (Polykovskiy et al., 2020).

Synthetic feasibility

The synthetic feasibility of each compound generated by AI was obtained with the retrosynthesis-associated fragment-based synthetic feasibility (RAFSF) score module. The fundamental idea of the module was that after cleaving synthetically meaningful bonds of the given compound, the bonds and the resulting fragments were searched from a bond/fragment space extracted from the ChEMBL (Mendez et al., 2019) or USPTO (Lowe, 2017) grants database in the same way. In the ChEMBL small molecule database, 1,917,863 molecules with molecular weights of less than 1,000 were used, as were 1,808,937 reactions from the USPTO grants database. To break bonds, we used the modified BRICS (Degen et al., 2008) module included in the rdkit (Landrum, 2016). If the bond/fragment from the given compound was not contained or rarely appeared in the bond/fragment space, a RAFSF score with a high value was assigned, meaning it was synthetically unfeasible. The RAFSF score is a value ranging from 1 (highly feasible synthesis) to 10 (highly unfeasible).

Protein preparation and modeling

The DFG-out WT FLT-3 protein structure was downloaded from the protein databank (www.rcsb.org) with PDB ID: 1RJB in the DFG-out conformation. Protein structures were prepared by correcting the bond orders, adding missing hydrogens, optimizing H-bonding with the protonation states of residues at pH 7.0, and restraining minimization for added hydrogens using the OPLS2005 forcefield of Protein Preparation Wizard (Sastry et al., 2013). The DFG-in conformation of the FLT-3 (D835Y) mutant was modeled using Modeler 9.25 (Sali and Blundell, 1993) with two templates, as reported previously (Ke et al., 2015). The first template was the DFG-out conformation of FLT-3 (PDB: 1RJB) with its DFG motif removed, and the other template was the DFG-in conformation of the colony-stimulating factor-1 receptor (CSF-1) crystal structure (PDB

id: 3LCD). These templates shared 93% and 63% sequence identity with the target protein, respectively. The model was subjected to loop refinement and minimization, followed by validation using a standard protocol discussed elsewhere (Sivakumar et al., 2013; Ke et al., 2015).

Ligand preparation and molecular docking

Hit compounds were initially optimized using the DFT method in Gaussian 16 with B3LYP functionals and the 6-31G** basis set (Tirado-Rives and Jorgensen, 2008). The antechamber obtained GAFF atom types with RESP charges from the Gaussian output file. The atom types and all needed parameters for the ligand were obtained from the above process along with parmed and tleap (Shirts et al., 2017). Molecular docking was carried out using Glide XP (Friesner et al., 2006) with default parameters; initially, the receptor was prepared with a grid box set covering the centroid of the active site, followed by flexible ligand sampling of the ligand docking.

Molecular dynamics simulations

The stabilities of the complexes were studied by MD simulations using Gromacs 2019 (Abraham et al., 2015). The Charmm36 force field (Huang and MacKerell, 2013) was used for the protein parameters. The protein–ligand complexes were solvated explicitly using the TIP3P water model inside the cubic box, and their sizes extended 0.1 nm away from the protein on the edges of the box in each direction. The overall charge of the system was neutralized by adding a 0.15 M salt (Na^+Cl^-). All simulations were carried out on GPU-enabled Linux clusters. The entire system was minimized with a maximum step size of 50,000 until the maximum force was less than 10 kJ/mol. The system was then equilibrated for 5 ns under NVT conditions with temperature coupling for two separate groups, protein–ligand and water-ions, at 300 K. The Lincs algorithm was used to constrain the bonds of the hydrogen atoms (Hess et al., 1997). A Berendsen thermostat and V-rescale were used to keep the temperature and pressure constant, respectively (Lemak and Balabaev, 1994; Bussi et al., 2007). The cutoff distances for Coulomb and van der Waals interactions were set as 1.2 nm. The particle mesh Ewald method (PME) was used to calculate the long-range electrostatic interactions (Darden et al., 1993). The final production run was carried out for 100 ns at a temperature of 300 K and a pressure of 1 bar.

Free energy calculation

The binding free energies for protein and ligand complexes were calculated in an explicit water environment by employing the alchemical method (Supplementary Figure S5). The final

snapshot from the MD simulations (100 ns) was used as the starting point for the free energy simulations. The alchemical method involves two steps: 1) decoupling of the ligand from the protein–ligand complex in an explicit water environment and 2) decoupling of the ligand from the water environment. The decoupling process includes turning off the van der Waals and electrostatic interactions responsible for complex formation (protein–ligand or water–ligand) with the help of the coupling parameter (λ). First, electrostatic interactions were turned off slowly, while the van der Waals interactions were still present. Then, the van der Waals interactions between the protein and ligand (water and ligand) were turned off using the coupling parameter (λ). The electrostatic interactions were turned off by changing λ (0 0.25 0.5 0.75 1.0) from 0 to 1 with a step size of $\Delta\lambda = 0.25$, and the van der Waals interactions were turned off with nonuniformly distributed values of λ (0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.65, 0.7, 0.75, 0.8, 0.85, 0.9, 0.95, 1.0). The same procedure was applied to decouple the ligands from the protein–ligand complex and ligands from the water environment. Therefore, 21 windows, each of 1 ns, were employed to decouple the ligand from the protein–ligand and water–ligand complexes. The free energy difference between the two end states was calculated using the Bennett acceptance ratio (BAR) method (Bennett, 1976). The BAR method is used to estimate the free energy difference between two states with the following equation:

$$\left\langle \frac{1}{1 + \exp\{\beta(\Delta U_{ij} - \Delta G)\}} \right\rangle_i = \left\langle \frac{1}{1 + \exp\{\beta(-\Delta U_{ij} + \Delta G)\}} \right\rangle_j \quad (3)$$

where β is the reciprocal of the thermodynamic temperature, ΔG is the free energy difference between states i and j , and $\Delta U_{ij} = U_j - U_i$ is the potential energy difference.

At each λ -point, the structures were subjected to energy minimization using the steepest descent method. Using Langevin dynamics, the resulting structures were equilibrated in an isothermal-isobaric (NPT) ensemble at 300 K and at a pressure of 1 bar. The rest of the simulation protocol was similar to the protocol followed in the classic MD section.

QM/MM approach

The final snapshots of protein and ligand complexes determined from MD simulations were optimized in the gas phase at the (B3LYP-D3/6-31G*)/Universal force field level of theory with the help of the Gaussian16 package. It has been found in earlier studies that density functionals such as M06-2X, B3LYP-D, and ω B97XD are suitable for investigating noncovalent interactions. Hence, in all calculations, the QM region was optimized with dispersion-corrected B3LYP with the Grimme empirical dispersion functional (B3LYP-D3)

using the 6-31G* basis set. The ligand and surrounding region within 4 Å were treated as the QM region, and the remaining parts were considered the MM region. We extracted only the QM region from the optimized geometries and added terminal hydrogens to calculate binding affinities. The resulting structures were used to calculate the interaction energies with the supermolecule approach at the B3LYP-D3/6-31G* level of theory.

$$IE = EC - (EM1 + EM2) \quad (4)$$

where IE is the interaction energy of the complex, EC is the energy of the complex, EM1 is the energy of the protein part of the complex, and EM2 is the energy of the ligand in the complex. All IEs were corrected for basis set superposition error (BSSE) using the counterpoise method suggested by Boys and Bernadi (Gutowski et al., 1993), as implemented in the Gaussian16 package (Frisch et al., 2016).

Selectivity

In total, 49 kinases were evaluated *via* selectivity score calculation, and those 49 kinases were previously used for the actual kinase panel assay. Representative PDB structures for the 49 kinases were extracted from the RCSB Protein Data Bank (<https://www.rcsb.org/>). The ligands were docked to the binding pocket of each PDB using AutoDock-Vina (Trott and Olson, 2010). The resulting docking score was rescaled to observe and compare compound trends. Quantitatively, the selectivity score was calculated to measure the overall selectivity across different kinase families. A lower selectivity score indicates better selectivity for the tested compound.

Selectivity score (S)

$$= \frac{\text{number of kinases with rescaled docking score less than 0.5}}{\text{total number of kinases tested}} \quad (5)$$

Cell culture and cell viability assay

MV4-11 cells were purchased from American Type Culture Collection (ATCC, VA, United States). The cells were passaged for less than 1 month, and mycoplasma infection was checked by PCR once a week. The growth medium was Iscove's Modified Dulbecco's Medium (IMDM; ThermoFisher, United States) supplemented with 10% fetal bovine serum (FBS; Corning, United States) and 1% penicillin/streptomycin (GenDEPOT, United States). The cells were maintained in a humidified atmosphere with 5% CO₂ at 37°C. Cell viability was determined using the WST-8 assay (Cyto XTM cell viability

assay kit; LPS solution, Daejeon, South Korea) in accordance with the manufacturer's protocol.

Synthesis of PCW-A1001

Unless otherwise stated, all reactions were performed under an inert (N₂) atmosphere. Reagents and solvents were reagent grade and purchased from Sigma-Aldrich, Alfa Aesar, and Combi-Blocks. Anhydrous solvents were purchased from Sigma-Aldrich and used as provided. Reactions were monitored by TLC and visualized with a UV lamp and/or KMnO₄ staining. Silica gel 60 (230–400 mesh, Merck) was used for flash column chromatography. ¹H and ¹³C NMR spectra were recorded on BRUKER Ultrashield 300 and 400 MHz NMR spectrometers at 25°C. Chemical shifts are reported in parts per million (ppm). Data for ¹H NMR are reported as follows: chemical shift (δ ppm) [multiplicity, coupling constant (Hz), integration]. Multiplicities are reported as follows: s = singlet, d = doublet, t = triplet, q = quartet, dd = doublet of doublets, m = multiplet. Data from ¹³C spectra are reported as chemical shifts (δ ppm). The residual solvent peak was used as an internal reference. Mass spectra were obtained on Acquity™ Waters A06UPD9BM and Agilent Technologies SG12109048 systems. Prior to biological testing, the final compound was confirmed to be > 98% pure by UPLC chromatography using a Waters ACQUITY H-class system fitted with a C₁₈ reverse-phase column (ACQUITY UPLC BEH C₁₈; 2.1 mm × 50 mm, Part No. 186002350) according to the following eluent conditions: (A) H₂O + 0.1% formic acid, (B) CH₃CN + 0.1% formic acid, (C) MeOH + 0.1% formic acid; (I) a gradient of 95% A to 95% B over 5 min; and (II) a gradient of 95% A to 95% C over 5 min.

Conclusion

In this work, we used an AI-assisted *de novo* drug design (LSTM) approach to identify a novel FLT-3 inhibitor that selectively targets the FLT-3 (D835Y) mutant. The deep learning model was pretrained on a known bioactive chemical space (ChEMBL22), and the generated hits were fine-tuned using our in-house FLT-3 inhibitors. The generated hits were further evaluated and filtered using various parameters focusing on their novelty, similarities, diversities, etc. We further evaluated the toxicities of the *de novo* molecules with our in-house program Pharmulator™. Among the screened hits, only 146 compounds passed the toxicity filters. The binding affinities, conformations and interaction patterns of these screened compounds were studied with WT FLT-3 and its mutant (D835Y). Since the FLT-3 (D835Y) mutant structure in the DFG-in conformation was unavailable, we modeled the protein to validate the compounds in terms of the binding interactions. The

stabilities of complexes were further validated qualitatively with MD simulations and quantitatively with free energy calculations.

The top compound, named PCW-A1001, was considered for synthesis and screening studies. The anticancer activity was tested against MV4-11 cells to verify the effectiveness of these compounds in AML treatment. PCW-A1001 was found to be a promising inhibitor of FLT3, and it showed an IC₅₀ of 764 nM against the FLT-3 (D835Y) mutant and 2.54 μM against WT FLT-3. PCW-A1001 also showed an IC₅₀ of 1.98 μM against MV4-11-cell line screening. We successfully implemented reverse network theory and AI-based *de novo* design strategies and identified a potential inhibitor of the FLT3/FLT3 (D835Y) mutant, PCW-A1001. AI generated a hit, PCW-A1001 exhibited better activity than the parent compound, PCW-1001. Further fine-tuning of PCW-A1001 is in progress to optimize the selectivity and activity and will be reported in due course.

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

SJ, DS, and SM—performed and analyzed the computational part, JC and SL—designed and analyzed the network part, MJ and SB—performed the AI part, ES, KL, and K-YJ—performed synthesis and characterization of the compound, MK and J-SK—performed the Cell assay, JH—analysis of experimental data, SW—conceived and designed the experiments, DS, SM, SB, K-YJ, and J-SK—writing-original draft, and SW—conceptualization and draft review.

Conflict of interest

Authors SJ, DS, SM, JC, SL, MJ, SB, JH, and SW were employed by PharmCADD.

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.

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.

References

- Abraham, M. J., Murtola, T., Schulz, R., Páll, S., Smith, J. C., Hess, B., et al. (2015). Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* 1 (2), 19–25. doi:10.1016/j.softx.2015.06.001
- Ambinder, A. J., and Levis, M. (2021). Potential targeting of FLT3 acute myeloid leukemia. *Haematologica* 106, 671–681. doi:10.3324/haematol.2019.240754
- Assouline, S., Cocolakis, E., and Borden, K. L. B. (2012). The development of novel therapies for the treatment of acute myeloid leukemia (AML). *Cancers* 4, 1161–1179. doi:10.3390/cancers4041161
- Auclair, D., Miller, D., Yatsula, V., Pickett, W., Carter, C., Chang, Y., et al. (2007). Antitumor activity of sorafenib in FLT3-driven leukemic cells. *Leukemia* 21, 439–445. doi:10.1038/sj.leu.2404508
- Awale, M., Sirockin, F., Stiefl, N., and Reymond, J.-L. (2019). Drug analogs from fragment-based long short-term memory generative neural networks. *J. Chem. Inf. Model.* 59, 1347–1356. doi:10.1021/acs.jcim.8b00902
- Banks, R. D., Blake, C. C. F., Evans, P. R., Haser, R., Rice, D. W., Hardy, G. W., et al. (1979). Sequence, structure and activity of phosphoglycerate kinase: A possible hinge-bending enzyme. *Nature* 279, 773–777. doi:10.1038/279773a0
- 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
- Berger, S. I., and Iyengar, R. (2009). Network analyses in systems pharmacology. *Bioinformatics* 25, 2466–2472. doi:10.1093/bioinformatics/btp465
- Brown, P., and Small, D. (2004). FLT3 inhibitors: A paradigm for the development of targeted therapeutics for paediatric cancer. *Eur. J. Cancer* 40, 707–721. doi:10.1016/j.ejca.2003.08.030
- Bussi, G., Donadio, D., and Parrinello, M. (2007). Canonical sampling through velocity rescaling. *J. Chem. Phys.* 126, 014101. doi:10.1063/1.2408420
- Chang Hsu, Y., Ke, Y.-Y., Shiao, H.-Y., Lee, C.-C., Lin, W.-H., Chen, C.-H., et al. (2014). Facile identification of dual FLT3–aurora A inhibitors: A computer-guided drug design approach. *ChemMedChem* 9, 953–961. doi:10.1002/cmdc.201300571
- Corey, E. J., and Wipke, W. T. (1969). Computer-assisted design of complex organic syntheses. *Science* 166, 178–192. doi:10.1126/science.166.3902.178
- 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
- Degen, J., Wegscheid-Gerlach, C., Zaliani, A., and Rarey, M. (2008). On the art of compiling and using “drug-like” chemical fragment spaces. *ChemMedChem* 3, 1503–1507. doi:10.1002/cmdc.200800178
- Del Zotto, G., Luchetti, F., and Zamai, L. (2001). CD135. *J. Biol. Regul. Homeost. Agents* 15, 103–106.
- Erikawa, D., Yasuo, N., and Sekijima, M. (2021). Mermaid: An open source automated hit-to-lead method based on deep reinforcement learning. *J. Cheminform.* 13, 94. doi:10.1186/s13321-021-00572-6
- Friesner, R. A., Murphy, R. B., Repasky, M. P., Frye, L. L., Greenwood, J. R., Halgren, T. A., et al. (2006). Extra precision Glide: Docking and scoring incorporating a model of hydrophobic enclosure for Protein–Ligand complexes. *J. Med. Chem.* 49, 6177–6196. doi:10.1021/jm051256o
- 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*. Willingford, CT.
- Gebru, M. T., and Wang, H.-G. (2020). Therapeutic targeting of FLT3 and associated drug resistance in acute myeloid leukemia. *J. Hematol. Oncol.* 13, 155. doi:10.1186/s13045-020-00992-1
- Grafone, T., Palmisano, M., Nicci, C., and Storti, S. (2012). An overview on the role of FLT3-tyrosine kinase receptor in acute myeloid leukemia: Biology and treatment. *Oncol. Rev.* 6, e8. doi:10.4081/oncol.2012.e8
- Grunwald, M. R., and Levis, M. J. (2013). FLT3 inhibitors for acute myeloid leukemia: A review of their efficacy and mechanisms of resistance. *Int. J. Hematol.* 97, 683–694. doi:10.1007/s12185-013-1334-8
- Guney, E., Menche, J., Vidal, M., and Barabási, A.-L. (2016). Network-based *in silico* drug efficacy screening. *Nat. Commun.* 7, 10331. doi:10.1038/ncomms10331
- Gupta, A., Müller, A. T., Huisman, B. J. H., Fuchs, J. A., Schneider, P., and Schneider, G. (2018). Generative recurrent networks for de novo drug design. *Mol. Inf.* 37, 1700111. doi:10.1002/minf.201700111
- Gutowski, M., van Duijneveldt-van de Rijdt, J. G. C. M., van Lenthe, J. H., and van Duijneveldt, F. B. (1993). Accuracy of the Boys and Bernardi function counterpoise method. *J. Chem. Phys.* 98, 4728–4737. doi:10.1063/1.465106
- Hayakawa, F., Towatari, M., Kiyoi, H., Tanimoto, M., Kitamura, T., Saito, H., et al. (2000). Tandem-duplicated Flt3 constitutively activates STAT5 and MAP kinase and introduces autonomous cell growth in IL-3-dependent cell lines. *Oncogene* 19, 624–631. doi:10.1038/sj.onc.1203354
- Heinrich, M. C., Griffith, D., McKinley, A., Patterson, J., Presnell, A., Ramachandran, A., et al. (2012). Crenolanib inhibits the drug-resistant PDGFRA D842V mutation associated with imatinib-resistant gastrointestinal stromal tumors. *Clin. Cancer Res.* 18, 4375–4384. doi:10.1158/1078-0432.CCR-12-0625
- Hess, B., Bekker, H., Berendsen, H. J. C., and Fraaije, J. G. E. M. (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
- Hessler, G., and Baringhaus, K.-H. (2018). Artificial intelligence in drug design. *Molecules* 23, E2520. doi:10.3390/molecules23102520
- Huang, J., and MacKerell, A. D. (2013). CHARMM36 all-atom additive protein force field: Validation based on comparison to NMR data. *J. Comput. Chem.* 34, 2135–2145. doi:10.1002/jcc.23354
- Kang, M., Pandit, N., Kim, A.-Y., Cho, S. J., Kwon, Y.-J., Ahn, J., et al. (2022). PCW-1001, a novel pyrazole derivative, exerts antitumor and radio-sensitizing activities in breast cancer. *Front. Oncol.* 12, 835833. doi:10.3389/fonc.2022.835833
- Kazi, J. U., and Rönnstrand, L. (2019). FMS-Like tyrosine kinase 3/FLT3: From basic science to clinical implications. *Physiol. Rev.* 99, 1433–1466. doi:10.1152/physrev.00029.2018
- Ke, Y.-Y., Singh, V. K., Coumar, M. S., Hsu, Y. C., Wang, W.-C., Song, J.-S., et al. (2015). Homology modeling of DFG-in FMS-like tyrosine kinase 3 (FLT3) and structure-based virtual screening for inhibitor identification. *Sci. Rep.* 5, 11702. doi:10.1038/srep11702
- Kindler, T., Lipka, D. B., and Fischer, T. (2010). FLT3 as a therapeutic target in AML: Still challenging after all these years. *Blood* 116, 5089–5102. doi:10.1182/blood-2010-04-261867
- Landrum, G. (2016). RDKit: Open-Source cheminformatics software. Available at: https://github.com/rdkit/rdkit/releases/tag/Release_2016_09_4.
- Lemak, A. S., and Balabaev, N. K. (1994). On the berendsen thermostat. *Mol. Simul.* 13, 177–187. doi:10.1080/08927029408021981
- Leung, A. Y. H., Man, C.-H., and Kwong, Y.-L. (2013). FLT3 inhibition: A moving and evolving target in acute myeloid leukaemia. *Leukemia* 27, 260–268. doi:10.1038/leu.2012.195
- Liang, D.-C., Shih, L.-Y., Hung, I.-J., Yang, C.-P., Chen, S.-H., Jaing, T.-H., et al. (2003). FLT3-TKD mutation in childhood acute myeloid leukemia. *Leukemia* 17, 883–886. doi:10.1038/sj.leu.2402928
- Lin, W.-H., Jiaang, W.-T., Chen, C.-W., Yen, K.-J., Hsieh, S.-Y., Yen, S.-C., et al. (2012). BPR1J-097, a novel FLT3 kinase inhibitor, exerts potent inhibitory activity against AML. *Br. J. Cancer* 106, 475–481. doi:10.1038/bjc.2011.564
- Liu, X. (2017). Deep recurrent neural network for protein function prediction from sequence. *arXiv* 2017, 08318. doi:10.48550/arXiv.1701.08318
- Lowe, D. (2017). *Chemical reactions from US patents (1976-Sep2016)*. doi:10.6084/m9.figshare.5104873.v1
- Luo, Y., Peng, J., and Ma, J. (2022). Next decade's AI-based drug development Features tight integration of data and computation. *Health Data Sci.* 2022, 1–3. doi:10.34133/2022/9816939

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmolb.2022.1072028/full#supplementary-material>

- 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
- McDonnell, L. M., Kernohan, K. D., Boycott, K. M., and Sawyer, S. L. (2015). Receptor tyrosine kinase mutations in developmental syndromes and cancer: Two sides of the same coin. *Hum. Mol. Genet.* 24, R60–R66. doi:10.1093/hmg/ddv254
- Mendez, D., Gaulton, A., Bento, A. P., Chambers, J., De Veij, M., Félix, E., et al. (2019). ChEMBL: Towards direct deposition of bioassay data. *Nucleic Acids Res.* 47, D930–D940. doi:10.1093/nar/gky1075
- Merk, D., Friedrich, L., Grisoni, F., and Schneider, G. (2018). De novo design of bioactive small molecules by artificial intelligence. *Mol. Inf.* 37, 1700153. doi:10.1002/minf.201700153
- O'Hare, T., Shakespeare, W. C., Zhu, X., Eide, C. A., Rivera, V. M., Wang, F., et al. (2009). AP24534, a pan-BCR-ABL inhibitor for chronic myeloid leukemia, potently inhibits the T315I mutant and overcomes mutation-based resistance. *Cancer Cell* 16, 401–412. doi:10.1016/j.ccr.2009.09.028
- Polykovskiy, D., Zhebrak, A., Sanchez-Lengeling, B., Golovanov, S., Tatanov, O., Belyaev, S., et al. (2020). Molecular sets (MOSES): A benchmarking platform for molecular generation models. *Front. Pharmacol.* 11, 565644. doi:10.3389/fphar.2020.565644
- Quentmeier, H., Reinhardt, J., Zaborski, M., and Drexler, H. G. (2003). FLT3 mutations in acute myeloid leukemia cell lines. *Leukemia* 17, 120–124. doi:10.1038/sj.leu.2402740
- Rosnet, O., Bühring, H. J., Marchetto, S., Rappold, I., Lavagna, C., Sainty, D., et al. (1996). Human FLT3/FLK2 receptor tyrosine kinase is expressed at the surface of normal and malignant hematopoietic cells. *Leukemia* 10, 238–248.
- Sali, A., and Blundell, T. L. (1993). Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* 234, 779–815. doi:10.1006/jmbi.1993.1626
- Sastry, G. M., Adzhigirey, M., Day, T., Annabhimoju, R., and Sherman, W. (2013). Protein and ligand preparation: Parameters, protocols, and influence on virtual screening enrichments. *J. Comput. Aided. Mol. Des.* 27, 221–234. doi:10.1007/s10822-013-9644-8
- Savage, D. G., and Antman, K. H. (2002). Imatinib mesylate — a new oral targeted therapy. *N. Engl. J. Med.* 346, 683–693. doi:10.1056/NEJMra013339
- Schittenhelm, M. M., Yee, K. W. H., Tyner, J. W., McGreevey, L., Haley, A. D., Town, A., et al. (2006). FLT3 K663Q is a novel AML-associated oncogenic kinase: Determination of biochemical properties and sensitivity to Sunitinib (SU11248). *Leukemia* 20, 2008–2014. doi:10.1038/sj.leu.2404374
- Scholl, S., Fleischmann, M., Schnetzke, U., and Heide, F. H. (2020). Molecular mechanisms of resistance to FLT3 inhibitors in acute myeloid leukemia: Ongoing challenges and future treatments. *Cells* 9, 112493. doi:10.3390/cells9112493
- Segler, M. H. S., Kogej, T., Tyrchan, C., and Waller, M. P. (2018). Generating focused molecule libraries for drug discovery with recurrent neural networks. *ACS Cent. Sci.* 4, 120–131. doi:10.1021/acscentsci.7b00512
- Shirts, M. R., Klein, C., Swails, J. M., Yin, J., Gilson, M. K., Mobley, D. L., et al. (2017). Lessons learned from comparing molecular dynamics engines on the SAMPL5 dataset. *J. Comput. Aided. Mol. Des.* 31, 147–161. doi:10.1007/s10822-016-9977-1
- Sivakumar, D., Gorai, B., and Sivaraman, T. (2013). Screening efficient BH3-mimetics to hBcl-B by means of peptidodynamic method. *Mol. Biosyst.* 9, 700–712. doi:10.1039/C2MB25195G
- Smith, B. D., Levis, M., Beran, M., Giles, F., Kantarjian, H., Berg, K., et al. (2004). Single-agent CEP-701, a novel FLT3 inhibitor, shows biologic and clinical activity in patients with relapsed or refractory acute myeloid leukemia. *Blood* 103, 3669–3676. doi:10.1182/blood-2003-11-3775
- Smith, C. C., Wang, Q., Chin, C.-S., Salerno, S., Damon, L. E., Levis, M. J., et al. (2012). Validation of ITD mutations in FLT3 as a therapeutic target in human acute myeloid leukaemia. *Nature* 485, 260–263. doi:10.1038/nature11016
- Stirewalt, D. L., and Radich, J. P. (2003). The role of FLT3 in hematopoietic malignancies. *Nat. Rev. Cancer* 3, 650–665. doi:10.1038/nrc1169
- Stone, R. M., De Angelo, J., Galinsky, I., Estey, E., Klimek, V., Grandin, W., et al. (2004). PKC 412 FLT3 inhibitor therapy in AML: Results of a phase II trial. *Ann. Hematol.* 83 (1), S61–S64. doi:10.1007/s00277-004-0850-2
- Tirado-Rives, J., and Jorgensen, W. L. (2008). Performance of B3LYP density functional methods for a large set of organic molecules. *J. Chem. Theory Comput.* 4, 297–306. doi:10.1021/ct700248k
- Trott, O., and Olson, A. J. (2010). AutoDock vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization and multithreading. *J. Comput. Chem.* 31, 455–461. doi:10.1002/jcc.21334
- Weininger, D. (1988). SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. *J. Chem. Inf. Model.* 28, 31–36. doi:10.1021/ci00057a005
- Wodicka, L. M., Ciceri, P., Davis, M. I., Hunt, J. P., Floyd, M., Salerno, S., et al. (2010). Activation state-dependent binding of small molecule kinase inhibitors: Structural insights from biochemistry. *Chem. Biol.* 17, 1241–1249. doi:10.1016/j.chembiol.2010.09.010
- Wu, Z., Wang, Y., and Chen, L. (2013). Network-based drug repositioning. *Mol. Biosyst.* 9, 1268–1281. doi:10.1039/C3MB25382A
- Yamamoto, Y., Kiyoi, H., Nakano, Y., Suzuki, R., Kidera, Y., Miyawaki, S., et al. (2001). Activating mutation of D835 within the activation loop of FLT3 in human hematologic malignancies. *Blood* 97, 2434–2439. doi:10.1182/blood.v97.8.2434
- Zarrinkar, P. P., Gunawardane, R. N., Cramer, M. D., Gardner, M. F., Brigham, D., Belli, B., et al. (2009). AC220 is a uniquely potent and selective inhibitor of FLT3 for the treatment of acute myeloid leukemia (AML). *Blood* 114, 2984–2992. doi:10.1182/blood-2009-05-222034
- Zhang, B., Li, J., and Lü, Q. (2018). Prediction of 8-state protein secondary structures by a novel deep learning architecture. *BMC Bioinforma.* 19, 293. doi:10.1186/s12859-018-2280-5
- Zhang, S., and Broxmeyer, H. E. (1999). p85 subunit of PI3 kinase does not bind to human Flt3 receptor, but associates with SHP2, SHIP, and a tyrosine-phosphorylated 100-kDa protein in Flt3 ligand-stimulated hematopoietic cells. *Biochem. Biophys. Res. Commun.* 254, 440–445. doi:10.1006/bbrc.1998.9959
- Zhavoronkov, A., Ivanenkov, Y. A., Aliper, A., Veselov, M. S., Aladinskiy, V. A., Aladinskaya, A. V., et al. (2019). Deep learning enables rapid identification of potent DDR1 kinase inhibitors. *Nat. Biotechnol.* 37, 1038–1040. doi:10.1038/s41587-019-0224-x



OPEN ACCESS

EDITED BY

Emilia Pedone,
Department of Biomedical Sciences,
National Research Council (CNR), Italy

REVIEWED BY

Chengwen Liu,
The University of Texas at Austin,
United States
Ramar Vanajothi,
Bharathidasan University, India

*CORRESPONDENCE

Ladislav Kalvoda,
ladislav.kalvoda@ujf.cvu.cz

SPECIALTY SECTION

This article was submitted to Biophysics,
a section of the journal
Frontiers in Molecular Biosciences

RECEIVED 06 October 2022

ACCEPTED 18 November 2022

PUBLISHED 07 December 2022

CITATION

Luštinec J, Koubský T and Kalvoda L
(2022), Ab-initio evaluation of acid
influence on chemical stability of
hydrophilic diglycolamides.
Front. Mol. Biosci. 9:1063022.
doi: 10.3389/fmolb.2022.1063022

COPYRIGHT

© 2022 Luštinec, Koubský and Kalvoda.
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.

Ab-initio evaluation of acid influence on chemical stability of hydrophilic diglycolamides

Jakub Luštinec, Tomáš Koubský and Ladislav Kalvoda*

Department of Solid State Engineering, Faculty of Nuclear Sciences and Physical Engineering, Czech Technical University in Prague, Prague, Czechia

Diglycolamides (DGA) form one of the most promising groups of organic ligands used in bio-inspired solvent extraction processes of lanthanide and actinide ions. Continuous experimental and theoretical research is still performed in order to further improve their application properties including their chemical stability in the real extraction environment. This work provides results of our theoretical approach focused on inclusion of an acid influence on the DGAs chemical structure, treated in frame of the density functional theory. Three different models describing the acid action are proposed and investigated in attempt to increase the resulting accuracy of the chemical stability predictions based on verified theoretical descriptors. The procedure is applied and tested on the set of selected hydrophilic DGA representatives. Comparison of the model results obtained with and without acid action shows that two types of protection effects may occur: a 'direct' protection, accompanied by an explicit change of the ligand stability indicators, and an 'indirect' one consisting in reaction of acid molecules with radicals preceding the contact of latter with the extracting ligands. The possibility of the direct acid protection route is supported by the significant decrease of the Fukui charges found with the acid models included. On the other hand, there is in general no significant difference of trends in the calculated chemical stability descriptors suggesting that an indirect mechanism must be also considered in order to explain the experimentally observed protective role of acids on the chemical stability of investigated DGA derivatives.

KEYWORDS

diglycolamides, hydrophilic DGA, density functional theory, radiolytic stability, acid influence

Introduction

Closing the nuclear fuel cycle and the maximal usage of uranium contained in the nuclear fuel are one of the biggest recent challenges related to a sustainable nuclear power plants operation. The spent nuclear fuel contains fission products, unused U, and minor actinides (MAs) which are, together with Pu, also responsible for the long term radiotoxicity of the nuclear waste. One of the possibilities for the spent nuclear fuel treatment is its partitioning and transmutation (Veliscek-Carolan, 2016). Partitioning consists of co-extraction of MAs and lanthanides from a liquid solution of nuclear waste

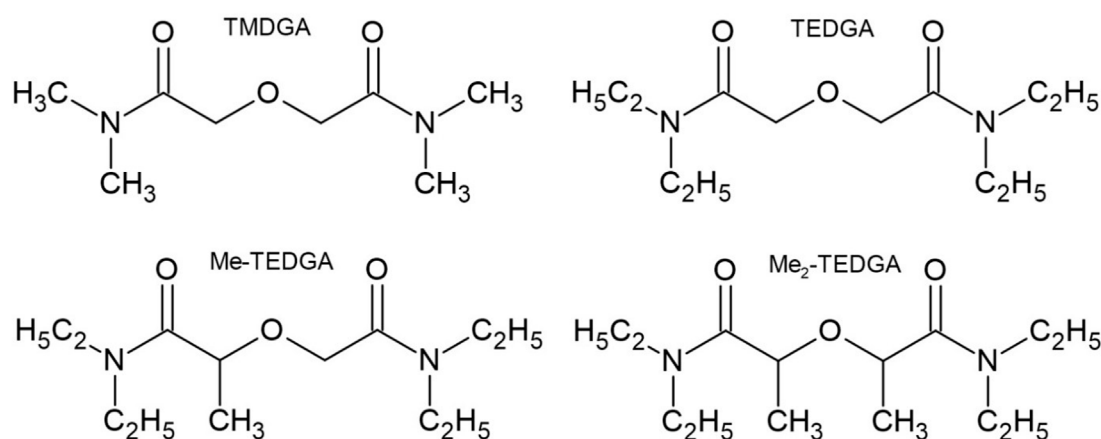


FIGURE 1
Studied hydrophilic DGA derivatives.

and the subsequent separation of these components from each other. After partitioning, the transmutation of MAs is performed in order to reduce their long radiative decay lifetime and toxicity. In the result, the radiotoxicity of MAs and remaining waste can be significantly reduced (Veliscek-Carolan, 2016).

Among other organic molecules used in solvent extraction procedures, one of the most promising groups is formed by diglycolamides (DGAs). DGAs are especially well applicable for extraction of trivalent actinide and lanthanide ions, mimicking thus some of the processes observed in biological systems (Mattocks and Cotruvo, 2020). The solubility of DGAs is determined by the optional presence and length of specific N-alkyl substituents. The DGA derivatives containing short N-alkyls are water soluble, e.g., tertamethyldiglycolamide (TMDGA) or tetraethyldiglycolamide (TEDGA). These hydrophilic molecules are used as aqueous stripping and back-holding agents (Sasaki et al., 2007) in ALSEP process (Actinide Lanthanide Separation) (Lumetta et al., 2014) or EXAm process (Extraction of Americium) (Rostaing et al., 2012).

In this work, we study the following DGAs representatives (Figure 1).

- TMDGA (N,N,N',N'-tetramethyl-diglycolamide; 2,2'-oxybis (N,N-dimethylacetamide))
- TEDGA (N,N,N',N'-tetraethyl-diglycolamide; 2,2'-oxybis (N,N-diethylacetamide))
- Me-TEDGA (2-(2-(diethylamino)-2-oxoethoxy)-N,N-diethylpropanamide)
- Me₂-TEDGA (2,2'-oxybis (N,N-diethylpropanamide)).

As indicated by experimental results (Wilden et al., 2018; Horne et al., 2019), addition of methyl group in Me-TEDGA and

Me₂-TEDGA results in remarkable radiolytic stability enhancement compared to the un-methylated TEDGA.

In our previous work (Koubský and Luštinec, 2018), the quantum mechanical indicators of radiolytic stability of the four above mentioned hydrophilic DGA derivatives were already evaluated and analysed in the environment of pure water. Especially, radical Fukui function, Fukui charges (condensed Fukui functions), and bond orders were found to be of key importance for the related radiolytic stability predictions. This work extends the theoretical treatment on the problem of acid influence implementation into the DFT calculations in order to improve the stability predictions. In particular, above mentioned verified stability indicators are calculated using three different acid inclusion models, and the obtained results compared and juxtaposed with the former acid-free results (Koubský and Luštinec, 2018). In addition, the condensed dual descriptor (CDD) Δf^A proposed by (Morell et al., 2005; Morell et al., 2006) and applied by Smirnova et al. (2020) for the radiolytic stability predictions of several extractants is evaluated and discussed.

Reference experimental studies

Two key experimental radiolytic stability studies of the hydrophilic DGA derivatives shown in Figure 1 were performed by Wilden et al. (2018) and Horne et al. (2019).

In the first study, Wilden and co-workers tested solutions of DGA derivatives in a warmed nitric acid environment and compared the results with the behaviour in a pure water environment. In addition to the steady state measurements, the pulsed radiolysis method followed by kinetics

measurements and mass spectroscopy was also applied, in order to get a deeper insight into the radiolysis process of the hydrophilic DGA derivatives. High rate constants for DGAs reactions with the hydroxyl radical are found, suggesting the important role of this radical in the radiolytic degradation mechanism in water environment. Observed decrease in dose constants with the increasing molecular weight of DGAs suggests, together with the measured rate constants, an electron transfer as the mechanism of the radical reaction. The radiolytic stability found for studied molecules follows the trend TMDGA < TEDGA < Me-TEDGA < Me₂-TEDGA, i.e., growths with the molecular weight of the tested derivative (Wilden et al., 2018).

The second experimental study performed by Horne et al. (2019) (using the setup similar to the setup used by Wilden and co-workers) deals with a neutral pH concentrated aqueous nitrate solutions of the four selected hydrophilic DGA derivatives. The authors conclude that the studied hydrophilic DGAs undergo a first-order decay; the observed degradation product distributions are similar to those found in Wilden et al. (2018) under pure water conditions (except for the additional appearance of NO_x adducts), and the radiolysis is driven by hydroxyl and nitrate radical oxidation chemistry, the latter then likely moderated by some secondary reactions scavenging the degradation products (Horne et al., 2019). The radiolysis rate of hydrophilic DGA representatives in aqueous nitrate solutions is found to be significantly reduced and less structurally sensitive compared to the acid-free solutions, similarly to the situation already observed for lipophilic DGA derivatives (Galan et al., 2015).

Degradation reaction mechanisms

As in our previous work (Koubský and Luštinec, 2018), we generally consider indirect radiolysis mechanisms to prevail, because of the low actual concentration of ligands used in extraction solutions (Wilden et al., 2018). Such indirect process consists in the primary radiolysis of solvent molecules, followed then by reactions of the radiolysis products with the ligand molecules. These water radiolysis products are represented by OH• and H• radicals. The overall reaction conditions anticipated in this work follow the experiments performed by Wilden et al. (2018), adding to the pure water environment [considered in the previous study (Koubský and Luštinec, 2018)] the influence of nitric acid *via* the proposed acid models.

The first degradation reaction mechanism considered here is supposed to start with the hydrogen abstraction followed by rupture of the ether bond C-O, in analogy with the lipophilic DGAs (Koubský et al., 2017). In the previous works (Koubský et al., 2017; Koubský and Luštinec, 2018), it was concluded that the hydrogen abstraction is more probable to occur on the

ether group than on the side chains. Therefore, the ether group is mainly investigated in this work. It is worth to notice that the methylation of the ether carbons that promotes the higher molecular stability, lowers also the number of ether hydrogens: four in case of the two non-methylated studied DGA derivatives TMDGA and TEDGA, three in case of Me-TEDGA, and two for Me₂-TEDGA.

The second possible degradation reaction mechanism follows the findings of Wilden et al. (2018). This mechanism is based on the known oxidation nature of the hydroxyl radical OH• that could cause the electron transfer from the amide group producing the DGA radical cation [DGA]^{•+}. Afterwards, the rupture of the ether C-O bonds or the amide C-N bonds occurs (Wilden et al., 2018).

Methods and computational settings

Applied acid models

The acid influence is implemented into DFT simulations of the selected hydrophilic DGA derivatives by setting up three different testing models. In the first two of them, the interaction of ligand with dissociated acid molecules is assumed. These two models are independent of the particular acid type used in the experiments. In the first, simplest model, the hydrogen cation H⁺ is added to the calculated system in order to create a complex with the extraction molecule. In the second tested model, H₃O⁺ cation (also applied in (Matveev et al., 2017)) is included in calculations instead of H⁺. Finally, in the third model, the interaction with undissociated HNO₃ molecule is tested. The latter model is also relevant since the DGAs extractants are commonly dissolved in nitric acid solutions concentrated enough to contain a significant amount of undissociated HNO₃ molecules (Wilden et al., 2018). This model explicitly includes the specific acid, in contrary to the first two models using general acid representations. The particular issue complicating mutual comparison of the results obtained by these models follows from the different total charge of the studied systems: it is equal to +1 for the first two models, and to zero for the last one. Thus, the behaviour of valence electrons and the related stability descriptors are affected by this difference.

Calculated stability indicators

The Frontier orbital theory of Fukui (1982) relates the molecule reactivity to the charge density $\rho(r)$ with respect to electrophilic and nucleophilic properties of the reaction. Further developed by Parr and Yang (1989), Fukui functions are practical tool for qualitatively measuring and displaying the reactive

regions of molecules. Fukui functions describe the sensitivity of charge density to losing or gaining electrons as follows

$$f^+(r) = \frac{1}{\Delta N} (\rho_{N+\Delta}(r) - \rho_N(r))$$

$$f^-(r) = \frac{1}{\Delta N} (\rho_N(r) - \rho_{N-\Delta}(r)).$$

Where $f^+(r)$ and $f^-(r)$ is respectively nucleophilic and electrophilic Fukui function, and ΔN is a change in the number of electrons. Radical Fukui function $f^0(r)$ is then obtained as the average of the nucleophilic and electrophilic Fukui function.

Another possible implementation of the Fukui theory is the condensed Fukui function giving Fukui charges. Fukui charges are calculated from the atomic charges q_A as follows

$$f_A^+ = q_A^{\text{anion}} - q_A$$

$$f_A^- = q_A - q_A^{\text{cation}}$$

Here f_A^+ and f_A^- is respectively the nucleophilic and electrophilic Fukui charge on the atom A. Radical Fukui charge f_A^0 on atom A is then given as the average of nucleophilic and electrophilic Fukui charges.

$$f_A^0 = \frac{1}{2} (f_A^+ + f_A^-)$$

The CDD (Morell et al., 2005; Morell et al., 2006) for atom A is defined as follows

$$\Delta f_A = f_A^+ - f_A^-$$

The sign of CDD indicates the vulnerability of the atomic site to the particular type of radical attack: the negative sign relates to electrophilic attack, the positive sign then to nucleophilic one.

Wiberg bond indices W_{AB} (Wiberg, 1968) are calculated from the electronic overlap between two atoms, A and B, as follows

$$W_{AB} = \sum_{\mu \in A} \sum_{\sigma \in B} P_{\mu\sigma}^2$$

Here μ and σ is atomic orbital on atom A and B, respectively, and $P_{\mu\sigma}$ is the corresponding density matrix element.

Computational settings

The DFT calculations were performed with DMol³ module from Materials Studio 8.0 (Delley, 1990; Delley, 2000) and Gaussian09 code (Frisch et al., 2013). The conformation analysis of the models combining extractants with the selected acid representation was performed using Gaussian09 code; the initial optimised extractants conformations were taken from the previous work (Koubský and Luštinec, 2018) where the geometry optimization was performed firstly with BLYP and subsequently with B3LYP functional.

The H⁺ and H₃O⁺ cations were then added in vicinity of the atoms possessing negative partial charge, as calculated in Koubský and Luštinec (2018). In the case of the third model employing the undissociated nitric acid molecule, eight different initial conformations were generated for each of the tested extractants, with the HNO₃ molecule placed gradually into eight different positions in vicinity of the carbonyl oxygens, ether oxygen and the two amide nitrogen atoms, which are all likely to create hydrogen bonds with the HNO₃ molecule.

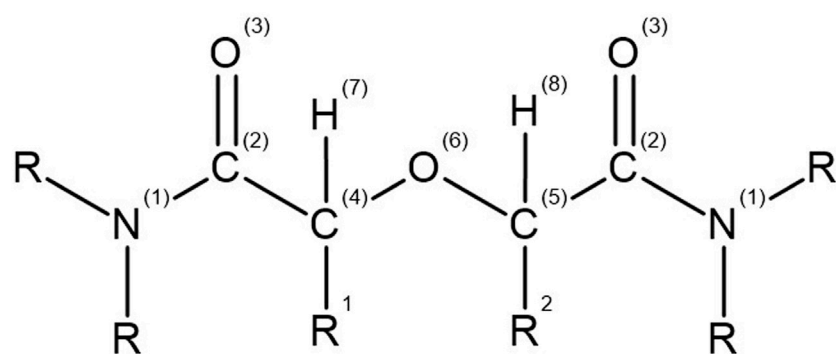
Geometry optimization of the initial conformations was performed using Gaussian09 code with the following settings: 6-31G (d,p) basis set (Petersson et al., 1988; Petersson and Al-Laham, 1991), PCM solvent model (Miertuš et al., 1981; Tomasi et al., 2005) with water taken as the solvent, GD3BJ dispersion correction (Grimme et al., 2010; Grimme et al., 2011), and B3LYP exchange and correlation functional (Beck, 1993). Gaussian09 code with NBO 6.0 (Glendening et al., 2013) was then used for the Natural population analysis and the Wiberg bond indices (Wiberg, 1968) calculations (the latter providing bond orders discussed in Section 3.4) with the same settings as the ones used in the conformation analysis.

Fukui functions, Fukui charges, and CDDs were calculated with DMol³ code using the following settings: DNP basis set (Delley, 1990), COSMO solvent model (Klamt and Schuurmann, 1993; Tomasi and Persico, 1994), GD2 dispersion correction (Grimme, 2006), and B3LYP exchange and correlation functional (Beck, 1993). The differences between PCM and COSMO implicit solvent models consists in the particular way in which the cavity containing the studied system is created. The specific choice of the implicit solvent model in the performed calculations was conditioned by the capabilities of the used software tool.

Results and discussion

At first, the geometrical optimization of the studied acid model systems was performed, leading to optimized conformations, indicating a localized direct interaction of the acid representatives with the extractant molecule, mediated by hydrogen atoms. These optimized conformation were then used in subsequent calculations of the studied stability indicators. The atomic denotation used in the calculations and discussion of the results achieved for the studied structures is shown in Figure 2.

As in the previous work (Koubský and Luštinec, 2018), arithmetic averages are used for the symmetrically equivalent atoms to simplify the stability indicators analysis and also to reduce the conformation dependence of the results. Structure of Me-TEDGA is unsymmetrical due to the methyl group bonded on one of the ether carbons C (4,5). For this reason, the atoms C (4,5) are considered as inequivalent; the remaining atoms symmetrical against oxygen O (6) are analyzed as being equivalent ones.



TMDGA: $R = \text{CH}_3$, $R^1 = \text{H}$, $R^2 = \text{H}$,
 TEDGA: $R = \text{C}_2\text{H}_5$, $R^1 = \text{H}$, $R^2 = \text{H}$,
 Me-TEDGA: $R = \text{C}_2\text{H}_5$, $R^1 = \text{CH}_3$, $R^2 = \text{H}$,
 Me₂-TEDGA: $R = \text{C}_2\text{H}_5$, $R^1 = \text{CH}_3$, $R^2 = \text{CH}_3$,

FIGURE 2

General chemical structure of the studied DGA derivatives with the atom labels indicated.

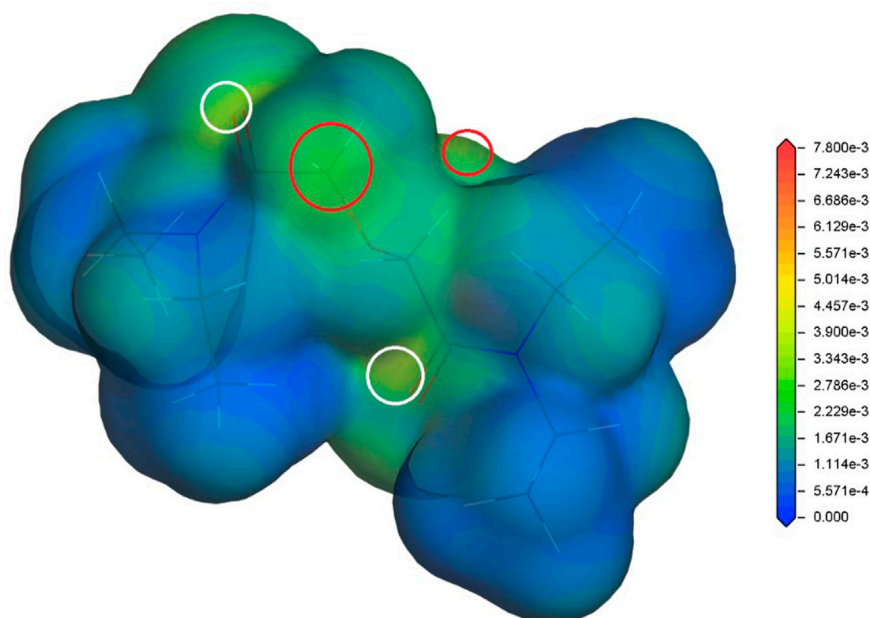


FIGURE 3

Radical Fukui function for TEDGA/acid-free model (Koubský and Luštinec, 2018), mapped on the electron density iso-surface $0.017 \text{ e}\text{\AA}^{-3}$; the red circles mark the FF maxima on hydrogens adjacent to ether group, the white circles identify maxima on amide groups (Dmol³, DNP, B3LYP).

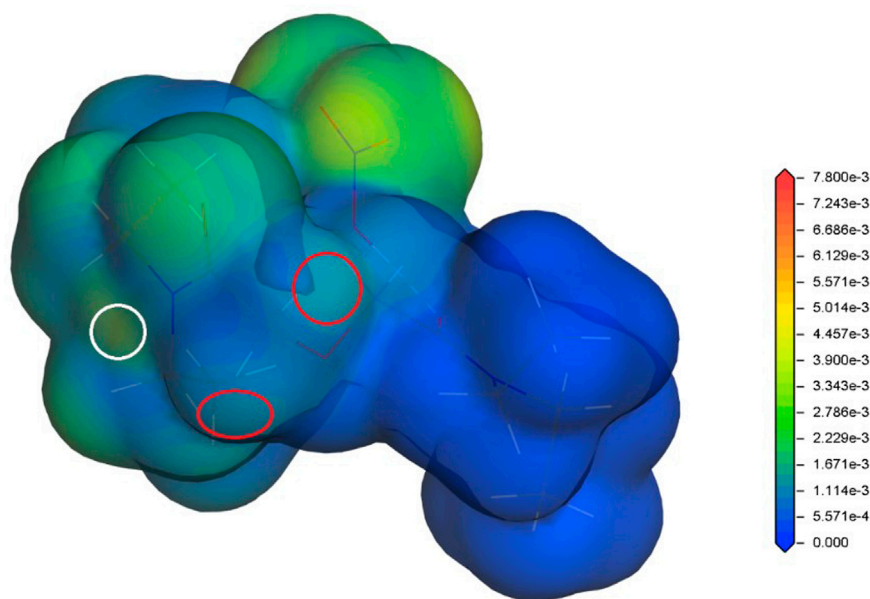


FIGURE 4

Radical Fukui function obtained for TEDGA/HNO₃ acid model, mapped on the electron density iso-surface 0.017 eÅ⁻³; the red circles mark the FF maxima on hydrogens adjacent to ether group, the white circles identify the maxima on amide groups (Dmol³, DNP, B3LYP).

TABLE 1 Values of the radical Fukui function in 10⁻³ eÅ⁻³ close to the ether-neighboring hydrogen atoms (red circles in Figure 3 and Figure 4); results for the acid-free model taken from (Koubský and Luštinec, 2018) (DMol³, B3LYP, COSMO).

Acid model	TMDGA	TEDGA	Me-TEDGA	Me ₂ -TEDGA
Acid-free	3.613	3.613	3.325	3.038
H ⁺ model	5.847	4.716	4.235	2.663
H ₃ O ⁺ model	5.287	4.516	2.463	2.053
HNO ₃ model	2.053	1.322	1.282	1.232

Radical Fukui function

In our previous studies (Koubský et al., 2017; Koubský and Luštinec, 2018), the radical Fukui function (FF) (Yang and Parr, 1985) has proved to be a radical stability descriptor relevant for the investigated hydrophilic and lipophilic DGA derivatives. Therefore, the radical FF is also evaluated for the proposed DGAs acid models. The calculated values of radical FF are mapped on electron density iso-surface, and the maxima identified. Similar to the previous study (Koubský and Luštinec, 2018), the main maxima appear in a close vicinity of the ether hydrogens and the amide groups. This trend is found for all tested combinations of DGA derivatives and acid representations. The results obtained for TEDGA are given in Figure 3 and Figure 4 as examples.

The trend of radical FF maxima located in vicinity of ether hydrogens (Table 1) agrees well with the expectations based on experimentally observed radiolytic stabilities of the derivatives. For all acid models, the calculated values of radical FF maxima decrease with the molecular weight of the derivative growing, the trend being even more pronounced than in the case of pure water environment considered in (Koubský and Luštinec, 2018). Within the three acid models considered here, the mentioned trend seems to be stronger for the H⁺ and H₃O⁺ models than for the HNO₃ one. However, as already stated, a direct quantitative comparison of the radical FF values obtained for the different acid models is not straightforward due to the different total charge present in the model systems.

Similar to the acid-free situation (Koubský and Luštinec, 2018), the maxima of radical FF on the amide groups are also

TABLE 2 Values of the radical Fukui function in 10^{-3} eÅ^{-3} close to one of the amide groups (white circles in Figure 3 and Figure 4); results for the acid-free model taken from (Koubský and Luštinec, 2018) (DMol³, B3LYP, COSMO)

Acid model	TMDGA	TEDGA	Me-TEDGA	Me ₂ -TEDGA
Acid-free	5.284	5.284	5.787	5.787
H ⁺ model	7.548	5.787	7.045	6.542
H ₃ O ⁺ model	5.284	5.032	7.800	5.535
HNO ₃ model	4.781	3.774	3.019	3.019

TABLE 3 Atomic radical Fukui charges based on Hirshfeld population analysis obtained for studied DGA derivatives and proposed acid models; results for the acid-free model taken from (Koubský and Luštinec, 2018); values for H⁸, H^(R1) and H^(R2) are identical to the value obtained for H⁷ where relevant (DMol³, B3LYP, COSMO).

Acid model	Ligand	N ⁽¹⁾	C ⁽²⁾	O ⁽³⁾	C ⁽⁴⁾	C ⁽⁵⁾	O ⁽⁶⁾	H ⁽⁷⁾
Acid-free	TMDGA	0.062	0.067	0.114	0.030	Eq. C ⁽⁴⁾	0.037	0.026
	TEDGA	0.057	0.063	0.111	0.030	Eq. C ⁽⁴⁾	0.033	0.025
	Me-TEDGA	0.057	0.064	0.113	0.021	0.029	0.033	0.023
	Me ₂ -TEDGA	0.054	0.064	0.111	0.022	Eq. C ⁽⁴⁾	0.036	0.020
H ⁺ model	TMDGA	0.096	0.071	0.097	0.020	Eq. C ⁽⁴⁾	0.023	0.025
	TEDGA	0.072	0.068	0.074	0.019	Eq. C ⁽⁴⁾	0.021	0.024
	Me-TEDGA	0.049	0.076	0.108	0.013	0.018	0.012	0.022
	Me ₂ -TEDGA	0.070	0.065	0.088	0.009	Eq. C ⁽⁴⁾	0.012	0.014
H ₃ O ⁺ model	TMDGA	0.065	0.072	0.079	0.025	Eq. C ⁽⁴⁾	0.033	0.031
	TEDGA	0.061	0.070	0.078	0.025	Eq. C ⁽⁴⁾	0.032	0.029
	Me-TEDGA	0.062	0.068	0.078	0.018	0.020	0.029	0.019
	Me ₂ -TEDGA	0.060	0.064	0.083	0.015	Eq. C ⁽⁴⁾	0.022	0.015
HNO ₃ model	TMDGA	0.038	0.019	0.067	0.009	Eq. C ⁽⁴⁾	0.022	0.011
	TEDGA	0.045	0.018	0.058	0.006	Eq. C ⁽⁴⁾	0.007	0.008
	Me-TEDGA	0.044	0.016	0.058	0.002	0.006	0.004	0.006
	Me ₂ -TEDGA	0.065	0.018	0.058	0.007	Eq. C ⁽⁴⁾	0.007	0.011

observed for all the tested acid models (Table 2). The values obtained for the HNO₃ acid model are reduced as a whole compared to the acid-free model suggesting that the presence of nitric acid molecule decreases the ability of DGA derivatives to react with radicals. The values obtained for different DGA derivatives show then a gradual descend with the molecular weight increasing, supporting thus the experimentally observed stability trend. For the H⁺ and H₃O⁺ models, the radical FF values on amide groups fluctuate and are dependent on the location of bonding model species. Also, the influence of different total charge of the model systems must be taken into account.

Radical Fukui charges and CDD

In order to simplify the analysis and discussion of the calculated results, the volumetric radical FF can be assigned to

individual atoms; the obtained condensed values are called atomic Fukui charges. As Table 3 shows, inclusion of acid does not significantly modify the trends found for the acid-free situation (Koubský and Luštinec, 2018). Since also the main features of results obtained with all tested acid models are similar, we discuss them on the example of H₃O⁺ acid model. (Although some minor differences can be identified in the case of H⁺ acid model, the arguments remain the same.) The HNO₃ acid model shows then a significant reduction of all calculated atomic Fukui charges (similar to the situation observed for the radical FF values, cf. Table 1 and Table 2) making analysis and a straightforward quantitative comparison with the other two models difficult.

Due to its high structural fragility, the ether group is the weakest part of DGA molecules when a radical attack is considered. In case of acid-free environment (Koubský and Luštinec, 2018), the values of radical Fukui charges located on the ether group (C^(4,5), O⁽⁶⁾, H⁽⁷⁾, H⁽⁸⁾, R¹ and R² chains) support with the greater (C^(4,5), H⁽⁷⁾) or lesser (O⁽⁶⁾) extend the

TABLE 4 Values of atomic CDD based on Hirshfeld population analysis calculated for the studied DGA derivatives and proposed acid models; results for the acid-free model taken from (Koubský and Luštinec, 2018); values for H^8 , $H^{(R1)}$ and $H^{(R2)}$ equal to the value obtained for H^7 where relevant (DMol³, B3LYP, COSMO).

Acid model	Ligand	N ⁽¹⁾	C ⁽²⁾	O ⁽³⁾	C ⁽⁴⁾	C ⁽⁵⁾	O ⁽⁶⁾	H ⁽⁷⁾
Acid-free	TMDGA	−0.040	0.057	−0.025	0.031	Eq. C ⁽⁴⁾	0.021	0.018
	TEDGA	−0.044	0.057	−0.020	0.033	Eq. C ⁽⁴⁾	0.016	0.017
	Me-TEDGA	−0.036	0.059	−0.018	0.017	0.036	0.015	0.014
	Me ₂ -TEDGA	−0.033	0.063	−0.017	0.016	Eq. C ⁽⁴⁾	−0.017	0.006
H ⁺ model	TMDGA	−0.034	0.059	−0.040	0.008	Eq. C ⁽⁴⁾	−0.021	0.015
	TEDGA	−0.033	0.073	−0.031	0.014	Eq. C ⁽⁴⁾	−0.004	0.018
	Me-TEDGA	−0.066	0.083	0.042	0.031	−0.008	0.001	0.012
	Me ₂ -TEDGA	−0.031	0.068	−0.045	0.012	Eq. C ⁽⁴⁾	0.011	0.010
H ₃ O ⁺ model	TMDGA	−0.033	0.086	0.014	0.016	Eq. C ⁽⁴⁾	−0.063	0.015
	TEDGA	−0.035	0.078	−0.003	0.014	Eq. C ⁽⁴⁾	−0.040	0.011
	Me-TEDGA	−0.035	0.067	−0.006	0.023	−0.001	−0.009	0.000
	Me ₂ -TEDGA	−0.028	0.079	0.044	0.011	Eq. C ⁽⁴⁾	−0.012	0.003
HNO ₃ model	TMDGA	−0.059	−0.031	−0.120	−0.015	Eq. C ⁽⁴⁾	−0.044	−0.017
	TEDGA	−0.084	−0.035	−0.106	−0.008	Eq. C ⁽⁴⁾	−0.013	−0.011
	Me-TEDGA	−0.084	−0.029	−0.109	−0.009	0.000	0.000	−0.009
	Me ₂ -TEDGA	−0.082	−0.027	−0.101	−0.006	Eq. C ⁽⁴⁾	−0.020	−0.017

TABLE 5 Calculated bond order (Wiberg bond indices) of the C⁽⁴⁾-O⁽⁶⁾ and equivalent C⁽⁵⁾-O⁽⁶⁾ bond; results for the acid-free model taken from (Koubský and Luštinec, 2018) (Gaussian, B3LYP, PCM, NBO).

Acid model	Bond	TMDGA	TEDGA	Me-TEDGA	Me ₂ -TEDGA
Acid-free	C ⁽⁴⁾ -O ⁽⁶⁾	0.904	0.904	0.888	0.884
	C ⁽⁵⁾ -O ⁽⁶⁾	Eq. to C ⁽⁴⁾ -O ⁽⁶⁾	Eq. to C ⁽⁴⁾ -O ⁽⁶⁾	0.900	Eq. to C ⁽⁴⁾ -O ⁽⁶⁾
H ⁺ model	C ⁽⁴⁾ -O ⁽⁶⁾	0.895	0.908	0.895	0.880
	C ⁽⁵⁾ -O ⁽⁶⁾	Eq. to C ⁽⁴⁾ -O ⁽⁶⁾	Eq. to C ⁽⁴⁾ -O ⁽⁶⁾	0.913	Eq. to C ⁽⁴⁾ -O ⁽⁶⁾
H ₃ O ⁺ model	C ⁽⁴⁾ -O ⁽⁶⁾	0.913	0.914	0.892	0.895
	C ⁽⁵⁾ -O ⁽⁶⁾	Eq. to C ⁽⁴⁾ -O ⁽⁶⁾	Eq. to C ⁽⁴⁾ -O ⁽⁶⁾	0.914	Eq. to C ⁽⁴⁾ -O ⁽⁶⁾
HNO ₃ model	C ⁽⁴⁾ -O ⁽⁶⁾	0.904	0.903	0.880	0.882
	C ⁽⁵⁾ -O ⁽⁶⁾	Eq. to C ⁽⁴⁾ -O ⁽⁶⁾	Eq. to C ⁽⁴⁾ -O ⁽⁶⁾	0.904	Eq. to C ⁽⁴⁾ -O ⁽⁶⁾

experimentally observed stability trend. With acid included, the analogical behaviour is observed for C^(4,5), H⁽⁷⁾ and the equivalent hydrogen atoms. For the ether oxygen O⁽⁶⁾, the tendency of atomic Fukui charge to decrease with the ligand molecular weight is remarkably enhanced, suggesting a possible positive influence of acid presence on the tested DGAs derivatives stabilization.

The maxima of radical FF on amide group are also reproduced in the radical Fukui charges located on atoms N⁽¹⁾, C⁽²⁾, and O⁽³⁾. Again, the same trends are observed as for the ether group in case of the acid-free model (Koubský and Luštinec, 2018), and remain qualitatively unchanged after the

inclusion of acid. The values of charges calculated with the HNO₃ model are all reduced compared to the acid-free results.

In order to get a deeper insight into the possible reaction mechanism, the CDD indicator values are calculated and evaluated. The calculated results are summarized in (Table 4).

In general, for the acid-free model, the CDD absolute values on almost all atoms show a decreasing trend with the ligand weight growing, conformal with the experimental stability trend. The signs then indicate that atoms N⁽¹⁾ and O⁽³⁾ are likely the only ones that are susceptible to an electrophilic attack. The remaining atoms might be vulnerable to a nucleophilic attack. This result

supports the reaction degradation mechanism proposed by Wilden et al. (2018) based on the electron transfer from the amide group. In addition, in the case of H_3O^+ acid model, the position $\text{O}^{(6)}$ is predicted to be sensitive to an electrophilic attack. The similar behaviour is observed for TMDGA and TEDGA in frame of the H^+ acid model, for Me_2 -TEDGA in the acid-free model, and for all tested derivatives in the case of the HNO_3 acid model. Thus, apparently, selection of a particular acid model affects significantly the afterward obtained CDD values.

Bond orders

For all tested ligands and acid models, the order of $\text{C}^{(4)}\text{-O}^{(6)}$ bond is found to be the lowest one of all bonds. The calculated $\text{C}^{(4)}\text{-O}^{(6)}$ bond order values are summarized in Table 5. Similar character of the $\text{C}^{(4)}\text{-O}^{(6)}$ bond was also identified for the lipophilic DGA derivatives (Koubský et al., 2017). For all tested models (with and without acid), presence of the methyl group(s) taking place in Me-TEDGA and Me_2 -TEDGA derivatives causes then a small drop in $\text{C}^{(4)}\text{-O}^{(6)}$ bond order. Again, such effect is also observed for the analogically modified lipophilic DGA representatives (Koubský et al., 2017).

Bond orders closely relate to the other frequently used partitioning quantity of electron density - atomic partial charges. However, the partial charges obtained with the here tested acid models are found to be strongly dependent on the particular total charge included in the model, and show also a high sensitivity to the particular position in the ligand where the acid-representing group is attached (as a result of the geometrical optimization of the system); no reliable interpretation of the partial charges thus can be achieved and their values are not discussed here.

Conclusion

Three simplified models of acid influence on the radiolytic stability of hydrophilic DGA representatives are proposed and applied in calculations of the selected chemical stability indicators: radical FF, radical Fukui charges, CDD, and bond orders. The results obtained for the individual acid models are compared and juxtaposed with the results obtained for models with no acid influence considered. The newly tested CDD indicator shows absolute values that are generally in agreement with the experimentally observed radiolytic stability trend ($\text{TMDGA} < \text{TEDGA} < \text{Me-TEDGA} < \text{Me}_2\text{-TEDGA}$ (Wilden et al., 2018)). For different acid models however, the CDD signs significantly vary, and the CCD indicator fails to provide a reliable stability description.

All the tested acid models provide results similar to the results of the acid-free model, with the main trends remaining unaffected. However, values of the tested indicators as a whole drop down with the acid models applied, suggesting that in addition to the direct

protection reflected by the local variations of calculated indicators, some indirect protection mechanism may also originate from acid presence, originating in a general decrease in the chemical reactivity of the ligands in the presence of acid.

Arguments for a direct acid protection effects follows from the two following obtained results: 1) the faster stabilizing reduction of the atomic Fukui charge values at the weakest atomic site of the tested DGA derivatives—the ether oxygen atom - found with the all acid models, and 2) the decrease of radical FF maxima on ether hydrogens and amide groups encountered with the HNO_3 acid model. The latter effect indicates that undissociated acid molecules may reduce the reactivity of the studied DGA derivatives with radicals, the final protection effect being then dependent on the specific nitric acid concentration applied.

Considering the similarity between the basic trends of radical FF and Fukui charges found for the acid models and the acid-free model, the obtained results in any case do not contradict the proposition made by Horne and co-workers (Wilden et al., 2018) that the significant indirect acid protection effect consists in the preferential reaction of acid with the products of solvent radiolysis, decreasing thus the subsequent direct radical attack rate of the solvent radiolysis products on the extractant molecules.

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

The calculations set up and performed by JL. All authors (JL, TK, LK) participated in evaluation, discussion and interpretation of the achieved results.

Funding

This research was funded by MEYSCR grant CZ.02.1.01/0.0/0.0/16_019/0000778 (JL and LK), MEYS CR institutional grant RVO14000 (JL, LK, and TK) and Grant Agency of the Czech Technical University in Prague, Grant No. SGS22/183/OHK4/3T/14 (JL). Support to the methodical part of the work (LK) from the CSF grant GA21-05259S is also acknowledged.

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.

References

- Beck, A. D. (1993). Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.* 98 (7), 5648–5652. doi:10.1063/1.464913
- Delley, B. (1990). An all-electron numerical method for solving the local density functional for polyatomic molecules. *J. Chem. Phys.* 92 (1), 508–517. doi:10.1063/1.458452
- Delley, B. (2000). From molecules to solids with the DMol 3 approach. *J. Chem. Phys.* 113 (18), 7756–7764. doi:10.1063/1.1316015
- Frisch, M. J., Trucks, G. W., Schlegel, H. B., Scuseria, G. E., Robb, M. A., Cheeseman, J. R., et al. (2013). *Gaussian 09, revision D.01*. Wallingford, CT: Gaussian, Inc.
- Fukui, K. (1982). Role of frontier orbitals in chemical reactions. *Science* 218 (4574), 747–754. doi:10.1126/science.218.4574.747
- Galan, H., Zarzana, C. A., Wilden, A., Nunez, A., Schmidt, H., Egberink, R. J. M., et al. (2015). Gamma-radiolytic stability of new methylated TODGA derivatives for minor actinide recycling. *Dalton Trans.* 44, 18049–18056. doi:10.1039/c5dt02484f
- Glendening, E., Badenhoop, J., Reed, A., Carpenter, J., Bohmann, J., Morales, C., et al. (2013). *Natural bond orbital analysis program: NBO 6.0*. Madison, WI: Theoretical Chemistry Institute, University of Wisconsin.
- Grimme, S., Antony, J., Ehrlich, S., and Krieg, H. (2010). A consistent and accurate *ab initio* parametrization of density functional dispersion correction (DFT-D) for the 94 elements H–Pu. *J. Chem. Phys.* 132 (15), 154104. doi:10.1063/1.3382344
- Grimme, S., Ehrlich, S., and Goerigk, L. (2011). Effect of the damping function in dispersion corrected density functional theory. *J. Comput. Chem.* 32 (7), 1456–1465. doi:10.1002/jcc.21759
- Grimme, S. (2006). Semiempirical GGA-type density functional constructed with a long-range dispersion correction. *J. Comput. Chem.* 27 (15), 1787–1799. doi:10.1002/jcc.20495
- Horne, G. P., Wilden, A., Mezyk, S. P., Twight, L., Hupert, M., Stärk, A., et al. (2019). Gamma radiolysis of hydrophilic diglycolamide ligands in concentrated aqueous nitrate solution. *Dalton Trans.* 48 (45), 17005–17013. doi:10.1039/c9dt03918j
- Klamt, A., and Schüürmann, G. (1993). COSMO: a new approach to dielectric screening in solvents with explicit expressions for the screening energy and its gradient. *J. Chem. Soc. Perkin Trans. 2* (5), 799–805. doi:10.1039/p29930000799
- Koubský, T., and Luštinec, J. (2018). Application of quantum mechanical simulations for studying the radiolytic stability of prospective extractants in the nuclear fuel cycle. *J. Radioanal. Nucl. Chem.* 318 (3), 2407–2413. doi:10.1007/s10967-018-6225-2
- Koubský, T., Fojtíková, J., and Kalvoda, L. (2017). Radical degradation stability of ether linkage in N, N, N', N'-tetraoctyldiglycolamide and related organic extractants: A density functional study. *Prog. Nucl. Energy* 94, 208–215. doi:10.1016/j.pnucene.2016.07.010
- Lumetta, G. J., Gelis, A. V., Carter, J. C., Niver, C. M., and Smoot, M. R. (2014). The actinide-lanthanide separation concept. *Solvent Extr. Ion Exch.* 32 (4), 333–347. doi:10.1080/07366299.2014.895638
- Mattocks, J. A., and Cotruvo, J. A. (2020). Biological, biomolecular, and bio-inspired strategies for detection, extraction, and separations of lanthanides and actinides. *Chem. Soc. Rev.* 49, 8315–8334. doi:10.1039/d0cs00653j
- Matveev, P., Mitrofanov, A., Petrov, V., Zhokhov, S., Smirnova, A., Ustynyuk, Y. A., et al. (2017). Testing a simple approach for theoretical evaluation of radiolysis products in extraction systems. A case of N, O-donor ligands for Am/Eu separation. *RSC Adv.* 7 (87), 55441–55449. doi:10.1039/c7ra11622e
- Miertuš, S., Scrocco, E., and Tomasi, J. (1981). Electrostatic interaction of a solute with a continuum. A direct utilization of AB initio molecular potentials for the prevision of solvent effects. *Chem. Phys.* 55 (1), 117–129. doi:10.1016/0301-0104(81)85090-2
- Morell, C., Grand, A., and Toro-Labbe, A. (2005). New dual descriptor for chemical reactivity. *J. Phys. Chem. A* 109 (1), 205–212. doi:10.1021/jp046577a
- Morell, C., Grand, A., and Toro-Labbe, A. (2006). Theoretical support for using the $\Delta f(r)$ descriptor. *Chem. Phys. Lett.* 425 (4–6), 342–346. doi:10.1016/j.cplett.2006.05.003
- Parr, R. G., and Yang, W. (1989). *Density functional theory of atoms and molecules*, sv. 1. New York: Oxford University Press.
- Petersson, G. A., and Al-Laham, M. A. (1991). A complete basis set model chemistry. II. Open-shell systems and the total energies of the first-row atoms. *J. Chem. Phys.* 94 (9), 6081–6090. doi:10.1063/1.460447
- Petersson, G. A., Bennett, A., Tensfeldt, T. G., Al-Laham, M. A., Shirley, W. A., and Mantzaris, J. (1988). A complete basis set model chemistry. I. The total energies of closed-shell atoms and hydrides of the first-row elements. *J. Chem. Phys.* 89 (4), 2193–2218. doi:10.1063/1.455064
- Rostaing, C., Poinssot, C., Warin, D., Baron, P., and Lorraina, B. (2012). Development and validation of the EXAm separation process for single Am recycling. *Procedia Chem.* 7, 367–373. doi:10.1016/j.proche.2012.10.057
- Sasaki, Y., Sugo, Y., Kitatsuji, Y., Kirishima, A., Kimura, T., and Choppin, G. R. (2007). Complexation and back extraction of various metals by water-soluble diglycolamide. *Anal. Sci.* 23 (6), 727–731. doi:10.2116/analsci.23.727
- Smirnova, A., Mitrofanov, A., Matveev, P., Baygildiev, T., and Petrov, V. (2020). A search of a quantitative quantum-chemical approach for radiolytic stability prediction. *Phys. Chem. Chem. Phys.* 22 (26), 14992–14997. doi:10.1039/d0cp01786h
- Tomasi, J., and Persico, M. (1994). Molecular interactions in solution: an overview of methods based on continuous distributions of the solvent. *Chem. Rev.* 94 (7), 2027–2094. doi:10.1021/cr00031a013
- Tomasi, J., Mennucci, B., and Cammi, R. (2005). Quantum mechanical continuum solvation models. *Chem. Rev.* 105 (8), 2999–3093. doi:10.1021/cr9904009
- Veliscek-Carolan, J. (2016). Separation of actinides from spent nuclear fuel: A review. *J. Hazard. Mater.* 318, 266–281. doi:10.1016/j.jhazmat.2016.07.027
- Wiberg, K. B. (1968). Application of the pople-santry-segal CNDO method to the cyclopropylcarbonyl and cyclobutyl cation and to bicyclobutane. *Tetrahedron* 24 (3), 1083–1096. doi:10.1016/0040-4020(68)88057-3
- Wilden, A., Mincher, B. J., Mezyk, S. P., Twight, L., Rosciolo-Johnson, K. M., Zarzana, C. A., et al. (2018). Radiolytic and hydrolytic degradation of the hydrophilic diglycolamides. *Solvent Extr. Ion Exch.* 36 (4), 347–359. doi:10.1080/07366299.2018.1495384
- Yang, W., and Parr, R. G. (1985). Hardness, softness, and the fukui function in the electronic theory of metals and catalysis. *Proc. Natl. Acad. Sci. U. S. A.* 82 (20), 6723–6726. doi:10.1073/pnas.82.20.6723



OPEN ACCESS

EDITED BY
Chandrabose Selvaraj,
Alagappa University, India

REVIEWED BY
Rachel Basques Caligiorno,
Grupo Santa Casa BH, Brazil
Balajee Ramachandran,
University of Notre Dame, United States

*CORRESPONDENCE
Pooja Vijayaraghavan,
vrpooja@amity.edu
Abhishek Sengupta,
asengupta@amity.edu

[†]These authors share first authorship

SPECIALTY SECTION
This article was submitted to Biophysics,
a section of the journal
Frontiers in Molecular Biosciences

RECEIVED 28 September 2022
ACCEPTED 18 November 2022
PUBLISHED 12 December 2022

CITATION
Kamboj H, Gupta L, Kumar P, Sen P,
Sengupta A and Vijayaraghavan P (2022),
Gene expression, molecular docking,
and molecular dynamics studies to
identify potential antifungal compounds
targeting virulence proteins/genes VelB
and THR as possible drug targets against
Curvularia lunata.
Front. Mol. Biosci. 9:1055945.
doi: 10.3389/fmolb.2022.1055945

COPYRIGHT
© 2022 Kamboj, Gupta, Kumar, Sen,
Sengupta and Vijayaraghavan. This is an
open-access article distributed under
the terms of the [Creative Commons
Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). 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.

Gene expression, molecular docking, and molecular dynamics studies to identify potential antifungal compounds targeting virulence proteins/genes VelB and THR as possible drug targets against *Curvularia lunata*

Himanshu Kamboj ^{1†}, Lovely Gupta ^{1†}, Pawan Kumar ²,
Pooja Sen ¹, Abhishek Sengupta ^{3*} and
Pooja Vijayaraghavan ^{1*}

¹Anti-mycotic Drug Susceptibility Laboratory, Amity Institute of Biotechnology, Amity University, Noida, India, ²School of Computational and Integrative Sciences, Jawaharlal Nehru University, New Delhi, India, ³Systems Biology and Data Analytics Research Laboratory, Amity Institute of Biotechnology, Amity University Uttar Pradesh, Noida, India

Curvularia lunata is a melanized fungus pathogenic to both plants and animals including humans, causing from mild, febrile to life-threatening illness if not well treated. In humans, it is an etiological agent of keratomycosis, sinusitis, and onychomycosis in immunocompromised and immunocompetent patients. The development of multiple-drug-resistant strains poses a critical treatment issue as well as public health problem. Natural products are attractive prototypes for drug discovery due to their broad-spectrum efficacy and lower side effects. The present study explores possible targets of natural antifungal compounds (α -pinene, eugenol, berberine, and curcumin) against *C. lunata* via gene expression analysis, molecular docking interaction, and molecular dynamics (MD) studies. Curcumin, berberine, eugenol, and α -pinene exhibited *in vitro* antifungal activity at 78 μ g/ml, 156 μ g/ml, 156 μ g/ml, and 1250 μ g/ml, respectively. In addition, treatment by these compounds led to the complete inhibition of conidial germination and hindered the adherence when observed on onion epidermis. Several pathogenic factors of fungi are crucial for their survival inside the host including those involved in melanin biosynthesis, hyphal growth, sporulation, and mitogen-activated protein kinase (MAPK) signalling. Relative gene expression of *velB*, *brn1*, *clm1*, and *pks18* responsible for conidiation, melanin, and cell wall integrity was down-regulated significantly. Results of molecular docking possessed good binding affinity of compounds and have confirmed their potential targets as THR and VelB proteins. The docked structures, having good binding affinity among all, were further refined, and rescored from their docked poses through 100-ns long MD

simulations. The MDS study revealed that curcumin formed a stable and energetically stabilized complex with the target protein. Therefore, the study concludes that the antifungal compounds possess significant efficacy to inhibit *C. lunata* growth targeting virulence proteins/genes involved in spore formation and melanin biosynthesis.

KEYWORDS

Curvularia lunata, molecular docking, molecular dynamics, bioactive molecules, virulence proteins

Introduction

The *Curvularia* genus is one of the major groups of opportunistic human pathogenic dematiaceous filamentous fungi (Alex et al., 2013). Within the genus, *Curvularia lunata* (teleomorph sexual state—*Cochliobolus lunatus*) is in the growing list of emerging fungal pathogens in humans (Giri et al., 2011; Chowdhary et al., 2014), whereas it is reported pathogenic to animals and plants (Beckett et al., 2017; Gupta et al., 2017; Bisht et al., 2018; Liu et al., 2019). The infections caused by *Curvularia* spp. include phaeohyphomycosis, non-dermatophytic onychomycosis, mycetoma, and infections in eyes, nails, sinuses, and wounds (Vineetha et al., 2016). Infections generally occur through direct inoculation of conidia or by inhalation, leaving it susceptible for invasion (Shrivastava et al., 2017). Patients with peritoneal, venous catheters, intravenous drug abusers, and cataract surgery patients are more prone to *C. lunata* infections (Alex et al., 2013).

Azoles (itraconazole and voriconazole) and polyenes (amphotericin B) are antifungals that have widely been used to control invasive human fungal infections for more than 4 decades (Jørgensen and Heick, 2021; Uma Maheshwari Nallal et al., 2021). The clinical use of azoles is of high priority since there are only a few available alternatives for prophylactic and therapeutic treatment of *C. lunata* infections (Chowdhary et al., 2014; Chang et al., 2019). *C. lunata* infections like foliar disease and leaf blight have been reported in plants (Liu et al., 2014; Garcia-Aroca et al., 2018), and to control these infections, a variety of fungicides were used in the crop fields. Excessive use of synthetic fungicides poses selective pressure on cross-kingdom pathogens and impacts antifungal drug resistance (Hof, 2001). Transferability of such drug-resistant isolates from farms to humans through the human–plant interaction stances a critical public health concern (Bengyella et al., 2017) as some of the azole fungicides possess chemical structures similar to medical azole (Snelders et al., 2012).

Natural compounds have received a renewed interest in their use as antimicrobials because of uncontrolled usage of synthetic drugs or fungicides (Nagoor Meeran et al., 2017). These can be exploited in controlling the growth of fungi consequently inhibiting secondary metabolite production. The effect of plant extracts has been investigated on *C. lunata* depicting alteration in growth, sporulation, and secondary metabolite

pathways (Ghany TM et al., 2015). Many investigations resulted in the screening of a wide variety of plant species/bioactive compounds for their antimicrobial activities and have revealed structurally unique biologically active compounds (Matasyoh et al., 2007). Their target identification can be approached via direct biochemical assays, molecular studies, or using computational methods. Furthermore, the mechanism of action of compounds can be generated by studying gene expression data in the presence or absence of it.

The pathogenicity of *C. lunata* involves a plethora of virulence factors including melanin pigment (Xu et al., 2007), siderophores (Wang et al., 2013), hydrophobins, and non-host-specific toxins (Gao et al., 2012). A number of genes and proteins are involved in these pathways including *brn1*, *clpks18*, *clvelB*, and *clm1* which are involved in melanin biosynthesis, hyphal growth, sporulation, non-ribosomal peptide synthetase, and cell wall integrity (Rižner and Wheeler, 2003; Gao et al., 2012; Fu et al., 2022). Melanin deposition in the cell wall of fungus protects it from host macrophage attack and phagolysosome oxidative burst of neutrophils (Rižner and Wheeler, 2003; Tóth et al., 2020). Multiple genes are involved in the melanin production pathway of the fungus directly or indirectly, including *clpks18* gene responsible for the synthesis of polyketide synthase enzyme (PKS) and *brn1* gene responsible for mechanical strength of appressorium required for penetration (Rižner and Wheeler, 2003). Trihydroxynaphthalene reductase (THR) is also an essential enzyme other than PKS of the DHN melanin biosynthesis pathway, and it represents an emerging target for the development of antimycotics. Secondary metabolite synthesis of this pathogen is regulated by genes involved in velvet protein biosynthesis (Gao et al., 2017), and the velvet-like B protein VelB plays a crucial role in controlling the production of conidia, cell wall composition, integrity, and host-specific methyl 5-(hydroxymethyl) furan-2-carboxylate toxin production. Studies have also reported involvement of VelB in pathogenicity as well as fungicide resistance in *C. lunata* (Wu et al., 2012; Gao et al., 2017).

The present study focuses on the identification of the potential virulence target of *C. lunata* via gene expression analysis, *in silico* approach, and *in vitro* evaluation of antifungal activity of natural bioactive compounds like α -pinene, eugenol, berberine, and curcumin.

Material and methodology

Sample collection and fungal isolation

Rice plant variety PUSA 1121 demonstrated typical symptoms of disease caused by *Curvularia* spp. including leaf lesions. The infected leaf samples were collected from Yamuna Nagar district, Haryana (GPS coordinates—30°02'38.1"N77°07'50.8"E). The collected samples were placed in sterile polythene bags properly tagged with date, time, and location (Naz et al., 2017). Sections of diseased leaf portions were surface-sterilized in 1% (w/v) sodium hypochlorite solution, rinsed in sterile distilled water, and incubated on fresh potato dextrose agar (PDA) for 96 h at 28 ± 2°C. Fungal isolates were identified macroscopically and microscopically (Sivanesan 1987; Cuervo-Parra et al., 2012) and further transferred to fresh PDA and incubated for 96 h at 28 ± 2°C.

Molecular identification

Genomic DNA was extracted from *Curvularia* spp. using the modified cetyltrimethylammonium bromide (CTAB) method (Lee et al., 1988; Wu et al., 2001). Molecular identification of the isolate was confirmed by the amplification and sequencing of the full-length 18 S internal transcribed spacer (ITS) region using the ITS1 (5'-TCC GTA GGT GAA CCT GCGG-3') and ITS4 (5'-TCC TCC GCT TAT TGA TATGC-3') primers (White et al., 1990). The PCR-amplified ITS region was sequenced by Sanger sequencing. The sequences obtained were compared to the sequences in the GenBank database (www.ncbi.nlm.nih.gov.in) using basic local alignment search tool (BLAST) analysis, and identification was confirmed when 99–100% sequence identity was observed.

Procurement of bioactive compounds

The compounds, namely, α -pinene, curcumin, berberine, and eugenol were procured from Sigma-Aldrich (India). The compounds were solubilised in dimethyl sulfoxide (DMSO) to make a stock solution of 100 mg/ml, except berberine (20 mg/ml in methanol). For working solution, stock solution was further diluted in potato dextrose broth (PDB). The final concentration of dimethyl sulfoxide (DMSO) never exceeded the amount with any detectable effect in assays (Szumilak et al., 2017).

Antifungal susceptibility testing

The conidia were harvested in sterile phosphate-buffered saline (1× PBS), observed, and counted using a haemocytometer under a light microscope. The final conidial suspension was adjusted to 10⁴ conidia/mL in PDB

(Amin and Abdalla, 1980; Xie et al., 2020). Minimum inhibitory concentration (MIC) of α -pinene, curcumin, berberine, and eugenol against *C. lunata* was determined using the broth microdilution method in a 96-well polystyrene plate according to CLSI protocol (CLSI, 2016; Alexander, 2017). Two-fold serial dilution was performed in a 96-well microplate to attain concentrations ranging from 5000 to 9.765 μ g/ml. Each well was inoculated with 100 μ L of the conidial suspension (as previously described in the section) except the negative control. The microplate was incubated at 28 ± 2°C for 5 days, and the growth in each well was compared with that of the positive control. The experiments were carried out in triplicate. The MIC value of a drug is determined as the lowest concentration with no visible growth relative to the drug-free control (Andrews, 2001).

Pathogenicity test for *C. lunata* on onion peel epidermis

Onion bulb scales were thoroughly rinsed with distilled water. The inner epidermis of onion bulb scales was peeled off and cut into 1 × 1 cm² strips. Sections of onion peel were floated on 4 ml distilled water in 60-mm Petri plates for treated and untreated samples. Freshly harvested conidia were washed with sterile water followed by centrifugation at 4500 g for 10 min and resuspended in sterile distilled water to the final concentration of 1 × 10⁴ conidia/ml. Only conidial suspension (10 μ L) was placed on a single strip as a positive control. MIC of α -pinene (1250 μ g/ml), curcumin (78 μ g/ml), berberine (156 μ g/ml), and eugenol (156 μ g/ml) was added with 10 μ L of the conidial suspension on individual strips of onion peel epidermis. Each experiment was independently conducted in triplicate. The strips were incubated at 28 ± 2°C for 24 h. After 24 h of inoculation, extra suspension was removed from the peel, and 30% methanol was applied to prevent further penetration during observation (Chida and Sisler, 1987; Gupta et al., 2019). The strips were stained with lactophenol cotton blue and observed under a light microscope (×40 magnification) to observe hyphal growth and penetration. For scanning electron microscopy (SEM), the strips were sputter-coated with gold and observed under Zeiss SEM, MA EVO-18 Special Edition (Liu W. et al., 2011; Gupta et al., 2019).

Biochemical assays

- 1) Melanin quantification: Isolation of melanin from *C. lunata* (treated and untreated) was performed by the modified method of Kumar et al. (2011). The fungus was cultured in the media supplemented with Inhibitory Concentration-50 (IC₅₀) of α -pinene, curcumin, berberine, and eugenol in a 12-well cell culture plate. Inhibitory Concentration-50 (IC₅₀) of

TABLE 1 Gene-specific primers used for qRT-PCR.

S. No.	Gene name	Gene reference ID	Primer sequence (5'-3')	Amplicon size (bp)
1	<i>velB</i>	KY435512.1	F: AGCATGGCTCACTACCAA R: GTCCACCATGAGGACAAA	270 bp
2	<i>brn1</i>	JQ698339.1	F: AACAGCCTTTCAATCCTCTC R: GTTCAAAGCCTTGATCTCCT	292 bp
3	<i>clm1</i>	HQ851366.1	F: GGCTACCAACAACCAGACC R: CTCTGGCCAAACCAAAATC	401 bp
4	<i>pks18</i>	MF114294.1	F- CGCCACCTCTGTTCTTCTT R- CCTCAACACCACAAGTCCA	185 bp
5	<i>GAPDH</i>	LT715821.1	F- CATTGGCCGTATCGTCTT R- GCCGTTGACAGTCAGGTT	339 bp

α -pinene, curcumin, berberine, and eugenol were 625 $\mu\text{g/ml}$, 39 $\mu\text{g/ml}$, 78 $\mu\text{g/ml}$, and 78 $\mu\text{g/ml}$, respectively, where 50 % growth of *C. lunata* was inhibited. The extracted melanin was resuspended in 100 mM borate buffer, and absorbance was recorded in the wavelength range (250–800 nm) on a UV-visible spectrophotometer. Also, 100 mM borate buffer was used as a blank. The experiment was conducted in triplicate.

- 2) Conidial cell surface hydrophobicity (CSH) and conidiation: Using two-phase partitioning with hexadecane as the hydrocarbon phase, hydrophobicity assay was conducted (Pihet et al., 2009; Hoda et al., 2020). In brief, *C. lunata* conidia were harvested in 1 \times PBS from treated and untreated samples, and their absorbance was set to 0.3 using a spectrophotometer (at wavelength 630 nm). Hexadecane (500 μL) was added to conidial suspension and vortexed for 2 min at an interval of 30 s; then, for the hydrophobic phase, separation tubes were kept at room temperature for 10 min. At 630 nm, absorbance of the aqueous phase was determined and compared to the initial absorbance, that is, 0.30.

Percentage reduction in cell surface hydrophobicity (%CSH) was calculated for treated as well as untreated *C. lunata* conidia using the formula:

$$\% \text{CSH} = \frac{A1 - A2}{A1} \times 100,$$

where A1 is the absorption before addition of hydrocarbon and A2 is the absorption after addition of hydrocarbon.

The effect of compounds on conidia formation was analysed by counting the number of spores using a haemocytometer (Abubakar and Likita 2021). Conidia were harvested from the 1 cm^3 mycelial mat of treated and untreated cultures and resuspended in 1 ml of 1 \times PBS supplemented with 0.25% Tween-20. A volume of 100 μl of conidial suspension was placed on the surface of the

counting chamber of the haemocytometer and covered with a cover slip. The number of conidia was counted from square grids in the counting unit of the haemocytometer. The conidia concentration was calculated.

Gene expression analysis

The *brn1*, *velB*, *clm1*, *pks18*, and glyceraldehyde-3-phosphate dehydrogenase (*GAPDH*) gene sequences were downloaded from the NCBI (<https://www.ncbi.nlm.nih.gov/pubmed>) database for designing the primers for expression studies. The primers were designed by Primer3 software (<http://primer3.ut.ee/>) and were analysed for potential hair pin formation and self-complementarity (<http://www.basic.northwestern.edu/biotools/oligocalc.html>). The details of primers are given in Table 1.

The expression of the genes of interest was quantified by quantitative real time-PCR (qRT-PCR) (Gupta et al., 2019, 2022). Mycelial cultures were harvested, and RNA was extracted using TRIzol™ reagent (Invitrogen). A measure of 2 μg of total RNA of each sample (treated and untreated) was used to synthesize first-strand cDNA by the oligo (dT)-18 primer using the Hi-cDNA Synthesis Kit (HiMedia). The qRT-PCR was performed using an ABI QuantStudio 3 system (Applied Biosystems, Streetsville, Canada), and amplification products were detected with SYBR Green Master Mix (G-Biosciences) for gene expression.

The relative quantification of individual gene expression was performed using the comparative threshold cycle method. The amplification program used for real time was 95°C for 3 min, 40 cycles at 95°C for 30 s, 60°C for 30 s, and 72°C for 30 s. To check the specificity of the PCR product, the melting curve was analysed at 95°C for 15 s, 60°C for 60 s, 72°C for 30 s, and holding stage 10 s. *GAPDH* gene was set as the reference gene. Relative expression was estimated using the $2^{-\Delta\Delta C_t}$ formula, where

$$\Delta\Delta Ct = \frac{[(Ct_{\text{target gene}})_{\text{sample}} - (Ct_{Tub})_{\text{sample}}]}{[(Ct_{\text{target gene}})_{\text{reference}} - (Ct_{Tub})_{\text{reference}}]}.$$

The results were analysed using ABI QuantStudio 3 software, and the genes were considered differentially expressed if they were at least two-fold up- or down-regulated.

Molecular docking and dynamics studies

The three-dimensional structure of the proteins (VelB and THR) was not available in the Protein Data Bank (PDB) hence FASTA sequence of virulence proteins velvet protein B (VelB) with accession number ARH19411 and 1,3,8-trihydroxynaphthalene reductase (THR) with accession number QTG11042—proteins of *C. lunata* were retrieved from the NCBI (<https://www.ncbi.nlm.nih.gov/>). To obtain the 3D structure of proteins, homology modelling was performed using SwissModel via the ExPaSy web server (<https://swissmodel.expasy.org/>). The Self-Optimized Prediction Method with Alignment (SOPMA) server was used to speculate the secondary structure of VelB and THR proteins (https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_sopma.html). The best template having maximum percentage identity with the target and modelled structure was then evaluated via PROCHECK. For model protein preparation such as charge assignment, solvation parameters, and fragmental volumes, Swiss-PdbViewer version 4.10 (SPBDV-4.10) was used (Morris et al., 2009). The 3D structures of the compounds (α -pinene, curcumin, berberine, and eugenol) were downloaded from the PubChem compound database (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4702940/>) in the spatial data file (SDF) format. The PubChem compound identifier (CID) was 969516 (curcumin), 2353 (berberine), 3314 (eugenol), and 6654 (α -pinene). The SDF file format was further converted to the PDB file format using the Open Babel tool (<https://openbabel.org/docs/dev/Installation/install.html>) for molecular docking via the AutoDock4.2.3 tool (O'Boyle et al., 2011). The absorption, distribution, metabolism, excretion, and toxicity profiles of plant-derived compounds were qualitatively measured by using the online SwissADME program (<http://www.swissadme.ch/index.php>) (Sharma et al., 2022).

Molecular docking was performed using the AutoDock4.2.3 tool (Morris et al., 2009; Karthika et al., 2021) to predict the binding and the structure of the intermolecular complex between drug targets and potential inhibitors. The Lamarckian genetic algorithm was utilized for protein–ligand interactions with the set parameters. The total number of poses was set to 50. Poses were further clustered using all atom root mean square deviation (RMSD) cut-off of 0.3 Å to remove redundancy. The default values were used for all other parameters for docking and scoring. The protein

structure was kept rigid in all steps. The molecular interactions of the best docking pose were visualised via Discovery Studio Visualizer programs (<http://accelrys.com/products/collaborative-science/biovia-discovery-studio/visualization-download.php>). The amino acid residues that displayed interactions with the ligand are documented in Table 2 and Table 3.

All-atom molecular dynamics (MD) simulation was performed to understand the conformational stability of VelB and THR proteins bound with the docked molecules in comparison to the unbound state of proteins. Atomic coordinates of VelB and THR protein complexes with docked molecules were used to generate the simulation trajectory using GROMACS v5.1.4 (Abraham et al., 2015). During the simulation complex preparation stage, the CHARMM27 (Bjellmar et al., 2010) force field was used for proteins, while the TIP3P water model was used to solvate the protein complex. The bound ligand parameters were generated as described in the literature (Zoete et al., 2011). The protein complex was placed in the centre of the cubical box with 10 Å edge-side filled with water molecules. The total charge of the simulation box was neutralised by adding 0.15 M counterions (Na^+Cl^-). All the MD simulations were performed under physiological conditions (Joung and Cheatham, 2008). The prepared simulation box was taken for energy minimization using steepest descent followed by conjugant gradients (50,000 steps for each). The system was further equilibrated through the constant number, volume, and temperature (NVT) and the constant number, pressure, and temperature (NPT) for 500 ps. The Berendsen thermostat (Berendsen et al., 1987) and the Parrinello–Rahman pressure (Parrinello and Rahman, 1980) algorithm were used to maintain the temperature and pressure, respectively. Final MD simulation was performed for 100 ns under the NPT ensemble condition with the step size of 0.2 fs. GROMACS modules and MD trajectory (McGibbon et al., 2015; Sankar et al., 2021) were employed to visualize the global structural order parameters: RMSD, radius of gyration (RoG), solvent-accessible surface area (SASA), and root mean square fluctuation (RMSF).

Statistical analysis

Statistical analysis was performed using the one-way analysis of variance (ANOVA) for the comparison of results of gene expression analysis via qRT-PCR and melanin and CSH percentage for biochemical assays. The experiment was conducted in technical and biological triplicate. Statistical analysis was also performed using GraphPad Prism software 8.0.2.263 version and Microsoft Excel. $p < 0.05$ was considered statistically significant.

Results

Rice plants with symptoms of blast disease specifically leaf lesions were collected from Yamuna Nagar, Haryana, India

TABLE 2 Molecular docking affinity of four compounds with THR protein of *C. lunata*.

Compound name	PubChem CID	THR protein		
		Binding affinity (Kcal/mol)	Hydrogen bond formation	Amino acid residues
Curcumin	969516	-10.80	4	TYR178, ILE41, ASN114, and ARG39
Berberine	2353	-9.62	3	ILE41 and SER164
Eugenol	3314	-6.25	3	ILE165, SER164, and PRO208
α -Pinene	6654	-6.13	0	-

TABLE 3 Molecular docking affinity of four compounds with VelB protein of *C. lunata*.

Compound name	PubChem CID	VelB protein		
		Binding affinity (Kcal/mol)	Hydrogen bond formation	Amino acid residues
Curcumin	969516	-8.03	4	SER124, SER119, and SER132
Berberine	2353	-6.98	0	-
Eugenol	3314	-5.21	1	GLU130
α -Pinene	6654	-4.39	0	-

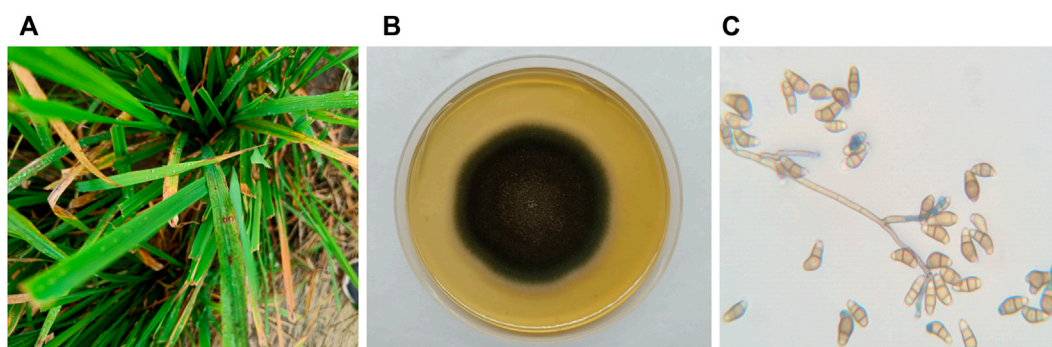


FIGURE 1
(A) *C. lunata* infection on the leaves of the rice plant; (B) *C. lunata* colony morphology on potato dextrose agar; and (C) *C. lunata* conidia.

(Figure 1). The fungal isolate was identified as *Curvularia* based on its colony morphology, viz., fluffy and velvety mycelia, greyish black colour, black pigment on the reverse side, and straight to pyriform conidia having three to four cells, with a large and curved central cell with smooth-walled in a lactophenol cotton blue mount. Morphological characteristics of the isolated fungus were in agreement with Ellis (1971) and Sivanesan (1984).

Furthermore, molecular characterisation and identification of the fungal isolate at the species level were conducted and confirmed by the amplification and sequencing of 18 S ITS1 and ITS4 regions. The obtained sequence showed 100% similarity with *C. lunata* from the GenBank database. The sequence was

submitted to the NCBI with GenBank accession number OL757869 (<https://www.ncbi.nlm.nih.gov/search/all/?term=OL757869>).

Antifungal susceptibility testing

The isolated fungal pathogen was susceptible to polyenes and azoles. The calculated MIC of bioactive compounds against *C. lunata* was in the range of 1250–78 μ g/ml: curcumin (78 μ g/ml), berberine (156 μ g/ml), eugenol (156 μ g/ml), and α -pinene (1250 μ g/ml). Among all, curcumin inhibited the pathogen

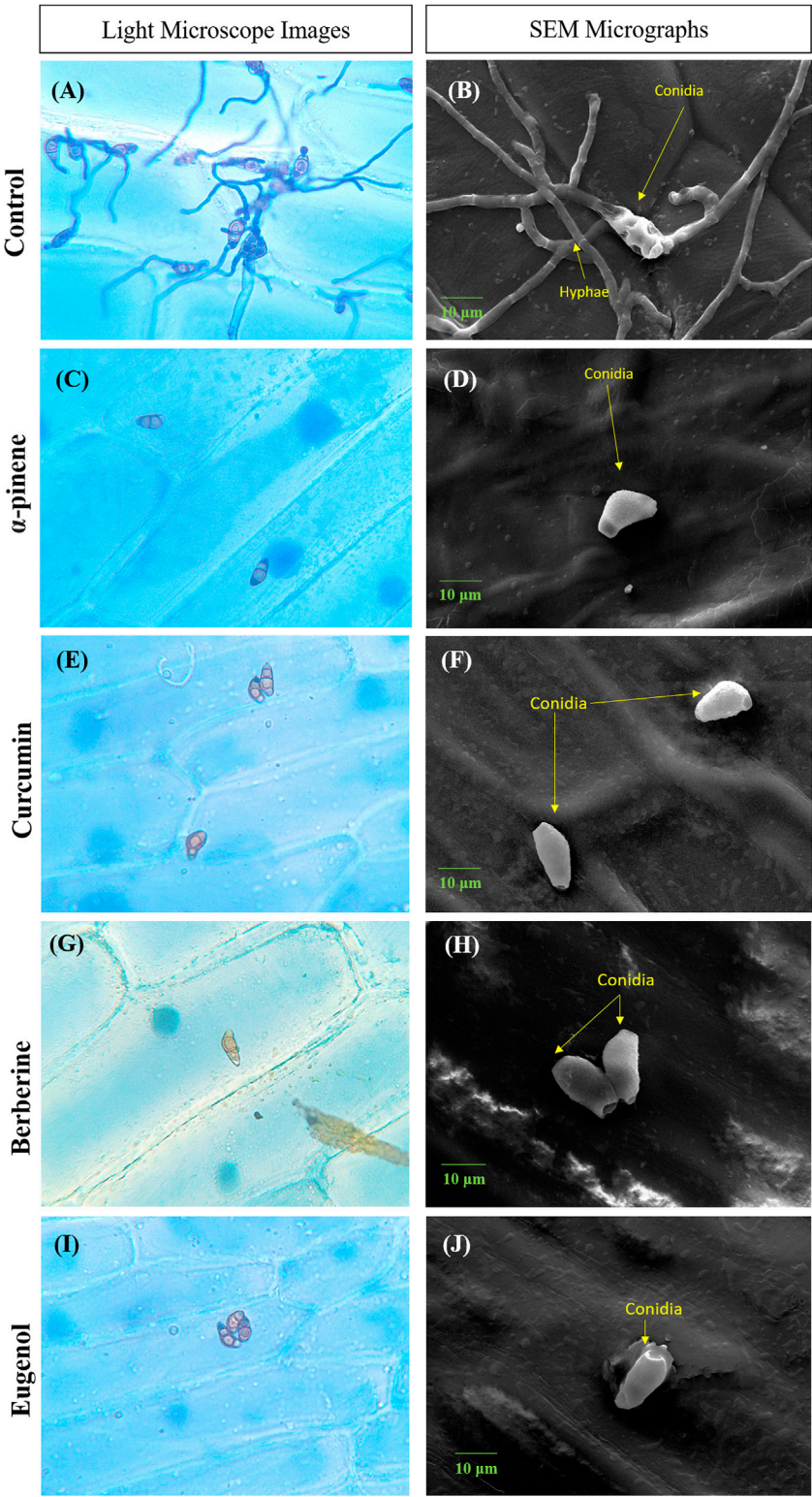


FIGURE 2
Microscopic images depicting conidia of *C.lunata* Control (A,B); α -pinene (C,D); curcumin (E,F); berberine (G,H); and eugenol (I,J). Scale bar = 10 μ m.

hyphal and conidial germination at a very low concentration. At IC_{50} of α -pinene, white fungal morphology was observed *in vitro*, whereas IC_{50} of other compounds inhibited the growth of fungi to 50% with minor changes in its morphology. IC_{50} of the compound referred when the growth of microorganism was suppressed by 50%.

Pathogenicity test of *C. lunata* conidia on onion peel epidermis

Conidia failed to adhere to onion peel *epidermis*, in the presence of bioactive compounds when observed after 24 h of incubation. No visible conidial germination was observed at MIC of α -pinene, curcumin, berberine, and eugenol. In the control sample, conidia with dense hyphal growth were observed on the onion peel under a light microscope as well as SEM as depicted in [Figure 2](#). Upon treatment, the shape of conidia was obovoidal to clavate, curved at subterminal ends, and treatment with bioactive compounds effectively prevented the germination of conidia.

Biochemical assays

- 1) Melanin content: The overall characteristic absorption of melanin was observed at 205 nm which was 0.604, 0.401, 0.274, 0.247, and 0.184 for control, berberine, eugenol, α -pinene, and curcumin, respectively ([Supplementary Figure S1](#); $p < 0.0001$). The melanin content in the compound-treated *C. lunata* culture showed significant reduction in the quantity as compared to that of control.
- 2) CSH percentage: A statistical decrease was observed in the biochemical CSH value of treated cultures as compared to the untreated positive control. The calculated CSH percentage was 87.05%, 79.57%, 75.54%, and 58.73% for α -pinene-, eugenol-, berberine-, and curcumin-treated *C. lunata* as compared to the control ([Supplementary Figure S2](#); $p < 0.05$). Curcumin affected the hydrophobicity of conidia more than the other compounds.
- 3) Culture was grown at IC_{50} of compounds to check the number of conidia formed, and a 50% reduction in conidiation was observed in treated cultures in comparison with the control.

Gene expression analysis

The effect of curcumin, α -pinene, berberine, and eugenol treatment on the expression of *brn1*, *velB*, *pks18*, and *clm1* of *C. lunata* was investigated by reverse transcription followed by qRT-PCR for differential gene expression. α -pinene, curcumin, berberine, and eugenol treatment led to a significant down-regulation of *velB*, *brn1*, *clm1*, and *pks18* gene transcripts in

comparison with the control (untreated) ([Figure 3](#)). The expression of secondary metabolite gene *brn1* was significantly down-regulated upon eugenol treatment, followed by curcumin, α -pinene, and berberine and *pks18* genes in curcumin and α -pinene. The *velB* gene was significantly down-regulated in eugenol, berberine, and curcumin. The relative expression of *clm1* gene was highly down-regulated in all treated samples as compared to the control. The complete expression data were normalized by the housekeeping gene *GAPDH*. Gene expression data expressed as $2^{-\Delta\Delta Ct}$ are the mean of at least three replicates \pm standard error.

Molecular docking studies and dynamics studies

Prediction of the secondary structure of THR ([Figure 4A](#)) resulted in 40.82% α -helix (h), 18.35% extended strand (e), 7.87% β -turn (t), and 32.96% random coil (c) elements, and the VelB sequence ([Figure 4B](#)) consisted of 15.18% α -helix (h), 17.26% extended strand (e), 4.46% β -turn (t), and 63.10% random coil (c) elements. Graphs were obtained to visualize the prediction and score curves for all predicted states using parameters such as window width and number of states.

Model accuracy assessment of modelled protein structures of THR and VelB of *C. lunata* was performed *via* SwissModel. The PDB ID 1YBV was used as the template to model the 3D protein structure of the THR protein sequence and PDB ID 4N6R chain B for VelB protein. The stereochemical quality and accuracy of the model were tested using PROCHECK. Results from PROCHECK were reported as the Ramachandran plot. For THR protein, 89.0% residues were in most favoured regions, 11.0% residues were in additional allowed regions, and no residues were in disallowed regions. Similarly, in protein VelB, 75.7% residues were in most favoured regions, 19.8% residues were in additional allowed regions, 2.1% residues were in generously allowed regions, and 2.5% residues were in disallowed regions. Both modelled proteins obtained by homology modelling were of good quality on the basis of the Ramachandran plot ([Figure 5](#)).

The binding affinity of four natural compounds with THR and VelB proteins of *C. lunata* are presented in [Tables 2](#) and [3](#), respectively. The binding interaction of four natural compounds with VelB and THR proteins are shown in [Figures 6](#), [7](#), respectively. The binding affinity and hydrogen bonds of curcumin were high with THR protein followed by berberine, eugenol, and α -pinene. With VelB protein, curcumin showed the highest binding affinity and hydrogen bonds as compared to berberine, eugenol, and α -pinene. The binding affinity of curcumin was -10.80 Kcal/mol with four hydrogen bonds for THR protein and -8.03 Kcal/mol with four hydrogen bonds for VelB protein. Berberine also showed good binding affinity for THR and VelB proteins, i.e., -9.62 Kcal/mol and -6.98 Kcal/mol,

FIGURE 3
Relative quantification of *brn1*, *velB*, *pks18*, and *clm1* gene expression in *C. lunata* (normalised to the house-keeping gene GAPDH). Data were reported as mean of fold changes with standard deviation from two independent experiments amplified in triplicate. $p \leq 0.05$ was considered statistically significant.

FIGURE 4
Predicted secondary structure validation of THR **(A)** and VelB **(B)** proteins of *C. lunata* using SOPMA. The blue line represents α -helices, red colour represents the extended strand, green colour represents β -turn, and magenta colour represents the random coil in graphical representation. The X-axis represents position of the amino acid; the Y-axis shows the score for each predicted state.

Structural order parameters of the THR complex were analysed with respect to the THR receptor to depict the structural changes upon curcumin binding. The RMSD

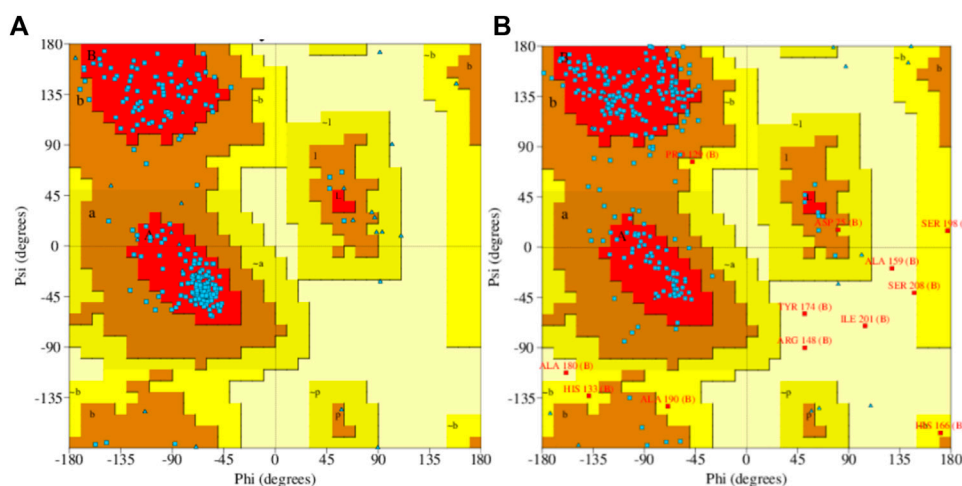


FIGURE 5
Ramachandran plot of predicted 3D structures of THR (A) and VelB (B) proteins of *C. lunata* using PROCHECK software.

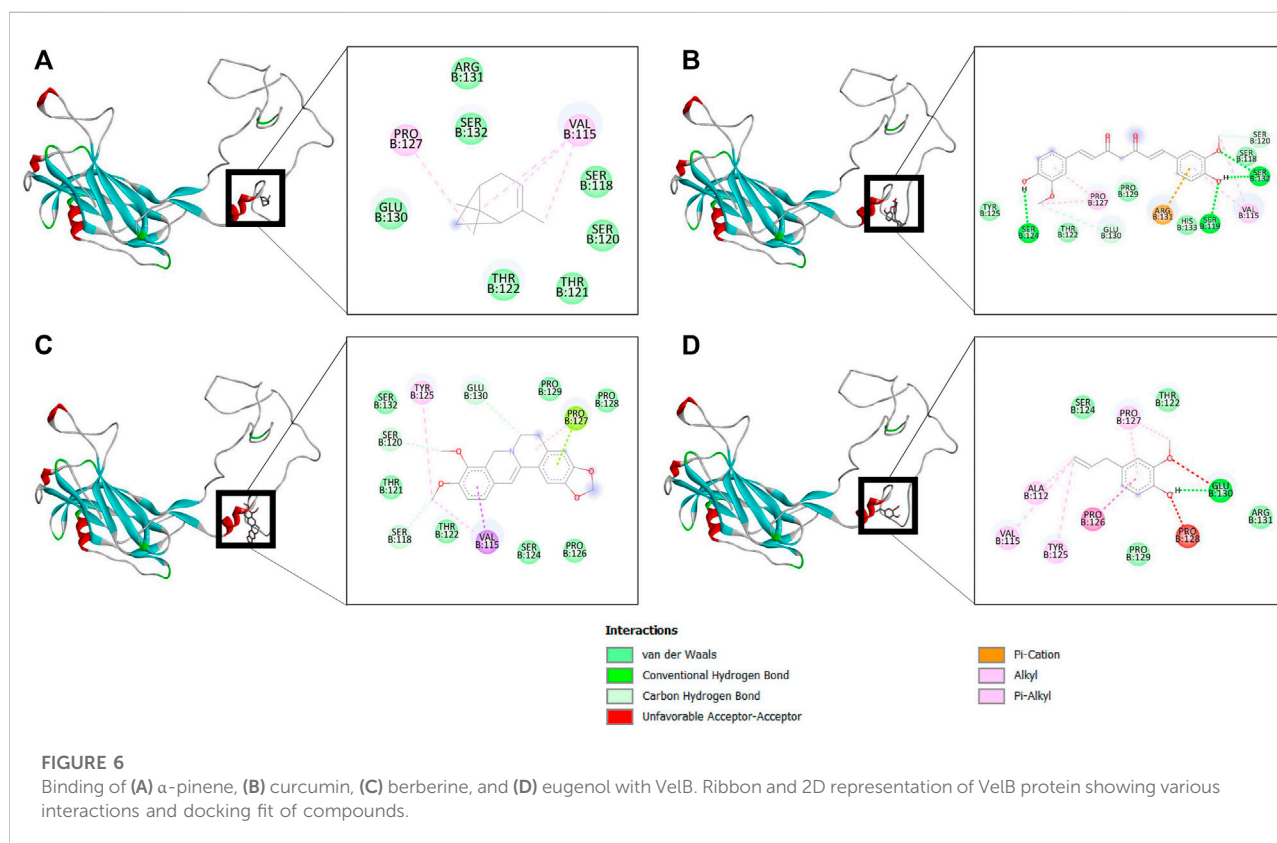
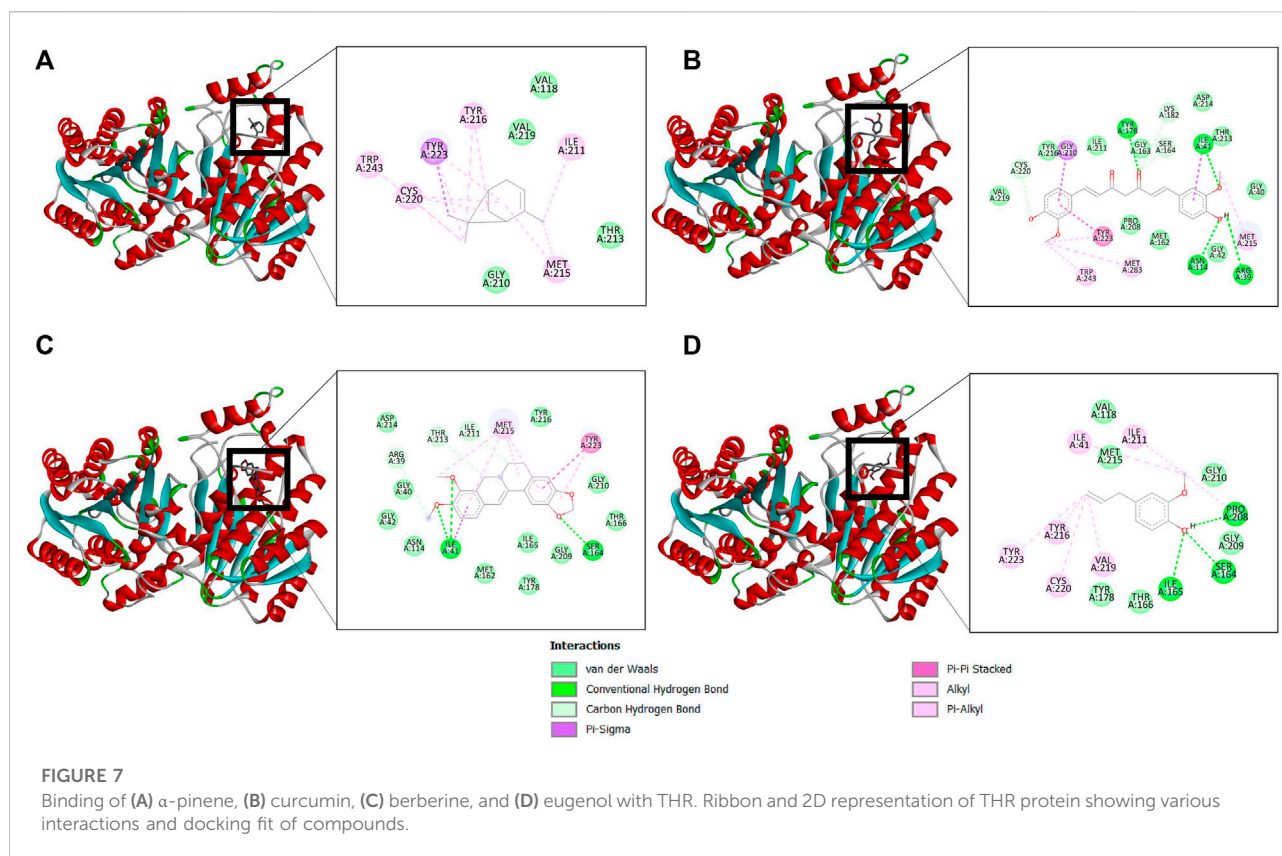


FIGURE 6
Binding of (A) α -pinene, (B) curcumin, (C) berberine, and (D) eugenol with VelB. Ribbon and 2D representation of VelB protein showing various interactions and docking fit of compounds.

distribution (Figure 8A) reflected that the inhibitor bound state of THR experiences comparatively less structural deviation along the simulation time and has 3.22 Å mean RMSD, while the THR

receptor has 4.58 Å mean RMSD. RoG defines the compactness of the structure, and Figure 8D shows that RoG remained consistent for the THR complex in comparison to the THR

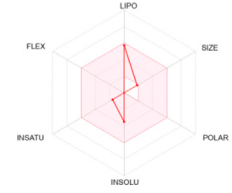
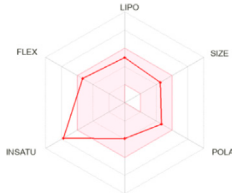
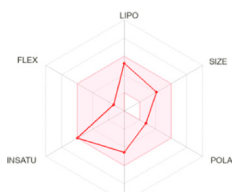
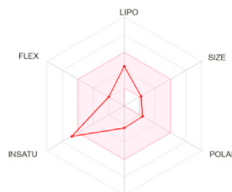


receptor which further highlights that curcumin binding at the active site of THR protein enhances the structural stability. The same was confirmed by SASA analysis (Figure 8B). Apart from the protein structure-level dynamics, the protein residue-level dynamics was attained by RMSF analysis (Figure 8C) which depicted that in the presence of an inhibitor, binding site residues 114 and 200–220 have less fluctuation and therefore stabilize the complex.

In the case of the VelB complex, structural order parameter analysis with respect to the VelB receptor showed that RMSD distribution showed a more zigzag pattern for receptor protein with the mean RMSD value of 12.81 Å, while the bound complex attained more stability in the presence of an inhibitor (Figure 9A). RoG (Figure 9D) and SASA analyses (Figure 9B) further explained the stability of the bound complex along the simulation. SASA achieved the plateau after 25 ns, while SASA for the unbound state receptor was still declining. Residue-level RMSF analysis also displayed that binding site residues 110–135 restricted the local fluctuation, thereby increasing the stability of the complex (Figure 9C). Curcumin binding stability at the binding site of respective proteins was further computed and analysed by monitoring the average distance between centre mean position of curcumin and selected binding site residues as shown in Supplementary Figure S3 for the THR–curcumin complex and Supplementary Figure S4 for the VelB–curcumin

complex. In case of the THR complex, binding site residues such as ARG39, ASN114, SER164, TYR178, and LYS182 retained the equilibrium distance after 25 ns. As ASN114, TYR178, and LYS182 were present at the beta and helix secondary structure, the distances from these residues were found to be more consistent compared to ARG39 and SER164 which were located in the turn of the secondary structure. This analysis showed that binding site residues maintained favourable distance for a stable molecular interaction. In case of the VelB complex, all binding site residues were located at the coiled coil secondary structure, and analysis of the centre mean distance from curcumin with binding site residues will further provide evidence of curcumin stability in the coiled coil binding region of VelB receptor protein. As it can be seen from Supplementary Figure S4, all six selected binding site residues (ARG131, PRO127, SER119, HIS133, SER118, and TYR125) maintained equilibrium distances after 25 ns simulation time. All distances have undergone shape drifts transiently at ~65 ns and thereafter returned to initial equilibrium distances. This calculation clearly suggests that despite the coiled coil region of the binding site, curcumin well oriented in the VelB-binding site. To further establish the fact that curcumin orientation in the protein-binding site of both proteins remains stable, curcumin RMSD was measured and plotted in a time evolution manner in Supplementary Figure S5. Similar to the distance in

TABLE 4 Physicochemical analysis of potential inhibitors of THR and VelB proteins of *C. lunata*.

Compounds	ADME properties (Lipinski's Rule of Five)	Molecular weight	Molecular formula	Radar diagram
α -Pinene	Molecular weight (g/mol)	136.23	$C_{10}H_{16}$	
	LogP	3.44		
	H-bond donor	0		
	H-bond acceptor	0		
	Violation	0		
Curcumin	Molecular weight (g/mol)	368.38	$C_{21}H_{20}O_6$	
	LogP	1.47		
	H-bond donor	2		
	H-bond acceptor	6		
	Violation	0		
Berberine	Molecular weight (g/mol)	336.36	$C_{20}H_{18}NO_4$	
	LogP	2.19		
	H-bond donor	0		
	H-bond acceptor	4		
	Violation	0		
Eugenol	Molecular weight (g/mol)	164.20	$C_{10}H_{12}O_2$	
	LogP	2.25		
	H-bond donor	1		
	H-bond acceptor	2		
	Violation	0		

Note: According to Lipinski's rule, molecular weight <500 g/mol, lipophilicity (MlogP) <5, hydrogen bond acceptor <5, and hydrogen bond donor <10 are considered.

Supplementary Figures S3, S4, curcumin stabilized its orientation in the respective binding site after 25 ns of simulation.

Apart from complex structural stability, interaction stability was assessed by calculating the number of hydrogen bonds formed between curcumin and receptor protein along the simulation time. The THR–curcumin complex formed on average one hydrogen bond with more than 90% occupancy, while the VelB–curcumin complex formed on average one hydrogen bond with more than 82% of simulation time (Figure 10).

Discussion

A fundamental problem with *Curvularia* infections is the risk of transferring a genetically evolved isolate from farms to humans. Such transferability through the human–plant interaction or intake of contaminated air could cause

resistance against available antifungal drugs since field isolates suffer from fungicide selective pressure and undergo virulence differentiation to adapt to adverse conditions (Bengyella et al., 2017). *Curvularia* spp. are sensitive to triazoles, and there are chances of interlocking lifestyle and fungicide pressure which may lead to the development of resistance in immunocompromised patients receiving azole therapy (Bengyella et al., 2017). Occurrence of antifungal resistance in fungal diseases along with a genetically evolved *Curvularia* spp. imparts the search for new therapy to control fungal infections though it is a major challenge of present-day treatment. There are known natural compounds that can control the disease caused by *C. lunata*; for example, essential oil of *Cymbopogon citratus* and extract of *Cinnamomum zeylani* are reported to inhibit conidial germination and reduce disease progress (Mishra et al., 2009; Mourão et al., 2017).

In the present study, natural compounds, namely, α -pinene, curcumin, berberine, and eugenol were studied for their

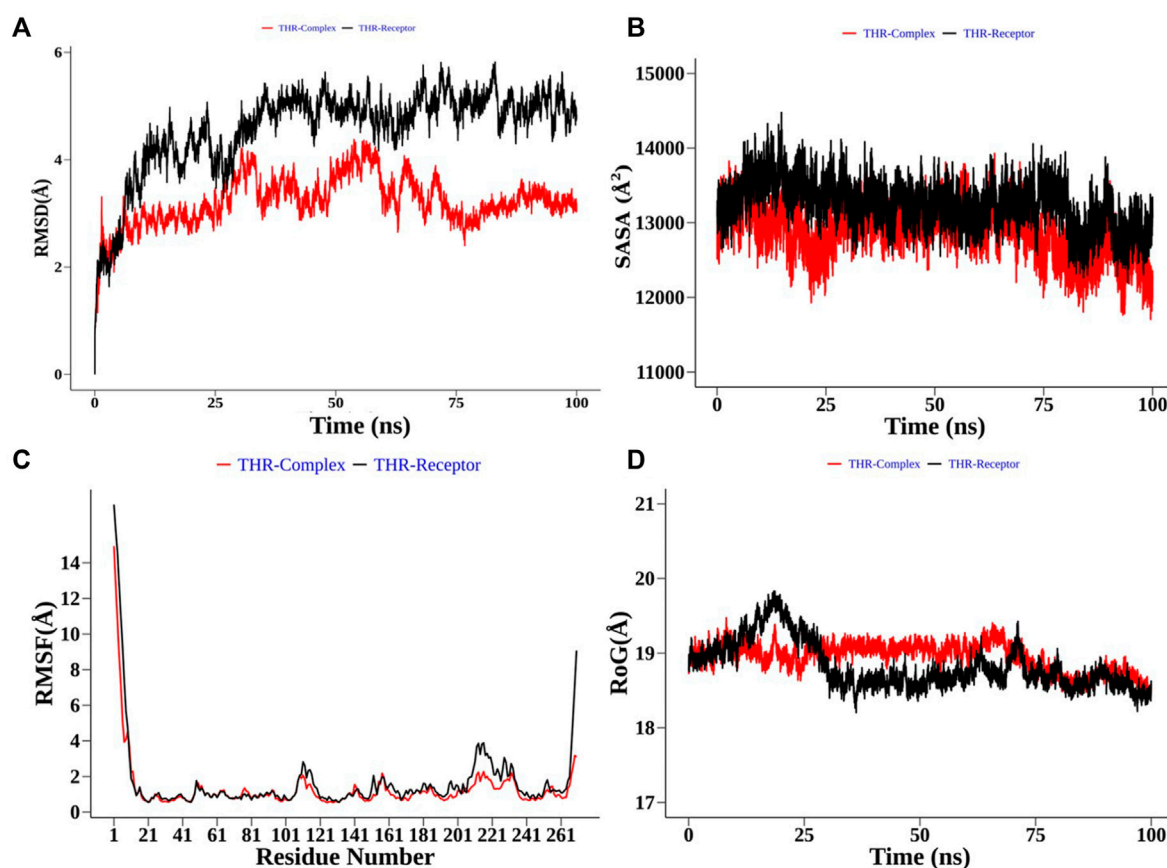


FIGURE 8
Structural order parameter analysis of the THR complex (curcumin) with respect to the THR receptor. (A) Root mean square deviation (RMSD), (B) solvent-accessible surface area (SASA), (C) root mean square fluctuation (RMSF), and (D) radius of gyration (RoG) analysis.

antifungal activity against *C. lunata*. The fungal growth was completely inhibited by curcumin (78 µg/ml), berberine (156 µg/ml), eugenol, (156 µg/ml) and α-pinene (1250 µg/ml). In the studies conducted elsewhere, α-pinene showed significant antifungal activity, with greater inhibitory activity against *Candida parapsilosis*, and proved to be effective in inhibiting fungal growth (Nóbrega et al., 2021). Previously reported MIC of berberine was 125 µg/ml against *C. lunata* (Kokkrua et al., 2020), whereas in the current study it was calculated as 156 µg/ml. Eugenol also inhibited the growth of fungi at 156 µg/ml, and similar results were reported for the antifungal activity of eugenol against *T. rubrum* strains (64–512 µg/ml) (de Oliveira Pereira et al., 2013). The extract of *Curcuma longa* also reported strong inhibitory effect on various pathogenic fungi at 1 mg/ml concentration (Chen et al., 2018), and antifungal effect of curcumin has been reported on *Aspergillus* spp. at 0.2 mg/ml (Martins et al., 2008; Gitika et al., 2019) and *Candida* spp. at 0.1–2 mg/ml (Narayanan et al., 2020).

The interaction between host plants and conidia of *C. lunata* begins with the adherence of conidia onto the leaf surface (Xie et al., 2020). Once the conidia adhere, they start to germinate and form appressoria. The cell wall of appressoria contains melanin which aids to provide mechanical strength for host tissue penetration. In *C. lunata*, sporulation and germination are crucial steps for spreading the disease (Xie et al., 2020). In the present study, our compounds suppressed sporulation as well as hyphal growth at IC₅₀ (curcumin, 39 µg/ml; berberine, 78 µg/ml; eugenol, 78 µg/ml; and α-pinene, 625 µg/ml). The spore germination and its pathogenicity were studied on the onion epidermis layer, where the compound-treated (α-pinene, curcumin, berberine, and eugenol) conidia failed to adhere and germinate on the surface. The invasive hyphae of *C. lunata* were observed when untreated conidia were observed under light and electron microscope. Another study on spore germination of *Curvularia maculans* reported that berberine affected the germination process of the fungus (Basha et al., 2002). Biochemical assay showed that curcumin exhibited a

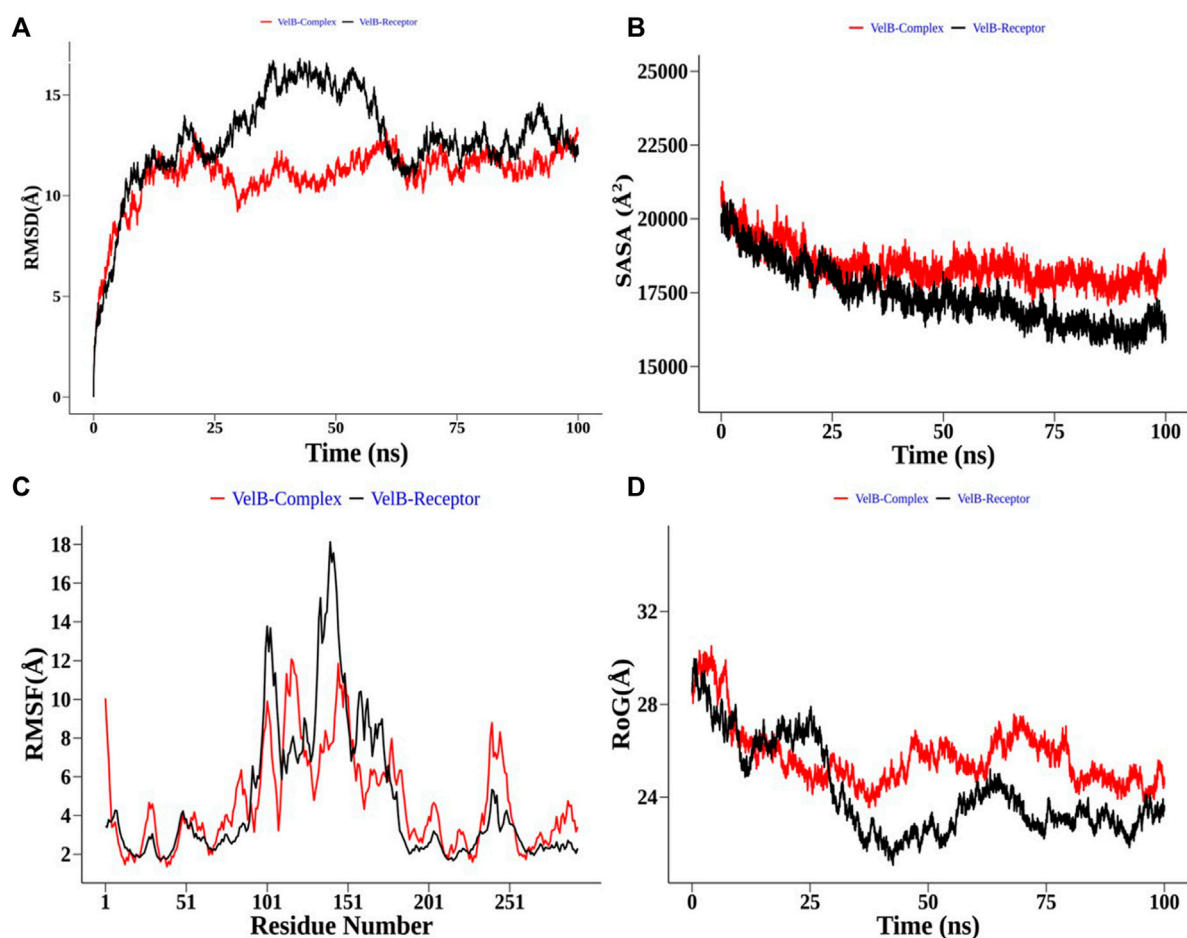


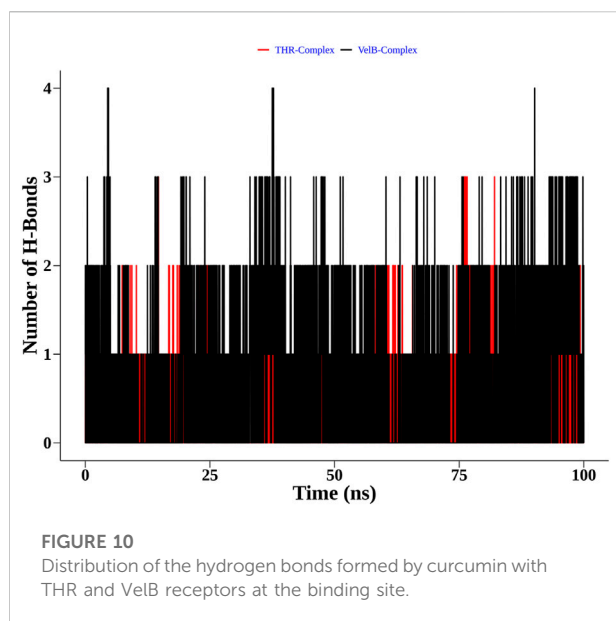
FIGURE 9

Structural order parameter analysis of the VelB complex (curcumin) with respect to the VelB receptor. (A) Root mean square deviation (RMSD), (B) solvent-accessible surface area (SASA), (C) root mean square fluctuation (RMSF), and (D) radius of gyration (RoG) analysis.

significant reduction in conidial hydrophobicity as compared to the control. It might be due to curcumin interacting with the conidial wall/membrane and disturbing the cell integrity (Srivastava et al., 2020).

Unlike other fungi, not much is studied about *C. lunata* pathogenesis at the molecular level. Melanin is one of the reported virulence factors which can improve mechanical strength of appressorium required for penetration inside the host (Rizner and Wheeler, 2003). It is produced from 1,8-dihydroxynaphthalene (DHN) via the pentaketide pathway in this fungus; crucial genes involved in the pathway are *pks18*, *scd1*, and *brn1* (Baker et al., 2006; Liu T. et al., 2011). According to Liu T. et al (2011) and Wang et al. (2020), *brn1* gene deletion led to improper DHN-melanin biosynthesis as well as

accumulation in the cell wall which also affects the production of other mycotoxins by this fungus. The current study observed significant down-regulation of *pks18* and *brn1* genes when the pathogen comes in contact with bioactive compounds. The velvet genes/proteins also play a significant role in regulating secondary metabolism, cell wall integrity pathway, and sporulation (Calvo, 2008; Gao et al., 2017). It has been reported that deletion of *velB* gene exhibited reduced growth rate and conidiation with increased aerial hyphae formation (Gao et al., 2017). In the present study, the expression of *velB* gene, which is a key member of velvet protein synthesis, was significantly down-regulated in the presence of α -pinene, curcumin, berberine, and eugenol. Mitogen-activated protein kinase gene *clm1* regulated cell



wall integrity, conidiophore formation, and cell-degrading enzyme activity (Liu T. et al., 2011; Ni et al., 2018). The presence of bioactive compounds (α -pinene, curcumin, berberine, and eugenol) showed that there was a significant down-regulation of *clm1* gene of *C. lunata*, which might result in decreased cell wall integrity as well as conidia formation. The qRT-PCR analysis showed that the expression level of *pks18*, *brn1*, *velB*, and *clm1* genes were reduced compared to that in the control. Hence, conidia were unable to maintain integrity for their growth and sporulation.

Reported mutant studies of *brn1*, MAPK gene, and toxin-related gene deletion led to a decrease in sporulation as well as virulence (Liu T. et al., 2011; Gao et al., 2012). In case of compound treatment, it was found that the conidia germination process was inhibited as shown in SEM images, and therefore the expression of *brn1*, *clm1*, and *velB* genes could be decreased. Another mutant study reported that $\Delta ClVelB$ showed high expression of *pks18*, *brn1*, and *cmr1* genes at 48 h and 60 h gene expression study (Gao et al., 2017). Also, *velB* gene deletion indicated the increase in osmotic resistance which suggests that *velB* gene involved in the regulation of cell wall integrity. Our study also suggested that the presence of compounds decreases *velB* gene expression which might result in decreased toxin production and cell wall integrity.

Furthermore, to understand the possible targets of compounds (α -pinene, curcumin, berberine, and eugenol), an *in silico* docking approach was used for virulence proteins involved in the melanin pathway and conidiation. The

virulence proteins THR and VelB were responsible for the melanin biosynthesis pathway, and conidiation and methyl 5-(hydroxymethyl) furan-2-carboxylate toxin production, respectively. Molecular docking of compounds with THR and VelB proteins showed significant negative binding affinity using AutoDock4.2.3. Polyketide synthase plays an important role in the pathogenicity, which synthesised THR, another crucial step for the biosynthesis of DHN-melanin in both mycelia and conidia (Lu et al., 2022). The velvet family protein is reported to play a key role in the regulation of secondary metabolism, fungal growth, and sporulation in many filamentous fungi (Gao et al., 2017). The main member of the velvet protein family is VelB protein in a few fungal species, according to the literature (Bayram et al., 2008; Wiemann et al., 2010; Yang et al., 2013; Gao et al., 2017). Curcumin is observed to have the highest negative binding energy and hydrogen bonds with both THR and VelB proteins. Berberine and eugenol also showed good binding affinity with both virulence proteins, whereas α -pinene did not form hydrogen bonds with THR as well as VelB proteins and also the binding affinity was comparatively lower. The docking interaction indicated that the protein-ligand complexes that had good binding affinity were those that formed the highest number of hydrogen bonds (Shamsi et al., 2022). Curcumin-bound complexes of THR and VelB receptor proteins were employed to understand the stability and dynamical behaviour of the complexes. Also, 100 ns MD simulation analysis highlights the gain in the structural stability after binding of the curcumin molecules. In both complexes, the curcumin molecule remains stable in the corresponding binding site of the protein and forms one hydrogen bond with more than 80% of the time. This result suggests the possibility of designing a high-affinity inhibitor carrying the curcumin scaffold entity.

Plant-derived compounds (α -pinene, curcumin, berberine, and eugenol) are effective against *C. lunata*. The natural bioactive compounds possess less toxicity and associated side effects, which makes them a suitable candidate for drug discovery. Among all, curcumin exhibited effective antifungal properties to inhibit the growth of the fungus. Conidia adherence and their germination were restricted by the activity of these compounds. It was shown that the compounds down-regulate *brn1*, *velB*, *pks18*, and *clm1* gene expression, leading to decreased cell wall integrity and sporulation. The correlation between the virulence gene down-regulation and *in silico* molecular docking interaction represented the potential antifungal activity of bioactive compounds and highlighted virulence proteins THR and VelB as a possible drug target.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/Supplementary Material.

Author contributions

HK and LG performed literature search, experimental part, and the manuscript drafting; PK performed molecular dynamics and simulation analysis; PS conducted the fungi identification; and PV and AS conceptualised the idea as well as critically analysed the results and manuscript.

Acknowledgments

The authors are grateful to AIARS-Amity University Uttar Pradesh for scanning electron microscopy, and the School of Computational and Integrative Sciences, Jawaharlal Nehru University, New Delhi, India, for providing the facilities to carry out the *in silico* research work. The authors are grateful to the Amity

University, Uttar Pradesh NOIDA campus for providing the facilities to carry out the research work.

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.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmolb.2022.1055945/full#supplementary-material>

References

- Abraham, M. J., Murtola, T., Schulz, R., Páll, S., Smith, J. C., Hess, B., et al. (2015). Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* 1 (2), 19–25. doi:10.1016/j.softx.2015.06.001
- Abubakar, M. N., and Likita, M. S. (2021). Effect of combined fungicides on the mycelial growth of *Curvularia lunata*. *Int. Innovative Agric. Bio. Res* 9, 1–9.
- Alex, D., Li, D., Calderone, R., and Peters, S. M. (2013). Identification of *Curvularia lunata* by polymerase chain reaction in a case of fungal endophthalmitis. *Med. Mycol. Case Rep.* 2, 137–140. doi:10.1016/j.mmcr.2013.07.001
- Alexander, B. D. (2017). *Reference method for broth dilution antifungal susceptibility testing of filamentous fungi*.
- Amin, E. N. E., and Abdalla, M. H. (1980). Survival of *Curvularia lunata* var. *aeria* in soil. *Mycopathologia* 71, 137–140. doi:10.1007/BF00473059
- Andrews, J. M. (2001). Determination of minimum inhibitory concentrations. *J. Antimicrob. Chemother.* 48, 5–16. doi:10.1093/jac/48.suppl_1.5
- Baker, S. E., Kroken, S., Inderbitzin, P., Asvarak, T., Li, B.-Y., Shi, L., et al. (2006). Two polyketide synthase-encoding genes are required for biosynthesis of the polyketide virulence factor, T-toxin, by *Cochliobolus heterostrophus*. *Mol. Plant. Microbe. Interact.* 19, 139–149. doi:10.1094/MPMI-19-0139
- Basha, S. A., Mishra, R. K., Jha, R. N., Pandey, V. B., and Singh, U. P. (2002). Effect of berberine and (±)-bicuculline isolated from *Corydalis chaerophylla* on spore germination of some fungi. *Folia Microbiol.* 47, 161–165. doi:10.1007/BF02817675
- Bayram, Ö., Krappmann, S., Ni, M., Bok, J. W., Helmstaedt, K., Valerius, O., et al. (2008). VelB/VeA/LaeA complex coordinates light signal with fungal development and secondary metabolism. *Science* 320, 1504–1506. doi:10.1126/science.1155888
- Beckett, A. R., Kahn, S. A., Seay, R., and Lintner, A. C. (2017). Invasive *Curvularia* infections in burn patients: A case series. *Surg. Infect. Case Rep.* 2, 76–79. doi:10.1089/crsi.2017.0018
- Bengyella, L., Yekwa, E., Waikhom, L. D., Nawaz, K., Iftikhar, S., Motloi, S., et al. (2017). Upsurge in *Curvularia* infections and global emerging antifungal drug resistance. *Asian J. Sci. Res.* 10, 299–307. doi:10.3923/ajsr.2017.299.307
- Berendsen, H. J. C., Grigera, J. R., and Straatsma, T. P. (1987). The missing term in effective pair potentials. *J. Phys. Chem.* 91, 6269–6271. doi:10.1021/j100308a038
- Bisht, S., Balodi, R., Ghatak, A., and Kumar, P. (2018). Determination of susceptible growth stage and efficacy of fungicidal management of *Curvularia* leaf spot of maize caused by *Curvularia lunata* (Wakker) Boedijn. *Maydica* 61, 5.
- Bjelkmar, P., Larsson, P., Cuendet, M. A., Hess, B., and Lindahl, E. (2010). Implementation of the CHARMM force field in GROMACS: Analysis of protein stability effects from correction maps, virtual interaction sites, and water models. *J. Chem. Theory Comput.* 6, 459–466. doi:10.1021/ct900549r
- Calvo, A. M. (2008). The VeA regulatory system and its role in morphological and chemical development in fungi. *Fungal Genet. Biol.* 45, 1053–1061. doi:10.1016/j.fgb.2008.03.014
- Chang, Y.-C., Graf, E., and Green, A. M. (2019). Invasive *Curvularia* infection in pediatric patients with hematologic malignancy identified by fungal sequencing. *J. Pediatr. Infect. Dis. Soc.* 8, 87–91. doi:10.1093/jpids/piy092
- Chen, C., Long, L., Zhang, F., Chen, Q., Chen, C., Yu, X., et al. (2018). Antifungal activity, main active components and mechanism of *Curcuma longa* extract against. *PLoS One* 13, e0194284. doi:10.1371/journal.pone.0194284
- Chida, T., and Sisler, H. D. (1987). Restoration of appressorial penetration ability by melanin precursors in *Pyricularia oryzae* treated with antipenetrants and in melanin-deficient mutants. *J. Pestic. Sci.* 12, 49–55. doi:10.1584/jpestics.12.49
- Chowdhary, A., Meis, J. F., Guarro, J., de Hoog, G. S., Kathuria, S., Arendrup, M. C., et al. (2014). ESCMID and ECMM joint clinical guidelines for the diagnosis and management of systemic phaeohyphomycosis: Diseases caused by black fungi. *Clin. Microbiol. Infect.* 20 (3), 47–75. doi:10.1111/1469-0691.12515
- Clinical and Laboratory Standards Institute (2016). *Performance standards for antimicrobial susceptibility testing, 26th ed CLSI document M100-S*. Wayne, PA: Clinical and Laboratory Standards Institute.
- Cuervo-Parra, J. A., Romero-Cortes, T., Ortiz-Yescas, G., and Ramirez-Lepe, M. (2012). *International congress food science and food biotechnology in developing countries*. Nayarit, México: Nuevo Vallarta. Isolation and molecular identification of *Curvularia lunata*/Cochliobolus lunatus causal agent of leaf spot disease of cocoa

- de Oliveira Pereira, F., Mendes, J. M., and de Oliveira Lima, E. (2013). Investigation on mechanism of antifungal activity of eugenol against *Trichophyton rubrum*. *Med. Mycol.* 51, 507–513. doi:10.3109/13693786.2012.742966
- Ellis, M. B. (1971). *Dematiaceae hyphomycetes*. Kew, Surrey, England: Commonwealth Mycological Institute, 608.
- Fu, D., Li, J., Yang, X., Li, W., Zhou, Z., Xiao, S., et al. (2022). Iron redistribution induces oxidative burst and resistance in maize against *Curvularia lunata*. *Planta* 256, 46. doi:10.1007/s00425-022-03963-7
- Gao, J.-X., Yu, C.-J., Wang, M., Sun, J.-N., Li, Y.-Q., and Chen, J. (2017). Involvement of a velvet protein ClVeB in the regulation of vegetative differentiation, oxidative stress response, secondary metabolism, and virulence in *Curvularia lunata*. *Sci. Rep.* 7, 46054. doi:10.1038/srep46054
- Gao, S., Liu, T., Li, Y., Wu, Q., Fu, K., and Chen, J. (2012). Understanding resistant germplasm-induced virulence variation through analysis of proteomics and suppression subtractive hybridization in a maize pathogen *Curvularia lunata*. *Proteomics* 12, 3524–3535. doi:10.1002/pmic.201200105
- Garcia-Aroca, T., Doyle, V., Singh, R., Price, T., and Collins, K. (2018). First report of *Curvularia* leaf spot of corn, caused by *Curvularia lunata*, in the United States. *Plant Health Prog.* 19, 140–142. doi:10.1094/PHP-02-18-0008-BR
- Ghany Tm, A., Rahman, M. A., Moustafa E, S., and Mohamed A, N. (2015). Efficacy of botanical fungicides against *Curvularia lunata* at molecular levels. *J. Plant Pathol. Microbiol.* 06. doi:10.4172/2157-7471.1000289
- Giri, D. K., Sims, W. P., Sura, R., Cooper, J. J., Gavrilov, B. K., and Mansell, J. (2011). Cerebral and renal phaeohyphomycosis in a dog infected with *bipolaris* species. *Vet. Pathol.* 48, 754–757. doi:10.1177/0300985810375239
- Gitika, A., Mishra, R., Panda, S. K., Mishra, C., Ranjan, P., and hoo, S. (2019). Evaluation of antifungal activity of curcumin against *Aspergillus flavus*. *Int. J. Curr. Microbiol. Appl. Sci.* 8, 2323–2329. doi:10.20546/ijcmas.2019.807.284
- Gupta, L., Cm, S., Sharma, A., Sk, A., Vermani, M., Konwar, D., et al. (2019). *Myristica fragrans* extract inhibits melanin biosynthesis, hyphal growth and appressorium formation in *Magnaporthe oryzae*. *Curr Res Environ Appl Mycol J. Fungal Biol.* 9, 255–270. doi:10.5943/cream/9/1/22
- Gupta, L., Sen, P., Bhattacharya, A. K., and Vijayaraghavan, P. (2022). Isoeugenol affects expression pattern of conidial hydrophobin gene RodA and transcriptional regulators MedA and SomA responsible for adherence and biofilm formation in *Aspergillus fumigatus*. *Arch. Microbiol.* 204, 214. doi:10.1007/s00203-022-02817-w
- Gupta, S., Dubey, A., and Singh, T. (2017). *Curvularia lunata* as, a dominant seed-borne pathogen in Dalbergia sissoo Roxb: Its location in seed and its phytopathological effects. *Afr. J. Plant Sci.* 11, 203–208. doi:10.5897/AJPS2017.1529
- Hof, H. (2001). Critical annotations to the use of azole antifungals for plant protection. *Antimicrob. Agents Chemother.* 45, 2987–2990. doi:10.1128/AAC.45.11.2987-2990.2001
- Jørgensen, L. N., and Heick, T. M. (2021). Azole use in agriculture, horticulture, and wood preservation – is it indispensable? *Front. Cell. Infect. Microbiol.* 11, 730297. doi:10.3389/fcimb.2021.730297
- Joung, I. S., and Cheatham, T. E. (2008). Determination of alkali and halide monovalent ion parameters for use in explicitly solvated biomolecular simulations. *J. Phys. Chem. B* 112, 9020–9041. doi:10.1021/jp8001614
- Karthika, A., Ramachandran, B., Chitra, J., Prabhu, D., Rajamanikandan, S., Veerapandian, M., et al. (2021). Molecular dynamics simulation of Toxin-Antitoxin (TA) system in *Acinetobacter baumannii* to explore the novel mechanism for inhibition of cell wall biosynthesis: Zeta Toxin as an effective therapeutic target. *J. Cell. Biochem.* 2021, 1832–1847. doi:10.1002/jcb.30137
- Kokkrua, S., Ismail, S. I., Mazlan, N., and Dethoup, T. (2020). Efficacy of berberine in controlling foliar rice diseases. *Eur. J. Plant Pathol.* 156, 147–158. doi:10.1007/s10658-019-01871-3
- Kumar, C. G., Mongolla, P., Pombala, S., Kamle, A., and Joseph, J. (2011). Physicochemical characterization and antioxidant activity of melanin from a novel strain of *Aspergillus bridgeri* ICTF-201. *Lett. Appl. Microbiol.* 53, 350–358. doi:10.1111/j.1472-765X.2011.03116.x
- Lee, S. B., Milgroom, M. G., and Taylor, J. W. (1988). A rapid, high yield mini-prep method for isolation of total genomic DNA from fungi. *Fungal Genet. Rep.* 35, 23. doi:10.4148/1941-4765.1531
- Liu, L. M., Huang, S. W., Wang, L., Hou, E. Q., and Xiao, D. F. (2014). First report of leaf blight of rice caused by *Cochliobolus lunatus* in China. *Plant Dis.* 98, 686. doi:10.1094/PDIS-03-13-0303-PDN
- Liu, T., Xu, S., Liu, L., Zhou, F., Hou, J., and Chen, J. (2011a). Cloning and characteristics of Brn1 gene in *Curvularia lunata* causing leaf spot in maize. *Eur. J. Plant Pathol.* 131, 211–219. doi:10.1007/s10658-011-9800-8
- Liu, W., Zhou, X., Li, G., Li, L., Kong, L., Wang, C., et al. (2011b). Multiple plant surface signals are sensed by different mechanisms in the rice blast fungus for appressorium formation. *PLoS Pathog.* 7, e1001261. doi:10.1371/journal.ppat.1001261
- Liu, Z., Liu, T., Chen, D., and Hou, J. (2019). First report of *Curvularia lunata* causing leaf spots on Partridge tea [*Mallotus oblongifolius* (Miq.) Müll. Arg.] in China. *J. Plant Pathol.* 101, 439. doi:10.1007/s42161-018-00208-z
- Lu, Z., Wang, S., Dou, K., Ren, J., and Chen, J. (2022). The interpretation of the role of a polyketide synthase ClPKS18 in the pathogenicity of *Curvularia lunata*. *Front. Microbiol.* 13, 853140. doi:10.3389/fmicb.2022.853140
- Martins, C. V. B., da Silva, D. L., Neres, A. T. M., Magalhaes, T. F. F., Watanabe, G. A., Modolo, L. V., et al. (2008). Curcumin as a promising antifungal of clinical interest. *J. Antimicrob. Chemother.* 63, 337–339. doi:10.1093/jac/dkn488
- Matasyoh, J. C., Kiplimo, J. J., Karubiu, N. M., and Hailstorks, T. P. (2007). Chemical composition and antimicrobial activity of essential oil of *Tarchonanthus camphoratus*. *Food Chem. x* 101, 1183–1187. doi:10.1016/j.foodchem.2006.03.021
- McGibbon, R. T., Beauchamp, K. A., Harrigan, M. P., Klein, C., Swails, J. M., Hernández, C. X., et al. (2015). MDTraj: A modern open library for the analysis of molecular dynamics trajectories. *Biophys. J.* 109, 1528–1532. doi:10.1016/j.bpj.2015.08.015
- Mishra, A. K., Mishra, A., Kehri, H. K., Sharma, B., and Pandey, A. K. (2009). Inhibitory activity of Indian spice plant *Cinnamomum zeylanicum* extracts against *Alternaria solani* and *Curvularia lunata*, the pathogenic dematiaceous moulds. *Ann. Clin. Microbiol. Antimicrob.* 8, 9. doi:10.1186/1476-0711-8-9
- 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
- Mourão, D. de S. C., Ferreira de Souza Pereira, T., Souza, D. J. de, Chagas Júnior, A. F., Dalcin, M. S., Veloso, R. A., et al. (2017). Essential oil of *Cymbopogon citratus* on the control of the *Curvularia* leaf spot disease on maize. *Med. (Basel)* 4, E62. doi:10.3390/medicines4030062
- Nagoor Meeran, M. F., Javed, H., Al Taei, H., Azimullah, S., and Ojha, S. K. (2017). Pharmacological properties and molecular mechanisms of thymol: Prospects for its therapeutic potential and pharmaceutical development. *Front. Pharmacol.* 8, 380. doi:10.3389/fphar.2017.00380
- Narayanan, V., Muddaiah, S., Shashidara, R., Sudheendra, U., Deepthi, N., and Samaranayake, L. (2020). Variable antifungal activity of curcumin against planktonic and biofilm phase of different *Candida* species. *Indian J. Dent. Res.* 31, 145–148. doi:10.4103/ijdr.IJDR_521_17
- Naz, F., Abbas, M. F., Rauf, C. A., Tariq, A., Mumtaz, A., Irshad, G., et al. (2017). First report of *Colletotrichum gloeosporioides* causing anthracnose on loquat in Pakistan. *Plant Dis.* 101, 1550. doi:10.1094/PDIS-11-16-1551-PDN
- Ni, X., Gao, J., Yu, C., Wang, M., Sun, J., Li, Y., et al. (2018). MAPKs and acetyl-CoA are associated with *Curvularia lunata* pathogenicity and toxin production in maize. *J. Integr. Agric.* 17, 139–148. doi:10.1016/S2095-3119(17)61697-6
- Nóbrega, J. R., Silva, D. de F., Andrade Júnior, F. P., Sousa, P. M. S., Figueiredo, P. T. R. de, Cordeiro, L. V., et al. (2021). Antifungal action of α -pinene against *Candida* spp. isolated from patients with otomycosis and effects of its association with boric acid. *Nat. Prod. Res.* 35, 6190–6193. doi:10.1080/14786419.2020.1837803
- 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
- Parrinello, M., and Rahman, A. (1980). Crystal structure and pair potentials: A molecular-dynamics study. *Phys. Rev. Lett.* 45, 1196–1199. doi:10.1103/PhysRevLett.45.1196
- Pihet, M., Vandeputte, P., Tronchin, G., Renier, G., Saulnier, P., Georgeault, S., et al. (2009). Melanin is an essential component for the integrity of the cell wall of *Aspergillus fumigatus* conidia. *BMC Microbiol.* 9, 177–188. doi:10.1186/1471-2180-9-177
- Rižner, T. L., and Wheeler, M. H. (2003). Melanin biosynthesis in the fungus *Curvularia lunata* (teleomorph: *Cochliobolus lunatus*). *Can. J. Microbiol.* 49, 110–119. doi:10.1139/w03-016
- Sankar, M., Ramachandran, B., Pandi, B., Mutharasappan, N., Ramasamy, V., Prabu, P. G., et al. (2021). *In silico* screening of natural phytochemicals towards identification of potential lead compounds to treat COVID-19. *Front. Mol. Biosci.* 8, 637122. doi:10.3389/fmolb.2021.637122
- Shamsi, A., Shahwan, M., Khan, M. S., Alhumaydhi, F. A., Alsagaby, S. A., Al Abdulmonem, W., et al. (2022). Mechanistic insight into binding of huperzine A with human serum albumin: Computational and spectroscopic approaches. *Molecules* 27, 797. doi:10.3390/molecules27030797
- Sharma, A., Goyal, S., Yadav, A. K., Kumar, P., and Gupta, L. (2022). *In-silico* screening of plant-derived antivirals against main protease, 3CLpro and

endoribonuclease, NSP15 proteins of SARS-CoV-2. *J. Biomol. Struct. Dyn.* 40, 86–100. doi:10.1080/07391102.2020.1808077

Shrivastava, A., Tadepalli, K., Goel, G., Gupta, K., and Kumar Gupta, P. (2017). Melanized fungus as an epidural abscess: A diagnostic and therapeutic challenge. *Med. Mycol. Case Rep.* 16, 20–24. doi:10.1016/j.mmcr.2017.04.001

Sivanesan, A. (1984). *Graminicolous species of bipolaris, Curvularia, Drechslera, exserohilum and their teleomorphs*. Wallingford: C.A.B International.

Snelders, E., Camps, S. M. T., Karawajczyk, A., Schaftenaar, G., Kema, G. H. J., van der Lee, H. A., et al. (2012). Triazole fungicides can induce cross-resistance to medical triazoles in *Aspergillus fumigatus*. *PLoS ONE* 7, e31801. doi:10.1371/journal.pone.0031801

Srivastava, A. K., Gupta, M., Singh, R. P., and Misra, P. (2020). Analysis on inhibiting pathogenic activity of fungi *Curvularia lunata* by essential oils. *Biosci. Biotechnol. Res. Commun.* 13, 2187–2192. doi:10.21786/bbrc/13.4/84

Szumilak, M., Galdyszynska, M., Dominska, K., Bak-Sypien, I., Merecz-Sadowska, A., Stanczak, A., et al. (2017). Synthesis, biological activity and preliminary *in silico* ADMET screening of polyamine conjugates with bicyclic systems. *Molecules* 22, 794. doi:10.3390/molecules22050794

Tóth, E. J., Varga, M., Takó, M., Homa, M., Jáger, O., Hermesz, E., et al. (2020). Response of human neutrophil granulocytes to the hyphae of the emerging fungal pathogen *Curvularia lunata*. *Pathogens* 9, 235. doi:10.3390/pathogens9030235

Uma Maheshwari Nallal, V., Padmini, R., Ravindran, B., Chang, S. W., Radhakrishnan, R., Almoallim, H. S. M., et al. (2021). Combined *in vitro* and *in silico* approach to evaluate the inhibitory potential of an underutilized allium vegetable and its pharmacologically active compounds on multidrug resistant *Candida* species. *Saudi J. Biol. Sci.* 28, 1246–1256. doi:10.1016/j.sjbs.2020.11.082

Vineetha, M., Palakkal, S., Sobhanakumari, K., Celine, M. I., and Letha, V. (2016). Verrucous onychomycosis caused by *Curvularia* in a patient with congenital pterygium. *Indian J. Dermatol.* 61, 701. doi:10.4103/0019-5154.193717

Wang, F., Gao, W., Sun, J., Mao, X., Liu, K., Xu, J., et al. (2020). NADPH oxidase CINOX2 regulates melanin-mediated development and virulence in *Curvularia*

lunata. *Mol. Plant. Microbe. Interact.* 33, 1315–1329. doi:10.1094/MPMI-06-20-0138-R

Wang, Y. J., Liu, T., Hou, J. M., and Zuo, Y. H. (2013). Isolation and identification of pathogenicity mutant of *Curvularia lunata* via restriction enzyme-mediated integration. *Indian J. Microbiol.* 53, 303–307. doi:10.1007/s12088-013-0385-0

White, T. J., Bruns, T., Lee, S., and Taylor, J. (1990). Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. *PCR Protoc. Elsevier*, 315–322. doi:10.1016/B978-0-12-372180-8.50042-1Refstyle

Wiemann, P., Brown, D. W., Kleigrew, K., Bok, J. W., Keller, N. P., Humpf, H.-U., et al. (2010). FfVel1 and FfLae1, components of a velvet-like complex in *Fusarium fujikuroi*, affect differentiation, secondary metabolism and virulence. *Mol. Microbiol.* 77, 972–994. doi:10.1111/j.1365-2958.2010.07263.x

Wu, D., Oide, S., Zhang, N., Choi, M. Y., and Turgeon, B. G. (2012). ChLae1 and ChVel1 regulate T-toxin production, virulence, oxidative stress response, and development of the maize pathogen *Cochliobolus heterostrophus*. *PLoS Pathog.* 8, e1002542. doi:10.1371/journal.ppat.1002542

Wu, Z.-H., Wang, T. H., Huang, W., and Qu, Y. B. (2001). A simplified method for chromosome DNA preparation from filamentous fungi. *Mycosystema* 20, 575–577.

Xie, S., Liu, J., Gu, S., Chen, X., Jiang, H., and Ding, T. (2020). Antifungal activity of volatile compounds produced by endophytic *Bacillus subtilis* DZSY21 against *Curvularia lunata*. *Ann. Microbiol.* 70, 2. doi:10.1186/s13213-020-01553-0

Xu, S., Chen, J., Liu, L., Wang, X., Huang, X., and Zhai, Y. (2007). Proteomics associated with virulence differentiation of *Curvularia lunata* in maize in China. *J. Integr. Plant Biol.* 49, 487–496. doi:10.1111/j.1744-7909.2007.00469.x

Yang, Q., Chen, Y., and Ma, Z. (2013). Involvement of BcVeA and BcVelB in regulating conidiation, pigmentation and virulence in *Botrytis cinerea*. *Fungal Genet. Biol.* 50, 63–71. doi:10.1016/j.fgb.2012.10.003

Zoete, V., Cuendet, M. A., Grosdidier, A., and Michelin, O. (2011). SwissParam: A fast force field generation tool for small organic molecules. *J. Comput. Chem.* 32, 2359–2368. doi:10.1002/jcc.21816



OPEN ACCESS

EDITED BY
Chandrabose Selvaraj,
Alagappa University, India

REVIEWED BY
Uchechukwu Chris Okoro,
University of Nigeria, Nsukka, Nigeria
Umesh Panwar,
Konkuk University, South Korea

*CORRESPONDENCE
Emmanuel Oluwadare Balogun,
✉ oluwadareus@yahoo.com,
✉ eobalogun@abu.edu.ng

SPECIALTY SECTION
This article was submitted to Biophysics,
a section of the journal
Frontiers in Molecular Biosciences

RECEIVED 14 October 2022
ACCEPTED 08 December 2022
PUBLISHED 19 December 2022

CITATION
Danazumi AU and Balogun EO (2022),
Microsecond-long simulation reveals
the molecular mechanism for the dual
inhibition of falcipain-2 and falcipain-3
by antimalarial lead compounds.
Front. Mol. Biosci. 9:1070080.
doi: 10.3389/fmolb.2022.1070080

COPYRIGHT
© 2022 Danazumi and Balogun. This is
an open-access article distributed
under the terms of the [Creative
Commons Attribution License \(CC BY\)](#).
The use, distribution or reproduction in
other forums is permitted, provided the
original author(s) and the copyright
owner(s) are credited and that the
original publication in this journal is
cited, in accordance with accepted
academic practice. No use, distribution
or reproduction is permitted which does
not comply with these terms.

Microsecond-long simulation reveals the molecular mechanism for the dual inhibition of falcipain-2 and falcipain-3 by antimalarial lead compounds

Ammar Usman Danazumi^{1,2} and
Emmanuel Oluwadare Balogun^{3,4,5,6*}

¹Faculty of Chemistry, Warsaw University of Technology, Warsaw, Poland, ²Groningen Research Institute of Pharmacy, University of Groningen, Groningen, Netherlands, ³Department of Biochemistry, Ahmadu Bello University, Zaria, Nigeria, ⁴Africa Centre of Excellence for Neglected Tropical Diseases and Forensic Biotechnology, Ahmadu Bello University, Zaria, Nigeria, ⁵Center for Discovery and Innovation in Parasitic Diseases, Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California, San Diego, San Diego, CA, United States, ⁶Department of Biomedical Chemistry, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan

The latest world malaria report revealed that human deaths caused by malaria are currently on the rise and presently stood at over 627,000 per year. In addition, more than 240 million people have the infection at any given time. These figures make malaria the topmost infectious disease and reiterate the need for continuous efforts for the development of novel chemotherapies. Malaria is an infectious disease caused majorly by the protozoan intracellular parasite *Plasmodium falciparum* and transmitted by mosquitoes. Reports abound on the central role of falcipains (cysteine protease enzymes) in the catabolism of hemoglobin for furnishing the plasmodium cells with amino acids that they require for development and survival in the hosts. Even though falcipains (FPs) have been validated as drug target molecules for the development of new antimalarial drugs, none of its inhibitory compounds have advanced beyond the early discovery stage. Therefore, there are renewed efforts to expand the collection of falcipain inhibitors. As a result, an interesting finding reported the discovery of a quinoliny oxamide derivative (QOD) and an indole carboxamide derivative (ICD), with each compound demonstrating good potencies against the two essential FP subtypes 2 (FP-2) and 3 (FP-3). In this study, we utilized microsecond-scale molecular dynamics simulation computational method to investigate the interactions between FP-2 and FP-3 with the quinoliny oxamide derivative and indole carboxamide derivative. The results revealed that quinoliny oxamide derivative and indole carboxamide derivative bound tightly at the active site of both enzymes. Interestingly, despite belonging to different chemical scaffolds, they are coordinated by almost identical amino acid residues via extensive hydrogen bond interactions in both FP-2 and FP-3. Our report provided molecular insights into the interactions between FP-2 and FP-3

with quinolinyl oxamide derivative and indole carboxamide derivative, which we hope will pave the way towards the design of more potent and druglike inhibitors of these enzymes and will pave the way for their development to new antimalarial drugs.

KEYWORDS

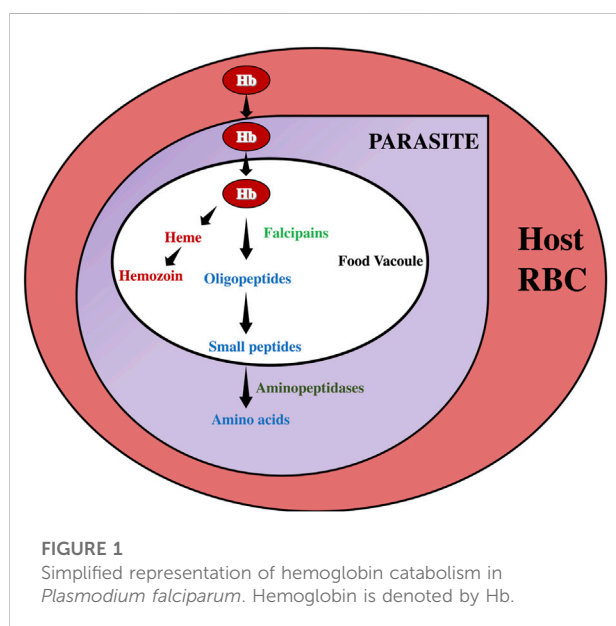
malaria, *Plasmodium falciparum*, falcipains, molecular dynamics simulation, enzyme inhibition

1 Introduction

Malaria continues to pose significant public health threats in the tropical and subtropical regions, given a global estimate of about 241 million cases in 2020, with yearly deaths reaching a record high of 627,000 in the same year. Malaria is an infectious disease caused by multiple species of intra-erythrocytic parasites belonging to the genus *Plasmodium*. The five species that are responsible for the disease in humans are *Plasmodium falciparum*, *P. vivax*, *P. malariae*, *P. ovale*, and *P. knowlesi*. From the standpoint of global health, *P. falciparum* is the most important because it is responsible for over 90% of global malaria cases and mortalities (Snow, 2015). The current situation is aggravated by the increased resistance to insecticides by the mosquito vectors as well as the spread of drug resistance to the available antimalarial drugs (WHO, 2022). This justified the unrelenting efforts towards the design and development of new drug candidates against malaria, along with expanding the plasmodial drug targets pool. Cysteine proteases have been identified as molecular targets that have become attractive for drug design not only against *Plasmodium sp* but also in related parasitic diseases (McKerrow et al., 1999; Pandey and Dixit, 2012; Siqueira-Neto et al., 2018).

Falcipains (FPs) are important cysteine proteases of *P. falciparum* that are central to acquisition by the parasite. To ensure the survival of *Plasmodium* in mammalian hosts, FPs catalyze the digestion of host hemoglobin in the food vacuole of *Plasmodium* to maintain the amino acid supply to the parasite (Hanspal et al., 2002). Figure 1 summarizes the pathway of hemoglobin catabolism in *P. falciparum*, highlighting the role of falcipains in ensuring amino acids supply for growth and development. There are two subfamilies of FPs, FP-1, and FP-2/3. While FP-1 is not relevant to the intra-erythrocytic human stage of the parasite, FP-2/3 are essential as gene deletions of FP2/3 are lethal to *P. falciparum* (Rosenthal, 2020). FPs have been genetically characterized, with FP-2 and FP-3 sharing 68% identity, and happened to be critical for the erythrocytic stage of the parasite's life cycle in the host (Marco and Miguel Coterón, 2012). Although FP-2 is the chief haemoglobinase of *Plasmodium falciparum* (Hanspal et al., 2002), concomitant inhibition of FP-2 and FP-3 is necessary to cut-off the parasite's amino acids supply and thus becomes an effective therapeutic target against *P. falciparum* (Ettari et al., 2021).

Several inhibitors of these targets have been reported, but none have yet reached clinical trials (Pant et al., 2018; Machin et al., 2019; Hernandez Gonzalez et al., 2022). Among others, a quinolinyl oxamide derivative (QOD), *N*-(2H-1,3-benzodioxol-5-yl)-*N'*-[2-(1-methyl-1,2,3,4-tetrahydroquinolin-6-yl) ethyl]ethanediamide and an indole carboxamide derivative (ICD), *N*-{3-[(biphenyl-4-yl carbonyl) amino]propyl}-1H-indole-2-carboxamide (herein referred to as compound QOD and ICD, respectively Figure 2A) were reported as potent dual inhibitors of FP-2 and FP-3 (Rana et al., 2020). However, the molecular/structural explanation for their dual inhibitory activities has not been established. However, there is a knowledge gap in the mechanisms of the dual inhibitory nature of the QOD and ICD due to lack of structural data, which has limited SAR studies for obtaining more efficient and more potent inhibitory derivatives. Further development of these compounds into active and non-toxic drugs depends heavily on deciphering the mechanism of their interaction with their molecular targets. This information is often harnessed from Biophysical techniques, including nuclear magnetic resonance spectroscopy, X-ray crystallography, or even cryo-electron microscopy (Batool et al., 2019). Although these techniques remain the gold standards for structure-based drug design, information on time-dependent dynamic interaction is difficult—and in some cases impossible to derive from such approaches. On the contrary, molecular dynamics (MD) simulations have emerged as a versatile computational method



that offers time-resolved dynamic behavior of biomolecules and have become an established method for studying the dynamics of protein-ligand interactions (Hollingsworth and Dror, 2018). Consequently, we used such an approach to unravel the mechanism of dual inhibition of FP-2 and FP-3 by antimalarial lead compounds QOD and ICD (Rana et al., 2020). To mimic the natural ligand binding, we placed the compounds away from the proteins and allowed them to diffuse freely in the simulation box to find their preferred binding site. By subjecting our systems to microsecond-long MD simulations, we gained atomic-scale insight into the binding mechanism of the compounds QOD and ICD to both FP-2 and FP-3. In addition to providing the molecular basis for the dual inhibitory activities of the compounds to these essential FPs, our findings will accelerate the optimization of QOD and ICD towards the development of new classes of antimalarials.

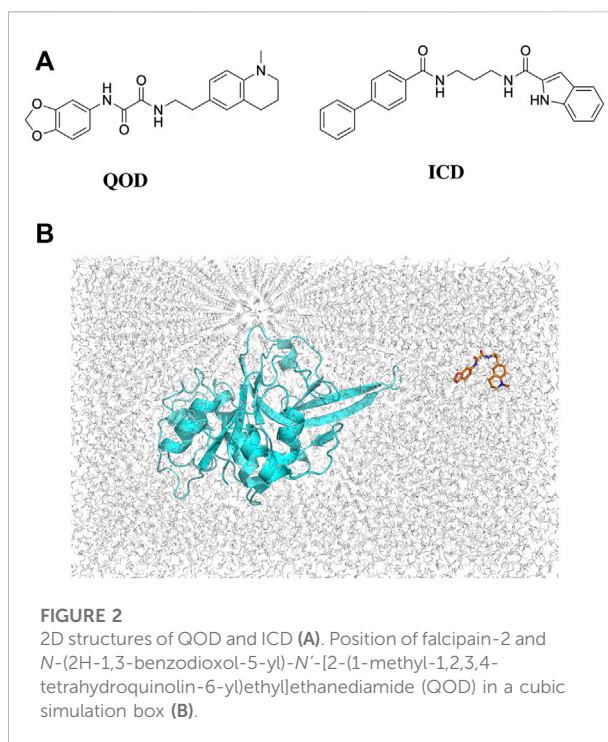
2 Methods

2.1 Structure preparation

The crystal structures of FP-2 and FP-3 were obtained from the protein databank (PDB) with the IDs 3BPF and 3BPM, respectively. Co-crystallized water molecules and other heteroatoms were removed, and Hydrogen atoms were added to the structures using the H++ server (Gordon et al., 2005). The 3D coordinates of the compounds QOD and ICD were collected from the MolPort database in SDF format, protonated, and converted to PDB format using the Schrodinger Maestro suite (Schrödinger, 2022). The cleaned-protonated proteins and ligands structures were then used to generate topology in the amber forcefield using AmberTools22 (Case et al., 2017). In each case, the ligand (i.e., QOD or ICD) is placed in a random position away from the protein (FP-2 or FP-3) in a cubic simulation box (Figure 2B) and solvated in the TIP3P water model (Price and Brooks, 2004). The systems were further neutralized with 150 mM NaCl, and ligand parameters were defined using GAFF forcefield (Wang et al., 2004).

2.2 Molecular dynamics simulations

All MD simulations were performed using GROMACS-2021.3 (Berendsen et al., 1995). Accordingly, each system was energy minimized for 5000 steps with the steepest descent algorithm followed by equilibration in the NVT ensemble for 100 ps to a temperature of 298 K using the velocity rescale temperature coupling (Bussi et al., 2007). The systems were then equilibrated in an NPT ensemble for 2 ns to a pressure of 1 atmosphere, using the Parrinello-Rahman barostat for pressure coupling (Parrinello and Rahman, 1998). This was followed by 1 μ s long final production mdrun in NPT ensemble, saving snapshots after every 10 ps. The complete



simulation input parameters are described in the [Supplementary Material](#).

2.3 MD simulation data analyses

To establish the stability of the systems, the trajectories generated from the 1 μ s simulations were used to calculate the root-mean-square deviation (RMSD) of the protein's backbone atoms. Unless told otherwise, the last 500 ns of the trajectories were used for other downstream analyses, such as the root-mean-square fluctuations (RMSF). Clustering analysis was performed to classify the different binding orientations of the ligands. The clustering was done using the GROMACS gmx cluster module, and the GROMOS method was chosen as the classification method (Daura et al., 1998). Clusters are separated by a 0.25 nm difference in RMSD. The calculation was performed from a snapshot every 10 ps.

2.4 Molecular mechanics generalized born and surface area continuum solvation (MM/GBSA) binding free energy calculation

The binding energies of the interaction between QOD, ICD, and FP-2 or FP-3 were calculated from the last 500 ns of the MD

simulation trajectories using the molecular mechanics generalized Born and surface area continuum solvation approach. In this method, the binding free energy is given by:

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

$$\Delta E_{MM} = \Delta E_{internal} + \Delta E_{electrostatic} + \Delta E_{vdw} \quad (2)$$

$$\Delta G_{sol} = \Delta G_{GB} + \Delta G_{SA} \quad (3)$$

(Hou et al., 2011)

The term ΔE_{MM} represents energy due to change in the gas phase, which is defined by internal (bond, angle, and dihedral energies, $\Delta E_{internal}$), electrostatic ($\Delta E_{electrostatic}$), and van der Waals energies (ΔE_{vdw}), while ΔG_{sol} and $T\Delta S$ account for the free energy of solvation and entropy due to binding-induced conformational changes, respectively. The solvation free energy is a term that is derived from the sum of electrostatic (ΔG_{GB} , estimated from the Generalised Born model) and non-electrostatic (ΔG_{SA}) solvation energy components, which is calculated using the solvent accessible surface area (Genheden and Ryde, 2015). The binding free energy calculation was performed using gmx_MMPBSA software, and the input parameters are described in the Supplementary Material (Valdés-Tresanco et al., 2021).

3 Results

3.1 Assessing the stability of the systems

The root-mean-square deviation (RMSD) is often used to measure global conformational changes in macromolecular structures and has become an increasingly popular method for assessing the convergence of molecular simulations. The RMSDs of the backbone atoms of our systems were calculated with reference to the energy-minimized structure, which is very close to the crystal structures. Although the systems were relatively stable even before the first 200 ns of the simulation, we observed some distortions around 500 ns simulation time, for example, in the FP-2-ICD complex (Figure 3A). This indicated that the system has not fully stabilized even at 500 ns, and therefore, those time frames should not be considered for analyses. In comparison, all systems with FP3 were remarkably quite stable throughout the simulation time, suggesting less dynamic interaction of FP3 with the ligands. Nevertheless, we deemed the first 500 ns as an extension of equilibration and only considered the last 500 ns for further structural analyses. The local protein flexibility can be followed using per-residue root-mean-square fluctuation (RMSF) and can especially be informative in describing ligand-induced flexibility. From Figure 3B, we can observe dramatic fluctuations around residues 107–120, 222–226 in the FP2-QOD system and residues 59–62, 79–83 in the FP-2-ICD complex, compared to the apo FP-2. These residues constitute part of the $\alpha 4/\beta 2$, $\beta 6-\beta 7$, $\alpha 2-\alpha 3$, $\alpha 3/$

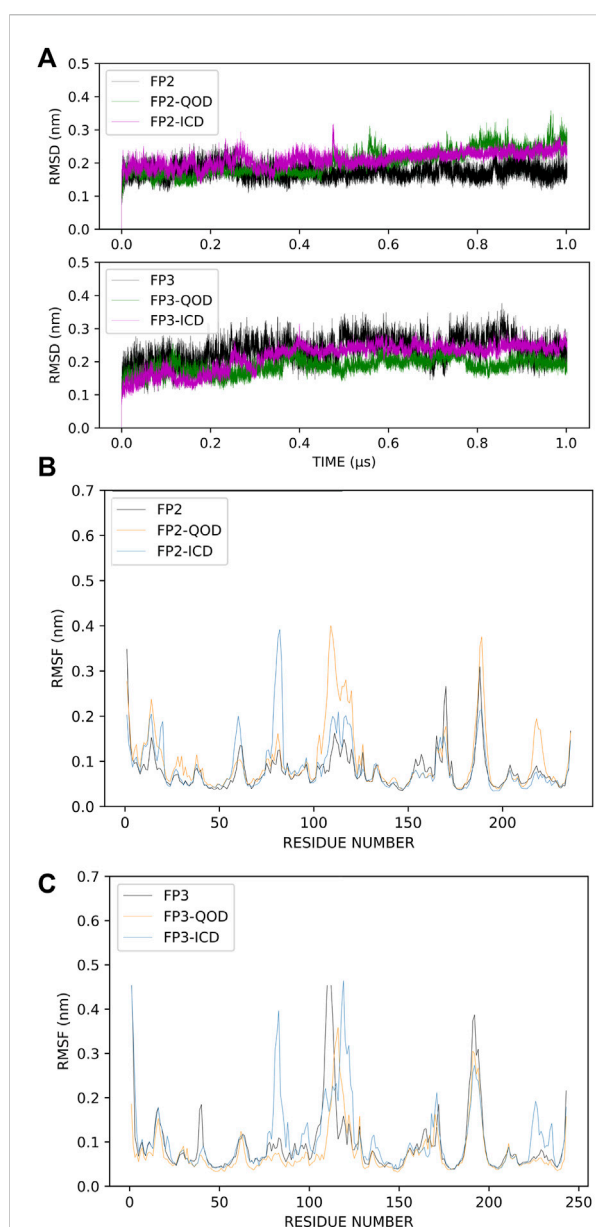
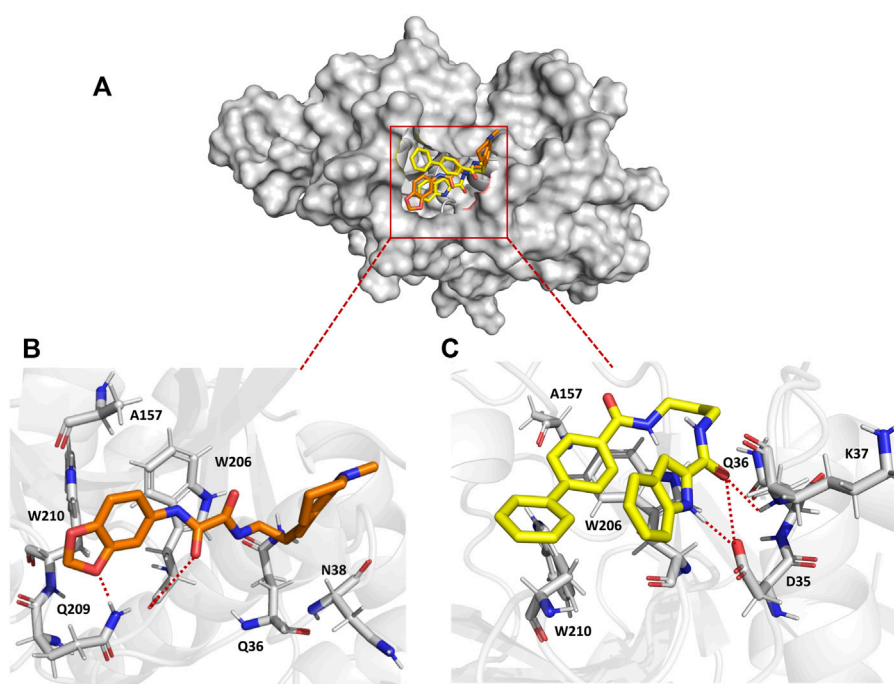


FIGURE 3

Root-mean-square deviations (RMSD) of the protein backbone atoms of the different simulation systems. Apo-FP2/FP3, FP2/FP3-QOD, FP2/FP3-ICD are represented in black, green and magenta lines, respectively. (A). Root-mean-square fluctuations (RMSF) of C- α atoms of FP-2 simulation systems. Apo-FP2, FP2-QOD, FP2-ICD are represented in black, orange and light blue lines, respectively. (B). Root-mean-square fluctuations (RMSF) of C- α atoms of FP-3 simulation systems. Apo-FP3, FP3-QOD, FP3-ICD are represented in black, orange and light blue lines, respectively. (C).

$\alpha 4$ loops (respectively) in FP-2 (Hogg et al., 2006), and therefore fluctuations are expected. Other observed regions of flexibility include residue 178 to 183 in FP2-QOD, which constitute a helical part of the protein. Therefore, the ligand, through its

**FIGURE 4**

Surface representation of QOD (orange) and ICD (yellow) bound to the active site of FP-2 (gray) (A). Interaction of QOD (B) and ICD (C) with the active site residues of FP-2. Hydrogen bonds are represented by red dotted lines.

interaction with the protein, could also contribute to the more pronounced fluctuations by influencing the global fluctuation of the system. Similarly, a higher fluctuation of the loops formed by residues 88–91, 105–130, and 232–242 (belonging to $\alpha 3/\alpha 4$ and $\beta 6$ – $\beta 7$ loops) was observed in the FP-3-ICD complex, which could be due to the same aforementioned reason (Figure 3C). On the contrary, the RMSF profile of the FP3-QOD system correlates very well with apo-FP3 RMSF, suggesting that QOD may not have a significant effect on the local flexibility of the protein.

3.2 QOD and ICD occupy the same pocket in the falcipain-2 active site

To understand the binding mechanism of the ligands to FP-2, RMSD-based clustering was applied to the generated trajectories to classify the different conformations visited during the simulation. The classification produced 278 and 188 clusters for FP2-QOD and FP-2-ICD, respectively. In each case, the largest clusters (31.13% and 41.96% of the total frames for FP2-QOD and FP2-ICD) happened to be the only cluster in which the ligand is bound to the target. Interestingly, both ligands occupy the same binding pocket at the active site of FP-2 (Figure 4A) and are coordinated by almost identical FP-2 residues (Figures 4B, C). QOD interacts

with Q36, N38, A157, W206, Q209, and W210, with hydrogen bonds formed by Q209 and W206 with one of the oxygen atoms of the dioxanyl ring and the carbonyl oxygen of the ethanediamide ligand backbone, being the major stabilizing interactions (Figure 4B). Similarly, ICD interacts with Q36, A157, W206, and W210, in addition to D35 and K37, both of which are responsible for hydrogen bonding with the formamide carbonyl oxygen of the ligand. An additional hydrogen bond is formed between the side chain carboxyl oxygen of D35 and the nitrogen atom of the indoyl ring of the ligand (Figure 4C). FP-2 is a member of the papain-like C1 cysteine proteases family that is characterized by catalytic quads of cysteine, histidine and asparagine, and glutamine (Martynov et al., 2015), represented by Q36, C42, H174 and N204 in FP-2 (Ettari et al., 2010). Our predicted interaction models suggest that access to these catalytic residues by the natural substrate is prevented by the ligands.

3.3 QOD and ICD interact with critical active site residue of FP-3

Similarly, the trajectories from the simulation of the FP3-QOD and FP3-ICD complexes were clustered to classify the

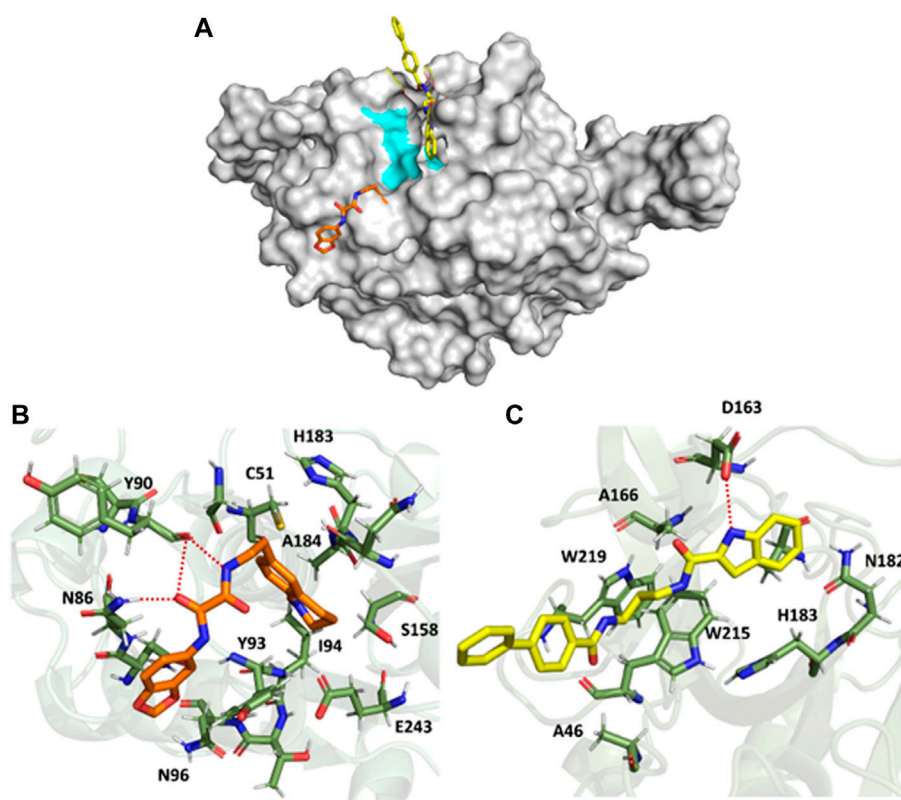


FIGURE 5

Surface representation of QOD (orange) and ICD (yellow) bound near the active site of FP-3 (gray). The critical catalytic residues are highlighted in cyan (A). FP-3 residues (green) interacting with QOD (B) and ICD (C). Hydrogen bonds are represented by red dotted lines.

TABLE 1 MM/GBSA binding energy calculations.

Ligand	Predicted binding energy (kcal/mol)	
	FP-2	FP-3
QOD	-18.59 ± 3.67	-32.81 ± 3.68
ICD	-21.36 ± 2.65	0.00 ± 2.67

NB: Binding energies are reported as mean \pm standard deviation. A detailed description of the contribution of the individual components of Eq. 1 to the total binding energies can be found in the supplementary datasheet.

different conformations sampled. Just like in FP-2, the largest clusters (67.54% and 26.62% of the total frames for FP3-QOD and FP3-ICD, respectively) happened to be the only clusters where both ligands are bound to FP-3. Unlike in FP-2, the ligands are not fully embedded in the FP-3 active site and do not assume the same binding conformation but are docked very close to the critical residues, with the indonyl ring of ICD directly blocking the catalytic N213 from the top (Figure 5A). The interaction of FP-3 with QOD is maintained by a hydrogen bond between the side chain nitrogen atom of N86 and amide-oxygen of the

ethanediamide backbone, as well as two hydrogen bonds formed between the backbone oxygen atom of Y90 and both N'-nitrogen and N-oxygen atom of the ligand ethanediamide backbone (Figure 5B). Other FP-3 residues that coordinate QOD include catalytic C51 and H183, Y93, I94, N96, S158, A184, and E243. The catalytic residues of the protein are not directly blocked in this interaction mode. However, the ligand's methyl-tetrahydroquinoliny ring is flexible and could move back and forth or even sideways to cause steric hindrance and prevent access to C51 and H183 for enzyme catalysis.

On the other hand, ICD forms only a single hydrogen bond with the side chain oxygen atom of D163 and the nitrogen atom of the ligand indonyl substituent. The amino acids A46, A166, N182, catalytic H183, W215, and W219 constituted the remaining residues interacting with ICD (Figure 5C). Notably, the ligand's biphenyl rings appeared not to contribute significantly to this interaction, given that the rings protruded outside and were almost completely excluded from the protein surface (Figures 5A, C).

3.4 The predicted binding energies are consistent with the predicted interactions, and the experimental results

The calculated binding energies of QOD and ICD against FP-2 returned comparable values (Table 1). This is not surprising, given that both ligands occupy the same binding pocket at the receptor's active site and are coordinated by almost identical residues (Figure 4). This is also consistent with the experimental IC_{50} values, which are 14.71 μ M and 12.16 μ M against FP-2, for QOD and ICD, respectively (Rana et al., 2020). The difference in the predicted binding energies of the ligands against FP-3 is surprisingly large (Table 1), considering their comparable IC_{50} values of 9.98 μ M and 8.56 μ M for QOD and ICD, respectively. However, this observation agrees, to some extent, with the predicted interactions since QOD exhibited more robust interaction with FP-3 in terms of the number of hydrogen bonds and the coordinating residues (Figure 5).

4 Discussion

Continuous efforts towards the design of new and improvement of the current antimalarial drugs are necessary to address the current demand for managing antimalarial drug resistance (Komatsuya et al., 2013; Suzuki et al., 2015) as well as meeting the WHO goal of eradicating malaria by 2030 in at least 35 countries (WHO, 2022). In line with this, we probed recently reported dual inhibitors of *P. falciparum* important cysteine proteases (FP-2 and FP-3), QOD, and ICD (Rana et al., 2020), using molecular dynamics simulations to unravel the mechanism of their inhibitions of these enzymes. We allowed the ligands to diffuse freely in the simulation box in an attempt to enable unbiased sampling of the ligands' preferred binding modes. After clustering the generated trajectories, we observed that both ligands preferred only single binding conformation, and those conformations represented the largest clusters sampled in all complexes. The ligands were unbound to the targets in all other clusters.

The predicted interaction model reveals that both compounds, QOD and ICD, docked to the same pocket in the

FP-2 active site and interacted with the active site residues in an almost identical fashion. This observation is not surprising considering that these ligands were selected from a screening using a pharmacophore model of the active site inhibitors of FP-3 (Kerr et al., 2009b; 2009a; Rana et al., 2020), that FP-2 and FP-3 share 68% identity and also the conservation of catalytic residues in the two enzymes (Pandey and Dixit, 2012). The comparable calculated binding energies of the ligands (-18.59 ± 3.67 and -21.36 ± 2.65 kcal/mol for QOD and ICD, respectively) against FP-2 corroborated with their experimental IC_{50} values (Rana et al., 2020) and also the interaction mode.

Conversely, QOD and ICD assumed different yet closely positioned docking poses very near to FP-3's catalytic residues. Compound QOD interacts with FP-3 more tightly than compound ICD, interacting with more residues, forming more hydrogen bonds, and the ligand fully embedded on the protein surface, compared to the latter, whose biphenyl substituent is wholly excluded from the interaction site and protein surface. In addition, the difference in their predicted binding energies was notably large, supporting stronger binding of QOD but incoherent with their experimental IC_{50} values (Rana et al., 2020). Notwithstanding, the QOD-FP-3 interaction model constitutes 67.54% of the total analyzed snapshots as opposed to the 26.62% for the ICD-FP-3 model, further pointing towards the more sustained and, therefore, stronger binding of QOD.

In conclusion, we have profiled QOD and ICD as active site inhibitors of FP-2 and FP-3. These compounds inhibit the activity of the FPs, most likely by preventing access to important catalytic residues in the enzymes' active sites. Inhibition of FP-2 and FP-3 is characterized by compromised amino acid metabolism in *P. falciparum* (Hanspal et al., 2002). Capitalizing on this, several ongoing research programs are currently trying to design both peptidyl and small molecule inhibitors of FPs as potential antimalarial drugs (Chakka et al., 2015; Previti et al., 2017; Himangini et al., 2018; Hernandez Gonzalez et al., 2022). However, most of these projects are still in the early stages of drug development. Therefore, we present our work as a framework for optimizing these lead compounds and hope that it will stimulate more efforts toward discovering potent antimalarial drugs. Owing to the potential globalization of vector-borne diseases due to climate change (Balogun et al., 2016), the search for new drug candidates must remain continuous.

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

AD and EB conceived the project, carried out the analyses, and wrote the manuscript.

Funding

This work was supported by an Africa Center of Excellence for Development (ACE Impacts) project through funding to the Africa Center of Excellence for Neglected Tropical Disease and Forensic Biotechnology (ACENTDFB), Ahmadu Bello University, Nigeria.

Acknowledgments

The authors thank Olukunmi Balogun (Department of Health Policy, National Center for Child Health and Development, Tokyo Japan) for critical reading of the manuscript and language corrections. We appreciate the University of Groningen Peregrine for providing us with computational resources to perform this work. EB is a recipient of the Emerging Global Leader (K43) Award and supported by the Fogarty International Center of the

National Institutes of Health under Award Number K43TW012015.

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.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmolb.2022.1070080/full#supplementary-material>

References

- Balogun, E. O., Nok, A. J., and Kita, K. (2016). Global warming and the possible globalization of vector-borne diseases: A call for increased awareness and action. *Trop. Med. Health* 44 (1), 38–43. doi:10.1186/s41182-016-0039-0
- Batool, M., Ahmad, B., and Choi, S. (2019). A structure-based drug discovery paradigm. *Int. J. Mol. Sci.* 20, 2783. doi:10.3390/IJMS20112783
- Berendsen, H. J. C., van der Spoel, D., and van Drunen, R. (1995). Gromacs: A message-passing parallel molecular dynamics implementation. *Comput. Phys. Commun.* 91, 43–56. doi:10.1016/0010-4655(95)00042-E
- Bussi, G., Donadio, D., and Parrinello, M. (2007). Canonical sampling through velocity rescaling. *J. Chem. Phys.* 126, 014101. doi:10.1063/1.2408420
- Case, D. A., Cerutti, D. S., Cheatham, T. E. I., Darden, T. A., Duke, R. E., Giese, T. J., et al. (2017). *Amber 2017 reference ManAMBER 2017*. San Francisco: University of California/University of California. Available at: <http://ambermd.org/doc12/Amber17.pdf>.
- Chakka, S. K., Kalamuddin, M., Sundararaman, S., Wei, L., Mundra, S., Mahesh, R., et al. (2015). Identification of novel class of falcipain-2 inhibitors as potential antimalarial agents. *Bioorg. Med. Chem.* 23, 2221–2240. doi:10.1016/j.bmc.2015.02.062
- Daura, X., Gademann, K., Jaun, B., Seebach, D., van Gunsteren, W. F., Mark, A. E., et al. (1998). Peptide folding: When simulation meets experiment. *Angew. Chem. Int. Ed. Engl.* 31, 236–240. doi:10.1002/(sici)1521-3773(19990115)38:1/2<236:aid-anie236>3.0.co;2-m
- Ettari, R., Bova, F., Zappalà, M., Grasso, S., and Micale, N. (2010). Falcipain-2 inhibitors. *Med. Res. Rev.* 30, 136–167. doi:10.1002/MED.20163
- Ettari, R., Previti, S., di Chio, C., and Zappalà, M. (2021). Falcipain-2 and falcipain-3 inhibitors as promising antimalarial agents. *Curr. Med. Chem.* 28, 3010–3031. doi:10.2174/0929867327666200730215316
- 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
- Gordon, J. C., Myers, J. B., Folta, T., Shoja, V., Heath, L. S., and Onufriev, A. (2005). H++: A server for estimating pKas and adding missing hydrogens to macromolecules. *Nucleic Acids Res.* 33, W368–W371. doi:10.1093/NAR/GKI464
- Hanspal, M., Dua, M., Takakuwa, Y., Chishti, A. H., and Mizuno, A. (2002). Plasmodium falciparum cysteine protease falcipain-2 cleaves erythrocyte membrane skeletal proteins at late stages of parasite development. *Blood* 100, 1048–1054. doi:10.1182/BLOOD-2002-01-0101
- Hernandez Gonzalez, J. E., Alberca, L. N., Masforrol González, Y., Reyes Acosta, O., Talevi, A., and Salas-Sarduy, E. (2022). Tetracycline derivatives inhibit plasmodial cysteine protease falcipain-2 through binding to a distal allosteric site. *J. Chem. Inf. Model* 62, 159–175. doi:10.1021/acs.jcim.1c01189
- Himangini, D. P., Sharma, V., and Kumar, S. (2018). Designing novel inhibitors against falcipain-2 of Plasmodium falciparum. *Bioorg. Med. Chem. Lett.* 28, 1566–1569. doi:10.1016/j.bmcl.2018.03.058
- Hogg, T., Nagarajan, K., Herzberg, S., Chen, L., Shen, X., Jiang, H., et al. (2006). Structural and functional characterization of Falcipain-2, a hemoglobinase from the malarial parasite Plasmodium falciparum. *J. Biol. Chem.* 281, 25425–25437. doi:10.1074/JBC.M603776200
- Hollingsworth, S. A., and Dror, R. O. (2018). Molecular dynamics simulation for all. *Neuron* 99, 1129–1143. doi:10.1016/j.neuron.2018.08.011
- 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
- Kerr, I. D., Lee, J. H., Farady, C. J., Marion, R., Rickert, M., Sajid, M., et al. (2009a). Vinyl sulfones as antiparasitic agents and a structural basis for drug design. *J. Biol. Chem.* 284, 25697–25703. doi:10.1074/JBC.M109.014340
- Kerr, I. D., Lee, J. H., Pandey, K. C., Harrison, A., Sajid, M., Rosenthal, P. J., et al. (2009b). Structures of falcipain-2 and falcipain-3 bound to small molecule inhibitors: Implications for substrate specificity. *J. Med. Chem.* 52, 852–857. doi:10.1021/JM8013663
- Komatsuya, K., Hata, M., Balogun, E. O., Hikosaka, K., Suzuki, S., Takahashi, K., et al. (2013). Synergy of ferrous ion on 5-aminolevulinic acid-mediated growth inhibition of Plasmodium falciparum. *J. Biochem.* 154 (6), 501–504. doi:10.1093/jb/mvt096

- Machin, J. M., Kantsadi, A. L., and Vakonakis, I. (2019). The complex of *Plasmodium falciparum* falcipain-2 protease with an (E)-chalcone-based inhibitor highlights a novel, small, molecule-binding site. *Malar. J.* 18, 388. doi:10.1186/S12936-019-3043-0
- Marco, M., and Miguel Coterón, J. (2012). Falcipain inhibition as a promising antimalarial target. *Curr. Top. Med. Chem.* 12, 408–444. doi:10.2174/156802612799362913
- Martynov, A. G., Elpidina, E. N., Perkin, L., and Oppert, B. (2015). Functional analysis of C1 family cysteine peptidases in the larval gut of *Tenebrio molitor* and *Tribolium castaneum*. *BMC Genomics* 16, 1–16. doi:10.1186/S12864-015-1306-X/TABLES/5
- McKerrow, J. H., Engel, J. C., and Caffrey, C. R. (1999). Cysteine protease inhibitors as chemotherapy for parasitic infections. *Bioorg Med. Chem.* 7, 639–644. doi:10.1016/S0968-0896(99)00008-5
- Pandey, K. C., and Dixit, R. (2012). Structure-function of falcipains: Malarial cysteine proteases. *J. Trop. Med.* 2012, 345195. doi:10.1155/2012/345195
- Pant, A., Kumar, R., Wani, N. A., Verma, S., Sharma, R., Pande, V., et al. (2018). Allosteric site inhibitor disrupting auto-processing of malarial cysteine proteases. *Sci. Rep.* 8 (1 8), 16193–16215. doi:10.1038/s41598-018-34564-8
- Parrinello, M., and Rahman, A. (1998). Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* 52, 7182–7190. doi:10.1063/1.328693
- Previti, S., Ettari, R., Cosconati, S., Amendola, G., Chouchene, K., Wagner, A., et al. (2017). Development of novel peptide-based michael acceptors targeting rhodesain and falcipain-2 for the treatment of neglected tropical diseases (NTDs). *J. Med. Chem.* 60, 6911–6923. doi:10.1021/acs.jmedchem.7b00405
- Price, D. J., and Brooks, C. L. (2004). A modified TIP3P water potential for simulation with Ewald summation. *J. Chem. Phys.* 121, 10096–10103. doi:10.1063/1.1808117
- Rana, D., Kalamuddin, M., Sundriyal, S., Jaiswal, V., Sharma, G., das Sarma, K., et al. (2020). Identification of antimalarial leads with dual falcipain-2 and falcipain-3 inhibitory activity. *Bioorg Med. Chem.* 28, 115155. doi:10.1016/J.BMC.2019.115155
- Rosenthal, P. J. (2020). Falcipain cysteine proteases of malaria parasites: An update. *Biochimica Biophysica Acta (BBA) - Proteins Proteomics* 1868, 140362. doi:10.1016/J.BBAPAP.2020.140362
- Schrödinger (2022). *Schrödinger release 2022-1*. New York, NY: Maestro, Schrödinger, LLC.
- Siqueira-Neto, J. L., Debnath, A., McCall, L. I., Bernatchez, J. A., Ndao, M., Reed, S. L., et al. (2018). Cysteine proteases in protozoan parasites. *PLoS Negl. Trop. Dis.* 12, e0006512. doi:10.1371/JOURNAL.PNTD.0006512
- Snow, R. W. (2015). Global malaria eradication and the importance of *Plasmodium falciparum* epidemiology in Africa. *BMC Med.* 13, 23–3. doi:10.1186/s12916-014-0254-7
- Suzuki, S., Hikosaka, K., Balogun, E. O., Komatsuya, K., Niikura, M., Kobayashi, F., et al. (2015). *In vivo* curative and protective potential of orally administered 5-aminolevulinic acid plus ferrous ion against malaria. *Antimicrob. agents Chemother.* 59 (11), 6960–6967. doi:10.1128/AAC.01910-15
- Valdés-Tresanco, M. S., Valdés-Tresanco, M. E., Valiente, P. A., and Moreno, E. (2021). Gmx_MMPBSA: A new tool to perform end-state free energy calculations with GROMACS. *J. Chem. Theory Comput.* 17, 6281–6291. doi:10.1021/acs.jctc.1c00645
- Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A., and Case, D. A. (2004). Development and testing of a general amber force field. *J. Comput. Chem.* 25, 1157–1174. doi:10.1002/JCC.20035
- WHO (2022). World health organization. Available at: <https://www.who.int/news-room/fact-sheets/detail/malaria> (Accessed September 23, 2022).



OPEN ACCESS

EDITED BY
Chandrabose Selvaraj,
Saveetha University, India

REVIEWED BY
Teodorico Castro Ramalho,
Universidade Federal de Lavras, Brazil
Bhuvaneshwari Sampath,
Anna University, India

*CORRESPONDENCE
Danislav S. Spassov,
✉ dspassov@pharmfac.mu-sofia.bg

SPECIALTY SECTION
This article was submitted to Biophysics,
a section of the journal
Frontiers in Molecular Biosciences

RECEIVED 10 October 2022
ACCEPTED 16 December 2022
PUBLISHED 10 January 2023

CITATION
Spassov DS, Atanasova M and
Doytchinova I (2023), A role of salt
bridges in mediating drug potency: A
lesson from the N-
myristoyltransferase inhibitors.
Front. Mol. Biosci. 9:1066029.
doi: 10.3389/fmolb.2022.1066029

COPYRIGHT
© 2023 Spassov, Atanasova and
Doytchinova. 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 role of salt bridges in mediating drug potency: A lesson from the N-myristoyltransferase inhibitors

Danislav S. Spassov*, Mariyana Atanasova and Irini Doytchinova

Department of Chemistry, Faculty of Pharmacy, Medical University of Sofia, Sofia, Bulgaria

The salt bridge is the strongest non-covalent interaction in nature and is known to participate in protein folding, protein-protein interactions, and molecular recognition. However, the role of salt bridges in the context of drug design has remained not well understood. Here, we report that a common feature in the mechanism of inhibition of the N-myristoyltransferases (NMT), promising targets for the treatment of protozoan infections and cancer, is the formation of a salt bridge between a positively charged chemical group of the small molecule and the negatively charged C-terminus of the enzyme. Substituting the inhibitor positively charged amine group with a neutral methylene group prevents the formation of the salt bridge and leads to a dramatic activity loss. Molecular dynamics simulations have revealed that salt bridges stabilize the NMT-ligand complexes by functioning as molecular clips that stabilize the conformation of the protein structure. As such, the creation of salt bridges between the ligands and their protein targets may find an application as a valuable tool in rational drug design.

KEYWORDS

salt bridge, ligand-protein complex, drug-protein interactions, NMT, protein conformation and drug potency, N-myristoyltransferase, conformational stabilization and inhibition, salt bridge and protein conformation

1 Introduction

The salt bridge is a non-covalent interaction that combines an electrostatic attraction between oppositely charged chemical groups or atoms and a hydrogen bond; hence, its strength exceeds the strength of a simple hydrogen bond (Donald et al., 2011; Ferreira de Freitas and Schapira, 2017). In proteins, the salt bridges occur most frequently between the positively charged basic amino acid residues Lys or Arg and the negatively charged acidic Asp or Glu residues (Kumar and Nussinov, 2002; Bosshard et al., 2004; Donald et al., 2011; Basu and Mukharjee, 2017; Ferreira de Freitas and Schapira, 2017). In this context, salt bridges are known to participate in protein-protein interaction, protein folding, protein recognition, protein conformational rigidity, and protein stability (Takano et al., 2000; Kumar and Nussinov, 2002; Bosshard et al., 2004; Basu and Mukharjee, 2017; Ferreira de Freitas and Schapira, 2017). However, much less is known about salt bridges and their significance in mediating the interaction between small molecule ligands and their protein targets. A substantial number of drug molecules

contain either positively or negatively charged chemical groups and, as such, might be capable of participating in salt bridges within their binding sites (Uddin et al., 2021). Indeed, systematic surveys of the protein-ligand complexes deposited in the protein databank have identified over a thousand unique small molecule ligands that form salt bridges with their protein targets (Ferreira de Freitas and Schapira, 2017; Kurczab et al., 2018). Such prevalence raises questions about the significance of salt bridges in drug-protein interactions and their role in mediating inhibitor potency.

Hitherto, little is known about the role of these salt bridges in protein-ligand complexes. In several cases, such as for certain inhibitors of the Epidermal growth factor receptor (EGFR) (Peng et al., 2013) and G-protein coupled receptors ligands (Ferreira de Freitas and Schapira, 2017), it has been shown that the salt bridge plays a crucial role in the ligand's activity. However, in the prevailing number of cases, the salt bridges, their role in protein-ligand interactions, their physical and chemical properties, and the mechanisms that mediate their effects on the potency of the compounds have remained largely unexplored. The incomplete understanding of the role of salt bridges in protein-ligand interactions represents an obstacle to drug design, as the researchers are unaware of their significance and how to use them to boost the inhibitors' potency.

This study began with the observation that two structurally unrelated inhibitors of N-myristoyltransferases (NMT)—IMP-1088 and DDD85646 that were developed independently of each other by high-throughput screening and fragment-based approaches form a salt bridge within their binding sites. IMP-1088 and DDD85646 are very potent NMT inhibitors with reported IC_{50} in the picomolar and low nanomolar range, respectively (Frearson et al., 2010; Mousnier et al., 2018). In preclinical models, the NMT inhibitors were found to be very effective for treating parasitic protozoan infections such as the African sleeping sickness and are promising therapeutics for the treatment of other protozoan diseases such as malaria and leishmaniasis (Frearson et al., 2010; Ritzefeld et al., 2018; Schlott et al., 2018). In addition, NMT inhibitors display potent anti-tumor activity and have produced complete anti-tumor responses in preclinical murine models (Thinon et al., 2016; Beauchamp et al., 2020). NMTs are enzymes that catalyze the myristoylation of selected cellular proteins, which contain a specific peptide sequence known as a myristoylation signal (Thinon et al., 2014). This sequence is located in the N-terminal region of the proteins and binds into a specially evolved pocket in the NMT active site, known as the peptide binding pocket (Thinon et al., 2014; Dian et al., 2020). NMT also uses a cofactor—Myristoyl-CoA (Myr-CoA), which binds to a site adjacent to the peptide binding pocket (Thinon et al., 2014; Dian et al., 2020). During myristoylation, the myristic acid is transferred from Myr-CoA to the N-terminus of the myristoylated proteins

(Thinon et al., 2014; Dian et al., 2020). The myristic acid, due to its hydrophobicity, is used for the attachment of the modified proteins to the cellular membranes but also facilitates the dimerization of interacting partners and regulates crucial cell signaling events (Patwardhan and Resh, 2010; Gaffarogullari et al., 2011; Spassov et al., 2018; Meinnel et al., 2020).

NMTs are an ancient enzyme family represented in human by two members—NMT1 and NMT2, which share a highly conserved catalytic domain (Ducker et al., 2005; Meinnel et al., 2020). IMP-1088 and DDD85646 are not selective against the two forms and inhibit them with almost identical potency (Frearson et al., 2010; Mousnier et al., 2018).

DDD85646 and IMP-1088 share a common mode of interaction with NMT (Figure 1). Both compounds contain a pyrazole ring that participates in a hydrogen bond with Ser405 in the catalytic center of NMT. In addition, the crystallographic structures of NMT complexes with these inhibitors have revealed an interaction between the C-terminal carboxyl group of NMT and the piperazine ring of DDD85646 or the dimethylamino group of IMP-1088, both of which contain a terminal basic nitrogen atom (Figure 1). Evidence indicates that in the DDD85646 series, the terminal basic nitrogen is crucial for activity. Specifically, this can be seen in examples of compounds where the nitrogen atom is substituted for a carbon atom (Figure 2). For example, a comparison between DDD86213 and DDD87749, two compounds intermediary in the development of DDD85646, reveals that substituting the nitrogen atom of the piperazine ring reduces the inhibitory potency of the ligand by 1,328-fold (from IC_{50} 7 nM to IC_{50} 9.3 μ M) (Figure 2) (Brand et al., 2017). In the case of the IMP-1088 series, intermediary compounds at which the nitrogen atom of the dimethylamino group is replaced with carbon have not been tested experimentally because the fragment that contains the dimethylamino group was identified by high throughput screening and was present in the initial hit of the series—IMP-72 (Mousnier et al., 2018). Irrespective of its crucial role in the activity of the NMT inhibitors, the interaction with the C-terminus of NMT has not been studied in detail, and the many aspects of this interaction have not been described previously. Here we report for the first time that the NMT inhibitors are positively charged in the physiological environment due to the protonation of the critical basic nitrogen atom and that the positively charged chemical groups of the inhibitors participate in a salt bridge with the negatively charged carboxyl group at the C-terminus of the NMT protein. Using Molecular dynamics (MD) simulations of protonated and non-protonated forms of the inhibitors, we demonstrate that the salt bridge has an unexpected role in stabilizing the NMT protein conformation and that this may be a significant factor in mediating its effects on NMT inhibitors' potency.

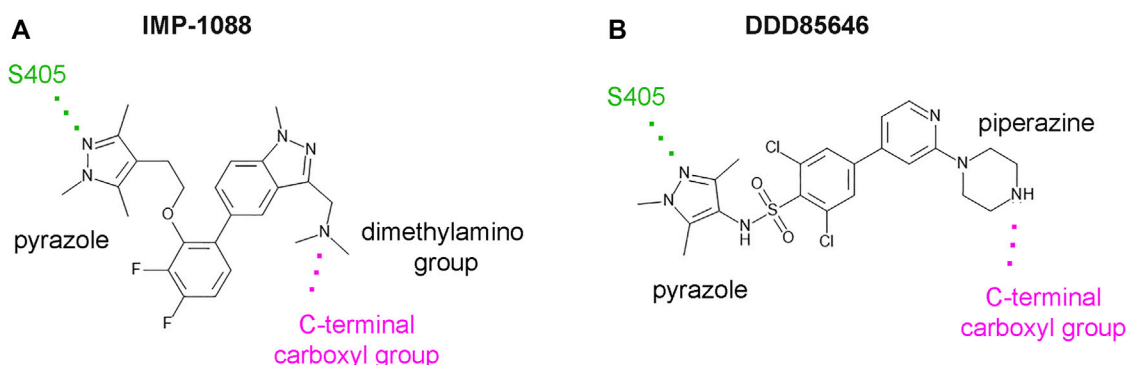


FIGURE 1

IMP-1088 and DDD85646 share a common mode of interaction in the active site of NMT. **(A)**. The main polar interactions between IMP-1088 and *Homo sapiens* NMT1 based on the crystallographic structure PDB 5MU6. **(B)**. The main polar interactions between DDD85646 and *Homo sapiens* NMT1 based on the crystal structure PDB 3IWE. Both inhibitors have a pyrazole ring that forms a hydrogen bond with S405 (shown in green). IMP-1088 and DDD85646 also interact with the C-terminal carboxyl group of NMT protein (shown in magenta) through a dimethylamino group and a piperazine ring, respectively.

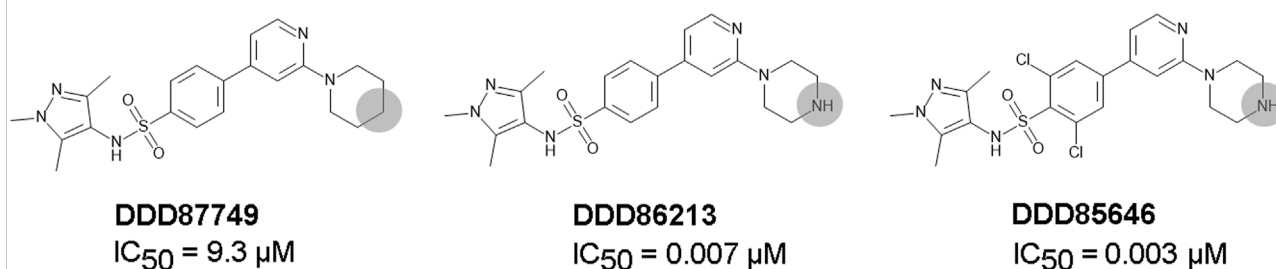


FIGURE 2

The role of the nitrogen atom of the piperazine ring in the potency of the DDD85646 series of compounds. IC_{50} towards *Homo sapiens* NMT1 is shown. Substitution of the indicated nitrogen atom with a carbon reduces potency by 1,328-fold.

2 Materials and methods

2.1 Visualization of protein-ligand interactions and image preparation

Protein-ligand interactions were visualized in PyMOL 1.6.0.0 (Schrödinger, New York, United States) (Schrödinger and DeLano, 2020) and YASARA v. 20.4.24 (IMBM, University of Graz, Austria) (Krieger and Vriend, 2014). The protonation state of the NMT inhibitors inside their complexes with N-myristoyltransferases was determined by YASARA and Protoss (University of Hamburg, Germany) (Bietz et al., 2014). pKa determinations were performed using Epic software from the Maestro package release 2018-4 (Schrödinger, New York, United States) and rely on extensions to the well-established Hammett and Taft approaches for pKa prediction, namely, mesomer standardization, charge cancellation, and charge

spreading to make the predicted results reflect the nature of the molecule itself rather than just for the particular chemical group (Shelley et al., 2007; Schrödinger and DeLano, 2020). The partial charges of the molecules were determined by using the Maestro package release 2018-4 (Schrödinger, New York, United States). Crystal structure images and superimpositions were prepared in PyMOL 1.6.0.0 (Schrödinger, New York, United States). The distances and angles of the salt bridge were determined in PyMOL based on the crystal structures of IMP-1088 and DDD85646 complexes with human NMT1 (PDB 5MU6 and PDB 3IWE, respectively). The 2D structures of the NMT inhibitors were created in MedChem Designer v.5.5 (SimulationsPlus, Lancaster, CA, United States). The nomenclature of the secondary structural elements in NMT, such as the names and the position of the α -helices, β -sheets, or connecting loops, was adopted from Dian et al. (Dian et al., 2020).

2.2 Molecular dynamics simulations

The protonated and unprotonated forms of IMP-1088 and DDD8646 were prepared using the crystal structures of their complexes with *Homo sapiens* NMT1 and Myr-CoA (PDB 5MU6 and 3IWE, respectively). The hydrogen atoms were added to the inhibitors' structures in YASARA. YASARA automatically assigns the protonation and charge of the ligands at pH 7.4 to generate the protonated charged forms of the NMT inhibitors. The unprotonated uncharged forms of the NMT inhibitors were generated after the removal of the hydrogen atom, participating in the salt bridge. The structure of DDD86213 was prepared by replacing the chlorine atoms from DDD85646 with hydrogen atoms and of DDD87749 by replacing the nitrogen donor atom in DDD86213 with a carbon atom using PyMOL. The NMT inhibitor structures were saved in mol2 format and prepared for MD simulations using the Antechamber program from the AMBER v. 18 package (UCSF, San Francisco, United States) (Case et al., 2005; Case et al., 2018) by assigning a net charge of 0 for the non-protonated forms and net charge +1, for the protonated forms. Myr-CoA was similarly prepared for MD simulations using the Antechamber program from the AMBER v. 18 package. The NMT protein used for MD simulations contained an N-terminal ACE cap and no C-terminal cap. The addition of a C-terminal cap was not possible because it could have interfered with the interaction with the ligand's positively charged group and was not necessary because the C-terminus of the native NMT protein contains a negatively charged carboxyl group at this position. The crystal structures of NMT-IMP-1088/DDD85646 complexes do not contain the full-length NMT protein and have truncated N-terminal parts (the first 114 amino acids). Thus the addition of the N-terminal ACE cap is necessary to eliminate the positively charged N-terminus at position 115 that is not normally present in the full-length protein. The ternary complexes, consisting of NMT protein, Myr-CoA, and either the protonated or the unprotonated forms of IMP-1088, DDD85646, DDD86231, or DDD87749 were solvated in saline (0.9% sodium chloride) in a truncated octahedral box, containing 15,891 water molecules (distance between protein and the edge of the box varied between 23.7–39.7 Å), energy minimized, heated to 310 K at constant volume for 1 ns, density equilibrated at 1 bar for 1 ns, equilibrated keeping constant T and p for 1 ns, using the Langevin thermostat (Adelman and Doll, 1974) and Berendsen barostat (Berendsen et al., 1984) and simulated for 1,000 ns by AMBER v. 18 (UCSF, San Francisco, United States) (Case et al., 2005; Case et al., 2018). During all steps of simulations, i.e., heating, density equilibration, preproduction, and production dynamic SHAKE algorithm (Ciccotti and Ryckaert, 1986) was used for constraining covalent bonds involving hydrogen with a 2 fs time step. The bonds to hydrogen were not constrained only during the energy minimization step. The systems were simulated with the ff14SB force field (Maier et al., 2015) under periodic boundary

conditions. Frames were saved every 1 ns to generate 1,000 frames for a total of 1 μ s duration of MD production simulations.

2.3 Analysis of results from MD simulations

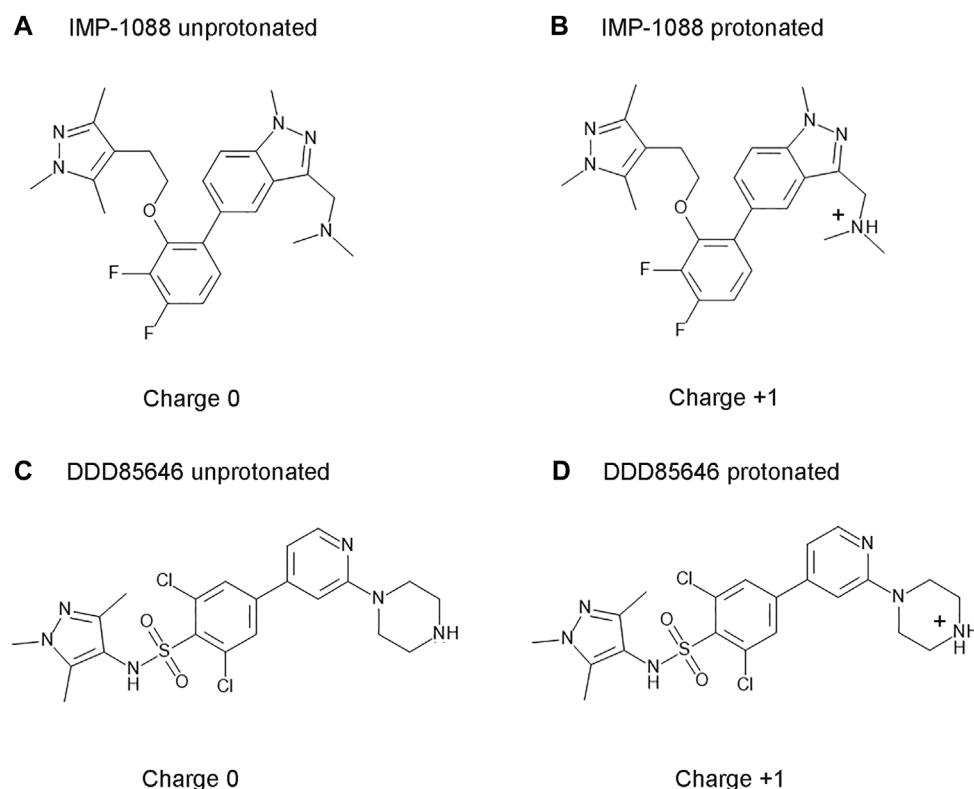
The data from the MD simulations were analyzed in VMD (Visual Molecular Dynamics, the University of Illinois at Urbana–Champaign, United States) (Humphrey et al., 1996). For determining the distance between the nitrogen atom of the ligand and the oxygen atom at the C-terminus of the NMT protein, the atoms were initially selected through the Graphic and Labels option in VMD, and the exported graphical data was used to generate the charts in GraphPad Prism.

RMSD (root-mean-square deviations) of the inhibitors, NMT protein, and Myr-CoA were determined by the RMSD trajectory tool in VMD. The NMT protein corresponds to resid 1 to 382 (includes all 382 amino acid residues present in the crystal structure PDB 3IWE), the A' α -helix- A'-loop region to resid 1 to 19, the NMT inhibitors to resid 383, and the Myr-CoA to resid 384. In all cases, the RMSD values were determined for the heavy atoms, e.g., excluding the hydrogen atoms. The exported RMSD graphical data was used to generate the graphs in GraphPad Prism.

3 Results

3.1 DDD85646 and IMP-1088 are positively charged at pH 7.4

IMP-1088 contains a dimethylamino group, and DDD85646 has a piperazine ring that can become protonated. As such, the compounds can exist in two forms—an unprotonated uncharged and a protonated charged form (Figure 3). The targets of these inhibitors - NMT1 and NMT2, are intracellular cytoplasmic proteins that exist in an environment where the pH has been estimated to be 7.0–7.4 (Flinck et al., 2018; Persi et al., 2018). The pKa values for trimethylamine and piperazine, determined experimentally, are 9.8 and 9.73, respectively (Khalili et al., 2009; Settimo et al., 2014), indicating that these basic groups exist almost entirely in a cation form in the physiological environment. In the context of the whole structure of the inhibitors, the predicted pKa for the dimethylamino group of IMP-1088 and the piperazine of DDD85646 by using the Epic package of Maestro (Schrodinger) was 8.39 and 8.78, respectively. Hence, in the physiological environment, most IMP-1088 and DDD85646 molecules are expected to be charged due to the protonation of their terminal nitrogen atom (Figure 3). The protonation of IMP-1088 and DDD85646 was also confirmed in YASARA (Krieger and Vriend, 2014) and Protoss (Bietz et al., 2014), and they both indicated the presence of the protonated

**FIGURE 3**

Structures and protonation forms of the NMT inhibitors IMP-1088 and DDD85646. **(A)** Unprotonated IMP-1088. **(B)** Protonated IMP-1088. **(C)** Unprotonated DDD85646. **(D)** Protonated DDD85646. IMP-1088 contains a dimethylamino group, and DDD85646 has a piperazine ring, and their protonation gives the molecules a positive +1 charge. At pH 7.4, IMP-1088 and DDD85646 are predicted to be in their protonated, positively charged forms.

forms of the inhibitors inside the binding sites of their complexes with NMT. DDD86213 has a predicted pKa value of 8.78, identical to DDD85646. In contrast, DDD87749 is not protonated and not charged due to the substitution of the terminal nitrogen atom with a carbon - the pKa for this compound lies outside the normal range of sampled pH range in Maestro. Therefore, DDD87749 models the interaction between the non-protonated inhibitors and NMT under physiological pH conditions.

3.2 IMP-1088 and DDD85646 form a salt bridge with the C-terminus of the NMT protein

N-myristoyltransferases are enzymes that contain two binding pockets—one for binding the cofactor Myr-CoA and the other for the substrate peptide (Dian et al., 2020). The crystallographic structures of NMT in complex with the NMT inhibitors IMP-1088 (PDB 5MU6) and DDD85646 (PDB 3IWE) reveal that the NMT inhibitors occupy the

peptide binding pocket of NMT, and the cofactor Myr-CoA binds to a distinct region proximal to the inhibitors' binding site (Figures 4A,B) (Frearson et al., 2010; Mousnier et al., 2018). In this binding mode, the NMT inhibitors and Myr-CoA occupy nearby sites but do not interact directly with each other (Frearson et al., 2010; Mousnier et al., 2018). To illuminate the role of the positively charged group of the NMT inhibitors, we analyzed the crystallographic structures of complexes of IMP-1088 and DDD85646 with human NMT1. In these structures, it is evident that the positively charged chemical group of the inhibitor forms a salt bridge with the negatively charged C-terminus of the NMT protein, which is located in the active site of the enzyme (Figures 4C,D) (Dian et al., 2020). The salt bridge's formation involves the free carboxyl group of the C-terminal amino acid residue - Gln496, and not its side chain, which is oriented in the opposite direction (Figures 4C,D). The average pKa value of the carboxyl group at the C-terminus of folded proteins is estimated to be 3.3 (Grimsley et al., 2009), indicating that in the cytoplasm of the cell (pH 7.0–7.4), the C-terminus is expected to be

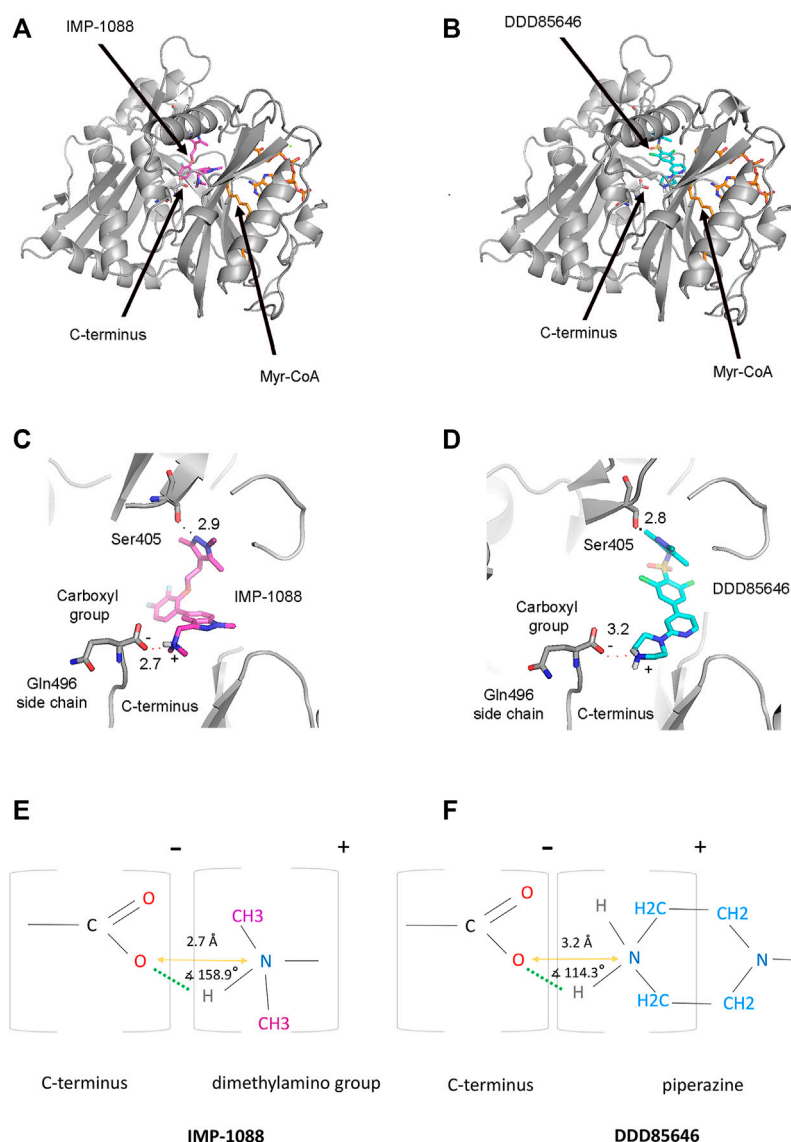


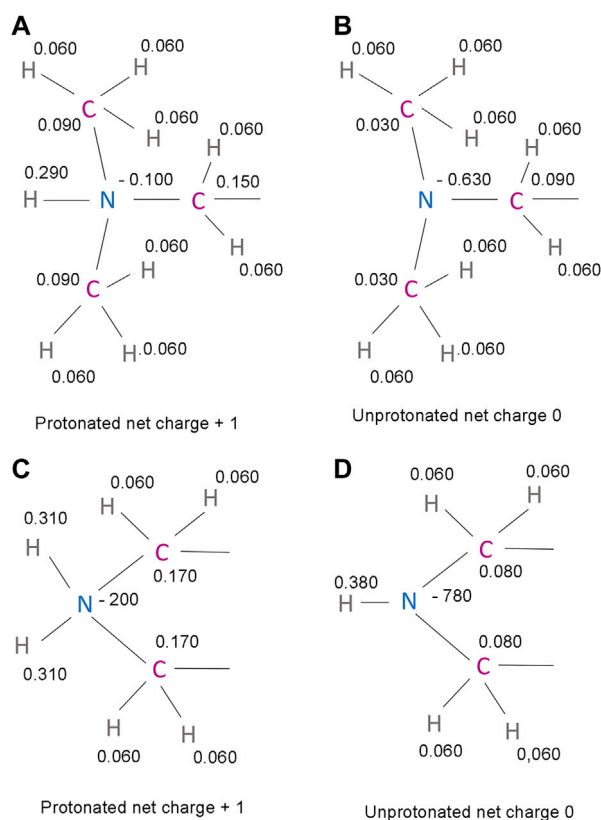
FIGURE 4

Formation of a salt bridge between the NMT inhibitors and the C-terminus of the NMT protein. **(A)** The crystal structure of *Homo sapiens* NMT1 in complex with IMP-1088 and Myr-CoA (PDB 5MU6). IMP-1088, shown in magenta, occupies the peptide-binding pocket of NMT and is located nearby the cofactor Myr-CoA, shown in orange. **(B)** The crystal structure of *Homo sapiens* NMT1 in complex with DDD85646 and Myr-CoA (PDB 3IWE). DDD85646 is shown in cyan and Myr-CoA in orange. **(C)** Formation of a salt bridge between IMP-1088 and the C-terminus of NMT in PDB 5MU6. The hydrogen bond with Ser405 is also indicated. **(D)** Formation of a salt bridge between DDD85646 and the C-terminus of NMT in PDB 3IWE. **(E)** A schematic representation of the salt bridge between the dimethylamino group of IMP-1088 and the C-terminal carboxyl group of NMT. **(F)** A schematic representation of the salt bridge between the amine group of the piperazine ring of DDD85646 and the carboxyl group at the C-terminus of NMT. The charge of the groups is indicated above the brackets. The length of the salt bridge, defined as the distance between the oxygen atom at the C-terminus of NMT and the nitrogen atom of the ligand (orange arrow), is shown in angstroms (Å). The angle between the ligand nitrogen atom, hydrogen atom, and the C-terminal carboxyl group oxygen atom is also indicated. Green dots indicate the hydrogen bond.

deprotonated, negatively charged, and therefore available for the formation of the salt bridge.

Generally, a salt bridge combines two non-covalent interactions - an electrostatic attraction between chemical groups of opposite charges and a hydrogen bond (Donald et al., 2011). Both of these interactions can be identified in

the complexes of NMT inhibitors. The electrostatic attraction involves the positively charged dimethylamino group of IMP-1088 or the amine group in the piperazine ring of DDD85646 and the negatively charged C-terminal carboxyl group of NMT (Figures 4E, F). A hydrogen bond between the protonated inhibitor's hydrogen atom and the

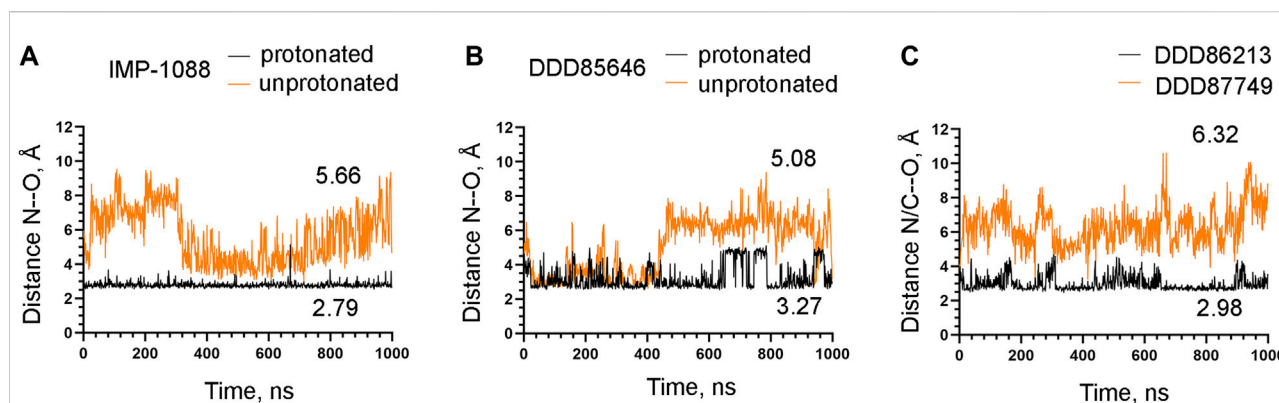
**FIGURE 5**

Predicted partial charges in the protonated and unprotonated forms of IMP-1088 and DDD85646 (A) The protonated dimethylamino group of IMP-1088. (B) The unprotonated dimethylamino group of IMP-1088. (C) The protonated amine group of DDD85646. A part of the piperazine ring is shown (D) The unprotonated amine group of DDD85646. A part of the piperazine ring is shown. Partial charges were determined using the Maestro software package release 2018-4 (Schrödinger, New York, United States of America) (Shelley et al., 2007; Schrödinger and DeLano, 2020).

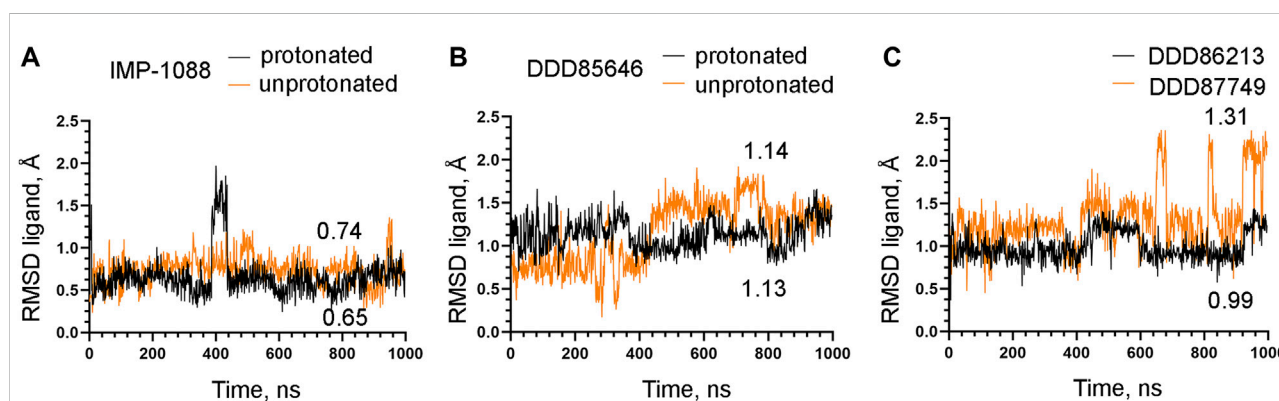
carboxyl group's oxygen atom at the C-terminus of NMT is also formed, fulfilling the second requirement for forming a salt bridge (Figures 4E,F). The length of the salt bridge in the IMP-1088 and DDD85646 complexes is 2.7 and 3.2 Å, respectively. (Figures 4E,F). Since the permissible distance for a salt bridge length is <4 Å, the measured distances are consistent with salt bridge formation between the inhibitors and NMT (Donald et al., 2011). Generally, the optimum angle between the nitrogen, hydrogen and oxygen atoms in the salt bridge is between 110° and 180° (Kurczab et al., 2018). The measured angles in the salt bridges of IMP-1088 and DDD85646 of 158.9° and 114.3° (Figures 4E,F) are in the range of values determined for different salt bridges in the crystallographic structures deposited in the protein databank (Kurczab et al., 2018). IMP-1088 may form a stronger salt bridge than DDD85646, considering that the length and the angle of the salt bridge in the complexes of this compound are more optimal for binding (Figures 4E,F). The distribution of partial charges

of the protonated and unprotonated dimethylamino group and piperazine ring of the NMT inhibitors were determined by using the Maestro software (Schrödinger, New York, United States) (Shelley et al., 2007; Schrödinger and DeLano, 2020) (Figure 5). The nitrogen atom of these groups is electronegative, and the positive charge of the protonated form is distributed among the nearby hydrogen and carbon atoms (Figure 5). The highest partial positive charge is observed on the hydrogen atom participating in the salt bridge formation (Figure 5). The distribution of partial charges is shown in the absence of interactions, and the formation of a salt bridge between the inhibitors and the electronegative C-terminal carboxyl group may affect the described distribution.

The interaction between the inhibitors and NMT also involves forming a hydrogen bond with Ser405 (Figure 1), multiple π - π stacking, and hydrophobic interactions with several aromatic residues, including Tyr296, His298, Phe188, and Phe311.

**FIGURE 6**

Stability of the interaction between the NMT inhibitors and the C-terminus of the NMT protein. (A) The distance between the ligand nitrogen atom and the C-terminal oxygen atom of NMT (N—O distance) in angstroms (Å) during the time course of MD simulations in nanoseconds (ns) of NMT complexes with protonated and unprotonated forms of IMP-1088. (B) The distance between the ligand nitrogen atom and the C-terminal oxygen atom of NMT (N—O distance) in angstroms (Å) vs. time of MD simulations in nanoseconds (ns) of NMT complexes with protonated and unprotonated forms of DDD85646. (C) The distance between the ligand nitrogen in DDD86213 or the corresponding carbon atom in DDD87749 and the C-terminal oxygen atom of NMT (N/C—O distance) in angstroms (Å) vs. time of MD simulations in nanoseconds (ns) of NMT complexes with DDD87749 and protonated DDD86213. The numbers in the plots indicate the average distance in angstroms (Å). Length < 4 Å indicates that the salt bridge is stable; distance > 4 Å demonstrates that the interaction between the ligand and the protein is disrupted.

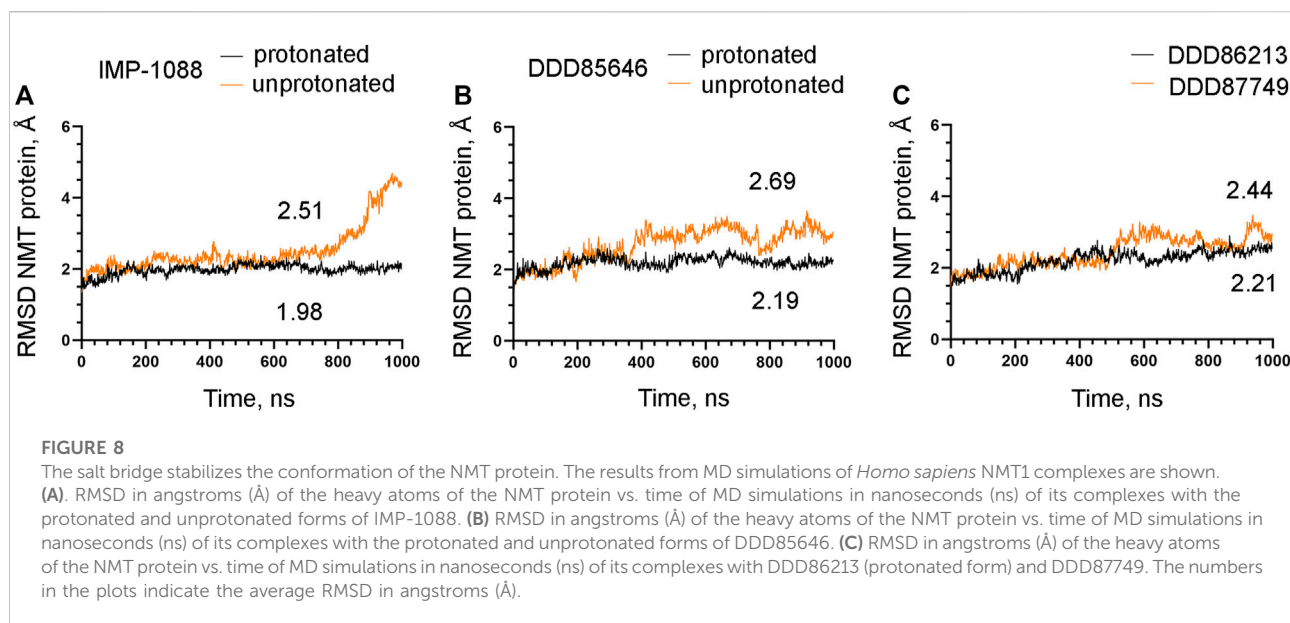
**FIGURE 7**

Stability of NMT inhibitor complexes by MD simulations. (A) RMSD in angstroms (Å) of the heavy atoms of the protonated and unprotonated IMP-1088 vs. time of MD simulations in nanoseconds (ns). (B) RMSD values in angstroms (Å) of the heavy atoms in the protonated and unprotonated DDD85646 forms vs. time of MD simulations in nanoseconds (ns). (C) RMSD in angstroms (Å) of the heavy atoms of DDD86213 (protonated form) and DDD87749 vs. time of MD simulations in nanoseconds (ns). The numbers in the plots indicate the average RMSD in angstroms (Å).

3.3 The salt bridge contributes to the stability of the NMT-ligand complexes

To examine the role of the salt bridge in the stability of NMT-ligand complexes, we performed MD simulations using the *Homo sapiens* NMT1 complexes with the unprotonated and protonated forms of IMP-1088 and DDD85646. An excellent representation of the stability of the salt bridge can be obtained by determining the distance between the nitrogen atom of the ligand (acting as a hydrogen bond donor) and the negatively charged

oxygen atom at the C-terminus of the NMT protein (acting as hydrogen bond acceptor) (Figures 4E,F), during the time course of MD simulations. The results with the protonated forms of IMP-1088 and DDD85646 indicated that the salt bridge is stable because its length did not exceed the permissible 4 Å distance for a salt bridge length during the entire duration of the MD simulation (Figure 6). Deprotonation of the small molecule eliminates the salt bridge and leads to substantial instability of the interaction between the inhibitors and the C-terminus of NMT (Figure 6). The unprotonated dimethylamino group of

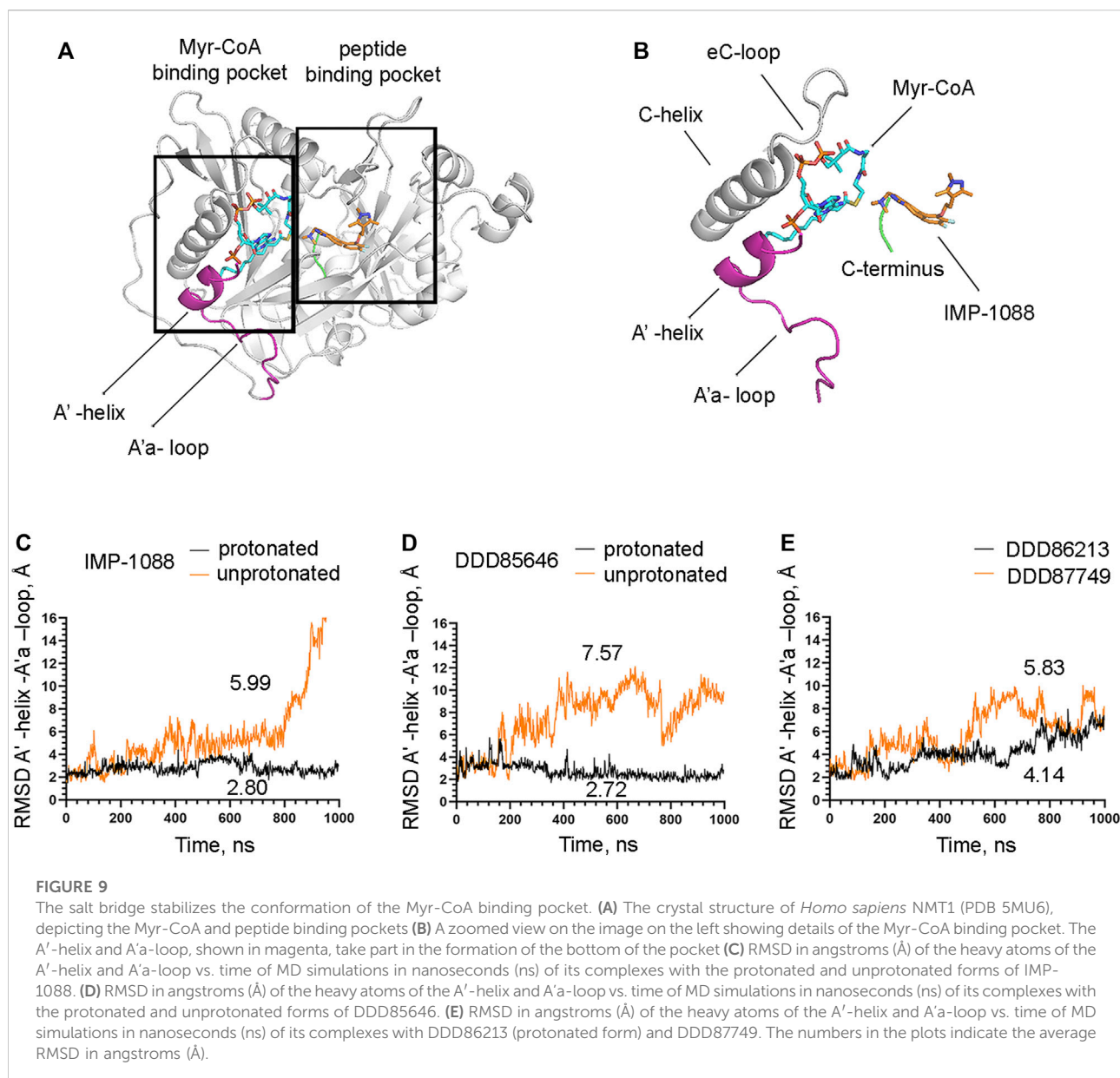


IMP-1088 lacks polar hydrogens (Figure 3A); hence, this form is incapable of forming a hydrogen bond with the C-terminus of NMT. This may explain why the unprotonated IMP-1088 forms particularly unstable complexes (Figure 6A). In contrast, the unprotonated piperazine ring of DDD85646 contains a polar hydrogen atom (Figure 3C) that may participate in hydrogen bonding with the C-terminus of NMT. This may explain why the DDD85646 complex with NMT displays relative stability at the beginning to up to 400 ns during the MD simulation (Figure 6B). However, later in the 400–1,000 ns period, the hydrogen bond is disrupted, suggesting that it cannot substitute for the salt bridge (Figure 6B). Similar results were obtained using the NMT complexes with DDD86213 and DDD87749 (Figure 2). For example, the protonated form of DDD86213 formed a stable salt bridge during MD simulations, and the interaction between the inhibitor and the C-terminus of NMT was disrupted due to the replacement of the nitrogen donor atom in DDD87749 with carbon (Figure 6C). As expected, the RMSD values of the heavy atoms for the unprotonated ligand molecules were increased compared to protonated forms (Figure 7). However, the RMSD values of the NMT inhibitors were lower than 1.5 Å, except in part for DDD87749, indicating that the small molecules remained largely restrained to their binding pockets, even though they have lost the interaction with the C-terminus of NMT (Compare Figure 6 and Figure 7). The increased distance between the inhibitor and the C-terminus of NMT in the non-protonated forms compared to protonated (Figure 6) is at least partly due to the conformational movement of the

C-terminal region relative to the protein backbone. For example, the average RMSD of the C-terminal amino acid (Gln496) was determined to be 1.10 Å and 3.04 Å in the complexes of protonated and unprotonated IMP-1088 complexes, 2.92 Å and 3.59 Å for the protonated and unprotonated DDD85646 and 1.66 Å and 2.90 Å in DDD86213 and DDD87749 complexes, respectively, indicating that the salt bridge restricts the movement of the C-terminus of NMT.

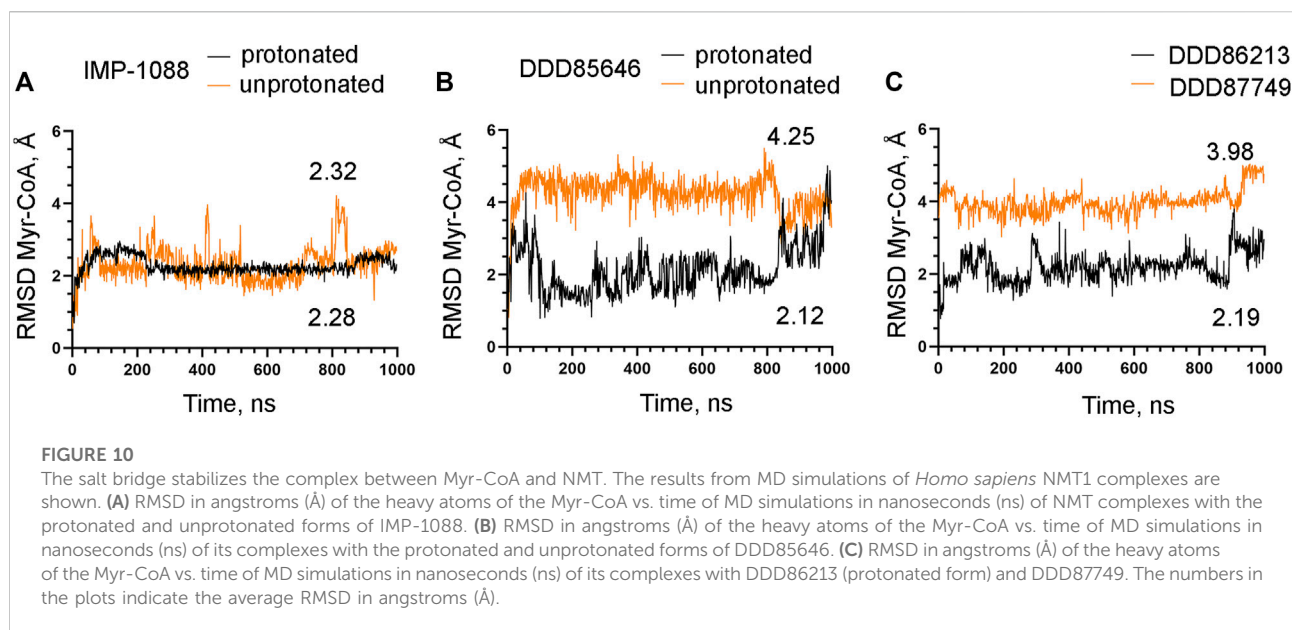
3.4 The salt bridge stabilizes the conformation of the NMT protein

The observation that the NMT inhibitor molecules are restrained into their binding pockets during MD simulations, irrespective of their protonation state (Figure 7), raises questions about how the salt bridge affects ligand potency. Since the salt bridge is an interaction between the inhibitor and the NMT protein, we hypothesize that its effects could also depend on its impact on the protein structure. To test this hypothesis, we compared the dynamic stability of the NMT protein during the MD simulations in complexes with either the protonated or the unprotonated forms of the inhibitors (Figure 8). The RMSD values of the heavy atoms of the NMT protein were significantly increased in the complexes of the unprotonated inhibitors compared to the complexes of the protonated forms (Figure 8), suggesting that the salt bridge stabilizes the conformation of the NMT protein. Similarly, DDD86213, which forms a salt bridge, stabilizes the NMT protein's conformation compared to



DDD87749, which does not (Figure 8). To investigate what specific conformation of the NMT protein was stabilized by the salt bridge, we observed movies of the protein dynamics during MD simulations. Most strikingly, the conformation and the position of the A'-helix and the A'a-loop, which form a part of the Myr-CoA binding pocket (Figures 9A,B), remained stable in the complexes containing protonated IMP-1088, DDD85646 and DDD86213; however, they were dramatically altered in the complexes containing the non-protonated forms of the inhibitors. RMSD calculations confirmed the extraordinary mobility of the A'-helix - A'a-loop region in the complexes with non-protonated inhibitors and its stabilization in the presence of the

protonated forms during MD simulations (Figures 9C-E). These results suggest that the salt bridge's presence stabilizes the conformation of the Myr-CoA binding pocket from which the A'-helix and the A'a-loop are part. Consistent with this, the RMSD values of the Myr-CoA cofactor in the NMT complexes with the protonated inhibitors were reduced compared to those with the non-protonated inhibitors (Figure 10). It was also observed that the eC-loop, located just above the Myr-CoA pocket (Figure 9B), adopts a partially helical conformation in some complexes during MD simulations (not shown). However, this was not related to the inhibitors' protonation state and is likely not associated with the salt bridge formation.



4 Discussion

Here we report that IMP-1088 and DDD85646, two potent NMT inhibitors, form a salt bridge within the active site of the N-myristoyltransferases. The salt bridge is mediated by the interaction between the positively charged groups (dimethylamino group or piperazine ring, respectively) of the inhibitors and the negatively charged carboxyl group at the C-terminus of the enzyme, which is located in the catalytic center of the enzyme. The fact that the salt bridge is formed in the complexes of two structurally unrelated NMT inhibitors, independently identified by high-throughput screening (Frearson et al., 2010; Mousnier et al., 2018; Spasov et al., 2022), implies that the salt bridge may play a special role in NMT inhibition. Indeed, preventing its formation by replacing the nitrogen atom of the ligand, involved in the salt bridge, with a carbon atom dramatically reduces the potency of the inhibitors (Figure 2) (Brand et al., 2017).

Along with the hydrogen, halogen, and chalcogen bonds, the salt bridge belongs to the group of non-covalent interactions, which depend on electrostatic attraction. However, recent findings based on the quantitative Kohn-Sham molecular orbital theory have indicated orbital interactions between the participating atoms, resembling the ones involved in the formation of the covalent bonds, demonstrating that a pure electrostatic model cannot fully describe these non-covalent interactions (Wang et al., 2016; de Azevedo Santos et al., 2021). Since the salt bridge can be viewed as a charge-enforced hydrogen bond, such orbital interactions may also participate during its formation. However, applying the Kohn-Sham molecular orbital theory for the formation of the salt bridge has not been reported in the literature, and the precise molecular orbital interactions during its formation remain to be determined.

Generally, the strength of the salt bridge is highly dependent on the environment (Takano et al., 2000). On the surface of the proteins, the salt bridges may not contribute significantly to interactions, as the gain in free binding energy due to the salt bridge's formation is insufficient to compensate for the energetic penalty of desolvating its charged groups (Takano et al., 2000; Ferreira de Freitas and Schapira, 2017). In contrast, buried salt bridges can significantly contribute to the binding (Ferreira de Freitas and Schapira, 2017). In this aspect, the salt bridges formed inside the NMT inhibitor complexes are buried into the enzyme's catalytic center and may possess significant strength.

The salt bridge is the strongest non-covalent interaction and participates in protein folding and protein-protein interactions, and it is known to contribute to protein conformational stability as well (Kumar and Nussinov, 2002; Bosshard et al., 2004; Donald et al., 2011; Basu and Mukharjee, 2017; Ferreira de Freitas and Schapira, 2017). However, the role of salt bridges in drug-protein interactions has remained not well understood. In a recent study, Kurczab et al. identified 1122 unique small molecule ligands from the Protein Databank that form salt bridges with their protein targets, indicating that salt bridges are frequent in drug-protein interactions (Kurczab et al., 2018). The set contains structures from different enzyme classes, including hydrolase, transferases, reductase, oxidoreductase, lyases, and certain G protein-coupled receptors (GPCRs) (Kurczab et al., 2018). In several cases, it has been experimentally shown that the salt bridge plays a critical role. For example, the Epidermal Growth Factor Receptor (EGFR) inhibitor, compound 10, forms a salt bridge with Asp831 in the kinase domain of EGFR through its terminal dimethylamino group (Peng et al., 2013; Ferreira de Freitas and Schapira, 2017). Replacing the nitrogen atom of this group with a carbon atom reduced the potency of this compound by over 800 folds (from IC_{50} 29 nM to IC_{50} 25 μ M) (Peng et al., 2013;

Ferreira de Freitas and Schapira, 2017). In addition, it has been reported that preventing the formation of the salt bridge between the aminergic ligands and the aminergic class A GPCRs by mutating the Asp3.32 residue, which is engaged in the salt bridge, to Ala reduces the binding 126-fold (Kurczab et al., 2018). In comparison, the absence of the salt bridge reduces the activity of the NMT inhibitors by over 1,300 fold (Figure 2) (Brand et al., 2017). Although the salt bridge is stronger than the hydrogen bond, such substantial differences in potency in the pairs of compounds that are capable of forming salt bridges or not are difficult to explain by a simple increase in the strength of the interaction between the ligand and the protein and may suggest the existence of a more complex mechanism that mediate or at least amplify the salt bridge's effects.

The mechanisms that mediate the salt bridge effect on the inhibitors' potency were investigated by MD simulations using NMT complexes with the protonated and unprotonated forms of the inhibitors. These forms of the inhibitors are best suited to study the role of the charge of the molecule and the salt bridge with minimum structural change. In DDD85646, this allows comparing compounds that form a hydrogen bond (unprotonated DDD85646) to compounds that form a salt bridge (protonated DDD85646). However, considering the pKa of the IMP-1088 and DDD85646, the unprotonated forms may rarely exist under the physiological pH, and thus the interaction between unprotonated forms and NMT may not occur under these conditions. To address this point, we also performed MD simulations of NMT complexes with DDD86213 and DDD87749, where the activity of the compounds under physiological pH has been determined (Figure 2). The results demonstrated the significance of adding a charge species to the DDD87749 scaffold and how this can be used to increase the ligands' potency. The inability of the inhibitors' non-protonated forms to participate in a salt bridge leads to the instability of their association with the C-terminus of NMT (Figure 6). This does not lead to the dissociation of the inhibitor molecules, at least in the 1 μ s time frame of the MD simulations, as they remain largely confined to their binding pockets (Figure 7). However, the salt bridge's presence stabilizes the NMT protein's conformation (Figure 8). Interestingly, the activity of the compounds IMP-1088 > DDD85646 > DDD86213 > DDD87749 corresponds to the compounds' ability to stabilize the NMT protein's conformation (Figure 8), implying that the stabilization of NMT conformation by the salt bridge may determine their potency. Among the protonated forms, DDD86213 has the weakest capacity to stabilize the NMT conformation (Figure 8), consistent with its lower inhibitory activity and that DDD86123 is an intermediate lead compound in the DDD85646 series and its activity has not been wholly optimized (Figure 2).

The salt bridge's effect on the protein conformation is particularly evident in the A'-helix and the A'-loop, which

forms part of the Myr-CoA binding pocket (Figure 9). By stabilizing the conformation of the Myr-CoA binding pocket, the salt bridge exerts a positive effect on the stability of the complex between NMT and its cofactor Myr-CoA (Figure 10). By doing so, the salt bridge may also indirectly exert a positive effect on the stability of the whole complex. For example, it has been reported that the binding affinity of DDD85646 to NMT is increased 33-fold in the presence of Myr-CoA (Frearson et al., 2010); hence the stability of the NMT-Myr-CoA complex may translate into increased stability of the interaction between the NMT inhibitors themselves and the NMT protein. In addition to the more pronounced conformational changes in the Myr-CoA binding pocket, smaller conformational changes exist in other protein regions, including the C-terminus of NMT. Thus the stabilization of regions of the protein structure by the salt bridge, other than the Myr-CoA binding pocket, may also play a role in the compounds' activity.

Salt bridges have been reported to function as molecular clips that stitch together large surface areas at interacting protein interfaces (Basu and Mukharjee, 2017). Similarly, the results reported here suggest that the formation of a salt bridge between the protein receptor and the ligand may stitch a particular protein region and by restricting its conformational movements to exert a stabilizing effect on the overall protein conformation. Therefore, the salt bridge's impact on the inhibitors' potency could be more complex than previously anticipated and may involve, at least in part, the stabilization of the protein receptor conformation. This finding may not be surprising in the light that salt bridges, in general, are known to contribute to protein conformational stability (Takano et al., 2000; Bosshard et al., 2004).

In the current work, the parameters for MD simulations were optimized for studying protein conformational dynamics. The systems were simulated with the ff14SB force field (Hornak et al., 2006; Maier et al., 2015; Case et al., 2018), specially parameterized and recommended for protein dynamics as a part of the AMBER v. 18 package (Case et al., 2005; Case et al., 2018). ff14SB (Maier et al., 2015) has evolved from the ff99SB force field and includes improvements in the torsional parameters for the backbone and side chains (Hornak et al., 2006; Case et al., 2018). Frames were saved every 1 ns to generate 1,000 frames for a total of 1 μ s duration of MD production simulations to detect significant events in the nanosecond time scale. These include protein conformational changes such as side chain rotations, backbone and loop motions, and ligand binding. Statistical inefficiency, principal component analysis, and wavelet analysis could be used to identify significant events during MD simulations to reduce the number of frames of MD simulations without losing statistical power (Heidari et al., 2016; Gonçalves et al., 2017). However, considering the number of frames analyzed in the nanosecond time range in the current study, most, if not all, of the significant events related

to protein conformational changes and ligand binding are expected to be captured during analysis.

During the MD simulations, some conformational changes in the NMT protein were observed that were not related to the protonation state of the inhibitors and hence are not dependent on the presence or the absence of the salt bridge; nevertheless, they might be of interest to researchers studying the N-myristoyltransferases. Specifically, it was observed that the eC-loop, positioned just on top of the Myr-CoA binding pocket, opens and adopts a partially helical composition in some MD simulations. An analogous loop, the Ab-loop, is located on top of the adjacent peptide-binding pocket (Dian et al., 2020). The Ab-loop has been reported to exist in open and closed conformations and controls the entry of the substrate peptide into the catalytic center (Dian et al., 2020). The observation of alternative conformations of the eC-loop suggests the possibility that this loop may also perform a similar role in regulating the entry of Myr-CoA into its binding pocket. However, this remains to be determined in future studies.

The salt bridges typically involve the amino acid side chains of Asp or Glu if the ligand is positively charged or the side chains of Lys or Arg if the ligand is negatively charged (Ferreira de Freitas and Schapira, 2017; Kurczab et al., 2018). The occurrence of a salt bridge with the carboxyl group at the C-terminus of the protein is probably unique to N-myristoyltransferases and possibly is a reflection of its unusual localization in the active site of the enzyme and its catalytic functions (Dian et al., 2020). According to Uniprot (UniProtConsortium, 2021), the C-terminal amino acid residue in *Homo sapiens* NMT1 and NMT2 is Gln. Gln is also the C-terminal residue in NMTs of the vertebrate species, including *Danio rerio*, *Xenopus laevis*, *Gallus gallus*, and *Mus musculus*. In the protozoan species, the identity of the C-terminal amino acid is not conserved—it is Val446 in *Trypanosoma brucei* NMT, Leu410, and Leu421 in *Plasmodium falciparum* and *Leishmania major* NMTs. This may not be surprising considering that the side chain of the C-terminal amino acid is orientated in the opposite direction of the active site and is not involved in the catalysis (Dian et al., 2020). However, the conservation of the C-terminal residue of NMT in all vertebrate species may suggest that the side chain of this residue may play a role that remains to be determined.

The positive charge of the NMT inhibitors may influence other pharmacological properties of theirs, such as cell membrane or BBB permeability, absorption, and distribution. It is estimated that about 40% of approved prescription drugs are positively charged and exist as organic cations at neutral pH (Uddin et al., 2021). The transport of such organic cations across the membrane could be dependent on or facilitated by uptake from transport carrier proteins (Thomas et al., 2004; Sprowl et al., 2016). Irrespective of the entry mechanism, the NMT inhibitors, as documented in numerous articles, are

undoubtedly capable of crossing the cell membrane and can effectively target the N-myristoyltransferases intracellularly both in cell culture models and *in vivo* (Thinon et al., 2014; Beauchamp et al., 2020).

In conclusion, the results suggest that salt bridges could be used as valuable tools in drug design. The inclusion of charged chemical groups in the ligand structures, where they can participate in salt bridges with the target proteins, analogous to the case of DDD87749 and DDD86213, could lead to a significant increase in activity. Moreover, the results also imply that conformational stabilization of the target protein structure could be a hallmark of the salt bridges' effects. In this aspect, studying the protein dynamics by MD simulations could be used to predict the outcome of introducing a new salt bridge in a specific protein-ligand complex.

Data availability statement

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

Author contributions

Conceptualization, DS, MA, and ID; methodology, DS, MA, and ID; investigation, DS and MA; writing—original draft preparation, DS; writing—review and editing, DS, MA, and ID; visualization, DS, MA, and ID; supervision, ID; project administration, ID; funding acquisition, DS, ID. All authors have read and agreed to the published version of the manuscript.

Funding

This project has received funding from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement No. 882247. The *in silico* calculations were performed in the Centre of Excellence for Informatics and ICT supported by the Science and Education for Smart Growth Operational Program and co-financed by the European Union through the European Structural and Investment funds (Grant No. BG05M2OP001-1.001-0003).

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

References

- Adelman, S. A., and Doll, J. D. (1974). Generalized Langevin equation approach for atom/solid-surface scattering: Collinear atom/harmonic chain model. *J. Chem. Phys.* 61, 4242–4245. doi:10.1063/1.1681723
- Basu, S., and Mukharjee, D. (2017). Salt-bridge networks within globular and disordered proteins: Characterizing trends for designable interactions. *J. Mol. Model* 23 (7), 206. doi:10.1007/s00894-017-3376-y
- Beauchamp, E., Yap, M. C., Iyer, A., Perinpanayagam, M. A., Gamma, J. M., Vincent, K. M., et al. (2020). Targeting N-myristoylation for therapy of B-cell lymphomas. *Nat. Commun.* 11 (1), 5348. doi:10.1038/s41467-020-18998-1
- 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
- Bietz, S., Urbaczek, S., Schulz, B., and Rarey, M. (2014). Protoss: A holistic approach to predict tautomers and protonation states in protein-ligand complexes. *J. Cheminform* 6, 12. doi:10.1186/1758-2946-6-12
- Bosshard, H. R., Marti, D. N., and Jelezarov, I. (2004). Protein stabilization by salt bridges: Concepts, experimental approaches and clarification of some misunderstandings. *J. Mol. Recognit.* 17 (1), 1–16. doi:10.1002/jmr.657
- Brand, S., Wyatt, P., Thompson, S., Smith, V., Bayliss, T., Harrison, J., et al. (2017). *N-myristoyl transferase inhibitors*. United States Patent and Trademark Office. US patent No. US 9,828,346 B2.
- Case, D. A., Ben-Shalom, I. Y., Brozell, S. R., Cerutti, D. S., Cheatham, T. E., III, Cruzeiro, V. W. D., et al. (2018). *Amber*. San Francisco, CA, USA: University of California. Available online: <https://ambermd.org/doc12/Amber18.pdf>.
- Case, D. A., Cheatham, T. E., 3rd, Darden, T., Gohlke, H., Luo, R., Merz, K. M., Jr., et al. (2005). The Amber biomolecular simulation programs. *J. Comput. Chem.* 26 (16), 1668–1688. doi:10.1002/jcc.20290
- Ciccotti, G., and Ryckaert, J. P. (1986). Molecular dynamics simulation of rigid molecules. *Comput. Phys. Rep.* 4 (6), 346–392. doi:10.1016/0167-7977(86)90022-5
- de Azevedo Santos, L., van der Lubbe, S. C. C., Hamlin, T. A., Ramalho, T. C., and Matthias Bickelhaupt, F. (2021). A quantitative molecular orbital perspective of the chalcogen bond. *ChemistryOpen* 10 (4), 390. doi:10.1002/open.202100066
- Dian, C., Perez-Dorado, I., Riviere, F., Asensio, T., Legrand, P., Ritzefeld, M., et al. (2020). High-resolution snapshots of human N-myristoyltransferase in action illuminate a mechanism promoting N-terminal Lys and Gly myristoylation. *Nat. Commun.* 11 (1), 1132. doi:10.1038/s41467-020-14847-3
- Donald, J. E., Kulp, D. W., and DeGrado, W. F. (2011). Salt bridges: Geometrically specific, designable interactions. *Proteins* 79 (3), 898–915. doi:10.1002/prot.22927
- Ducker, C. E., Upson, J. J., French, K. J., and Smith, C. D. (2005). Two N-myristoyltransferase isozymes play unique roles in protein myristoylation, proliferation, and apoptosis. *Mol. Cancer Res.* 3 (8), 463–476. doi:10.1158/1541-7786.MCR-05-0037
- Ferreira de Freitas, R., and Schapira, M. (2017). A systematic analysis of atomic protein-ligand interactions in the PDB. *Medchemcomm* 8 (10), 1970–1981. doi:10.1039/c7md00381a
- Flinck, M., Kramer, S. H., and Pedersen, S. F. (2018). Roles of pH in control of cell proliferation. *Acta Physiol. (Oxf)* 223 (3), e13068. doi:10.1111/apha.13068
- Frearson, J. A., Brand, S., McElroy, S. P., Cleghorn, L. A., Smid, O., Stojanovski, L., et al. (2010). N-myristoyltransferase inhibitors as new leads to treat sleeping sickness. *Nature* 464 (7289), 728–732. doi:10.1038/nature08893
- Gaffarogullari, E. C., Masterson, L. R., Metcalfe, E. E., Traaseth, N. J., Balatri, E., Musa, M. M., et al. (2011). A myristoyl/phosphoserine switch controls cAMP-dependent protein kinase association to membranes. *J. Mol. Biol.* 411 (4), 823–836. doi:10.1016/j.jmb.2011.06.034
- Gonçalves, M. A., Santos, L. S., Prata, D. M., Peixoto, F. C., da Cunha, E. E. F., and Ramalho, T. C. (2017). Optimal wavelet signal compression as an efficient alternative to investigate molecular dynamics simulations: Application to thermal and solvent effects of MRI probes. *Theor. Chem. Accounts* 136 (15), 15. doi:10.1007/s00214-016-2037-z
- Grimsley, G. R., Scholtz, J. M., and Pace, C. N. (2009). A summary of the measured pK values of the ionizable groups in folded proteins. *Protein Sci.* 18 (1), 247–251. doi:10.1002/pro.19
- Heidari, Z., Roe, D. R., Galindo-Murillo, R., Ghasemi, J. B., and Cheatham, T. E., 3rd (2016). Using wavelet analysis to assist in identification of significant events in molecular dynamics simulations. *J. Chem. Inf. Model* 56 (7), 1282–1291. doi:10.1021/acs.jcim.5b00727
- 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 (3), 712–725. doi:10.1002/prot.21123
- Humphrey, W., Dalke, A., and Schulten, K. (1996). Vmd: Visual molecular dynamics. *J. Mol. Graph* 14 (1), 33–38. doi:10.1016/0263-7855(96)00018-5
- Khalili, F., Henni, A., and East, A. L. L. (2009). pKa values of some piperazines at (298, 303, 313, and 323) K. *J. Chem. Eng. Data* 54 (10), 2914–2917. doi:10.1021/je900005c
- Krieger, E., and Vriend, G. (2014). YASARA View - molecular graphics for all devices - from smartphones to workstations. *Bioinformatics* 30 (20), 2981–2982. doi:10.1093/bioinformatics/btu426
- Kumar, S., and Nussinov, R. (2002). Close-range electrostatic interactions in proteins. *Chembiochem* 3 (7), 604–617. doi:10.1002/1439-7633(20020703)3:7<604::AID-CBIC604>3.0.CO;2-X
- Kurczab, R., Sliwa, P., Rataj, K., Kafel, R., and Bojarski, A. J. (2018). Salt bridge in ligand-protein complexes-systematic theoretical and statistical investigations. *J. Chem. Inf. Model* 58 (11), 2224–2238. doi:10.1021/acs.jcim.8b00266
- Maier, J. A., Martinez, C., Kasavajhala, K., Wickstrom, L., Hauser, K. E., and Simmerling, C. (2015). ff14SB: Improving the accuracy of protein side chain and backbone parameters from ff99SB. *J. Chem. Theory Comput.* 11 (8), 3696–3713. doi:10.1021/acs.jctc.5b00255
- Meinell, T., Dian, C., and Giglione, C. (2020). Myristoylation, an ancient protein modification mirroring eukaryogenesis and evolution. *Trends Biochem. Sci.* 45 (7), 619–632. doi:10.1016/j.tibs.2020.03.007
- Mousnier, A., Bell, A. S., Swieboda, D. P., Morales-Sanfrutos, J., Perez-Dorado, I., Brannigan, J. A., et al. (2018). Fragment-derived inhibitors of human N-myristoyltransferase block capsid assembly and replication of the common cold virus. *Nat. Chem.* 10 (6), 599–606. doi:10.1038/s41557-018-0039-2
- Patwardhan, P., and Resh, M. D. (2010). Myristoylation and membrane binding regulate c-Src stability and kinase activity. *Mol. Cell Biol.* 30 (17), 4094–4107. doi:10.1128/MCB.00246-10
- Peng, Y. H., Shiao, H. Y., Tu, C. H., Liu, P. M., Hsu, J. T., Amancha, P. K., et al. (2013). Protein kinase inhibitor design by targeting the asp-phe-gly (DFG) motif: The role of the DFG motif in the design of epidermal growth factor receptor inhibitors. *J. Med. Chem.* 56 (10), 3889–3903. doi:10.1021/jm400072p
- Persi, E., Duran-Frigola, M., Damaghi, M., Roush, W. R., Aloy, P., Cleveland, J. L., et al. (2018). Systems analysis of intracellular pH vulnerabilities for cancer therapy. *Nat. Commun.* 9 (1), 2997. doi:10.1038/s41467-018-05261-x
- Ritzefeld, M., Wright, M. H., and Tate, E. W. (2018). New developments in probing and targeting protein acylation in malaria, leishmaniasis and African sleeping sickness. *Parasitology* 145 (2), 157–174. doi:10.1017/S0031182017000282
- Schlott, A. C., Holder, A. A., and Tate, E. W. (2018). N-myristoylation as a drug target in malaria: Exploring the role of N-myristoyltransferase substrates in the inhibitor mode of action. *ACS Infect. Dis.* 4 (4), 449–457. doi:10.1021/acsinfectdis.7b00203
- Schrödinger, L., and DeLano, W. (2020). PyMOL. Available at: <http://www.pymol.org/pymol>.
- Settimo, L., Bellman, K., and Knettel, R. M. (2014). Comparison of the accuracy of experimental and predicted pKa values of basic and acidic compounds. *Pharm. Res.* 31 (4), 1082–1095. doi:10.1007/s10955-013-1232-z
- Shelley, J. C., Cholleti, A., Frye, L. L., Greenwood, J. R., Timlin, M. R., and Uchimaya, M. (2007). Epik: A software program for pK(a) prediction and protonation state generation for drug-like molecules. *J. Comput. Aided Mol. Des.* 21 (12), 681–691. doi:10.1007/s10822-007-9133-z

- Spasov, D. S., Atanasova, M., and Doytchinova, I. (2022). Novel hits for N-myristoyltransferase inhibition discovered by docking-based screening. *Molecules* 27, 5478. doi:10.3390/molecules27175478
- Spasov, D. S., Ruiz-Saenz, A., Piple, A., and Moasser, M. M. (2018). A dimerization function in the intrinsically disordered N-terminal region of src. *Cell Rep.* 25 (2), 449–463. doi:10.1016/j.celrep.2018.09.035
- Sprowl, J. A., Ong, S. S., Gibson, A. A., Hu, S., Du, G., Lin, W., et al. (2016). A phosphotyrosine switch regulates organic cation transporters. *Nat. Commun.* 7, 10880. doi:10.1038/ncomms10880
- Takano, K., Tsuchimori, K., Yamagata, Y., and Yutani, K. (2000). Contribution of salt bridges near the surface of a protein to the conformational stability. *Biochemistry* 39 (40), 12375–12381. doi:10.1021/bi000849s
- Thinon, E., Morales-Sanfrutos, J., Mann, D. J., and Tate, E. W. (2016). N-myristoyltransferase inhibition induces ER-stress, cell cycle arrest, and apoptosis in cancer cells. *ACS Chem. Biol.* 11 (8), 2165–2176. doi:10.1021/acscchembio.6b00371
- Thinon, E., Serwa, R. A., Broncel, M., Brannigan, J. A., Brassat, U., Wright, M. H., et al. (2014). Global profiling of co- and post-translationally N-myristoylated proteomes in human cells. *Nat. Commun.* 5, 4919. doi:10.1038/ncomms5919
- Thomas, J., Wang, L., Clark, R. E., and Pirmohamed, M. (2004). Active transport of imatinib into and out of cells: Implications for drug resistance. *Blood* 104 (12), 3739–3745. doi:10.1182/blood-2003-12-4276
- Uddin, M. E., Talebi, Z., Chen, S., Jin, Y., Gibson, A. A., Noonan, A. M., et al. (2021). *In vitro* and *in vivo* inhibition of MATE1 by tyrosine kinase inhibitors. *Pharmaceutics* 13 (12), 2004. doi:10.3390/pharmaceutics13122004
- UniProtConsortium (2021). UniProt: The universal protein knowledgebase in 2021. *Nucleic Acids Res.* 49 (D1), D480–D489. doi:10.1093/nar/gkaa1100
- Wang, B., Jiang, W., Dai, X., Gao, Y., Wang, Z., and Zhang, R. Q. (2016). Molecular orbital analysis of the hydrogen bonded water dimer. *Sci. Rep.* 6, 22099. doi:10.1038/srep22099



OPEN ACCESS

EDITED BY

Chandrabose Selvaraj,
Saveetha University, India

REVIEWED BY

Gunassekaran Gowri Rangaswamy,
Kyungpook National University, Republic
of Korea
Bhuvaneswari Sampath,
Anna University, India
Noor Ahmad Shaik,
King Abdulaziz University, Saudi Arabia

*CORRESPONDENCE

S. Gowtham Kumar,
✉ gowtham_phd@yahoo.com
K. Langeswaran,
✉ dr.langeswaran@gmail.com

SPECIALTY SECTION

This article was submitted to Biophysics,
a section of the journal
Frontiers in Molecular Biosciences

RECEIVED 15 December 2022

ACCEPTED 23 January 2023

PUBLISHED 09 February 2023

CITATION

Kirubhanand C, Merciline Leonora J,
Anitha S, Sangeetha R, Nachammai KT,
Langeswaran K and Gowtham Kumar S
(2023), Targeting potential receptor
molecules in non-small cell lung cancer
(NSCLC) using in silico approaches.
Front. Mol. Biosci. 10:1124563.
doi: 10.3389/fmolb.2023.1124563

COPYRIGHT

© 2023 Kirubhanand, Merciline Leonora,
Anitha, Sangeetha, Nachammai,
Langeswaran and Gowtham Kumar. 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.

Targeting potential receptor molecules in non-small cell lung cancer (NSCLC) using in silico approaches

C. Kirubhanand¹, J. Merciline Leonora², S. Anitha³, R. Sangeetha⁴,
K. T. Nachammai⁵, K. Langeswaran^{5*} and S. Gowtham Kumar^{6*}

¹Department of Anatomy, All India Institute of Medical Sciences, Nagpur, Maharashtra, India, ²PG and Research Department of Physics, Government Arts College, Madurai, Tamil Nadu, India, ³Department of Physics, ArulmiguPalaniandavar College of Arts and Science, Palani, Tamil Nadu, India, ⁴Department of Physics, Mannar Thirumalai Naicker College, Madurai, Tamil Nadu, India, ⁵Department of Biotechnology, Science Campus, Alagappa University, Karaikudi, Tamil Nadu, India, ⁶Faculty of Allied Health Sciences, Chettinad Hospital & Research Institute, Chettinad Academy of Research and Education, Kelambakkam, Tamil Nadu, India

Introduction: Non-Small Cell Lung Cancer is the most prevalent type of cancer in lung cancer. Chemotherapy, radiation therapy, and other conventional cancer treatments have a low success rate. Thus, creating new medications is essential to halt the spread of lung cancer.

Methods: In this study bioactive nature of lochnericine against Non-Small Cell Lung Cancer (NSCLC) was analyzed using various computational approaches such as quantum chemical calculations, molecular docking, and molecular dynamic simulation. Furthermore, the MTT assay shows the anti-proliferation activity of lochnericine.

Results and Discussion: Using Frontier Molecular Orbital (FMO), the calculated band gap energy value associated with bioactive compounds and the molecule's potential bioactivity is confirmed. The H38 hydrogen atom and O1 oxygen atom in the molecule are effectively electrophilic, and potential nucleophilic attack sites were confirmed through analysis of the Molecular electrostatic potential surface. Furthermore, the electrons within the molecule were delocalized, which confers bioactivity on the title molecule and was authorized through Mulliken atomic charge distribution analysis. A molecular docking study revealed that lochnericine inhibits non-small cell lung cancer-associated targeted protein. The lead molecule and targeted protein complex were stable during molecular dynamics simulation studies till the simulation period. Further, lochnericine demonstrated remarkable anti-proliferative and apoptotic features against A549 lung cancer cells. The current investigation powerfully suggests that lochnericine is a potential candidate for lung cancer.

KEYWORDS

lochnericine, quantum chemical calculations, molecular docking, molecular dynamics, anticancer study, non-small lung cancer

1 Introduction

Developed and developing nations consider cancer a significant health concern with increasing incidence and mortality rates. The tumor is caused by complex biological processes, namely uninhibited cell proliferation, cell death resistance, neo-angiogenesis, and metastasis (Hausman, 2019). Incursion and metastasis are the frequent causes of cancer-related mortality, resulting in additional invasion sites and severe organ damage. Cases of cancer are

rapidly becoming a risk factor for various malignancies, most notably lung, liver, colorectal, and breast cancers. Because of the disease's high notoriety, therapy has been a never-ending fight with few results (Mullard, 2020; Banaganapalli et al., 2022). One of the deadliest and most ubiquitous forms of cancer in people is lung cancer. The most prevalent type of cancer is NSCLC (non-small cell lung cancer), estimated as 85% among other cancers. Nearly 90% of non-small cell lung cancer occurrences are caused by smoking. Tumor-infiltrating lymphocytes (TILs), in general, and notably CD8⁺ TILs, are associated with a favourable prognosis in non-small cell lung cancer (NSCLC) (Gueguen et al., 2021). Traditional cancer treatments, such as surgery, radiation, and chemotherapy, have a low success rate. Thus, developing new drugs is crucial to stop lung cancer spread (Schabath and Cote, 2019).

In cancer treatment, natural products may be used in addition to traditional anticancer medications. Recently, researchers worldwide have focused on developing novel drugs with medicinal potential derived from natural sources such as plants. Plants and their phytoconstituents have been used for medical purposes since antiquity (Ekiert and Szopa, 2020). Natural-source chemicals attract the interest of scientists worldwide. Humans have extracted cancer chemotherapeutic chemicals from flora (Deng et al., 2020). Lochnericine is a significant monoterpene indole alkaloid (MIA), found in the roots of Madagascar periwinkle (*Catharanthus roseus*) and formed by the stereoselective C6, C7-epoxidation of tabersonine. The stereoselective C6 produces Lochnericine and C7-epoxidation of tabersonine and can be subsequently metabolized to produce more complex MIAs. Although the enzymes in charge of its downstream modifications have been identified, those in charge of lochnericine production have not.

Growth factors, namely hormones, mainly influence the migration of normal cells and the division of cells. Tyrosine kinase EGRF (Epidermal Growth Factor Receptor) influences cell proliferation, cell division, and tumour development. It is a representative transmembrane receptor which triggers signalling pathways *via* ligand-provoked dimerization (Kavitha et al., 2015; Zhou and Yao, 2016; Shaik et al., 2019). In human cancers, the most prevalently mutated oncogene is the Kirsten rat sarcoma (KRAS) gene involved in carcinogenesis, which accounts for more than twenty per cent of lung cancer, particularly NSCLC (Reck et al., 2021). There is a strong connotation between various solid tumours NTRK (Neurotrophic Tropomyosin Receptor Kinase) gene fusion, including NSCLC. In cancers covering NSCLC, ALK activation occurs through fusion gene formation, the preliminary actuating mutation in ALK. TRIM 1 is designated by cancer cells to encourage tumorigenic development and is upregulated in malignant cells. In tumour cells, it indorses elevated proteasome activity (Wang et al., 2016). This study investigated a lochnericine molecule using DFT quantum chemical calculations, including molecular structure, vibrational (FMOs), and other vibrational analyses. The computational studies likely docking analysis was also performed using bioinformatics tools to check the molecule's repressive behaviour against lung cancer-associated targeted proteins. Computer-Aided Drug Design (CADD) has emerged as an effective method for determining potential lead compounds and improving the development of new treatments for various diseases. Molecular dynamics simulations determine the stability of protein-ligand complexes. Furthermore, *in vitro* cytotoxicity testing was carried out to confirm the anti-lung cancer activity.

2 Materials and methods

2.1 Quantum chemical calculations

The most stable molecular structure of lochnericine was optimized using the Gaussian 09 software, and the DFT/B3LYP method with LANL2DZ was used to optimize the structure. The Los Alamos National Laboratory 2 double zeta (LANL2DZ) for transition metals and all-electron basis sets for all other non-transition metal atoms are used more frequently in computations on systems including transition metals. (Frisch et al., 2009). The molecule's vibrational wavenumbers were computed and assigned using the VEDA 4.0 software (Jamroz, 2004). The optimized molecular structure, vibrational wavenumbers, FMOs, MEP surface, and Mulliken atomic charge distributions of Lochnericine molecule were visualized using GaussView 05 software (Dennington et al., 2009). All the quantum chemical calculations were performed without regard for the potential energy surface to grasp the interception nature of the Lochnericine molecule.

2.2 Bioinformatics study

2.2.1 Target and ligand preparation

In the present study, target receptor molecules, including EGFR [PDB ID: 2ITY], KRAS [PDB ID: 7LGI], Ntrk [PDB ID: 7VKO], and ALK [PDB ID: 2XP2], were obtained from the RCSB Protein Data Bank (Sussman et al., 1998). TRIM11 protein sequences were obtained from UniProt (UniProt ID: Q96F44) (UniProt, 2019). The 3D protein structure of the other protein, TRIM11, was modelled using the Swiss model (Schwede et al., 2003). Procheck in the SAVES server was used to verify the three-dimensional protein structure (Laskowski et al., 2016). All water molecules and irregular residues were purified from the primary protein structure. In the PyRx workspace folders, the protein receptor was saved as a PDBQT file after adding missing hydrogens and charges. Using the PubChem database, the lochnericine molecule was downloaded. The ligand was imported and prepared. These ligands were then converted to Auto Dock Ligand format (PDBQT).

2.2.2 Molecular docking

Molecular docking was performed using Vina version 2.0 in PyRX (Dallakyan and Olson, 2015). In the PyRX interface, rigid docking was carried out for the targeted receptor molecules. The ligand and proteins were then subjected to docking to get the binding affinity with each other. Lamarckian genetic algorithm conformational search with the default parameters was utilized. Further, the grid on the targeted protein's ligand-binding site was situated in the centre of the binding site. Protein-lead molecule interactions were analyzed using Discovery Studio Visualize software.

2.2.3 ADME prediction

ADME prediction was performed in the SWISS ADME tool (Tripathi et al., 2019). This tool can forecast the pharmacokinetics and drug-like characteristics of active compounds. The numbers of rotatable bonds, hydrogen bond donors, acceptors, and molecular weight, were also predicted. The log *p* values represent the lipophilicity of the compounds. The logarithmic *S* value represents the solubility of water. Human gastrointestinal absorption (HIA) and blood-brain

barrier (BBB) permeation were predicted pharmacokinetic properties in swiss adme.

2.2.4 Molecular dynamics simulation

In simulation studies using molecular dynamics (MD), the atomic and molecular motions within the protein structure were examined. Atoms and molecules can interact for 100 ns in the GROMACS (GROningenMACHINE for Chemical Simulations) (Van Der Spoel et al., 2005). Macromolecular structure-to-function interactions may now be efficiently understood using an advanced technology known as molecular dynamics simulations (Awan et al., 2021; Bima et al., 2022). MD simulations were performed by using the GROMOS96 43a1 force field. Targeted receptor molecule and ligand complex topology was created by parametrizing the compound through the PRODRG server (Schüttelkopf and Van Aalten, 2004; Sangavi and Langeswaran, 2021). The simple point charge (SPC216) water model's step for salvation is the cubic box centre to edges. To balance the MD simulations, counter ions (Na^+ and Cl^-) were also included in the simulation systems (Shaik et al., 2022). Adding counter-ions to the system to neutralize it. Two methods exist for adding ions: Solve the solution and then add ions to replace the solvent's randomly distributed molecules. To solve a problem, distribute ions in accordance with the macromolecule's electrostatic potential. The simulation box interacts with an endless number of its periodic images when periodic boundary conditions are utilized, and the Coulomb energy is calculated using grid-based techniques. As a result, if a simulation system is charged, the electrostatic energy will accumulate to infinity. We must neutralize the system by introducing counter-ions to address this problem and allow simulations to determine the electrostatic energy accurately. Using MD simulations at the atomistic level has the advantage of allowing for capturing and evaluating various motional contributions to the overall complex dynamics of the targeted receptor molecule. Potential steric obstacles between the atoms in the solvated system were removed using the energy minimization step and the 1,500 steepest descent followed by conjugate gradient techniques. The NVT and NPT ensemble's two-step equilibration was carried out for 100 ps at constant volume, with gradual heating from 0 to 300 K, and at a pressure of 1 atm. The temperature is maintained with a Berendsen thermostat. The Parrinello-Rahman barostat is used to keep the system's pressure constant. The simulation of each system took place for 100 ns? Xmgrace software was used to examine the flexibility and stability of the protein and ligand complex (Turner, 2005).

2.3 Cell line studies

2.3.1 Materials and drug preparation

Sigma, St Louis, MO, United States, supplied the dimethyl sulfoxide (DMSO) and Lochnericine. All other chemicals and solvents were purchased from SISCO Research Laboratories (SRL). Lochnericine was synthesized in a serum-free RPMI medium after being dissolved in dimethyl sulfoxide (DMSO) at a final concentration of less than 0.1 percent (v/v). It did not affect cell viability) and filtered through a 0.045 mm syringe filter before being stored at 4°C.

2.3.2 Cell proliferation assay by MTT method

The MTT assay was employed to assess the anti-proliferative efficacy of lochnericine against A549 lung cancer cells. Through mitochondrial enzymes, the translation of MTT (3-(4,5-

dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide) to the formation of MTT is the critical mechanism involved in this cytotoxicity assay. NCCS Pune supplied the lung cancer cell line A549 for this investigation. In 24 well tissue culture plates, 2×10^5 cells were cultured with 0.5 mL medium/well, 37°C temperature was maintained with CO_2 5%, and relative humidity was about 95%. For 48 h, cells were treated with various concentrations of lochnericine. The cells were incubated at 37°C for 4 h after adding 200 μL of MTT solution (5 mg/mL), and cells were harvested. Through the gentle aspiration method, the medium and unread MTT were detached. At room temperature, the plates were shaken well for 5 min by adding 500 μL of DMSO. The crystallized dye was extracted by adding 1 mL of DMSO to the well. At 570 nm absorbance, the quantity of blue dye formed was estimated. The proliferation of the cells was examined by microscopic visualization at 48X magnification. The control and drug-treated cells were treated with 10% methanol for 5 min. Then cells were viewed under an inverted microscopeto confirm the anti-proliferative efficacy of the drug.

2.4 Statistical analysis

The mean and standard deviation data were displayed (SD). The noteworthyvariances were calculated using a one-way analysis of variance (ANOVA). A commercial software application (SPSS version 20) was employed for statistical analysis. SPSS offers data analysis for descriptive and bivariate statistics, numerical result forecasts, and predictions for classifying groups.

3 Results and discussion

3.1 Molecular structure analysis

The Lochnericine molecule's molecular structure was optimized using the DFT/B3LYP method with the cc-pVTZ basis set (Valarmathi et al., 2020a). Figure 1 depicts the optimized molecular structure of the lochnericine molecule, and its energy value was determined to be 1,111.43 a. u. Bond angle, bond length, and dihedral calculations collectively known as lochnericinesstructural parameters were also displayed in Table 1. The C1 point group has been noticed in the molecular geometry of the lochnericine molecule. Lochnericine has been pointed to as a centrosymmetric structure due to its IR and Raman active vibrational modes. The optimized molecular structure of the Lochnericine molecule is located at a local minimum on the potential energy surface due to the absence of negative vibrational wavenumbers (Valarmathi et al., 2020b).

3.2 Vibrational spectral analysis

49 atoms and 141 normal vibration modes are present in the lochnericine, and all belong to the same symmetry species. The calculated values oflochnericinevibrational frequencies, IR intensity, and activity of Raman scattering have been represented in Table 2. C1 point group symmetry exists in the Lochnericine molecule. Figure 2 depicts the theoretically simulated infrared and Raman spectra of the

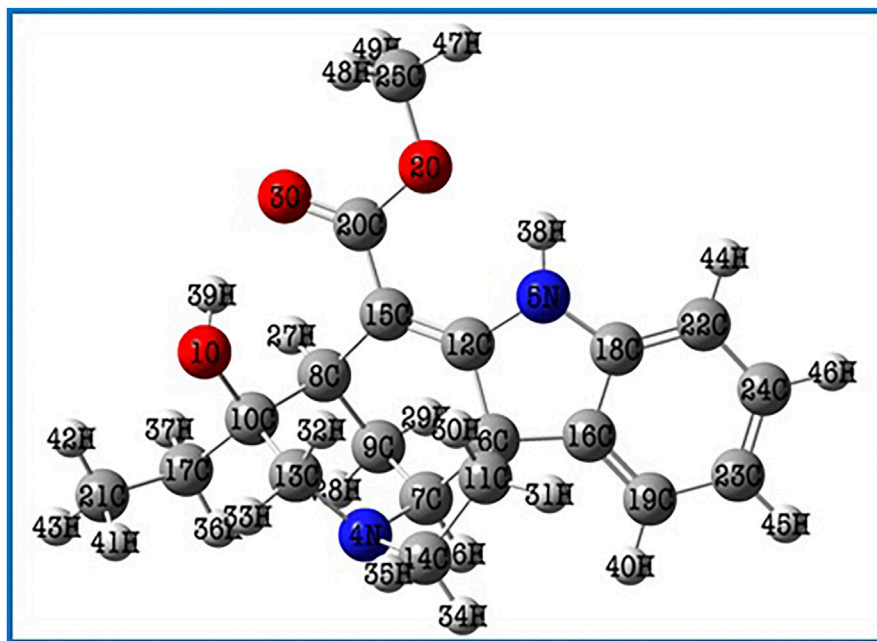


FIGURE 1
The optimized molecular structure of the Lochnericine molecule.

Lochnericine bioactive compound. The experimentally determined vibrational wavenumbers and theoretically calculated vibrational wavenumber values agreed well (Areans et al., 1988; Karabacak et al., 2008; Almansouret et al., 2021). The molecule's vibrational frequencies were calculated using the VEDA 4.0 programme, which many researchers have recognized as applicable (Zhou et al., 2021).

3.3 Frontier molecular orbitals (FMOs) analysis

The highest occupied molecular orbital (HOMO) and the lowest unoccupied molecular orbital (LUMO) are the two molecular orbitals that are used to analyse how a molecule interacts with other species (LUMO). The HOMO energy represents the capacity to provide electrons, whereas the LUMO energy represents the capacity to carry electrons (Koyambo-Konzapa et al., 2021). Electric and optical properties, UV-Vis spectra, and quantum chemistry all depend on FMOs. (Babu et al., 2016). The molecule's kinetic stability and chemical reactivity—both crucial components in determining its electrical properties—are represented by the HOMO-LUMO gap. Small HOMO-LUMO gaps are linked to soft molecules that have strong chemical reactivity and limited kinetic stability. Figure 3 depicts the Lochnericine molecule's FMOs. The FMOs' related molecular features were calculated using Koopman's theory (Mohamed Asath et al., 2017). Red and green are used to represent the positive and negative phases, respectively, in Figure 3. The calculated low energy gap value (4.18 eV), which also explains the intramolecular charge transfer interaction that affects the lochnericine molecule's biological activity, supports the increased chemical reactivity of the compound (Saravanan and Blachandran, 2014). The electron affinity A) is the term used to describe the energy

produced when an electron is introduced to an empty orbital, while the ionisation energy I) (5.49 eV) is used to describe the energy necessary to remove an electron from a full orbital (1.31 eV). The molecule is susceptible to electrophilic and nucleophilic reactions due to its anticipated high ionisation energy and low electron affinity. The chemical potential $\mu = 3.40$ eV, global softness, $S = 2.09$ eV, $\mu = 3.40$ eV, global hardness, $\eta = 0.47$ eV, and global electrophilicity index, $\psi = 12.29$ eV, of the molecule were also computed. Based on the anticipated values of higher hardness and reduced softness, the molecule is probably stable. The molecule's predicted chemical potential and electrophilicity index provide additional evidence that it has chemical stability comparable to that of compounds with potential bioactivity.

3.4 MEP surface analysis

Red, yellow, light blue, and blue on the MEP surface (Figure 4) represent, respectively, regions that are slightly electron-rich, slightly electron-rich, slightly electron-deficient, and slightly electron-deficient. It was revealed that the oxygen atoms lone pair electrons cause the regions around them to be electron-rich (red). Moreover, the oxygen atom O1 has a stronger electronegative potential, and all of the other oxygen atoms are classified as being in the slightly electron-rich zone. The slightly electron-deficient (light blue) portion of the molecule contained all of the hydrogen atoms. The molecule's electron-deficient (blue) portion is identified as hydrogen atom H38. The hydrogen atom H38, which is shown as being electron-poor (blue), makes up the molecule. The neutral electrostatic potential envelopes were positioned around the molecule (green). Potential electrophilic and nucleophilic attack sites are the hydrogen atom H38 and the oxygen atom O1.

TABLE 1 The optimized structural parameters of Lochnericine molecule.

Structural parameters	Cc-pVTZ	Structural parameters	Cc-pVTZ	Structural parameters	Cc-pVTZ
Bond length (Å)		Bond length (Å)		Bond angle (degree)	
O1-C10	1.46	C19-H40	1.08	C7-C9-H29	109.73
O1-H39	0.98	C21-H41	1.09	C8-C9-H28	111.07
O2-C20	1.39	C21-H42	1.09	C8-C9-H29	110.09
O2-C25	1.47	C21-H43	1.09	H28-C9-H29	107.54
O3-C20	1.25	C22-C24	1.40	O1-C10-C8	109.90
N4-C7	1.49	C22-H44	1.08	O1-C10-C13	105.09
N4-C13	1.50	C23-C24	1.39	O1-C10-C17	108.86
N4-C14	1.48	C23-H45	1.08	C8-C10-C13	108.18
N5-C12	1.38	C24-H46	1.08	C8-C10-C17	111.52
N5-C18	1.41	C25-H47	1.08	C13-C10-C17	113.05
N5-H38	1.01	C25-H48	1.09	C6-C11-C14	103.66
C6-C7	1.57	C25-H49	1.09	C6-C11-H30	110.59
C6-C11	1.59	Bond Angle (Degree)		C6-C11-H31	110.84
C6-C12	1.53	C10-O1-H39	110.44	C14-C11-H30	111.06
C6-C16	1.52	C20-O2-C25	116.16	C14-C11-H31	112.93
C7-C9	1.52	C7-N4-C13	117.00	H30-C11-H31	107.75
C7-H26	1.09	C7-N4-C14	102.38	N5-C12-C6	106.95
C8-C9	1.54	C13-N4-C14	108.96	N5-C12-C15	128.50
C8-C10	1.55	C12-N5-C18	110.11	C6-C12-C15	124.10
C8-C15	1.53	C12-N5-H38	120.30	N4-C13-C10	114.67
C8-H27	1.09	C18-N5-H38	125.83	N4-C13-H32	111.81
C9-H28	1.09	C7-C6-C11	102.68	N4-C13-H33	107.24
C9-H29	1.09	C7-C6-C12	111.08	C10-C13-H32	105.28
C10-C13	1.53	C7-C6-C16	122.69	C10-C13-H33	110.43
C10-C17	1.54	C11-C6-C12	111.67	H32-C13-H33	107.18
C11-C14	1.55	C11-C6-C16	108.00	N4-C14-C11	104.61
C11-H30	1.09	C12-C6-C16	100.79	N4-C14-H34	109.25
C11-H31	1.09	N4-C7-C6	106.00	N4-C14-H35	110.85
C12-C15	1.37	N4-C7-C9	112.28	C11-C14-H34	110.66
C13-H32	1.09	N4-C7-H26	106.78	C11-C14-H35	113.12
C13-H33	1.09	C6-C7-C9	111.91	H34-C14-H35	108.25
C14-H34	1.09	C6-C7-H26	108.64	C8-C15-C12	119.26
C14-H35	1.09	C9-C7-H26	110.92	C8-C15-C20	116.13
C15-C20	1.44	C9-C8-C10	109.46	C12-C15-C20	123.79
C16-C18	1.41	C9-C8-C15	108.53	C6-C16-C18	107.52
C16-C19	1.38	C9-C8-H27	109.55	C6-C16-C19	132.32
C17-C21	1.53	C10-C8-C15	115.79	C18-C16-C19	119.76
C17-H36	1.09	C10-C8-H27	106.98	C10-C17-C21	113.94

(Continued on following page)

TABLE 1 (Continued) The optimized structural parameters of Lochnericine molecule.

Structural parameters	Cc-pVTZ	Structural parameters	Cc-pVTZ	Structural parameters	Cc-pVTZ
Bond length (Å)		Bond length (Å)		Bond angle (degree)	
C17-H37	1.10	C15-C8-H27	106.34	C10-C17-H36	109.13
C18-C22	1.39	C7-C9-C8	108.09	C10-C17-H37	108.11
C19-C23	1.40	C7-C9-H28	110.30	C21-C17-H36	109.20
C21-C17-H37	109.00	C13-N4-C7-H26	-166.63	C16-C6-C11-H30	116.61
H36-C17-H37	107.21	C14-N4-C7-C6	-41.38	C16-C6-C11-H31	-2.83
N5-C18-C16	108.93	C14-N4-C7-C9	-163.88	C7-C6-C12-N5	-155.20
N5-C18-C22	129.02	C14-N4-C7-H26	74.32	C7-C6-C12-C15	17.83
C16-C18-C22	121.99	C7-N4-C13-C10	41.81	C11-C6-C12-N5	90.80
C16-C19-C23	119.02	C7-N4-C13-H32	-77.94	C11-C6-C12-C15	-96.15
C16-C19-H40	120.87	C7-N4-C13-H33	164.83	C16-C6-C12-N5	-23.67
C23-C19-H40	120.09	C14-N4-C13-C10	157.26	C16-C6-C12-C15	149.36
O2-C20-O3	119.93	C14-N4-C13-H32	37.51	C7-C6-C16-C18	143.01
O2-C20-C15	115.03	C14-N4-C13-H33	-79.70	C7-C6-C16-C19	-44.29
O3-C20-C15	125.02	C7-N4-C14-C11	45.781	C11-C6-C16-C18	-98.09
C17-C21-H41	111.68	C7-N4-C14-H34	-72.74	C11-C6-C16-C19	74.59
C17-C21-H42	110.30	C7-N4-C14-H35	168.03	C12-C6-C16-C18	19.12
C17-C21-H43	110.37	C13-N4-C14-C11	-78.76	C12-C6-C16-C19	-168.18
H41-C21-H42	107.91	C13-N4-C14-H34	162.70	N4-C7-C9-C8	55.58
H41-C21-H43	107.44	C13-N4-C14-H35	43.48	N4-C7-C9-H28	-66.03
H42-C21-H43	109.01	C18-N5-C12-C6	20.30	N4-C7-C9-H29	175.65
C18-C22-C24	117.59	C18-N5-C12-C15	-152.32	C6-C7-C9-C8	-63.50
C18-C22-H44	121.51	H38-N5-C12-C6	179.74	C6-C7-C9-H28	174.88
C24-C22-H44	120.88	H38-N5-C12-C15	7.11	C6-C7-C9-H29	56.57
C19-C23-C24	120.46	C12-N5-C18-C16	-7.65	H26-C7-C9-C8	174.97
C19-C23-H45	119.71	C12-N5-C18-C22	170.06	H26-C7-C9-H28	53.36
C24-C23-H45	119.81	H38-N5-C18-C16	-165.68	H26-C7-C9-H29	-64.94
C22-C24-C23	121.13	H38-N5-C18-C22	12.02	C10-C8-C9-C7	-65.36
C22-C24-H46	119.16	C11-C6-C7-N4	20.63	C10-C8-C9-H28	55.77
C23-C24-H46	119.69	C11-C6-C7-C9	143.37	C10-C8-C9-H29	174.78
O2-C25-H47	104.89	C11-C6-C7-H26	-93.80	C15-C8-C9-C7	61.88
O2-C25-H48	110.06	C12-C6-C7-N4	-98.87	C15-C8-C9-H28	-176.97
O2-C25-H49	110.21	C12-C6-C7-C9	23.85	C15-C8-C9-H29	-57.96
H47-C25-H48	111.17	C12-C6-C7-H26	146.67	H27-C8-C9-C7	177.61
H47-C25-H49	111.15	C16-C6-C7-N4	142.03	H27-C8-C9-H28	-61.24
H48-C25-H49	109.27	C16-C6-C7-C9	-95.23	H27-C8-C9-H29	57.76
Dihedral Angle (Degree)		C16-C6-C7-H26	27.59	C9-C8-C10-O1	175.37
C25-O2-C20-O3	2.73	C7-C6-C11-C14	6.67	C9-C8-C10-C13	61.13
C25-O2-C20-C15	-176.46	C7-C6-C11-H30	-112.43	C9-C8-C10-C17	-63.79

(Continued on following page)

TABLE 1 (Continued) The optimized structural parameters of Lochnericine molecule.

Structural parameters	Cc-pVTZ	Structural parameters	Cc-pVTZ	Structural parameters	Cc-pVTZ
Bond length Å)		Bond length Å)		Bond angle (degree)	
C20-O2-C25-H47	177.04	C7-C6-C11-H31	128.11	C15-C8-C10-O1	52.33
C20-O2-C25-H48	57.37	C12-C6-C11-C14	125.77	C15-C8-C10-C13	-61.90
C20-O2-C25-H49	-63.22	C12-C6-C11-H30	6.67	C15-C8-C10-C17	173.16
C13-N4-C7-C6	77.6534	C12-C6-C11-H31	-112.78	H27-C8-C10-O1	-65.99
C13-N4-C7-C9	-44.84	C16-C6-C11-C14	-124.27	H27-C8-C10-C13	179.76
H27-C8-C10-C17	54.83	C6-C11-C14-N4	-32.14	C10-C17-C21-H41	-65.10
C9-C8-C15-C12	-22.20	C6-C11-C14-H34	85.41	C10-C17-C21-H42	54.90
C9-C8-C15-C20	147.91	C6-C11-C14-H35	-152.90	C10-C17-C21-H43	175.43
C10-C8-C15-C12	101.32	H30-C11-C14-N4	86.63	H36-C17-C21-H41	57.20
C10-C8-C15-C20	-88.56	H30-C11-C14-H34	-155.80	H36-C17-C21-H42	177.22
H27-C8-C15-C12	-139.99	H30-C11-C14-H35	-34.12	H36-C17-C21-H43	-62.24
H27-C8-C15-C20	30.11	H31-C11-C14-N4	-152.18	H37-C17-C21-H41	174.04
O1-C10-C13-N4	-165.87	H31-C11-C14-H34	-34.62	H37-C17-C21-H42	-65.93
O1-C10-C13-H32	-42.53	H31-C11-C14-H35	87.05	H37-C17-C21-H43	54.59
O1-C10-C13-H33	72.84	N5-C12-C15-C8	152.85	N5-C18-C22-C24	-176.69
C8-C10-C13-N4	-48.49	N5-C12-C15-C20	-16.45	N5-C18-C22-H44	2.67
C8-C10-C13-H32	74.83	C6-C12-C15-C8	-18.62	C16-C18-C22-C24	0.75
C8-C10-C13-H33	-169.78	C6-C12-C15-C20	172.06	C16-C18-C22-H44	-179.87
C17-C10-C13-N4	75.52	C8-C15-C20-O2	-165.54	C16-C19-C23-C24	1.35
C17-C10-C13-H32	-161.14	C8-C15-C20-O3	15.30	C16-C19-C23-H45	-178.84
C17-C10-C13-H33	-45.76	C12-C15-C20-O2	4.07	H40-C19-C23-C24	-179.03
O1-C10-C17-C21	-44.95	C12-C15-C20-O3	-175.07	H40-C19-C23-H45	0.75
O1-C10-C17-H36	-167.30	C6-C16-C18-N5	-8.39	C18-C22-C24-C23	-0.37
O1-C10-C17-H37	76.39	C6-C16-C18-C22	173.69	C18-C22-C24-H46	179.35
C8-C10-C17-C21	-166.38	C19-C16-C18-N5	177.82	H44-C22-C24-C23	-179.74
C8-C10-C17-H36	71.25	C19-C16-C18-C22	-0.08	H44-C22-C24-H46	-0.02
C8-C10-C17-H37	-45.04	C6-C16-C19-C23	-172.94	C19-C23-C24-C22	-0.67
C13-C10-C17-C21	71.44	C6-C16-C19-H40	7.45	C19-C23-C24-H46	179.60
C13-C10-C17-H36	-50.91	C18-C16-C19-C23	-0.97	H45-C23-C24-C22	179.52
C13-C10-C17-H37	-167.21	C18-C16-C19-H40	179.41	H45-C23-C24-H46	-0.19

3.5 Mulliken atomic charge distribution analysis

Mulliken's atomic charge distribution influences the molecule's dipole moment, electrostatic potential, electronegativity equalization, electronic structure, vibrational modes, and polarizability (Geetha et al., 2021). Because of its attachment to the two electronegative oxygen atoms, the carbon atom C20 has a higher positive charge value, as shown in Figure 5 representation of the predicted Mulliken atomic charge distribution values (O2 and O3). The highest electronegative charge

values are found in the oxygen atom O1 and the nitrogen atom N5. The carbon atoms' negative charge values significantly promote electron delocalization within the molecule.

3.6 Molecular docking

In drug designing, molecular docking is essential to determine the binding mode and efficacy of the active lead molecules. Molecular docking approaches were performed for all targeted

TABLE 2 Theoretical IR and Raman frequencies and their assignments for Lochnericine molecule.

Mode	Wavenumber (cm ⁻¹)	Assignment	Mode	Wavenumber (cm ⁻¹)	Assignment
1	3627	ν NH (100)	41	1456	δ CH ₂ (28) + β OH (14)
2	3506	ν OH (100)	42	1450	δ CH ₂ (28) + β OH (13)
3	3227	ν CH (99)	43	1446	β CH (18) + β NH (12)
4	3214	ν CH (98)	44	1421	ω CH ₂ (29)
5	3201	ν CH (97)	45	1416	β CH (23) + ω CH ₂ (18)
6	3200	ν_{as} CH ₃ (97)	46	1402	β CH (22) + ω CH ₂ (13)
7	3193	ν CH (97)	47	1392	β CH (22) + ω CH ₂ (12)
8	3172	ν_{as} CH ₃ (95)	48	1385	β CH (38)
9	3156	ν_{as} CH ₃ (94)	49	1379	β CH (23) + β NH (11)
10	3145	ν_{as} CH ₂ (96)	50	1377	β CH (28) + β NH (10)
11	3135	ν_{as} CH ₂ (96)	51	1367	β CH (14) + ω CH ₂ (26)
12	3129	ν_{as} CH ₂ (97)	52	1352	β CH (27) + ω CH ₂ (13)
13	3113	ν_{as} CH ₃ (96)	53	1339	β CH (18) + ω CH ₂ (28)
14	3104	ν_{as} CH ₂ (95)	54	1330	t CH ₂ (29)
15	3087	ν_s CH ₂ (94)	55	1322	β CH (27) + β NH (20)
16	3085	ν_{as} CH ₂ (93)	56	1301	β CH (23) + t CH ₂ (17)
17	3084	ν_s CH ₃ (94)	57	1297	t CH ₂ (32)
18	3072	ν_s CH ₂ (93)	58	1279	β NH (22) + t CH ₂ (16)
19	3057	ν CH (92)	45	1262	β NH (26) + t CH ₂ (13)
20	3053	ν CH (92)	60	1259	β NH (33) + t CH ₂ (12)
21	3051	ν CH (93)	61	1237	β CH (29) + β OH (22)
22	3048	ν_s CH ₃ (92)	62	1227	β CH (25) + t CH ₂ (12)
23	3024	ν CH (92)	63	1217	β CH (22) + t CH ₂ (10)
24	3021	ν_s CH ₂ (92)	64	1209	β CH (27) + β OH (18)
25	1678	ν CC (50) + β CH (32)	65	1202	β CH ₃ (36) + β CH (14)
26	1662	ν CC (39) + β NH (18)	66	1196	β CH ₃ (29) + β CH (12)
27	1652	ν CC (39) + β CH (17) + β NH (11)	67	1183	β CH (27) + β NH (12)
28	1640	ν C=O (87)	68	1175	β CH ₃ (23) + β CH (11)
29	1560	δ CH ₂ (39)	69	1165	β CH ₃ (32)
30	1555	β CH ₃ (37)	70	1155	ρ CH ₂ (23) + β CH (12)
31	1544	δ CH ₂ (42)	71	1145	ρ CH ₂ (24) + β CH (11)
32	1543	δ CH ₂ (36)	72	1139	ρ CH ₂ (29) + ω CH ₂ (22)
33	1539	β CH ₃ (39)	73	1107	ρ CH ₂ (28) + t CH ₂ (18)
34	1537	β CH ₃ (41)	74	1099	ρ CH ₂ (19) + t CH ₂ (14)
35	1535	δ CH ₂ (29)	75	1084	ρ CH ₂ (23) + β CH ₃ (13)
36	1531	β CH (32)	76	1067	ρ CH ₂ (23) + β CH ₃ (13)
37	1527	δ CH ₂ (38)	77	1059	ρ CH ₂ (29) + β CH ₃ (11)
38	1518	β CH ₃ (46)	78	1052	Ring deformation
39	1517	β CH (38)	79	1046	ν CC (39) + ν CO (14)

(Continued on following page)

TABLE 2 (Continued) Theoretical IR and Raman frequencies and their assignments for Lochnericine molecule.

Mode	Wavenumber (cm ⁻¹)	Assignment	Mode	Wavenumber (cm ⁻¹)	Assignment
40	1492	β CH ₃ (43)	80	1026	ν CC (25)+ β CH ₃ (12)
81	1008	γ CH (42)	112	528	γ OH (23)
82	1007	γ CH (38)	113	509	ρ CH ₂ (23)+ γ OH (13)
83	996	γ CH (34)+ν CC (23)+ ν CO (11)	114	491	τ Ring
84	976	γ CH (37)	115	476	ρ CH ₂ (28)+ γ CH (18)
85	969	γ CH (29)+ β NH (11)	116	453	ρ CH ₂ (29)+ γ CH (11)
86	957	γ CH (29)	117	438	ρ CH ₂ (12)+ γ OH (23)
87	945	γ CH (39)+ν CC (21)	118	416	ρ CH ₂ (32)
88	942	γ CH (37)+ ν CC (18)	119	395	γ OH (23)
89	905	t CH ₂ (19) + β NH (10)	120	362	τCCCH (12)+ τCCCC (13)
90	894	γ CH (27)	121	355	τCCCC (13)+ τCCCN (12)
91	890	ρ CH ₂ (26)+ν CO (10)	122	328	τCCCH (13)+ τCCNO (11)
92	880	ρ CH ₂ (23)	123	321	τCCCC (13)+ τCCNO(11)
93	875	ρ CH ₂ (25) + ν CC (12)	124	307	τCCCH (12)+ τCCCC (13)
94	853	Ring Breathing	125	303	τ Ring
95	824	ρ CH ₂ (23)+ t CH ₂ (13)	126	285	τ Ring
96	804	Ring Breathing	127	267	τCCCH(12)+ τCCCC(13)
97	792	γ CH (38)	128	251	τCCCH(12)+ τCCCC(12)
98	783	γ CH (27)+ γ NH (28)	129	228	ρ CH ₂ (23)
99	776	γ CH (29)	130	219	τ CH ₃ (32)
100	745	ρ CH ₂ (16) + β NH (12)	131	179	τ CH ₃ (34)
101	740	γ NH (29)	132	153	τCCCH(12)+ τCCCC(13)
102	727	γ NH (32)	133	147	τCOCO(11)+ τCCCC(11)
103	706	Ring deformation	134	132	Ring deformation
104	687	Ring deformation	135	118	Ring deformation
105	663	Ring deformation	136	103	τ CH ₃ (23)
106	656	ρ CH ₂ (21) + β NH (13)	137	94	τ CH ₃ (12)
107	628	Ring deformation	138	77	Ring deformation
108	600	ρ CH ₂ (26) + β OH (12)	139	70	τ CH ₃ (14)
109	594	τ Ring	140	67	τCOCO(12)+ τCCCC(11)
110	582	γ CH (38)+ β CH ₃ (18)	141	55	τCOCO(12)+ τCCCC(10)
111	537	γ OH (23)			

ν, Stretching; ν_s, symmetric stretching; ν_{as}, asymmetric stretching; β, in plane bending; γ, out of plane bending; ρ, rocking; ω, Wagging; δ, Scissoring; t, twisting.

receptor molecules. Receptor molecules docking score and interacting residues were tabulated in Table 3. EGFR receptor (PDB ID 2ITY) molecule interacts with the active lead molecules that lead to the formation of hydrogen bonds, Pi-Pi stacked, and Pi-Alkyl. Lead molecule interacts with glycine residue at the position of 724 and forms a conventional hydrogen bond. Leu 747 residue forms Pi-Alkyl interaction. Phenylalanine 723 amino acid forms the Pi-Pi Stacked. EGFR_ Lochnericine has good interaction, and the docking

score is -7.52 Kcal/Mol. 2 days and 3 days interactions between EGFR_Lochnericine were depicted in Figure 6A. A docking protocol was carried out between KRAS and Lochnericine, and the docking score between KRAS and lead molecules is -9.10 Kcal/Mol. Arginine amino acid at the position of 97 interacts with small molecules and forms the hydrogen bond. Another hydrogen bond is formed between Lysine 101 and Lochnericine. Pi-Alkyl formed by the Tyrosine 137 amino acid residues and 2 days and 3 days

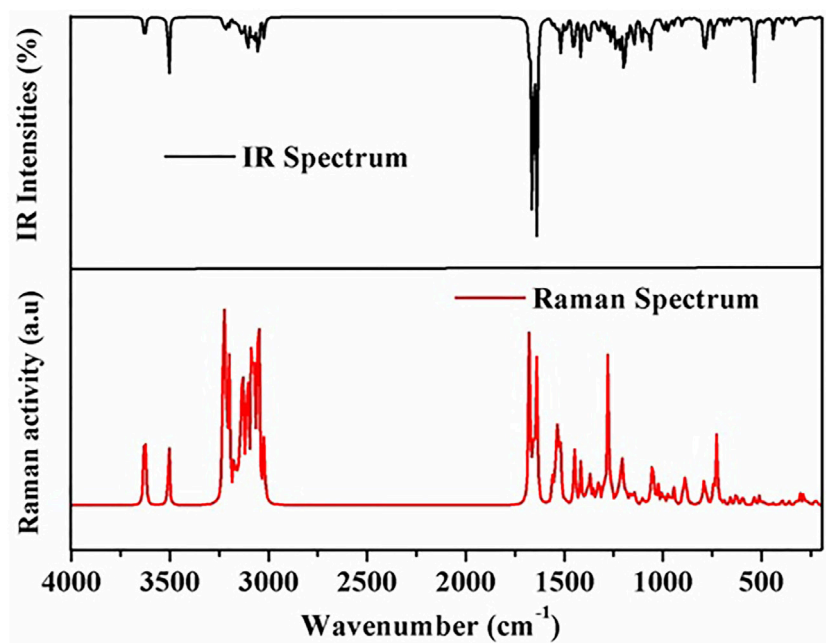


FIGURE 2
Theoretically simulated infrared and Raman spectra of the Lochnericine molecule.

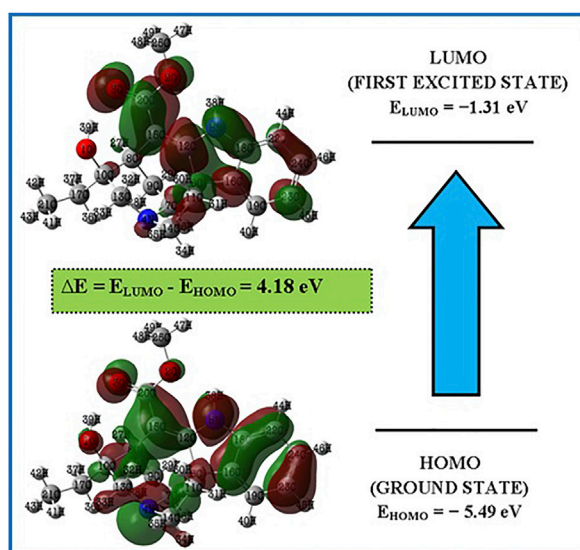


FIGURE 3
FMOs of the Lochnericine molecule.

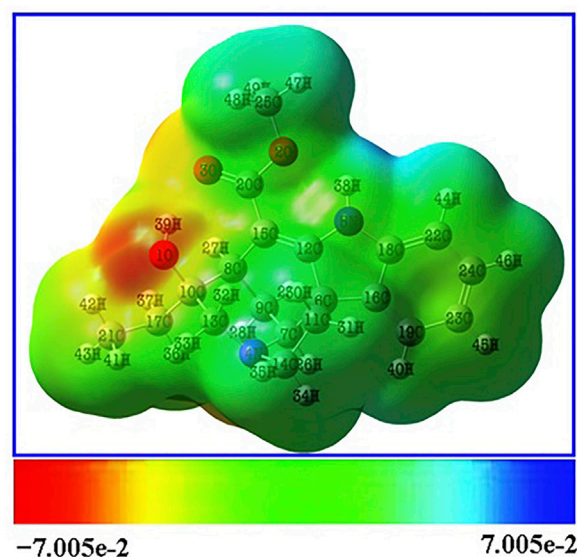


FIGURE 4
MEP surface of the Lochnericine molecule.

interactions between KRAS_Lochnericine were depicted in **Figure 6B**. Ntrk protein structure was downloaded from the RCSB PDB (PDB ID: 7VKO). The lead molecule with Glycine 667 amino acid residue forms the carbon-hydrogen bond. Further, Methionine 592 residue forms another carbon-hydrogen bond. Alkyl and Pi-Alkyl interactions were formed between lead molecule and amino acid residues (Valine 524, Tyrosine 591, Leucine 516, Alanine 542, Phenylalanine 589 and Lysine 544).

Ntrk_Lochnericine docking score is -8.24 Kcal/Mol, and the 2 days and 3 days interactions between Ntrk_Lochnericine are depicted in **Figure 6C**. ALK three-dimensional proteins with PDB ID: 2XP2 were downloaded and prepared for docking. Alkyl and Pi-Alkyl interactions are formed by amino acid residue Leu 1,256, Ala 1,148, Val 1,130, Ile 1,171 and Lys 1,150, Glu 1,167, Asn 1,254 and Arg 1,253 amino acid residue interact with Lochnericine forms carbon-hydrogen bond. The docking score is -11.59 Kcal/Mol.

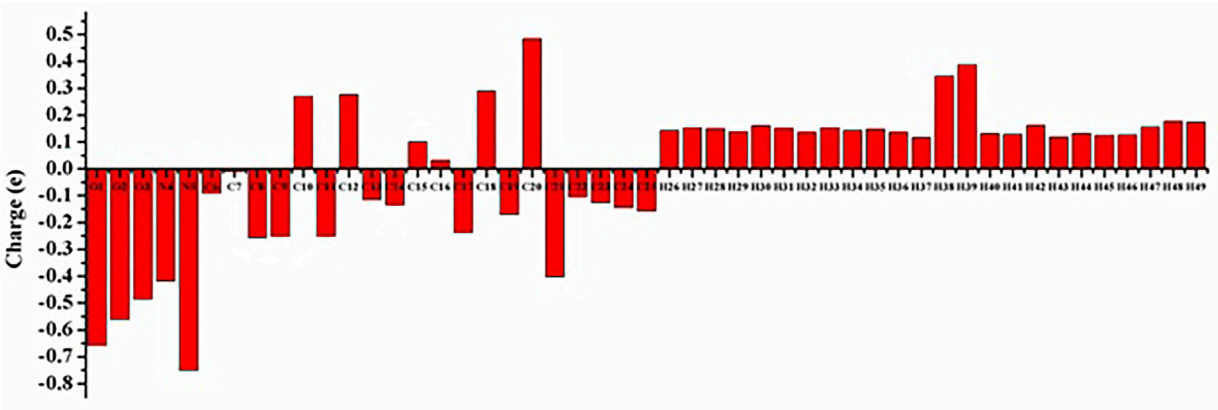


FIGURE 5
Mulliken atomic charge distribution of the Lochnericine molecule.

TABLE 3 Mulliken Atomic charges for optimized geometry.

Atom	Mulliken atomic charges	Atom	Mulliken atomic charges
O1	−0.6579	H26	0.1422
O2	−0.5613	H27	0.1517
O3	−0.4867	H28	0.1495
N4	−0.4188	H29	0.1378
N5	−0.7511	H30	0.1608
C6	−0.0910	H31	0.1503
C7	−0.0117	H32	0.1359
C8	−0.2577	H33	0.1530
C9	−0.2506	H34	0.1421
C10	0.2697	H35	0.1469
C11	−0.2508	H36	0.1354
C12	0.2759	H37	0.1159
C13	−0.1154	H38	0.3454
C14	−0.1352	H39	0.3875
C15	0.1008	H40	0.1318
C16	0.0315	H41	0.1284
C17	−0.2383	H42	0.1627
C18	0.2895	H43	0.1179
C19	−0.1709	H44	0.1311
C20	0.4833	H45	0.1241
C21	−0.4034	H46	0.1266
C22	−0.1034	H47	0.1551
C23	−0.1271	H48	0.1775
C24	−0.1444	H49	0.1734
C25	−0.1584		

2 days and 3 days interactions between ALK_Lochnericine were depicted in Figure 6D. TRIM11 protein was modelled, and protein structure was assessed using the Saves server. TRIM11 forms Pi-Alkyl interaction with phenylalanine 407 amino acid residue. Tyrosine 343 forms a carbon-hydrogen bond, and the docking score between TRIM11_Lochnericine is -7.13 Kcal/Mol. 2 days and 3 days interactions between TRIM11_Lochnericine were depicted in Figure 6E. The docking score and interacting residues between the targeted receptor molecules and lead compound were tabulated in Table 4.

3.7 ADME prediction

In the drug-designing process, ADME prediction is essential for determining the lead molecule's efficacy. These obey all the five rules of Lipinski. Several pharmacokinetic properties, pharmacokinetics, drug-likeness, and water solubility properties were tabulated in

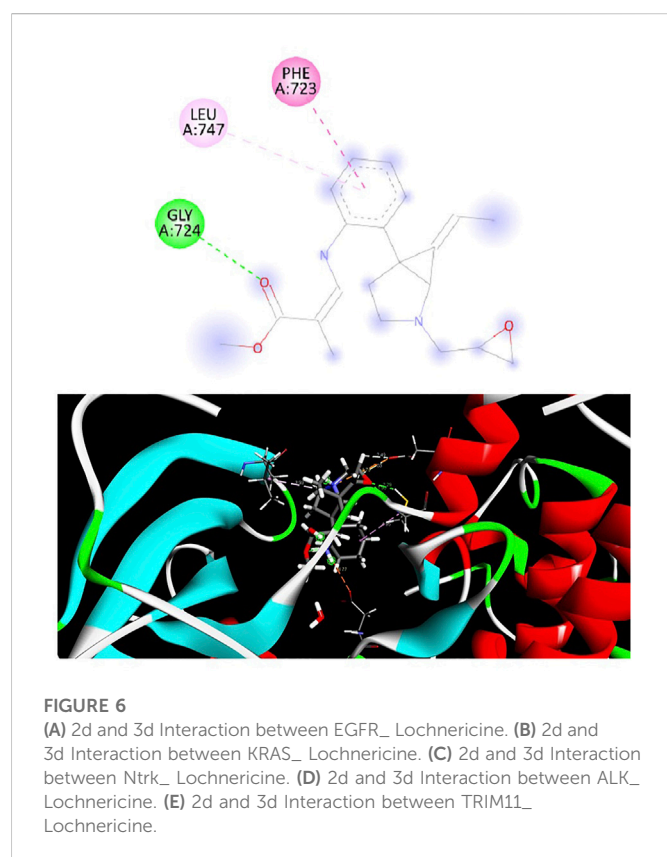


Table 5. Further, lead compound lipophilicity, size, insoluble, unsaturation, flexibility, and polar nature are shown in Figure 7.

3.8 Molecular dynamics simulation

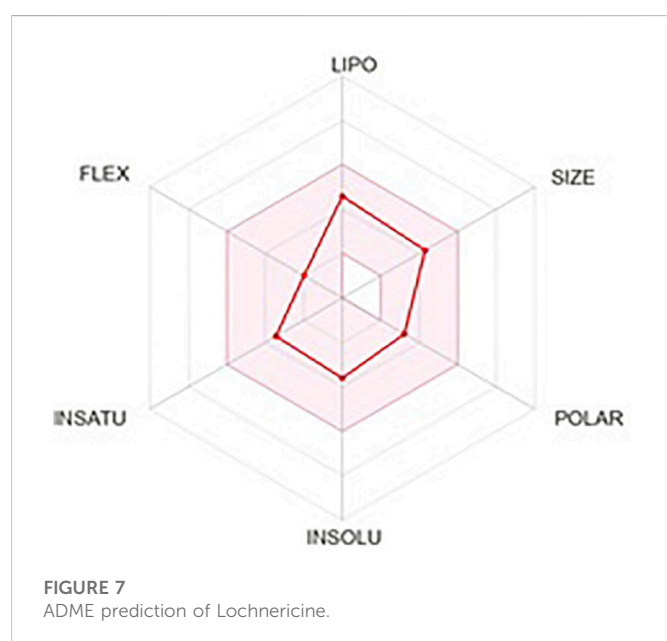
Gromacs package's molecular dynamics simulations were used to examine the stability of targeted protein-ligand complexes. Based on the docking score, interaction residues, and binding affinity between the targeted proteins and the Lochnericine. For 50 ns, molecular dynamics simulations were performed. Typically, MD simulation is used to predict the behaviour of macromolecules, and it relies on classical mechanics and Newton's equation of motion to determine the speed and position of each atom in the system under consideration. As a result, MD performs a more concise structural assessment. The difference between the initial and final structural conformational positions was determined using the RMSD (Root Mean Square Deviation). The results of MDs of protein-ligand complexes are shown in Figure 8. The average range of RMSD deviation was observed between all the protein-ligand molecules. This indicates a good RMSD deviation. All protein receptor molecules indicate favourable RMSD deviation throughout the entire simulation period. RMSF (Root Mean Square Fluctuation) assesses and computes the average deviation of a targeted protein over time from an amino acid residue reference position. The average fluctuation rate between targeted protein-ligand complexes ranges between all the protein-ligand molecules. The fluctuation between the targeted receptor molecules and Lochnericine is depicted in Figure 8. The smallest fluctuation with good stability and the fluctuation rate that occurs in the loop or disallowed region have no implications on the complexes' stability. Hydrogen bonds in protein-ligand complexes are calculated using cut-offs for the angle between the Hydrogen Donor-Acceptor and the distance between the Donor-Acceptor. Hydrogen bond interactions between targeted proteins and Lochnericine complexes are represented in Figure 8. EGFR (PDB ID-2ITY) with the lead compound represented in the black colour. 7LGI with the active compound represented in the green colour. The red colour indicates the 7VKO protein with lochnericine. TRIM11 with lochnericine represent in blue colour. Throughout the simulation runs, the GROMACS command line `gmx_rmsd` was used to compute the RMSD values of each targeted protein-ligand complex's stability profile. RMSD often interprets the deviation for a set of atoms (protein, ligand, or even ligand-protein complex) from the

TABLE 4 Interacting residues between receptor molecules and lead compound.

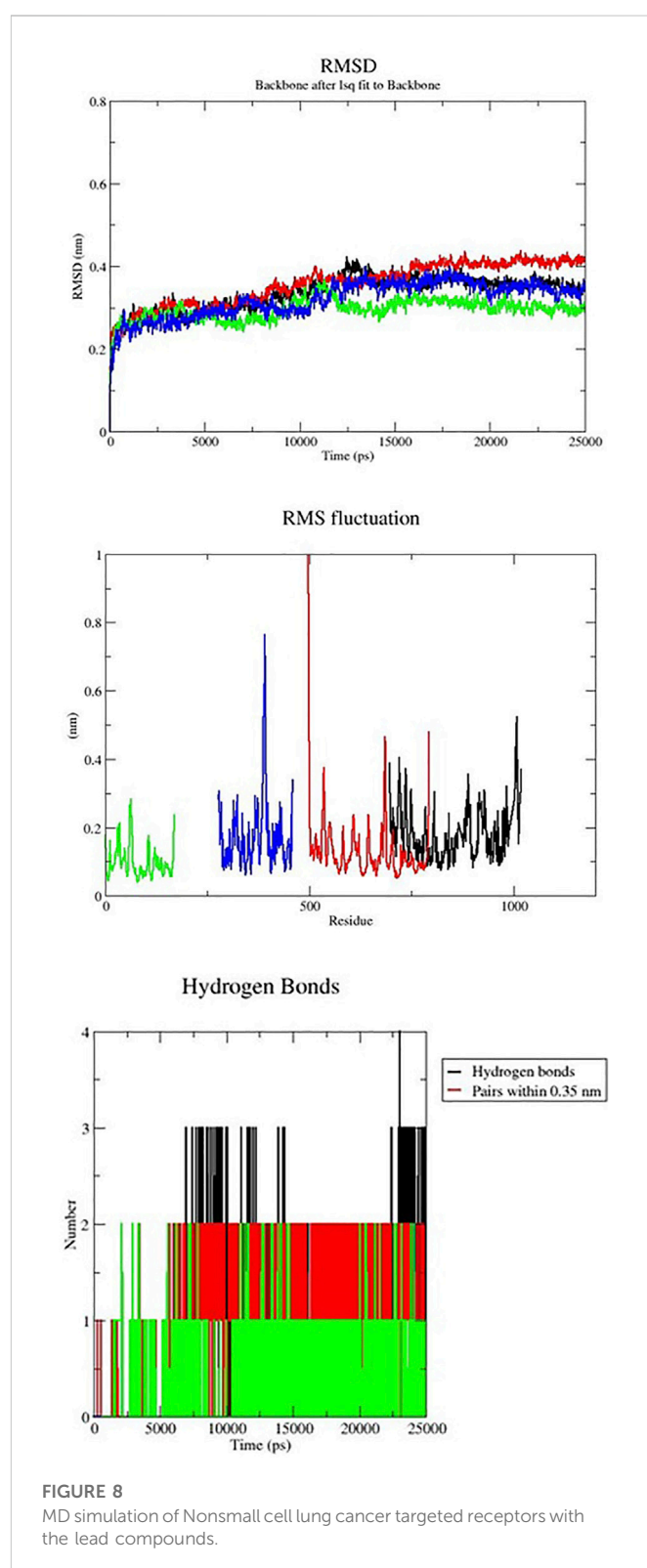
S.No	Protein_lead compound	Interacting residues	Docking score
1	EGFR_ Lochnericine	GLY_724, LEU_74 and PHE_723	-7.52 Kcal/Mol
2	KRAS_ Lochnericine	ARG_97, LYS_101 and TYR_137	-9.10 Kcal/Mol
3	Ntrk_ Lochnericine	TYR_591, VAL_524, LEU_516, MET_592, ALA_542, PHE_589, LYS_544, GLY_667 and ASP_668	-8.24 Kcal/Mol
4	ALK_Lochnericine	LEU_1256, VAL_1130, ALA_1148, ILE_1171, LYS_1150, ARG_1253, ASN_1254, GLU_1167 and ASP_1270	-11.59 Kcal/Mol
5	TRIM11_Lochnericine	PHE_407 and TYR_343	-7.13 Kcal/Mol

TABLE 5 ADME prediction of Lochnericine.

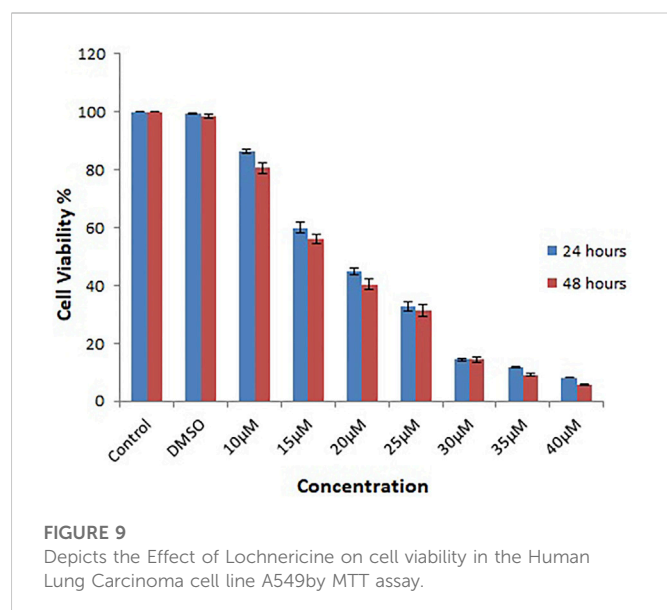
ADME prediction of lochnericine	
Number of rotatable bonds	3
Number of Hydrogen bond acceptors	4
Number of Hydrogen bond donor	1
Lipophilicity Log $P_{O/W}$	3.08
Log S	-3.59
GI absorption	High
BBB permeant	Yes
Bioavailability Score	0.55



corresponding initial reference structure. Less RMSD values would therefore be associated with considerable stability due to changes in the examined molecule's structure. Additionally, ligands show lower RMSD values corresponding to the ligand-protein complexes. A lower RMSD value denotes higher stability for the protein-ligand complexes. All targeted receptor molecules with lead compound lesser deviation within the range of 0.4 nm. All the potential receptors of Non-small cell lung cancer protein flexibility and target Residue Root-Mean-Square Fluctuation (RMSF) for each ligand-bound protein residue root-mean-square fluctuation (RMSF) profile were calculated in order to acquire more insights into the stability of the complex binding site. The GROMACS "gmx_rmsf" command line was used to calculate the specific backbone RMSF of each protein. This flexibility validation criterion shows the contribution of specific protein residues to the structural variations of the ligand/protein complex. The RMSF calculates the average deviation for each residue from its reference position inside the reduced initial structures over time. All the targeted receptor molecules show an acceptable fluctuation range except the disallowed region. The targeted protein-ligand



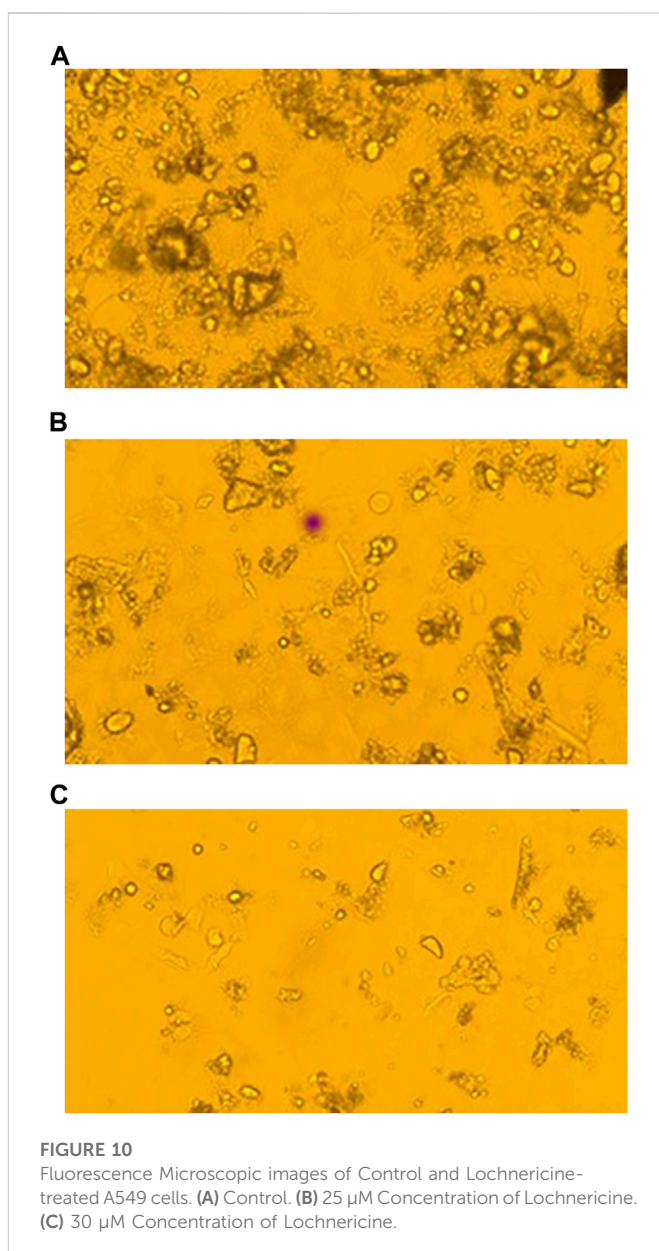
complex fluctuates within the range of 0.50 nm. The lowest values for the deviation below 0.40 nm and the lesser deviation below 1 nm were found while considering the examined complexes. The above RMSD and RMSF analysis shows that the targeted protein-ligand complexes have good stability and flexibility. The stability of ligand-protein complexes and the corresponding



conformational changes were thought to be best understood by examining the hydrogen bond network connections between the receptors and lochnericine. It was helpful to investigate the recognized ligand-protein hydrogen bond interactions and their relative frequency using the GROMACS command line `gmxbond`, which analyses the hydrogen bonds (H-bonds) between all conceivable donors D and acceptors A. EGFR receptor molecule exhibit a maximum number of 4 hydrogen bonds between the targeted receptor molecule was observed. All other receptors possess the average number of 3 hydrogen bonds between the receptor and lochnericine. All the resultant complexes show better hydrogen bond interaction throughout the simulation period. Furthermore, lochnericine induced a stabilized favoured hydrogen bond association with the targeted receptor, which was sustained for significant MD simulation. This preferential hydrogen bond pair interaction with the targeted receptors introduces the potential activity of lochnericine to block the receptor molecules of non-small-cell lung cancer.

3.9 Anti-proliferative estimation by MTT assay

Several chemicals and natural compounds' cytotoxic efficacy against malignant cells have been considered primary studies for the antitumor potential of the compounds. The conversion of MTT (3-(4, 5-dimethylthiazol-2-yl)-2, 5-diphenyltetrazolium bromide) to MTT-formation using mitochondrial enzymes in the MTT assay is a notable assay for the valuation of the proliferation of cancer cells (Sangavi et al., 2022). Treatment Lochnericine for 24 and 48 h declined cell proliferation in a dosage and time-dependent manner, as shown in Figure 9, 10. In the present study, Lochnericine condensed cell feasibility in a concentration-dependent manner. Lochnericine inhibition of cell growth might be linked to the induction of cell death. Thus, Lochnericine suppressive impact on A549 cells significantly supports Lochnericine's anti-proliferation capability.



4 Conclusion

The optimization geometry for the lochnericine bioactive molecule was carried out using the DFT/B3LYP functional technique and the cc-PVTZ basis set. Simulated Lochnericine vibrational spectra, including infrared and Raman spectra, showed that the computed vibrational wavenumbers significantly correspond with those from earlier works of literature. The determined band gap energy value of 4.18 eV abutment lochnericine bioactivity. H38 hydrogen and O1 oxygen atom in the molecule are possible electrophilic and nucleophilic attack sites, according to a molecular electrostatic potential surface analysis. An examination of the atomic charge distribution of Mulliken verified the electron delocalization that led to the molecule's bioactivity. non-small cell lung cancer potential targeted receptor molecules were selected and analyzed using a computational approach. During molecular docking studies, all the

targeted receptor has a strong binding affinity for drugs containing lochnericine. Further targeted receptor and lochnericinedocked complexes showed good stability and lesser fluctuation during the simulation time. The protein-ligand complexes are more stable when their RMSD value is smaller. Every targeted receptor molecule has a lead compound deviation that seems to be smaller than 0.4 nm. Within a 0.50 nm range, the targeted protein-ligand complex fluctuates. While accounting for the studied complexes, the lowest values for the deviation below 0.40 nm and the lesser deviation below 1 nm were obtained. The targeted protein-ligand complexes exhibit good stability and flexibility, according to RMSD and RMSF analysis. Throughout the simulation, better hydrogen bond interaction can be seen in the resulting complexes with lochnericine.

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

CK, SG, and KL designed the study, analyzed the data, and wrote the manuscript. JM, SA, and RS performed the experiments. KTN

validated the data. All the authors have read and agreed to the published version of the manuscript.

Acknowledgments

Dr. KL thankfully acknowledges MHRD-RUSA 2.0 F.24/51/ 2014-U, Policy (TN Multi-Gen), Department of Education, Government of India for providing infrastructural facilities to carry out this research work.

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.

References

- Almansour, A. I., Arumugam, N., Soliman, S. M., Krishnamoorthy, B. S., Halet, J. F., Vishnu Priya, R., et al. (2021). Stereoselective synthesis structure and DFT studies on fluoro- and nitro-substituted spirooxindole-pyrrolidine heterocyclic hybrids. *J. Mol. Struct.* 1237, 130396. doi:10.1016/j.molstruc.2021.130396
- Areans, J. F., Lopez, I., Wooley, M. S., Otero, J. C., and Marcos, J. I. (1988). Vibrational spectrum and internal rotation in 2-methylpyrazine. *J. Chem. Soc. Faraday Trans. Chem. Phys.* 284, 53–65. doi:10.1039/f29888400053
- Awan, Z. A., Bahattab, R., Kutbi, H. I., Noor, A. O. J., Al-Nasser, M. S., Shaik, N. A., et al. (2021). Structural and molecular interaction studies on familial hypercholesterolemia causative PCSK9 functional domain mutations reveals binding affinity alterations with LDLR. *Int. J. Peptide Res. Ther.* 27 (1), 719–733. doi:10.1007/s10989-020-10121-8
- Babu, B., Chandrasekaran, J., Mohanbabu, B., Matsushita, Y., and Saravanakumar, M. (2016). Growth physicochemical and quantum chemical investigations on 2-amino 5-chloropyridinium 4-carboxybutanoate-an organic crystal for biological and optoelectronic device applications. *RSC Adv.* 6, 110884–110897. doi:10.1039/c6ra15791b
- Banaganapalli, B., Mallah, B., Alghamdi, K. S., Albaqami, W. F., Alshaer, D. S., Alrayes, N., et al. (2022). Integrative weighted molecular network construction from transcriptomics and genome wide association data to identify shared genetic biomarkers for COPD and lung cancer. *Plos one* 17 (10), e0274629. doi:10.1371/journal.pone.0274629
- Bima, A. I. H., Elsamanoudy, A. Z., Alghamdi, K. S., Shinawi, T., Mujalli, A., Kaipa, P. R., et al. (2022). Molecular profiling of melanocortin 4 receptor variants and agouti-related peptide interactions in morbid obese phenotype: A novel paradigm from molecular docking and dynamics simulations. *Biologia* 77 (5), 1481–1496. doi:10.1007/s11756-022-01037-3
- Dallakyan, S., and Olson, A. J. (2015). Small-molecule library screening by docking with PyRx. *Methods Mol. Biol. Clift. NJ* 1263, 243–250. doi:10.1007/978-1-4939-2269-7_19
- Deng, L. J., Qi, M., Li, N., Lei, Y. H., Zhang, D. M., and Chen, J. X. (2020). Natural products and their derivatives: Promising modulators of tumor immunotherapy. *J. Leukoc Biol.* 108 (2), 493–508. doi:10.1002/JLB.3MR0320-444R
- Dennington, R., Keith, T., and Milam, J. (2009). *Gauss view version 5*. Shawnee KS: Semichem Inc.
- Ekiert, H. M., and Szopa, A. (2020). Biological activities of natural products. *Molecules* 25 (23), 5769. doi:10.3390/molecules25235769
- Frisch, M. J., Trucks, G. W., Schlegel, H. B., Scuseria, G. E., Robb, M. A., Cheeseman, J. R., et al. (2009). *Gaussian 09 revision C. 02*. Wallingford CT: Gaussian Inc.
- Geetha, R., Meera, M. R., Vijayakumar, C., Premkumar, R., and Milton Franklin Benial, A. (2021). Synthesis spectroscopic characterization molecular docking and *in vitro* cytotoxicity investigations on 8-amino-6-methoxy quinolinium picrate: A novel breast cancer drug. *J. Biomol. Struct. Dyn.* 2021, 1–14. doi:10.1080/07391102.2021.2024259
- Gueguen, P., Metoikidou, C., Dupic, T., Lawand, M., Goudot, C., Baulande, S., et al. (2021). Contribution of resident and circulating precursors to tumor-infiltrating CD8+ T cell populations in lung cancer. *Sci. Immunol.* 6 (55), eabd5778. doi:10.1126/sciimmunol.abd5778
- Hausman, D. M. (2019). What is cancer? *Perspect. Biol. Med.* 62 (4), 778–784. doi:10.1353/pbm.2019.0046
- Jamroz, M. H. (2004). *Vibrational energy distribution analysis VEDA 4.0 program*. Warsaw: VEDA analysis.
- Karabacak, M., Karagöz, D., and Kurt, M. (2008). Experimental (FT-IR and FT-Raman spectra) and theoretical (*ab initio* HF, DFT) study of 2-chloro-5-methylaniline. *J. Mol. Struct.* 892 (1–3), 25–31. doi:10.1016/j.molstruc.2008.04.054
- Kavitha, M., Iravathy, G., Adi Maha, L. M., Ravi, V., Sridhar, K., Vijayanand, R. P., et al. (2015). Correlation between EGFR gene mutations and lung cancer: A hospital-based study. *Asian Pac. J. Cancer Prev.* 16 (16), 7071–7076. doi:10.7314/apjcp.2015.16.16.7071
- Koyambo-Konzapa, S. J., MbesseKongbonga, G. Y., Premkumar, R., Duvalier RamlinaVamhindi, B. S., Nsangou, M., and Milton Franklin Benial, A. (2021). Spectroscopic quantum chemical molecular docking and molecular dynamics investigations of hydroxylic indole-3-pyruvic acid: A potent candidate for nonlinear optical applications and alzheimer's drug. *J. Biomol. Struct. Dyn.* 40, 10651–10664. doi:10.1080/07391102.2021.1947380
- Laskowski, R. A., MacArthur, M. W., and Thornton, J. M. (2016). Procheck: Validation of protein-structure coordinates. *Int. Tables Crystallogr.* 21.4, 684–687. doi:10.1107/97809553602060000882
- Mohamed Asath, R., Premkumar, R., Mathavan, T., and Milton Franklin Benial, A. (2017). Spectroscopic and molecular docking studies on NN-di-tert-butoxycarbonyl (Boc)-2-amino pyridine: A potential bioactive agent for lung cancer treatment. *J. Mol. Struct.* 1143, 415–423. doi:10.1016/j.molstruc.2017.04.117
- Mullard, A. (2020). Addressing cancer's grand challenges. *Nat. Rev. Drug Discov.* 19 (12), 825–826. doi:10.1038/d41573-020-00202-0
- Reck, M., Carbone, D. P., Garassino, M., and Barlesi, F. (2021). Targeting KRAS in non-small-cell lung cancer: Recent progress and new approaches. *Ann. Oncol.* 32 (9), 1101–1110. doi:10.1016/j.annonc.2021.06.001

- Sangavi, P., and Langeswaran, K. (2021). Anti-tumorigenic efficacy of tangeretin in liver cancer-an *in-silico* approach. *Curr. computer-aided drug Des.* 17 (3), 337–343. doi:10.2174/1573409916666200219120254
- Sangavi, P., Langeswaran, K., and Sangeetha, R. (2022). Identification and validation of hydroxychavicol from betel leaf as a promising breast cancer inhibitor: An *in vitro* and *in silico* analysis. *Curr. Enzyme Inhib.* 18 (2), 127–134. doi:10.2174/1573408018666220210141311
- Saravanan, S., and Blachandran, V. (2014). Quantum mechanical study and spectroscopic (FT-IR FT-Raman UV-Visible) study potential energy surface scan Fukui function analysis and HOMO–LUMO analysis of 3-tert-butyl-4-methoxyphenol by DFT methods. *Spectrochim. Acta* 130, 604–620. doi:10.1016/j.saa.2014.04.058
- Schabath, M. B., and Cote, M. L. (2019). Cancer progress and priorities: Lung cancer. *Cancer Epidemiol. Biomarkers Prev.* 28 (10), 1563–1579. doi:10.1158/1055-9965.EPI-19-0221
- Schüttelkopf, A. W., and Van Aalten, D. M. (2004). ProdrG: A tool for high-throughput crystallography of protein-ligand complexes. *Acta Crystallogr. Sect. D. Biol. Crystallogr.* 60 (8), 1355–1363. doi:10.1107/S0907444904011679
- Schwede, T., Kopp, J., Guex, N., and Peitsch, M. C. (2003). SWISS-MODEL: An automated protein homology-modeling server. *Nucleic acids Res.* 31 (13), 3381–3385. doi:10.1093/nar/gkg520
- Shaik, N. A., Al-Kreathy, H. M., Ajabnoor, G. M., Verma, P. K., and Banaganapalli, B. (2019). Molecular designing, virtual screening and docking study of novel curcumin analogue as mutation (S769L and K846R) selective inhibitor for EGFR. *Saudi J. Biol. Sci.* 26 (3), 439–448. doi:10.1016/j.sjbs.2018.05.026
- Shaik, N. A., Al-Saud, N. B. S., Aljuhani, T. A., Jamil, K., Alnuman, H., Aljeaid, D., et al. (2022). Structural characterization and conformational dynamics of alpha-1 antitrypsin pathogenic variants causing alpha-1-antitrypsin deficiency. *Front. Mol. Biosci.* 9, 1051511. doi:10.3389/fmolb.2022.1051511
- Sussman, J. L., Lin, D., Jiang, J., Manning, N. O., Prilusky, J., Ritter, O., et al. (1998). Protein Data Bank (PDB): Database of three-dimensional structural information of biological macromolecules. *Acta Crystallogr. Sect. D. Biol. Crystallogr.* 54 (6-1), 1078–1084. doi:10.1107/s0907444998009378
- Tripathi, P., Ghosh, S., and Talapatra, S. N. (2019). Bioavailability prediction of phytochemicals present in Calotropis procera (Aiton) R. Br. by using Swiss-ADME tool. *World Sci. News* 131, 147–163.
- Turner, P. J. (2005). *XMGRACE version 5.1*. 19. Beaverton OR: Center for Coastal and Land-Margin Research Oregon Graduate Institute of Science and Technology.
- UniProt (2019). UniProt: A worldwide hub of protein knowledge. *Nucleic acids Res.* 47 (1), D506–D515. doi:10.1093/nar/gky1049
- Valarmathi, T., Premkumar, R., and Milton Franklin Benial, A. (2020b). Quantum chemical and molecular docking studies on 14-bis(methylamino)anthraquinone: A DFT approach. *AIP Conf. Proc.* 2270, 040001.
- Valarmathi, T., Premkumar, R., and Milton Franklin Benial, A. (2020a). Spectroscopic and molecular docking studies on 1- hydroxyanthraquinone: A potent ovarian cancer drug. *J. Mol. Struct.* 1213, 128163. doi:10.1016/j.molstruc.2020.128163
- Van Der Spoel, D., Lindahl, E., Hess, B., Groenhof, G., Mark, A. E., and Berendsen, H. J. (2005). Gromacs: Fast flexible and free. *J. Comput. Chem.* 26 (16), 1701–1718. doi:10.1002/jcc.20291
- Wang, X., Shi, W., Shi, H., Lu, S., Wang, K., Sun, C., et al. (2016). TRIM11 overexpression promotes proliferation, migration and invasion of lung cancer cells. *J. Exp. Clin. Cancer Res.* 35 (1), 100. doi:10.1186/s13046-016-0379-y
- Zhou, C., and Yao, L. D. (2016). Strategies to improve outcomes of patients with EGFR-mutant non-small cell lung cancer: Review of the literature. *J Thorac Oncol.* 11 (2), 174–186. doi:10.1016/j.jtho.2015.10.002
- Zhou, Z., Liu, Y., Ren, Q., Yu, D., and Lu, H. (2021). Synthesis crystal structure and DFT study of a novel compound *N*-(4-(24-dimorpholinopyrido[23-*d*]pyrimidin-6-yl)phenyl) pyrrolidine-1-carboxamide. *J. Mol. Struct.* 1235, 130261. doi:10.1016/j.molstruc.2021.130261



OPEN ACCESS

EDITED BY

Sanjeev Kumar Singh,
Alagappa University, India

REVIEWED BY

Qin Xu,
Shanghai Jiao Tong University, China
Chandrabose Selvaraj,
Saveetha University, India

*CORRESPONDENCE

Lilly M. Saleena,
✉ saleenam@srnmist.edu.in

SPECIALTY SECTION

This article was submitted to Biophysics,
a section of the journal
Frontiers in Molecular Biosciences

RECEIVED 14 December 2022

ACCEPTED 23 February 2023

PUBLISHED 17 March 2023

CITATION

Achudhan AB, Kannan P and Saleena LM
(2023), Functional metagenomics
uncovers nitrile-hydrolysing enzymes in a
coal metagenome.
Front. Mol. Biosci. 10:1123902.
doi: 10.3389/fmolb.2023.1123902

COPYRIGHT

© 2023 Achudhan, Kannan and Saleena.
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.

Functional metagenomics uncovers nitrile-hydrolysing enzymes in a coal metagenome

Arunmozhi Bharathi Achudhan, Priya Kannan and Lilly M. Saleena*

Department of Biotechnology, School of Bioengineering, SRM Institute of Science and Technology,
Kattankulathur, Tamil Nadu, India

Introduction: Nitriles are the most toxic compounds that can lead to serious human illness through inhalation and consumption due to environmental pollution. Nitrilases can highly degrade nitriles isolated from the natural ecosystem. In the current study, we focused on the discovery of novel nitrilases from a coal metagenome using *in silico* mining.

Methods: Coal metagenomic DNA was isolated and sequenced on the Illumina platform. Quality reads were assembled using MEGAHIT, and statistics were checked using QUAST. Annotation was performed using the automated tool SqueezeMeta. The annotated amino acid sequences were mined for nitrilase from the unclassified organism. Sequence alignment and phylogenetic analyses were carried out using ClustalW and MEGA11. Conserved regions of the amino acid sequences were identified using InterProScan and NCBI-CDD servers. The physicochemical properties of the amino acids were measured using ExPASy's ProtParam. Furthermore, NetSurfP was used for 2D structure prediction, while AlphaFold2 in Chimera X 1.4 was used for 3D structure prediction. To check the solvation of the predicted protein, a dynamic simulation was conducted on the WebGRO server. Ligands were extracted from the Protein Data Bank (PDB) for molecular docking upon active site prediction using the CASTp server.

Results and discussion: *In silico* mining of annotated metagenomic data revealed nitrilase from unclassified *Alphaproteobacteria*. By using the artificial intelligence program AlphaFold2, the 3D structure was predicted with a per-residue confidence statistic score of about 95.8%, and the stability of the predicted model was verified with molecular dynamics for a 100-ns simulation. Molecular docking analysis determined the binding affinity of a novel nitrilase with nitriles. The binding scores produced by the novel nitrilase were approximately similar to those of the other prokaryotic nitrilase crystal structures, with a deviation of ± 0.5 .

KEYWORDS

functional metagenomics, nitrilase, nitriles, artificial intelligence, unclassified microorganisms

Introduction

Cyanide-containing compounds are known as nitriles and are widely distributed in the natural environment. They are generated by different plants in various forms, such as ricinine, phenyl acetonitrile, cyanogenic glycosides, and β -cyanoalanine (Sewell et al., 2003). Anthropogenic activities have substantially influenced the production of vast quantities of nitrile compounds. Nitriles are naturally poisonous and are recognised to be a leading cause

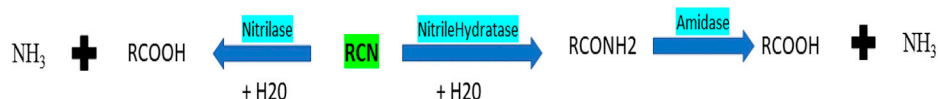


FIGURE 1
Mechanism of nitrilase, nitrile hydratase, and amidase.

of environmental pollution, which is detrimental to human health (Li et al., 2013). Most of the cyanide in soil and water comes from effluents that contain a variety of inorganic cyanides and nitriles. Contamination is caused by using herbicides with the nitrile group, such as 2,6-dichlorobenzonitrile and bromoxynil (3,5-dibromo-4-hydroxybenzonitrile). Nitrile pollution can also be caused by the exhaust from cars (Nigam et al., 2017). The majority of nitrile poisoning symptoms include abdominal pain, seizures, breathing problems, sore throat, difficulty falling asleep, and damage to the kidneys (Kupke et al., 2016; Tanii, 2017).

Nitrile compounds can be degraded by using microbes or chemicals. Chemical degradation of nitriles involves harsh reaction conditions and generates excess waste (Wang, 2015). Nitrile-hydrolysing enzymes can convert various nitriles to acids. Enzymes that hydrolyse nitrile include nitrilases (EC 3.5.5.1), nitrile hydratases (EC 4.2.1.84), and amidases (EC 3.5.1.4) (Figure 1). These enzymes are utilised extensively in the production of amides and organic acids, both of which are extremely valuable to the manufacturing industry (Egelkamp et al., 2019).

The benzonitrile analogues chloroxynil, bromoxynil, and ioxynil are efficiently degraded by the soil *Actinobacteria* *Rhodococcus rhodochrous* PA-34, *Rhodococcus* sp. NDB 1165, and *Nocardia globerula* NHB-2, and these nitrile degraders should be studied for the bioremediation of benzonitrile herbicide-contaminated soils (Vesela et al., 2010). Nitrilase enzymes are used as good biocatalysts in a wide range of synthetic processes, leading to a huge rise in demand. Hydrolysis of the ricinine nitrile group was first discovered in the soil-isolated bacterial strain belonging to the genus *Pseudomonas*. Many nitrilases have been purified, characterised, immobilised, gene cloned, overexpressed in host strains, and used in industrial plants (Amrutha and Nampoothiri, 2022; Gong et al., 2012).

Advanced bioinformatics tools and techniques are used more often than traditional methods to find or screen a new nitrilase gene. This helps in identifying the protein's suitable substrates (Jones et al., 2020; Klasberg et al., 2016). The maximum synthesis of propionic acids, which have applications in the food and chemical industries, was demonstrated by nitrilase from *Bacillus* sp. (BITSN007) in a study on the biotransformation of nitrile compounds to valuable acids (Biocatalysis of Different Nitriles to Valuable Acids, 2021). The Tyr141Ala mutation in the nitrilase from *P. fluorescens* EBC191 led to a nitrilase variant that can convert aromatic and aliphatic substrates (Brunner et al., 2018). There is a need for new nitrile-degrading enzymes, particularly those with the wide substrate-catalysing properties required for environmental remediation. The study employs shotgun sequencing of lignite samples collected from the coal mine. After extracting the coal metagenomic data, we identified an unclassified bacterium that codes for the nitrilase enzyme. The primary amino acid sequence

was searched for conserved regions and domain findings. The secondary and tertiary structures were also identified for the nitrilase protein, with different types of nitriles (substrates) used to analyse their binding efficiency by molecular docking analysis. Binding energy was also calculated for other reference prokaryotic crystal structures and compared to the predicted structure. This study focuses on identifying nitrilase enzymes from metagenomic data and exploring their binding affinity with a wide range of nitriles.

Materials and methods

Sample isolation and sequencing

A lignite sample from the coal mine in Neyveli, India (11°35'34.44"N and 79°29'29.04"E), was collected. The metagenomic DNA was isolated from the lignite sample using the PowerMax soil DNeasy kit (QIAGEN). The isolated metagenomic DNA was then subjected to shotgun sequencing on an Illumina HiSeqTM 2000 platform to generate paired-end sequences.

Metagenomics data analysis

The forward and reverse end reads in the FASTQ format were used as the input in the FASTQC tool (Andrews, 2010). The generated output HTML files were merged using MultiQC (Ewels et al., 2016) to create a single HTML file report containing the quality statistics of the reads. The forward and reverse FASTQ files are the input for the MEGAHIT assembler (Li et al., 2015). A k-mer value of 99 and a minimum contig length of 200 parameters were assigned. The output was generated as contigs in a single FASTA file. The obtained contigs were analysed in the QUAST tool (Gurevich et al., 2013) for the number and size of the contigs.

Taxonomical and functional findings

The contigs in FASTA format were annotated using the SqueezeMeta tool, an automated pipeline (Tamames and Puente-Sánchez, 2019). First, protein-coding genes were predicted from the contigs using Prodigal, and amino acid and nucleotide sequences were generated in the FASTA files. The results of these annotated nucleotide sequences were automatically loaded as input into Diamond, which searched the GenBank nr database for taxonomical classification and the KEGG database for functional annotation. The term "Nitrilase" was searched using the grep script in tab-separated value files. KEGG IDs and contig IDs were noted for

identifying taxonomy and extracting the nitrilase nucleotide and amino acid sequences.

Sequence alignment and phylogenetic analysis

The identified amino acid sequence in FASTA format was uploaded in BlastP (Johnson et al., 2008) and ran against the NCBI database of protein reference sequences to find similar sequences. Similar sequences were chosen based on the >50% identity of the matches with the query sequence. These sequences above the threshold were downloaded in FASTA format in a single file along with the query sequence. This FASTA file was uploaded using MEGA 11 software (Tamura et al., 2021) and was aligned using Clustal Omega. The evolutionary history of these sequences was created using the neighbour-joining method by selecting the phylogeny tab in the software application.

Conserved region analysis

The amino acid sequence in FASTA format was submitted with the default parameters in the NCBI-CDD (Marchler-Bauer et al., 2017) and InterProScan (Quevillon et al., 2005) databases to predict the homologous superfamily, conserved domain, conserved region, Gene Ontology, and NCBI-CDD from the query sequence of amino acids.

Physiochemical properties of nitrilase enzymes

The amino acid sequence was pasted in FASTA format into a query box of the ExPASy's ProtParam (Gasteiger et al., 2003) server and submitted to identify the physical and chemical properties of the protein sequence. The server page quantifies the number of amino acids, the molecular weight, the number of negatively charged residues, the instability index, the theoretical pI, and the grand average of hydropathicity.

Structure prediction

The secondary structure was predicted by uploading an amino acid sequence in FASTA format to the NetSurfP tool with the default parameters (Høie et al., 2022). The amino acid sequences were used in the software package Chimera X version 1.4 (Pettersen et al., 2021; Goddard et al., 2018) to perform the AlphaFold2 tool for 3D structure prediction. This was predicted in the software programme by choosing the AlphaFold2 option from the Structure Prediction tab under Tools. The amino acid sequence was uploaded in the query box and submitted with the default parameters. The predicted tertiary structure was uploaded into the PROCHECK tool (Laskowski et al., 1993) to create the Ramachandran plot, which checks to validate the stereochemical quality of the protein structure.

Molecular dynamic simulation

The predicted structure was submitted for molecular dynamics simulation on the WebGRO server (Bekker et al., 1993) to check its stability. Using the GROMOS96 43a1 force field settings, the complex system was solvated using a simple point charge (SPC) water model in a triclinic periodic box. The complex system was maintained at a salt concentration of 0.15 M by adding a suitable amount of Na⁺ and Cl⁻ counterions. Using the steepest descent approach, energy reduction was achieved in 5,000 steps. Constant amount, volume temperature (NVT/NPT), and pressure equilibrium types were used. The temperature was set to 300 K, and the pressure was set to 1.0 bar. The simulation time was 100 ns and was conducted with 1,000 frames per simulation. Finally, the simulation result was analysed based on the time-dependent root mean square deviation (RMSD) of the given structure and the root mean square fluctuation (RMSF) of each residue.

Protein and ligand preparation

The homepage of the Protein Data Bank (PDB) (Burley et al., 2017) was searched for the 3D nitrilase protein structure. X-ray diffraction was selected using the filter option, and the crystal structure of nitrilase proteins was retrieved in PDB format. Nitriles were downloaded from the PubChem database (Kim et al., 2020) in SDF format. The ligands were converted to the PDB format using PyMOL (Schrödinger, 2000), and protein and ligand formats were changed to the PDBQT format for molecular docking using the AutoDockTools-1.5.7 tool.

Active site predictions

The protein structures' active site residues were predicted by uploading the PDB files of nitrilase proteins in the computed atlas of the protein surface topography -3.0 (CASTp) server (Tian et al., 2018). The alpha shape theory's pocket algorithm calculated the active pockets or binding sites, with large pockets with high volumes likely to contain enzyme-binding sites for the interaction between proteins and ligands.

Molecular docking analysis

In AutoDockTools-1.5.7, the protein structures in the PDB format were used as input to create a grid file by placing a grid box in the protein's predicted active site for the binding of ligand molecules. The inputs of a protein, a ligand in PDBQT format, and a grid file in text format were used to perform a molecular docking analysis in AutoDock Vina (Eberhardt et al., 2021) with the default parameters. The output was the protein-ligand complex in PDB format. The complex in PDB format was used to study the interaction of amino acids with ligands using the LigPlot + tool (Laskowski and Swindells, 2011).

✓1. Nitrilase Coal metagenome	A	R	R	G	A	S	V	V	A	F	F	E	V	W	L	P	G	Y	P	L	F	L	W	I	G	D	A	E	W	Q	R	P	Y	R	E	R	Y	A	E	N	S	A	V	I	D	G	P	E	H	R	R	
✓2. carbon-nitrogen hydrolase family protein Rhizobium sp. YK2	A	E	K	G	A	Q	I	V	A	F	F	E	V	W	L	P	G	Y	P	V	F	L	W	L	G	D	E	A	W	Q	A	G	Y	R	S	K	Y	I	E	N	S	A	V	F	G	G	P	E	H	E	R	
✓3. carbon-nitrogen hydrolase protein A9Z06 14010 Rhizobium sp. YK2	A	E	K	G	A	Q	I	V	A	F	F	E	V	W	L	P	G	Y	P	V	F	L	W	L	G	D	E	A	W	Q	A	G	Y	R	S	K	Y	I	E	N	S	A	V	F	G	G	P	E	H	E	R	
✓4. carbon-nitrogen hydrolase family protein Martellella alba	A	R	K	G	V	K	L	I	A	F	F	E	I	W	L	P	G	Y	P	V	F	L	W	L	G	D	E	A	W	Q	T	E	Y	R	R	N	F	I	E	N	S	A	R	L	G	G	P	E	H	E	K	
✓5. carbon-nitrogen hydrolase family protein Bradyrhizobium sp. Ce-3	A	R	K	G	A	R	V	I	A	F	F	E	V	W	L	P	G	Y	P	V	F	L	W	L	G	D	E	Q	W	Q	A	D	Y	R	T	Y	I	K	N	S	A	E	H	G	G	P	E	H	K	R		
✓6. nitrilase Bradyrhizobium sp. Ce-3	A	R	K	G	A	R	V	I	A	F	F	E	V	W	L	P	G	Y	P	V	F	L	W	L	G	D	E	Q	W	Q	A	D	Y	R	T	Y	I	K	N	S	A	E	H	G	G	P	E	H	K	R		
✓7. carbon-nitrogen hydrolase family protein Microbacterium sp. SK110	A	A	G	A	D	L	I	A	F	F	E	V	W	L	P	G	Y	P	V	F	L	W	L	G	T	D	E	W	Q	A	D	Q	R	E	H	Y	I	A	N	S	A	V	H	G	G	P	E	H	V	R		
✓8. Sulfotabacter sp. 20 GPM-1509mcarbon-nitrogen hydrolase family protein	A	E	Q	G	A	R	L	I	A	F	F	E	L	W	L	P	G	Y	P	I	F	L	W	L	G	D	E	T	W	Q	A	S	R	R	A	R	Y	I	E	N	S	A	M	L	N	G	G	P	E	H	Q	A
✓9. carbon-nitrogen hydrolase family protein Sulfotabacter sp. DFL-14	A	E	Q	G	T	R	L	I	A	F	F	E	L	W	L	P	G	Y	P	I	F	L	W	L	G	D	E	V	W	Q	A	S	R	R	A	R	Y	I	E	N	S	A	M	L	N	G	G	P	E	H	Q	A

FIGURE 2

Multiple sequence alignment of the selected eight sequences with the identified novel nitrilase sequence. The highlighted residues are the conserved amino acid sequences. The identified nitrilase sequence from the coal metagenome showed similarity to the 39th site of the other selected protein sequences.

Results

Metagenomic data analysis

FASTQC and MULTIQC determined the paired-end reads to be within the Phred score value of 36, considering that the raw reads to be of standard quality (Bin Kwong et al., 2017). The quality reads of 150-bp length from 32 GB of data were processed for assembly. The statistics of the assembled contigs were evaluated using the QUAST tool, resulting in 226 contigs with >50,000 bp, 1,240 contigs with >25,000 bp, 5,265 contigs with >10,000 bp, 11,596 contigs with >5000 bp, and 73,144 contigs with >1000 bp.

Taxonomical and functional findings

The phyla observed in significant numbers belong to *Proteobacteria* (76%), *Actinobacteria* (8%), *Firmicutes* (8%), *Spirochaetes* (2%), *Bacteroidetes* (1.5%), *Chloroflexi* (0.8%), *Planctomycetes* (0.5%), *Cyanobacteria* (0.2%), [Thermi] (0.1%), *Fusobacteria* (0.1%), and *Acidobacteria* (0.1%). Using “grep,” the results from SqueezeMeta were searched for nitrilase, and then organisms involved in nitrilase enzymes were identified. In total, 27 organisms were identified, of which four were from unclassified microorganisms and belonged to the KEGG ID K01501 with the metabolic pathway number EC:3.5.5.1. For the novel protein approach, long amino acid sequences and unclassified microorganisms were preferred. A single amino acid sequence of 261 base pairs encoding an unclassified *Alphaproteobacteria* was selected.

Sequence alignment and phylogenetic analysis

Using BlastP, the target protein sequence was compared to the NCBI database of protein reference sequences. Eight sequences above the threshold were downloaded from NCBI-BLAST, and a phylogenetic tree was built along with the identified nitrilase sequence using the MEGA11 tool. Figure 2 shows the multiple sequence alignment of all the sequences. The branch lengths marked next to the branches on the tree are shown to scale and are in the same units as the evolutionary distances, which are used to estimate the phylogenetic tree (Figure 3). The scale value of 0.050 shows the genetic change between the protein sequences. The branch of the

nitrilase coal metagenome is the longest, with a length of 0.30, and is therefore known to have greater genetic change than other protein sequences. The software application calculated evolutionary distances using the Poisson correction method and amino acid substitutions per site. This analysis involved nine protein sequences. There were a total of 309 positions in the final data set, and 139 residues were conserved among the protein sequences.

Conserved region and domain findings

According to the analysis of the InterProScan database and NCBI-CDD results, a carbon–nitrogen hydrolase domain was found in the 4–19 amino acid position in the protein sequence, indicating it to be a member of the nitrilase superfamily. InterProScan classifies three domains: cellular component, molecular function, and biological process. The predicted biological process for the identified protein is involved in the metabolic process of the nitrogen compounds (GO: 0006807), and the molecular function is involved in catalytic activity (GO:0003824). Based on the unclassification, amino acid sequence length, conserved region, and domain analysis, it is considered to be a novel nitrilase enzyme.

Physiochemical characteristics of amino acid sequences

Using the ProtParam tool, the physiochemical characteristics of the novel protein sequence were calculated. The compositions of the amino acids are shown in Figure 4. The result shows that the negatively charged aspartic acid and glutamic acid were higher than the positively charged arginine and lysine. The protein has an acidic nature because of its 4.86 isoelectric points. The total number of atoms is around 4,010, including 1,275 carbon, 1,990 hydrogen, 354 nitrogen, 385 oxygen, and 6 sulphur atoms. Proteins with a GRAVY score below zero are hydrophilic, while proteins with a GRAVY score above zero are hydrophobic. The novel protein has a GRAVY value of −0.126, which indicates that the protein is inherently hydrophilic. The protein has an instability index of 53.02, which indicates that it is unstable because a protein with an instability index of 40 or less was stable in the test tube. The sequence of amino acids was also used to estimate the protein’s half-life. For yeast cells, the half-life period is > 20 h (*in vivo*), and for *E. coli*, the half-life period is >10 h (*in vivo*) (Gasteiger et al., 2005).

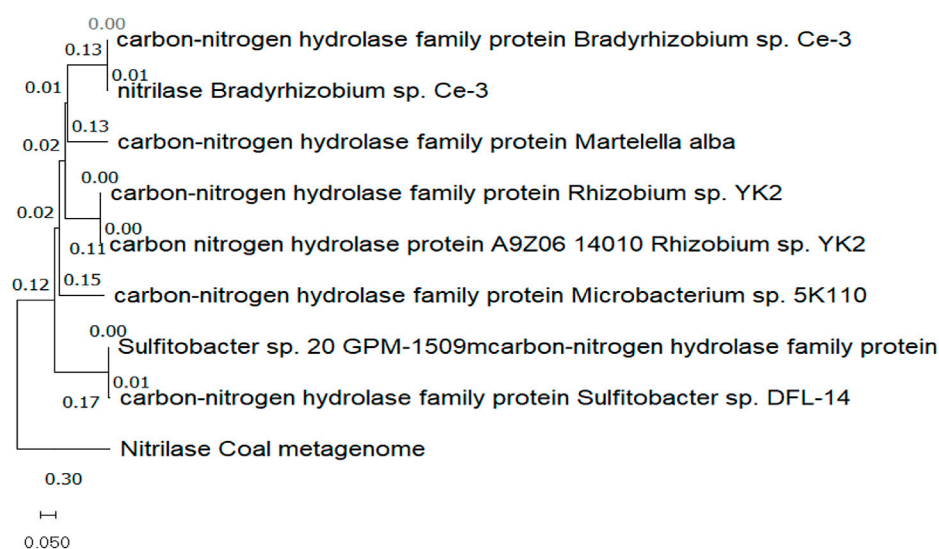


FIGURE 3

Phylogenetic tree showing the evolutionary relationship between nitrilase from the coal metagenome and other eight selected protein sequences from the NCBI database.

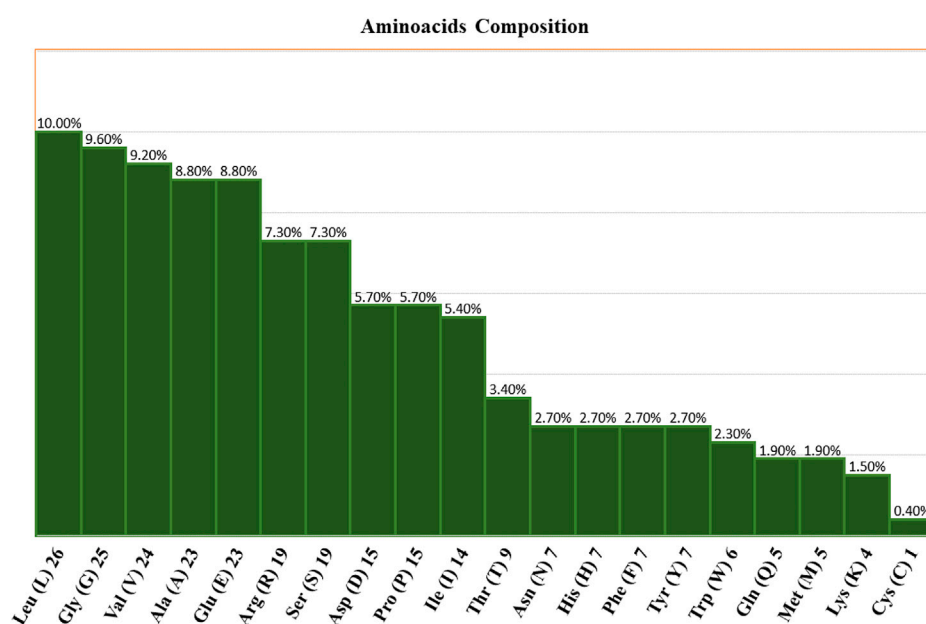


FIGURE 4

Amino acid compositions of the nitrilase enzyme.

Structure predictions

The results from the NetSurfP version 2.0 server revealed that the nitrilase protein has 10 helices, 16 strands, and 23 coils in its secondary structure (Figure 5).

Using the amino acid sequence with the default parameters, the nitrilase protein structure was predicted in AlphaFold2

(Figure 6). In AlphaFold2, the predicted local distance difference test (pLDDT) calculates the mean confidence value ranging from 0 to 100. Here, the predicted nitrilase's mean pLDDT score was 95.8. A mean confidence value of 90 and above is said to agree with an experimental structure (Jumper et al., 2021). The Ramachandran plot was utilised in the PROCHECK tool, and the modelled protein was evaluated and

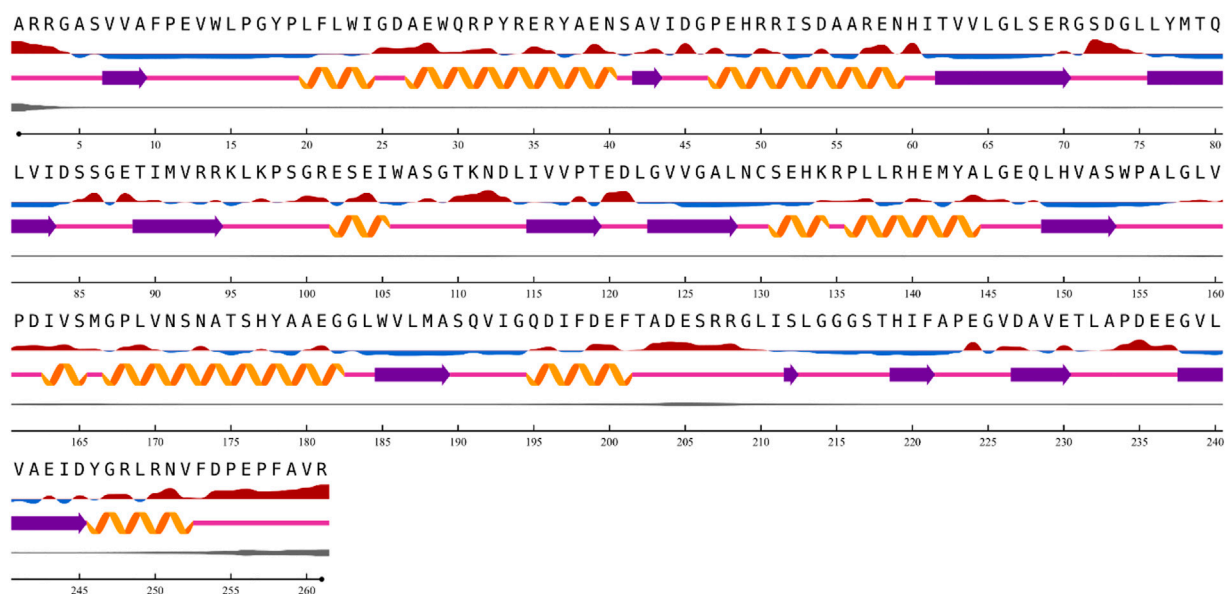
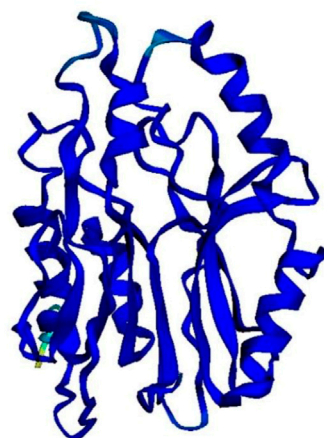


FIGURE 5
Prediction of nitrilase secondary structure.



pLDDT: ■ Very low (<50) ■ Low (60) ■ OK (70) ■ Confident (80) ■ Very high (>90)

FIGURE 6
Prediction of a novel nitrilase structure using AlphaFold2.

validated (Supplementary Material S1). In the model, 92.7% of residues were found in the most favoured regions (A, B, and L), 6.8% in the additional allowed regions (a, b, l, and p), 0.5% in the generously allowed regions (~a, ~b, ~l, and ~p), and no residues in the disallowed regions. As a result, the percentage distribution of amino acid residues revealed that the predicted nitrilase structure is of high quality.

Molecular dynamic simulation

To analyse the flexibility and stability of the best-predicted protein structure provided by AlphaFold2, a time-dependent molecular dynamic simulation was conducted at 100 ns employing the GROMACS forcefield on the WebGRO server. The RMSD value of the nitrilase protein structure showed

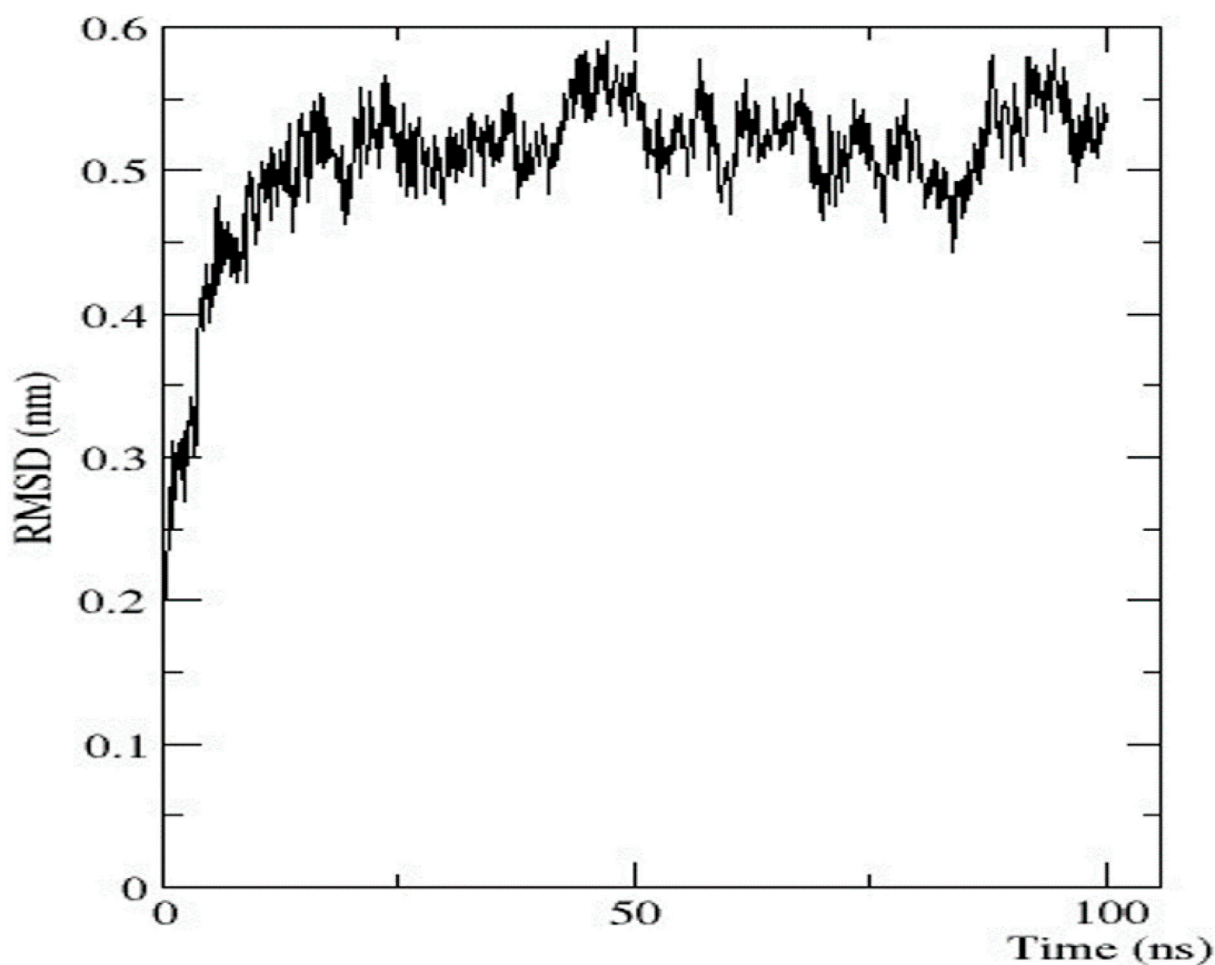


FIGURE 7
Graphical representation of RMSD.

equilibrium with an average of 0.2 nm. The largest oscillation in RMSD was observed in the 0–10 ns range. Afterwards, the 10–ns RMSD value was stabilised up to 100 ns, with an average value of 0.5 nm (Figure 7). The RMSF value was then assessed to analyse the structural flexibility of the atoms in the backbone of the proteins. The obtained data showed that there were fluctuations in the residues present in the loops of the protein structure ($\text{RMSF} \leq 0.5$ nm), and this indicates that the complex was flexible in these loop regions (Supplementary Material S2). The conformational changes in the loop structure caused by the flexibility have no impact on the protein structure (Li et al., 2011).

Protein and ligand preparation

The predicted nitrilase protein and the retrieved crystal structures of 3WUY (Zhang et al., 2014), 3IW3, and 3IVZ (Raczynska et al., 2011) were used to check the binding scores individually by molecular docking analysis and compared. The three structures were determined by X-ray diffraction analysis. The protein structure (3WUY), which was identified from

Synechocystis sp. PCC 6803 substr. Kazusa, comprises 349 amino acids and has a resolution of 3.10 Å. The other two protein structures, 3IW3 and 3IVZ, were identified in *Pyrococcus abyssi* GE5 and have a length of 262 amino acid sequences with resolutions of 1.80 Å and 1.57 Å, respectively. For the protein–ligand interaction study, nitriles such as acrylonitrile, benzonitrile, dichlobenil, fumaronitrile, malanonitrile, and succinonitrile were retrieved. Proteins and ligand molecules were converted to the PDBQT format and were ready for docking analysis.

Active site predictions

The CASTp server identified the active site for the predicted nitrilase protein and the crystal structure of the proteins. The surface area measurement and cavity volumes were predicted. For the predicted protein, the area of the active site was 677.110 Å², and the volume was 645.046 Å³. The area of the active site for 3WUY, 3IW3, and 3IVZ was 983.426 Å², 61.447 Å², and 188.905 Å² and the volume was 1,129.287 Å³,

TABLE 1 Docking scores of proteins binding with ligands.

Nitrile	Predicted nitrilase	3WUY	3IW3	3IVZ
Acrylonitrile	-2.9	-3.2	-2.9	-3.5
Benzonitrile	-4.9	-5.4	-5.0	-4.7
Dichlobenil	-5.4	-5.9	-5.2	-4.9
Fumaronitrile	-3.8	-3.6	-3.6	-3.3
Malononitrile	-3.2	-3.4	-3.0	-3.1
Succinonitrile	-3.6	-3.9	-3.4	-3.4

31.456 Å³, and 60.553 Å³, respectively. All the pockets chosen for the proteins have a high surface area and volume for the specific enzyme-binding site (Barman et al., 2020).

Molecular docking analysis

Using AutoDockTools-1.5.7, a grid file was generated for all the proteins. When the model overlapped with the template, an acceptable range of RMSD was determined to be 2.0, which is regarded as satisfactory docking (Castro-Alvarez et al., 2017). The docking analysis was completed using AutoDock Vina, as

tabulated in Table 1. The docking results produced nine ligand-binding poses, of which the one with the lowest RMSD (value 0) was selected, indicating a true binding pose. The protein–ligand interaction showed that these proteins' docking scores are almost similar when they bind to six different ligands.

The protein–ligand complex of the predicted nitrilase protein was analysed to determine the nature of the interactions between the amino acids and the ligands using LigPlot + (Figure 8).

Nitrile (C≡N) conversion involves a nucleophilic attack on the electrophilic carbon (electrophilic) by the nucleophilic group present in the side chain of amino acids in the active site, resulting in the formation of carboxyl groups (COOH). Cysteine, serine, threonine, tyrosine, glutamic acid, aspartic acid, lysine, arginine, and histidine have nucleophilic R groups and can act as nucleophilic donors. Arginine and tyrosine amino acids serve as functional converters in benzonitrile and succinonitrile. Threonine and glutamic acid play a functional role in the catalysis of fumaronitrile and serine in malononitrile. The nucleophilic attack on acrylonitrile and dichlobenil is mediated by histidine and threonine, respectively.

Discussion

Microbial nitrilase, an enzyme of the nitrilase superfamily, is a great option for numerous industrial applications and

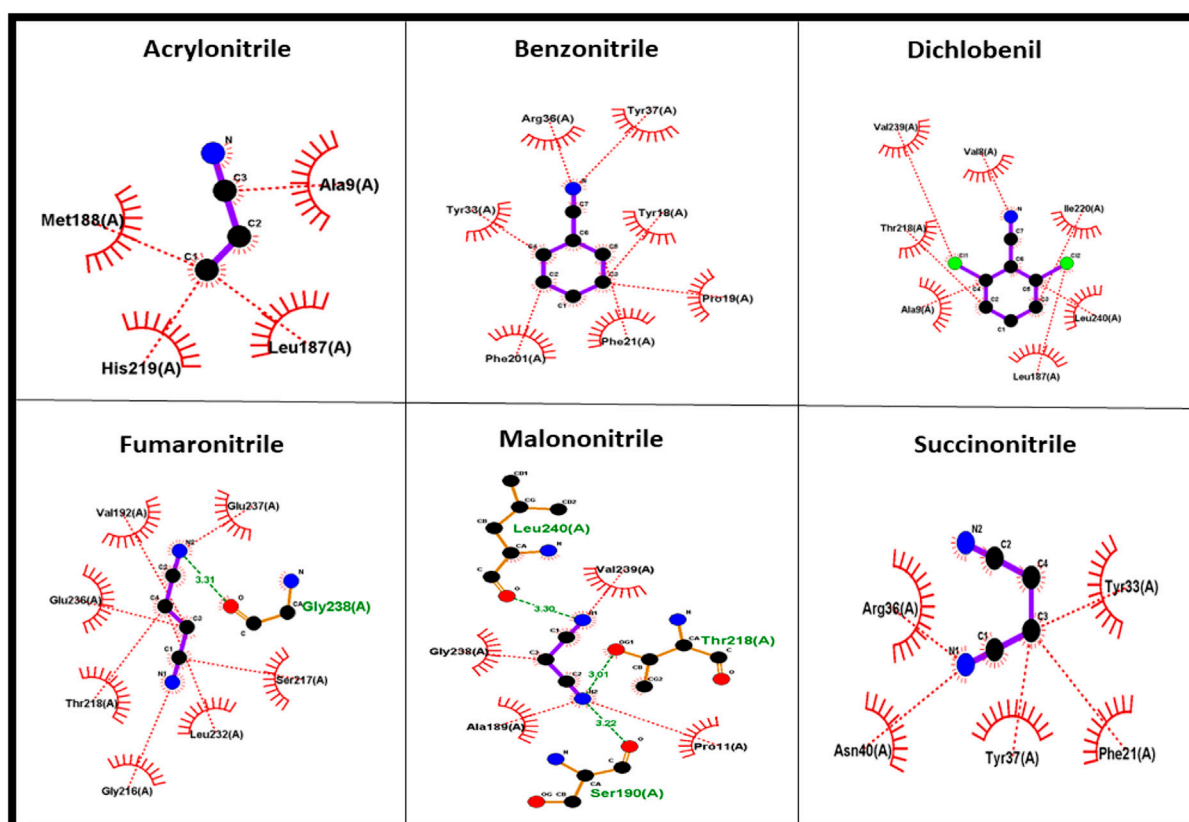


FIGURE 8
Amino acid interaction with nitrile ligands.

bioremediation procedures. Functional metagenomics enabled the identification of the novel nitrilase enzyme from environmental sources, which represent a non-culturable source of the enzyme. With the help of artificial intelligence and machine learning in metagenomics, novel enzyme candidates can be identified for potential use in bioremediation and therapeutics. Ligand molecules are bound to this active site of the protein, and the identification of their protein–ligand binding efficacy will lead to drug discovery, which is beneficial for the advancement of green chemistry.

Data availability statement

The data presented in the study are deposited in the NCBI SRA repository with the link <http://www.ncbi.nlm.nih.gov/bioproject/941822> and accession number - PRJNA941822.

Author contributions

AA—conceptualization, data curation, formal analysis, methodology, validation, visualization, writing—original draft, and writing—reviewing and editing. LS—conceptualization, investigation, project administration, resources, software, supervision, and validation. PK—writing—review and editing.

References

- Amrutha, M., and Nampoothiri, K. M. (2022). *In silico* analysis of nitrilase-3 protein from *Corynebacterium glutamicum* for bioremediation of nitrile herbicides. *J. Genet. Eng. Biotechnol.* 20 (1), 51. doi:10.1186/s43141-022-00332-5
- Andrews, S. (2010). *FastQC: A quality control tool for high throughput sequence data*. Cambridge, United Kingdom: Babraham Bioinformatics, Babraham Institute.
- Barman, U. D., Saha, S. K., Kader, M. A., Jamal, M. A. H. M., Sharma, S. P., Samad, A., et al. (2020). Clinicopathological and prognostic significance of GPC3 in human breast cancer and its 3D structure prediction. *Netw. Model. Anal. Heal. Inf. Bioinforma.* 9 (1), 24. doi:10.1007/s13721-020-00234-x
- Bekker, H., Berendsen, H. J. C., Van Druenen, R., and Van Der Spoel, D. (1993). *Gromacs: A parallel computer for molecular dynamics simulations*. SINGAPORE: World Scientific Publishing.
- Bin Kwong, Q., Ong, A. L., and Tammi, M. (2017). Inspection of sequence quality. *Bioinformatics* 2017, 49–66. doi:10.1142/9789813144750_0003
- BIOCATALYSIS OF DIFFERENT NITRILES TO VALUABLE ACIDS (2021). “Smally sinha and vinod kumar nigam,” in *Department of bio-engineering* (Mesra, Ranch: Birla Institute of Technology), 2662–2669. doi:10.13040/IJPSR.0975-8232.12(5). 2662-69
- Brunner, S., Eppinger, E., Fischer, S., Gröning, J., and Stolz, A. (2018). Conversion of aliphatic nitriles by the arylacetone nitrilase from *Pseudomonas fluorescens* EBC191. *World J. Microbiol. Biotechnol.* 34 (7), 91. doi:10.1007/s11274-018-2477-9
- Burley, S. K., Berman, H. M., Kleywegt, G. J., Markley, J. L., Nakamura, H., and Velankar, S. (2017). Protein Data Bank (PDB): The single global macromolecular structure archive. *Methods Mol. Biol.* 1607, 627–641. doi:10.1007/978-1-4939-7000-1_26
- Castro-Alvarez, A., Costa, A. M., and Vilarrasa, J. (2017). The performance of several docking programs at reproducing protein-macrolide-like crystal structures. *Molecules* 22, 136. doi:10.3390/molecules22010136
- Eberhardt, J., Santos-Martins, D., Tillack, A. F., and Forli, S. (2021). AutoDock Vina 1.2.0: New docking methods, expanded force field, and Python bindings. *J. Chem. Inf. Model.* 61 (8), 3891–3898. doi:10.1021/acs.jcim.1c00203
- Ewels, P., Magnusson, M., Lundin, S., and Käller, M. (2016). MultiQC: Summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* 32 (19), 3047–3048. doi:10.1093/bioinformatics/btw354
- Gasteiger, B. A., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M. R., and Appel, R. D. (2005). “Protein identification and analysis tools on the ExPASy server,” in *Proteomics protoc. Handbook* (Humana Press), 571–667.
- Gasteiger, E., Gattiker, A., Hoogland, C., Ivanyi, I., Appel, R. D., and Bairoch, A. (2003). ExPASy: The proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res.* 31 (13), 3784–3788. doi:10.1093/nar/gkg563
- Goddard, T. D., Huang, C. C., Meng, E. C., Pettersen, E. F., Couch, G. S., Morris, J. H., et al. (2018). UCSF ChimeraX: Meeting modern challenges in visualization and analysis. *Protein Sci.* 27 (1), 14–25. doi:10.1002/pro.3235
- Gong, J. S., Lu, Z. M., Li, H., Shi, J. S., Zhou, Z. M., and Xu, Z. H. (2012). Nitrilases in nitrile biocatalysis: Recent progress and forthcoming research. *Microb. Cell. Fact.* 11, 142–218. doi:10.1186/1475-2859-11-142
- Gurevich, A., Saveliev, V., Vyahhi, N., and Tesler, G. (2013). Quast: Quality assessment tool for genome assemblies. *Bioinformatics* 29 (8), 1072–1075. doi:10.1093/bioinformatics/btt086
- Hoie, M. H., Kiehl, E. N., Petersen, B., Nielsen, M., Winther, O., Nielsen, H., et al. (2022). NetSurfP-3.0: Accurate and fast prediction of protein structural features by protein language models and deep learning. *Nucleic Acids Res.* 50 (W1), W510–W515. doi:10.1093/nar/gkac439
- Johnson, M., Zaretskaya, I., Raytselis, Y., Merezukh, Y., McGinnis, S., and Madden, T. L. (2008). NCBI blast: A better web interface. *Nucleic Acids Res.* 36, W5–W9. doi:10.1093/nar/gkn201
- Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., et al. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature* 596 (7873), 583–589. doi:10.1038/s41586-021-03819-2
- Kim, S., Chen, J., Cheng, T., Gindulyte, A., He, J., He, S., et al. (2020). PubChem in 2021: New data content and improved web interfaces. *Nucleic Acids Res.* 49 (D1), D1388–D1395. doi:10.1093/nar/gkaa971
- Klasberg, S., Bitard-Feildel, T., and Mallet, L. (2016). Computational identification of novel genes: Current and future perspectives. *Bioinform. Biol. Insights* 10, 121–131. doi:10.4137/BBIS.39950
- Kupke, F., Herz, C., Hanschen, F. S., Platz, S., Odongo, G. A., Helmig, S., et al. (2016). Cytotoxic and genotoxic potential of food-borne nitriles in a liver *in vitro* model. *Sci. Rep.* 6, 37631–37711. doi:10.1038/srep37631

Acknowledgments

The authors would like to thank their organization for providing the opportunity to pursue this research. The authors thank the supervisor and co-authors for their cooperation and encouragement.

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.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmolb.2023.1123902/full#supplementary-material>

- Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993). Procheck: A program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* 26 (2), 283–291. doi:10.1107/S0021889892009944
- Laskowski, R. A., and Swindells, M. B. (2011). LigPlot+: Multiple ligand-protein interaction diagrams for drug discovery. *J. Chem. Inf. Model.* 51 (10), 2778–2786. doi:10.1021/ci200227u
- Jones, L. B., Wang, X., Gullapalli, J. S., and Kunz, D. A. (2020). Characterization of the Nit6803 nitrilase homolog from the cyanotroph *Pseudomonas fluorescens* NCIMB 11764. *Biochem. Biophys. Rep.* 25, 100893. doi:10.1016/j.bbrep.2020.100893
- Li, C., Li, Y., Cheng, X., Feng, L., Xi, C., and Zhang, Y. (2013). Immobilization of *Rhodococcus rhodochrous* BX2 (an acetonitrile-degrading bacterium) with biofilm-forming bacteria for wastewater treatment. *Bioresour. Technol.* 131, 390–396. doi:10.1016/j.biortech.2012.12.140
- Li, D., Liu, C. M., Luo, R., Sadakane, K., and Lam, T. W. (2015). Megahit: An ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* 31 (10), 1674–1676. doi:10.1093/bioinformatics/btv033
- Li, M. H., Luo, Q., Xue, X. G., and Li, Z. S. (2011). Molecular dynamics studies of the 3D structure and planar ligand binding of a quadruplex dimer. *J. Mol. Model.* 17 (3), 515–526. doi:10.1007/s00894-010-0746-0
- Marchler-Bauer, A., Bo, Y., Han, L., He, J., Lanczycki, C. J., Lu, S., et al. (2017). CDD/SPARCLE: Functional classification of proteins via subfamily domain architectures. *Nucleic Acids Res.* 45 (D1), D200–D203. doi:10.1093/nar/gkw1129
- Nigam, V. K., Arfi, T., Kumar, V., and Shukla, P. (2017). Bioengineering of nitrilases towards its use as green catalyst: Applications and perspectives. *Indian J. Microbiol.* 57 (2), 131–138. doi:10.1007/s12088-017-0645-5
- Pettersen, E. F., Goddard, T. D., Huang, C. C., Meng, E. C., Couch, G. S., Croll, T. I., et al. (2021). UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Sci.* 30 (1), 70–82. doi:10.1002/pro.3943
- Quevillon, E., SilVentoinen, V., Pillai, S., Harte, N., Mulder, N., Apweiler, R., et al. (2005). InterProScan: Protein domains identifier. *Nucleic Acids Res.* 33, W116–W120. doi:10.1093/nar/gki442
- Raczynska, J. E., Vorgias, C. E., Antranikian, G., and Rypniewski, W. (2011). Crystallographic analysis of a thermoactive nitrilase. *J. Struct. Biol.* 173 (2), 294–302. doi:10.1016/j.jsb.2010.11.017
- Egelkamp, R., Zimmermann, T., Schneider, D., Hertel, R., and Daniel, R. (2019). Impact of nitriles on bacterial communities. *Front. Environ. Sci.* 7, 1–14. doi:10.3389/fenvs.2019.00103
- Schrödinger, L. (2000). The PyMOL molecular graphics system. Version 1.2r3pre.
- Sewell, B. T., Berman, M. N., Meyers, P. R., Jandhyala, D., and Benedik, M. J. (2003). The cyanide degrading nitrilase from *Pseudomonas stutzeri* AK61 is a two-fold symmetric, 14-subunit spiral. *Structure* 11 (11), 1413–1422. doi:10.1016/j.str.2003.10.005
- Tamames, J., and Puente-Sánchez, F. (2019). SqueezeMeta, A highly portable, fully automatic metagenomic analysis pipeline. *Front. Microbiol.* 9, 3349. doi:10.3389/fmicb.2018.03349
- Tamura, K., Stecher, G., and Kumar, S. (2021). MEGA11: Molecular evolutionary genetics analysis version 11. *Mol. Biol. Evol.* 38 (7), 3022–3027. doi:10.1093/molbev/msab120
- Tanii, H. (2017). Allyl nitrile: Toxicity and health effects. *J. Occup. Health* 59 (2), 104–111. doi:10.1539/joh.16-0147-RA
- Tian, W., Chen, C., Lei, X., Zhao, J., and Liang, J. (2018). CASTp 3.0: Computed atlas of surface topography of proteins. *Nucleic Acids Res.* 46 (W1), W363–W367. doi:10.1093/nar/gky473
- Veselá, A. B., Franc, M., Pelantova, H., Kubac, D., VejVoda, V., Sulc, M., et al. (2010). Hydrolysis of benzonitrile herbicides by soil actinobacteria and metabolite toxicity. *Biodegradation* 21 (5), 761–770. doi:10.1007/s10532-010-9341-4
- Wang, M. X. (2015). Enantioselective biotransformations of nitriles in organic synthesis. *Acc. Chem. Res.* 48 (3), 602–611. doi:10.1021/ar500406s
- Zhang, L., Yin, B., Wang, C., Jiang, S., Wang, H., Yuan, Y. A., et al. (2014). Structural insights into enzymatic activity and substrate specificity determination by a single amino acid in nitrilase from *Syechocystis* sp. PCC6803. *J. Struct. Biol.* 188 (2), 93–101. doi:10.1016/j.jsb.2014.10.003



OPEN ACCESS

EDITED BY

Chandrabose Selvaraj,
Alagappa University, India

REVIEWED BY

Ramar Vanajothi,
Bharathidasan University, India
Quratulain Hasan,
Kamineni Hospitals, India

*CORRESPONDENCE

Eman Alefishat,
✉ eman.alefishat@ku.ac.ae
Zuhier Awan,
✉ zawan@kau.edu.sa
Babajan Banaganapalli,
✉ bbabajan@kau.edu.sa

[†]These authors have contributed equally to this work

RECEIVED 11 April 2023

ACCEPTED 12 June 2023

PUBLISHED 04 July 2023

CITATION

Shaik NA, Al-Shehri N, Athar M, Awan A, Khalili M, Al Mahadi HB, Hejazy G, Saadah OI, Al-Harhi SE, Elango R, Banaganapalli B, Alefishat E and Awan Z (2023), Protein structural insights into a rare PCSK9 gain-of-function variant (R496W) causing familial hypercholesterolemia in a Saudi family: whole exome sequencing and computational analysis. *Front. Physiol.* 14:1204018. doi: 10.3389/fphys.2023.1204018

COPYRIGHT

© 2023 Shaik, Al-Shehri, Athar, Awan, Khalili, Al Mahadi, Hejazy, Saadah, Al-Harhi, Elango, Banaganapalli, Alefishat and Awan. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). 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.

Protein structural insights into a rare PCSK9 gain-of-function variant (R496W) causing familial hypercholesterolemia in a Saudi family: whole exome sequencing and computational analysis

Noor Ahmad Shaik^{1,2†}, Najla Al-Shehri^{3†}, Mohammad Athar^{4,5}, Ahmed Awan³, Mariam Khalili⁶, Hadiyah Bassam Al Mahadi⁷, Gehan Hejazy³, Omar I. Saadah⁸, Sameer Eida Al-Harhi⁹, Ramu Elango^{1,2}, Babajan Banaganapalli^{1,2*}, Eman Alefishat^{6,10,11*} and Zuhier Awan^{3*}

¹Department of Genetic Medicine, Faculty of Medicine, King Abdulaziz University, Jeddah, Saudi Arabia,

²Princess Al-Jawhara Center of Excellence in Research of Hereditary Disorders, King Abdulaziz University,

Jeddah, Saudi Arabia, ³Department of Clinical Biochemistry, Faculty of Medicine, King Abdulaziz

University, Jeddah, Saudi Arabia, ⁴Department of Medical Genetics, Faculty of Medicine, Umm Al-Qura

University, Makkah, Saudi Arabia, ⁵Science and Technology Unit, Umm Al-Qura University, Makkah, Saudi

Arabia, ⁶Department of Pharmacology, College of Medicine, Khalifa University, Abu Dhabi, United Arab

Emirates, ⁷Department of Research and Development, Al Borg Medical Laboratories, Jeddah, Saudi

Arabia, ⁸Department of Pediatrics, Pediatric Gastroenterology Unit, Faculty of Medicine, King Abdulaziz

University, Jeddah, Saudi Arabia, ⁹Department of Clinical Pharmacology, Faculty of Medicine, King

Abdulaziz University, Jeddah, Saudi Arabia, ¹⁰Department of Biopharmaceutics and Clinical Pharmacy,

Faculty of Pharmacy, The University of Jordan, Amman, Jordan, ¹¹Center for Biotechnology, Khalifa

University, Abu Dhabi, United Arab Emirates

Familial hypercholesterolemia (FH) is a globally underdiagnosed genetic condition associated with premature cardiovascular death. The genetic etiology data on Arab FH patients is scarce. Therefore, this study aimed to identify the genetic basis of FH in a Saudi family using whole exome sequencing (WES) and multidimensional bioinformatic analysis. Our WES findings revealed a rare heterozygous gain-of-function variant (R496W) in the exon 9 of the *PCSK9* gene as a causal factor for FH in this family. This variant was absent in healthy relatives of the proband and 200 healthy normolipidemic controls from Saudi Arabia. Furthermore, this variant has not been previously reported in various regional and global population genomic variant databases. Interestingly, this variant is classified as “likely pathogenic” (PP5) based on the variant interpretation guidelines of the American College of Medical Genetics (ACMG). Computational functional characterization suggested that this variant could destabilize the native *PCSK9* protein and alter its secondary and tertiary structural features. In addition, this variant was predicted to negatively influence its ligand-binding ability with *LDLR* and Alirocumab antibody molecules. This rare *PCSK9* (R496W) variant is likely to expand our understanding of the genetic basis of FH in Saudi Arabia. This study also provides computational structural insights into the genotype-protein phenotype relationship of *PCSK9* pathogenic variants and contributes to the development of personalized medicine for FH patients in the future.

KEYWORDS

familial hypercholesterolemia, cardiovascular diseases, whole exome sequence, sanger sequence, pcsk9

1 Introduction

Familial hypercholesterolemia (FH) is a globally underdiagnosed genetic condition characterized by life-long elevation of low-density lipoprotein cholesterol (LDL-C) in plasma (≥ 190 mg/dL) (Alhabib et al., 2021). FH results from inherited pathogenic mutations in genes that regulate hepatic LDL-C clearance and cholesterol metabolism (Mlinaric et al., 2020; Alhabib et al., 2021). FH begins to manifest at birth, and if untreated, the chronic elevation of lipids in the blood leads to plaque formation in arterial walls, subsequently accelerating atherosclerosis and increasing the risk of developing premature coronary artery disease (Alallaf et al., 2017; Hooper et al., 2018). Early detection and therapeutic intervention are critical to delay or prevent cardiovascular morbidity and mortality in patients with FH (Blanco-Vaca et al., 2018).

Majority of clinically diagnosed FH patients carry autosomal dominant loss-of-function (LOF) mutations in the low-density lipoprotein receptor (*LDLR*) and apolipoprotein-B (*APOB*) genes or gain-of-function (GOF) mutations in the proprotein convertase subtilisin/kexin type 9 (*PCSK9*) gene (Alallaf et al., 2017; Awan et al., 2019; Al-Waili et al., 2020). A rare form of autosomal recessive FH caused by loss-of-function (LOF) mutations in the low-density lipoprotein receptor adaptor protein 1 (*LDLRAP1*) gene has been reported (Paththinige et al., 2017; Chan et al., 2019). Interestingly, most of the mutations occur in *LDLR* (80%), whereas only 5% occur in *APOB* and 3% occur in *PCSK9* (Alallaf et al., 2017). Recent evidence supports the presence of other genes with numerous pathogenic variants, with either a causative or contributory role in FH pathogenesis. These genes include adenosine triphosphate-binding cassette transporters G5 and G8 (*ABCG5* and *ABCG8*), lipase A (*LIPA*), apolipoprotein E (*APOE*), signal transducing adaptor family member 1 (*STAP1*), cholesteryl ester transfer protein (*CETP*), and sterol regulatory element binding transcription factor 2 (*SREBP2*) (de Grooth et al., 2004; Paththinige et al., 2017; Blanco-Vaca et al., 2018; Ibrahim et al., 2021).

The current global prevalence of FH is estimated to be 1:300 (Beheshti et al., 2020), and its occurrence among founder populations and specific ethnicities is even higher (Alallaf et al., 2017; Sturm et al., 2018). This is particularly true for genetically isolated populations, such as in Saudi Arabia, where the high rate of consanguineous marriage plays a significant role in the higher incidence of genetic disorders within the population (Alallaf et al., 2017; Awan et al., 2019a). As a culturally distinct population, Saudi Arabia has been reported to have a 3-fold higher prevalence of FH (Awan et al., 2022). However, genetic data on FH, particularly founder FH mutations, among Saudi Arabian patients is scarce (Sturm et al., 2018; Al-Allaf et al., 2017; Alallaf et al., 2017; Awan et al., 2021b). In recent years, few whole exome sequencing (WES) and target gene panel testing-based studies have reported a few rare variants in *LDLR* (Awan et al., 2022; Chan et al., 2018; Beheshti et al., 2020) and *APOB* (Banaganapalli

et al., 2017; Awan et al., 2021a) genes, however, the role of *PCSK9* gene variants in Saudi FH patients is not yet reported.

Molecular-level understanding and functional characterization of the effect of variants on candidate proteins are essential for the development of suitable therapeutic strategies (Banaganapalli et al., 2017). However, the experimental characterization of every variant is a time-consuming, expensive, and complex process that requires a variety of skills and infrastructure. However, the expansion of computational biology applications and artificial intelligence algorithms in genomics has led to the development of various web servers and software that can perform better molecular, structural, and functional analyses of plausible disease-causing genes and variants within a short time with limited resources (Awan et al., 2021a). For example, graph-based algorithms have been shown to aid in disease-causative gene selection by removing irrelevant and redundant genes using different criteria (Harling-Lee et al., 2022) (Saberi-Movahed et al., 2022). Although numerous FH-causative *LDLR*, *APOB*, and *PCSK9* variants have been reported in the literature, detailed bioinformatics-based genotype-protein phenotype characterization is lacking (Awan et al., 2019a; Meshkov et al., 2021; Guo et al., 2020a).

The main goal of this study is to identify the pathogenic variant causing FH in Saudi families through whole exome sequencing. Moreover, this study also aims to utilize computational biology methods to understand the relationship between variant macromolecular structures and their function, and role in FH.

2 Materials and methods

2.1 Recruitment of FH patients and their families

In this study, a Saudi Arabian family was clinically examined in the dyslipidemia clinic at King Abdulaziz University Hospital (KAUH) using the combined Simon Broome Register and the Dutch Lipid Clinic Network (DLCN) criteria for FH diagnosis. The family was then referred for molecular testing. This research work was approved by the Institutional Research Ethics Committee. All participants analyzed in this study gave their written and informed consent after briefing them about the study design, potential risks of any discomfort, and benefits. After careful interviews with the family members, collected a detailed family history of FH and other cardiovascular abnormalities. Plasma lipid profile data (LDL-C, high-density lipoprotein cholesterol (HDL-C), triglycerides, and total cholesterol) of the family were collected from health records. For molecular testing, approximately 5 mL (ml) of venous blood samples from whole family and 200 Saudi control participants were collected. All control subjects had a normal lipid profile (triglycerides < 1.70 mmol/L, total cholesterol < 5 mmol/L, HDL-C 1.40 – 1.55 mmol/L, and LDL-C 2.50 – 4.11 mmol/L) as per their health records. They were recruited after thoroughly inquiring

about their clinical and family history during personal interviews.

2.2 Molecular genetic analysis

2.2.1 DNA isolation

DNA (deoxyribonucleic acid) was isolated from white blood cells of the peripheral blood samples following the manufacturer's protocol (Haven Scientific DNABler-Blood Kit), and then the purity and concentration of DNA based on optical density ratios of 260/230 and 260/280 were assessed using Nanodrop (DeNovix DS-11 Series) Spectrophotometer.

2.2.2 Whole exome sequencing (WES)

In brief, genomic DNA (100 ng/μl) was used to prepare DNA libraries (Ion AmpliSeq™ Library Kit 2.0), which were sequenced on the Ion GeneStudio™ S5 System with 30X coverage. All raw sequence reads were aligned against the human genome reference (GRC38, NCBI) using the Torrent Suite Software 5.4. Finally, a variant caller plug-in is used for variant calling. Then, the output data was filtered with minor allele frequency (MAF) = < 0.01 in global databases, such as the genome aggregation database (gnomAD), 1,000 genomes, and searched in local databases such as the Greater Middle Eastern Genome (GME) and the King Abdullah International Medical Research Center (KAIMRC) Genome Database for Arab population frequencies. Variants with Phred scores of >20 were selected for analysis (Alves-Bezerra and Cohen, 2017; Bener and Mohammad, 2017; Gaboon et al., 2020; Kamar et al., 2021) (Supplementary Figure S1). From this list, heterozygous variants in the coding regions of known FH causative genes were retained for downstream analyses.

2.2.3 Variant validation and segregation analysis using dideoxy sequencing method

The WES identified a potential FH variant which was further validated using the deoxy—Sanger sequencing method. In this context, oligonucleotide primer sets (forward: 5'-TGTTCTTTAAGCCCTCCTCTC-3' and reverse: 5'-AGAGCTGGAGTCTGGAGGAT-3') spanning at least 100 bp upstream and downstream of the variant location were designed using the Primer-BLAST website hosted by the National Center for Biotechnology Information (NCBI) (<https://www.ncbi.nlm.nih.gov/tools/primer-blast/>). These primers were tested using the PCR primer stats tool (Supplementary Table S1) (https://www.bioinformatics.org/sms2/pcr_primer_stats.html) to assess the melting temperature (>50°C), guanine-cytosine (GC) ratio (>40%), primer length (18–22 bp), self-annealing, and hairpin loop formation parameters. The target gene region was amplified by polymerase chain reaction (PCR) (Veriti™ 96-Well Thermal Cycler, Applied Biosystems, United States) and bidirectionally sequenced using the Sanger method (Applied Biosystems SeqStudio Genetic Analyzer) with both forward and reverse primers. To identify and annotate the variant position and nature, the generated sample sequences were aligned against the reference candidate gene sequence using the BioEdit 7.2 computational program. Based on the sequencing results, we

were able to determine variant segregation patterns among the family members and control subjects (Awan et al., 2022).

2.3 Bioinformatic analysis of FH variant

2.3.1 Sequence based annotation

Variant pathogenicity was determined based on the prediction scores of SIFT (Sorting Intolerant from Tolerant), PolyPhen2 (Polymorphism Phenotyping), and Loss-of-Function Tool (loFtool) computational tools. In this context, we entered query variant details such as reference mRNA sequence or nucleotide sequence position in the Variant Effect Predictor (VEP) tool hosted on the Ensembl web server. From the VEP output, we selected the prediction scores of the SIFT, PolyPhen-2, and loFtools. In addition, the variant was queried on the Franklin webserver (<https://franklin.genoox.com/clinical-db/home>) to classify it based on the ACMG criteria with the existing population, computational, functional, segregation, *de novo*, and allelic data (Harrison et al., 2019).

2.3.2 Structure based annotations

2.3.2.1 3D protein modeling

The crystal structure of PCSK9 (PDB code 2P4E, resolution 1.98) was initially obtained for molecular modeling from the Protein Data Bank (PDB). The Swiss model server was used to build mutant models using the PCSK9 crystal structure as a template, and the quality of the generated model was assessed using the global model quality estimation (GMQE) score. In addition, an artificial intelligence program named AlphaFold generated 3D structure was used to improve the accuracy and dependability of the protein structure predictions.

AlphaFold, a DeepMind advanced deep learning system, predicts protein folding and generates accurate 3D protein structure models by combining deep neural networks and unique computational approaches. The AlphaFold models were compared to the PCSK9 crystal structure and mutant models created with the Swiss model server. To evaluate the stereochemical quality of the wild-type and mutant models, we used multiple programs, including Procheck, Verify3D, and ERRAT (<https://saves.mbi.ucla.edu/>), as reported by Laskowski et al. (1996). Additionally, the altered models were improved using Swiss PDB Viewer 3.5 and energy minimization (steepest descent). Finally, using PyMOL software, all of the created protein models, including those generated by AlphaFold, were displayed and analyzed (Grell et al., 2006).

2.3.2.2 Secondary structure and protein stability analysis

The secondary structures of wild-type and mutant proteins were generated using the PDBsum web server with amino acid coordinates as inputs to analyze the differences in their secondary structures, such as α-helices, β-pleated sheets, and loops. The impact of the genetic variant on the stability of the corresponding protein was analyzed using the Multi-AgEnt Stability pRedictiOn (MAESTRO) web server by providing 3D structures of both native and variant protein forms (Laimer et al., 2015; Awan et al., 2021a).

2.3.2.3 Molecular docking with LDLR and monoclonal antibody

The ClusPro server, which relies on the Fast Fourier Transform (FFT) algorithm-based docking program (PIPER16) (Kozakov et al.,

2006), was used to perform molecular docking between *LDLR* (PDB ID: 3P5B) and the query protein. This docking step provides ten candidate models for both proteins (wildtype and mutant), with cluster scores ranging from 0 to 9 as the best values, ranking them from the heaviest “0” cluster size to the lightest “9” (Kozakov et al., 2017; Vajda et al., 2017; Desta et al., 2020). For antibody docking, Alirocumab (CID 88214187), a fully human monoclonal antibody, was used to perform molecular docking with query PCKS9, in both native and mutant conditions, using the DockThor webserver, the blind docking approach with grid centers on the x, y, and z-axes of 37.42, 24.71, and 34.51, respectively, with a maximum grid size of 40 for the three-dimensional grid (Reyes-Soffer et al., 2017). This web server relies on phenotypic crowding-based multiple solution steady-state genetic algorithms that adhere to the following parameters: 24 docking runs, 1,000,000 evaluations per docking run, a population of 1,000 individuals, and a maximum of 100 leaders on each docking procedure. The affinity of the interaction between the query protein and ligand molecules was represented in the form of a total energy (E_{total}) score that calculates the sum of the van der Waals and electrostatic potentials between the 1–4 atom pairs (Guedes et al., 2021).

3 Results

3.1 Clinical presentation

A 59-year-old woman from the western region of Saudi Arabia was clinically diagnosed with FH in 2018 with a positive family history (II.2). She initially visited the KAUH emergency room with a complaint of shortness of breath. She was diagnosed with coronary ischemia and referred to the dyslipidemia clinic, where she was

treated for hypercholesterolemia, since her late twenties. She had smoked for 30 years, had a corneal wheel in her eyes, and no Achilles tendon growth was found (Figure 1). She was clinically diagnosed as heterozygous FH (HeFH). Her (I.1) father had premature coronary artery disease (CAD) and died in his 40s. However, we could not access any of his clinical records. The biochemical tests of the proband (II.2) show high levels of LDL-C (225 mg/dl), total cholesterol (355.2 mg/dl), and triglycerides (142.6 mg/dl), with a normal HDL-C level (66.6 mg/dl). She had been on statins (atorvastatin/20 g per day), which inhibit cholesterol synthesis, since her clinical diagnosis. No evidence of xanthomas (cholesterol deposits) was found on her tendons, elbows and knees. She is the mother of four daughters: III.2 (39 years old), III.4 (37 years old), III.6 (21 years old), and III.7 (16 years old). Despite having a high lipid profile, two sisters (III.6 and III.7) refused genetic testing due to privacy concerns. Furthermore, we could not test the offspring of III.2 (IV.1 and IV.2), III.4 (IV.3, IV.4, and IV.5), and III.6 (IV.6, IV.7, and IV.8) because of the lack of interest from those families. The abnormal plasma lipid profiles of III.6 and III.7 suggest that they are likely to have familial hypercholesterolemia (FH). They may have inherited the disease from their mother, II.2, in an autosomal dominant manner. However, this family is lost to clinical follow-up; hence, we could not ascertain our clinical assumptions and their current health statuses.

3.2 Genetic analysis

3.2.1 Whole exome sequence (WES) results

The exome sequencing of the proband generated approximately 67,610 variants, of which only approximately

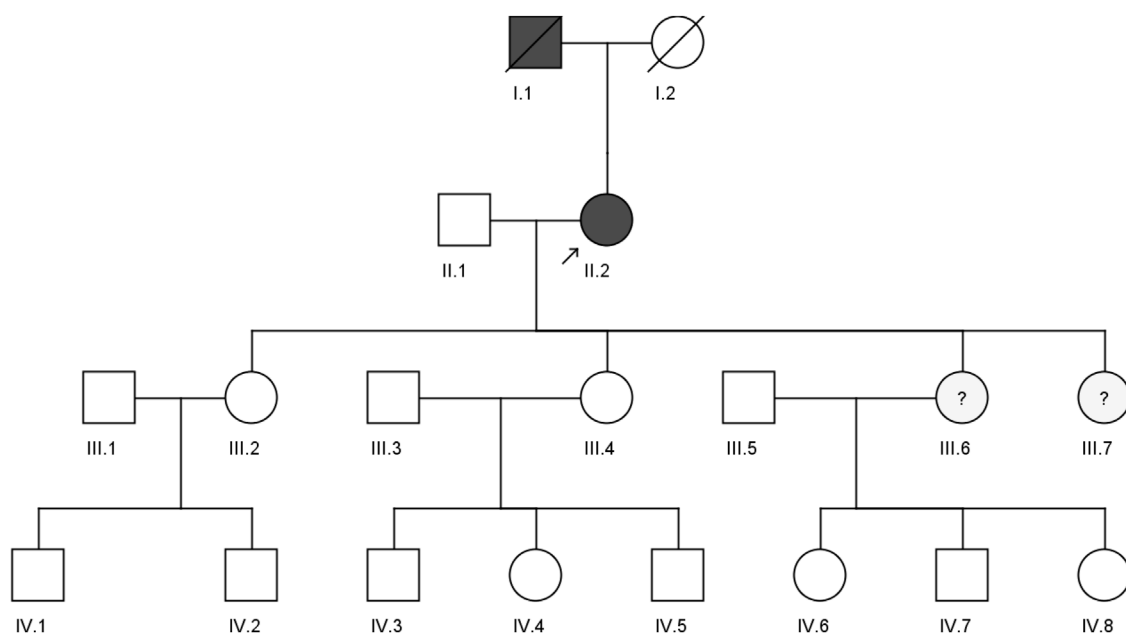


FIGURE 1

Pedigree of a four-generation Saudi family with a clinically diagnosed FH patient. The arrow indicates the index case, which was screened by both whole exome sequencing and sanger sequencing, while II.1, III.2, and III.4 were screened by the sanger sequencing method.

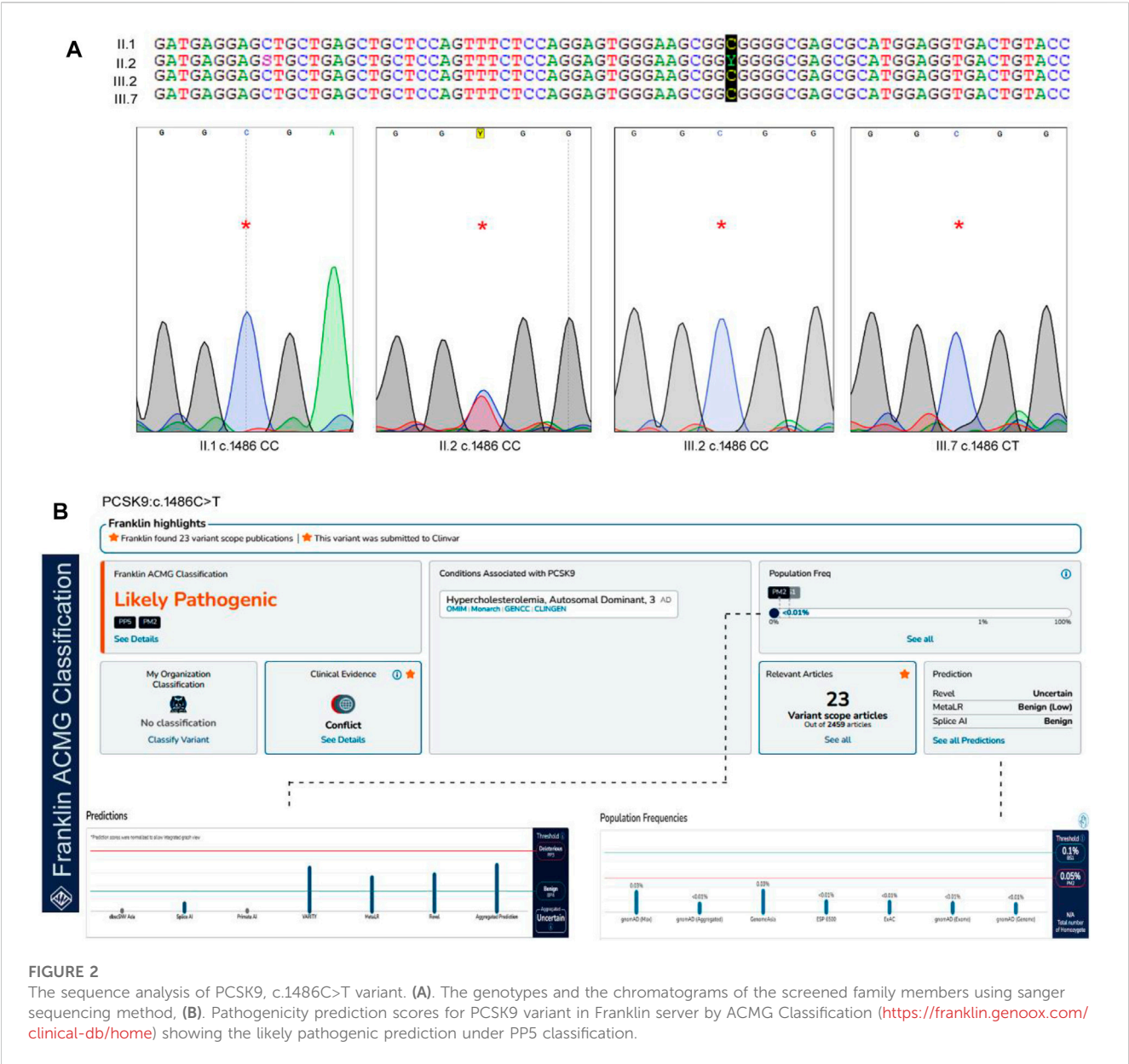


FIGURE 2 The sequence analysis of PCSK9, c.1486C>T variant. (A). The genotypes and the chromatograms of the screened family members using sanger sequencing method. (B). Pathogenicity prediction scores for PCSK9 variant in Franklin server by ACMG Classification (<https://franklin.genoox.com/clinical-db/home>) showing the likely pathogenic prediction under PP5 classification.

20,202 were localized in the coding region. These variants were filtered based on: (i) mapping to known FH causative genes (*LDLR*, *APOB*, *PCSK9*, *LDLRAP1*, *APOE*, *ABCG5*, *ABCG8*, and *LIPA*); (ii) zygosity; (iii) minor allele frequency of less than 1% (MAF <0.01) in the global population databases; and (iv) Phred score of >20. Based on the filtering criteria mentioned above, we identified a rare heterozygous missense variant at c.1486C>T (rs374603772) in the *PCSK9* gene in exon 9, causing a substitution of the amino acid residue R to W at the 496th position in the protein. In the gnomAD database, the global MAF of this variant was 0.00004 (12 heterozygotes of 139,471 individuals). In South Asian populations MAF is 0.0002 (8/15,254), and 0.00003 (4/63,319) in Europeans (non-Finnish). Surprisingly, this variant has not yet been reported in other global populations. Additionally, this variant was not detected in the 9497 exomes of GME or the 1,563 exomes of the KAIMRC database, the majority of them Arab nationals, if not Saudi Arabians.

3.2.2 Sanger sequencing results

The Sanger sequencing results of the family confirmed that the proband (II.2) was a heterozygous carrier of the c.1486C>T variant, but none of the screened family members (II.1, III.2, and III.4), nor the 200 healthy control samples had this variant (Figure 2). We could not confirm the variant segregation pattern in this family because the rest of the family members refused to participate in this study, although two of the proband's daughters (III.6 and III.7) had abnormal lipid profile that fulfilled the laboratory test diagnostic criteria for FH.

3.2.3 Genotype-phenotype correlations

From the biochemical and genetic data mentioned in Table 1, we can observe that individuals with the c.1486C/T genotype (II.2) have significantly elevated cholesterol, triglycerides, and LDL-C levels compared to individuals with the c.1486C/C genotype (II.1, III.2, and III.4). This suggests that the c.1486C/T genotype is associated

TABLE 1 Biochemical findings of FH in the index case and the family members.

Family members	Genotype	Cholesterol (mg/dL)		HDL-C (mg/dL)		Triglycerides (mg/dL)		LDL-C (mg/dL)	
II.1 ●	c.1486C/C	101.8	Reference: below 200 mg/dL	25.4	Reference: 60 mg/dL or above	158.7	Reference: above 150 mg/dL	45	Reference: below 100 mg/dL
II.2* ●	c.1486C/T	355.2		66.6		318		225	
III.2 ●	c.1486C/C	178.1		47.1		75.8		116	
III. 4 ●	c.1486C/C	188.6		67.1		55.8		140	
III.6	-	219.2		47.17		32.5		272.4	
III.7	-	296.7		51.58		97.5		243.32	

Note: LDL-C: Low-density lipoprotein cholesterol, TC: total cholesterol, TG: triglycerides, HDL-C: High-density lipoprotein cholesterol. * Exome sequenced individuals. ● Sanger sequenced individuals.

with a more severe lipid profile in the proband. However, further analysis and larger sample sizes are needed to draw more definitive conclusions about the genotype-phenotype correlation.

3.3 Bioinformatic analysis

3.3.1 Pathogenicity prediction

In our study, we utilized two widely used computational tools, SIFT and PolyPhen2, to assess the potential impact of the c.1486C>T variant on protein function. SIFT predicts the tolerance of an amino acid substitution based on sequence conservation, while PolyPhen2 predicts the potential pathogenicity of a variant by considering multiple sequence and structure-based features. The c.1486C>T variant was assigned a SIFT prediction value of 0.08, which suggests that it is a tolerated variant. On the other hand, PolyPhen2 assigned a prediction value of 0.798 to the variant, indicating a higher likelihood of being deleterious. This discrepancy in the predictions suggests the need for further evaluation and interpretation. In our whole-exome analysis, we incorporated these SIFT and PolyPhen2 predictions to prioritize and annotate variants. The predictions from these tools provided valuable information regarding the potential functional impact of the c.1486C>T variant. However, it is important to note that computational predictions alone may not provide conclusive evidence and require independent evaluation. According to the ACMG guidelines, this variant is likely to be pathogenic under the PP5 classification. PP5 criteria refer to pathogenic variants, that require independent evaluation. This variant is extremely rare in the GnomAD databases, but it has previously been reported in the ClinVar and Uniprot databases with no supporting clinical evidence (Harrison et al., 2019).

3.3.2 3D protein modeling and secondary structure

Procheck Ramachandran plot analysis of the wild-type and mutant PCSK9 protein models showed that approximately 99.8% and 99.4% of the amino acids fell in the allowed regions, while 0.2% and 0.6% were in the disallowed regions, respectively. The overall structural quality of the protein models was confirmed using the ERRAT (Wild-type, 91.818; Mutant, 87.037) and Verify3D (Wildtype, 95.5% A.A. 3D score of = 0.2; Mutant, 94.2% A.A. 3D

score of = 0.2) scores (Supplementary Figure S2; Supplementary Table S2). Secondary structure analysis was performed using the PDBSUM web server on Swiss modeller built PCSK9 molecular models (wild-type and mutant forms) with a QGMEN score of >0.82. The native PCSK9 secondary structure is characterized by 12 beta chains, 1 beta-alpha-beta unit, 6 beta hairpins, 9 beta bulges, 38 strands, 14 helices, 7 helix-helix interactions, 47 beta turns, 7 gamma turns, and 12 disulfide bonds. In comparison to the native PCSK9, the R496W variant carrying protein has lost the 1 sheet and 1 helix-helix interaction but gained 2 beta hairpins, 3 beta bulges, 3 strands, 2 helices, 14 beta turns, and 3 gamma turns (Table 2). These findings suggest that the R496W variant of PCSK9 introduces structural alterations in the protein. The loss of a beta sheet and helix-helix interaction indicates potential disruptions in the overall protein folding and stability. On the other hand, the gained beta hairpins, beta bulges, strands, helices, beta turns, and gamma turns suggest that the mutation could influence the local structural elements and potentially affect the protein's functional properties.

3.3.3 Protein stability

We analyzed the changes in the stability of the mutant PCSK9 protein using the MAESTRO webserver. This computational program relies on statistical scoring functions (SSFs) and different machine learning approaches to quantify the extent of protein destabilization based on free energy values ($\Delta\Delta G_{pred}$). $\Delta\Delta G_{pred}$ values below 0.0 indicate a stable protein, whereas if above 0.0 indicate an unstable protein. The $\Delta\Delta G_{pred}$ value of the R496W variant was -0.032 kcal/mol, suggesting that R-to-W substitution may destabilize the PCSK9 protein. The confidence estimation score (Cpred) of this prediction was 0.93. On a scale of 0.0–1.0, any Cpred score closer to 1 corresponds to a perfect consensus of all prediction (support vector machines, artificial neural networks, and multiple linear regression agents) agents. (Laimer et al., 2015).

3.3.4 Molecular docking between PCSK9 and LDLR

The ClusPro software provided the cluster energy scores for ten docking models of the two proteins complexes, incorporating LDLR as the receptor and each of wildtype and mutant PCSK9 as ligands. The resulting complex of wild-type PCSK9 and LDLR complex has

TABLE 2 The difference secondary structure elements SSE of the wildtype and mutant PCSK9 proteins.

Secondary structure elements SSE	Wildtype PCSK9 protein	Mutant PCSK9 protein
	Total	Total
Beta Sheets	12	10
Beta alpha beta units	1	1
Beta hairpins	6	7
Beta bulges	9	9
Strands	38	36
Helices	14	13
Helix-Helix interacts	7	5
Beta turns	47	57
Gamma turns	7	9
Disulphide bond	12	12

23 hydrogen bonds between *LDLR* and the wild-type *PCSK9* protein. The number of hydrogen bonds between *LDLR* and the mutant *PCSK9* protein was reduced to 15 ([Supplementary Table S3](#)). The cluster energy value increased from $-1,340.4$ kJ/mol in the wildtype to $-1,214.6$ kJ/mol in the mutant *PCSK9-LDLR* complex, when focusing on the selected models (o), and the best candidate docking models ([Supplementary Table S3](#)). The molecular complexes of *LDLR* and *PCSK9* were stabilized by the formation of >20 hydrogen bonds at 3 bond distance.

The interaction between *LDLR* and *PCSK9* is mediated by specific amino acids. In the wild-type *PCSK9-LDLR* complex, key amino acid interactions included Ser153(P)-Asp299, Leu298(L), Ile154(P)-Leu298(L), Pro155(P)-Leu298(L), Asp238(P)-Asn295(L), Ile369(P)-Asn301(L), Ser372(P)-Leu318(L), Try374(P)-Leu318(Cys319, Pro320(L)), Cys375(P)-Leu318(L), Thr377(P)-Asn309(Asp310, Cys308(L)), Cys378(P)-Leu318(Val307, Cys308(L)), Phe379(P)-Val307(Cys308, Asn301, His306(L)), and Val380(P)-His306(L). In the mutant *PCSK9-LDLR* complex, the key amino acid interactions were found to be Glu84(P)-Lys811(L), Ser89(P)-Gln242(L), Arg93(P)-Gln242(L), Arg96(P)-Ser244(L), Arg104(P)-His769(L), Gly106(P)-Leu772(L), Gln254(P)-His87(L), Val277(P)-Lys283(L), Arg476(P)-Phe801(L), Pro479(P)-Asn812(L), Glu482(P)-Lys816(L), Gln554(P)-His837(L), Gln555(P)-Glu835(L), Thr573(P)-Asp748, Thr749(L), and His602(P)-Asp834(L).

These detailed amino acid interactions provide insights into the specific residues involved in the interaction between *LDLR* and *PCSK9*. The changes in hydrogen bonding and amino acid interactions in the mutant *PCSK9-LDLR* complex compared to the wild-type complex may have implications for the stability and functionality of the complex.

3.3.5 Molecular docking with alirocumab

The DockThor webserver was used to perform molecular docking between the antibody Alirocumab and the *PCSK9* protein (wildtype and mutant). The protein-antibody docking prediction values represented as total energy (Etotal), which is the sum of the van der Waals and electrostatic potential between protein-ligand atom pairs and torsion term of the ligand ([Santos et al., 2020](#)) ([Guedes et al., 2021](#)). The docking of wildtype *PCSK9* protein and the Alirocumab complex had a total energy of -64.497 kJ/mol (Etotal) with a score of -5.934 kJ/mol, while the mutant-monoclonal antibody complex had a total energy of -62.084 kJ/mol with a score of -6.839 kJ/mol ([Table 3](#); [Supplementary Tables S4, S5](#)).

4 Discussion

In this study, we detected a rare heterozygous c.1486C>T p. (R496W) gain of function variant in the exon 9 of the *PCSK9* gene in the proband (II.2), which was not found in the screened healthy family members or the 200 healthy normolipidemic Saudi Arabian controls. The *PCSK9* variants are usually classified into 2 types: gain-of-function (GOF) and loss-of-function (LOF), where the GOF variant causes a decrease in *LDLR* on hepatocytes, leading to the phenotypes of FH, whereas the LOF is involved in the low LDL-C levels that lower the risk of developing coronary heart disease with no known adverse effects on human health ([Bayona et al., 2020](#); [Guo et al., 2020b](#)).

The c.1486C>T variant was predicted to be a likely pathogenic (PP5) mutation according to the ACMG guidelines. These guidelines propose a five-tier categorization to evaluate the

TABLE 3 The docking scores calculated from the total energies of wildtype *PCSK9*, mutant *PCSK9* docking with Alirocumab.

Name	Score	Total energy kJ/mol	Van der Waals energy kJ/mol	Electrostatic energy kJ/mol
Wildtype <i>PCSK9</i> -Alirocumab	-5.934	-64.49	0.84	-40.48
Mutant <i>PCSK9</i> -Alirocumab	-6.839	-62.08	-5.27	-31.99

TABLE 4 A list of FH causative PCSK9 variants reported from Middle Eastern countries.

No.	Country	Variant ID	Chromosomal location	Coding regions change	Amino acid change	FH zygosity	Reference
1	Saudi Arabia	-	1:55,512,313	c.517delC	p.P173fs	HeFH	Al-Allaf et al. (2015)
2		rs509504	1:55,523,033	c.1026A>G	p.Q342H	HoFH	
3		rs540796	1:55,524,197	c.1380A>G	p.Y243C	HoFH	
4		rs562556	1:55,524,237	c.1420G>A	p.V474I	HoFH	
5		-	1:55,529,113	c.1935delG	p.L645fs	HeFH	
6		rs505151	1:55,529,187	c.2009G>A	p.G670E	HoFH	
7	Tunisia	rs505151	1:55,529,187	c.2009G>A	p.G670E	HoFH	Alhababi and Zayed (2018)
8		rs533273863	1: 55046643	c.520C>T	p.P174S	HeFH	
9		-	-	c.1545T > G	p.F515L	HeFH	Slimani et al. (2015)
10		rs72555377	1:55,039,880	c.65_66insGCTGCT	p.L22_L23dup	HeFH	
11	Lebanon	rs72555377	1:55,039,880	c.65_66insGCTGCT	p.L22_L23dup	HeFH	Abifadel et al. (2009)
12		rs11591147	1: 55,039,447	c.137G>T	p.R46L	HeFH	Bamimore et al. (2015)
13	Oman	rs562556	1: 55058564	c.1420G>A	I474V	HeFH	Al-Waili et al. (2013)

Mendelian disease variants into “pathogenic” (P), “likely pathogenic” (LP), “uncertain significance” (VUS), “likely benign” (LB), and “benign” (B). This classification considers allele frequency data, reputable sources, functional and computational data, case-level data, and nature and location of the variants (Harrison et al., 2019). The c.1486C>T variant was reported in FH patients from Italy (Pisciotta et al., 2006), Japan (Ohta et al., 2016), Spain (Ibarretxe et al., 2018), and Turkey (Eroğlu, 2018). All these studies have performed detailed clinical characterization of the c.1486C>T but did not report its functional impact on protein structure and function.

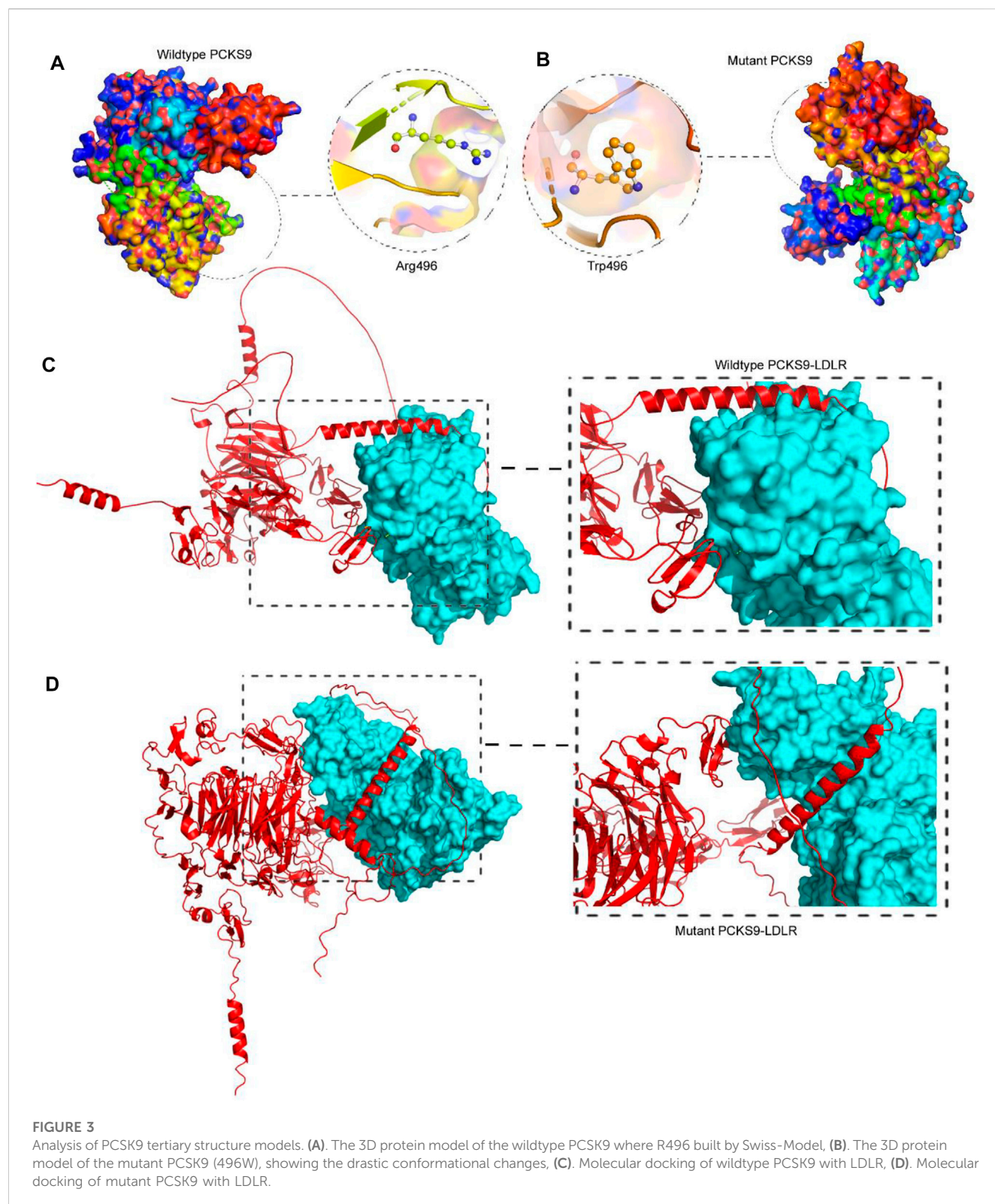
Furthermore, our findings suggest that c.1486C>T is an FH causative variant, as it was not found in Saudi and Middle Eastern regional population genetic databases like GME and KAIMRC and has a very low frequency of 12 het carriers in 139,471 exomes of the gnomAD (Table 4). Eight of these carriers were from South Asia (of 15,254 exomes) and 4 among the non-Finnish Europeans (of 63,319 exomes).

PCSK9 is a proteolytic enzyme and a member of the proprotein convertase family of serine proteases, that is mainly expressed in the liver (Guo et al., 2020b). The GOF variant c.1486C>T (p.R496W) is located in the C-terminal's first domain, which is the second reported domain with the highest GOF variants, as the majority of which are mapped to the subtilisin-like catalytic region of the Peptidase S8 domain, whereas the majority of LOF variants are in both regions of the Peptidase S8 domain (Bayona et al., 2020). Furthermore, variant c.1486C>T is a missense mutation that results in the substitution of arginine (R), an amino acid with a positively charged side chain, with tryptophan (W), a hydrophobic, aromatic side chain, in codon 496, resulting in different secondary structure elements (SSE), such as the absence of two beta sheets, two strands, one helix, and two helix-helix interactions, while gaining one beta-hairpin and two gamma turns. As a result, this mutation alters the folding pattern,

exhibiting tremendous conformational changes in the tertiary structure by the absence of the protein's two-chain structure (chain P-chain A) into a malfunctioning protein with highly unstable PCSK9 (Shaik and Banaganapalli, 2019; Awan et al., 2021a).

The role of the PCSK9 GOF variants in accumulating the cholesterol in the plasma was first reported in 2003 (Abifadel et al., 2003). It is responsible for the degradation of LDLR on the cell membrane of the hepatocytes by binding to the extracellular domain of LDLR (Awan et al., 2021b). Therefore, LDLR is a transmembrane receptor that binds to circulating LDL-C to form the LDLR-LDL-C complex. This complex is then internalized through clathrin-coated pits into endosomes, where the LDLR releases the LDL-C for degradation in the lysosomes and returns to the cell membrane. The prevention of LDLR recycling and hepatic LDL-C clearance by hepatocytes, and chronic elevation of LDL-C in the plasma leads to plaque formation in arterial walls, accelerates atherosclerosis, and eventually leads to the development of premature CAD (McGowan et al., 2019; Seidah and Prat, 2021).

Molecular docking is the analysis of binding between biological and chemical structures using computational tools that focus on electrostatic potentials, physicochemical complementarity, and binding energy and presents them as 3D models with the best-predicted score. To compare the output of the binding of the mutant PCSK9 with LDLR, molecular docking was performed between the wild-type PCSK9 and LDLR as a reference. The lower energy was produced by the docking process between PCSK9 and LDLR, yielding ten candidate models (Figure 3). The cluster energy score increased from −1,340.4 kJ/mol in the wildtype to −1,214.6 kJ/mol in the mutant PCSK9-LDLR complex, suggesting that the wildtype complex docked more effectively than the mutant because low binding energies correlate with higher binding affinity (Shaik and Banaganapalli, 2019; Desta et al., 2020) (Figure 4).



As previously stated, PCSK9 mutations that result in loss play an important role in lowering LDL levels in the blood. As a result, human monoclonal anti-PCSK9 antibodies have proven effective in lowering LDL-C levels and atherosclerotic cardiovascular disease (ASCVD) risk, particularly in

individuals with severe phenotypes or resistance to lipid-lowering therapies (LLT). This was accomplished by binding to PCSK9 in plasma and preventing it from binding with LDLR. We performed molecular docking between both the wild type and the mutant PCSK9 and Alirocumab, a known

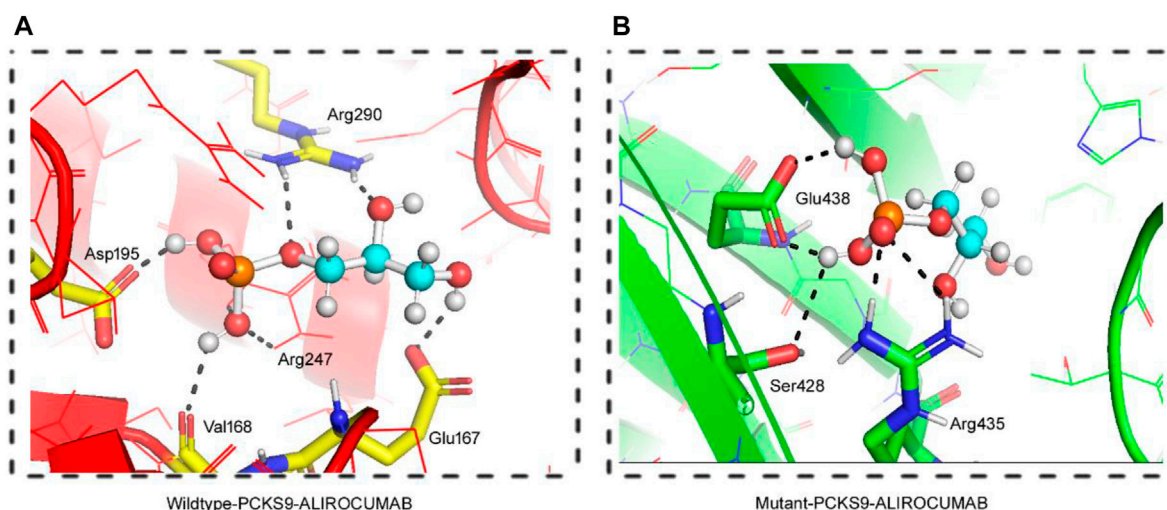


FIGURE 4

3D Visualization of the interaction between PCSK9 proteins and Alirocumab. Alirocumab with the wildtype PCSK9 protein (A), and with the mutant PCSK9 protein (B) provided by the DockThor-VS web server, where the ball-and-stick indicating interactions of the Alirocumab with different amino acid residues in the wildtype and mutant PCSK9 proteins.

PCSK9 inhibitor, to evaluate the change in the binding and the effectiveness of the PCSK9 inhibitor against the mutant protein. This scoring system is based on the sum of the van der Waals and electrostatic potentials of PCSK9 binding to Alirocumab (Dhankhar et al., 2020; Dhankhar et al., 2021; Dalal et al., 2022; Kumari et al., 2022). The increase in docking scores between mutant PCSK9 and Alirocumab showed a damaging effect of variant c.1486C>T on the binding potential, which decreases the Van der Waals energy, indicating molecules moving apart. This study found that the c.1486C>T variant of PCSK9 reduced the efficacy of alirocumab.

This study has few limitations. We could not recruit more families with FH to assess the role of PCSK9 (R496W) as a hotspot mutation among Saudi Arabs. However, given the rare occurrence of PCSK9 variant-carrier FH families in any ethnic population this seems to be an unrealistic limitation. Moreover, we could not ascertain the biological impact of the PCSK9 (R496W) variant using *in-vitro* functional biology assays because of the refusal of the participants to provide fresh tissue samples. However, to complement this lacuna, we used computational analysis as a primary functional characterization tool before undertaking technically complicated, time-consuming and expensive *in vitro* and *in vivo* methods.

In conclusion, this is the first c.1486C>T GOF variant identified in FH patients from Saudi Arabia. This study emphasizes the significance of genetic testing in identifying rare or novel FH mutations in underrepresented populations, which has the potential to reduce the burden of cardiovascular disease (CVD) in risk-group countries. Furthermore, this research provides comprehensive computational and structural insights into the genotype-protein phenotype correlation of the PCSK9 pathogenic variant with a PCSK9 inhibitor monoclonal antibody.

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.

Ethics statement

The studies involving human participants were reviewed and approved by the ethics committee of King Abdulaziz University, Saudi Arabia. Informed consent was obtained from each tested individual prior to genetic testing. The patients/participants provided their written informed consent to participate in this study.

Author contributions

Conceptualization: NS, NA-S, and ZA; data curation: NS, NA-S, BB, and HA; Formal Analysis: NS, NA-S, BB, and RE; funding acquisition: NS and EA; methodology: NS, AA, NA-S, BB, and ZA; project administration: NS; software: NA and BB; supervision: NS, ZA, and JM; writing original draft and editing: NS, NA-S, MA, MK, OS, SA, MN, BB, EA, and ZA. All authors contributed to the article and approved the submitted version.

Funding

This publication is based upon work supported by the Khalifa University (KU) and King Abdulaziz University (KAU) Join Research Program Award No. KAUJRP-1M-2021.

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.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fphys.2023.1204018/full#supplementary-material>

References

- Abifadel, M., Rabès, J.-P., Devillers, M., Munnich, A., Erlich, D., Junien, C., et al. (2009). Mutations and polymorphisms in the proprotein convertase subtilisin kexin 9 (PCSK9) gene in cholesterol metabolism and disease. *Hum. Mutat.* 30, 520–529. doi:10.1002/humu.20882
- Abifadel, M., Varret, M., Rabès, J. P., Allard, D., Ouguerram, K., Devillers, M., et al. (2003). Mutations in PCSK9 cause autosomal dominant hypercholesterolemia. *Nat. Genet.* 34, 154–156. doi:10.1038/ng1161
- Al-Allaf, F. A., Athar, M., Abduljaleel, Z., Taher, M. M., Khan, W., Ba-Hammam, F. A., et al. (2017). 1. Founder mutation identified in the LDLR gene causing familial hypercholesterolemia associated with increased risk of coronary heart disease. *Gene* 565, 76–84. doi:10.1016/j.gene.2015.03.064
- Al-Allaf, F., Athar, M., Alashwal, A., Abduljaleel, Z., Mohiuddin, T., Bouazzaoui, A., et al. (2017). 1. Founder mutation identified in the LDLR gene causing familial hypercholesterolemia associated with increased risk of coronary heart disease. *J. Saudi Heart Assoc.* 29, 318. doi:10.1016/j.jsha.2017.06.013
- Al-Waili, K., Al-Rasadi, K., Zadjali, F., Al-Hashmi, K., Al-Mukhaini, S., Al-Kindi, M., et al. (2020). Clinical and genetic characteristics of familial hypercholesterolemia at sultan qaboos university hospital in Oman. *Oman Med. J.* 35, e141. doi:10.5001/omj.2020.59
- Al-Waili, K., Al-Zidi, W. A., Al-Abri, A. R., Al-Rasadi, K., Al-Sabti, H. A., Shah, K., et al. (2013). Mutation in the PCSK9 gene in Omani Arab subjects with autosomal dominant hypercholesterolemia and its effect on PCSK9 protein structure. *Oman Med. J.* 28, 48–52. doi:10.5001/omj.2013.11
- Alallaf, F., H Nazar, F. A., Alnefaie, M., Almaymuni, A., Rashidi, O. M., Alhabib, K., et al. (2017). The spectrum of familial hypercholesterolemia (FH) in Saudi Arabia: Prime time for patient FH registry. *open Cardiovasc. Med. J.* 11, 66–75. doi:10.2174/1874192401711010066
- Alhababi, D., and Zayed, H. (2018). Spectrum of mutations of familial hypercholesterolemia in the 22 Arab countries. *Atherosclerosis* 279, 62–72. doi:10.1016/j.atherosclerosis.2018.10.022
- Alhabib, K. F., Al-Rasadi, K., Almigbal, T. H., Batais, M. A., Al-Zakwani, I., Al-Allaf, F. A., et al. (2021). Familial hypercholesterolemia in the arabian gulf region: Clinical results of the gulf FH registry. *PLoS One* 16, e0251560. doi:10.1371/journal.pone.0251560
- Alves-Bezerra, M., and Cohen, D. E. (2017). Triglyceride metabolism in the liver. *Compr. Physiol.* 8, 1–8. doi:10.1002/cphy.c170012
- Awan, Z. A., Bahattab, R., Kutbi, H. I., Noor, A. O. J., Al-Nasser, M. S., Shaik, N. A., et al. (2021a). Structural and molecular interaction studies on familial hypercholesterolemia causative PCSK9 functional domain mutations reveals binding affinity alterations with LDLR. *Int. J. Peptide Res. Ther.* 27, 719–733. doi:10.1007/s10989-020-10121-8
- Awan, Z. A., Bondagji, N. S., and Bamimore, M. A. (2019a). Recently reported familial hypercholesterolemia-related mutations from cases in the Middle East and North Africa region. *Curr. Opin. Lipidol.* 30, 88–93. doi:10.1097/MOL.0000000000000586
- Awan, Z. A., Rashidi, O. M., Al-Shehri, B. A., Jamil, K., Elango, R., Al-Aama, J. Y., et al. (2021b). Saudi familial hypercholesterolemia patients with rare LDLR stop gain variant showed variable clinical phenotype and resistance to multiple drug regimen. *Front. Med. (Lausanne)* 8, 694668. doi:10.3389/fmed.2021.694668
- Awan, Z., Batran, A., Al-Allaf, F. A., Alharbi, R. S., Hejazy, G. A., Jamalail, B., et al. (2022). Identification and functional characterization of 2 Rare LDLR stop gain variants (p.C231* and p.R744*) in Saudi familial hypercholesterolemia patients. *Panminerva Med.* 2022. doi:10.23736/S0031-0808.22.04612-2
- Bamimore, M. A., Zaid, A., Banerjee, Y., Al-Sarraf, A., Abifadel, M., Seidah, N. G., et al. (2015). Familial hypercholesterolemia mutations in the Middle eastern and north african region: A need for a national registry. *J. Clin. Lipidol.* 9, 187–194. doi:10.1016/j.jacl.2014.11.008
- Banaganapalli, B., Rashidi, O., Saadah, O. I., Wang, J., Khan, I. A., Al-Aama, J. Y., et al. (2017). Comprehensive computational analysis of GWAS loci identifies CCR2 as a candidate gene for celiac disease pathogenesis. *J. Cell. Biochem.* 118, 2193–2207. doi:10.1002/jcb.25864
- Bayona, A., Arrieta, F., Rodríguez-Jiménez, C., Cerrato, F., Rodríguez-Nóvoa, S., Fernández-Lucas, M., et al. (2020). Loss-of-function mutation of PCSK9 as a protective factor in the clinical expression of familial hypercholesterolemia: A case report. *Med. Baltim.* 99, e21754. doi:10.1097/MD.00000000000021754
- Beheshti, S. O., Madsen, C. M., Varbo, A., and Nordestgaard, B. G. (2020). Worldwide prevalence of familial hypercholesterolemia: Meta-analyses of 11 million subjects. *J. Am. Coll. Cardiol.* 75, 2553–2566. doi:10.1016/j.jacc.2020.03.057
- Bener, A., and Mohammad, R. R. (2017). Global distribution of consanguinity and their impact on complex diseases: Genetic disorders from an endogamous population. *Egypt. J. Med. Hum. Genet.* 18, 315–320. doi:10.1016/j.ejmhg.2017.01.002
- Blanco-Vaca, F., Martín-Campos, J. M., Pérez, A., and Fuentes-Prior, P. (2018). A rare STAP1 mutation incompletely associated with familial hypercholesterolemia. *Clin. Chim. Acta* 487, 270–274. doi:10.1016/j.cca.2018.10.014
- Chan, D. C., Pang, J., Hooper, A. J., Bell, D. A., Bates, T. R., Burnett, J. R., et al. (2018). A comparative analysis of phenotypic predictors of mutations in familial hypercholesterolemia. *J. Clin. Endocrinol. Metabolism* 103, 1704–1714. doi:10.1210/jc.2017-02622
- Chan, M. L.-Y., Cheung, C.-L., Lee, A. C.-H., Yeung, C.-Y., Siu, C.-W., Leung, J. Y.-Y., et al. (2019). Genetic variations in familial hypercholesterolemia and cascade screening in East Asians. *Mol. Genet. genomic Med.* 7, e00520. doi:10.1002/mgg3.520
- Dalal, V., Golemi-Kotra, D., and Kumar, P. (2022). Quantum mechanics/molecular mechanics studies on the catalytic mechanism of a novel esterase (FmtA) of *Staphylococcus aureus*. *J. Chem. Inf. Model* 62, 2409–2420. doi:10.1021/acs.jcim.2c00057
- De Grooth, G. J., Smilde, T. J., Van Wissen, S., Klerck, A. H., Zwinderman, A. H., Fruchart, J. C., et al. (2004). The relationship between cholesteryl ester transfer protein levels and risk factor profile in patients with familial hypercholesterolemia. *Atherosclerosis* 173, 261–267. doi:10.1016/j.atherosclerosis.2003.11.020
- Desta, I. T., Porter, K. A., Xia, B., Kozakov, D., and Vajda, S. (2020). Performance and its limits in rigid body protein-protein docking. *Structure* 28, 1071–1081. doi:10.1016/j.str.2020.06.006
- Dhankhar, P., Dalal, V., Mahto, J. K., Gurjar, B. R., Tomar, S., Sharma, A. K., et al. (2020). Characterization of dye-decolorizing peroxidase from *Bacillus subtilis*. *Arch. Biochem. Biophys.* 693, 108590. doi:10.1016/j.abb.2020.108590
- Dhankhar, P., Dalal, V., Singh, V., Sharma, A. K., and Kumar, P. (2021). Structure of dye-decolorizing peroxidase from *Bacillus subtilis* in complex with veratryl alcohol. *Int. J. Biol. Macromol.* 193, 601–608. doi:10.1016/j.ijbiomac.2021.10.100
- Eroglu, Z., Tetik Vardarlı, A., Düzgün, Z., Gündüz, C., Bozok Çetintaş, V., and Kayıkçıoğlu, M. (2018). Case-control study on PCSK9 R496W (rs374603772) and D374Y (rs137852912) mutations in Turkish patients with primary dyslipidemia. *Anatol. J. Cardiol.* 19, 334–340. doi:10.14744/AnatolJCardiol.2018.86648
- Gaboon, N. E. A., Banaganapalli, B., Nasser, K., Razeeth, M., Alsaedi, M. S., Rashidi, O. M., et al. (2020). Exome sequencing and metabolomic analysis of a chronic kidney disease and hearing loss patient family revealed RMND1 mutation induced sphingolipid metabolism defects. *Saudi J. Biol. Sci.* 27, 324–334. doi:10.1016/j.sjbs.2019.10.001
- Grell, L., Parkin, C., Slate, L., and Craig, P. A. (2006). EZ-Viz, a tool for simplifying molecular viewing in PyMOL. *Biochem. Mol. Biol. Educ.* 34, 402–407. doi:10.1002/bmb.2006.494034062672
- Guedes, I. A., Barreto, A. M. S., Marinho, D., Krempser, E., Kuenemann, M. A., Sperandio, O., et al. (2021). New machine learning and physics-based scoring functions for drug discovery. *Sci. Rep.* 11, 3198. doi:10.1038/s41598-021-82410-1

- Guo, Q., Feng, X., and Zhou, Y. (2020a). PCSK9 variants in familial hypercholesterolemia: A comprehensive synopsis. *Front. Genet.* 11, 1020. doi:10.3389/fgene.2020.01020
- Guo, Q., Feng, X., and Zhou, Y. (2020b). PCSK9 variants in familial hypercholesterolemia: A comprehensive synopsis. *Front. Genet.* 11, 1020. doi:10.3389/fgene.2020.01020
- Harling-Lee, J. D., Gorzynski, J., Yebra, G., Angus, T., Fitzgerald, J. R., and Freeman, T. C. (2022). A graph-based approach for the visualisation and analysis of bacterial pangenomes. *BMC Bioinforma.* 23, 416. doi:10.1186/s12859-022-04898-2
- Harrison, S. M., Biesecker, L. G., and Rehm, H. L. (2019). Overview of specifications to the ACMG/AMP variant interpretation guidelines. *Curr. Protoc. Hum. Genet.* 103, e93. doi:10.1002/cphg.93
- Hooper, A. J., Burnett, J. R., Bell, D. A., and Watts, G. F. (2018). The present and the future of genetic testing in familial hypercholesterolemia: Opportunities and caveats. *Curr. Atheroscler. Rep.* 20, 31. doi:10.1007/s11883-018-0731-0
- Ibarretxe, D., Rodríguez-Borjabad, C., Feliu, A., Bilbao, J. Á., Masana, L., and Plana, N. (2018). Detecting familial hypercholesterolemia earlier in life by actively searching for affected children: The DECOPIN project. *Atherosclerosis* 278, 210–216. doi:10.1016/j.atherosclerosis.2018.09.039
- Ibrahim, S., Defesche, J. C., and Kastelein, J. J. P. (2021). Beyond the usual suspects: Expanding on mutations and detection for familial hypercholesterolemia. *Expert Rev. Mol. Diagnostics* 21, 887–895. doi:10.1080/14737159.2021.1953985
- Kamar, A., Khalil, A., and Nemer, G. (2021). The digenic causality in familial hypercholesterolemia: Revising the genotype–phenotype correlations of the disease. *Front. Genet.* 11, 572045. doi:10.3389/fgene.2020.572045
- Kozakov, D., Brenke, R., Comeau, S. R., and Vajda, S. (2006). Piper: An FFT-based protein docking program with pairwise potentials. *Proteins* 65, 392–406. doi:10.1002/prot.21117
- Kozakov, D., Hall, D. R., Xia, B., Porter, K. A., Padhorny, D., Yueh, C., et al. (2017). The ClusPro web server for protein–protein docking. *Nat. Protoc.* 12, 255–278. doi:10.1038/nprot.2016.169
- Kumari, R., Kumar, V., Dhankhar, P., and Dalal, V. (2022). Promising antivirals for PLpro of SARS-CoV-2 using virtual screening, molecular docking, dynamics, and MMPBSA. *J. Biomol. Struct. Dyn.* 41, 4650–4666. doi:10.1080/07391102.2022.2071340
- Laimer, J., Hofer, H., Fritz, M., Wegenkittl, S., and Lackner, P. (2015). MAESTRO—multi agent stability prediction upon point mutations. *BMC Bioinforma.* 16, 116. doi:10.1186/s12859-015-0548-6
- Laskowski, R. A., Rullmann, J. A., Macarthur, M. W., Kaptein, R., and Thornton, J. M. (1996). AQUA and PROCHECK-NMR: Programs for checking the quality of protein structures solved by NMR. *J. Biomol. NMR* 8, 477–486. doi:10.1007/BF00228148
- Mcgowan, M. P., Moriarty, P. M., and Barton Duell, P. (2019). Diagnosis and treatment of heterozygous familial hypercholesterolemia. *J. Am. Heart Assoc.* 8. doi:10.1161/JAHA.119.013225
- Meshkov, A., Ershova, A., Kiseleva, A., Zotova, E., Sotnikova, E., Petukhova, A., et al. (2021). The LDLR, APOB, and PCSK9 variants of index patients with familial hypercholesterolemia in Russia. *Genes (Basel)* 12, 66. doi:10.3390/genes12010066
- Mlinaric, M., Bratanic, N., Dragos, V., Skarlovnik, A., Cevc, M., Battelino, T., et al. (2020). Case report: Liver transplantation in homozygous familial hypercholesterolemia (HoFH)—Long-Term follow-up of a patient and literature review. *Front. Pediatr.* 8, 567895. doi:10.3389/fped.2020.567895
- Ohta, N., Hori, M., Takahashi, A., Ogura, M., Makino, H., Tamanaha, T., et al. (2016). Proprotein convertase subtilisin/kexin 9 V4I variant with LDLR mutations modifies the phenotype of familial hypercholesterolemia. *J. Clin. Lipidol.* 10, 547–555. doi:10.1016/j.jacl.2015.12.024
- Paththinige, C. S., Sirisena, N. D., and Dissanayake, V. (2017). Genetic determinants of inherited susceptibility to hypercholesterolemia - a comprehensive literature review. *Lipids health Dis.* 16, 103. doi:10.1186/s12944-017-0488-4
- Pisciotta, L., Oliva, C. P., Cefalù, A. B., Noto, D., Bellocchio, A., Fresa, R., et al. (2006). Additive effect of mutations in LDLR and PCSK9 genes on the phenotype of familial hypercholesterolemia. *Atherosclerosis* 186, 433–440. doi:10.1016/j.atherosclerosis.2005.08.015
- Reyes-Soffer, G., Pavlyha, M., Ngai, C., Thomas, T., Holleran, S., Ramakrishnan, R., et al. (2017). Effects of PCSK9 inhibition with alirocumab on lipoprotein metabolism in healthy humans. *Circulation* 135, 352–362. doi:10.1161/CIRCULATIONAHA.116.025253
- Saberi-Movahed, F., Rostami, M., Berahmand, K., Karami, S., Tiwari, P., Oussalah, M., et al. (2022). Dual regularized unsupervised feature selection based on matrix factorization and minimum redundancy with application in gene selection. *Knowledge-Based Syst.* 256, 109884. doi:10.1016/j.knsys.2022.109884
- Santos, K. B., Guedes, I. A., Karl, A. L. M., and Dardenne, L. E. (2020). Highly flexible ligand docking: Benchmarking of the DockThor program on the LEADS-PEP protein–peptide data set. *J. Chem. Inf. Model.* 60, 667–683. doi:10.1021/acs.jcim.9b00905
- Seidah, N. G., and Prat, A. (2021). The multifaceted biology of PCSK9. *Endocr. Rev.* 43, 558–582. doi:10.1210/edrev/bnab035
- Shaik, N. A., and Banaganapalli, B. (2019). Computational molecular phenotypic analysis of PTPN22 (W620R), IL6R (D358A), and TYK2 (P1104A) gene mutations of rheumatoid arthritis. *Front. Genet.* 10, 168. doi:10.3389/fgene.2019.00168
- Slimani, A., Hrira, M. Y., Najah, M., Jomaa, W., Maatouk, F., Hamda, K. B., et al. (2015). PCSK9 polymorphism in a Tunisian cohort: Identification of a new allele, L8, and association of allele L10 with reduced coronary heart disease risk. *Mol. Cell. Probes* 29, 1–6. doi:10.1016/j.mcp.2014.09.001
- Sturm, A. C., Knowles, J. W., Gidding, S. S., Ahmad, Z. S., Ahmed, C. D., Ballantyne, C. M., et al. (2018). Clinical genetic testing for familial hypercholesterolemia: JACC scientific expert panel. *J. Am. Coll. Cardiol.* 72, 662–680. doi:10.1016/j.jacc.2018.05.044
- Vajda, S., Yueh, C., Beglov, D., Bohnuud, T., Mottarella, S. E., Xia, B., et al. (2017). New additions to the ClusPro server motivated by CAPRI. *Proteins* 85, 435–444. doi:10.1002/prot.25219

Frontiers in Molecular Biosciences

Explores biological processes in living organisms
on a molecular scale

Focuses on the molecular mechanisms
underpinning and regulating biological processes
in organisms across all branches of life.

Discover the latest Research Topics

[See more →](#)

Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

Contact us

+41 (0)21 510 17 00
frontiersin.org/about/contact



Frontiers in Molecular Biosciences

