

Computational drug discovery of medicinal compounds for cancer management

Edited by

Khurshid Ahmad, Sibhghatulla Shaikh, Mohammad Ehtisham Khan
and Faez Iqbal Khan

Published in

Frontiers in Chemistry



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 978-2-8325-4180-7
DOI 10.3389/978-2-8325-4180-7

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

Computational drug discovery of medicinal compounds for cancer management

Topic editors

Khurshid Ahmad — Yeungnam University, Republic of Korea
Sibhghatulla Shaikh — Yeungnam University, Republic of Korea
Mohammad Ehtisham Khan — Jazan University, Saudi Arabia
Faez Iqbal Khan — Xi'an Jiaotong-Liverpool University, China

Citation

Ahmad, K., Shaikh, S., Khan, M. E., Khan, F. I., eds. (2023). *Computational drug discovery of medicinal compounds for cancer management*.
Lausanne: Frontiers Media SA. doi: 10.3389/978-2-8325-4180-7

Table of contents

- 05 **Editorial: Computational drug discovery of medicinal compounds for cancer management**
Sibhghatulla Shaikh, Khurshid Ahmad, Mohammad Ehtisham Khan and Faez Iqbal Khan
- 08 **Carrier-free supramolecular nanoassemblies of pure LSD1 inhibitor for effective anti-tumor therapy**
Boao Li, Xiangyu Zhang and Jibin Li
- 17 **Identification of compelling inhibitors of human norovirus 3CL protease to combat gastroenteritis: A structure-based virtual screening and molecular dynamics study**
Shan He, Alaa F. Nahhas, Alaa Hamed Habib, Mohammed Ali Alshehri, Saleh Alshamrani, Saeed A. Asiri, Mashael M. Alnamshan, Nawal Helmi, Ibtesam Al-Dhuayan, Jawaher Almulhim, Ahmed M. Alharbi, Dongxiao Su, Ankita Kumari and Abdul Rahaman
- 28 **Identification of novel natural drug candidates against BRAF mutated carcinoma; An integrative in-silico structure-based pharmacophore modeling and virtual screening process**
F. A. Dain Md Opo, Ahad Amer Alsaiaari, Mohammad Habibur Rahman Molla, Md Afsar Ahmed Sumon, Khaled A. Yaghmour, Faysal Ahammad, Farhan Mohammad and Jesus Simal-Gandara
- 49 **Combined 3D-QSAR, molecular docking and dynamics simulations studies to model and design TTK inhibitors**
Noureen Ashraf, Asnuzilawati Asari, Numan Yousaf, Matloob Ahmad, Mahmood Ahmed, Amir Faisal, Muhammad Saleem and Muhammad Muddassar
- 68 ***In-silico* molecular interactions among the secondary metabolites of *Caulerpa* spp. and colorectal cancer targets**
Nazli Mert-Ozupek, Gizem Calibasi-Kocal, Nur Olgun, Yasemin Basbinar, Levent Cavas and Hulya Ellidokuz
- 82 **Carvacrol instigates intrinsic and extrinsic apoptosis with abrogation of cell cycle progression in cervical cancer cells: Inhibition of Hedgehog/GLI signaling cascade**
Afza Ahmad, Rohit Kumar Tiwari, Mohd Saeed, Hadba Al-Amrah, Ihn Han, Eun-Ha Choi, Dharmendra K. Yadav and Irfan Ahmad Ansari
- 100 **Chemical, biological and *in silico* assessment of date (*P. dactylifera* L.) fruits grown in Ha'il region**
Abdulmohsen Khalaf Dhahi Alsukaibi, Khalaf M. Alenezi, Ashanul Haque, Irfan Ahmad, Mohd Saeed, Mahima Verma, Irfan Ahmad Ansari and Ming-Fa Hsieh
- 113 **Combining machine learning and structure-based approaches to develop oncogene PIM kinase inhibitors**
Haifa Almkadi, Gada Ali Jadkarim, Arif Mohammed, Majid Almansouri, Nasreen Sultana, Noor Ahmad Shaik and Babajan Banaganapalli

- 124 **Mechanism exploration and prognosis study of Astragali Radix-Spreading hedyotis herb for the treatment of lung adenocarcinoma based on bioinformatics approaches and molecular dynamics simulation**
Junfeng Guo, Yuting Zhao, Xuanyu Wu, Ganggang Li, Yuwei Zhang, Yang Song and Quanyu Du
- 141 **Increased inflammatory cytokines and oxidative stress enhanced antibody production in breast and prostate cancer patients with COVID-19 related depression**
Naif K. Binsaleh, Subuhi Sherwani, Reem Eltayeb, Husam Qanash, Abdulrahman S. Bazaid, Maryam Althobiti, Mohannad S. Hazzazi and Saravanan Rajendrasozhan
- 151 **Predicting the effects of rare genetic variants on oncogenic signaling pathways: A computational analysis of HRAS protein function**
Sadaqat Ali, Usman Ali, Adeem Qamar, Imran Zafar, Muhammad Yaqoob, Qurat ul Ain, Summya Rashid, Rohit Sharma, Hiba-Allah Nafidi, Yousef A. Bin Jordan and Mohammed Bourhia
- 172 ***Amomum subulatum*: A treasure trove of anti-cancer compounds targeting TP53 protein using *in vitro* and *in silico* techniques**
Sadaqat Ali, Asifa Noreen, Adeem Qamar, Imran Zafar, Quratul Ain, Hiba-Allah Nafidi, Yousef A. Bin Jordan, Mohammed Bourhia, Summya Rashid and Rohit Sharma
- 189 **Identifying promising GSK3 β inhibitors for cancer management: a computational pipeline combining virtual screening and molecular dynamics simulations**
Libo Hua, Farah Anjum, Alaa Shafie, Amal Adnan Ashour, Abdulraheem Ali Almalki, Ali Abdullah Alqarni, Hamsa Jameel Banjar, Sarah Abdullah Almaghrabi, Shan He and Nenggui Xu
- 199 **Unlocking the potential of approved drugs for the allosteric inhibition of tropomyosin-receptor kinase A using molecular docking and molecular dynamics studies**
Rua M. Mukhtar, Nihal Abdelmoniem, Hisham A. Elrufaie, Alaa Edris, Hiba Ghaboosh, Mohanad A. Mahgoub, Elrashied A. E. Garelnabi, Wadah Osman, Asmaa E. Sherif, Ahmed Ashour, Kholoud F. Ghazawi, Waad A. Samman, Aisha A. Alhaddad, Rawan Bafail, Sabrin R. M. Ibrahim, Gamal A. Mohamed and Abdulrahim A. Alzain
- 211 **The combination of multi-approach studies to explore the potential therapeutic mechanisms of imidazole derivatives as an MCF-7 inhibitor in therapeutic strategies**
Maryam Rashid, Ayesha Maqbool, Nusrat Shafiq, Yousef A. Bin Jordan, Shagufta Parveen, Mohammed Bourhia, Hiba-Allah Nafidi and Rashid Ahmed Khan



OPEN ACCESS

EDITED AND REVIEWED BY
Michael Kassiou,
The University of Sydney, Australia

*CORRESPONDENCE

Sibhghatulla Shaikh,
✉ sibhghat.88@gmail.com
Khurshid Ahmad,
✉ ahmadkhurshed2008@gmail.com
Mohammad Ehtisham Khan,
✉ ehtisham.nanotech@gmail.com
Faez Iqbal Khan,
✉ khanfaeziqbal@gmail.com

RECEIVED 23 November 2023

ACCEPTED 01 December 2023

PUBLISHED 07 December 2023

CITATION

Shaikh S, Ahmad K, Khan ME and Khan FI
(2023), Editorial: Computational drug
discovery of medicinal compounds for
cancer management.
Front. Chem. 11:1343183.
doi: 10.3389/fchem.2023.1343183

COPYRIGHT

© 2023 Shaikh, Ahmad, Khan and Khan.
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: Computational drug discovery of medicinal compounds for cancer management

Sibhghatulla Shaikh^{1*}, Khurshid Ahmad^{1*},
Mohammad Ehtisham Khan^{2*} and Faez Iqbal Khan^{3*}

¹Department of Medical Biotechnology, Yeungnam University, Gyeongsan, Republic of Korea,

²Department of Chemical Engineering Technology, College of Applied Industrial Technology, Jazan

University, Jazan, Saudi Arabia, ³Department of Biological Sciences, School of Science, Xi'an Jiaotong-Liverpool University, Suzhou, Jiangsu, China

KEYWORDS

computational methods, drug discovery, cancer, medicinal compounds, druglikeness

Editorial on the Research Topic

Computational drug discovery of medicinal compounds for cancer management

Cancer remains a major public health concern, with it ranking as the leading cause of death worldwide. Despite significant advances in biotechnology, developing practical and innovative small-molecule drugs remains a hard, time-consuming, and costly process. This endeavor necessitates the collaboration of experts from a variety of disciplines, including computational biology, drug metabolism, and clinical research. Hence, there is a pressing need for novel drug development processes that save time and cost while improving overall efficiency. Computer-aided drug design (CADD) methodologies are becoming increasingly crucial in drug discovery, particularly in their ability to identify promising drug candidates cost-effectively. In this area of study, we launched a Research Topic in Frontiers in Chemistry journal titled “Computational Drug Discovery of Medicinal Compounds for Cancer Management.” This Research Topic attracted the interest of researchers, and a large number of manuscripts were submitted. Among these submissions, 15 Original Research articles have been published covering a wide range of CADD topics and elucidating advanced *in silico* methodologies applicable to the field of drug discovery.

The following is a summary of the published articles:

Dain Md Opo et al. used structure-based pharmacophore modeling and virtual screening to identify potentially natural lead compounds that can inhibit BRAF and thus inhibiting cancer. Promising candidate compounds targeting the upregulated BRAF gene have been identified using *in silico* drug design methodologies and computational tools. The study identified four potential compounds by utilizing these computational methods. The investigation suggests that these compounds may be useful against a range of cancers by specifically targeting the overexpressed BRAF gene.

Ashraf et al. presented a comprehensive study on potential inhibitors of Tyrosine Threonine Kinase (TTK), a target in a variety of human cancers including breast,

colorectal, and thyroid carcinomas. They used computational techniques like 3D-QSAR modeling and structure-based alignment. The study identified key factors influencing compound activity (electrostatic, steric, HBA, HBD, and hydrophobic fields). New compounds with predicted binding modes and structural stability were designed using molecular docking and molecular dynamics (MD) simulations, displaying promising TTK protein binding affinity as potential TTK inhibitor candidates.

Li et al. investigated LSD1, a protein methylation oxidase linked to gene expression and tumor initiation. Compound 17i showed promise as an LSD1 inhibitor and anti-tumor agent, but solubility issues hampered clinical validation. A carrier-free 17i nano assembly made with DSPE-PEG2000 improved therapeutic efficacy. In CT-26 mice with colorectal tumors, these nano assemblies matched 17i's cytotoxicity *in vitro* but had better therapeutic efficacy, anti-tumor immune response, and lower systemic toxicity. A novel drug nanoassembly method may improve the effects of poorly soluble anti-tumor compounds.

He et al. computationally screened a library of 700 antiviral compounds against the 3CL protease (3CLP) of human noroviruses. The researchers discovered three compounds that had higher binding energy with 3CLP than the control (Dipeptidyl inhibitor 7). These hits' estimated physicochemical and ADME properties were in the favorable range. According to the findings, these compounds could be used as 3CLP inhibitors in the treatment of gastroenteritis.

Mert-Ozupek et al. conducted an insilico screening of bioactive compounds from *Caulerpa* spp. against the colorectal cancer enzymes glucose-6-phosphate dehydrogenase and 6-phosphogluconate dehydrogenase. Caulerpin, monomethyl caulerpinate, and caulersin bind to these enzymes strongly. They proposed that the identified compounds be tested for their potential efficacy against enzymes in the pentose phosphate pathway.

Ahmad et al. investigated the anticancer and apoptotic effects of carvacrol (CAR) on C33A cervical cancer cells. The antiproliferative and apoptotic effects of CAR were mostly observed in C33A cervical cancer cells *in vitro*. ROS production in C33A mitochondria triggered a chain of events that resulted in mitochondrial apoptosis. CAR also influences extrinsic or death receptor pathway signaling. CAR inhibited hedgehog signaling, causing apoptosis and inhibiting cell proliferation in cervical carcinoma cells. These findings imply that CAR could be used to treat cervical cancer.

Guo et al. explored 422 targets and 29 active ingredients from *Astragali Radix* (AR) and *Spreading Hedyotis* Herb. They demonstrated how AR-SH reduces lung adenocarcinoma (LUAD) symptoms by targeting EGFR, MAPK1, and KRAS. Molecular docking and dynamics simulations showed that AR-SH's main active components bind to the right proteins, especially EGFR, suggesting it is more effective than Gefitinib. These findings show that AR-SH can improve LUAD treatment and prognosis.

Almukadi et al. identified novel and efficacious therapeutics for PIM-1 kinase by employing structure-based and machine-learning approaches. Four potential molecules were discovered to modulate PIM-1. Additionally, the MD simulation study

revealed that these compounds interacted with the PIM kinase stably.

Alsukaibi et al. investigated the phytochemical and biological properties of two date fruit cultivars from Saudi Arabia's Ha'il region, Shishi M1 and Majdool M2. *In vitro*, both cultivars inhibited HCT-116 colon cancer cells. Procyanidin B2 and luteolin-7-O-rutinoside were identified as active constituents by computational analysis.

Ali et al. used insilico techniques to predict how rare genetic variations would affect HRAS protein function. Fifty nonsynonymous single nucleotide polymorphisms (nsSNPs) were discovered, 23 of which were in the HRAS gene exon, implying that they are potentially harmful. Ten of the twenty-three tested substances were the most dangerous. This study lends credence to the notion that nsSNPs may increase HRAS expression and activate carcinogenic signaling pathways in cancer.

Ali et al. utilized phytocompounds targeting TP53 from *Amomum subulatum* seed extract, focusing on major alkaloids and saponins. The antioxidant activity was confirmed by DPPH analysis, particularly in methanol, BHT, and n-hexane extracts. Additionally, the computational analysis revealed that top phytocompounds had strong binding affinities to TP53, implying potential anti-cancer actions. This study presents novel cancer treatment drug discovery insights using *A. subulatum* seed compounds.

Binsaleh et al. investigated the effect of depression on breast cancer and prostate cancer patients during the COVID-19 pandemic. Cancer patients, especially those suffering from depression, had higher levels of proinflammatory cytokines and oxidative stress markers than healthy people. Elevated levels of specific serum antibodies in cancer patients suggested increased oxidative stress. The findings highlight the importance of addressing mental health concerns in cancer patient care and disease management.

Rashid et al. investigated the potential of imidazole derivatives as MCF-7 inhibitors for the treatment of breast cancer. Using Flare's machine learning, they developed a 3D-QSAR and activity atlas model that classified compound datasets as active or inactive in comparison to a reference drug. Molecular docking analysis discovered active compound interactions with TTK, HER2, GR, NUDT5, MTHFS, and NQO2. The most promising cancer inhibitor was identified as compound C10, paving the way for new approaches to breast cancer treatment.

Hua et al. computationally screened 4,222 anti-cancer compounds against GSK3 β . They observed that two potent compounds, BMS-754807 and GSK429286A, had high affinities for binding to GSK3 β . These compounds also have promising drug-like properties. This study proposed that BMS-754807 and GSK429286A undergo experimental validation to assess their potential as anti-cancer agents.

Mukhtar et al. screened 9,923 compounds from the ChEMBL database against Tropomyosin-receptor kinase A (TrkA). Among the screened compounds, Delanzomib and Tibalosin, the two leading compounds demonstrated great potential. At 200 ns MD simulations, these compounds demonstrated stable interactions with the TrkA protein. This study proposed that additional

research be conducted to determine the viability of Delanzomib and Tibalotin as TrkA inhibitors.

Finally, the authors and editors of this Research Topic hope that the Research Topic of articles will highlight the advances made in the use of computational methodologies for facilitating the design of pharmaceutical compounds directed at various protein targets implicated in cancer management. Furthermore, it is expected that these articles will contribute to a deeper and more comprehensive understanding of the intricate biological processes that underpin cancer, potentially leading to novel therapeutic interventions for this disease. We hope that these articles will serve as a source of motivation, information, and guidance for researchers and scholars working in this field.

Author contributions

SS: Writing–original draft, Writing–review and editing. KA: Writing–original draft, Writing–review and editing. MK: Writing–original draft, Writing–review and editing. FK: Writing–original draft, Writing–review and editing.

Funding

The authors declare financial support was received for the research, authorship, and/or publication of this article. The author, FK, would like to thank the XJTLU Research Development Fund (RDF-22-02-090) for its 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.



OPEN ACCESS

EDITED BY
Khurshid Ahmad,
Yeungnam University, South Korea

REVIEWED BY
Piotr G Rychahou,
University of Kentucky, United States
Qiwen Wang,
Zhejiang University, China
Barnali Maiti,
VIT University, India
Praveen Nekkar Rao,
University of Waterloo, Canada

*CORRESPONDENCE
Xiangyu Zhang,
xiangyuzhangsyphu@163.com
Jibin Li,
leejibin@126.com

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

RECEIVED 06 August 2022
ACCEPTED 14 September 2022
PUBLISHED 30 September 2022

CITATION
Li B, Zhang X and Li J (2022), Carrier-
free supramolecular nanoassemblies of
pure LSD1 inhibitor for effective anti-
tumor therapy.
Front. Chem. 10:1012882.
doi: 10.3389/fchem.2022.1012882

COPYRIGHT
© 2022 Li, Zhang and Li. 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.

Carrier-free supramolecular nanoassemblies of pure LSD1 inhibitor for effective anti-tumor therapy

Boao Li¹, Xiangyu Zhang^{2*} and Jibin Li^{1*}

¹Department of Colorectal Surgery, Liaoning Cancer Hospital, Shenyang, China, ²State Key Laboratory of Natural and Biomimetic Drugs, School of Pharmaceutical Sciences, Peking University, Beijing, China

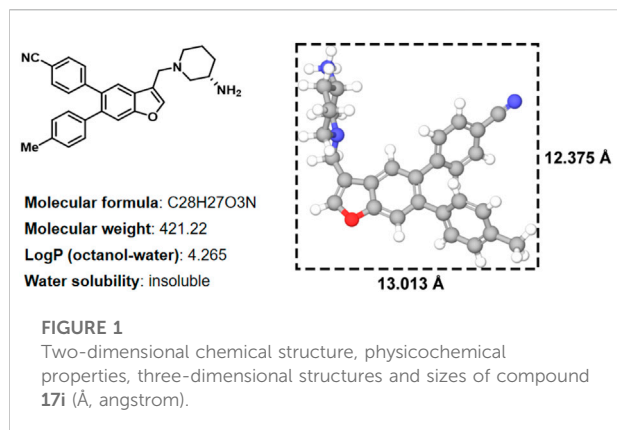
The LSD1 protein is an oxidase that regulates protein methylation, which regulates gene expression and triggers tumors. Previously, inhibiting LSD1 has been found to be an effective treatment strategy for opposing tumors caused by overexpression of LSD1. Our recent study found that compound **17i** was a suitable LSD1 inhibitor with potential anti-tumor activity. However, its extremely insoluble nature limits further validation of its anti-tumor activity at the clinical level. In this study, a unique carrier-free supramolecular nanoassemblies of pure compound **17i** is expected to enhance therapeutic efficacy. Aqueous-insoluble compound **17i** was mixed with a small quantity of DSPE-PEG₂₀₀₀ into an organic solvent and was prepared as nanoassemblies in water via the one-step nanoprecipitation method. The **17i** nanoassemblies have a similar effect on its cytotoxicity when compared with **17i** solution *in vitro*. Importantly, the PEGylated **17i** nanoassemblies exhibit significant superiorities over **17i** solutions in therapeutic efficiency, anti-tumor immune response and systemic toxicity in BALB/c mice bearing CT-26 colorectal tumors. We envision that the fabrication of pure drug nanoassemblies offers an efficient platform for reforming the undesirable characteristics of drug-like compounds to potentiate the anti-tumor therapeutic effect.

KEYWORDS

drug delivery, supramolecular nanoassemblies, therapeutic efficiency, systemic toxicity, drug-like compound

Introduction

In histone lysine specific demethylase 1 (LSD1), the methyl groups are removed from the lysine residues (H3K4 and H3K9) by an oxidative enzyme (Shi et al., 2004; Suzuki and Miyata, 2011; Schmitt et al., 2013; Sorna et al., 2013; Tortorici et al., 2013; Zheng et al., 2013). As an epigenetic regulator, gene expression and cancer initiation are influenced by LSD1 (Ma et al., 2015; Zheng et al., 2015; Vianello et al., 2016). Therefore, inhibiting LSD1 is an effective strategy for anti-tumor treatment (Wu et al., 2016; Mould et al., 2017). In our previous study, various LSD1 inhibitors were reported (Wang X. et al., 2020; Zhang et al., 2021), in which the compound **17i**



(Figure 1) ($IC_{50} = 0.065 \mu M$) exhibited a significant effect on target LSD1. The solubility of compound 17i in organic solvents was good. However, it has been tested that 17i is almost insoluble in water. This impeded further validation of its anti-tumor activity at the clinical level.

The wide application of nanotechnology in the medical field has significantly enriched the delivery strategies of anti-tumor drugs (Sun M. et al., 2019; Sun et al., 2022). Especially, the construction of a novel nanodrug delivery system can significantly improve the druggability of chemotherapeutics, prolong the systemic circulation time and enhance tumor-specific accumulation by increasing permeability and retention (EPR) effect, thus augmenting the anti-tumor effect and reduce the side effects (Sun B. et al., 2019; Wang Q. et al., 2020; Yang et al., 2020; Zhang et al., 2020; Zhao et al., 2021). Herein, to solve the dilemma of clinical transformation of compound 17i, we put forward a fascinating nanoassembly based on compound 17i for effective anti-tumor therapy. Firstly, we confirm tight interactions between LSD1 and compound 17i through molecular docking and molecular dynamic simulation, elaborating that compound 17i had effective inhibitory activity for LSD1. The supramolecular nanoassemblies of 17i were fabricated by a one-step nanoprecipitation approach. A small amount of 1,2-distearoyl-sn-glycero-3-phosphoethanolamine-N-[methoxy (polyethyleneglycol)-2000] (DSPE-PEG₂₀₀₀) (Supplementary Figure S3) was attached on the surface to improve surface hydrophilicity of nanoassemblies and extend the period of blood circulation. The PEGylated 17i nanoassemblies exhibited comparable cytotoxicity when compared with 17i solution *in vitro*, but showed particular superiorities in terms of efficient anti-tumor therapy and anti-tumor immune response and less side effect in BALB/c mice bearing CT-26 tumors. As far as we are aware, this is the first time pure LSD1 inhibitor has been fabricated into nanoassemblies without the addition of carrier excipients. Such a potent nanoplatform holds promising clinical application prospects for drug-like compounds.

Materials and methods

Materials

Compound 17i was self-prepared. 2-(4-Amidinophenyl)-6-indolecarbamide dihydrochloride (DAPI) and 3-(4,5-Dimethyl-2-thiazolyl)-2,5-diphenyl-2H-tetrazolium bromide (MTT) were obtained from Dalian Meilun Biotech Co., Ltd. (Dalian, China). Cell culture media RPMI 1640, penicilline-streptomycin, and fetal bovine serum were available from GIBCO, (Carlsbad, United States). 1,2-distearoyl-sn-glycero-3-phosphoethanolamine-N-[methoxy (polyethyleneglycol)-2000] (DSPE-PEG₂₀₀₀) was purchased from Shanghai Advanced Vehicle Technology Co., Ltd. All other chemical components and solvents applied in this study are of analytical grade.

Molecular docking

To predict the binding mode of the target molecule with the binding site, we performed molecular docking using Glide 9.7 module (Friesner et al., 2004; Friesner et al., 2006), which the protein structure PDB 5YJB (residues 172–833) used for docking. The inhibitor 17i was docked into the binding pocket of LSD1 using the standard precision module to get initial binding predictions and docking scores.

MD simulation

The 100 ns MD simulations were carried out of LSD1-17i complex by using Desmond v3.8 (Wang et al., 2019). This system is dissolved in a cubic box (8 Å × 30 Å × 8 Å) with a simple point charge (SPC) water, adding an appropriate amount of Na⁺ counter ions to achieve neutralization. Based on the OPLS-2005 force field, the energy of the system was minimized. Finally, the 100 ns MD simulations were performed in NPT ensemble. Other parameters are default.

Preparation of supramolecular 17i nanoassemblies

Nanoassemblies based on pure 17i compound were prepared by the one-step nanoprecipitation method. 8 mg 17i was dissolved in 1 ml methanol to acquire 17i methanol solution. Then, 500 μL mixtures were dropped slowly into the 2 ml aqueous solution under the stirring for 5 min. Following that, methanol was removed from the colloidal solution in a vacuum at 37°C. The PEGylated 17i nanoassemblies were prepared in above protocol using a mixed methanol solution of 17i and DSPE-PEG₂₀₀₀ (15%, w/w). The prepared nanoassemblies would be stored at 4°C. In addition, the 17i reagents were dissolved in

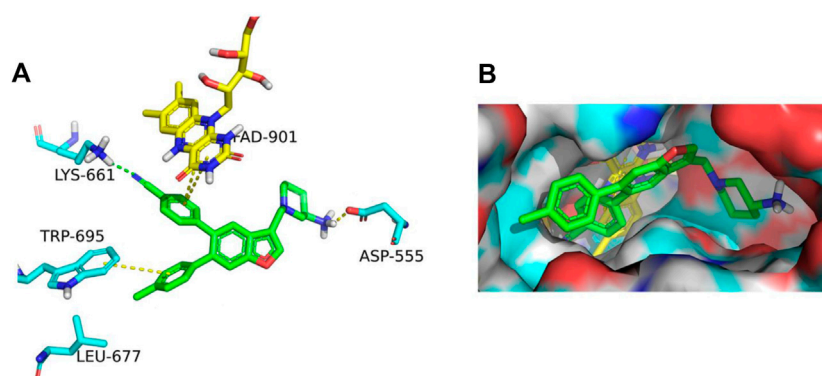


FIGURE 2
(A) In LSD1, **17i** (green) binds to its active site. (B) **17i** represents the binding surface of the LSD1 pocket (PDB: 5YJB).

0.5 ml acetonitrile. Then, the mother liquor was diluted with PBS (1:9), prepared as the **17i** solution.

Characterization of supramolecular **17i** nanoassemblies

The dynamic light scattering particle size of **17i** nanoassemblies was measured through a Zetasizer (Malvern Co., UK). The prepared nanoassemblies were diluted with phosphate buffer solution (PBS) and the particle size was measured three times. The **17i** nanoassemblies were diluted 1/20 with deionized water and dropped onto a copper mesh (300 mesh) covered by carbon film. After natural drying, negative staining was performed with 2% phosphotungstate acid. Transmission electron microscopy (TEM) (Hitachi, HT7700, Japan) was used to observe the appearance and morphology of **17i** nanoassemblies.

Binding conformation and binding energy of two molecules of **17i** together was calculated with molecular docking. Docking was performed using Glide module in Schrödinger. One molecule of **17i** was selected as the receptor and receptor grid was set to cover the whole receptor molecule, and then another molecule of **17i** was docked onto **17i** receptor, and binding energy were calculated.

Molecular dynamic (MD) simulation was performed with Materials Studio Program (Accelrys Inc.). First, amorphous cell module was used to construct molecular aggregation models in which 16 **17i** molecules and 11170 water molecules were put into cubic box with side length equals 7ns. Then 50000 steps energy minimization were performed followed with 50 ns molecular dynamic with compass force field in temperature of 298K, pressure of 1.01325 bar and NPT ensemble. Root mean square derivation (RMSD) value were calculated using forcite module.

Simulation study of assembly

Computational simulations of intermolecular interactions between **17i** molecules were performed. The two-dimensional (2D) structure of **17i** was established by Marvin sketch software, and the three-dimensional (3D) structure of **17i** was optimized by Sybyl 6.9.1 software package. In the previous work, we have introduced the runtime simulation environment and other method parameters in detail.

Cell viability

CT-26 cells were cultured with RPMI-1640 cells containing 10% FBS and 100 $\mu\text{g/ml}$ 3×10^3 cells were cultured in 96-well plates for 12 h to evaluate the cellular viability of **17i** nanoparts. Next, the medium was replaced with different concentrations of **17i** solutions and **17i** nanocomposites containing 10% alcohol.

24 or 48 h later, MTT solution (1 mg/ml) was placed in an incubator for further incubation for about 4 h. Drained from supernatant, DMSO was then added to each well, and the purple crystals were fully dissolved by slowly shaking for 5 min. The absorbance value of each well was measured at 492 nm by the multifunctional microplate analyzer.

Animal studies

All animal protocols were evaluated and approved by the Animal Laboratory Ethics Committee of the Liaoning Cancer Hospital. The BALB/c mice bearing CT-26 tumors were established. PBS, **17i** solution and PEGylated **17i** nanoassemblies (20 mg/kg **17i**) were intraperitoneal-injected into the mice separately. The drug was administered every other day for five times, and the tumor volume was measured

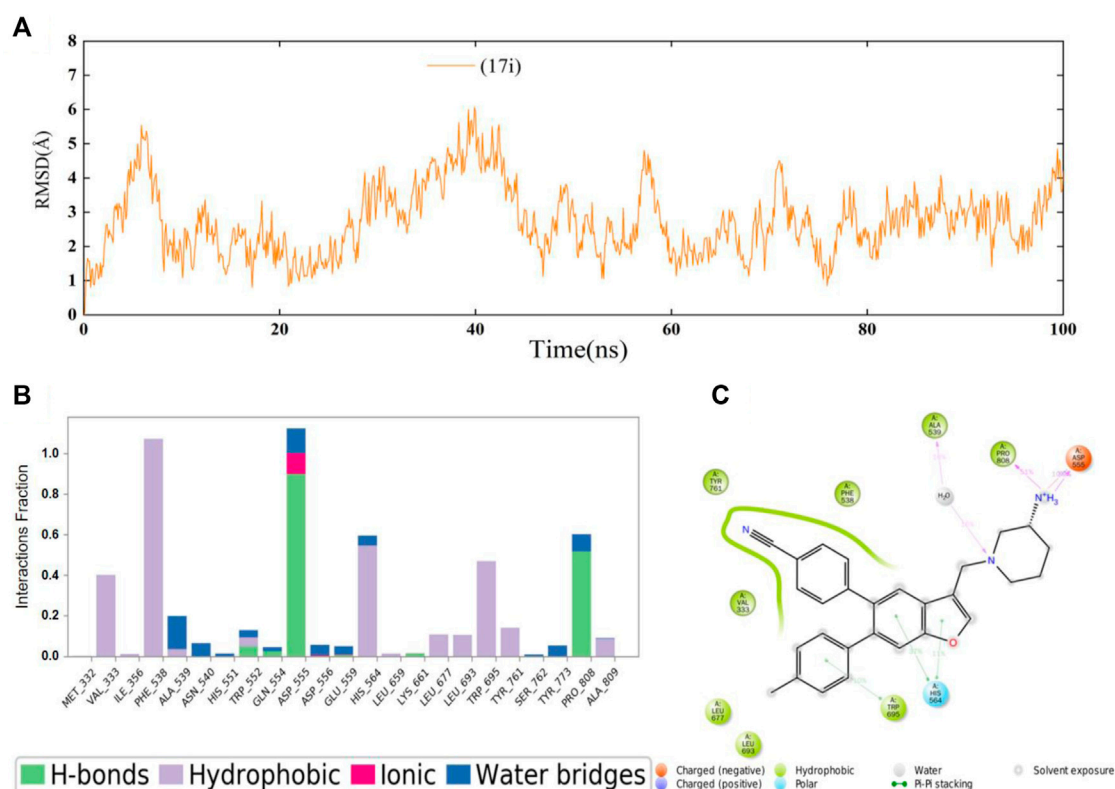


FIGURE 3

(A) Time evolution of the RMSD value of compound **17i** analyzed by 100 ns molecular dynamic (MD) simulations. (B) Statistical protein-ligand contacts of compound **17i** and (C) Two-dimensional interactions of compound **17i** during the whole MD simulations.

and weighed daily. On the last day of the efficacy trial, animals were killed, and tumor tissues were isolated, weighed, and photographed to compare the effects of different preparations on tumor growth.

Flow cytometry analysis

Tumor tissues after different treatments were extracted and collected from CT-26-bearing BALB/c mice. Tumor tissue were conducted to prepare a single-cell suspension. Then, the cells were stained with fluorescence-labeled antibodies CD3, CD4 and CD8 in compliance with the instructions of manufacturers. The proportion of stained cells from tumor tissues were measured using flow cytometry and data were analyzed using FlowJo software.

Statistical analysis

Statistical analysis was carried out using Graphpad Prism software. All data results were showed as mean \pm standard

deviation (SD). Student's T-test was used to analyze differences between groups. The $p < 0.05$ was deemed statistically significant (Sun B. et al., 2019; Sun M. et al., 2019; Wang Q. et al., 2020; Sun et al., 2022).

Results

Computational simulation studies

The binding affinity of inhibitor **17i** in LSD1 (PDB code 5YJB) was first evaluated using a combined docking and molecular dynamics protocol. Figure 2 illustrates the possible binding schema between **17i** and LSD1 was predicted by glide 9.7 software with a high binding affinity (SP score = -9.109 kcal/mol), consisting of its bioactivity ($IC_{50} = 65$ nM).

Next, 100 ns MD simulations of the protein-ligand complex (LSD1 protein and compound **17i**) were performed to predefined binding modes using Desmond v3.8. As depicted in Figure 3A, the LSD1-**17i** complex reached equilibrium at about 80 ns and the fluctuation of RMSD values (\AA) was found to be 2–4 \AA , indicating that compound **17i** was stabilized favorably with the

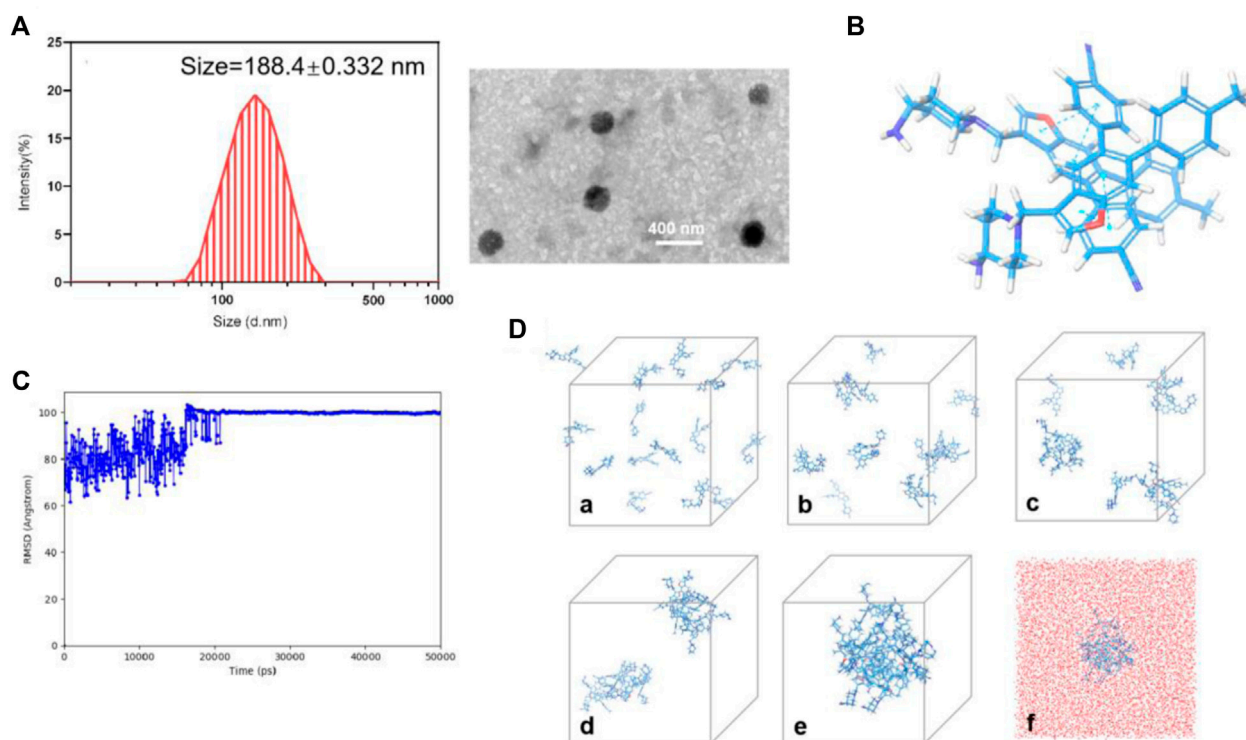


FIGURE 4

(A) Particle size distribution and TEM images of PEGylated **17i** nanoassemblies. (B) Docking diagram of the assembly mechanism of compound **17i**. (C) Root Mean Square Deviation (RMSD) values of the nanoassembly system. (D) The aggregation process of **17i** in water at (a) 0 ns, (b) 2.5 ns, (c) 6.1 ns, (d) 14 ns, (e) 50 ns and (f) 50 ns contained water beads (red beads).

active site during the binding process. Meanwhile, the contributions of amino acid interactions were also analyzed during the MD simulation in Figures 3B,C. Hydrogen bonds were formed between the piperidine-3-amine moiety of **17i** and Asp555 and Pro808, accounting for 100 and 51% respectively. The nitrogen atom on the piperidine ring also interacted with Ala539 through water molecules to form the hydrogen bond, counting for 16%. Surprisingly, the Lys661-mediated hydrogen bond interaction was weak (less than 10%). There was little that Lys661 contributed to the activity of the compound **17i** during the binding process. In addition, some new residues not determined by molecular docking were observed, such as His564. It was positioned in proximity of compound **17i** and participated in π - π stacking interactions with the ring of benzofuran ring.

Preparation and characterization of **17i** nanoassemblies

For the fabrication of **17i** nanoassemblies, one-step nanoprecipitation was used. The hydrophobic **17i** molecules spontaneously assembled into uniform **17i** nanoassemblies

formed without the aid of any carrier excipients. We constructed PEGylated **17i** nanoassemblies with a small quantity DSPE-PEG₂₀₀₀ (15 wt%). Obviously, nanoassemblies composed of **17i** molecules were obviously the main components, and PEGylated **17i** nanoassemblies loaded drugs more than 80 wt% efficiently.

17i molecules themselves were the main component of nanoassemblies, and the drug loading efficiency of the PEGylated **17i** nanoassemblies was more than 80 wt%. The hydrated particle size of the nanoassemblies by dynamic light scattering (DLS) was 188.4 ± 0.332 nm. The image showed the irregular spheres and a particle size of <200 nm in dehydrated diameter, as determined by transmission electron microscopy (TEM) (Figures 4A,B). The critical aggregation concentration (CAC) of **17i**-based nanoassemblies was equal to 1 μ g/ml compound **17i**. Subsequently, the assembly mechanisms of pure compound **17i** were then evaluated using molecular docking. Multiple intermolecular interactions led to the formation of hydrogen bonds, hydrophobic interactions, and stacking interactions in this nanosystem (Figure 4C). It was assumed that the hydrophobic interactions among compound **17i** molecules could drive to gather, and the hydrophilic amino groups in **17i** molecules were exposed and coexisted stably with water by hydrogen bonds.

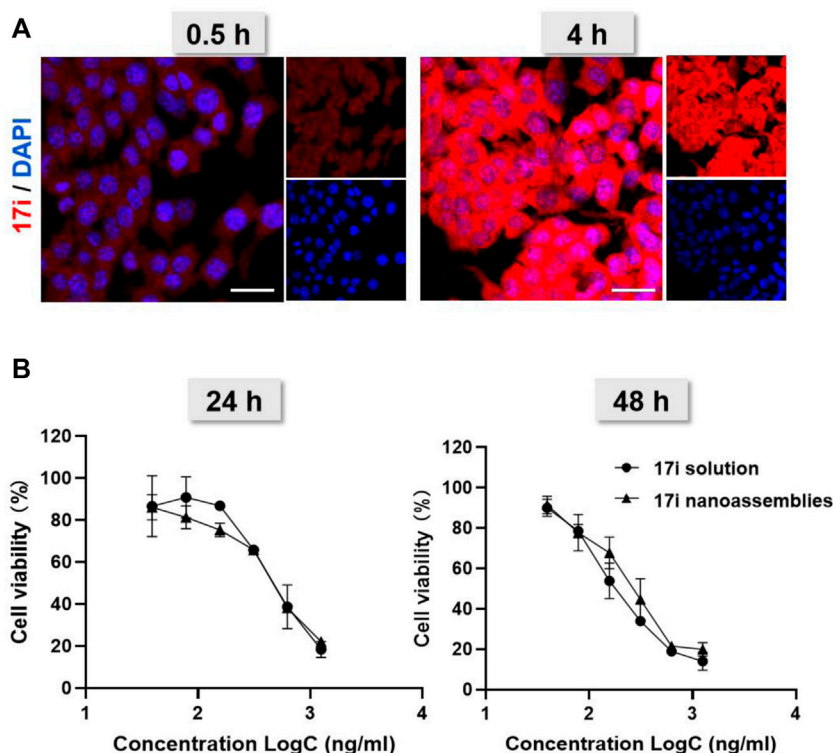


FIGURE 5

(A) Cellular uptake: confocal laser scanning microscopy (CLSM) images of CT-26 cells after being treated for 0.5 and 4 h with **17i** nanoassemblies. Scale bar = 20 μ m. (B) Viability of CT-26 cells after treated with various concentrations of **17i** solution and **17i** nanoassemblies for 24 and 48 h, respectively (n = 3).

Next, we investigated the cellular uptake of PEGylated **17i** nanoassemblies by measuring the amount of **17i** after incubation with CT-26 murine colorectal tumor cells. **Figure 5A** shows that, with the passage of time, **17i** nanoassemblies treated with PEGylated nanoparticles produced stronger red intracellular fluorescence. In addition, a MTT assay was conducted to determine the cytotoxicity of **17i** solution and PEGylated **17i** nanoassemblies *in vitro* (**Figure 5B**). For CT-26 cells, the cytotoxicity of PEGylated **17i** nanoassemblies was comparable to that of **17i** compound solutions, demonstrating that **17i** nanoassemblies had a negligible effect on its *in vitro* cytotoxicity. The wound healing assay and apoptosis induction was also used to investigate the anti-cancer activity *in vitro* (**Supplementary Figure S4**). Similar to results of the cytotoxicity experiment, the anti-cancer ability of **17i** nanoassemblies make it a potent candidate for further *in vivo* evaluation.

In vivo antitumor studies

This part evaluated the anti-tumor activity of CT-26 *in vivo*. PBS, **17i** solution and PEGylated **17i** nanoassemblies (20 mg kg⁻¹ of **17i**) were treated by intraperitoneal administration for a total of five times.

As shown in **Figure 6**, the PBS group could not inhibit the rapid growth of the tumor. In contrast, both **17i** solution and PEGylated **17i** nanoassemblies had an anti-tumor effect, and the tumor growth rate was significantly slowed down. The PEGylated **17i** nanoassemblies had a much stronger anti-tumor effect than the **17i** solution. Like the intraperitoneal administrated model, the anti-tumor activity of **17i** nanoassemblies in the intravenous administrated model was better than the other groups (**Supplementary Figure S5**). In addition, as illustrated in **Figure 7**, In immunofluorescence staining and flow cytometry, C57 mice bearing CT-26 were significantly infiltrated by CD8⁺ T cells in tumor regions after receiving PEGylated **17i** nanoassemblies. It is speculated that the appropriate particle size of PEGylated **17i** nanoassemblies was more conducive to drugs enrichment in the tumor site *via* the EPR effect. An enhanced anti-tumor efficacy was demonstrated by the presence of enriched **17i** at tumor sites, resulting in an increased number of CD8⁺ T cells infiltrating the tumor site. The PEGylated nanoassemblies showed stronger anti-tumor activity than the solution, which is determined by the pharmacologic advantages of nanoassemblies: 1) extremely high drug loading (more than 80%); 2) long circulation time in the body; 3) strong ability of tumor-targeted accumulation. These factors lead to the large

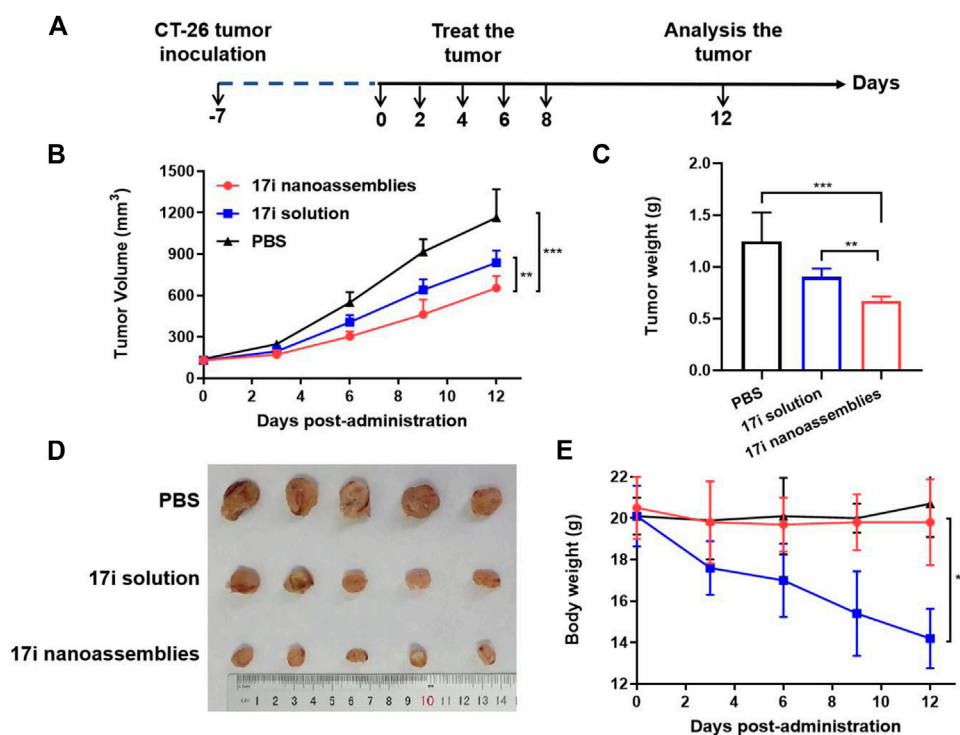


FIGURE 6

In vivo anti-tumor efficacy of 17i solution and PEGylated 17i nanoassemblies against CT-26 tumors. (A) Therapeutic protocol on mouse CT26 subcutaneous tumor xenograft. (B) Tumor volume growth curve after different treatments. (C) Tumor weight (G) of mice after 12 days in different treatments (n = 5). (D) After repeated administration, excised tumors of different groups are shown. (E) Changes in body weight of mice during different treatments (n = 5).

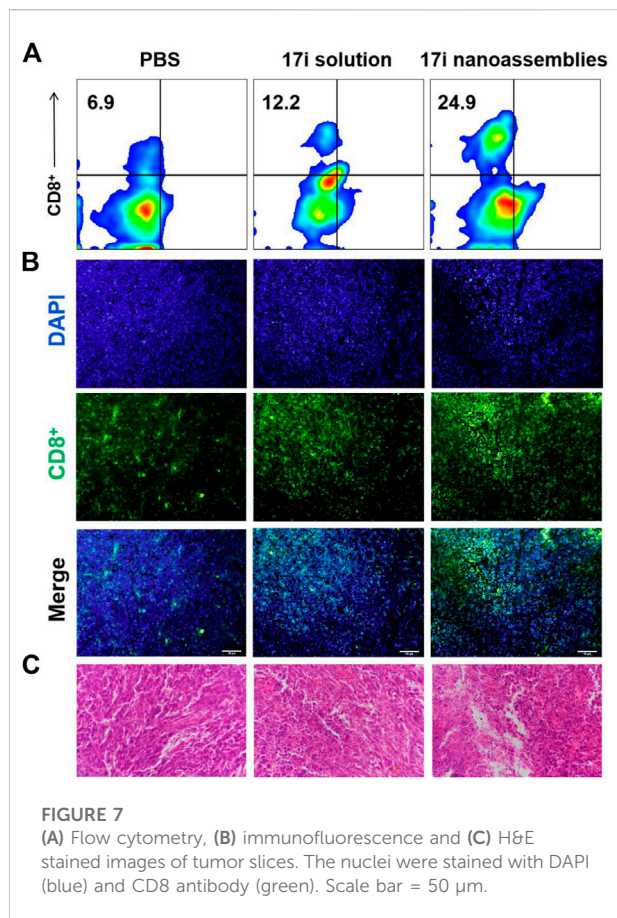
area of apoptosis and necrosis in tumor tissue, showing a great anti-tumor effect. In our previous study, the body weight of the nude mice after the treatment of compound 17i at a high dose (20 mg/kg/d) was loss to some extent during the treatment (Zhang et al., 2021). In this study, H&E stained images are provided, and blood tests are conducted to demonstrate the safety of PEGylated 17i nanoassemblies in Supplementary Figure S6, S7. This is mainly due to their biophysical targeting properties and little distribution to other tissues.

Discussion

In the clinical setting, poorly soluble drug molecules often have low bioavailability issues and absorption problems. Thus, almost 70% of potential drugs were discarded because of their poorly soluble. As the number of poorly soluble drugs increases from discovery, developing technology to enhance their solubility as well as control their release is one of the many challenges facing the pharmaceutical industry. Reducing the size of insoluble molecules is an effective solubilization method. However, there is no simple and reliable technology to manufacture and stabilize nanoparticles in

aqueous solution. Polymer carriers are used to load insoluble drug molecules, but they still face problems such as low drug loading capacity. In addition, their pharmacokinetic characteristics are closer to the properties of carrier which also affect key parameters such as distribution and drug release. Further evaluation and modification of carriers are always needed.

Carrier-free supramolecular nanoassemblies of pure compound 17i is expected to enhance its solubility without causing problems such as low drug loading capacity as mentioned above. The drug content is greater than 80%, and the preparation process is controllable and simple. Importantly, the preparation process does not involve the use of organic solvents for dissolving hydrophobic precursors, thereby solving safety issues (such as possible toxicity) from carriers or organic solvents. On the one hand, 17i nanoassemblies showed similar *in vitro* antitumor activity to free drugs. On the other hand, PEGylated 17i nanoassemblies showed longer circulation times than free drugs. We speculate that its nanostructure are not substrates for some enzymes, thus the adverse metabolism of 17i nanoassemblies caused by enzymes are reduced, which prolong the blood circulation time. In addition, we found that 17i nanoassemblies exhibit satisfactory targeting ability, possibly due to the EPR effect. We also observed that tumors regions



in C57 mice were significantly infiltrated by CD8⁺ T cells after treatment of PEGylated 17i nanoassemblies in immunofluorescence staining and flow cytometry. Obviously, 17i nanoassemblies elevated tumor immune response when killing tumor cells. The principle of this pure nano-drugs may open up the way to provide inspiration for maximizing the therapeutic potential of drug-like compounds.

Conclusion

In this paper, insoluble LSD1 inhibitor, pure compound 17i, was prepared into nanoassemblies by one-step nano-precipitation method without the addition of carrier material, and DSPE-PEG₂₀₀₀ was modified on the surface to improve stability and prolong blood circulation time. The characterization of the DLS and TEM proved that the ~200 nm PEGylated 17i nanoassemblies were successfully constructed. Multiple intermolecular forces in assemblies were observed using computational simulation. The PEGylated 17i nanoassemblies had a negligible impact on the cytotoxicity of 17i solution. As expected, the PEGylated 17i nanoassemblies exhibited distinct advantages over 17i solution in terms of therapeutic efficiency, anti-tumor immune response and side effects *in vivo*. This is the first time that the one-step

nanoprecipitation method overcomes the problem of low water solubility of anti-tumor drug-likeness LSD1 inhibitor. Such a simple and practical nanoplatform of pure LSD1 inhibitor holds a promising application prospect for clinical cancer therapy.

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

All animal protocols were evaluated and approved by the Animal Laboratory Ethics Committee of the Liaoning Cancer Hospital.

Author contributions

BL: Data curation, writing and original draft. XZ: Conceptualization and investigation. JL: Writing, conceptualization, investigation, review, editing and funding acquisition.

Funding

This work was supported by the Guiding plan of Liaoning Natural Science Foundation (No. 20180550277).

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/fchem.2022.1012882/full#supplementary-material>

References

- Friesner, R. A., Banks, J. L., Murphy, R. B., Halgren, T. A., Klicic, J. J., Mainz, D. T., et al. (2004). Glide: A new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J. Med. Chem.* 47, 1739–1749. doi:10.1021/jm0306430
- Friesner, R. A., Murphy, R. B., Repasky, M. P., Frye, L. L., Greenwood, J. R., Halgren, T. A., et al. (2006). Extra precision glide: Docking and scoring incorporating a model of hydrophobic enclosure for Protein–Ligand complexes. *J. Med. Chem.* 49, 6177–6196. doi:10.1021/jm051256o
- Ma, L.-Y., Zheng, Y.-C., Wang, S.-Q., Wang, B., Wang, Z.-R., Pang, L.-P., et al. (2015). Design, synthesis, and structure-activity relationship of novel LSD1 inhibitors based on pyrimidine-thiourea hybrids as potent, orally active antitumor agents. *J. Med. Chem.* 58, 1705–1716. doi:10.1021/acs.jmedchem.5b00037
- Mould, D. P., Alli, C., Bremberg, U., Cartic, S., Jordan, A. M., Geitmann, M., et al. (2017). Development of (4-cyanophenyl)glycine derivatives as reversible inhibitors of lysine specific demethylase 1. *J. Med. Chem.* 60, 7984–7999. doi:10.1021/acs.jmedchem.7b00462
- Schmitt, M. L., Hauser, A.-T., Carlino, L., Pippel, M., Schulz-Fincke, J., Metzger, E., et al. (2013). Nonpeptidic propargylamines as inhibitors of lysine specific demethylase 1 (LSD1) with cellular activity. *J. Med. Chem.* 56, 7334–7342. doi:10.1021/jm400792m
- Shi, Y., Lan, F., Matson, C., Mulligan, P., Whetstone, J. R., Cole, P. A., et al. (2004). Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. *Cell* 119, 941–953. doi:10.1016/j.cell.2004.12.012
- Sorna, V., Theisen, E. R., Stephens, B., Warner, S. L., Bearss, D. J., Vankayalapati, H., et al. (2013). High-throughput virtual screening identifies novel N'-(1-Phenylethylidene)-benzohydrazides as potent, specific, and reversible LSD1 inhibitors. *J. Med. Chem.* 56, 9496–9508. doi:10.1021/jm400870h
- Sun, B., Luo, C., Zhang, X., Guo, M., Sun, M., Yu, H., et al. (2019a). Probing the impact of sulfur/selenium/carbon linkages on prodrug nanoassemblies for cancer therapy. *Nat. Commun.* 10, 3211. doi:10.1038/s41467-019-11193-x
- Sun, M., Jiang, H., Liu, T., Tan, X., Jiang, Q., Sun, B., et al. (2022). Structurally defined tandem-responsive nanoassemblies composed of dipeptide-based photosensitive derivatives and hypoxia-activated camptothecin prodrugs against primary and metastatic breast tumors. *Acta Pharm. Sin. B* 12, 952–966. doi:10.1016/j.apsb.2021.08.008
- Sun, M., Zhang, X., Gao, Z., Liu, T., Luo, C., Zhao, Y., et al. (2019b). Probing a dipeptide-based supramolecular assembly as an efficient camptothecin delivering carrier for cancer therapy: Computational simulations and experimental validations. *Nanoscale* 11, 3864–3876. doi:10.1039/C8NR07014H
- Suzuki, T., and Miyata, N. (2011). Lysine demethylases inhibitors. *J. Med. Chem.* 54, 8236–8250. doi:10.1021/jm201048w
- Tortorici, M., Borrello, M. T., Tardugno, M., Chiarelli, L. R., Pilotto, S., Ciossani, G., et al. (2013). Protein recognition by short peptide reversible inhibitors of the chromatin-modifying LSD1/CoREST lysine demethylase. *ACS Chem. Biol.* 8, 1677–1682. doi:10.1021/cb4001926
- Vianello, P., Botrugno, O. A., Cappa, A., Dal Zuffo, R., Dessanti, P., Mai, A., et al. (2016). Discovery of a novel inhibitor of histone lysine-specific demethylase 1A (KDM1A/LSD1) as orally active antitumor agent. *J. Med. Chem.* 59, 1501–1517. doi:10.1021/acs.jmedchem.5b01209
- Wang, Q., Sun, M., Li, D., Li, C., Luo, C., Wang, Z., et al. (2020a). Cytochrome P450 enzyme-mediated auto-enhanced photodynamic cancer therapy of co-nanoassembly between clopidogrel and photosensitizer. *Theranostics* 10, 5550–5564. doi:10.7150/thno.42633
- Wang, X., Zhang, C., Zhang, X., Yan, J., Wang, J., Jiang, Q., et al. (2020b). Design, synthesis and biological evaluation of tetrahydroquinoline-based reversible LSD1 inhibitors. *Eur. J. Med. Chem.* 194, 112243. doi:10.1016/j.ejmech.2020.112243
- Wang, Y., Hu, B., Peng, Y., Xiong, X., Jing, W., Wang, J., et al. (2019). *In silico* exploration of the molecular mechanism of cassane diterpenoids on anti-inflammatory and immunomodulatory activity. *J. Chem. Inf. Model.* 59, 2309–2323. doi:10.1021/acs.jcim.8b00862
- Wu, F., Zhou, C., Yao, Y., Wei, L., Feng, Z., Deng, L., et al. (2016). 3-(Piperidin-4-ylmethoxy)pyridine containing compounds are potent inhibitors of lysine specific demethylase 1. *J. Med. Chem.* 59, 253–263. doi:10.1021/acs.jmedchem.5b01361
- Yang, Y., Sun, B., Zuo, S., Li, X., Zhou, S., Li, L., et al. (2020). Trisulfide bond-mediated doxorubicin dimeric prodrug nanoassemblies with high drug loading, high self-assembly stability, and high tumor selectivity. *Sci. Adv.* 6, 1725. doi:10.1126/sciadv.abc1725
- Zhang, X., Huang, H., Zhang, Z., Yan, J., Wu, T., Yin, W., et al. (2021). Design, synthesis and biological evaluation of novel benzofuran derivatives as potent LSD1 inhibitors. *Eur. J. Med. Chem.* 220, 113501. doi:10.1016/j.ejmech.2021.113501
- Zhang, X., Li, N., Zhang, S., Sun, B., Chen, Q., He, Z., et al. (2020). Emerging carrier-free nanosystems based on molecular self-assembly of pure drugs for cancer therapy. *Med. Res. Rev.* 40, 1754–1775. doi:10.1002/med.21669
- Zhao, D., Tao, W., Li, S., Chen, Y., Sun, Y., He, Z., et al. (2021). Apoptotic body-mediated intercellular delivery for enhanced drug penetration and whole tumor destruction. *Sci. Adv.* 7, 880. doi:10.1126/sciadv.abg0880
- Zheng, Y.-C., Duan, Y.-C., Ma, J.-L., Xu, R.-M., Zi, X., Lv, W.-L., et al. (2013). Triazole-dithiocarbamate based selective lysine specific demethylase 1 (LSD1) inactivators inhibit gastric cancer cell growth, invasion, and migration. *J. Med. Chem.* 56, 8543–8560. doi:10.1021/jm401002r
- Zheng, Y.-C., Ma, J., Wang, Z., Li, J., Jiang, B., Zhou, W., et al. (2015). A systematic review of histone lysine-specific demethylase 1 and its inhibitors. *Med. Res. Rev.* 35, 1032–1071. doi:10.1002/med.21350



OPEN ACCESS

EDITED BY

Mohammad Ehtisham Khan,
Jazan University, Saudi Arabia

REVIEWED BY

Muniba Faiza,
Nazarbayev University, Kazakhstan
Mohd Shuaib,
Central University of Punjab, India

*CORRESPONDENCE

Dongxiao Su,
dongxsu@126.com
Ankita Kumari,
anki93.ag@gmail.com
Abdul Rahaman,
rahaman_knabdul@gmail.com

SPECIALTY SECTION

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

RECEIVED 02 September 2022

ACCEPTED 12 September 2022

PUBLISHED 30 September 2022

CITATION

He S, Nahhas AF, Habib AH, Alshehri MA,
Alshamrani S, Asiri SA, Alnamshan MM,
Helmi N, Al-Dhuayan I, Almulhim J,
Alharbi AM, Su D, Kumari A and
Rahaman A (2022), Identification of
compelling inhibitors of human
norovirus 3CL protease to combat
gastroenteritis: A structure-based virtual
screening and molecular
dynamics study.
Front. Chem. 10:1034911.
doi: 10.3389/fchem.2022.1034911

COPYRIGHT

© 2022 He, Nahhas, Habib, Alshehri,
Alshamrani, Asiri, Alnamshan, Helmi, Al-
Dhuayan, Almulhim, Alharbi, Su, Kumari
and Rahaman. This is an open-access
article distributed under the terms of the
Creative Commons Attribution License
(CC BY). The use, distribution or
reproduction in other forums is
permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original
publication in this journal is cited, in
accordance with accepted academic
practice. No use, distribution or
reproduction is permitted which does
not comply with these terms.

Identification of compelling inhibitors of human norovirus 3CL protease to combat gastroenteritis: A structure-based virtual screening and molecular dynamics study

Shan He^{1,2,3}, Alaa F. Nahhas⁴, Alaa Hamed Habib⁵,
Mohammed Ali Alshehri⁶, Saleh Alshamrani⁶, Saeed A. Asiri⁶,
Mashaal M. Alnamshan⁷, Nawal Helmi^{8,9}, Ibtesam Al-Dhuayan¹⁰,
Jawaher Almulhim¹¹, Ahmed M. Alharbi¹², Dongxiao Su^{1*},
Ankita Kumari^{13*} and Abdul Rahaman^{13*}

¹School of Chemistry and Chemical Engineering, Guangzhou University, Guangzhou, China, ²Institute for Nano Scale and Technology, College of Science and Engineering, Flinders University, Adelaide, SA, Australia, ³Suzhou Ultra-Water-Cleaning Tech, Pty, Ltd., Suzhou, Jiangsu, China, ⁴Biochemistry Department, Faculty of Science, King Abdulaziz University, Jeddah, Saudi Arabia, ⁵Department of Physiology, Faculty of Medicine, King Abdulaziz University, Jeddah, Saudi Arabia, ⁶Department of Clinical Laboratory Sciences, Faculty of Applied Medical Sciences, Najran University, Najran, Saudi Arabia, ⁷Department of Biology, College of Science, Imam Abdulrahman Bin Faisal University, Dammam, Saudi Arabia, ⁸Department of Biochemistry, College of Science, University of Jeddah, Jeddah, Saudi Arabia, ⁹Department of Medical Laboratory Technology, College of Applied Medical Sciences, University of Jeddah, Jeddah, Saudi Arabia, ¹⁰Department of Biology, College of Science, Imam Abdulrahman Bin Faisal University, Dammam, Saudi Arabia, ¹¹Department of Biological Sciences, King Faisal University, Alahsa, Saudi Arabia, ¹²Department of Clinical Laboratory Sciences, College of Applied Medical Sciences, University of Hail, Hail, Saudi Arabia, ¹³School of Food Science and Engineering, South China University of Technology, Guangzhou, China

Human noroviruses (NV) are the most prevalent cause of sporadic and pandemic acute gastroenteritis. NV infections cause substantial morbidity and death globally, especially amongst the aged, immunocompromised individuals, and children. There are presently no authorized NV vaccines, small-molecule therapies, or prophylactics for humans. NV 3CL protease (3CLP) has been identified as a promising therapeutic target for anti-NV drug development. Herein, we employed a structure-based virtual screening method to screen a library of 700 antiviral compounds against the active site residues of 3CLP. We report three compounds, Sorafenib, YM201636, and LDC4297, that were revealed to have a higher binding energy (BE) value with 3CLP than the control (Dipeptidyl inhibitor 7) following a sequential screening, in-depth molecular docking and visualization, physicochemical and pharmacological property analysis, and molecular dynamics (MD) study. Sorafenib, YM201636, and LDC4297 had BEs of -11.67, -10.34, and -9.78 kcal/mol with 3CLP, respectively, while control had a BE of -6.38 kcal/mol. Furthermore, MD simulations of the two best compounds and control were used to further optimize the interactions, and a 100 ns MD simulation revealed that they form stable complexes with 3CLP. The estimated physicochemical,

drug-like, and ADMET properties of these hits suggest that they might be employed as 3CLP inhibitors in the management of gastroenteritis. However, wet lab tests are a prerequisite to optimize them as NV 3CLP inhibitors.

KEYWORDS

noroviruses, gastroenteritis, protease, natural compounds, molecular dynamics

Introduction

Human noroviruses (NV), which are members of the Caliciviridae family, are the leading cause of acute gastroenteritis globally, with substantial morbidity and a significant economic burden (Koo et al., 2010; Lopman et al., 2016). NV infections are difficult to combat because of their easy food and waterborne transmission, their genomic diversity, and environmental stability (Hall, 2012). The situation is aggravated further by the absence of diagnostics and NV-specific treatments and prophylactics, such as vaccinations (Kaufman et al., 2014; Rocha-Pereira et al., 2014). Therefore, the discovery of anti-NV small-molecule therapies and prophylactics, as well as efficient vaccinations, is an imperative and unmet medical need.

The NV genome is made up of a positive-strand RNA with three open reading frames encoding: i) polyprotein, ii) minor capsid protein, and iii) main capsid protein. The polyprotein is processed by a virally encoded 3CL protease (3CLP), a cysteine protease having cysteine139-histidine30-glutamate54 catalytic triad, an extended binding cleft, and a major substrate selectivity for a P₁ glutamine (or glutamate) residue, yielding six nonstructural proteins required for NV replication (Hussey et al., 2011; Muhaxhiri et al., 2013). NV 3CLP is important in the virus's life cycle, making it ideal for the development/discovery of anti-norovirus treatments and prophylactics (Chang et al., 2019; Netzler et al., 2019). Peptidyl and macrocyclic transition state inhibitors, as well as transition state mimics, are only a few of the 3CLpro inhibitors that have been found to exhibit anti-NV activity (Deng et al., 2013; Galasiti Kankanamalage et al., 2015; Damalanka et al., 2016; Weerawarna et al., 2016). In addition, a dipeptidyl transition state inhibitor of the NV 3CLpro has also been shown to work in a mouse model (Galasiti Kankanamalage et al., 2015).

Traditional drug development is time-consuming and costly, taking an average of 10–15 years to reach the market and costing an average of 58.8 billion USD in 2015 (Mullard, 2016). These figures represent a dramatic 10% rise over previous years for both the biotechnology and pharmaceutical sectors. The high failure rate and high expense of this conventional approach to drug development have necessitated the adoption of computer-assisted drug development (Schaduengrat et al., 2020). The various adverse effects of drugs that result in severe toxicity necessitate the screening of drug likeness and

physicochemical properties at the early stage of drug development process to maximize success and minimize the time spent screening candidates (Hughes et al., 2011). Here, we aimed to find novel NV 3CLP inhibitors using the *in silico* approach to combat gastroenteritis.

Methodology

Retrieval and preparation of 3CLP and compound library

The 3D structure of 3CLP, which has the PDB ID 5T6F, was taken from the protein data bank (PDB) (Galasiti Kankanamalage et al., 2017). In order to clean up the complex, the hetero atoms and water molecules had to be removed. Whereupon, using the steepest descent method with an RMS gradient of 0.1, energy minimization of protein was performed for 1,000 steps. This study employs a unique collection of 700 compounds known to target HCV protease, HIV protease, Integrase, Reverse Transcriptase, and other enzymes, including some FDA-approved compounds. The antiviral compounds were retrieved in 'sdf' format, prepared by minimization, and converted to 'pdbqt' format for virtual screening (VS.).

Structure-based virtual screening

The prepared library of compounds in pdbqt format was used for VS. against the active site residues of the 3CLP with the PyRx 0.8 program (Dallakyan and Olson, 2015). The grid center of the protein was set as X = 4.963, Y = 67.188, and Z = -6.787.

Molecular docking simulations

The top 20 compounds, including the control (Dipeptidyl inhibitor 7) (Supplementary Table S1), were subjected to an in-depth molecular docking simulation following the virtual screening to optimize the binding conformations of these compounds. Autodock4.2 was used to do a docking analysis with the default settings and the same grid center as virtual screening.

Prediction of physicochemical, drug-likeness, and ADMET properties

By identifying lead molecules, computational approaches help improve the success rate of experimental drug trials. The computational prediction of pharmacokinetic and ADMET properties of small molecules provides the clue to narrowing down the screening and their potential to be drug-like molecules. The efficacy and safety profiles of the selected hits and their pharmacokinetics were predicted using SwissADME (Daina et al., 2017) and the DataWarrior tool (López-López et al., 2019).

The attributes of the bioactivity score predict the overall potential of the three best selected hits to be an effective lead candidate. An online tool 'Molinspiration chemoinformatics' (<https://www.molinspiration.com>), was used to evaluate the drug score of selected hits in relation to various human receptors such as ion channels, GPCRs, enzymes, kinases, proteases, and nuclear receptors. On the whole, a greater bioactivity score specifies that the active compound is more likely to be active.

Molecular dynamics simulation

GROMACS 5.1.2 (Van Der Spoel et al., 2005) was used for MD simulations on 3CLP -control, 3CLP Protease-Sorafenib, and 3CLP -YM201636 at 300 K, with the GROMOS96 43a1 force-field (Pol-Fachin et al., 2009). The PRODRG server was used to produce the compound's topology and force-field parameters (Schüttelkopf and Van Aalten, 2004). Charges were manually

corrected in the topology file, new compound atoms were added to the complex topology files, and all of the compounds' attributes were included in the system topology. 3CLP-control, 3CLP-Sorafenib, and 3CLP -YM201636 were immersed in a 'cubic box' of water molecules with an initial diameter of 8 nm using the 'gmxditconf' module for boundary conditions and the 'gmxditconf' module for solvation. Adding Na⁺ and Cl⁻ ions to preserve neutrality and a physiological concentration using the gmxditconf module (0.15 M) neutralized the charges on the complexes. PyMOL and VMD have been used to generate all visualizations of the 3D models (Humphrey et al., 1996; Amarnath Jonniya et al., 2021).

Results and discussion

Proteases are a type of enzyme that plays an important role in a several biological processes in living organisms ranging from viruses to mammals. NV 3CLP is a major viral target for anti-NV drug development due to its critical role in viral replication (Thorne and Goodfellow, 2014). This study screened 700 antiviral compounds against NV 3CLP. Sorafenib, LDC4297, and YM201636 were identified as potential lead compounds after sequential screening and interaction analysis of the complexes, as they interacted strongly with 3CLP (Figure 1). Sorafenib interacted with Glu54, Ile109, Gln110, Arg112, Val114, Leu132, Gly133, Thr134, Gly137, Ala158, His30, Ala159, Ala160, and Lys162 residues of 3CLP with a binding energy (BE) of -11.67 kcal/mol (Figure 2; Table 1). LDC4297 had a BE of -9.78 kcal/mol, and interacted with

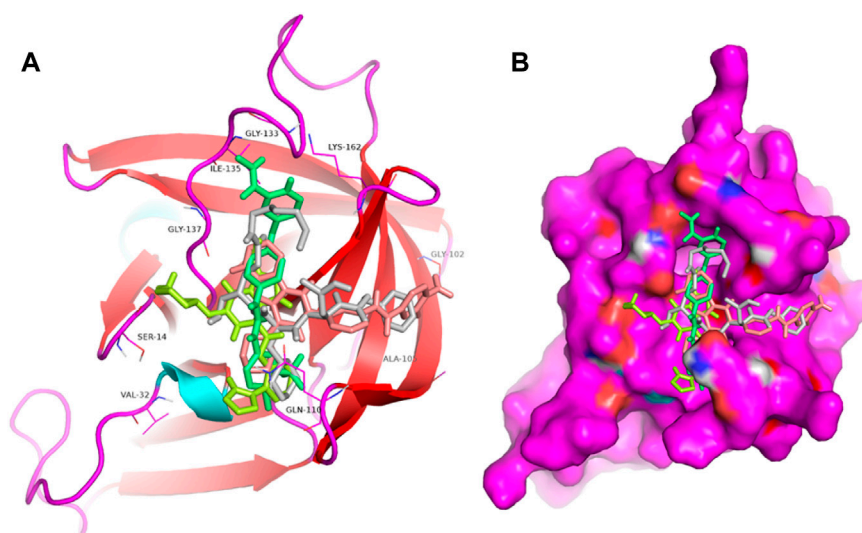


FIGURE 1
Structural alignment of screened lead compounds (sorafenib, LDC4297, YM201636) and dipeptidyl inhibitor seven in the 3CLP binding pocket (A,B).

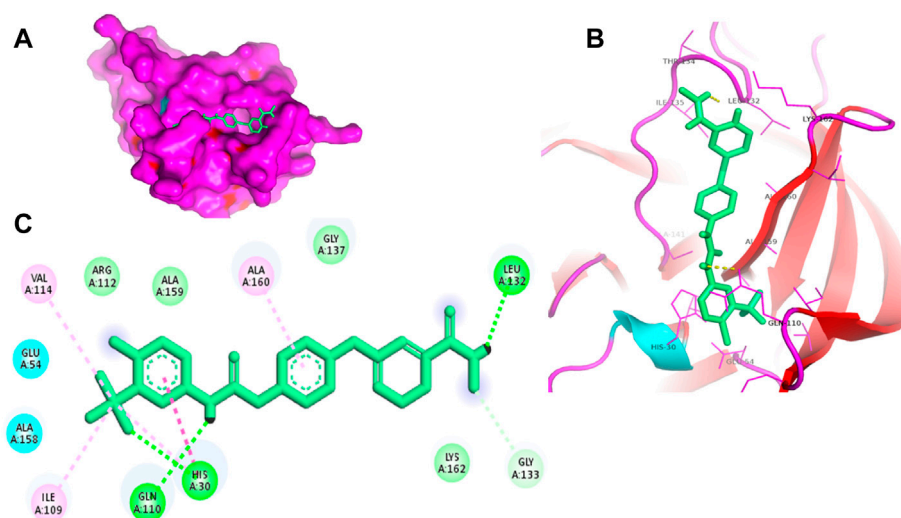


FIGURE 2

Surface view of sorafenib in the 3CLP binding pocket (A), 3D (B) and 2D (C) interacting residues of 3CLP with sorafenib.

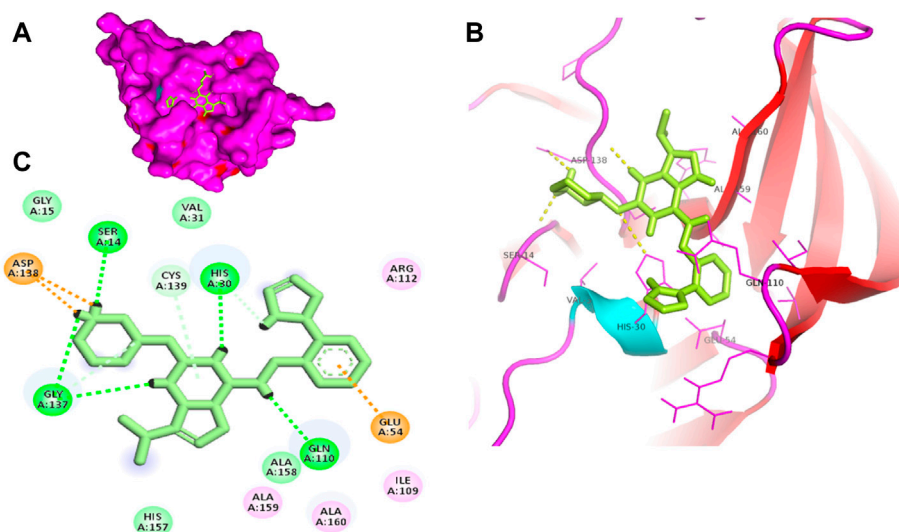
TABLE 1 BE of lead compounds with NV 3CLP.

Compounds	Structure	Binding energy (kcal/mol)	H-bonds interacting residues
Sorafenib		-11.67	His30, Gln110, and Leu132
LDC4297		-9.78	Ala158, and Ala160
YM201636		-10.34	His30 and Ala160
Dipeptidyl inhibitor 7 ^a		-6.38	His30, Gln110, Thr134, Ala158, and Ala160

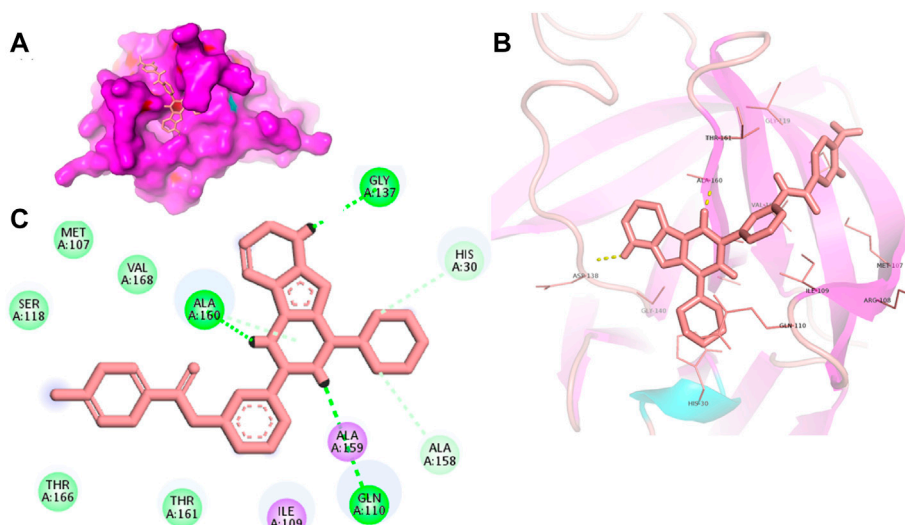
^aReference inhibitor.

Thr28, His30, Glu54, Arg108, Ile109, Gln110, Arg112, Val114, Ile135, Pro136, Gly137, Cys139, Ala158, Ala159, Ala160, Thr161, Lys162, and Val168 residues of 3CLP (Figure 3; Table 1). Further, YM201636 interacted with His30, Ala105, Met107, Arg108,

Ile109, Gln110, Ser118, Thr134, Ile135, Pro136, His157, Ala158, Ala159, Ala160, Thr161, Lys162, and Val168 residues, and has a BE of -10.34 kcal/mol with 3CLP (Figure 4; Table 1). The 3CLP protease residues Ala158, Ala160, Val168, and

**FIGURE 3**

Surface view of LDC4297 in the 3CLP binding pocket (A), 3D (B) and 2D (C) interacting residues of 3CLP with LDC4297.

**FIGURE 4**

Surface view of YM201636 in the 3CLP binding pocket (A), 3D (B) and 2D (C) interacting residues of 3CLP with YM201636.

Ile109 have been shown to be important in inhibitor binding (Galasiti Kankanamalage et al., 2017). Consistent with this, the hits Sorafenib, LDC4297, and YM201636 have been found to interact with these 3CLP residues.

NV 3CLP's mode of action is similar to that of comparable cysteine proteases, in which Cys139 functions as a nucleophile, His30 acts as a general acid/base, and Glu54 aids in the alignment of His30 and stimulates the deprotonation of Cys139 (Chang

et al., 2019). Interestingly, this study showed that the lead compounds sorafenib, LDC4297, and YM201636 interacted with these residues (His30, Glu54, and Cys139) of NV 3CLP, possibly inhibiting the 3CLP.

Moreover, to obtain a better picture of 3CLP interacting residues with leads (sorafenib, LDC4297, and YM201636), 3CLP interacting residues with its co-crystallized inhibitor (dipeptidyl inhibitor seven; PDB ID: 5T6F) were analyzed, which showed

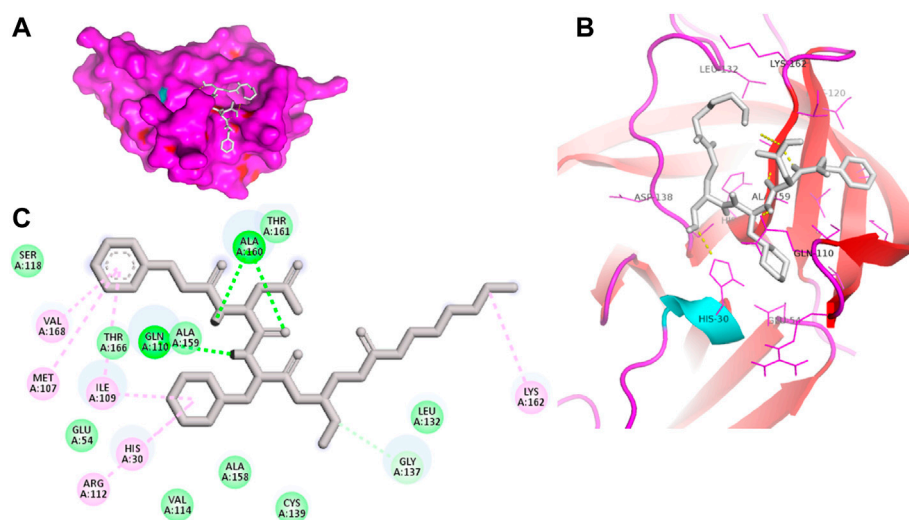


FIGURE 5

Surface view of dipeptidyl inhibitor seven in the 3CLP binding pocket (A), 3D (B) and 2D (C) interacting residues of 3CLP with dipeptidyl inhibitor 7.

TABLE 2 Bioactivity score of top three hits.

Properties		Top 3 hits		
		Sorafenib	YM201636	LDC4297
Ligands	GPCR	0.18	0.03	0.04
	Nuclear receptor	-0.07	-0.81	-0.55
Ion channel modulator	0	-0.38	-0.27	
Inhibitors	Kinase	0.44	0.39	0.36
	Protease	0.11	-0.32	-0.23
	Enzyme	0.08	0.27	-0.11

that Gln110, Arg112, Val114, His130, Thr134, Glu54, Met107, Arg108, Ile109, Ile135, Pro136, Gly137, Cys139, Lys152, His157, Ala158, Ala159, Ala160, Thr161, Lys162, and Val168 residues were important in binding with dipeptidyl inhibitor 7 (Figure 5). Interestingly, Ile109, Gln110, Ala158, Ala159, Ala160, and Lys162 were the common binding residues of 3CLP with sorafenib, LDC4297, YM201636, and the dipeptidyl inhibitor 7 (Figure 2, 3, 4, and 5), revealing that the binding mode of these compounds in the 3CLP catalytic pocket was similar to that of the reference inhibitor.

The intention of ligand-protein docking is to anticipate the most probable binding modes of the ligand with the catalytic pocket residues of the target protein, with a high BE value (more

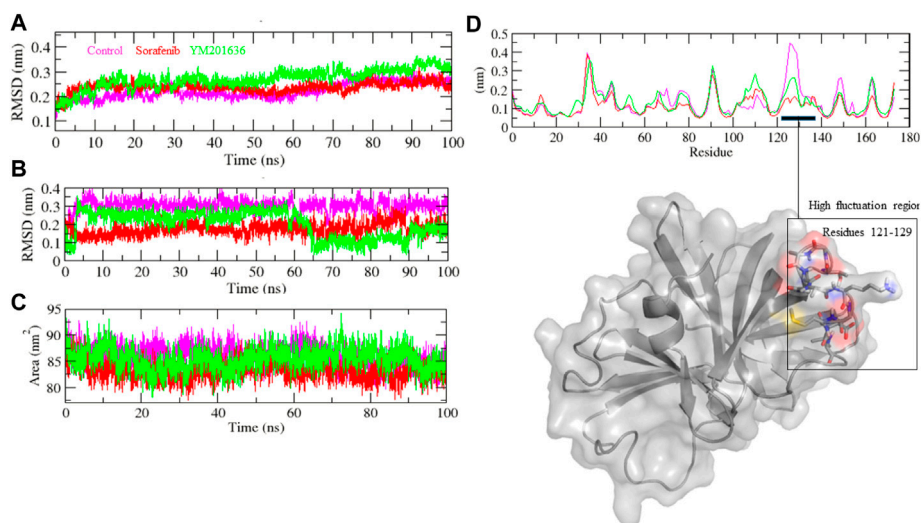
negative) implying an effective interaction between ‘inhibitor-protein’ complexes (Meng et al., 2011). Sorafenib, LDC4297, and YM201636 have higher BEs than the control dipeptidyl inhibitor 7 (Table 1), indicating that these leads bind to the NV 3CLP strongly.

Despite advances in developing effective antiviral therapies, the currently available antiviral agents have various issues, including high prices, drug resistance, safety, and effectiveness limits (Antonelli and Turriziani, 2012). Further, we predicted the physicochemical, drug-likeness, and ADMET properties of these selected hits. Table 2 shows the bioactivity scores predicted by the Molinspiration web tool for sorafenib, LDC4297, and

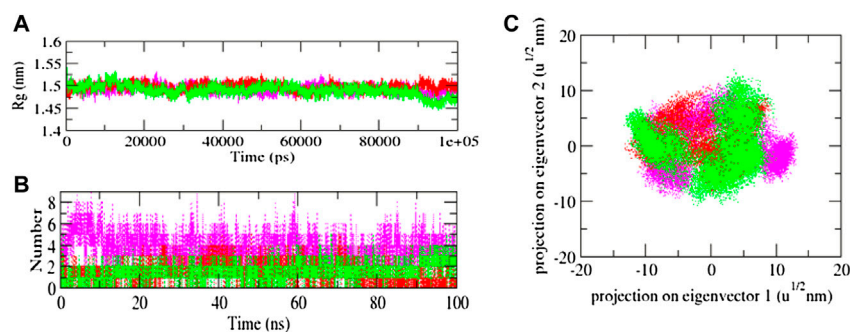
TABLE 3 Physicochemical and druglikeness properties of selected hits.

Selected hits		Sorafenib	YM201636	LDC4297	
Physicochemical Properties	Molwt	464.83	467.488	432.53	
	cLogP	4.1428	2.7535	1.7767	
	cLogS	-6.689	-7.308	-4.913	
	H-Acceptors	7	10	9	
	H-Donors	3	2	2	
	TPSA	92.35	132.29	94.19	
Lipophilicity	iLOGP	3.42	3.08	4.04	
	XLOGP3	4.07	2.46	3.43	
	WLOGP	6.32	2.94	2.45	
	MLOGP	2.91	1.18	2.63	
	Silicos-IT Log P	3.78	2.47	1.65	
	Consensus Log P	4.1	2.43	2.84	
Water Solubility	ESOL	Log S	-5.11	-4.49	-4.68
		Solubility (mg/ml)	3.62E-03	1.52E-02	8.97E-03
		Class	Moderately soluble	Moderately soluble	Moderately soluble
	Ali	Log S	-5.71	-4.88	-5.09
		Solubility (mg/ml)	8.98E-04	6.13E-03	3.53E-03
		Class	Moderately soluble	Moderately soluble	Moderately soluble
Pharmacokinetics	GI absorption	L	H	H	
	BBB permeant	N	N	N	
	Pgp substrate	N	Y	Y	
	Inhibitor	CYP1A2	Y	Y	Y
		CYP2C19	Y	Y	N
		CYP2C9	Y	Y	Y
		CYP2D6	Y	Y	Y
		CYP3A4	Y	Y	Y
	log Kp (cm/s)	-6.25	-7.41	-6.5	
Druglikeness	Violations	Lipinski	0	0	0
		Ghose	1	1	0
		Veber	0	0	0
		Egan	1	1	0
		Muegge	0	0	0
	Bioavailability Score	0.55	0.55	0.55	
	Mutagenic	N	N	N	
	Tumorigenic	N	N	N	
	Reproductive Effective	N	N	N	
	Irritant	N	N	N	
	Druglikeness score	-5.1185	2.2644	4.0968	

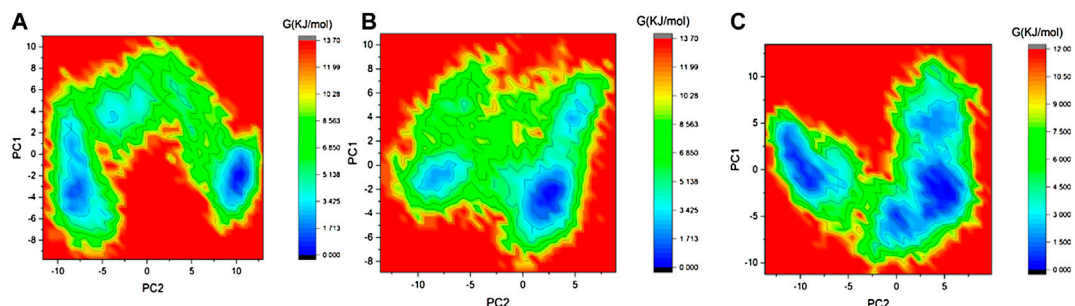
(Y=Yes, N=No, None, H=High, L = Low).

**FIGURE 6**

Structural stability studies of complexes. 3CLP-Sorafenib, 3CLP-YM201636, and 3CLP-control complexes were depicted in red, green, and pink color, respectively. RMSD plot of backbone of protein with complexes (A), RMSD plot of ligands within pocket of protein (B), SASA plot of complexes (C), and RMSF plot of backbone of protein (D).

**FIGURE 7**

Radius of gyration plot of complexes (A), Hydrogen bond interaction (B), and 2D plot of projection (C).

**FIGURE 8**

GFE landscape of 3CLP-control (A), 3CLP-Sorafenib (B), and 3CLP-YM201636 (C) complexes.

YM201636. The chemical compound is active if its bioactivity score is greater than 0.0; moderately active if it is between -5.0 and 0.0; and inactive if it is less than -5.0. As per calculated values, the identified hits, Sorafenib, YM201636, and LDC4297 are physiologically active/moderately active substances and meet the criteria mentioned here.

The pharmacokinetics viability and druglikeness characteristics of identified hits computed using the SwissADME and DataWarrior tools indicated that they should be promising lead candidates. Table 3 displays the values of the different properties, viz., physicochemical properties, pharmacokinetics, and druglikeness of the Sorafenib, YM201636, and LDC4297.

Evaluation of pharmacokinetic properties for a success therapeutic in the early stages of drug design by *in silico* ADME assays is critical to achieve druglikeness and decrease risk attrition in the advanced stages. The predicted compounds had acceptable drug-likeness and pharmacokinetic features, and they could not be P-glycoprotein substrates (P-gp). The cytochrome P450 monooxygenase (CYP) enzyme superfamily, which includes cytochrome CYP1A2, CYP2C19, CYP2C9, CYP2D6, and CYP3A4, is crucial in drug metabolism in the liver, and drug biotransformation by O-type oxidation processes, particularly those of 2D6, 2C9, and 3A4, has been anticipated. Sorafenib, YM201636, and LDC4297 were shown to be CYP1A2, CYP2D6, and CYP2C9 inhibitors rather than CYP2C19 inhibitors. These three compounds had substantial lipophilicity and were moderately soluble in water.

Finally, based on BE, sorafenib and YM201636 including the control were chosen for MD simulation studies to assess the stability of complexes. The root mean square deviation (RMSD) is a protein stability metric; the slighter the deviations, the more stable the protein structure (Aier et al., 2016). 3CLP-control, 3CLP-Sorafenib, and 3CLP-YM201636 had RMSD average values of 0.21, 0.23, and 0.32 nm, respectively. The RMSD figure revealed that 3CLP-control and 3CLP-Sorafenib binding increased 3CLP stability and resulted in fewer structural aberrations from its normal conformation. The bound structure of the 3CLP-YM201636 complex is highly deviated (Figure 6A); it showed that the catalytic pocket of 3CLP not forming well interaction with YM201636, therefore it showed high deviation. In addition, the ligand RMSD also exhibits that control and Sorafenib bind better than YM201636 and they showed more stability (Figure 6A).

The 3CLP-Sorafenib and 3CLP-YM201636 backbones displayed consistent fluctuations in the 3CLP catalytic pocket, most probably due to differing orientations, with the biggest fluctuation regions found between 25–30 and 120–125 residues (Figure 6D). The vibrations around the equilibrium are not random; rather, they are determined by the flexibility of the local structure. The average fluctuation of all residues throughout the simulation, as well as the root

mean square fluctuation (RMSF) of 3CLP during binding of 3CLP-control, 3CLP-Sorafenib, and 3CLP-YM201636, were plotted as a function of 3CLP protease residue numbers. The RMSF plot revealed that 3CLP has residual variants in various protein domain areas. Due to their close interaction with the 3CLP. 3CLP-control, and 3CLP-Sorafenib have been shown to reduce the residual fluctuations of the protease.

The solvent-accessible surface area (SASA) of a protein is the area of its surface that is involved in the interaction with its solvent molecules. Average SASA values for 3CLP-control, 3CLP-Sorafenib, and 3CLP-YM201636 complexes were recorded during the 100 ns simulation. The SASA values for the 3CLP-control, 3CLP-Sorafenib, and 3CLP-YM201636 complexes were 88.02, 84.01, and 85.30 nm², respectively (Figure 6C). SASA analysis showed that upon binding of Sorafenib, surface exposure has been reduced. Further to gain insight of the complex stability/compactness profile in a biological system, we applied the Radius of gyration (Rg). The 3CLP-control, 3CLP-Sorafenib, and 3CLP-YM201636 complexes had average Rg values 1.50, 1.53, and 1.47 nm, respectively. Stable Rg trajectories were observed within the catalytic pocket of 3CL protease (Figure 7A).

The hydrogen bond is vital to the stability of the ligand-target complex (Hubbard and Kamran Haider, 2001). Between protein and ligand, hydrogen bonds were formed within 0.35 nm. The stability of docked complexes was tested using 100 ns simulations of 3CLP-control, 3CLP-Sorafenib, and 3CLP-YM201636 in a solvent environment. 3CLP-control, 3CLP-Sorafenib form 3–6 hydrogen bonds with the 3CLP catalytic pocket, whereas 3CLP-YM201636 forms 1–2 hydrogen bonds with the 3CLP catalytic pocket (Figure 7B).

The Gibbs free energy (GFE) landscape of the 3CLP-control, 3CLP-Sorafenib, and 3CLP-YM201636 complexes has been plotted. The blue color represents the location with the least amount of energy. The 3CLP-control and 3CLP-Sorafenib complexes have two distinct global energy minima basins (in blue) (Figures 8A,B), but the 3CLP-YM201636 complexes have three global energy minima states (Figure 8C). More blue spots imply changes in the protein structure followed by a thermodynamically more favorable state, whereas increased blue areas suggest more stability.

Conclusion

Given the clinical relevance of NV, the study presented here focuses on a high-throughput virtual screen of an antiviral natural compound library against the NV 3CLP using computational technique. The lead compounds sorafenib, LDC4297, and YM201636 met ADMET criteria and interacted with key 3CLP residues. Based on BE, sorafenib and YM201636 were chosen for

MD simulation studies, and these compounds displayed stability with the 3CLP. Therefore, these compounds have the potential to be useful in the development of 3CLP inhibitors, and further testing in wet laboratory is warranted.

Data availability statement

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

Author contributions

Conceptualization, AK, AR, and DS; Formal analysis, SH, AFN, AHH, AMA, and DS; Methodology, SH, AMA, SA, and MAA; Writing—original draft, SH, AK, and AR; Writing—review & editing, SAA, MMA, NH, IA-D, and JA. All authors discussed, edited, and approved the final version.

Funding

The study was funded by the S&T projects of Guangzhou City (project No: 202102020731), Guangzhou City University Alliance Fundamental Research Fund (Fund No: 20210210486), the S&T projects of China's Ministry

(QN2021163001L), and National Natural Science Foundation of China (32150410363).

Conflict of interest

SH is employed by the company Suzhou Ultra-Water-Cleaning Tech, Pty, 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/fchem.2022.1034911/full#supplementary-material>

References

- Aier, I., Varadwaj, P. K., and Raj, U. (2016). Structural insights into conformational stability of both wild-type and mutant EZH2 receptor. *Sci. Rep.* 6, 34984. doi:10.1038/srep34984
- Amarnath Jonniya, N., Sk, M. F., and Kar, P. (2021). Elucidating specificity of an allosteric inhibitor WNK476 among With-No-Lysine kinase isoforms using molecular dynamic simulations. *Chem. Biol. Drug Des.* 98, 405–420. doi:10.1111/cbdd.13863
- Antonelli, G., and Turriziani, O. (2012). Antiviral therapy: Old and current issues. *Int. J. Antimicrob. Agents* 40, 95–102. doi:10.1016/j.ijantimicag.2012.04.005
- Chang, K. O., Kim, Y., Lovell, S., Rathnayake, A. D., and Groutas, W. C. (2019). Antiviral drug discovery: Norovirus proteases and development of inhibitors. *Viruses* 11, 197. doi:10.3390/v11020197
- Daina, A., Michielin, O., and Zoete, V. (2017). SwissADME: A free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. *Sci. Rep.* 7, 42717–42813. doi:10.1038/srep42717
- Dallakyan, S., and Olson, A. J. (2015). "Small-molecule library screening by docking with PyRx," in *Chemical biology* (New York, NY: Humana Press), 243–250.
- Damalanka, V. C., Kim, Y., Alliston, K. R., Weerawarna, P. M., Galasiti Kankanamalage, A. C., Lushington, G. H., et al. (2016). Oxadiazole-based cell permeable macrocyclic transition state inhibitors of norovirus 3CL protease. *J. Med. Chem.* 59, 1899–1913. doi:10.1021/acs.jmedchem.5b01464
- Deng, L., Muhaxhiri, Z., Estes, M. K., Palzkill, T., Prasad, B. V., and Song, Y. (2013). Synthesis, activity and structure-activity relationship of noroviral protease inhibitors. *Medchemcomm* 4, 1354. doi:10.1039/c3md00219e
- Galasiti Kankanamalage, A. C., Kim, Y., Rathnayake, A. D., Damalanka, V. C., Weerawarna, P. M., Doyle, S. T., et al. (2017). Structure-based exploration and exploitation of the S4 subsite of norovirus 3CL protease in the design of potent and permeable inhibitors. *Eur. J. Med. Chem.* 126, 502–516. doi:10.1016/j.ejmech.2016.11.027
- Galasiti Kankanamalage, A. C., Kim, Y., Weerawarna, P. M., Uy, R. A., Damalanka, V. C., Mandadapu, S. R., et al. (2015). Structure-guided design and optimization of dipeptidyl inhibitors of norovirus 3CL protease. Structure-activity relationships and biochemical, X-ray crystallographic, cell-based, and *in vivo* studies. *J. Med. Chem.* 58, 3144–3155. doi:10.1021/jm5019934
- Hall, A. J. (2012). Noroviruses: The perfect human pathogens? *J. Infect. Dis.* 205, 1622–1624. doi:10.1093/infdis/jis251
- Hubbard, R. E., and Kamran Haider, M. (2001). "Hydrogen bonds in proteins: Role and strength," in *eLS* (John Wiley & Sons).
- Hughes, J. P., Rees, S., Kalindjian, S. B., and Philpott, K. L. (2011). Principles of early drug discovery. *Br. J. Pharmacol.* 162, 1239–1249. doi:10.1111/j.1476-5381.2010.01127.x
- Humphrey, W., Dalke, A., and Schulten, K. (1996). VMD: Visual molecular dynamics. *J. Mol. Graph.* 14, 33–38. doi:10.1016/0263-7855(96)00018-5
- Hussey, R. J., Coates, L., Gill, R. S., Erskine, P. T., Coker, S. F., Mitchell, E., et al. (2011). A structural study of norovirus 3C protease specificity: Binding of a designed active site-directed peptide inhibitor. *Biochemistry* 50, 240–249. doi:10.1021/bi1008497
- Kaufman, S. S., Green, K. Y., and Korba, B. E. (2014). Treatment of norovirus infections: Moving antivirals from the bench to the bedside. *Antivir. Res.* 105, 80–91. doi:10.1016/j.antiviral.2014.02.012
- Koo, H. L., Ajami, N., Atmar, R. L., and Dupont, H. L. (2010). Noroviruses: The leading cause of gastroenteritis worldwide. *Discov. Med.* 10 (50), 61–70.
- López-López, E., Naveja, J. J., and Medina-Franco, J. L. (2019). DataWarrior: An evaluation of the open-source drug discovery tool. *Expert Opin. Drug Discov.* 14, 335–341. doi:10.1080/17460441.2019.1581170
- Lopman, B. A., Steele, D., Kirkwood, C. D., and Parashar, U. D. (2016). The vast and varied global burden of norovirus: Prospects for prevention and control. *PLoS Med.* 13, e1001999. doi:10.1371/journal.pmed.1001999

- Meng, X. Y., Zhang, H. X., Mezei, M., and Cui, M. (2011). Molecular docking: A powerful approach for structure-based drug discovery. *Curr. Comput. Aided Drug Des.* 7, 146–157. doi:10.2174/157340911795677602
- Muhaxhiri, Z., Deng, L., Shanker, S., Sankaran, B., Estes, M. K., Palzkill, T., et al. (2013). Structural basis of substrate specificity and protease inhibition in Norwalk virus. *J. Virol.* 87, 4281–4292. doi:10.1128/jvi.02869-12
- Mullard, A. (2016). Biotech R&D spend jumps by more than 15%. *Nat. Rev. Drug Discov.* 15, 447. doi:10.1038/nrd.2016.135
- Netzler, N. E., Enosi Tuipulotu, D., and White, P. A. (2019). Norovirus antivirals: Where are we now? *Med. Res. Rev.* 39, 860–886. doi:10.1002/med.21545
- Pol-Fachin, L., Fernandes, C. L., and Verli, H. (2009). GROMOS96 43a1 performance on the characterization of glycoprotein conformational ensembles through molecular dynamics simulations. *Carbohydr. Res.* 344, 491–500. doi:10.1016/j.carres.2008.12.025
- Rocha-Pereira, J., Neyts, J., and Jochmans, D. (2014). Norovirus: Targets and tools in antiviral drug discovery. *Biochem. Pharmacol.* 91, 1–11. doi:10.1016/j.bcp.2014.05.021
- Schaduangrat, N., Lampa, S., Simeon, S., Gleeson, M. P., Spjuth, O., and Nantasenamat, C. (2020). Towards reproducible computational drug discovery. *J. Cheminform.* 12, 9. doi:10.1186/s13321-020-0408-x
- Schuttelkopf, A. W., and Van Aalten, D. M. (2004). PRODRG: A tool for high-throughput crystallography of protein-ligand complexes. *Acta Crystallogr. D. Biol. Crystallogr.* 60, 1355–1363. doi:10.1107/s0907444904011679
- Thorne, L. G., and Goodfellow, I. G. (2014). Norovirus gene expression and replication. *J. Gen. Virol.* 95, 278–291. doi:10.1099/vir.0.059634-0
- 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
- Weerawarna, P. M., Kim, Y., Galasiti Kankanamalage, A. C., Damalanka, V. C., Lushington, G. H., Alliston, K. R., et al. (2016). Structure-based design and synthesis of triazole-based macrocyclic inhibitors of norovirus protease: Structural, biochemical, spectroscopic, and antiviral studies. *Eur. J. Med. Chem.* 119, 300–318. doi:10.1016/j.ejmech.2016.04.013



OPEN ACCESS

EDITED BY

Khurshid Ahmad,
Yeungnam University, South Korea

REVIEWED BY

Qazi Mohammad Sajid Jamal,
Qassim University, Saudi Arabia
Mohd Athar,
University of Cagliari, Italy

*CORRESPONDENCE

Foysal Ahammad,
foysalgebt@gmail.com
Farhan Mohammad,
mohammadfarhan@hbku.edu.qa
Jesus Simal-Gandara,
jsimal@uvigo.es

SPECIALTY SECTION

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

RECEIVED 08 July 2022

ACCEPTED 07 September 2022

PUBLISHED 04 October 2022

CITATION

Dain Md Opo FA, Alsaiani AA,
Rahman Molla MH, Ahmed Sumon MA,
Yaghmour KA, Ahammad F,
Mohammad F and Simal-Gandara J
(2022), Identification of novel natural
drug candidates against BRAF mutated
carcinoma; An integrative in-silico
structure-based pharmacophore
modeling and virtual screening process.
Front. Chem. 10:986376.
doi: 10.3389/fchem.2022.986376

COPYRIGHT

© 2022 Dain Md Opo, Alsaiani, Rahman
Molla, Ahmed Sumon, Yaghmour,
Ahammad, Mohammad and Simal-
Gandara. This is an open-access article
distributed under the terms of the
[Creative Commons Attribution License
\(CC BY\)](#). The use, distribution or
reproduction in other forums is
permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original
publication in this journal is cited, in
accordance with accepted academic
practice. No use, distribution or
reproduction is permitted which does
not comply with these terms.

Identification of novel natural drug candidates against BRAF mutated carcinoma; An integrative in-silico structure-based pharmacophore modeling and virtual screening process

F. A. Dain Md Opo^{1,2}, Ahad Amer Alsaiani³,
Mohammad Habibur Rahman Molla¹, Md Afsar Ahmed Sumon⁴,
Khaled A. Yaghmour⁵, Foysal Ahammad^{1,6*},
Farhan Mohammad^{6*} and Jesus Simal-Gandara^{7*}

¹Department of Biological Science, Faculty of Sciences, King Abdulaziz University (KAU), Jeddah, Saudi Arabia, ²Embryonic Stem Cell Research Unit, King Fahd Medical Research Center (KFMRC), KAU, Jeddah, Saudi Arabia, ³Clinical Laboratories, Science Department, College of Applied Medical Science, Taif University, Taif, Saudi Arabia, ⁴Department of Marine Biology, Faculty of Marine Sciences, King Abdulaziz University, Jeddah, Saudi Arabia, ⁵Family Medicine Department, Faculty of Medicine, King Abdulaziz University, Jeddah, Saudi Arabia, ⁶Division of Biological and Biomedical Sciences (BBS), College of Health and Life Sciences (CHLS), Hamad Bin Khalifa University (HBKU), Doha, Qatar, ⁷Nutrition and Bromatology Group, Department of Analytical Chemistry and Food Science, Faculty of Food Science and Technology, University of Vigo, Ourense, Spain

The BRAF gene is responsible for transferring signals from outside of the cell to inside of the nucleus by converting a protein namely B-Raf through the RAS/MAPK pathway. This pathway contribute to cell division, proliferation, migration, and apoptotic cell death of human and animal. Mutation in this gene may cause the development of several cancers, including lung, skin, colon, and neuroblastoma. Currently, a few available drugs are being used that has developed by targeting the BRAF mutated protein, and due to the toxic side effects, patients suffer a lot during their treatment. Therefore this study aimed to identify potentially lead compounds that can target and block the expression of BRAF and subsequently inhibit the cancer. The hits were generated through the pharmacophore model-based virtual screening, molecular docking, pharmacophore model validation, ADME (absorption, distribution, metabolism, and excretion) analysis molecular dynamics (MD) simulation to find more suitable candidate against the overexpress BRAF gene. The pharmacophore

Abbreviations: BRAF, v-raf murine sarcoma viral oncogene homolog B1; AUC, Area under curve; CAAD, Computer Aided Drug Discovery; CRC, Colorectal cancer; EGFR, Epidermal Growth Factor Receptor; IC₅₀, Inhibitory concentration; IDB, Intermediate Debug File; KRAS, Kirsten rat sarcoma viral oncogene homolog; LC₅₀, Lethal Concentration 50; LD₅₀, Lethal Dose 50; MOL; MM-GBSA, Molecular mechanics with generalised Born and surface area solvation; MAPK, mitogen-activated protein kinase; SDF, Structure data file.

based screening initially identified 14 k possible hits from online database which were further screened by ligand scout advance software to get hit compound. Based on molecular docking score of ZINC70454679 (-10.6 kcal/mol), ZINC253500968 (-9.4 kcal/mol), ZINC106887736 (-8.6 kcal/mol), and ZINC107434492 (-8.1 kcal/mol), pharmacophore feature and toxicity evaluation, we selected four possible lead compounds. The dynamic simulation with Schrodinger Maestro software was used to determine the stability of the potential lead candidates with target protein (PDB ID: 5VAM). The results showed that the newly obtained four compounds were more stable than the control ligand (Pub Chem ID: 90408826). The current results showed that the ZINC70454679, ZINC253500968, ZINC106887736, and ZINC107434492 compounds may be able to work against several cancers through targeting the BRAF overexpressed gene. To develop a novel drug candidate, however the evaluation of the web lab based experimental work are necessary to evaluate the efficiency of the each compound against the BRAF target gene.

KEYWORDS

pharmacophore modeling, virtual screening, molecular docking, molecular dynamics simulation, BRAF, B-Raf

1 Introduction

BRAF also known as the proto oncogene highly responsible for the signal transduction inside the cells for growing the cell number through maintaining the signaling pathway known as MAP/ERK pathway (McCubrey et al., 2007) (Guo et al., 2020). BRAF participates in cell division by activating phosphorylation by binding to Ras-GTP and eventually producing ADP, phosphorylated protein. (Cope et al., 2018). EGF (Epidermal Growth Factor) bind to the cytoplasmic serine and activate the EGFR receptor. In the presence of the two adaptor protein (SOS and GRB2) EGFR knock KRAS to release the GDP. This KRAS allow to bind cystolic BRAF and activate the MEK kinase. Finally through simulating transcription factors contribute in cellular proliferation, differentiation, apoptosis and cell survival (Fanelli et al., 2020).

Genetic mutations of BRAF gene are more common and responsible for developing cardiovascular defects, retardation of mental growth, and also lead to the development of several cancers (A. Richards and Garg, 2010). Mutations in this gene are responsible for more than 80% of skin cancers known as melanomas; others are lung cancer, colon cancer, and also neuroblastoma (Hussain et al., 2015). BRAF mutation in position V600E, which carried about 80% of alteration and V600 K about 10–20%, were responsible for development of cancer in young people, mainly the tumors appear in the parts of body that were not commonly exposed to sunlight (Menzies et al., 2012) (Ascierto et al., 2012) (Luu and Price, 2019). Smokers as well as non-smokers can be radially affected by the cancer, although the cancer in smokers can develop more aggressively and quickly. It has been reported that the BRAF mutation developed in lung adenocarcinoma in people who were never addicted to smoking. The treatment of the lung adenomas is difficult in the case of this

mutation as it has been observed as a resistance mutation (Cardarella et al., 2013) (Nguyen-Ngoc et al., 2015). The incidence of colon cancer due to such mutations is higher in females, those over the age of 50, and those with no history of genetically colon cancer. The mutation in chromosome seven from valine to glutamine at position 600 was developed for right-sided colon cancer (Barras, 2015) (Grassi et al., 2021). Both BRAF and KRAF mutations were linked to the development of CRC in two ways: one activated the expression of the KRAS/mTOR/AKT and the other caused instability in cell cycle regulation. (Morkel et al., 2015) (Merz et al., 2021).

The combination target therapy with encorafenib, binimetinib and cetuximab are in the clinical trial phases and showed to more effective rather than the using two drugs (irinotecan + cetuximab) (Roviello et al., 2020) (Geel and Iersel, 2022). One of the aggressive tumors, thyroid cancer, was also developed by the mutation in the BRAF gene. Most of the BRAF mutations occur in the position of the T1799A and others, including the mutation in the K601E in thyroid cancer (Rowe et al., 2007) (Tran et al., 2020). Through examining the total 75 samples, among whom 17 patients developed KRAS mutation and 26 were examined for BRAF mutation, it has been identified that BRAF mutation may lead to developing ovarian cancer in females (Turashvili et al., 2018). Two common mutations were identified, including BRAF in codon 599, and at codon 12 and 13, the KRAS mutation. This mutation is less common (less than 3%) in carcinomas of the stomach, esophagus, and glioma (Ayatollahi et al., 2018).

The BRAF positive mutated patients were under chemotherapy or immunotherapy besides using the targeted therapy. Combination with two drugs (combine therapy) and three drugs (triple therapy) are common in the treatment of BRAF mutations and are also in clinical trial phase (Eroglu and

Ribas, 2016) (Patel et al., 2020). Several drugs, such as vemurafenib, dabrafenib, and encorafenib, currently available to treat BRAF mutated cancer based on targeting the mutations V600E and V600K, two types of possible mutations in several cancers. Drugs known as checkpoint inhibitors are being used in triple therapy. Use of these drugs for target therapy has been shown to produce several side effects, including urine in blood, fever, joint pain, skin ulceration, and so on (Proietti et al., 2020) (Tanda et al., 2020). The number of other drugs that can be used during treatment are limited due to drug-drug interactions. Due to their long-time use, most of the BRAF/KRAS mutated tumors are showing resistance to these treatments. A 60 year old female patient was identified the BRAF mutation and treatment with the vemurafenib showed less efficient. A new mutation was also observed after the 11 months of treatment and through the multiple organ failure patient died after 12 months (Wang et al., 2022).

So the development of new drugs with less side effects and also possible to overcome resistance are the first choice for researchers, caused by the BRAF mutations. In our study, we focused on computational drug design to develop more efficient compounds that can be used as drugs through further experiments and validation results. Currently computer based drug discovery are the popular tool for designing a new compounds against the specific target area. For rapid lead compounds identification this pathway follow the pharmacophore modeling, molecular docking, virtual screening, ADMET (absorption, distribution, metabolism, excretion, and toxicity) analysis, molecular dynamics (MD) simulation, and MM-GBSA method (Opo et al., 2021) (Bouback et al., 2021). Molecular docking result usually express the binding possibility between the ligand and receptor, which is the important part for drug efficacy. ADMET analysis by the online database and tool showed the possibility of toxic effect of a lead compound inside the body are more easier to determine rather the conventional method, whereas the toxicity development from the blood sample, stool or urine might create a risk for drug failure (Valasani et al., 2014). As the CADD approach are more convenient, cheap in comparison to the conventional drug design this study aimed to discover lead compound against the BRAF mutations. The identified potentially lead compounds through the in-silico drug design might be able to reduce the BRAF mutated carcinoma.

2 Materials and methods

2.1 Pharmacophore modelling

To interact with natural molecules, a ligand with a protein structure was retrieved (PDB ID: 5VAM), as well as a three-dimensional structure (Nishiguchi et al., 2017). For identification of the protein structure screening has been performed based on

the organism source, X-Ray diffraction method, and refinement resolution also the release date. The attached ligand IC₅₀ was already established by several experimental analysis and the toxicity of the attached ligand was low with higher LD₅₀ value (2000 mg/kg). PubChem database was used to get the chemical ID of the attached ligand 92J to the target protein (Pub Chem ID: 90408826) (S. Kim et al., 2021). A structure-based pharmacophore model was created using Ligand Scout 4.4.8 advanced software. This powerful software created the interaction between inhibitors and crucial amino acids at the active sites in our target protein. Different pharmacophore properties, such as hydrogen bond donors, charge transfer, hydrophilic and hydrophobic areas, and hydrogen bond acceptors, were used to interpret ligand-receptor interactions. Other parameters such as the quantity of aromatic rings, hybridization state, binding pattern, and receptor molecule distance have been discovered using stepwise algorithms. Using ligand scout, we provided hydrophilic characteristics to the protein to improve the measurement of protein binding. The number of active sites was also measured by using the CASTp (sts.bioe.uic.edu/castp/) software for further analysis (Supplementary Figure S1).

2.2 Pharmacophore model verification

A set of active compounds (Supplementary Table S1) were identified from the ChEMBL database (<https://www.ebi.ac.uk/chembl/>) with an active IC₅₀ value (Gaulton et al., 2017). The DUD-E decoy set (obtained from the DUD-E decoys database) was used to evaluate the known active compounds in order to more accurately distinguish between the active and inactive compounds (Mysinger et al., 2012). All the active known compounds and the extracted decoy set were transferred to the ligand scout 4.4.8 advance software to make an "idb" file. Models were generated from the protein-ligand complex through the screening of all active compounds in correspondence of the 4,094 decoy sets. The quality of our selected structure based model was assessed by the AUC value, GH score, and enrichment factor (Wolber and Langer, 2005).

2.3 Pharmacophore based virtual screening

A freely assessable database was used to identify the potential lead compounds, including the ZINC Pharmer (<http://zincpharmer.csb.pitt.edu/pharmer.html>) and ambinter data base (<https://www.ambinter.com/#search>) (Koes & Camacho, 2012). Both databases were the available source for the determination of the physical and chemical properties such as 2D and 3D structure determination, the boiling point, the melting point, molecular weight, and

biological activity of the compounds (Opo et al., 2021). The screening was performed in the Zinc Purchasable database and natural database based on the pharmacophore features generated by the ligand Scout software and previously saved as 'pml' file. The chosen compounds had the most similar pharmacophore features to our query compounds. The selected compounds were then subjected to a series of tests, including molecular weight, hydrogen bond donor, hydrogen bond acceptor, and LogP value, all of which were based on Lipinski's rule of five. All the selected compounds were preserved with their Canonical SMILES ID obtained from PubChem database (<https://pubchem.ncbi.nlm.nih.gov/>) and proceeded to further study (S. Kim et al., 2021). The database generated from the Zinc and ambinter was validated based on structure based pharmacophore features. A freely accessible ZINC database and also an ambinter database were used to find the most similar compounds. We identified our specific protein structure and a previously prepared library with 14 k compounds was inserted into the Ligand Scout 4.4.8 advance software. The library was screened based on the created pharmacophore features, with the addition of the 1 h bond donor feature. Fitted hit compounds were further subjected to validation based on the relative pharmacophore fit score.

2.4 Protein and ligand preparation

The selected protein structures were prepared for docking purposes. The downloaded 'sdf' file was opened by the discovery studio and removed the water molecule and also the hetatm. The addition of any necessary bond and deletion of the water molecules was not part of the structural refinement process. The desired protein structure (PDB ID: 5VAM) was obtained and analyzed for the R value-free (0.223), resolution (2.0Å), and observed R-value (0.194). We discovered that a few bonds in the currently selected protein were missing; therefore, we used BIOVA Discovery Studio Tool 16.1.0 to construct a new bond by using the force field (CHARMm). Generally, this force field contribute distinctive effects including electronegativity, stereo electrical effects, polarization, bond stretching and angle bending, on the other hand, are characterized by simple harmonic motion (Hwang et al., 2020).

2.5 Grid generation and active site identification

The active site of our protein has been identified and analyzed by the UniProtKB and PrankWeb (<https://prankweb.cz/>) (Gray et al., 2021) (Jendele et al., 2019). The number of active pockets was also determined using CASTp (CASTp 3.0: Computed Atlas of Surface Topography of Proteins (uic.edu)) (Supplementary Figure S1, Supplementary Table S2) (Tian et al., 2018). The presence of

hydrogen bonds, lipophilic or hydrophilic interactions, and ionizable charges all affect the protein and ligand's binding affinity. The PyRx software was used to generate the grid by selecting the active sites of the proteins (Dallakyan and Olson, 2015). The server-generated binding sites were utilized to create a receptor grid box in center with the following coordinates: X = -29.1124, Y = 42.6919, and Z = 8.227 and with the exhaustiveness of 8.

2.6 Binding affinity determination by docking

All the selected hit compounds "sdf" files were downloaded from the PubChem database. The compounds and also previously prepared the protein 3D structure were transferred to the PyRx software and docking was conducted by AutoDock Vina. A prominent tool being used in drug design for selecting drugs against various animal diseases and identifying new therapeutic candidates (Dallakyan and Olson, 2015). The compounds were then submitted to the BIOVA Discovery Studio Visualizer Tool 16.1.0 for analysis based on the binding affinity and RMSD value. The validation of the docking has been performed several times with the above mentioned grid generation for the all selected ligands.

2.7 ADME profile evaluation

The metabolism and pharmacokinetic properties of a drug are important parameters in determining drug efficacy (Benedetti et al., 2009). Approximately fifty percent of drug candidates fail due to their lack of efficacy and toxicity at the time of the drug development, so the ADME profile analysis is crucial part before drug development. (Opo et al., 2021). Usually elimination of drugs from the body occur through urine and feces, several physiochemical features such as hydrophobicity, lipophilicity, gastrointestinal environment, and blood brain barrier have a direct impact on the ADME profile before elimination of drugs. The bioavailability of a medicine also are being affected by its sex, age, disease state, lipophilicity, hydrophobicity, microbiota, body enzymes, and administration method (Stillhart et al., 2020). For evaluating the ADME profile, such as solubility, GIT absorption, and bioavailability in the case of the ligand, we used the freely available Swiss-ADME server (<http://www.swissadme.ch/>). Swiss ADME sever is a popular online database for determination of the compound physicochemical and pharmacokinetic properties (Daina et al., 2017).

2.8 Evaluation of toxicity

In-silico approaches for analyzing the safety profile of the required chemicals have been developed by computational

research (Bouback et al., 2021). Otherwise, these substances could have a negative impact on discovery of new compounds and lead to the failure of drug discovery in the middle of research. The toxicity profile such as hepatic failure, carcinogenicity, immunological response, membrane potential route was easily quantified and qualitatively determined to see the possibility of toxicity before going to the lab based experiment. The computer aided toxicity measurement tools (Toxicity Estimation Software Tool, TEST version 4.2.1) usually commonly used to estimate a chemical's harmful effect based on its molecular structure. In our study we measured the fathead minnow LC_{50} (96 h), 48-h daphnia magna LC_{50} , developmental toxicity, oral rat LD_{50} , bioaccumulation factor, and water solubility (at 25°C). Freely access database ProTox-II server ([https://tox-new.charite.de/prottox II/](https://tox-new.charite.de/prottox-II/)) was used to detect hepatotoxicity, carcinogenicity, mutagenicity, immunogenicity, and numerous toxicological pathways for selected antagonist (Banerjee et al., 2018).

2.9 Protein and ligand preparation for simulation

The simulation of the protein ligand complex tells us the binding pattern and characteristics between the atoms and amino acid residues (Opo et al., 2021). The 100ns dynamic simulation was used to validate our ligand binding to the protein, which had been obtained through the docking studies. The stability of the complex must be assessed to see the possible effect inside the body, as well as the projection of every atom bonding behavior both of ligand and protein molecules during a given time period. Using the Linux command, we conducted our dynamic simulation through utilizing software Schrödinger Release 2020-3 (Academic version) (Bowers et al., 2006). The water model was used to solve the ligand and protein interaction, as well as provide the orthorhombic box shape boundary. By combining the Na⁺ and Cl⁻ with a 0.15 M salt concentration, the complicated atom buffer box calculation approach was applied. The simulation was run with an ambient temperature of 300 K and a pressure of 1.01325 bar, with a record interval time of 50 ps. The OPLS-2005 force field was used to execute the simulation (Shivakumar et al., 2010).

2.9.1 Trajectory file analysis from ligand protein interaction

The MD simulation's quality was confirmed, and the simulation scenario was investigated utilizing Schrödinger package's simulation interaction diagram (SID). The Simulation Interaction Diagram (SID) of the Desmond module was used to evaluate all of the simulation's data sets (Bowers et al., 2006). Depending on the RMSD, RMSF value, and ligand-protein complex, the simulation trajectory file offered

information about the integrity of the protein-ligand interaction complex. The ligand torsion profile has been evaluated to find the rotatable bond were present in the ligand during the simulation trajectories (Jin et al., 2020). Radius of gyration has been used to evaluate the structural compression changes and intra molecular hydrogen bond analysis was performed to identify the presence of internal hydrogen bonds within a ligand molecule.

2.9.2 MM-GBSA analysis

A common technique for determining the free binding energy of ligands is the calculation of molecular mechanics with generalized born surface area (MM/GBSA). Typically this analysis based on the receptor ligand complex that are more precise unlike many docking studies grading algorithms and computationally fewer taxing other molecular free energy techniques (Genheden and Ryde, 2015). We estimated the binding free energy of four potentially leads compound and control ligand using the Schrödinger Prime MM/GBSA package (released 2020-3) (Bouback et al., 2021).

3 Results

3.1 Protein analysis based on pharamacophore features

The 3D structure of a protein is important to facilitate the possible drug interaction with the biological activity and is necessary to predict the possibility of efficacy prior to synthesis. The protein was bound to a single ligand, and the structure was determined by x-ray diffraction with a resolution of 2.10, R value free (0.223), R value observed (0.194), and R value work (0.192). The IC_{50} value was calculated from the several assays and was minimum 0.4nM with maximum 1.8 nM and the toxicity of the attached ligand was low with higher LD_{50} value (2000 mg/kg). For determining an active series of inhibitors, it is important to look for enough interaction to attain better biological activity than the current one. The important chemical characteristics were generated using Ligand Scout 4.4.8 advanced critical molecular design software, which was based on a pharmacophore model. Total seven chemical features were observed, including three hydrophobic bonds, three H-bond acceptors, and one H-bond donor without the inclusion of exclusion volume (Figure 1).

Analysis of the interaction with the protein ligand contact indicated the number of hydrophobic interactions were most predominant type of bond among the twelve amino acids. The red arrows represented the interaction of the H-bond acceptors ASP594, HOH917, HOH972, and CYS532. One H-bond donor bond was formed with the GLU501 position of the amino acids (Supplementary Figure S1B).

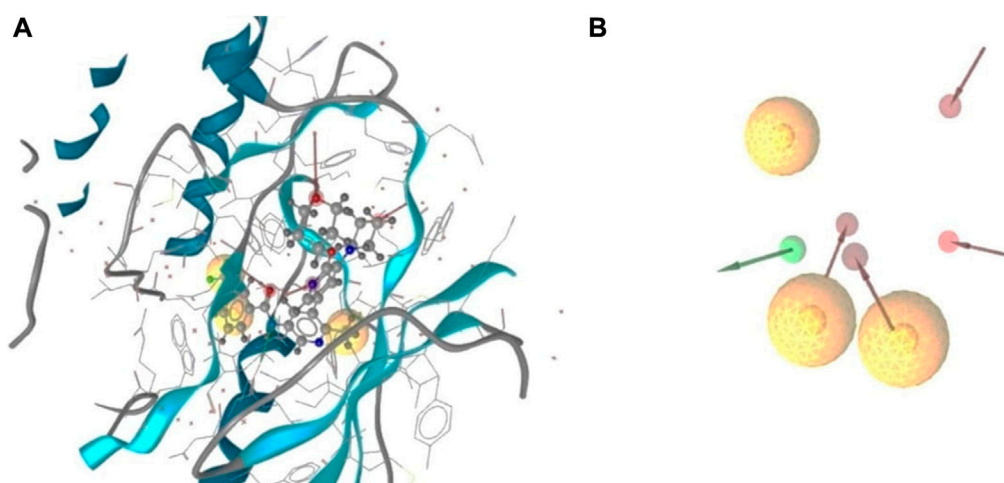


FIGURE 1

3D structure based on pharmacophore model of BRAF protein ligand complex. Arrangement of the pharmacophore features along with the selected protein structure (A), and the observed pharmacophores in the absence of the protein chain (B). Three hydrophobic interactions represented by yellow spheres, red arrows demonstrated H-bond acceptor, and one green arrows depicted the presence of the H-bond donor.

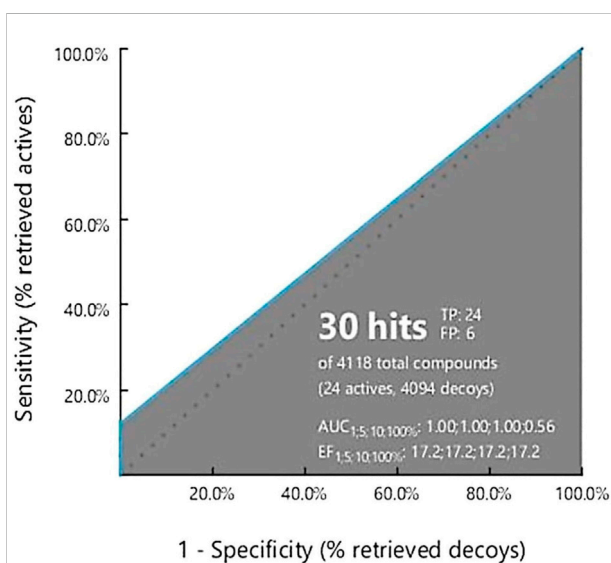


FIGURE 2

Ligand Scout 4.3 Advance software was used to create a receiver operating characteristic (ROC) curve. The total number of active decoy sets was determined using the dude decoy database's predefined decoy sets.

3.2 Pharmacophore model validation

Validation is necessary to evaluate the model quality and to obtain an accurate pharmacophore analysis. Validation of the derived pharmacophore model was performed using 24 active known BRAF antagonists in correspondence with 4,094 decoy

molecules obtained from the online decoy database. The quality of the curve is represented by the area under the curve (AUC) and the EF value. The early enrichment factor (EF1%) was 17.2, referring to an excellent curve, and the average AUC value was 0.89, indicating good to excellent results (Figure 2).

3.3 Dataset generation

The development of data sets is critical for distinguishing the lead compounds. The ZINC and ambinter database are the most commercially available database, with 730 million compounds including natural and chemical compounds, as well as 3D structures and current clinical development conditions (Irwin et al., 2020) (Bouback et al., 2021). The Ligand Scout 4.4.8 advance tool was used to produce pharmacophore features and was submitted to the online database for further screening to identify the potentially active lead compounds. We followed the rule of five in the case of screening the database, the RMSD value was less than or equal to one.

3.4 Pharmacophore based virtual screening

Ligand Scout 4.4.8 advanced software was used to create pharmacophore characteristics, which were then transferred to the ZINC database through a 'mol' file. We add one H bond features to get the more suitable drug candidate after screening. The search has been completed based on the following rules: The Rule of Five. A total of 155 hits were retrieved when the RMSD

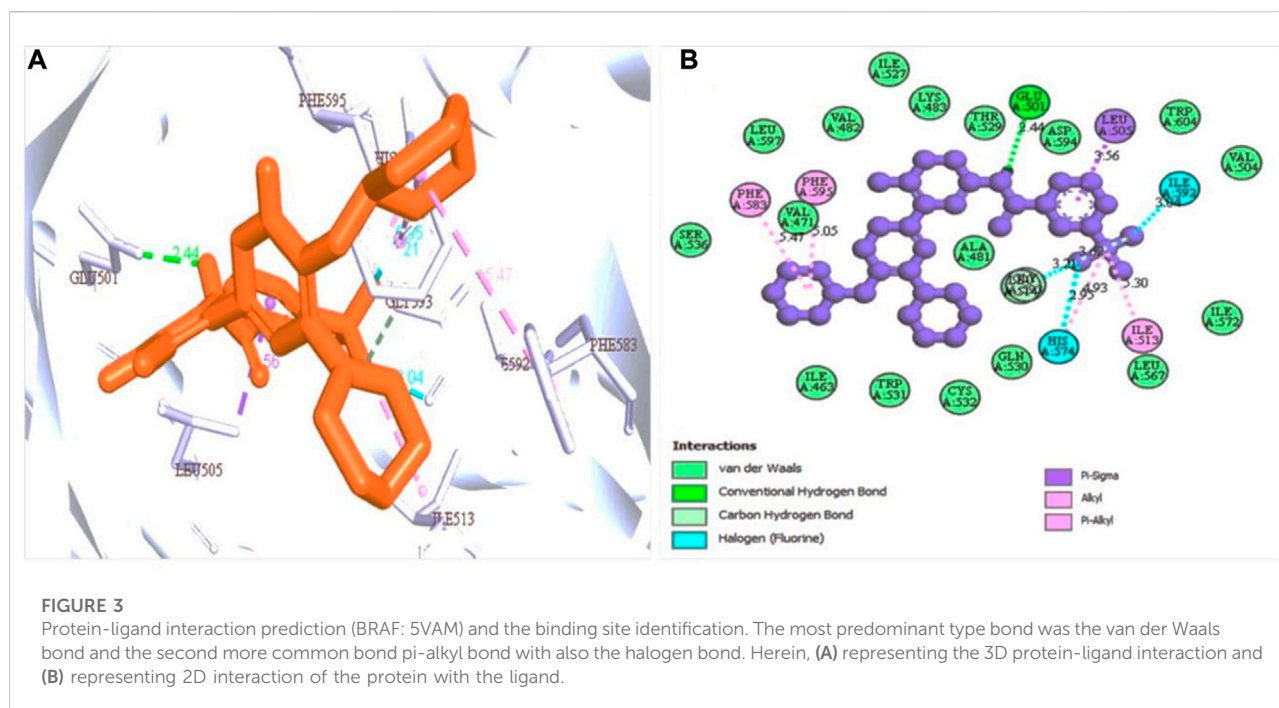
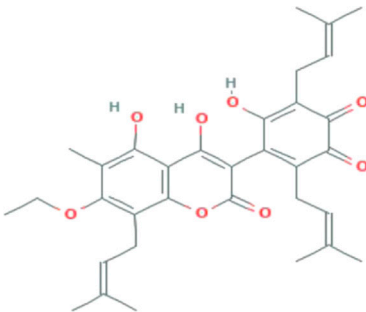
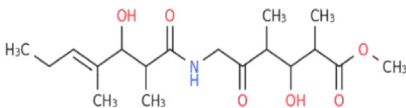


TABLE 1 The binding score generated from the docking with the protein (PDB ID: 5VAM) along together with the compound structure, molecular formula. The compound were selected based on the binding energy and also by evaluating toxicity.

ZINC ID	Compound structure	Binding affinity (kcal/mol)	Molecular formula
ZINC70454679		-10.6	C ₃₀ H ₃₄ O ₆
ZINC253500968		-9.4	C ₃₄ H ₄₄ O ₁₉
ZINC106887736		-8.6	C ₃₃ H ₃₈ O ₈

(Continued on following page)

TABLE 1 (Continued) The binding score generated from the docking with the protein (PDB ID: 5VAM) along together with the compound structure, molecular formula. The compound were selected based on the binding energy and also by evaluating toxicity.

ZINC ID	Compound structure	Binding affinity (kcal/mol)	Molecular formula
			
ZINC107434492		-8.1	C ₂₃ H ₃₂ N ₂ O ₃

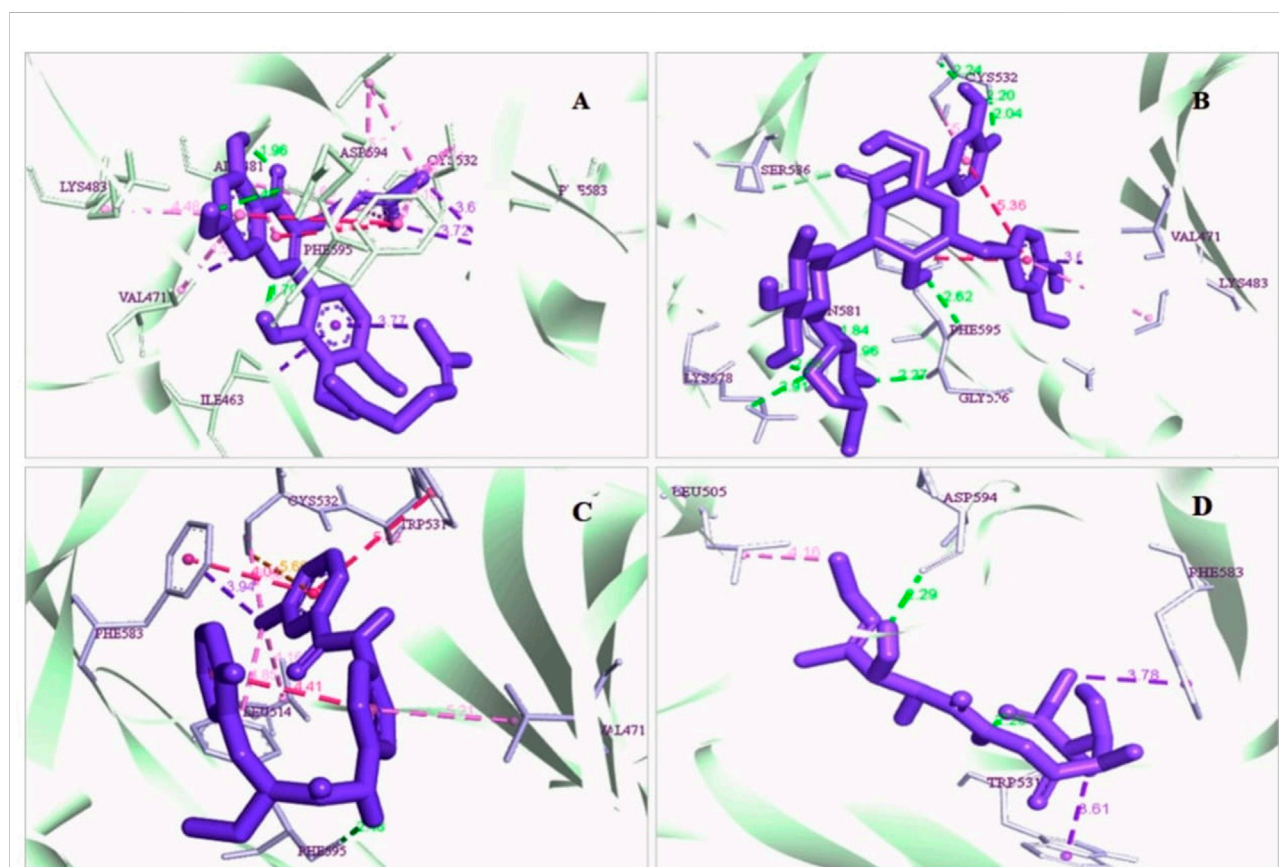


FIGURE 4

3D interaction of the selected antagonist with the protein complex (PDB ID: 5VAM). Our ligands (A) ZINC70454679, (B) ZINC253500968, (C) ZINC106887736, and (D) ZINC107434492 shown the better interaction with the 5VAM protein. Based on the binding affinity score and also the toxicity analysis, four compounds were selected.

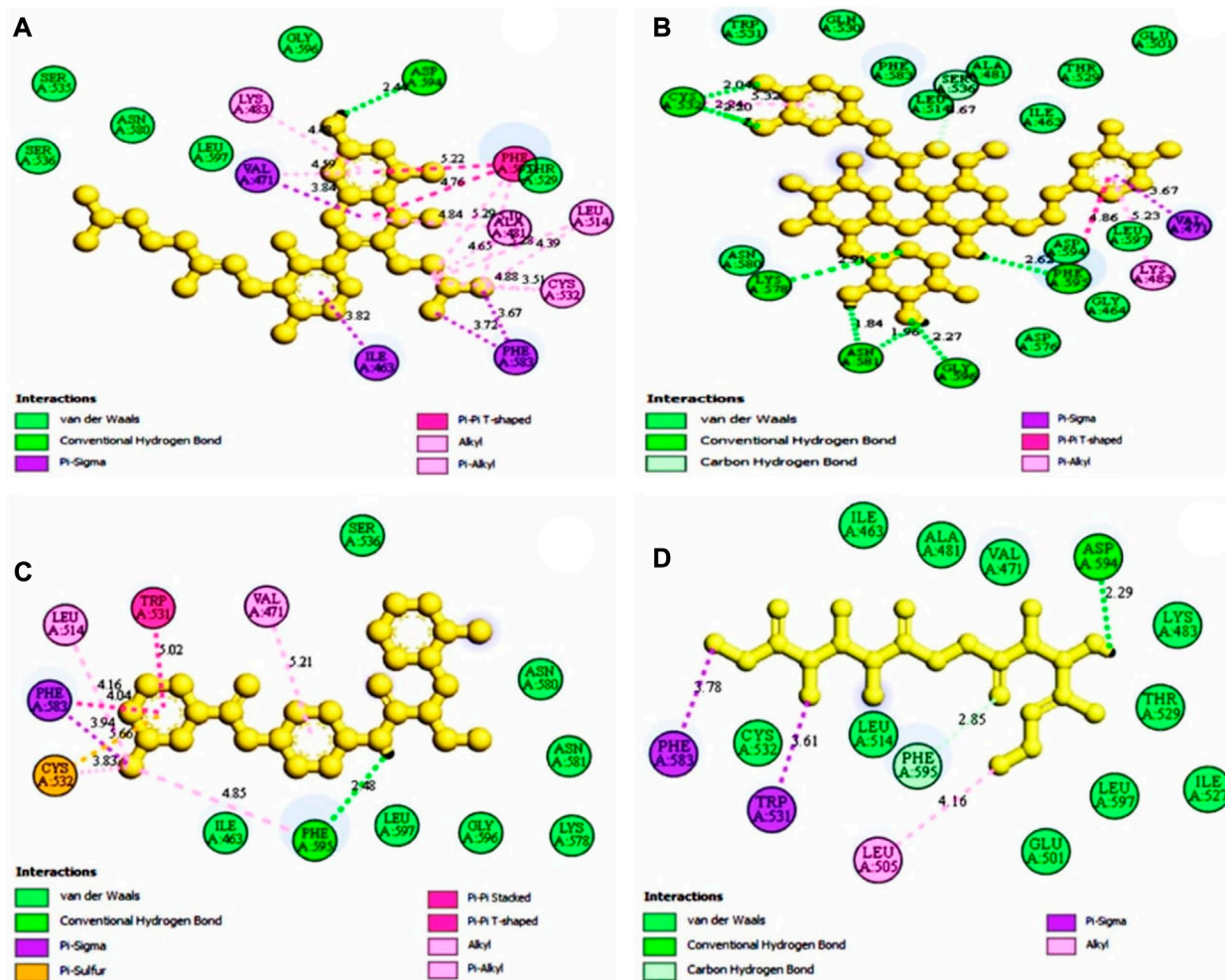


FIGURE 5

The selected antagonist's 2D interaction with the protein complex (PDB ID: 5VAM). Our ligands (A) ZINC70454679, (B) ZINC253500968 (C) ZINC106887736, and (D) ZINC107434492 had the best protein interaction. Four compounds were chosen based on the docking score as well as the toxicity analysis.

value was set to around 1, with a relative pharmacophore fit score of 0.82. The molecules were then docked with the Autodock vina and selected the compounds with the highest binding energy for further investigation and through initially toxicity analysis.

3.5 Binding site identification and ligand-protein interaction

Based on the structure generated by X-ray crystallography, the selected protein has one attached ligand and separate attachment sites for interacting with the target ligand. A total of seven bond formations with the active sites were observed with multiple amino acid residues indicated by the discovery studio

program by analysis of the protein-ligand complex (Figure 3A). The number of active sites has also been determined based on the CASTp software (Supplementary Figure S1A).

3.6 Molecular docking

Docking is a technique used in drug development to assess the binding affinity of a protein and its ligand (Salmaso and Moro, 2018). With the addition of one ligand, the BRAF protein was linked to two chains, and we selected the protein through the removal of the water and hetatm. The protein was prepared by combining the force field (CHARMm) and the receptor grid was generated in the PyRx software based on the previously identified

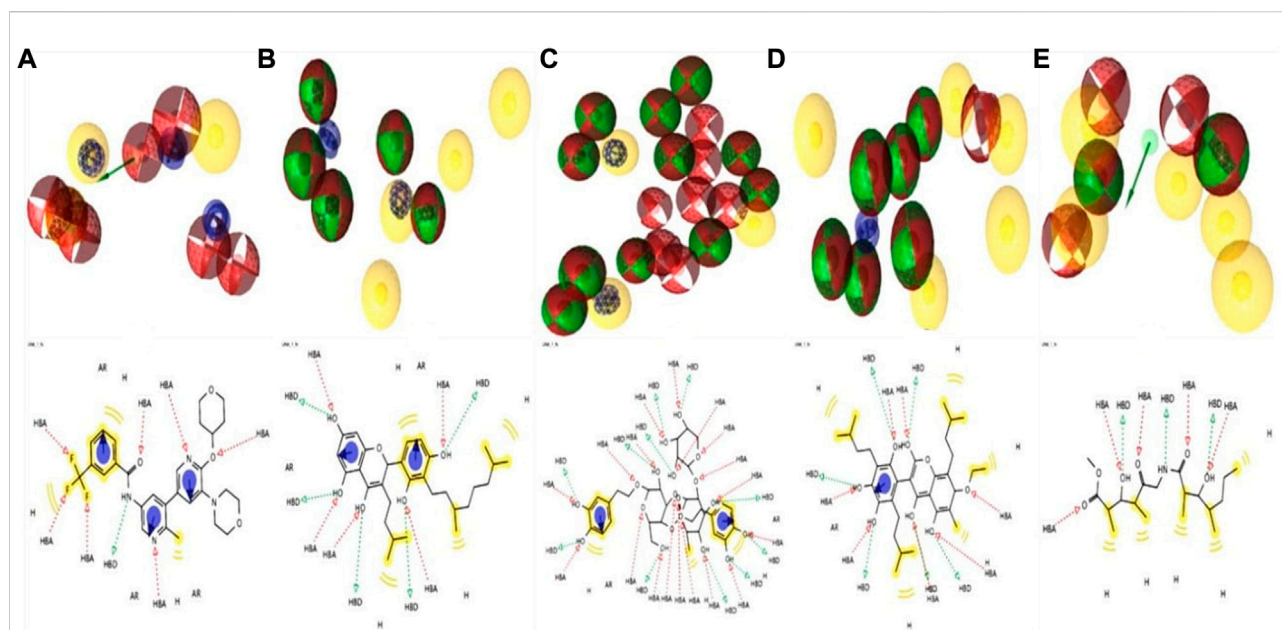


FIGURE 6

Analysis of four compounds with the target BRAF protein using 2D and 3D pharmacophore characteristics. The pharmacophore features of the (A) ligand (Pub Chem ID 90408826) coupled to the protein (PDB ID: 5VAM) were less than the (B) ZINC70454679, (C) ZINC253500968, (D) ZINC106887736, and (E) ZINC107434492 our selected four antagonist.

active sites (Dallakyan and Olson, 2015). The hits identified through obtained compound library screening as well as the selected known antagonist were sent for docking. The binding affinity score for known antagonist were shown in [Supplementary Table S1](#). For generated hits were selected

based on the best binding affinity containing ligands with fewer side effects. These selected potentially lead compounds were considered for further interaction evaluation ([Table 1](#)). The docking for the each compound validated to get the exact binding scenario of our selected four compounds, which were showed

TABLE 2 Different features of the four selected compounds we chose were identified. The table depicts the several physical, chemical, pharmacokinetic, and drug likeness aspects.

Properties	Parameters	ZINC70454679	ZINC253500968	ZINC106887736	ZINC107434492
Physico-chemical properties	MW (g/mol)	490.59	756.70	562.65	357.44
	Heavy atoms	36	53	41	25
	Arom. heavy atoms	16	12	10	0
	Rotatable bonds	8	13	9	12
	H-bond acceptors	5	19	8	6
	H-bond donors	4	11	3	3
	Molar Refractivity	147.01	174.81	162.34	94.98
Lipophilicity	Log $P_{o/w}$	4.52	2.95	4.94	2.58
Water Solubility	Log S (ESOL)	Poor	Soluble	poor	Soluble
Pharmacokinetics	GI absorption	Low	Low	Low	High
	CYP3A4 inhibitor	No	Yes	Yes	No
	BBB permeant	No	No	No	No
Drug likeness	Lipinski, Violation	Yes	Yes, 3	Yes, 1	Yes
	Bioavailability Score	0.55	0.17	0.56	0.55
Medi. Chemistry	Synthetic accessibility	4.78	7.28	5.27	4.40

that all antagonists would be able to bind to the target protein (Supplementary Figure S2).

3.7 Identification of the protein-ligand interaction

The protein-ligand interaction is important to observe the possibility of achieving better biological functions (Opo et al., 2021). In our experiment, we discovered that the higher the binding affinity, the greater the interaction with the amino acids' various targets. In the interaction analysis, ZINC70454679 showed the formation of six bonds with the various amino acids, such as six van der Waals bonds (SER536, SER535, ASN580, LEU597, GLY596, THR529), one conventional hydrogen bond (ASP594), five pi-sigma bonds (ILE463, VAL471, PHE583), pi-pi T shaped (PHE595), three alkyl bonds (ALA481, LEU514, CYS532), and four pi-alkyl bonds (PHE595, ALA481, VAL471, LYS483). In ZINC253500968, seven conventional hydrogen bonds were formed and interacted with GLY596, ASN581, CYS532, PHE595, one carbon hydrogen bond (SER536), one Pi-Sigma bond (VAL471), one Pi-Pi T-shaped bond (PHE595), and two Pi-Alkyl bonds (CYS532, LYS483), but the maximum amino acids showed van der Waals bonds (TRP531, GLN530, PHE583, LEU514, ALA481, ILE463, THR529, GLU501, ASN580, ASP536, GLY464, ASP594, LEU597). ZINC106887736 has been shown to interact with several amino acids such as van der Waals bonds (SER536, ASN580, ASN581, LYS578, GLY596, LEU597), conventional hydrogen bonds (PHE595), Pi-Sigma (PHE583), Pi-Sulfur (CYS532), Pi-Pi Stacked (PHE583), Pi-Pi T-shaped (TRP531), two alkyl bonds (LEU514, CYS532), and Pi-alkyl bonds (VAL471). ZINC107434492, on the other hand, formed a van der Waals bond with ten amino acid residues (LEU514, ILE463, ALA481, VAL471, LYS483, LEU597, THR529, GLY596, GLU501, ILE527, CYS532), one conventional hydrogen bond (ASP594), one carbon hydrogen bond (PHE595), two Pi-Sigma bonds (TRP531, PHE583) and one alkyl bond (LEU505) with the BRAF protein (Figure 4 and Figure 5).

3.8 Pharmacophore features analysis

Lead development screening is an important aspect of the biopharmaceutical industry prior to the development of a medication, and these features predict the possibility of binding with the macromolecule. The analysis of pharmacophore features predicts the H, AR, HBA or HBD, PI, and NI characteristics among the compounds, which are essential parts of predicting binding capacity among the proteins (Batool et al., 2019). By using the rule of five, we were able to interpret the drug-likeness and non-drug aspects of the top four higher binding energy molecules: ZINC70454679, ZINC253500968, ZINC106887736, and ZINC107434492. The pharmacophore characteristics

generated by the examined ligands were superior to or comparable to the antagonist attached to the protein (PDB ID: 5VAM) (Figure 6).

3.9 Pharmacokinetic (ADME) evaluation

For computational drug design, it enabled us to get the absorption, distribution, metabolism, excretion, and toxicity analysis before going to establish a molecule as a drug candidate. From administration to excretion by sweat, urine, or stool, the key pathways for a drug showing efficacy inside the body are absorption, distribution, metabolism, and excretion (Watanabe et al., 2019). For higher bioavailability, the drug's volume of distribution to the tissue and target site must be increased, and to lessen side effects and toxic effects, the drug should be washed out easily through following the metabolic pathway. We evaluated ADME properties such as lipophilicity, water-solubility, drug-likeness, and medicinal chemistry by using the online Swiss ADME database (<http://www.swissadme.ch/>) (Daina et al., 2014). The characteristics of the drugs provide us with important information regarding the formulation (tablet, ointment, capsule, injection, and inhaler) and its route of administration (Table 2).

3.10 Toxicity prediction

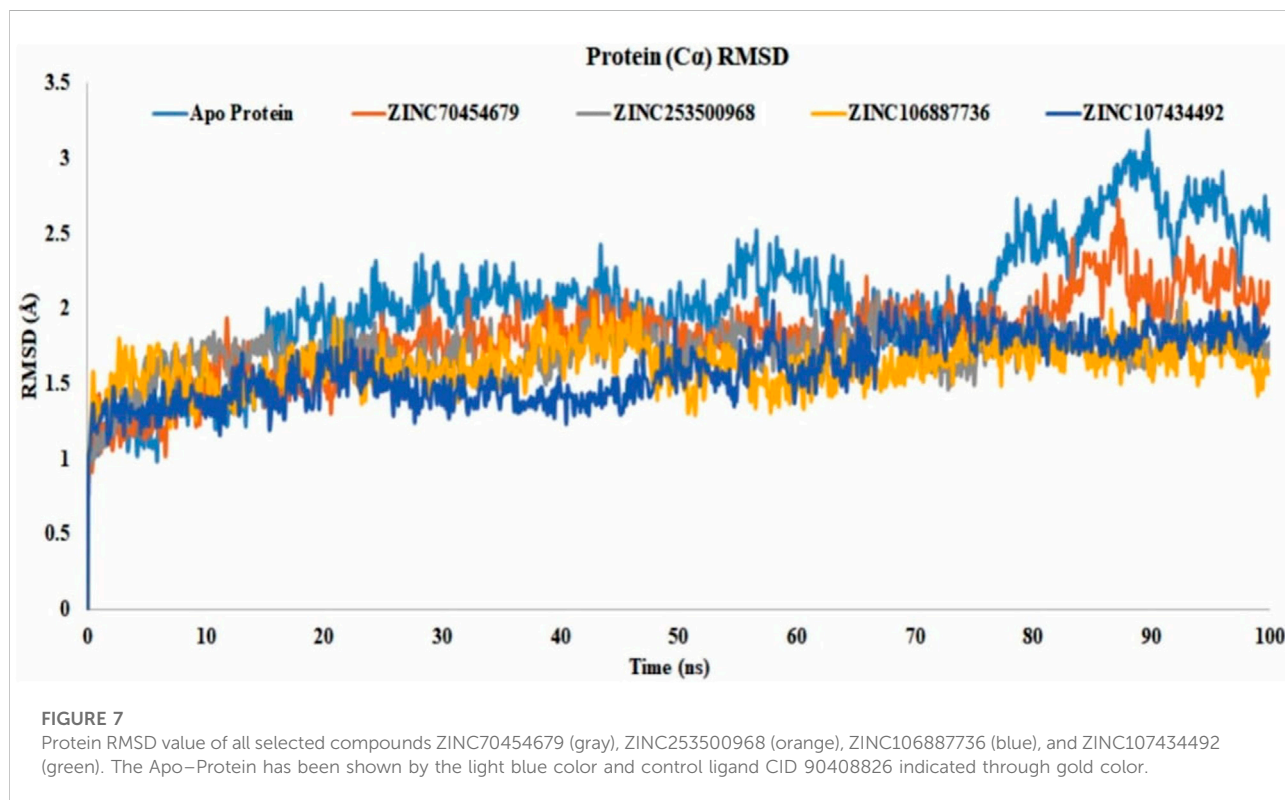
Because of its accuracy, efficiency, and availability for both synthetic and natural chemicals, toxicity analysis is a common technique to choose a suitable therapeutic candidate using computer-based drug discovery. TEST and ProTox-II are two free tools that can be used to test a compound's toxicity. The drug candidate must be chosen based on toxicity, as the less toxic drugs are better for disease intervention. Table 3 showed the results of the cytotoxicity, mutagenicity, carcinogenicity, hepatotoxicity, and LD₅₀ (mg/kg) tests based on software analysis. Three compounds, such as ZINC70454679, ZINC253500968, and ZINC106887736, were shown to have immunologic reactions, except ZINC107434492. Other toxicity data was not available for these compounds, although some data was missing in the case of ZINC106887736.

3.11 Protein ligand complex structure analysis

The interaction between the protein and ligand with the same environmental factors inside the human body is predicted by molecular simulation. It also tells us how many different sorts of bonds there are and how they interact with the different amino acids throughout time. The concentration of the ion, pH all were kept near to the same environment of human body before

TABLE 3 Various toxicities (Organ Toxicity, Toxicity Class, Tox21-Nuclear receptor signaling pathways, Tox21-Stress response pathway, Fathead minnow LC50 (96 h), Developmental toxicity, Water solubility, Oral rat LD50, and Bioaccumulation factor of selected four compounds) were investigated.

Endpoint	Target	ZINC70454679	ZINC253500968	ZINC106887736	ZINC107434492
Organ Toxicity	Hepatotoxicity	Inactive	Inactive	Inactive	Inactive
Toxicity Endpoints	Carcinogenicity	Inactive	Inactive	Inactive	Inactive
	Immunotoxicity	Active	Active	Active	Inactive
	Mutagenicity	Inactive	Inactive	Inactive	Inactive
	Cytotoxicity	Inactive	Inactive	Inactive	Inactive
	LD ₅₀ (mg/kg)	159	5,000	300	8,300
	Toxicity Class	3	5	3	6
Tox21-Nuclear receptor signaling pathways	Androgen Receptor (AR)	Inactive	Inactive	Inactive	Inactive
	Aryl hydrocarbon Receptor (AhR)	Inactive	Inactive	Inactive	Inactive
Tox21-Stress response pathway	Heat shock factor response element	Inactive	Inactive	Inactive	Inactive
	Mitochondrial Membrane Potential (MMP)	Inactive	Active	Active	Inactive
	Phosphoprotein (Tumor Suppressor) p53	Inactive	Inactive	Inactive	Inactive
Fathead minnow LC50 (96 h)	mg/L	267.57	N/A	N/A	18.40
48-h <i>Daphnia magna</i> LC ₅₀	mg/L	11.58	149.64	N/A	85.90
Developmental toxicity	value	1.19	N/A	N/A	0.72
Oral rat LD ₅₀	mg/kg	151.54	N/A	N/A	124.33
Mutagenicity	Result	Negative	Negative	N/A	Negative
Water Solubility (25°C)	mg/L (predicted Value)	489.55	6,263.21	N/A	805.91



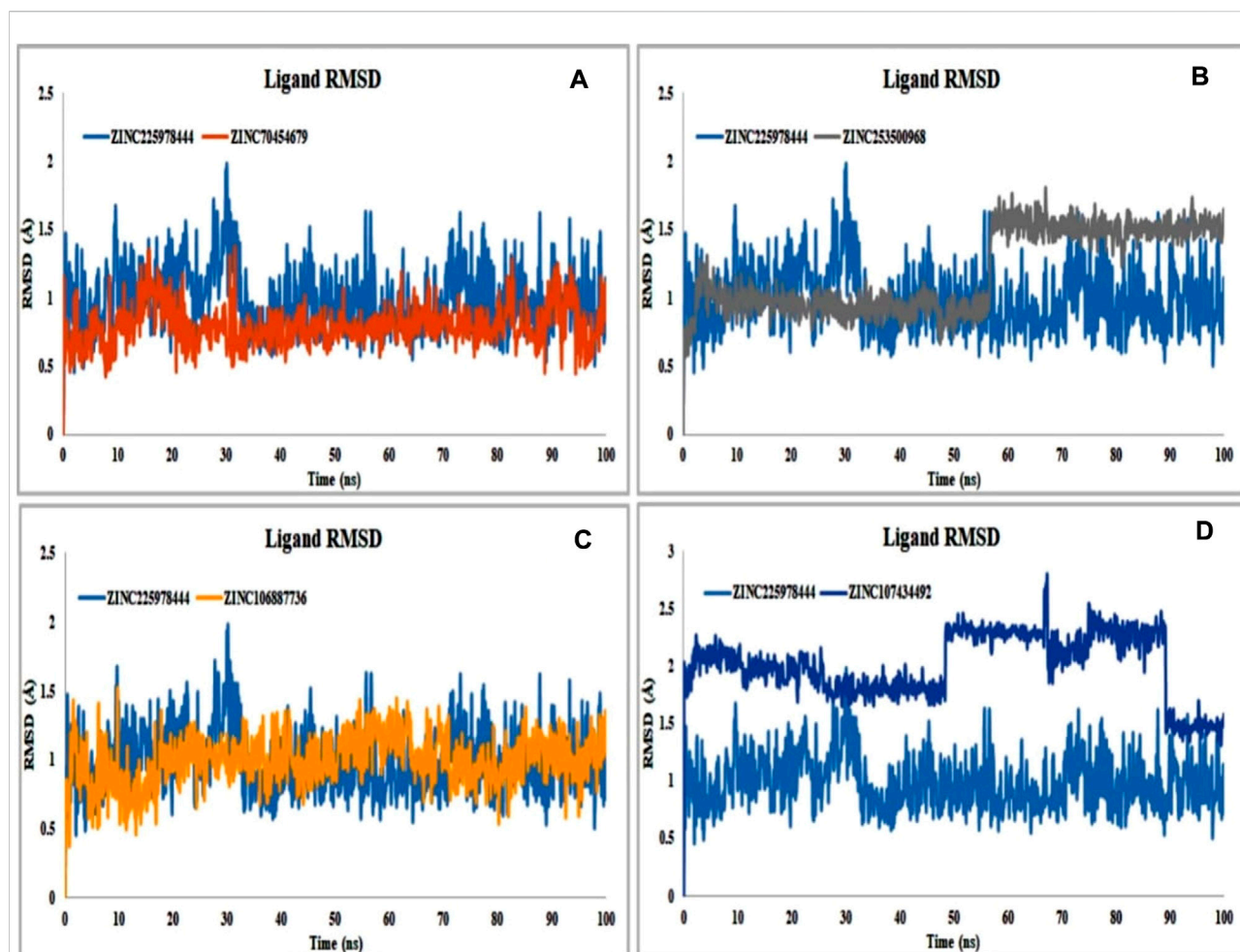


FIGURE 8

Protein compatibility-RMSD value determined from the ligand interaction. Several colors such as (A) ZINC70454679 (gold), (B) ZINC253500968 (grey), (C) ZINC106887736 (orange), and (D) ZINC107434492 (blue) indicate the number of ligands and their expression patterns in comparison to control ligand (CID: 90408826).

proceeding to simulation. The 'pdb' files of the compounds were chosen for simulation based on the binding score generated by docking. The protein secondary structure elements were analyzed in each trajectory frame at the time of simulation (Supplementary Figure S3).

3.11.1 Analysis of the protein RMSD

The RMSD value showed us the number of atoms that were not fitted properly. Most of the compounds have shown that they were stable with the interaction between the protein and ligands. The values of more than 3 Å indicated the conformational changes of the protein and the system were unstable. The analysis of all selected proteins ZINC70454679, ZINC253500968, ZINC106887736, and ZINC107434492 revealed that most of the 100 ns are stable, with the exception of the apo protein, which fluctuated at

89.6 ns and again at 90.5 ns. The selected compound has shown lower fluctuations (Figure 7) in contrast to the control protein (5VAM).

3.11.2 Ligand RMSD analysis

Binding of the ligand with the protein and their stability is the important parameters for the proper efficacy of a drugs. The selected compound ZINC225978444 was found to be the most unstable in the interaction with the protein-ligand complex in our experiment. In 49.2 ns it showed instability and again was stable until 66.6 ns and again unstable from 66.2 to 68.7 ns. Finally, it comes to the stability of the 89.5 ns through slight instability at 88.7 ns. In 56.6 ns, the compound ZINC253500968 showed slight instability and again came to stable 57.3 ns. On the other hand, all other

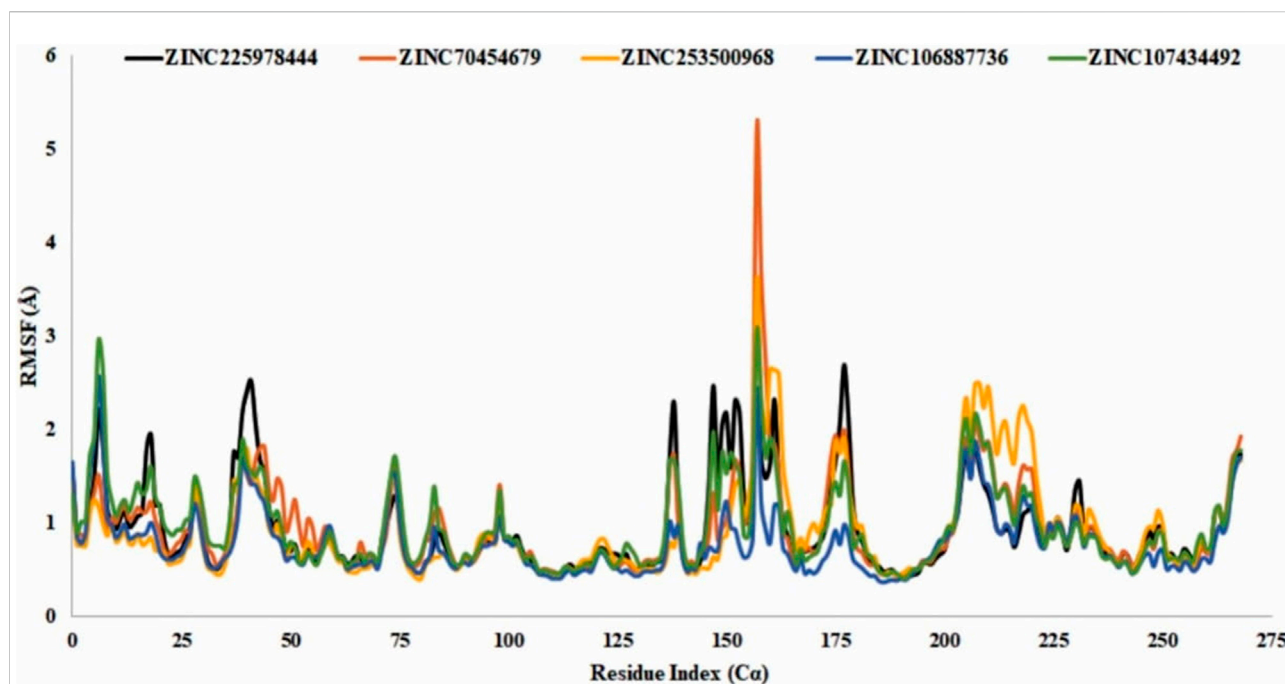


FIGURE 9

RMSF value identification of all the selected compounds from the obtained Ca value. The color of the graph indicated the compounds RMSF value such as control ligand CID 90408826 (black), ZINC70454679 (orange), ZINC253500968 (gold), ZINC106887736 (blue), and ZINC107434492 (green). N- and C-terminal showed fluctuation more than the other but the value with the 3Å.

compounds showed good stability within the protein-ligand interaction complex (Figure 8).

3.11.3 RMSF analysis

RMSF analysis showed that the local conformational changes in the protein and the compounds were used as antagonists. The local fluctuations of the protein with the interaction of our selected compounds were determined by the Ca residue index (Figure 9). In our experiment, all the selected compounds, CID 90408826 (BRAf: 5VAM), ZINC253500968, ZINC106887736, and ZINC107434492, showed a stable RMSF value within the 1–3 Å except ZINC70454679, which showed a little fluctuation at position 157 amino acid residue (PHE610). ZINC253500968 showed a slight fluctuation at the same amino acid position of 157 and then came to a stable position again.

3.11.4 Identification of protein-ligand interaction

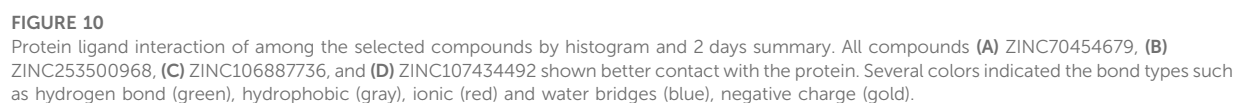
For consideration of a compound as a drug molecule, it should have the properties to bind with the target protein by several bonds, such as conventional hydrogen bonds, hydrophobic, hydrophilic interactions, pi-sigma interactions, pi-sigma bonds, etc (Varma et al., 2010). The majority of the amino acid residues in all compounds came into contact with the ligands during the various interactions. In ZINC70454679, three

amino acids did not come into contact, such as GLY466, GLU533, SER535 and six amino acid residues (GLN461, ARG462, GLU533, TYR538, and ARG662) did not bind with the protein in the case of ZINC253500968 (Figure 10).

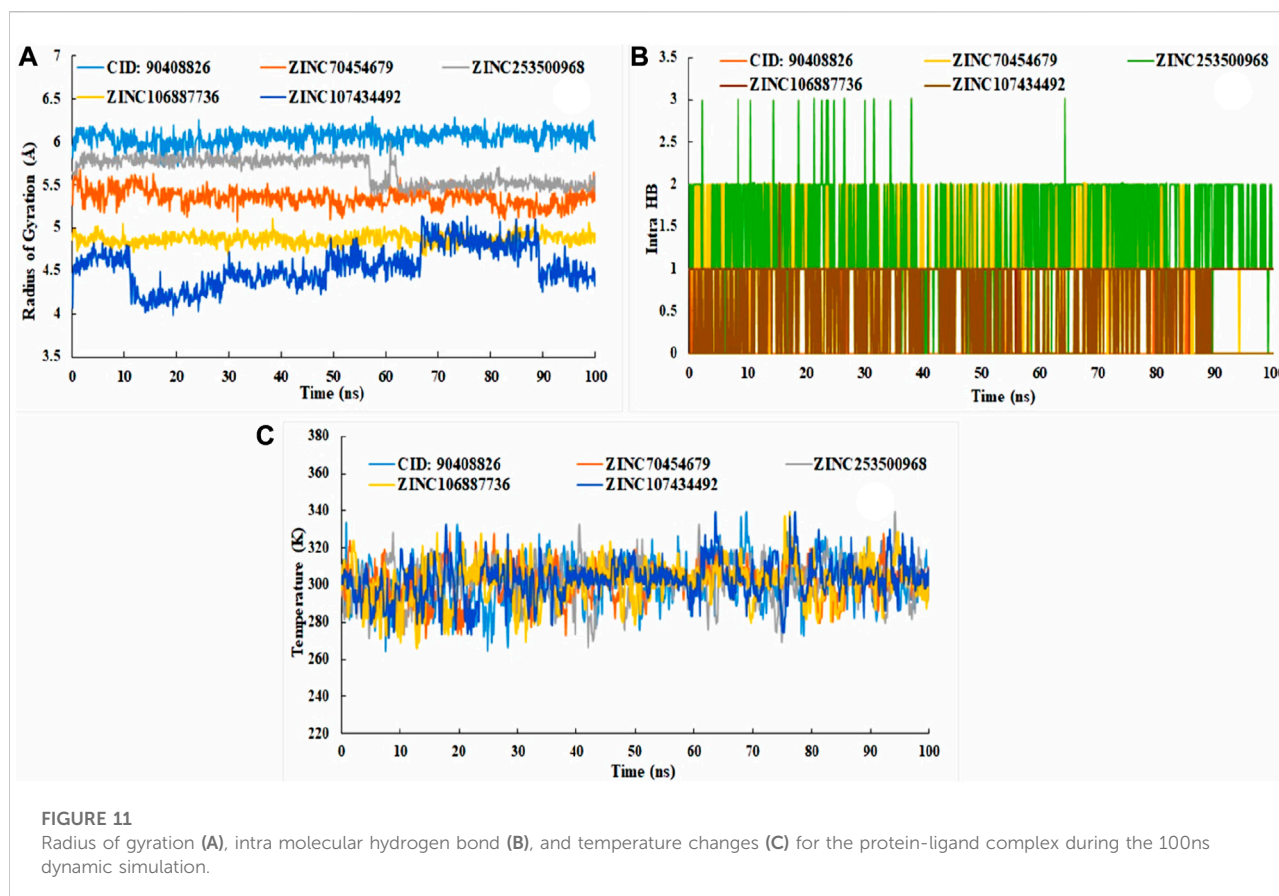
To comprehend how the selected four antagonists' structural evolution were changed across the simulation trajectories analysis from 0 to 100ns, the torsional conformations of each rotatable bond in the ligand were determined (Supplementary Figure S4). Gyration analysis showed that all the compounds were compressed throughout the simulation time except ZINC253500968. Structural transformation occurred from 10 to 40ns as sudden dropped was observed for the ZINC253500968. In case of other compounds sharp, sudden dropped and peak were not observed, which indicated the low structural change (Figure 11A). The number of the intra molecular hydrogen bond was present overall compounds and the higher in ZINC253500968 (Figure 11B). The temperature variations has been mentioned during the simulation and the showed the fluctuations was low during 100ns simulation time (Figure 11C).

3.11.5 MM-GBSA analysis

Usually MM/GBSA analysis are being used to determine the binding free energy of the selected antagonist from the protein-ligand complex from the trajectory simulation file. Analysis of the



-20.30 kcal/mol, -22.64 kcal/mol respectively for ZINC70454679, ZINC253500968, ZINC106887736, and ZINC107434492. The result depicted that all four potentially lead compounds maintained good interaction with the protein complex. At the same time screening, physical and chemical

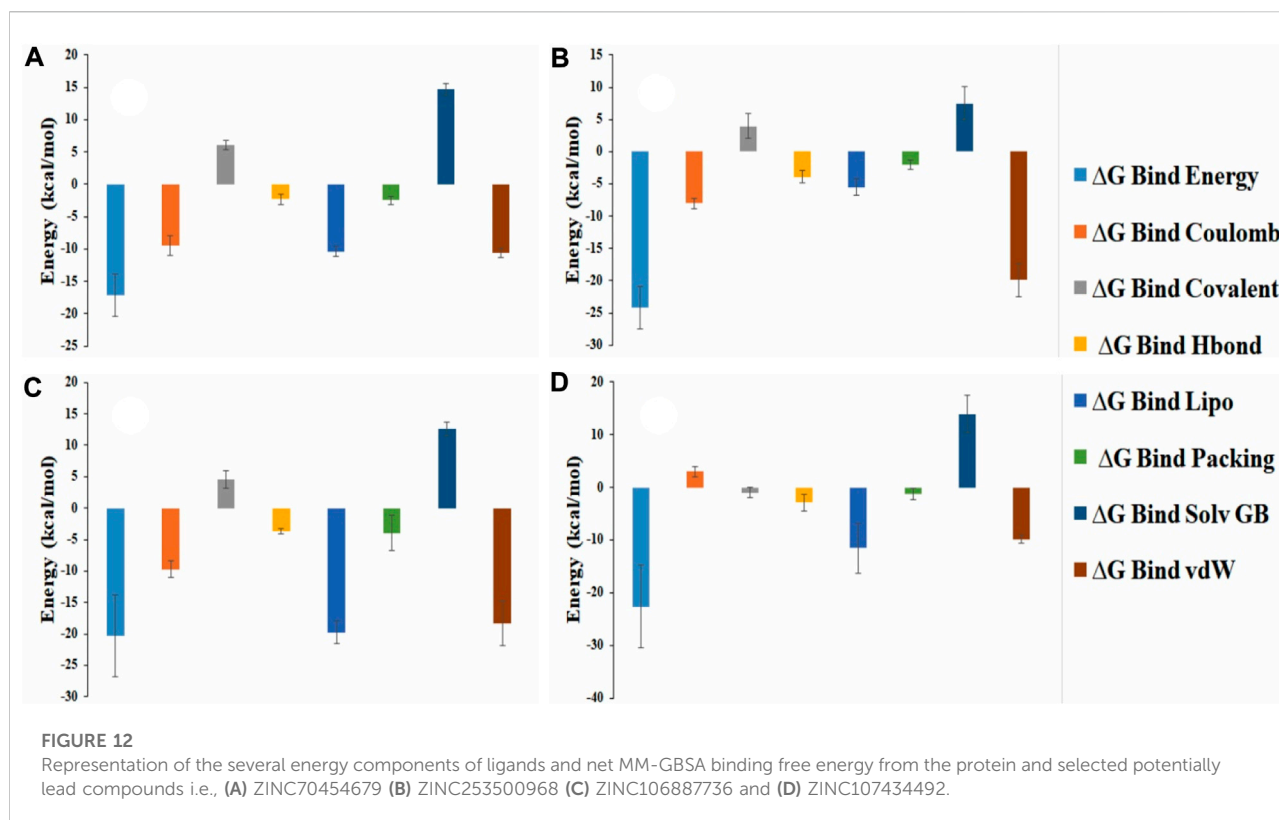


components of our selected ligands were indicated a significant contribution of coulomb energy and Van Der wall interaction energy.

4 Discussion

BRAF mutation in the metastatic colorectal cancer showed poor chemotherapeutic response and shorter the survival rate for patients. V600E mutation in BRAF overexpressed carcinoma consist of near about 80% and other 20% remain in V600 K. Activation of the mitogen-activated protein kinase pathway are responsible for accelerating the RAF (Rapidly accelerated Fibro sarcoma and conduct signal to the signal regulated kinase (MEK), and finally participate cell proliferation and survival through activating the ERK kinase (Extra cellular signal Regulated Kinase) (Leonetti et al., 2018).

Holderfield et al. (2014). It has been identified BRAF mutations as the most frequent mutations related to human carcinomas such as thyroid cancer, ovarian cancer, hepatic carcinoma, and hairy cell leukaemia. The most common mutation has been observed at V600E by sequencing (Yan et al., 2022). The discovery against this BRAF mutation target are in some clinical trial phase and currently using the drugs showing the side effects after administration to the patients (A. Kim and Cohen, 2016) (Holderfield et al., 2014). However, no drugs are available with fewer side effects and to cure cancer as well. Therefore, our study aim was to find potentially lead compounds through computer based drug design that would be effective against the overexpression of the BRAF protein. For computer aided drug design, the BRAF protein structure identified from the online protein database screening and selected protein based on the resolution, R-value free and R-value observed (Ormö et al., 1996). The



ligand attached to the protein were also evaluated by the toxicity software Swiss ADME and also by the ProTox II database (Daina et al., 2017) (Opo et al., 2021) (Rella et al., 2006). The active antagonists were currently available on the market, as well as the literature search was considered for the virtual screening, molecular docking, and also the comparison with the selected compounds. The ZINC and Ambinter databases were further screened for getting the natural compounds with the generation of the pharmacophore features from the Ligand Scout 4.4.8 advanced software (Wolber and Langer, 2005). We arranged all the structures for antagonists with their IC_{50} values and further generated the ROC curve from the ligand scout software, and our obtained ROC curve indicated the satisfactory identification capability. The obtained compounds were docked with the PyRx tool, and compounds were selected based on the docking results (Dallakyan and Olson, 2015).

All the selected four compounds in our in-silico drug design, PubChem ID: 90408826, ZINC253500968, ZINC106887736, and ZINC107434492, indicated the least toxicity based on the evaluation of the ADME profile. Although immunotoxicity is

more common in the cases of control ligand (PubChem ID: 90408826), ZINC253500968, and ZINC106887736, the ADME profiling of ZINC107434492 revealed no toxicity. The compound ZINC253500968 violated three of the five Lipinski rules but was not harmful to humans or animals due to its low toxicity. For the further protein ligand complex stability evaluation of lead compounds, we used molecular dynamic simulation for 100 ns. The trajectory files obtained from the simulation were analyzed based on the RMSD, RMSF value, protein–ligand interaction, intra molecular hydrogen bond, radius of gyration, ligand torsion profile were been evaluated and showed the stability of our four lead compounds. As our potentially lead drug candidates having lower toxicities profile so it could be provided an opportunity to develop lower toxic drug for the researcher and possible to treat BRAF overexpression related cancer. The overall workflow by the in-silico drug design has been mentioned in Figure 13, from the starting of the selection of protein, selected antagonist and molecular dynamic simulation analysis. The majority of the patients were identified as having mutations in BRAF-V600E and were most predominant in thyroid carcinoma, colon cancer,

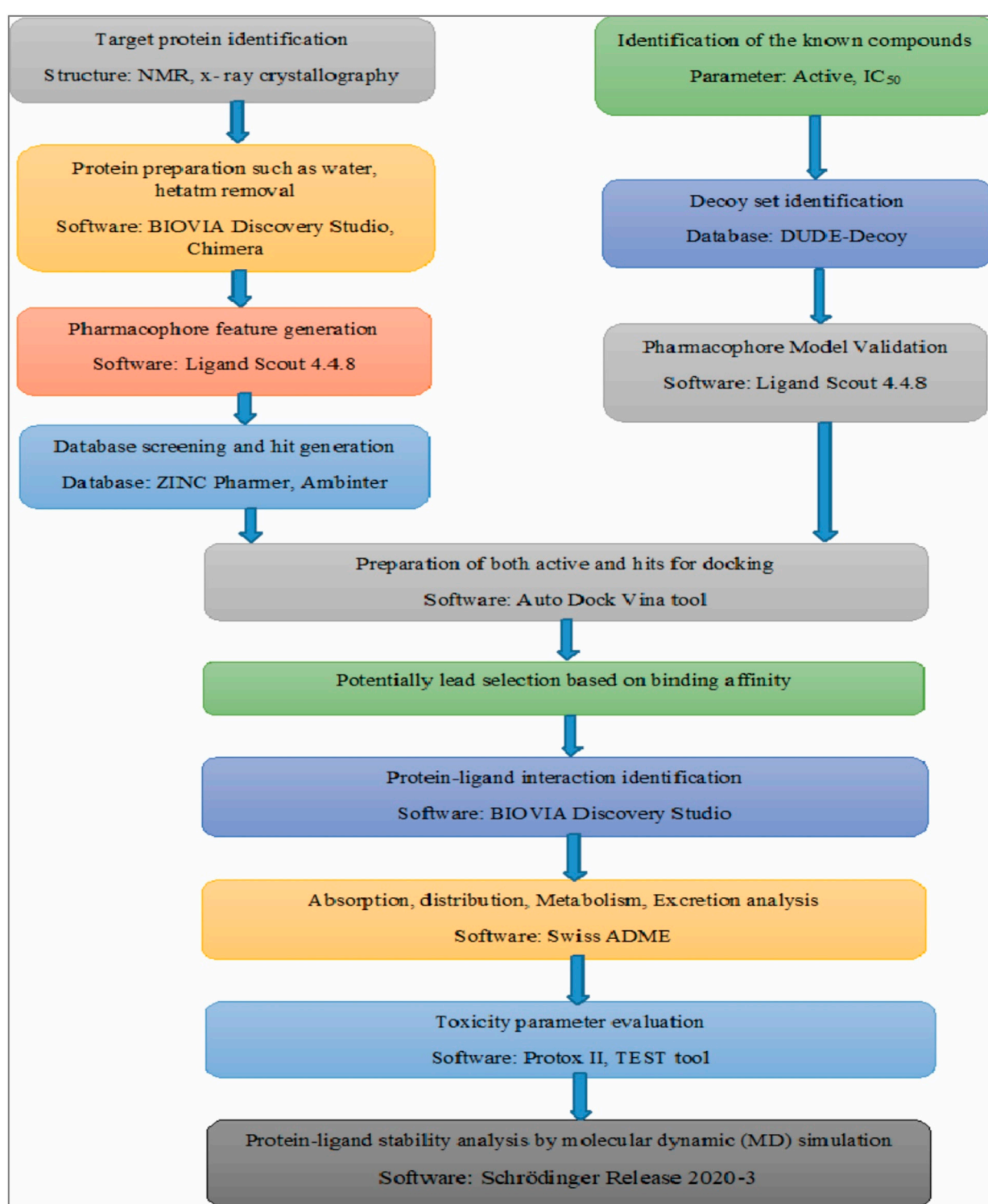


FIGURE 13

Overall the workflow in our computer aided drug design. The figure has mentioned from the beginning of the starting of the protein selection, virtual screening, protein-ligand interaction and stability analysis.

and skin cancer (Tufano et al., 2012) (Lasota et al., 2014) (Ascierto et al., 2012). As a result, the development of a BRAF antagonist will alter treatment options in cancer treatment from the early to late-stage carcinoma and may aid in overcoming drug resistance.

5 Conclusion

In this study, four identified compounds were selected ZINC70454679, ZINC253500968, ZINC106887736, and ZINC107434492 through the virtual screening as a potential lead candidates for BRAF protein overexpression related carcinoma. It may be able to increase apoptosis in several cancer cells by targeting the BRAF protein. The higher binding affinity with the protein showed the docking score from -8.1 to -10.6 kcal/mol and have higher possibilities to bind the target area. The stability of the protein and four ligand complexes were validated through using the dynamic simulation and trajectory file analysis indicated the four key amino acid residues i.e., PHE583, CYS532, VAL471, LEU597, ILE463 based on the interactions. The binding energy was calculated based on the MM-GBSA method and predicted that the lower binding energy due to more stable hydrogen bonds among the protein-ligand complex. Based on the evaluation ADME and toxicity profile of potentially lead compounds, they have lower toxic effects and ZINC107434492 is the most suitable candidate for further analysis as it had no toxicity. Evaluating the *in-silico* toxicity profile of the other available marketed drugs against the BRAF overexpression cancer such as sorafenib, TAK-632 our selected antagonist would have the more possibility to reduce the side effects currently possible anti-cancer treatments. The use of virtual screening, molecular docking, pharmacophore model validation, ADMET profile analysis, protein-ligand binding analysis by discovery studio, and dynamic simulation revealed that these compounds should go for further *in-vitro* as well as *in-vivo* work, which may be able to discover new BRAF antagonists.

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

FAD, FA, MHR, JS-G and FM, designed and performed experimental work, wrote the original draft, analysis result.

FAD, KAY, FM and AAA, MAA securitized the result, analyzed the draft time to time, and provided the direction to work, provided the motivation. FAD, JS-G, FM, MRO, FA, AAA and KAY, read the manuscript, provided the valuable information, guided to work smoothly, analyzed the data. All authors have read and agreed to the published version of the manuscript.

Funding

This work was partly supported by intramural funding to MF from the College of Health and Life Sciences (CHLS), Hamad Bin Khalifa University, Qatar Foundation. FA received support from the College of Health and Life Sciences (CHLS), Hamad Bin Khalifa University. Also, the study was partially supported by the Deanship of Scientific Research (DSR) (Grant Number: IFPRC-115-130-2020), King Abdulaziz University (KAU), Jeddah, Saudi Arabia.

Acknowledgments

We thank to FA for providing manuscript analysis. We also thank to BioSol Centre (www.biosolcent.re.org) for performing molecular dynamics simulation and also for their technical 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/fchem.2022.986376/full#supplementary-material>

References

- Ascierto, P. A., Kirkwood, J. M., Grob, J. J., Simeone, E., Grimaldi, A. M., Maio, M., et al. (2012). The role of BRAF V600 mutation in melanoma. *J. Transl. Med.* 10 (1), 85. doi:10.1186/1479-5876-10-85
- Ayatollahi, H., Tavassoli, A., Jafarian, A. H., Alavi, A., Shakeri, S., Shams, S. F., et al. (2018). KRAS codon 12 and 13 mutations in gastric cancer in the northeast Iran. *Iran. J. Pathol.* 13 (2), 167–172. doi:10.30699/ijp.13.2.167
- Banerjee, P., Dehnbostel, F. O., and Preissner, R. (2018). Prediction is a balancing act: Importance of sampling methods to balance sensitivity and specificity of predictive models based on imbalanced chemical data sets. *Front. Chem.* 6, 362. doi:10.3389/fchem.2018.00362
- Barras, D. (2015). BRAF mutation in colorectal cancer: An update. *Biomark. Cancer* 7 (1), BIC.S25248. doi:10.4137/BIC.S25248
- Batool, M., Ahmad, B., and Choi, S. (2019). A structure-based drug discovery paradigm. *Int. J. Mol. Sci.* 20 (11), 2783. doi:10.3390/IJMS20112783
- Benedetti, M. S., Whomsley, R., Poggesi, I., Cawello, W., Mathy, F. X., Delporte, M. L., et al. (2009). Drug metabolism and pharmacokinetics. *Drug Metab. Rev.* 41 (3), 344–390. doi:10.1080/1083745090289129510.1080/10837450902891295
- Bouback, T. A., Pokhrel, S., Albeshri, A., Aljohani, A. M., Samad, A., Alam, R., et al. (2021). Pharmacophore-based virtual screening, quantum mechanics calculations, and molecular dynamics simulation approaches identified potential natural antiviral drug candidates against MERS-CoV S1-NTD. *Molecules* 26 (16), 4961. doi:10.3390/MOLECULES26164961
- Bowers, K. J., Chow, E., Xu, H., Dror, R. O., Eastwood, M. P., Gregersen, B. A., et al. (2006). Molecular dynamics—Scalable algorithms for molecular dynamics simulations on commodity clusters. *Proc. 2006 ACM/IEEE Conf. Supercomput.* 06, 84. doi:10.1145/1188455.1188544
- Cardarella, S., Ogino, A., Nishino, M., Butaney, M., Shen, J., Lydon, C., et al. (2013). Clinical, pathologic, and biologic features associated with BRAF mutations in non-small cell lung cancer. *Clin. Cancer Res.* 19 (16), 4532–4540. doi:10.1158/1078-0432.CCR-13-0657
- Cope, N., Candelora, C., Wong, K., Kumar, S., Nan, H., Grasso, M., et al. (2018). Mechanism of BRAF activation through biochemical characterization of the recombinant full-length protein. *ChemBioChem* 19 (18), 1988–1997. doi:10.1002/CBIC.201800359
- Daina, A., Michielin, O., and Zoete, V. (2014). Ilogp: A simple, robust, and efficient description of n-octanol/water partition coefficient for drug design using the GB/SA approach. *J. Chem. Inf. Model.* 54 (12), 3284–3301. doi:10.1021/ci500467k
- Daina, A., Michielin, O., and Zoete, V. (2017). SwissADME: A free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. *Sci. Rep.* 7, 42717. doi:10.1038/srep42717
- Dallakyan, S., and Olson, A. J. (2015). Small-molecule library screening by docking with PyRx. *Methods Mol. Biol.* 1263, 243–250. doi:10.1007/978-1-4939-2269-7_19
- Eroglu, Z., and Ribas, A. (2016). Combination therapy with BRAF and MEK inhibitors for melanoma: Latest evidence and place in therapy. *Ther. Adv. Med. Oncol.* 8 (1), 48–56. doi:10.1177/1758834015616934
- Fanelli, G. N., Dal Pozzo, C. A., Depetris, I., Schirripa, M., Brignola, S., Biason, P., et al. (2020). The heterogeneous clinical and pathological landscapes of metastatic BRAF-mutated colorectal cancer. *Cancer Cell Int.* 20 (11), 30–12. doi:10.1186/S12935-020-1117-2
- Gaulton, A., Hersey, A., Nowotka, M. L., Patricia Bento, A., Chambers, J., Mendez, D., et al. (2017). The ChEMBL database in 2017. *Nucleic Acids Res.* 45 (D1), D945–D954. doi:10.1093/NAR/GKW1074
- Geel, R. M. J. M. van, and Iersel, L. B. J. V. (2022). Combined targeted therapy for BRAF mutant metastatic colorectal cancer: Are we there yet? *Dig. Med. Res.* 5 (0), 5. doi:10.21037/DMR-22-15
- Genheden, S., and Ryde, U. (2015). The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. *Expert Opin. Drug Discov.* 10 (5), 449–461. doi:10.1517/17460441.2015.1032936
- Grassi, E., Corbelli, J., Papiani, G., Barbera, M. A., Gazzaneo, F., and Tamperi, S. (2021). Current therapeutic strategies in BRAF-mutant metastatic colorectal cancer. *Front. Oncol.* 11, 601722. doi:10.3389/fonc.2021.601722
- Gray, M. E., Johnson, Z. R., Modak, D., Tamilselvan, E., Tyska, M. J., Sotomayor, M., et al. (2021). A crowdsourcing open platform for literature curation in UniProt. *PLoS Biol.* 19 (12), e3001464. doi:10.1371/JOURNAL.PBIO.3001464
- Guo, Y., Pan, W., Liu, S., Shen, Z., Xu, Y., and Hu, L. (2020). ERK/MAPK signalling pathway and tumorigenesis. *Exp. Ther. Med.* 19 (3), 1997–2007. doi:10.3892/ETM.2020.8454
- Holderfield, M., Deuker, M. M., McCormick, F., and McMahon, M. (2014). Targeting RAF kinases for cancer therapy: BRAF mutated melanoma and beyond. *Nat. Rev. Cancer* 14 (7), 455–467. doi:10.1038/NRC3760
- Hussain, M. R. M., Baig, M., Mohamoud, H. S. A., Ulhaq, Z., Hoessli, D. C., Khogeer, G. S., et al. (2015). BRAF gene: From human cancers to developmental syndromes. *Saudi J. Biol. Sci.* 22 (4), 359–373. doi:10.1016/J.SJBS.2014.10.002
- Hwang, S. B., Lee, C. J., Lee, S., Ma, S., Kang, Y. M., Cho, K. H., et al. (2020). Pmff: Development of a physics-based molecular force field for protein simulation and ligand docking. *J. Phys. Chem. B* 124 (6), 974–989. doi:10.1021/ACS.JPCB.9B10339/ASSET/IMAGES/LARGE/JP9B10339_0001
- Irwin, J. J., Tang, K. G., Young, J., Dandarchuluun, C., Wong, B. R., Khurelbaatar, M., et al. (2020). ZINC20 – a free ultra large-scale chemical database for ligand discovery. *J. Chem. Inf. Model.* 60 (12), 6065–6073. doi:10.1021/ACS.JCIM.0C00675
- Jendele, L., Krivak, R., Skoda, P., Novotny, M., and Hoksza, D. (2019). PrankWeb: A web server for ligand binding site prediction and visualization. *Nucleic Acids Res.* 47 (W1), W345–W349. doi:10.1093/NAR/GKZ424
- Jin, Z., Wang, Y., Yu, X. F., Tan, Q. Q., Liang, S. S., Li, T., et al. (2020). Structure-based virtual screening of influenza virus RNA polymerase inhibitors from natural compounds: Molecular dynamics simulation and MM-GBSA calculation. *Comput. Biol. Chem.* 85, 107241. doi:10.1016/J.COMBIOCHEM.2020.107241
- Kim, A., and Cohen, M. S. (2016). The discovery of vemurafenib for the treatment of BRAF-mutated metastatic melanoma. *Expert Opin. Drug Discov.* 11 (9), 907–916. doi:10.1080/17460441.2016.1201057
- Kim, S., Chen, J., Cheng, T., Gindulyte, A., He, J., He, S., et al. (2021). PubChem in 2021: New data content and improved web interfaces. *Nucleic Acids Res.* 49 (D1), D1388–D1395. doi:10.1093/NAR/GKAA971
- Koes, D. R., and Camacho, C. J. (2012). ZINCPharmer: Pharmacophore search of the ZINC database. *Nucleic Acids Res.* 40 (W1), W409–W414. doi:10.1093/nar/gks378
- Lasota, J., Kowalik, A., Wasag, B., Wang, Z. F., Felisiak-Golabek, A., Coates, T., et al. (2014). Detection of the BRAF V600E mutation in colon carcinoma – critical evaluation of the immunohistochemical approach. *Am. J. Surg. Pathol.* 38 (9), 1235–1241. doi:10.1097/PAS.0000000000000229
- Leonetti, A., Facchinetti, F., Rossi, G., Minari, R., Conti, A., Friboulet, L., et al. (2018). BRAF in non-small cell lung cancer (NSCLC): Pickaxing another brick in the wall. *Cancer Treat. Rev.* 66, 82–94. doi:10.1016/J.CTRV.2018.04.006
- Luu, L.-J., and Price, T. J. (2019). BRAF mutation and its importance in colorectal cancer. *Adv. Mol. Underst. Colorectal Cancer* 7, 9–12. doi:10.5772/INTECHOPEN.82571
- McCubrey, J. A., Steelman, L. S., Chappell, W. H., Abrams, S. L., Wong, E. W. T., Chang, F., et al. (2007). Roles of the raf/mek/erk pathway in cell growth, malignant transformation and drug resistance. *Biochimica Biophysica Acta - Mol. Cell Res.* 1773 (8), 1263–1284. doi:10.1016/J.BBAMCR.2006.10.001
- Menzies, A. M., Haydu, L. E., Visintin, L., Carlino, M. S., Howle, J. R., Thompson, J. F., et al. (2012). Distinguishing clinicopathologic features of patients with V600E and V600K BRAF-mutant metastatic melanoma. *Clin. Cancer Res.* 18 (12), 3242–3249. doi:10.1158/1078-0432.CCR-12-0052
- Merz, V., Gaule, M., Zecchetto, C., Cavaliere, A., Casalino, S., Pesoni, C., et al. (2021). Targeting KRAS: The elephant in the room of epithelial cancers. *Front. Oncol.* 11, 638360. doi:10.3389/fonc.2021.638360
- Morkel, M., Riemer, P., Bläker, H., and Sers, C. (2015). Similar but different: Distinct roles for KRAS and BRAF oncogenes in colorectal cancer development and therapy resistance. *Oncotarget* 6 (25), 20785–20800. doi:10.18632/ONCOTARGET.4750
- Mysinger, M. M., Carchia, M., Irwin, J. J., and Shoichet, B. K. (2012). Directory of useful decoys, enhanced (DUD-E): Better ligands and decoys for better benchmarking. *J. Med. Chem.* 55 (14), 6582–6594. doi:10.1021/jm300687e
- Nguyen-Ngoc, T., Bouchaab, H., Adjei, A. A., and Peters, S. (2015). BRAF alterations as therapeutic targets in non-small-cell lung cancer. *J. Thorac. Oncol.* 10 (10), 1396–1403. doi:10.1097/JTO.0000000000000644
- Nishiguchi, G. A., Rico, A., Tanner, H., Aversa, R. J., Taft, B. R., Subramanian, S., et al. (2017). Design and discovery of N-(2-Methyl-5'-morpholino-6'-(tetrahydro-2H-pyran-4-yl)oxy)-[3, 3'-bipyridin]-5-yl)-3-(trifluoromethyl)benzamide (RAF709): A potent, selective, and efficacious RAF inhibitor targeting RAS mutant cancers. *J. Med. Chem.* 60 (12), 4869–4881. doi:10.1021/ACS.JMEDCHEM.6B01862
- Opo, F. A. D. M., Rahman, M. M., Ahammad, F., Ahmed, I., Bhuiyan, M. A., and Asiri, A. M. (2021). Structure based pharmacophore modeling, virtual screening, molecular docking and ADMET approaches for identification of natural anti-cancer agents targeting XIAP protein. *Sci. Rep.* 11 (1), 4049–4117. doi:10.1038/s41598-021-83626-x

- Orm6, M., Cubitt, A. B., Kallio, K., Gross, L. A., Tsien, R. Y., and Remington, S. J. (1996). Crystal structure of the *Aequorea victoria* green fluorescent protein. *Sci. (New York, N.Y.)* 273 (5280), 1392–1395. doi:10.1126/SCIENCE.273.5280.1392
- Patel, H., Yacoub, N., Mishra, R., White, A., Yuan, L., Alanazi, S., et al. (2020). Current advances in the treatment of BRAF-mutant melanoma. *Cancers* 12 (2), 482. doi:10.3390/CANCERS12020482
- Proietti, I., Skroza, N., Michelini, S., Mambrin, A., Balduzzi, V., Bernardini, N., et al. (2020). BRAF inhibitors: Molecular targeting and immunomodulatory actions. *Cancers* 12 (7), 1823–1913. doi:10.3390/CANCERS12071823
- Rella, M., Rushworth, C. A., Guy, J. L., Turner, A. J., Langer, T., and Jackson, R. M. (2006). Structure-based pharmacophore design and virtual screening for novel Angiotensin Converting Enzyme 2 inhibitors. *J. Chem. Inf. Model.* 46 (2), 708–716. doi:10.1021/ci0503614
- Richards, A., and Garg, V. (2010). Genetics of congenital heart disease. *Curr. Cardiol. Rev.* 6 (2), 91–97. doi:10.2174/157340310791162703
- Roviello, G., D'Angelo, A., Petrioli, R., Roviello, F., Cianchi, F., Nobili, S., et al. (2020). Encorafenib, binimetinib, and cetuximab in BRAF V600E-mutated colorectal cancer. *Transl. Oncol.* 13 (9), 100795. doi:10.1016/j.TRANON.2020.100795
- Rowe, L. R., Bentz, B. G., and Bentz, J. S. (2007). Detection of BRAF V600E activating mutation in papillary thyroid carcinoma using PCR with allele-specific fluorescent probe melting curve analysis. *J. Clin. Pathology* 60 (11), 1211–1215. doi:10.1136/JCP.2006.040105
- Salmaso, V., and Moro, S. (2018). Bridging molecular docking to molecular dynamics in exploring ligand-protein recognition process: An overview. *Front. Pharmacol.* 9, 923. doi:10.3389/fphar.2018.00923
- Shivakumar, D., Williams, J., Wu, Y., Damm, W., Shelley, J., and Sherman, W. (2010). Prediction of absolute solvation free energies using molecular dynamics free energy perturbation and the opls force field. *J. Chem. Theory Comput.* 6 (5), 1509–1519. doi:10.1021/ct900587b
- Stillhart, C., Vučićević, K., Augustijns, P., Basit, A. W., Batchelor, H., Flanagan, T. R., et al. (2020). Impact of gastrointestinal physiology on drug absorption in special populations—An UNGAP review. *Eur. J. Pharm. Sci.* 147, 105280. doi:10.1016/j.EJPS.2020.105280
- Tanda, E. T., Vanni, I., Boutros, A., Andreotti, V., Bruno, W., Ghiorzo, P., et al. (2020). Current state of target treatment in BRAF mutated melanoma. *Front. Mol. Biosci.* 7, 154. doi:10.3389/fmolb.2020.00154
- 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
- Tran, T. V., Dang, K. X., Pham, Q. H., Nguyen, U. D., Trinh, N. T. T., Hoang, L. Van, et al. (2020). Evaluation of the expression levels of BRAF V600E mRNA in primary tumors of thyroid cancer using an ultrasensitive mutation assay. *BMC Cancer* 20 (1), 368–369. doi:10.1186/s12885-020-06862-w
- Tufano, R. P., Teixeira, G. V., Bishop, J., Carson, K. A., and Xing, M. (2012). BRAF mutation in papillary thyroid cancer and its value in tailoring initial treatment: A systematic review and meta-analysis. *Med. (United States)* 91 (5), 274–286. doi:10.1097/MD.0B013E31826A9C71
- Turashvili, G., Grisham, R. N., Chiang, S., DeLair, D. F., Park, K. J., Soslow, R. A., et al. (2018). *BRAF*^{V600E} mutations and immunohistochemical expression of VE1 protein in low-grade serous neoplasms of the ovary. *Histopathology* 73 (3), 438–443. doi:10.1111/HIS.13651
- Valasani, K. R., Vangavaragu, J. R., Day, V. W., and Yan, S. S. (2014). Structure based design, synthesis, pharmacophore modeling, virtual screening, and molecular docking studies for identification of novel cyclophilin D inhibitors. *J. Chem. Inf. Model.* 54 (3), 902–912. doi:10.1021/ci5000196
- Varma, A. K., Patil, R., Das, S., Stanley, A., Yadav, L., and Sudhakar, A. (2010). Optimized hydrophobic interactions and hydrogen bonding at the target-ligand interface leads the pathways of drug-designing. *PLoS ONE* 5 (8), e12029. doi:10.1371/JOURNAL.PONE.0012029
- Wang, L., Lu, Q., Jiang, K., Hong, R., Wang, S., and Xu, F. (2022). BRAF V600E mutation in triple-negative breast cancer: A case report and literature review. *Oncol. Res. Treat.* 45 (1–2), 54–61. doi:10.1159/000520453
- Watanabe, R., Ohashi, R., Esaki, T., Kawashima, H., Natsume-Kitatani, Y., Nagao, C., et al. (2019). Development of an *in silico* prediction system of human renal excretion and clearance from chemical structure information incorporating fraction unbound in plasma as a descriptor. *Sci. Rep.* 9 (1), 18782. doi:10.1038/S41598-019-55325-1
- Wolber, G., and Langer, T. (2005). LigandScout: 3-D pharmacophores derived from protein-bound ligands and their use as virtual screening filters. *J. Chem. Inf. Model.* 45 (1), 160–169. doi:10.1021/ci049885e
- Yan, N., Guo, S., Zhang, H., Zhang, Z., Shen, S., and Li, X. (2022). BRAF-mutated non-small cell lung cancer: Current treatment status and future perspective. *Front. Oncol.* 12, 863043. doi:10.3389/FONC.2022.863043



OPEN ACCESS

EDITED BY

Sibhghatulla Shaikh,
Yeungnam University, South Korea

REVIEWED BY

Malay Kumar Rana,
Indian Institute of Science Education
and Research Berhampur (IISER), India
Hanine Hadni,
Sidi Mohamed Ben Abdellah University,
Morocco

*CORRESPONDENCE

Asnuzilawati Asari,
asnuzilawati@urmt.edu.my
Muhammad Muddassar,
mmuddassar@comsats.edu.pk

[†]These authors have contributed equally
to this work

SPECIALTY SECTION

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

RECEIVED 26 July 2022

ACCEPTED 13 October 2022

PUBLISHED 02 November 2022

CITATION

Ashraf N, Asari A, Yousaf N, Ahmad M,
Ahmed M, Faisal A, Saleem M and
Muddassar M (2022), Combined 3D-
QSAR, molecular docking and dynamics
simulations studies to model and design
TTK inhibitors.
Front. Chem. 10:1003816.
doi: 10.3389/fchem.2022.1003816

COPYRIGHT

© 2022 Ashraf, Asari, Yousaf, Ahmad,
Ahmed, Faisal, Saleem and Muddassar.
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.

Combined 3D-QSAR, molecular docking and dynamics simulations studies to model and design TTK inhibitors

Noureen Ashraf^{1†}, Asnuzilawati Asari^{2*†}, Numan Yousaf^{1†},
Matloob Ahmad³, Mahmood Ahmed⁴, Amir Faisal⁵,
Muhammad Saleem⁶ and Muhammad Muddassar^{1*}

¹Department of Biosciences, COMSATS University Islamabad, Islamabad, Pakistan, ²Faculty of Science and Marine Environment, Universiti Malaysia Terengganu, Kuala Nerus, Terengganu, Malaysia,

³Department of Chemistry, Government College University, Faisalabad, Pakistan, ⁴Department of Chemistry, Division of Science and Technology, University of Education, Lahore, Pakistan,

⁵Department of Biology, Syed Babar Ali School of Science and Engineering, Lahore University of Management Sciences, Lahore, Pakistan, ⁶School of Biological Sciences, University of the Punjab, Lahore, Pakistan

Tyrosine threonine kinase (TTK) is the key component of the spindle assembly checkpoint (SAC) that ensures correct attachment of chromosomes to the mitotic spindle and thereby their precise segregation into daughter cells by phosphorylating specific substrate proteins. The overexpression of TTK has been associated with various human malignancies, including breast, colorectal and thyroid carcinomas. TTK has been validated as a target for drug development, and several TTK inhibitors have been discovered. In this study, ligand and structure-based alignment as well as various partial charge models were used to perform 3D-QSAR modelling on 1H-Pyrrolo[3,2-c] pyridine core containing reported inhibitors of TTK protein using the comparative molecular field analysis (CoMFA) and comparative molecular similarity indices analysis (CoMSIA) approaches to design better active compounds. Different statistical methods i.e., correlation coefficient of non-cross validation (r^2), correlation coefficient of leave-one-out cross-validation (q^2), Fisher's test (F) and bootstrapping were used to validate the developed models. Out of several charge models and alignment-based approaches, Merck Molecular Force Field (MMFF94) charges using structure-based alignment yielded highly predictive CoMFA ($q^2 = 0.583$, $\text{Pred}r^2 = 0.751$) and CoMSIA ($q^2 = 0.690$, $\text{Pred}r^2 = 0.767$) models. The models exhibited that electrostatic, steric, HBA, HBD, and hydrophobic fields play a key role in structure activity relationship of these compounds. Using the contour maps information of the best predictive model, new compounds were designed and docked at the TTK active site to predict their plausible binding modes. The structural stability of the TTK complexes with new compounds was confirmed using MD simulations. The simulation studies revealed that all compounds formed stable complexes. Similarly, MM/PBSA method based free energy calculations showed that these compounds bind with reasonably good affinity to the TTK protein. Overall molecular modelling results suggest that newly designed compounds can act as lead compounds for the optimization of TTK inhibitors.

KEYWORDS

TTK inhibitors, molecular docking, 3D-QSAR, MD simulations, MMPBSA

Introduction

The dual specificity kinase TTK (Also known as monopolar spindle 1 or MPS1) is the core component of spindle assembly checkpoint that ensures accurate segregation of chromosomes during mitosis. TTK controls the bipolar attachment of chromosomes to spindle microtubules by regulating the spindle assembly checkpoint (Wei et al., 2005; Lindberg and Meijer, 2021; Xing et al., 2021). TTK is activated at the unattached kinetochores and recruits' components of the mitotic checkpoint complex (MCC), thereby initiating SAC (Fisk and Winey, 2001; Stucke et al., 2002; Laufer et al., 2014; Huang et al., 2021). MCC hinders metaphase to anaphase transition by inhibiting the activation of Anaphase Promoting Complex/Cyclosome (APC/C) until all the kinetochores are correctly attached to the microtubules, which is prerequisite for accurate chromosome division (Wengner et al., 2016). TTK is composed of 857 amino acids with double lobed protein kinase structure. It comprises C terminal catalytic and activation loops having residues ranging from 515-794. The N terminal lobe (Glu 516-Met602) is smaller than the C terminal lobe (Asn606-Gln794) and has six beta sheets and one alpha helix. The larger C-terminal lobe, however, is more complex and consists of 2 beta sheets, 7 alpha helices together with the activation, catalytic and p + 1 loops. Both lobes join the hinge region through the amino acid residues Glu603 and Gly605. In its dormant state TTK is catalytically inactive as the activation loop is locked. However, phosphorylation at the activation loop enables the TTK to adopt an active conformation and elevate its catalytic activity (Wang et al., 2009). Besides its role in mitosis, it also plays a role in meiosis, cell transformation and cytokinesis (Maia et al., 2015; Sugimoto et al., 2017b). TTK overexpression is detected in many cancer types including, breast, hepatocellular and thyroid carcinomas (Maia et al., 2015; Lu and Ren, 2021). Overexpression of TTK is associated with high serum AFP (alpha-fetoprotein) levels, large tumor size, advanced TNM stage (tumor, nodes, and metastases), and distant metastases. Enhanced expression can also lead to centrosome duplication, genomic instability, mitotic check point failure, abrogated kinetochore attachment, incorrect spindle stress, and chromosomal misalignment (Liu and Winey, 2012; Liu et al., 2015a; Liu et al., 2015b; Sugimoto et al., 2017a). Due to its major role in mitotic checkpoint and overexpression in different malignancies, TTK is considered a potential anti-cancer drug target.

Studies involving RNA interference-mediated knockdown or chemical inhibition of TTK have validated it as a target for cancer therapeutics (Schmidt et al., 2005; Brough et al., 2011; Daniel et al., 2011). Several TTK inhibitors, therefore, have been discovered during the last decade. This includes NMS-P715,

CCT251455, CFI-402257, BOS172722, S81694, BAY1161909 and BAY 1217389 with last five progressing to clinical evaluations (Chen et al., 2018). Similarly, some other small-molecule inhibitors i.e., Diaminopyridine, pyrrolopyrimidine and quinazolines containing compounds have shown low nano-molar activities with reasonably well growth inhibition of cell lines (Kusakabe et al., 2012; Bursavich et al., 2013).

In the current study Pyrrolo pyridine derivatives were used to develop 3D-QSAR models for designing of TTK inhibitors with improved activity. To the best of our knowledge so far, no QSAR modeling and docking simulations have been performed on this class of compounds. CoMFA and CoMSIA models were developed using different alignment schemes and charge models which were then validated using various statistical methods. The information derived from the models were exploited in designing of new compounds that are predicted to have better biological activities than the existing compounds in this class. The stability of binding modes and interactions of newly designed compounds with TTK protein were confirmed by MD Simulations.

Materials and methods

Data collection

Different reported inhibitors of TTK protein sharing similar scaffolds but different biological activities were retrieved from the literature (Naud et al., 2013). The IC₅₀ values of all inhibitors were converted into pIC₅₀ values. The 39 retrieved compounds were randomly divided into two groups: training (28 compounds) and test (11 compounds) datasets (Puzyn et al., 2011). The pIC₅₀ values were used as dependent variable while CoMFA and CoMSIA descriptors were taken as independent variables (Balasubramanian et al., 2014).

Structure preparation and alignment

The 2D structures of inhibitors were sketched by 2D builder tool of Maestro implemented in Schrödinger's suite (Bhachoo and Beuming, 2017). The structures of all compounds were minimized by Conjugate gradient and Powell methods, while the partial charges were computed by Gasteiger Huckel (GH), Gasteiger Marsili (GM), Pullman, and MMFF94 charges (Sainy and Sharma, 2015; Shiri et al., 2016). The compound with the highest biological activity among all the inhibitors was selected as a template for ligand and structure-based conformer alignment of all compounds.

CoMFA and CoMSIA field calculations

The CoMFA electrostatic and steric fields were calculated through SYBYL software using a 3D grid having a 2.0 Å spacing (Ghosh et al., 2021). A fixed energy value of 30 kcal/mol was set to avoid energy clashes. A carbon with sp^3 hybridization and an atom with +1.0 charge were used as steric and electrostatic probes, respectively. A probe atom having a radius of 1.0 Å was used to calculate the CoMSIA fields. The attenuation factor (α) with a default value 0.3 was used to calculate the distance dependent similarities. The Eq. 1 was used to calculate the indices. All computations were carried out in the same way as the CoMFA analysis (Hu et al., 2009; Li et al., 2017).

$$A_{F,K}^q(j) = \sum \omega_{\text{probe},k} w_{i,k} e^{-\alpha r^2} i^q \quad (1)$$

A^q = similarity index

K = physiochemical properties of CoMFA fields descriptors

ω_{probe} = the probe atom

i = summation index of molecule

$w_{i,k}$ = observed value k of a specific property of the atom

r = atomic radius

The efficiency of SAR model was determined by Partial Least Square regression. The CoMFA and CoMSIA descriptors were selected as dependent variables while IC₅₀ value was selected as an independent variable in PLS regression (Tahir et al., 2018). The cross validation using the leave-one-out method was used to select the best model that had high prediction power. The cross-validation (q^2) analysis is defined by Eq. 2.

$$q^2 = 1 - \frac{\sum_y (y_{\text{pred}} - y_{\text{obs}})^2}{\sum_y (y_{\text{obs}} - y_{\text{mean}})^2} \quad (2)$$

y_{pred} = predicted values

y_{obs} = experimental values

y_{mean} = mean values

For non-cross validation, the column filtering was set to 2.0. Standard error estimation (SEE) values were also calculated along cross and non-cross validation. To evaluate the effectiveness of the generated models, bootstrapping was used up to 100 runs. Predictive r^2 was used to express the predictive ability of the developed models, that was based on the test set compounds. The predictive r^2 was calculated using Eq. 3.

$$r_{\text{pred}}^2 = \frac{(SD - \text{PRESS})}{SD} \quad (3)$$

SD = sum of squared deviations between pIC₅₀ values of the test set and mean pIC₅₀ values of the training set

PRESS = sum of squared deviations between the test molecules observed and expected activities

Designing of new compounds

Based on the information obtained from the contour maps of best predictive CoMFA and CoMSIA models, ten new compounds were designed by substitution of specific electrostatic, steric, hydrophobic, hydrogen bond donor, and hydrogen bond acceptor groups to enhance their inhibitory activities against TTK protein. The newly designed compounds belong to the synthetic class of compounds and their biological activities were predicted using the best predictive models (Lorca et al., 2018; Ghosh et al., 2021).

Molecular docking

The co-crystal structure of TTK (PDB ID: 4C4J) was prepared by protein preparation wizard implemented in Maestro. The receptor was preprocessed by adding hydrogens, removing water, adding charges and fixing residues side chain atoms. The unnecessary ligands and chains were removed while the tautomeric states were generated at pH 7.0. The structure of the receptor was further optimized and minimized by OPLS_2005 forcefield [34]. The grid was generated by selecting the co-crystal ligand to perform site-specific docking. To soften the potential of non-polar sections of the receptor, the van der Waals radii of the receptor atom were scaled to 1.0 and the partial charge cutoff value was set to 0.25. The values for the X, Y, and Z coordinates were 0.8, 17.52, and 45.37 respectively. After grid generation, newly designed compounds were prepared by LigPrep tool of Maestro prior to docking [35]. Different ionization states were generated at pH 7 by using Epik [35]. The stereoisomers of compounds with specified chirality were generated by using OPLS_2005 forcefield. The prepared ligands were then docked to the prepared receptor by using the Glide docking tool and the binding poses were analyzed based on the glide gscore.

MD simulations

The binding poses of each compound were used to make complexes with the TTK protein. The stability of each protein-ligand complex was estimated by running MD simulation using NAMD (Acun et al., 2018). All the complexes were prepared by using Leap module of AMBER21 tools (Case et al., 2021). The parameters of the ligands were generated by antechamber program by semi-empirical calculation. The PDB4amber module was used to convert the amino acid residues to amber format. The forcefield parameters for protein and ligands were AMBER ff14SB force field and general amber forcefield, respectively (Duan et al., 2003). The parameters of ligands and receptor were connected by tleap program. All the complexes were solvated in a water box

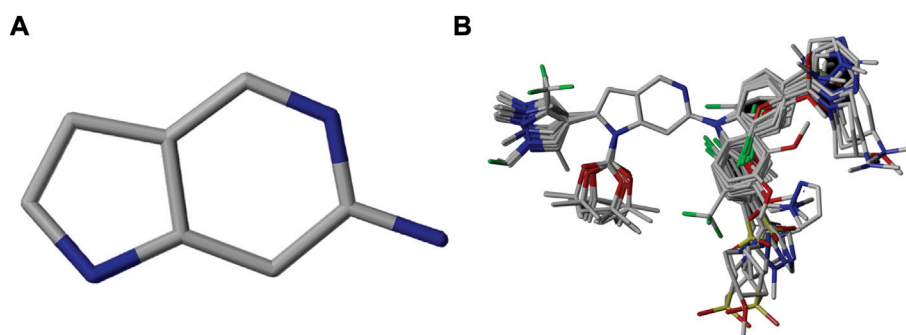


FIGURE 1

Alignment of dataset compounds: (A) Common Substructure and (B) Structure-based alignment of docked compounds.

of size 10 Å using TIP3P water model. To neutralize the system, counter ions Na^+ and Cl^- were added by using Leap. The systems were minimized by conjugating gradient and steepest descent method for 10,000 times. The water equilibration was done for 5,000 steps, followed by the three-temperature equilibration from 0 to 200 K, 200–250 K, and 250–300 K for 5,000 steps. After equilibration of the system at different temperatures, the production of systems was run for 25 ns with constant temperature 310 K and pressure of 1 atm using NPT ensemble. The trajectories of all the systems were analyzed to get RMSD, RMSF, Radius of gyration, SASA, PCA, by using VMD tcl commands, CPPTRAJ (Roe et al., 2013) and R package.

Binding free energy calculation

The binding free energy of the system was calculated by using molecular mechanics-based scoring methods MM/PBSA (Sun et al., 2014). The calculations were based on a total of 300 snapshots of the complex, taken at 2 ps interval from the last 2 ns stable MD trajectories. The binding free energy was determined as the difference between the total free energy (ΔG_{com}) of the ligand-receptor complex and the sum of free energy of individual receptors (ΔG_{pro}) and ligand (ΔG_{lig}) using the equation provided below:

$$\Delta G_{\text{bind}} = \Delta H - T\Delta S = \Delta G_{\text{com}} - [\Delta G_{\text{pro}} + \Delta G_{\text{lig}}]$$

The ΔG for the complex, receptor and ligand can be calculated by the following equation:

$$\Delta G = \Delta E_{\text{MM}} + \Delta G_{\text{sol}} - T\Delta S$$

ΔE_{MM} = Molecular Mechanics Energy

ΔG_{sol} = Solvation Free Energy

$T\Delta S$ = Entropy at given Temperature

ADMET analysis

The physicochemical properties i.e., molecular weight, Hydrogen bond donors and acceptors along with the ADMET properties of the newly designed compounds were predicted by QikProp tool of Maestro (Koç et al., 2021).

Results and discussion

An essential stage in '3D-QSAR' is the systematized assortment of compounds and their division into training and test datasets. Compounds and their biological activities in terms of pIC50 values are mentioned in Supplementary Table S1. They were classified into two categories with respect to their activity range from high to low while maintaining structural variations. All the selected compounds possess a common sub-structure 1H-Pyrrolo[3,2-c] pyridine as shown in Figure 1A. These compounds have mainly hydrophobic (halogens Cl and Br) and hydrophilic substituents (amine and amides) attached to the core scaffold 1H-Pyrrolo[3,2-c] pyridine. Hydrogen bond donors like NH and OH, hydrogen bond acceptors like N, O and F and steric groups like CH3 and Cl have been attached to enhance the activities of the compounds. By changing the substituent at main scaffold, the activity of the compounds predicted by the developed models, was affected. The quality of the models is affected by multiple factors like the conformation of the molecules and their assigned partial charges (Muddassar et al., 2009; Wang et al., 2015). Therefore, different conformations of dataset molecules using ligand and structure-based approaches were generated along with different charge models i.e., GH, GM, MMFF. For structural alignment, compounds were aligned on a common sub-structure to get the best predictive "CoMFA" and "CoMSIA" models (Figure 1).

The best models were obtained with Merck Molecular Force Field charges using structure-based conformation alignment as shown in Figure 1B. The correlation coefficient q^2 of leave-one-out cross validation for CoMFA fields was 0.589, with 3 optimum

TABLE 1 Statistical parameters of structure based CoMFA and CoMSIA models with different charge schemes.

	Gasteiger Huckel charges (GH)		Gasteiger Marsili Charges (GM)		Pullman charges (PM)		Merck molecular force field (MMFF94)	
Parameters	CoMFA	CoMSIA	CoMFA	CoMSIA	CoMFA	CoMSIA	CoMFA	CoMSIA
N	3	3	3	3	3	3	3	3
q ²	0.583	0.705	0.584	0.665	0.575	0.663	0.589	0.690
r ² (NoV)	0.891	0.946	0.888	0.918	0.893	0.950	0.902	0.931
SEE	0.090	0.075	0.090	0.068	0.093	0.087	0.088	0.109
F	65.535	139.223	63.112	89.516	66.594	152.924	73.624	108.296
Pred (r ²)	0.638	0.814	0.767	0.804	0.619	0.721	0.751	0.767
r ² _{bs}	0.928	0.959	0.913	0.937	0.942	0.930	0.919	0.941
SD _{bs}	0.215	0.168	0.216	0.179	0.206	0.201	0.216	0.187
Fields contribution								
Steric (S)	0.714	0.126	0.724	0.139	0.670	0.128	0.683	0.129
Electrostatic(E)	0.286	0.227	0.276	0.192	0.330	0.216	0.317	0.231
Hydrophobic (H)	-----	0.263	-----	0.258	-----	0.272	-----	0.251
Donor (D)	-----	0.171	-----	0.166	-----	0.172	-----	0.177
Acceptor (A)	-----	0.213	-----	0.245	-----	0.212	-----	0.212

N, "Optimal number of components; q², cross-validated correlation coefficient; r², determination coefficient; r² nov, non-cross validated correlation coefficient; SEE, standard error of estimate; F, Fischer's test F-value; Pred-r², predictive r² for test set compounds; r² bs, r² obtained after 100 bootstrapping runs; and SD_{bs}, bootstrapping standard deviation.

number of components, the standard error of estimation was 0.088, non-cross validated coefficient (r²ncv) = 0.902, F-value = 73.624 and r²pred = 0.751 as mentioned in Table 1. The electrostatic and steric fields contributed 68.3% and 31.7% respectively to the model. However, ligand-based conformations yielded the poor predictive models. Powell method generated conformation with MMFF charges produced best CoMFA model with q² = 0.268 value for 3 optimum number of components (other data shown in Supplementary Table S2). Similarly Conjugate Gradient conformation method with MMFF charges using 2 optimum number of components yielded q² = 0.191 value for steric and electrostatic fields (Supplementary Table S3). In the ligand-based alignment technique, the effects of different charges on the models are shown in Supplementary Table S3. The reasons for superior performance of one charge method over the other in "CoMFA" and "CoMSIA" predictive models are still unknown, as the literature

shows variable performance of these charge models on compounds targeting different proteins. As for as CoMSIA models are concerned, structure-based alignment with Merck Molecular Force Field charges produced q² = 0.690 with N = 3, SEE = 0.109, F-value = 108.296, r²ncv = 0.931, and r²pred = 0.767 shown in Table 1. The CoMSIA fields like steric, electrostatic, hydrophobic, hydrogen bond donor and acceptor contributions were 12.9%, 23.1%, 25.1%, 17.7% and 21.2% respectively. The results exhibited that electrostatic and hydrophobic interactions and hydrogen bond donors played major role in CoMSIA model. In CoMSIA modeling GH, GM, PM, and MMFF94 charges did not significantly influence the quality of models. Using the best predictive CoMFA and CoMSIA models, the biological activities of the training and test dataset compounds were predicted as shown in Figures 2A–B, respectively. The scattered plots show that the predicted values are similar to the

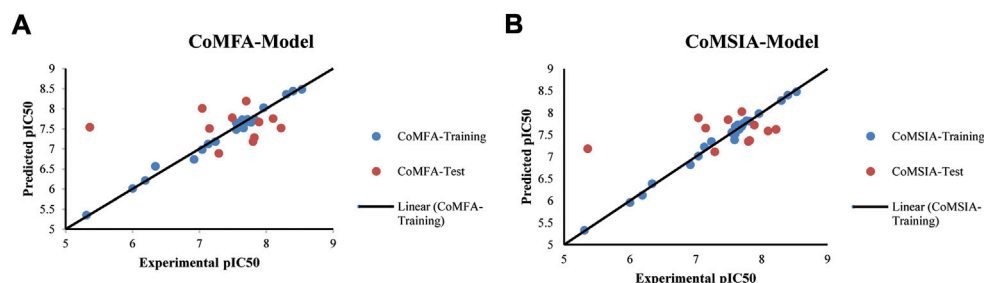
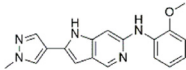
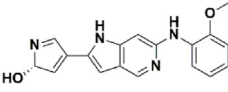
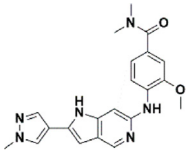
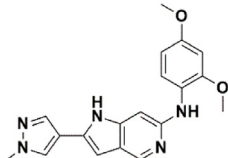
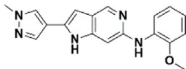
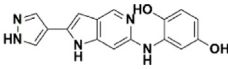
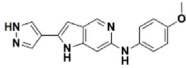
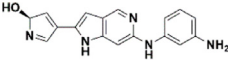
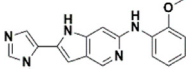
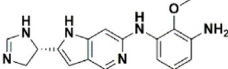
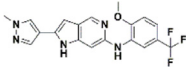
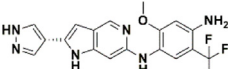
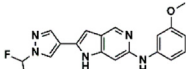
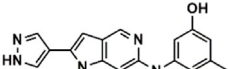
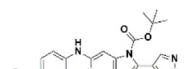
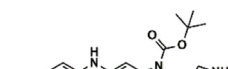


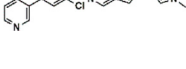
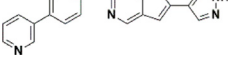


FIGURE 2
Correlation plots between experimental and predicted biological activities (A) From CoMFA Model (B) From CoMSIA Model.

TABLE 2 Comparison of parent and modified compounds activities.

Parent compounds	Actual pIC50	Modified compounds	Predicted-pIC50
9 	6.92	NDC1 	7.04
19 	7.63	NDC2 	6.96
9 	6.92	NDC3 	7.36
2 	7.29	NDC4 	7.10
15 	6.19	NDC5 	7.24
16 	5.36	NDC6 	7.13
10 	6.34	NDC7 	7.13
26 	7.03	NDC8 	7.913
37 	7.56	NDC9 	8.062
39 	7.68	NDC10 	8.063

experimental values except one compound (Outlier). Outliers can occur as a result of incorrectly measured inhibitory concentrations, variable binding confirmations, or major physicochemical variances. Similarly, external validation (higher r^2_{pred} values of test set compounds) of both models shows their highly predictive nature. Internal validations such as r^2_{ncv} , F-values, and r^2_{bs} values revealed their reliability and precision to design and improve new compounds. As any individual field can influence the quality of the model, therefore models with good statistical significance were

used to design new compounds for improved activity as shown in Table 2.

CoMFA contour maps

Contour maps of the best predictive models were generated on the most active compound, and then this 3D information was exploited to create new compounds predicted

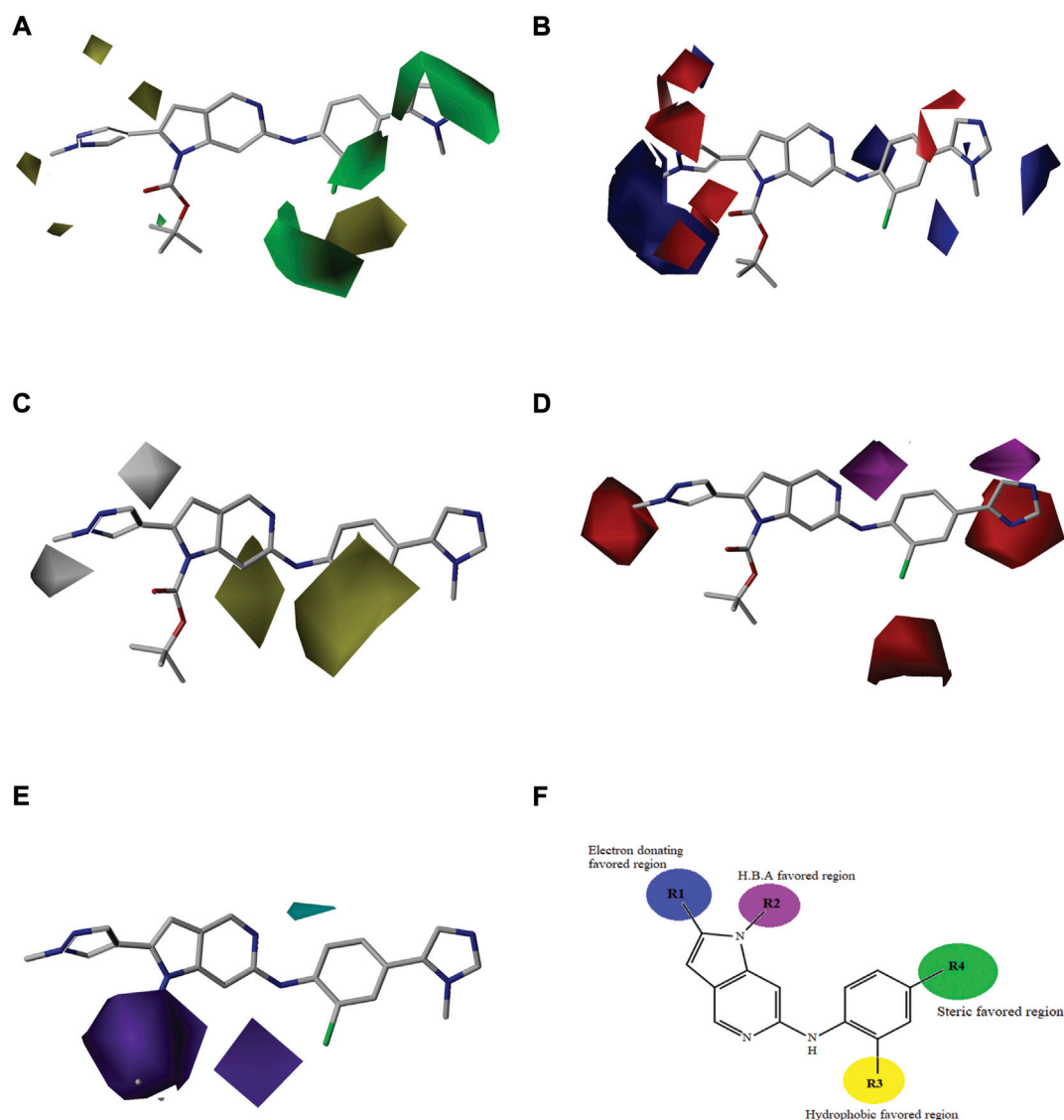


FIGURE 3

"Structure-based model of the most active compound represented through contour maps (36). (A) Contour maps of CoMFA steric field; (B) Contour maps of CoMFA electrostatic fields; (C) Contour maps of CoMSIA hydrophobic field; (D) Contour maps of CoMSIA hydrogen bond acceptor fields; (E) Contour maps of CoMSIA hydrogen bond donor fields; (F) Structure activity relationship representation of contour groups."

to have improved biological activities. The contour maps of CoMFA fields for the best model are shown in Figures 3A,B. The steric contour maps are shown in Figure 3A, the green contour denotes the favored area for bulky group substitution, whereas the yellow contour shows the disfavored area for bulky group substitutions. The replacement of bulky groups at R4 position will increase the activity of compounds. For example, compound 27 ($pIC_{50} = 8.1$) with azetidine amide at R4 is predicted to be more active than compound 17 ($pIC_{50} = 7.13$) which has nothing at same position. Figure 3B shows the electrostatic field contour maps. The red and blue contours

represent the effect of the electrostatic field on the biological activity of compounds. The large blue contour near R1 position shows that the substitution of electron donating group will increase the activity of compound that's why the activity of compound 1 ($pIC_{50} = 7.60$) having electron donating nitrogen at R1, is better than compound 10 ($pIC_{50} = 6.34$) that has electron withdrawing difluoromethyl at the same position. Similarly, the red contour near the R2 indicates that the replacement with electron withdrawing group will increase the bioactivity of the compounds. These observations are in agreement with previously published results (Vaidya et al., 2017).

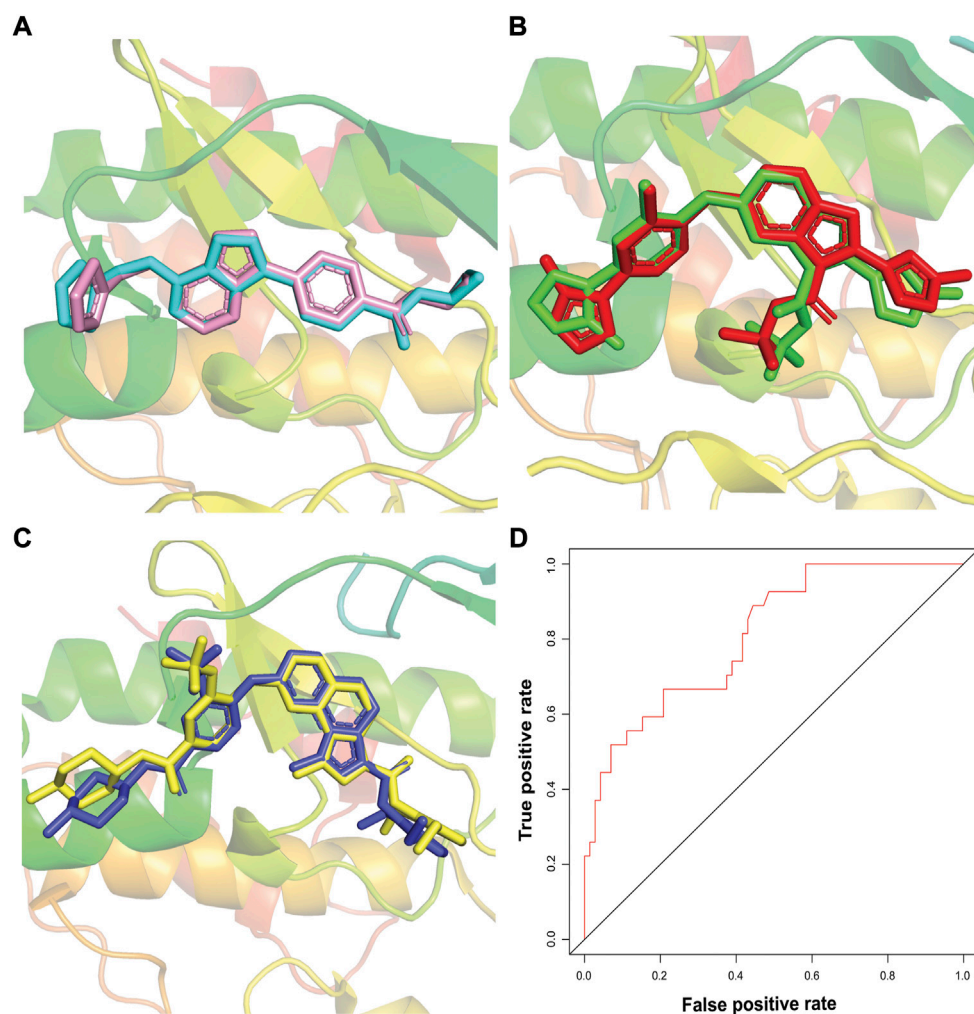


FIGURE 4

Docking protocol validation studies. The redocking of co-crystal ligands of three TTK X-rays crystal structures, **(A)** redocked pose of ligand in PDB ID: 3WZK is shown in cyan sticks, **(B)** redocked pose of ligand in PDB ID: 4C4J is shown in green sticks, **(C)** redocked pose of ligand in PDB ID: 5AP7 is shown in yellow sticks. The RMSD of redocked poses was less than 1Å. **(D)** The estimation of docking accuracy by AUC curve with a value of 0.80.

CoMSIA contour maps

The contour maps of CoMFA and CoMSIA showed the similarity in steric and electrostatic fields. The remaining fields of CoMSIA i.e., hydrophobic, HBA and HBD are shown in Figures 3C,D,E. Figure 3C shows the hydrophobic contour, where yellow contour at the R3 position indicates that the substitution of the hydrophobic group is favorable to increase the activity while white contour near R1 position shows that the activity can be increased by replacing the hydrophilic group at this position. Therefore, compounds 22, 23, 28–30 and 33–39 with hydrophobic groups at the R3 position showed significant predicted biological activities. Figure 3D indicates the hydrogen bond acceptors contour. Magenta contour shows the area which is favorable for hydrogen bond acceptor group substitution while red region is favorable for

hydrogen bond donor group substitution to increase the activity of compounds. Similarly, the purple contour in Figure 5E shows the disfavored area for hydrogen bond donor group substitution. So, the substitution of hydrogen bond donor groups at R1 and R4 position will increase the biological activity of compounds, while the substitution of hydrogen bond acceptor groups at R2 position will increase the activity. In the contour maps, hydrogen bond acceptors and donors shared 80% for favored regions while 20% for unfavored regions to increase the biological activity of compounds. The structure-activity relationship diagram (Figure 3F) was obtained from the CoMFA and CoMSIA contour maps. In order to design new compounds with better biological activities, the regions R1, R2, R3, and R4 are favorable for substitutions of electron donating groups, hydrogen bond acceptor groups, hydrophobic groups and bulky groups, respectively. In order to design new compounds,

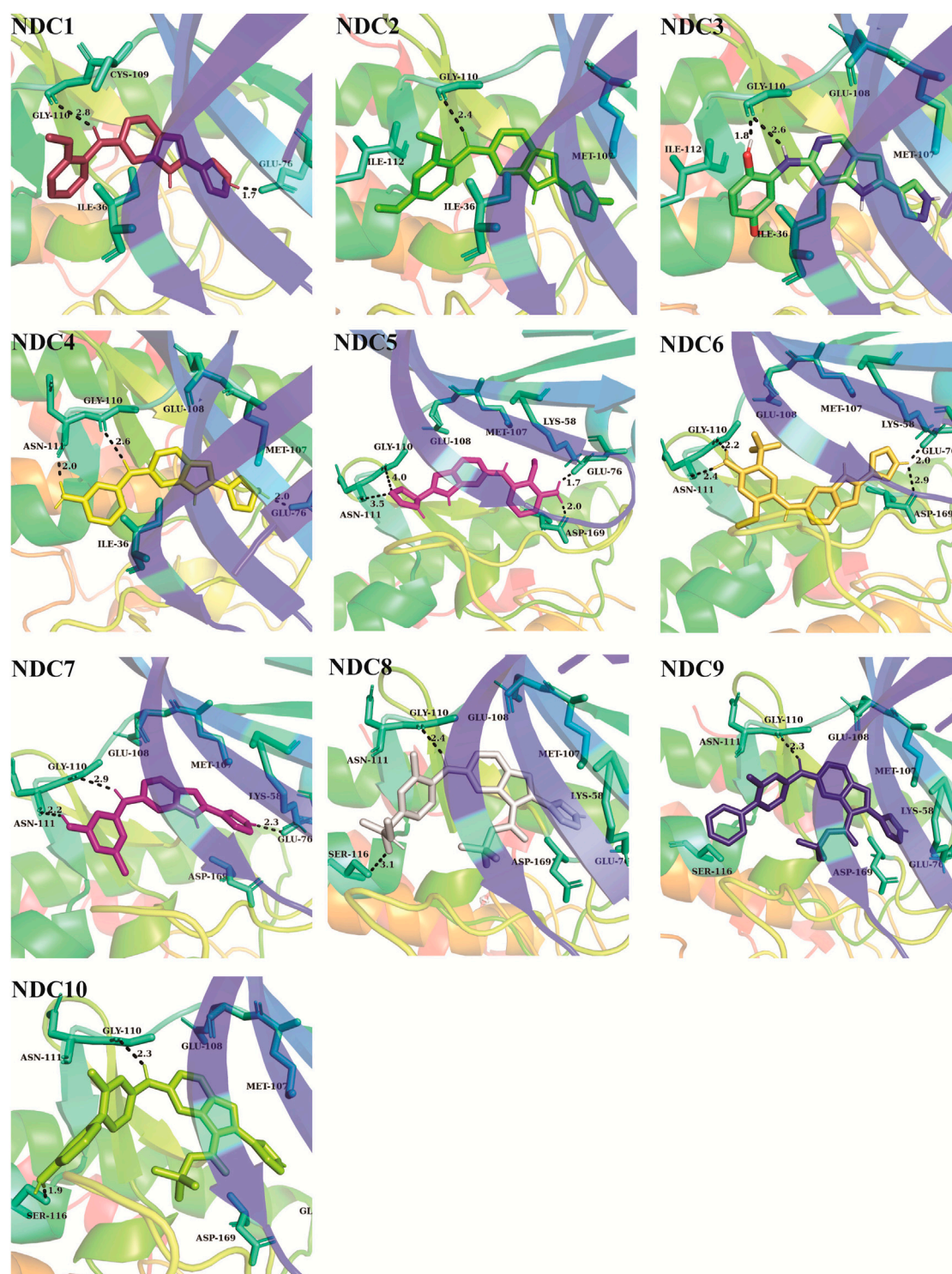


FIGURE 5

The binding interactions of newly designed compounds with the key residues of TTK binding pocket. The hydrogen bonds are denoted with black dash lines. The distance between the compounds and binding site residues is measured in Å.

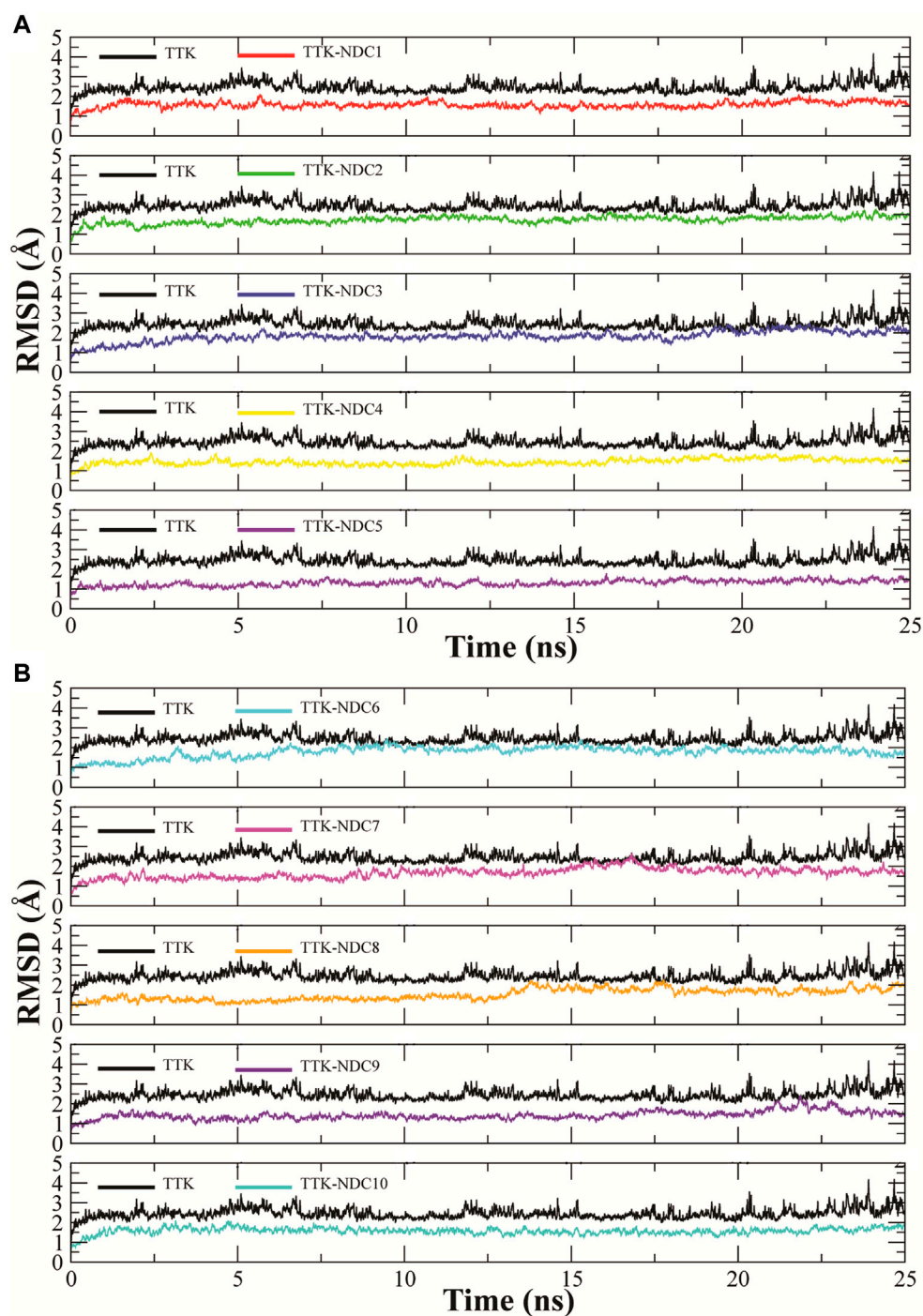


FIGURE 6

Root Mean Square Deviations in backbone of TTK bound to newly designed compounds; (A) TTK-NDC1 (red), TTK-NDC2 (green), TTK-NDC3 (blue), TTK-NDC4 (yellow), TTK-NDC5 (violet) (B) TTK-NDC6 (cyan), TTK-NDC7 (magenta), TTK-NDC8 (orange), TTK-NDC9 (indigo) and TTK-NDC10 (turquoise).

different modifications in the parent structures have been introduced based on the best CoMFA and CoMSIA models contours. For example, pyrazole ring of compound 9 showed favorable region for electron donating groups, so by replacing the methyl group with hydroxyl group, new compounds showed better predicted activity

than parent compound. Similarly, a hydroxyl group was added to the compound 2 to get a new molecule with better activity. All new compounds were designed by adding specific groups at the favorable electron donating, hydrogen bond acceptor and steric group regions for better activities.

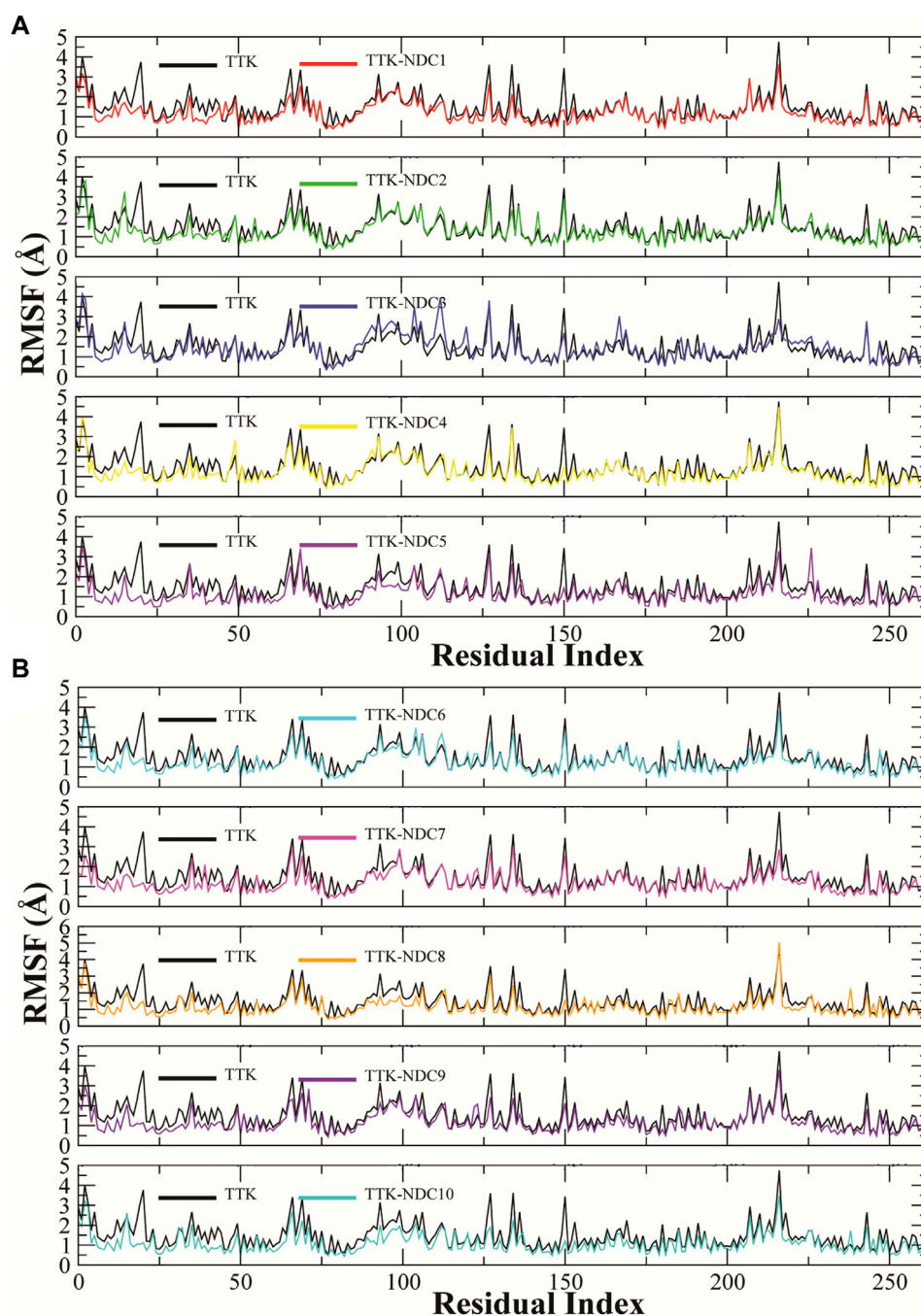


FIGURE 7

(A,B). Root Mean Square Fluctuations of amino acid residues of TTK protein and its complexes to compare the flexibility of protein structures.

Binding mode elucidation of newly designed compounds

Newly designed compounds NDC1-10 were docked into the active site of TTK protein to identify their plausible binding modes. Prior to the docking of newly designed compounds, the glide

docking protocol was validated by calculating the RMSD of redocked poses of co-crystal ligands (Muddassar et al., 2010). The co-crystal ligands were extracted from the co-crystal structures (PDB IDs: 3WZK, 4C4J, 5AP7) and docked again at the same binding position where cocrystal ligands were making the hydrogen bonding interactions with the hinge region residues. The

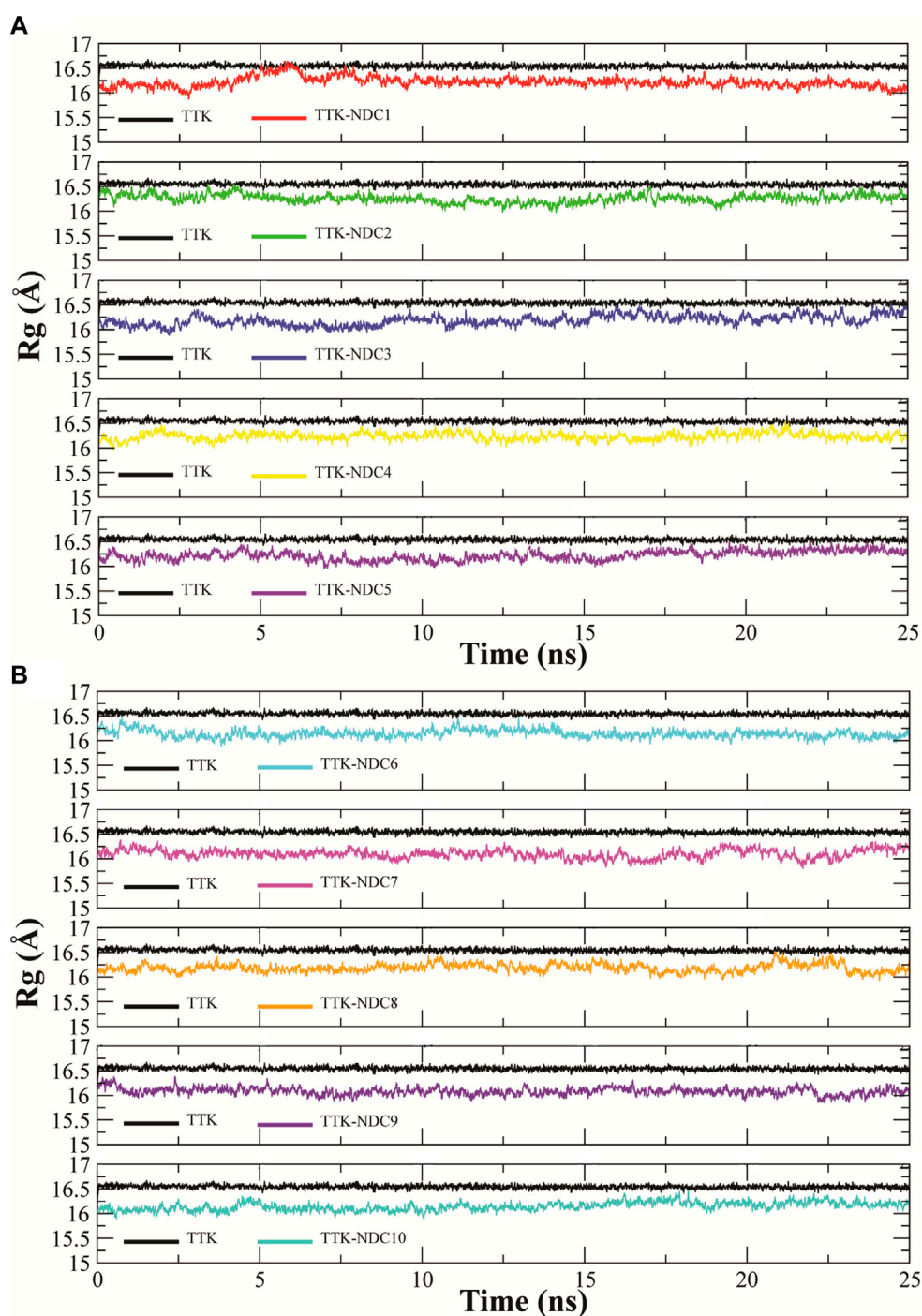


FIGURE 8

(A,B). Radius of gyration of Ca atoms of TTK protein with bound compounds to analyze the relative compactness of protein complexes.

docked pose was then aligned on the native ligand which showed identical interaction with <1.0 Å deviation from original pose. The redocking of representative co-crystal ligands can be observed in Figures 4A–C. Moreover, the accuracy of glide tool was estimated by

area under curve studies. A decoy dataset of 917 compounds was used along with active compounds of TTK. The AUC curve value of 0.80 showed that the true positive rate was higher than the false positive results produced by the glide scoring scheme as shown in

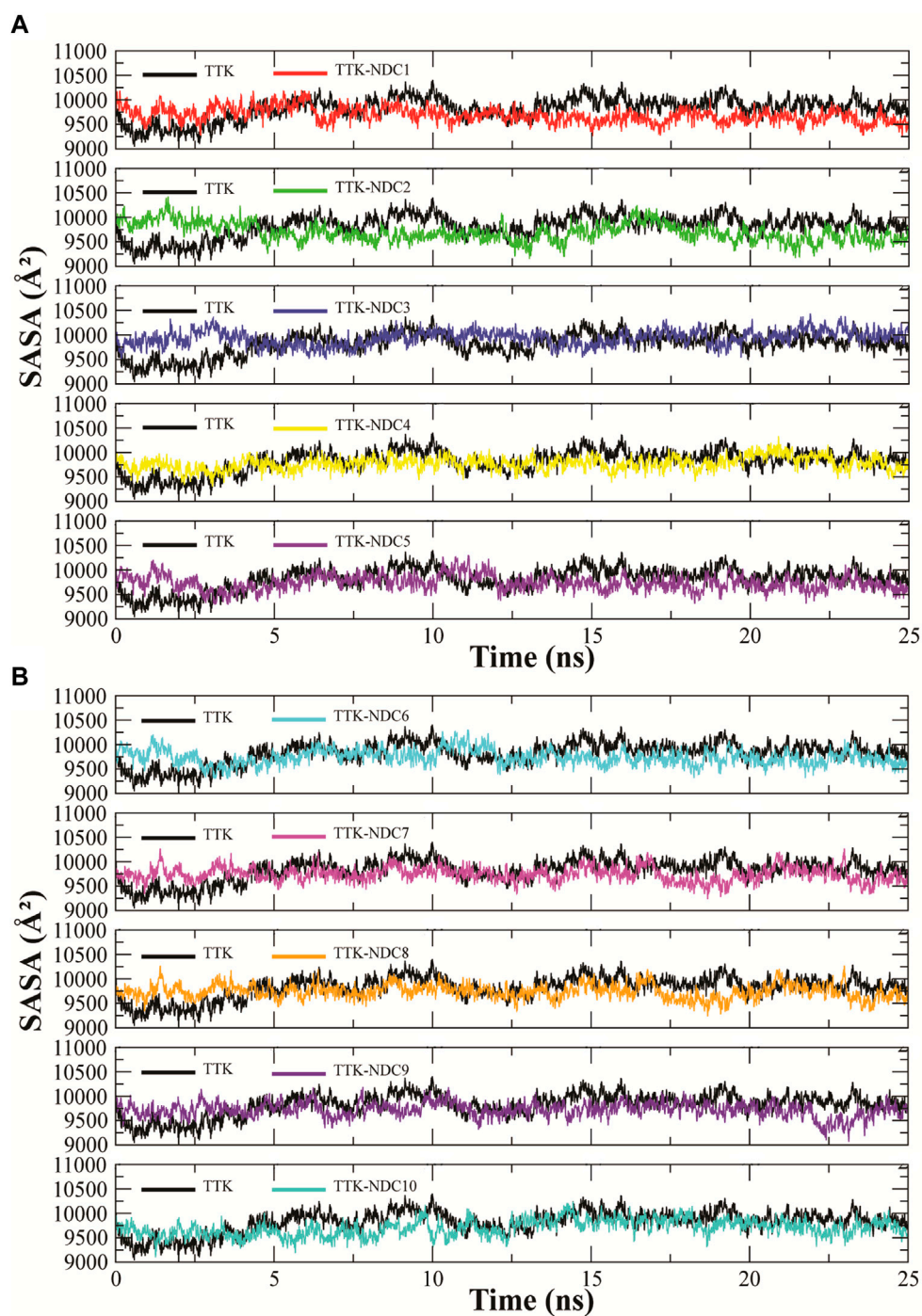
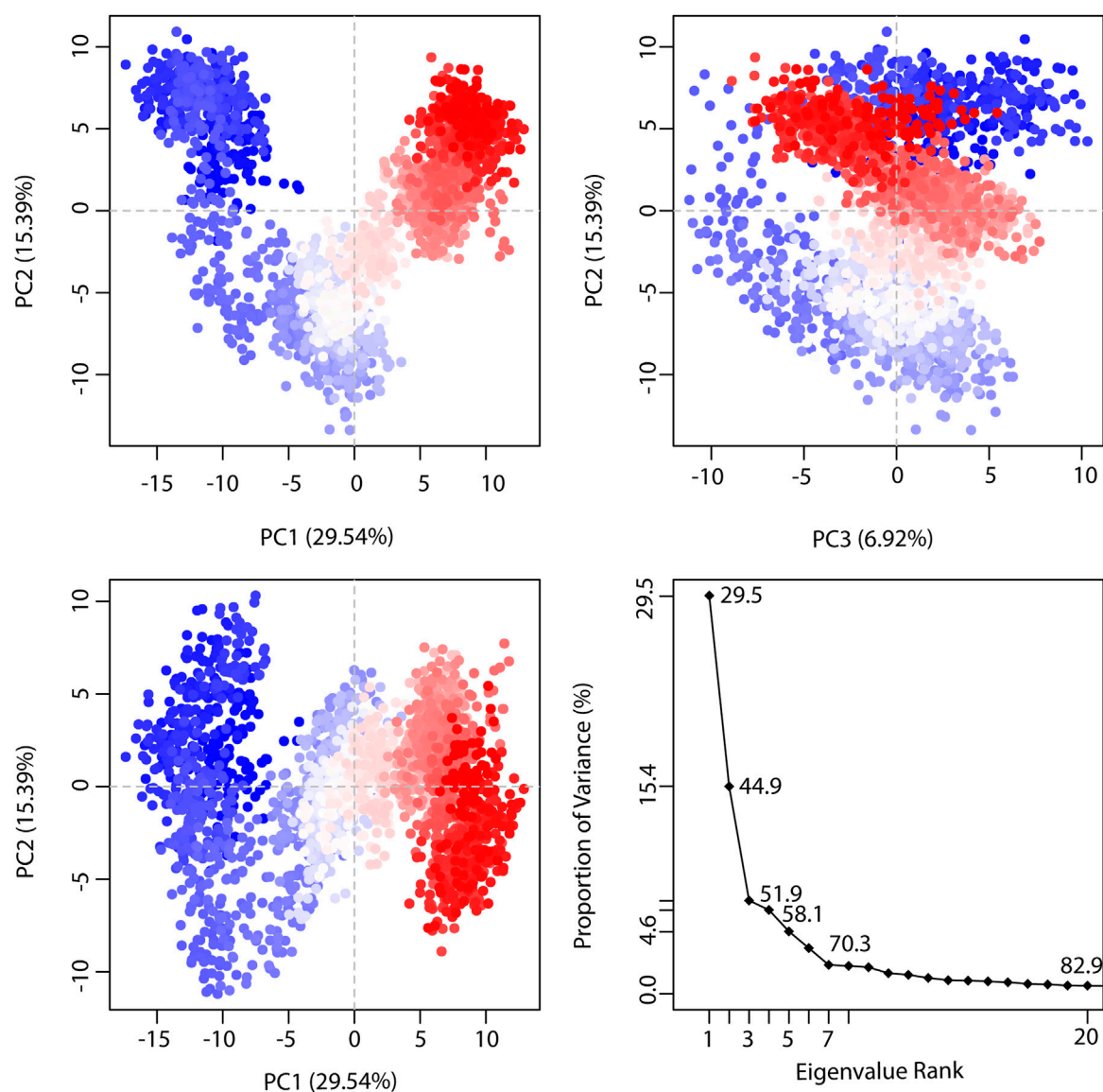


FIGURE 9

(A,B). Solvent accessible surface area (SASA) calculation of TTK protein and its complexes throughout the simulation to find the exposed surface of protein to solvent during simulation.

Figure 4D. After validation of docking method, newly designed compounds were docked in the active site of TTK protein. All compounds showed good binding affinities in terms of glide scores

given in [Supplementary Table S4](#). The binding interactions of the newly designed compounds were analyzed and it was observed that all compounds were making hydrogen bonds with the hinge region

**FIGURE 10**

The representation of proportion of variance % (TTK-NDC1) against eigenvalue calculated by Principal Component Analysis. Three PCs are showing the fluctuating regions. The fluctuations in PC1, PC2, and PC3 are 29.54%, 15.39% and 6.92% respectively. The overall fluctuations are 51.85%.

residues especially Gly110. The other interacting residues were Ile36, Lys58, Glu76, Met107, Cys109, Asn111, Ile112, Ser116, Lys120, Asp169, Met176 and Pro178. In case of hydrogen bonding, NDC1 made two hydrogen bonds with Glu76 and Gly110 with a distance of 1.7 and 2.8 Å respectively. NDC2 made one hydrogen bond with Gly110, while NDC3 was making two hydrogen bonds with Gly110. The same bonding pattern was observed in all complexes i.e., hydrogen bonding with Gly110, Glu76. Moreover, the residues Ile36, Ile112, Met105, Met176 and Pro178 were involved in hydrophobic interactions with the newly designed

compounds. The hydrogen bonds and the distances between the ligands and key residues are given in [Figure 5](#).

MD simulation analysis

MD simulations were carried out to estimate the steady nature and stability of the protein and ligand complexes. The protein-ligand complex stability was estimated by the Root Mean Square Deviation (RMSD) of the complexes in 25 ns long simulation. The RMSD

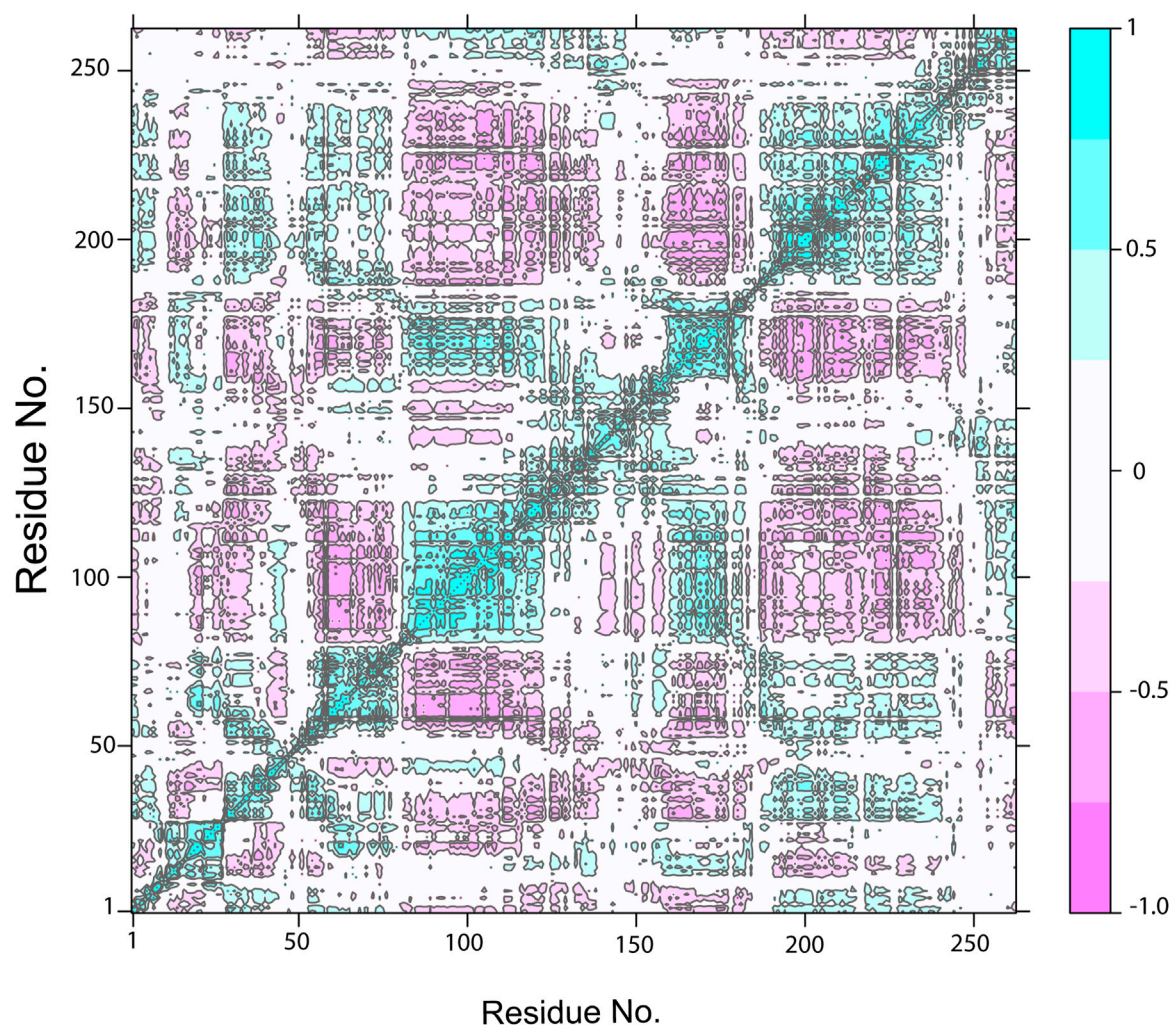


FIGURE 11

The Dynamic Cross-Correlation map of TTK-NDC1 complex. The positive and negative correlation among the residues is shown by cyan and purple color, respectively.

trajectories of ten complexes are shown in [Figures 6A,B](#). The ideal range for the stable complex in terms of RMSD is 2–3 Å. It can be observed that all the complexes showed a stable RMSD value i.e., less than ~2.5 Å as compare to the apo TTK which showed higher deviation in confirmation than complexes. All the complexes equilibrated at ~ 2 ns and then got stability till the end of the simulation. Some complexes showed higher stability than others i.e., TTK-NDC4 and TTK-NDC5 with a RMSD values less than ~1.5 Å. Additionally, the behavior of apo protein was also tested which showed that the protein complexes were more stable than apo protein.

The fluctuations in the amino acid residues were calculated by Root Mean Square Fluctuation (RMSF). The residues with high RMSF values showed higher flexibility, or the residues that form loop regions showed higher RMSF values. Similarly, the residues

with lower RMSF values remained rigid during the simulation. [Figures 7A,B](#) describes the RMSF plots of all complexes. The C and N terminals showed highest RMSF values while the loop regions also showed relatively higher values than the rigid residues. All the complexes showed almost the same trend in RMSF values, with two regions having major fluctuations except for TTK-NDC3 and TTK-NDC6 complex. The major fluctuations were observed in the regions 85 to 115 and 210 to 225 residues. These correspond to the loop regions in the TTK protein while the other residues remained rigid having only a minor fluctuation in RMSF values.

Rg (radius of gyration) is used to show the change in the structure compactness of subjected protein during simulations. The compactness shows that bound small molecules did not induce any conformational in the protein over the simulation time period ([Seeliger and de Groot, 2010](#)). Rg analysis represents how the

TABLE 3 Components of the binding free energies of TTK and designed compounds complexes.

Complexes	ΔE_{vdW}	ΔE_{ele}	EPB	ΔG_{NP}	ΔG_{DIS}	ΔG_{gas}	ΔG_{solv}	ΔG_{bind}
TTK-NDC1	-59.30 ± 0.34	-5.85 ± 0.30	18.54 ± 0.38	-28.92 ± 0.05	56.32 ± 0.20	-65.16 ± 0.46	45.95 ± 0.54	-19.20 ± 0.81
TTK-NDC2	-59.19 ± 0.30	0.39 ± 0.21	16.74 ± 0.33	-30.64 ± 0.13	61.17 ± 0.17	-58.84 ± 0.38	47.27 ± 0.40	-11.57 ± 0.43
TTK-NDC3	-54.15 ± 0.20	-4.09 ± 0.23	18.27 ± 0.24	-27.20 ± 0.04	53.56 ± 0.09	-58.25 ± 0.31	44.63 ± 0.29	-13.61 ± 0.42
TTK-NDC4	-58.18 ± 0.20	-2.38 ± 0.18	16.47 ± 0.27	-28.00 ± 0.03	56.23 ± 0.08	-60.56 ± 0.24	44.70 ± 0.34	-15.86 ± 0.46
TTK-NDC5	-60.40 ± 0.25	-8.55 ± 0.23	30.06 ± 0.42	-28.38 ± 0.04	57.43 ± 0.09	-68.95 ± 0.32	59.10 ± 0.46	-9.85 ± 0.53
TTK-NDC6	-64.34 ± 0.33	-2.05 ± 0.17	27.69 ± 0.45	-33.11 ± 0.06	62.10 ± 0.11	-66.41 ± 0.34	56.68 ± 0.50	-9.73 ± 0.55
TTK-NDC7	-56.79 ± 0.26	-2.57 ± 0.20	14.23 ± 0.24	-27.97 ± 0.04	54.40 ± 0.10	-59.36 ± 0.30	40.65 ± 0.33	-18.71 ± 0.38
TTK-NDC8	-29.34 ± 0.21	-19.18 ± 0.45	48.81 ± 0.51	-39.58 ± 0.09	70.59 ± 0.12	-78.53 ± 0.51	79.82 ± 0.56	1.28 ± 0.40
TTK-NDC9	-78.76 ± 0.28	-5.04 ± 0.26	26.06 ± 0.28	-38.91 ± 0.08	75.29 ± 0.11	-83.80 ± 0.42	63.07 ± 0.35	-20.72 ± 0.55
TTK-NDC10	-72.07 ± 0.34	-5.22 ± 0.18	27.94 ± 0.37	-36.57 ± 0.12	73.83 ± 0.20	-77.30 ± 0.39	65.21 ± 0.47	-12.08 ± 0.41

secondary structures are compactly packed in 3D structure of proteins. The Rg plots of all complexes and apo TTK are shown in Figures 8A,B. TTK-NDC1 (black) had shown highest Rg during ~5–8 ns, with Rg value reaching 16.6 Å, while TTK-NDC2 (red) had shown the second highest value of ~16.6 Å during first 5 ns simulations. The remaining complexes showed the stable Rg values throughout the simulation period. The stable Rg values indicated that the protein remained compact, and less unfolding was observed in all protein-ligand complexes as compare to the apo protein which showed higher Rg values than complexes throughout the simulation period (Grutsch et al., 2014).

SASA (solvent accessible surface area) is the surface area of a biomolecule that is accessible to a solvent. It determines how much an amino acid is exposed to its environment. Lower SASA values represent compact structure of protein while high values represent unfolded structures. SASA values of all TTK protein and its complexes were analyzed to predict the changes in the structure of the protein Figures 9A,B. The Figure shows that the proteins with ligands NDC2, NDC3 and NDC5 have higher SASA values while proteins with rest of the ligands have lowest

SASA values. These results indicate that ligand binding can affect the protein's tertiary structure. Increased values of SASA represent distortion in the structures.

PCA (Principal Component Analysis) characterizes the dynamic behavior of proteins (David and Jacobs, 2014). It helps to identify collective motions of the trajectories during MD simulations. In the graph of TTK-NDC1 (Figure 10), eigenvalues of the proteins were plotted against the corresponding eigenvector index for the first twenty modes of motion. The eigenvalues represent eigenvector fluctuations in hyperspace. In simulations overall movement of the proteins is controlled by eigenvectors with higher eigenvalues. In our systems, the first five eigenvectors exhibited dominant movements with a higher eigenvalue (29.5–70.3%), whereas the remaining eigenvectors had low eigenvalues. The plotted first three PC1, PC2 and PC3 covered the more than 50% of total variations. The Figure 10 plots shows that PC1 clusters possessed highest variability of 29.54%, PC2 depicted the variability of 15.39%, while PC3 exhibited minimal variability which is 6.92%. Minimal variability suggests that PC3 has the most stabilized protein ligand binding and occupies less region in phase

TABLE 4 Predicted physicochemical properties of the newly designed molecules.

Compounds	MW	HBD	HBA	QPlogPo/w	QPlogHERG	QPCaco	QPlogBB	QPlogKhsa
NDC1	320.35	3	5	13.329	-6.193	620.519	-0.945	0.086
NDC2	349.391	2	4	10.346	-6.155	2019.982	-0.427	0.545
NDC3	307.311	5	4	15.776	-5.889	66.578	-1.9	-0.152
NDC4	305.338	4	5	16.231	-6.195	145.681	-1.599	-0.155
NDC5	324.385	5	4	15.72	-5.619	366.227	-1.176	-0.038
NDC6	388.351	4	4	14.646	-5.949	291.298	-1.059	0.217
NDC7	307.354	5	3	14.015	-5.518	189.325	-1.36	0.091
NDC8	504.73	8	8	-0.743	-7.367	1.276	-1.203	-0.007
NDC9	470.573	4	6	15.944	-5.648	401.322	-1.072	0.759
NDC10	482.584	4	6	16.905	-7.032	244.531	-1.612	0.947

"QPlogPo/w recommended range = "-2.0 to 6.5," QPlogHERG recommended range = "<-5," QPCaco2 recommended range = "<25 poor," "> 500 great," QPlogBB recommended range = "-3.0 to 1.2," QPlogKhsa recommended range = "-1.5 to 1.5."

space; hence its structure is compact as compare to PC1 and PC2. Through simple clustering in PC subspace, the PCA analysis revealed conformational changes in all clusters, blue regions showed most significant movement, white regions show intermediate movement while red regions show that there is less movement of flexibility. The PCA plots of remaining complexes are given in [Supplementary Figures S1A–S1I](#).

The cross-correlation map showed the pairwise correlation of NDC1 with the TTK protein by the value of pairwise cross-correlation coefficient ([Figure 11](#)). The correlated residues are more than 0.8 and are shown in cyan color, while the anti-correlated residues (<-0.4) are indicated with magenta color. The high percentage of pairwise correlated residues indicated the stable binding of the ligand with the TTK protein. The cross-correlation maps of other complexes are given in [Supplementary Figures S2A–S2I](#).

The binding free energy estimation

The MM/PBSA is a significant method to estimate the binding free energy of protein-ligand complexes. The ΔG_{bind} values for all complexes were estimated using this method. The ΔG is the outcome of contribution of various protein-ligand interactions such as van der Waals energy (ΔE_{vdW}), electrostatic energy (ΔE_{ele}) and EPB (electrostatic contribution to solvation free energy by Poisson-Boltzmann) energy. The ΔE_{vdW} of NDC9 and NDC10 complexes are found to be -78.76 kcal/mol and -72.07 kcal/mol respectively and contributing more in binding affinities as compared to other designed compounds. Whereas NDC8 complex having -29.34 kcal/mol showing its limited contribution whereas in remaining complexes it contributed more. In case of ΔE_{ele} contribution, the energy component is -5.85 kcal/mol in NDC1, 0.39 kcal/mol in NDC2, -4.09 kcal/mol in NDC3, -2.38 kcal/mol in NDC4, -8.55 kcal/mol in NDC5, -2.05 kcal/mol in NDC6, -2.57 kcal/mole in NDC7, -19.18 kcal/mol in NDC8, -5.04 kcal/mol in NDC9, and -5.22 kcal/mol in NDC10. ΔE_{ele} energy contribution of NDC8 complex is highest among all other complexes. Moreover, the PB contribution of all complexes are showing that NDC8 has higher PB value than other complexes. The calculated binding free energies of all the complexes are shown in [Table 3](#). The binding free energies ΔG_{bind} of NDC1 (-19.20 kcal/mol), NDC7 (-18.71 kcal/mol) and NDC9 (-20.72 kcal/mol) are quite better than other complexes. The differences in the binding energies are due to the difference in the contribution of electrostatic, polar, and non-polar energies in the protein-ligand complexes.

Calculations of physicochemical properties

QikProp software was used to estimate the physicochemical parameters ([Table 4](#)). With almost one

rule violation, the majority of newly created molecules followed the Lipinski's rule. The predicted octanol/water partition coefficient 'QPlogPo/w' values range (10.346 – 16.905), HERG K⁺ channels "QPlogHERG" blocking IC₅₀ values range (-7.032 to -5.518), caco-2 cell permeability "QPPCaco" values range (66.578 – 2019.982), brain/blood partition coefficient "QPlogBB" values range (-1.9 to -0.427), and human serum albumin binding "QPkhsa" values range (-0.155 – 0.947) are within the acceptable ranges for 95 percent oral drugs ([Kumar et al., 2016](#)). Within the recommended ranges, physicochemical qualities such as 'QPlogPo/w and QPlogHERG' showed smooth diffusion of drug and protection against unexpected cardiac arrest ([Kumar et al., 2016](#)).

Conclusion

TTK is an important mitotic kinase whose loss of function results in chromosomal segregation defects that can lead to aneuploidy and cell death, making it an attractive drug target for cancer. Using different partial charges and alignment methods, structure-based 3D-QSAR models on MMFF94 charges yielded best CoMFA and CoMSIA models. Using these predictive models and contour maps information ten new compounds were designed and their biological activities were predicted. The newly designed compounds showed better predicted activities than their parent compounds, demonstrating that structure-based approaches using MMFF94 charges can be used to design better active TTK inhibitors. Further MD simulations described the stability of protein-ligand complexes. Similarly computational binding free energy calculations suggest that newly designed compounds can bind to TTK protein with better binding affinity than reported compounds.

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

NA, NY, and AF retrived data and performed experiment. AA, MA, MS and MM contributed to conception and design of the study. MA performed the statistical analysis. NA, NY, AF, MM wrote the first draft of the manuscript. MA, AA, and MS, proof read all sections of the manuscript.

Acknowledgments

Authors are thankful to Higher Education Commission of Pakistan for providing funds vide project no. 8094/Balochistan/NRPU/R&D/HEC/2017 to purchase software. We also thank the CERTARA company for providing us the evaluation version of SYBYL-X software.

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

- Acun, B., Hardy, D. J., Kale, L. V., Li, K., Phillips, J. C., and Stone, J. E. (2018). Scalable molecular dynamics with NAMD on the Summit system. *IBM J. Res. Dev.* 62 (4), 1–9. doi:10.1147/jrd.2018.2888986
- Balasubramanian, P. K., Balupuri, A., and Cho, S. J. (2014). A CoMFA study of phenoxypyridine-based JNK3 inhibitors using various partial charge schemes. *J. Chosun Nat. Sci.* 7, 45–49. doi:10.13160/ricns.2014.7.1.45
- Bhachoo, J., and Beuming, T. J. M. P.-P. I. (2017). Investigating protein–peptide interactions using the Schrödinger computational suite. *Methods Mol. Biol.* 1561, 235–254. doi:10.1007/978-1-4939-6798-8_14
- Brough, R., Frankum, J. R., Sims, D., Mackay, A., Mendes-Pereira, A. M., Bajrami, I., et al. (2011). Functional viability profiles of breast cancer. *Cancer Discov.* 1, 260–273. doi:10.1158/2159-8290.CD-11-0107
- Bursavich, M. G., Dastrup, D., Shenderovich, M., Yager, K. M., Cimbor, D. M., Williams, B., et al. (2013). Novel Mps1 kinase inhibitors: From purine to pyrrolopyrimidine and quinazoline leads. *Bioorg. Med. Chem. Lett.* 23, 6829–6833. doi:10.1016/j.bmcl.2013.10.008
- Case, D. A., Aktulga, H. M., Belfon, K., Ben-Shalom, I., Brozell, S. R., Cerutti, D., et al. (2021). Amber 2021: Reference manual. *Covers Amber20 AmberTools21*. doi:10.13140/RG.2.2.15902.66881
- Chen, Y., Yu, W., Jiang, C.-C., and Zheng, J.-G. J. M. (2018). Insights into resistance mechanisms of inhibitors to Mps1 C604Y mutation via a comprehensive molecular modeling study. *Molecules* 23, 1488. doi:10.3390/molecules23061488
- Daniel, J., Coulter, J., Woo, J.-H., Wilsbach, K., and Gabrielson, E. J. P. O. T. N. A. O. S. (2011). High levels of the Mps1 checkpoint protein are protective of aneuploidy in breast cancer cells. *Proc. Natl. Acad. Sci. U. S. A.* 108, 5384–5389. doi:10.1073/pnas.1007645108
- David, C. C., and Jacobs, D. J. (2014). Principal component analysis: A method for determining the essential dynamics of proteins. *Methods Mol. Biol.* 1084, 193–226. doi:10.1007/978-1-62703-658-0_11
- Duan, Y., Wu, C., Chowdhury, S., Lee, M. C., Xiong, G., Zhang, W., et al. (2003). A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *J. Comput. Chem.* 24, 1999–2012. doi:10.1002/jcc.10349
- Fisk, H. A., and Winey, M. (2001). The mouse Mps1p-like kinase regulates centrosome duplication. *Cell* 106, 95–104. doi:10.1016/s0092-8674(01)00411-1
- Ghosh, S., Keretsu, S., and Cho, S. J. (2021). Designing of the N-ethyl-4-(pyridin-4-yl) benzamide based potent ROCK1 inhibitors using docking, molecular dynamics, and 3D-QSAR. *PeerJ* 9, e11951. doi:10.7717/peerj.11951
- Grutsch, S., Fuchs, J. E., Freier, R., Kofler, S., Bibi, M., Asam, C., et al. (2014). Ligand binding modulates the structural dynamics and compactness of the major birch pollen allergen. *Biophys. J.* 107, 2972–2981. doi:10.1016/j.bpj.2014.10.062
- Hu, R., Barbault, F., Delamar, M., and Zhang, R. (2009). Receptor- and ligand-based 3D-QSAR study for a series of non-nucleoside HIV-1 reverse transcriptase inhibitors. *Bioorg. Med. Chem.* 17, 2400–2409. doi:10.1016/j.bmc.2009.02.003
- Huang, M., Huang, Y., Guo, J., Yu, L., Chang, Y., Wang, X., et al. (2021). Pyrido [2, 3-d] pyrimidin-7 (8H)-ones as new selective orally bioavailable Threonine Tyrosine Kinase (TTK) inhibitors. *Eur. J. Med. Chem.* 211, 113023. doi:10.1016/j.ejmech.2020.113023
- Koç, E., Üngördü, A., and Candan, F. J. S. C. (2021). Antioxidant activities of *Alyssum virgatum* plant and its main components. *Struct. Chem.* 33, 267–279. doi:10.1007/s11224-021-01856-1
- Kumar, A., Ito, A., Hirohama, M., Yoshida, M., and Zhang, K. Y. (2016). Identification of new SUMO activating enzyme 1 inhibitors using virtual screening and scaffold hopping. *Bioorg. Med. Chem. Lett.* 26, 1218–1223. doi:10.1016/j.bmcl.2016.01.030
- Kusakabe, K.-I., Ide, N., Daigo, Y., Itoh, T., Higashino, K., Okano, Y., et al. (2012). Diaminopyridine-based potent and selective Mps1 kinase inhibitors binding to an unusual flipped-peptide conformation. *ACS Med. Chem. Lett.* 3, 560–564. doi:10.1021/ml3000879
- Laufer, R., Ng, G., Liu, Y., Patel, N. K. B., Edwards, L. G., Lang, Y., et al. (2014). Discovery of inhibitors of the mitotic kinase TTK based on N-(3-(3-sulfamoylphenyl)-1H-indazol-5-yl)-acetamides and carboxamides. *Bioorg. Med. Chem.* 22, 4968–4997. doi:10.1016/j.bmc.2014.06.027
- Li, S., Fan, J., Peng, C., Chang, Y., Guo, L., Hou, J., et al. (2017). New molecular insights into the tyrosyl-tRNA synthase inhibitors: CoMFA, CoMSIA analyses and molecular docking studies. *Sci. Rep.* 7, 1–13. doi:10.1038/s41598-017-10618-1
- Lindberg, M. F., and Meijer, L. (2021). Dual-specificity, tyrosine phosphorylation-regulated kinases (DYRKs) and cdc2-like kinases (CLKs) in human disease, an overview. *Int. J. Mol. Sci.* 22, 6047. doi:10.3390/ijms22116047
- Liu, X., Liao, W., Yuan, Q., Ou, Y., and Huang, J. (2015a). TTK activates Akt and promotes proliferation and migration of hepatocellular carcinoma cells. *Oncotarget* 6, 34309–34320. doi:10.18632/oncotarget.5295
- Liu, X., and Winey, M. (2012). The MPS1 family of protein kinases. *Annu. Rev. Biochem.* 81, 561–585. doi:10.1146/annurev-biochem-061611-090435
- Liu, Y., Lang, Y., Patel, N. K., Ng, G., Laufer, R., Li, S.-W., et al. (2015b). The discovery of orally bioavailable tyrosine threonine kinase (TTK) inhibitors: 3-(4-(heterocyclyl) phenyl)-1 H-indazole-5-carboxamides as anticancer agents. *J. Med. Chem.* 58, 3366–3392. doi:10.1021/jm501740a
- Lorca, M., Valdes, Y., Chung, H., Pessoa-Mahana, C. D., and Mella, J. (2018). 3D-QSAR on a series of pyrimidinyl-piperazine-carboxamides based Fatty Acid Amide Hydrolase (FAAH) inhibitors as a useful tool to obtain novel endocannabinoid enhancers, Preprints 2018.
- Lu, N., and Ren, L. J. B. (2021). TTK (threonine tyrosine kinase) regulates the malignant behaviors of cancer cells and is regulated by microRNA-582-5p in ovarian cancer. *Bioengineered* 12, 5759–5768. doi:10.1080/21655979.2021.1968778
- Maia, A. R., De Man, J., Boon, U., Janssen, A., Song, J.-Y., Omerzu, M., et al. (2015). Inhibition of the spindle assembly checkpoint kinase TTK enhances the efficacy of docetaxel in a triple-negative breast cancer model. *Ann. Oncol.* 26, 2180–2192. doi:10.1093/annonc/mdv293

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/fchem.2022.1003816/full#supplementary-material>

- Muddassar, M., Jang, J. W., Hong, S. K., Cho, Y. S., Kim, E. E., Keum, K. C., et al. (2010). Identification of novel antitubercular compounds through hybrid virtual screening approach. *Bioorg. Med. Chem.* 18, 6914–6921. doi:10.1016/j.bmc.2010.07.010
- Muddassar, M., Pasha, F., Neaz, M., Saleem, Y., and Cho, S. (2009). Elucidation of binding mode and three dimensional quantitative structure–activity relationship studies of a novel series of protein kinase B/Akt inhibitors. *J. Mol. Model.* 15, 183–192. doi:10.1007/s00894-008-0416-7
- Naud, S., Westwood, I. M., Faisal, A., Sheldrake, P., Bavetsias, V., Atrash, B., et al. (2013). Structure-based design of orally bioavailable 1H-Pyrrolo[3, 2-c]pyridine inhibitors of mitotic kinase monopolar spindle 1 (MPS1). *J. Med. Chem.* 56, 10045–10065. doi:10.1021/jm401395s
- Puzyn, T., Mostrag-Szlichtyng, A., Gajewicz, A., Skrzyński, M., and Worth, A. P. (2011). Investigating the influence of data splitting on the predictive ability of QSAR/QSPR models. *Struct. Chem.* 22, 795–804. doi:10.1007/s11224-011-9757-4
- Roe, D. R., and Cheatham, T. E. J. O. C. T., III (2013). PTRAJ and CPPTRAJ: Software for processing and analysis of molecular dynamics trajectory data. *J. Chem. Theory Comput.* 9, 3084–3095. doi:10.1021/ct400341p
- Sainy, J., and Sharma, R. (2015). QSAR analysis of thiolactone derivatives using HQSAR, CoMFA and CoMSIA. *SAR QSAR Environ. Res.* 26, 873–892. doi:10.1080/1062936x.2015.1095238
- Schmidt, M., Budirahardja, Y., Klompaker, R., and Medema, R. H. J. E. R. (2005). Ablation of the spindle assembly checkpoint by a compound targeting Mps1. *EMBO Rep.* 6, 866–872. doi:10.1038/sj.embor.7400483
- Seeliger, D., and De Groot, B. L. (2010). Conformational transitions upon ligand binding: Holo-structure prediction from apo conformations. *PLoS Comput. Biol.* 6, e1000634. doi:10.1371/journal.pcbi.1000634
- Shiri, F., Rakhshani-Morad, S., Samzadeh-Kermani, A., and Karimi, P. (2016). Computer-aided molecular design of some indolinone derivatives of PLK4 inhibitors as novel anti-proliferative agents. *Med. Chem. Res.* 25, 2643–2665. doi:10.1007/s00044-016-1638-3
- Stucke, V. M., Silljé, H. H., Arnaud, L., and Nigg, E. A. (2002). Human Mps1 kinase is required for the spindle assembly checkpoint but not for centrosome duplication. *EMBO J.* 21, 1723–1732. doi:10.1093/emboj/21.7.1723
- Sugimoto, Y., Sawant, D. B., Fisk, H. A., Mao, L., Li, C., Chettiar, S., et al. (2017b). Novel pyrrolopyrimidines as Mps1/TTK kinase inhibitors for breast cancer. *Bioorg. Med. Chem.* 25, 2156–2166. doi:10.1016/j.bmc.2017.02.030
- Sugimoto, Y., Sawant, D. B., Fisk, H. A., Mao, L., Li, C., Chettiar, S., et al. (2017a). Novel pyrrolopyrimidines as Mps1/TTK kinase inhibitors for breast cancer. *Bioorg. Med. Chem.* 25, 2156–2166. doi:10.1016/j.bmc.2017.02.030
- Sun, H., Li, Y., Shen, M., Tian, S., Xu, L., Pan, P., et al. (2014). Assessing the performance of MM/PBSA and MM/GBSA methods. 5. Improved docking performance using high solute dielectric constant MM/GBSA and MM/PBSA rescoring. *Phys. Chem. Chem. Phys.* 16, 22035–22045. doi:10.1039/c4cp03179b
- Tahir, A., Alharthy, R. D., Naseem, S., Mahmood, N., Ahmed, M., Shahzad, K., et al. (2018). Investigations of structural requirements for brd4 inhibitors through ligand- and structure-based 3D QSAR approaches. *Molecules* 23, 1527. doi:10.3390/molecules23071527
- Vaidya, A., Jain, A. K., Kumar, B. P., Sastry, G., Kashaw, S. K., and Agrawal, R. K. (2017). CoMFA, CoMSIA, kNN MFA and docking studies of 1, 2, 4-oxadiazole derivatives as potent caspase-3 activators. *Arabian J. Chem.* 10, S3936–S3946. doi:10.1016/j.arabjc.2014.05.034
- Wang, F., Yang, W., Shi, Y., and Le, G. (2015). Structural analysis of selective agonists of thyroid hormone receptor β using 3D-QSAR and molecular docking. *J. Taiwan Inst. Chem. Eng.* 49, 1–18. doi:10.1016/j.jtice.2014.11.009
- Wang, W., Yang, Y., Gao, Y., Xu, Q., Wang, F., Zhu, S., et al. (2009). Structural and mechanistic insights into Mps1 kinase activation. *J. Cell. Mol. Med.* 13, 1679–1694. doi:10.1111/j.1582-4934.2008.00605.x
- Wei, J.-H., Chou, Y.-F., Ou, Y.-H., Yeh, Y.-H., Tyan, S.-W., Sun, T.-P., et al. (2005). TTK/hMps1 participates in the regulation of DNA damage checkpoint response by phosphorylating CHK2 on threonine 68. *J. Biol. Chem.* 280, 7748–7757. doi:10.1074/jbc.m410152200
- Wengner, A. M., Siemeister, G., Koppitz, M., Schulze, V., Kosemund, D., Klar, U., et al. (2016). Novel Mps1 kinase inhibitors with potent antitumor activity. *Mol. cancer Ther.* 15, 583–592. doi:10.1158/1535-7163.mct-15-0500
- Xing, C., Zhou, X., Chen, C., Sun, W., Zheng, Q., and Liang, D. (2021). Studies of interaction mechanism between pyrido [3, 4-d] pyrimidine inhibitors and Mps1. *Molecules* 26, 5075. doi:10.3390/molecules26165075



OPEN ACCESS

EDITED BY
Sibhghatulla Shaikh,
Yeungnam University, South Korea

REVIEWED BY
Shakir Khan,
Harvard Medical School, United States
Yusuf Sert,
Bozok University, Turkey

*CORRESPONDENCE
Hulya Ellidokuz,
hulya.ellidokuz@deu.edu.tr

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

RECEIVED 16 September 2022
ACCEPTED 22 November 2022
PUBLISHED 06 December 2022

CITATION
Mert-Ozupek N, Calibasi-Kocal G,
Olgun N, Basbinar Y, Cavas L and
Ellidokuz H (2022), *In-silico* molecular
interactions among the secondary
metabolites of *Caulerpa* spp. and
colorectal cancer targets.
Front. Chem. 10:1046313.
doi: 10.3389/fchem.2022.1046313

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

In-silico molecular interactions among the secondary metabolites of *Caulerpa* spp. and colorectal cancer targets

Nazli Mert-Ozupek¹, Gizem Calibasi-Kocal², Nur Olgun³,
Yasemin Basbinar², Levent Cavas⁴ and Hulya Ellidokuz^{5*}

¹Department of Basic Oncology, Institute of Health Sciences, Dokuz Eylül University, İzmir, Türkiye,
²Department of Translational Oncology, Institute of Oncology, Dokuz Eylül University, İzmir, Türkiye,
³Department of Pediatric Oncology, Institute of Oncology, Dokuz Eylül University, İzmir, Türkiye,
⁴Department of Chemistry, Faculty of Sciences, Dokuz Eylül University, İzmir, Türkiye, ⁵Department of Preventive Oncology, Institute of Oncology, Dokuz Eylül University, İzmir, Türkiye

Caulerpa spp. secrete more than thirty different bioactive chemicals which have already been used in cancer treatment research since they play a pivotal role in cancer metabolism. Colorectal cancer is one of the most common cancer types, thus using novel and effective chemicals for colorectal cancer treatment is crucial. In the cheminformatics pipeline of this study, ADME-Tox and drug-likeness tests were performed for filtering the secondary metabolites of *Caulerpa* spp. The ligands which were selected from the ADME test were used for *in silico* molecular docking studies against the enzymes of the oxidative branch of the pentose phosphate pathway (glucose-6-phosphate dehydrogenase and 6-phosphogluconate dehydrogenase), which is of great importance for colorectal cancer, by using AutoDock Vina. Pharmacophore modeling was carried out to align the molecules. Molecular dynamic simulations were performed for each target to validate the molecular docking studies and binding free energies were calculated. According to the ADME test results, 13 different secondary metabolites were selected as potential ligands. Molecular docking studies revealed that vina scores of caulerpin and monomethyl caulerpinate for G6PDH were found as -10.6 kcal mol⁻¹, -10.5 kcal mol⁻¹, respectively. Also, the vina score of caulersin for 6PGD was found as -10.7 kcal mol⁻¹. The highest and the lowest binding free energies were calculated for monomethyl caulerpinate and caulersin, respectively. This *in silico* study showed that caulerpin, monomethyl caulerpinate, and caulersin could be evaluated as promising marine phytochemicals against pentose phosphate pathway enzymes and further studies are recommended to investigate the detailed activity of these secondary metabolites on these targets.

KEYWORDS

Caulerpa, colorectal cancer, *in silico*, pentose phosphate pathway, oncoinformatics

1 Introduction

Caulerpa, a green siphonous macroalgae, belongs to the Caulerpaceae family with 97 species. It was first described by J V Lamouroux in 1809 and derived from the Greek words caulos (stalk/stem) and erpos (creep) (Mehra et al., 2019). Especially, *C. taxifolia* and *C. cylindracea* (previously known as *C. racemosa* var. *cylindracea*) have attracted attention for the last 30 years due to their invasive properties (Montefalcone et al., 2015) and, more importantly, the various properties of bioactive (especially secondary metabolites) chemicals for defense, communication, growth and development regulation, reproduction, competition, and infection (Erb and Kliebenstein, 2020; Ramawat and Goyal, 2020). *Caulerpa* secrete linear or monocyclic terpenoids that have aldehyde and enol-acetate functional groups (Mehra et al., 2019). The structure of caulerpin ($C_{24}H_{18}N_2O_4$), a secondary metabolite and pigment of the *C. cylindracea* species, was first described by Aguilar-Santos in 1970 (Aguilar-Santos, 1970). Its molecular weight is 398.418 g/mol and its characterized structure is 'dimethyl-6,13-dihydrodibenzo [b,i] phenazin-5,12-dicarboxylate methyl ester. Caulersin ($C_{21}H_{14}N_2O_3$) is another bis-indole alkaloid which is isolated from *C. serrulata* (Su et al., 1997) and from *C. racemosa* (Yang et al., 2014). Its isomers are caulersin A, B, and C. The molecular weight of caulersin is 342.1004 g/mol (Su et al., 1997). It is characterized by its "central troponoid bridging" bisindole structure (Su et al., 1997). Caulerpenyne is a sesquiterpenoid-structured secondary metabolite which has some bioactivities such as antiproliferative and apoptotic activities (Cavas et al., 2006) and inhibitors of lipoxygenase (Cengiz et al., 2011) and 5-lipoxygenase (Richter et al., 2014), etc. Secondary metabolites of genus *Caulerpa* are responsible for complex modulation network induced in AMPK, ER Stress, mitochondrial stress, PTP1B inhibition and cell cycle stop pathways, metabolic reprogramming in cancer cells, apoptosis and cell cycle arrest in cancer metabolism (Mehra et al., 2019).

Colorectal cancer (CRC) is the second and third most common diagnosed cancer type in women and men in the world, respectively (Dekker et al., 2019). According to the World Health Organization data, in 2018, CRC caused 1.80 million new cases and 862,000 deaths all around the world (WHO, 2018). Since CRC is a common and fatal cancer type, using novel and effective chemicals for treatment is essential.

The pentose phosphate pathway (PPP) is the pivotal pathway for ribonucleotide synthesis and is the main source of NADPH (the reduced form of Nicotinamide Adenine Dinucleotide Phosphate) (Stincone et al., 2015), which is of great importance for fatty acid synthesis and reactive oxygen species scavenging. PPP branches from glycolysis, and it plays a key role in cancer cells (Patra and Hay, 2014). The oxidative phase of PPP is initiated with hexokinase, which converts glucose to glucose 6-phosphate

(G6P). G6P oxidizes (dehydrogenated) to 6-phosphogluconolactone by the rate-limiting enzyme (glucose 6-phosphate dehydrogenase (G6PDH)) to yield NADPH (by reducing $NADP^+$) (Stincone et al., 2015). The other NADPH source of PPP is the conversion of 6-phosphogluconate into ribose (ribulose) 5-phosphate by 6-phosphogluconate dehydrogenase (6PGD). Eventually, the regeneration of GSH, synthesis of DNA, fatty acids, and sterols is achieved by producing 2 mol of NADPH per mole of G6P entering the oxidative phase of PPP (Patra and Hay, 2014). In many solid tumors, overexpression of PPP (especially the enzyme 6PGD) has already been observed (Patra and Hay, 2014; Jin and Zhou, 2019). Furthermore, targeting oxidative phase of PPP for mutant *KRAS* colorectal carcinomas prevents the recurrence (Gao et al., 2019). Thus, targeting the PPP is a potentially new target for CRC treatment.

In-silico computer-aided methods are commonly used to predict and elucidate the molecular-level behavior of a compound (Dege et al., 2022). Molecular docking is a convenient *in-silico* method which can be used to evaluate the binding affinity of the ligand on the receptor and can predict the position of the these molecules (Trott and Olson, 2010; Mert Ozupek and Cavas, 2017). ADME provides information about *in-silico* ADME behavior which is important for medicinal chemistry (Bocci et al., 2017; Dege et al., 2022; Gokce et al., 2022; Pantaleão et al., 2022). Drug-likeness analysis using *in-silico* is of great importance for evaluating the pharmacokinetic features of fast and cheap (Gokce et al., 2022; Hasan et al., 2022).

In this study, the anticancer activity of phytochemicals of *Caulerpa* spp. were tested on potential targets (G6PDH and 6PGD) against CRC by using *in-silico* pharmacokinetic and pharmacodynamic tools. The aim of this study was to investigate the potential *Caulerpa*-based phytochemicals against fundamental targets (G6PDH and 6PGD) for colorectal cancer treatment.

2 Materials and methods

2.1 Ligand preparation

The secondary metabolites found in *Caulerpa* spp. Were selected from the literature and organism-specific natural product lists of PubChem; Lotus-the natural products occurrence database (<https://pubchem.ncbi.nlm.nih.gov>). The three-dimensional (3D) structures of chemicals found in *Caulerpa* spp. were extracted from the PubChem Database. The Canonical smiles formats of the secondary metabolites were drawn using ACD/ChemSketch software. To optimize the geometry and minimize the energy for the selected ligands (secondary metabolites), Open Babel (O'Boyle et al., 2011) minimization tool was used. As force field, uff (universal force

field) was selected. Conjugate gradients were selected as optimization algorithm, and total number of steps was set as 200.

2.2 Protein preparation

The proteins used in the molecular modeling studies were retrieved from the RCSB Protein Data Bank (<https://www.rcsb.org/>). The crystal structure of receptors against human colorectal cancer G6PDH (PDBID: 6E08; resolution: 1.90 Å) and 6PGD (PDBID: 4GWK; resolution: 1.534 Å) were extracted. The crystal structures were rebuilt and both water and small molecules were removed. To perform energy minimization and geometry optimization, polar hydrogens were added, and non-polar hydrogens were merged into the molecules by using AutoDock Tools-1.5.6. Before the docking studies, Kollman charges were added, and the related receptors were saved in the PDBQT format.

2.3 *In silico* analysis of pharmacokinetic ADME, drug-likeness and toxicity test

The drug-likeness of the compounds found in *Caulerpa* spp. was calculated using SwissADME (<http://www.swissadme.ch/>) (Daina et al., 2017). The molecular structures of compounds were converted into SMILES format. Only the ligands that could Lipinski's five rule variations (calculated Log P (CLog P) should be less than five, polar surface area, the number of hydrogen bond donors should be less than five, hydrogen acceptors should be less than ten and the molecular weight should be less than 500) with no more than one violation were used for molecular docking experiments. Toxicity Estimation Software Tool (T.E.S.T.) (Martin et al., 2008) and ProTox-II (http://tox.charite.de/protoc_II/; Banerjee et al., 2018) server were used to determine the toxicity estimation of secondary metabolites of *Caulerpa* which were selected from ADME results.

2.4 Molecular docking studies

To examine the selected ligands on related receptors, molecular docking experiments were carried out with AutoDock Vina. After the minimization process, the grid box resolution was set at 29.7030, 17.7197, and 29.5355 along the x, y, and z points, respectively, for G6PDH (PDBID: 6E08). To define the binding site for conducting the docking for 6PGD (PDBID: 4GWK), grid box resolution was set at 21.6405, 23.5892, and -2.9280 along the x, y, and z points, respectively. The grid dimensions of all receptors were adjusted to 25 × 25 × 25 for all molecular docking studies. DHEA and 6aminonicotinamide (6ANA) were used as control ligands.

The results of the graphical representations were prepared via Maestro Schrödinger.

2.5 Pharmacophore model generation and pharmacophore screening

The compounds of *Caulerpa* spp. that showed binding affinities less than -10 kcal/mol (threshold value), were analyzed for pharmacophoric features using the PharmaGist web server (Schneidman-Duhovny et al., 2008). For further studies, ZINCPharmer server (Koes and Camacho, 2012) was used to visualize the best pairwise alignment of ligands (compounds from *Caulerpa* spp.) with the pivot molecule (DHEA or 6ANA). Scores were calculated for each pharmacophore feature by PharmaGist server.

2.6 Molecular dynamic simulation for docking validation

The molecular docking simulations of caulerpin, monomethyl caulerpinate and caulersin with glucose-6-phosphate dehydrogenase (G6PDH) and 6-phosphogluconate dehydrogenase (6PGD) proteins were performed using a web-based MD simulation package WebGRO for Macromolecular Simulations (<https://simlab.uams.edu/>) Simlab, the University of Arkansas for Medical Sciences (UAMS), Little Rock, United States provided by GROMACS-2019.2 (Abraham et al., 2015). PRODRG server (Schüttelkopf and Van Aalten, 2004) was used for the generation of the ligand topology files. GROMOS96 43a1 force field was used for the approximation of the protein-ligand (G6PDH-caulerpin; G6PDH-monomethylcaulerpinate; 6PGD-caulersin) interactions. The triclinic box was filled with SPC water and 0.15 M NaCl (counter ions) to neutralize the system for each ligand-protein complex. The equilibration type was NVT/NPT and Parrinello-Danadio-Bussi thermostat and Parrinello-Rahmanbarostat were used to control the temperature (300 K) and the pressure (atmospheric pressure-1 bar). 5,000 steepest descent was used to minimize the energy of the system. Each protein-ligand complex (G6PDH-caulerpin; G6PDH-monomethyl caulerpinate; 6PGD-caulersin) was simulated for 100 ns. H bonds, the Radius of gyration (Rg), Root Mean Square Deviation (RMSD) and SASA were tested to estimate the complex stability.

2.7 Calculation of binding free energy by MM/PB(GB)SA

The best docking poses for each ligand (caulerpin, monomethyl caulerpinate, and caulersin) were rescored. In this study, the binding free energy of the ligands was

TABLE 1 List of pharmacokinetic properties of 36 compounds from *Caulerpa* spp.

Properties	Physicochemical properties								Lipophilicity	Water solubility	Pharmacokinetics	Drug-likeness	Medicinal chemistry
Parameters	Molecular weight (g/mol)	Number of heavy atoms	Number of aromatic heavy atoms	Number of rotatable bonds	Number of H-bond acceptors	Number of H-bond donors	Molar refractivity	TPSA (Å)	Log P _{0/w}	LogS (ESOL)	GI absorption	Lipinski/violation	Synthetic accessibility
Compound													
Caulerpin	398.41	30	22	4	4	2	116.54	84.18	2.98	-5.30	High	Yes/0	2.32
Caulerpenyne	374.43	27	0	10	6	0	103.18	78.90	3.79	-4.19	High	Yes/0	4.69
Caulersin	342.35	26	21	2	3	2	102.72	74.95	2.61	-4.97	High	Yes/0	2.34
10,11-epoxycaulerpenyne	390.43	28	0	10	7	0	102.67	91.43	3.95	-3.30	High	Yes/0	5.22
Flexilin	320.42	23	0	11	4	0	94.12	52.60	3.69	-4.24	High	Yes/0	3.90
Trans-phytol	296.53	21	0	13	1	1	98.94	20.23	4.71	-5.98	Low	Yes/1	4.30
Alpha tocopherol quinone	446.71	32	0	15	3	1	140.05	54.37	5.83	-7.14	Low	Yes/1	5.74
Taraxerol	426.72	31	0	0	1	1	134.88	20.23	4.77	-8.34	Low	Yes/1	6.04
Beta-sitosterol	414.71	30	0	6	1	1	133.23	20.33	4.79	-7.90	Low	Yes/1	6.30
Palmitic acid	256.42	18	0	14	2	1	80.80	37.30	3.85	-5.02	High	Yes/1	2.31
Sulfoquinovosyldiacylglycerol	834.15	57	0	37	12	4	228.17	197.33	0	-7.12	Low	No/2	9.02
Racemosin C	372.37	28	18	2	4	3	105.83	95.18	2.15	-4.65	High	Yes/0	3.71
Caulerchlorin	374.82	27	22	2	2	2	110.27	57.88	2.97	-5.83	High	Yes/1	2.26
Racemosin A	345.33	26	12	2	4	2	101.16	92.34	2.97	-5.83	High	Yes/1	2.26
Racemosin B	314.34	24	20	2	2	2	96.45	57.88	2.74	-5.23	High	Yes/0	3.30
Caulerprenylol B	248.36	18	6	3	2	2	75.08	40.46	2.93	-3.64	High	Yes/0	3.30
Caulerprenylol A	258.36	19	6	0	2	2	80.88	40.46	2.91	-3.59	High	Yes/0	4.14
AmBiosome	924.08	65	0	3	18	12	239.06	319.61	3.76	-5.37	Low	No/3	10
Monomethyl caulerpinate	384.38	29	22	3	4	3	112.22	95.18	2.11	-5.09	High	Yes/0	2.21
4',5'-dehydrodiodictyonema A	461.68	33	0	18	4	1	142.05	72.47	5.00	-6.81	High	Yes/1	5.28
Racemobutenolid A	308.5	22	0	11	2	0	96.95	26.30	4.63	-5.47	High	Yes/1	4.65
Racemobutenolid B	308.5	22	0	11	2	0	96.95	26.30	4.63	-5.47	High	Yes/1	4.65
(23E)-3β-hydroxy-stigmasta-5,23-dien-28-one	426.67	31	0	5	2	1	133.21	37.30	4.54	-6.58	Low	Yes/1	6.04
(3b,24R)-stigmasta-5,28-diene-3,24-diol	430.66	31	0	6	3	2	130.08	57.53	4.13	-6.01	High	Yes/1	6.05
(3β,24S)-stigmasta-5,28-diene-3,24-diol	430.66	31	0	6	3	2	130.08	57.53	4.04	-6.01	High	Yes/1	6.05
(22E)-3β-hydroxy-cholesta-5,22-dien-24-one	398.62	29	0	4	2	1	123.6	37.30	4.28	-6.12	High	Yes/1	5.82

(Continued on following page)

TABLE 1 (Continued) List of pharmacokinetic properties of 36 compounds from *Caulerpa* spp.

Properties	Physicochemical properties								Lipophilicity	Water solubility	Pharmacokinetics	Drug-likeness	Medicinal chemistry
Parameters	Molecular weight (g/mol)	Number of heavy atoms	Number of aromatic heavy atoms	Number of rotatable bonds	Number of H-bond acceptors	Number of H-bond donors	Molar refractivity	TPSA (Å)	Log P _{0/w}	LogS (ESOL)	GI absorption	Lipinski/violation	Synthetic accessibility
Fucosterol	410.67	30	0	4	1	1	132.54	20.23	4.28	-6.12	High	Yes/1	5.82
24R,28S-epoxyfucosterol	426.67	31	0	4	2	1	132.02	32.76	4.87	-6.63	Low	Yes/1	6.35
24S,28R-epoxyfucosterol	426.67	31	0	4	2	1	132.02	32.76	4.97	-6.63	Low	Yes/1	6.35
(3β,23E)-stigmasta-5,23-dien-3,28-diol	428.69	31	0	5	2	2	134.18	40.46	4.82	-6.78	High	Yes/1	6.37
α-tocoxyleneoxy	552.87	40	12	14	3	1	176.97	38.69	0	-10.13	Low	No/2	6.14
Cacospongionolide C	324.5	23	0	12	3	1	98.11	46.53	4.29	-5.30	High	Yes/1	4.81
α-tocospiro A	462.7	33	0	13	4	1	139.58	63.60	5.18	-6.53	Low	Yes/0	6.88
α-tocospirone	462.7	33	0	12	4	1	139.58	63.60	5.45	-6.99	Low	Yes/0	6.65
Furocaulerpin	272.34	20	5	5	3	0	79.86	39.44	3.79	-4.11	High	Yes/0	4.08
Trifaritin	390.56	28	0	15	4	0	118.16	52.60	5.14	-5.87	High	Yes/1	4.46
Caulerpicin	622.10	44	0	39	2	2	203.36	49.33	9.57	-12.63	Low	No/2	5.88

^aThe bold values indicate the chemicals that fit Lipinski's Rule of 5.

identified to determine the performance of MM/PB(GB)SA by using Amber package (Fast Amber Rescoring for PPI inhibitors-farPPI; <http://cadd.zju.edu.cn/farppi>; (Wang et al., 2019). The input files were generated using AutoDock Tools. The force field parameter was set as GAFF2 (for ligand) + ff14SB (for a receptor) and the rescoring procedure was set as PB3 (radii = parse, $\gamma = 0.00542$, $\beta = 0.9200$). AM1-BCC method was used to calculate the partial charge of the ligands by using antechamber module of Amber.

3 Results

3.1 Drug-likeness analysis, ADME and toxicity test analysis of ADME properties

Lipophilicity, water solubility, drug-likeness, medicinal chemistry (leadlikeness) values of 36 metabolites from *Caulerpa* spp. Were obtained using the SwissADME server. The results reveal that the logP of 31 compounds were in the range of 0–5, on the other hand, five of the secondary metabolites (transphytol, alpha-tocospiroA, alpha-tocospirone, trifarin and caulerpicin) were not in the range of Lipinski's Rule of five (LRO5): $2 \leq \log P \leq 5$. According to the rule of 5, the molecular weight should be $200 \leq MW \leq 500$. The molecular weights of the 32 compounds were in the acceptable range. However, the MW of sulfoquinovosyldiacyl glycerol, amBiosome, alpha-tocoxyleneoxy and caulerpicin do not satisfying the Lipinski Ro5. The number of H-bond acceptors (≤ 10) and donors (≤ 5) for 34 (except sulfoquinovosyldiacyl glycerol and amBiosome) and 35 (except sulfoquinovosyldiacyl glycerol) compounds falling in acceptable range, respectively. All the compounds (except sulfoquinovosyldiacyl glycerol) were found to be the range of topological polar surface area (TPSA; < 140). The minimum and the maximum numbers of rotatable bonds were found to be 0 and 39, respectively (Table 1). Only the chemicals that were acceptable for LRO5 with no violation were selected for the cheminformatic pipeline and further pharmacodynamic studies. Considering all the obtained results, 13 *Caulerpa*-based phytochemicals (caulerpin, caulerpenyne, caulersin, 10,11-epoxycaulerpenyne, flexilin, racemosin C, racemosin B, caulerprenylol B, caulerprenylol A, monomethyl caulerpinate, α -tocospiro A, α -tocospirone and furocaulerpin) were chosen and used in subsequent steps.

In the cheminformatic pipeline of the study, computational based *in-silico* toxicity was also used. T.E.S.T. tool and ProTox-II servers were used to identify the adverse effects and toxicity of the 13 selected compounds to evaluate several toxicological parameters (acute toxicity, carcinogenicity, cytotoxicity, hepatotoxicity, immunotoxicity, predicted median lethal dose; LD50 and mutagenicity). ProTox-II results revealed that caulerpin, caulerpenyne, caulersin, flexilin, racemosin C, racemosin B and monomethyl caulerpinate belonging to the

toxicity class 4, LD50 range from 500 to 1760 mg/kg, these would be harmful in case oral delivery. (Table 2).

3.2 Molecular docking studies

In this study, molecular docking studies were applied for the investigation of anticancer activity of caulerpin, caulerpenyne, 10,11-epoxycaulerpenyne, caulersin, flexilin, racemosin C, racemosin B, caulerprenylol B, caulerprenylol A, monomethyl caulerpinate, α -tocospiro A, α -tocospirone and furocaulerpin. For anticancer studies, G6PDH and 6PGD, which are crucial for CRC, were selected as receptors.

3.2.1 Glucose 6-phosphate dehydrogenase

G6PDH is a cytosolic rate-limiting enzyme that converts G6P into 6-phosphoglucono- δ -lactone in the pentose phosphate pathway. In this study, human G6PDH (PDBID:6E08) was selected as a target. Both caulerpin and monomethyl caulerpinate, which have docking scores less than $-10 \text{ kcal mol}^{-1}$ were selected as ligands with the highest activity. According to the results, the lowest and the highest binding energies on G6PDH were found as -10.6 and -5.8 kcal/mol for caulerpin and 6-aminonicotinamide, respectively (Table 3). Inside the binding cavity of G6PDH, the methyl ester group of caulerpin makes H-bond with Lys171. Also, the indole ring of caulerpin docked in G6PDH makes pi-pi staking with Phe253 (Figure 1A). Furthermore, the indole ring of monomethyl caulerpinate makes pi-pi stacking with Phe253 (Figure 1B).

3.2.2 6-Phosphogluconate dehydrogenase

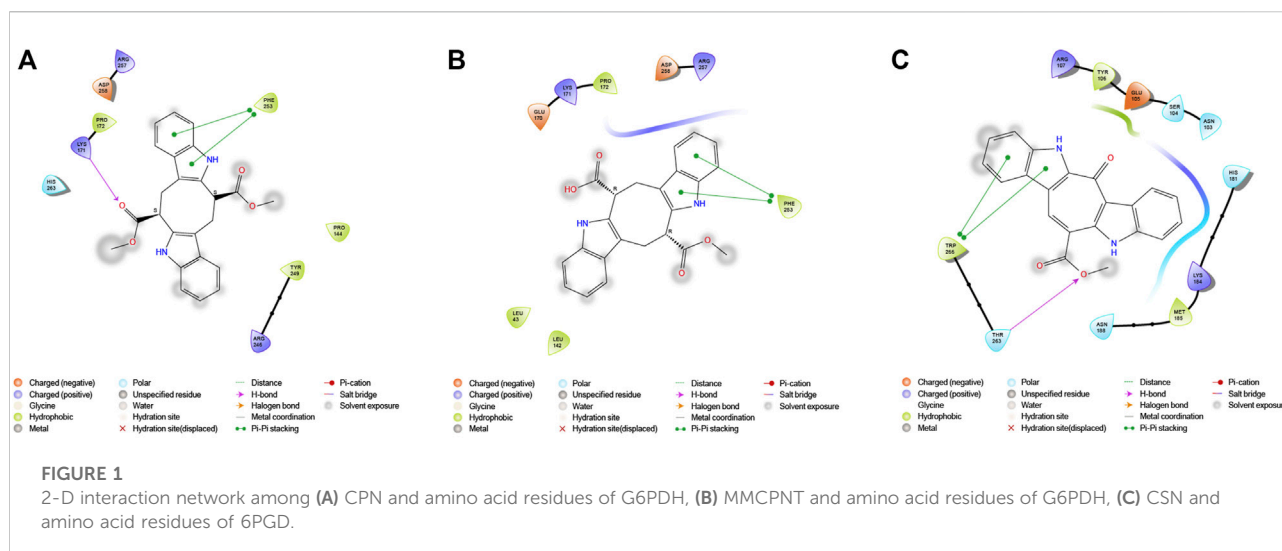
The molecular docking studies on 6PGD reveal that the lowest binding score was found as -10.7 kcal/mol for caulersin as given in Table 3. Also, the highest binding energy was found as -5.8 for 6-aminonicotinamide. Inside the binding cavity of 6PGD, Trp266 forms pi-pi stacking with both the pirole ring of the indole ring of caulersin. Also, Thr263 makes hydrogen bond with double bond oxygen of the methyl ester group of the central traponoid of caulersin. Related information is given in Figure 1C.

3.3 Docking validation by molecular dynamic simulations

In the MD simulations, RMSD (Root Mean Square Deviation) values, the starting position of the backbone of all amino acid residues, were calculated using WebGRO to clarify the stability and overall conformational dynamics of receptor-ligands (G6PDH-caulerpin, G6PDH-monomethyl caulerpinate, 6PGD-caulersin). The results reveal that the average RMSD values for caulerpin-G6PDH, monomethyl caulerpinate-

TABLE 2 List of toxicity properties of selected *Caulerpa*-based phytochemicals.

Endpoint		Caulerpin	Caulerpenyne	Caulersin	10,11-Epoxycaulerpenyne	Flexilin	Racemosin C	Racemosin B	Caulerprenylol B	Caulerprenylol A	Monomethyl caulerpinate	α -tocospiro A	α -tocospirone	Furocaulerpin
Organ toxicity	Hepatotoxicity	IA	IA	IA	IA	IA	IA	IA	IA	IA	IA	IA	IA	IA
Toxicity	Carcinogenicity	IA	IA	IA	IA	IA	IA	IA	IA	IA	IA	IA	IA	IA
	Cytotoxicity	IA	IA	IA	IA	IA	IA	IA	IA	IA	IA	IA	IA	IA
	Immunotoxicity	IA	IA	IA	A	IA	A	A	IA	IA	IA	IA	IA	IA
	LD50 (mg/kg)	1760	500	500	2,000	710	1760	4,425	5,500	4400	1760	300	300	5000
	Mutagenicity	A	IA	A	A	IA	A	IA	IA	IA	IA	IA	IA	IA
	Toxicity class	4	4	4	5	4	4	4	5	5	4	3	3	5
	AR	IA	IA	IA	IA	IA	IA	IA	IA	IA	IA	IA	IA	IA
	AhR	IA	IA	IA	IA	IA	IA	A	IA	IA	IA	IA	IA	IA
	PPARgamma	IA	IA	IA	IA	IA	IA	IA	IA	IA	IA	IA	IA	IA
	P53	IA	IA	IA	IA	IA	IA	IA	IA	IA	IA	IA	IA	IA
	Heat shock protein	IA	IA	IA	IA	IA	IA	IA	IA	IA	IA	IA	IA	IA
Bioconcentration factor	Log10	0.74	N/A	N/A	N/A	0.93	N/A	1.10	2.20	1.81	0.36	1.24	1.55	N/A
<i>Daphnia magna</i> toxicity (48 h)	mg/L	0.67	0.16	0.71	4.27 E-02	0.76	2.15	0.78	3.43	5.74	0.66	0.85	0.70	0.19
Developmental toxicity value		0.90	0.68	0.88	0.68	0.68	0.93	0.82	0.79	0.74	0.97	0.77	0.76	0.62
Fathead minnow (LC50 96h)	mg/L	8.61 E-03	0.32	7.15 E-03	0.12	0.30	3.98 E-03	6.69 E-02	13.52	0.53	3.33 E-02	1.20	0.68	0.44
Mutagenicity (AMES)		N/A	-	N/A	+	-	+	+	-	+	N/A	-	-	-
Oral rat LD50)	mg/L	215.30	N/A	759.69	N/A	8193.16	268.66	N/A	764.83	1857.98	288.29	143.00	268.08	N/A



G6PDH and caulersin-6PGD were found as 0.31, 0.46, and 0.43 nm respectively. All the values were comparable and in the physiological environment, indicating the stability of ligand-protein interaction. RMSF (Root Mean Square Fluctuation) values, the standard deviation of atomic positions of each amino acid residues, were also calculated. 0.06, (Figures 2A,E,I). The results from RMSD showed that CPN, MMCNPNT and CSN remained positioned at the active sites of the G6PDH and 6PGD with stable interactions.

Radius of gyration (Rg) computes the structural compactness and dynamic adaptability of the ligand-protein complex about the x-, y- and z-axes, as a function of time. In Figures 2B,F,J, Rg values of CPN, MMCNPNT and CSN with G6PDH and 6PGD receptors ranged between -2.35 and 2.45 nm, -2.30–2.43, and -2.32–2.60 nm, respectively. The overall Rg results revealed that, G6PDH-CPN receptor-ligand complex had minimum structural compactness variations and this result indicates the stability of the complexes. SASA is an approximate structural stability of the ligand-protein interaction that is accessible to a solvent (water) with respect to simulation time (100 ns). It was observed that the frequencies of SASA of all G6PDH complexes were decreased around 210 nm² (Figures 2C,G), on the other hand, SASA result of 6PGD-CSN complex was restricted around 190 nm² (Figure 2K). The maximum numbers of H-bonds of caulerpin-G6PDH, monomethyl caulerpinate-G6PDH, and caulersin-6PGD per time frame were found to be 2, 3 and 4, respectively. Furthermore, H-bond formation dynamics between ligands and proteins reveal that for all complexes, at least one H-bond was found as long-lived all through the simulation (100 ns) (Figures 2D, H, L).

3.4 Pharmacophore modeling


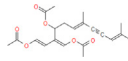

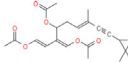
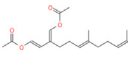


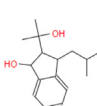

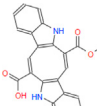
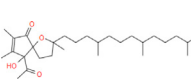
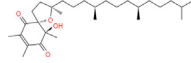
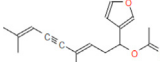
In this study, combined structure- and ligand-based pharmacophore modeling was performed to evaluate *Caulerpa*-based phytochemicals with potential activity against G6PDH and 6PGD. PharmaGist server was used for pharmacophore modeling to enlighten the three-dimensional pharmacophoric features of top hit ligands for each receptor. Pharmacophore modeling is of great importance for specific receptors to elucidate if the interaction blocks or triggers a biological response.

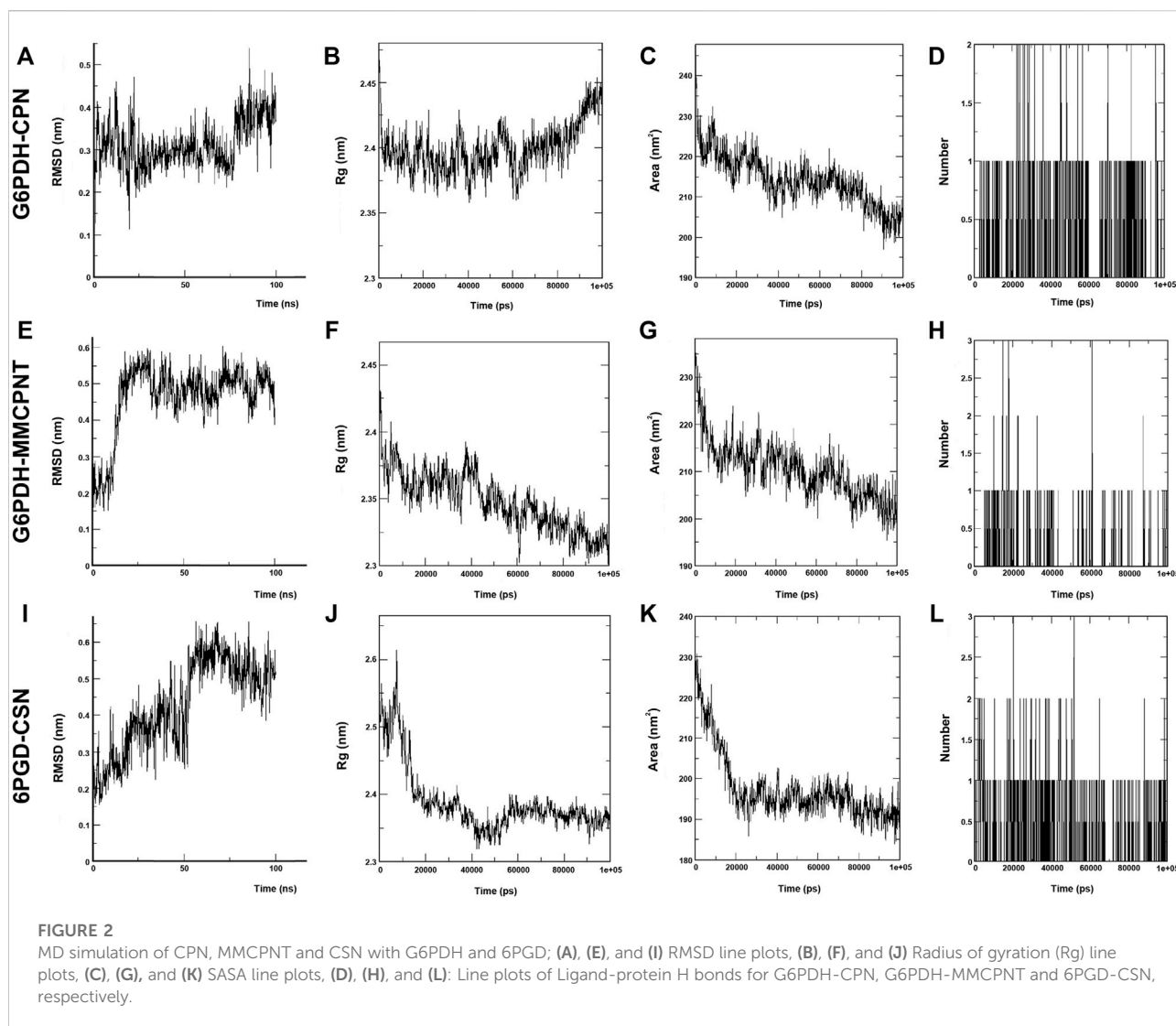
In this study, top hits (caulerpin and monomethyl caulerpinate for G6PDH and caulersin for 6PGD) were used for each compound in the same orientation at the same binding pocket. The pairwise structural alignment details are given in Table 4. DHEA and 6-aminonicotinamide (6ANA) were used as pivot molecules for G6PDH and 6PGD, respectively. The hit compounds, “caulerpin and monomethyl caulerpinate” and “caulersin” were modeled for G6PDH and 6PGD, respectively. The results revealed that both caulerpin and monomethyl caulerpinate shared the maximum feature number with DHEA. The pairwise structural alignment of DHEA and 6ANA with the top hits is shown in Figure 3.

3.5 Binding free energy calculations

The binding free energies of caulerpin and monomethyl caulerpinate with G6PDH using MM-PB(GB)SA were calculated as -38.43 and -40.94 kcal/mol, respectively (Figures 4A,B). Also, the binding free energy of caulersin with 6PGD using MM-PB (GB) SA was calculated as -20.20 kcal/mol (Figure 4C).

TABLE 3 Docking scores with G6PDH and 6PGD and the top 13 selected compounds from *Caulerpa* spp.

Docking score (kcal/mol) compounds		Glucose-6-phosphate dehydrogenase (G6PDH) (PDBID:6E08)	6-Phosphogluconate dehydrogenase (6PGD) (PDBID:4GWK)
Caulerpin		-10.6	-8.7
Caulerpenyne		-9.5	-6.8
Caulersin		-9.5	-10.7
10,11-epoxycaulerpenyne		-7.9	-6.5
Flexilin		-6.8	-5.9
Racemosin C		-9.8	-9.1
Racemosin B		-9.2	-9.5
Caulerprenylol B		-8.3	-8.3
Caulerprenylol A		-8.8	-7.7
Monomethyl caulerpinate		-10.5	-9.8
α -tocospiro A		-9.8	-7.3
α -tocospirone		-8.3	-7.4
Furocaulerpin		-7.6	-6.9
DHEA		-7.8	-6.2
6 aminonicotinamide		-5.8	-5.8



4 Discussion

In this study, 13 of 36 different secondary metabolites of *Caulerpa* (caulerpin, caulerpenyne, 10,11-epoxycaulerpenyne, caulersin, flexilin, racemosin C, racemosin B, caulerprenylol B, caulerprenylol A, monomethyl caulerpinate, α -tocospirone A, α -tocospirone and furocaulerpin) against crucial targets (glucose-6-phosphate dehydrogenase and 6-phosphogluconate dehydrogenase) for colorectal cancer were carried out by using *in-silico* pharmacokinetic and pharmacodynamic methods. Caulerpin and monomethyl caulerpinate were found to be the most effective metabolites against G6PDH. Caulersin had the lowest affinity score against 6PGD. The pentose phosphate pathway is fundamental for colorectal cancer, thus caulerpin, monomethyl caulerpinate and caulersin play important roles in colorectal cancer treatment through the pentose phosphate pathway.

ADME is an important medicinal chemistry tool that provides information about *in-silico* ADME behavior (Bocci et al., 2017; Pantaleão et al., 2022). The detailed pharmacokinetic ADME-Tox and drug-likeness results are given in Table 1 and Table 2. The physicochemical properties (molecular weight (g/mol), the number of heavy atoms, the number of aromatic heavy atoms, the number of rotatable bonds, number of H-bond acceptors, the number of H-bond donors, molar refractivity and TPSA (Å)) were calculated for 36 *Caulerpa*-based phytochemicals. ADME results reveal that the compounds with low molecular weight (≤ 500 g/mol) tend to absorb well (Daina et al., 2017). In our study, 4 of 36 phytochemicals (sulfoquinovosyldiacyl glycerol, ambiosome, α -tocoxyleneoxy and caulerpicin) from *Caulerpa* spp. have high molecular weight (low absorption capacity). The flexibility of bioactive molecules is determined using the number of rotatable bonds (Daina et al., 2017) which

TABLE 4 Pairwise structural alignment showing common pharmacophoric features of secondary metabolites (pivot molecule) and top hit compounds against G6PDH and 6PGD

Score	Features	Spatial features	Aromatic	Hydrophobic	Donors	Acceptors	Negatives	Positives	Molecules
2.430	4	4	0	3	0	1	0	0	DHEA(pivot molecule of G6PDH)-caulerpin (hit compound of G6PDH)
2.431	4	4	0	3	0	1	0	0	DHEA(pivot molecule of G6PDH)-monomethyl caulerpinate (hit compound of G6PDH)
6.016	3	3	1	0	1	1	0	0	6ANA (pivot molecule of 6PGD)-caulersin(hit compound of 6PGD)

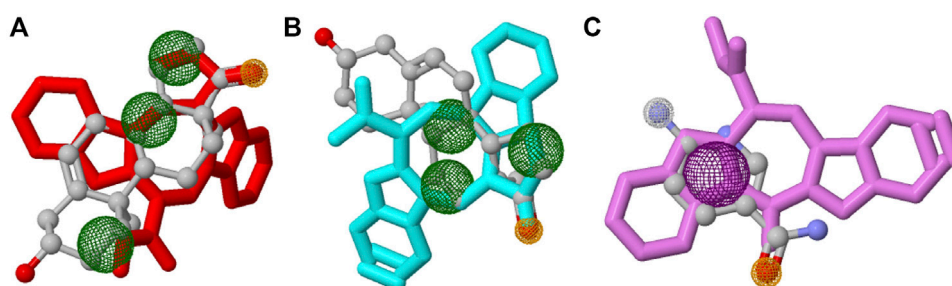


FIGURE 3

Structural alignment of pivot molecule, (A) DHEA (gray) with CPN (red), (B) DHEA (gray) with MMCPNT (blue), and (C) 6ANA (gray) with CSN (violet). DHEA and 6ANA were displayed in ball and stick style, CPN, MMCPNT and CSN were shown in sticks style. (A,B) yellow spheres represent hydrogen bond acceptors, green spheres represent hydrophobic features, (C) white spheres represent hydrogen bond donors, yellow spheres represent hydrogen bond acceptors, and purple spheres specify aromatic features.

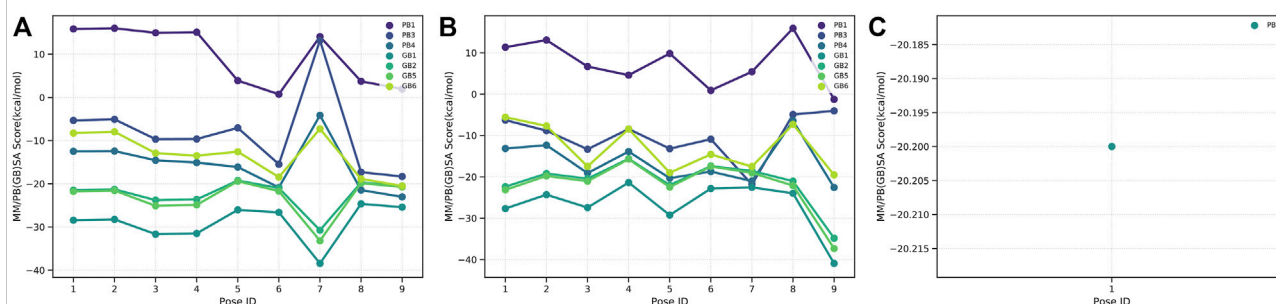


FIGURE 4

MM-PB(BB)SA results of (A) CPN, (B) MMCPNT, and (C) CSN.

should be between 0 and 9. In this study, all the chemicals except sulfoquinovosyldiacyl glycerol and caulerpin are in the range of this value. Topological polar surface area (TPSA) is based on the fragmental system of phosphorous atoms and polar sulfur. TPSA value should be between 20 and 130 Å for polarity (Daina et al., 2017). In this study, all the 13 phytochemicals were found in this range. Lipophilicity (logP) is crucial for clarifying the effect of

chemicals' absorption, distribution, transportation on physiological systems. In this study, all the samples (36) out of 5 phytochemicals were in the range of logP ($-2 \leq \log P \leq 5$). For water solubility, logS (ESOL) was tested. Daina et al. defined the scale of water solubility as insoluble < -10 < poorly < -6 < moderately < -4 < soluble < -2 < very < 0 < highly (Daina et al., 2017). The results reveal that caulerpin, caulerpenyne,

caulersin, flexilin, racemosin C, racemosin B, monomethyl caulerpinate, and furocaulerpin were found as moderately soluble; 10,11-epoxycaulerpenyne, caulerprenylol B and caulerprenylol A were found as soluble and α -tocospiro A and α -tocospirone were found as poorly soluble.

Before the clinical trials of drug candidates, *in-silico* toxicity measurement procedure is quite important for better selecting the lead compound (Han et al., 2019). These computational-based toxicity measurement procedures are accurate, accessible, rapid, and common. Both ProTox-II and T.E.S.T servers (freely accessible) were used to identify the adverse effects and toxicity (acute toxicity, carcinogenicity, cytotoxicity, hepatotoxicity, immunotoxicity and mutagenicity) of selected phytochemicals from ADME results. Our results reveal that the toxicity classes of the selected phytochemicals were found to be more than three. For all selected phytochemicals, hepatotoxicity, carcinogenicity, cytotoxicity androgen receptor results were found as inactive.

Drug design and discovery is a step-by-step process, which is costly for companies (Opo et al., 2021). The bioavailability and drug-likeness analysis using *in-silico* is of great importance for evaluating the pharmacokinetic features of fast and cheap (Hasan et al., 2022).

In this study, the selected metabolites (ligands) have the potential anticancer activity against the selected receptor targets for CRC. According to the results, among the commercial drugs (DHEA and 6ANA), for G6PDH, the best docking energy was exhibited by caulerpin with a vina score of -10.6 kcal/mol, while for G6PDH, the other best docking energy was exhibited by monomethyl caulerpinate with a vina score of -10.5 kcal/mol. For 6PGD, the best docking score among the secondary metabolites of *Caulerpa* was exhibited by caulersin (-10.7 kcal/mol). (Table 3). In the literature, limited studies are related to the activities of caulerpin on different targets *in-silico*. In the study by Lorenzo et al. (2015), *in-silico* molecular docking study of caulerpin and its nine analogs against monoamine oxidase B was carried out. Their results reveal that moldock energy, predicted probability (%) and drug-like score of caulerpin were found as -152% , 58% , and 0.77 , respectively. However, the analogs of caulerpin which have non-polar and polar groups showed different moldock energy, predicted probability (%) and drug-like scores. The methods that they used (Volsurf descriptors, structure-based methodology and Random Forest algorithm) are crucial for finding the good drug candidates (caulerpin and its analogs) against monoamine oxidase B (Lorenzo et al., 2015). In the literature, Vitale et al. (2018) carried out an *in silico* molecular docking evaluation of caulerpin against PPAR α and PPAR γ . Their study reveals that the main interaction between ligand (caulerpin) and protein is a hydrophobic interaction. Also, according to their molecular dynamics results, caulerpin makes intermolecular H-bonds with S289 (VI) and S342/R288 (V) (Vitale et al., 2018). Furthermore, antiviral activity of caulerpin against SARS-CoV-2 was tested using *in-silico* tools (Abdelrheem

et al., 2020; Ahmed et al., 2020; Çavaş et al., 2020; El-Mageed et al., 2021).

MM-PBSA analysis is an important and popular method in drug candidate filtration since it is an easy method, and the speed-accuracy balance of the information is high. Estimation of binding free energies with MM-PB(GB) SA for the ligands is automated with the farPPI web server. In our study, the highest and the lowest binding free energies were calculated for monomethyl caulerpinate and caulersin, respectively.

Caulerpin also has some biological activities such as anticancer activity on melanoma cells (Rocha et al., 2007), HIF-1 activation and inhibition of mitochondrial respiration (Liu et al., 2009), antiviral activities against bovine viral diarrhea virus in cattle and herpes simplex virus (Macedo et al., 2012; Pinto et al., 2012; Zhang et al., 2015), pain-sensitizing and spasmolytic effect (Cavalcante-Silva et al., 2014), antituberculosis activity (Canché Chay et al., 2014), antiproliferative activity (Movahhedin et al., 2014), monoamine oxidase inhibitory activity against Alzheimer's and Parkinson's disease (Lorenzo et al., 2015), activity on cisplatin-resistant ovarian cancer and inhibition respiratory complex II activity (Ferramosca et al., 2016), AMPK α 1 pathway activation in colorectal cancer cells (Yu et al., 2017) and PPAR α and PPAR γ agonist activity on hepatocellular cell line (Vitale et al., 2018). Caulersin is a known human protein tyrosine phosphatase-1B inhibitor, which regulates insulin signaling negatively (Yang et al., 2014). The anticancer activity of caulerpin, caulersin, caulerpenyne and 10,11-epoxycaulerpenyne for colorectal cancer data reveal that caulerpin and caulersin are promising anticancer agents against CRC targets and G6PDH and 6PGD could be important targets for CRC.

Inhibition of PPP-enzymes is related with AMPK-activation, HIF-1 α degradation, impaired folate metabolism and PP2A-activation (Meskers et al., 2022). G6PDH is the main NADPH production and redox homeostasis contributor (Ghergurovich et al., 2020). The expression level of G6PDH is upregulated and negatively correlated with patients with cancer (Ghergurovich et al., 2020). In different CRC cell lines, the expression levels of G6PDH and 6PGD are quite different. In the study of Polat et al. (2021), the highest and the lowest G6PDH levels were found in HT29 and Caco-2 cell lines, respectively (Polat et al., 2021). Thus, the effectiveness of the selected secondary metabolites could be changed depending on the colorectal cancer cell type.

For the prospects, the not only the *in-silico* analysis but also the *in-vitro* experiments of these targets for caulerpin, monomethyl caulerpinate and caulersin should be performed.

5 Conclusion

Global warming and human activities change the ecosystem structures. Alien members of Genus *Caulerpa* are widely studied marine algae due to their invasive properties. The present paper proposes an alternative utilization method in medicinal

chemistry. The secondary metabolites of *Caulerpa* spp. attract attention due to their bioactivities. Using the ADME-tox and drug-likeness tests, 13 of 36 secondary metabolites were selected and molecular docking, and molecular dynamics analysis were performed. Caulerpin, monomethyl caulerpinate, and caulersin were found the hit compounds of *Caulerpa* spp. Against G6PDH and 6PGD, which may play pivotal roles in CRC. Thus, instead of eradication of these algae, the secondary metabolites proposed in this paper (caulerpin, monomethyl caulerpinate) might further be evaluated as promising agents that can be obtained from pharmacy of nature.

Data availability statement

The datasets generated and/or analysed during the current study are available from the corresponding author on reasonable request.

Author contributions

YB, LC, and HE made significant contributions to the concept or design of the work, NM-O made contributions to the acquisition and analysis data of the study, HE, NO, and GC-K made contributions to the interpretation of data for the work. All authors participated in the drafted and revised the manuscript critically. All authors read and approved the final manuscript. All authors agreed to be responsible for all aspects of the work are appropriately investigated and resolved.

References

- Abdelrhheem, D. A., Ahmed, S. A., Abd El-Mageed, H. R., Mohamed, H. S., Rahman, A. A., Elsayed, K. N. M., et al. (2020). The inhibitory effect of some natural bioactive compounds against SARS-CoV-2 main protease: insights from molecular docking analysis and molecular dynamic simulation. *Journal of Environmental Science and Health Part A* 55, 1373–1386. doi:10.1080/10934529.2020.1826192
- 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
- Aguilar-Santos, G. (1970). Caulerpin, a New Red Pigment from Green Algae of the Genus *Caulerpa*. *J. Chem. Soc. Perkin 1*, 6, 842–843. doi:10.1039/J39700000842
- Ahmed, S. A., Abdelrhheem, D. A., El-Mageed, H. R. A., Mohamed, H. S., Rahman, A. A., Elsayed, K. N. M., et al. (2020). Destabilizing the structural integrity of COVID-19 by caulerpin and its derivatives along with some antiviral drugs: An in silico approaches for a combination therapy. *Struct. Chem.* 31, 2391–2412. doi:10.1007/s11224-020-01586-w
- Banerjee, P., Eckert, A. O., Schrey, A. K., and Preissner, R. (2018). ProTox-II: A webserver for the prediction of toxicity of chemicals. *Nucleic Acids Res.* 46, W257–W263. doi:10.1093/nar/gky318
- Bocci, G., Carosati, E., Vayer, P., Arrault, A., Lozano, S., and Cruciani, G. (2017). ADME-Space: A new tool for medicinal chemists to explore ADME properties. *Sci. Rep.* 7, 6359–27. doi:10.1038/s41598-017-06692-0
- Canché Chay, C. I., Cansino, R. G., Espitia Pinzón, C. I., Torres-Ochoa, R. O., and Martínez, R. (2014). Synthesis and anti-tuberculosis activity of the marine natural product caulerpin and its analogues. *Mar. Drugs* 12, 1757–1772. doi:10.3390/md12041757
- Cavalcante-Silva, L. H. A., Falcão, M. A. P., Vieira, A. C. S., Viana, M. D. M., De Araújo-Júnior, J. X., Sousa, J. C. F., et al. (2014). Assessment of mechanisms involved in antinociception produced by the alkaloid caulerpine. *Molecules* 19, 14699–14709. doi:10.3390/molecules190914699
- Calas, L., Baskin, Y., Yurdakoc, K., and Olgun, N. (2006). Antiproliferative and newly attributed apoptotic activities from an invasive marine alga: *Caulerpa racemosa* var. *cylindracea*. *J. Exp. Mar. Biol. Ecol.* 339, 111–119. doi:10.1016/j.jembe.2006.07.019
- Çavaş, L., Dag, C., Carmena-barreño, M., Martínez-cortés, C., Pedro, J., and Pérez-sánchez, H. (2020). Secondary Metabolites from *Caulerpa cylindracea* (Sonder) Could Be Alternative Natural Antiviral Compounds for COVID-19: A Further in Silico Proof. doi:10.26434/chemrxiv.13353473.v1
- Cengiz, S., Cavas, L., Yurdakoc, K., and Pohnert, G. (2011). The Sesquiterpene Caulerpenyne from *Caulerpa* spp. is a Lipooxygenase Inhibitor. *Mar. Biotechnol.* 13, 321–326. doi:10.1007/s10126-010-9303-1
- Daina, A., Michielin, O., and Zoete, V. (2017). SwissADME: A free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. *Sci. Rep.* 7, 42717–13. doi:10.1038/srep42717
- Dege, N., Gökce, H., Doğan, O. E., Alpaslan, G., Ağar, T., Muthu, S., et al. (2022). Quantum computational, spectroscopic investigations on N-(2-((2-chloro-4, 5-dicyanophenyl)amino)ethyl)-4-methylbenzenesulfonamide by DFT/TD-DFT with different solvents, molecular docking and drug-likeness researches. *Colloids and Surfaces A: Physicochemical and Engineering Aspects* 638, 128311. doi:10.1016/j.colsurfa.2022.128311
- Dekker, E., Tanis, P. J., Vleugels, J. L. A., Kasi, P. M., and Wallace, M. B. (2019). Colorectal cancer. *The Lancet* 394, 1467–1480. doi:10.1016/S0140-6736(19)32319-0

Funding

This study was funded by Dokuz Eylül University Scientific Research Projects (Grant number: 2018.KB.SAG.079). NMO is supported by the Council of Higher Education (YOK) 100/2000 Doctoral Project and Scientific and Technological Research Council of Turkey (TUBITAK, 2211-A).

Acknowledgments

The authors would like to thank Caner Karaca from Dokuz Eylül University for improvement of qualities of figures in this study.

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.

- El-Mageed, H. R. A., Abdelreheem, D. A., Ahmed, S. A., Rahman, A. A., Elsayed, K. N. M., Ahmed, S. A., et al. (2021). Combination and tricomination therapy to destabilize the structural integrity of COVID-19 by some bioactive compounds with antiviral drugs: insights from molecular docking study. *Struct. Chem.* 32, 1415–1430. doi:10.1007/s11224-020-01723-5
- Erb, M., and Kliebenstein, D. J. (2020). Plant Secondary Metabolites as Defenses, Regulators, and Primary Metabolites: The Blurred Functional Trichotomy. *Plant Physiol.* 184, 39–52. doi:10.1104/PP.20.00433
- Ferramosca, A., Conte, A., Guerra, F., Felling, S., Rimoli, M. G., Mollo, E., et al. (2016). Metabolites from invasive pests inhibit mitochondrial complex II: A potential strategy for the treatment of human ovarian carcinoma? *Biochem. Biophys. Res. Commun.* 473, 1133–1138. doi:10.1016/j.bbrc.2016.04.028
- Gao, W., Xu, Y., Chen, T., Du, Z., Liu, X., Hu, Z., et al. (2019). Targeting oxidative pentose phosphate pathway prevents recurrence in mutant Kras colorectal carcinomas. *PLoS Biol.* 17, e3000425. doi:10.1371/journal.pbio.3000425
- Ghergurovich, J. M., Esposito, M., Chen, Z., Wang, J. Z., Bhatt, V., Lan, T., et al. (2020). Glucose-6-phosphate dehydrogenase is not essential for K-Ras-driven tumor growth or metastasis. *Cancer Res.* 80, 3820–3829. doi:10.1158/0008-5472.can-19-2486
- Gokce, H., Sen, F., Sert, Y., Abdel-Wahab, B. F., Kariuki, B. M., and El-Hiti, G. A. (2022). Quantum Computational Investigation of (E)-1-(4-methoxyphenyl)-5-methyl-N'-(3-phenoxybenzylidene)-1H-1, 2, 3-triazole-4-carbohydrazide. *Molecules* 27, 2193. doi:10.3390/molecules27072193
- Han, Y., Zhang, J., Hu, C. Q., Zhang, X., Ma, B., and Zhang, P. (2019). In silico ADME and toxicity prediction of ceftazidime and its impurities. *Front. Pharmacol.* 10, 434–442. doi:10.3389/fphar.2019.00434
- Hasan, R., Alsaifi, A. A., Fakhurji, B. Z., Habibur, M., Molla, R., Asseri, A. H., et al. (2022). Application of Mathematical Modeling and Computational Tools in the Modern Drug Design and Development Process. *Molecules* 27, 4169. doi:10.3390/molecules27134169
- Jin, L., and Zhou, Y. (2019). Crucial role of the pentose phosphate pathway in malignant tumors. *Oncol. Lett.* 17, 4213–4221. doi:10.3892/ol.2019.10112
- Koes, D. R., and Camacho, C. J. (2012). ZINCPharmer: Pharmacophore search of the ZINC database. *Nucleic Acids Res.* 40, W409–W414. doi:10.1093/nar/gks378
- Liu, Y., Morgan, J. B., Coothankandaswamy, V., Liu, R., Jakabsons, M. B., Mahdi, F., et al. (2009). The *Caulerpa* pigment caulerpin inhibits HIF-1 activation and mitochondrial respiration. *J. Nat. Prod.* 72, 2104–2109. doi:10.1021/np9005794
- Lorenzo, V. P., Filho, J. M. B., Scotti, L., and Scotti, M. T. (2015). Combined structure- and ligand-based virtual screening to evaluate caulerpin analogs with potential inhibitory activity against monoamine oxidase B. *Rev. Bras. Farmacogn.* 25, 690–697. doi:10.1016/j.bjp.2015.08.005
- Macedo, N. R. P. V., Ribeiro, M. S., Villaca, R. C., Ferreira, W., Pinto, A. M., Teixeira, V. L., et al. (2012). Caulerpin as a potential antiviral drug against herpes simplex virus type 1. *Rev. bras. farmacogn.* 22, 861–867. doi:10.1590/S0102-695X2012005000072
- Martin, T. M., Harten, P., Venkatapathy, R., Das, S., and Young, D. M. (2008). A hierarchical clustering methodology for the estimation of toxicity. *Toxicol. Mech. Methods* 18, 251–266. doi:10.1080/15376510701857353
- Mehra, R., Bhushan, S., Bast, F., and Singh, S. (2019). Marine macroalga *Caulerpa*: role of its metabolites in modulating cancer signaling. *Mol. Biol. Rep.* 46, 3545–3555. doi:10.1007/s11033-019-04743-5
- Mert Ozupek, N., and Cavas, L. (2017). Triterpene glycosides associated antifouling activity from *Holothuria tubulosa* and *H. polii*. *Reg. Stud. Mar. Sci.* 13, 32–41. doi:10.1016/j.rsma.2017.04.003
- Meskers, C. J. W., Franczak, M., Smolenski, R. T., Giovannetti, E., and Peters, G. J. (2022). Are we still on the right path(way)? the altered expression of the pentose phosphate pathway in solid tumors and the potential of its inhibition in combination therapy. *Expert Opin. Drug Metab. Toxicol.* 18, 61–83. doi:10.1080/17425255.2022.2049234
- Montefalcone, M., Morri, C., Parravicini, V., and Bianchi, C. N. (2015). A tale of two invaders: divergent spreading kinetics of the alien green algae *Caulerpa taxifolia* and *Caulerpa cylindracea*. *Biol. Invasions* 17, 2717–2728. doi:10.1007/s10530-015-0908-1
- Movahhedini, N., Barar, J., Azad, F. F., Barzegari, A., and Nazemiyeh, H. (2014). Phytochemistry and biologic activities of caulerpa peltata native to Oman sea. *Iran. J. Pharm. Res.* 13, 515–521.
- 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, 33–14. Available at: <https://jcheminf.biomedcentral.com/track/pdf/10.1186/1758-2946-3-33>. doi:10.1186/1758-2946-3-33
- Opo, F. A. D., Rahman, M. M., Ahammad, F., Ahmed, I., Bhuiyan, M. A., and Asiri, A. M. (2021). Structure based pharmacophore modeling, virtual screening, molecular docking and ADMET approaches for identification of natural anti-cancer agents targeting XIAP protein. *Sci. Rep.* 11, 4049–18. doi:10.1038/s41598-021-83626-x
- Pantaleão, S. Q., Fernandes, P. O., Gonçalves, J. E., Maltarollo, V. G., and Honório, K. M. (2022). Recent Advances in the Prediction of Pharmacokinetics Properties in Drug Design Studies: A Review. *ChemMedChem* 17, e202100542. doi:10.1002/cmdc.202100542
- Patra, K. C., and Hay, N. (2014). The pentose phosphate pathway and cancer. *Trends Biochem. Sci.* 39, 347–354. doi:10.1016/j.tibs.2014.06.005
- Pinto, A. M. V., Leite, J. P. G., Ferreira, W. J., Cavalcanti, D. N., Villaca, R. C., Giongo, V., et al. (2012). Marine natural seaweed products as potential antiviral drugs against bovine viral diarrhoea virus. *Rev. bras. farmacogn.* 22, 813–817. doi:10.1590/S0102-695X2012005000060
- Polat, I. H., Tarrado-Castellarnau, M., Benito, A., Hernandez-Carro, C., Centelles, J., Marin, S., et al. (2021). Glutamine modulates expression and function of glucose 6-phosphate dehydrogenase via nrf2 in colon cancer cells. *Antioxidants* 10, 1349. doi:10.3390/antiox10091349
- Ramawat, K. G., and Goyal, S. (2020). Co-evolution of Secondary Metabolites During Biological Competition for Survival and Advantage: An Overview. *Ref. Ser. Phytochem.*, 3–17. doi:10.1007/978-3-319-96397-6_45
- Richter, P., Schubert, G., Schaible, A. M., Cavas, L., Werz, O., and Pohnert, G. (2014). Caulerpin and related bis-enol esters are novel-type inhibitors of human 5-lipoxygenase. *ChemMedChem* 9, 1655–1659. doi:10.1002/cmdc.201402065
- Rocha, F. D., Soares, A. R., Houghton, P. J., Pereira, R. C., Kaplan, M. A. C., and Teixeira, V. L. (2007). Potential Cytotoxic Activity of Some Brazilian Seaweeds on Human Melanoma Cells. *Phytother. Res.* 21, 170–175. doi:10.1002/ptr.2038
- Schneidman-Duhovny, D., Dror, O., Inbar, Y., Nussinov, R., and Wolfson, H. J. (2008). PharmaGist: a webserver for ligand-based pharmacophore detection. *Nucleic Acids Res.* 36, 223–228. doi:10.1093/nar/gkn187
- 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, 1355–1363. doi:10.1107/S0907444904011679
- Stinccone, A., Prigione, A., Cramer, T., Wamelink, M. M. C., Campbell, K., Cheung, E., et al. (2015). The return of metabolism: Biochemistry and physiology of the pentose phosphate pathway. *Biol. Rev.* 90, 927–963. doi:10.1111/brv.12140
- Su, J. Y., Zhu, Y., Zeng, L. M., and Xu, X. H. (1997). A new bisindole from alga *Caulerpa serrulata*. *J. Nat. Prod.* 60, 1043–1044. doi:10.1021/np970149x
- Trott, O., and Olson, A. (2010). AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* 31, 455–461. doi:10.1002/jcc.21334
- Vitale, R. M., D'aniello, E., Gorbi, S., Martella, A., Silvestri, C., Giuliani, M. E., et al. (2018). Fishing for targets of alien metabolites: A novel peroxisome proliferator-activated receptor (PPAR) agonist from a marine pest. *Mar. Drugs* 16, 431–446. doi:10.3390/md16110431
- Wang, Z., Wang, X., Li, Y., Lei, T., Wang, E., Li, D., et al. (2019). FarPPI: A webserver for accurate prediction of protein-ligand binding structures for small-molecule PPI inhibitors by MM/PB(GB)SA methods. *Bioinformatics* 35, 1777–1779. doi:10.1093/bioinformatics/bty879
- Yang, H., Liu, D. Q., Liang, T. J., Li, J., Liu, A. H., Yang, P., et al. (2014). Racemosin C, a novel minor bisindole alkaloid with protein tyrosine phosphatase-1B inhibitory activity from the green alga *Caulerpa racemosa*. *J. Asian Nat. Prod. Res.* 16, 1158–1165. doi:10.1080/10286020.2014.965162
- Yu, H., Zhang, H., Dong, M., Wu, Z., Shen, Z., Xie, Y., et al. (2017). Metabolic reprogramming and AMPKα1 pathway activation by caulerpin in colorectal cancer cells. *Int. J. Oncol.* 50, 161–172. doi:10.3892/ijo.2016.3794
- Zhang, M.-Z., Chen, Q., and Yang, G.-F. (2015). A review on recent developments of indole-containing antiviral agents. *Eur. J. Med. Chem.* 89, 421–441. doi:10.1016/j.ejmech.2014.10.065



OPEN ACCESS

EDITED BY

Faez Iqbal Khan,
Xi'an Jiaotong-Liverpool University,
China

REVIEWED BY

Anupam Dhasmana,
The University of Texas Rio Grande
Valley, United States
Mohd Anees Ahmed,
Cornell University, United States
Mohd Adil,
Dalhousie University, Canada

*CORRESPONDENCE

Eun-Ha Choi,
✉ choipdp@gmail.com
Dharmendra K. Yadav,
✉ dharmendra30oct@gmail.com

†PRESENT ADDRESS

Rohit Kumar Tiwari, Department of
Clinical Research, School of Allied
Health Sciences, Sharda University,
Greater Noida, Uttar Pradesh

SPECIALTY SECTION

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

RECEIVED 07 October 2022

ACCEPTED 05 December 2022

PUBLISHED 11 January 2023

CITATION

Ahmad A, Tiwari RK, Saeed M, Al-Amrah H,
Han I, Choi E-H, Yadav DK and Ansari IA (2023),
Carvacrol instigates intrinsic and extrinsic
apoptosis with abrogation of cell cycle
progression in cervical cancer cells: Inhibition
of Hedgehog/GLI signaling cascade.
Front. Chem. 10:1064191.
doi: 10.3389/fchem.2022.1064191

COPYRIGHT

© 2023 Ahmad, Tiwari, Saeed, Al-Amrah, Han, Choi, Yadav and Ansari.
This is an open-access article
distributed under the terms of the
[Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/)
(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.

Carvacrol instigates intrinsic and extrinsic apoptosis with abrogation of cell cycle progression in cervical cancer cells: Inhibition of Hedgehog/GLI signaling cascade

Afza Ahmad¹, Rohit Kumar Tiwari^{1†}, Mohd Saeed²,
Hadba Al-Amrah³, Ihn Han⁴, Eun-Ha Choi^{4*},
Dharmendra K. Yadav^{5*} and Irfan Ahmad Ansari¹

¹Department of Biosciences, Integral University, Lucknow, India, ²Department of Biology, College of Sciences, University of Hail, Hail, Saudi Arabia, ³Department of Biological Sciences, Faculty of Science, King Abdulaziz University, Jeddah, Saudi Arabia, ⁴Plasma Bioscience Research Center, Applied Plasma Medicine Center, Department of Electrical & Biological Physics, Kwangwoon University, Seoul, South Korea, ⁵Department of Pharmacy and Gachon Institute of Pharmaceutical Science, College of Pharmacy, Gachon University, Incheon, South Korea

Recent times have seen a strong surge in therapeutically targeting the hedgehog (HH)/GLI signaling pathway in cervical cancer. HH signaling pathway is reported to be a crucial modulator of carcinogenesis in cervical cancer and is also associated with recurrence and development of chemoresistance. Moreover, our previous reports have established that carvacrol (CAR) inhibited the proliferation of prostate cancer cells *via* inhibiting the Notch signaling pathway and thus, it was rational to explore its antiproliferative effects in cervical cancer cell lines. Herein, the present study aimed to investigate the anticancer and apoptotic potential of CAR on C33A cervical cancer cells and further explore the underlying mechanisms. We found that CAR significantly suppressed the growth of C33A cells, induced cell cycle arrest, and enhanced programmed cell death along with augmentation in the level of ROS, dissipated mitochondrial membrane potential, activation of caspase cascade, and eventually inhibited the HH signaling cascade. In addition, CAR treatment increased the expression of pro-apoptotic proteins (Bax, Bad, Fas-L, TRAIL, FADD, cytochrome c) and concomitantly reduced the expression of anti-apoptotic proteins (Bcl-2 and Bcl-xL) in C33A cells. CAR mediates the activation of caspase-9 and -3 (intrinsic pathway) and caspase-8 (extrinsic pathway) accompanied by the cleavage of PARP in cervical cancer cells. Thus, CAR induced apoptosis by both the intrinsic and extrinsic apoptotic pathways. CAR efficiently inhibited the growth of cervical cancer cells *via* arresting the cell cycle at G0/G1 phase and modulated the gene expression of related proteins (p21, p27, cyclin D1 and CDK4). Moreover, CAR inhibited the HH/GLI signaling pathway by down regulating the expression of SMO, PTCH and GLI1 proteins in cervical carcinoma cells. With evidence of the above results, our data revealed that CAR treatment suppressed the growth of

HPV⁺C33A cervical cancer cells and further elucidated the mechanistic insights into the functioning of CAR.

KEYWORDS

beta-glucan, antioxidant activity, apoptosis, anticancer, cervical cancer, ROS generation

1 Introduction

With approximately 600,000 newly diagnosed cases and 342,000 demises reported globally during 2020, cervical carcinoma remains the fourth most routinely diagnosed cancers in women and the fourth leading cause of mortality and morbidity in females (Sung et al., 2021). Persistent infection with Human papilloma Virus (HPV) (subtypes 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, and 59) along with other risk factors, including increased parity, infection with HIV and smoking causes cervical cancer (Bosch et al., 2002). Regional discrepancies in the burden of cervical cancer are bleak and reveal the availability, coverage, and quality of preventative interventions and the prevalence of risk factors. Approximately 9/10 women who dies from cervical cancer live in low- and middle-income countries (L/MICs). Inequities are broadening since high-income countries have witnessed a steep decline in the incidence rates, along with some nations moving ahead towards eliminating cervical carcinoma in the upcoming decades (Simms et al., 2019). In contrast, the incidence rate has increased in some sub-Saharan African countries. In these nations, the rate of incidence have either increased or remained mostly at high levels in various eastern European and west-Asian countries (Arbyn et al., 2011; Vaccarella et al., 2017; Arbyn et al., 2020).

Although the management of worldwide screening programs has reduced the occurrence and mortality of cervical carcinoma, however, the incidence of this dreaded disease within the young female population remains a grievous public health concern. Moreover, the systemic use of chemotherapeutic drugs leads to the development of drug resistance which eventually results in poor gynecological outcomes (Moss and Kaye, 2002; Lin et al., 2016). However, there is a constant need of novel drug development which could combat with drug resistance and adverse aftermath associated with the current treatment (Datta et al., 2019).

Phytocompounds are usually non-toxic, proven effective against numerous diseases, and considered a safe, cheap and effective alternative against cancer (Lai and Roy, 2004). Carvacrol (CAR) is a phenolic monoterpenoid abundantly present in the essential oils of oregano and thyme and is well-known for exerting multiple pharmacological effects such as antimicrobial, anticancer, insecticidal, anti-angiogenic, and anti-tumor activity (Sökmen et al., 2004; Baser, 2008). Notably, the Food and Drug Administration (FDA) has authorized the usage of CAR as a food supplement which

testifies its non-toxic nature (Zotti et al., 2013). Furthermore, it is reported that CAR exerts cytotoxic effects on breast, lung, and colon cancer cells; however, the effect of CAR on the proliferation and apoptosis of cervical cancer and its underlying mechanism is not deciphered yet (Dai et al., 2016).

Infection with HPV is regarded as an initial strike that causes cervical carcinoma. Despite that, this factor is not individually sufficient for cancer development. Several additional cellular alterations are needed to commend the action of HPV. Correspondingly, in the present report, we have investigated the effect of CAR on the functionality of Hedgehog (HH) signaling in cervical cancer cells (Samarzija and Beard, 2012). The HH signaling cascade has been demonstrated to play an imperative role in the proliferation, metastasis, recurrence, invasion, drug resistance, and radioresistance of cervical cancer (Liu and Wang, 2019). The binding of HH ligand to its receptor patched (PTCH) activates the HH pathway. This binding relieves the repression from its second receptor Smoothened (SMO), which relocates to the cell membrane and drives a range of reactions leading to the translocation of transcription activators encoded by Glioma associated oncogenes (GLI1, GLI2 and GLI3) into the nucleus and subsequent transcription of target genes (Wu et al., 2020). Previous studies have established that molecular alterations in HH signaling cascade leads to various cancers such as medulloblastoma, basal cell carcinoma, small cell lung cancer, and prostate cancer. Reportedly, HH pathway is found to be hyper activated in cervical cancer and is associated with poor prognosis (Wang et al., 2010).

In this report, we studied the effect of CAR on regulating HH signaling cascade in HPV⁺ C33A cervical cancer cells *via* apoptosis induction and abrogation of cell cycle progression. However, to the best of our knowledge, we are reporting for the first time that CAR suppresses the progression of cervical cancer by inhibiting the HH signaling pathway.

2 Materials and methods

2.1 Reagents and chemicals

Carvacrol (CAR), DAPI (4, 6-diamidino-2-phenylindole), propidium iodide (PI), and 2, 7-dichlorodihydrofluorescein diacetate (DCFH-DA) were purchased from Sigma (St. Louis, MO, United States). Caspase-9, -8 and -3 colorimetric assay kit with catalogue numbers K119, K113-25 and K106-100 were

procured from BioVision, United States. Acridine orange, Ethidium bromide, RPMI-1640, fetal bovine serum (FBS), 1% antibiotic-antimycotic solution, RNase A, 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyl tetrazolium bromide (MTT) and HiPurATM Total RNA Miniprep Purification Kit were purchased from Himedia India, Ltd., Mumbai, India. JC-1 mitochondrial membrane potential (MMP) assay kit was purchased from G-Biosciences, United States. All the primer sequences utilized during the study were procured from IDT, United States. FITC Annexin V Apoptosis Detection Kit was procured from BD Bioscience, PharMingen (San Diego, United States of America). DyNAmoColorFlash SYBR Green qPCR Kit and Verso cDNA synthesis kit were obtained from Thermo-Scientific, United States.

2.2 Cell line maintenance

Human cervical cancer cell line (C33A) was procured from the national repository division of the National Centre for Cell Sciences (NCCS), Pune, India. C33A cells were grown and maintained in RPMI-1640 completed with FBS (10%) and antibiotic-antimycotic solution (1%) under optimal culture conditions (temperature: 37 °C and 5% CO₂).

2.3 Methods

2.3.1 Cell proliferation assay

Cell viability was assessed by using MTT assay as described previously (Ansari et al., 2021). Briefly, C33A cells were cultured in 96-well plates (5×10^3 cells/well). C33A cells were maintained in the presence of CAR for 24 and 48 h. 10 μ l MTT working dye (5 mg/ml) was added to the cultured C33A cells and then incubated for an additional 4 h. After that, media was removed and 100 μ l DMSO was added to dissolve MTT crystals. Absorbance intensity was analyzed at 570 nm by an ELISA microplate reader (Bio-Rad, United States of America). Viability of C33A cells was calculated as a ratio of the optical density of treated and untreated cells.

2.3.2 LDH assay

Human cervical cancer cells (C33A) were cultured into a 96-well plate with a growth medium. LDH activity was determined in CAR-treated C33A cells according to the manufacturer's protocol. The LDH activity in treated and untreated cells was determined by evaluating the optical density of the cells at 490 nm using the ELISA reader (Bio-Rad, United States).

2.3.3 Observation of cell morphology

C33A cells plated on 96-well plate were treated with CAR doses (25, 50, 75 and 90 μ M) for 24 and 48 h. The CAR-treated and control cells were then observed for

morphological alterations within the C33A cervical cancer cells using a fluorescence microscope (Thermo-Scientific, United States of America), and photomicrographs were captured.

2.3.4 Colony formation assay

C33A cells were allowed to attach in each well on a 6-well plate at low counts of approximately 400–500 cells/well and were subsequently used for colony formation assay. Initially, the stated number of C33A cells was exposed to a previously stated concentration of CAR and were left undisturbed under optimum tissue culture conditions for 2 weeks. The colonies were treated and stained using crystal violet stain (0.1%). Aggregates of 40 or more cells were included in the counting as an individual colony.

2.3.5 ROS assay

The intracellular levels of ROS were determined in C33A cervical cancer cells after treatment with various doses of CAR by H₂DCFDA staining protocol, as described briefly (Ahmad and Ansari, 2021). Briefly, following drug treatment for 12 h, cells were subsequently re-exposed to H₂DCFDA (25 μ M) for 30 min in darkness at 37°C. The fluorescent micrographs were captured under the green fluorescence channel of the FLoid Imaging Station, Thermo-Fischer Scientific, United States.

In addition, ROS generation post-CAR exposure was also quantified through flow cytometric evaluation, as stated previously (Ahmad and Ansari, 2021). 5×10^5 C33A cells were allowed to adhere in each well of a 6-well plate and after that, exposed to state concentrations of CAR as stated above. The cells were then pelleted, resuspended in 25 μ M H₂DCFDA and incubated briefly for 30 min in the dark at RT. Post-incubation, the cells were re-pelleted and re-suspended in PBS (1X). The suspension was eventually analyzed using the FITC channel of the FACSCalibur flow cytometer (BD Biosciences, United States of America).

2.3.6 GSH analysis

Reduced GSH or glutathione levels were quantified using a commercially available GSH kit (BioVision, Mountain View, CA, United States of America) following the supplier's instructions. Briefly, CAR-treated and untreated C33A cells were exposed to ice-cold GSH buffer (100 μ l) for homogenization. The resulting homogenate was placed in a separate test tube containing chilled HClO₄ (10 ml) and vortexed for nearly 1 min. Subsequently, the homogenate was pelleted (13,000 \times g for 2 min). The supernatant was collected and mixed with KOH in a ratio of 2:1, and after 5 min, the suspension was re-pelleted at the same force (13,000 \times g), followed by the collection of supernatant for the remaining assay protocol. During the concluding steps, 10 ml of supernatant from different groups was diluted by reconstitution in 80 ml of assay buffer. Finally, the absorbance of fluorescence intensity was read at an excitation/emission ratio

of 340/450 nm using a fluorimeter (Thermo-Fischer Scientific, United States of America).

2.3.7 DAPI/PI staining

C33A cells treated with different CAR doses (25, 50, 75, and 90 μM) were collected and fixed in ice-cold methanol for 15 min at -20°C . The cells were stained with DAPI and PI for 30 min at 37°C and analyzed for blue and red merged fluorescence using FLoid Imaging Station, Thermo-Fischer Scientific, United States.

2.3.8 Qualitative staining for apoptosis assessment

CAR-treated and untreated C33A cells at the above-stated concentrations for 24 h were carefully washed using tissue culture grade PBS (1X) and then treated with solution constituted by equal concentrations (100 $\mu\text{g}/\text{ml}$) of Acridine orange (AO) and Ethidium bromide (EtBr) for 30 min. The cells were carefully washed after incubation, visualized, and their red/green merged fluorescence was recorded using FLoid Imaging Station, Thermo-Fischer Scientific, United States.

2.3.9 Assessment of mitochondrial membrane potential ($\Delta\Psi\text{m}$)

C33A cells were exposed to the different concentrations of CAR (25, 50, 75, and 90 μM) for 24 h and then stained with JC-1 dye (200 μM) which gets accumulated within mitochondria following a potential-dependent trend. After that, the cells were pelleted and washed with pre-warmed 1X MMP buffer, and image acquisition was performed by FLoid Imaging Station, Thermo-Fischer Scientific, United States of America. Furthermore, alteration in $\Delta\Psi\text{m}$ was quantified using a FACS Calibur flow cytometer (BD Biosciences, United States of America) as per the instructions of the manufacturer JC-1 MMP assay kit. Increased green fluorescence with concomitantly reduced levels of red fluorescence signified dissipated $\Delta\Psi\text{m}$.

2.3.9 Measurement of caspase activities

CAR treatment was given to C33A cells as per above-stated concentration for 24 h in a 96-well plate. The activities of caspase-3, -8 and -9 was assessed using colorimetric kit available commercially and following the instructions supplied by the manufacturer. The observations were interpreted as percentage (%) change in the activities of stated key caspases in comparison with CAR untreated control C33A cells.

2.3.10 Assessment of caspase inhibitors pre-treatment

Cervical cancer cells were initially treated with inhibitors (50 μM ; 2 h) specific for caspase-3, -8 and -9, namely Z-DEVD-FMK, Z-IETD-FMK and Z-LEHD-FMK, respectively. Then, C33A cells were treated with the above -stated concentrations

of CAR for 24 h. Finally, the viability of cells was calculated using MTT assay as stated in 2.3.1.

2.3.11 PARP estimation

Quantitative assessment of cleaved PARP levels was assessed in C33A cells using a Human PARP ELISA kit as per the manufacturer's instruction. The absorbance of cleaved PARP was analyzed using a spectrophotometer (Bio-Rad, United States) at 450 nm.

2.3.12 Cytochrome-c release assay

The total concentration of cytochrome-c present within the total protein content of CAR-treated and untreated C33A cells was quantified using an ELISA kit (Thermo-Fischer Scientific, United States) by following the instruction from the manufacturer.

2.3.13 Apoptosis quantification

Apoptosis instigated in C33A cells upon exposure to CAR was quantified through flow cytometry as per the manufacturers' protocol. Approximately 5×10^5 C33A cells were exposed to the above-stated concentrations of CAR for 24 h. Subsequently, the cells (including the detached ones) were collected and pelleted. Subsequently, the pellets were re-suspended in Annexin V-FITC and PI solution of apoptosis detection kit according to the manufacturer's recommendations (BD Biosciences, United States). The suspension was then promptly evaluated through FACS Calibur (BD Biosciences, United States), and the interpretation of results was made as a percentage (%) live cells, early apoptotic and late apoptotic or necrotic cells based on Annexin V-, PI-; Annexin V+, PI- and Annexin V+, PI + staining respectively. Moreover, total apoptosis induced in C33A cells was expressed as a percentage of early and late apoptotic C33A cells.

2.3.14 Assessment of cell cycle progression

As previously described, the analysis of cell cycle progression with CAR-treated C33A cells was quantified using a flow cytometer (Ahmad et al., 2022). For the assessment, C33A cells (5×10^5 cells/well) were treated with the above-stated CAR concentrations for 24 h. After that, the cells from each group were trypsinized and fixed using chilled methanol (-20°C for 15 min). Subsequently, the cells were treated with RNase A for 30 min at RT, followed by incubation with PI for 1 h. Eventually, the samples were evaluated through FACSCalibur (BD Biosciences, United States of America).

2.3.15 Real-time PCR (qPCR) analysis

After treatment of 1×10^6 C33A cells/group with the above-stated concentrations of CAR, the total RNA content was isolated using a commercially available RNA isolation kit. The extracted RNA (2 μg) was used to prepare cDNA using the Verso cDNA synthesis kit per the manufacturer's instructions. qPCR analysis

TABLE 1 List of primers used for qPCR.

S. No.	Target gene	Sequence of primers	
		Forward (5'-3')	Reverse (3'-5')
1	GAPDH	GAAATCCCATCACCATCTTCCAGG	GAGCCCCAGCCTTCTCCATG
2	Bcl2	GATTGTGGCCTTCTTTGAG	CAAAGTGAAGAGAGTCTTC
3	Bcl-X _L	CAGAGCTTTGAACAGGTAG	GCTCTCGGGTGTGTATTG
4	Bax	GCCCTTTTGCTTCAGGGTTT	TCCAATGTCCAGCCCATGAT
5	c-myc	AGCGACTCTGAGGAGGAACAAG	GTGGCACCTCTTGAGGACCA
6	Cyclin D1	CCGTCCATGCGGAAGATC	GAAGACCTCCTCCTCGCACT
7	PTCH1	GGGTGGCACAGTCAAGAACAG	TACCCCTTGAAGTGTCTGTACA
8	SMO1	CTATTCACTCCCGCACCAAC	CAGTCAGCCACAGGTTCTC
9	GLI1	GAAGTCATACTCACGCCTCGAA	CAGCCAGGGAGCTTACATACAT
10	Fas	CGGACCCAGAATACCAAGTG	CCAAGTTAGATCTG
11	Fas-L	GGGG	GTGGCCTAT
		ATGTT TCAGCTCTTCC-3	TTG CTT CTCCA
12	CDK4	CCTGGCCAGAATCTACAGCTA	ACATCTCGAGGCCAGTCATC
13	Bad	CCTCAGGCCTATGCAAAAAG	AAACCCAAAACCTCCGATGG

was performed on ABI-7500 real-time PCR (Applied Biosystems) as per the stated instructions of DyNamoColorFlash SYBR Green qPCR Kit. The sequence of all the primers involved in the investigation was optimized using the NCBI pick tool, as listed in [table 1](#). The normalizations were made using GAPDH as a housekeeping gene and the results were interpreted using the $2^{-\Delta\Delta CT}$ method.

2.3.16 In silico analysis

To determine the binding effect of the protein and ligands, GLI (PDB ID: 2GLI) and SMO (PDB ID: 4JKV) were docked with Carvacrol (PubChem ID:10,364), Itraconazole (PubChem ID: 55,283), and Cyclopamine (PubChem ID: 442,972). All these 3-dimensional structures were retrieved from the PDB database (<https://www.rcsb.org/>) and PubChem (<https://pubchem.ncbi.nlm.nih.gov/>). AutoDock Vina four was used for molecular docking. AutoDock Vina is a flexible molecular docking program that has generated nine different docked poses for protein-ligand complexes. The best-docked position was chosen from nine conformations depending on the interacting residues, such as hydrogen bonds with a high binding affinity (kcal/mol). The protein-ligand interaction of docked complexes was presented in two dimensions for interaction analysis of the protein-ligand complex by using LigPlus (Ahmad et al., 2022). The proteins, ligands, and their binding pockets for the protein's three-dimensional structure were generated using PyMol (Ahmad et al., 2014; Hassan et al., 2014; Baig et al., 2016; Khan et al., 2016; Khan et al., 2021; Khan et al., 2022a; Khan et al., 2022b).

2.3.17 Statistical inferences

The quantitative observations reported are the mean \pm SEM of individual experiments performed thrice in triplicate. Statistical significance was calculated after applying one-way ANOVA followed by Dunnett *post hoc* and two-tailed, paired Student's t-test as per the suitability. *, ** and *** represents $p < 0.05$; $p < 0.01$ and $p < 0.001$ respectively in comparison with untreated control.

3 Results

3.1 CAR inhibits proliferation and clonogenic potential of C33A cervical cancer cells

We studied the antiproliferative effects of CAR in human cervical cancer cells. C33A cells were cultured with increasing concentrations of CAR (15, 25, 50, 75, and 90 μ M) for 24 h. CAR significantly inhibited the viability of C33A cells to $86.47 \pm 2.62\%$, $64.56 \pm 2.25\%$, $50.75 \pm 3.28\%$, $36.39 \pm 2.68\%$, and $22.56 \pm 3.02\%$, respectively, as compared to the control cells, and the population of viable C33A cells declined with increase in CAR concentration ([Figure 1A](#)). This effect was more pronounced after treatment of CAR for 48 h and the viability was further reduced to $67.82 \pm 3.72\%$, $54.15 \pm 2.50\%$, $46.81 \pm 3.92\%$, $28.50 \pm 2.87\%$, and $12.56 \pm 1.82\%$ ([Figure 1A](#)) respectively as compared to the untreated control cells. Subsequently, the

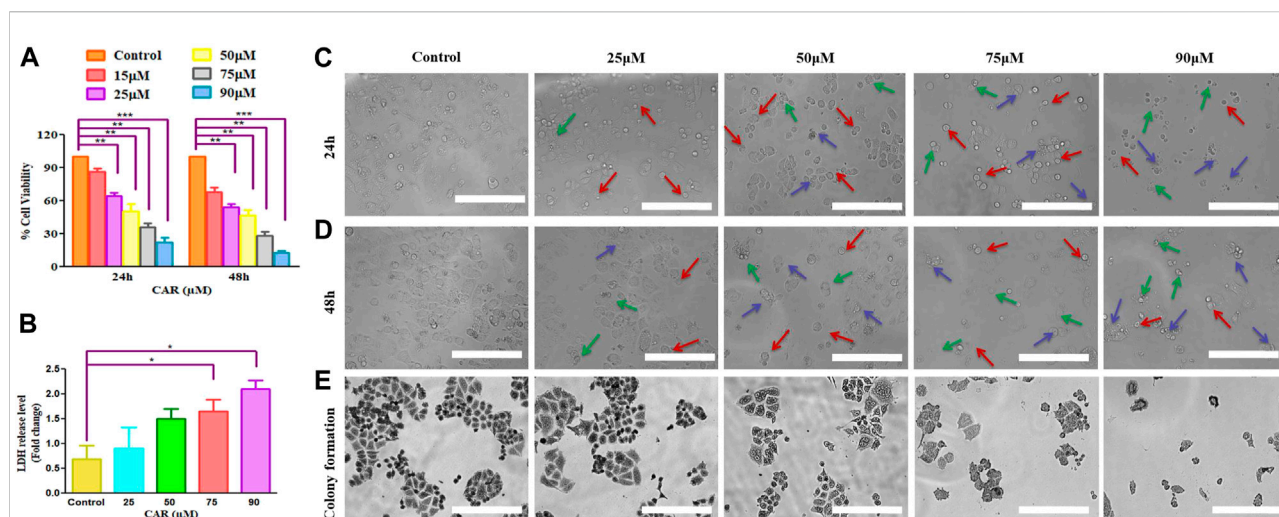


FIGURE 1

CAR effectively suppressed the growth and colony formation of cervical cancer cells. (A) Bar graph represents the percent (%) viability of C33A cells as compared to the vehicle-treated cells after treatment with CAR for 24 and 48 h (B) Bar graph represents the percent (%) LDH released as analyzed by LDH assay in CAR-treated C33A cell. (C,D) Morphological alterations in C33A cells after treatment with CAR (25–90 μ M) for 24 and 48 h as indicated by red, blue and green arrows representing swelling, shrinkage and disintegration of cell organelles. (E) CAR treatment reduced the number of colonies in a dose-dependent manner after treatment with various doses of CAR. Photomicrographs were captured at 20X magnification (scale bar = 100 μ m). Each value in the bar graph represents the mean \pm SEM of three independent experiments. Significant difference among the treatment groups were analyzed by one-way ANOVA followed by Dunnett *post-hoc* test (* p < 0.05, ** p < 0.01, *** p < 0.001 represent significant difference compared with control).

release of LDH in CAR-treated C33A cells was quantified. We observed a significant release of LDH to approximately 2.08 folds in C33A cells post-exposure to CAR. Thus, these results of MTT and LDH assays suggested that CAR suppressed the growth and proliferation of cervical cancer cells.

The morphological analysis of CAR-treated C33A cells also showed several morphological alterations compared to the untreated cells. It was observed that the treatment with increasing doses of CAR induced substantial morphological aberrations such as cell shrinkage, rounding of cells, and blebbing of the plasma membrane and disintegration of cell organelles in cervical cancer cells (Figures 1C,D). Thus, CAR exerts antiproliferative effects by inhibiting the proliferation of cervical cancer cells.

Moreover, we performed colony formation assay to study the effect of CAR on cervical cancer C33A cells. Our data showed that C33A cells treated with CAR at indicated concentrations (25–90 μ M) for 24 h demonstrated small and few colonies relative to the control cells. Thus, these findings implicated that CAR treatment suppresses cervical cancer cell growth and clonogenic potential (Figure 1E).

3.2 Assessment of ROS and GSH levels in CAR-treated C33A cells

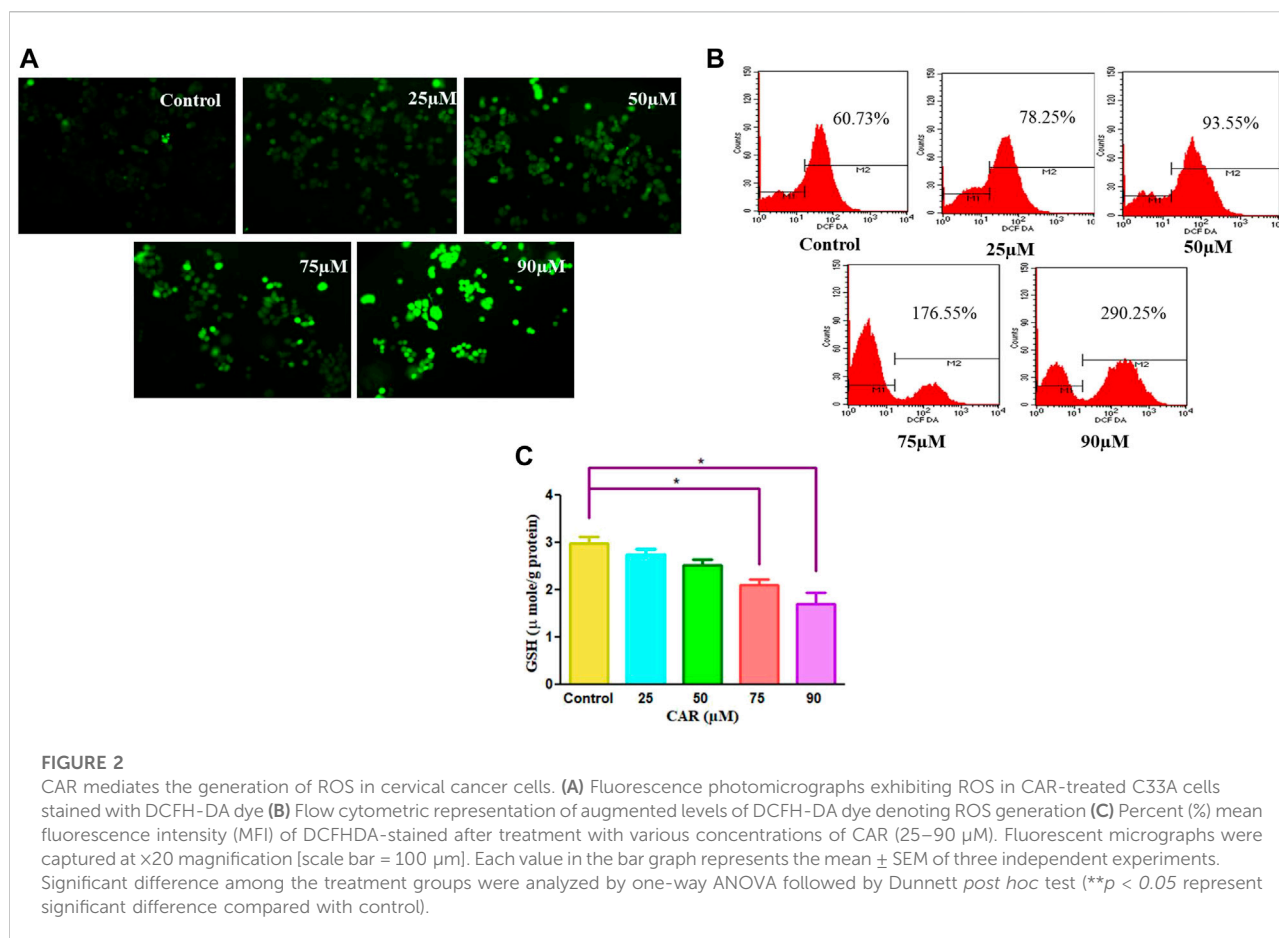
Several physiological and biochemical processes occurring during homeostatic conditions are responsible for ROS

production. It is established that ROS play an imperative role in the pathobiology of multiple diseases (Shamas-Din et al., 2013). ROS-inducing property of CAR in C33A cells was assessed qualitatively and quantitatively by fluorescence microscopy and fluorometrically, respectively. It was observed that the amount ROS levels increased by nearly 5-fold in C33A cells, whereas in the untreated cells, the levels of ROS were found to be 60.73% which augmented to 290.25% in treated C33A cells (Figures 2A,B).

GSH is a well-reputed member of the antioxidant family involved in imparting protection against ROS-mediated injury to the cell by inhibiting lipid peroxidation and eliminating hydrogen peroxide (H_2O_2). Thus, to find out the effects of CAR on the cellular redox environment, GSH levels were studied. It was noted that CAR treatment reduced the GSH levels in cervical cancer cells (Figure 2C).

3.3 CAR promotes apoptosis in C33A cells

To delineate the practical implications of CAR treatment as a plausible therapeutic against cervical cancer, cell-based apoptosis was investigated. We primarily studied the morphological alterations under the microscope occurred within the CAR-treated C33A cells. As shown in Figure 3A, treatment with CAR induced condensation, fragmentation and margination of chromatin around the nucleus of C33A cells, which is indicative



of programmed cell death evaluated by DAPI/PI staining. An enhanced number of C33A cells were observed exhibiting bright blue and red fluorescence indicating condensed or fragmented nuclei, considered as marked characteristics of apoptosis after the treatment with CAR (Figure 3A).

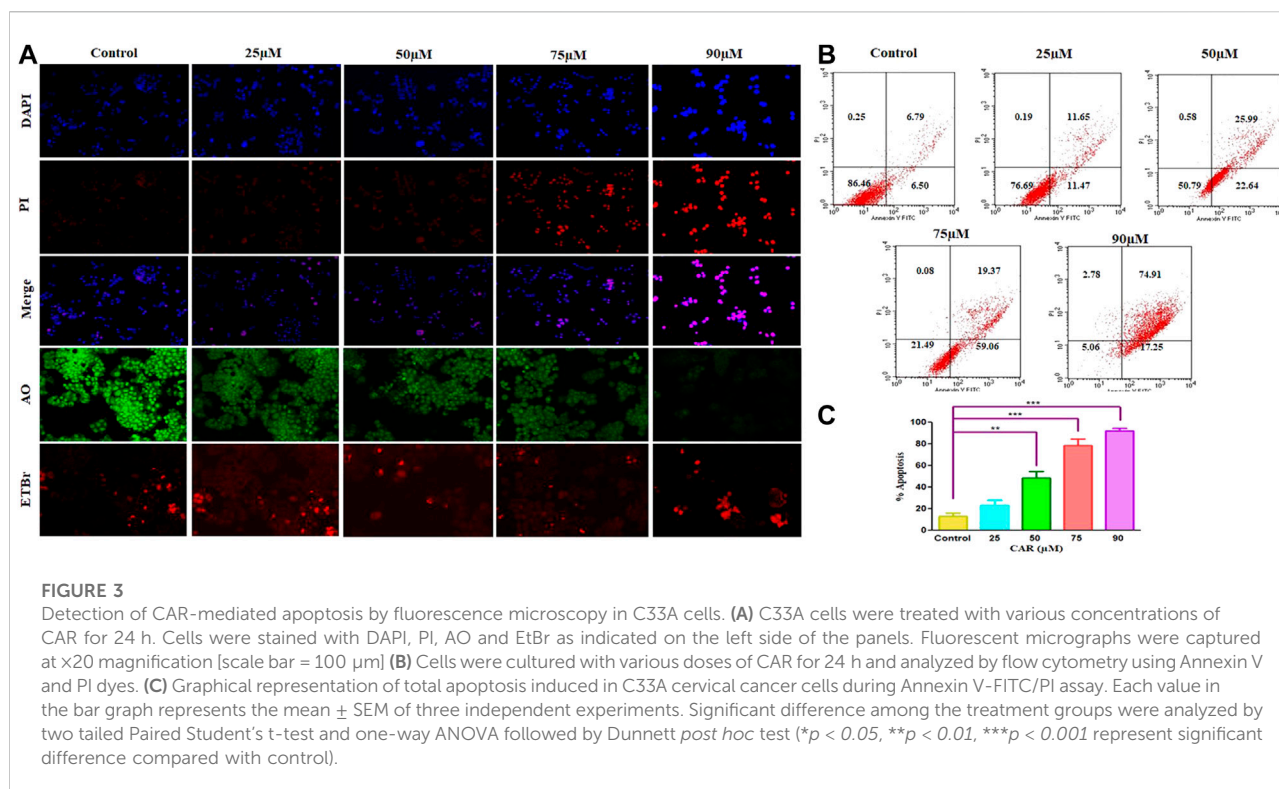
Furthermore, to investigate the underlying reason for the increase in cell death in cervical carcinoma cells, AO/EB staining was performed in CAR-treated and untreated cervical cancer cells. AO and EB dyes were used to distinguish between live and dead cells based on membrane integrity. AO is reported to infuse within the live DNA and subsequently gives green fluorescence to the cells. Contrastingly, EB uptake occurs only in non-viable cells, where it also infuses in the cellular DNA and gives the nucleus of dead cells characteristic red fluorescence. The fluorescent micrographs, untreated control cells displayed normal morphology of nucleus exhibiting bright green and diffused red fluorescence, whereas CAR-treated cells exhibited bright red fluorescence and diffuse green fluorescence indicating condensed and fragmented nuclei as shown in Figure 3A. Thus, treatment with CAR induced cell death in cervical carcinoma cells.

Moreover, apoptosis was also reaffirmed by performing Annexin V-FITC/PI cells. The total amount of apoptosis in

CAR-treated C33A cells was calculated by summation of cells undergoing early and late apoptosis characterized by Annexin V-FITC⁺, PI⁻ and Annexin V-FITC⁺ and PI⁺ respectively. The observations indicated that apoptosis escalated within CAR-treated C33A cells, which was dependent on the concentration and time of exposure to CAR compared with untreated control (Figures 3B,C).

3.4 CAR exposure augmented caspase activities in C33A cells

Apoptosis, or programmed cell death, is pivotal for maintaining a homeostatic environment within a multicellular organism. Mechanistically apoptotic cell death is primarily classified as intrinsic- and extrinsic apoptosis. The intrinsic apoptosis pathway is also known as the mitochondrial pathway and is characterized by the dissipated potential of mitochondrial membrane leading to apoptosome formation, caspase activation, and activation of downstream effector caspases (caspase-3) (Birben et al., 2012; Zorov et al., 2014). The activities of caspase-3, -8, and -9 was considerably elevated



by $41.36 \pm 3.30\%$, $67.46 \pm 4.15\%$, $97.54 \pm 4.20\%$, and $158.67 \pm 2.47\%$; $13.54 \pm 4.55\%$, $29.76 \pm 3.93\%$, $55.05 \pm 2.84\%$, and $\pm 96.30 \pm 3.75\%$ and $32.98 \pm 4.12\%$, $54.06 \pm 4.56\%$, $77.25 \pm 4.80\%$, and $122.51 \pm 2.10\%$, respectively in comparison with CAR untreated control C33A cells (Figure 4A). The data demonstrated that treatment of CAR increased caspase-8, -9 and -3 activities in a dose-proportional manner. Cells were pre-exposed with specific caspase inhibitors to affirm that the activation of caspase is mediated by the treatment of CAR in C33A cells. The result revealed that pretreatment with caspase inhibitors completely abrogated the CAR-induced apoptosis in C33A cells (Figure 4B), which directly indicated that CAR-instigated apoptosis in C33A cells strongly correlates with the activation of key caspases.

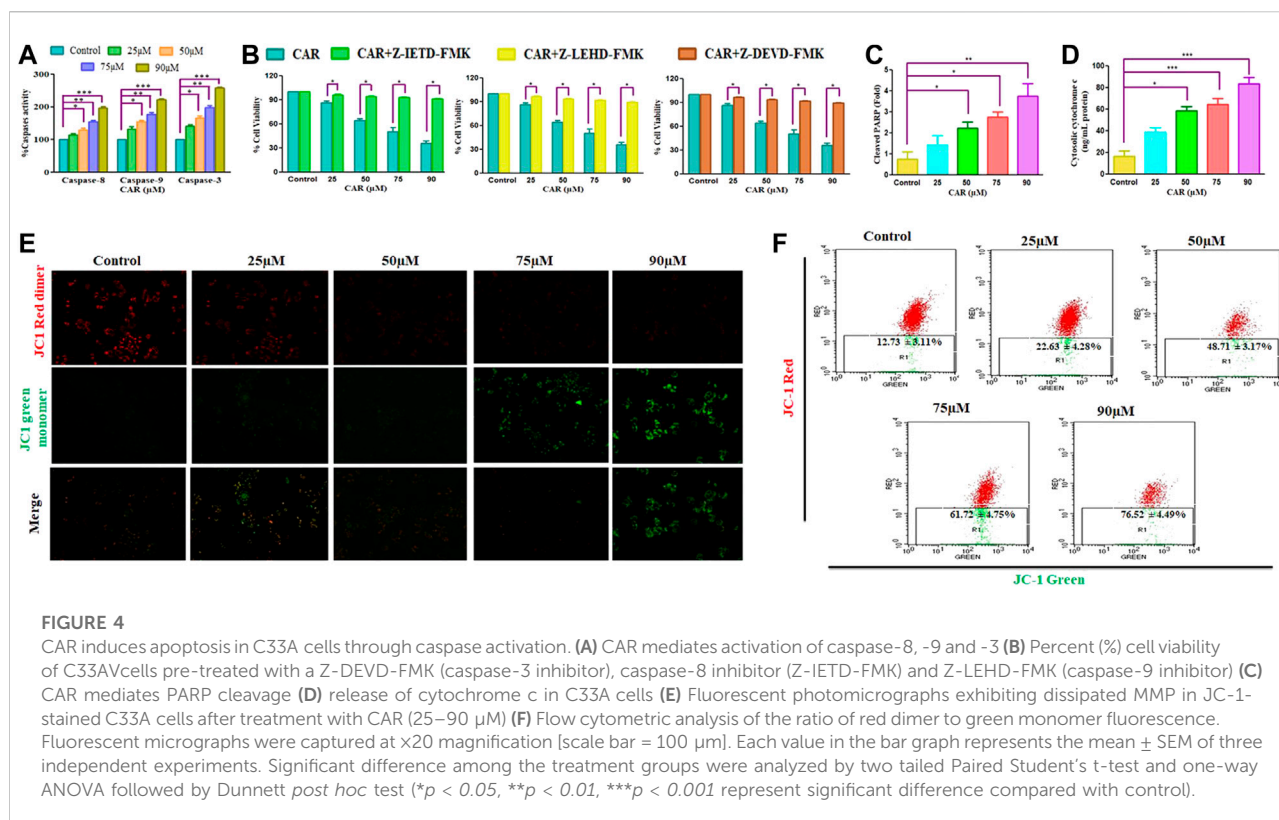
Dissipated mitochondrial membrane potential ($\Delta\Psi\text{m}$) is a major event occurring in mitochondria that leads to apoptosis (Shamas-Din et al., 2013). The $\Delta\Psi\text{m}$ disruption results in the altered potential of the mitochondrial membrane resulting in imbalanced oxidation-reduction potential within mitochondria. To examine whether CAR could disrupt $\Delta\Psi\text{m}$, membrane-permeant JC-1 dye was used to detect alteration in $\Delta\Psi\text{m}$ in CAR-treated C33A cells. JC-1 is highly sensitive to mitochondrial potential, which gives red and green fluorescence for polarized and non-polarized mitochondria. JC-1 exhibits a shift from green to red fluorescence corresponding to approximately 529–590 nm. As a result, the dissipation of $\Delta\Psi\text{m}$ is characterized by reduced red: green ratio of JC-1 mediated fluorescence intensity due to the generation of

J-aggregates. The merged fluorescent photomicrographs demonstrated that treatment with CAR (25, 50, 75, and 90 μM) for 24 h resulted in mitochondrial depolarization suggested by the decrease in the red to green intensity ratio in C33A cervical carcinoma cells. With the increase in CAR concentration, $\Delta\Psi\text{m}$ decreased in cervical cancer C33A cells (Figure 4E). Furthermore, $\Delta\Psi\text{m}$ dissipation was also quantitatively measured in CAR-treated cells by flow cytometry (Figure 4F). In C33A cells, 12.73% and 76.52% cell population was found positive for dissipated $\Delta\Psi\text{m}$ in CAR untreated control and treated cells respectively, showing a 5.01 fold increase in the depolarized cell population. The results showed that CAR treatment dissipated $\Delta\Psi\text{m}$ in C33A cells proportionally with its concentration (Figures 4E,F).

Moreover, treatment of CAR significantly enhanced the level of PARP by 3.94 folds, a well-known marker of apoptosis in a dose-related manner, along with an enhanced level of cytochrome c by four folds in C33A cells, validating the involvement of mitochondria in CAR-induced apoptosis (Figures 4C,D).

3.4 Crosstalk between cell survival/death genes in CAR treated C33A cells

The pro- (Bad, Bax) and anti-apoptotic genes (Bcl- X_L , Bcl-2, Mcl-1) are members of the Bcl-2 protein family that tightly regulates the mitochondrial viability, caspase activation and the release of cytochrome c (Wang et al., 2019). Therefore, we performed real-



time PCR to evaluate the mRNA expression of Bcl-2 family proteins. As demonstrated in Figures 5A,B, an increased expression of pro-apoptotic genes with a concomitant decline in anti-apoptotic gene expression after CAR exposure in C33A cells. In treated C33A cells, Bcl-2 and Bcl-X_L expression level was decreased (by 0.29 and 0.26 folds), while Bax and Bad expression were elevated (by 4.50 and 5.14 folds) (Figures 5C,D). The exposure of C33A cells with CAR considerably elevated pro-apoptotic gene expression and significantly reduced anti-apoptotic gene expression. These findings revealed that members of the Bcl-2 family played a central role during CAR-mediated apoptotic cell death in C33A cells.

The extrinsic pathway is initiated through transmembrane death receptors through the involvement of FasL and TRAIL. These activate the death receptor with subsequent downstream activation of FADD and caspase-8, resulting in the onset of apoptosis. After treatment of C33A cells with increased mRNA levels of FasL, TRAIL and FADD by 5.26, 4.26 and 5.46 folds expression in CAR-treated C33A cervical cancer cells (Figures 5E–G).

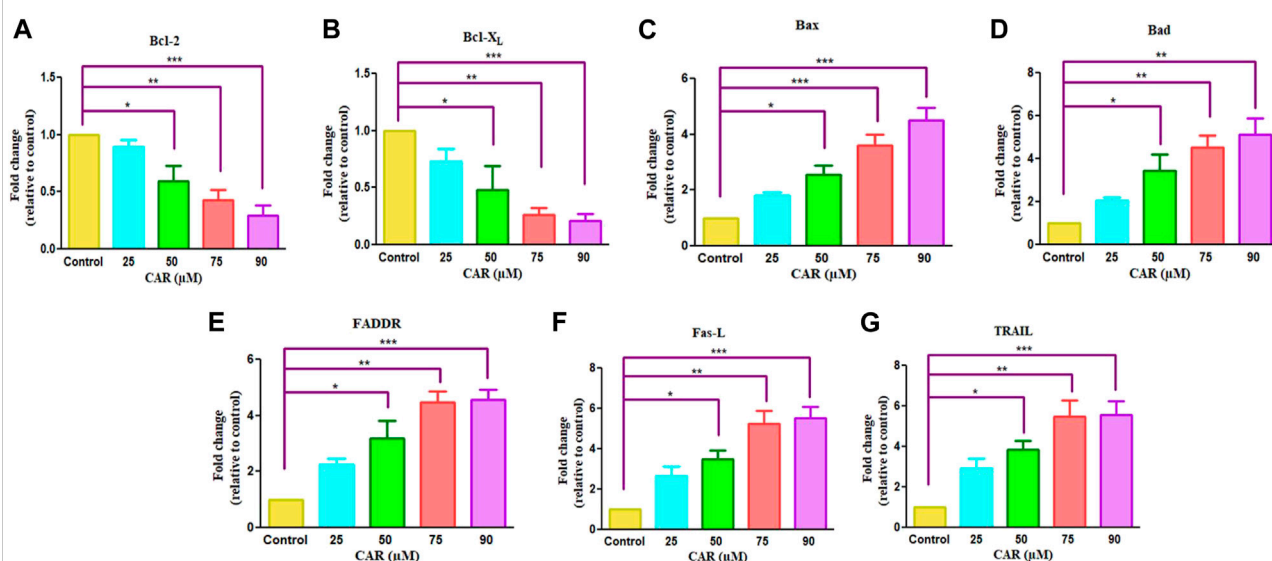
3.6 Quantification of G0/G1 phase population in CAR-treated C33A cells

Cell cycle arrest is a plausibly effective therapeutic target for the clinical management of various cancers (Chen et al., 2002). Exposure

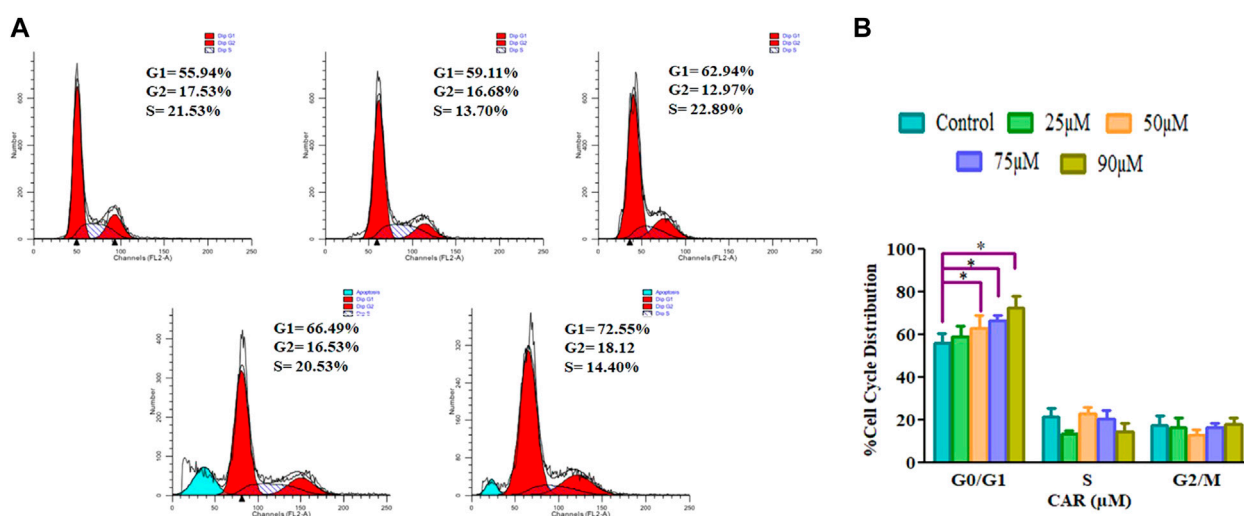
of C33A cells to CAR increased the number of C33A cells by $59.11 \pm 4.10\%$, $62.94 \pm 5.19\%$, $66.49 \pm 3.27\%$ and $72.55 \pm 4.55\%$ in the G0/G1 phase of the cell cycle, respectively comparatively with untreated control C33A cells ($55.94 \pm 3.79\%$) (Figures 6A,B). The qRT-PCR analysis also reaffirmed the above results by indicating that CAR exposure decreased cyclin D1, c-myc and cyclin-dependent kinase 4 (CDK4) expression in C33A cells. Furthermore, our data also exhibited that CAR treatment dose-proportionally enhanced the mRNA levels of p21 in C33A cells as compared to untreated groups (Figures 7A–D).

3.7 CAR impeded HH/GLI signaling and mediated loss of C33A cell viability

To investigate the involvement of CAR-mediated inhibition of the HH signaling pathway in the underlying molecular mechanism of CAR on C33A cervical cancer cells, we studied the mRNA expression of key components of the HH signaling pathway by qRT-PCR analysis. As the results shown in Fig., CAR treatment (25–90 μ M) for 24 h reduced GLI1 and SMO mRNA levels in C33A cells dose-dependently. The reduction in mRNA levels of GLI1 and SMO was found to be 0.84 ± 0.04 , 0.65 ± 0.19 , and 0.24 ± 0.05 fold; 0.79 ± 0.03 , 0.57 ± 0.13 , and 0.34 ± 0.03 fold for 25, 50, 75, and 90 μ M of CAR treatment respectively in

**FIGURE 5**

Effect of CAR on the mRNA expression of intrinsic and extrinsic apoptotic signaling molecules in C33A cells. Fold change in the mRNA expression of antiapoptotic proteins and proapoptotic proteins such as (A) Bcl-2, (B) Bcl-X_L, (C) Bax, (D) Bad, (E) FADD, (F) Fas-L and (G) TRAIL genes were analyzed by real-time PCR using SYBR Green dye. Target gene expression is normalized to GAPDH mRNA expression and the results are expressed as fold change from control. Data reported are mean \pm SEM of three separate experiments each of which were performed in triplicates. Significant difference among the treatment groups were analyzed by two tailed Paired Student's t-test and one-way ANOVA followed by Dunnett *post hoc* test (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$ represent significant difference compared with control).

**FIGURE 6**

CAR-induced cell cycle arrest in cervical cancer cells (A) Cell cycle distribution of propidium iodide-stained C33A cells treated with CAR (25–90 μ M) for 24 h observed by flow cytometric analysis. Data shown are representative of three independent experiments. (B) Graphical representation of percent cell cycle distribution in C33A cervical carcinoma cells as determined by flow cytometric analysis. Data reported are mean \pm S.E.M of three individual experiments performed in triplicate. Significant difference among the treatment groups were analyzed by one-way ANOVA followed by Dunnett *post-hoc* test (** $p < 0.05$ represent significant difference compared with control).

comparison to the control cells (Figures 8A,B). Moreover, CAR treatment also reduced the mRNA expression of PTCH1 by 0.82 ± 0.05 , 0.54 ± 0.12 , and 0.27 ± 0.06 fold in C33A cervical

cancer cells (Figure 8C). Thus, the results suggested that CAR inhibited hyper activated HH signaling in C33A cervical cancer cells.

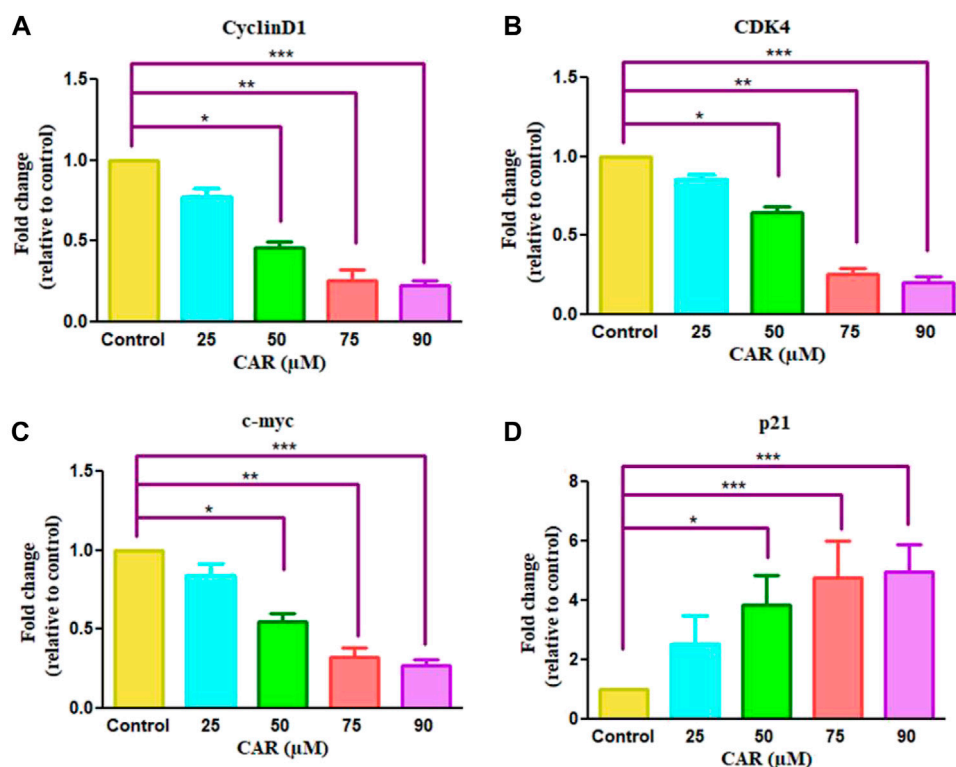


FIGURE 7

Effect of CAR on the mRNA expression of cell cycle regulatory genes in C33A cells. Fold change in the mRNA expression of antiapoptotic proteins and pro-apoptotic proteins such as (A) cyclinD1 (B) CDK4 (C) c-myc and (D) p21 genes were analyzed by real-time PCR using SYBR Green dye. Target gene expression is normalized to GAPDH mRNA expression and the results are expressed as fold change from control. Data reported are mean \pm SEM of three separate experiments each of which were performed in triplicates. Significant difference among the treatment groups were analyzed by two tailed Paired Student's t-test and one-way ANOVA followed by Dunnett *post hoc* test (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$ represent significant difference compared with control).

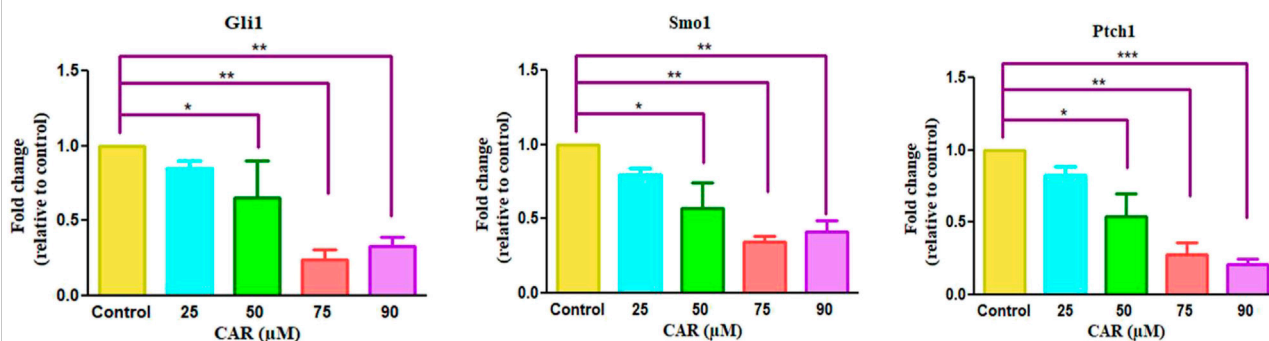
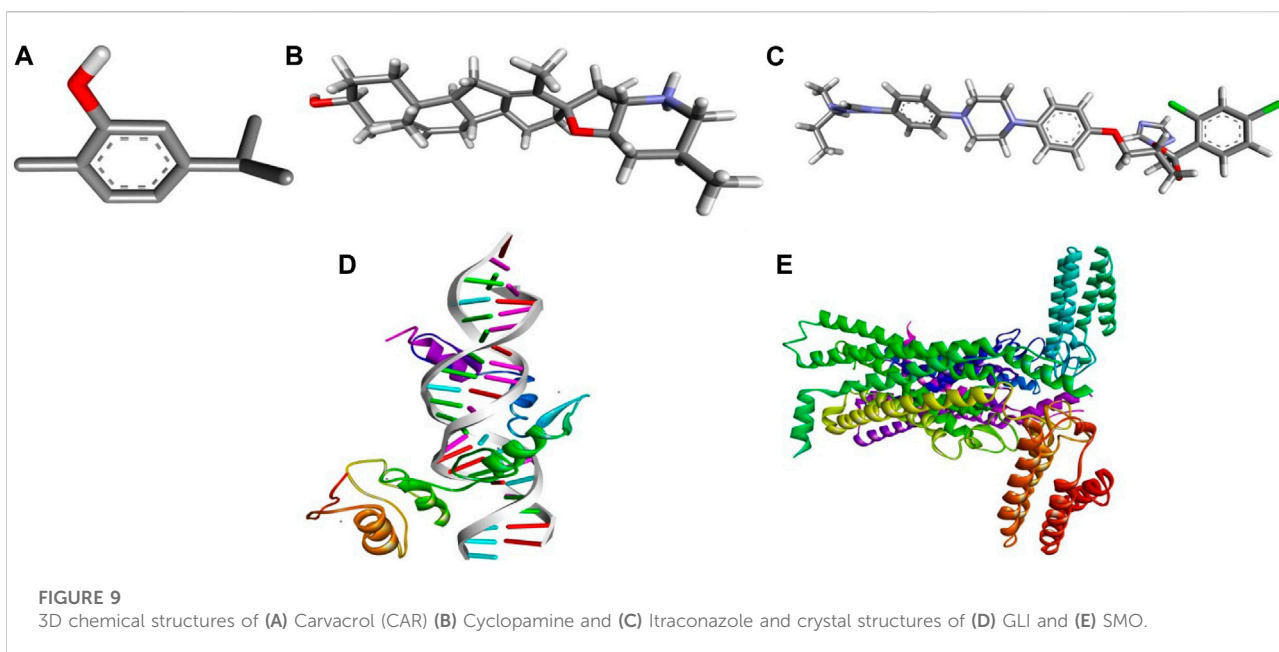


FIGURE 8

Effect of CAR on mRNA expression of key components of HH signaling pathway in cervical carcinoma cells. The graphs show fold change in mRNA expression of GLI1, SMO1 and PTCH1 relative to control in C33A cells after treatment with CAR (25–90 μM) as measured by qPCR using SYBR Green dye. Target gene expression is normalized to GAPDH mRNA expression and the results are expressed as fold change from control. Data reported are mean \pm SEM of three separate experiments each of which were performed in triplicates. Significant difference among the treatment groups were analyzed by two tailed Paired Student's t-test and one-way ANOVA followed by Dunnett *post hoc* test (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$ represent significant difference compared with control).



3.4 CAR showed high binding affinity toward GLI and SMO protein

Cyclopamine is a steroidal alkaloid with intrinsic teratogenic and anticancer attributes, which results from its ability to impede HH signaling. Cyclopamine inhibits HH pathway activation by binding directly to SMO (Deng et al., 2020). Furthermore, itraconazole is an antifungal drug primarily used as an antimycotic. Recent reports have demonstrated the efficacy of Cyclopamine in impeding HH signaling by inhibiting smoothened receptors (SMO), glioma-associated oncogene homologs (GLI), and subsequent downstream effectors. The 3D structures of ligands of interest, such as carvacrol (CAR), Cyclopamine and itraconazole, are shown in Figures 9A–C. The crystal structures of proteins such as GLI and SMO are depicted in Figures 9D,E. Initially, different types of interaction studies were performed between CAR, Cyclopamine and itraconazole with GLI and SMO. In this study, we have taken Cyclopamine and itraconazole as a reference to compare the interaction of CAR with GLI and SMO. It was observed that CAR showed high binding affinity towards SMO with a binding energy value of -6.9 kcal/mol.

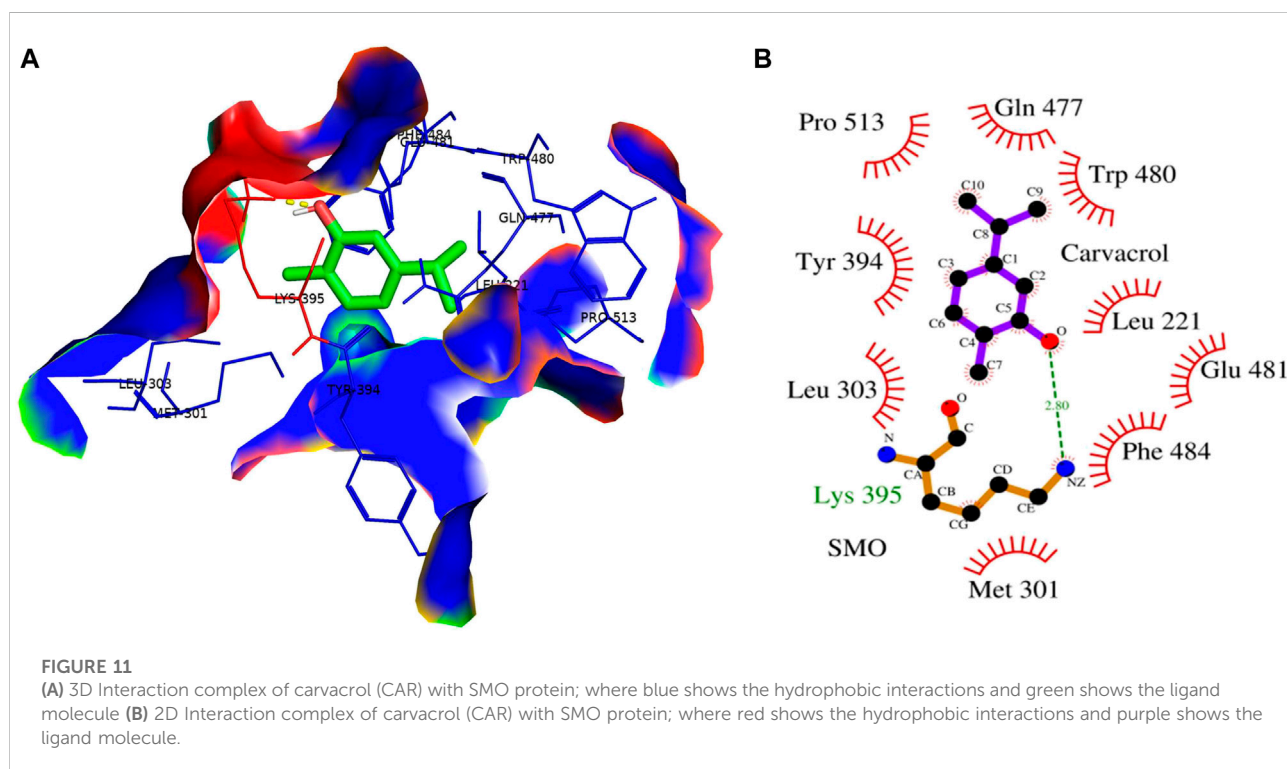
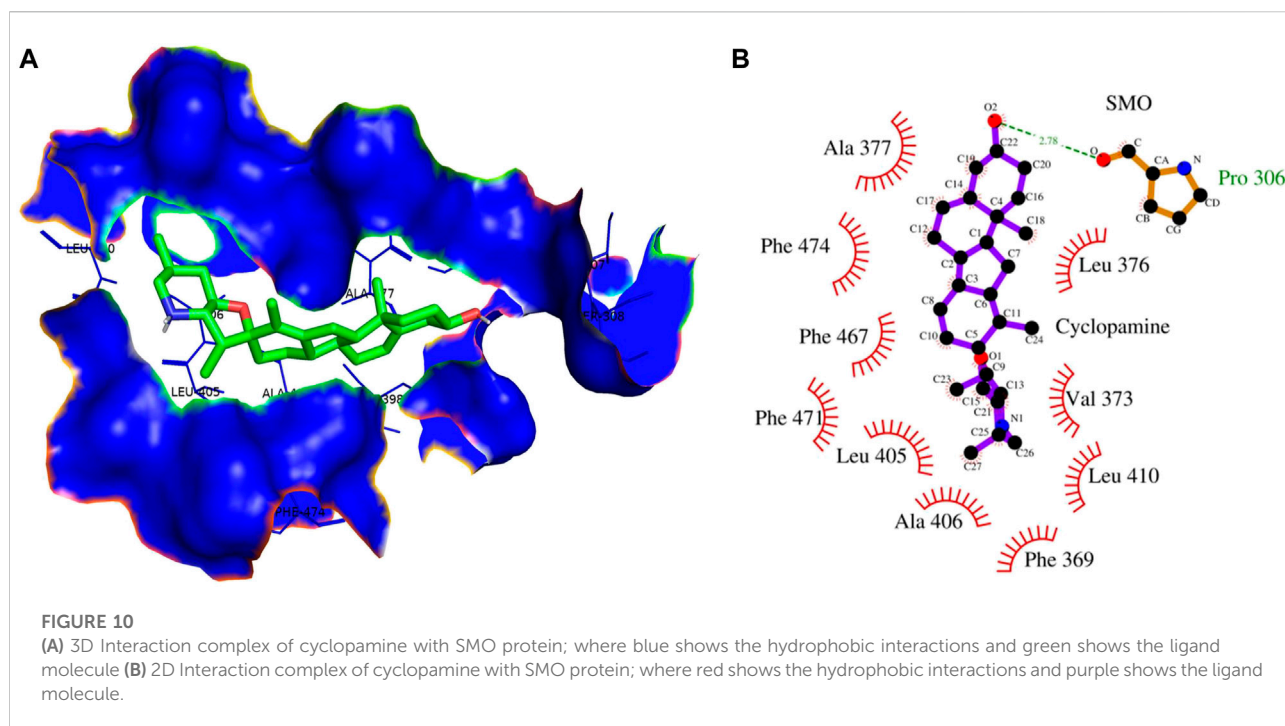
In contrast, Cyclopamine exhibited a binding energy value of -9.3 kcal/mol (Figures 10A,B and Figures 11A,B). Similarly, CAR also showed high binding affinity towards GLI protein with a binding energy of -6.4 kcal/mol. In contrast, itraconazole showed binding energy of -6.4 kcal/mol (Figures 12A,B and Figures 13A,B). However, it is critical to note that the CAR exhibited higher binding energies with SMO and GLI protein, as compared to their standard inhibitors Cyclopamine and itraconazole. Moreover, the amino acid residues of SMO and

GLI proteins interacting with Cyclopamine and itraconazole are similar to CAR (Table 2). Furthermore, *in silico* findings concluded that CAR interacted with the GLI-DNA complex to deregulate the HH signaling pathway. However, the exact mechanism has yet to be discovered. Therefore, we tried to corroborate our qPCR results with *in silico* results to establish that CAR downregulated HH signaling pathway and exhibited strong binding affinity towards HH components (SMO and GLI). Furthermore, these results support real-time PCR results and provide a strong rationale for why CAR strongly inhibited the HH signaling in cervical carcinoma cells. The plausible mechanism of action of CAR against cervical cancer is summarized in Figure 14.

4 Discussion

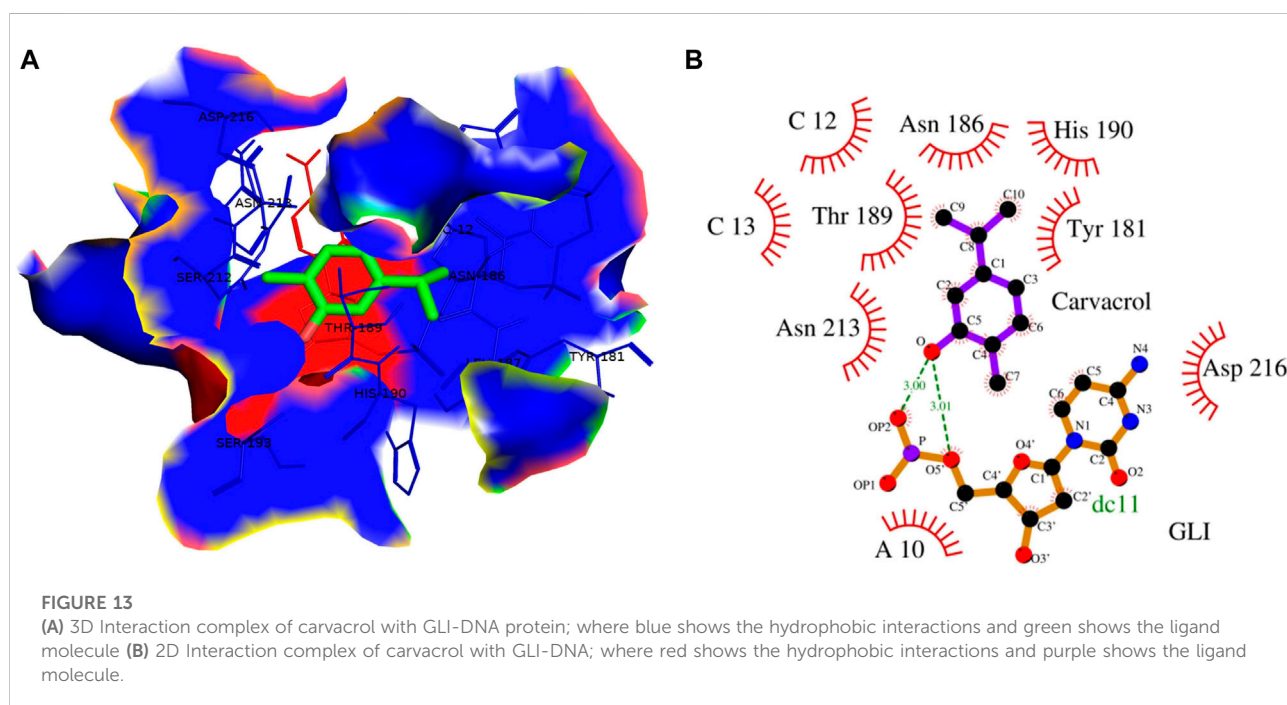
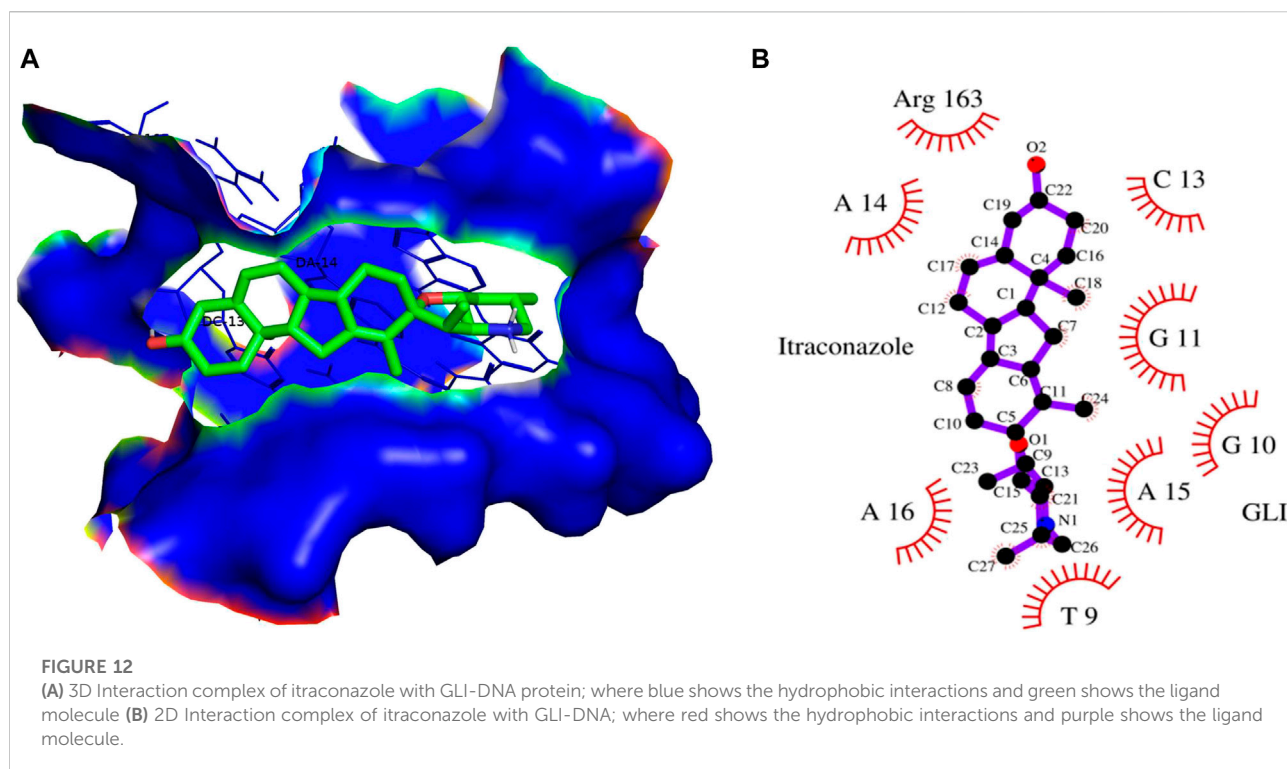
Cervical cancer is the most common gynecological malignancy, which is a reason behind cancer-associated fatalities among women worldwide. Even though numerous screening programs and anticancer therapeutic strategies have improved the prognosis of cervical cancer patients, they still suffer from metastasis and recurrence with a low survival rate. Thus, there is an urge to develop novel drugs or therapeutic interventions to improve patient prognosis.

Earlier reports have outlined the efficacy of CAR in impeding the growth, proliferation and instigation of apoptosis in several cancer cells. In view of these inferences, the investigators of this study focused on elucidating the efficacy of CAR in impeding growth and instigating apoptosis in HPV-negative C33A cells. Rapid proliferation and apoptosis resistance are essential



biological attributes of malignant tumors, and most cancer therapies aim to mitigate proliferation and provoke apoptosis on transformed cells. The findings from our practical

investigation suggested that CAR was competent in suppressing the proliferation rate of cervical cancer cells. It was observed that CAR exposure significantly reduced the

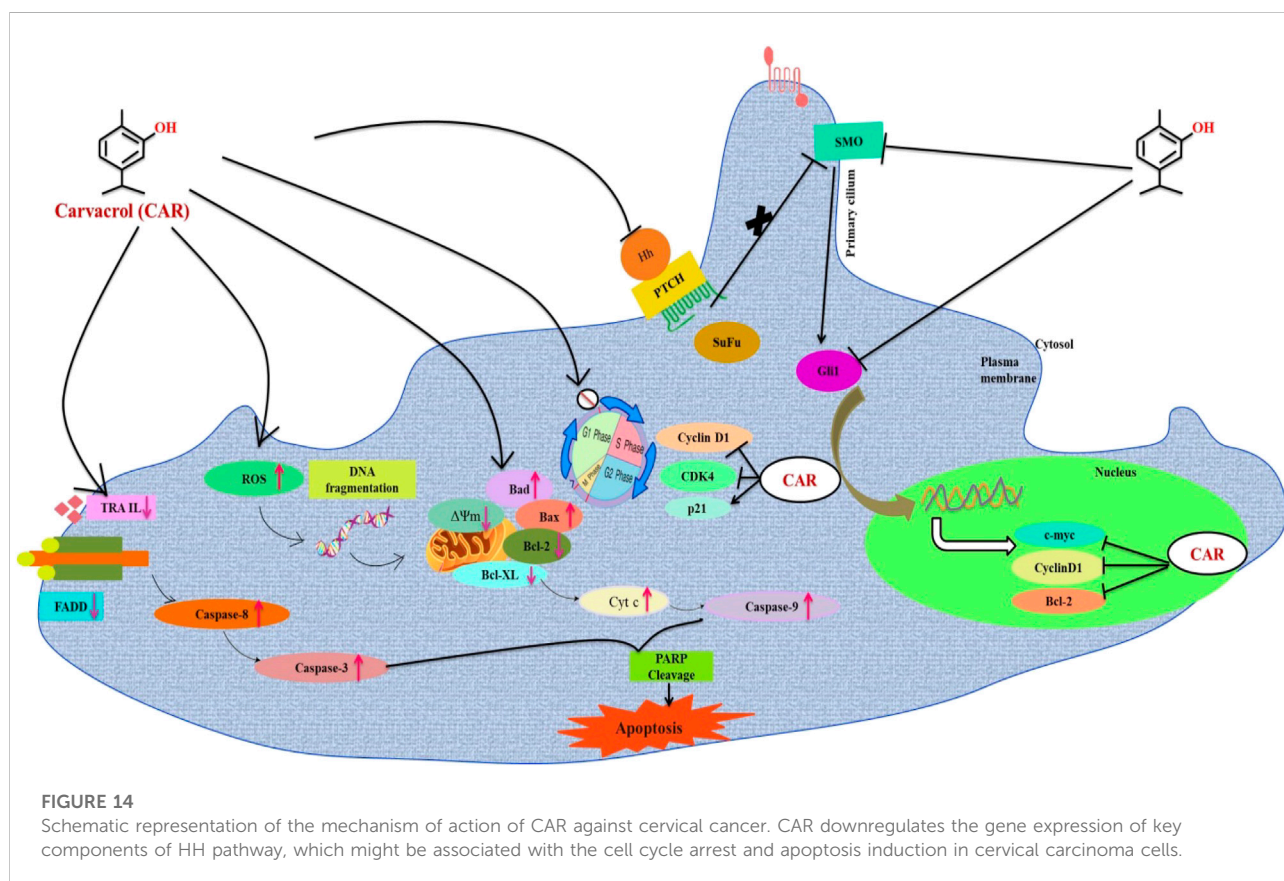


viability of C33A cells as witnessed by MTT and LDH assays, which were well supported by clonogenic assay. Treatment with CAR induces various morphological aberrations within the

cervical cancer cells, such as cell shrinkage, blebbing in the cell membrane and detachment of cells from the surface, which collectively indicates apoptosis induction. To our

TABLE 2 Binding energy of carvacrol, cyclopamine and itraconazole with Gli1 and Smo1.

Interacting molecules	Binding energy (Kcal/mole)	Interacting residues
SMO-Cyclopamine	−9.3	Leu ³⁷⁶ , Val ³⁷³ , Leu ⁴¹⁰ , Phe ³⁶⁹ , Ala ⁴⁰⁶ , Leu ⁴⁰⁵ , Gln ⁴⁷⁷ , Gln ⁴⁷⁷ , Phe ⁴⁷¹ , Phe ⁴⁶⁷ , Phe ⁴⁷⁴ , Ala ³⁷⁷ , Pro ³⁰⁶
SMO-Carvacrol	−6.9	Gln ⁴⁷⁷ , Trp ⁴⁸⁰ , Leu ²²¹ , Glu ⁴⁸¹ , Phe ⁴⁸⁴ , Met ³⁰¹ , Leu ³⁰³ , Tyr ³⁹⁴ , Pro ⁵¹³ , Lys ³⁹⁵
GLI-Itraconazole	−9.5	Adenine ^{14,15,16} , guanine ^{10,11} , thymine ⁹ , cytosine ¹³ , Arg ¹⁶³
GLI-Carvacrol	−6.4	Cytosine ^{12,13} , adenine ¹⁰ are the nucleotide, and Tyr ¹⁸¹ , Asn ¹⁸⁶ , Thr ¹⁸⁹ , His ¹⁹⁰ , Asn ²¹³ , Asp ²¹⁶



understanding, no study has demonstrated the inhibition of the HH signaling pathway mediating apoptosis induction and cell cycle arrest in cervical cancer cells.

Subsequently, the mechanistic insight of CAR-mediated growth inhibition of cancerous cells was further investigated. From a biochemical perspective, apoptosis is a complex physiological process which is characterized by depolarization as a result of excess ROS generation (Riedl and Salvesen, 2007). Previously, we have reported that CAR could be a plausible therapeutic intervention against HPV-positive cervical cancer cells by targeting ROS-mediated apoptosis and cell cycle arrest (Khosravi-Far and Esposti, 2004). To decipher the functioning of CAR at the molecular level behind its impeding effect on C33A cells, fluorescent microscopy, FACS and qRT-PCR based investigations to ascertain whether the treatment of

CAR altered the levels of mitochondrial ROS levels. The results of DAPI/PI and AO/EtBr double staining demonstrated peculiar attributes of apoptosis, including condensation and fragmentation of the nuclei. Further, the increased number of Annexin V-stained cells and cleaved PARP levels in cervical cancer cells indicate that CAR acts as an apoptosis-inducing agent.

Intrinsic apoptosis is instigated by the depolarization of mitochondria, which releases apoptogenic factors such as cytochrome c into the cytosol, eventually mediating the activation of caspase-3 (Jimeno et al., 2008). We found that CAR treatment instigates mitochondria-centered intrinsic apoptosis pathway by activating caspase-3 and -9, dissipation of $\Delta\Psi_m$ with concomitant release of cytochrome c in cervical cancer cells. Additionally, CAR treatment upregulated the

mRNA expression of pro-apoptotic proteins and downregulated the expression of anti-apoptotic proteins in cervical cancer cells.

The extrinsic apoptosis is activated through transmembrane death receptors including the targets for FasL and TRAIL. DR4 and DR5 activation primarily culminates in DISC (death-inducing signaling complex) formation by interacting with FADD *via* the death domain. Activation of FADD subsequently results in the recruitment of caspase-8 facilitated by interaction with death effector domains (DEDs), resulting in the activation of caspase-8 (Traverso et al., 2013). In this report, CAR mediates the activation of caspase-8 and increases the mRNA expression level of extrinsic signaling molecules such as FasL, FADD and TRAIL in C33A cervical cancer cells.

Glutathione (GSH) is pivotal for regulating several processes, such as cell differentiation, cell proliferation, and apoptosis modulation. Intriguingly, alterations within the cellular levels of GSH often lead to the onset and progression of various ailments, including cancer. Elevated levels of GSH are frequently observed in tumors, which subsequently aids in developing chemotherapeutic resistance in neoplastic cells (Circu and Aw, 2012). In the present report, we found a significant depletion of endogenous GSH in C33A cells after the treatment with CAR. Moreover, previous reports have established that GSH depletion is an imperative marker for apoptosis induction in response to various apoptotic stimuli (Thayyullathil et al., 2011). Thus, our data demonstrated that CAR mediates apoptosis in cervical cancer cells, which supports the principle that GSH depletion may favor apoptotic cell death.

Multiple studies have established that targeting drugs/agents responsible for exerting cell cycle arrest could be a plausible therapeutic for treating and managing different cancers, including cervical cancer (Vishnoi et al., 2016). Abrogation of the cell cycle at the G1 phase prevents DNA repair and inhibits entry into S phase. Thus, the G1 checkpoint appeared as a propitious therapeutic target for cancer treatment. The results from this study exhibited that CAR induced a powerful growth-suppressing effect on cervical cancer C33A cells by restricting their progression to G0/G1 phase. Moreover, it is also known that both cyclins and CDKs are pre-requisite for regulating the progression of the cell cycle and their deactivation results in cell cycle arrest. We found inhibitory effects of CAR on cyclin D1 and CDK4 in cervical cancer cells suggesting its interference with cell cycle regulatory proteins. CDK inhibitors, namely p21/WAF1 and p27/KIP1 families of proteins, regulate the CDK activity. Our findings indicate that CAR enhanced the p21 and p27 expression levels in HPV-negative cervical cancer C33A cells. Taken together, the present observations substantiated that CAR was competent in impeding the progression of the cell cycle at G0/G1 phase *via* altering the expression of genes involved in cell cycle regulation in cervical cancer cells. These results further suggested that CAR treatment mediates the inhibition of C33A cells by obstructing the replication of DNA followed by its repair and thereby instigating cell cycle arrest and apoptosis.

The HH/GLI signaling represents an evolutionarily conserved signaling cascade involved in regulating normal development and determination of cell fate. It has been observed that HH signaling contributes to increased chemoresistance, stemness and metastasis. Various components of HH signaling are present in the advanced stages of cervical cancer, indicating that constitutive HH signaling is associated with the development of cervical cancer along with chemoresistance and recurrence (Li et al., 2019). We further studied the CAR-mediated inhibition of HH signaling in C33A cells. The mRNA expression of key proteins such as PTCH1, SMO, and GLI1 was downregulated significantly, indicating that CAR may prevent cervical cancer by modulating hedgehog signaling powerfully.

These findings further intrigued us to investigate the mechanism of CAR-mediated cytotoxic effects in comparison to cyclopamine and itraconazole. The interactions between CAR, cyclopamine and itraconazole with SMO and GLI proteins were studied by *in silico* techniques. It was found that the binding energies (BE) of cyclopamine and SMO protein was -9.3 kcal/mol, which is nearby and comparable to BE of CAR and SMO (-6.9 kcal/mol). Similarly, the BE of itraconazole and GLI protein was -9.5 kcal/mol, whereas CAR and GLI were -6.3 kcal/mol and were also observed near each other. Furthermore, our *in silico* studies have substantiated that the amino acid residues involved in the interaction of CAR with SMO and GLI were considerably similar as compared to the reference drugs (cyclopamine and itraconazole). Thus, our *in silico* findings corroborated RT-PCR results and concluded that CAR could inhibit the HH signaling cascade in cervical cancer cells. After an exhaustive literature review, the authors of the present manuscript are confident that this is the report which demonstrated the CAR-mediated apoptosis *via* inhibition of HH signaling cascade in the C33A cell line. Nevertheless, a more exhaustive study using appropriate *in vivo* models is further warranted to establish the pre-clinical efficacy of CAR against cervical cancer.

5 Conclusion

Collectively our findings conclude that CAR principally exerted antiproliferative and apoptotic effects in C33A (HPV-) cervical cancer C33A cells *in vitro* by inducing the production of ROS inside mitochondria of C33A cells, which provoked multiple cellular events leading to mitochondrion-centered intrinsic apoptosis. Furthermore, CAR also modulates the key signaling components of extrinsic or death receptor pathways. Thus, CAR could suppress cell proliferation and induce apoptosis in cervical carcinoma cells by inhibiting HH signaling pathway. The investigators believe that this is a novel report elucidating the best of our understanding. This is the first report demonstrating the antiproliferative and apoptotic potential of CAR against human cervical carcinoma C33A cells *via* targeting the HH signaling pathway. Nevertheless, our present report is still restricted to the

in vitro evidence; yet due to the significant anticancer potential of CAR, we propose it as a suitable drug for cervical cancer therapy after sufficient pre-clinical and clinical trials.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the corresponding authors after due discussion.

Author contributions

AA and RT contributed in analysis, surveying literature, writing, visualization, draft preparation. MS and HAA contributed in data validation. IH, EHC, and DY contributed in conceptualization, data curation, validation and funding. IA contributed in conceptualization, validation, data curation and editing of the manuscript.

Funding

The authors are thankful to the Basic Science Research Program of the National Re-search Foundation of Korea

References

- Ahmad, A., and Ansari, I. A. (2021). Carvacrol exhibits chemopreventive potential against cervical cancer cells via caspase-dependent apoptosis and abrogation of cell cycle progression. *Anticancer. Agents Med. Chem.* 21 (16), 2224–2235. doi:10.2174/1871520621999201230201258
- Ahmad, K., Khan, S., Adil, M., Saeed, M., and Srivastava, A. K. (2014). Structure based molecular inhibition of Caspase-8 for treatment of multi-neurodegenerative diseases using known natural compounds. *Bioinformation* 10 (4), 191–195. doi:10.6026/97320630010191
- Ahmad, A., Tiwari, R. K., Saeed, M., Ahmad, I., and Ansari, I. A. (2022). Glycyrrhizin mediates downregulation of Notch pathway resulting in initiation of apoptosis and disruption in the cell cycle progression in cervical cancer cells. *Nutr. cancer* 74 (2), 622–639. doi:10.1080/01635581.2021.1895234
- Ansari, I. A., Ahmad, A., Imran, M. A., Saeed, M., and Ahmad, I. (2021). Organosulphur compounds induce apoptosis and cell cycle arrest in cervical cancer cells via downregulation of HPV E6 and E7 oncogenes. *Anticancer. Agents Med. Chem.* 21 (3), 393–405. doi:10.2174/1871520620999200818154456
- Arbyn, M., Antoine, J., Mägi, M., Smailyte, G., Stengrevics, A., Suteu, O., et al. (2011). Trends in cervical cancer incidence and mortality in the Baltic countries, Bulgaria and Romania. *Int. J. Cancer* 128 (8), 1899–1907. doi:10.1002/ijc.25525
- Arbyn, M., Weiderpass, E., Bruni, L., de Sanjosé, S., Saraiya, M., Ferlay, J., et al. (2020). Estimates of incidence and mortality of cervical cancer in 2018: a worldwide analysis. *Lancet Glob. Health* 8 (2), e191–e203. doi:10.1016/S2214-109X(19)30482-6
- Baig, M. H., Ahmad, K., Roy, S., Ashraf, J. M., Adil, M., Siddiqui, M. H., et al. (2016). Computer aided drug design: Success and limitations. *Curr. Pharm. Des.* 22 (5), 572–581. doi:10.2174/1381612822666151125000550
- Baser, K. H. (2008). Biological and pharmacological activities of carvacrol and carvacrol bearing essential oils. *Curr. Pharm. Des.* 14 (29), 3106–3119. doi:10.2174/138161208786404227
- (NRF) funded by the Ministry of Education (2021R1A6A1A03038785 and 2020R1I1A1A01073071).
- ## Acknowledgments
- AA is highly grateful to the Department of Science Technology (DST) for providing DST Inspire fellowship (IF170529).
- ## 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.
- Birben, E., Sahiner, U. M., Sackesen, C., Erzurum, S., and Kalayci, O. (2012). Oxidative stress and antioxidant defense. *World Allergy Organ. J.* 5 (1), 9–19. doi:10.1097/WOX.0b013e3182439613
- Bosch, F. X., Lorincz, A., Muñoz, N., Meijer, C. J., and Shah, K. V. (2002). The causal relation between human papillomavirus and cervical cancer. *J. Clin. pathology* 55 (4), 244–265. doi:10.1136/jcp.55.4.244
- Chen, J. K., Taipale, J., Cooper, M. K., and Beachy, P. A. (2002). Inhibition of Hedgehog signaling by direct binding of cyclopamine to Smoothened. *Genes Dev.* 16 (21), 2743–2748. doi:10.1101/gad.1025302
- Circu, M. L., and Aw, T. Y. (2012). Glutathione and modulation of cell apoptosis. *Biochim. Biophys. Acta - Mol. Cell Res.* 1823 (10), 1767–1777. doi:10.1016/j.bbamcr.2012.06.019
- Dai, W., Sun, C., Huang, S., and Zhou, Q. (2016). Carvacrol suppresses proliferation and invasion in human oral squamous cell carcinoma. *Oncotargets. Ther.* 9, 2297–2304. doi:10.2147/OTT.S98875
- Datta, N. R., Stutz, E., Gomez, S., and Bodis, S. (2019). Efficacy and safety evaluation of the various therapeutic options in locally advanced cervix cancer: A systematic review and network meta-analysis of randomized clinical trials. *Int. J. Radiat. Oncol. Biol. Phys.* 103 (2), 411–437. doi:10.1016/j.ijrobp.2018.09.037
- Deng, H., Huang, L., Liao, Z., Liu, M., Li, Q., and Xu, R. (2020). Itraconazole inhibits the Hedgehog signaling pathway thereby inducing autophagy-mediated apoptosis of colon cancer cells. *Cell Death Dis.* 11 (7), 539. doi:10.1038/s41419-020-02742-0
- Hassan, M., Ahmad, K., Adi, M., Zainul, A., Imran, M., Lohan, M., et al. (2014). Drug discovery and *in silico* techniques: a mini-review. *Enz. Eng.* 04. doi:10.4172/2329-6674.1000123
- Jimeno, A., Rudek, M. A., Kulesza, P., Ma, W. W., Wheelhouse, J., Howard, A., et al. (2008). Pharmacodynamic-guided modified continuous reassessment method-based, dose-finding study of rapamycin in adult patients with solid tumors. *J. Clin. Oncol. Official J. Am. Soc. Clin. Oncol.* 26 (25), 4172–4179. doi:10.1200/JCO.2008.16.2347

- Khan, F. I., Wei, D. Q., Gu, K. R., Hassan, M. I., and Tabrez, S. (2016). Current updates on computer aided protein modeling and designing. *Int. J. Biol. Macromol.* 85, 48–62. doi:10.1016/j.ijbiomac.2015.12.072
- Khan, F. I., Kang, T., Ali, H., and Lai, D. (2021). Remdesivir strongly binds to RNA-dependent RNA polymerase, membrane protein, and main protease of SARS-CoV-2: Indication from molecular modeling and simulations. *Front. Pharmacol.* 12, 710778. doi:10.3389/fphar.2021.710778
- Khan, F. I., Hassan, F., and Lai, D. (2022). *In silico* studies on psilocybin drug derivatives against SARS-CoV-2 and cytokine storm of human interleukin-6 receptor. *Front. Immunol.* 12, 794780. doi:10.3389/fimmu.2021.794780
- Khan, F. I., Rehman, M. T., Sameena, F., Hussain, T., AlAjmi, M. F., Lai, D., et al. (2022). Investigating the binding mechanism of topiramate with bovine serum albumin using spectroscopic and computational methods. *J. Mol. Recognit.* 35 (7), e2958. doi:10.1002/jmr.2958
- Khosravi-Far, R., and Esposti, M. D. (2004). Death receptor signals to the mitochondria. *Cancer Biol. Ther.* 3 (11), 1051–1057. doi:10.4161/cbt.3.11.1173
- Lai, P. K., and Roy, J. (2004). Antimicrobial and chemopreventive properties of herbs and spices. *Curr. Med. Chem.* 11 (11), 1451–1460. doi:10.2174/0929867043365107
- Li, K., Fang, D., Xiong, Z., and Luo, R. (2019). Inhibition of the hedgehog pathway for the treatment of cancer using Itraconazole. *Onco. Targets. Ther.* 12, 6875–6886. doi:10.2147/OTT.S223119
- Lin, C. J., Kuan, C. H., Wang, L. W., Wu, H. C., Chen, Y., Chang, C. W., et al. (2016). Integrated self-assembling drug delivery system possessing dual responsive and active targeting for orthotopic ovarian cancer theranostics. *Biomaterials* 90, 12–26. doi:10.1016/j.biomaterials.2016.03.005
- Liu, C., and Wang, R. (2019). The roles of hedgehog signaling pathway in radioresistance of cervical cancer. *Dose. Response.* 17 (4), 155932581988529. doi:10.1177/1559325819885293
- Moss, C., and Kaye, S. B. (2002). Ovarian cancer: Progress and continuing controversies in management. *Eur. J. cancer.* 38, 1701–1707. doi:10.1016/s0959-8049(02)00161-2
- Riedl, S. J., and Salvesen, G. S. (2007). The apoptosome: Signalling platform of cell death. *Nat. Rev. Mol. Cell Biol.* 8 (5), 405–413. doi:10.1038/nrm2153
- Samarzija, I., and Beard, P. (2012). Hedgehog pathway regulators influence cervical cancer cell proliferation, survival and migration. *Biochem. biophysical Res. Commun.* 425 (1), 64–69. doi:10.1016/j.bbrc.2012.07.051
- Shamas-Din, A., Kale, J., Leber, B., and Andrews, D. W. (2013). Mechanisms of action of Bcl-2 family proteins. *Cold Spring Harb. Perspect. Biol.* 5 (4), a008714. doi:10.1101/cshperspect.a008714
- Simms, K. T., Steinberg, J., Caruana, M., Smith, M. A., Lew, J. B., Soerjomataram, I., et al. (2019). Impact of scaled up human papillomavirus vaccination and cervical screening and the potential for global elimination of cervical cancer in 181 countries, 2020–99: a modelling study. *Lancet Oncol.* 20 (3), 394–407. doi:10.1016/S1470-2045(18)30836-2
- Sökmen, M., Serkedjieva, J., Daferera, D., Gulluce, M., Polissiou, M., Tepe, B., et al. (2004). *In vitro* antioxidant, antimicrobial, and antiviral activities of the essential oil and various extracts from herbal parts and callus cultures of *Origanum acutidens*. *J. Agric. Food Chem.* 52 (11), 3309–3312. doi:10.1021/jf049859g
- Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., et al. (2021). Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *Ca. A Cancer J. Clin.* 71 (3), 209–249. doi:10.3322/caac.21660
- Thayyullathil, F., Chathoth, S., Kizhakkayil, J., Galadari, A., Hago, A., Patel, M., et al. (2011). Glutathione selectively inhibits Doxorubicin induced phosphorylation of p53Ser¹⁵, caspase dependent ceramide production and apoptosis in human leukemic cells. *Biochem. Biophys. Res. Commun.* 411 (1), 1–6. doi:10.1016/j.bbrc.2011.05.156
- Traverso, N., Ricciarelli, R., Nitti, M., Marengo, B., Furfaro, A. L., Pronzato, M. A., et al. (2013). Role of glutathione in cancer progression and chemoresistance. *Oxid. Med. Cell. Longev.* 2013, 1–10. doi:10.1155/2013/972913
- Vaccarella, S., Laversanne, M., Ferlay, J., and Bray, F. (2017). Cervical cancer in africa, Latin America and the caribbean and asia: Regional inequalities and changing trends. *Int. J. Cancer* 141 (10), 1997–2001. doi:10.1002/ijc.30901
- Vishnoi, K., Mahata, S., Tyagi, A., Pandey, A., Verma, G., Jadli, M., et al. (2016). Cross-talk between human papillomavirus oncoproteins and hedgehog signaling synergistically promotes stemness in cervical cancer cells. *Sci. Rep.* 6, 34377. doi:10.1038/srep34377
- Wang, Z., Li, Y., Ahmad, A., Azmi, A. S., Banerjee, S., Kong, D., et al. (2010). Targeting Notch signaling pathway to overcome drug resistance for cancer therapy. *Biochimica Biophysica Acta - Rev. Cancer* 1806 (2), 258–267. doi:10.1016/j.bbcan.2010.06.001
- Wang, L., Liu, Y., Zhou, Y., Wang, J., Tu, L., Sun, Z., et al. (2019). Zoledronic acid inhibits the growth of cancer stem cell derived from cervical cancer cell by attenuating their stemness phenotype and inducing apoptosis and cell cycle arrest through the Erk1/2 and Akt pathways. *J. Exp. Clin. Cancer Res.* 38 (1), 93. doi:10.1186/s13046-019-1109-z
- Wu, Z., Zou, B., Zhang, X., and Peng, X. (2020). Eupatilin regulates proliferation and cell cycle of cervical cancer by regulating hedgehog signalling pathway. *Cell Biochem. Funct.* 38 (4), 428–435. doi:10.1002/cbf.3493
- Zorov, D. B., Juhaszova, M., and Sollott, S. J. (2014). Mitochondrial reactive oxygen species (ROS) and ROS-induced ROS release. *Physiol. Rev.* 94 (3), 909–950. doi:10.1152/physrev.00026.2013
- Zotti, M., Colaïanna, M., Morgese, M. G., Tucci, P., Schiavone, S., Avato, P., et al. (2013). Carvacrol: From ancient flavoring to neuromodulatory agent. *Mol. (Basel, Switz.)* 18 (6), 6161–6172. doi:10.3390/molecules18066161



OPEN ACCESS

EDITED BY

Sibghatulla Shaikh,
Yeungnam University, Republic of Korea

REVIEWED BY

Shahnawaz Ali,
King's College London, United Kingdom
Girish Kumar Gupta,
Sri Sai Group of Institutes Badhani, India
Sadaf Hasan,
Grossman School of Medicine, New York
University, United States

*CORRESPONDENCE

Khalaf M. Alenezi,
✉ k.alenezi@uoh.edu.sa
Ashanul Haque,
✉ a.haque@uoh.edu.sa
Ming-Fa Hsieh,
✉ mfhsieh@cycu.edu.tw

SPECIALTY SECTION

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

RECEIVED 05 January 2023

ACCEPTED 22 February 2023

PUBLISHED 03 March 2023

CITATION

Alsukaibi AKD, Alenezi KM, Haque A,
Ahmad I, Saeed M, Verma M, Ansari IA and
Hsieh M-F (2023), Chemical, biological
and *in silico* assessment of date (*P.
dactylifera* L.) fruits grown in Ha'il region.
Front. Chem. 11:1138057.
doi: 10.3389/fchem.2023.1138057

COPYRIGHT

© 2023 Alsukaibi, Alenezi, Haque, Ahmad,
Saeed, Verma, Ansari and Hsieh. 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.

Chemical, biological and *in silico* assessment of date (*P. dactylifera* L.) fruits grown in Ha'il region

Abdulmohsen Khalaf Dhahi Alsukaibi¹, Khalaf M. Alenezi^{1*},
Ashanul Haque^{1*}, Irfan Ahmad², Mohd Saeed³, Mahima Verma⁴,
Irfan Ahmad Ansari⁴ and Ming-Fa Hsieh^{5*}

¹Department of Chemistry, College of Science, University of Ha'il, Hail, Saudi Arabia, ²Department of Clinical Laboratory Science, College of Applied Medical Sciences, King Khalid University, Abha, Saudi Arabia, ³Department of Biology, College of Science, University of Ha'il, Hail, Saudi Arabia, ⁴Department of Biosciences, Integral University, Lucknow, India, ⁵Department of Biomedical Engineering, Chung Yuan Christian University, Taoyuan City, Taiwan

Background: Dates palm (*Phoenix dactylifera* L.) fruits are among the most widely used fruits in the Middle East and African nations. Numerous researchers confirmed the presence of phytochemicals in *P. dactylifera* L. fruit and its by-products with broad-ranging biological activities.

Objectives: In the present work, phytochemical and biological assessments of two different cultivars of date fruit (*Shishi M1* and *Majdool M2* grown in the Ha'il region of Saudi Arabia) have been carried out.

Methods: Date fruits were extracted and analyzed by gas chromatography-mass spectrometry (GC-MS), liquid chromatography-mass spectrometry (LC-MS) and Fourier-transform infrared spectroscopy (FT-IR) techniques. The lyophilized methanolic extracts were analyzed for their *in-vitro* antiproliferative and cytotoxicity against colon cancer (HCT116) cell line. To identify the possible constituents responsible for the bioactivity, *in-silico* molecular docking and molecular dynamics (MD) simulation studies were carried out.

Results: Both cultivars exhibited *in-vitro* anticancer activity (IC₅₀ = 591.3 µg/mL and 449.9 µg/mL for **M1** and **M2**, respectively) against colon cancer HCT-116 cells. The computational analysis results indicated procyanidin B2 and luteolin-7-O-rutinoside as the active constituents.

Conclusion: Based on these results, we conclude that these cultivars could be a valuable source for developing health promoter phytochemicals, leading to the development of the Ha'il region, Saudi Arabia.

KEYWORDS

date fruits, extraction, GC-MS, Ha'il region, molecular docking, molecular dynamics, *P. dactylifera* L.

Introduction

Natural products (NPs) play an instrumental role in drug design and remain an inspiration for discovering new drug candidates (Newman and Cragg, 2016). Being the largest source of new pharmacophores, 60%–70% of drugs used today are based directly or indirectly on NPs (Atanasov et al., 2021). Indeed, enormous diversity in their chemical structure, broad-ranging bioactivity, low toxicity, and ability to bind with different proteins (targets) gives

natural compounds an edge over synthetic ones (Ali et al., 2010). In the quest for new drug candidates, research is being carried out on the extraction, isolation, and identification of bioactive compounds found in plants, animals and microbes (Altemimi et al., 2017). Among a large pool of NPs, the dates palm (*Phoenix dactylifera* L.), a member of the Asteraceae family, has garnered an immense interest (Al-Alawi et al., 2017; Maqsood et al., 2020). Date fruits of the date palm tree (*P. dactylifera* L.) is one of the most consumed fruits worldwide, especially in the Middle East and Asian countries (Complexity, 2022). Apart from their high nutritional and commercial values (Mia et al., 2020), date fruits and their by-products have also attracted researchers due to their potential health benefits (Maqsood et al., 2020). Their antibacterial, antifungal, antiviral, antidiabetic, anticancer, anti-inflammatory, antioxidant, antiangiogenic and other protective effects along with negligible side effects, are particularly interesting (Vayalil, 2002; Maqsood et al., 2020). It has been demonstrated that date fruits are rich in carbohydrates, protein, fibres, minerals, vitamins, phenolic acids, flavonoids, and other phytochemicals responsible for bioactivities. The chemical composition depends on various factors, including the type of cultivar, geographical location, irrigation method, ripening stage, processing time, extracting solvents, etc. (Borochov-Neori et al., 2013). Based on this knowledge, various groups investigated the chemical composition of date fruits and seeds native to different regions (Vayalil, 2012; Al-Alawi et al., 2017; Mia et al., 2020; Echegaray et al., 2021; Ibrahim et al., 2021). The group led by Aviram (Borochov-Neori et al., 2013) has conducted studies on the chemical and biological analysis of several varieties of date fruits (Maqsood et al., 2020). They, along with others, confirmed that the phytoconstituents and bioactivity of the date fruits are the function of the parameters mentioned above. For example, a pilot study showed that *Medjool* or *Hallawi* varieties of date fruits vary in phenolics, catechins and quercetin derivative content and antioxidant effect (Rock et al., 2009). The same group reported anti-atherogenic properties of acetone extracts of *Hallawi* in addition to the eight other variants (Borochov-Neori et al., 2013). As per the group, phenolic compounds exerted anti-atherogenic properties via low-density lipoprotein (LDL) oxidation and serum-mediated cholesterol efflux. On the other hand, alcoholic extract of the Tunisian variety was found to inhibit α -glucosidase and α -amylase enzymes with low IC_{50} values (El Abed et al., 2017). Zhang et al. (2013) performed an extensive chemical and biochemical profiling of *Ajwa* date fruits. They identified several new compounds such as bis (2-ethylhexyl) terephthalate and bis (2-ethylheptyl) phthalate in addition to glycoside, terpenoids, triglyceride, phthalates, etc. They noted that the aqueous and organic extract exerts dose-dependent antioxidant and anti-inflammatory effects.

In Saudi Arabia, more than 400 varieties of date fruit are cultivated, which vary in appearance, nutrition, and nutritional value (Zhang et al., 2017). Several researchers reported that the varieties such as *Barni*, *Khalas*, and *Ajwa* show unique biological activities (Eid et al., 2013; Assirey, 2015; Hamad et al., 2015). Recently, Amir and co-workers (Alghamdi et al., 2018) studied the nutritional value of several varieties of date fruits found in Ha'il province; however, the biological activity of the date fruits remains unclear. Prompted by this, we carried out extraction, characterization, *in-vitro* antiproliferative and cytotoxicity assay of two varieties of date fruits (*Shishi M1* and *Majdool M2*) grown in the Ha'il region of Saudi Arabia). Then, to further identify the possible potential constituent present, ligand-based virtual screening

was performed. This is followed by molecular dynamics simulation studies to identify the stability of promising compounds with possible receptor.

Materials and methods

General

All solvents used for isolation and purification were of ACS reagent grade (Sigma-Aldrich Chemical Co., St. Louis, MO, United States). Lyophilization was carried out on BenchTop Manifold Freeze Dryer (MILLROCK, United States) for 24 h at a condenser temperature of -45°C equipped with Edward pump. Attenuated-total-reflectance IR spectra were recorded on pure samples on diamond using a Shimadzu IRSpirit-T spectrometer.

Sample collection, extraction, and sample preparation

Two different varieties of date fruits (*Tamr* stage, Supplementary Table S1) grown in Ha'il province were collected from the local market. Authors (AKDA and KMA) and local farmers authenticated the samples, and a voucher specimen was deposited. The samples were stored and kept in a -20°C freezer. First, three pieces of date fruits from each variant (Supplementary Table S1) were pitted to remove seeds and cut into small pieces. Then, cold extraction was performed by shaking and mixing fruit materials in methanol (MeOH) overnight at room temperature, followed by filtration. The residue was further extracted twice with MeOH for 1 h. Finally, the extracts were combined and concentrated using rotatory evaporation at room temperature. The resulting viscous honey-like liquid was lyophilized to afford light yellow water-soluble powder and stored at -20°C till further analysis.

Chromatographic studies

Date fruit extracts were analyzed by LC-MS system using a reverse phase C_{18} column (Accucore, 150×4.6 , $2.6 \mu\text{m}$). The LC-MS system comprised a Waters Alliance 2695 HPLC pump, an autosampler, a vacuum degasser, and a column compartment attached to a XEVO-TQD detector with electrospray ionization (ESI). The following gradient of solvents were used: acetonitrile (mobile phase A) and 5 mM acetic acid (mobile phase B); ratio of A to B, 0–1 min, 5:95; 1–10 min, 5:95 to 30:70; 10–16 min, 30:70 to 60:40; 16–24 min, 60:40 to 80:20; 24–32 min, 80:20; 32–40 min, 5:95. In all cases, the columns were reequilibrated between injections with the equivalent of 5 mL of the mobile phase. During the full scan by MS/MS, mass acquisition was set from 150 to 2000 Da. This method utilized ESI-LC/MS/MS operating in MRM mode. The ESI settings were the following capillary voltage, 3.5 kV; cone voltage, 40 V; the flow of desolvation gas (Argon gas), 650 L/h; flow of cone gas, 30 L/h. Gas Chromatography-Mass Spectrometry (GC-MS) analysis was carried out using Agilent 8890/5977B Series (Agilent 5977B EI/CI MSD) spectrometer. The segments were recognized by examination of their delay times and mass spectra with those of the NIST 11 mass spectral database.

Cell culture and maintenance

The human colon cancer cell HCT116 was acquired from American Type Culture Collection (ATCC). McCoy's 5A media supplemented with 10% v/v Fetal Bovine Serum (FBS), and 1% antibiotic-antimycotic solution (1 mL contains 10,000 U Penicillin, 10 mg Streptomycin and 25 µg Amphotericin B) was used to grow and maintain HCT116 cells. A humidified environment constituted the standard conditions for cell culture at 37°C with 5% CO₂.

Cell viability assay

To determine the cytotoxicity of **M1** and **M2** extracts on colon cancer HCT-116 cell line, MTT assay was used. In a 96-well plate, the cells (5×10^3 cells/well) were cultured for 24 h. The cells were treated with **M1** and **M2** at varied concentrations (100, 500, 1,000, and 5,000 µg/mL) for 24 h, respectively. Each well received 10 µL of MTT solution (5 mg/mL) and was subjected to further 3 h incubation at 37°C. In order to dissolve the purple formazan crystals, 100 µL of dimethyl sulfoxide (DMSO) was added to each well. A microplate reader measured absorbance at 570 nm (Bio-Rad, United States). The cell viability was expressed as a percentage (%) over the untreated control. For calculating the IC₅₀ value, GraphPad Prism Professional software was used.

Morphological analysis

The effects of **M1** and **M2** extract treatment on the morphology of HCT-116 cells were investigated using a phase contrast microscope. Briefly, HCT-116 cells (5×10^3) were cultivated in a 96-well plate before **M1** and **M2** (100, 500, 1,000, and 5,000 µg/mL) treatment. The alteration in morphology of **M1** and **M2**-treated cells was then examined using a phase contrast microscope (Labomed, United States).

Trypan blue exclusion assay

Trypan blue dye exclusion assay was performed further to confirm the **M1** and **M2**-mediated cytotoxicity in HCT-116 cells. A hemocytometer and a microscope were used to count the cells (5×10^4) after they had been co-cultured with and without **M1** and **M2** (100, 500, 1,000, and 5,000 µg/mL), respectively for 24 h. The proportion of dead cells in each treatment set from studies done in triplicates was used to express the results.

Lactate dehydrogenase release assay

In accordance with the manufacturer's instructions, the lactate dehydrogenase (LDH) release assay kit was used to measure the level of cellular cytotoxicity. First, **M1** and **M2** were applied to the HCT-116 cells using a 96-well plate at different doses (100, 500, 1,000, and 5,000 µg/mL) for 24 h. LDH release kit was then used to detect released LDH in both the **M1** and **M2**-treated HCT-116 cells in the incubation medium.

Computational details

Computational calculations were carried out on a Dell workstation (Galax GeForce GTX 1660 Ti) equipped with 8-core processors, 64 GB Ram, and NVIDIA graphics card.

Receptor and ligands preparation

The co-crystal structure of Bcl-2 complex (PDB ID: 5JSN) was selected for virtual screening and molecular dynamics studies. In the crystallographic structure of this complex, there is a gap at position 33–86, which was fixed by homology modelling using the Swiss modeller tool. Based on literature reports (Mia et al., 2020), a total of ninety-four ($n = 94$) phytoconstituents of different chemical classes (Supplementary Table S2) were selected. Chemical structures (.sdf format) of the compounds were retrieved from the NCBI PubChem database (Wang et al., 2009). The downloaded files were converted to .pdb format using the Open Babel software. The ligand files were prepared using AutoDock Tools 1.5.7 (the Scripps Research Institute, La Jolla, CA, United States) software and finally written as .pdbqt file format for docking studies (Ahamad et al., 2021a).

Active site prediction

The Bcl-2 protein (PDB ID: 5JSN) was given as input to identify the active site, which gives significant insight into recognizing surface structural pockets, shape and volume of every pocket, internal cavities of protein and surface areas. Next, the active site and the interactive residues were selected using PDBsum and CASTp online tools (Laskowski et al., 2018; Tian et al., 2018). The ligands were prepared using AutoDock Tools (ADT), and saved in pdbqt format (Trott and Olson, 2010).

Protein preparation and grid generation

The 3D structure of Bcl-2 was prepared using the ADT protein preparation wizard. The polar and missing hydrogen atoms were added, while water molecules and hetero-atoms were deleted (Forli et al., 2016). Energy minimization was performed with a default constraint of 0.3 Å root mean square (RMS) and charges were assigned. After protein preparation, clean structure was saved as pdbqt file. Grid box (84 Å × 82 Å × 84 Å) was generated around the centroid of compounds with assigned X, Y, and Z axis.

Virtual screening and binding affinity calculation

To identify the potential compounds found in *P. dactylifera* L., a dataset of ninety-four compounds was utilized for the virtual screening. The pdbqt files were provided as input and screened against Bcl-2 (Duffy and Avery, 2012). Top two compounds (ranked based on the binding energy scores and the docking poses) were selected for further studies (Trott and Olson, 2010; Forli et al., 2016). The compounds with favourable binding poses

were identified with the help of the lowest free energy (ΔG), defined using the equation as follows,

$$\Delta G = \Delta G_{\text{complex}} - (\Delta G_{\text{enzyme}} + \Delta G_{\text{ligand}})$$

Where ($\Delta G_{\text{complex}}$), ($\Delta G_{\text{receptor}}$), and (ΔG_{ligand}) are the average values of Gibbs free energy for the complex, receptor, and ligand, respectively. The stability of the docked complex between the receptor-ligand exhibits more negative scores, revealing the high potency of the inhibitor. All the other docking parameters were kept default, and the docked complexes final visualization was performed using PyMOL tool (DeLano, 2002). The active pocket of Bcl-2 and docked pose of the top-ranked compounds were compared to find interactive orientations.

Molecular dynamics (MD) simulation

MD simulations were performed for the best-docked complexes with maximum binding affinity scores using GROningen MAchine for Chemical Simulations (GROMACS) version 5.18.3. Package (Abraham et al., 2015). The topology of Bcl-2 was generated using GROMOS9643a1 force field (Van Der Spoel et al., 2005). Due to the lack of suitable force field parameters for a drug-like molecule in the GROMACS software, the PRODRG server was used for the generation of molecular topologies and coordinate files (Schüttelkopf and Van Aalten, 2004). All the systems were solvated using a simple point charge model (SPC/E) in a cubic box. To neutralize the system 0.15 M counter ions (Na^+ and Cl^-) were added. The energy minimization of all the neutralized systems was performed using the steepest descent and conjugate gradients (50,000 steps for each). The constant number of particles, volume, and temperature (NVT) ensemble and constant number of particles, pressure, and temperature (NPT) ensemble were run for system equilibration (Ahmad et al., 2021b). Steepest descent followed by conjugate gradient algorithms was utilized on enzyme-ligand complexes. The NVT ensemble was employed at a constant temperature of 300 K and a constant pressure of 1 bar. The SHAKE algorithm was used to confine the H atoms at their equilibrium distances and periodic boundary conditions. Moreover, the Particle Mesh Ewald (PME) method defines long-range electrostatic forces (Abraham et al., 2015). The cut-offs for van der Waals and coulombic interactions were set as 1.0 nm. LINC algorithm was used to constrain the bonds and angles. Using the NPT ensemble, production runs were performed for 500 ns, with time integration. The energy, velocity, and trajectory were updated at the time interval of 10 ps. The analysis is performed by using Ca-atom deviations of the protein calculated using root mean square deviations (RMSD). The relative fluctuations of each amino acid were defined with root mean square fluctuations (RMSF). To measure the compactness of a given molecule radius of gyration (R_g) is implemented, and the solvent accessible surface area (SASA) was employed to know the electrostatic contributions of molecular solvation (Ahmad et al., 2018; Ahmad et al., 2019).

Results

Extraction and characterization of *Shishi* (M1) and *Majdool* (M2) fruits

Aqueous, organic or mixture solvents can extract a natural product, depending upon the analyte of interest (Nematallah et al., 2018). In the

past, it has been demonstrated that when date fruits are extracted with an organic solvent, it yields bioactive compounds able to inhibit colon, liver and cervical cancerous cell lines *in vitro* (Mansour et al., 2011; Ravi, 2017). Especially, using a polar solvent such as methanol allows the extraction of various components (MHM et al., 2015). It was found that the alcoholic extract of the date fruits effectively inhibits α -glucosidase and α -amylase enzymes with low IC_{50} values in both *in-vitro* and *in-vivo* (El Abed et al., 2017). In 2016, Khan et al. (2016) demonstrated that the methanolic extract of *Ajwa* Date (Saudi origin) inhibits breast cancer (MCF-7) cell lines *via* cell cycle arrest and apoptosis. Motivated by this, we selected methanol as the extracting solvent in this study too. Cold extraction of finely cut **M1** (36.11 gm) and **M2** (46.22 gm) followed by concentration and lyophilization yielded **M1** and **M2** as a light-yellow water-soluble powder (9.1 and 15 gm of **M1** and **M2**, respectively). To identify the extract components, chromatographic (LC-MS & GC-MS) and spectroscopic (FT-IR) techniques (Supplementary Figures S1, S2) were employed.

Lyophilized products were analyzed by LC-MS (negative mode) using gradient mobile phase as it allows easy detection of phenolic acids & flavonoids as they contain acidic hydroxy group. It has been reported that the methanolic extract contains phenolic acids, flavonoid diglucosides, monoglucosides, acylated monoglucosides, free aglycones, lipids and others when analyzed under similar conditions (Farg et al., 2014). Table 1 collects the identities and molecular/fragment ions of some major components present in **M1** and **M2** as identified by comparing LC-MS (negative mode, Figure 1) results with the literature. Both varieties exhibit similar chromatograms, with **M1** having relatively more fraction than **M2**.

GC-MS analysis further indicated the presence of several phytochemical belongings of different classes. For example, quinic acid, oleic acid, trans-13-octadecenoic acid, stearic acid, O-caffeoyl shikimic acid, luteolin, trihydroxy-octadecenoic acid, stearic acid linoleic acid, 6-hydroxy 7 methoxy coumarin, 4-hydroxy 6-methylcoumarin and amino acids were tentatively identified (Supplementary Table S3). These components and other metabolites have been well-identified in different varieties of date fruits (Farg et al., 2016; Abdul-Hamid et al., 2019; Perveen and Bokahri, 2020; Souda et al., 2020; Ibrahim et al., 2021).

The FTIR spectrum of the methanolic extracts (Supplementary Figure S1) is also consistent with previous literature (Alam et al., 2022). It has been reported that IR spectrum of dates extracts exhibits multiple peaks responsible for functionalities present in lipid ($2,960\text{--}2,850\text{ cm}^{-1}$), amide ($3,299\text{--}3,399\text{ cm}^{-1}$ and $1,591\text{--}1,529\text{ cm}^{-1}$ for amine and $1,619\text{--}1,691$ for carbonyl) and carbohydrates ($900\text{--}1,200\text{ cm}^{-1}$). As it is clear, the IR spectra of **M1** and **M2** are identical. The spectrum shows a stretching vibrations band at $3,280\text{ cm}^{-1}$ attributed to -OH group, bands at $2,888$ and $2,930\text{ cm}^{-1}$ attributed to $\text{C}_{\text{sp}^3}\text{-H}$ stretching vibration, aromatic $\text{C}=\text{C}$ stretching vibrations at $1,622\text{ cm}^{-1}$ and C-O deformation vibrations of aliphatic alcohols at $1,009\text{ cm}^{-1}$ (Alam et al., 2022).

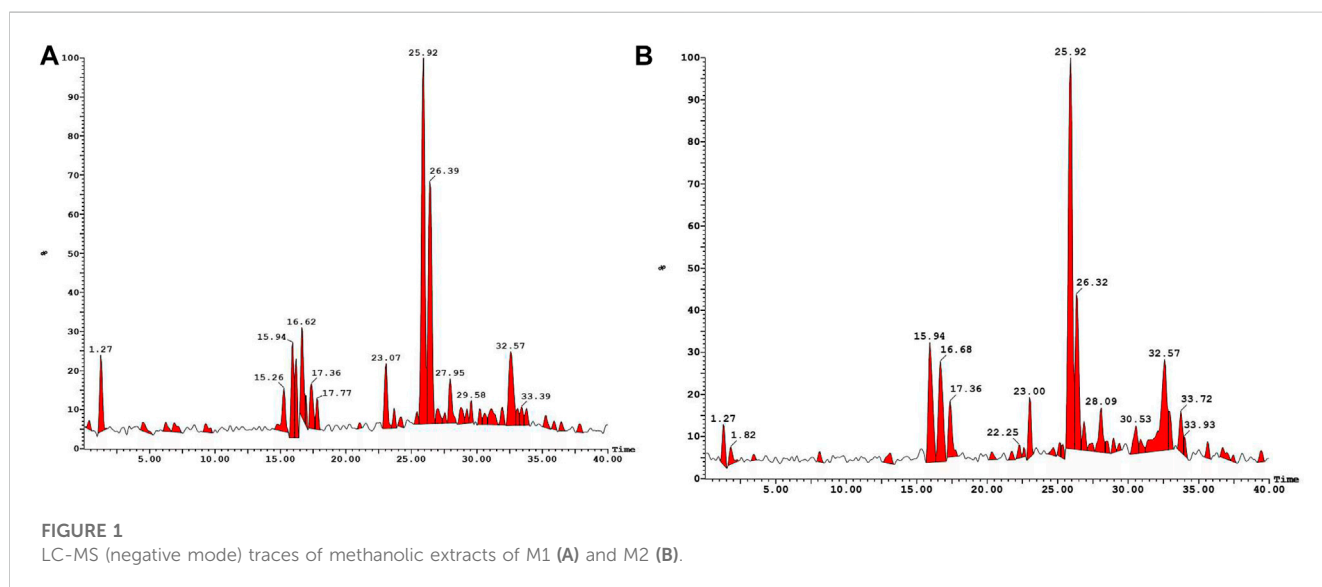
Biological studies

Antiproliferative and cytotoxic effect

Using the MTT test, the antiproliferative and cytotoxic effects of the **M1** and **M2** date extract were assessed against colon cancer HCT116 cells for 24 h (Figures 2A, B). The **M1** and **M2** extracts

TABLE 1 LC-MS (negative mode) results of M1 and M2.

S. No	M-H	Mol. Form.	Identification	Ref.
1.	180.1	C ₆ H ₁₂ O	β-D-Glucopyranose	Najm et al. (2021)
2.	341.1	C ₁₅ H ₁₈ O ₉	Caffeic acid hexoside	Farag et al. (2014)
3.	179.1	C ₁₆ H ₁₅ O ₈	O-Caffeoyl shikimic acid	Otify et al. (2019)
4.	322.0	C ₁₂ H ₁₉ O ₁₀	Anhydro dihexose	Otify et al. (2019)
5.	425.3	C ₁₇ H ₂₉ O ₁₂	Acyl sucrose	Najm et al. (2021)
6.	463.1	C ₂₁ H ₂₀ O ₁₂	Isoquercetin	Farag et al. (2014)
7.	476.0	C ₁₂ H ₂₁ O ₁₂	Isorhamnetin-3-O-glucoside	Najm et al. (2021)
8.	476.0	C ₂₃ H ₄₃ NO ₇ P	Sphingolipid conjugate I	Otify et al. (2019)
9.	311.3	C ₁₈ H ₃₁ O ₄	Dihydroxy linolenic acid	Farag et al. (2014)
10.	277.4	C ₁₈ H ₂₉ O ₂	Linolenic acid	Farag et al. (2014)
11.	255.6	C ₁₆ H ₃₁ O ₂	Palmitic acid	Farag et al. (2014)
12.	279.4	C ₁₈ H ₃₁ O ₂	Linoleic acid	Farag et al. (2014)
13.	283.2	C ₁₈ H ₃₆ O ₂	Stearic acid	Eid et al. (2013)
14.	195.0	C ₁₀ H ₁₀ O ₄	ferulic acid	Nematallah et al. (2018)
15.	326.3	C ₁₈ H ₃₂ O ₅	Trihydroxy octadecadienoic acid	Nematallah et al. (2018)
16.	311.4	C ₁₈ H ₃₂ O ₄	Dihydroxy octadecadienoic acid	Nematallah et al. (2018)
17.	594.9	C ₂₇ H ₃₀ O ₁₅	Luteolin-7-O-rutinoside	Nematallah et al. (2018)



exhibited strong and dose-dependent cytotoxic potential in HCT116 cells. The % cell viability of M1- and M2-treated HCT116 cells were found to be $88.80 \pm 1.33\%$, $63.26 \pm 3.47\%$, $45.24 \pm 2.80\%$, and $15.28 \pm 1.53\%$; and $83.07 \pm 2.37\%$, $59.35 \pm 4.72\%$, $28.90 \pm 1.49\%$ and $10.63 \pm 1.47\%$ at a dose of 100, 500, 1,000, and 5,000 $\mu\text{g/mL}$, respectively. IC_{50} values were determined to be 591.3 $\mu\text{g/mL}$ and 449.9 $\mu\text{g/mL}$ for M1 and M2, respectively,

revealing the inhibitory potential (Figures 2C, D). Our findings thus demonstrated that both the M1 and M2 inhibit colon cancer cell proliferation in a dose-dependent manner.

Morphological alterations

Under a phase contrast microscope, the images of control and M1 & M2-treated HCT-116 cells revealed discernible morphological

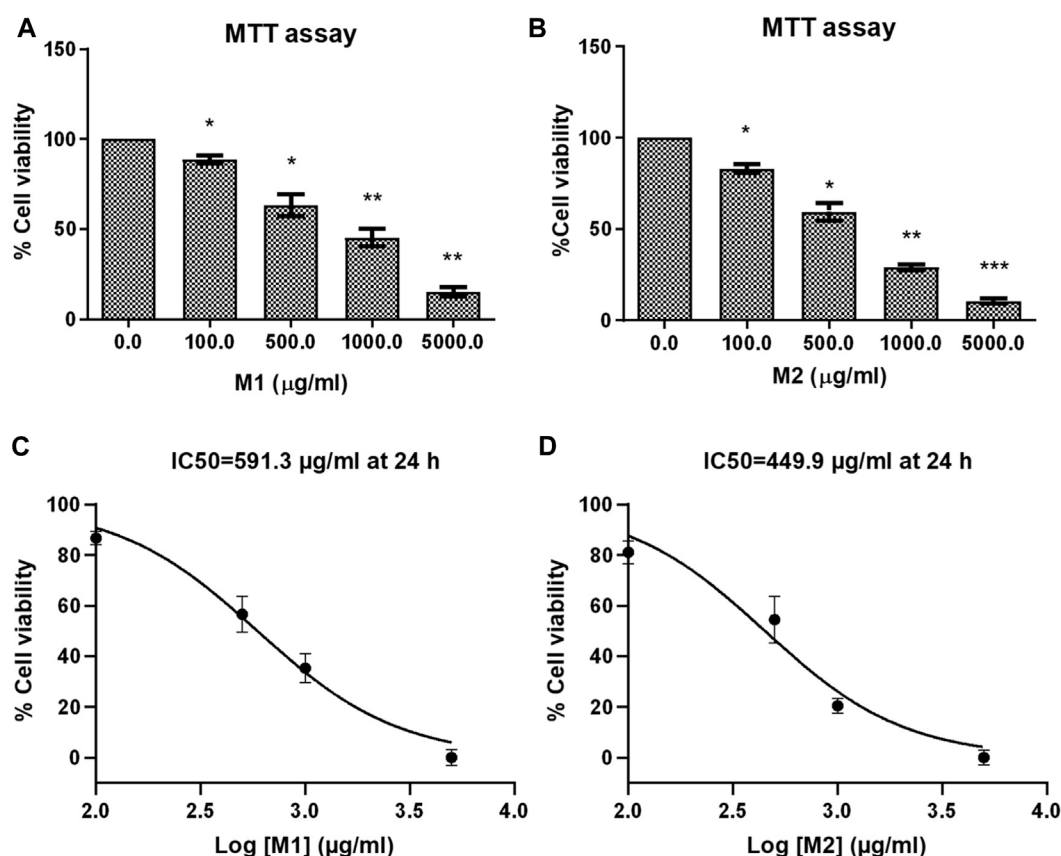


FIGURE 2

Effect of date extracts M1 and M2 on HCT-116 cells. (A, B) Percent (%) cell viability of HCT-116 cells treated with different doses of M1 and M2 (100–5,000 μg/mL) for 24 h. The results shown are the mean ± SEM of three independent experiments performed in triplicate (ns > 0.01, **p* < 0.01, ***p* < 0.001, and ****p* < 0.0001 represent significant differences compared with control). (C, D) Graph showing IC₅₀ of M1 and M2 against HCT-116 colon cancer cell at 24 h.

alterations. The control cells showed increased cell growth and intact cell shape. However, in a dose-dependent manner (100, 500, 1,000, and 5,000 μg/mL), significant morphological modifications were observed in the M1 and M2-treated HCT-116 cells (Figures 3A, B). Moreover, M1 and M2-treated HCT-116 cells showed increased detachment and cytoplasmic shrinkage, which led to a rise in the number of floating cells. The findings thus support the hypothesis that treatment with M1 and M2 causes cytotoxicity in HCT-116 colon cancer cells.

M1 and M2 causes cell death in HCT-116 cells

Trypan blue dye exclusion assay was used to assess how M1 and M2-treated HCT-116 cells lost viability. Figures 4A, B illustrates the considerable increase in cell mortality in HCT-116 cells after exposure to M1 and M2 at various doses (100, 500, 1,000, and 5,000 μg/mL) for 24 h. This result supported the cytotoxic action of M1 and M2 on colon cancer cells.

Release of cellular LDH in HCT-116 cells

LDH release assay displayed that treatment with both the M1 and M2 in HCT-116 cells mediated significant release of LDH, which showed the degree of cellular membrane damage post-treatment. Higher M1 and M2 concentrations were found to be significantly more cytotoxic, as evidenced by increased

cytotoxicity in HCT-116 cells (Figures 5A, B). The percentage cytotoxicity in M1-treated HCT-116 cells, after 24 h of treatment, was found to be 112.08% ± 3.42%, 145.78% ± 3.88%, 174.07% ± 2.25%, and 190.03% ± 2.64% at 100, 500, 1,000, and 5,000 μg/mL dose, respectively. Similarly, after 24 h of treatment with M2, HCT-116 cells exhibited percent cytotoxicity of 120.53% ± 3.13%, 144.07% ± 3.00%, 168.46% ± 4.29%, 192.15% ± 1.98% at 100, 500, 1,000, and 5,000 μg/mL dose, respectively. Thus, our results suggest that both the M1 and M2 were able to decrease the viability and proliferation in colon cancer cells.

In Silico studies

Virtual screening

It has been demonstrated that the *Phoenix dactylifera* L. extract exhibits anticancer activity by modulating Bcl-2-family proteins (Khan et al., 2021). To identify the critical component(s) responsible for the anticancer activity, we conducted exhaustive *in silico* studies. To this end, ligand-based virtual screening was performed using ninety-four compounds found in *P. dactylifera* against the receptor (PDB: 5JSN). It has been reported that the small molecule may interact with the various

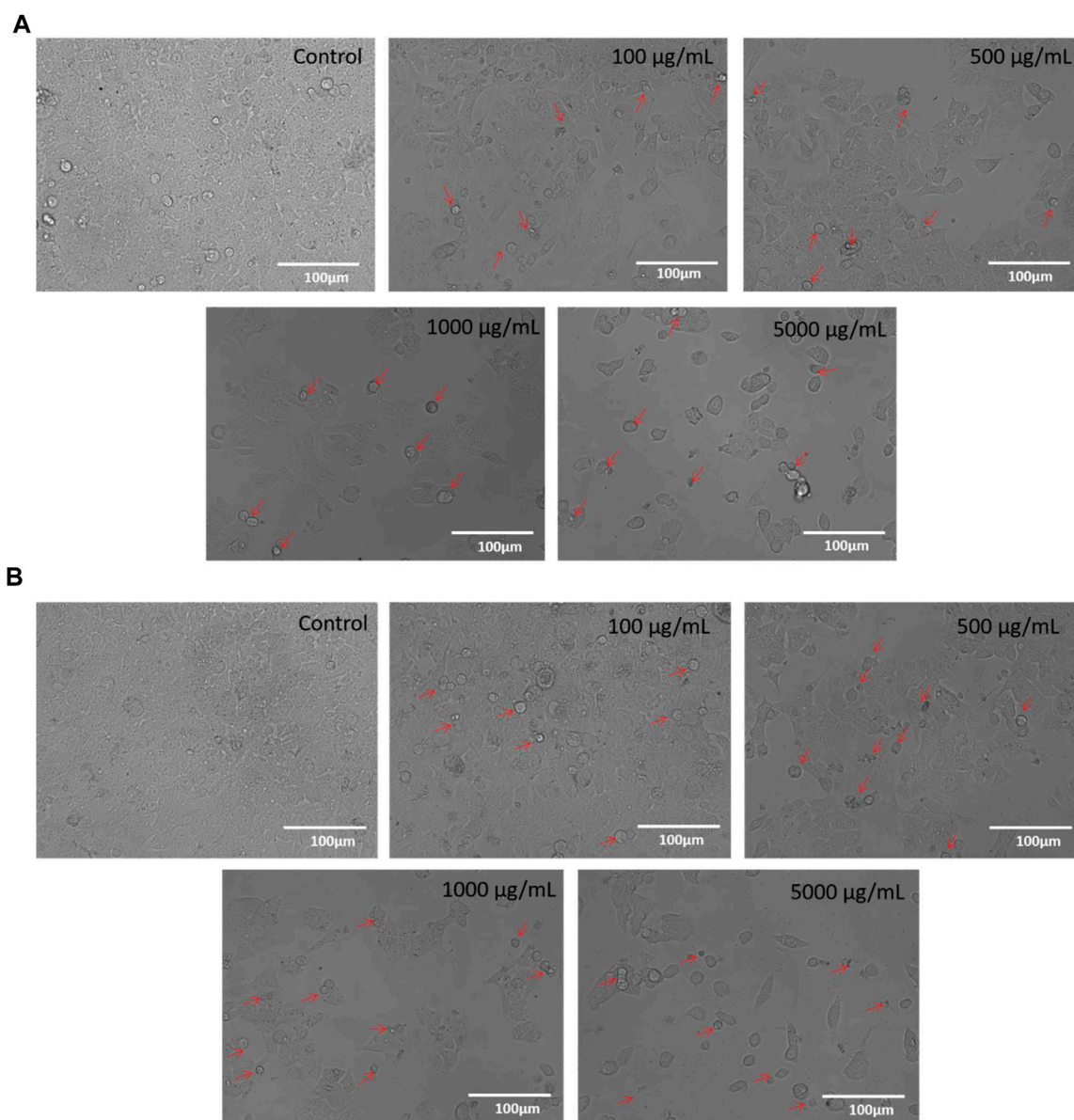


FIGURE 3

(A): Phase-contrast images of HCT-116 cells treated with either vehicle or different doses of M1 (100–5,000 µg/mL) for 24 h. The photomicrographs shown are the representatives of three independent experiments. (B): Phase-contrast images of HCT-116 cells treated with either vehicle or different doses of M2 (100–5,000 µg/mL) for 24 h. The photomicrographs shown are the representatives of three independent experiments.

receptors of Bcl-2 protein *via* multiple non-covalent interactions (Khosravi et al., 2022; Taghizadeh et al., 2022). Among others, Lys22, Arg26, Asp102, Ser105, Arg106, Arg109, Phe112, Val156, Val159, Asp163, Glu160, and Glu209 which participates in H-bonds and steric interactions (Khosravi et al., 2022). Based on the free binding energies and docking poses, virtual screening of the ligands resulted in procyanidin B2 and luteolin-7-*O*-rutinoside as the most potent candidates (Table 2). As depicted in Figure 6A, procyanidin B2 interacted with various amino acid residues *via* H-bonding (Ala100, Arg107, Asn143, Gly145, and Arg146) and other non-covalent interactions (such as hydrophobic and Van der Waal's) with a total binding energy of -9.3 kcal/mol. On the other hand, luteolin-7-*O*-rutinoside formed H-bond with Asp111, Asn143 and

Arg146 amino acids and yielded a binding energy of -9.1 kcal/mol (Figure 6B). Overall results revealed that the proposed two compounds have an edge over the Bcl-2 complexes attributable to more potent binding abilities.

Molecular dynamics (MD) simulations

To understand the complex stability and interaction profile of the most promising hit compounds inside the active site of Bcl-2, MD simulations of Bcl-2-native, procyanidin B2 and luteolin-7-*O*-rutinoside complexes were performed on a 500 nanosecond (ns) scale. In addition, structural parameters, including RMSD, RMSF, SASA, and Rg were

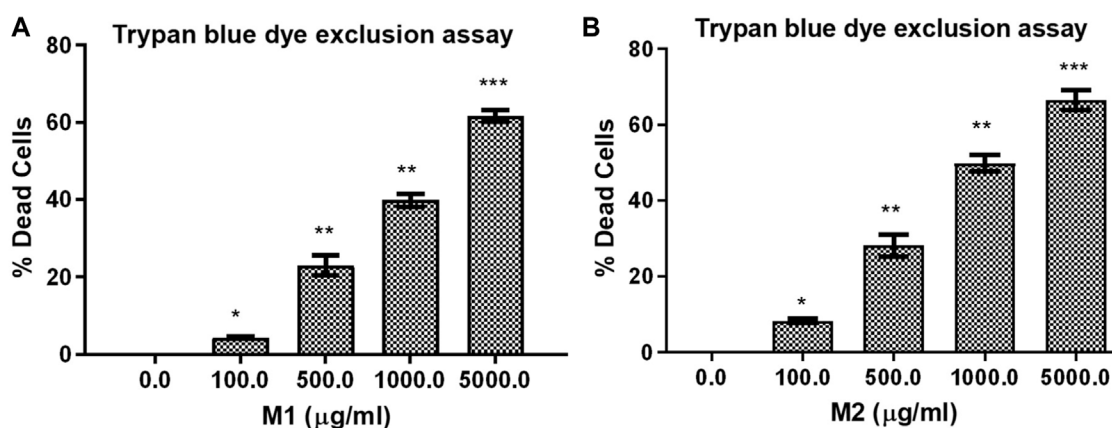


FIGURE 4

Trypan blue dye exclusion assay. Percent (%) dead cells in HCT-116 cells treated with different doses of (A) M1 and (B) M2 (100–5,000 µg/ml) for 24 h. The results shown are the mean ± SEM of three independent experiments performed in triplicate (ns > 0.01, **p* < 0.01, ***p* < 0.001, and ****p* < 0.0001 represent significant difference compared with control).

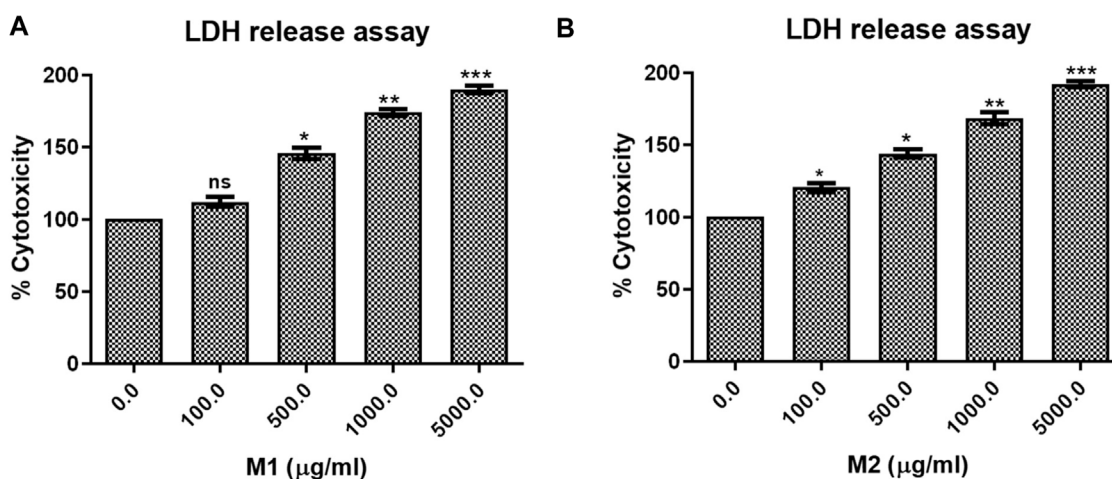
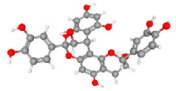
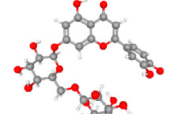


FIGURE 5

LDH release assay. Percent cytotoxicity in HCT-116 cells treated with different doses of (A) M1 and (B) M2 (100–5,000 µg/ml) for 24 h. The results shown are the mean ± SEM of three independent experiments performed in triplicate (ns > 0.01, **p* < 0.01, ***p* < 0.001, and ****p* < 0.0001 represent significant differences compared to control).

TABLE 2 Ligands name, 3D structures, SMILE format and the virtual screening outputs.

S. No.	Ligand	3D structure	SMILE format	Binding score (kcal/mol)	H-bond residues
1	Procyanidin B2		<chem>C1C(OC2=C1C(=CC3=C2C4C(C(O3)OC5=CC(=CC(=C45)O)O)C6=CC(=C(C=C6)O)O)O)C7=CC(=C(C=C7)O)O</chem>	−9.3	Ala100, Arg107, Asn143, Gly145, and Arg146
2	Luteolin-7-O-rutinoside		<chem>C1=CC(=C(C=C1C2=CC(=O)C3=C(C=C(C=C3O2)OC4C(C(C(O4)COC5C(C(C(C(O5)CO)O)O)O)O)O)O)O</chem>	−9.1	Asp111, Asn143 and Arg146

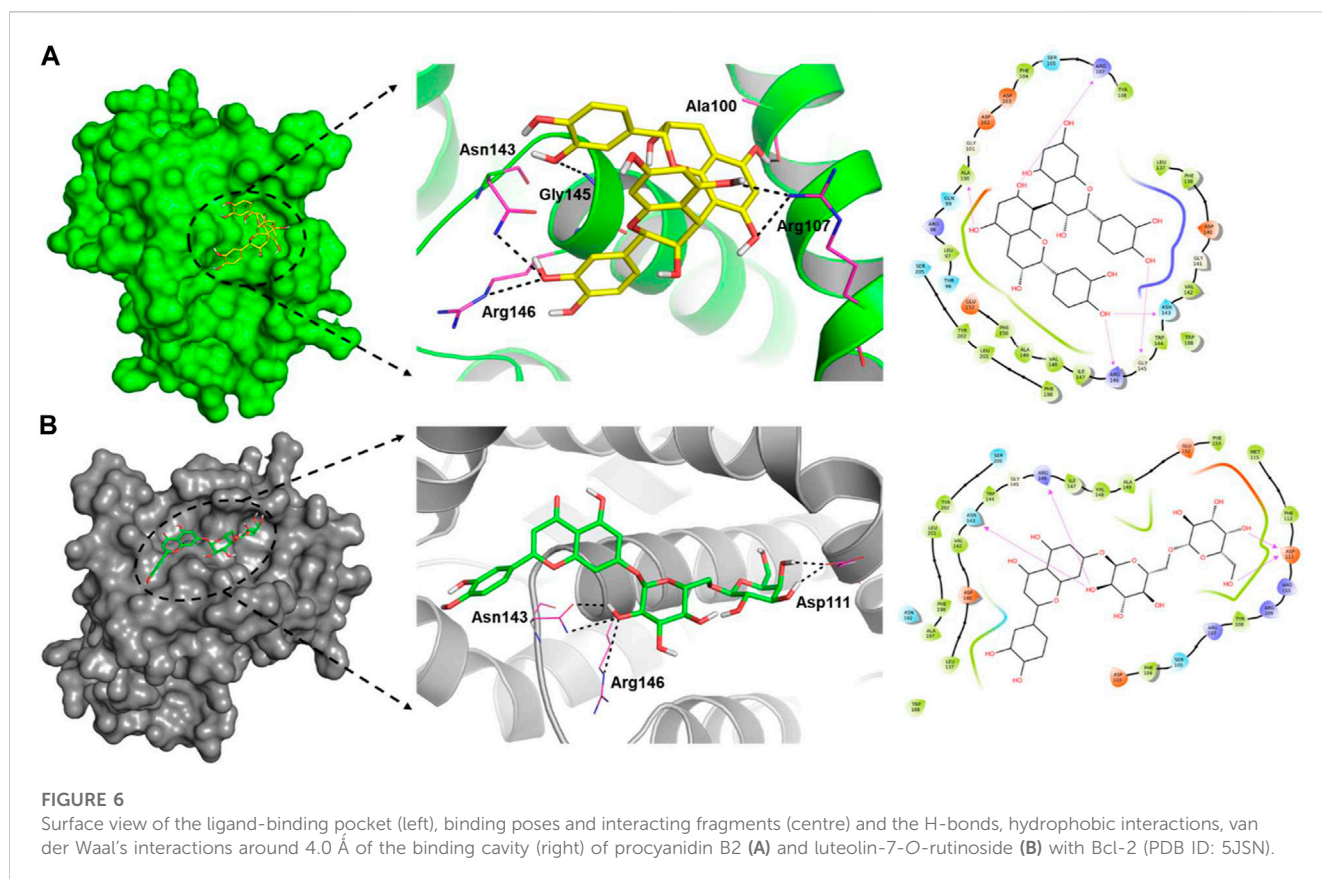


TABLE 3 The average RMSD, Rg, and SASA of the native and ligand-protein complexes.

System	RMSD (nm)	Rg (nm)	SASA (nm) ²
Native	0.92	1.56	111.73
Procyanidin B2	0.63	1.65	112.77
Luteolin-7-O-rutinoside	0.71	1.57	115.39

evaluated as a function of time and discussed in the following subsections.

RMS-deviation and RMS-fluctuations

The docked complexes were subjected to RMSD analysis to assess the residual flexibility of the Bcl-2 receptor. It was noted that the native protein exhibits higher RMSD fluctuation and reaches equilibrium between 0.8 nm and 1.0 nm. However, in the presence of procyanidin B2, it reached an equilibrium at 0.6 nm and showed steady RMSD (average RMSD value 0.92 nm, Table 3), which remained stable over the 500 ns MD simulation (Figure 7A). Similarly, luteolin-7-O-rutinoside and Bcl-2 complexes showed stable equilibrium at 0.6 nm–0.7 nm. Furthermore, they displayed minimal fluctuation over the 500 ns MD simulation. The average RMSD of Luteolin-7-O-rutinoside and Bcl-2 complexes was 0.71 nm. Overall, both procyanidin B2 and luteolin-7-O-rutinoside complexes exhibit stable RMSD values and have a

stable binding with Bcl-2 under the given simulation conditions. This also indicates that the studied compounds reached stable and reliable dynamic equilibriums, which bolstered the docking results.

Furthermore, RMSF analysis was implemented to identify the flexible and rigid regions of the complexes and to measure the average atomic flexibility of the C α -atoms of native Bcl-2 and docked complexes. In the case of native Bcl-2, amino acids residues such as Pro46 ~0.53 nm, Gly47 ~0.71 nm, Ile48 ~0.64 nm, Arg63 ~0.58 nm, Asp64 ~0.65 nm, Pro65 ~0.73 nm and Val66 ~0.67 nm showed higher fluctuations (Figure 7B). However, fluctuation at 104–112, 162–163 and 201–207 amino acids residue also was found to be higher while other amino acids remain stable. For example, in a complex with procyanidin B2, RMS fluctuations were found in the region Gly79 ~0.60 nm and Ala80 ~0.57 nm, which is acceptable as these amino acids did not participate in the binding. Similarly, the complex with luteolin-7-O-rutinoside showed RMS-fluctuations at Gly54 ~0.57 nm, Ala61 ~0.55 nm and Arg63 ~0.66 nm values. Overall, the RMSF displayed the highest degree of flexibility, exhibiting stable active site residues interaction compared to the native protein.

Hydrogen bond monitoring

To underpin the stability of the ligand-protein complex, the number of H-bond was monitored by analyzing the MD trajectories

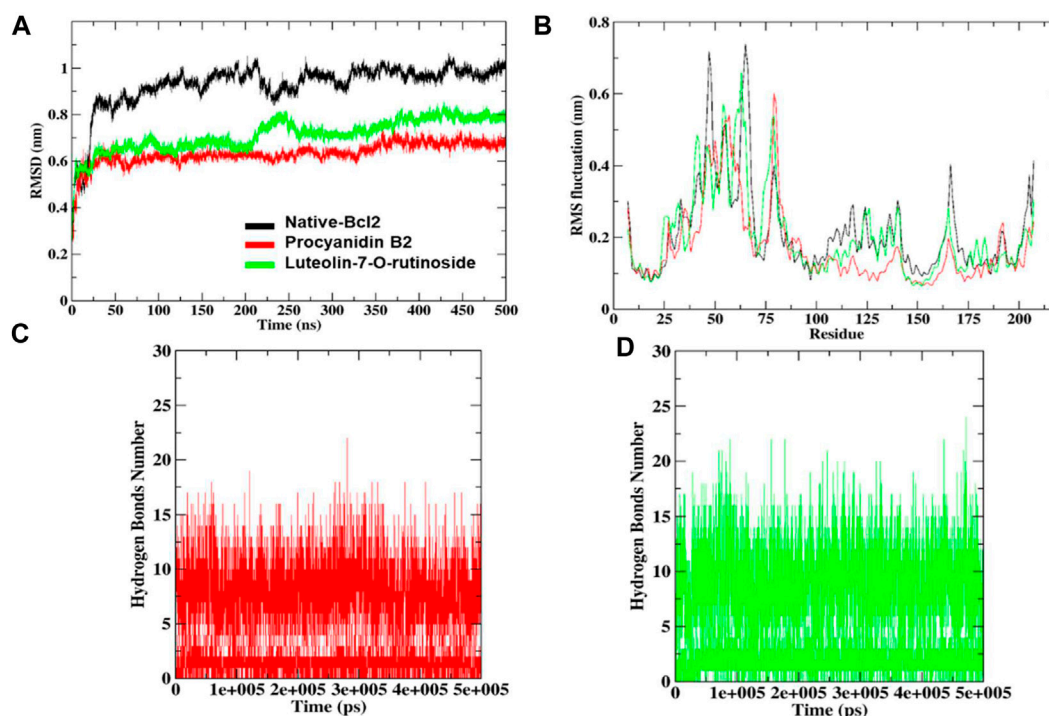


FIGURE 7 RMSD and RMSF analysis of the complexes of native protein Bcl-2, complexes of shortlisted ligands (A), combined RMS fluctuations (B). The number of hydrogen bonds formed by procyanidin B2 (C) and luteolin-7-O-rutinoside (D).

(Figures 7C, D). As can be seen, both compounds procyanidin B2 and luteolin-7-O-rutinoside formed 17 and 22 hydrogen bonds, respectively, which increased/remained the same during the 500 ns MD simulation.

Radius of gyration (Rg) and solvent accessible surface area (SASA)

Rg helps determine protein folding and unfold upon ligand binding, thus giving an idea about the stability of the complex during the simulation. A higher Rg indicates a less compact structure, while a lower Rg means more compactness (Sharma et al., 2022). We found that the average Rg values for the native Bcl-2 protein (1.56 nm) and luteolin-7-O-rutinoside complex (1.57 nm) were almost similar, indicating that the protein will likely maintain a relatively steady value and is stably folded (Figure 8A; Table 2). However, in the case of the procyanidin B2 complex, the average Rg value was 1.65 nm, indicating unfolded structure.

SASA was also conducted to ascertain the interactions between the protein-ligand complex and solvent during the 500 ns MD simulation (Figure 8B; Table 2). It was noted that the average SASA value for the complexes (112.77 and 115.39 nm²) of procyanidin B2 and luteolin-7-O-rutinoside, respectively, was better than the native Bcl-2 protein (111.73 nm²).

Discussion

It has been long understood that the phytochemicals found in *Phoenix dactylifera* L. target and inhibit several important biochemical pathways contributing to disease development (Farang et al., 2014; Lamia and Mukti, 2021). The ethnopharmacological significance of *P. dactylifera* L., such as antioxidant, anti-inflammatory anticancer, antimicrobial, etc., is now well established (El Abed et al., 2018). The amount of phytoconstituents, and thus the bioactivity, depends on several factors, including the part of the plant (fruits, seeds, etc.), stage, geographical location, and others. For example, it has been demonstrated that date fruit seeds extract shows anticancer activity against pancreatic (Habib et al., 2014), colorectal (Rezaei et al., 2015), liver (Al-Sheddi, 2019), lung (Al-Sheddi, 2019), and breast (Al-Sheddi, 2019) and other (Al-Zubaidy et al., 2016; Hilary et al., 2021; Habib et al., 2022; Khan et al., 2022) cancer cell lines. On the other hand, Siddiqui et al. (2019) reported that the pulp extract of the *Ajwa* variety exhibit antiproliferative activity against human liver cancer cells (HepG2, IC₅₀ = 20.03 and 16.78 mg/mL at 24 and 48 h periods, respectively). Moreover, Khan et al. (2021) demonstrated the apoptosis-inducing potential of *Ajwa* date pulp extract against human triple-negative breast cancer cells (MDA-MB-231, IC₅₀ = 17.45 and 16.67 mg/mL at 24 and 48 h, respectively). Khattak et al. (2020) have reported the antiproliferative property of Emirati date fruits extract on human triple-negative breast cancer cell line MDA-MB-231. Besides, the antioxidant and apoptotic potentials of the

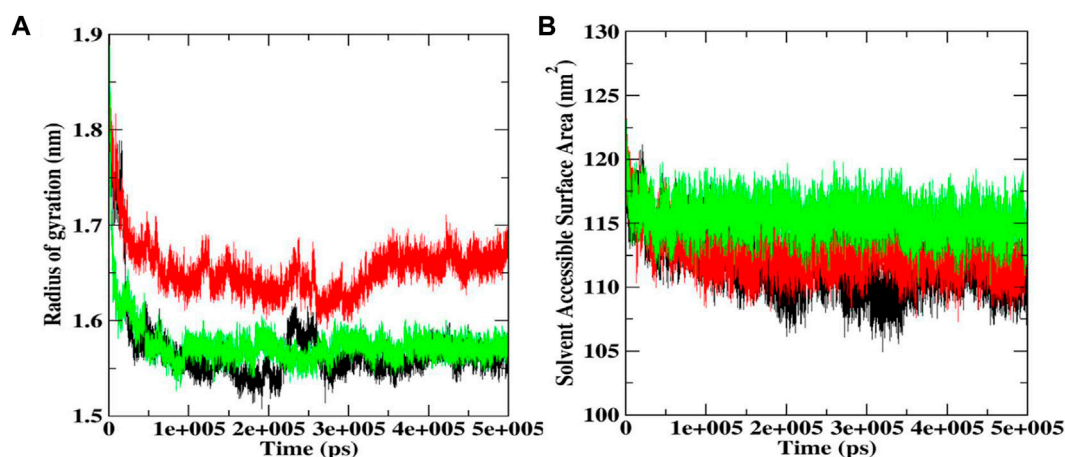


FIGURE 8

Rg (A) and SASA plot (B) during 500 ns MD simulations docked complexes of native Bcl-2, and complexes with procyanidin B2 and luteolin-7-O-rutinoside.

whole fruit (flesh and pit extracts) is also known (Shahbaz et al., 2022). In addition to the above-mentioned factors, the polarity of extracting solvents also plays an important role; therefore both aqueous and organic solvent systems have been investigated in the past. In a study, it was found that the aqueous extract of a number of date varieties (Saudi Arabian origin) was less bioactive than the methanolic counterparts (Zhang et al., 2017). In a remarkable study, Khan et al. (2016) noted that the methanolic extract of *Ajwa* date fruits exhibit strong anticancer effect on human breast adenocarcinoma (MCF7) (Khan et al., 2016). Besides, other researchers also noted the antitumor activity of methanolic extracts (Mansour et al., 2011; Thouri et al., 2019). Therefore, in the present study, we selected methanol as the solvent to extract date fruits of *Shishi* (M1) and *Majdool* (M2) cultivars grown in Ha'il region of Saudi Arabia.

As mentioned, (*vide-infra*), the methanolic extract concentrates were subjected to lyophilization and the resulting water-soluble products were used for further studies without any further purification. LC-MS (negative mode) and GC-MS analyses of the extract revealed the presence of various phytochemicals in both varieties. We noted that the chromatograms of M1 have more peaks than M2; therefore, the former has relatively more constituents. Among the main constituents identified were flavonoids, sphingolipids, and fatty acids classes of phytochemicals. Several researchers already report the presence of these constituents in a wide variety of date fruits (see references in the result section). At the same time, we firmly believe the presence of other components escaped the detection. MTT assay of the extracts against colon cancer cells (HCT-116) revealed a dose-dependent inhibitory nature of the compounds (IC_{50} = 591.3 μ g/mL and 449.9 μ g/mL for M1 and M2 at 24, respectively).

It has been demonstrated that the *P. dactylifera* L. extract exhibits anticancer activity by modulating Bcl-2-family proteins which is also expressed in the HCT-116 cell line (El-Far et al., 2021). Considering this, attempts have been made to identify the principal agent(s) present in the extract using computational approaches, including virtual screening and MD studies. Ligand-based virtual screening identified procyanidin B2 and luteolin-7-O-rutinoside as the most

probable candidates since they could bind with Bcl-2 protein efficiently through various amino acids. MD simulation study further strengthens this observation. Considering the earlier reported values and inhibition mechanism on other cell lines, we believe that the anticancer potential of both *Shishi* and *Majdool* date extracts against colon cancer cells is interesting and requires further biochemical investigation.

Conclusion

In conclusion, the anticancer activity of methanolic extract of two varieties of dates fruits (*Shishi* M1 and *Majdool* M2) grown in the Ha'il region of Saudi Arabia has been compared. The results of GC-MS and Ft-IR studies indicated the presence of various components in the M1 and M2 extracts, which are responsible for dose-dependent cytotoxicity against colon cancer cells (HCT116 cells) through morphological modifications, including cellular membrane damage. The IC_{50} value was 591.3 μ g/mL and 449.9 μ g/mL for M1 and M2, respectively. Furthermore, Trypan blue dye exclusion assay further supported the cytotoxic action of M1 and M2 on colon cancer cells. Extensive virtual screening combined with MD simulations studies at 500 ns revealed that procyanidin B2 and luteolin-7-O-rutinoside could be possible agents for the bioactivities. Overall, our data strongly suggest that the consumption of date fruits might prove helpful against colon cancer. Moreover, we also believe that these varieties of date fruits could be utilized as a source of bioactive phytochemicals, leading to the development of Ha'il, Saudi Arabia.

Data availability statement

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

Author contributions

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

Acknowledgments

This research has been funded by Scientific Research Deanship at University of Ha'il—Saudi Arabia through project number “RD-21 085”.

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

- Abdul-Hamid, N. A., Abas, F., Ismail, I. S., Tham, C. L., Maulidiani, M., Mediani, A., et al. (2019). Metabolites and biological activities of phoenix dactylifera L. pulp and seeds: A comparative MS and nmr based metabolomics approach. *Phytochem. Lett.* 31, 20–32. doi:10.1016/j.phytol.2019.03.004
- 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, 19–25. doi:10.1016/j.softx.2015.06.001
- Ahamad, S., Hassan, M. I., and Dwivedi, N. (2018). Designing of phenol-based β -carbonic anhydrase1 inhibitors through QSAR, molecular docking, and MD simulation approach. *3 Biotech.* 8, 256. doi:10.1007/s13205-018-1278-z
- Ahamad, S., Islam, A., Ahmad, F., Dwivedi, N., and Hassan, M. I. (2019). 2/3D-QSAR, molecular docking and MD simulation studies of FtsZ protein targeting benzimidazoles derivatives. *Comput. Biol. Chem.* 78, 398–413. doi:10.1016/j.compbiolchem.2018.12.017
- Ahamad, S., Kanipakam, H., Birla, S., Ali, M. S., and Gupta, D. (2021a). Screening Malaria-box compounds to identify potential inhibitors against SARS-CoV-2 Mpro, using molecular docking and dynamics simulation studies. *Eur. J. Pharmacol.* 890, 173664. doi:10.1016/j.ejphar.2020.173664
- Ahamad, S., Kanipakam, H., Kumar, V., and Gupta, D. (2021b). A molecular journey to check the conformational dynamics of tau tubulin kinase 2 mutations associated with Alzheimer's disease. *RSC Adv.* 11, 1320–1331. doi:10.1039/d0ra07659g
- Al-Alawi, R. A., Al-Mashiqri, J. H., Al-Nadabi, J. S., Al-Shihi, B. I., and Baqi, Y. (2017). Date palm tree (Phoenix dactylifera L.): Natural products and therapeutic options. *Front. Plant Sci.* 8, 845. doi:10.3389/fpls.2017.00845
- Al-Sheddi, E. (2019). Anticancer potential of seed extract and pure compound from Phoenix dactylifera on human cancer cell lines. *Pharmacogn. Mag.* 15, 494–499. doi:10.4103/pm.pm_623_18
- Al-Zubaidy, N., Al-Zubaidy, A., and Sahib, H. (2016). The anti-proliferative activity of phoenix dactylifera seed extract on MCF-7 breast cancer cell line. *Int. J. Pharm. Sci. Rev. Res.* 41, 358–362.
- Alam, M. Z., Ramachandran, T., Antony, A., Hamed, F., Ayyash, M., and Kamal-Eldin, A. (2022). Melanin is a plentiful bioactive phenolic compound in date fruits (Phoenix dactylifera L.). *Sci. Rep.* 12, 6614–6712. doi:10.1038/s41598-022-10546-9
- Alghamdi, A. A., Awadelkarem, A. M., Hossain, A., Ibrahim, N. A., Fawzi, M., and Ashraf, S. A. (2018). Nutritional assessment of different date fruits (Phoenix dactylifera L.) varieties cultivated in Hail province Saudi Arabia. *Biosci. Biotechnol. Res. Commun.* 11, 263. doi:10.21786/bbrc/11.2/11
- Ali, I., Rahisuddin, S. K., Haque, A., and El-Azzouny, A. (2010). Natural products: Human friendly anticancer medications. *Egypt Pharma J.* 9, 133–179.
- Altemimi, A., Lakhssassi, N., Baharlouei, A., Watson, D. G., and Lightfoot, D. A. (2017). Phytochemicals: Extraction, isolation, and identification of bioactive compounds from plant extracts. *Plants* 6, 42. doi:10.3390/plants6040042
- Assirey, E. R. (2015). Nutritional composition of fruit of 10 date palm (Phoenix dactylifera L.) cultivars grown in Saudi Arabia. *J. Taibah Univ. Sci.* 9, 75–79. doi:10.1016/j.jtusci.2014.07.002
- Atanasov, A. G., Zotchev, S. B., Dirsch, V. M., and Supuran, C. T. (2021). Natural products in drug discovery: Advances and opportunities. *Nat. Rev. Drug Discov.* 20, 200–216. doi:10.1038/s41573-020-00114-z
- Borochoy-Neori, H., Judeinstein, S., Greenberg, A., Volkova, N., Rosenblat, M., and Aviram, M. (2013). Date (Phoenix dactylifera L.) fruit soluble phenolics composition and anti-atherogenic properties in nine Israeli varieties. *J. Agric. food Chem.* 61, 4278–4286. doi:10.1021/jf400782v
- Complexity, O.-T. O. O. E. (2022). Dates, fresh or dried. Available at <https://oec.world/en/profile/hs/dates-fresh-or-dried#:~:text=Trade%20in%20Dates%2C%20fresh%20or>.
- Delano, W. L. (2002). Pymol: An open-source molecular graphics tool. *CCP4 Newsl. protein Crystallogr.* 40, 82–92.
- Duffy, S., and Avery, V. M. (2012). Development and optimization of a novel 384-well anti-malarial imaging assay validated for high-throughput screening. *Am. J. Trop. Med. Hyg.* 86, 84–92. doi:10.4269/ajtmh.2012.11-0302
- Echegaray, N., Gullón, B., Pateiro, M., Amarowicz, R., Misihairabgwi, J. M., and Lorenzo, J. M. (2021). Date fruit and its by-products as promising source of bioactive components: A review. *Food Rev. Int.* 2021, 1–22. doi:10.1080/87559129.2021.1934003
- Eid, N. M., Al-Awadi, B., Vauzour, D., Oruna-Concha, M. J., and Spencer, J. P. (2013). Effect of cultivar type and ripening on the polyphenol content of date palm fruit. *J. Agric. food Chem.* 61, 2453–2460. doi:10.1021/jf303951e
- El Abed, H., Chakroun, M., Abdelkafi-Koubaa, Z., Drira, N., Marrakchi, N., Mejdoub, H., et al. (2018). Antioxidant, anti-inflammatory, and antitumoral effects of aqueous ethanolic extract from Phoenix dactylifera L. parthenocarpic dates. *BioMed Res. Int.* 2018, 1542602. doi:10.1155/2018/1542602
- El Abed, H., Chakroun, M., Fendri, I., Makni, M., Bouaziz, M., Drira, N., et al. (2017). Extraction optimization and *in vitro* and *in vivo* anti-postprandial hyperglycemia effects of inhibitor from Phoenix dactylifera L. parthenocarpic fruit. *Biomed. Pharmacother.* 88, 835–843. doi:10.1016/j.biopha.2017.01.129
- El-Far, A. H., Ragab, R. F., and Mousa, S. A. (2021). *Date palm bioactive compounds: Nutraceuticals, functional nutrients, and pharmaceuticals, in the date palm genome*, (Germany: Springer)
- Farag, M. A., Handoussa, H., Fekry, M. I., and Wessjohann, L. A. (2016). Metabolite profiling in 18 Saudi date palm fruit cultivars and their antioxidant potential via UPLC-qTOF-MS and multivariate data analyses. *Food & Funct.* 7, 1077–1086. doi:10.1039/c5fo01570g
- Farag, M. A., Mohsen, M., Heinke, R., and Wessjohann, L. A. (2014). Metabolomic fingerprints of 21 date palm fruit varieties from Egypt using UPLC/PDA/ESI-qTOF-MS and GC-MS analyzed by chemometrics. *Food Res. Int.* 64, 218–226. doi:10.1016/j.foodres.2014.06.021
- Forli, S., Huey, R., Pique, M. E., Sanner, M. F., Goodsell, D. S., and Olson, A. J. (2016). Computational protein–ligand docking and virtual drug screening with the AutoDock suite. *Nat. Protoc.* 11, 905–919. doi:10.1038/nprot.2016.051
- Habib, H. M., El-Fakharany, E. M., Souka, U. D., Elsebae, F. M., El-Ziney, M. G., and Ibrahim, W. H. (2022). Polyphenol-rich date palm fruit seed (Phoenix dactylifera L.) extract inhibits labile iron, enzyme, and cancer cell activities, and DNA and protein damage. *Nutrients* 14, 3536. doi:10.3390/nu14173536
- Habib, H., Platat, C., Almaqbali, F., and Ibrahim, W. (2014). Date seed (Phoenix dactylifera) extract reduces the proliferation of pancreatic cancer cells, DNA damage and superoxide-dependent iron release from ferritin *in vitro* (829.20). *FASEB J.* 28, 829. doi:10.1096/fasebj.28.1_supplement.829.20

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/fchem.2023.1138057/full#supplementary-material>

- Hamad, I., Abdelgawad, H., Al Jaouni, S., Zinta, G., Asard, H., Hassan, S., et al. (2015). Metabolic analysis of various date palm fruit (Phoenix dactylifera L.) cultivars from Saudi Arabia to assess their nutritional quality. *Molecules* 20, 13620–13641. doi:10.3390/molecules200813620
- Hilary, S., Kizhakkayil, J., Souka, U., Al-Meqbaali, F., Ibrahim, W., and Platat, C. (2021). *In-vitro* investigation of polyphenol-rich date (Phoenix dactylifera L.) seed extract bioactivity. *Front. Nutr.* 8, 667514. doi:10.3389/fnut.2021.667514
- Ibrahim, S. A., Ayad, A. A., Williams, L. L., Ayivi, R. D., Gyawali, R., Krastanov, A., et al. (2021). Date fruit: A review of the chemical and nutritional compounds, functional effects and food application in nutrition bars for athletes. *Int. J. Food Sci. Technol.* 56, 1503–1513. doi:10.1111/ijfs.14783
- Khan, F., Ahmed, F., Pushparaj, P. N., Abuzenadah, A., Kumosani, T., Barbour, E., et al. (2016). Ajwa date (Phoenix dactylifera L.) extract inhibits human breast adenocarcinoma (MCF7) cells *in vitro* by inducing apoptosis and cell cycle arrest. *PLoS one* 11, e0158963. doi:10.1371/journal.pone.0158963
- Khan, M. A., Siddiqui, S., Ahmad, I., Singh, R., Mishra, D. P., Srivastava, A. N., et al. (2021). Phytochemicals from Ajwa dates pulp extract induce apoptosis in human triple-negative breast cancer by inhibiting AKT/mTOR pathway and modulating Bcl-2 family proteins. *Sci. Rep.* 11, 10322–10414. doi:10.1038/s41598-021-89420-z
- Khan, M. A., Singh, R., Siddiqui, S., Ahmad, I., Ahmad, R., Upadhyay, S., et al. (2022). Anticancer potential of phoenix dactylifera L. Seed extract in human cancer cells and pro-apoptotic effects mediated through caspase-3 dependent pathway in human breast cancer MDA-MB-231 cells: An *in vitro* and *in silico* investigation. *BMC complementary Med. Ther.* 22, 68–19. doi:10.1186/s12906-022-03533-0
- Khattak, M. N. K., Shanableh, A., Hussain, M. I., Khan, A. A., Abdulwahab, M., Radeef, W., et al. (2020). Anticancer activities of selected Emirati Date (Phoenix dactylifera L.) varieties pits in human triple negative breast cancer MDA-MB-231 cells. *Saudi J. Biol. Sci.* 27, 3390–3396. doi:10.1016/j.sjbs.2020.09.001
- Khosravi, A., Sharifi, I., Tavakkoli, H., Molaakbari, E., Bahraminegad, S., Salarkia, E., et al. (2022). Cytotoxicity of amphotericin B and AmBisome: *In silico* and *in vivo* evaluation employing the chick embryo model. *Front. Pharmacol.* 13, 1. doi:10.3389/fphar.2022.860598
- Lamia, F. S., and Mukti, R. F. (2021). Bangladeshi wild date palm fruits (Phoenix sylvestris): Promising source of anti-cancer agents for hepatocellular carcinoma treatment. *Int. J. Appl. Sci. Biotechnol.* 9, 32–37. doi:10.3126/ijasbt.v9i1.36110
- Laskowski, R. A., Jabłońska, J., Pravda, L., Vařeková, R. S., and Thornton, J. M. (2018). PDBsum: Structural summaries of PDB entries. *Protein Sci.* 27, 129–134. doi:10.1002/pro.3289
- Mansour, R. B., Lassoued, S., Dammak, I., Elgaied, A., Besbes, S., Attia, H., et al. (2011). Cytotoxicity evaluation and antioxidant activity of date seed oil from “Deglet-nour Tunisian cultivar” (Phoenix dactylifera L.). *Nat. Prod.* 7, 16–20.
- Maqsood, S., Adiamo, O., Ahmad, M., and Mudgil, P. (2020). Bioactive compounds from date fruit and seed as potential nutraceutical and functional food ingredients. *Food Chem.* 308, 125522. doi:10.1016/j.foodchem.2019.125522
- Mhm, A. E.-A., El-Meslami, A., Yassin, F., and Khalil, S. (2015). Identification phenolic and biological activities of methanolic extract of date palm pollen (Phoenix dactylifera). *J. Microb. Biochem. Technol.* 7, 047–050. doi:10.4172/1948-5948.1000180
- Mia, M. a.-T., Mosaib, M. G., Khalil, M. I., Islam, M. A., and Gan, S. H. (2020). Potentials and safety of date palm fruit against diabetes: A critical review. *Foods* 9, 1557. doi:10.3390/foods9111557
- Najm, O. A., Addnan, F. H., Mohd-Manzor, N. F., Elkadi, M. A., Abdullah, W. O., Ismail, A., et al. (2021). Identification of phytochemicals of phoenix dactylifera L. Cv Ajwa with UHPLC-ESI-QTOF-MS/MS. *Int. J. Fruit Sci.* 21, 848. doi:10.1080/15538362.2021.1939227
- Nematallah, K. A., Ayoub, N. A., Abdelsattar, E., Meselhy, M. R., Elmazar, M. M., El-Khatib, A. H., et al. (2018). Polyphenols LC-MS2 profile of Ajwa date fruit (Phoenix dactylifera L.) and their microemulsion: Potential impact on hepatic fibrosis. *J. Funct. foods* 49, 401. doi:10.1016/j.jff.2018.08.032
- Newman, D. J., and Cragg, G. M. (2016). Natural products as sources of new drugs from 1981 to 2014. *J. Nat. Prod.* 79, 629–661. doi:10.1021/acs.jnatprod.5b01055
- Otifi, A. M., El-Sayed, A. M., Michel, C. G., and Farag, M. A. (2019). Metabolites profiling of date palm (phoenix dactylifera L.) commercial by-products (pits and pollen) in relation to its antioxidant effect: A multiplex approach of MS and nmr metabolomics. *Metabolomics* 15, 119–217. doi:10.1007/s11306-019-1581-7
- Perveen, K., and Bokahri, N. A. (2020). Comparative analysis of chemical, mineral and *in-vitro* antibacterial activity of different varieties of date fruits from Saudi Arabia. *Saudi J. Biol. Sci.* 27, 1886–1891. doi:10.1016/j.sjbs.2019.11.029
- Ravi, L. (2017). Bioactivity of Phoenix dactylifera seed and its phytochemical analysis. *Int. J. Green Pharm. (IJGP)* 11. doi:10.22377/ijgp.v11i02.1037
- Rezaei, M., Khodaei, F., and Hooshmand, N. (2015). Date seed extract diminished apoptosis event in human colorectal carcinoma cell line. *MOJ Toxicol.* 1, 00017. doi:10.15406/mojt.2015.01.00017
- Rock, W., Rosenblat, M., Borochov-Neori, H., Volkova, N., Judeinstein, S., Elias, M., et al. (2009). Effects of date (phoenix dactylifera L., Medjool or Hallawi variety) consumption by healthy subjects on serum glucose and lipid levels and on serum oxidative status: A pilot study. *J. Agric. food Chem.* 57, 8010–8017. doi:10.1021/jf901559a
- 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, 1355–1363. doi:10.1107/s0907444904011679
- Shahbaz, K., Asif, J. A., Liszen, T., Nurul, A. A., and Alam, M. K. (2022). Cytotoxic and antioxidant effects of phoenix dactylifera L. (Ajwa date extract) on oral squamous cell carcinoma cell line. *BioMed Res. Int.* 2022, 1–16. doi:10.1155/2022/5792830
- Sharma, P., Joshi, T., Mathpal, S., Joshi, T., Pundir, H., Chandra, S., et al. (2022). Identification of natural inhibitors against Mpro of SARS-CoV-2 by molecular docking, molecular dynamics simulation, and MM/PBSA methods. *J. Biomol. Struct. Dyn.* 40, 2757–2768. doi:10.1080/07391102.2020.1842806
- Siddiqui, S., Ahmad, R., Khan, M. A., Upadhyay, S., Husain, I., and Srivastava, A. N. (2019). Cytostatic and anti-tumor potential of Ajwa date pulp against human hepatocellular carcinoma HepG2 cells. *Sci. Rep.* 9, 245. doi:10.1038/s41598-018-36475-0
- Souda, B., Rami, R., Jalloul, B., and Mohamed, D. (2020). Roasted date palm seeds (phoenix dactylifera) as an alternative coffee: Chemical composition and bioactive properties. *Biomass Convers. Biorefinery* 12, 3771. doi:10.1007/s13399-020-00896-7
- Taghizadeh, M. S., Niazi, A., Moghadam, A., and Afsharif, A. (2022). Experimental, molecular docking and molecular dynamic studies of natural products targeting overexpressed receptors in breast cancer. *Plos one* 17, e0267961. doi:10.1371/journal.pone.0267961
- Thouri, A., La Barbera, L., Canuti, L., Vegliante, R., Jelled, A., Flamini, G., et al. (2019). Antiproliferative and apoptosis-inducing effect of common Tunisian date seed (var. Korkobbi and Arechti) phytochemical-rich methanolic extract. *Environ. Sci. Pollut. Res.* 26, 36264–36273. doi:10.1007/s11356-019-06606-9
- 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, W363–W367. doi:10.1093/nar/gky473
- 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
- Vayalil, P. K. (2002). Antioxidant and antimutagenic properties of aqueous extract of date fruit (Phoenix dactylifera L. Arecaceae). *J. Agric. Food Chem.* 50, 610–617. doi:10.1021/jf010716t
- Vayalil, P. K. (2012). Date fruits (phoenix dactylifera linn): An emerging medicinal food. *Crit. Rev. food Sci. Nutr.* 52, 249–271. doi:10.1080/10408398.2010.499824
- Wang, Y., Xiao, J., Suzek, T. O., Zhang, J., Wang, J., and Bryant, S. H. (2009). PubChem: A public information system for analyzing bioactivities of small molecules. *Nucleic acids Res.* 37, W623–W633. doi:10.1093/nar/gkp456
- Zhang, C.-R., Aldosari, S. A., Vidyasagar, P. S., Nair, K. M., and Nair, M. G. (2013). Antioxidant and anti-inflammatory assays confirm bioactive compounds in Ajwa date fruit. *J. Agric. food Chem.* 61, 5834–5840. doi:10.1021/jf401371v
- Zhang, C.-R., Aldosari, S. A., Vidyasagar, P. S., Shukla, P., and Nair, M. G. (2017). Health-benefits of date fruits produced in Saudi Arabia based on *in vitro* antioxidant, anti-inflammatory and human tumor cell proliferation inhibitory assays. *J. Saudi Soc. Agric. Sci.* 16, 287–293. doi:10.1016/j.jssas.2015.09.004



OPEN ACCESS

EDITED BY

Khurshid Ahmad,
Yeungnam University, Republic of Korea

REVIEWED BY

Danishuddin Nan,
Independent researcher, Republic of
Korea
Xiaodong Ma,
Anhui University of Chinese Medicine,
China
C. George Priya Doss,
VIT University, India

*CORRESPONDENCE

Noor Ahmad Shaik,
✉ nshaik@kau.edu.sa
Nasreen Sultana,
✉ nasreensci@gmail.com
Babajan Banaganapalli,
✉ bbabajan@kau.edu.sa

[†]These authors have contributed
equally to this work

SPECIALTY SECTION

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

RECEIVED 04 January 2023

ACCEPTED 09 February 2023

PUBLISHED 10 March 2023

CITATION

Almukadi H, Jadkarim GA, Mohammed A,
Almansouri M, Sultana N, Shaik NA and
Banaganapalli B (2023), Combining
machine learning and structure-based
approaches to develop oncogene PIM
kinase inhibitors.
Front. Chem. 11:1137444.
doi: 10.3389/fchem.2023.1137444

COPYRIGHT

© 2023 Almukadi, Jadkarim, Mohammed,
Almansouri, Sultana, Shaik 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.

Combining machine learning and structure-based approaches to develop oncogene PIM kinase inhibitors

Haifa Almukadi^{1†}, Gada Ali Jadkarim², Arif Mohammed³,
Majid Almansouri⁴, Nasreen Sultana^{5*}, Noor Ahmad Shaik^{2,6*} and
Babajan Banaganapalli^{2,6*†}

¹Department of Pharmacology and Toxicology, Faculty of Pharmacy, King Abdulaziz University, Jeddah, Saudi Arabia, ²Department of Genetic Medicine, Faculty of Medicine, King Abdulaziz University, Jeddah, Saudi Arabia, ³Department of Biology, College of Science, University of Jeddah, Jeddah, Saudi Arabia, ⁴Department of Clinical Biochemistry, Faculty of Medicine, King Abdulaziz University, Jeddah, Saudi Arabia, ⁵Department of Biotechnology, Acharya Nagarjuna University, Guntur, India, ⁶Princess Al-Jawhara Al-Brahim Center of Excellence in Research of Hereditary Disorders, King Abdulaziz University, Jeddah, Saudi Arabia

Introduction: PIM kinases are targets for therapeutic intervention since they are associated with a number of malignancies by boosting cell survival and proliferation. Over the past years, the rate of new PIM inhibitors discovery has increased significantly, however, new generation of potent molecules with the right pharmacologic profiles were in demand that can probably lead to the development of Pim kinase inhibitors that are effective against human cancer.

Method: In the current study, a machine learning and structure based approaches were used to generate novel and effective chemical therapeutics for PIM-1 kinase. Four different machine learning methods, namely, support vector machine, random forest, k-nearest neighbour and XGBoost have been used for the development of models. Total, 54 Descriptors have been selected using the Boruta method.

Results: SVM, Random Forest and XGBoost shows better performance as compared to k-NN. An ensemble approach was implemented and, finally, four potential molecules (CHEMBL303779, CHEMBL690270, MHC07198, and CHEMBL748285) were found to be effective for the modulation of PIM-1 activity. Molecular docking and molecular dynamic simulation corroborated the potentiality of the selected molecules. The molecular dynamics (MD) simulation study indicated the stability between protein and ligands.

Discussion: Our findings suggest that the selected models are robust and can be potentially useful for facilitating the discovery against PIM kinase.

KEYWORDS

PIM kinase, classification models, virtual screening, molecular docking, cancer drug treatment

Introduction

Proto-oncogene PIM-1 kinase is a member of the serine/threonine protein kinase family (Narlik-Grassow et al., 2014). PIM kinases are involved in cancer cell survival, proliferation, and tumor growth and are overexpressed in a number of hematological malignancies, in addition to solid cancers such as pancreatic, prostate, and colon cancers (Amson et al., 1989;

Li et al., 2006; Nawijn et al., 2011). PIM-1, PIM-2, and PIM-3 are the three highly homologous genes that make up the PIM family. This kinase family is highly homologous with the kinase domains, especially in the linker region and the ATP-binding sites (Warfel and Kraft, 2015). These enzymes are constitutively expressed in tumors and are becoming more widely acknowledged as crucial survival signal mediators in malignancies, stress responses, and neurological development. PIM-1 kinase is a genuine oncogene that is the focus of drug development research initiatives since it has been linked to the emergence of leukemias, lymphomas, and prostate cancer (Li et al., 2011; Le et al., 2015; Huang et al., 2022). PIM kinases regulate the network of signaling pathways that are critical for tumorigenesis and development, making them attractive drug targets (Drygin et al., 2012; Tursynbay et al., 2016).

The crystal structure of PIM-1 has been published by numerous independent groups in both the presence and the absence of its inhibitors (Wang et al., 2013; Nonga et al., 2021). Structural research on PIM-1 has found a number of distinctive characteristics that set it apart from other kinases with known structures. The catalytic domain of PIM-1 kinase spans amino acid positions 38 to 290 and includes a conserved glycine loop motif at positions 45 to 50, phosphate-binding sites at positions 44 to 52 and 67, and a proton acceptor site at position 167. The hunt for small-molecule ATP-competitive inhibitors with the potential to develop into novel targeted oncology treatments has been sparked by the involvement of the PIM kinases in important cancer hallmarks. The majority of PIM-1 inhibitors have failed to evolve into a new anticancer medication despite having excellent biochemical potency, largely because they were found to have subpar pharmacological qualities (Dakin et al., 2012; Drygin et al., 2012; Ogawa et al., 2012; Vivek et al., 2017; Zhao et al., 2017; Park et al., 2021). Due to their therapeutic value in cancer, the discovery of PIM-1 inhibitors has increasingly attracted much attention in past few years. The rate of new PIM inhibitor discovery has increased significantly, and there has been demand for a new generation of potent molecules with the right pharmacologic profiles that can probably lead to the development of PIM kinase inhibitors that are effective against human cancer.

This work was undertaken to develop machine learning-based classification models to identify a new class of PIM-1 inhibitors. Under this approach, four different machine learning methods were applied to develop the classification models. These models were further used to screen chemical libraries to retrieve novel potent PIM-1 inhibitors. In addition, we also carried out molecular docking and molecular dynamics simulations to investigate the interaction and stability within the catalytic site of PIM-1 kinase. This multistage approach allows us to screen large chemical libraries efficiently and effectively in a reasonable time. Moreover, it can also help us identify novel chemical scaffolds for potent PIM-1 inhibitors.

Materials and methods

Data collection and model building

All chemical compounds with activity against PIM-1 were collected from the literature and the ChEMBL database (Gaulton

et al., 2012). Inorganic and duplicate compounds were removed from the list. Generally, compounds with $IC_{50} \leq 10 \mu M$ will likely be “active,” predicting a large number of active molecules. However, such a high fraction of active compounds cannot be expected from any experimental platform. Therefore, in order to make the most efficient use of costly experimental validation, the optimal model should identify compounds with affinity higher than $10 \mu M$. The higher the value, the higher the drug dose needed to achieve the required potency and, thus, the higher the chance of “off-target” activity. To address this issue, we chose to set the decision boundary at $IC_{50} \leq 1 \mu M$ for active molecules. Molecular descriptors were calculated using the PaDEL software (Yap, 2011). A two-tier selection procedure was applied to select the best descriptors. First, we randomly selected one descriptor from a pair showing >0.85 correlation. Second, descriptors were reduced using the Boruta method (Kursa et al., 2010). We used four different machine learning methods, namely, Support Vector Machine (SVM) (Mitchell, 1997), random forest (Breiman, 2001), Extreme Gradient Boosting (XGBoost) (Chen and Guestrin, 2016), and kappa nearest neighbor (kNN) (Voulgaris and Magoulas, 2008), to build the classification models. All the classification experiments and calculations were conducted using the R.3.0.2 environment (<http://www.R-project.org/>) and Python (<http://www.python.org/>) platform. The compounds used in training and test sets are given in Supplementary Tables S1 and S2, respectively.

Model validation

A receiver operating characteristic (ROC) plot and area under the curve (AUC) were used to assess the performance of the model (Hanley and McNeil, 1983; Park et al., 2004). In Table 1, the terms precision (Eq. 1), recall (Eq. 2), accuracy (Eq. 3), and F1 score (Eq. 4) are defined along with their relationships to the statistical performance calculations used to assess the quality of the model.

$$\text{Precision} = \frac{\text{True positive}}{\text{True positive} + \text{False Negative}}, \quad (1)$$

$$\text{Recall} = \frac{\text{True positive}}{\text{True positive} + \text{False Negative}}, \quad (2)$$

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}, \quad (3)$$

$$F1 = 1. \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (4)$$

Applicability domain

In order to highlight the region of the chemical space that contains the chemicals for which the model is expected to make accurate predictions, a well-validated predictive model needs to have a defined applicability domain (AD) (Rakhimbekova et al., 2020). Any predictive model must verify its constraints in terms of its structural domain and response space. As a result, determining a model's AD and evaluating the accuracy of its predictions are both challenging tasks. These QSAR models typically use the training set to cover a certain chemical space. The model's predictions are accurate if any query compound falls within this definition of AD. If not, the prediction might not conform to the

TABLE 1 Evaluation metrics for the test set.

Method	Descriptors	Precision	Recall	Accuracy (Q)	F1 score	AUC
XGBoost	All descriptors	0.82	0.81	0.83	0.97	0.89
	Boruta	0.81	0.79	0.85	0.80	0.88
	MACCS	0.80	0.76	0.81	0.77	0.92
Random forest	All descriptors	0.85	0.81	0.86	0.98	0.91
	Boruta	0.86	0.81	0.87	0.83	0.92
	MACCS	0.80	0.76	0.82	0.78	0.90
SVM	All descriptors	0.74	0.73	0.78	0.86	0.83
	Boruta	0.75	0.72	0.78	0.71	0.82
	MACCS	0.70	0.73	0.70	0.69	0.82
kNN	All descriptors	0.77	0.75	0.80	0.75	0.84
	Boruta	0.72	0.67	0.75	0.68	0.78
	MACCS	0.81	0.76	0.82	0.77	0.82

TABLE 2 Probability scores and docking scores of the selected compounds.

Compound ID	Classifier probability				Binding energy
	XGBoost	Random forest	SVM	kNN	
CHEMBL303779	0.82	0.74	0.84	0.78	−8.34
CHEMBL690270	0.76	0.70	0.92	0.85	−7.56
CHEMBL748285	0.72	0.74	0.68	0.63	−9.78
EBM-MPC	0.81	0.75	0.71	0.71	−8.45

model's presumptions. Principal component analysis (PCA) (Sushko et al., 2010) has been employed in our work to define the AD of the compounds used in this study.

$$Tc = \frac{C}{A + B - C} \quad (5)$$

Y-randomization

To test the robustness of the proposed models, y-randomization was applied. This technique involves randomly mixing up the values of the target variable in the training set (Rücker et al., 2007; Lipiński and Szurmak, 2017). The same parameters used in the initial model are then applied to a new prediction generated with the scrambled data. Every estimate of the model's accuracy was recorded. In total, 50% of the compounds in the training set were resampled and used in a 500-run y-randomization test.

Similarity calculations

The Tanimoto coefficient (Tc) (Eq. 5) was computed using MACCS-166 fingerprints to quantify chemical similarity. The active and inactive chemicals in the training set were compared against false and true positive compounds in systematic pairwise similarity computations.

Substructure analyses

Molecular substructures related to PIM activity were analyzed using the distribution of MACSS fingerprints in active and inactive compounds (Eq. 6).

$$Frequency = \frac{\sum_i^N FP(1|0)}{N} \times 100. \quad (6)$$

Analysis of probability scores

Additionally, the probability scores of the developed classification models were examined. In general, a molecule is defined as inactive if its probability score is lower than 0.5, while a compound with a probability score of 0.5 is considered active (Ponzoni et al., 2019). The more this score approaches 1, the more confident we are in our prediction. Here, we examined the probability score distributions for TP (true positive), TN

(true negative), FP (false positive), and FN (false negative) results.

Chemical database screening

The developed models were used to screen the hits against PIM-1. The NCI library and Maybridge databases were used for virtual screening. The National Cancer Institute maintains a repository of compounds that have been evaluated as potential anticancer agents. These compounds represent unique structural diversity based on synthetic and natural products. The Maybridge library consists of a highly diverse set of over 53,000 lead-like compounds. Maybridge Hit-to-Lead was designed for medicinal chemistry, allowing SAR development and hit-to-lead optimization. The following filters were used to select the hits: Filter 1: compounds predicted to be active by all the validated models; Filter 2: compounds having a probability score; and Filter 3: compounds falling within the chemical space of the training set. These compounds were further processed for molecular docking, followed by molecular dynamics simulations. Finally, compounds with the best affinity and conformance within the active site were selected and analyzed.

Molecular docking

Molecular docking was implemented to identify the best physical confirmation of inhibitor binding within the active site of PIM-1 kinase. The PIM kinase enzyme structure was taken from the Protein Data Bank (PDB ID: 5KZI). All of the docking simulations for this work were performed using AutoDock Vina (Trott and Olson, 2009) with a 1 Å spacing, default exhaustiveness, and full ligand flexibility. The grid resolution was internally set to 1 Å. We set the number of binding modes to 10 and exhaustiveness to 8. A cubical grid of size 60 × 60 × 60 Å with 0.375 Å spacing was used around the active sites of the protein. To acquire the structure in the PDBQT format, polar hydrogen atoms were added using AutoDock Tools 92.

Molecular dynamics simulations

Selected best compounds were further subjected to molecular dynamics (MD) simulations using Groningen Machine for Chemical Simulations (GROMACS v5.1.5) (Pronk et al., 2013). The parameters and coordinate files for PIM-1 kinase and selected potential hit compounds were generated using the CHARMM27 forcefield in GROMACS and PRODRG, respectively. The TIP3P water model was used for each simulation system, which was neutralized by the addition of Na⁺ ions in a dodecahedron periodic box. Energy minimization was performed for 50,000 nstep using the steepest descent algorithm to avoid steric clashes. Equilibration of each system was performed in two stages: the first phase was carried out with a constant number of particles, volume, and temperature (NVT) ensemble for 500 ps at 300 K, using the V-rescale thermostat (Bussi et al., 2007); and in the second phase, the pressure of each system was equilibrated for 500 ps at a constant number of particles, pressure, and temperature

(NPT) at 1 bar using a Parrinello–Rahman barostat (Parrinello and Rahman, 1981). Each equilibrated system was simulated for 30 ns under periodic boundary conditions to avoid edge effects. Electrostatic interactions were handled by the particle mesh Ewald (PME) method, while the heavy-atom bonds were restrained using the LINCS algorithm.

Results

Model development and evaluation

In total, 54 descriptors from the set of 240 were eventually selected using the Boruta method (Supplementary Table S3). All these descriptors belonged to 12 different classes. The descriptors include autocorrelation, information content, atom-type electrotopological state, Burden modified eigenvalues, molecular distance edge, carbon type, and molecular linear free energy relation. The models were trained using four machine learning methods (SVM, random forest, XGBoost, and kNN). Evaluation metrics for the developed models are given in Table 1, including accuracy, recall, precision, F1 Score (a measure of a model's accuracy, which takes into account both precision and recall), and Area Under the Curve (AUC) values. SVM, random forest, and XGBoost performed better than kNN according to these metrics in combination with the selected descriptor set. Among the three, random forest achieved the best accuracy, at 0.87 for the test set (with selected descriptors), as compared to SVM (0.78) and XGBoost (0.84). In addition, these models also had significant AUC values (Figure 1).

Applicability domain and y-randomization

An applicability domain (AD) analysis was performed to check the reliability of the generated classification models. Figure 2 shows a scatter plot of the PC1 and PC2 coordinates derived from the set of selected PIM-1 compound descriptors. The training and test compounds share similar PC1 and PC2 coordinates, suggesting that predictions were within the applicability domain (AD) of both the training and test sets. To check the robustness of the developed models, y-randomization tests were performed (Rücker et al., 2007). Y-randomization test accuracies were found to be lower, and none of the random trials achieved higher scores than our main models (Figure 3). The average accuracy across all randomly generated models was found to be less than 0.58. This confirms that the selected models are robust and reliable and were not generated by chance correlations. A pairwise comparison of the compounds in each cluster was found to reflect reasonable Tanimoto coefficient similarities between them.

Probability analyses

Probability scores of the selected models, reflecting the probability of belonging to each class, were also analyzed. It is known that a compound with a probability score of ≥0.5 is classified

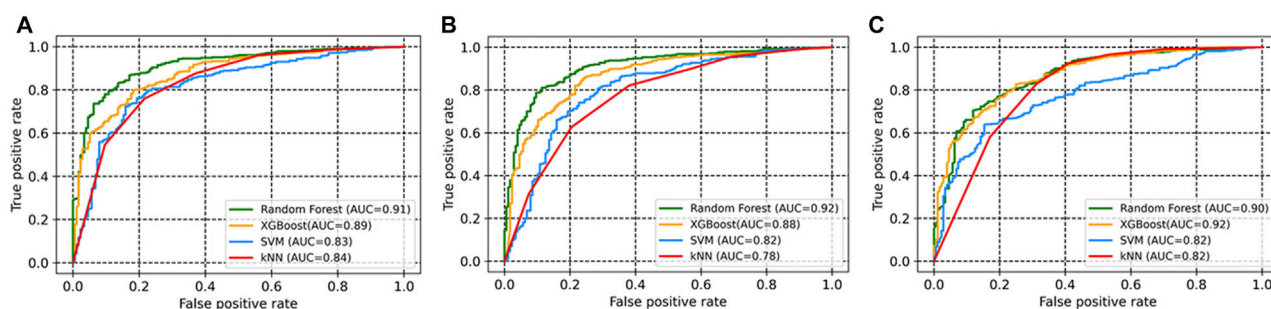


FIGURE 1

ROC curves of the models based on four machine learning approaches for (A) all descriptors; (B) selected descriptors (Boruta method); (C) MACCS fingerprints.

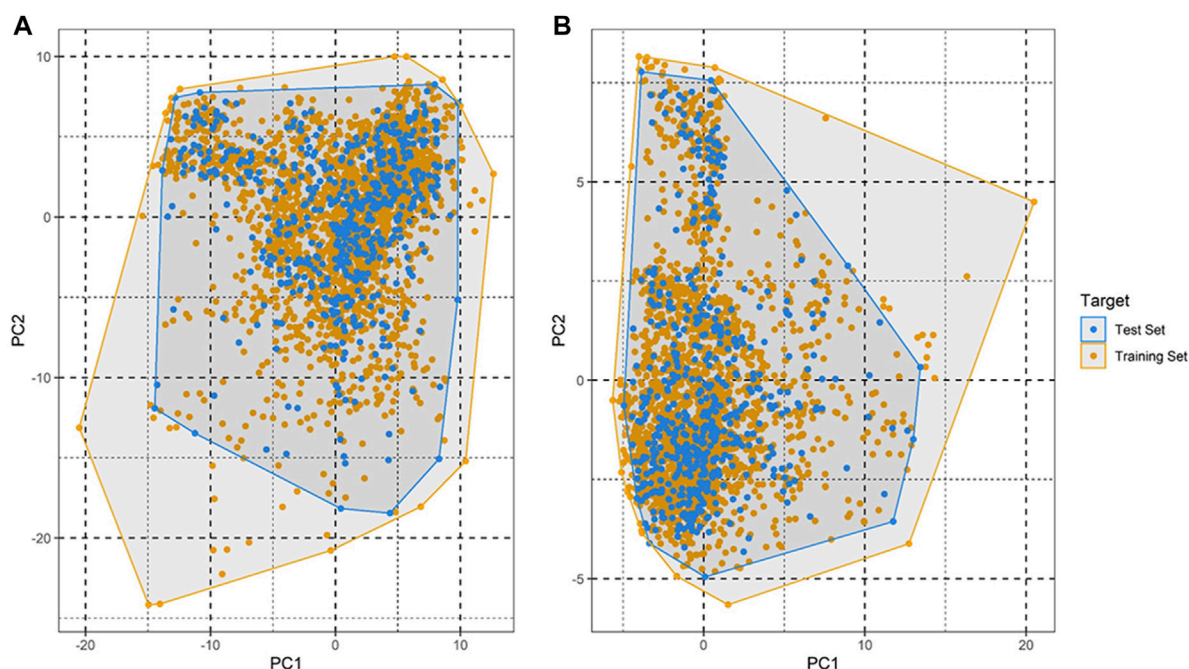


FIGURE 2

Applicability domain plot based on principal component analysis (PCA) for (A) training set and (B) test set.

as active, whereas a molecule with a probability below <0.5 is classified as inactive. As this score approaches 1, the higher the value, the higher the model's confidence in the prediction is (Minerali et al., 2020; Esposito et al., 2021). In our study, we analyzed the distribution of probability scores among TN (true negative), FP (false positive), TP (true positive), and FN (false negative) results. For the SVM model, compounds with a probability score of more than 0.80 (an average value) were more likely to be active, whereas compounds with a probability score of 0.36 were more likely to be inactive. In the case of the random forest model, a compound with a probability score of more than 0.87 was more likely to be active, whereas a compound with a probability score of 0.24 was more likely to be inactive. Random forest achieved

values of 0.95 and 0.11 for active and inactive compounds, respectively, indicating greater success in predicting compound activity with the desired probability score (Supplementary Figure S1). False positive compounds were predicted with probability scores of 0.63, 0.65, and 0.69 for the random forest, XGBoost, and SVM models, respectively. In contrast, false negative compounds were found to have probability scores of 0.31, 0.42, and 0.14 for the random forest, SVM, and XGBoost models, respectively. Each predictive model's effectiveness in the early recognition of hits was visually evaluated using a cumulative gain plot (Table 2). The cumulative gain curve is an evaluation curve that evaluates the model's performance and contrasts the outcomes with a random selection. It displays the percentage of targets identified

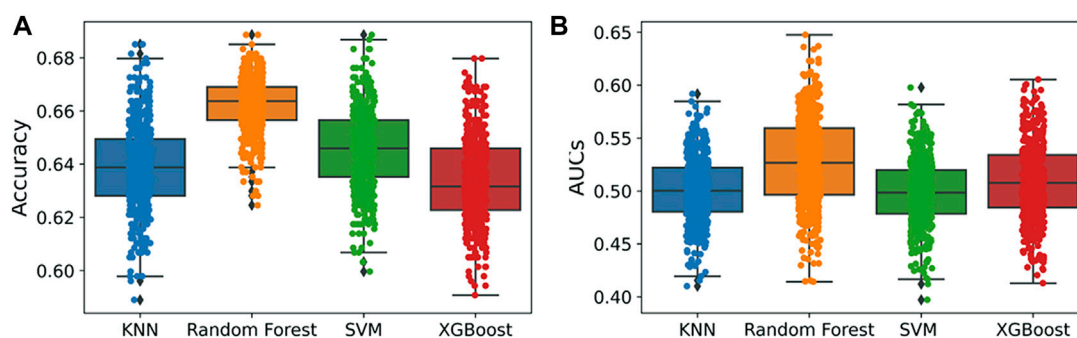


FIGURE 3
Y-randomization models. (A) Accuracy; (B) AUC values. A total of 500 y-randomization runs were performed.

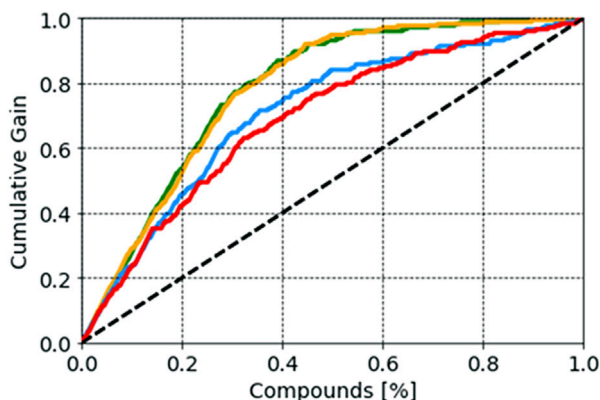


FIGURE 4
Probabilistic distribution plot showing cumulative gain for the developed models.

when taking into account a particular portion of the population that has the highest likelihood of being a target based on the model. The comparison showed that the XGBoost and random forest methods performed better than SVM and kNN in terms of early recognition of hits (Figure 4).

MACCS fingerprint analyses

Molecular substructures related to the PIM-1 activity of the compounds can be identified by analyzing the bits in the MACCS fingerprints. We analyzed the MACCS fingerprints showing a reasonable difference between active and inactive compounds (Supplementary Table S4). The occurrence of MACCS fingerprints differed significantly between active and inactive compounds in the training dataset, suggesting that the substructures represented by these features may be closely related to PIM-1 activity. Descriptions and the number of occurrences of these substructures are listed in Supplementary Table S4. It was found that MACCS38, MACCS52, MACCS92, MACCS98, MACCS107, MACCSFP142, etc. are prevalent in active molecules. This is consistent with previous studies, which shows that

compounds with such functional groups have therapeutic potential against PIM kinase (Tsuganezawa et al., 2012; El-Hawary et al., 2018; Park et al., 2021).

Database screening and molecular interaction analyses

The NCI and Maybridge databases were used to screen the potential hits from validated models. Commonly predicted active compounds with high probability scores were selected and further filtered out within the applicability domain (AD) of the training set. These compounds were further subjected to molecular docking simulation (Table 2). Finally, four compounds (CHEMBL303779, CHEMBL690270, CHEMBL748285, and N-[(1-ethylbenzimidazol-2-yl)methyl]-3-(4-methoxyphenyl)-1H-pyrazole-4-carboxamide (EBM-MPC)) were observed to have reasonable binding affinity and stable interaction with the catalytic residues in the active site (Table 3 and Figure 5). A literature survey revealed that Leu44, Lys67, Glu121, and Asp186 are crucial for the interaction of inhibitors (Tsuganezawa et al., 2012; El-Hawary et al., 2018; Park et al., 2021). It can be observed in Figure 4 that CHEMBL690270, CHEMBL303779, and EBM-MPC form hydrogen bond interactions with Lys67 and hydrophobic interactions with Asp186 (Figure 6). In contrast, CHEMBL748285 forms hydrogen bonds with Asp186 (Figure 6). The quinazoline ring of compounds was involved in multiple π -alkyl interactions. In addition, a number of hydrophobic contacts, particularly residues Leu44, Gly47, Phe49, Ile104, and Leu120, stabilize interaction with hits. PIM inhibitors fall into two broad categories: ATP mimetics, which form hydrogen bonds with the glutamate residue that serves as the hinge (Glu121), and non-ATP mimetics, which bind far from the hinge or interact with the hinge through hydrophobic interactions with a number of residues in the specific hydrophobic pocket that serves as the hinge environment (El-Hawary et al., 2018; Park et al., 2021). The Tanimoto coefficient (Tc) similarity score of these selected hits was found to be ≤ 0.5 with high-activity compounds (Figure 5B).

MD simulation analyses

By analyzing 100-ns MD trajectories, the structural changes to PIM-1 upon inhibitor binding were studied. We examined the RMSD

TABLE 3 Binding mode analysis of the four selected inhibitors.

Compound	Hydrogen bonding	Hydrophobic interaction	H-bond range (Å)	Hydrophobic interaction range (Å)
CHEMBL303779	Lys67 and Arg122	Gly45, Gly47, Gly48, Phe49, Ala65, Lys67, Ile104, Leu120, Glu121, Arg122, Pro123, Val126, and Leu174	2.7–3.2	3.3–4.9
CHEMBL690270	Lys67 and Asp186	Leu44, Gly45, Phe49, Lys67, Ile104, Val126, Asp128, Glu171, Asn172, Leu174, and Asp186	2.4–2.6	3.3–4.7
EBM-MPC	Lys67 and Glu121	Gly47, Val52, Lys67, Ile104, Leu120, Glu121, Pro123, Val126, Leu174, and Asp186	2.8–3.0	3.6–4.89
CHEMBL748285	Asn172 and Asp186	Leu44, Val52, Phe49, Asn172, Leu174, Leu182, Leu184, and Asp186	1.6–3.1	3.6–4.4

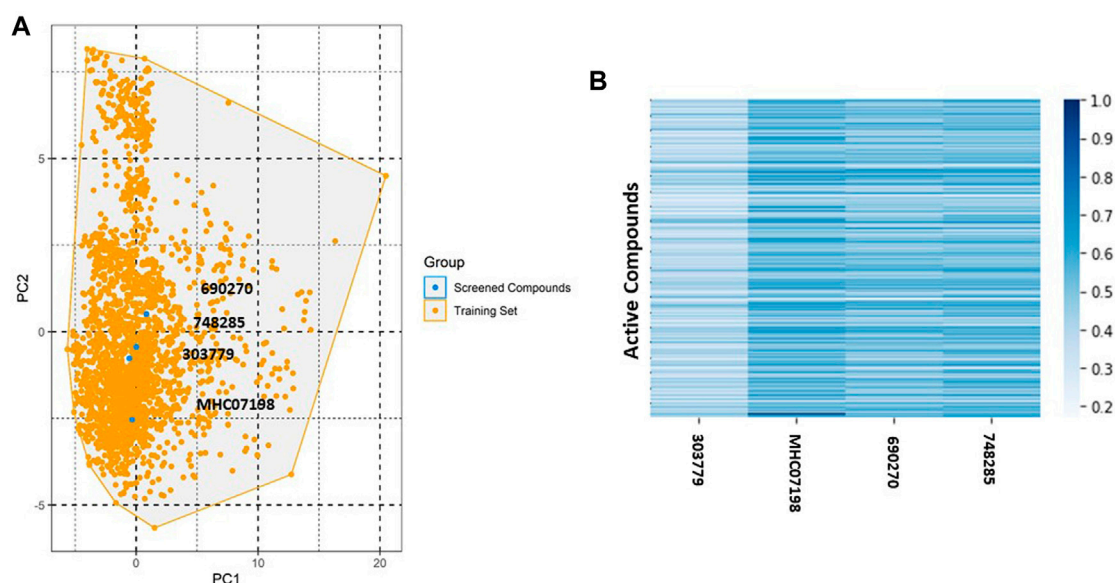


FIGURE 5

Chemical space and similarity analyses for selected compounds. (A) Chemical space of selected compounds; (B) heat map of the distance matrix for the selected compounds and active compounds in the training set.

of the protein backbone and the RMSF of the protein's alpha-carbon atoms. As shown in Figure 6, all the systems exhibited stability throughout the 100-ns simulation. The average RMSD value for all four systems was observed to be below 0.31 nm, which indicated that simulated complexes displayed RMSD values below the threshold. The average RMSD values further showed that the CHEMBL690270 PIM-1 complex displayed less deviation (0.26 nm), whereas CHEMBL303779 and CHEMBL748285 demonstrated similar average values of 0.34 nm (Figure 7A). RMSF is a significant value, used to characterize each residue's fluctuation rate upon ligand binding. It was observed that the inhibitor binding residues (Leu44, Phe49, Lys67, Glu121, and Asp186) did not fluctuate significantly (Figure 7B).

Discussion

This study was designed with the aim of building a classification model to predict potential hits for PIM-1 kinase. Four different machine learning approaches were used to build the models. Our

proposed models performed well in terms of accuracy, F1 score, precision, and recall. We used the area under the receiver operating characteristic curve approach to compare classifiers. The ROC curve is a graphical representation that contrasts a classifier's true positive rate and false positive rate at various threshold levels. The area under this curve, or AUC, is thus a useful metric for assessing machine learning algorithms, since it shows the degree of separability (Parrinello and Rahman, 1981). A ROC curve with a higher AUC value implies greater sensitivity in identifying active molecules and specificity in rejecting inactive compounds (Figure 1). In addition, our study also distinguished and ranked the top 18 variables, including 2D autocorrelation, Burden modified eigenvalues, and topological charge. These descriptors have the capacity to distinguish between active and inactive compounds.

QSAR Classification models must undergo an extensive validation process, and the reliability of those models must be objectively determined. The OECD guidelines state that a model must have a clearly defined domain of applicability (Dwyer et al., 2013). Additionally, the dataset for such models with a defined AD

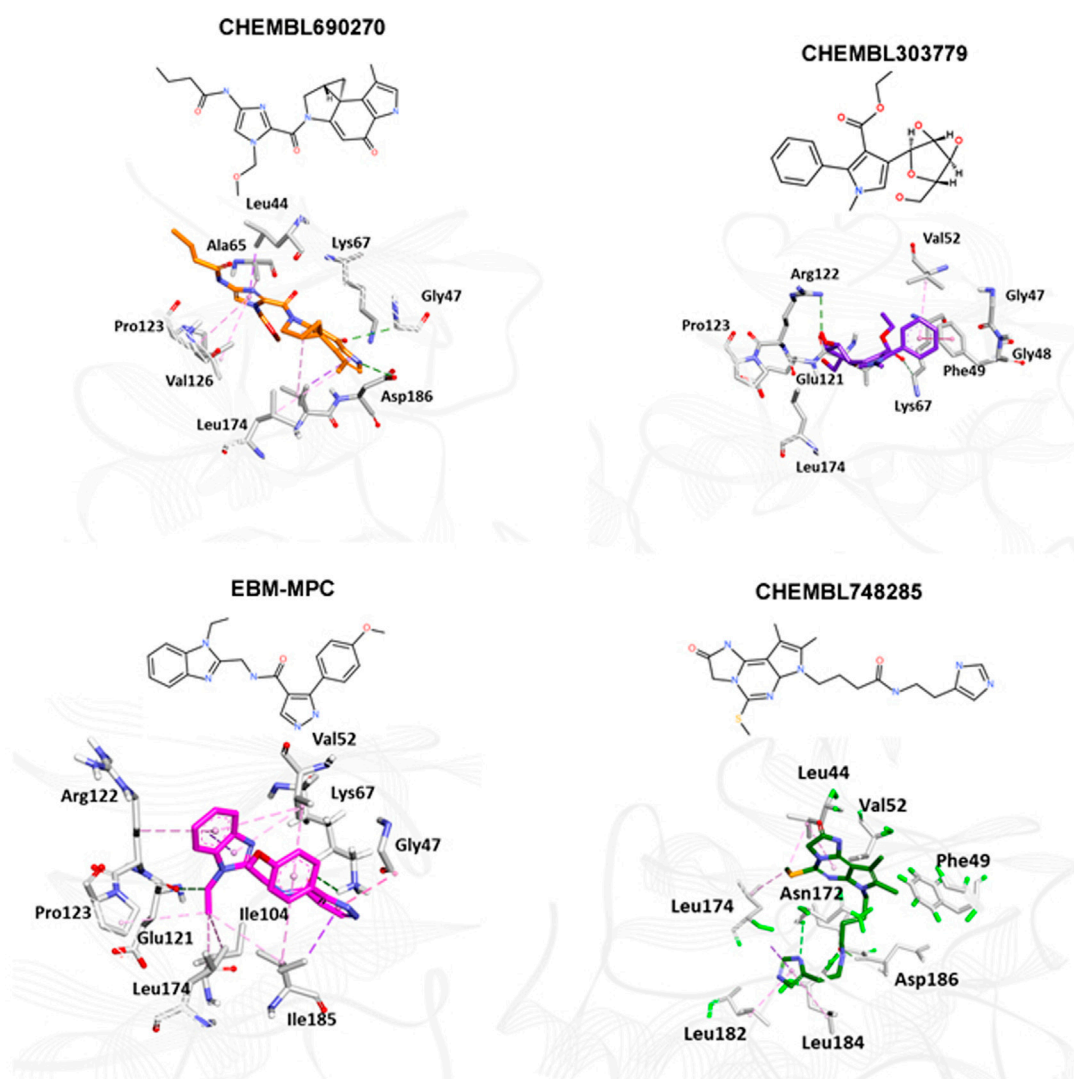


FIGURE 6

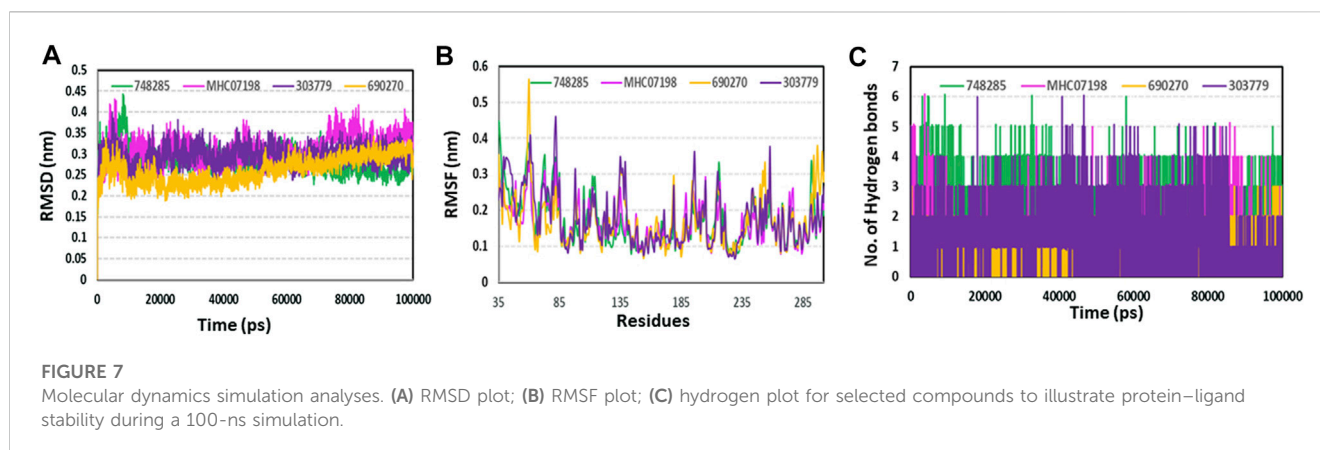
Binding mode analyses of selected compounds within the active site of PIM-1 kinase. Active site residues are shown as gray sticks; the protein backbone is shown as a light gray wire; hydrogen bonds are shown with a green dashed line.

should cover a broad chemical space and a diverse range of structural types. The AD of PIM kinase inhibitors has been defined using a principal component analysis-based approach for model development. A sufficient level of assurance in the produced models can be seen in the 2D plot obtained from the first two PCs, which represents the training and test set compounds, illustrating their structural variety and similar chemical space (Figure 2). To assess the likelihood of a random correlation for a chosen descriptor, y-randomization was utilized. This technique is used to assess the reliability or robustness of QSAR models and is recognized as one of the most effective validation processes (Rücker et al., 2007). By comparing a developed model's performance to the average measure of 500 random models, which are obtained by using the same parameters as those used to construct the original model along with a randomly scrambled target variable class, the statistical significance of the developed model can be examined. The results of the y-randomization tests demonstrated that the models created

for this study did not exhibit these connections by chance and that a true structure–activity relationship existed (Figure 3).

Fingerprints describe the molecular makeup of a compound. The description of each molecule is given as a string of binary substructures called a fingerprint. The corresponding fingerprint bit is set to 1 if the specified substructure is present in the given molecule; otherwise, it is set to 0. In our study, we used MACCS fingerprints to represent the presence of structures and their representative substructures in active and inactive compounds. These molecules contained MACCS65, MACCS128, and MACCS90. Compounds having such substructures were found to exhibit reasonable levels of activity toward PIM-1 kinase (Akué-Gédu et al., 2010; Dwyer et al., 2013; Hu et al., 2015; Wurz et al., 2015; Li et al., 2016).

To identify potent PIM-1 inhibitors, virtual screening of the NCI and Maybridge databases was performed using the validated models. To gain structural insight relevant to the inhibitory activities of the newly identified inhibitors, their binding modes in the binding site of



PIM-1 were examined. Figure 6 shows the most stable binding configurations of selected four compounds derived via docking simulations with potent inhibitors. These compounds appear to be accommodated in a similar way in the binding site of PIM1 (Xia et al., 2009; Abdelaziz et al., 2018; Ibrahim et al., 2022). The necessity of the interactions with the hinge region and Gly-loop residues (Qian et al., 2005; Pogacic et al., 2007; Tsuganezawa et al., 2012; Casuscelli et al., 2013; Fan et al., 2016; Abdelaziz et al., 2018; Bima et al., 2022; Ibrahim et al., 2022; Shaik et al., 2022) for tight binding to PIM-1 was also implicated with potent inhibitors (Xia et al., 2009; Ibrahim et al., 2022). Moreover, these four compounds can also interact with the activation loop including the Asp186 residue. A hydrophobic cavity is formed among the Ala65, Ile104, Phe187, Val52, Lys67, and Leu120 residues, and this maintains molecular stability through various hydrophobic forces. Similar interactions have also been noted in earlier published investigations, highlighting the significance of these amino acids for the assembly of PIM-1 inhibitor complexes (Tsuganezawa et al., 2012; El-Hawary et al., 2018; Park et al., 2021). Residue Lys67 is known to be significant in stabilizing the interaction with the compound and to play an important role in the catalytic activity of PIM-1 (Pogacic et al., 2007; Fan et al., 2016). In our study, we found that all four compounds interacted with Lys67, either with hydrogen bonds or through hydrophobic contact. Compared to the currently available PIM-1 inhibitors, the four selected compounds exhibit low Tanimoto coefficient (T_c) similarities, highlighting their structural novelty and druggability. Moreover, all these compounds were found to have a similar chemical boundary (Figure 5). Therefore, models constructed using these selected descriptors have good interpretability and reliability.

Molecular docking studies were conducted to analyze the binding mode of inhibitors at the PIM-1 catalytic domain. Notably, these inhibitors are positioned in the active site, between the residues Leu44, Gly45, Phe49, Lys67, Ile104, Lys67, Leu172, Leu174, and Asp186 (Table 3). These inhibitors were found to have stabilized the complex with hydrogen and hydrophobic interactions with residues, namely, Lys67 and Asp186. This is consistent with earlier research that revealed that these amino acid residues were essential for the catalytic activity of PIM-1 kinase (Qian et al., 2005; Banaganapalli et al., 2016; Shaik et al., 2021; Bima et al., 2022; Shaik et al., 2022).

Although molecular docking has strong computational capabilities, its predictions of the shape of the protein–ligand binding are frequently inaccurate. Thus, in this study, we

performed 100-ns MD simulations to test the stability of the chosen compounds in the PIM-1 binding pocket. It was determined that selected compounds remained stable in the binding pocket, as analyzed through the RMSD, RMSF, and hydrogen bonds. Most notably, stable hydrogen bonds with the residues Lys67 and Asp186 were observed in the complexes with the compounds (namely, ChEMBL748285, and ChEMBL690270).

Conclusion

The PIM kinase family has become a focus of attention in drug discovery. In particular, the search for inhibitors simultaneously targeting PIM-1 isoforms is of great interest because it opens new horizons toward the discovery of new chemicals capable of therapeutically modulating many biochemical pathways involved in the emergence and development of various cancers. In the present study, ensemble learning based on four different machine learning approaches, together with molecular docking and molecular dynamics simulation, was successfully utilized to identify novel scaffold inhibitors against PIM kinase. By combining machine learning and structure-based approaches, it was possible to evaluate the quantitative contributions of the molecules to the activity. This permitted the guided design of four new molecules, predicted to be potential PIM-1 inhibitors. The molecular docking analyses showed that the active inhibitors were able to interact with the amino acids (Lys67, Asp186, Leu44, Glu171, etc.) crucial for catalytic activity of PIM kinase. The interactions were found to be stable, as investigated through 100-ns molecular dynamics simulation.

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

HA, NS, and BB: conceptualization; HA, GJ, and BB: data curation; BB and GJ: formal analysis; HA: funding acquisition;

HA, GJ, and BB: methodology; HA: project administration; BB, NS, and HA: resources; BB: software; NS: supervision; HA and NS: validation; BB: visualization; HA, BB, AM, MA, NS, and GJ: writing—original draft and review.

Funding

This project was funded by the Deanship of Scientific Research (DSR) at King Abdulaziz University, under Grant no. G:207-249-1441. The authors therefore acknowledge the DSR for technical and financial support.

References

- Abdelaziz, M. E., M El-Miligy, M., Fahmy, S. M., Mahran, M. A., and A Hazzaa, A. (2018). Design, synthesis and docking study of pyridine and thieno[2, 3-b] pyridine derivatives as anticancer PIM-1 kinase inhibitors. *Bioorg Chem.* 80, 674–692. doi:10.1016/j.bioorg.2018.07.024
- Akué-Gédu, R., Nauton, L., Théry, V., Bain, J., Cohen, P., Anizon, F., et al. (2010). Synthesis, Pim kinase inhibitory potencies and *in vitro* antiproliferative activities of diversely substituted pyrrolo[2, 3-a]carbazoles. *Bioorg Med. Chem.* 18, 6865–6873. doi:10.1016/j.bmc.2010.07.036
- Amson, R., Sigaux, F., Przedborski, S., Flandrin, G., Givol, D., and Telerman, A. (1989). The human protooncogene product p33pim is expressed during fetal hematopoiesis and in diverse leukemias. *Proc. Natl. Acad. Sci. U. S. A.* 86, 8857–8861.
- Bananapalli, B., Mohammed, K., Khan, I. A., Al-Aama, J. Y., Elango, R., and Shaik, N. A. (2016). A computational protein phenotype prediction approach to analyze the deleterious mutations of human MED12 gene. *J. Cell Biochem.* 117, 2023–2035. doi:10.1002/jcb.25499
- 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
- Breiman, L. (2001). Random forests. *Mach. Learn.* 45, 5–32. doi:10.1023/A:1010933404324
- Bussi, G., Donadio, D., and Parrinell, M. (2007). Canonical sampling through velocity rescaling. *J. Chem. Phys.* 126, 014101. doi:10.1063/1.2408420
- Casuscelli, F., Ardini, E., Avanzi, N., Casale, E., Cervi, G., D'Anello, M., et al. (2013). Discovery and optimization of pyrrolo[1, 2-a]pyrazinones leads to novel and selective inhibitors of PIM kinases. *Bioorg Med. Chem.* 21, 7364–7380. doi:10.1016/j.bmc.2013.09.054
- Chen, T., and Guestrin, C. (2016). “XGBoost: A scalable tree boosting system,” in Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, San Francisco, March, 2016.
- Dakin, L. A., Block, M. H., Chen, H., Code, E., Dowling, J. E., Feng, X., et al. (2012). Discovery of novel benzylidene-1, 3-thiazolidine-2, 4-diones as potent and selective inhibitors of the PIM-1, PIM-2, and PIM-3 protein kinases. *Bioorg Med. Chem. Lett.* 22, 4599–4604. doi:10.1016/j.bmcl.2012.05.098
- Drygin, D., Haddach, M., Pierre, F., and Ryckman, D. M. (2012). Potential use of selective and nonselective pim kinase inhibitors for cancer therapy. *J. Med. Chem.* 55, 8199–8208. doi:10.1021/jm3009234
- Dwyer, M. P., Keertika, K., Paruch, K., Alvarez, C., Labroli, M., Poker, C., et al. (2013). Discovery of pyrazolo[1, 5-a]pyrimidine-based pim inhibitors: A template-based approach. *Bioorg Med. Chem. Lett.* 23, 6178–6182. doi:10.1016/j.bmcl.2013.08.110
- El-Hawary, S. S., Sayed, A. M., Mohammed, R., Khanfar, M. A., Rateb, M. E., Mohammed, T. A., et al. (2018). New pim-1 kinase inhibitor from the Co-culture of two sponge-associated actinomycetes. *Front. Chem.* 6, 538. doi:10.3389/fchem.2018.00538
- Esposito, C., Landrum, G. A., Schneider, N., Stieff, N., and Riniker, S. (2021). GHOST: Adjusting the decision threshold to handle imbalanced data in machine learning. *J. Chem. Inf. Model* 61, 2623–2640. doi:10.1021/acs.jcim.1c00160
- Fan, Y. B., Li, K., Huang, M., Cao, Y., Li, Y., Jin, S. Y., et al. (2016). Design and synthesis of substituted pyrrolo[3, 2-d]-1, 2, 3-triazines as potential Pim-1 inhibitors. *Bioorg Med. Chem. Lett.* 26, 1224–1228. doi:10.1016/j.bmcl.2016.01.032
- Gaulton, A., Bellis, L. J., Bento, A. P., Chambers, J., Davies, M., Hersey, A., et al. (2012). ChEMBL: A large-scale bioactivity database for drug discovery. *Nucleic Acids Res.* 40, D1100–D1107. doi:10.1093/nar/gkr777
- Hanley, J. A., and McNeil, B. J. (1983). A method of comparing the areas under receiver operating characteristic curves derived from the same cases. *Radiology* 148, 839–843. doi:10.1148/radiology.148.3.6878708
- Hu, H., Wang, X., Chan, G. K., Chang, J. H., Do, S., Drummond, J., et al. (2015). Discovery of 3, 5-substituted 6-azaindazoles as potent pan-Pim inhibitors. *Bioorg Med. Chem. Lett.* 25, 5258–5264. doi:10.1016/j.bmcl.2015.09.052
- Huang, J., Yuan, Y., Zhu, X., Li, G., Xu, Y., Chen, W., et al. (2022). Identification of pim-1 kinase inhibitors by pharmacophore model, molecular docking-based virtual screening, and biological evaluation. *Curr. Comput. Aided. Drug. Des.* 18, 240–246. doi:10.2174/1573409918666220427120524
- Ibrahim, M. H., Harras, M. F., Mostafa, S. K., Mohyeldin, S. M., Kamaly, O., Altwaijry, N., et al. (2022). Development of novel cyanopyridines as PIM-1 kinase inhibitors with potent anti-prostate cancer activity: Synthesis, biological evaluation, nanoparticles formulation and molecular dynamics simulation. *Bioorg Chem.* 129, 106122. doi:10.1016/j.bioorg.2022.106122
- Kursa, M. B., Jankowski, A., and Rudnicki, W. R. (2010). Boruta—a system for feature selection. *Fundam. Inf.* 101, 271–285. doi:10.3233/fi-2010-288
- Le, B. T., Kumarasiri, M., Adams, J. R., Yu, M., Milne, R., Sykes, M. J., et al. (2015). Targeting pim kinases for cancer treatment: Opportunities and challenges. *Future. Med. Chem.* 7, 35–53. doi:10.4155/fmc.14.145
- Li, J., Loveland, B. E., and Xing, P. X. (2011). Anti-Pim-1 mAb inhibits activation and proliferation of T lymphocytes and prolongs mouse skin allograft survival. *Cell Immunol.* 272, 87–93. doi:10.1016/j.cellimm.2011.09.002
- Li, K., Li, Y., Zhou, D., Fan, Y., Guo, H., Ma, T., et al. (2016). Synthesis and biological evaluation of quinoline derivatives as potential anti-prostate cancer agents and Pim-1 kinase inhibitors. *Bioorg Med. Chem.* 24 (8), 1889–1897. doi:10.1016/j.bmc.2016.03.016
- Li, Y. Y., Popivanova, B. K., Nagai, Y., Ishikura, H., Fujii, C., and Mukaida, N. (2006). Pim-3 a proto-oncogene with serine/threonine kinase activity, is aberrantly expressed in human pancreatic cancer and phosphorylates bad to block bad-mediated apoptosis in human pancreatic cancer cell lines. *Cancer Res.* 66, 6741–6747. doi:10.1158/0008-5472.can-05-4272
- Lipiński, P. F. J., and Szurmak, P. (2017). SCRAMBLE'N'GAMBLE: A tool for fast and facile generation of random data for statistical evaluation of QSAR models. *Chem. Pap.* 71, 2217–2232. doi:10.1007/s11696-017-0215-7
- Minerali, E., Foil, D. H., Zorn, K. M., Lane, T. R., and Ekins, S. (2020). Comparing machine learning algorithms for predicting drug-induced liver injury (DILI). *Mol. Pharm.* 17, 2628–2637. doi:10.1021/acs.molpharmaceut.0c00326
- Mitchell, T. (1997). *Machine learning*. New York: McGraw-Hill.
- Narlik-Grassow, M., Blanco-Aparicio, C., and Carnero, A. (2014). The PIM family of serine/threonine kinases in cancer. *Med. Res. Rev.* 34, 136–159.
- Nawijn, M., Alendar, A., and Berns, A. (2011). For better or for worse: The role of pim oncogenes in tumorigenesis. *Nat. Rev. Cancer* 11 (11), 23–34. doi:10.1038/nrc2986
- Nonga, O. E., Lavogina, D., Enkvist, E., Kestav, K., Chaikuad, A., Dixon-Clarke, S. E., et al. (2021). Crystal structure-guided design of bisubstrate inhibitors and photoluminescent probes for protein kinases of the PIM family. *Molecules* 26, 4353. doi:10.3390/molecules26144353
- Ogawa, N., Yuki, H., and Tanaka, A. (2012). Insights from Pim1 structure for anti-cancer drug design. *Expert Opin. Drug Discov.* 7, 1177–1192. doi:10.1517/17460441.2012.727394
- Park, H., Jeon, J., Kim, K., Choi, S., and Hong, S. (2021). Structure-based virtual screening and de novo design of PIM1 inhibitors with anticancer activity from natural products. *Pharm. (Basel)* 14, 275. doi:10.3390/ph14030275

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.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fchem.2023.1137444/full#supplementary-material>

- Park, S. H., Goo, J. M., and Jo, C. H. (2004). Receiver operating characteristic (ROC) curve: Practical review for radiologists. *Korean J. Radiol.* 5, 11–18. doi:10.3348/kjr.2004.5.1.11
- Parrinello, M., and Rahman, A. (1981). Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* 52, 7182–7190. doi:10.1063/1.328693
- Pogacic, V., Bullock, A. N., Fedorov, O., Filippakopoulos, P., Gasser, C., Biondi, A., et al. (2007). Structural analysis identifies imidazo[1, 2-b]pyridazines as PIM kinase inhibitors with *in vitro* antileukemic activity. *Cancer Res.* 67, 6916–6924. doi:10.1158/0008-5472.can-07-0320
- Ponzoni, I., Sebastián-Pérez, V., Martínez, M. J., Roca, C., Cruz Pérez, C. D., Cravero, F., et al. (2019). QSAR classification models for predicting the activity of inhibitors of beta-secretase (BACE1) associated with alzheimer's disease. *Sci. Rep.* 9, 9102. doi:10.1038/s41598-019-45522-3
- Pronk, S., Páll, S., Schulz, R., Larsson, P., Bjelkmar, P., Apostolov, R., et al. (2013). Gromacs 4.5: A high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* 29, 845–854. doi:10.1093/bioinformatics/btt055
- Qian, K. C., Wang, L., Hickey, E. R., Studts, J., Barringer, K., Peng, C., et al. (2005). Structural basis of constitutive activity and a unique nucleotide binding mode of human Pim-1 kinase. *J. Biol. Chem.* 280, 6130–6137. doi:10.1074/jbc.m409123200
- Rakhimbekova, A., Madzhidov, T. I., Nugmanov, R. I., Gimadiev, T. R., Baskin, I. I., and Varnek, A. (2020). Comprehensive analysis of applicability domains of QSPR models for chemical reactions. *Int. J. Mol. Sci.* 21, 5542. doi:10.3390/ijms21155542
- Rücker, C., Rücker, G., and Meringer, M. (2007). γ -Randomization and its variants in QSPR/QSAR. *J. Chem. Inf. Model* 47, 2345–2357. doi:10.1021/ci700157b
- 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
- Shaik, N. A., Saud Al-Saud, N. B., Abdulhamid Aljuhani, T., 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
- Sushko, I., Novotarskyi, S., Körner, R., Pandey, A. K., Cherkasov, A., Li, J., et al. (2010). Applicability domains for classification problems: Benchmarking of distance to models for ames mutagenicity set. *J. Chem. Inf. Model.* 50, 2094–2111. doi:10.1021/ci100253r
- Trott, O., and Olson, A. J. (2009). AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* 31, 455–461. doi:10.1002/jcc.21334
- Tsuganezawa, K., Watanabe, H., Parker, L., Yuki, H., Taruya, S., Nakagawa, Y., et al. (2012). A novel Pim-1 kinase inhibitor targeting residues that bind the substrate peptide. *J. Mol. Biol.* 417, 240–252. doi:10.1016/j.jmb.2012.01.036
- Tursynbay, Y., Zhang, J., Li, Z., Tokay, T., Zhumadilov, Z., Wu, D., et al. (2016). Pim-1 kinase as cancer drug target: An update. *Biomed. Rep.* 4, 140–146. doi:10.3892/br.2015.561
- Vivek, A., Bharti, S. K., and Kumar Budhwani, A. (2017). 3D-QSAR and virtual screening studies of thiazolidine-2, 4-dione analogs: Validation of experimental inhibitory potencies towards PIM-1 kinase. *J. Mol. Struct.* 1133, 278–293. doi:10.1016/j.molstruc.2016.12.006
- Voulgaris, Z., and Magoulas, G. D. (2008). “Extensions of the k nearest neighbour methods for classification problems,” in Proceedings of the 26th IASTED International Conference on Artificial Intelligence and Applications (AIA '08), Anaheim, CA, February, 2008 (ACTA Press), 23–28.
- Wang, X., Magnuson, S., Pastor, R., Fan, E., Hu, H., Tsui, V., et al. (2013). Discovery of novel pyrazolo[1, 5-a]pyrimidines as potent pan-Pim inhibitors by structure- and property-based drug design. *Bioorg Med. Chem. Lett.* 23, 3149–3153. doi:10.1016/j.bmcl.2013.04.020
- Warfel, N. A., and Kraft, A. S. (2015). PIM kinase (and Akt) biology and signaling in tumors. *Pharmacol. Ther.* 151, 41–49. doi:10.1016/j.pharmthera.2015.03.001
- Wurz, R. P., Pettus, L. H., Jackson, C., Wu, B., Wang, H. L., Herberich, B., et al. (2015). The discovery and optimization of aminooxadiazoles as potent Pim kinase inhibitors. *Bioorg Med. Chem. Lett.* 25 (4), 847–855. doi:10.1016/j.bmcl.2014.12.067
- Xia, Z., Knaak, C., Ma, J., Beharry, Z. M., McInnes, C., Wang, W., et al. (2009). Synthesis and evaluation of novel inhibitors of Pim-1 and Pim-2 protein kinases. *J. Med. Chem.* 52, 74–86. doi:10.1021/jm800937p
- Yap, C. W. (2011). PaDEL-descriptor: An open source software to calculate molecular descriptors and fingerprints. *J. Comput. Chem.* 32, 1466–1474. doi:10.1002/jcc.21707
- Zhao, W., Qiu, R., Li, P., and Yang, J. (2017). PIM1: A promising target in patients with triple-negative breast cancer. *Med. Oncol.* 34, 142. doi:10.1007/s12032-017-0998-y



OPEN ACCESS

EDITED BY

Khurshid Ahmad,
Yeungnam University, Republic of Korea

REVIEWED BY

Shakir Khan,
Massachusetts General Hospital and
Harvard Medical School, United States
Jinshan Tang,
Jinan University, China
Manikanta Murahari,
KL University, India

*CORRESPONDENCE

Yang Song,
✉ songyang1989@cdutcm.edu.cn
Quanyu Du,
✉ quanyu@cdutcm.edu.cn

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

SPECIALTY SECTION

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

RECEIVED 21 December 2022

ACCEPTED 17 March 2023

PUBLISHED 29 March 2023

CITATION

Guo J, Zhao Y, Wu X, Li G, Zhang Y, Song Y
and Du Q (2023), Mechanism exploration
and prognosis study of Astragali Radix-
Spreading hedyotis herb for the
treatment of lung adenocarcinoma based
on bioinformatics approaches and
molecular dynamics simulation.
Front. Chem. 11:1128671.
doi: 10.3389/fchem.2023.1128671

COPYRIGHT

© 2023 Guo, Zhao, Wu, Li, Zhang, Song
and Du. 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.

Mechanism exploration and prognosis study of Astragali Radix-Spreading hedyotis herb for the treatment of lung adenocarcinoma based on bioinformatics approaches and molecular dynamics simulation

Junfeng Guo^{1†}, Yuting Zhao^{2†}, Xuanyu Wu¹, Ganggang Li¹,
Yuwei Zhang¹, Yang Song^{1*} and Quanyu Du^{1*}

¹Hospital of Chengdu University of Traditional Chinese Medicine, Chengdu, China, ²Laboratory of Metabolomics and Drug-Induced Liver Injury, Frontiers Science Center for Disease-related Molecular Network, West China Hospital, Sichuan University, Chengdu, China

Background: Herb pair of Astragali Radix (AR) and Spreading Hedyotis Herb (SH) has been frequently prescribed in clinical for the treatment of lung cancer owing to its favorable efficacy. Yet, the mechanism under the therapeutic effects remained unveiled, which has limited its clinical applications, and new drug development for lung cancer.

Methods: The bioactive ingredients of AR and SH were retrieved from the Traditional Chinese Medicine System Pharmacology Database, with the targets of obtained components predicted by Swiss Target Prediction. Genes related to lung adenocarcinoma (LUAD) were acquired from GeneCards, OMIM and CTD databases, with the hub genes of LUAD screened by CTD database. The intersected targets of LUAD and AR-SH were obtained by Venn, with David Database employed to perform Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analyses. Survival analysis of the hub genes of LUAD was carried out using TCGA-LUAD dataset. Molecular docking of core proteins and active ingredients was performed by Auto-Dock Vina software, followed by molecular dynamics simulations of protein-ligand complexes with well-docked conformations.

Results: 29 active ingredients were screened out with 422 corresponding targets predicted. It is revealed that AR-SH can act on various targets such as EGFR, MAPK1, and KARS by ursolic acid (UA), Astragaloside IV(ASIV), and Isomucronulatol 7,2'-di-O-glucoside (IDOG) to alleviate the symptoms of LUAD. Biological processes involved are protein phosphorylation, negative regulation of apoptotic process, and pathways involved are endocrine resistance, EGFR tyrosine kinase inhibitor resistance, PI3K-Akt, and HIF-1 pathway. Molecular docking analysis indicated that the binding energy of most of the screened active ingredients to proteins encoded by core genes was less than -5.6 kcal/mol, with some active ingredients showing even lower binding energy to EGFR than Gefitinib. Three ligand-receptor complexes

including EGFR-UA, MAPK1-ASIV, and KRAS-IDOG were found to bind relatively stable by molecular dynamics simulation, which was consistent with the results of molecule docking.

Conclusion: We suggested that the herb pair of AR-SH can act on targets like EGFR, MAPK1 and KRAS by UA, ASIV and IDOG, to play a vital role in the treatment and the enhancement of prognosis of LUAD.

KEYWORDS

Astragali Radix, spreading hedyotis herb, prognosis, network pharmacology, molecular docking, molecular dynamics simulation

1 Introduction

Lung cancer, the major cause of cancer-related mortality around the world, resulted in 1.6 million deaths each year with a poor 5-year survival rate of only 19% (Bray et al., 2018; Siegel et al., 2019). Lung cancer can be categorized into small cell lung cancer (SCLC) (15%) and non-small cell lung cancer (NSCLC) (85%) based on the pathological characteristics and differentiation degree of cancer cells, where the latter is further divided into adenocarcinoma, squamous cell carcinoma, and large cell carcinoma (Sher et al., 2008). Lung adenocarcinoma (LUAD) is one of the common types of lung cancer, accounting for approximately 40% of all lung cancers, which originates from small airway epithelial, type II alveolar cells that secrete mucus and other substances (Noguchi et al., 1995; Zappa and Mousa 2016). Expression of mutated oncogenes in cells can lead to the activation of downstream signaling molecules that drive the abnormal proliferation and differentiation of cells to form tumor cells eventually. Various target agents have been developed that are effective and have low toxicity, but the therapeutic effect of targeted therapy remains unsatisfactory (Eguchi et al., 2008; Yue et al., 2018; Zhong et al., 2021). Although great efforts have been made over the decades, LUAD remains a persistent disease, making it increasingly imperative to search for more effective therapies and drugs for LUAD.

Traditional Chinese Medicine (TCM) pays attention to the enhancement of healthy Qi in patient and individualized therapy for each person in the treatment of lung cancer, which has been widely applied in clinical practice. The advantages of TCM therapy for cancer are extensive, including the improved survival quality of patients (Duflos et al., 2002; Efferth et al., 2007), enhanced physical fitness of patients, alleviation of clinical symptoms, minimum side effects, reduced side effects by radiotherapy, prolonged survival with tumor (Tian and Liu, 2010), and extended survival time (Liao et al., 2017). Modern pharmacological research have demonstrated that TCM and its extracts can act on tumor cells through multiple targets to inhibit the proliferation and migration of tumor cells, playing an essential role in all stages of tumor therapy. TCM has shown potent therapeutic effects to enhance efficacy and reduce toxicity in the complementary treatment of lung cancer, but the underlying molecular mechanisms are too complex and have yet to be revealed.

Through data mining, we discovered that herb pair of AR-SH was most frequently used in the treatment of lung cancer in clinical (Chen et al., 2022). AR, one of the most commonly used tonic herbs in clinical practice, can strengthen the spleen, and enhance the body, where modern pharmacological studies have shown that AR has a wide range of effects including hepatoprotective, diuretic,

hypotensive, and immunomodulatory functions (Bedir et al., 2000). Extracts of AR has been widely used as alternative therapies in the treatment of various diseases, including fatigue, anorexia, anemia, fever, allergies, gastric ulcers, and cancer (Astragalus, 2003; Fu et al., 2014). SH is a famous herb with heat-clearing and detoxifying properties, possessing several biological activities, such as neuroprotection (Kim et al., 2001) and antitumor activity (Lee et al., 2011). The anti-tumor effect of SH is generally recognized. SH has been shown to inhibit angiogenesis of tumor (Lin et al., 2011), combat HepG2 cancer cells through inducing apoptosis (Li et al., 2016), effectively kill human colorectal cancer cells (Lin et al., 2015) and breast cancer cells (Liu et al., 2010).

Network pharmacology is a popular method for predicting the underlying mechanism of herbal medicines. Prognostic analysis is an essential way to evaluate the efficacy of antitumor drugs, and molecular docking (Wang and Zhu 2016), and molecular dynamics simulations (De Vivo et al., 2016) can be used to validate and complement the network pharmacological results. In this study, a systemic pharmacology strategy (Figure 1) integrating network pharmacology, molecular docking and molecular dynamics simulations, was employed to explore the active components of AR-SH, and their corresponding targets and signaling pathways in the treatment of LUAD, with prognostic analysis used to examine key targets of AR-SH so as to provide scientific evidence for the complementary therapeutic effect of AR-SH in LUAD.

2 Materials and methods

2.1 Collection of active compounds and targets prediction of AR-SH

The active compounds of AR-SH were searched using the Traditional Chinese Medicine Systems Pharmacology (TCMSP, <https://old.tcmsp-e.com/tcmsp.php>) with the criteria set as oral-bioavailability (OB) $\geq 30\%$ and drug-likeness (DL) ≥ 0.18 . In addition, bioactive ingredients of AR-SH were supplemented from relevant literature. Pubchem database (<https://pubchem.ncbi.nlm.nih.gov/>) were employed to acquire the 2D structure of the ingredients, which were further uploaded to Swiss Target Prediction (<http://www.swisstargetprediction.ch/>) for target prediction with screening standard as Probability ≥ 0.1 . We calculated the similarity matrix of the molecules by Morgan Fingerprint in the RDKit toolkit, and the similarity was evaluated using the Tanimoto score (Hert et al., 2004; Rogers and Hahn 2010).

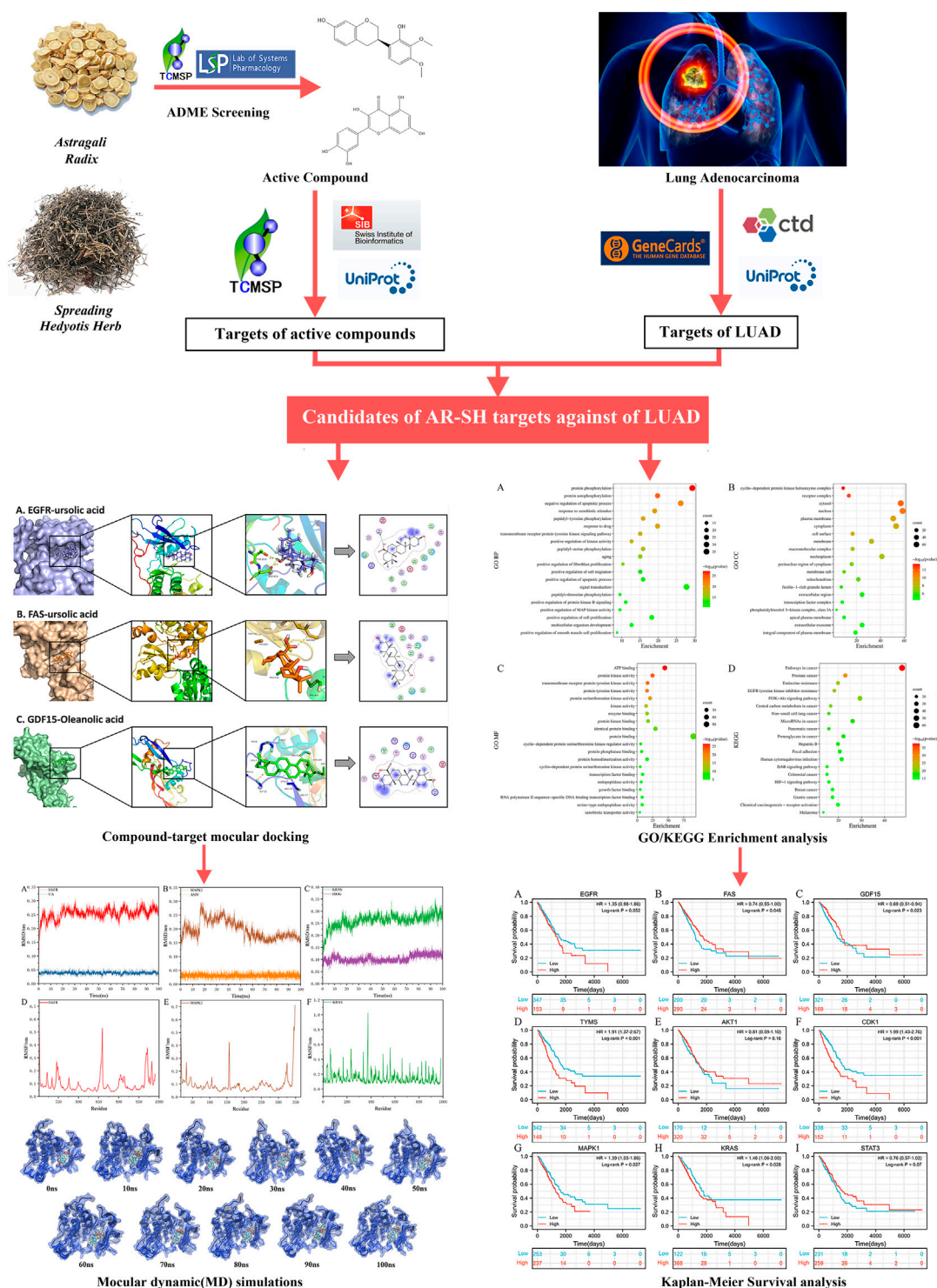


FIGURE 1
Flow chart of the employed systemic pharmacology strategy.

2.2 Acquisition of LUAD genes and screening of hub genes

LUAD related genes were obtained using the keyword “lung adenocarcinoma” in various databases, including GeneCard

(<https://www.genecards.org/>), Online Mendelian Inheritance in Man (OMIM, <https://omim.org/>) and Comparative Toxicogenomics Database (CTD, <http://ctdbase.org/>). The genes retrieved from the databases were integrated and de-duplicated, and the protein names were normalized using the

Uniprot (<https://www.uniprot.org/>) database. Through the CTD database, the common targets with the highest “Inference Score” and “References” were selected as hub genes of LUAD.

2.3 Gene ontology (GO) and kyoto encyclopedia of genes and genomes (KEGG) pathway enrichment analyses

The screened targets of AR-SH and LUAD genes were imported into the Venny2.1 online platform to capture the common targets of active compounds and LUAD. GO enrichment analysis of the intersected targets was performed in terms of the biological process (BP), cellular component (CC), and molecular function (MF) based on the David database. KEGG was selected for target pathway annotation analysis with P set less than 0.05, and the top 20 KEGG signal pathways were ranked according to the results in descending order of enrichment value.

2.4 Survival analysis of hub genes of LUAD

Gene expression data and survival information obtained from the Cancer Genome Atlas (TCGA) database were assessed by Kaplan-Meier survival analysis, and a log-rank test was performed using the survival package version 2.44-1.1 in R software.

2.5 The molecular docking of active compounds of AR-SH with core proteins of LUAD.

The 2D structure of the screened ingredients were downloaded from the PubChem database, which were imported into Chem3D software to draw the 3D structure of the compounds and optimize the energy of the ligand structure. The processed structure was saved in PDB format, and AutoDockTools-1.5.6 software was then applied to add charge and display rotatable keys, with the final structure saved in PDBQT format.

Next, the protein crystal structures encoded by hub genes were obtained from the PDB database (<https://www.rcsb.org/>), which were imported into PyMOL software to remove solvent and ligand. AutoDockTools-1.5.6 software was then employed to add hydrogen atoms, and the structure was saved in PDBQT format, with their active pockets searched. Molecular docking was performed by adjusting the X-Y-Z coordinates and grid size of the protein and optimizing the position of the protein structure binding sites. Processed active compound and the protein were docked for ten times by AutoDock Vina with the minimum binding energy of each docking taken as the final result. Docking results of the clinically used epidermal growth factor receptor-tyrosine kinase inhibitor (EGFR-TKI) Gefitinib with the core proteins were compared with those of the screened compounds with the core proteins. Docked ligand-protein complexes with lower docking binding affinity and research value for each protein were selected for further detailed demonstration.

2.6 Molecular dynamic (MD) simulation

The conformations of core protein-ligand complexes with lower docking binding affinity and research significance in the molecular docking results were further analyzed by MD simulations. MD simulation was carried out using GROMACS (version 2021-2). Protein topology file was generated using the AMBER99SB-ILDN force field, whereas ligand topology file was generated by ACPYPE script using the AMBER14SB force field. MD simulation was carried out in a dodecahedral box filled with TIP3 water molecules, and periodic bounding conditions were applied. The system was neutralized with NaCl counter ions. Energy minimization was achieved using the steepest descent algorithm, with cutoff of 1.4 nm for Coulomb interactions and Van der Waals interactions.

Before the simulation, each system was equilibrated for 100 ps at 310 K for NVT (constant atomic number, volume, and temperature) using a V-rescale thermostat (Bussi et al., 2007) and for 100 ps at 1.0 bar for NPT (constant atomic number, pressure, and temperature) via a Parrinello-Rahman barometer. The protein backbone was inhibited, while the solvent and countercharge ions were allowed to move during the equilibrium phase. The LINCS algorithm was used for all binding constraints. The particle-mesh Ewald (PME) method was used for long-range electrostatic processing. During the simulation, the positional constraints were removed. Finally, simulations were performed for 100 ns for each system under periodic boundary conditions at 310 K temperature and 1.0 bar pressure, and snapshots of the trajectories were taken every 10 ns.

2.7 Free binding energy calculations

The calculation of the free binding energy of protein-ligand complexes is an important way to verify the strength of intermolecular interactions, providing insight into the relative importance of various chemical energies that contribute to the overall stability. The molecular mechanics Poisson-Boltzmann surface area (MM-PBSA) method is a simple technique for quantifying the binding free energy of a ligand docked to an acceptor (Miller et al., 2012). The g_mmpbsa (Miller et al., 2012) tool was used to calculate the binding affinity of simulated protein-ligand complexes.

In general, Formula 1 can be used to calculate the free binding energy of a protein to a ligand in a solvent (Kollman et al., 2000):

$$\Delta G_{\text{bind}} = G_{\text{complex}} - (G_{\text{protein}} + G_{\text{ligand}}) \quad (1)$$

G_{protein} and G_{ligand} denote the total free energy of the isolated protein and ligand in the solvent, respectively, and G_{complex} represents the total free energy of the protein-ligand complex. In addition, the free energy of each entity can be obtained using Formula 2:

$$G_x = (E_{\text{MM}}) - TS + (G_{\text{solvation}}) \quad (2)$$

X denotes protein or ligand or protein-ligand complex. (E_{MM}) represents the average molecular mechanical potential energy in

TABLE 1 Active components of AR-SH.

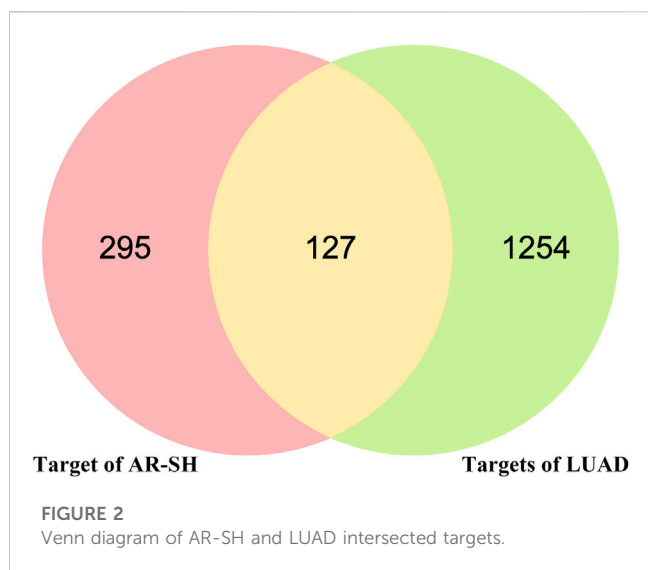
Num	PubChem CID	Molecule name	Source	MW	OB (%)	DL
1	10380176	(R)-Isomucronulatol	AR	302.35	67.67	0.26
2	15976101	(24S)-24-Propylcholesta-5-ene-3 β -ol	AR	428.82	36.23	0.78
3	11869658	3-Epioleanolic acid	SH	456.78	32.03	0.76
4	14077830	Astrapterocarpan	AR	300.33	64.26	0.42
5	5316760	1,7-Dihydroxy-3,9-dimethoxy pterocarpene	AR	314.31	39.05	0.48
6	2782115	2-(Chloromethyl)-4-(4-nitrophenyl)-1,3-thiazole	AR	254.69	\	\
7	162906151	2,3-dimethoxy-6-methyanthraquinone	SH	282.31	34.86	0.26
8	10514946	2-methyl-3-methoxyanthraquinone	SH	252.28	37.83	0.21
9	15689655	3,9-di-O-methylnissoin	AR	314.36	53.74	0.48
10	162842488	5'-hydroxyiso-muronulatol-2',5'-di-O-glucoside	AR	642.67	41.72	0.69
11	15689652	7-O-methylisomucronulatol	AR	316.38	74.69	0.3
12	101679160	9,10-dimethoxypterocarpan-3-O- β -D-glucoside	AR	462.49	36.74	0.92
13	222284	beta-sitosterol	SH	414.79	36.91	0.75
14	108213	Bifendate	AR	418.38	31.1	0.67
15	5280448	Calycosin	AR	284.28	47.75	0.24
16	6037	FA	AR	441.45	68.96	0.71
17	5280378	formononetin	AR	268.28	69.67	0.21
18	73299	hederagenin	AR	414.79	36.91	0.75
19	15689653	Isomucronulatol 7,2'-di-O-glucoside	AR	626.67	49.28	0.62
20	5281654	isorhamnetin	AR	316.28	49.6	0.31
21	5318869	Jaranol	AR	314.31	50.83	0.29
22	5280863	kaempferol	AR	286.25	41.88	0.24
23	64971	Mairin	AR SH	456.78	55.38	0.78
24	10494	Oleanolic acid	SH	456.78	29.02	0.76
25	5281330	Poriferasterol	SH	412.77	43.83	0.76
26	5280343	quercetin	AR SH	302.25	46.43	0.28
27	5280794	Stigmasterol	SH	412.77	43.83	0.76
28	64945	ursolic acid	SH	456.78	16.77	0.75
29	122130319	AstragalosideIV	AR	785.09	17.74	0.15

vacuum. ($G_{\text{solvation}}$) denotes the solvation free energy (Kollman et al., 2000; Kumari et al., 2014). TS represents the entropic contribution of the free energy in vacuum, where T and S denote temperature and entropy, respectively. The TS term is the conformational entropy term associated with complex, and isolated protein is calculated in the vacuum environment. Instead of considering absolute binding free energy, we focused on the contribution of individual residues of protein and ligands to the individual components of E_{MM} and

$G_{\text{solvation}}$ terms. The change in entropy term was neglected owing to that it does not affect the relative binding energy of ligands.

The molecular mechanics potential energy E_{MM} is the vacuum potential energy and includes both bonded and non-bonded interactions. It is calculated using molecular mechanics (MM) force field parameters, as in Formula 3

$$E_{\text{MM}} = E_{\text{bonded}} + E_{\text{nonbonded}} = E_{\text{bonded}} + (E_{\text{vdW}} + E_{\text{elec}}) \quad (3)$$



The value of E_{bonded} energy can be taken as zero under the assumption that the bound and unbound forms of protein and ligand conformations in the single trajectory method are similar (Homeyer and Gohlke 2012). Non-bonded interactions ($E_{\text{nonbonded}}$) include electrostatic (E_{elec}) and van der Waals (E_{vdW}) interactions.

The free energy of dissolution is the energy required to transfer the solute from the vacuum to the solvent. In the MM-PBSA method, the free energy of dissolution is calculated using the following solvent model, as in **Formula 4**:

$$G_{\text{solvation}} = G_{\text{PB}} + G_{\text{SA}} \quad (4)$$

G_{PB} and G_{SA} denote the electrostatic and non-electrostatic contributions to the free energy of dissolution, respectively. The electrostatic term G_{polar} was calculated by solving the Poisson-Boltzmann (PB) equation (Wang et al., 2004), and the G_{SA} term was calculated using the solvent accessible surface area (SASA). We also performed studies related to the energy decomposition of each residue, which help to estimate the MM-PBSA binding energy of the ligand in the protein-ligand complex.

3 Results

3.1 The active compounds and targets of AR-SH

The active compounds and targets of AR-SH were obtained by searching the TCMSP platform and the Swiss Target Prediction database, respectively. Among them, 29 active compounds of AR were retrieved with 361 targets predicted, and 2 compounds supplemented from literature were Astragalus polysaccharide (Bamodu et al., 2019) and AstragalosideIV (Zhang et al., 2018; Chen et al., 2021). 7 active compounds of SH were retrieved, with 227 targets predicted, and 3 compounds supplemented from literature included ursolic acid, Mairin and oleanolic acid (Liang et al., 2022), as shown in **Table 1**. 21 of the 29 obtained active compounds were found to possess a diversity index less than 0.8 and

TABLE 2 The inference score and reference score of the hub genes.

Gene symbol	Inference score	References
EGFR	41.93	38
FAS	39.02	23
GDF15	35.62	22
TYMS	35.01	24
AKT1	34.17	34
CDK1	33.91	26
MAPK1	32.37	39
KRAS	32.07	33
STAT3	30.93	29
MMP9	28.3	29

an average score of 0.178, which indicated a favorable diversity, as shown in **Supplementary Table S1**.

3.2 Acquisition of LUAD-related genes and screening of hub genes of LUAD

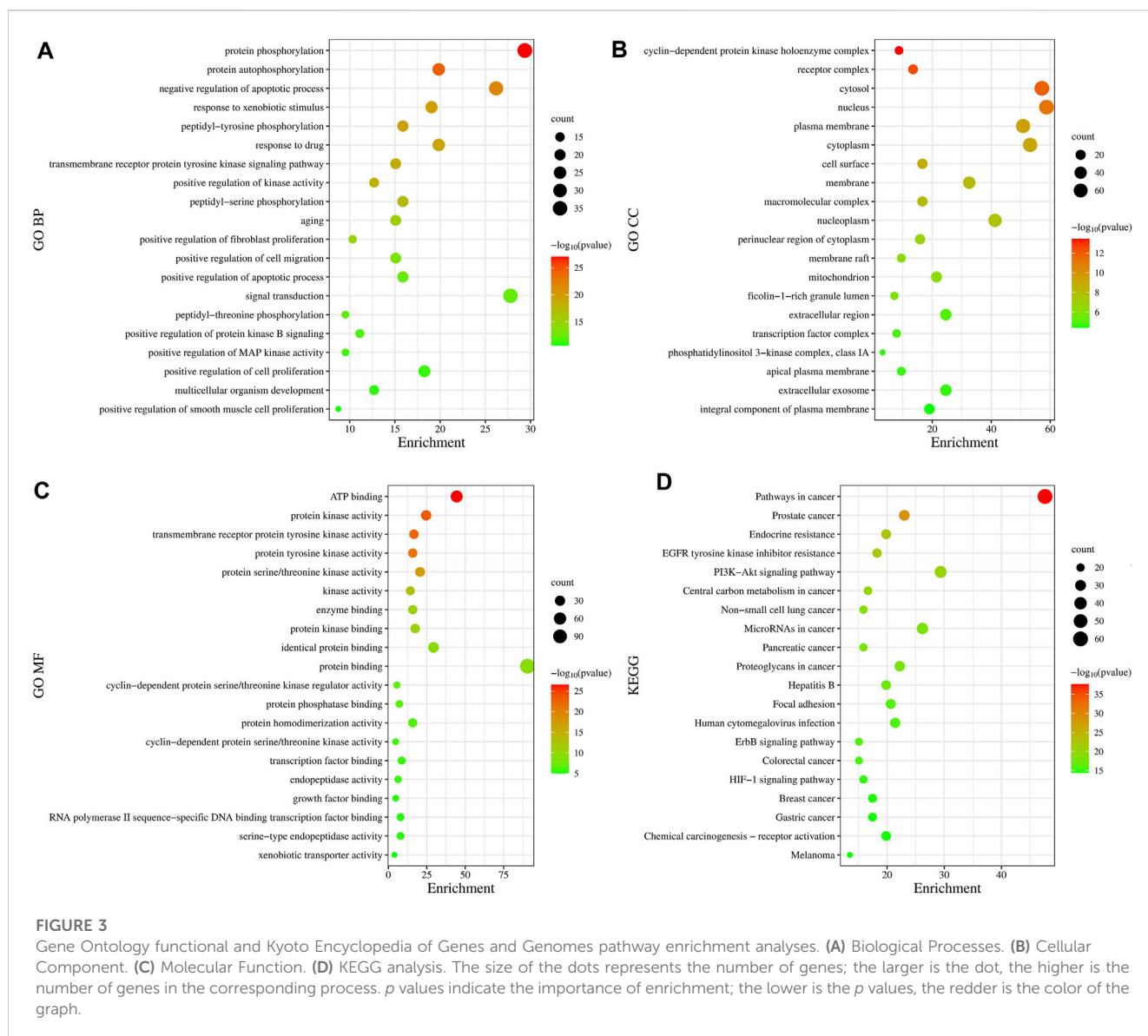
Using Gene Cards, OMIM and CTD databases, 1091, 227 and 157 LUAD-related genes were acquired respectively, with a total 1381 LUAD-related genes obtained after de-duplication. The 422 active compounds targets and 1381 LUAD genes were analyzed by Venn, and 127 common targets were obtained, which maybe the potential targets of AR-SH for LUAD treatment, as shown in **Figure 2**.

Based on CTD database, 10 hub genes were selected as docking targets for the next simulation experiments according to their "Inference Score" and "References" scores as well as lung adenocarcinoma-related research hotspots. As shown in the **Table 2**, the proteins coded by the 10 hub genes are Epidermal growth factor receptor (EGFR), Fas cell surface death receptor (FAS), Growth differentiation factor 15 (GDF15), Thymidylate synthetase (TYMS), AKT serine/threonine kinase 1 (AKT1), Cyclin dependent kinase 1 (CDK1), Mitogen-activated protein kinase 1 (MAPK1), KRAS proto-oncogene, GTPase (KRAS), Signal transducer and activator of transcription 3 (STAT3), and Matrix metalloproteinase-9 (MMP9).

3.3 GO and KEGG pathway enrichment analysis

GO is a bioinformatics analysis tool that defines the input genes by describing the function of the gene and the relationship between the enriched terms. GO functional analysis divides the gene functions into three parts: cellular component (CC), molecular function (MF), and biological process (BP), among which, BP can best reflect changes in biological function within the body.

In total, 775 GO entries of the GO functional enrichment analysis were obtained from DAVID database, including



556 entries in BP, 89 entries in CC, and 130 entries related to MF. **Figure 3A** shows that the potential targets were mainly enriched in BP such as protein phosphorylation, negative regulation of apoptotic process, response to xenobiotic stimulus, peptide-tyrosine phosphorylation and response to drugs. The involved terms of CC shown in **Figure 3B** are cyclin-dependent protein kinase holoenzyme complex, receptor complex, cytoplasm and plasma membrane accounted for a significant proportion. As for MF, ATP binding, protein kinase activation, transmembrane receptor protein tyrosine kinase activity, protein tyrosine kinase activity and protein serine/threonine kinase activity were ranked in the top, as shown in **Figure 3C**. Using KEGG pathway enrichment analysis, 147 pathways were screened out based on the threshold of $p < 0.05$. As shown in **Figure 3D**, the pathways with the highest significance, involve a variety of cancer pathways, including non-small cell lung cancer, prostate cancer, pancreatic cancer, colorectal cancer, etc. Cancer-related cell alterations, including endocrine resistance, EGFR tyrosine

kinase inhibitor resistance, central carbon metabolism and proteoglycans in cancer are also enriched significantly. Signaling pathways including PI3K-Akt and HIF-1 also are shown to be important.

3.4 Survival analysis of important targets

TCGA-LUAD dataset consisting of 526 LUAD samples and 59 normal samples were obtained from the TCGA database. We divided the LUAD samples into high and low expression groups according to the expression levels of the ten hub genes, and further investigated the correlation between the expression of the ten hub genes and the prognosis of LUAD patients by Kaplan-Meier survival analysis. As shown in **Figure 4**, the expression of FAS ($p = 0.046$), GDF15 ($p = 0.023$), TYMS ($p < 0.001$), CDK1 ($p < 0.001$), MAPK1 ($p = 0.027$) and KRAS ($p = 0.028$) showed significant correlation with prognosis. The survival analysis revealed that the correlation

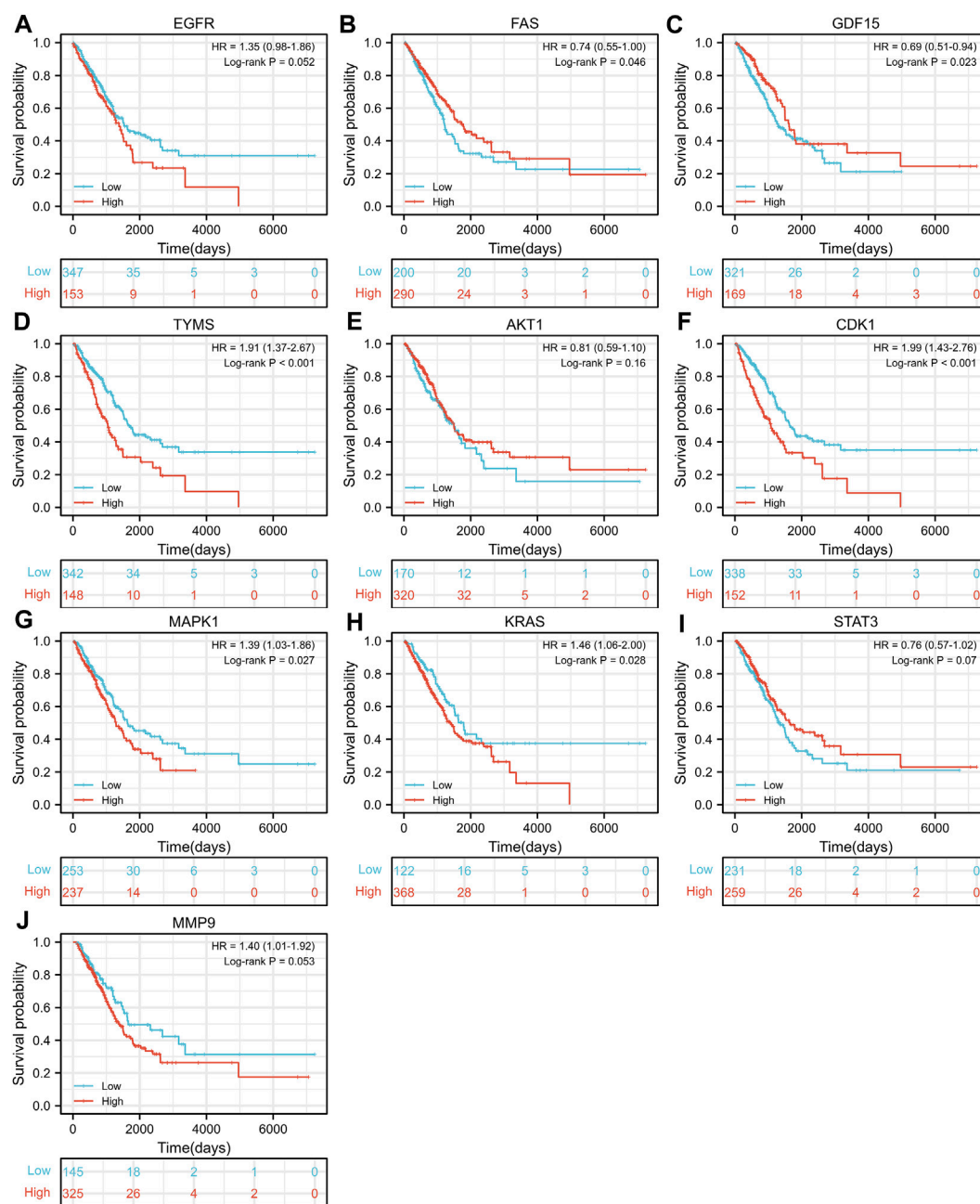


FIGURE 4
Kaplan-Meier Survival analysis of the correlation between expression of important target genes (A–J) and prognosis of LUAD in TCGA database.

between the expression of other hub genes and the survival is not statistically significant. However, EGFR has been proven to be a determinant driving lung adenocarcinoma growth and treatment response *in vivo* (Foggetti et al., 2021), thus subsequent analysis will also be performed for EGFR.

3.5 Molecular docking

The molecular docking results are illustrated in Figure 5 (the unit of measurement are kcal/mol). The redder the color, the lower

binding energy and stronger affinity of the ligand-protein complex. The bluer the color, the higher the binding energy and weaker affinity of the ligand-protein complex. It is generally accepted that a compound with a binding energy less than -5.6 kcal/mol to the receptor protein indicates a strong binding (Hsin et al., 2016). The molecular docking results showed that the binding energy of most of the screened active compounds to core proteins was lower than -5.6 kcal/mol, with the binding energy of some active compounds to important targets being even less than that of Gefitinib. So, we assume that the active compounds of AR-SH can effectively treat LUAD *via* multiple targets. The complexes

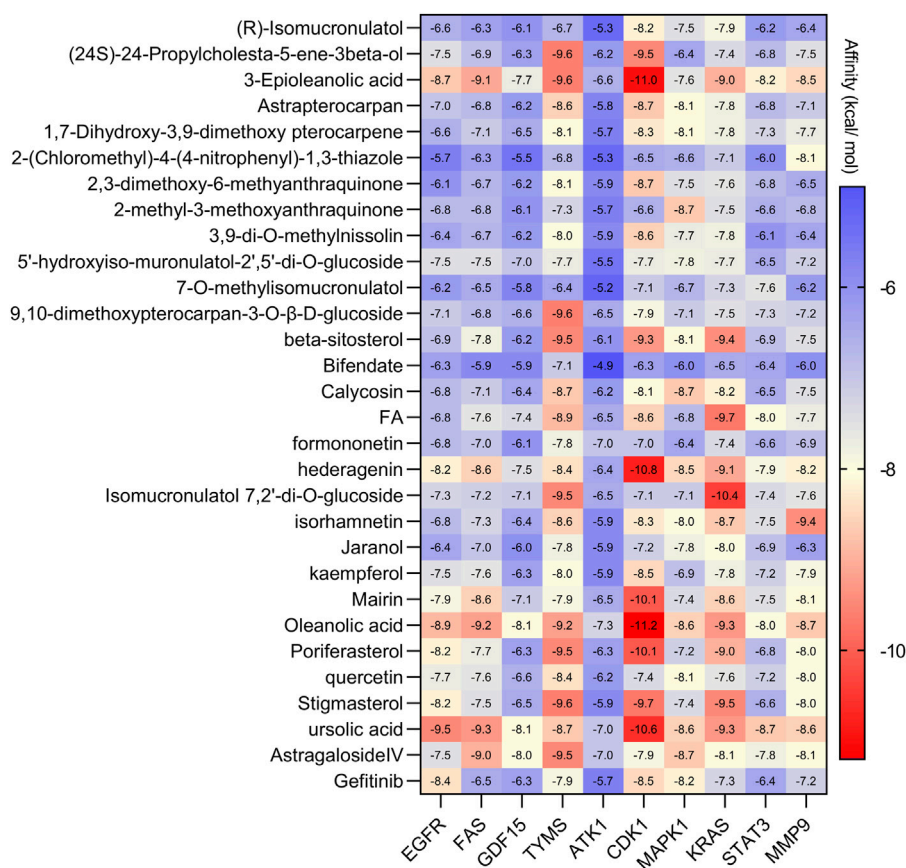


FIGURE 5

Docking results of active ingredients of AR-SH with core proteins.

with lower binding energy and better conformation in each group of docking results, which were shown in Figure 5, were selected for detailed demonstration to investigate the stability of the binding, as shown in Figure 6; Figure 7 The box co-ordinates and grid size information of the protein-ligand binding sites were shown in Supplementary Table S2.

We observed that UA and Gefitinib enter the same EGFR binding cavity. UA interacted with Thr-854 and Asp855 via hydrogen bonding and forms hydrophobic contacts with Ala722, Phe723, Val726, Ala743, Leu792, Met793 and Leu844. The interaction of UA with these residues may be the reason for its action on the target EGFR. Gefitinib forms hydrogen bonds with Asp855, as well as two π -H bonds with Leu718 and interacts with Val726, Ala743, Leu792, Met793, Leu844 and Met1002 through hydrophobic bonds.

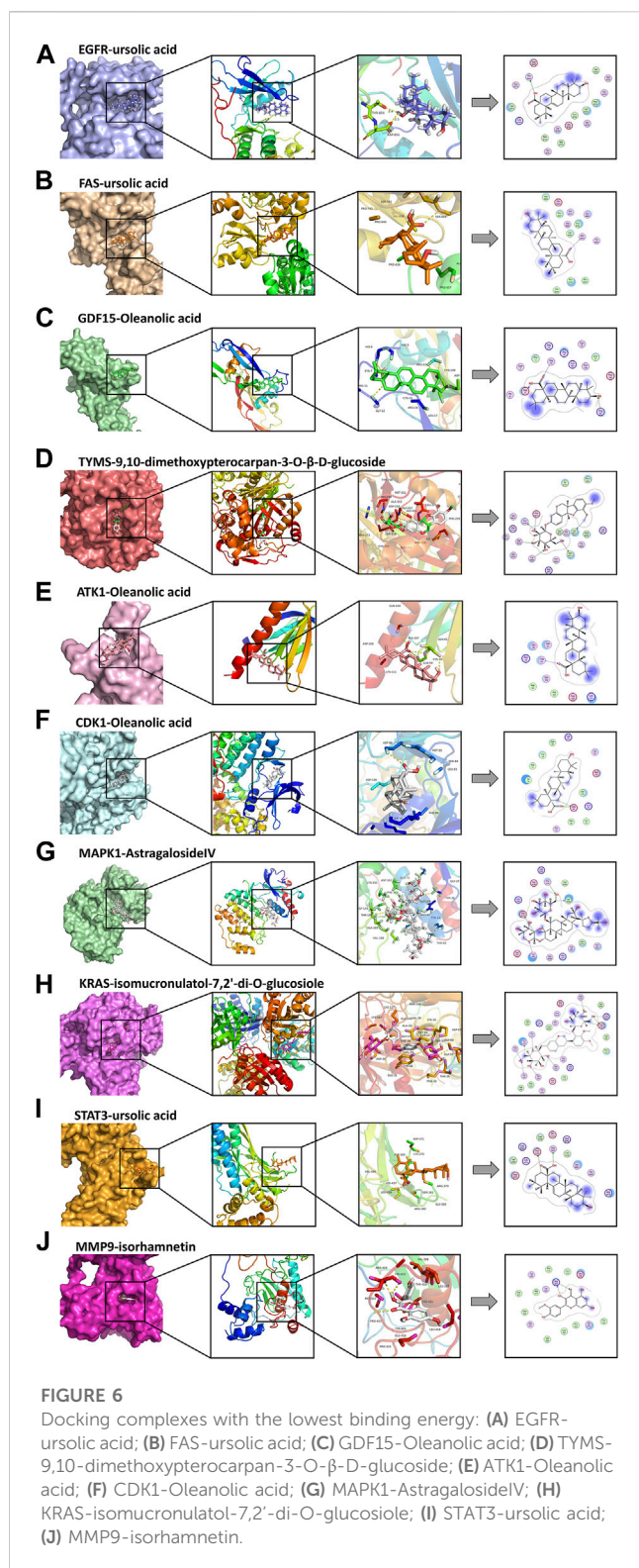
3.6 MD simulation

EGFR-ursolic acid (UA), MAPK1-AstragalosideIV (ASIV) and KRAS-Isomucronulatol 7,2'-di-O-glucoside (IDOG), which are the complexes with favorable conformations and research value in the docking results, were selected for further MD simulations. MD simulations can provide a digital environmental condition like those of human cells for us, involving temperature, pressure, solvents and

ions, to investigate the effects of temperature and environmental conditions on the binding process. Therefore, data obtained from MD simulations can offer valuable insights into the mechanism, dynamics, and nature of ligand-protein interactions (Wang et al., 2001).

For the information of the equilibrium time of each simulated protein-ligand complex during the MD simulation, the Root Mean Square Deviation (RMSD) of the protein backbone was calculated. RMSD is a valuable parameter for estimating changes or variations in molecular conformation, whose plots are commonly used to assess the time it takes for a system to reach structural equilibrium and to estimate the duration of the run. During the period of dramatic change in the initial structural conditions, a sudden increase of the RMSD values of the simulated complexes including the reference is expected, because the protein is rigid and would return to its dynamic motion when it is solvated in the water box in the crystal structure.

As shown in Figure 8A–C, the horizontal coordinates represent the time, while the vertical coordinates represent the specific values of RMSD. Sharp fluctuations of the RMSD of the three receptor-ligand complexes were witnessed in the initial stage. As the simulation proceeds, the RMSD of the three complexes tends to be smooth and stable after 20 ns. EGFR and UA were stable near 0.25 nm and 0.04 nm, MAPK1 and ASIV were stable near 0.17 nm and 0.03 nm after 50 ns, and KRAS and IDGO were stable near 0.25 nm and 0.1 nm. This



phenomenon suggests that the three complexes were relatively stable in stimulated conditions (Martínez 2015). It is necessary to note that the higher the RMSD value, the more unstable the complexes (Zhao et al., 2015). Therefore, EGFR-UA exhibited greater stability.

Root Mean Square Fluctuation (RMSF) is used to examine areas with high levels of volatility, where a higher RMSF value indicates a less

stable protein-ligand complex. As shown in Figures 8D–F, KRAS and the tail of MAPK1 exhibit high RMSF values, which may due to the presence of a large number of tightly coiled structures (e.g. α -helix and β -sheet). In addition, the lower RMSF value can be caused by the loss of the corresponding structures in the complex.

Hydrogen bonding facilitates the binding capability between proteins and ligands, and the number of hydrogen bonding can reflect the induced binding affinity (Dichiara et al., 2020). As shown in Figures 8G–I, KRAS forms 6 hydrogen bonds with the ligand on average, MAPK1 forms an average of 3 hydrogen bonds with the ligand, and EGFR creates 2 hydrogen bonds on average with ligands, all of which contribute to the stable binding of the complexes.

The radius of gyration (R_g) is directly associated with the tertiary structure and overall conformational state that has been utilized to determine whether a structure has a stable, compact and folded conformation. The Larger R_g value, the more flexible proteins, and the more unstable the complexes of ligand-protein. In contrast, lower R_g values indicate densely and tightly packed protein structures (Islam et al., 2021; Dey et al., 2022). As shown in Figure 8J, EGFR and MAPK1 exhibit low R_g values, from which we thought that they are stable.

The solvent accessible surface area (SASA) can be used to describe the effective interaction between ligand complexes and receptors (Geierhaas et al., 2007), which represents the interconnection between the water molecules and the surface of the complex submerged in water molecules. SASA is based on the ratio of the total area to energy. Compounds with high SASA values form unstable protein-ligand complexes due to their easy access to solvent, while complexes with low SASA values are considered to be stable (Patel et al., 2021). Through the hydrophobic interactions in non-polar amino acids, the SASA value of the complex can be maximally reduced (Shivanika et al., 2022). In Figure 8K, EGFR and MAPK1 possess low SASA values, indicating their better stability.

Molecular mechanics Poisson-Boltzmann surface area (MM/PBSA) is an effective and reliable method for calculating the free binding energy of small inhibitors to their protein targets (Wang et al., 2017). The free binding energies of the three complexes and their changes within 100 ns of simulation are shown in Figure 8L. The average free binding energy of KRAS is -53.08 kJ/mol, the average free binding energy of MAPK1 is -85.81 kJ/mol, and the average free binding energy of EGFR is -139.21 kJ/mol. From Table 3 we could propose the complex of EGFR-UA with the best binding energy, of which the E_{vdW} and E_{ele} were both lower than that of MAPK1-ASIV. Although the E_{vdW} and E_{ele} of KRAS-IDOG were the lowest, the highest E_{PB} hindered the binding of receptor and ligand.

The contribution of protein residues to free binding energy was calculated. As shown in Figures 8M–O, in the 100 ns simulation, the EGFR complex has more amino acid residues that can provide binding energy compared to the KRAS and MAPK1 complexes, indicating higher binding of ligands and receptors of the EGFR complex. Residues contributing to the free binding energy of the EGFR complex were LYS716, LYS746, ARG795, ARG841, LYS897 and SER972. Residues contributing to the binding energy of the MAPK1 complex were VAL36, ILE81, THR102 and LEU153. ASN26 VAL44 and ARG149 made significant energy contributions to the KRAS complex. These amino acid residues contributing to the free binding energy play a pivotal role in the interaction with the ligand and are the active sites for binding, which is consistent with the molecular docking results.

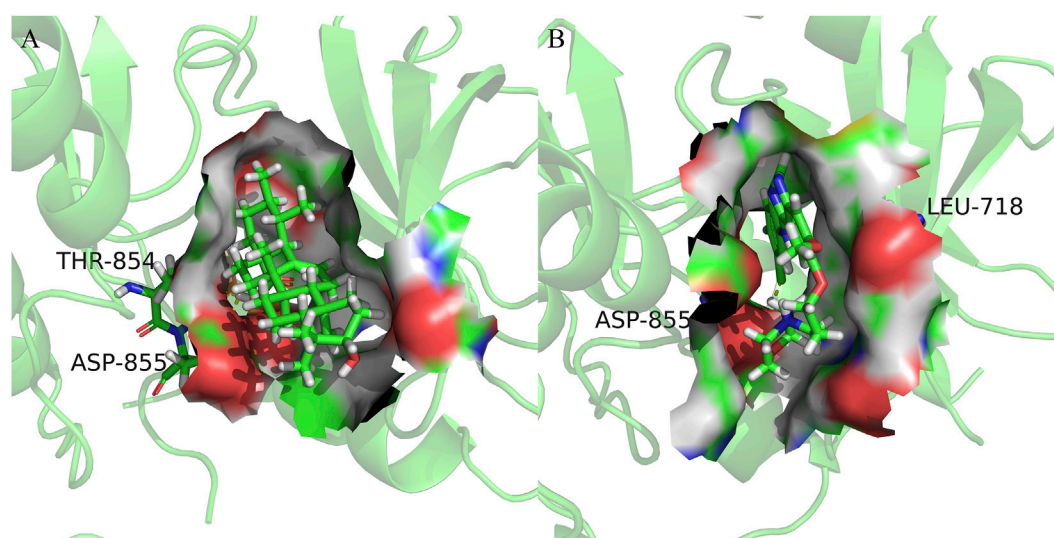


FIGURE 7

Cartoon representation of EGFR in complex with: (A) Ursolic acid; (B) Gefitinib. The binding site is shown as surface representation with the ligands shown as sticks.

The secondary structure analyses of 100 ns simulated trajectory were shown in **Figures 8P–R**. We identified that the number of secondary structures of the three complexes including α -Helix, β -Sheet, β -Bridge, Bend, Turn and Coil is kept relatively steady, and the fluctuations get smaller as the simulation proceeds, indicating that the complexes are relatively stable.

100 ns MD simulations analysis revealed that all three complexes were stable during the simulation. Based on the RMSD, RMSF, RG, SASA, hydrogen bonding number and free binding energy, the EGFR complex emerged as the most stable complex, followed by MAPK1, which might be because of Van der Waals forces and electrostatic potential energy. KRAS complex was slightly less stable may be attributed to its more flexible protein and denser convoluted structures.

The simulation trajectory of the EGFR-UA complex with the best stability was chosen for visualization. As displayed in **Figure 9**, UA can interact flexibly and gradually stabilize in the docking pocket of EGFR, with the snapshots of complexes obtained every 10 ns. As the simulation proceeded, the stability of the complex system did not change, and the ligand and protein were in a relatively static state of motion.

Free energy landscape (FEL) diagram was drawn to study the relationship between structural transitions or conformational changes of proteins and free binding energy through appropriate conformational sampling procedures. RMSD and Gyrate were selected to construct 3D landscape maps to detect and explore their steady-state structures. As we can see in **Figure 10**, the FEL plot of EGFR-UA has a minimum in a single lowest energy well, and the free energy values are below 0 kJ/mol, indicating that the system has good stability.

4 Discussion

Targets chemotherapy remains to play a leading role in the treatment for the majority of patients with advanced-stage LUAD,

and EGFR-TKI is the first line drug for lung cancer patients harboring an EGFR mutation in routine clinical practice (Ciuleanu et al., 2012). EGFR-TKI has been confirmed to significantly prolong disease free survival (DFS) but not overall survival (OS) of patients compared to conventional chemotherapeutic agents (Wu et al., 2022). Nowadays, Chinese herbal medicines have been demonstrated by several studies to increase therapeutic efficiency and reduce the adverse effects of chemotherapy drugs (Tseng et al., 2016; Yang et al., 2019a; Zhang et al., 2020; Wei et al., 2022).

Herb pair of AR and SH were found to be most frequently used in the treatment of lung cancer in clinical (CHEN.H.F et al., 2022). AR, one of the most commonly used tonic herbs in clinical practice, can strengthen the spleen, and enhance the body, where modern pharmacological studies have shown that AR has a wide range of effects including hepatoprotective, diuretic, hypotensive, and immunomodulatory functions (Bedir et al., 2000). AR may inhibit the progression and metastasis of LUAD by regulating immune system such as modulating macrophage polarization (Xu et al., 2018). SH possesses heat-clearing and detoxifying properties, with several biological activities, such as neuroprotection (Kim et al., 2001) and antitumor activity (Lee et al., 2011). But there is no systematic study on the bioactive ingredients of AR-SH and the underlying mechanism of AR-SH compounds in the treatment of LUAD by now. Therefore, a network pharmacology strategy and molecular docking approach as well as molecular dynamics simulations were adopted to identify the potential targets and elucidate mechanisms of action of AR-SH in the treatment of LUAD.

A total of 29 active compounds were acquired from TCMSP using ADME parameters, and literature, with 422 targets obtained. 1381 LUAD-related targets were collected from GeneCards, OMIM and CTD databases. There are 127 common targets of AR-SH and LUAD. Among the

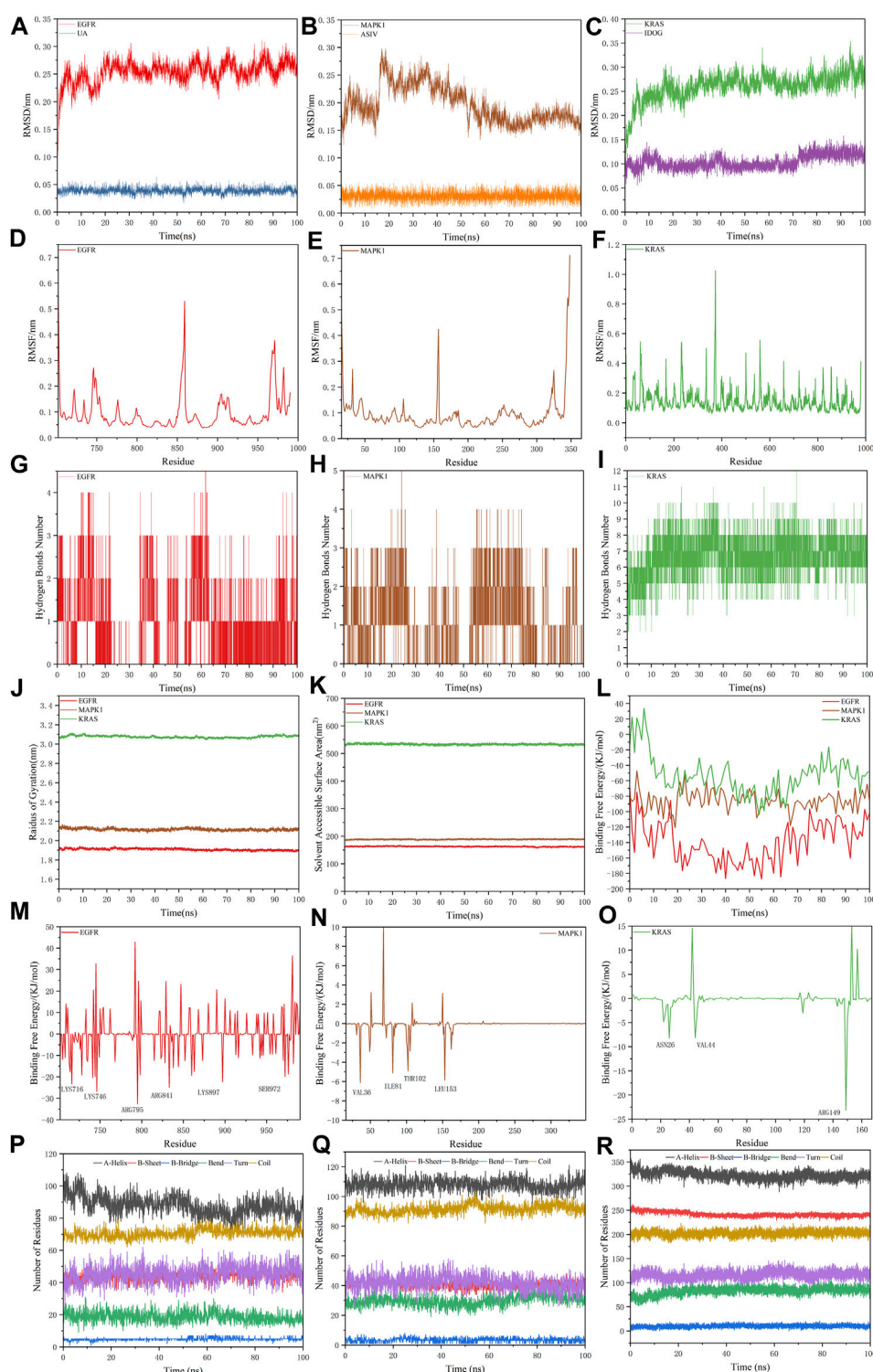


FIGURE 8

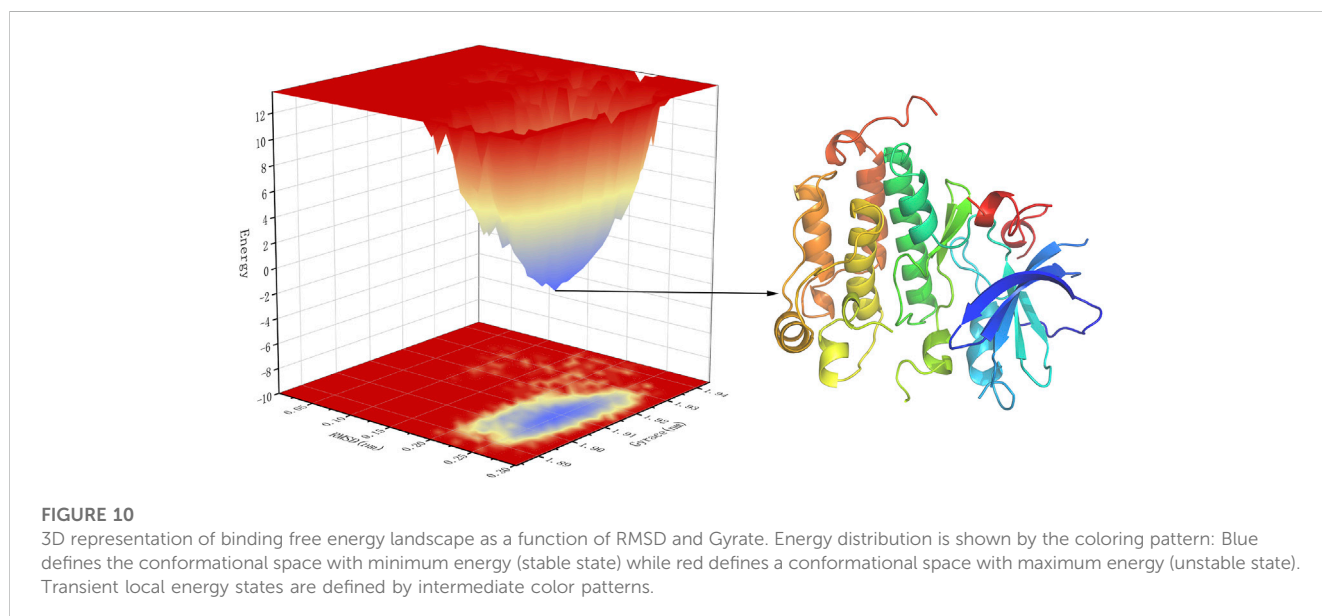
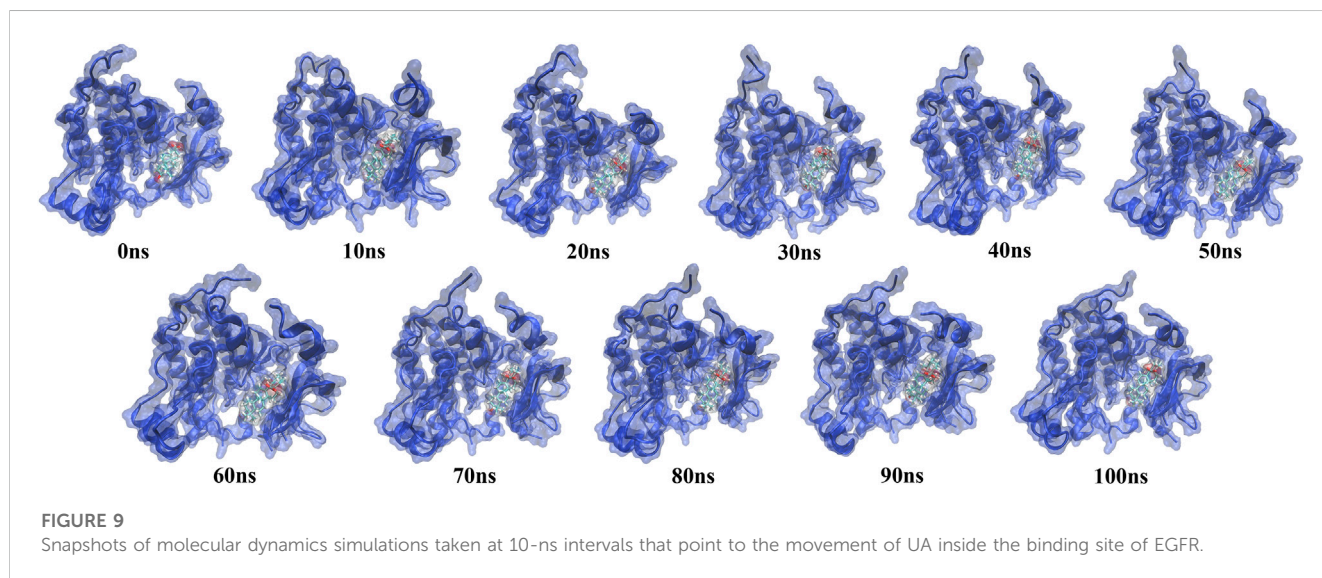
Molecular dynamics simulations. (A–C) The RMSD plot of EGFR-UA, MAPK1-ASIV and KRAS-IDOG. (D–F) The RMSF plot of EGFR-UA, MAPK1-ASIV and KRAS-IDOG. (G–I) The hydrogen bond numbers of EGFR-UA, MAPK1-ASIV and KRAS-IDOG. (J,K) Rg and SASA plots of EGFR-UA, MAPK1-ASIV and KRAS-IDOG. (L) Binding Free Energy plots of EGFR-UA, MAPK1-ASIV and KRAS-IDOG. (M–O) Binding energy contribution plots of amino acid residues of EGFR-UA, MAPK1-ASIV and KRAS-IDOG. (P–R) The secondary structure analysis plot of EGFR-UA, MAPK1-ASIV and KRAS-IDOG.

bioactive compounds, UA is a natural pentacyclic triterpenoid with anticancer activity against a variety of cancers *in vitro* and *in vivo* (Shanmugam et al., 2011; Yang et al., 2016; Yang et al.,

2019b). Wang et al. (2020) proved that UA can suppress the proliferation of various lung cancer cells, including human NSCLC cells H460, H1975, A549, H1299 and H520.

TABLE 3 Binding free energies of complexes in kJ/mol.

Complexes	G_{bind} (\pm SEM)	E_{MM} (\pm SEM)	E_{vdW} (\pm SEM)	E_{ele} (\pm SEM)	E_{PB} (\pm SEM)	E_{SA} (\pm SEM)
EGFR-UA	-139.21 ± 2.596	-212.0 ± 3.003	-167.8 ± 1.153	-44.15 ± 2.801	99.70 ± 2.732	-26.94 ± 0.1091
MAPK1-ASIV	-85.81 ± 1.284	-164.2 ± 1.078	-145.2 ± 0.8670	-18.97 ± 0.6718	97.60 ± 1.176	-19.23 ± 0.05296
KRAS-IDOG	-53.08 ± 2.546	-368.6 ± 3.087	-247.2 ± 2.245	-121.3 ± 1.926	348.3 ± 2.051	-32.83 ± 0.08525



Yang et al. (2019b) found that UA can inhibit the expression of CT45A2, and suppress the proliferation and motility of tumor cells while promoting apoptosis in NSCLC carrying the EGFR T790M mutation with this mutation being the main cause of drug resistance to EGFR. ASIV is a naturally occurring tetracyclic triterpene saponin that has been shown to be free of any significant hepatotoxic or nephrotoxic effects. Studies have

shown that ASIV enhances the Bax/Bcl-2 ratio and induces intrinsic apoptosis in a variety of cancer cells, including cells of colorectal, breast, lung, vulvar squamous cell carcinoma, and hepatocellular carcinoma (Jia et al., 2019; Sun et al., 2019; Zhao et al., 2019; Zheng et al., 2019; Cui et al., 2020). Li et al. (2017a) suggested that ASIV can inhibit glioma progression by interfering with the MAPK/ERK signaling pathway, which is consistent with

the high and stable MAPK1-ASIV binding ability in the present study, making ASIV a promising anti-cancer candidate. IDOG showed a strong binding affinity with KARS, indicating it may be a potentially active compound against cancer. Further studies should be conducted to investigate the anti-tumor effects of IDOG. In addition to the tumor suppressive function, we found that UA and ASIV can also co-regulate immune function. UA could reduce Th1 cytokine expression (IL-2, IL-6, IL-12, IFN- γ and TNF- α) and induce Th2 (IL4, IL5) cytokine expression (Raphael and Kuttan 2003; Ahmad et al., 2006). ASIV was thought to act as an immune adjuvant (Hong et al., 2011) to enhance cellular immune function by activating the NF- κ B/ MAPK signaling pathway (Li et al., 2017b). The cooperative anticancer activity of UA and ASIV, and the modulation of the immune system demonstrated the synergistic effect of the AR-SH drug pair.

GO enrichment analysis showed that the biological processes involved in AR-SH treatment of LUAD mainly include protein phosphorylation, negative regulation of apoptotic process, response to xenobiotic stimulus, peptide-tyrosine phosphorylation and response to drugs. KEGG pathway analysis associated with AR-SH against LUAD includes pathways of a variety of cancers, such as non-small cell lung cancer, prostate cancer, pancreatic cancer, and colorectal cancer. Cancer-related cellular alterations include endocrine resistance, EGFR tyrosine kinase inhibitor resistance, central carbon metabolism, and proteoglycans in cancer. PI3K-Akt, HIF-1 and other signaling pathways were also engaged. The emergence of drug resistance remains a major issue for EGFR-TKIs treatment of lung cancer. PI3K-Akt, an important signaling pathway present in normal human cells, is involved in a variety of physiological and pathological processes and plays a central regulatory role in cell growth and proliferation. Furthermore, PI3K-Akt pathway can also affect the development of NSCLC by inducing apoptosis, inhibiting cell proliferation, invasion and migration, and regulating tumor angiogenesis (Wang et al., 2019; Chen et al., 2020; Hu et al., 2020). HIF-1 is a key transcriptional activator that mediates the adaptive response of the organism to hypoxia, which regulates gene expression through changes in intracellular oxygen concentration and exerts an influential role in tumor cell hypoxia adaptation, energy metabolism, tumor angiogenesis, and invasion and metastasis, with its expression of HIF-1 closely related to invasive metastasis of lung cancer (Hua et al., 2020). During rapid tumor cell multiplication in patients with non-small cell lung cancer, tumor cells are in a relatively hypoxic state, making HIF-1 α more likely to be activated and stay in a highly expressed stage.

Based on the reference "Inference Score" and "References" scores of each target in the CTD database and the hot spots of LUAD-related research, 10 targets were selected and considered as core targets of AR-SH for LUAD treatment, including EGFR, FAS, GDF15, TYMS, AKT1, CDK1, MAPK1, KRAS, STAT3 and MMP9. Survival analysis of core targets revealed a significant correlation between the expression of FAS ($p = 0.046$), GDF15 ($p = 0.023$), TYMS ($p < 0.001$), CDK1 ($p < 0.001$), MAPK1 ($p = 0.027$) and KRAS ($p = 0.028$) and prognosis of LUAD. However, it is undeniable that EGFR is a determinant driving lung adenocarcinoma growth and treatment response *in vivo* (Foggetti et al., 2021). EGFR and KRAS are the two most frequently mutated oncogenic driver genes (Rodenhuis et al., 1987; Lynch et al., 2004; Paez et al., 2004)

that occur in the presence of multiple identified tumor suppressor gene alterations (Cancer Genome Atlas Research Network, 2014; Politi and Herbst 2015; Campbell et al., 2016; Skoulidis and Heymach 2019). EGFR, a receptor-type tyrosine kinase, is overexpressed and/or mutated in LUAD and controls tumor growth through signaling regulation. The expression of EGFR is closely associated with neo-angiogenesis, tumor invasion and metastasis (Cancer Genome Atlas Research Network, 2014), whose mutations are a major causative factor for LUAD in East Asian countries (accounting for approximately 60% of LUAD) (Dong et al., 2018). Remarkable advances have been made in the treatment of advanced NSCLC with molecularly targeted EGFR-TKIs, yet patients are highly susceptible to drug resistance (Tan et al., 2017; Dong et al., 2018). Mutation of KRAS was first initiated in lung cancer in the 1980s (Santos et al., 1984), which is a gene that is hard to target. Mutations in the KRAS gene directly trigger the EGFR-Ras-Raf-MAPK pathway in the EGF signaling pathway, followed by activation and overexpression of MAPK1 to further promote tumor cell migration and invasion, increase cell viability and participate in epithelial mesenchymal transition, allowing the rapid progress of LUAD and rendering targeted drugs against the EGFR upstream pathway ineffective (Lee et al., 2014). There exist a close relationship between mutations of KRAS and MAPK1 and the resistance of NSCLC to EGFR-TKI targeted drugs such as Gefitinib and Erlotinib, which can cause sustained activation of the EGFR signaling pathway and accelerate tumor cell proliferation (Zer et al., 2016).

What's more, as one of the most commonly used tonic herbs, AR is not negligible for its modulating effect on the immune system. We found that the core targets EGFR, MAPK1 and KRAS are closely related to the regulation of immune system function. In the EGFR-positive genetic state, tumors exhibit a relatively immunosuppressive microenvironment, as evidenced by a decrease in CD8⁺ T cells and an increase in regulatory T cells (Treg) (Xiao et al., 2023). Of note, KRAS-mutant tumors showed a marked immune activation status in LUAD, exemplified by an elevated abundance of CD8⁺T cells, Cytotoxic T Lymphocyte cells (CTL), and Follicular helper T cells (Tfh), and reduced immunosuppressed M2-macrophage. In primary lung cancer, a retrospective study found KRAS-mutant tumors had a significantly higher PD-L1 expression, high CD8⁺T cells infiltration and higher TMB than EGFR-mutant tumors (Liu et al., 2020). Therefore, the high response rate of KRAS-positive tumors to immunotherapy may be related to the activated immune microenvironment (Lee et al., 2018). Zfp831, as a downstream molecule of MAPK1, directly binds to the Tfh cells signature gene Bcl6 and thus promotes Tfh cells differentiation (Wan et al., 2021). Consequently, we suggest that AR can modulate the immune system by acting on the core targets EGFR, MAPK1 and KRAS, thus improving the tumor immune microenvironment.

Molecular docking results showed that the binding affinities of the screened active compounds to the core targets ranged from -4.9 to -11.2 kcal/mol, with most of the ingredients exhibiting binding energies less than -5.6 kcal/mol. The binding energy of UA and oleanolic acid of SH to EGFR was even less than that of Gefitinib, while ASIV and IDOG of AR showed stronger binding to MAPK1 and KARS, respectively. It is drug pair formed from AR and SH that the effect could be reached, therefore, the synergistic effects of AR-SH on the targets may be responsible for the treatment of LUAD. Three docked complexes including UA-EGFR,

ASIV-MAPK1 and IDOG-KRAS exhibited favorable docking conformations and low binding energies, and molecular dynamics simulations further suggested stability of the binding of docked complexes, with hydrogen bonding being the most critical factor for their stable binding. It was found that KRAS protein functions as a molecular switch, as it activates and regulates the downstream MAPK pathway in response to upstream EGFR, which amplifies the signaling efficiency of the MAPK pathway in KRAS mutations, ultimately controlling tumor cell proliferation and metastasis, and thus promoting tumor growth (Ponsioen et al., 2021). In this study, we found that the active ingredients of AR-SH can stably bind to EGFR, MAPK1 and KRAS to trigger or suppress their protein functions, thus contributing to the treatment of LUAD.

Although molecular dynamics simulation can be used to describe the motion of ligand-protein complex in one system, it lacks the ability to simultaneously show the interactions between the compound and other proteins which are unavoidable in human body. So, many unknown variables that cannot be controlled are stand in the way, which may have an impact accuracy of result. Despite these limitations, the molecular level analysis in this study provides a reference and guidance for further exploration of the mechanism of AR-SH for LUAD treatment. What's more, we also found that the main active compounds of AR-SH were not acquired from databases but from literature supplementation, so we suggest that we should not rely on the database alone for active ingredient mining.

5 Conclusion

In conclusion, we suggested that the herb pair of AR-SH can act on targets like EGFR, MAPK1 and KRAS by UA, ASIV and IDOG, to play a vital role in the treatment and the enhancement of prognosis of LUAD.

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**.

References

- Ahmad, S. F., Khan, B., Bani, S., Suri, K. A., Satti, N. K., and Qazi, G. N. (2006). Amelioration of adjuvant-induced arthritis by ursolic acid through altered Th1/Th2 cytokine production. *Pharmacol. Res.* 53 (3), 233–240. doi:10.1016/j.phrs.2005.11.005
- Astragalus (2003). Astragalus membranaceus, Monograph. *Altern. Med. Rev.* 8 (1), 72–77.
- Bamodu, O. A., Kuo, K. T., Wang, C. H., Huang, W. C., Wu, A. T. H., Tsai, J. T., et al. (2019). Astragalus polysaccharides (PG2) enhances the M1 polarization of macrophages, functional maturation of dendritic cells, and T cell-mediated anticancer immune responses in patients with lung cancer. *Nutrients* 11 (10), 2264. doi:10.3390/nu11102264
- Bedir, E., Pugh, N., Calis, I., Pasco, D. S., and Khan, I. A. (2000). Immunostimulatory effects of cycloartane-type triterpene glycosides from astragalus species. *Biol. Pharm. Bull.* 23 (7), 834–837. doi:10.1248/bpb.23.834
- Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R. L., Torre, L. A., and Jemal, A. (2018). Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA A Cancer J. Clin.* 68 (6), 394–424. doi:10.3322/caac.21492
- Bussi, G., Donadio, D., and Parrinello, M. (2007). Canonical sampling through velocity rescaling. *J. Chem. Phys.* 126 (1), 014101. doi:10.1063/1.2408420
- Campbell, J. D., Alexandrov, A., Kim, J., Wala, J., Berger, A. H., Pedamallu, C. S., et al. (2016). Distinct patterns of somatic genome alterations in lung adenocarcinomas and squamous cell carcinomas. *Nat. Genet.* 48 (6), 607–616. doi:10.1038/ng.3564
- Cancer Genome Atlas Research Network (2014). Comprehensive molecular profiling of lung adenocarcinoma. *Nature* 511 (7511), 543–550. doi:10.1038/nature13385
- Chen, H. F., Lei, X. D., Dang, H. F., Li, Y. H., Wang, H., and Xia, X. J. (2022). The study of the law of drug use for the treatment of lung cancer by Chinese medicine. *Gansu Med. J.* 41 (02), 134–136. doi:10.15975/j.cnki.gsy.2022.02.023
- Chen, M., Zhu, L. L., Su, J. L., Li, G. L., Wang, J., and Zhang, Y. N. (2020). Prucalopride inhibits lung cancer cell proliferation, invasion, and migration through blocking of the PI3K/

Author contributions

JG and YZ contributed equally to this manuscript. QD and JG designed the study. JG completed the data analysis and manuscript drafting. YZ completed the language revision and translation. XW, GL, and YZ contributed to the processing of the figures. QD and YS made the final revisions to the manuscript. All authors read and approved the final manuscript.

Funding

This study was supported by the Science and Technology Planning Project of Sichuan Provincial Department of Science and Technology (NO. 2022YFS0405); Science and Technology Research Project of Sichuan Administration of Traditional Chinese Medicine (NO. 2020JC0048); Research Innovation Project of Chengdu University of Traditional Chinese Medicine (NO.KY-2023038).

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/fchem.2023.1128671/full#supplementary-material>

- AKT/mTOR signaling pathway. *Hum. Exp. Toxicol.* 39 (2), 173–181. doi:10.1177/0960327119883409
- Chen, T., Yang, P., and Jia, Y. (2021). Molecular mechanisms of astragaloside-IV in cancer therapy (Review). *Int. J. Mol. Med.* 47 (3), 4846. doi:10.3892/ijmm.2021.4846
- Ciuleanu, T., Stelmakh, L., Cicenias, S., Miliauskas, S., Grigorescu, A. C., Hillenbach, C., et al. (2012). Efficacy and safety of erlotinib versus chemotherapy in second-line treatment of patients with advanced, non-small-cell lung cancer with poor prognosis (TITAN): A randomised multicentre, open-label, phase 3 study. *Lancet Oncol.* 13 (3), 300–308. doi:10.1016/s1470-2045(11)70385-0
- Cui, X., Jiang, X., Wei, C., Xing, Y., and Tong, G. (2020). Astragaloside IV suppresses development of hepatocellular carcinoma by regulating miR-150-5p/ β -catenin axis. *Environ. Toxicol. Pharmacol.* 78, 103397. doi:10.1016/j.etap.2020.103397
- 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
- Dey, D., Hossain, R., Biswas, P., Paul, P., Islam, M. A., Ema, T. I., et al. (2022). Amentoflavone derivatives significantly act towards the main protease (3CL(PRO)/M(PRO)) of SARS-CoV-2: In silico admet profiling, molecular docking, molecular dynamics simulation, network pharmacology. *Mol. Divers* 2022, 1–15. doi:10.1007/s11030-022-10459-9
- Dichiaro, M., Amata, B., Turnaturi, R., Marrazzo, A., and Amata, E. (2020). Tuning properties for blood-brain barrier permeation: A statistics-based analysis. *ACS Chem. Neurosci.* 11 (1), 34–44. doi:10.1021/acscchemneuro.9b00541
- Dong, J. K., Lei, H. M., Liang, Q., Tang, Y. B., Zhou, Y., Wang, Y., et al. (2018). Overcoming erlotinib resistance in EGFR mutation-positive lung adenocarcinomas through repression of phosphoglycerate dehydrogenase. *Theranostics* 8 (7), 1808–1823. doi:10.7150/thno.23177
- Duflos, A., Kruczyński, A., and Barret, J. M. (2002). Novel aspects of natural and modified vinca alkaloids. *Curr. Med. Chem. Anticancer Agents* 2 (1), 55–70. doi:10.2174/1568011023354452
- Efferth, T., Li, P. C., Konkimalla, V. S., and Kaina, B. (2007). From traditional Chinese medicine to rational cancer therapy. *Trends Mol. Med.* 13 (8), 353–361. doi:10.1016/j.molmed.2007.07.001
- Eguchi, T., Kodera, Y., Nakanishi, H., Yokoyama, H., Ohashi, N., Ito, Y., et al. (2008). The effect of chemotherapy against micrometastases and isolated tumor cells in lymph nodes: An *in vivo* study. *Vivo* 22 (6), 707–712.
- Foggetti, G., Li, C., Cai, H., Hellyer, J. A., Lin, W. Y., Ayeni, D., et al. (2021). Genetic determinants of EGFR-driven lung cancer growth and therapeutic response *in vivo*. *Cancer Discov.* 11 (7), 1736–1753. doi:10.1158/2159-8290.CD-20-1385
- Fu, J., Wang, Z., Huang, L., Zheng, S., Wang, D., Chen, S., et al. (2014). Review of the botanical characteristics, phytochemistry, and pharmacology of Astragalus membranaceus (Huangqi). *Phytother. Res.* 28 (9), 1275–1283. doi:10.1002/ptr.5188
- Geierhaas, C. D., Nickson, A. A., Lindorff-Larsen, K., Clarke, J., and Vendruscolo, M. (2007). BPPred: A web-based computational tool for predicting biophysical parameters of proteins. *Protein Sci.* 16 (1), 125–134. doi:10.1110/ps.062383807
- Hert, J., Willett, P., Wilton, D. J., Acklin, P., Azzouli, K., Jacoby, E., et al. (2004). Comparison of topological descriptors for similarity-based virtual screening using multiple bioactive reference structures. *Org. Biomol. Chem.* 2 (22), 3256–3266. doi:10.1039/b409865j
- Homeyer, N., and Gohlke, H. (2012). Free energy calculations by the molecular mechanics Poisson-Boltzmann surface area method. *Mol. Inf.* 31 (2), 114–122. doi:10.1002/minf.201100135
- Hong, F., Xiao, W., Ragupathi, G., Lau, C. B., Leung, P. C., Yeung, K. S., et al. (2011). The known immunologically active components of Astragalus account for only a small proportion of the immunological adjuvant activity when combined with conjugate vaccines. *Planta Med.* 77 (8), 817–824. doi:10.1055/s-0030-1250574
- Hsin, K. Y., Matsuoka, Y., Asai, Y., Kamiyoshi, K., Watanabe, T., Kawaoka, Y., et al. (2016). systemsDock: a web server for network pharmacology-based prediction and analysis. *Nucleic Acids Res.* 44 (1), W507–W513. doi:10.1093/nar/gkw335
- Hu, T., Shen, H., Huang, H., Yang, Z., Zhou, Y., and Zhao, G. (2020). Cholesterol-lowering drug pitavastatin targets lung cancer and angiogenesis via suppressing prenylation-dependent Ras/Raf/MEK and PI3K/Akt/mTOR signaling. *Anticancer Drugs* 31 (4), 377–384. doi:10.1097/cad.0000000000000885
- Hua, Q., Mi, B., Xu, F., Wen, J., Zhao, L., Liu, J., et al. (2020). Hypoxia-induced lncRNA-AC020978 promotes proliferation and glycolytic metabolism of non-small cell lung cancer by regulating PKM2/HIF-1 α axis. *Theranostics* 10 (11), 4762–4778. doi:10.7150/thno.43839
- Islam, R., Parves, M. R., Paul, A. S., Uddin, N., Rahman, M. S., Mamun, A. A., et al. (2021). A molecular modeling approach to identify effective antiviral phytochemicals against the main protease of SARS-CoV-2. *J. Biomol. Struct. Dyn.* 39 (9), 3213–3224. doi:10.1080/07391102.2020.1761883
- Jia, L., Lv, D., Zhang, S., Wang, Z., and Zhou, B. (2019). Astragaloside IV inhibits the progression of non-small cell lung cancer through the akt/GSK-3 β / β -catenin pathway. *Oncol. Res.* 27 (4), 503–508. doi:10.3727/096504018x15344989701565
- Kim, Y., Park, E. J., Kim, J., Kim, Y., Kim, S. R., and Kim, Y. Y. (2001). Neuroprotective constituents from Hedyotis diffusa. *J. Nat. Prod.* 64 (1), 75–78. doi:10.1021/np000327d
- Kollman, P. A., Massova, I., Reyes, C., Kuhn, B., Huo, S., Chong, L., et al. (2000). Calculating structures and free energies of complex molecules: Combining molecular mechanics and continuum models. *Acc. Chem. Res.* 33 (12), 889–897. doi:10.1021/ar000033j
- Kumari, R., Kumar, R., and Lynn, A. (2014). g_mmpbsa—a GROMACS tool for high-throughput MM-PBSA calculations. *J. Chem. Inf. Model* 54 (7), 1951–1962. doi:10.1021/ci500020m
- Lee, B., Lee, B., Han, G., Kwon, M. J., Han, J., and Choi, Y. L. (2014). KRAS mutation detection in non-small cell lung cancer using a peptide nucleic acid-mediated polymerase chain reaction clamping method and comparative validation with next-generation sequencing. *Korean J. Pathol.* 48 (2), 100–107. doi:10.4132/KoreanJPathol.2014.48.2.100
- Lee, C. K., Man, J., Lord, S., Cooper, W., Links, M., Gebbski, V., et al. (2018). Clinical and molecular characteristics associated with survival among patients treated with checkpoint inhibitors for advanced non-small cell lung carcinoma: A systematic review and meta-analysis. *JAMA Oncol.* 4 (2), 210–216. doi:10.1001/jamaoncol.2017.4427
- Lee, H. Z., Bau, D. T., Kuo, C. L., Tsai, R. Y., Chen, Y. C., and Chang, Y. H. (2011). Clarification of the phenotypic characteristics and anti-tumor activity of Hedyotis diffusa. *Am. J. Chin. Med.* 39 (1), 201–213. doi:10.1142/s0192415x11008750
- Li, B., Meng, T., Hao, N., Tao, H., Zou, S., Li, M., et al. (2017b). Immune regulation mechanism of Astragaloside IV on RAW264.7 cells through activating the NF- κ B/ MAPK signaling pathway. *Int. Immunopharmacol.* 49, 38–49. doi:10.1016/j.intimp.2017.05.017
- Li, B., Wang, F., Liu, N., Shen, W., and Huang, T. (2017a). Astragaloside IV inhibits progression of glioma via blocking MAPK/ERK signaling pathway. *Biochem. Biophys. Res. Commun.* 491 (1), 98–103. doi:10.1016/j.bbrc.2017.07.052
- Li, Y. L., Zhang, J., Min, D., Hongyan, Z., Lin, N., and Li, Q. S. (2016). Anticancer effects of 1,3-dihydroxy-2-methylanthraquinone and the ethyl acetate fraction of Hedyotis diffusa willd against HepG2 carcinoma cells mediated via apoptosis. *PLoS One* 11 (4), e0151502. doi:10.1371/journal.pone.0151502
- Liang, Y. N., Hou, B. L., Wu, K. N., Wu, D. Y., Pei, G., and Wang, Z. (2022). Study on chemical components from Hedyotis diffusa Willd and their anti-tumour activity. *Nat. Prod. Res. Dev.* 34 (08), 1281–1288+1300. doi:10.16333/j.1001-6880.2022.8.002
- Liao, Y. H., Li, C. I., Lin, C. C., Lin, J. G., Chiang, J. H., and Li, T. C. (2017). Traditional Chinese medicine as adjunctive therapy improves the long-term survival of lung cancer patients. *J. Cancer Res. Clin. Oncol.* 143 (12), 2425–2435. doi:10.1007/s00432-017-2491-6
- Lin, J., Li, Q., Chen, H., Lin, H., Lai, Z., and Peng, J. (2015). Hedyotis diffusa Willd. extract suppresses proliferation and induces apoptosis via IL-6-inducible STAT3 pathway inactivation in human colorectal cancer cells. *Oncol. Lett.* 9 (4), 1962–1970. doi:10.3892/ol.2015.2956
- Lin, J., Wei, L., Xu, W., Hong, Z., Liu, X., and Peng, J. (2011). Effect of Hedyotis diffusa willd extract on tumor angiogenesis. *Mol. Med. Rep.* 4 (6), 1283–1288. doi:10.3892/mmr.2011.577
- Liu, C., Zheng, S., Jin, R., Wang, X., Wang, F., Zang, R., et al. (2020). The superior efficacy of anti-PD-1/PD-L1 immunotherapy in KRAS-mutant non-small cell lung cancer that correlates with an inflammatory phenotype and increased immunogenicity. *Cancer Lett.* 470, 95–105. doi:10.1016/j.canlet.2019.10.027
- Liu, Z., Liu, M., Liu, M., and Li, J. (2010). Methylanthraquinone from Hedyotis diffusa Willd induces Ca(2+)-mediated apoptosis in human breast cancer cells. *Toxicol. Vitro* 24 (1), 142–147. doi:10.1016/j.tiv.2009.08.002
- Lynch, T. J., Bell, D. W., Sordella, R., Gurubhagavatula, S., Okimoto, R. A., Brannigan, B. W., et al. (2004). Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small-cell lung cancer to gefitinib. *N. Engl. J. Med.* 350 (21), 2129–2139. doi:10.1056/NEJMoa040938
- Martínez, L. (2015). Automatic identification of mobile and rigid substructures in molecular dynamics simulations and fractional structural fluctuation analysis. *PLoS One* 10 (3), e0119264. doi:10.1371/journal.pone.0119264
- Miller, B. R., 3rd, McGee, T. D., Jr., Swails, J. M., Homeyer, N., Gohlke, H., and Roitberg, A. E. (2012). MMPBSA.py: An efficient program for end-state free energy calculations. *J. Chem. Theory Comput.* 8 (9), 3314–3321. doi:10.1021/ct300418h
- Noguchi, M., Morikawa, A., Kawasaki, M., Matsuno, Y., Yamada, T., Hirohashi, S., et al. (1995). Small adenocarcinoma of the lung. Histologic characteristics and prognosis. *Cancer* 75 (12), 2844–2852. doi:10.1002/1097-0142(19950615)75:12<2844::aid-cnrc2820751209>3.0.co;2-#
- Paez, J. G., Jänne, P. A., Lee, J. C., Tracy, S., Greulich, H., Gabriel, S., et al. (2004). EGFR mutations in lung cancer: Correlation with clinical response to gefitinib therapy. *Science* 304 (5676), 1497–1500. doi:10.1126/science.1099314
- Patel, C. N., Goswami, D., Jaiswal, D. G., Parmar, R. M., Solanki, H. A., and Pandya, H. A. (2021). Pinpointing the potential hits for hindering interaction of SARS-CoV-2 S-protein with ACE2 from the pool of antiviral phytochemicals utilizing molecular docking and molecular dynamics (MD) simulations. *J. Mol. Graph Model* 105, 107874. doi:10.1016/j.jmgm.2021.107874
- Politi, K., and Herbst, R. S. (2015). Lung cancer in the era of precision medicine. *Clin. Cancer Res.* 21 (10), 2213–2220. doi:10.1158/1078-0432.Ccr-14-2748
- Ponsioen, B., Post, J. B., Buissant des Amorie, J. R., Laskaris, D., van Ineveld, R. L., Kersten, S., et al. (2021). Quantifying single-cell ERK dynamics in colorectal cancer

- organoids reveals EGFR as an amplifier of oncogenic MAPK pathway signalling. *Nat. Cell Biol.* 23 (4), 377–390. doi:10.1038/s41556-021-00654-5
- Raphael, T. J., and Kuttan, G. (2003). Effect of naturally occurring triterpenoids glycyrrhizic acid, ursolic acid, oleanolic acid and nomilin on the immune system. *Phytomedicine* 10 (6–7), 483–489. doi:10.1078/09447110332231421
- Rodenhuis, S., van de Wetering, M. L., Mooi, W. J., Evers, S. G., van Zandwijk, N., and Bos, J. L. (1987). Mutational activation of the K-ras oncogene. A possible pathogenetic factor in adenocarcinoma of the lung. *N. Engl. J. Med.* 317 (15), 929–935. doi:10.1056/nejm198710083171504
- Rogers, D., and Hahn, M. (2010). Extended-connectivity fingerprints. *J. Chem. Inf. Model* 50 (5), 742–754. doi:10.1021/ci100050t
- Santos, E., Martin-Zanca, D., Reddy, E. P., Pierotti, M. A., Della Porta, G., and Barbacid, M. (1984). Malignant activation of a K-ras oncogene in lung carcinoma but not in normal tissue of the same patient. *Science* 223 (4637), 661–664. doi:10.1126/science.6695174
- Shanmugam, M. K., Manu, K. A., Ong, T. H., Ramachandran, L., Surana, R., Bist, P., et al. (2011). Inhibition of CXCR4/CXCL12 signaling axis by ursolic acid leads to suppression of metastasis in transgenic adenocarcinoma of mouse prostate model. *Int. J. Cancer* 129 (7), 1552–1563. doi:10.1002/ijc.26120
- Sher, T., Dy, G. K., and Adjei, A. A. (2008). Small cell lung cancer. *Mayo Clin. Proc.* 83 (3), 355–367. doi:10.4065/83.3.355
- Shivanika, C., Deepak Kumar, S., Venkataraghavan Ragunathan, R., Sumitha, A., and Brindha Devi, P. (2022). Molecular docking, validation, dynamics simulations, and pharmacokinetic prediction of natural compounds against the SARS-CoV-2 main-protease. *J. Biomol. Struct. Dyn.* 40 (2), 585–611. doi:10.1080/07391102.2020.1815584
- Siegel, R. L., Miller, K. D., and Jemal, A. (2019). Cancer statistics, 2019. *CA Cancer J. Clin.* 69 (1), 7–34. doi:10.3322/caac.21551
- Skoulidis, F., and Heymach, J. V. (2019). Co-occurring genomic alterations in non-small-cell lung cancer biology and therapy. *Nat. Rev. Cancer* 19 (9), 495–509. doi:10.1038/s41568-019-0179-8
- Sun, P., Liu, Y., Wang, Q., and Zhang, B. (2019). Astragaloside IV inhibits human colorectal cancer cell growth. *Front. Biosci. (Landmark Ed.)* 24 (3), 597–606. doi:10.2741/4738
- Tan, C. S., Cho, B. C., and Soo, R. A. (2017). Treatment options for EGFR mutant NSCLC with CNS involvement—Can patients BLOOM with the use of next generation EGFR TKIs? *Lung Cancer* 108, 29–37. doi:10.1016/j.lungcan.2017.02.012
- Tian, X. Y., and Liu, L. (2010). Effect and advantage of orally taking Chinese herbal medicine for treatment of lung cancer. *China J. Chin. Materia Medica* 35 (21), 2795–2800.
- Tseng, C. Y., Lin, C. H., Wu, L. Y., Wang, J. S., Chung, M. C., Chang, J. F., et al. (2016). Potential combinational anti-cancer therapy in non-small cell lung cancer with traditional Chinese medicine Sun-Bai-pi extract and cisplatin. *PLoS One* 11 (5), e0155469. doi:10.1371/journal.pone.0155469
- Wan, S., Ni, L., Zhao, X., Liu, X., Xu, W., Jin, W., et al. (2021). Costimulation molecules differentially regulate the ERK-Zfp831 axis to shape T follicular helper cell differentiation. *Immunity* 54 (12), 2740–2755.e6. doi:10.1016/j.immuni.2021.09.018
- Wang, C., Greene, D., Xiao, L., Qi, R., and Luo, R. (2017). Recent developments and applications of the MMPBSA method. *Front. Mol. Biosci.* 4, 87. doi:10.3389/fmolb.2017.00087
- Wang, G., and Zhu, W. (2016). Molecular docking for drug discovery and development: A widely used approach but far from perfect. *Future Med. Chem.* 8 (14), 1707–1710. doi:10.4155/fmc-2016-0143
- 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 (9), 1157–1174. doi:10.1002/jcc.20035
- Wang, M., Yu, H., Wu, R., Chen, Z. Y., Hu, Q., Zhang, Y. F., et al. (2020). Autophagy inhibition enhances the inhibitory effects of ursolic acid on lung cancer cells. *Int. J. Mol. Med.* 46 (5), 1816–1826. doi:10.3892/ijmm.2020.4714
- Wang, Q., Liu, Z., Du, K., Liang, M., Zhu, X., Yu, Z., et al. (2019). Babaodan inhibits cell growth by inducing autophagy through the PI3K/AKT/mTOR pathway and enhances antitumor effects of cisplatin in NSCLC cells. *Am. J. Transl. Res.* 11 (8), 5272–5283.
- Wang, W., Donini, O., Reyes, C. M., and Kollman, P. A. (2001). Biomolecular simulations: Recent developments in force fields, simulations of enzyme catalysis, protein-ligand, protein-protein, and protein-nucleic acid noncovalent interactions. *Annu. Rev. Biophys. Biomol. Struct.* 30, 211–243. doi:10.1146/annurev.biophys.30.1.211
- Wei, J., Liu, Z., He, J., Liu, Q., Lu, Y., He, S., et al. (2022). Traditional Chinese medicine reverses cancer multidrug resistance and its mechanism. *Clin. Transl. Oncol.* 24 (3), 471–482. doi:10.1007/s12094-021-02716-4
- Wu, Y. L., John, T., Grohe, C., Majem, M., Goldman, J. W., Kim, S. W., et al. (2022). Postoperative chemotherapy use and outcomes from ADAURA: Osimertinib as adjuvant therapy for resected EGFR-mutated NSCLC. *J. Thorac. Oncol.* 17 (3), 423–433. doi:10.1016/j.jtho.2021.10.014
- Xiao, G., Li, L., Tanzhu, G., Liu, Z., Gao, X., Wan, X., et al. (2023). Heterogeneity of tumor immune microenvironment of EGFR/ALK-positive tumors versus EGFR/ALK-negative tumors in resected brain metastases from lung adenocarcinoma. *J. Immunother. Cancer* 11 (3). doi:10.1136/jitc-2022-006243
- Xu, F., Cui, W. Q., Wei, Y., Cui, J., Qiu, J., Hu, L. L., et al. (2018). Astragaloside IV inhibits lung cancer progression and metastasis by modulating macrophage polarization through AMPK signaling. *J. Exp. Clin. Cancer Res.* 37 (1), 207. doi:10.1186/s13046-018-0878-0
- Yang, K., Chen, Y., Zhou, J., Ma, L., Shan, Y., Cheng, X., et al. (2019a). Ursolic acid promotes apoptosis and mediates transcriptional suppression of CT45A2 gene expression in non-small-cell lung carcinoma harbouring EGFR T790M mutations. *Br. J. Pharmacol.* 176 (24), 4609–4624. doi:10.1111/bph.14793
- Yang, K., Li, R. Y., Yang, X. Y., Cui, Q. F., Wang, F. Y., Lin, G. Q., et al. (2019b). Co-Administration of shexiang baoxin pill and chemotherapy drugs potentiated cancer therapy by vascular-promoting strategy. *Front. Pharmacol.* 10, 565. doi:10.3389/fphar.2019.00565
- Yang, L. J., Tang, Q., Wu, J., Chen, Y., Zheng, F., Dai, Z., et al. (2016). Inter-regulation of IGFBP1 and FOXO3a unveils novel mechanism in ursolic acid-inhibited growth of hepatocellular carcinoma cells. *J. Exp. Clin. Cancer Res.* 35, 59. doi:10.1186/s13046-016-0330-2
- Yue, D., Xu, S., Wang, Q., Li, X., Shen, Y., Zhao, H., et al. (2018). Erlotinib versus vinorelbine plus cisplatin as adjuvant therapy in Chinese patients with stage IIIA EGFR mutation-positive non-small-cell lung cancer (EVAN): A randomised, open-label, phase 2 trial. *Lancet Respir. Med.* 6 (11), 863–873. doi:10.1016/s2213-2600(18)30277-7
- Zappa, C., and Mousa, S. A. (2016). Non-small cell lung cancer: Current treatment and future advances. *Transl. Lung Cancer Res.* 5 (3), 288–300. doi:10.21037/tlcr.2016.06.07
- Zer, A., Ding, K., Lee, S. M., Goss, G. D., Seymour, L., Ellis, P. M., et al. (2016). Pooled analysis of the prognostic and predictive value of KRAS mutation status and mutation subtype in patients with non-small cell lung cancer treated with epidermal growth factor receptor tyrosine kinase inhibitors. *J. Thorac. Oncol.* 11 (3), 312–323. doi:10.1016/j.jtho.2015.11.010
- Zhang, C., Cai, T., Zeng, X., Cai, D., Chen, Y., Huang, X., et al. (2018). Astragaloside IV reverses MNNG-induced precancerous lesions of gastric carcinoma in rats: Regulation on glycolysis through miRNA-34a/LDHA pathway. *Phytother. Res.* 32 (7), 1364–1372. doi:10.1002/ptr.6070
- Zhang, Y., Lou, Y., Wang, J., Yu, C., and Shen, W. (2020). Research status and molecular mechanism of the traditional Chinese medicine and antitumor therapy combined strategy based on tumor microenvironment. *Front. Immunol.* 11, 609705. doi:10.3389/fimmu.2020.609705
- Zhao, Y., Wang, L., Wang, Y., Dong, S., Yang, S., Guan, Y., et al. (2019). Astragaloside IV inhibits cell proliferation in vulvar squamous cell carcinoma through the TGF- β /Smad signaling pathway. *Dermatol Ther.* 32 (4), e12802. doi:10.1111/dth.12802
- Zhao, Y., Zeng, C., and Massiah, M. A. (2015). Molecular dynamics simulation reveals insights into the mechanism of unfolding by the A130T/V mutations within the MID1 zinc-binding Bbox1 domain. *PLoS One* 10 (4), e0124377. doi:10.1371/journal.pone.0124377
- Zheng, Y., Dai, Y., Liu, W., Wang, N., Cai, Y., Wang, S., et al. (2019). Astragaloside IV enhances taxol chemosensitivity of breast cancer via caveolin-1-targeting oxidant damage. *J. Cell Physiol.* 234 (4), 4277–4290. doi:10.1002/jcp.27196
- Zhong, W. Z., Wang, Q., Mao, W. M., Xu, S. T., Wu, L., Wei, Y. C., et al. (2021). Gefitinib versus vinorelbine plus cisplatin as adjuvant treatment for stage II-IIIa (N1-N2) EGFR-mutant NSCLC: Final overall survival analysis of CTONG1104 phase III trial. *J. Clin. Oncol.* 39 (7), 713–722. doi:10.1200/jco.20.01820



OPEN ACCESS

EDITED BY

Khurshid Ahmad,
Yeungnam University, Republic of Korea

REVIEWED BY

Juan Manuel Guzman-Flores,
University of Guadalajara, Mexico
Jalaluddin Ashraf,
Samarkand State Medical Institute,
Uzbekistan

*CORRESPONDENCE

Naif K. Binsaleh,
✉ n.binsaleh@uoh.edu.sa

RECEIVED 22 March 2023

ACCEPTED 10 April 2023

PUBLISHED 20 April 2023

CITATION

Binsaleh NK, Sherwani S, Eltayeb R,
Qanash H, Bazaid AS, Althobiti M,
Hazzazi MS and Rajendrasozhan S (2023),
Increased inflammatory cytokines and
oxidative stress enhanced antibody
production in breast and prostate cancer
patients with COVID-19
related depression.
Front. Chem. 11:1192074.
doi: 10.3389/fchem.2023.1192074

COPYRIGHT

© 2023 Binsaleh, Sherwani, Eltayeb,
Qanash, Bazaid, Althobiti, Hazzazi and
Rajendrasozhan. 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.

Increased inflammatory cytokines and oxidative stress enhanced antibody production in breast and prostate cancer patients with COVID-19 related depression

Naif K. Binsaleh^{1*}, Subuhi Sherwani², Reem Eltayeb¹,
Husam Qanash¹, Abdulrahman S. Bazaid¹, Maryam Althobiti³,
Mohannad S. Hazzazi^{4,5} and Saravanan Rajendrasozhan⁶

¹Department of Medical Laboratory Sciences, Faculty of Applied Medical Sciences, University of Ha'il, Ha'il, Saudi Arabia, ²Department of Biology, Faculty of Sciences, University of Ha'il, Ha'il, Saudi Arabia, ³Department of Clinical Laboratory Science, College of Applied Medical Science, Shaqra University, Shaqra, Saudi Arabia, ⁴Department of Medical Laboratory Sciences, Faculty of Applied Medical Sciences, King Abdulaziz University, Jeddah, Saudi Arabia, ⁵Hematology Research Unit, King Fahd Medical Research Center, King Abdulaziz University, Jeddah, Saudi Arabia, ⁶Department of Chemistry, Faculty of Sciences, University of Ha'il, Ha'il, Saudi Arabia

Cancer management is highly dependent on the immune status of the patient. During the COVID-19 pandemic, a large number of people suffered from anxiety and depression, especially cancer patients. The effect of depression on breast cancer (BC) and prostate cancer (PC) patients, during the pandemic has been analyzed in this study. Levels of proinflammatory cytokines (IFN- γ , TNF- α , and IL-6) and oxidative stress markers malondialdehyde (MDA) and carbonyl content (CC) were estimated in patients' serum samples. Serum antibodies against *in vitro* hydroxyl radical (\bullet OH) modified pDNA (\bullet OH-pDNA-Abs) were estimated using direct binding and inhibition ELISA. Cancer patients showed increased levels of proinflammatory cytokines (IFN- γ , TNF- α , and IL-6) and oxidative stress markers (MDA and CC levels), which were further significantly enhanced in cancer patients with depression compared to normal healthy (NH) individuals. Increased levels of \bullet OH-pDNA-Abs were detected in breast cancer (0.506 ± 0.063) and prostate cancer (0.441 ± 0.066) patients compared to NH subjects. Serum antibodies were found to be significantly elevated in BC patients with depression (BCD) (0.698 ± 0.078) and prostate cancer patients with depression (PCD) (0.636 ± 0.058). Inhibition ELISA also exhibited significantly high percent inhibition in BCD ($68.8\% \pm 7.8\%$) and PCD ($62.9\% \pm 8.3\%$) subjects compared to BC ($48.9\% \pm 8.1\%$), and PC ($43.4\% \pm 7.5\%$) subjects. Cancer is characterized by enhanced oxidative stress and increased inflammation, which may be exaggerated with COVID-19 related depression. High oxidative stress and compromised antioxidant homeostasis exerts alterations in DNA, leading to formation of neo-antigens, subsequently leading to the generation of antibodies. COVID-19 pandemic related depression needs to be addressed globally for improved cancer patient care and cancer disease management.

KEYWORDS

breast cancer, prostate cancer, COVID-19, depression, oxidative stress, antibody, inflammatory cytokines

1 Introduction

Cancer is a chronic inflammatory disease and one of the leading causes of death worldwide. According to the latest data by WHO (World Health Organization), one in eleven females and one in eight males died due to cancer worldwide in 2022 (World Health Organization, 2020). During the COVID-19 pandemic, due to burden on the healthcare system, many cancer patients suffered delays in treatment, arising from limited access to necessary medical facilities.

At the start of 2020, the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) spread around the world at a very fast pace, and was declared a pandemic by the WHO (Cucinotta and Vanelli, 2020). The pandemic caused a negative impact on social life styles, healthcare systems and economic conditions of individuals worldwide leading to distress, stress and anxiety (Choi et al., 2020; Daly and Robinson, 2022). Infected individuals present mild to severe symptoms, which may include respiratory distress, and in severe cases multi-organ failure or even death. The primary reason for inflammatory changes is the excessive production of cytokines (Sherwani and Khan, 2020). Extended periods of lockdown, social distancing norms and other mandatory protective practices at home and in the workplace impacted levels of physical activity as well as the mental health of individuals, subsequently leading to anxiety and depression among a large population (Trabelsi et al., 2021; van der Werf et al., 2021). Several studies have already been published on depression during unprecedented COVID-19 pandemic (Wan Mohd Yunus et al., 2021; Bueno-Notivol et al., 2021). Depression rates were found to increase in several countries impacting the mental health status of populations (Bueno-Notivol et al., 2021). All age groups were affected due to an overall increase in physiological stress, fear, anxiety and depression (Rodríguez-Hidalgo et al., 2020; Saraswathi et al., 2020).

The burden of a cancer diagnosis itself causes immense psychological and physiological stress in patients, which might have been further aggravated due to the stress and uncertainty of the pandemic too. Increased oxidative stress is found in cancer patients as well as in depressed individuals. Increased generation of reactive species cause non-repairable damage in cells, causing cell death (Hussain et al., 2016). Inflammatory cytokines directly and indirectly play an important role in enhancing oxidative stress and through multiple pathways support tumor cell progression (Kartikasari et al., 2021). Extensive production of reactive species and disease induced redox homeostasis imbalance (Hussain et al., 2016) may cause alterations and/or degradation of biomolecules (DNA, proteins, lipid etc.)

Taking all these into account, this study has been designed to analyze the inflammatory and oxidative stress, as well as antibodies against hydroxyl radical ($\cdot\text{OH}$) treated plasmid-DNA in males with BC and PC patients alone or with depression (BCD and PCD). The estimation of humoral immune status in these cancer patients with or without COVID-related depression might provide an insight into the immune imbalance associated with COVID-19 and better cancer management and palliative care.

2 Materials and methods

Complete and incomplete Freund's adjuvant, Dialysis tubing cellulose membrane pUC18 plasmid DNA, hydrogen peroxide, PBS

solution, catalase, mannitol, EDTA, SOD, IgG-alkaline phosphatase conjugate, p-nitrophenyl phosphate (Aldrich Sigma, United States). Cytokine kits (IL-6, IFN- γ and TNF- α) were purchased from R&D Systems, Minneapolis, MN, United States.

2.1 Free radical modification of plasmid DNA

Plasmid DNA (1 mg/mL) was treated with hydrogen peroxide (10 mM) and the reaction mixture was exposed to UV light at 254 nm for 30–45 min at room temperature under dark conditions. The reaction was carried out in 10 mM PBS solution at pH 7.4. The reaction induced formation of hydroxyl radicals and the reaction mixture was dialyzed against PBS solution to remove unwanted free radicals. The resultant modified plasmid DNA was stored at -80°C for further experiments (Khan et al., 2011).

2.2 Analysis of pDNA modification by spectroscopic studies

UV analysis: Both native and $\cdot\text{OH}$ modified pDNA were screened under UV-spectrophotometer (UV-1700, Shimadzu, Kyoto Japan) to determine change in UV intensities at 260 nm wavelength. Quenching studies using catalase (500 units/mL), mannitol (100 mM), EDTA (100 mM) and SOD (500 units/mL) were performed for the changes in the percent modifications of UV intensities at 260 nm.

2.2.1 Thermal stability

Native and modified pDNA were analysed for their stability. All the samples were treated with varying temperatures ranging from 30°C to 96°C . UV spectrophotometer was connected with a thermal chamber to increase the temperature of the sample at a rate of $1.0^{\circ}\text{C}/\text{min}$. All absorbance was analysed and recorded at 260 nm.

2.2.2 Fluorimetry

Fluorescence intensities were observed for both native and modified pDNA. A wavelength range of 350–600 nm was used to record the fluorescence spectra for both samples on spectrofluorometer (RF-5301, Shimadzu, Tokyo Japan) (Binsaleh et al., 2022).

2.3 Sera samples collection

Nintey sera samples were collected from patients and normal healthy subjects from June 2021–December 2021. The research study was conducted according to the Declaration of Helsinki (1964). Serum samples collected from normal healthy male ($n = 15$) and female ($n = 15$) individuals were from individuals who did not show any signs or symptoms of any disease. Fifteen serum samples were collected for each of the following groups; prostate cancer (PC), breast cancer (BC), prostate cancer with depression (PCD), and breast cancer with depression (BCD). Samples were collected from volunteers with their full written consents as per the approval of the Research Ethics Committee at the University of Ha'il; study protocol H-2021-122. Patients and healthy controls

were checked using similar diagnostic procedures and the disease was confirmed by an oncologist based on the recommended procedures for prostate cancer (e.g., physical examination, antigen tests, and histopathology analysis, etc.) and breast cancer (e.g., physical examination, antigen tests, mammography, histopathology, etc.). All cancer patients included in this study exhibited metastasis at the time of diagnosis.

Demographic (age and gender) and clinical data such as fasting blood glucose (FBG), glycated hemoglobin (HbA1c), basal metabolic index (BMI), C-reactive protein (CRP)) data fasting blood glucose (FBG), glycated hemoglobin (HbA1c), basal metabolic index (BMI), C-reactive protein (CRP) etc. were also collected. FBG, HbA1c, and BMI were analyzed by the recommended prescribed procedures applied in the diagnostic clinics. Levels of CRP were measured using the latex agglutination reaction method.

Volunteers were completed a modified version of a self rating questionnaire to screen the individuals and determine their level of depression (Kroenke and Spitzer, 2002). The depression index score was evaluated as the total score from 20 questions divided by 80 (the maximum possible score). The questionnaire was discussed with a psychiatrist in the College of Medicine.

Serum samples were isolated from blood collected (3–5 mL) from volunteers and serum complement proteins were deactivated by heating the samples at 56°C for half an hour and then stored at –80°C.

Exclusion criteria for the individuals and patients included individuals aged less than 18 years, pregnancy or lactation, patients using antibiotics, alcohol drinkers and smokers. Individuals with any previous history of COVID-19 disease or associated complications were excluded from this study.

2.4 Detection of carbonyl compounds

Amount of protein bound carbonyl was detected in the sera samples of cancer patients and healthy individuals, as described by a published procedure (Levine et al., 1990). Briefly, 100 µL of serum sample was mixed with 400 µL of dinitrophenyl hydrazine (DNPH). Control samples were devoid of DNPH. The mixture was incubated for 1 h at 25°C and then precipitation of DNP-hydrazones was done by adding 500 µL of trichloroacetic acid (4% w/v). The sample was centrifuged for 3–5 min at 13,000 g. To remove the non-reacted DNPH, pellet was dissolved in ethanol-ethylacetate (1:1, v/v) and then centrifuged. The centrifugation was repeated about 3–4 times, and pellet was dissolved in 0.6 mL guanidinium HCl (6M, pH 2.3). To dissolve the hydrazones completely, the sample mixture was frozen at –20°C and thawed. A 200 µL from the aliquot was read at 379 nm using ELISA reader (MR9600-415 Accuris, NJ, United States). The results were evaluated as nanomoles of carbonyl per mg of protein using a $\epsilon_{379\text{ nm}} = 22,000\text{ M}^{-1}\text{cm}^{-1}$.

2.5 Serum malondialdehyde contents

Serum MDA levels were identified using a commercially available ELISA kit (Elabscience, United States). Lipid oxidative stress was assessed by measuring MDA levels in serum samples. Samples were analysed as instructed in the ELISA kit manual. Absorbance of test

samples was measured at 532 nm on a microplate reader. MDA contents were calculated with the following formula:

$$\text{MDA} = \frac{X_1}{X_2} \times C$$

X_1 ; OD of test sample—OD of control. X_2 ; OD of standard—OD of blank. C; Standard concentration (10 nmol/mL).

2.6 Cytokine estimation

Detection and analysis of levels of inflammatory cytokines IL-6, IFN- γ and TNF- α in serum samples of cancer patients and NH subjects, was done using quantitative ELISA (sandwich) (R&D System, Minneapolis, MN, United States), with a sensitivity of less than 0.5 pg/mL for all cytokines. Each sample was assayed in three different wells.

2.7 Antibodies raised against native and •OH-pDNA

Antibodies against native and •OH-pDNA were raised in female rabbits as mentioned previously (Alouffi et al., 2018; Khan et al., 2019a). Briefly, first dose of antigens (50 µg) were emulsified with equal volume of Freund's adjuvant (complete), and injected intramuscularly into the experimental animals. Subsequently, similar amount of antigens (50 µg) mixed with equal volume of Freund's adjuvant (incomplete) were injected into the animals. A dose of about 400 µg antigens was received by each animal. Blood samples were collected regularly, and serum samples were separated as mentioned above and stored at –80°C. Blood samples were also collected before immunization as pre-immune sera to be used as negative controls.

2.8 Serum IgG purification

Pre-immune and immunized experimental animals serum samples were used to isolate purified IgGs using Protein-A Agarose column (Sigma-Aldrich, United States) by using a published procedure (Goding, 1978). A ratio of 1:1 (1 mL) serum sample and phosphate buffer saline (pH 7.4) were applied to Protein-A Agarose column (5 mL). Then column was cleaned 2–3 times with PBS buffer to remove unbound IgG. Acetic acid (0.6%) and NaCl (0.9%) were used to elute the bound IgGs from the column. Eluted samples were neutralized by adding 1 mL of Tris-HCl (1M, pH 8.5). Eluted fractions were analyzed on a spectrophotometer at wavelength of 251 nm and 278 nm, to calculate the concentration of IgG. The optical density of 1.4 at 280 nm is equivalent to 1.0 mg/mL of purified IgG.

2.9 Enzyme-linked immunosorbent assay

Direct binding ELISA was performed to detect the presence of antibodies against native and pDNA in healthy individuals and patients' sera (Sherwani et al., 2022; Sherwani and Khan, 2022). Levels of antibodies against both native and modified pDNA were screened in immunized animals, as described elsewhere (Khan et al., 2020).

TABLE 1 Clinical and demographic presentation of cancer and healthy subjects.

Groups	Gender	Age	Menopausal status (number)	Disease duration	Fasting blood glucose (mg/dL)	HbA1c (%)	BMI (kg/m ²)	Waist-to-hip ratio
	M/F							
NH-M	15	48 ± 12.1	—	—	84.3 ± 7.1	5.34 ± 0.08	23.61 ± 2.87	0.78 ± 0.073
NH-F	15	47 ± 11.7	8	—	85.5 ± 6.3	5.37 ± 0.09	23.21 ± 2.93	0.79 ± 0.076
Prostate cancer	15	49 ± 12.6	—	8 ± 4.1	87.1 ± 8.5	5.53 ± 0.11	24.11 ± 3.04	0.81 ± 0.074
Breast cancer	15	48 ± 13.3	7	7 ± 3.6	86.9 ± 8.5	5.51 ± 0.13	24.91 ± 2.93	0.82 ± 0.079
Prostate cancer with depression	15	58 ± 14.5	—	9 ± 4.5	87.8 ± 8.4	5.55 ± 0.12	25.11 ± 3.09	0.81 ± 0.073
Breast cancer with depression	15	51 ± 16.3	6	8 ± 4.7	88.8 ± 7.6	5.53 ± 0.13	26.71 ± 2.98	0.82 ± 0.078

All the patients included were of metastasis stage at the time of diagnosis. All the patients were on chemotherapy with other medications. Subcategories of any cancer is not defined. Each sample was run in triplicate. All data is given as mean ± standard deviation (SD).

Competition ELISA was also done to evaluate recognition of specific antibodies from healthy control and cancer patient sera samples, as well as immunized antibodies to native and modified pDNA (Khan et al., 2020; Binsaleh et al., 2022; Sherwani et al., 2022). Briefly, both the antigens (pDNA and [•]OH-pDNA) at a concentration of 5 µg/mL were added to the 96 well ELISA plates and kept at room temperature for 3–4 h. ELISA plates were washed with TBS-T (Tris Buffer Saline with Tween 20) and plates were blocked with bovine serum albumin (1%). Immune complexes, formed in the test tubes using sera samples (1:100 dilutions) and varying concentrations (0–10 µg/mL) of antigens, were added and incubated for 4 h and then overnight in a refrigerator at 4°C. These immune complexes (100 µL) were transferred to the ELISA plates and incubated at room temperature for 3–4 h. ELISA plates were washed with TBS-T and anti-human IgG-alkaline phosphate conjugate was added for 1 h and again ELISA plates were washed. Then the substrate p-nitrophenyl phosphate was added, and within 20 min the microplates were read at 410 nm, using a microplate reader (MR9600-415 Accuris, NJ, United States).

2.10 Statistical analysis

Results are presented as mean ± SD. Normality tests and multiple comparison were performed using a software SPSS (16.0, Chicago, United States). Significance was calculated using the Student's t-test. All significances were represented as *p*-value of <0.05.

3 Results

3.1 Clinical characterization

Clinical investigations were carried out for patients with cancer alone or with depression as well (Table 1). There was slight increase in FBG and HbA1c levels in cancer patients compared to their respective healthy controls, however, the changes were not significant. Similarly, BMI and waist-to-hip ratios of cancer patients were found to be slightly higher than the healthy

controls. No significant differences were observed in FBG, HbA1c, BMI and waist-to-hip ratios in patients with cancer alone or with depression.

Furthermore, clinical analyses were conducted for the sera samples of prostate and breast cancer patients for CRP, cytokines levels (IFN-γ, TNF-α and IL-6), and oxidative stress level markers (MDA and carbonyl contents) (Table 2). A significant increase (*p* < 0.001) in the levels of CRP was observed in prostate and breast cancer patients, which was further enhanced with accompanying depression. Also, a substantial increase (*p* < 0.001) in the levels of IL-6 was observed in prostate and breast cancer patients when compared to the healthy controls. However, no respectable difference in the IL-6 levels was observed in patients with cancer alone or cancer with depression. In contrast, there was a significant increase in the levels of IFN-γ and TNF-α in cancer patients with depression compared to the cancer patients (prostate and breast cancer) alone.

Oxidative stress markers were estimated in the serum samples to understand the patients' oxidative stress levels. Prostate and breast cancer patients exhibited significantly higher levels of MDA and carbonyl contents compared to the healthy individuals. The levels of these markers were found to further increase in cancer patients with depression (Table 2).

3.2 Hydroxyl radical modification of pDNA

Native pDNA was modified using hydroxyl radicals in *in vitro* reaction. Modified pDNA was analyzed for structural alterations due to the presence of [•]OH radicals. Native and [•]OH-pDNA were analyzed using spectrophotometer at a wavelength 260 nm. The UV intensity was observed to increase in [•]OH-pDNA as compared to native pDNA (Table 3). This may be due to the exposure of the nitrogenous bases of the pDNA, upon free radical damage.

The stability of the native and modified DNA samples was also subjected to high temperatures (30°C–96°C) and UV intensities of treated samples were observed. Increased fluorescence intensity was recorded at 520 nm after the hydroxyl modification of pDNA (Table 3). This may be due to the breakdown and unwinding of pDNA and exposure of nitrogenous bases.

TABLE 2 Oxidative stress, inflammatory marker, and inflammatory cytokines analysis of cancer patients and healthy subjects.

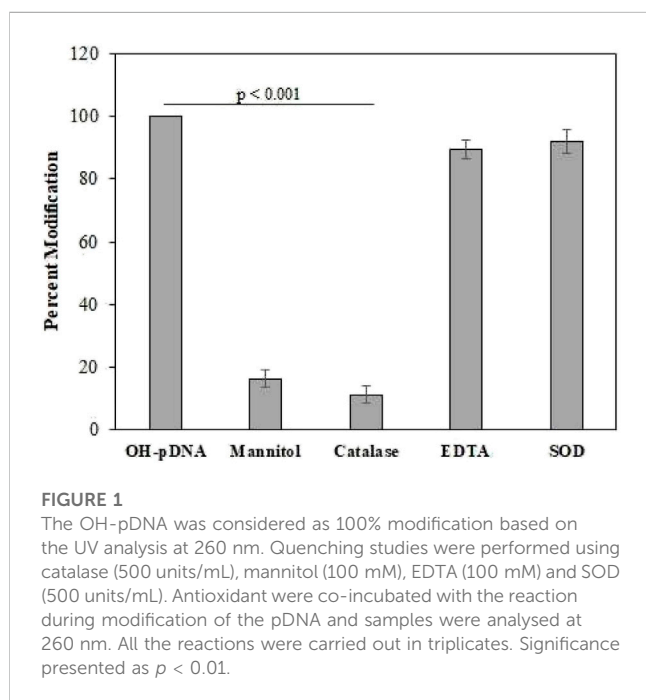
Groups	MDA (nmol/mL)	Carbonyl content (nmol/mg protein)	C-reactive protein (mg/L)	IFN- γ (pg/mL)	TNF- α (pg/mL)	IL-6 (pg/mL)
HC-M ($n = 15$)	2.61 \pm 0.16	0.78 \pm 0.084	0.92 \pm 0.077	3.0 \pm 0.21	0.94 \pm 0.09	1.97 \pm 0.16
HC-F ($n = 15$)	2.58 \pm 0.17	0.93 \pm 0.089	0.93 \pm 0.071	3.2 \pm 0.29	0.90 \pm 0.11	1.99 \pm 0.17
Prostate cancer ($n = 15$)	4.11 \pm 0.12*	2.84 \pm 0.16*	5.01 \pm 0.47***	4.0 \pm 0.32*	1.33 \pm 0.18*	7.3 \pm 0.61***
Breast cancer ($n = 15$)	4.39 \pm 0.21*	2.22 \pm 0.18*	5.24 \pm 0.54***	4.2 \pm 0.29*	1.39 \pm 0.13*	7.1 \pm 0.65***
Prostate cancer with depression ($n = 15$)	5.41 \pm 0.19**	3.64 \pm 0.25**	7.13 \pm 0.66***	4.9 \pm 0.30**	2.73 \pm 0.21***	7.5 \pm 0.62***
Breast cancer with depression ($n = 15$)	5.97 \pm 0.23**	3.92 \pm 0.23**	7.67 \pm 0.62***	5.3 \pm 0.35**	2.99 \pm 0.22***	6.8 \pm 0.59***

Each sample was run in three different wells. Values are presented as mean \pm standard deviation. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Statistical values were evaluated by comparing healthy controls with gender matched patients.

TABLE 3 Characterization of hydroxyl radical modified plasmid DNA.

Groups	Native p-DNA	\bullet OH-pDNA
Absorbance at 260 (37°C)	2.04 \pm 0.09	2.88 \pm 0.014**
Percent hyperchromacity at 260 nm (95°C)	20.9 \pm 0.13	34.7 \pm 0.19**
Fluorescence (Exc. 520 nm)	589 \pm 8.8 AU	837 \pm 11.4***

Significant changes were representing as ** $p < 0.01$; *** $p < 0.001$. Exc. and nm represent excitation and nanometre. For statistical analysis native p-DNA, values were compared with \bullet OH-pDNA, values.



3.3 Quenching studies

Quenching studies were conducted using various antioxidants (mannitol, catalase, and SOD) and a metal chelating agent (EDTA), used in pDNA modification reaction mixtures (Figure 1). After the reactions all the samples were dialyzed against the PBS and the

percent modifications were observed using an Absorbance of 260 nm. The modified sample, devoid of any quenching agent, was considered as positive control (100% modification). Results showed highest decrease in the modification by catalase and mannitol, both specific antioxidants for \bullet OH and H_2O_2 molecules. This shows the involvement of H_2O_2 and \bullet OH in pDNA modification.

3.4 Antigenicity of pDNA and \bullet OH modified pDNA

Antigens pDNA and OH-pDNA were administered to female rabbits for up to 8 weeks (details are provided in the methods section). Blood samples were collected before the administration of antigens (pre-immune sera) and afterwards at different intervals. Immunoglobulin G was isolated from pre-immune and immune rabbit serum samples using Protein A-Agarose column. Purity was assessed for different IgG samples on 7% SDS-polyacrylamide using gel electrophoresis (data not shown).

Direct binding ELISA was used to assess the antibody production against respective antigens. IgGs raised against modified p-DNA, exhibited significantly high titre ($>1:25,600$) (Figure 2A). IgGs raised against native p-DNA also showed high titre ($\sim 1:12,800$) (Figure 2B). However, antibody production against the \bullet OH-pDNA showed almost two-fold higher titre than native pDNA in immune sera samples of experimental animals. Pre-immune sera IgG exhibited low number of antibodies against both antigens.

Induced antibodies were isolated as mentioned above and specificities were analyzed against their immunogens using

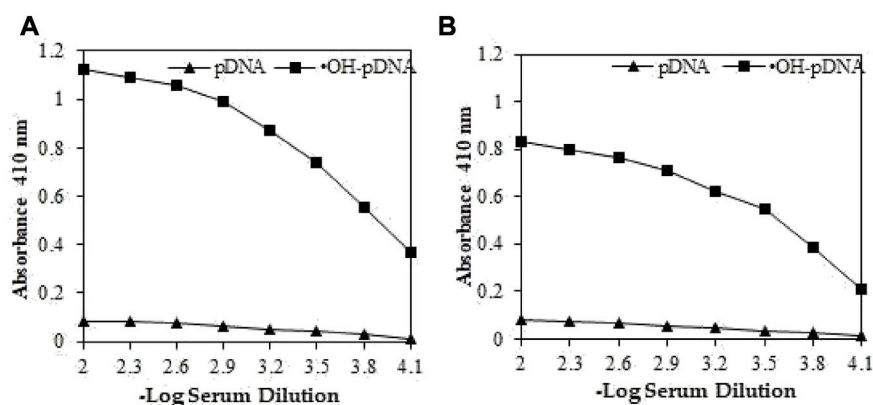


FIGURE 2

Direct binding ELISA of OH-pDNA (A) and native pDNA (B) from preimmune (▲) and immune sera (■) sera samples from immunized animals. Microtitre plates were coated with 5 µg/mL of respective antigens.

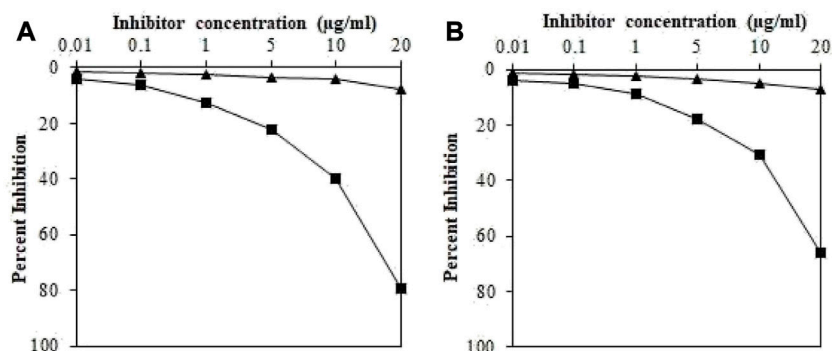


FIGURE 3

Inhibition ELISA of IgGs raised against OH-pDNA (A) and pDNA (B) (■) and IgGs from pre-immune sera (▲) samples. The 96-well plates were coated with pDNA and OH-pDNA (5 µg/mL). All the samples were run in triplicates.

inhibition ELISA. The maximum percent inhibition (MPI) for IgGs against \bullet OH-pDNA was determined to be 79.3% at 20 µg/mL (Figure 3A), whereas, IgGs against pDNA showed 66.1% (Figure 3B). Moreover, IgGs against \bullet OH-pDNA showed 50% inhibition at 12.1 µg/mL. However, IgGs against pDNA exhibited 50% inhibition at 17.7 µg/mL of the antigen.

These results indicate high production of antibodies (IgGs) against \bullet OH-pDNA in experimental animals as compared to native pDNA. Raised antibodies were also highly specific for the modified antigen as compared to the native antigen.

3.5 Detection of serum antibodies against \bullet OH-pDNA

The sera samples (1:100 dilutions) of all subjects were analyzed for the presence of antibodies (IgG) against native and \bullet OH-pDNA antigens (Figure 4). Direct binding ELISA was used to screen for the

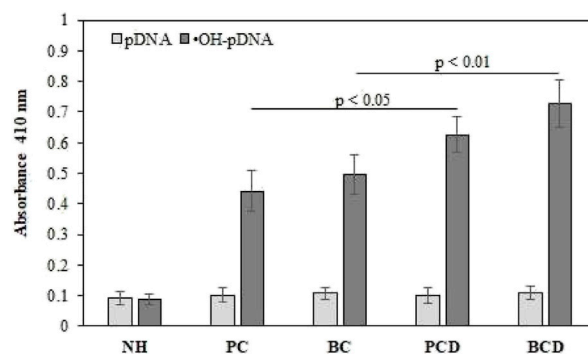
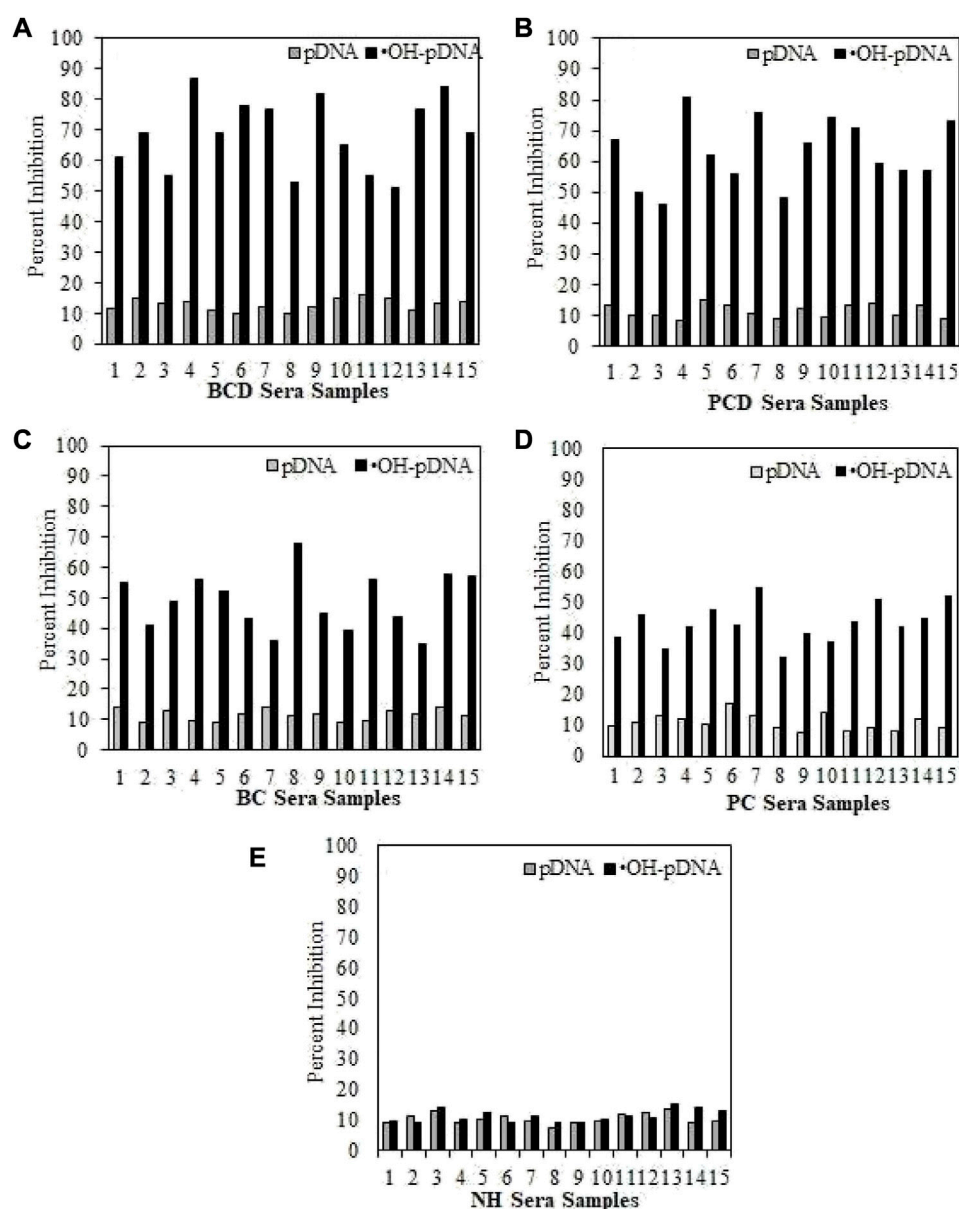


FIGURE 4

Direct binding ELISA of serum antibodies against native and \bullet OH-pDNA antigen. ELISA plates were coated with the respective antigen (5 µg/mL). All the samples were tested in triplicates. Significance was represent as $p < 0.05$ and $p < 0.01$.

**FIGURE 5**

Inhibition ELISA for cancer and normal human subjects against native and modified antigens (10 $\mu\text{g/mL}$). maximum percent inhibition at 10 $\mu\text{g/mL}$ were assessed for serum samples from BCD (A), PCD (B), BC (C), PC (D) and NH serve as controls (E). Each sample was tested in triplicates and the values given are in mean \pm SD.

levels of the antibodies against $\bullet\text{OH-pDNA}$ (anti- $\bullet\text{OH-pDNA}$ antibodies). Direct binding data showed highest binding of serum antibodies against modified antigen in breast cancer patients with depression (OD; 0.698 ± 0.078), followed by prostate cancer patients with depression (0.636 ± 0.058), breast cancer patient alone (0.506 ± 0.063) or prostate cancer patients (0.441 ± 0.066) alone. Normal human subjects did not show respectable binding against either of the antigens.

Serum antibody-antigen interaction specificities were tested by inhibition ELISA. In inhibition ELISA, serum samples (1:100 dilutions) were co-incubated with varying concentrations of antigens (0–10 $\mu\text{g/mL}$) for 4 h at room temperature and

overnight at 4°C. Then the reaction mixture was added to the antigen coated wells. The maximum percent of inhibition at 10 $\mu\text{g/mL}$ for both antigens are presented in **Figure 5** for all the samples. Inhibition against modified antigen yielded highest percent inhibition in breast cancer patients with depression ($68.8\% \pm 7.8\%$) (**Figure 5A**), followed by prostate cancer patients with depression ($62.9\% \pm 8.3\%$) (**Figure 5B**), breast cancer patients ($48.9\% \pm 8.1\%$) (**Figure 5C**), and prostate cancer patients ($43.4\% \pm 7.5\%$) (**Figure 5D**). Native antigen showed negligible inhibition in all the groups. Normal human subjects did not exhibit significant inhibition with any of the antigens (**Figure 5E**).

4 Discussion

Cancer disease has multifactorial etiology with concomitant involvement of several molecular components. Breast and prostate cancers are major cancer types in female and males, respectively. Cancer patients have faced several challenges during the unprecedented COVID-19 pandemic. Some of the major issues included non-availability of timely and specialized medical healthcare, limited availability of prescribed medications, and delays in follow-up procedures, etc. COVID-19 related depression further added to complications in such patients. Unnatural stress environment, which all the human beings were subject to during extended lockdown, social distancing, work from home policies extensively affected the mental health of individuals. These conditions induced anxiety and depression in large populations (Trabelsi et al., 2021; van der Werf et al., 2021). Depression was found to increase many-fold during the pandemic period, with associated negative impacts on cancer patients (Hussain et al., 2016; Bueno-Notivol et al., 2021).

Several factors may contribute to the development of cancers including lifestyle factors leading to hormone level imbalances, family history, exposure to carcinogens and pollutants, altered redox homeostasis, etc. (Anand et al., 2008). Cancer cells disrupt signaling pathways to alter oxidative metabolism (Schiliro and Firestein, 2021). Thus, there are increased levels of free radicals. Excessive free radicals have the potential to induce nucleic acid, protein and lipid degradation as well as post-translational modifications, leading to tumorigenesis (Wang et al., 2017). Serum samples of cancer patients in this study exhibited significantly high levels of oxidative stress markers such as MDA and carbonyl contents.

Cancer is a dangerous disease and patients are extremely fearful about its diagnosis and prognosis, causing high levels of distress (Smith, 2015). Long durations of anxiety and psychological distress in cancer patients lead to depression, with subsequent higher rates of mortality (Colleoni et al., 2000; Pinquart and Duberstein, 2010). One study showed various levels of depression can increase mortality rate by up to 39% (Satin et al., 2009). Considering these findings, we have designed this study, with two major types of cancer patients, i.e., breast and prostate alone or along with depression, for which clinical and immunological investigations were carried out to determine the impact and role of depression in such patients.

Cytokines play an important role in imbalance of oxidative levels and promote free radical production, which can alter DNA structure and lead to DNA mutations (Kartikasari et al., 2021). Furthermore, inflammatory cytokines are directly involved in epigenetic (Yasmin et al., 2015) and post-translational modifications (Louis and Bohjanen, 2017). In our results, cancer patients with depression exhibited higher levels of inflammatory cytokines (IFN- γ , TNF- α , and IL-6), as compared to the cancer patients without any depression. Other clinical factors (FBG, HbA1c, and BMI) were also analyzed, however, no remarkable changes were observed between patients and the healthy controls. An important inflammatory marker, CRP, was also found at significantly elevated levels in BC, PC, BCD, and PCD compared to NH. However, the levels of CRP were higher in breast and prostate cancer patients with depression as compared to BC and PC patients, respectively.

Plasmid DNA was modified using hydroxyl radical, and the structural alterations were observed using various biophysical characterizations (spectroscopy and fluorometry). Significant changes were found after the \bullet OH modification of pDNA, as compared to the native form.

Modified and native antigens were introduced into experimental animals to check their antigenicity. \bullet OH-pDNA exhibited a twofold higher antibody titre after eight doses of the antigen as compared to native pDNA. Antibody raised against \bullet OH-pDNA was found to be more specific to its antigen and gave 50% inhibition at lower concentrations as compared to its native form. These results showed \bullet OH-pDNA is more immunogenic and produced higher amounts of antibodies in animal model.

Previous studies showed the presence of autoantibodies against various antigens in cancer patients with depression (Khan et al., 2019b; Khan and Khan, 2020). In this study pDNA and \bullet OH-pDNA were used as antigens and cancer patients' sera samples were screened for the presence of antibodies against these antigens. Direct binding ELISA showed, antigen \bullet OH-pDNA was recognized by higher levels of antibodies in cancer patients. Highest levels of anti- \bullet OH-pDNA Abs were detected in breast cancer patients with depression followed by prostate cancer patients with depression. Comparatively, breast cancer and prostate cancer patients showed significantly lower levels of antibodies against the modified antigen. However, this level was still higher than the level of antibodies in normal human subjects. Serum anti- \bullet OH-pDNA Abs were tested for their specificity using inhibition ELISA, which further ascertained direct binding results. Highest levels of percent inhibition were detected in BCD followed by PCD.

The findings from our study indicate that COVID-19 related depression in male (Prostate cancer) and female (breast cancer) cancer patients caused increased levels of oxidative stress, resulting in alterations in nucleic acid molecules. Additionally, higher levels of proinflammatory cytokines in depressed cancer patients further exaggerated destruction due to molecular alterations. Modified antigen showed high recognition of serum antibodies compared to the native antigen. Cancer patients with depression showed the higher levels of anti- \bullet OH-pDNA Abs compared to the cancer patients without depression.

5 Conclusion

Cancer management has been extremely challenging during COVID-19 pandemic as there have been a lack of medical facilities and oncologists or physicians as well as limited availability of prescribed medications. Additionally, COVID-19 induced depression in a large population, including cancer patients. Increased serum levels of oxidative stress markers, inflammatory cytokines and an inflammatory marker were detected in BCD and PCD as compared to BC and PC patients. High levels of serum antibodies against \bullet OH-pDNA were detected in BCD and PCD as compared to BC and PC patients. Furthermore, inhibition ELISA also substantiated results of binding studies for the presence of antibodies against hydroxyl radical modified pDNA in BCD and PCD, as compared to BC and PC subjects. Thus, elevated levels of oxidative stress and inflammatory conditions, concomitant

with depression in cancer patients cause increased generation of circulatory antibodies against autoantigens, which could lead to disease progression and development of additional complications. Thus, it is high time to address the impact of COVID-19 pandemic related depression on the mental health of cancer patients in an effort to provide focused medical care, treatment options and better overall disease management.

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.

Ethics statement

The studies involving human participants were reviewed and approved by Research Ethics Committee at the University of Ha'il; study protocol H-2021-122. The patients/participants provided their written informed consent to participate in this study. The animal study was reviewed and approved by Research Ethics Committee at the University of Ha'il; study protocol H-2021-122.

Author contributions

NB and SS designed the study. NB, SS, RE, and SR conducted the clinical and immunological part of the study. HQ, AB, MA, and MH conducted the modification and characterization of the antigens.

References

- Alouffi, S., Sherwani, S., Al-Mogbel, M. S., Sherwani, M., and Ali Khan, M. (2018). Depression and smoking augment the production of circulating autoantibodies against glycated HSA in rheumatoid arthritis patients. *Int. Arch. Allergy Immunol.* 177 (2), 1–11. doi:10.1159/000489896
- Anand, P., Kunnumakara, A. B., Sundaram, C., Harikumar, K. B., Tharakan, S. T., Lai, O. S., et al. (2008). Cancer is a preventable disease that requires major lifestyle changes. *Pharm. Res.* 25, 2097–2116. doi:10.1007/s11095-008-9661-9
- Binsaleh, N. K., Eltayeb, R., Qanash, H., Aziz, M. A., Albaradie, R., and Khan, M. W. A. (2022). Presence of circulatory autoantibodies against ROS-modified histone H1 protein in lymphoma patients. *Front. Genet.* 13, 909903. doi:10.3389/fgene.2022.909903
- Bueno-Notivol, J., Gracia-García, P., Olaya, B., Lasheras, I., López-Antón, R., and Santabàrbara, J. (2021). Prevalence of depression during the COVID-19 outbreak: A meta-analysis of community-based studies. *Int. J. Clin. Health Psychol.* 21 (1), 100196. doi:10.1016/j.ijchp.2020.07.007
- Choi, E. P., Hui, B. P., and Wan, E. Y. (2020). Depression and anxiety in Hong Kong during COVID-19. *Int. J. Environ. Res. Public Health.* 17 (10), 3740. doi:10.3390/ijerph17103740
- Colleoni, M., Mandala, M., Peruzzotti, G., Robertson, C., Bredart, A., and Goldhirsch, A. (2000). Depression and degree of acceptance of adjuvant cytotoxic drugs. *Lancet* 356, 1326–1327. doi:10.1016/s0140-6736(00)02821-x
- Cucinotta, D., and Vanelli, M. (2020). WHO declares COVID-19 a pandemic. *Acta Biomed.* 91 (1), 157–160. doi:10.23750/abm.v91i1.9397
- Daly, M., and Robinson, E. (2022). Depression and anxiety during COVID-19. *Lancet* 399 (10324), 518. doi:10.1016/s0140-6736(22)00187-8
- Goding, J. W. (1978). Use of staphylococcal protein A as an immunological reagent. *J. Immunol. Methods.* 20, 241–253. doi:10.1016/0022-1759(78)90259-4
- Hussain, T., Tan, B., Yin, Y., Blachier, F., Tossou, M. C. B., and Rahu, N. (2016). Oxidative stress and inflammation: What polyphenols can do for us? *Oxid. Med. Cell. Longev.* 2016, 1–9. doi:10.1155/2016/7432797
- Kartikasari, A. E., Huertas, C. S., Mitchell, A., and Plebanski, M. (2021). Tumor-induced inflammatory cytokines and the emerging diagnostic devices for cancer detection and prognosis. *Front. Oncol.* 11, 692142. doi:10.3389/fonc.2021.692142
- Khan, M. W. A., Banga, K., Mashal, S. N., and Khan, W. A. (2011). Detection of autoantibodies against reactive oxygen species modified glutamic acid decarboxylase-65 in type 1 diabetes associated complications. *BMC Immunol.* 12, 19. doi:10.1186/1471-2172-12-19
- Khan, M. W. A., Otaibi, A. A., Sherwani, S., Khan, W. A., Alshammari, E. M., Saleem, M., et al. (2020). Glycation and oxidative stress increase autoantibodies in the elderly. *Molecules* 25, 3675. doi:10.3390/molecules25163675
- Khan, W. A., and Khan, M. W. A. (2020). Depression augments the production of high-affinity antibodies against estrogen metabolite-receptor complex in prostate cancer patients. *J. Mens. Health.* 16 (4), e72–e83. doi:10.31083/jomh.v16i4.268
- Khan, W. A., Khan, M. W. A., Sherwani, S., and Siddiqui, W. A. (2019a). Depression enhanced the production of autoantibodies against 16 α -hydroxyestrone-estrogen receptor adduct in breast cancer. *Int. Immunopharmacol.* 66, 251–259. doi:10.1016/j.intimp.2018.11.018
- Khan, W. A., Zaman, G. S., Alouffi, S., and Khan, M. W. A. (2019b). Depression and its related parameters increased the production of autoantibodies against 16 α -hydroxyestrone-albumin complex in systemic lupus erythematosus. *Int. Immunopharmacol.* 71, 215–223. doi:10.1016/j.intimp.2019.03.036
- Kroenke, K., and Spitzer, R. L. (2002). The PHQ-9: A new depression diagnostic and severity measure. *Psychiatr. Ann.* 32, 509–515. doi:10.3928/0048-5713-20020901-06
- Levine, L. R., Garland, D., Oliver, N. C., Amici, A., Climent, I., Lenz, A. G., et al. (1990). Determination of carbonyl content in oxidatively modified proteins. *Methods Enzymol.* 186, 464–478. doi:10.1016/0076-6879(90)86141-h
- Louis, I. V.-S., and Bohjanen, P. R. (2017). Post-transcriptional regulation of cytokine and growth factor signaling in cancer. *Cytokine Growth Factor Rev.* 33, 83–93. doi:10.1016/j.cytogr.2016.11.004

Statistical analysis done by NB. HQ, AB, MA, and MH reviewed and edit the manuscript.

Funding

This research has been funded by the Scientific Research Deanship at University of Ha'il—Saudi Arabia [project number (RG-21111)].

Acknowledgments

All authors thank the Scientific Research Deanship at University of Ha'il—Saudi Arabia [project number (RG-21111)].

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.

- Pinquart, M., and Duberstein, P. R. (2010). Depression and cancer mortality: A meta-analysis. *Psychol. Med.* 40, 1797–1810. doi:10.1017/s0033291709992285
- Rodriguez-Hidalgo, A. J., Pantaleón, Y., Dios, I., and Falla, D. (2020). Fear of COVID-19, stress, and anxiety in university undergraduate students: A predictive model for depression. *Front. Psychol.* 11, 591797. doi:10.3389/fpsyg.2020.591797
- Saraswathi, I., Saikarthik, J., Senthil, K. K., Madhan Srinivasan, K., Ardhanaari, M., and Gunapriya, R. (2020). Impact of COVID-19 outbreak on the mental health status of undergraduate medical students in a COVID-19 treating medical college: A prospective longitudinal study. *PeerJ* 8, e10164. doi:10.7717/peerj.10164
- Satin, J. R., Linden, W., and Phillips, M. J. (2009). Depression as a predictor of disease progression and mortality in cancer patients: A meta-analysis. *Cancer* 115, 5349–5361. doi:10.1002/cncr.24561
- Schilero, C., and Firestein, B. L. (2021). Mechanisms of metabolic reprogramming in cancer cells supporting enhanced growth and proliferation. *Cells* 10 (5), 1056. doi:10.3390/cells10051056
- Sherwani, S., and Khan, M. W. A. (2022). Autoantibodies against ROS-human serum albumin-A potent immunological marker in depressed individuals with smoking history. *Rev. Romana Med. Lab.* 30 (4), 399–411. doi:10.2478/rrlm-2022-0039
- Sherwani, S., and Khan, M. W. A. (2020). Cytokine response in SARS-CoV-2 infection in the elderly. *J. Inflamm. Res.* 13, 737–747. doi:10.2147/jir.s276091
- Sherwani, S., Khan, M. W. A., Mallik, A., Saleem, M., Raafat, M., Shati, A. A., et al. (2022). Seroprevalence of anti-S1-RBD antibodies in pre-pandemic and pandemic subjects from Hail region, KSA. *Front. Public Health.* 10, 874741. doi:10.3389/fpubh.2022.874741
- Smith, H. R. (2015). Depression in cancer patients: Pathogenesis, implications and treatment (Review). *Oncol. Lett.* 9 (4), 1509–1514. doi:10.3892/ol.2015.2944
- Trabelsi, K., Ammar, A., Masmoudi, L., Boukhris, O., Chtourou, H., Bouaziz, B., et al. (2021). Globally altered sleep patterns and physical activity levels by confinement in 5056 individuals: ECLB COVID-19 international online survey. *Biol. Sport* 38 (4), 495–506. doi:10.5114/biolsport.2021.101605
- van der Werf, E. T., Busch, M., Jong, M. C., and Hoenders, H. J. R. (2021). Lifestyle changes during the first wave of the COVID-19 pandemic: A cross-sectional survey in The Netherlands. *BMC Public Health* 21, 1226. doi:10.1186/s12889-021-11264-z
- Wan Mohd Yunus, W. M., Badri, S. K., Panatik, S. A., and Mukhtar, F. (2021). The unprecedented movement control order (lockdown) and factors associated with the negative emotional symptoms, happiness, and work-life balance of Malaysian University students during the coronavirus disease (COVID-19) pandemic. *Front. psychiatry* 11, 566221. doi:10.3389/fpsyg.2020.566221
- Wang, S., Wang, N., Zheng, Y., Zhang, J., Zhang, F., and Wang, Z. (2017). Caveolin-1: An oxidative stress-related target for cancer prevention. *Oxid. Med. Cell Longev.* 2017, 7454031. doi:10.1155/2017/7454031
- World Health Organization (2020). Cancer. Available at: <https://www.who.int/news-room/fact-sheets/detail/cancer> (Accessed March 22, 2023).
- Yasmin, R., Siraj, S., Hassan, A., Khan, A. R., Abbasi, R., and Ahmad, N. (2015). Epigenetic regulation of inflammatory cytokines and associated genes in human malignancies. *Mediat. Inflamm.* 2015 (1), 1–8. doi:10.1155/2015/201703



OPEN ACCESS

EDITED BY

Khurshid Ahmad,
Yeungnam University, Republic of Korea

REVIEWED BY

F. A. Dain Md Opo,
King Abdulaziz University, Saudi Arabia
Abdul Malik,
King Saud University, Saudi Arabia

*CORRESPONDENCE

Mohammed Bourhia,
✉ bourhiamohammed@gmail.com
Rohit Sharma,
✉ rohitsharma@bhu.ac.in

RECEIVED 24 February 2023

ACCEPTED 10 April 2023

PUBLISHED 21 April 2023

CITATION

Ali S, Ali U, Qamar A, Zafar I, Yaqoob M, Ain Qu, Rashid S, Sharma R, Nafidi H-A, Bin Jordan YA and Bourhia M (2023), Predicting the effects of rare genetic variants on oncogenic signaling pathways: A computational analysis of HRAS protein function. *Front. Chem.* 11:1173624. doi: 10.3389/fchem.2023.1173624

COPYRIGHT

© 2023 Ali, Ali, Qamar, Zafar, Yaqoob, Ain, Rashid, Sharma, Nafidi, Bin Jordan and Bourhia. 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.

Predicting the effects of rare genetic variants on oncogenic signaling pathways: A computational analysis of HRAS protein function

Sadaqat Ali¹, Usman Ali², Adeem Qamar³, Imran Zafar⁴, Muhammad Yaqoob⁵, Qurat ul Ain⁶, Summya Rashid⁴, Rohit Sharma^{7*}, Hiba-Allah Nafidi⁸, Yousef A. Bin Jordan⁹ and Mohammed Bourhia^{10*}

¹Medical Department, DHQ Hospital Bhawalnagr, Punjab, Pakistan, ²Basic Health Unit, Punjab, Pakistan, ³Department of Pathology, Sahiwal Medical College Sahiwal, Punjab, Pakistan, ⁴Department of Bioinformatics and Computational Biology, Virtual University of Pakistan, Punjab, Pakistan, ⁵Department of Life Sciences, ARID University-Barani Institute of Sciences Burewala Campus, Punjab, Pakistan, ⁶Department of Chemistry, Government College Women University, Faisalabad, Pakistan, ⁷Department of Rasa Shastra and Bhaishajya Kalpana, Faculty of Ayurveda, Institute of Medical Sciences, Banaras Hindu University, Varanasi, Uttar Pradesh, India, ⁸Department of Food Science, Faculty of Agricultural and Food Sciences, Laval University, Quebec City, QC, Canada, ⁹Department of Pharmaceutics, College of Pharmacy, King Saud University, Riyadh, Saudi Arabia, ¹⁰Laboratory of Chemistry and Biochemistry, Faculty of Medicine and Pharmacy, Ibn Zohr University, Agadir, Morocco

The HRAS gene plays a crucial role in regulating essential cellular processes for life, and this gene's misregulation is linked to the development of various types of cancers. Nonsynonymous single nucleotide polymorphisms (nsSNPs) within the coding region of HRAS can cause detrimental mutations that disrupt wild-type protein function. In the current investigation, we have employed *in-silico* methodologies to anticipate the consequences of infrequent genetic variations on the functional properties of the HRAS protein. We have discovered a total of 50 nsSNPs, of which 23 were located in the exon region of the HRAS gene and denoting that they were expected to cause harm or be deleterious. Out of these 23, 10 nsSNPs ([G60V], [G60D], [R123P], [D38H], [I46T], [G115R], [R123G], [P110L], [A59L], and [G13R]) were identified as having the most deleterious effect based on results of SIFT analysis and PolyPhen2 scores ranging from 0.53 to 69. The DDG values -3.21 kcal/mol to 0.87 kcal/mol represent the free energy change associated with protein stability upon mutation. Interestingly, we identified that the three mutations (Y4C, T58I, and Y12E) were found to improve the structural stability of the protein. We performed molecular dynamics (MD) simulations to investigate the structural and dynamic effects of HRAS mutations. Our results showed that the stable model of HRAS had a significantly lower energy value of -18756 kJ/mol compared to the initial model of -108915 kJ/mol. The RMSD value for the wild-type complex was 4.40 Å, and the binding energies for the G60V, G60D, and D38H mutants were -107.09 kcal/mol, -109.42 kcal/mol, and -107.18 kcal/mol, respectively as compared to wild-type HRAS protein had -105.85 kcal/mol. The result of our investigation presents convincing corroboration for the potential functional significance of nsSNPs in augmenting HRAS expression and adding to the activation of malignant oncogenic signalling pathways.

KEYWORDS

cancer, genetic variants, mutations, oncogenic signaling pathways, SNPs, nsSNP, HRAS

1 Introduction

The HRAS gene is identified as the Harvey rat sarcoma viral oncogene homolog and is responsible for encoding a GTPase protein of small stature that belongs to the RAS family. A plethora of cellular processes, comprising proliferation, differentiation, and survival, are subject to regulation by this particular intracellular signalling pathway (Rajalingam et al., 2007). Genetic alterations in RAS genes, notably in HRAS, are among the most common mutations detected in human cancers (Kawazu et al., 2013). The aberrant functioning of the HRAS protein instigates the activation of downstream signalling pathways, namely, MAPK/ERK and PI3K/AKT, which are critical in promoting cell proliferation and survival. Notably, these pathways are often disrupted in cancer, thereby underscoring their significant contribution to the disease's pathogenesis (De Luca et al., 2012; Asati et al., 2016). Several studies have confirmed the involvement of HRAS mutations in various cancers, including bladder, colon, head and neck, lung, and thyroid cancers (Ngan et al., 2022). For instance, activating HRAS mutations have been found in up to 10% of thyroid cancers, linked with aggressive disease and poor prognosis (Garcia-Rostan et al., 2003).

Similarly, in bladder cancer, HRAS mutations have been detected in 1%–2% of cases and are associated with high-grade tumours and advanced disease (Nagata et al., 2016). Identifying and characterizing HRAS mutations are vital for cancer diagnosis and treatment. HRAS mutations may serve as biomarkers for cancer diagnosis or prognosis or as targets for cancer therapies to inhibit RAS signalling (Kompier et al., 2010). Research on HRAS and its role in cancer remains an active area of investigation, with ongoing efforts to identify new mutations and decipher their functional consequences.

The HRAS gene and its protein product have significant roles in tumour genesis by regulating fundamental cellular processes such as growth, differentiation, and viability. In its wild-type cellular context, the expression and activity of HRAS are under the precise control of several signaling pathways, especially the RAS-MAPK axis, which is vital for cellular differentiation and growth. However, genetic mutations in the HRAS gene can disrupt the balance of the RAS-MAPK pathway, leading to unregulated cellular proliferation and neoplastic transformation (Rezatabar et al., 2019; Ullah et al., 2022). Such HRAS mutations have been consistently observed in several types of cancer, including squamous cell carcinoma of the head and neck, bladder cancer, and thyroid carcinoma (Jefferies and Foulkes, 2001; Gilardi et al., 2020). Moreover, the involvement of HRAS in cancer is not restricted to the RAS-MAPK pathway alone, as it extensively interacts with other pivotal signaling pathways and cellular processes, including but not limited to the PI3K-AKT pathway, Wnt signaling pathway and cytoskeleton. These intricate interactions facilitate the promotion of a multitude of oncogenic processes, such as cell survival, invasion, and metastasis. Therefore, the role of HRAS in cancer is multifaceted and cannot be limited to a single pathway or mechanism (Rezatabar et al., 2019; Shorning et al.,

2020). The intricate protein-protein interaction network between HRAS and its downstream effectors in cancer cells comprises several signaling molecules, kinases, and transcription factors that are essential for malignant transformation and disease progression. A comprehensive understanding of this network can provide crucial insights into the mechanisms of HRAS-driven cancer and facilitate the development of innovative therapeutic interventions targeted at HRAS and its downstream effectors (Khan et al., 2020; Odeniyide et al., 2022).

Proteins exhibit nsSNPs which can result in alterations in the amino acid sequence. It is well-established that such modifications have been linked to the initiation and advancement of cancer (Masoodi et al., 2013; Wang et al., 2019). These variations can arise in genes involved in cell growth regulation, such as oncogenes or tumour suppressor genes. They may impede normal cellular processes such as cell division and programmed cell death, associated with malignant transformation. One of the prototypical oncogenes affected by nsSNPs is the RAS gene family, encompassing HRAS, which encodes small GTPases involved in signalling pathways that control cellular proliferation, differentiation, and survival (Khan and Bisen, 2013; Makrides et al., 2017). Mutations in RAS genes, including nsSNPs, can activate these pathways and subvert normal cellular regulation, thereby instigating tumorigenesis. HRAS regulates several essential cellular processes, including cell differentiation, division, and programmed cell death. Mutations in HRAS have been observed in various cancers, including bladder cancer, pancreatic cancer, and lung cancer. nsSNPs within the coding region of HRAS may yield deleterious mutations that impair the normal function of the HRAS protein (Backwell and Marsh, 2022). Identifying and characterising nsSNPs in cancer-associated genes, including HRAS, is critical for comprehending the mechanisms underlying cancer pathogenesis and devising personalized cancer therapies (Pang, 2018). Computational tools can aid in predicting the impact of nsSNPs on protein function and can facilitate a deeper understanding of the contribution of these genetic variations to cancer onset and progression (Jubb et al., 2017; Ahmad et al., 2022).

The main objective of our research is identifying and characterising nsSNPs in the HRAS gene and their potential impact on the structure and function of the HRAS protein. Specifically, we explore *in silico* approaches to predict the effects of rare genetic variants on HRAS protein function, identify the most deleterious nsSNPs, and investigate the consequences of these mutations on the stability, flexibility, and compaction of the HRAS protein using molecular dynamics simulations. In this study, we also analyze the binding energies of the wild-type and mutant HRAS protein with docked complexes to understand the potential impact of these mutations on the activation of oncogenic signalling pathways. Our research provides compelling evidence for the potential functional role of nsSNPs in up-regulating HRAS expression and contributing to the development of various types of cancers.

2 Material and methods

2.1 Collecting and preparing SNP data

The ensuing discourse delineates the origins and manipulation of Single Nucleotide Polymorphism (SNP) data garnered from sundry databases like dbSNP (<https://www.ncbi.nlm.nih.gov/snp/>), ENSEMBLE (<https://ensemblgenomes.org/>), SNP500 cancer (<https://pubmed.ncbi.nlm.nih.gov/>), GeneCards (<https://www.genecards.org/>), and UniPort (<https://www.uniprot.org/>), which are periodically refreshed with new information. In particular, the ENSEMBLE repository was availed to obtain the nucleotide and protein sequences germane to the HRAS gene as per the methods of earlier researchers (Buljan et al., 2018; Zafar et al., 2022). This undertaking holds the promise of research prospects for exploring genetic variations and their potential implications (Rajaram et al., 2001).

2.2 Prediction of deleterious nsSNPs

The SIFT tool (<https://sift.bii.a-star.edu.sg/>) was employed to forecast the impact of non-synonymous SNPs on the mutant protein (Bromberg and Rost, 2007). This technique bifurcated the SNPs into two categories, intolerant and tolerant, based on homologous alignment (Seal et al., 2014). Precisely, amino acids with normalization probabilities falling beneath the designated threshold value were ascertained as intolerant, whilst those with a tolerance index measuring over >0.05 were considered tolerant (Dakal et al., 2017). The implications of this approach hold potential research prospects in assessing genetic variations and their resultant phenotypic outcomes (Pauls et al., 2013).

2.3 Structural homology-based approach: Coding of nsSNPs

The PolyPhen2 tool (<http://genetics.bwh.harvard.edu/pph2/>) was leveraged to predict the pernicious ramifications of nsSNPs on the structural and functional aspects of proteins. This prediction was based on the naive Bayesian algorithm, which involves classifying scores from 0 to 1. Mutations were partitioned into three categories, depending on the scores, with those possessing a score closest to 1 being identified as probably damaging and exhibiting a significant impact on protein structure. Implementing this tool could potentially open up research opportunities in the realm of genetic variability and its influence on protein conformation and function.

2.4 Categorization of functional nsSNPs

Identifying functional nsSNPs was executed with the aid of online servers, including SNP&GO (<https://snps-and-go/>), PhD-SNP (<https://snps.biofold.org/>), PROVEAN (<https://bio.tools/provean>), PANTHER (<http://www.pantherdb.org/>), and P-Mut (<https://bio.tools/pmut>). In particular, PhD-SNP relied on support vector systems to classify and depict the effects of non-

synonymous SNPs on proteins. This method divided nsSNPs into deleterious or neutral categories (Li et al., 2006; Pauls et al., 2013). Furthermore, the ROVEAN tool was utilized to identify damaging SNPs by classifying mutations into deleterious or neutral depending on a threshold score of -2.5 . At the same time, SNP&GO relied on the support vector machine algorithm (Sharma et al., 2022). Lastly, P-MUT employed the neural networking algorithm to segregate mutants into disease or neutral based on probability statistics of the sequences (Shinwari et al., 2022). These techniques could potentially unlock research opportunities in the sphere of genetic variations and their ramifications on protein function and structure (Singh et al., 2007).

2.5 Identification of nsSNPs on the coding area of protein

The SNP SnpEff and SnpSift (<http://pcingola.github.io/SnpEff/>) tool box was implemented to foretell the impact of nsSNPs on the coding region of the HRAS protein as per earlier researchers (Hossain et al., 2020). Apart from exhibiting the outcomes in conservation scores, the program also identified the protein homeostasis landscape (Han et al., 2020). The SNP effect used several software and tools to discover the propensity, including the aggregation tendency of the mutant via TANGO (<https://switchlab.org/software/>), amyloid propensity through WALTZ (<https://switchlab.org/software/>), and chaperone binding with LIMBO as per investigation of the earlier researcher (De Baets et al., 2012). These sophisticated methodologies can foster new research avenues for exploring the complex interplay between genetic variations and protein homeostasis.

2.6 Influence of non-synonymous mutations on protein stability (INPS)

The I-MUTANT 3.0 suit (https://bio.tools/i-mutant_suite) was utilized to prognosticate the impact of mutations on the stability of the HRAS protein, with the MUpPro program (<https://mupro.proteomics.ics.uci.edu/>) being employed to validate the outcomes (Venkata Subbiah et al., 2020). Both servers rely on the same algorithm and are designed to evaluate the influence of mutations on protein stability, whether it is enhancing or diminishing it (Dehouck et al., 2011). Furthermore, the SRide server (<http://sride.enzim.hu/>) was leveraged to pinpoint the stabilizing residues of the native and mutant proteins (Kotha, 2010). The integration of these advanced computational techniques holds great promise for driving further research in protein stability and its response to genetic variations.

2.7 Conservational analysis of HRAS protein

The ConSurf server (https://consurf.tau.ac.il/consurf_index.php) was deployed to thoroughly analyze each amino acid's conservation and evolutionary aspects in the HRAS protein (Kumar et al., 2021). The outcomes were presented as diverse conservation scores, ranging from 1 to 9. Scores from 1 to

3 corresponding to variable positions, while scores from 4 to 6 indicated amino acid positions that were moderately conserved. Furthermore, scores falling within the range of 7–9 represented highly conserved positions of amino acids. Utilizing this cutting-edge technology can pave the way for further protein evolution and conservation research, thereby expanding our knowledge of the complex interplay between genetics and protein function (Gourbal et al., 2018).

2.8 Molecular docking

We investigated the effects of mutations on the structure and function of the HRAS protein employing molecular docking analysis (Hossain et al., 2020). To achieve this, they acquired the three-dimensional configuration of the HRAS protein (PDB ID: 6MQT) from the Protein Data Bank. They transferred it into the MOE software (<https://www.chemcomp.com/Products.htm>) as per the method of the earlier researcher (Bhattacharya et al., 2017).

For the docking process, we excluded heteroatoms, ligands, and aqueous molecules from the structure. We conducted structural refinement employing definite parameters, such as energy minimization concerning 0.1 gradients, addition of hydrogen atoms, and utilization of the MMFF94X force field. We recognized an active site within the protein, encompassing a critical area of interacting residues.

Utilizing the MOE software, we performed molecular docking simulations for both standard and mutated molecules, storing the results in mdb format for further analysis (Ahmad et al., 2015). The highest-ranking postures underwent further refinement and calculation of binding free energies (ΔG) by employing the scoring function (GBVI/WSA dg). The scoring function is grounded on several molecular interactions, such as pi, hydrogen, and hydrophobic interactions. It presents a dependable scoring method that yields the docking score of the correct binding postures (Jin et al., 2023).

We meticulously surveyed the docked complex's MOE database to understand the mode of binding interactions of the wild-type and mutated complex (Niranjan et al., 2021). This exploration enabled the research team to identify possible impacts of mutations on the protein's structure and function, providing insightful discoveries for further investigations in the field.

2.9 Molecular dynamics (MD) simulation

We employed the Schrodinger 2021.2 software suite for their computational investigations (Kutzner et al., 2022). The initial structures were drawn using Maestro 12.8 and subsequently ionized with Epik 3.2 program at a pH of 7.4 using Ligprep 3.4 (<https://www.schrodinger.com/products/ligprep>) as per the analysis of (Santana-Romo et al., 2020). This protocol was carried out to produce the requisite starting structures for molecular dynamics (MD) simulations.

We perform calculations to produce MD simulation trajectories across various intervals during the simulation run (Galindo-Murillo et al., 2015). After docking, the conformational study of three complexes was executed with the MacroModel 10.8 module (Kellici et al., 2019). This module employs a torsional sampling

approach for all conformational search: a Monte Carlo multiple minimum method (Li and Scheraga, 1987). The highest energy conformers were removed using a 21 kJ mol⁻¹ energy limit. Each conformer was reduced for a maximum of 2,000 steps using the Polak-Ribiere conjugate gradient technique, with a gradient convergence threshold of 0.001 kJ mol⁻¹ Å⁻¹, and the OPLS3 force field was utilized for this process. The OPLS3 force field is advantageous for small molecules because it delivers accurate energy minimization potential functions (Harder et al., 2016).

The team employed MD simulation to appraise the stability of the enzyme-inhibitor complex and explore the conformational aspects of protein-ligand interactions (Amaral et al., 2017). The conformational variations and stability index of secondary structural components of the simulated complexes were assessed utilizing data reduction techniques such as root-mean-square deviation (RMSD), root-mean-square fluctuation (RMSF), the radius of gyration (Rg), and beta-factor values as per earlier researcher (Kumar et al., 2019). We employed a suite of computational software to perform molecular dynamics simulations and scrutinize the conformational aspects of protein-ligand interactions. They used diverse techniques to appraise the stability and conformational variations of the simulated complexes.

3 Results and discussion

3.1 Compilation of a single nucleotide polymorphism (SNP) library

We aimed to construct a comprehensive SNP database for the HRAS gene, a human gene that encodes for the protein HRAS (Hossain et al., 2020). The team utilized various bioinformatics tools and databases to achieve this goal, including the Ensemble genome browser, Gene card, Uniprot, and NCBI db-SNP (Phillips, 2009). These resources provide access to various genetic and protein-related information, which the team could use to identify and analyze SNPs in the HRAS gene (Kohl et al., 2015). Our analysis was on non-synonymous SNPs (nsSNPs), which are genetic variations that change the amino acid sequence of the protein encoded by the HRAS gene (Chai et al., 2022). These variations are more likely to impact the structure and function of the HRAS protein, potentially leading to changes in protein activity and, in turn, contributing to disease development.

We specifically looked for nsSNPs located within the HRAS gene's exon region, which is the coding region translated into protein (Tarek et al., 2021). They identified the location and nature of each nsSNP and determined its frequency in different populations. However, during their analysis, we observed that many SNPs were located in the intron region of the HRAS gene (Estep et al., 2006). Intron regions are non-coding regions of the gene and traditionally were thought to be non-functional. However, recent studies have suggested that intronic SNPs could also have a role in regulating gene expression and splicing and may contribute to disease susceptibility (Meyer et al., 2008). Therefore, we also analyzed the potential impact of these intronic SNPs on HRAS gene expression and splicing as per earlier investigations (Vornholt et al., 2021). The distribution of SNPs in a different region of human HRAS gene is represented in Figure 1.

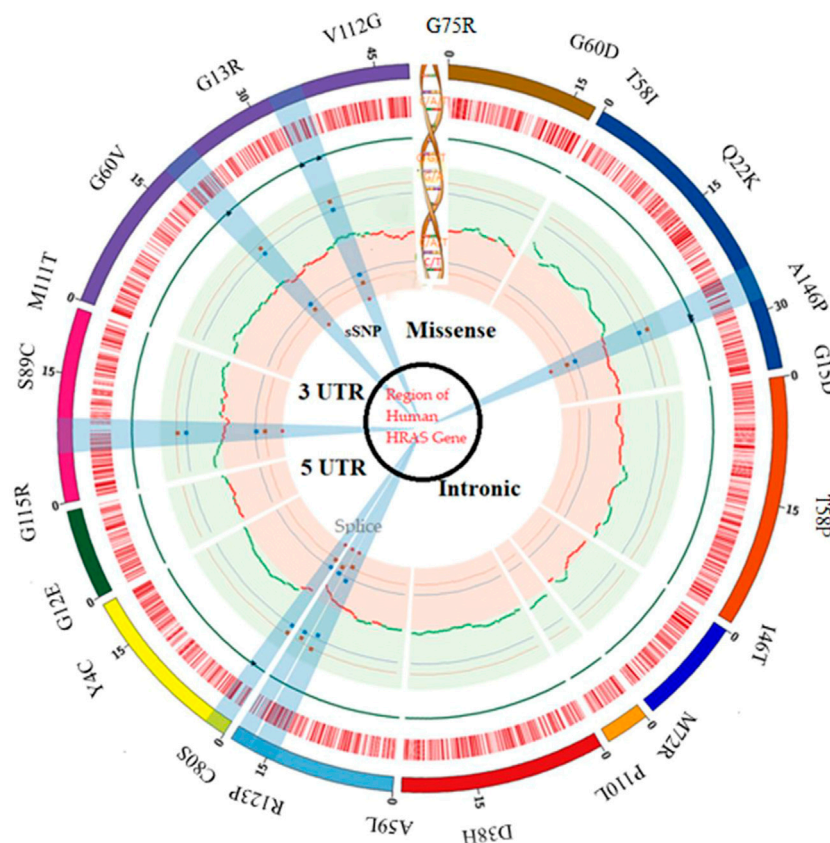


FIGURE 1

SNP distribution in a different region of the human HRAS gene.

Our work involved using various bioinformatics tools and databases to map a comprehensive SNP dataset for the HRAS gene, focusing on non-synonymous SNPs in the exon region (Rivera et al., 2005). This dataset could provide valuable insights into the genetic variations contributing to disease development and help develop personalized medicine approaches (Hamburg and Collins, 2010).

3.2 Evaluation of non-synonymous (nsSNP) SNPs

We use the SIFT (Sorting Intolerant from Tolerant) algorithm to predict the impact of non-synonymous SNPs (nsSNPs) on the structure of the HRAS protein (Chai et al., 2022). The SIFT algorithm uses homologous protein sequences to determine which amino acid substitutions are likely to be tolerated and which are likely deleterious (Ng and Henikoff, 2001). The algorithm's output is a tolerance index (TI) score, ranging from 0 to 1, with lower scores indicating a greater likelihood of harmful consequences. We submitted 50 nsSNPs to the SIFT algorithm for analysis (Savas et al., 2004). The of 23 SNPs were identified with TI scores ranging from 0 to 0.04, with 28 having a TI score of 0, indicating highly deleterious consequences. This means that these

nsSNPs are likely to impact the structure and function of the HRAS protein significantly and the result matched with earlier research (Chai et al., 2022). Multiple SNPs had a TI score of 0.01, one had a score of 0.02, and one had a score of 0.004, suggesting only minor importance.

We noted a high frequency of substitutions involving cytosine and thymine or guanine and adenine, while substitutions involving adenine and thymine or adenine and cytosine were rare (Kim et al., 2019). This information could provide important insights into the mechanisms by which these nsSNPs affect HRAS protein function and may be helpful in developing targeted therapies for diseases caused by HRAS mutations as per the investigation of the earlier researcher (Chai et al., 2022). Overall, the use of the SIFT algorithm provides valuable information for understanding the impact of nsSNPs on the HRAS protein and could potentially contribute to the development of new treatments for diseases caused by HRAS mutations (Hossain et al., 2020).

3.3 Identification of functional modifications in coding nsSNPs

We describe the results of further analysis of 50 selected nsSNPs submitted to a server and utilized the

TABLE 1 Screening of most deleterious SNPs by using different software.

No.	Variant ID	Alleles	Amino acid changes	SIFT	Polyphen2	Provean deleterious	P-mut disease
1	rs756190012	C/T	G75R	0.03	1	-7.893	0.69
2	rs730880460	C/A/T	G60V	0	1	-6.760	0.66
3	rs121917758	G/A	T58I	0	1	D -5.798	0.66
4	rs104894231	C/G/T	A146P	0	1	-4.427	0.53
5	rs730880460	C/A/T	G60D	0	1	-8.689	0.69
6	rs770492627	T/G	T58P	0	0.999	-5.797	0.53
7	rs755488418	A/C	M72R	0	0.999	-5.873	0.66
8	rs755322824	G/C	S89C	0	0.999	-4.711	0.64
9	rs764622691	T/C	Y4C	0	0.999	-6.949	0.66
10	rs730880464	C/G	R123P	0	0.998	-5.560	0.69
11	rs750680771	C/G/T	D38H	0	0.998	-6.461	0.66
12	rs1564789700	A/G	I46T	0	0.998	-4.596	0.64
13	rs1554885139	C/T	G15D	0	0.996	-5.631	0.63
14	rs1564789063	A/G	M111T	0	0.995	-4.376	0.64
15	rs121917757	G/A/T	Q22K	0	0.993	-3.341	0.62
16	rs917210997	C/T	G115R	0	0.993	-7.193	0.69
17	rs1370566417	A/T	C80S	0	0.992	-7.008	0.64
18	rs1204223913	G/A	G60V	0	0.99	-9.167	0.69
19	rs727504747	GC/AG	A59L	0	0.988	-4.756	0.69
20	rs1427823770	A/C	V112G	0	0.97	-6.172	0.67
21	rs104894228	C/A/G	G13R	0	0.975	-6.676	0.66
22	rs898057728	G/A/S	S65R	0	0.947	-4.536	0.67
23	rs727503094	GC/AG	G12E	0.01	0.942	-6.114	0.66

PolyPhen2 algorithm to predict the impact of nsSNPs on protein structure. The algorithm provides a score between 0 and 1, with a higher score indicating a greater likelihood of deleterious consequences. Out of the 50 nsSNPs submitted, only 10 had a probabilistic score greater than 0.97, indicating that they were probably damaging nsSNPs. The total of 10 nsSNPs had scores higher than 0.83, which were classified as possibly damaging. We exactly identified seven nsSNPs (G75R, P34S, G60D, G60V, T58I, G60D, and A146P) with a maximum score of 1, indicating a high probability of being damaging. Most of the remaining mutations had scores in the range of 0.98 to -0.99. We also compared the results obtained from SIFT and PolyPhen2 and observed that ten nsSNPs were identified as common between the SIFT and PolyPhen2 analyses, despite using different methods to obtain results. The SIFT algorithm utilizes structural detail to yield results, while PolyPhen2 is based on structure and has shown a good correlation with the SIFT approach. Finally, we observed that most nsSNPs had a SIFT tolerance index of 0.00, indicating that they were highly intolerant to variation, while their PolyPhen2 scores were >0.90 and based on these

findings, we concluded that these alterations may be responsible for disease.

3.4 Phenotypic impact of mutations

The SNP-effect tool assesses the potential effects of genetic variants on protein structure and function. We also evaluate the phenotypic impact of alterations within the HRAS molecule. Specifically, the approach evaluated chaperone binding propensity, aggregation propensity, and amyloid tendency. Chaperones are a class of proteins that assist in properly folding and assembly of other proteins. A protein's chaperone binding propensity can provide insight into its stability and folding efficiency. Aggregation propensity refers to the tendency of proteins to form aggregates or clumps, which can interfere with proper cellular function. The amyloid tendency measures a protein's ability to form amyloid fibrils, which are associated with several diseases, including Alzheimer's and Parkinson's and Cancer. The results of the SNP-effect analysis showed that the selected mutations had little to no impact on chaperone binding propensity,

TABLE 2 Non-synonymous single nucleotide polymorphisms (nsSNPs) and their predicted effects on protein function.

Variant ID	Position	Wild type	Mutant	DDG value
rs730880460	60	G	D	−2.19
rs121917758	58	T	I	0.28
rs730880460	60	G	V	−1.22
rs764622691	4	Y	C	0.62
rs730880464	123	R	P	−1.13
rs750680771	38	D	H	−1.57
rs1564789700	46	I	T	−3.21
rs1554885139	15	G	D	−0.4
rs917210997	115	G	R	−0.98
rs369106578	123	R	G	−1.01
rs1204223913	110	P	L	−0.87
rs727504747	59	A	L	−0.32
rs104894228	13	G	R	−1.27
rs1564789552	64	Y	H	−0.94
rs727503094	12	G	E	0.87

aggregation propensity, or amyloid tendency. However, it is essential to note that this approach only evaluates a limited set of characteristics and does not comprehensively assess all possible effects. The findings suggest that the selected mutations may have other potential impact on the structure and properties of the HRAS protein, which could have significant implications for biological organisms. Further experimental studies would be necessary to fully understand the impact of these mutations and their potential role in disease. Screening of most deleterious SNPs using different software's are mentioned in [Table 1](#); the HRAS gene encodes HRAS protein and plays an essential role in various cellular processes, including cell proliferation, differentiation, and survival. SNPs in the HRAS gene have been associated with multiple diseases, including cancer, and can impact protein structure and function. Therefore, identifying deleterious SNPs in the HRAS gene is crucial for understanding disease mechanisms and developing potential treatments.

Screening potentially deleterious SNPs in the HRAS gene is essential to understanding the possible impact of genetic variations on the protein's function ([Chai et al., 2022](#)). Several software tools have been developed to analyze the potential effect of SNPs on protein structure and function ([Paniri et al., 2021](#)). We applied SIFT, PolyPhen2, and SNP-effect to predict the impact of nsSNPs on the HRAS protein. The SIFT algorithm is a sequence-based tool that utilizes sequence homology to predict the potential effect of nsSNPs on protein function. SIFT analysis identified 23nsSNPs with TI scores ranging from 0.53 to 69, suggesting highly deleterious consequences. PolyPhen2 is a tool that combines sequence-based and structure-based predictions to predict the effect of nsSNPs on protein structure and function. We found that G75R, P34S, G60D, A59L, G60V, I46T, D38H, T58I, and A146P had a maximum score

of 1, while most mutations had scores in the range of 0.98 to −0.99. A total of 10 nsSNPs were identified as standard between the SIFT and PolyPhen2 analyses, despite using different methods to obtain results.

The SNP-effect tool is a computational pipeline that assesses the effect of SNPs on protein properties, such as chaperone binding propensity, aggregation propensity, and amyloid tendency ([Ji et al., 2021](#)). We found that the selected alternate variants did not significantly impact these characteristics. However, they suggested that variations in protein structure and properties resulting from these SNPs could still significantly impact biological organisms. In summary, the use of multiple software tools to predict the effects of nsSNPs on the HRAS protein provides a more comprehensive understanding of the potential consequences of genetic variations, where earlier researchers ([Hossain et al., 2020](#); [Chai et al., 2022](#)) also explore and indicate exact predictions. The SIFT and PolyPhen2 analyses identified several nsSNPs with a high likelihood of deleterious consequences. In contrast, the SNP-effect study showed that the selected nsSNPs did not significantly affect the protein's chaperone binding, aggregation propensity, or amyloid tendency. These findings could have implications for understanding the role of HRAS mutations in developing various diseases.

3.5 Effect of mutation on stability of HRAS

The stability of a protein is crucial for its proper function, and destabilizing mutations can lead to the misfolding and aggregation of the protein, resulting in various diseases ([Gómez et al., 2018](#)). We explored the effect of different mutations on the stability of the HRAS protein and assessed using I-MUTANT and MUPRO for final results and validations. The Delta Gibbs free energy (DDG) values were calculated to determine the stability of the protein, and a DDG value lower than 0 indicated a destabilizing mutation. The I-MUTANT (<https://folding.biofold.org/i-mutant/i-mutant2.0.html>) server predicted that most modifications can decrease the stability of the HRAS protein, with DDG values ranging from −3.21 kcal/mol to 0.87 kcal/mol. The most significant effect was observed with the I46T mutation, which had a DDG value of −3.21 kcal/mol. Only three mutations, Y4C, T58I, and Y12E, showed perfection in the structural stability of the protein. The MUPRO server (<https://mupro.proteomics.ics.uci.edu/>) provided similar results, except for the Y4C, G12E, and T58I mutations, which showed a decrease in stability. The G15D mutation showed an increase in strength in the MUPRO prediction, while I-MUTANT predicted a reduction in stability and the results are summarized in [Table 2](#).

The differences in predictions between the two servers can be attributed to their different calculation methods. I-MUTANT uses a support vector regression algorithm based on various structural properties of the protein, such as solvent accessibility and secondary structure, to predict the DDG values. In contrast, MUPRO employs a neural network-based approach that incorporates sequence and structural information, as well as evolutionary conservation, to predict the effects of mutations on protein stability. The results of this study indicate that several mutations within the HRAS gene can destabilise the protein, potentially resulting in disease.

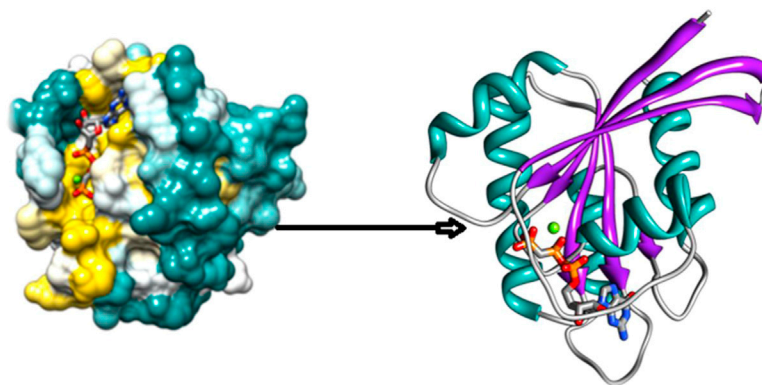


FIGURE 2
3D Structure of HRAS protein.

Identifying these destabilizing mutations can provide insights into the molecular mechanisms of HRAS-associated diseases and may aid in developing new therapeutic strategies.

Our results indicated that most of the mutations decreased the stability of the HRAS protein, while only a few improved it. The mutation with the most significant effect on stability was I46T, with a DDG value of -3.21 kcal/mol. The I-MUTANT server predicted that 15 mutations, including Y4C, G12E, G13R, G15D, D38H, I46T, T58I, A59L, G75R, P34S, G60D, G60V, G115R, R123G, and R123P, decreased the stability of the protein, while Y4C, T58I, and Y12E improved it. The MUpro server provided similar results, except for Y4C, G12E, and T58I mutations, which showed decreased stability, contrary to the I-MUTANT prediction. Additionally, MUpro predicted I-MUTANT predicted a decrease in increased stability for the G15D mutation, and I-MUTANT predicted decreased stability. The DDG values of most mutations ranged from -3.21 kcal/mol to 0.87 kcal/mol, indicating reduced protein stability with a DDG value lower than 0. The results suggest that mutations can significantly affect the HRAS protein's stability, which may impact the biological organism. Further research may be required to understand the specific effects of each mutation on HRAS protein stability and its overall impact on biological systems.

3.6 Conservation analysis

The ConSurf server (https://consurf.tau.ac.il/consurf_index.php) is a valuable tool for determining the evolutionary conservation of protein residues across a set of homologous sequences. The conservation scores of HRAS protein residues were analyzed using ConSurf to evaluate the impact of the 10 deleterious mutations on the protein structure and function. Out of the 10 deleterious mutations, 6 missense mutations (G13R, D38H, A59L, G60V, G60D, G115R, R123P, and R123G) were located in highly conserved regions (7–8–9). This finding suggests that mutations in these regions could significantly affect the function and structure of the HRAS protein. Additionally, G13R, D38H, and I46T were predicted to be exposed, while mutants such as A59L, G60V, G60D,

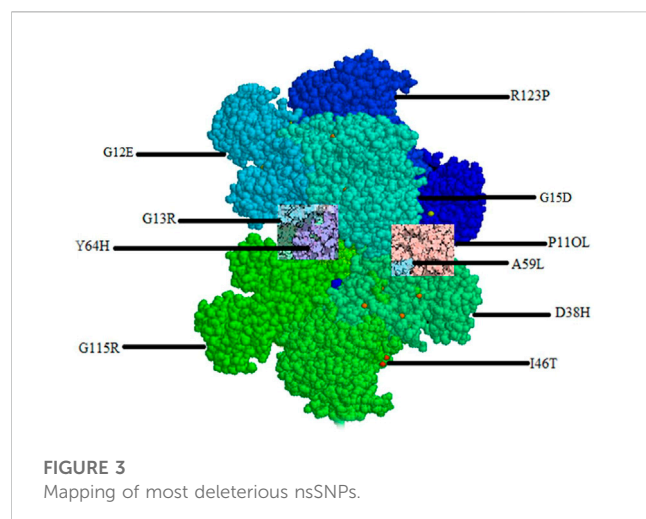
Q60D, R123P, and R123G were expected to be functional and revealed mutations.

The conservation analysis showed that one mutation was located in a variable region (1–2–3) and one in an average part, indicating that these mutations may have a milder effect on the protein structure and function. These findings suggest that highly conserved regions of HRAS protein are more sensitive to mutations that could impact the protein function. These results provide insights into the functional and structural effects of the 10 selected deleterious mutations on the HRAS protein. By identifying the regions that are highly conserved and sensitive to mutations, this study can help researchers better understand the consequences of HRAS mutations and may lead to new treatments for diseases associated with HRAS mutations.

3.7 3D structures

The Protein Data Bank (<https://www.rcsb.org/>) provides an extensive collection of experimentally determined protein structures that can be used for structural analysis (Protein Data Bank, 2019). In this study, we obtained the wild-type entire structure of the HRAS protein with its PDB ID: (6MQT) from the PDB. The protein structure was analyzed to identify features such as active sites, protein-protein interface sites, domain motifs, and ligand-binding affinities. The three-dimensional structure of HRAS was visualized in Figure 2, where the protein's helices, beta-sheets, and coils were represented by yellow, cyan, and green colors, respectively. The structure of HRAS showed that it consists of five alpha-helices and six beta-strands, which are arranged in a characteristic fold called the G-domain as per earlier researcher investigations (Korzeniecki and Priefer, 2021). The G-domain contains the nucleotide-binding site, which is responsible for the hydrolysis of GTP to GDP, and plays a crucial role in regulating HRAS activity (Liu et al., 2019).

The protein structure was further analyzed to identify potential active sites and protein-protein interface sites, which can be targeted for drug design as per the results of researchers (Velazquez et al., 2018; Lin et al., 2020). In addition, the ligand-binding affinities of



HRAS were predicted to identify potential small molecule inhibitors that can be used to target the protein in various diseases. The three-dimensional structure of HRAS provides essential insights into this protein's function and regulation. It can be used to guide the design of new therapeutics for the treatment of HRAS-related diseases, as shown in an earlier exploration by [Ahmad et al. \(2022\)](#).

3.8 Mapping of most deleterious nsSNPs on HRAS gene

The mapping of the 10 most deleterious nsSNPs on the HRAS protein structure using mutagenesis techniques in Pymol software (as shown in [Figure 3](#)) provides a visual representation of the location and distribution of these mutations on the protein. The mutations were distributed throughout the protein structure, with several mutations located in the Group-I (G12E, G13R, G15D) and Group-II (Q61L, A59L) regions, which are known to play a crucial role in HRAS activation. Mutations in these regions can potentially lead to impaired GTP hydrolysis, affecting the normal functioning of HRAS. Moreover, several mutations were located in or near the protein's active site (D38H, G60D, G60V), which could interfere with HRAS's ability to interact with its downstream effectors and may impair its biological functions. The mutagenesis techniques used in Pymol also showed that some of the mutations (I46T, D38H, A59L, G60V) could form hydrogen bonds with neighbouring residues, suggesting a possible alteration of the protein's conformation and potential effects on protein stability. Mapping these deleterious mutations in HRAS protein structure provides insights into how these mutations could potentially impact the protein's structure and function, providing a foundation for further experimental investigation.

3.9 Structural analysis of HRAS protein

The native complete structure of HRAS was retrieved from the Protein Data Bank, and the Swiss Model server (<https://swissmodel.expasy.org/>) was used to predict the mutated form using homology modelling approaches. The ten most deleterious mutations,

indicated by all the analyzing tools, were mapped in their respective region of HRAS using mutagenesis techniques in Pymol software, as shown in [Figure 4](#). The mutated models were generated to further investigate the effect of these mutations on the HRAS protein structure, and energy minimization was carried out using Schrödinger. The energy minimization process minimized the energy and force acting on each atom in a gathering of atoms to obtain the most thermodynamically stable HRAS structure. The final and stable model of HRAS had an energy value of $-18,756$ kJ/mol, which was significantly lower than the energy value of the initial model, which was -108915 kJ/mol. This indicates that the mutated models had a more stable conformation than the initial models, after the energy minimisation process.

The deleterious mutated models of HRAS, including rs730880460 (G60V), rs730880460 (G60D), rs730880464 (R123P), rs750680771 (D38H), rs1564789700 (I46T), rs917210997 (G115R), rs369106578 (R123G), rs1204223913 (P110L), rs727504747 (A59L), and rs104894228 (G13R), were generated using Pymol software and compared with the native HRAS protein structure. The mutations were observed to cause structural changes in different regions of HRAS, and the energy minimization process helped achieve more stable configurations. The results suggest that these mutations could potentially affect the function of HRAS and contribute to cancer development.

Each mutation can have different effects on the structure and function of the protein. The specific effects of each mutation can depend on a variety of factors, such as the location of the mutation in the protein, the surrounding amino acid residues, and the protein's function as per the investigation of the earlier researcher ([Cain et al., 2020](#)). For example, the G60V mutation, predicted to be deleterious in our analysis, is located in the Group-II region of the protein and can cause problems in protein folding due to the larger size of the mutant residue compared to the wild-type residue. This can prevent the mutant residue from fitting correctly in the core region of the protein, potentially leading to the destabilization of the protein structure.

Similarly, the G115R mutation, also predicted to be deleterious, can lead to an incorrect conformation and disturbance of the local structure of the protein due to the larger size of the mutant residue, as a result, compare with ([Wang S et al., 2022](#)), and this result in loss of protein function. The G13R mutation located in the G-domain in the G-domain can cause loss of interaction because the mutant residue is more minor and has a different hydrophobicity compared to the wild-type residue as per earlier researchers ([O'Bryan, 2019](#)). This can affect the interaction between HRAS and its downstream effectors proteins, potentially leading to downstream signalling defects. The D38H mutation can cause loss of interaction or repulsion due to changes in charge, as the positively charged histidine positively charged histidine replaces the negatively charged aspartic acid. This can affect the interaction of HRAS with its upstream activators, potentially leading to downstream signaling defects.

3.10 Molecular docking of HRAS

Molecular docking is a computational method used to predict the binding mode and affinity of small molecules to a protein as per earlier researchers ([Luo et al., 2019](#)). In this study, molecular

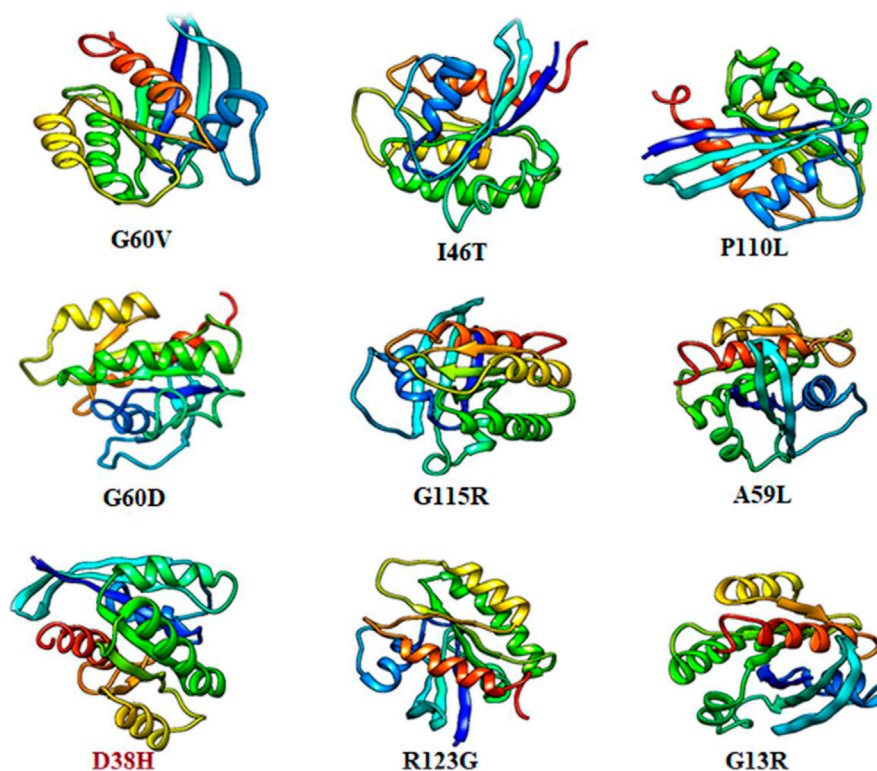


FIGURE 4
Mutant structure of HRAS protein.

docking was applied to understand the effect of three deleterious mutations, G60V, G60D, and D38H, on the binding pocket of the HRAS protein. The HRAS protein structure with PDB ID: 6MQT was imported into MOE software (<https://www.chemcomp.com/Products.htm>), and the docked complexes of the wild-type and mutated protein with ligands were generated. The docked complexes were analyzed for their docked score, hydrogen bonds, and pi-interactions within the 4.5 Å. The active binding site residues GLU62, GLY10, THR58, ASP33, ILE36, GLU63, TYR64, ALA11, TYR96, GLY13, and LYS16 were identified to be involved in pi-interactions and generate hydrogen bonds. The docked complexes showed significant binding affinity and interactions with the active binding site residues.

When the three deleterious mutations, G60V, G60D, and D38H, were docked into the same binding pocket, the D38H residue was found to be involved in the binding interactions. With G60V, three hydrogen bonds were generated by the D38H residue, while one hydrogen bond with D38H and two with G60D were observed. This suggests that the D38H residue plays an important role in stabilizing the conformation of the mutated HRAS protein and results were compared with the investigation of the earlier researcher (Chai et al., 2022). The docked complexes were further subjected to MD simulation to analyze the stability and conformation of the wild-type and mutated complexes as shown in Figure 5. The MD simulation analysis revealed that the wild-type and mutated complexes were stable during the simulation. The RMSD and RMSF values were calculated, and it was observed that the

mutated complexes had higher RMSD and RMSF values compared to the wild-type complex. This indicates that the mutated complexes had a slightly different conformation than the wild-type complex.

The -protein residues are shown as sticks, and the ligands are shown in space-filling mode. The wild-type protein and the three mutated proteins are shown in different colors. The two-dimensional plot of interacting target residues shows the residues of the protein that are protein residues involved in important interactions with the ligands. The residues involved in hydrogen bonding, Pi, and hydrophobic interactions are shown as circles in different colors. The Non-mutant protein is shown in blue color, and the residues involved in interactions with the ligands are labelled with their residue numbers. The plot in Figure 5 shows that the residues involved in interactions with the ligands are distributed throughout the protein's active site. The residues involved in hydrogen bonding are mainly located in the loops and helices of the protein. The residues involved in Pi interactions are primarily situated in the helices and strands of the protein. The hydrophobic residues are mainly located in the core of the protein as mentioned in Figures 5A-D. For the G60V mutation, the plot shows that the mutation affects the hydrogen bonding with residue D38H and G60D mutation; the plot shows that the mutation affects the hydrogen bonding with residue D38H and generates two new hydrogen bonds with residue G60D. In D38H mutation, the plot shows that the mutation affects the hydrophobic interactions with the residues in the core of the protein. The plot provides valuable

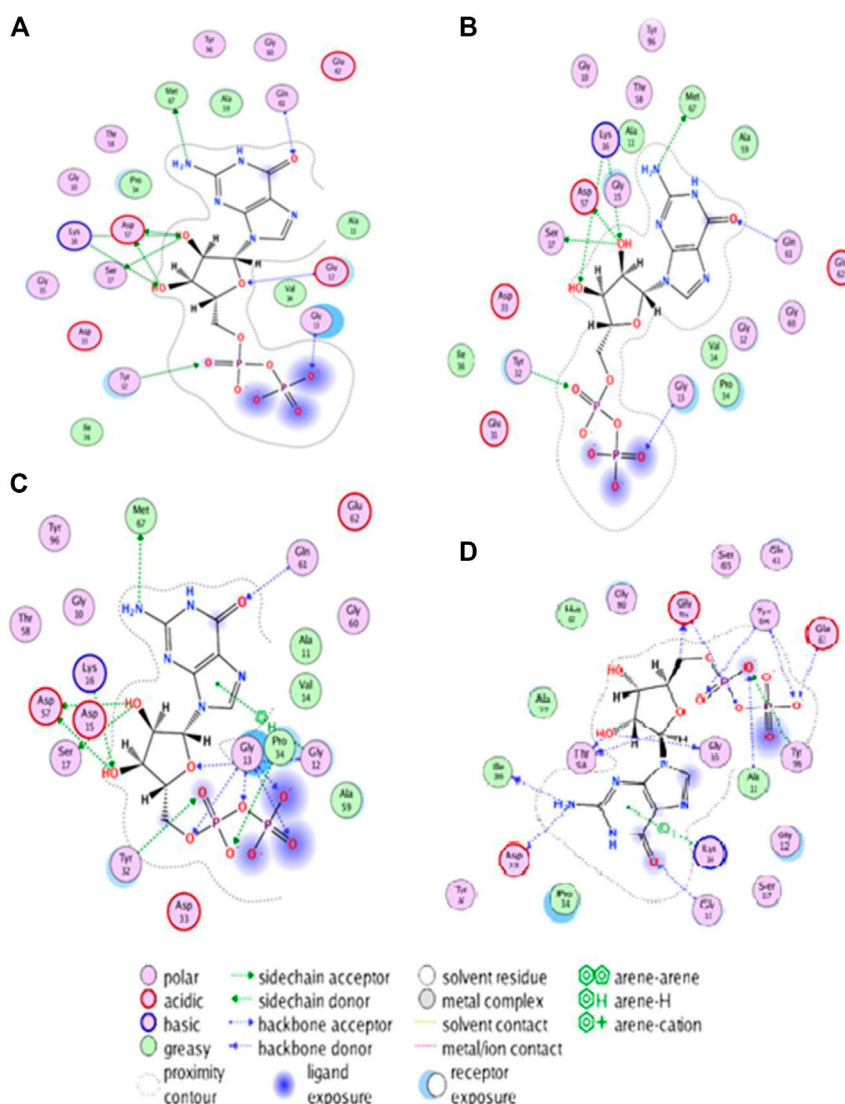


FIGURE 5

Visualizing structural changes in active site of HRAS protein due to mutations: Insights from hydrogen bonding and hydrophobic interactions.

insights into the specific residues of the protein involved in the interactions with the ligands and how the mutations affect these interactions.

The 3D interaction study of the docked complexes of G60V, G60D, and D38H with Non-mutant protein was performed to visualize the interactions and understand the binding mechanism of the mutated proteins with the Non-mutant protein as shown in Figure 6. The results showed that the mutated residues interacted with different residues compared to the Non-mutant protein. In the case of G60V, the mutated residue interacted with residues GLY13, ILE36, and GLU63. The GLU63 precipitate, which was involved in Pi-interactions and hydrogen bonding in the Non-mutant protein, did not form any interactions with the G60V mutated residue. Instead, GLY13 and ILE36 residues formed new interactions with the G60V mutated residue. This suggests that the G60V mutation might have altered the interaction pattern in the active site of the HRAS protein, which could affect the protein function.

In the case of G60D, the mutated residue interacted with residues ALA11, GLY10, and TYR96. The ALA11 residue, involved in Pi-interactions and hydrogen bonding in the Non-mutant protein, formed new interactions with the G60D mutated residue. Similarly, GLY10 and TYR96 residues also included new interactions with the G60D mutated residue. This indicates that the G60D mutation might have altered the binding pattern of the HRAS protein with ligands. In the case of D38H, the mutated residue interacted with residues GLU62, GLY10, and TYR64. The GLU62 residue, involved in Pi-interactions and hydrogen bonding in the Non-mutant protein, formed new interactions with the D38H mutated residue.

Similarly, GLY10 and TYR64 residues also formed new interactions with the D38H mutated residue. This suggests that the D38H mutation might have affected the ligands' binding in the HRAS protein's active site. Overall, the 3D interaction study revealed that the mutated residues interacted with different

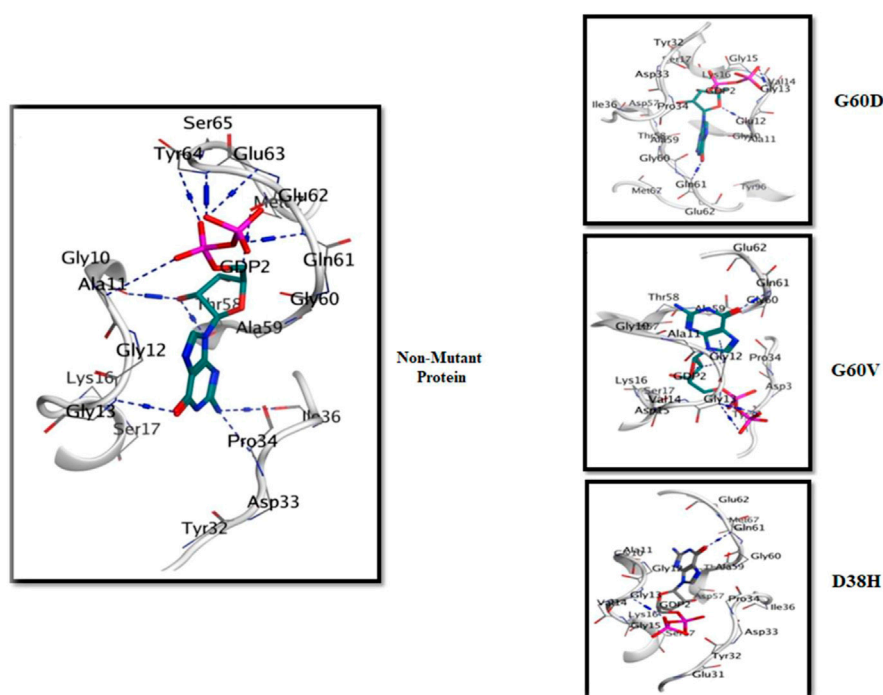


FIGURE 6

The 3D interaction of the docked complexes (G60V, G60D and D38H with Non-mutant protein).

residues compared to the Non-mutant protein, which could affect the protein-ligand interactions and the function of the HRAS protein.

3.11 Molecular dynamic simulations

The RMSD analysis is an essential tool for understanding the structural stability of a protein-ligand complex over a given time interval. The present study focused on analyzing the root-mean-square deviation (RMSD) values of protein-ligand complexes of HRAS in both APO and docked states, to examine the impact of three mutations, namely, G60V, G60D, and D38H, on the stability of the complexes. The findings of this study revealed that the wild-type HRAS complex demonstrated a greater degree of fluctuation compared to both the APO and mutated complexes. These results provide valuable insights into the effects of specific mutations on the stability of protein-ligand complexes and contribute to a better understanding of the dynamics of HRAS proteins in different states. The mean RMSD value of the wild-type complex was 4.40 Å, indicating that the protein underwent secondary structure changes with high loop regions compared to the average RMSD values of the mutated models. The mutated models (G60V, G60D, and D38H) decreased mean square calculation upon mutation, indicating a more stable environment and compactness of the whole system. The RMSD of the apoprotein was the most durable among all the protein models (Roy et al., 2022), with a value of 3.14 Å, as shown in Figure 7.

The results suggest that the mutations in the HRAS protein can lead to a more stable protein-ligand complex structure. The RMSD

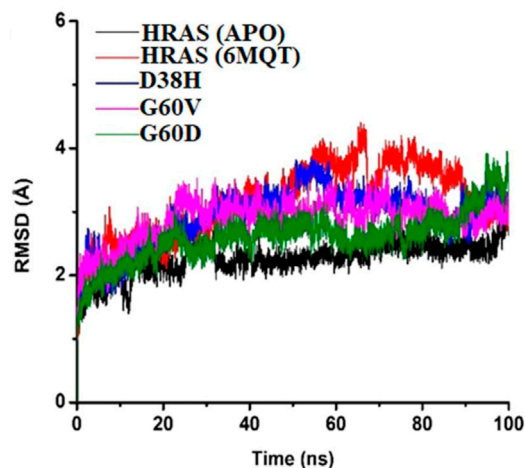


FIGURE 7

The RMSD of the APO protein was The Most Stable among All the Protein Models, with a Value of 3.14 Å as Shown in Figure 7.

analysis also revealed that the ligand remained fixed throughout the simulation time interval, indicating that the mutations did not cause any significant changes in the ligand's position or displacement. The RMSD analysis supports the conclusion that the mutations have a stabilizing effect on the protein-ligand complex. The minor fluctuations observed in the wild-type complex may indicate a higher degree of flexibility, potentially leading to changes in the protein-ligand complex's structural framework.

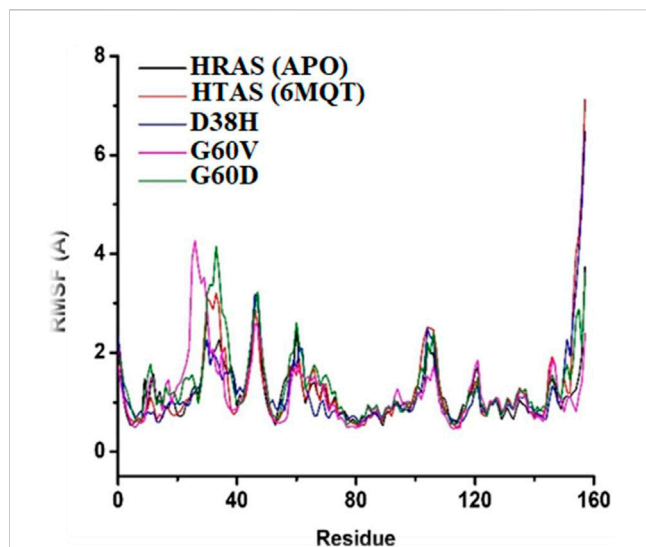


FIGURE 8

Conformational changes in structure that occurred in the docked protein complex at different time intervals.

The RMSF plot was generated to understand the fluctuation and stability of protein residues over the simulation period of 100 ns. The RMSF graph showed that the residues of the normal HRAS complex exhibited higher instability in comparison to the mutated and APO models, indicating more flexibility and conformational changes in the normal complex. The residues that showed higher RMSF values in the wild-type HRAS complex include Ser157, G12E, ILE46, and ASP47, which experienced significant structural changes concerning APO protein. On the other hand, the mutated complexes (G60V, G60D, and D38H) showed decreased RMSF values compared to the wild-type HRAS complex, indicating higher stability and less conformational changes, which are shown in Figure 8. The residues that exhibited high RMSF values in the mutated models were similar to those in the wild-type HRAS complex, including ILE46 and ASP47, but with lower fluctuations.

Notably, the residues ASN26 and ASP33 of the mutated complexes showed higher RMSF values than the wild-type HRAS complex, indicating more fluctuations and conformational changes. These residues also showed significant structural changes during the simulation time, which may affect the stability and binding affinity of the mutated HRAS complex with ligands. The RMSF plot indicated that the mutated HRAS complexes were more stable than the wild-type complex, as evidenced by the lower RMSF values. However, some residues still experienced fluctuations and conformational changes, which could affect the stability and functionality of the protein-ligand complex.

The RMSF analysis provides information about the flexibility and fluctuation of each residue in the protein structure. The RMSF analysis revealed that the fluctuations occurred at the residual level, leading to the system stability. The RMSF values showed changes occurring in regions other than the active site, which resulted in relatively more significant fluctuations. The loops in the protein structure are flexible regions, and upon ligand binding during MD simulations, they start changing their configuration at different intervals. The fluctuations in the loops and other flexible regions

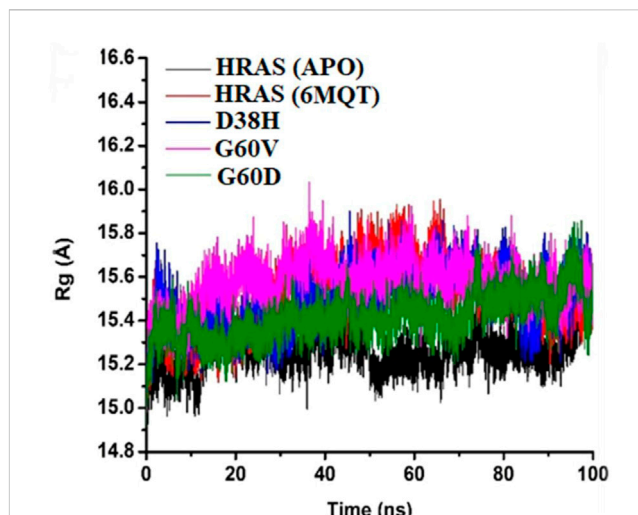


FIGURE 9

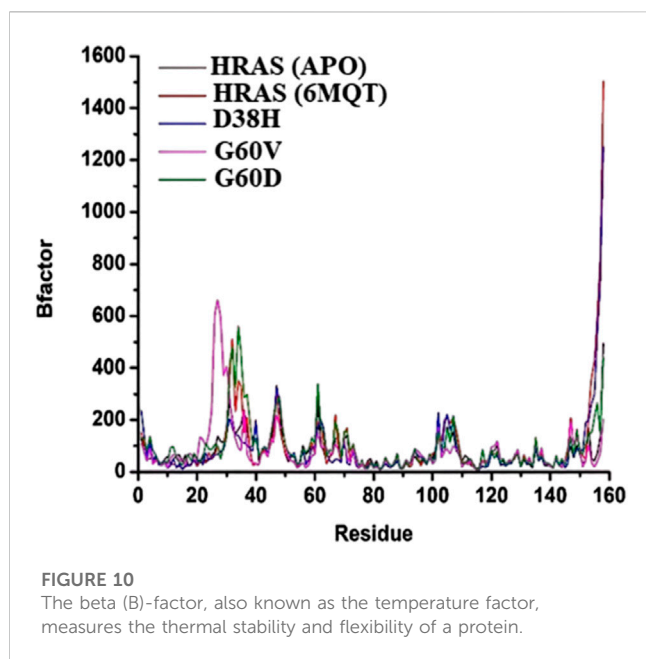
The protein Structure's Stability throughout Simulation Time Interval.

can be observed in the RMSF analysis. It can provide valuable information about the changes in the protein structure that occur during the simulation.

The radius of gyration measures the compactness of the protein structure (Ahmed et al., 2020). It provides information about the average distance of all the atoms in the protein structure from the centre of mass. In this study, the radius of gyration analysis showed that the protein structure remained stable throughout the simulation time intervals. Although the radius of gyration values did not show any significant difference among all the proteins, it still provided helpful information about the compactness of the protein structure. The RMSF and radius of gyration analyses offered valuable information about the stability and structural changes in the protein-ligand complex during the MD simulations as mentioned in Figure 9. These analyses can be helpful in understanding the dynamic behaviour of the protein-ligand complex and provide insights into the binding mechanism of the ligand to the protein target.

The beta (B)-factor, also known as the temperature factor, measures a protein's thermal stability and flexibility as per earlier researcher (Mao et al., 2020). It is calculated from the atomic displacement parameters obtained from X-ray crystallography experiments. The B-factor reflects a protein's atomic vibrations and thermal motion, with higher values indicating greater mobility and lower stability (Wang W et al., 2022). The B-factor is often used to identify regions of a protein that are flexible or disordered (Vander Meersche et al., 2021). The root-mean-square fluctuation (RMSF) is another measure of a protein's thermal stability and flexibility, based on the atomic changes of a protein over time (Khan et al., 2021). It indicates the amount of localized atomic fluctuations in a protein, which contribute to its overall vibration movement and thermal stability.

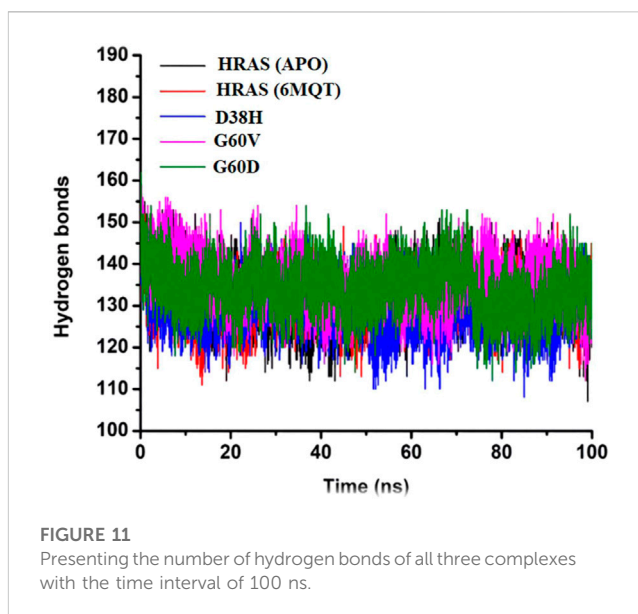
The study used the B-factor and RMSF to investigate the thermal stability and flexibility of different HRAS models, including the APO (no ligand bound), wild-type, and mutant (G60V, G60D, and D38H) HRAS models. The average B-factor values were



determined for each model, and the RMSF results were compared with the B-factor data. The results showed that the average B-factor values were highest for the G60D and wild-type HRAS models, indicating that these proteins had greater mobility and lower stability as mentioned in Figure 10. The lowest B-factor value was observed for the apo HRAS model, suggesting that this protein was the most stable. The G60V and D38H HRAS models had intermediate B-factor values, indicating moderate thermal instability.

The RMSF results also showed that the thermal instability was higher at specific residues in the different HRAS models. These results were consistent with the B-factor data, as shown in Figure 11. In particular, residues 7, 13, 27, 38, and 50 had higher RMSF and B-factor values, indicating that these residues were more flexible and less stable in all the HRAS models. The study suggests that the B-factor and RMSF analyses can provide valuable insights into proteins' thermal stability and flexibility. The results can be used to identify protein regions that are likely to be flexible or disordered, which could be important for understanding protein structure and function and drug design and development.

Hydrogen bonds are weak interactions between a hydrogen atom bonded to an electronegative atom (such as oxygen or nitrogen) and another electronegative atom in a nearby molecule (Karas et al., 2020). In biological systems, hydrogen bonds play a crucial role in determining the specificity and directionality of molecular recognition between molecules such as proteins, nucleic acids, and carbohydrates (Vladilo and Hassanali, 2018). In the study, the number of hydrogen bonds in the APO protein and all four HRAS complexes (including the G60V, G60D, and D38H mutants) were analyzed over time. The average number of hydrogen bonds was recorded for each system, and the results were presented in a time-dependent manner. The goal was to test the degree of intermolecular association across the simulation period and to investigate the effect of mutations on hydrogen bonding in the active site of the HRAS protein.



The results showed that the average number of hydrogen bonds was highest for the D38H mutant (138.9) and lowest for the G60V mutant (127.5), with the APO protein (133.6) and G60D mutant (136) having intermediate values. This suggests that the mutations can affect the number of hydrogen bonds formed within the active site of the HRAS protein. The hydrogen bond analysis also revealed that the number of hydrogen bonds decreased upon the D38H mutation and increased significantly upon the G60V, G60D, and D38H mutations. This suggests that the modifications can affect the hydrogen bonding patterns within the active site of the HRAS protein and potentially alter its structure and function. Figure 11 visually represents the number of hydrogen bonds formed in all three complexes over time. The data in the figure shows that the number of hydrogen bonds fluctuates over time, indicating that the interactions between the molecules are dynamic and can change with time. The analysis of hydrogen bonding patterns in the HRAS protein provides valuable insights into the molecular recognition and specificity of the protein. The results suggest that mutations can significantly affect the number of hydrogen bonds formed in the protein's active site, which could impact its function and potentially lead to disease.

3.12 MM-GBSA analysis

In molecular docking and drug design studies, it is essential to estimate a ligand's binding affinity to its target protein. One commonly used method for calculating binding energies is the molecular mechanics/generalized born surface area (MM-GBSA) approach (Pang et al., 2021). This method combines molecular mechanics force fields with implicit solvent models and can provide estimates of the relative binding free energies of different ligands to a target protein. In the study mentioned, the MM-GBSA approach was used to calculate the binding energies of the G60V, G60D, and D38H mutants and the normal HRAS protein. The results showed that all four systems had highly favourable binding energies, with the mutants having the

TABLE 3 Summary of the calculated binding free energies utilizing the MM/GBSA algorithm.

Protein complex	Coulomb solute	Coulomb solvent	VDW solute	VDW_Solvent	ΔG binding	Solvent GB
G60V	−4,908.6	−637.9	−299.437	−448.99198	−5,067.99	−633.999
G60D	−4,909.7	−685.6	−298.622	−86.0249567	−5,068.6	−683.333
D38H	−4,723.4	−620.7	−296.619	−80.1992073	−4,882.6	−620.903
HRAS-Normal	−4,758.06	−681.1	−144.27	−383.163312	−4,902.33	−678.231

highest. The table presented in the study (Table 3) showed that the binding energies for the G60V, G60D, and D38H mutants were −107.09, −109.42, and −107.18 kcal/mol, respectively. The Wild-type HRAS protein had a binding energy of −105.85 kcal/mol.

In addition to examining the RMSD values, this study delved deeper into the binding energies of the protein-ligand complexes by scrutinizing their components. This comprehensive analysis facilitated a meticulous investigation of the specific interactions that play a role in determining the overall binding affinity, including van der Waals forces, electrostatic interactions, and hydrogen bonding. The outcomes of this investigation furnished valuable insights into the fundamental nature of the protein-ligand interactions, unravelling the intricacies of the underlying mechanisms that govern the binding affinity of these complexes. These findings have far-reaching implications for developing novel therapeutics targeting HRAS and allied proteins. An enhanced understanding of the binding energies can optimize drug efficacy and specificity and pave the way for developing more potent and targeted treatments. The results showed that the favourable Van der Waals and Coulombic interactions were the main contributors to the high binding energies in all four systems. Van der Waals interactions are attractive forces between atoms and molecules nearby, while Coulombic interactions are electrostatic forces between charged particles. The study also noted that Coulombic interactions reduced the binding affinity of some ligands towards the active site residues. This is because electrostatic repulsion between negatively charged ligands and negatively charged residues in the active site can decrease the strength of the binding. MM-GBSA calculations performed in this study provided insights into the binding energies and components of the G60V, G60D, and D38H mutants, as well as the normal HRAS protein.

Table 3 illustrates the prime MM-GBSA calculations for the G60V, G60D, and D38H mutants and the normal HRAS protein. The results exhibit various constituents of the binding energies, comprising Coulomb solute, Coulomb solvent, VDW (van der Waals) solute, VDW solvent, ΔG (change in Gibbs free energy) binding, and Solvent GB (Generalized Born) energy. Coulomb solute and Coulomb solvent indicate the electrostatic interaction energies between the protein and solvent molecule. In contrast, VDW solute and VDW solvent imply the van der Waals interaction energies between the protein and solvent molecules. ΔG binding represents the overall binding energy of the protein-ligand complex, which shows the free energy change upon binding the ligand to the protein. Solvent GB energy reflects the energy contribution from the solvent Generalized Born model.

The outcomes manifest that all four systems exhibit highly favourable binding energies, with the mutants displaying slightly higher binding energies than the normal HRAS protein. This

suggests the mutations could intensify ligands' binding to the protein's active site. The Coulomb solute and Coulomb solvent energies were identified to be the highest in all four systems, indicating that electrostatic interactions play a significant role in the binding of ligands to the protein. The VDW solute and VDW solvent energies were also substantial, with the VDW solvent energy being the highest for the normal HRAS protein. This suggests that the solvent molecules have a part in stabilizing the protein-ligand complex. Regarding the particular mutants, the G60D mutant was observed to have the highest Coulomb solvent energy, while the D38H mutant had the lowest Coulomb solute and VDW solute energies. This implies that the mutations may affect the electrostatic and van der Waals interactions between the protein and solvent molecules, potentially influencing the binding of ligands to the active site.

The ΔG binding free energies of all the systems, including the wild-typesystem, were determined to be quite favourable, ranging from −4,902.33 to −5,068.6 kcal/mol, indicating that the binding of ligands to the HRAS protein is energetically stable. The favourable binding energies were due to the highly favourable Coulombic and Van der Waals interactions between the ligands and the protein. The Coulombic interactions between the ligands and the protein were more favourable in all the systems, resulting in lower binding energies. The electrostatic solvation energy of Generalized Born was compensated by the Coulombic interactions, which were less favourable. This suggests that the Coulombic interactions play a significant role in ligand binding towards the active site residues in the HRAS protein.

All four complexes displayed a clear pattern, which was considerably similar, indicating that the mutations did not significantly affect the binding energy. This observation is intriguing and prompts further investigation into the conformational characteristics of the systems, i.e., the interaction between the ligands and the receptor. The molecular docking and dynamics results may explain the *in vitro* findings logically. Overall, the MM-GBSA calculations provide valuable insights into the thermodynamic stability of the ligand-protein complex, which is crucial for understanding the binding affinity and specificity of the ligand towards the protein. The favourable binding energies and strong Coulombic interactions observed in this study suggest that the ligands are likely to bind tightly to the active site residues of the HRAS protein, which may have implications for the design of novel HRAS inhibitors with improved efficacy and specificity.

3.13 3D modeling and structural analysis of HRAS protein

The study aimed to investigate the impact of mutations on the structure and stability of the HRAS protein. To achieve this, the native

structure of the HRAS protein was obtained from the Protein Data Bank (PDB), and homology modeling approaches were used to predict the mutated structure of the protein. The mutations were identified using various Single Nucleotide Polymorphism (SNP) databases, including rs730880460 (G60V), rs730880460 (G60D), rs730880464 (R123P), rs750680771 (D38H), rs1564789700 (I46T), rs917210997 (G117R), rs369106578 (R123G), rs1204223913 (P110L), rs727504747 (A59L), rs104894228 (G13R), and rs1564789552 (Y64H). The PyMOL software was utilized to generate the mutated models of HRAS.

The energy minimization process was carried out using Schrödinger, a method that minimizes the energy and force load applied upon every atom in a gathering of atoms to obtain the best thermodynamically stable HRAS structure. The final and stable model of the HRAS protein was obtained after energy minimization, and the energy value was reported to be $-17,755$ kJ/mol, significantly lower than the initial energy value of -107916 kJ/mol. This indicates that the energy minimization process significantly improved the stability of the protein structure. The results of this study suggest that mutations can dramatically alter the structure and stability of the HRAS protein. The homology modelling approach successfully predicted the mutated models of the protein, and the energy minimization process further improved the stability of the predicted structures. The energy minimization process revealed a remarkable reduction in the energy value of the predicted protein structures. This significant decrease in energy strongly suggests that the systems are thermodynamically stable and likely to represent the native conformation of the protein. The observed reduction in energy value serves as a reliable indicator of the stability and accuracy of the predicted structures, bolstering confidence in the findings of this study.

The study is significant in providing insights into the impact of mutations on the stability of the HRAS protein. The mutated models generated in this study can be used for further studies to understand the effect of mutations on protein function and interactions. Overall, the results of this study provide a foundation for developing new HRAS inhibitors that can target the mutated forms of the protein with improved efficacy and specificity.

3.13.1 Residue substitutions and HRAS protein function

In the case of R123P, substituting R with P at position 123 may result in a significant loss of interaction due to the marked disparity in size and hydrophobicity between the wild type and mutant residues. The smaller size of the mutant residue may significantly impede the protein's interactions with other molecules. In D38H, the mutation of D to H at position 38 can potentially disturb the protein's interaction with other parts of the protein. The loss of the charge of the wild type residue due to this mutation may significantly impact the protein's functionality. In I46T, substituting I with T at position 46 can cause the core region of the protein to lose its hydrophobic interactions. This is because the mutant residue is, more minor, but more hydrophobic than the wild-type residue. In G117R, the mutation of G to R at position 117 could result in a protein folding problem. Introducing a charge in a buried region of the protein due to the mutant residue, which is normally flexible, may disrupt the local structure of the protein. In R123G, the mutation of R to G at position 123 could cause the protein to lose its ability to interact with other molecules due to the charge difference among mutant and wild-type residues. In G60V, substituting P with L

at position 110 can result in the loss of external interactions with other proteins in a pathway due to the difference in size between the mutant and wild-type residues. The native residue is embedded in the core region, which may make it difficult for the more considerable mutant residue to fit. In A59L, the mutation of A to L at position 59 can potentially cause the mutant residue, which is larger than the wild-type residue, to not fit in the protein's core region. This could result in a loss of external interactions. In G13R, substituting G with R at position 13 may significantly impede the protein's interaction with other molecules. Introducing a charge due to the mutant residue may cause a repelling effect between the mutant and adjacent residues, which are typically found on the protein's surface. In Y64H, converting Y to H at position 64 may result in a loss of external interactions at this point has potential result in the loss of hydrophobic interactions with other molecules.

4 Discussion

Single nucleotide polymorphisms (SNPs) are a type of genetic variation in the human genome that can result in amino acid substitutions, leading to alterations in protein structure and function (Stalin et al., 2022). Non-synonymous SNPs (nsSNPs) result in an amino acid substitution, and they have been implicated in many genetic disorders. Therefore, it is essential to understand the functional impact of nsSNPs on protein structure and function. Bioinformatics approaches have become an essential tool for predicting the practical impact of nsSNPs (Yazar and Özbek, 2021). These approaches can be broadly classified into two categories: *in silico* and experimental. Experimental approaches involve biochemical and biophysical techniques, such as X-ray crystallography, NMR spectroscopy, and site-directed mutagenesis, to directly measure the functional impact of nsSNPs. However, these methods are time-consuming, expensive, and require specialized expertise. In contrast, *in silico* approaches use computational methods to predict the functional impact of nsSNPs, making them a faster and more cost-effective way to screen for potentially deleterious variants.

SIFT (Sorting Intolerant from Tolerant) and PolyPhen-2 (Polymorphism Phenotyping v2) are widely used *in silico* tools for predicting the functional effects of nsSNPs. SIFT uses a sequence homology-based approach to predict the impact of an amino acid substitution on protein function. It compares the protein sequence of interest to a database of related protein sequences and determines the degree of conservation of the substituted amino acid. If the amino acid is highly conserved, it is predicted to be intolerant to substitution, and the variant is classified as damaging. PolyPhen-2, conversely, uses a combination of sequence- and structure-based approaches to predict the impact of amino acid substitutions on protein structure and function. It considers factors such as the physicochemical properties of the substituted amino acid, the local protein structure, and the conservation of the substituted amino acid across different species. Variants are classified as damaging if they are predicted to affect protein function, stability, or interaction with other molecules. Both SIFT and PolyPhen-2 have been shown to have high accuracy in predicting the functional impact of nsSNPs. However, they use different algorithms and approaches, which can lead to differences in their predictions. Therefore, it is recommended to use both tools to increase the accuracy of predictions. In addition, other tools, such as PROVEAN, MutPred,

and MutPred-LOF, are also available and can be used in combination to improve the accuracy of predictions.

Other tools, such as SNP&GO, P-Mut, Phd SNP, and ROVEAN, have also been developed to predict the functional effects of nsSNPs. These tools use different algorithms and approaches to predict the impact of nsSNPs on protein structure and function. They can provide complementary information to other tools. In recent research, a comprehensive analysis of nsSNPs in the HRAS gene was performed using various *in silico* tools to identify potentially deleterious SNPs that may be associated with the disease. After searching the SNP databases, 50 hits were found, and the mutants' rsIDs were submitted to SIFT and PolyPhen-2 for functional analysis. Of the 50 SNPs, 24 were found to be non-tolerable by SIFT, and 28 were predicted to be possible or probably damaging by PolyPhen-2. The results were then validated using other tools, and 11 SNPs were identified as deleterious by all of the tools. The impact of these deleterious SNPs on protein structure was analyzed using software such as PyMOL and SNP effect. In conclusion, *in silico* approaches and tools have become powerful tools for identifying and characterizing the functional effects of nsSNPs and predicting their potential association with various diseases. While no single tool can accurately predict the practical effects of nsSNPs, a combination of tools can provide more accurate and reliable predictions. Moreover, detailed analysis of the impact of deleterious SNPs on protein structure and function can provide important insights into the development of genetic disorders and may lead to the development of novel therapeutic approaches.

Bioinformatics is an essential tool for analysing genetic variations, including non-synonymous single nucleotide polymorphisms (nsSNPs) (Wang et al., 2020). *In-silico* approaches have been developed to predict the functional impact of nsSNPs on protein structure and function and to distinguish between neutral and deleterious variants. One of the widely used approaches for predicting the deleterious effect of nsSNPs is SIFT which predicts whether an amino acid substitution is likely to affect protein function based on sequence conservation. Another approach is PolyPhen (Polymorphism Phenotyping), which is a tool for predicting the functional effect of amino acid substitutions on protein structure and function (Seifi and Walter, 2018). It uses both sequence-based and structure-based features to predict the impact of a variant on protein function. In addition, machine learning-based methods such as Random Forest and Support Vector Machines (SVM) have also been developed to classify nsSNPs as deleterious or neutral (Ge et al., 2021). These methods use a combination of sequence, structural and evolutionary features to predict the functional impact of nsSNPs.

In a recent study by Behairy et al. (2022), various SNPs of the HRAS gene were assessed to identify those potentially deleterious and associated with disease development using computational approaches. We searched for nsSNPs against HRAS in SNP databases and found 138 hits. The rsIDs of these mutants were submitted to two widely used computational tools, SIFT and PolyPhen2, to determine the functional effects of the nsSNPs. SIFT identified 15nsSNPs as non-tolerable, while PolyPhen2 showed 23 nsSNPs as possible and probably damaging. To validate the results, the authors submitted the rsIDs of the 23 nsSNPs to several other tools, including SNP&GO, P-Mut, Phd SNP, and ROVEAN. Among these, 10 SNPs, with rsIDs of rs730880460, rs730880460, rs730880464, rs750680771,

rs1564789700, rs917210997, rs369106578, rs1204223913, rs727504747, rs104894228, and rs1564789552, were identified as deleterious by all the tools. The authors highlighted that the association of these damaging nsSNPs with disease development has not been reported in any other study yet. Therefore, further research is required to validate the functional significance of these nsSNPs in HRAS and their association with disease development.

The authors emphasized that combining multiple algorithms frequently is a powerful tool for selecting candidate functional nsSNPs. In a previous study by Falahi et al. (2021), it was reported that among various *In-silico* tools, Polyphen 2 and SNAP show better performance for identifying functional nsSNPs. Thus, using multiple computational tools and integrating their results can provide more reliable predictions of functional nsSNPs, which can aid in understanding the molecular basis of diseases caused by nsSNPs. Overall, the study by Behairy et al. (2022), highlights the importance of using computational approaches to identify potential deleterious nsSNPs in genes and their association with disease development. The study also underscores the significance of using multiple computational tools to validate the functional relevance of nsSNPs and the need for further research to establish their association with disease development.

In our study, we utilized SNP effect to assess the impact of SNPs on the aggregation tendency, amyloid propensity, and chaperone binding of HRAS protein. The outcomes of SNP effect revealed that rs1204223913 increases the aggregation propensity of HRAS protein with a dTANGO score of 547.61, while rs1564789700 decreases the aggregation propensity with a dWALTZ score of -106.36. However, we observed that most of the variations did not affect the molecular phenotype of the protein. Although these variations may convey some damaging mutation on HRAS protein, they seem unrelated to the protein's aggregation tendency, amyloid propensity, or chaperone binding tendency, according to the outcomes of the SNP effect. To further investigate the impact of deleterious nsSNPs on the protein structure of HRAS, we retrieved the tertiary structure of HRAS from PDB with PDB id (6MQT). We mapped all 10 deleterious nsSNPs using PyMOL software. Our findings indicate that these nsSNPs are located in different regions of the protein structure and may affect the protein function by altering its stability or interactions with other molecules.

Several previous studies have also investigated the impact of nsSNPs on protein structure and function. For instance, a survey by Taghvaei et al. (2021) utilized *in silico* tools to predict the effects of nsSNPs on protein function and reported that these tools aid in identifying deleterious nsSNPs associated with human diseases. Additionally, a study by Teng et al. (2009) utilized molecular dynamics simulations to investigate the impact of nsSNPs on protein stability and interactions. Our study highlights the importance of using computational approaches to identify deleterious nsSNPs and their potential effects on protein structure and function. Our findings also suggest that combining multiple *in silico* tools can provide a more accurate prediction of the functional impact of nsSNPs on proteins.

Using computational methods, our investigation aimed to detect harmful single nucleotide polymorphisms (SNPs) in the HRAS gene. By examining several SNP databases and employing different

analytical tools, we identified 10 damaging SNPs. Our study is significant because mutations in HRAS are associated with various diseases, including cancer. Recognizing potentially harmful SNPs in HRAS can help us understand the genetic basis of these diseases and facilitate the development of personalized treatments. Moreover, using computational methods to identify functional nsSNPs can considerably reduce the time and cost required for experimental validation. Our findings emphasize combining various algorithms and tools to pinpoint candidate functional nsSNPs. In conclusion, our research offers a valuable contribution to the genetics field and can assist in developing personalized medical interventions.

5 Conclusion

We investigate the potential impact of nsSNPs in the HRAS gene on the structure and function of the HRAS protein, as well as their potential contribution to the development of various types of cancers. Specifically to identify and characterize nsSNPs within the coding region of HRAS that can cause detrimental mutations, disrupt normal protein function, and activate oncogenic signaling pathways. To decrease expenses and enhance the efficiency of genetic association studies, We applied *in silico* approaches, including SIFT analysis, PolyPhen2 scores, and TI scores, to predict the potential impact of rare genetic variants on HRAS protein function, and identified 50 nsSNPs in total, of which 23 were located in the exon region of the HRAS gene and were likely to be deleterious. Among these 23 nsSNPs, 10 had the most destructive impacts, including G60V, G60D, R123P, D38H, I46T, G115R, R123G, P110L, A59L, and G13R, with DDG values ranging from -3.21 to 0.87 kcal/mol. We conducted molecular dynamics (MD) simulations to analyze the stability, flexibility, and compaction of the HRAS protein to investigate the consequences of specific non-synonymous single nucleotide polymorphisms (nsSNPs). Our predicted results indicated that the stable model of HRAS had a significantly lower energy value than the initial model, suggesting that these nsSNPs may alter the stability of the protein. Furthermore, we analyzed the binding energies of both the wild-type and mutant HRAS protein with docked complexes to understand the potential impact of these mutations on the activation of oncogenic signalling pathways. Our findings indicated that the G60V, G60D, and D38H mutants had higher binding energies than the wild-type HRAS protein, potentially leading to the activation of oncogenic signalling pathways and contributing to the development of various types

of cancers. Our systematic study analysis provides essential insights into the potential functional role of nsSNPs in the HRAS gene in cancer development. It could inform future studies aimed at developing targeted therapies for cancer treatment.

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

All authors contributed to the study's conception and design. IZ, RS, MB, supervised this manuscript, and finally conceived and designed the study protocols. Material preparation, data collection, analysis, and manuscript writing were performed by SA, UA, AQ, MY, QUA, SR, H-AN, and YABJ, reviewed and proofread the manuscript. All authors read and approved the final manuscript.

Acknowledgments

The authors sincerely thank the Researchers Supporting Project, King Saud University, Riyadh, Saudi Arabia, for funding this work through project number (RSP 2023R457).

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

- Ahmad, A., Ahad, A., Rao, A. Q., and Husnain, T. (2015). Molecular docking based screening of neem-derived compounds with the NS1 protein of Influenza virus. *Bioinformation* 11 (7), 359–365. doi:10.6026/97320630011359
- Ahmad, H. M., Abrar, M., Izhar, O., Zafar, I., Rather, M. A., Alanazi, A. M., et al. (2022). Characterization of fenugreek and its natural compounds targeting AKT-1 protein in cancer: Pharmacophore, virtual screening, and MD simulation techniques. *J. King Saud University-Science* 34 (6), 102186. doi:10.1016/j.jksus.2022.102186
- Ahmed, M. C., Crehuet, R., and Lindorff-Larsen, K. (2020). Computing, analyzing, and comparing the radius of gyration and hydrodynamic radius in conformational ensembles of intrinsically disordered proteins. *Intrinsically Disord. Proteins Methods Protoc.* 2141, 429–445. doi:10.1007/978-1-0716-0524-0_21
- Amaral, M., Kokh, D., Bomke, J., Wegener, A., Buchstaller, H., Eggenweiler, H., et al. (2017). Protein conformational flexibility modulates kinetics and thermodynamics of drug binding. *Nat. Commun.* 8 (1), 2276. doi:10.1038/s41467-017-02258-w

- Asati, V., Mahapatra, D. K., and Bharti, S. K. (2016). PI3K/Akt/mTOR and Ras/Raf/MEK/ERK signaling pathways inhibitors as anticancer agents: Structural and pharmacological perspectives. *Eur. J. Med. Chem.* 109, 314–341. doi:10.1016/j.ejmech.2016.01.012
- Backwell, L., and Marsh, J. A. (2022). Diverse molecular mechanisms underlying pathogenic protein mutations: Beyond the loss-of-function paradigm. *Annu. Rev. genomics Hum. Genet.* 23, 475–498. doi:10.1146/annurev-genom-111221-103208
- Behairy, M. Y., Soltan, M. A., Adam, M. S., Refaat, A. M., Ezz, E. M., Albogami, S., et al. (2022). Computational analysis of deleterious SNPs in NRAS to assess their potential correlation with carcinogenesis. *Front. Genet.* 13, 872845. doi:10.3389/fgene.2022.872845
- Bhattacharya, R., Rose, P. W., Burley, S. K., and Prlić, A. (2017). Impact of genetic variation on three dimensional structure and function of proteins. *PLoS one* 12 (3), e0171355. doi:10.1371/journal.pone.0171355
- Bromberg, Y., and Rost, B. (2007). Snap: Predict effect of non-synonymous polymorphisms on function. *Nucleic acids Res.* 35 (11), 3823–3835. doi:10.1093/nar/gkm238
- Buljan, M., Blattmann, P., Aebersold, R., and Boutros, M. (2018). Systematic characterization of pan-cancer mutation clusters. *Mol. Syst. Biol.* 14 (3), e7974. doi:10.15252/msb.20177974
- Cain, A. K., Barquist, L., Goodman, A. L., Paulsen, I. T., Parkhill, J., and van Opijnen, T. (2020). A decade of advances in transposon-insertion sequencing. *Nat. Rev. Genet.* 21 (9), 526–540. doi:10.1038/s41576-020-0244-x
- Chai, C.-Y., Maran, S., Thew, H.-Y., Tan, Y.-C., Rahman, N. M. A. N. A., Cheng, W.-H., et al. (2022). Predicting deleterious non-synonymous single nucleotide polymorphisms (nsSNPs) of HRAS gene and in silico evaluation of their structural and functional consequences towards diagnosis and prognosis of cancer. *Biology* 11 (11), 1604. doi:10.3390/biology11111604
- Dakal, T. C., Kala, D., Dhiman, G., Yadav, V., Krokhotin, A., and Dokholyan, N. V. (2017). Predicting the functional consequences of non-synonymous single nucleotide polymorphisms in IL8 gene. *Sci. Rep.* 7 (1), 6525. doi:10.1038/s41598-017-06575-4
- De Baets, G., Van Durme, J., Reumers, J., Maurer-Stroh, S., Vanhee, P., Dopazo, J., et al. (2012). SNPeff 4.0: On-line prediction of molecular and structural effects of protein-coding variants. *Nucleic acids Res.* 40 (D1), D935–D939. doi:10.1093/nar/gkr996
- De Luca, A., Maiello, M. R., D'Alessio, A., Pergameno, M., and Normanno, N. (2012). The RAS/RAF/MEK/ERK and the PI3K/AKT signalling pathways: Role in cancer pathogenesis and implications for therapeutic approaches. *Expert Opin. Ther. targets* 16 (2), S17–S27. doi:10.1517/14728222.2011.639361
- Dehouck, Y., Kwasigroch, J. M., Gilis, D., and Rooman, M. (2011). PoPMuSiC 2.1: A web server for the estimation of protein stability changes upon mutation and sequence optimality. *BMC Bioinforma.* 12 (2), 151–212. doi:10.1186/1471-2105-12-151
- Estep, A. L., Tidyman, W. E., Teitell, M. A., Cotter, P. D., and Rauen, K. A. (2006). HRAS mutations in costello syndrome: Detection of constitutional activating mutations in codon 12 and 13 and loss of wild-type allele in malignancy. *Am. J. Med. Genet. Part A* 140 (1), 8–16. doi:10.1002/ajmg.a.31078
- Falahi, S., Karaji, A. G., Koohyanizadeh, F., Rezaeiemanesh, A., and Salari, F. (2021). A comprehensive *in silico* analysis of the functional and structural impact of single nucleotide polymorphisms (SNPs) in the human IL-33 gene. *Comput. Biol. Chem.* 94, 107560. doi:10.1016/j.compbiolchem.2021.107560
- Galindo-Murillo, R., Roe, D. R., and Cheatham, T. E., III (2015). Convergence and reproducibility in molecular dynamics simulations of the DNA duplex d (GCACGAACGAACGACG). *Biochimica Biophysica Acta (BBA)-General Subj.* 1850 (5), 1041–1058. doi:10.1016/j.bbagen.2014.09.007
- Gámez, A., Yuste-Checa, P., Brasil, S., Briso-Montiano, Á., Desviat, L. R., Ugarte, M., et al. (2018). Protein misfolding diseases: Prospects of pharmacological treatment. *Clin. Genet.* 93 (3), 450–458. doi:10.1111/cge.13088
- Garcia-Rostan, G., Zhao, H., Camp, R. L., Pollan, M., Herrero, A., Pardo, J., et al. (2003). Ras mutations are associated with aggressive tumor phenotypes and poor prognosis in thyroid cancer. *J. Clin. Oncol.* 21 (17), 3226–3235. doi:10.1200/jco.2003.10.130
- Ge, F., Muhammad, A., and Yu, D.-J. (2021). DeepnsSNPs: Accurate prediction of non-synonymous single-nucleotide polymorphisms by combining multi-scale convolutional neural network and residue environment information. *Chemom. Intelligent Laboratory Syst.* 215, 104326. doi:10.1016/j.chemolab.2021.104326
- Gilardi, M., Wang, Z., Proietto, M., Chilà, A., Calleja-Valera, J. L., Goto, Y., et al. (2020). Tipifarnib as a precision therapy for HRAS-mutant head and neck squamous cell carcinomas. *Mol. cancer Ther.* 19 (9), 1784–1796. doi:10.1158/1535-7163.mct-19-0958
- Gourbal, B., Pinaud, S., Beckers, G. J., Van Der Meer, J. W., Conrath, U., and Netea, M. G. (2018). Innate immune memory: An evolutionary perspective. *Immunol. Rev.* 283 (1), 21–40. doi:10.1111/immr.12647
- Hamburg, M. A., and Collins, F. S. (2010). The path to personalized medicine. *N. Engl. J. Med.* 363 (4), 301–304. doi:10.1056/nejmp1006304
- Han, X., Zhou, Z., Fei, L., Sun, H., Wang, R., Chen, Y., et al. (2020). Construction of a human cell landscape at single-cell level. *Nature* 581 (7808), 303–309. doi:10.1038/s41586-020-2157-4
- Harder, E., Damm, W., Maple, J., Wu, C., Reboul, M., Xiang, J. Y., et al. (2016). OPLS3: A force field providing broad coverage of drug-like small molecules and proteins. *J. Chem. theory Comput.* 12 (1), 281–296. doi:10.1021/acs.jctc.5b00864
- Hossain, M. S., Roy, A. S., and Islam, M. S. (2020). *In silico* analysis predicting effects of deleterious SNPs of human RASSF5 gene on its structure and functions. *Sci. Rep.* 10 (1), 14542. doi:10.1038/s41598-020-71457-1
- Jefferies, S., and Foulkes, W. (2001). Genetic mechanisms in squamous cell carcinoma of the head and neck. *Oral Oncol.* 37 (2), 115–126. doi:10.1016/s1368-8375(00)00065-8
- Ji, S. S., German, C. A., Lange, K., Sinsheimer, J. S., Zhou, H., Zhou, J., et al. (2021). Modern simulation utilities for genetic analysis. *BMC Bioinforma.* 13 (10), 228–313. doi:10.1186/s12859-021-04086-8
- Jin, H., Wu, C., Su, R., Sun, T., Li, X., and Guo, C. (2023). Identifying dopamine D3 receptor ligands through virtual screening and exploring the binding modes of hit compounds. *Molecules* 28 (2), 527. doi:10.3390/molecules28020527
- Jubb, H. C., Pandurangan, A. P., Turner, M. A., Ochoa-Montano, B., Blundell, T. L., and Ascher, D. B. (2017). Mutations at protein-protein interfaces: Small changes over big surfaces have large impacts on human health. *Prog. biophysics Mol. Biol.* 128, 3–13. doi:10.1016/j.pbiomolbio.2016.10.002
- Karas, L. J., Wu, C. H., Das, R., and Wu, J. I. C. (2020). Hydrogen bond design principles. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* 10 (6), e1477. doi:10.1002/wcms.1477
- Kawazu, M., Ueno, T., Kontani, K., Ogita, Y., Ando, M., Fukumura, K., et al. (2013). Translating mutations of RAC guanosine triphosphatases in human cancers. *Proc. Natl. Acad. Sci.* 110 (8), 3029–3034. doi:10.1073/pnas.1216141110
- Kellici, T. F., Ntountaniotis, D., Liapakis, G., Tzakos, A. G., and Mavromoustakos, T. (2019). The dynamic properties of angiotensin II type 1 receptor inverse agonists in solution and in the receptor site. *Arabian J. Chem.* 12 (8), 5062–5078. doi:10.1016/j.arabj.2016.11.014
- Khan, Z., and Bisen, P. S. (2013). Oncoapoptotic signaling and deregulated target genes in cancers: Special reference to oral cancer. *Biochimica Biophysica Acta (BBA)-Reviews Cancer* 1836 (1), 123–145. doi:10.1016/j.bbcan.2013.04.002
- Khan, I., Rhett, J. M., and O'Bryan, J. P. (2020). Therapeutic targeting of RAS: New hope for drugging the “undruggable”. *Biochimica Biophysica Acta (BBA)-Molecular Cell. Res.* 1867 (2), 118570. doi:10.1016/j.bbamcr.2019.118570
- Khan, M. T., Ali, A., Wang, Q., Irfan, M., Khan, A., Zeb, M. T., et al. (2021). Marine natural compounds as potent inhibitors against the main protease of SARS-CoV-2—A molecular dynamic study. *J. Biomol. Struct. Dyn.* 39 (10), 3627–3637. doi:10.1080/07391102.2020.1769733
- Kim, H. S., Jeong, Y. K., Hur, J. K., Kim, J.-S., and Bae, S. (2019). Adenine base editors catalyze cytosine conversions in human cells. *Nat. Biotechnol.* 37 (10), 1145–1148. doi:10.1038/s41587-019-0254-4
- Kohl, S., Zobor, D., Chiang, W.-C., Weisschuh, N., Staller, J., Menendez, I. G., et al. (2015). Mutations in the unfolded protein response regulator ATF6 cause the cone dysfunction disorder achromatopsia. *Nat. Genet.* 47 (7), 757–765. doi:10.1038/ng.3319
- Kompier, L. C., Lurkin, I., van der Aa, M. N., van Rhijn, B. W., van der Kwast, T. H., and Zwarthoff, E. C. (2010). FGFR3, HRAS, KRAS, NRAS and PIK3CA mutations in bladder cancer and their potential as biomarkers for surveillance and therapy. *PLoS one* 5 (11), e13821. doi:10.1371/journal.pone.0013821
- Korzeniacki, C., and Priefer, R. (2021). Targeting KRAS mutant cancers by preventing signaling transduction in the MAPK pathway. *Eur. J. Med. Chem.* 211, 113006. doi:10.1016/j.ejmech.2020.113006
- Kotha, S. (2010). *Analysis of missense mutations in adenosine deaminase using Pathogenic-Or-Not-Pipeline (PON-P)*
- Kumar, N., Sood, D., Tomar, R., and Chandra, R. (2019). Antimicrobial peptide designing and optimization employing large-scale flexibility analysis of protein-peptide fragments. *ACS omega* 4 (25), 21370–21380. doi:10.1021/acsomega.9b03035
- Kumar, A. P., Verma, C. S., and Lukman, S. (2021). Structural dynamics and allostery of rab proteins: Strategies for drug discovery and design. *Briefings Bioinforma.* 22 (1), 270–287. doi:10.1093/bib/bbz161
- Kutzner, C., Knip, C., Cherian, A., Nordstrom, L., Grubmüller, H., de Groot, B. L., et al. (2022). GROMACS in the cloud: A global supercomputer to speed up alchemical drug design. *J. Chem. Inf. Model.* 62 (7), 1691–1711. doi:10.1021/acs.jcim.2c00044
- Li, Z., and Scheraga, H. A. (1987). Monte Carlo-minimization approach to the multiple-minima problem in protein folding. *Proc. Natl. Acad. Sci.* 84 (19), 6611–6615. doi:10.1073/pnas.84.19.6611
- Li, Y., Wang, Y., Li, Y., and Yang, L. (2006). Prediction of the deleterious nsSNPs in ABCB transporters. *FEBS Lett.* 580 (30), 6800–6806. doi:10.1016/j.febslet.2006.11.047

- Lin, S.-M., Lin, S.-C., Hsu, J.-N., Chang, C.-k., Chien, C.-M., Wang, Y.-S., et al. (2020). Structure-based stabilization of non-native protein-protein interactions of coronavirus nucleocapsid proteins in antiviral drug design. *J. Med. Chem.* 63 (6), 3131–3141. doi:10.1021/acs.jmedchem.9b01913
- Liu, P., Wang, Y., and Li, X. (2019). Targeting the untargetable KRAS in cancer therapy. *Acta Pharm. Sin. B* 9 (5), 871–879. doi:10.1016/j.apsb.2019.03.002
- Luo, J., Wei, W., Waldispühl, J., and Moitessier, N. (2019). Challenges and current status of computational methods for docking small molecules to nucleic acids. *Eur. J. Med. Chem.* 168, 414–425. doi:10.1016/j.ejmech.2019.02.046
- Makrides, V., Dolgodilina, E., and Virgintino, D. (2017). Blood-brain barrier transporters and neuroinflammation: Partners in neuroprotection and in pathology. *Blood Brain Barrier Inflamm.*, 103–151. doi:10.1007/978-3-319-45514-3_6
- Mao, S., Cheng, X., Zhu, Z., Chen, Y., Li, C., Zhu, M., et al. (2020). Engineering a thermostable version of D-allulose 3-epimerase from *Rhodospirillum rubrum* via site-directed mutagenesis based on B-factors analysis. *Enzyme Microb. Technol.* 132, 109441. doi:10.1016/j.enzmictec.2019.109441
- Masoodi, T. A., Al Shammari, S. A., Al-Muammar, M. N., Alhamdan, A. A., and Talluri, V. R. (2013). Exploration of deleterious single nucleotide polymorphisms in late-onset Alzheimer disease susceptibility genes. *Gene* 512 (2), 429–437. doi:10.1016/j.gene.2012.08.026
- Meyer, K. B., Maia, A.-T., O'Reilly, M., Teschendorff, A. E., Chin, S.-F., Caldas, C., et al. (2008). Allele-specific up-regulation of FGFR2 increases susceptibility to breast cancer. *PLoS Biol.* 6 (5), e108. doi:10.1371/journal.pbio.0060108
- Nagata, M., Muto, S., and Horie, S. (2016). Molecular biomarkers in bladder cancer: Novel potential indicators of prognosis and treatment outcomes. *Dis. markers* 2016, 8205836. doi:10.1155/2016/8205836
- Ng, P. C., and Henikoff, S. (2001). Predicting deleterious amino acid substitutions. *Genome Res.* 11 (5), 863–874. doi:10.1101/gr.176601
- Ngan, H.-L., Law, C.-H., Choi, Y. C. Y., Chan, J. Y.-S., and Lui, V. W. Y. (2022). Precision drugging of the MAPK pathway in head and neck cancer. *NPJ Genomic Med.* 7 (1), 20. doi:10.1038/s41525-022-00293-1
- Niranjan, V., Rao, A., Janaki, B., Uttarkar, A., Setlur, A. S., Chandrashekar, K., et al. (2021). Molecular docking and interaction studies of identified abscisic acid receptors in *Oryza sativa*: An *in-silico* perspective on comprehending stress tolerance mechanisms. *Curr. Genomics* 22 (8), 607–619. doi:10.2174/1389202923666211222161006
- O'Bryan, J. P. (2019). Pharmacological targeting of RAS: Recent success with direct inhibitors. *Pharmacol. Res.* 139, 503–511. doi:10.1016/j.phrs.2018.10.021
- Odeniyide, P., Yohe, M. E., Pollard, K., Vaseva, A. V., Calizo, A., Zhang, L., et al. (2022). Targeting farnesylation as a novel therapeutic approach in HRAS-mutant rhabdomyosarcoma. *Oncogene* 41 (21), 2973–2983. doi:10.1038/s41388-022-02305-x
- Pang, J., Gao, S., Sun, Z., and Yang, G. (2021). Discovery of small molecule PLpro inhibitor against COVID-19 using structure-based virtual screening, molecular dynamics simulation, and molecular mechanics/Generalized Born surface area (MM/GBSA) calculation. *Struct. Chem.* 32 (2), 879–886. doi:10.1007/s11224-020-01665-y
- Pang, C. S. M. (2018). Developing a computational approach to investigate the impacts of disease-causing mutations on protein function UCL. London: University College London.
- Paniri, A., Hosseini, M. M., and Akhavan-Niaki, H. (2021). First comprehensive computational analysis of functional consequences of TMPRSS2 SNPs in susceptibility to SARS-CoV-2 among different populations. *J. Biomol. Struct. Dyn.* 39 (10), 3576–3593. doi:10.1080/07391102.2020.1767690
- Pauls, S. U., Nowak, C., Bálint, M., and Pfenninger, M. (2013). The impact of global climate change on genetic diversity within populations and species. *Mol. Ecol.* 22 (4), 925–946. doi:10.1111/mec.12152
- Phillips, C. (2009). SNP databases. *Single Nucleotide Polymorphisms Methods Protoc.* 578, 43–71. doi:10.1007/978-1-60327-411-1_3
- Protein Data Bank/Berman, H. M., Bhikadiya, C., Bi, C., Chen, L., and Costanzo, L. D. (2019). Protein Data Bank: The single global archive for 3D macromolecular structure data. *Nucleic acids Res.* 47 (D1), D520–D528. doi:10.1093/nar/gky949
- Rajalingam, K., Schreck, R., Rapp, U. R., and Albert, S. (2007). Ras oncogenes and their downstream targets. *Biochimica Biophysica Acta (BBA)-Molecular Cell. Res.* 1773 (8), 1177–1195. doi:10.1016/j.bbmr.2007.01.012
- Rajaram, S. (2001). "Prospects and promise of wheat breeding in the 21 st century Wheat in a Global Environment," in Proceedings of the 6th International Wheat Conference, Budapest, Hungary, 5–9 June 2000.
- Rezatabar, S., Karimian, A., Rameshknia, V., Parsian, H., Majidinia, M., Kopi, T. A., et al. (2019). RAS/MAPK signaling functions in oxidative stress, DNA damage response and cancer progression. *J. Cell. physiology* 234 (9), 14951–14965. doi:10.1002/jcp.28334
- Rivera, A., Fisher, S. A., Fritsche, L. G., Keilhauer, C. N., Lichtner, P., Meitinger, T., et al. (2005). Hypothetical LOC387715 is a second major susceptibility gene for age-related macular degeneration, contributing independently of complement factor H to disease risk. *Hum. Mol. Genet.* 14 (21), 3227–3236. doi:10.1093/hmg/ddi353
- Roy, R., Mishra, A., Poddar, S., Nayak, D., and Kar, P. (2022). Investigating the mechanism of recognition and structural dynamics of nucleoprotein-RNA complex from Peste des petits ruminants virus via Gaussian accelerated molecular dynamics simulations. *J. Biomol. Struct. Dyn.* 40 (5), 2302–2315. doi:10.1080/07391102.2020.1838327
- Santana-Romo, F., Lagos, C. F., Duarte, Y., Castillo, F., Moglie, Y., Maestro, M. A., et al. (2020). Innovative three-step microwave-promoted synthesis of N-propargyltetrahydroquinoline and 1, 2, 3-triazole derivatives as a potential factor Xa (FXa) inhibitors: Drug design, synthesis, and biological evaluation. *Molecules* 25 (3), 491. doi:10.3390/molecules25030491
- Savas, S., Kim, D. Y., Ahmad, M. F., Shariff, M., and Ozelik, H. (2004). Identifying functional genetic variants in DNA repair pathway using protein conservation analysis. *Cancer Epidemiol. Biomarkers Prev.* 13 (5), 801–807. doi:10.1158/1055-9965.801.13.5
- Seal, A., Gupta, A., Mahalaxmi, M., Aykhal, R., Singh, T. R., and Arunachalam, V. (2014). Tools, resources and databases for SNPs and indels in sequences: A review. *Int. J. Bioinforma. Res. Appl.* 10 (3), 264–296. doi:10.1504/ijbra.2014.060762
- Seif, M., and Walter, M. A. (2018). Accurate prediction of functional, structural, and stability changes in PITX2 mutations using *in silico* bioinformatics algorithms. *PLoS one* 13 (4), e0195971. doi:10.1371/journal.pone.0195971
- Sharma, K., Hema, K., Bhatraju, N. K., Kukreti, R., Das, R. S., Gupta, M. D., et al. (2022). The deleterious impact of a non-synonymous SNP on protein structure and function is apparent in hypertension. *J. Mol. Model.* 28, 14–17. doi:10.1007/s00894-021-04997-6
- Shinwari, K., Rehman, H. M., Liu, G., Bolkov, M. A., Tuzankina, I. A., and Cheresnev, V. A. (2022). Novel disease-associated missense single-nucleotide polymorphisms variants predication by algorithms tools and molecular dynamics simulation of human TCIRG1 gene causing congenital neutropenia and osteopetrosis. *Front. Mol. Biosci.* 9, 879875. doi:10.3389/fmolb.2022.879875
- Shorning, B. Y., Dass, M. S., Smalley, M. J., and Pearson, H. B. (2020). The PI3K-AKT-mTOR pathway and prostate cancer: At the crossroads of AR, MAPK, and WNT signaling. *Int. J. Mol. Sci.* 21 (12), 4507. doi:10.3390/ijms21124507
- Singh, M., Popowicz, G. M., Krajewski, M., and Holak, T. A. (2007). Structural ramification for acetyl-lysine recognition by the bromodomain of human BRG1 protein, a central ATPase of the SWI/SNF remodeling complex. *Chembiochem* 8 (11), 1308–1316. doi:10.1002/cbic.200600562
- Stalin, A., Lin, D., Josephine Princy, J., Feng, Y., Xiang, H., Ignacimuthu, S., et al. (2022). Computational analysis of single nucleotide polymorphisms (SNPs) in PPAR gamma associated with obesity, diabetes and cancer. *J. Biomol. Struct. Dyn.* 40 (4), 1843–1857. doi:10.1080/07391102.2020.1835724
- Taghvaei, S., Saremi, L., and Babaniamansour, S. (2021). Computational analysis of Gly482Ser single-nucleotide polymorphism in PPARGC1A gene associated with CAD, NAFLD, T2DM, obesity, hypertension, and metabolic diseases. *PPAR Res.* 2021, 1–12. doi:10.1155/2021/5544233
- Tarek, M. M., Yahia, A., El-Nakib, M. M., and Elhefnawi, M. (2021). Integrative assessment of CIP2A overexpression and mutational effects in human malignancies identifies possible deleterious variants. *Comput. Biol. Med.* 139, 104986. doi:10.1016/j.combiomed.2021.104986
- Teng, S., Madej, T., Panchenko, A., and Alexov, E. (2009). Modeling effects of human single nucleotide polymorphisms on protein-protein interactions. *Biophysical J.* 96 (6), 2178–2188. doi:10.1016/j.bpj.2008.12.3904
- Ullah, R., Yin, Q., Snell, A. H., and Wan, L. (2022). RAF-MEK-ERK pathway in cancer evolution and treatment Seminars in cancer biology. *Semin. Cancer Biol.* 85, 123. doi:10.1016/j.semcancer.2021.05.010
- Vander Meersche, Y., Cretin, G., de Brevern, A. G., Gelly, J.-C., and Galochkina, T. (2021). Medusa: Prediction of protein flexibility from sequence. *J. Mol. Biol.* 433(11), 166882. doi:10.1016/j.jmb.2021.166882
- Velazquez, H. A., Riccardi, D., Xiao, Z., Quarles, L. D., Yates, C. R., Baudry, J., et al. (2018). Ensemble docking to difficult targets in early-stage drug discovery: Methodology and application to fibroblast growth factor 23. *Chem. Biol. Drug Des.* 91 (2), 491–504. doi:10.1111/cbdd.13110
- Venkata Subbiah, H., Ramesh Babu, P., and Subbiah, U. (2020). *In silico* analysis of non-synonymous single nucleotide polymorphisms of human DEFB1 gene. *Egypt. J. Med. Hum. Genet.* 21 (1), 66–69. doi:10.1186/s43042-020-00110-3
- Vladilo, G., and Hassanali, A. (2018). Hydrogen bonds and life in the universe. *Life* 8 (1), 1. doi:10.3390/life8010001
- Vornholt, E., Drake, J., Mamdani, M., McMichael, G., Taylor, Z. N., Bacanu, S. A., et al. (2021). Identifying a novel biological mechanism for alcohol addiction associated with circRNA networks acting as potential miRNA sponges. *Addict. Biol.* 26 (6), e13071. doi:10.1111/adb.13071
- Wang, Q., Mehmood, A., Wang, H., Xu, Q., Xiong, Y., and Wei, D.-Q. (2019). Computational screening and analysis of lung cancer related non-synonymous single nucleotide polymorphisms on the human Kirsten rat sarcoma gene. *Molecules* 24 (10), 1951. doi:10.3390/molecules24101951

- Wang, Z., Huang, C., Lv, H., Zhang, M., and Li, X. (2020). *In silico* analysis and high-risk pathogenic phenotype predictions of non-synonymous single nucleotide polymorphisms in human Crystallin beta A4 gene associated with congenital cataract. *PLoS one* 15 (1), e0227859. doi:10.1371/journal.pone.0227859
- Wang, S., Jain, A., Novales, N. A., Nashner, A. N., Tran, F., and Clarke, C. F. (2022). Predicting and understanding the pathology of single nucleotide variants in human COQ genes. *Antioxidants* 11 (12), 2308. doi:10.3390/antiox11122308
- Wang, W., Su, X., Liu, D., Zhang, H., Wang, X., and Zhou, Y. (2022). Predicting DNA-binding protein and coronavirus protein flexibility using protein dihedral angle and sequence feature. *Proteins Struct. Funct. Bioinforma.* 91, 497–507. doi:10.1002/prot.26443
- Yazar, M., and Özbek, P. (2021). *In silico* tools and approaches for the prediction of functional and structural effects of single-nucleotide polymorphisms on proteins: An expert review. *OMICS A J. Integr. Biol.* 25 (1), 23–37. doi:10.1089/omi.2020.0141
- Zafar, I., Rubab, A., Aslam, M., Ahmad, S. U., Liyaqat, I., Malik, A., et al. (2022). Genome-wide identification and analysis of GRF (growth-regulating factor) gene family in *Camila sativa* through *in silico* approaches. *J. King Saud University-Science* 34 (4), 102038. doi:10.1016/j.jksus.2022.102038



OPEN ACCESS

EDITED BY

Khurshid Ahmad,
Yeungnam University, Republic of Korea

REVIEWED BY

Sevgi Gezici,
University of Gaziantep, Türkiye
Shoib Baba,
Department of Education, India

*CORRESPONDENCE

Mohammed Bourhia,
✉ bourhiamohammed@gmail.com
Rohit Sharma,
✉ rohisharma@bhu.ac.in

RECEIVED 26 February 2023

ACCEPTED 12 April 2023

PUBLISHED 26 April 2023

CITATION

Ali S, Noreen A, Qamar A, Zafar I, Ain Q,
Nafidi H-A, Bin Jordan YA, Bourhia M,
Rashid S and Sharma R (2023), *Amomum
subulatum*: A treasure trove of anti-
cancer compounds targeting
TP53 protein using *in vitro* and *in
silico* techniques.
Front. Chem. 11:1174363.
doi: 10.3389/fchem.2023.1174363

COPYRIGHT

© 2023 Ali, Noreen, Qamar, Zafar, Ain,
Nafidi, Bin Jordan, Bourhia, Rashid and
Sharma. 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.

Amomum subulatum: A treasure trove of anti-cancer compounds targeting TP53 protein using *in vitro* and *in silico* techniques

Sadaqat Ali¹, Asifa Noreen², Adeem Qamar³, Imran Zafar⁴,
Quratul Ain⁵, Hiba-Allah Nafidi⁶, Yousef A. Bin Jordan⁷,
Mohammed Bourhia^{8*}, Summya Rashid⁴ and Rohit Sharma^{9*}

¹Medical Department, DHQ Hospital Bhawalnagr, Punjab, Pakistan, ²Department of Chemistry, Riphah International University, Faisalabad, Pakistan, ³Department of Pathology, Sahiwal Medical College Sahiwal, Punjab, Pakistan, ⁴Department of Bioinformatics and Computational Biology, Virtual University of Pakistan, Punjab, Pakistan, ⁵Department of Chemistry, Government College Women University, Faisalabad, Pakistan, ⁶Department of Food Science, Faculty of Agricultural and Food Sciences, Laval University, Quebec City, QC, Canada, ⁷Department of Pharmaceuticals, College of Pharmacy, King Saud University, Riyadh, Saudi Arabia, ⁸Laboratory of Chemistry and Biochemistry, Faculty of Medicine and Pharmacy, Ibn Zohr University, Laayoune, Morocco, ⁹Department of Rasa Shastra and Bhaishajya Kalpana, Faculty of Ayurveda, Institute of Medical Sciences, Banaras Hindu University, Varanasi, Uttar Pradesh, India

Cancer is a primary global health concern, and researchers seek innovative approaches to combat the disease. Clinical bioinformatics and high-throughput proteomics technologies provide powerful tools to explore cancer biology. Medicinal plants are considered effective therapeutic agents, and computer-aided drug design (CAAD) is used to identify novel drug candidates from plant extracts. The tumour suppressor protein TP53 is an attractive target for drug development, given its crucial role in cancer pathogenesis. This study used a dried extract of *Amomum subulatum* seeds to identify phytochemicals targeting TP53 in cancer. We apply qualitative tests to determine its phytochemicals (Alkaloid, Tannin, Saponin, Phlobatinin, and Cardic glycoside), and found that alkaloid composed of $9.4\% \pm 0.04\%$ and Saponin $1.9\% \pm 0.05\%$ crude chemical constituent. In the results of DPPH Analysis *Amomum subulatum* Seeds founded antioxidant activity, and then we verified via observing methanol extract (79.82%), BHT (81.73%), and n-hexane extract (51.31%) found to be positive. For Inhibition of oxidation, we observe BHT is 90.25%, and Methanol (83.42%) has the most significant proportion of linoleic acid oxidation suppression. We used diverse bioinformatics approaches to evaluate the effect of *A. subulatum* seeds and their natural components on TP53. Compound-1 had the best pharmacophore match value (53.92), with others ranging from 50.75 to 53.92. Our docking result shows the top three natural compounds had the highest binding energies (-11.10 to -10.3 kcal/mol). The highest binding energies (-10.9 to -9.2 kcal/mol) compound bonded to significant sections in the target protein's active domains with TP53. Based on virtual screening, we select top phytochemicals for targets which highly fit based on pharmacophore score and observe these compounds exhibited potent antioxidant activity and inhibited cancer cell inflammation in the TP53 pathway. Molecular Dynamics (MD) simulations indicated that the ligand was bound to the protein with some significant conformational changes in the protein structure. This study provides novel insights into the development of innovative drugs for the treatment of cancer disorders.

KEYWORDS

Amomum subulatum, TP53, cancer, anti-cancer compounds, pharmacophore modeling, high-throughput techniques, bioinformatics

1 Introduction

Cancer is a multifaceted ailment that has the potential to impact any anatomical structure, and it is the result of a combination of genotypic, environmental, and lifestyle determinants (Pandya et al., 2021; Mazhar et al., 2023). Although cancer may manifest in any individual, specific predisposing factors, such as advanced age, familial history, and exposure to carcinogenic substances, can augment the likelihood of developing the disease. Cancerous cells are atypical cells that multiply and proliferate in an unbridled fashion (Hossain et al., 2022). They can generate neoplasms, infiltrate adjacent tissues, and disseminate to other body regions via a phenomenon known as metastasis. Cancerous cells can disrupt the normal functioning of organs and tissues in the body, resulting in various symptoms and complications. Managing cancer is challenging because it isn't a single disease but a group of illnesses with different causes, such as evidence from clinical presentations and treatment options (Lu et al., 2020). The treatment of cancers varies depending on their type and stage. Cutaneous carcinoma and haematological malignancies require special treatments. Research in the field of cancer is ongoing, with new therapies and curative methods being developed continuously. The most promising areas of cancer research currently include immunotherapy, which uses the immune system to target and destroy cancer cells, and personalized medicine, which tailors treatment to the unique genomic and molecular characteristics of an individual's cancer. These innovative approaches can potentially transform cancer treatment, leading to more effective and precise therapies that may improve outcomes and increase survival rates (Kaminski et al., 2020). Prevention is equally critical in combatting the ubiquitous impact of cancer, and this includes efforts to reduce exposure to known carcinogens, such as tobacco smoke and UV radiation, as well as promoting healthy lifestyles, such as regular exercise and a balanced diet (Balwan and Kour, 2021; Klein et al., 2022). In-depth, cancer is a complex and challenging disease. Still, with ongoing research, prevention efforts, and advances in treatment provides a hope for improving outcomes and reducing the impact of this disease on individuals and communities worldwide.

The p53 gene is responsible for encoding a vital transcription factor protein called p53 (Sharma et al., 2023), plays a crucial role in regulating essential cellular mechanisms such as DNA repair, cell cycle arrest, and programmed cell death (apoptosis). Upon sensing signals of DNA damage or similar stressful stimuli, the p53 protein can galvanize or inhibit the expression of sundry downstream genes, predicated on the prevailing cellular ambience. The p53 protein interacts with a vast spectrum of alternative proteins to form a labyrinthine network of interplays, which help to govern its activity and fuse its reaction to distinct stress signals (Chen L et al., 2020). These interactions and networks are complex and context-dependent, with a range of downstream targets and regulators that can influence its activity and function. Some essential proteins and networks associated with p53 include MDM2,

ATM/ATR, p21, BAX, and NF- κ B (Karimian et al., 2019; Samavarchi Tehrani et al., 2019; Marei et al., 2021). Anomalies in the TP53 gene have the potential to compromise the efficacy of the p53 protein, thereby inciting uncontrolled cellular multiplication and a heightened susceptibility to cancer. However, therapeutic techniques that target the p53 pathway are being pioneered to ameliorate curative alternatives for cancer patients. These comprise minuscule molecule medications that can rejuvenate or stabilize the p53 protein, gene manipulation methodologies that introduce operative copies of the TP53 gene into neoplastic cells, and immunotherapeutic manoeuvres that utilize the immune system to attack cancer cells bearing TP53 gene aberrations. A comprehensive comprehension of the interplays and networks linked with p53 is indispensable for the evolution of precision therapies and the augmentation of clinical outcomes for patients who have cancer or related ailments (Otohinoyi et al., 2022).

Herbal medicine utilizes natural compounds that interact harmoniously with our bodies, while conventional medicine relies on synthesized and artificial molecules. Herbal medicine practitioners stress the importance of combining various medicinal ingredients instead of isolating a single component (Xijun et al., 2016; Ali et al., 2018). Herbal medications are often preferred by individuals who cannot tolerate pharmaceuticals or experience adverse reactions to them. Natural medicine target the underlying cause of symptoms in certain conditions, whereas medications only alleviate the symptoms. Synthetic treatments are associated with a range of adverse effects, from minor to severe, and cause harm to the body's internal organs and external features, including the skin, hair, and teeth. Previous research {Sharifi-Rad, 2020 #2712} has shown that herbal medicine may be slower in producing results, but the outcome is healthier when using natural compounds and treatments. Artificial processing of natural substances can result in toxicity concerns, and ensuring that the proper dosage is administered is essential to guarantee safety. However, if herbal medicine is used carefully, it may be possible to develop more effective and less toxic pharmaceuticals than those solely relying on refined medications (Saad et al., 2017).

Traditional medicine heavily depends on plants and their bioactive constituents to treat various ailments, including cancer (Matowa et al., 2020). An estimated two-thirds to three-quarters of the global population utilize herbal remedies for therapy, leading to an upsurge in interest in studying phytomedicines and their active biological properties. Phytochemicals are non-nutrient bioactive molecules found in plant-based foods like fruits and vegetables and have been associated with reduced risks of chronic diseases (Wu et al., 2021). Plants contain approximately 25,000 terpenoids, 12,000 alkaloids, 8,000 phenolics, and other compounds, providing a bountiful source of active molecules. Understanding how essential compounds like flavonoids, chlorogenic acids, alkaloids, carotenoids, minerals, and toxic substances affect health results can be learned from data on these compounds. Recent studies by (de Carvalho and Conte-Junior, 2021) indicate that the potential health advantages of phytochemicals found in fruits and vegetables

could be more significant than previously assumed. Antioxidants, for example, can help mitigate oxidative stress caused by free radicals, which can contribute to the progression of chronic diseases. However, many phytochemicals remain unidentified, requiring identification and measurement before assessing their health risks. *Amomum subulatum* is a short-lived herbal plant widely used globally for its culinary and medicinal properties. Its health-promoting effects have been documented in Ayurveda and traditional Chinese medicine, and *in-vivo* and *in-vitro* studies have validated its anticancer potential.

The domain of cancer research has undergone a revolution through *in-silico* approaches, which present a cost-effective and time-efficient alternative to traditional experimental methods (Rahman et al., 2022). *In-silico* approaches involve describing and simulating biological processes, predicting drug-target interactions, and using computational tools and algorithms to develop new antitumor drugs. Researchers can explore vast chemical regions and identify potential cancer-targeting drugs by utilizing molecular docking, molecular dynamics models, and virtual screening methods. These methods also enable the prediction of medication metabolic and toxicity characteristics, thereby reducing the risk of failure during clinical studies. Moreover, *in-silico* methods facilitate the detection of genomic and epigenetic alterations in cancer cells, which can aid in developing more personalized treatments. These techniques have shown tremendous potential in identifying new targets and creating innovative cancer therapies. In organic plant molecules, bioinformatics plays a critical role in detecting and treating cancer (Tabrez et al., 2022). Using computational tools and databases, scientists can proficiently explore the chemical space of natural plant compounds and their interactions with cancerous targets. Bioinformatics approaches also allow the discovery of potential synergistic amalgamations of plant compounds that enhance efficacy and mitigate toxicity (Fokunang and Fokunang, 2022). By scrutinizing the gene expression profiles of cancerous cells, bioinformatics can help recognize biomarkers that predict the response to plant-based therapies (Ahmad et al., 2022; Satpathy, 2022). Additionally, bioinformatics approaches can expedite the repurposing of existing drugs for cancer treatment and the detection of new targets for drug development. All in all, bioinformatics offers a powerful platform for leveraging the therapeutic potential of natural plant compounds for cancer therapy.

The principal aim of our study is to explore the feasibility of employing novel phytochemicals originating from the *A. subulatum* seeds as anti-cancer drugs by targeting the TP53 receptor. To attain this objective, we implement computational techniques for structure prediction and validation, molecular docking, and simulation studies to scrutinize the interaction between the selected phytochemicals and TP53. The investigation includes cherry-picking a set of phytochemicals from the *A. subulatum* plant with the potential to combat cancer. It will then assess their three-dimensional structures and accuracy verification through computational methods. The molecular docking analyses are carried out to scrutinize the binding interactions between the chosen phytochemicals and TP53. The primary goal is to recognize new phytochemicals with high binding affinity and specificity towards TP53. The outcomes

obtained from the molecular docking studies are further endorsed through simulation studies to ascertain that the binding interactions are steady and biologically significant. To evaluate the potential of phytochemicals as anti-cancer medications, their ADMET characteristics, including uptake, diffusion, metabolism, elimination, and toxicity, are closely examined. Based on the results of these studies, phytochemicals with optimal pharmacokinetic and toxicity profiles are identified. In addition, by investigating the interactions of these phytochemicals with the TP53 receptor and analyzing their ADMET characteristics, researchers aim to gain new insights into the potential of phytochemicals derived from the *A. subulatum* seeds as anti-cancer agents. The findings of this study could pave the way for developing more sophisticated and effective cancer treatments.

2 Materials and methods

2.1 Plant materials

The *A. subulatum* seeds were randomly gathered from a particular location of the agricultural lands in Faisalabad. They were washed and dried at room temperature to ensure an optimal outcome and prevent impurities. For powder extraction, 75 g of the seeds were pulverized, mixed with Methanol, and agitated on an orbital shaker for 7 mins using the maceration method. As per the method of the earlier researcher (Kumar et al., 2022), we used maceration permits the solvent to infiltrate the plant material and dissolve the targeted components. The aqueous filtrate, which consists of the desired compounds extracted from the plant material, was attained by sieving the mixture to eliminate unwanted particles or impurities. To precisely measure the number of active compounds in the *A. subulatum* seeds, the extracted material was converted into a solid form utilizing an evaporation technique, which included heating the methanol solution to evaporate the solvent and leaving behind the solid components of the *A. subulatum* seed extraction. The aim of obtaining a solid form of the extraction is to simplify weighing and quantifying, thus making it more convenient for analysis.

2.2 Phytochemical Screening

We performed Phytochemical Screening to determine the presence of particular chemical compounds in the *A. subulatum* seeds, utilizing traditional techniques by the approach employed by preceding investigators (Cabero Pérez, 2020). These constituents consist of tannins, flobatannins, saponins, steroids, and terpenoids, which are all important bioactive compounds frequently present in plants and can offer various health advantages. To discern the presence of tannins, the pulverized, desiccated seeds of *A. subulatum* were subjected to boiling in water, and the resultant admixture was filtered. A minute quantity of 0.1% FeCl₃ was then appended, and the manifestation of a brown-green tint denoted the existence of tannins. Tannins are a polyphenolic compounds recognized for their antioxidative and antimicrobial characteristics (Leite et al.,

2021). To identify flobatonins, a 0.5 g specimen of the seeds was boiled with 1% aqueous HCl and the presence of flobatonins was indicated by the emergence of a crimson precipitate. Flobatonins are a type of tannin that is present in various plant species and are acknowledged for their astringent properties. To detect saponins, 2 g of the specimen was heated in distilled water to extract the compounds. After filtration, 5 mL of distilled water was added to 10 mL of the filtrate, and the mixture was vigorously shaken to produce stable foam. Olive oil droplets were added to the frothy mixture and promptly mixed to form an emulsion. Saponins are a glycoside compound with diverse pharmacological activities, including anti-inflammatory and anti-cancer effects (El Aziz et al., 2019). 0.5 g of the ethanol extract was added into test tube with 2 mL of H₂SO₄ and 2 mL of acetic anhydride to screen for steroids. The change in colour from violet to green indicated the presence of steroids, a type of lipid molecule with diverse physiological functions, including the regulation of metabolism, immune response, and development (Morales-Lázaro et al., 2019). For terpenoids analysis, 5 mL of each extract was delicately mixed with 2 mL of chloroform and 3 mL of concentrated sulfuric acid to generate a layer. The reddish-brown staining of the interface suggested that terpenoids are a large class of natural compounds known for their diverse biological activities, including anti-inflammatory, anti-cancer, and anti-microbial effects (Fan et al., 2023).

2.3 DPPH analyses

DPPH radical scavenging is a commonly employed procedure for gauging the antioxidative efficacy of natural substances, including plant-based extracts as per earlier researchers (Amarowicz et al., 2004). DPPH is a stable radical that exhibits a unique absorption peak at 517 nm and can be rendered neutral by antioxidants, leading to a decrement in absorption intensity. To evaluate the antioxidant potential of *A. subulatum* seeds, their ability to neutralize DPPH (2,2-diphenyl-1-picrylhydrazyl) radicals were investigated. To achieve this, a solution of DPPH in Methanol was mixed and resulting mixture was then incubated for 30 min, after which the absorbance was measured at a wavelength of 517 nm. The scavenging ability of the extract against DPPH radicals was inferred from the reduction in absorption of the reaction mixture, and the percentage of DPPH radical scavenging activity was calculated accordingly. In this investigation, the synthetic antioxidant BHT was utilized as a positive control to compare the antioxidative activity of *A. subulatum* seeds with a recognized antioxidant. A higher percentage of DPPH radical scavenging activity corresponded to a more tremendous antioxidative potential of the extract (Pyrzynska and Pękal, 2013).

2.4 Reducing power determination

To assess the plant specimen's ability to reduce, a modified rendition of the technique, initially presented by Matanjun et al. (2008), was utilized. Various dissolvable extractions of *A. subulatum*

seeds were concocted at convergences that fluctuated between 2.5 and 1.0 mg/mL. Sodium phosphate buffer (5.0 mL) and potassium ferricyanide (5.0 mL, 1.0% in 0.2 M, pH 6.6) were added to every extraction. The resultant mixture was then left to incubate at a temperature of 50°C for 20 min. After incubation, the mixture was centrifuged at 980 g and 5°C for 10 min. The resulting supernatant was then treated with 5 mL of 10% trichloroacetic acid via applying second round of centrifugation. The top layer of the resultant solution (roughly 5 mL) was diluted with 5 mL of distilled water and 1.0 mL of 0.1% ferric chloride. The spectrophotometer was then used to measure the absorbance of the solution at a wavelength of 700 nm. This protocol was conducted in order to ascertain the plant sample's reducing power (Pasaribu et al. (2021).

2.5 Hemolytic activity

The hemolytic efficacy of *A. subulatum* seeds was assessed using a customized methodology as delineated by (Pasaribu et al. (2021). To prevent coagulation, 3 mL of human blood was collected into a heparinized tube and centrifuged at 850 g for 5 min. The resulting supernatant was discarded, and the red blood cells (RBCs) were washed three times with 5 mL of chilled sterile isotonic phosphate-buffered saline (PBS) solution at pH 7.4. The RBCs were then suspended in 20 mL of chilled PBS and counted using a hemacytometer. The RBC count was adjusted to 7.068×10^8 cells per mL to prepare for the assay. Plant extracts (20 μ L) were instilled into 2 mL Eppendorf tubes, succeeded by adding diluted blood cell suspension (180 μ L). The samples were incubated for 35 min at 37°C, followed by placing the tubes on ice for 5 min and centrifuging for 5 min at 1,310 g. After centrifugation, the supernatant (100 μ L) was withdrawn from each tube and diluted with chilled PBS (900 μ L). The mixture from each Eppendorf (200 μ L) supernatant was then added to 96-well plates. For each test, 0.1% Triton X-100 was utilized as a positive control and PBS as a negative control. For hemolysis %, the absorbance was measured at 576 nm via applying formula: % hemolysis = (sample absorbance/control absorbance) \times 100. To maintain consistency during the assay, all Eppendorf tubes were kept on ice.

2.6 Structure prediction

The amino acid sequence of TP53, having accession number (P04637), was obtained from the UniProt database (<https://www.uniprot.org/>) in the FASTA format. Then we performed a BLASTp query against the Protein DataBank (<https://www.rcsb.org/search>) to locate fitting templates (6IU7, 6IUA, 6MY0, 5Z78, 5ZCJ, 6I3V, 7LIN, and 7LIO) for the desired protein. Utilizing SWISS model (<https://swissmodel.expasy.org/>) and ITASSER (<https://zhanggroup.org/I-TASSER/>) server, multiple 3D structures were prognosticated and validated. To appraise the quality of the envisaged 3D structures, numerous online validation tools, including ERRAT (<https://saves.mbi.ucla.edu/>), Verify3D, and Rampage (<http://www.scfbio-iitd.res.in/software/proteomics/protasv.jsp>), were employed. Chimaera was

afterwards to optimize the 3D structures of the selected compounds intermingled with TP53. While conducting the simulation studies, the protein secondary structure elements (PSSE) were thoroughly scrutinized and determined. These techniques facilitated the foreknowledge and validation of TP53's 3D structures and the revelation of undiscovered phytocompounds possessing potential anti-cancer properties.

2.7 Docking analysis

For in-silico analysis, we utilized a reservoir of autochthonous compounds obtained from the Asinex database (<https://www.asinex.com/>). We select phytocompounds based on their pharmacophore fit score using AutoDock Vina and employed LigandScout for pharmacophore modelling (Wolber and Langer, 2005; Heider et al., 2022). We utilized AutoDock Vina for targeted molecular docking and meticulously examined and visualized the docking investigations employing UCSF Chimera v1.12 and Discovery Studio. We identified 2D structures from databases like PubChem (<https://pubchem.ncbi.nlm.nih.gov/>) to create plant-based molecules and employed ChemDraw and Chimera. We performed 100 docking runs for each docking experiment and added polar hydrogen atoms to all targeted proteins. We chose TP53 as the grid size for docking investigations. To ensure drug-likeness, we rigorously examined Lipinski's Rule of Five (RO5) as per the method of earlier researcher Chen X et al. (2020), by utilizing the online mcule server (<https://mcule.com/>) and computed the drug characteristics of all specified compounds. To evaluate the ADMET properties (Absorption, Distribution, Metabolism, Excretion, and Toxicity), we utilized the AdmetSAR tool (<http://lmmd.ecust.edu.cn/admetSar2>) to scrutinize the bioavailability of the selected compounds.

2.8 Molecular dynamic simulation

Molecular Dynamic (MD) simulations have investigated the interplay between the protein and ligand (Salo-Ahen et al., 2020). The preeminent docked complex was handpicked, and various modules from the Schrodinger suite were utilized to execute the MD simulations (Páll et al., 2020). An NPT ensemble was implemented to prepare a rudimentary simulation milieu for the docked systems, held at a temperature of 300 K for a duration of 100 nanoseconds (ns) as per the method of earlier researcher Prunotto (2020). In protein-ligand docking, the docked conformers underwent root mean square deviation (RMSD) and root mean square fluctuation (RMSF) plot analysis to evaluate the stability and fluctuations of the protein-ligand complex during the MD simulation. The RMSD analysis quantifies the structural aberration between the docked complex and the MD trajectory, whereas the RMSF analysis gauges the residue-wise fluctuation of the complex throughout the simulation as per the methods of earlier researchers (Rather et al., 2020; Tabti et al., 2023). The MD simulation outcomes were assessed utilizing various visualization tools, including VMD and Chimera, to acquire perspicacity into the protein-ligand

interactions and recognize prospective binding sites and vital residues implicated in the interaction.

3 Results and discussion

3.1 Phytochemical Screening of *Amomum subulatum* seeds

This study scrutinised the existence of phytochemicals in *A. subulatum* seeds by conducting various qualitative tests to assess their plausible therapeutic advantages. Qualitative phytochemical analysis was executed, which divulged the presence of tannins, saponin, steroids, and cardiac glycosides in the seeds, whereas phlorotannins, terpenoids, and alkaloids were not detected. Phytochemicals such as saponin and tannins are renowned for their inherent antibiotic properties and are frequently employed to combat pathogenic strains. Alkaloids exhibit confirmed antioxidant activities and have been noted as efficacious therapeutic agents in ethnomedicine. Tannin has been implemented as an active compound in pharmaceuticals and beverages, primarily because of its antioxidant effects, as per the investigation by Saxena et al. (2013). Flavonol glycosides, among others, are potent inhibitors of lipid peroxidation. Steroids can scavenge free radicals and convert them into more stable molecules, arresting the chain reaction. Cardiac glycosides have been utilized to treat moderate to severe myocardial infarctions by inhibiting the Na⁺/K⁺ pump and increasing Ca⁺⁺ concentrations. These phytochemicals offer myriad health benefits, such as anticancer properties, cholesterol-lowering effects, promoting strong bones, and boosting the immune system (Felth et al., 2009). Saponins possess many activities, including anti-inflammatory, antifungal, hemolytic, fungistatic, molluscicidal, and foaming properties. The results of this study are consistent with previous studies (Wang et al., 2011; Abbas et al., 2015), suggesting that *A. subulatum* seeds are rich in phytochemical constituents that may be responsible for the antioxidant and anticancer activities of plant-based foods. The present study on *A. subulatum* has identified several therapeutically active components in the plant, as exhibited in Table 1. Alkaloids, tannins, saponins, and cardiac glycosides were detected in the plant, while steroids and terpenoids were absent.

In this study, the crude chemical constituents of *A. subulatum* seeds were quantitatively estimated, as shown in Table 2. At the same time, qualitative parameters were evaluated to distinguish between closely related plant species or varieties with similar pharmacological activities. The plant contained alkaloids, tannins, and saponins, which possess medicinal properties and are used in various antibiotics to treat common pathogenic strains. Previous research has also demonstrated the presence of alkaloids in leafy vegetables, such as the bitter leaf, which have been reported to alleviate headaches associated with hypertension.

Alkaloids are a group of nitrogen-containing chemical substances that frequently have medicinal effects. Alkaloids have a wide range of molecular structures in numerous organisms. *Amomum subulatum* has an estimated alkaloid concentration of

TABLE 1 Qualitative analysis of the phytochemicals of the *Amomum subulatum* seeds extracts.

Plant	Alkaloid	Tannin	Saponin	Steroid	Phlobatinin	Terpenoid	Cardic glycoside
<i>Amomum subulatum</i>	+	+	+	—	+	—	+

TABLE 2 Quantitative estimation of the percentage crude chemical constituents in *Amomum subulatum* seeds extracts.

Chemical constituents of <i>Amomum subulatum</i>	Percentage (%) of crude chemical constituent
Alkaloids	9.4 ± 0.04
Saponin	1.9 ± 0.05

9.40.04%, which is a substantial quantity. This shrub contains alkaloids like piperine, amomumine, and aphylline. Another family of chemical substances discovered in vegetation are saponins. They have a range of biological functions and are distinguished by their capacity to produce steady froths when agitated with water. Researchers have discovered that saponins have antifungal, antibacterial, and anti-inflammatory properties. According to reports, *A. subulatum* contains 1.90.05% saponin. Cardamonin, quercetin, and kaempferol are a few saponins that can be found in this plant. Further research is required to fully comprehend the molecular components' precise impacts and possible applications in *A. subulatum*, which may have therapeutic qualities.

3.2 Percentage yield of plant extracts

The study found that the amount of plant compounds present varied between 2.72 and 3.97 mg/100 g of dried plant material. The methanol extract of *A. subulatum* seeds, which is known to be effective in extracting phytochemical components, yielded the highest amount of compounds. However, the choice of solvent can significantly impact the number of recovered substances since different solvents can extract distinct phytochemical components depending on their charges. Therefore, it is crucial to carefully consider the choice of the solvent when removing plant chemicals to ensure the best output and effectiveness of the extracted components. The percentage yield of plant extracts, which represents the number of separated chemicals produced from a specific volume of starting plant material, is an essential factor to consider in the extraction process. Since various solvents with different phases can extract distinct phytochemical components, the solvent option for plant extraction can substantially impact the number of recovered substances. Plant extract output in this situation differed according to the extraction fluid. For instance, the output for n-hexane was only 1.5, which isn't very high. This is because n-hexane is a non-polar solvent that works well to remove non-polar substances like triglycerides and essential oils. However, removing polar substances like flavonoids and phenolics might not be as efficient.

On the other hand, the output of chloroform was 3.5, which is a significant value. A reasonably neutral fluid like chloroform can be used to remove a variety of chemicals from plant matter successfully.

The use of chloroform should be avoided if feasible because it is a dangerous chemical. With yields of 2.5, acetone and butanol can remove plant chemicals with a tolerable efficiency level. While butanol is a less polar solvent that is effective in extracting non-polar substances like terpenoids and alkaloids, acetone is a polar solvent that removes polar substances like flavonoids and phenolics. According to [Figure 1](#), Methanol had the most incredible output out of the 4. Various phytochemical components can be extracted from plant material using the neutral liquid Methanol. Methanol is poisonous and explosive, so it should be closely regulated and supervised.

3.2.1 Total flavonoids content

[Table 3](#) (A) presents the cumulative flavonoid matter (CFM) detected in arid essences of *A. subulatum* seeds through varying solvents employed for the extraction phase. CFM is expressed in catechin equivalencies (CE) per 100 g of arid core, where our results divulge that CFM values are contingent on the solvent selected for the extraction process. Our data indicate that the CFM for the n-hexane extract was found to be the least at 4.47 ± 0.05 mg/100 g dry extract, whereas the CFM for the methanol extract was the highest at 127.51 ± 0.76 mg/100 g dry extract. The chloroform, ethyl acetate, and n-butanol extracts exhibited CFM values of 32.17 ± 0.52 , 17.48 ± 0.17 , and 25.98 ± 0.50 mg/100 g dry extract, respectively. Our results signify that Methanol is the most proficient solvent for flavonoid extraction from *A. subulatum* seeds, whereas n-hexane is the most inefficient. Flavonoids are vital secondary metabolites in plants that augment plant colouration and exhibit diverse salutary biological activities, including anti-inflammatory, anti-allergic, and anti-cancer properties ([Ekalu and Habila, 2020](#)). Our findings suggest that the solvent preference for plant extraction can significantly influence the cumulative flavonoid content in the extract. Therefore, it is imperative to cautiously consider the solvent selection for each extraction to guarantee optimal efficiency and yield of the extracted flavonoids.

3.2.2 Total phenolics content

Our research study explored the total phenolic content (TPC) of extracts obtained from *A. subulatum* seeds using different solvent mediums, primarily focusing on methanol extraction. TPC was quantified in milligrams of gallic acid equivalents (GAE) per 100 g of dry plant material. Findings showed that the TPC varied considerably across different solvent systems, with the highest TPC

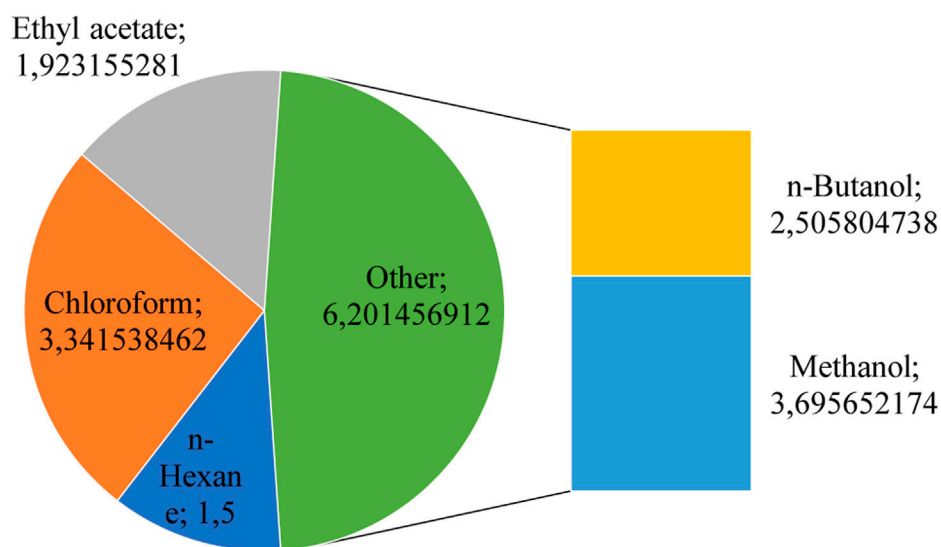


FIGURE 1
Percentage yield of *Amomum subulatum* seeds extracts.

TABLE 3 (A) explore the Total flavonoid contents, (B) Total phenolic contents, (C) DPPH percentage scavenging assay, (D) linoleic acid percentage inhibition and oxidation and (E) explore the hemolytic activity in the percentage of hemolysis in extracts of *Amomum subulatum* seeds extracts.

Labels	3-A	3-B	3-C	3-D	3-E
Sample	Total Flavonoid Contents (CE Mg/100 G) Dry Extracts	Total Phenolic Contents (GAE Mg/100 G) In Dry Seeds Extracts	DPPH Percentage Scavenging	Linoleic Acid Percentage Inhibition and Oxidation	Percentage Of Hemolysis
n-Hexane	4.47 ± 0.05	23.25 ± 0.102	51.31 ± 0.38	23.68 ± 0.53	1.44 ± 0.019
Chloroform	32.17 ± 0.52	111.55 ± 0.136	60.02 ± 0.58	56.92 ± 0.61	3.11 ± 0.024
Ethyl acetate	17.48 ± 0.17	26.95 ± 0.056	72.41 ± 0.76	45.70 ± 0.32	4.19 ± 0.043
n-Butanol	25.98 ± 0.50	125.09 ± 0.101	68.96 ± 0.66	73.50 ± 0.53	6.10 ± 0.05
Methanol	127.51 ± 0.76	134.39 ± 0.589	79.82 ± 0.54	83.42 ± 0.87	8.50 ± 0.072
BHT/Titron X 100	—	—	81.73 ± 0.79	90.25 ± 0.90	99.64 ± 0.92

being recorded in the methanol extract (134.39 ± 0.589 mg GAE/100 g). In contrast, the n-hexane extract recorded the lowest TPC (23.25 ± 0.102 mg GAE/100 g) as elaborated in Table 3 Column (B). Phenolic compounds are well-known for their remarkable antioxidant properties. The higher TPC content in the extracts significantly contributes to improved, reducing power and radical scavenging effect on DPPH radicals (Muzolf-Panek and Stuper-Szablewska, 2021). We used the Folin-Ciocalteu method to measure TPC, a sensitive, interference-free, and rapid method for quantifying the number of phenolics present in plant extracts. Phenolics are extensively found in plants and have been associated with diverse biological activities, including antioxidant activity (Karak, 2019). They act as reducing agents, hydrogen donors, and oxygen quenchers, thereby effectively decreasing oxidative stress. Our analysis concludes that the TPC of the extracts varied significantly across different solvent systems, with methanol

extraction resulting in the highest TPC and phenolic content significantly contributing to their antioxidant activities.

3.3 DPPH analysis of *Amomum subulatum* seeds for antioxidant activity

The DPPH test is a popular technique for evaluating the antioxidant activity of natural substances. It considers a material's capacity to bind free radicals and convert them to a rigid state. The more free radicals the material scavenges, the more antioxidant activity it exhibits. The DPPH% scavenging technique was used in this study to assess the antioxidant potential of different formulations of *A. subulatum* seeds. The results demonstrate variations in the antioxidant activity of *A. subulatum* seed products. The methanol extract (79.82%) and BHT (81.73%), a

synthetic antioxidant used as a standard, had the highest percentage of scavenging activity. The least quantity of scavenging activity was demonstrated by the n-hexane extract (51.31%). These results suggest that the methanol seed extract of *A. subulatum* has considerable antioxidant activity.

The TPC and TFC assays reveal that phenolic and flavonoid compounds are what give the products their antioxidant properties. The preparation's high content of total phenols may be responsible for their better results in terms of reducing power and radical cooling effect on DPPH radicals. The antioxidant qualities of phenolic compounds are due to their ability to give electrons, which allows them to combat free radicals and serve as reducing agents. Our findings of this study point out the significant antioxidant activity of *A. subulatum* seed products, with the methanol extract having the highest antioxidant capacity. These discoveries depend on developing organic antioxidants and using *Amomum subulatum* seeds in foods and medications.

The proportion of DPPH free radical scavenging activity of various solvent preparations of *A. subulatum* seeds, such as n-hexane, chloroform, ethyl acetate, n-butanol, Methanol, and BHT, is shown in Table 3 Column (C) as a synthetic antioxidant. The methanol extract had the most significant proportion of antioxidant action (79.82%), followed by BHT (81.73%). The reducing activity of the n-hexane extract was the lowest (51.31%). According to the study, the antioxidant potential of the compounds, especially phenolics, relies on their capacity to function as hydrogen-donating species. Their lowering ability may also influence the antioxidant activity of the compounds. The findings indicate that various liquid preparations of *A. subulatum* seeds had different antioxidant effects on DPPH (Rafique et al., 2020). The total phenol concentration, flavonoid content, radical scavenging activity, and % suppression of linoleic acid oxidation were all at their highest levels in the methanol extract. Due to their capacity to serve as a supply of antioxidant compounds, the results imply that the seeds of *A. subulatum* have strong antioxidant activity.

The results demonstrate that the DPPH test is an accurate technique for detecting phenolic and flavonoid components in natural goods and the antioxidant capabilities of mass preparations. In percentage words, the extract's capacity to absorb free radicals was concentration-dependent, and this capacity increased with extract volume and hydroxylation level. The methanol extract had the highest scavenging activity, but it had slightly lower antioxidant activity than the synthetic antioxidant BHT. The study concludes that *A. subulatum* seeds have a variety of antioxidant impacts on DPPH, with the methanol extract having the strongest lowering action. It is advised to use the DPPH test as a valuable tool to assess the antioxidant capabilities of mass products. The study highlights the potential of *A. subulatum* seeds as an antioxidant chemical source (Pyrzynska and Pękal, 2013).

3.4 Inhibition of oxidation

The current study work's findings imply that preparation from *A. subulatum* seeds has antioxidant qualities that can

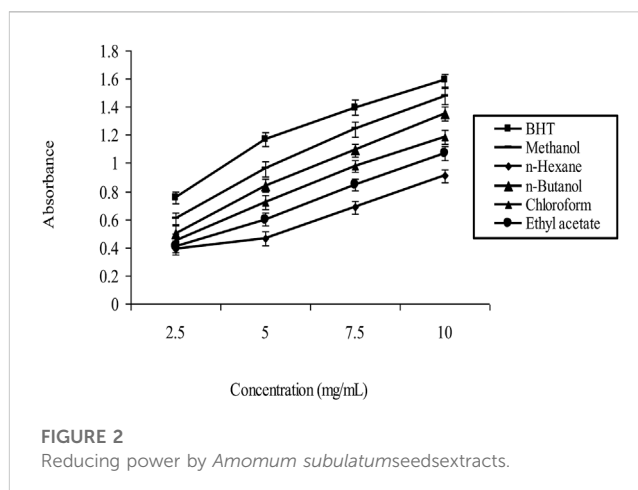


FIGURE 2
Reducing power by *Amomum subulatum* seed extracts.

prevent oxidative damage, a phenomenon that can harm biological cells. Linoleic acid was used as a model system in the research to evaluate the degree to which compounds inhibited oxidation. According to the study's findings, the various *A. subulatum* seed preparations inhibited oxidation to differing degrees. According to Table 3 Column (D), methanol extract had the most significant proportion of oxidation suppression, suggesting the most potent antioxidant activity. This outcome might result from the methanol extract having a higher quantity of phytochemical components.

In contrast, n-hexane extract exhibited the lowest percentage of suppression, likely due to the non-polar solvent's reduced phenolic and other phytochemical components content. It's important to note that BHT, a manufactured antioxidant frequently employed in the food business, was used to compare the antioxidant activity of the compounds to that of BHT. According to the findings, only the methanol extract's antioxidant activity was marginally lower than that of BHT's, while that of all other extracts was noticeably lower. The findings imply that the extract from *Amomum subulatum* seeds has antioxidant qualities that can prevent lipid degradation. The of phytochemical components in the extract and the extraction liquid affect how much suppression occurs.

The proportion suppression of linoleic acid oxidation in various liquids is shown in Table 3 Column (D), along with a positive control of BHT (butylated hydroxytoluene). A mixed fatty acid called linoleic acid is vulnerable to oxidation, which creates free radicals that can harm cells and tissues through oxidative stress. N-hexane, formaldehyde, ethyl acetate, n-butanol, and Methanol were among the solvents examined. Methanol (83.42%) has the most significant proportion of linoleic acid oxidation suppression, followed by n-butanol (73.50%), ethyl acetate (45.70%), chloroform (56.92%), and n-hexane (23.68%). The percentage reduction for the positive control, BHT, is 90.25%. These findings imply that the examined solvents can block linoleic acid oxidation to varying degrees, with Methanol and n-butanol exhibiting the greatest inhibiting action. The well-known antioxidant BHT is most suppressed, demonstrating its potent antioxidant activity. These findings may significantly impact the application of solvents and antioxidants in the culinary, medicinal, and skin care sectors.

3.5 Reducing power

According to the study's results, *A. subulatum* seeds' natural antioxidant properties appear significant. They can be used to prevent the harm that free radicals and lipid breakdown cause to living cells. To evaluate how well the various formulas stopped lipid breakdown, linoleic acid was used as a model system. As shown in [Figure 2](#), our findings revealed that the methanol extract had the largest% oxidation decrease, suggesting the most potent antioxidant activity. The greater quantity of phytochemical components in the methanol solution may be responsible for this outcome. Due to the low content of phenolic and other phytochemical components in this non-polar fluid, the n-hexane extract had the lowest proportion of oxidation suppression. These findings imply that the number of phytochemical components in the extract, which can impact the extract's antioxidant activity, can be influenced by the pH of the liquid used during the extraction process.

It is also notable that BHT, a synthetic antioxidant frequently used in the food business, was used to measure the goods' antioxidant activity. The findings revealed that all compositions had considerably reduced antioxidant activity than BHT, except the methanol extract, which had marginally lower antioxidant training than BHT. This contrast demonstrates the effectiveness of using natural antioxidants, such as those present in *A. subulatum*, instead of manufactured ones. The combined findings imply that the seeds combination of *A. subulatum* may be used as a natural supply of antioxidants to prevent lipid degradation and free radical harm. Additional investigation is required to determine the most efficient extraction technique and pinpoint the particular phytochemical components accountable for the purported antioxidant activity.

3.6 Hemolytic activity for cytotoxicity assay

According to our findings, *A. subulatum* seeds have naturally existing antioxidants that can stop lipid peroxidation, a process that can be harmful to living cells. When using linoleic acid as a model system to assess the ability of the different extracts to avoid oxidation, the study results showed that the methanol extract had the highest antioxidant activity. The greater concentration of phytochemical components in the methanol extract may cause the extract's enhanced antioxidant activity, which is noted in [Table 3](#) Column 3 (E). The decreased antioxidant activity observed in the n-hexane extract is presumably due to the lower concentration of phenolic and other phytochemical components in this non-polar liquid. These findings suggest that the extraction solvent may affect the amount of phytochemical components in the extract and, as a result, on the extract's antioxidant activity.

Comparing the antioxidant activity of the molecules to that of BHT, a synthesised antioxidant, provides a helpful insight into the potential use of natural antioxidants as an alternative to manufactured antioxidants. The potential that natural antioxidants could be used as effective alternatives to synthetic antioxidants is raised by the methanol extract's slightly decreased antioxidant activity compared to BHT. However, the significantly reduced antioxidant activity discovered in the other formulations suggests that additional research is needed to determine the most effective extraction method and to identify the same phytochemical

elements responsible for the reported antioxidant activity. According to our research findings, *A. subulatum* seeds contain natural antioxidants that can stop lipid degradation and protect cells from damage caused by free radicals. More research is necessary to fully understand the potential of these natural antioxidants and determine the most efficient extraction and application techniques.

3.7 Virtual screening analyses

Computer-aided drug design (CADD) has emerged as a promising approach for discovering novel plant-based cancer treatments. Computational drug design involves using computational methods to identify potential drug candidates from libraries of natural plant compounds. Natural compounds with anticancer properties can be starting points for drug design studies. The TP53 protein, which plays a crucial role in cancer, is a key therapeutic target. In this study, homology-based modelling was used to create 3D structures of the TP53 protein based on its sequence and protein templates with high sequence identity and query coverage were selected, as shown in [Figure 3](#). The top 10 lead hits from virtual screening analyses were subjected to pharmacophore modelling to identify the 3D arrangement of chemical features necessary for these compounds to interact with the TP53 protein. The resulting pharmacophore models could guide the design of new small molecule inhibitors of the TP53 protein for the treatment of cancer.

Pharmacophore refers to a 3D arrangement of chemical features necessary for a molecule to interact with a specific biological target. In the case of virtual screening of P53 protein, the top 10 lead hits are likely small molecules that have been predicted to bind to the protein based on their expected shape, electrostatic properties, and other characteristics. Based on previous studies of the P53 protein, it is known that the protein has several pockets and grooves on its surface that are important for interactions with other molecules. These pockets and grooves have specific shapes and electrostatic properties that can be used to guide the design of small molecule inhibitors. The pharmacophore matches findings for the chosen substances revealed encouraging outcomes, as shown in [Table 4](#). Out of all the implications, Compound-1 had the best pharmacophore match value (53.92). In summary, the pharmacophore of the top 10 lead hits from virtual screening analyses of P53 protein likely includes features that allow them to interact with specific pockets and grooves on the protein's surface.

A pharmacophore is a molecular model that describes the spatial arrangement of the functional groups in a molecule that is responsible for its biological activity. It can identify small molecules that share similar structural features to known ligands or bind to a specific target. In this case, a pharmacophore model was generated for the top 50 lead hits identified in the virtual screening study, and the highest-scoring compounds among the top 10 were selected. The table lists with the highest score for each of the top 10 compounds identified in the virtual screening study. The score reflects the fit of the compounds to the pharmacophore model, with higher scores indicating a better fit. Based on the table, the top 10 lead hits have scores ranging from 50.75 to 53.92, with compound-1 having the highest score. These compounds were selected as the most promising hits based on their high scores

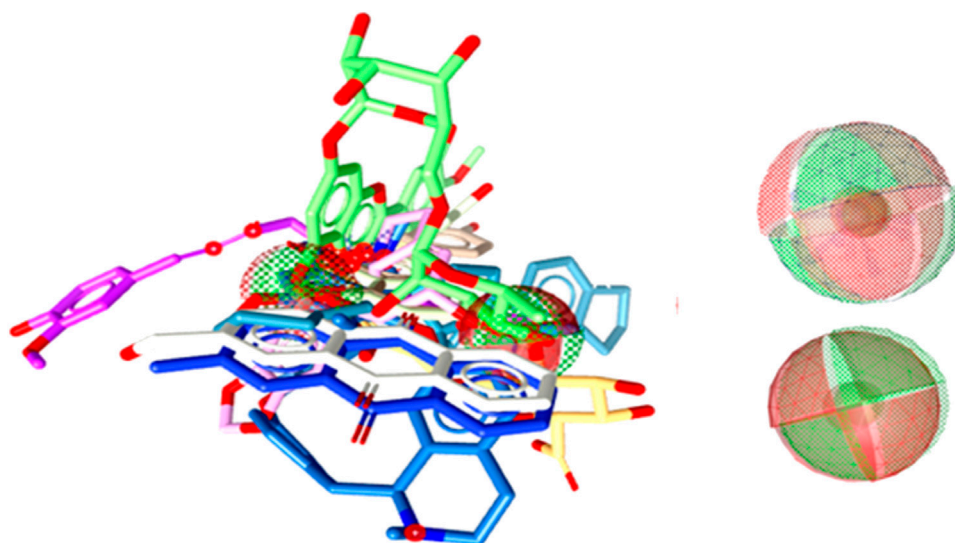


FIGURE 3
3D representation of pharmacophore modelling.

TABLE 4 Pharmacophore of top 10 lead hits from virtual screening analyses and Calculated binding affinities of 10 lead hits docked compounds against P53.

Sr #	Top leads hit from virtual screening	Highest score of top 10 leads hits	Binding affinities (kcal/mol)
1	Compound-1	53.92	−11.10
2	Compound-2	51.05	−10.7
3	Compound-3	51.09	−10.3
4	Compound-4	51.07	−8.4
5	Compound-5	51.78	−6.9
6	Compound-6	50.75	−7.7
7	Compound-7	51.06	−7.6
8	Compound-8	51.07	−7.10
9	Compound-9	51.80	−7.6
10	Compound-10	50.75	−7.7

and are thus expected to have a higher likelihood of binding to the target protein with high affinity. It is important to note that a high pharmacophore score does not guarantee that a compound will be an effective drug. Further studies, such as molecular docking and molecular dynamics simulations, are necessary to assess the binding and stability of the compounds to the target protein and evaluate their potential as drug candidates.

3.8 Molecular docking analyses

The top three compounds identified by a simulated screening of possible antitumor compounds underwent further investigation based on their drug-likeness, binding preferences, and energies. The top ten compounds' binding values, ranging

from −11.10 to −6.9 kcal/mol, assessed each substance's ability to attach to the TP53 protein. The best three matches had the lowest binding strengths and most favourable binding energies of all the produced anchored complexes. As seen in [Figure 4](#), these substances were discovered to bind to the same active areas of the TP53 protein. This indicates that these substances might block the protein's action and might even be turned into effective antitumor medications.

Based on the docking findings, the top three natural compounds (Compound-1, Compound-2, and Compound-3) showed the highest binding energies (−11.10 to −10.3 kcal/mol) in contrast to the other compounds looked at ([Table 4](#)). Additionally, the top three compounds interacted with critical areas, suggesting a potential for particular binding to the target protein, P53. We also investigated these compounds' binding energies and concentrations and their drug-like properties. The compounds with the lowest binding

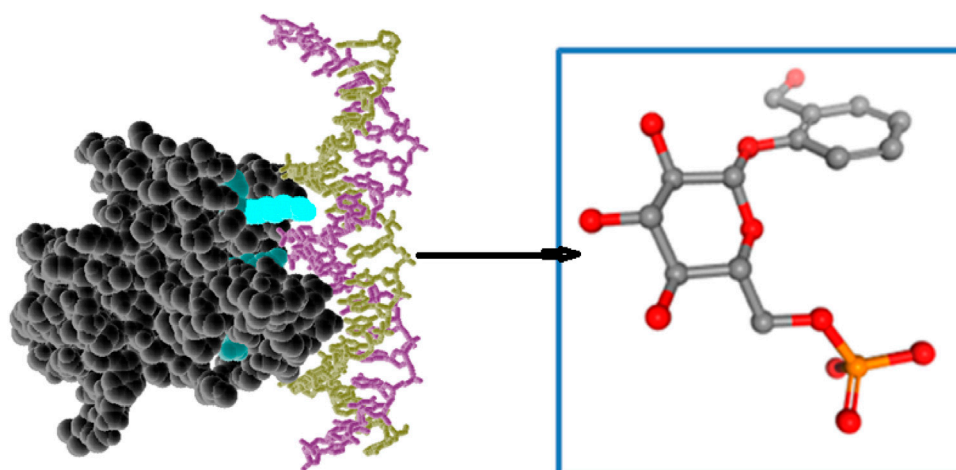


FIGURE 4

All 3 ligands attached at same binding pockets.

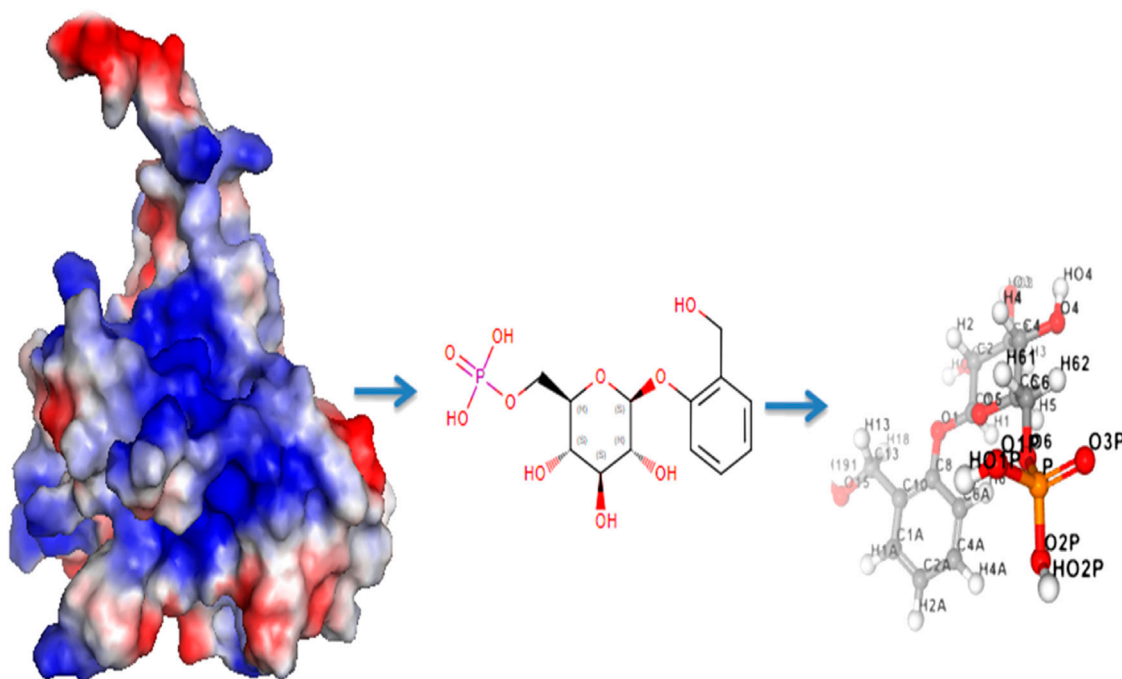


FIGURE 5

3D docked possess of leading hits.

strengths were found to be the most effective at clinging to the active portions of the P53 protein. The selected compounds displayed potential drug-like properties. The P53 protein sequence was also analysed, and the best protein models with the most significant amounts of identity and query coverage were selected for homology-based modelling. Simulation studies of the P53 protein were conducted to identify the top 50 lead matches, and their

pharmacophore characteristics were analyzed. We explore that the top 10 lead hits have a pharmacophore fit with particular pockets and loops on the protein's surface. The findings suggest the possibility of developing these top three natural compounds into influential P53 protein-targeting antitumor drugs.

The binding affinity parameter measures how strongly a ligand—a particle or other small substance—binds to a target protein. The bound

TABLE 5 The binding affinities (kcal/mol) of top-ranked 3 compounds against P53.

Compounds	Binding affinities (kcal/mol)
Compound-1	−10.9
Compound-2	−9.6
Compound-3	−9.2

sensitivities of ten lead hits against the P53 protein were discovered in this instance. The binding strengths were calculated by considering the drug's and protein's intermolecular interactions, including hydrogen bonds, hydrophobic, and electrostatic interactions as shown in [Figure 5](#). The binding energies are expressed in kcal/mol units, with lower numbers indicating stronger binding. With a value of −11.10 kcal/mol, Compound 1 had the highest binding affinity in this instance, suggesting that out of the 10 compounds examined, it attaches to the P53 protein the most firmly. Compound-3 had a slightly lower binding affinity than compound-2, which also had a value of −10.7 kcal/mol, with a value of −10.3 kcal/mol. Compound 4 had the lowest binding affinity of the 10 compounds, with a weight of −8.4 kcal/mol; subsequent compounds had progressively lower binding affinities. More than binding, strengths can affect a drug's or ligand's effectiveness; when developing new medicines, thought must also be given to variables like metabolism and toxicity.

The top three natural compounds (Compounds 1, 2, and 3) from the docking analysis of the top 10 lead hits against the P53 protein were discovered to be the most active based on their binding affinities and common chemistry interactions ([Table 5](#)). These compounds had the highest binding energies (−10.9 to −9.2 kcal/mol) and bonded to significant sections in the target protein's active domains. Fewer correlations and weaker binding energies were observed for the other compounds under study. Overall, the results indicate that these top 3 natural ingredients could be used to develop novel cancer therapy drugs. The potent binding sites and common chemical interactions that the bound compounds disclose can be used to create effective inhibitors against the P53 protein.

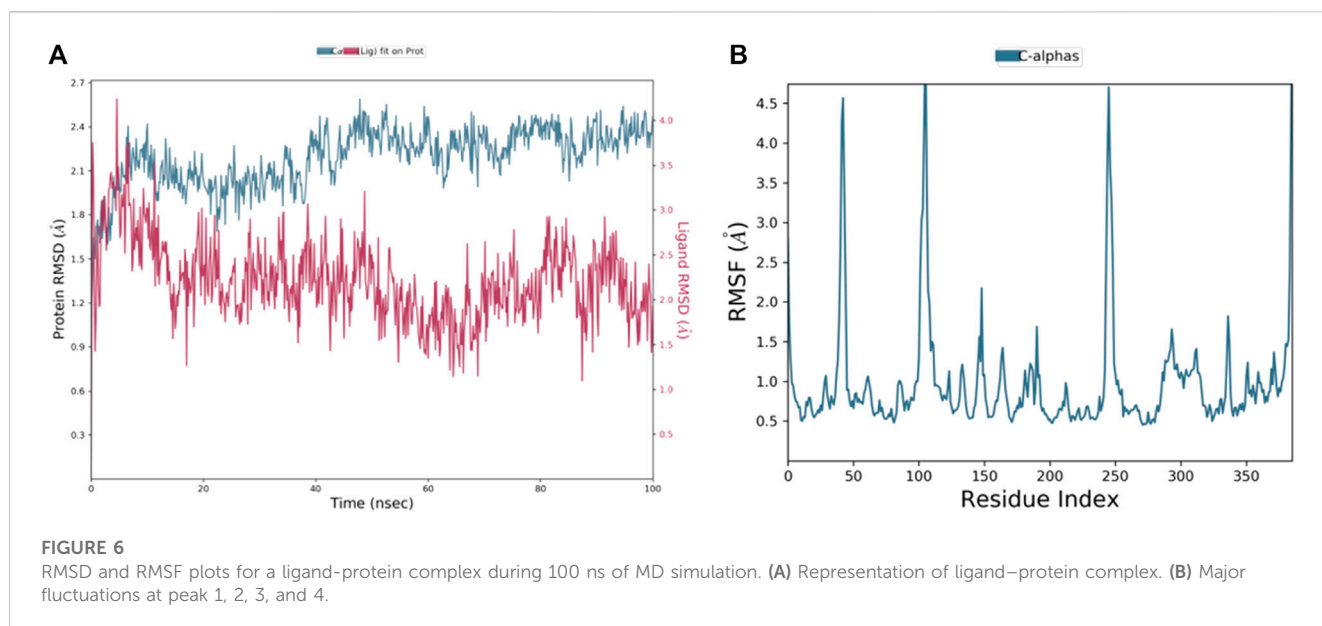
Using Lipinski's rule of five, which evaluates the drug-likeness of small compounds, the morphological attributes of the best three hits were examined. The molecules had a molecular weight within the

permissible range of 500 g/mol, 10 hydrogen acceptors, 5 hydrogen donors, and 5 LogP values. The LogP number determines the absorption and membrane permeability of the substances. The best three hits' LogP values varied from 2.06 to 9.02 and were within the permissible absorption range. Additionally, the compounds have a sufficient number of flexible bonds, hydrogen bond acceptors, and donors, which suggests that they have excellent drug action. The best three results' physical characteristics are shown in [Table 6](#), and it was found that these substances met most of the expected characteristics and lacked cancer potential. According to a study of the ADMET characteristics, the gut can readily take the chemicals. The best three results are good prospects for antitumor medicines based on their physical characteristics and drug-likeness traits.

The details mentioned in [Table 6](#) provide the physicochemical properties of the top 3 selected compounds, where Compound-1, Compound-2 and Compound-3 molecular weight (g/mol) represents the sum of the atomic weights of all atoms in a molecule, which are composed of 420.3877, 403.4048, and 526.7162 g/mol, respectively. LogP (o/w) is a measure of the solubility of a compound in water (o) compared to octanol (w). The three Compound's LogP values are 2.0649, 3.8989, and 9.0248, respectively. H-bond acceptors and H-bond donors are related to the hydrogen bonding capacity of a compound with other molecules. Compound-1, Compound-2, and Compound-3 have 6, 6, and 2 H-bond acceptors, and 3, 3, and 2 H-bond donors, respectively. Rotatable bonds indicate the number of bonds that can rotate around their axis within a molecule. Compound-1, Compound-2, and Compound-3 have 5, 5, and 1 rotatable bond, respectively. Polar Surface Area (PSA) measures the size of the polar or charged atoms and groups of atoms in a molecule. Compound-1, Compound-2, and Compound-3 have 74.9200, 92.0500, and 30.2300 PSA values, respectively. The number of atoms and rings in a molecule can influence the properties and interactions of a compound. Compound-1, Compound-2, and Compound-3 have 54, 50, and 91 atoms, and 5, 4, and 6 rings, respectively. The values of these physicochemical properties indicate that the top 3 compounds are drug-like and satisfy Lipinski's rule of five, which suggests they may have a good chance of becoming a viable drug candidate. The ADMET analysis indicates that

TABLE 6 The analyses of physicochemical Properties of the top 3 selected compounds.

Properties	Compound-1	Compound-2	Compound-3
Molecular weight (g/mol)	420.3877	403.4048	526.7162
Logp (o/w)	2.0649	3.8989	9.0248
H-bond acceptors	6	6	2
H-bonds donors	3	3	2
Rotatable bonds	5	5	1
PSA	74.9200	92.0500	30.2300
Atoms	54	50	91
Rings	5	4	6



these compounds are non-carcinogenic and can be easily absorbed by the intestine.

3.9 MD simulation

Molecular dynamics (MD) simulations are an effective tool for investigating the dynamic behavior of proteins at biologically relevant timescales (Rather et al., 2020). In this study, MD simulations were used to examine how the presence of ligands affects the stability and conformational changes of the TP53 protein. The RMSD plot is a widely used method for evaluating the structural stability of proteins during MD simulations, which measures the deviation of the protein backbone conformation from the starting structure as a function of time. According to the results of this study, the RMSD plot for the TP53 protein displayed an initial increase in the C- α backbone from 1 ns to 25 ns, indicating significant structural rearrangement. This was followed by a gradual decrease in RMSD from 26th ns to 30th ns, indicating that the protein was becoming more stable. Peaks in the RMSD values from 35 ns to 40 ns suggested that the protein was undergoing further conformational changes. The gradual decrease in RMSD from 65 ns to 100 ns suggested that the protein had finally achieved a stable conformation as shown in Figure 6.

As shown in Figure 6A, the ligand's RMSD variations were congruent with those of the protein C-backbone, proving that the ligand was stably attached to the protein throughout the experiment. The bound complex's average RMSD over the course of the 100 ns exercise was 3.2 Å, showing that the protein-ligand complex was largely steady. Because the bound complex's average RMSF was 4 Å, some structural changes may have occurred in the protein during the experiment. Another popular technique for examining the kinetics of proteins in MD models is the RMSF image. Each protein residue's departure from its typical location is tracked during the exercise. In this research, the TP53 protein's

RMSF histogram for positions 1–95 revealed a rise, with a highest value of 5 Å. This indicates that these residues were more malleable during the exercise and experienced more structural changes. Other positions showed modest variations with an average RMSF of 3 Å. Residues with a value of 4.8 showed an increase in peaks 3 and 4 Å, suggesting that these residues were experiencing substantial structural changes. Altogether, the outcomes of the MD models indicate that the chosen substances can attach to TP53 and maintain its shape, thereby preventing the development of cancer cells. The protein-ligand complex's stability was examined using a number of metrics, such as the RMSD, RMSF, and ligand characteristics, and it was found to be a stable, closely bonded structure with a high degree of density. These discoveries shed light on the protein-ligand interaction process and may help create new cancer treatments.

The departure of the shape of the ligand-protein complex from the initial structure over time is depicted by the RMSD (Root Mean Square Deviation) diagram for a ligand-protein complex during a 100 ns Molecular Dynamics (MD) simulation. The RMSD, which is used to track the stability and progress of the simulation, is a measurement of the average distance between the atoms of the simulation structure and the reference structure (typically the beginning structure). The RMSD figure frequently exhibits an early rise, a peak, and a steady rise over time. While the peak shows the stability of the complex throughout the exercise, the initial rise results from the system relaxing from its original structure. The steady rise over time results from the system's tiny variations adding up, which can have a significant systemic impact.

The main peaks and valleys of the ligand-protein combination during the experiment are depicted in the RMSF (Root Mean Square Fluctuation) Figure 6B. The complex's most malleable areas or those undergoing the greatest structural changes during the exercise are indicated by the peaks in the RMSF image. The average departure of each residue's location from its normal position throughout the experiment is represented by the RMSF, which is computed for each residue in the protein or ligand. The complex's area of high flexibility

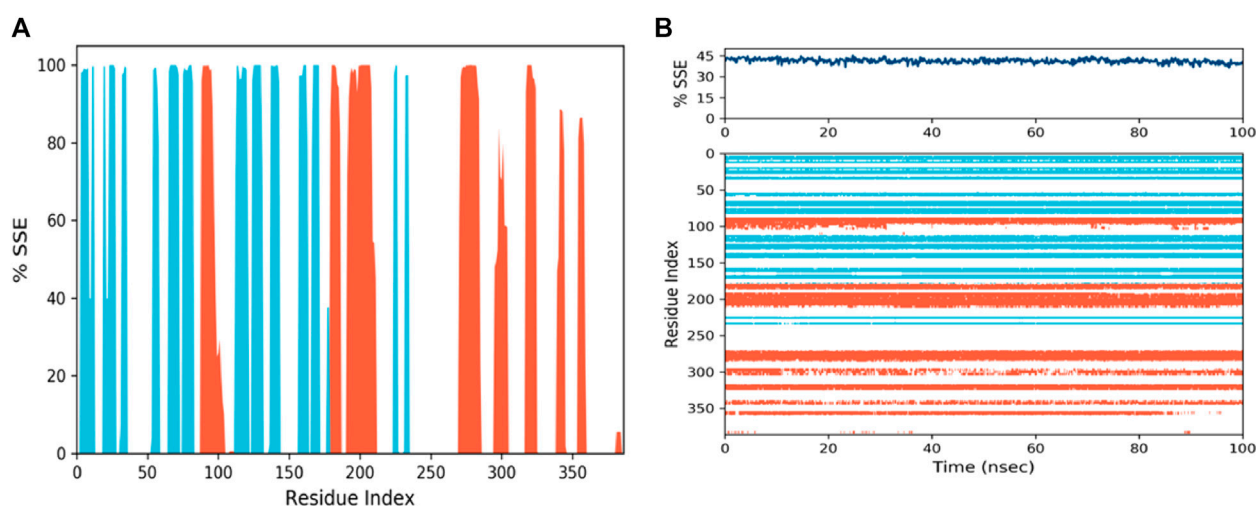


FIGURE 7

Visualization of the secondary structures of alpha helices, beta-strands, and other secondary structures in protein structure. (A) SSE distribution by residue index: red peaks = helices, blue peaks = beta-strands. (B) SSE composition for each trajectory frame and assignment for each residue over time.

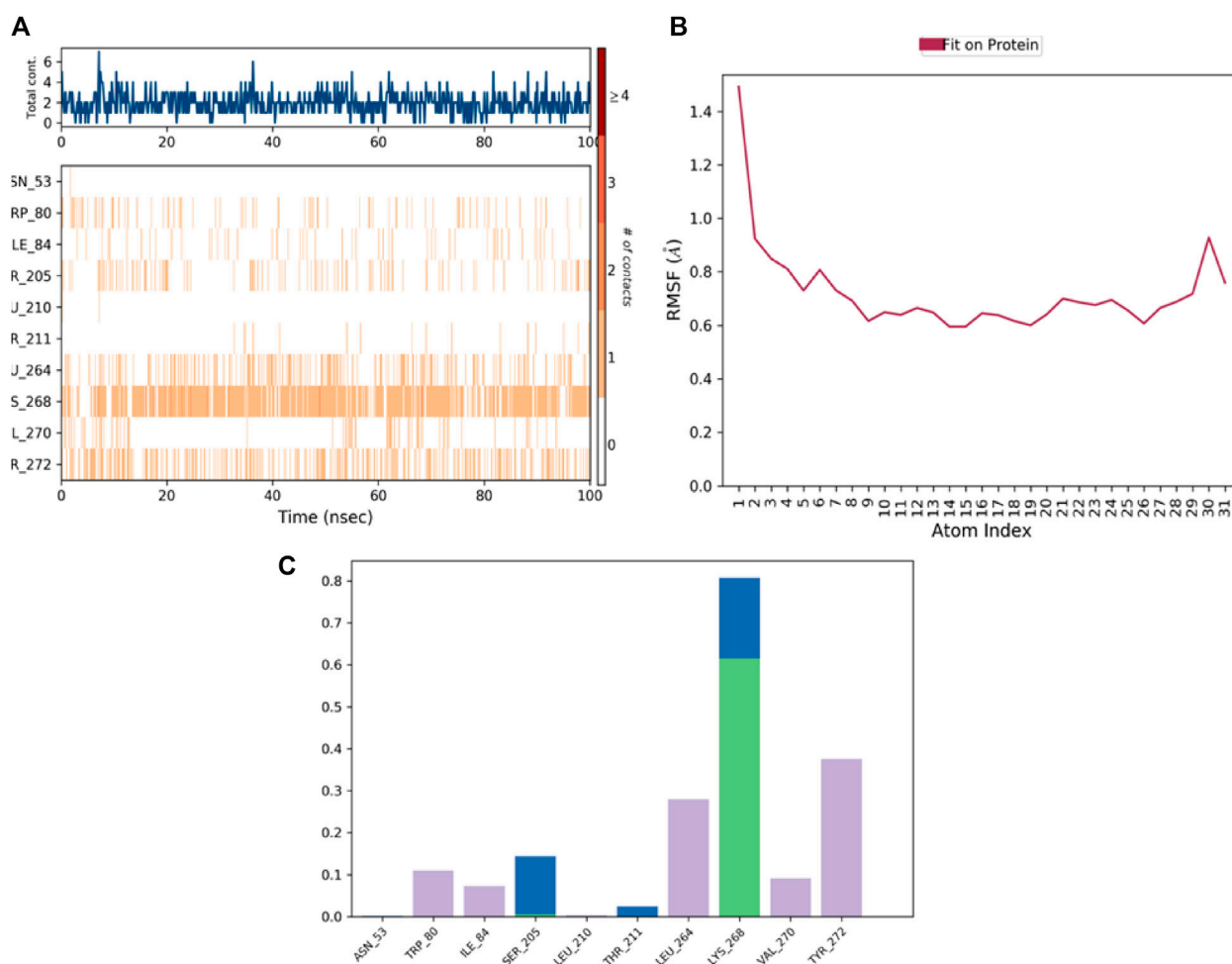


FIGURE 8

A schematic of ligand-ligand interactions. (A) Protein-ligand contacts categorized by interaction. (B) Timeline of H-bond and hydrophobic contacts between protein and ligand. (C) Schematic diagram of ligand atom interactions with protein residues.

or movement is indicated by peak 1 Å in the RMSF image, while the region of moderate flexibility is demonstrated by peak 2 Å, Peaks 3 Å, and 4 Å, could be localised structurally changing areas that interact with fluid molecules or other regions of the complex. The analysis of these peaks can reveal details about the dynamics of the complex and point out areas crucial to its stability or functionality.

The root mean square distance of all parts from the protein's centre of mass is known as the radius of gyration (Rg), which gauges how dense a protein's structure is. It shows how crowded and how much room the protein's auxiliary components fill. Proteins with bigger Rgs have structures that are longer or more open, whereas proteins with smaller Rgs have structures that are closer together. A closely bonded complex in the context of protein-ligand binding denotes intimate interaction between the protein and the ligand, resulting in a lower total Rg of the complex than the isolated protein or ligand. This could mean a high propensity for the protein and receptor to bind. The stability and kinetics of protein-ligand interactions are frequently assessed using MD models and the RMSD (Root Mean Square Deviation) and RMSF (Root Mean Square Fluctuation) studies. The RMSF measures the average variation in each molecule or peptide throughout the experiment, while the RMSD determines how much the protein-ligand combination deviates from its original structure. The protein-ligand complex is physically stable, and the protein and ligand are closely attached if the RMSD/RMSF of the protein-ligand complex stays constant throughout the exercise. On the other hand, if the complex is experiencing structural alterations or the protein-ligand association is feeble, the RMSD/RMSF ratios are significant.

We analyze the phytochemicals based on the absence of significant structural changes in the RMSD/RMSF plot during the simulation. This indicates that the compounds are structurally stable and not undergoing any major conformational changes, per earlier researchers (Hafiz Muhammad et al., 2022). This is a significant result as it suggests that the phytochemicals aren't degrading or becoming unstable during the simulation. Therefore, the observed interactions with the protein likely represent their actual behaviour *in vivo*. The study's results suggest that the protein-ligand complex is tightly bound, structurally stable, and compact, indicating a strong binding affinity between the protein and ligand. These findings are consistent with previous research and provide insights into the structure-function relationships of the complex.

3.10 Protein structure prediction analyses

The study of secondary structure elements (SSE) offers details about the protein's molecular alterations during the MD simulation. As you pointed out, the findings in Figure 7 shows that the protein's helices and strands experienced a major shape shift, which raises the possibility that the protein endured a substantial structural reorganisation. The protein was highly dynamic and many areas experienced significant structural changes, as evidenced by the fact that 43.78% of the protein's SSE were impacted during the 100 ns simulation. The study of several ligand about radius of gyration, intramolecular hydrogen

bonds, molecular surface area, solvent-accessible surface area, and polar surface area, showed ligand is firmly attached to the protein despite these structural alterations (PSA). The fact that these characteristics barely changed suggests that the ligand was firmly attached to the protein and that its interactions with the protein were very steady. According to these results, the protein-ligand complex is highly stable and closely bonded even though the protein underwent substantial structural changes during the MD simulation. This aligns with earlier research's findings and emphasises the importance of comprehending the kinetics and stability of protein-ligand interactions to create more successful drug design methods. Overall, data offers insightful information about the stability and structural changes of a protein-ligand complex during an MD simulation, and these discoveries can be used to guide future studies into the creation of new medicines.

The secondary structure components of the protein are represented visually in the protein structure by the SSE distribution by residue index (Figure 7A). In the diagram, alpha helices are represented by the crimson peaks and beta strands by the blue peaks. Plotting the spread of these components along the protein chain enables a rapid evaluation of the protein's general secondary structure. The SSE makeup for each motion frame throughout the experiment is shown in the figure below (Figure 7B). This diagram displays the spread of beta strands, alpha helices, and other secondary structures at various modelling time points. It makes it possible to see how the SSE makeup varies over time and can be helpful in pinpointing instances of instability or structural change. The image at Figure 7 bottom tracks each residue's SSE designation over time. This image thoroughly examines how each residue's SSE evolves throughout the exercise. The image can help find the stability of particular secondary structures or locate specific acids that may be engaged in structural shifts. Overall, the SSE representation offers essential knowledge about the secondary structure components of the protein and how they evolve throughout the MD simulation. Researchers can learn more about the protein's stability and structural changes that may be crucial for ligand binding and function by examining changes in the SSE makeup.

The information regarding the ligand components involved in interactions and their interaction with protein residues, specifically amino acids, is depicted in (Figure 8A). This information, which illuminates the critical relationships between the ligand and protein, can help develop novel ligands or improve already existing ones. The ligand features graphic shows the ligand's radius of gyration (rGyr), molecular surface area (MolSA), solvent accessible surface area (SASA), and polar surface area (PSA) as shown in (Figure 8B). These traits can provide details regarding the ligand's stability and protein interactions. A safe ligand, for instance, should have a minimum radius of gyration to form intramolecular hydrogen bonds. The ligand's molecular surface area can be used to estimate its size. Its polar surface area and liquid accessible surface area can be used to determine its solubility and polarity. Figure 8C shows the interactions in protein-ligand junctions and categorises them by category. They illustrate the ligand's interactions with specific protein areas containing amino

acids. Among the different types of interactions that can occur are electrostatic interactions, hydrophobic contacts, and hydrogen bonds. The hue labelling in the picture represents the number of distinct connections of a specific peptide with the ligand. Darker orange indicates that a peptide has been exposed to the receptor more than once. The links between the ligand and the protein can be improved by identifying key amino acid sequences implicated in ligand binding.

4 Conclusion

The study investigated the presence of phytochemicals and antioxidant properties in the *Amomum subulatum* seeds. We confirmed the presence of critical phytochemical groups (alkaloids, tannin, saponin, phlobatinin, and cardiac glycoside) responsible for the plant's medicinal properties. The *A. subulatum* plant's ability to scavenge reactive oxygen species (ROS) makes it a potential candidate for therapeutic use, particularly in treating cancer. In this study, our results revealed that the extract comprised $9.4\% \pm 0.04\%$ alkaloid and $1.9\% \pm 0.05\%$ saponin. We also performed DPPH analysis and found that the methanol extract (79.82%), BHT (81.73%), and n-hexane extract (51.31%) exhibited considerable antioxidant activity.

Furthermore, we assessed the extract's ability to inhibit oxidation and observed that Methanol (83.42%) and BHT (90.25%) had the most significant inhibitory effects. We found that Compound-1 had the best pharmacophore match value (53.92), with others ranging from 50.75 to 53.92, where docking results showed that the top three natural compounds had the highest binding energies (-11.10 to -10.3 kcal/mol) and bound to significant regions in the target protein's active domains. The Molecular Dynamics (MD) simulations indicated that the ligand was tied to the protein with substantial conformational changes in the protein structure. The study also identified three compounds that exhibit promising effects against cancer cells, with marked effects on the TP53 protein and the P53 pathway.

The computational models provide insight into protein-ligand interactions and binding affinities. They may serve as a model for developing innovative, less toxic, and highly effective drugs for

cancer treatment. The study's findings highlight the potential of natural compounds for the development of novel cancer therapies.

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

SA, AN, AQ, IZ, and QA: writing—original draft, investigation, and formal analysis; RS, SR: editing; reviewing and investigation H-AN, YB and MB; RS and IZ: editing and supervision. All authors approved the final version of the manuscript.

Acknowledgments

The authors would like to extend their sincere appreciation to the researchers supporting project, King Saud University, Riyadh, Saudi Arabia for funding this work through the project number (RSP 2023R457).

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

- Abbas, G., Rauf, K., and Mahmood, W. (2015). Saponins: The phytochemical with an emerging potential for curing clinical depression. *Nat. Prod. Res.* 29 (4), 302–307. doi:10.1080/14786419.2014.942661
- Ahmad, H. M., Abrar, M., Izhar, O., Zafar, I., Rather, M. A., Alanazi, A. M., et al. (2022). Characterization of fenugreek and its natural compounds targeting AKT-1 protein in cancer: Pharmacophore, virtual screening, and MD simulation techniques. *J. King Saud University-Science* 34 (6), 102186. doi:10.1016/j.jksus.2022.102186
- Ali, M., Khan, T., Fatima, K., Ali, Q. u. A., Ovais, M., Khalil, A. T., et al. (2018). Selected hepatoprotective herbal medicines: Evidence from ethnomedicinal applications, animal models, and possible mechanism of actions. *Phytotherapy Res.* 32, 199–215. doi:10.1002/ptr.5957
- Amarowicz, R., Pegg, R., Rahimi-Moghaddam, P., Barl, B., and Weil, J. (2004). Free-radical scavenging capacity and antioxidant activity of selected plant species from the Canadian prairies. *Food Chem.* 84 (4), 551–562. doi:10.1016/s0308-8146(03)00278-4
- Balwan, W. K., and Kour, S. (2021). Lifestyle diseases: The link between modern lifestyle and threat to public health. *Saudi J. Med. Pharm. Sci.* 7 (4), 179–184. doi:10.36348/sjms.2021.v07i04.003
- Cabero Pérez, M. J., and Patil, M. (2020). Qualitative tests for preliminary phytochemical screening: An overview. *Int. J. Chem. Stud.* 8 (2), 603–608. doi:10.22271/chemi.2020.v8.i2i.8834
- Chen, L., Liu, S., and Tao, Y. (2020). Regulating tumor suppressor genes: Post-translational modifications. *Signal Transduct. Target. Ther.* 5 (1), 90. doi:10.1038/s41392-020-0196-9
- Chen, X., Li, H., Tian, L., Li, Q., Luo, J., and Zhang, Y. (2020). Analysis of the physicochemical properties of acaricides based on Lipinski's rule of five. *J. Comput. Biol.* 27 (9), 1397–1406. doi:10.1089/cmb.2019.0323
- de Carvalho, A. P. A., and Conte-Junior, C. A. (2021). Health benefits of phytochemicals from Brazilian native foods and plants: Antioxidant, antimicrobial, anti-cancer, and risk factors of metabolic/endocrine disorders control. *Trends Food Sci. Technol.* 111, 534–548. doi:10.1016/j.tifs.2021.03.006

- Ekalu, A., and Habila, J. D. (2020). Flavonoids: Isolation, characterization, and health benefits. *Beni-Suef Univ. J. Basic Appl. Sci.* 9 (1), 45–14. doi:10.1186/s43088-020-00065-9
- El Aziz, M., Ashour, A., and Melad, A. G. (2019). A review on saponins from medicinal plants: Chemistry, isolation, and determination. *J. Nanomed. Res.* 8 (1), 282–288. doi:10.15406/jnmr.2019.08.00199
- Fan, M., Yuan, S., Li, L., Zheng, J., Zhao, D., Wang, C., et al. (2023). Application of terpenoid compounds in food and pharmaceutical products. *Fermentation* 9 (2), 119. doi:10.3390/fermentation9020119
- Felth, J., Rickardson, L., Rosén, J., Wickström, M., Fryknaäs, M., Lindskog, M., et al. (2009). Cytotoxic effects of cardiac glycosides in colon cancer cells, alone and in combination with standard chemotherapeutic drugs. *J. Nat. Prod.* 72 (11), 1969–1974. doi:10.1021/np900210m
- Fokunang, E. T., and Fokunang, C. N. (2022). Overview of the advancement in the drug discovery and contribution in the drug development process. *J. Adv. Med. Pharm. Sci.* 24 (10), 10–32. doi:10.9734/jamps/2022/v24i10580
- Hafiz Muhammad, A., Muhammad, A., Osheen, I., Imran, Z., Mohd Ashraf, R., Amer, M. A., et al. (2022). Characterization of fenugreek and its natural compounds targeting AKT-1 protein in cancer: Pharmacophore, virtual screening, and MD simulation techniques. *J. King Saud Univ. - Sci.* 34, 102186. doi:10.1016/j.jksus.2022.102186
- Heider, J. r., Kilian, J., Garifulina, A., Hering, S., Langer, T., and Seidel, T. (2022). Apo2ph4: A versatile workflow for the generation of receptor-based pharmacophore models for virtual screening. *J. Chem. Inf. Model.* 63 (1), 101–110. doi:10.1021/acs.jcim.2c00814
- Hossain, M. S., Karuniawati, H., Jairoun, A. A., Urbi, Z., Ooi, D. J., John, A., et al. (2022). Colorectal cancer: A review of carcinogenesis, global epidemiology, current challenges, risk factors, preventive and treatment strategies. *Cancers* 14 (7), 1732. doi:10.3390/cancers14071732
- Kaminski, M. F., Robertson, D. J., Senore, C., and Rex, D. K. (2020). Optimizing the quality of colorectal cancer screening worldwide. *Gastroenterology* 158 (2), 404–417. doi:10.1053/j.gastro.2019.11.026
- Karak, P. (2019). Biological activities of flavonoids: An overview. *Int. J. Pharm. Sci. Res.* 10 (4), 1567–1574. doi:10.13040/IJPSR.0975-8232
- Karimian, A., Mir, S. M., Parsian, H., Refieyan, S., Mirza-Aghazadeh-Attari, M., Yousefi, B., et al. (2019). Crosstalk between Phosphoinositide 3-kinase/Akt signaling pathway with DNA damage response and oxidative stress in cancer. *J. Cell. Biochem.* 120 (6), 10248–10272. doi:10.1002/jcb.28309
- Klein, W. M., O'Connell, M. E., Bloch, M. H., Czajkowski, S. M., Green, P. A., Han, P. K., et al. (2022). Behavioral research in cancer prevention and control: Emerging challenges and opportunities. *JNCI J. Natl. Cancer Inst.* 114 (2), 179–186. doi:10.1093/jnci/djab139
- Kumar, P. S., Ranganathan, V., Vijayakaran, K., and Elamaran, A. (2022). Efficient protocols of extraction, isolation and structural determination of active compounds from medicinal plants. *Med. Plants-International J. Phytomedicines Relat. Industries* 14 (1), 7–19. doi:10.5958/0975-6892.2022.00002.8
- Leite, L. S. F., Pham, C., Bilatto, S., Azeredo, H. M., Cranston, E. D., Moreira, F. K., et al. (2021). Effect of natural acid and cellulose nanocrystals on antioxidant and antimicrobial properties of gelatin films. *ACS Sustain. Chem. Eng.* 9 (25), 8539–8549. doi:10.1021/acssuschemeng.1c01774
- Lu, R.-M., Hwang, Y.-C., Liu, I.-J., Lee, C.-C., Tsai, H.-Z., Li, H.-J., et al. (2020). Development of therapeutic antibodies for the treatment of diseases. *J. Biomed. Sci.* 27 (1), 1–30. doi:10.1186/s12929-019-0592-z
- Marei, H. E., Althani, A., Afifi, N., Hasan, A., Cecaci, T., Pozzoli, G., et al. (2021). p53 signaling in cancer progression and therapy. *Cancer Cell. Int.* 21 (1), 703. doi:10.1186/s12935-021-02396-8
- Matanjun, P., Mohamed, S., Mustapha, N. M., Muhammad, K., and Ming, C. H. (2008). Antioxidant activities and phenolics content of eight species of seaweeds from north Borneo. *J. Appl. Phycol.* 20, 367–373. doi:10.1007/s10811-007-9264-6
- Matowa, P. R., Gundidza, M., Gwanzura, L., and Nhachi, C. F. (2020). A survey of ethnomedicinal plants used to treat cancer by traditional medicine practitioners in Zimbabwe. *BMC Complementary Med. Ther.* 13 (6), 278–313. doi:10.1186/s12906-020-03046-8
- Mazhar, T., Haq, I., Ditta, A., Mohsan, S. A. H., Rehman, F., Zafar, I., et al. (2023). The role of machine learning and deep learning approaches for the detection of skin cancer. *Healthcare* 11, 415. doi:10.3390/healthcare11030415
- Morales-Lázaro, S. L., González-Ramírez, R., and Rosenbaum, T. (2019). Molecular interplay between the sigma-1 receptor, steroids, and ion channels. *Front. Pharmacol.* 10, 419. doi:10.3389/fphar.2019.00419
- Muzolf-Panek, M., and Stuper-Szablewska, K. (2021). Comprehensive study on the antioxidant capacity and phenolic profiles of black seed and other spices and herbs: Effect of solvent and time of extraction. *J. Food Meas. Charact.* 15 (5), 4561–4574. doi:10.1007/s11694-021-01028-z
- Otohinoyi, D., Kuchi, A., Wu, J., and Hicks, C. (2022). Integrating genomic information with tumor-immune microenvironment in triple-negative breast cancer. *Int. J. Environ. Res. Public Health* 19 (21), 13901. doi:10.3390/ijerph192113901
- Páll, S., Zhmurov, A., Bauer, P., Abraham, M., Lundborg, M., Gray, A., et al. (2020). Heterogeneous parallelization and acceleration of molecular dynamics simulations in GROMACS. *J. Chem. Phys.* 153 (13), 134110. doi:10.1063/5.0018516
- Pandya, G., Kirtonia, A., Singh, A., Goel, A., Mohan, C. D., Rangappa, K. S., et al. (2021). A comprehensive review of the multifaceted role of the microbiota in human pancreatic carcinoma. *Seminars in cancer biology. Semin. Cancer Biol.* 86, 682. doi:10.1016/j.semcancer.2021.05.027
- Pasaribu, G., Winarni, I., Gusti, R. E. P., Maharani, R., Fernandes, A., Harianja, A. H., et al. (2021). Current challenges and prospects of Indonesian non-timber forest products (NTFPs): A review. *Forests* 12 (12), 1743. doi:10.3390/f12121743
- Prunotto, A. (2020). *Characterization of protein-membrane interfaces through a synergistic computational-experimental approach*. Lausanne, Switzerland: Lausanne, EPFL.
- Przyrnska, K., and Pękal, A. (2013). Application of free radical diphenylpicrylhydrazyl (DPPH) to estimate the antioxidant capacity of food samples. *Anal. methods* 5 (17), 4288–4295. doi:10.1039/c3ay40367j
- Rafique, R., Khan, K. M., Chigurupati, S., Wadood, A., Rehman, A. U., Salar, U., et al. (2020). Synthesis, *in vitro* α -amylase inhibitory, and radicals (DPPH & ABTS) scavenging potentials of new N-sulfonylhydrazide substituted indazoles. *Bioorg. Chem.* 94, 103410. doi:10.1016/j.bioorg.2019.103410
- Rahman, M. M., Islam, M. R., Rahman, F., Rahaman, M. S., Khan, M. S., Abrar, S., et al. (2022). Emerging promise of computational techniques in anti-cancer research: At a glance. *Bioengineering* 9 (8), 335. doi:10.3390/bioengineering9080335
- Rather, M. A., Dutta, S., Guttula, P. K., Dhandare, B. C., Yusufzai, S., and Zafar, M. I. (2020). Structural analysis, molecular docking and molecular dynamics simulations of G-protein-coupled receptor (kisspeptin) in fish. *J. Biomol. Struct. Dyn.* 38 (8), 2422–2439. doi:10.1080/07391102.2019.1633407
- Saad, B., Zaid, H., Shanak, S., and Kadan, S. (2017). "Introduction to medicinal plant safety and efficacy," in *Anti-diabetes and anti-obesity medicinal plants and phytochemicals: Safety, efficacy, and action mechanisms* (Springer International Publishing), 21–55. doi:10.1007/978-3-319-54102-0_2
- Salo-Ahen, O. M., Alanko, I., Bhadane, R., Bonvin, A. M., Honorato, R. V., Hossain, S., et al. (2020). Molecular dynamics simulations in drug discovery and pharmaceutical development. *Processes* 9 (1), 71. doi:10.3390/pr9010071
- Samavarchi Tehrani, S., Mahmoodzadeh Hosseini, H., Yousefi, T., Abolghasemi, M., Qujeq, D., Maniati, M., et al. (2019). The crosstalk between trace elements with DNA damage response, repair, and oxidative stress in cancer. *J. Cell. Biochem.* 120, 1080–1105. doi:10.1002/jcb.27617
- Satpathy, R. (2022). "Application of bioinformatics techniques to screen and characterize the plant-based anti-cancer compounds," in *Handbook of research on natural products and their bioactive compounds as cancer therapeutics* (Hershey, Pennsylvania: IGI Global), 466–484.
- Saxena, M., Saxena, J., Nema, R., Singh, D., and Gupta, A. (2013). Phytochemistry of medicinal plants. *J. Pharmacogn. phytochemistry* 1 (6), 168–182.
- Sharma, H., Yadav, K., and Sharma, R. D. (2023). "Role of alternative splicing in health and diseases," in *Transcription and translation in health and disease* (Elsevier), 19–36.
- Tabrez, S., Zughaibi, T. A., Hoque, M., Suhail, M., Khan, M. I., and Khan, A. U. (2022). Targeting glutaminase by natural compounds: Structure-based virtual screening and molecular dynamics simulation approach to suppress cancer progression. *Molecules* 27 (15), 5042. doi:10.3390/molecules27155042
- Tabti, K., Ahmad, I., Zafar, I., Sbati, A., Maghat, H., Bouachrine, M., et al. (2023). Profiling the Structural determinants of pyrrolidine derivative as gelatinases (MMP-2 and MMP-9) inhibitors using *in silico* approaches. *Comput. Biol. Chem.* 104, 107855. doi:10.1016/j.compbiolchem.2023.107855
- Wang, S., Meckling, K. A., Marcone, M. F., Kakuda, Y., and Tsao, R. (2011). Can phytochemical antioxidant rich foods act as anti-cancer agents? *Food Res. Int.* 44 (9), 2545–2554. doi:10.1016/j.foodres.2011.05.021
- Wolber, G., and Langer, T. (2005). LigandScout: 3-D pharmacophores derived from protein-bound ligands and their use as virtual screening filters. *J. Chem. Inf. Model.* 45 (1), 160–169. doi:10.1021/ci049885e
- Wu, B., Vanamala, J. K., Chopra, S., and Reddivari, L. (2021). Near-isogenic lines as powerful tools to evaluate the effect of individual phytochemicals on health and chronic diseases. *Plant Biotechnol. Exp. Future Prospects* 1, 249–256. doi:10.1007/978-3-030-68345-0_17
- Xijun, W., Aihua, Z., Hui, S., Ying, H., and Guangli, Y. (2016). Discovery and development of innovative drug from traditional medicine by integrated chinmedomics strategies in the post-genomic era. *TrAC Trends Anal. Chem.* 76, 86–94. doi:10.1016/j.trac.2015.11.010



OPEN ACCESS

EDITED BY

Khurshid Ahmad,
Yeungnam University, Republic of Korea

REVIEWED BY

Anshul Tiwari,
Vanderbilt University, United States
Shaheer Hasan Khan,
Aligarh Muslim University, India
Mohammad Aslam,
Fujian Agriculture and Forestry University,
China
Brijesh Singh Yadav,
Agricultural Research Organization
(ARO), Israel

*CORRESPONDENCE

Farah Anjum,
✉ farahanjum@tu.edu.sa
Nenggui Xu,
✉ ngxu8018@163.com

RECEIVED 05 April 2023

ACCEPTED 05 May 2023

PUBLISHED 22 May 2023

CITATION

Hua L, Anjum F, Shafie A, Ashour AA,
Almalki AA, Alqarni AA, Banjer HJ,
Almaghrabi SA, He S and Xu N (2023),
Identifying promising GSK3 β inhibitors for
cancer management: a computational
pipeline combining virtual screening and
molecular dynamics simulations.
Front. Chem. 11:1200490.
doi: 10.3389/fchem.2023.1200490

COPYRIGHT

© 2023 Hua, Anjum, Shafie, Ashour,
Almalki, Alqarni, Banjer, Almaghrabi, He
and Xu. 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.

Identifying promising GSK3 β inhibitors for cancer management: a computational pipeline combining virtual screening and molecular dynamics simulations

Libo Hua¹, Farah Anjum^{2*}, Alaa Shafie², Amal Adnan Ashour³,
Abdulraheem Ali Almalki², Ali Abdullah Alqarni³,
Hamsa Jameel Banjer², Sarah Abdullah Almaghrabi^{4,5}, Shan He^{6,7,8}
and Nenggui Xu^{1*}

¹South China Research Center for Acupuncture and Moxibustion, Medical College of Acupuncture Moxibustion and Rehabilitation, Guangzhou University of Chinese Medicine, Guangzhou, China,

²Department of Clinical Laboratory Sciences, College of Applied Medical Sciences, Taif University, Taif, Saudi Arabia, ³Department of Oral and Maxillofacial Surgery and Diagnostic Sciences, Faculty of Dentistry, Taif University, Taif, Saudi Arabia, ⁴Department of Medical Laboratory Technology, Faculty of Applied Medical Sciences, King Abdulaziz University, Jeddah, Saudi Arabia, ⁵Center for Innovations in Personalized Medicine (CIPM), King Abdulaziz University, Jeddah, Saudi Arabia, ⁶School of Food and Pharmacy, Zhejiang Ocean University, Zhoushan, China, ⁷Institute for Nano Scale and Technology, College of Science and Engineering, Flinders University, Bedford Park, SA, Australia, ⁸College of Engineering, Information Technology and Environment, Charles Darwin University, Darwin, NT, Australia

Glycogen synthase kinase-3 (GSK3 β), a serine/threonine protein kinase, has been discovered as a novel target for anticancer drugs. Although GSK3 β is involved in multiple pathways linked to the etiology of various cancers, no specific GSK3 β inhibitor has been authorized for cancer therapy. Most of its inhibitors have toxicity effects therefore, there is a need to develop safe and more potent inhibitors. In this study, a library of 4,222 anti-cancer compounds underwent rigorous computational screening to identify potential candidates for targeting the binding pocket of GSK3 β . The screening process involved various stages, including docking-based virtual screening, physicochemical and ADMET analysis, and molecular dynamics simulations. Ultimately, two hit compounds, BMS-754807 and GSK429286A, were identified as having high binding affinities to GSK3 β . BMS-754807 and GSK429286A exhibited binding affinities of -11.9 , and -9.8 kcal/mol, respectively, which were greater than that of the positive control (-7.6 kcal/mol). Further, molecular dynamics simulations for 100 ns were employed to optimize the interaction between the compounds and GSK3 β , and the simulations demonstrated that the interaction was stable and consistent throughout the study. These hits were also anticipated to have good drug-like properties. Finally, this study suggests that BMS-754807 and GSK429286A may undergo experimental validation to evaluate their potential as cancer treatments in clinical settings.

KEYWORDS

glycogen synthase kinase-3, cancer, virtual screening, molecular dynamics, drug-likeness

1 Introduction

Cancer, a hyperproliferative condition, is characterized by excessive cell division and, eventually, metastasis. Protein kinases are essential regulators of many biological processes and are targets for a wide range of human disorders (Greten and Grivennikov, 2019). Glycogen synthase kinase-3 (GSK3 β), a serine/threonine protein kinase, has been discovered as a novel target for anticancer drugs. It was originally thought to be the most significant enzyme involved in the metabolism of glycogen, but it is now largely recognized as a regulator of various cellular processes, such as the activity of several metabolic and signaling proteins when it phosphorylates (Cohen and Frame, 2001; Dickey et al., 2011). It promotes tumor cell survival in several cancers by using different pro-survival pathways regulated by NF- κ B (Zhang et al., 2014; Saud et al., 2016), Hh/Gli (Trnski et al., 2015), mTOR (Pal et al., 2014), and STAT3 (Gao et al., 2017). GSK3 β inhibitors are now being used to treat a variety of conditions, including Alzheimer's disease, diabetes, and cancer (Klamer et al., 2010; Medina and Avila, 2010; Zeng et al., 2014). Several inhibitors of GSK3 β have been developed and progressed to early-stage clinical trials for various types of cancer (Sahin et al., 2019). One such inhibitor is Tideglusib, which was initially developed to target tau phosphorylation in Alzheimer's disease. However, studies have shown that it can increase proapoptotic proteins in murine models of human neuroblastoma, indicating its potential efficacy in cancer treatment (Mathuram et al., 2016). Another GSK3 β inhibitor, LY2090314, has demonstrated antiproliferative properties in preclinical studies involving melanoma and neuroblastoma. This ATP-competitive inhibitor has shown promising results in clinical trials for cancer treatment (Palomo and Martinez, 2017; Kunnimalaiyaan et al., 2018). Additionally, Solasodine, a naturally occurring aglycone of glycoalkaloid, has been shown to inhibit the GSK-3 pathway and induce apoptosis in various types of malignancies (Zhuang et al., 2017). Although GSK3 β is involved in multiple pathways linked to the etiology of various cancers, no specific GSK3 β inhibitor has been authorized for cancer therapy.

GSK3 β is a 433-residue protein with three different structural domains. The first 134 residues form a 7-strand beta-barrel in the N-terminal domain. Residues 135–151 form a brief linker connecting the N-terminal domain to the alpha-helical domain. Residues 152–342 comprise the alpha-helical domain, and residues 343–433 make up the C-terminal domain. The ATP-binding site is located between the N-terminal and alpha-helical domains (Jacobs et al., 2012).

Drug development is a multidisciplinary, costly, and time-taking process. Computer-assisted drug discovery (CADD) is a constructive approach to drug development that hastens the process and decreases expenses. By reducing the need for animal models in pharmacological research and aiding in the rational design of safe drug candidates, CADD supports medicinal chemists and pharmacologists throughout drug discovery (Paul et al., 2021). The use of CADD has proven crucial to several projects across a range of contexts and research environments. CADD has played a substantial role in the identification and optimization of successful compounds that have moved to further stages of the drug development pipeline or commercialization (Talele et al., 2010). CADD has limitations due to the accuracy of computational models, the scarcity of structural data, the limited chemical diversity, the complexity of

drug targets, and the lack of experimental validation (Sliwoski et al., 2014).

To identify possible candidates for targeting the binding pocket of GSK3 β , we employed computational approaches to screening a diverse library of therapeutically active potential candidates for targeting the binding pocket of GSK3 β .

2 Methodology

2.1 Retrieval and preparation of GSK3 β protein

The GSK3 β protein (PDB: 4AFJ) was retrieved from the PDB database (Gentile et al., 2012). Heteroatoms, water molecules, and co-crystallized ligands were extracted, and the protein was saved in.pdb format. The heteroatoms were proto-oncogene frat, a 30 amino acid short peptide, SO₄, GOL, and SJJ. The clean protein was finally prepared using Discovery Studio 2021 for further studies.

2.2 Compound library preparation and virtual screening

The process of identifying new compounds with specific bioactivity has been transformed by virtual screening methods, which use computer simulations to assess large structure libraries against a biological target (Macalino et al., 2015). This study employed a library of 4,222 anti-cancer compounds (retrieved from <https://www.selleckchem.com>), including both FDA-approved drugs and naturally occurring substances. The collection included compounds that were cell-permeable, therapeutically active, and diverse in terms of their chemical structure. The compound library was downloaded in 'sdf' format and then processed in 'PyRx 0.8' program (Dallakyan and Olson, 2015). The compounds library was minimized utilizing 'UFF' force field as the energy minimization parameter and finally saved in the 'pdbqt' format for further analysis. The grid coordinates of the GSK3 β were set as X = 104.048, Y = 26.822, and Z = -12.474. The best hits were carefully chosen based on the binding affinity and interaction analysis toward GSK3 β .

2.3 Physicochemical and ADMET properties

The DataWarrior tool was used to predict the drug-likeness and physicochemical properties of the top ten compounds, and the ProTox-II (Banerjee et al., 2018) and pkCSM web servers (Pires et al., 2015) were used for ADMET and pharmacokinetic properties.

2.4 MD simulations

The GROMACS 2021.4 software and GROMOS96 43a1 force-field were used to perform MD simulations of three complexes: GSK3 β -control, GSK3 β -BMS-754807, and GSK3 β -GSK429286A, all at 300 K (Pol-Fachin et al., 2009). The topology and force-field

TABLE 1 Binding affinity of top 10 screened compounds.

S. No	Ligand	Binding affinity (kcal/mol)
1	YM201636	−10.6
2	OSI-906	−10.3
3	BMS-754807	−10.2
4	Tipifarnib	−10
5	VX-809	−10
6	INCB28060	−9.8
7	GSK429286A	−9.5
8	Limonin	−9.1
9	Icotinib	−8.7
10	AZ628	−8
11	AR-AO-14418 (Positive control)	−7.5

factors of the compounds were generated using the PRODRG server, and their atoms were combined in the complex topology files. Na⁺ and Cl[−] ions were introduced to neutralize the charges on the GSK3 β protein complexes using the 'gmX_genion' module (0.15 M) (Schuttelkopf and van Aalten, 2004). The 'particle-mesh Ewald' method was employed to analyze the interactions of GSK3 β with these selected compounds. The system (for MD simulation) was minimized employing the 'steepest descent' method (1,500 steps) and equilibrated over a 100-ps period at constant volume in two stages: NVT and NPT ensembles. The total simulation of 100 ns was conducted at 300 K. Trajectories were analyzed using GROMACS modules, and 3D models were created using VMD (Humphrey et al., 1996) and PyMOL (Yuan et al., 2017).

3 Results and discussion

GSK3 β dysfunction has been reported in various cancer types (Domoto et al., 2020), and has been identified as being at the

crossroads of various biochemical pathways, including cancer-related pathways (Duda et al., 2020). Here in this study, a unique collection of 4,222 anti-cancer compounds were screened against the active pocket of GSK3 β . Based on binding affinity, we chose the top ten screened compounds for further visual inspection and interaction studies (Table 1).

The drug-likeness and physicochemical attributes of these top 10 compounds were anticipated using the DataWarrior tool, which employs several parameters including LogP, LogS, H-bond donors and acceptors, relative PSA, and the presence of structures with particular pharmacological properties. Both GSK429286A and BMS-754807 retain an adequate range of drug-likeness properties (Table 2). GSK429286A was discovered to be a selective inhibitor of Rho-associated coiled-coil protein kinase 1 (ROCK1) and ROCK2. These kinases are involved in a several cellular activities, which include cell motility, contraction, and adhesion, and have been associated to cancer, cardiovascular disease, and neurological disorders (Kim et al., 2021). BMS-754807 efficiently and irreversibly inhibits both insulin receptor (IR) family kinases and the insulin-like growth factor 1 receptor (IGF-1R). These kinases have a high binding affinity for it (K_i, 2 nmol/L), which is important for controlling cell growth, survival, and metabolism. BMS-754807 is now in phase I clinical trials for the treatment of multiple types of human cancer (Carboni et al., 2009).

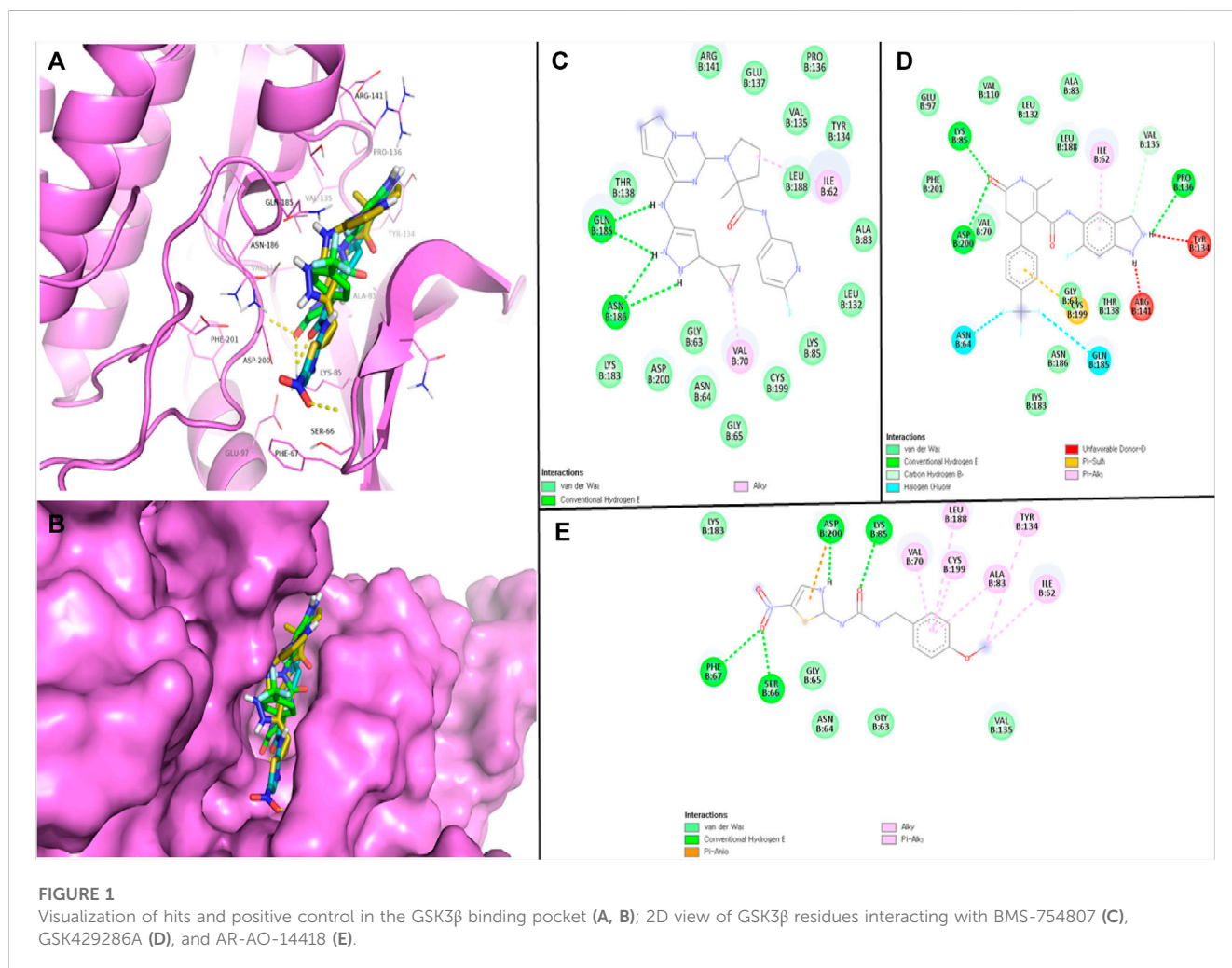
The robust stability of the hit compounds within the GSK3 β active site was attributed to the presence of various Van der Waals and H-bonding interactions (Table 3). These interactions aided in the intercalation of the compounds into the active site, increasing their binding affinity and overall stability (Figures 1A–E). BMS-754807 interacted with several residues like Ile62, Gly63, Asn64, Gly65, Val70, Ala83, Lys85, Leu132, Tyr134, Val135, Pro136, Glu137, Thr138, Arg141, Lys183, Gln185, Asn186, Leu188, Cys199, and Asp200 and residues of GSK3 β . The Van der Waals interaction involves a set of amino acid residues, namely, Gly63, Asn64, Gly65, Ala83, Lys85, Leu132, Tyr134, Val135, Pro136, Glu137, Thr138, Arg141, Lys183, Leu188, Cys199, and Asp200; while Gln185 and Asn186 residues were involved in H-bonding (Figure 1C). Further, GSK429286A interacted with Ile62, Gly63, Asn64, Ala83, Lys85, Glu97, Val110, Leu132, Tyr134, Val135,

TABLE 2 Physicochemical assessment of top 10 screened compounds.

Molecule name	Mol Weight	cLogP	cLogS	H-acceptors	H-donors	Relative PSA	Drug likeness	Mut	Tum	Rep	Irr
Icotinib	391.41	2.9067	−3.673	7	1	0.23718	−12.415	NO	NO	NO	NO
INCB28060	412.41	2.2964	−3.284	7	1	0.24682	4.3136	NO	NO	NO	NO
OSI-906	421.49	4.0356	−6.927	6	2	0.2212	1.7312	NO	NO	NO	NO
GSK429286A	432.37	3.3102	−5.084	6	3	0.24809	−1.5348	NO	NO	NO	NO
AZ628	451.51	4.6632	−6.584	7	2	0.21932	−3.6106	NO	NO	NO	NO
VX-809	452.40	5.632	−6.513	7	2	0.26229	−1.4961	NO	NO	NO	NO
BMS-754807	461.49	2.3302	−5.859	10	3	0.31262	6.8809	NO	NO	NO	NO
YM201636	467.47	2.7535	−7.308	10	2	0.32145	2.2644	NO	NO	NO	NO
Limonin	470.51	1.0279	−4.073	8	0	0.32001	−3.0035	NO	NO	NO	NO
Tipifarnib	489.39	4.03	−5.185	5	1	0.14038	2.4006	NO	NO	NO	NO

TABLE 3 H-bonded and other interactions residues.

Compounds	H-bonded residues	Number of H-bond	Other interactions
BMS-754807	Asn186, and Gln185	2	Ile62, Gly63, Asn64, Gly65, Val70, Ala83, Lys85, Leu132, Tyr134, Val135, Pro136, Glu137, Thr138, Arg141, Lys183, Leu188, Cys199, and Asp200
GSK429286A	Lys85, Pro136, and Asp200	3	Ile62, Gly63, Asn64, Ala83, Glu97, Val110, Leu132, Tyr134, Val135, Thr138, Arg141, Val70, Lys183, Gln185, Asn186, Leu188, Cys199, and Phe201
AR-AO-14418	Ser66, Phe67, Lys85, and Asp200	4	Ile62, Gly63, Asn64, Gly65, Val70, Ala83, Tyr134, Val135, Lys183, Leu188, and Cys199



Pro136, Thr138, Arg141, Val70, Lys183, Gln185, Asn186, Leu188, Cys199, Asp200, and Phe201 residues of GSK3β. Gly63, Val70, Ala83, Lys85, Glu97, Val110, Leu132, Val135, Thr138, Lys183, Asn186, Leu188, and Phe201 residues participates in Van der Waals interactions; while Lys85, Pro136, and Asp200 residues were involved in H-bonding (Figure 1D). Several amino acid residues of the GSK3β have already been described to play an essential role in inhibitor binding. These residues include Ile62, Gly63, Asn64, Val70, Ala83, Lys85, Glu97, Leu132, Tyr134, Val135, Leu188, Cys199, Asp200, and Phe201 (Mishra et al., 2019). It is

noteworthy that the compounds (BMS-754807, and GSK429286A) have been observed to bind with these GSK3β residues.

N-(4-methoxybenzyl)-N0-(5-nitro-1,3-thiazol-2-yl) urea (AR-AO-14418) is a selective inhibitor of GSK3β (Bhat et al., 2003), and was used as a positive control in this study. *In vitro*, AR-AO-14418 inhibited GSK3β without significantly inhibiting other kinases, reducing tau phosphorylation at Ser-396. *In vivo*, it induced antidepressant-like effects in rats by decreasing immobility time and both spontaneous and amphetamine-induced activity (Gould et al., 2004). AR-AO-14418 interacted with Ile62, Gly63,

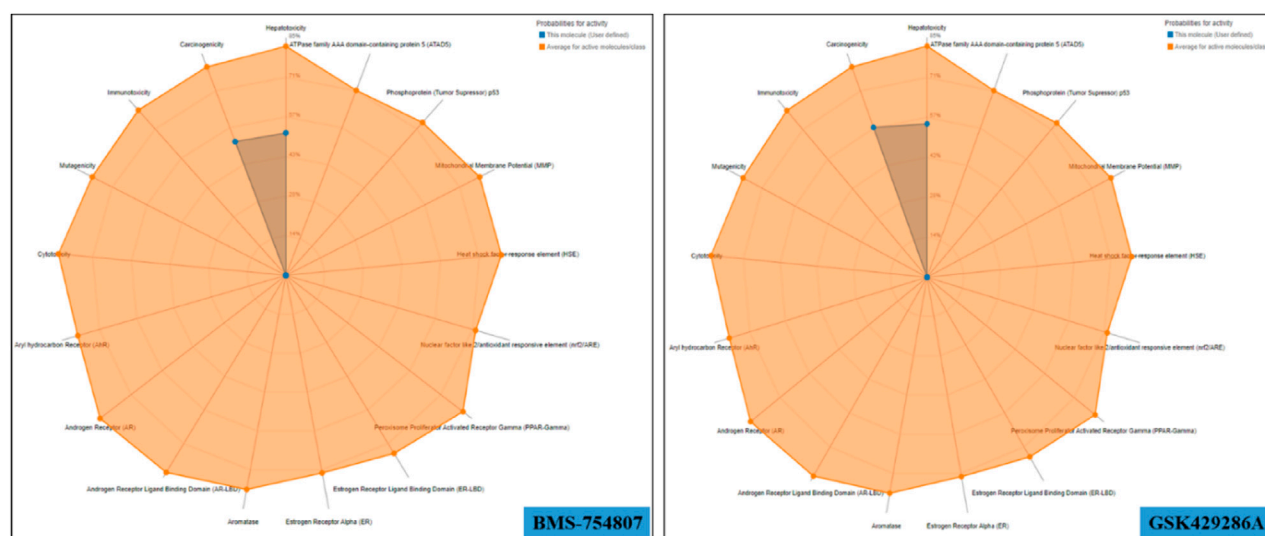


FIGURE 2
Toxicity radar chart of BMS-754807 and GSK429286A.

Asn64, Gly65, Ser66, Phe67, Val70, Ala83, Lys85, Tyr134, Val135, Lys183, Leu188, Cys199, and Asp200 residues of GSK3 β (Figure 1E). Interestingly, Ile62, Gly63, Asn64, Val70, Ala83, Lys85, Tyr134, Val135, Lys183, Leu188, Cys199, and Asp200 GSK3 β residues were observed to be common interacting residues with both the hits (BMS-754807, and GSK429286A) as well as the AR-AO-14418 (Figure 1C-1E). In addition, Lys85, and Asp200 were the common H-bonded residues with GSK429286A and the AR-AO-14418.

BMS-754807 oral toxicity prediction results were as follows: pLD50: 500 mg/kg, pToxicity Class: 4, average similarity: 39.97%, pAccuracy: 23%; while GSK429286A oral toxicity prediction results were as follows: pLD50: 50 mg/kg, pToxicity Class: 2, Average similarity: 45.44%, pAccuracy: 54.26%. BMS-754807 falls into the category of 'harmful if swallowed' while GSK429286A falls into the category of 'fatal if swallowed' as its LD50 value is greater than that of BMS-754807. In addition, multiple toxicity endpoints, including acute toxicity, hepatotoxicity, cytotoxicity, carcinogenicity, mutagenicity, immunotoxicity, etc., were within the acceptable range for these two compounds (Figure 2). Further, multiple pharmacokinetic properties including ADMET of both GSK429286A and BMS-754807 were predicted by the pkCSM web server, which results that both compounds are satisfactorily appropriate in several parameters of ADMET (Table 4).

MD simulation studies were carried out to evaluate complex stability. Protein stability can be measured using the root mean square deviation (RMSD), where a lower RMSD value indicates a more stable protein structure. GSK3 β -control, GSK3 β -BMS-754807, and GSK3 β -GSK429286A had RMSD average values of 0.42, 0.31, and 0.35 nm, respectively. The RMSD plot revealed that GSK3 β -BMS-754807 and GSK3 β -GSK429286A complex showed more binding stability than the control compound. The 'GSK3 β -control' complex showed high deviation from its original conformation, it showed that the

active site pocket of GSK3 β formed quite stable interaction with both of the selected compounds. In addition, the ligand RMSD exhibits GSK3 β -BMS-754807 and GSK3 β -GSK429286A high deviation, and interestingly, the GSK3 β -control complex showed low deviation (Figure 3A-B).

The fluctuation of each residue during the simulation was averaged, and the root mean square fluctuation (RMSF) of GSK3 β was calculated while binding to GSK3 β -control, GSK3 β -BMS-754807, and GSK3 β -GSK429286A. These values were plotted against the residue numbers of GSK3 β . The GSK3 β -control and GSK3 β -GSK429286A backbones presented steady fluctuations in the catalytic pocket of GSK3 β , presumably due to different orientations and the GSK3 β -BMS-754807 complex indicated high fluctuation in region 230–260 residues (Figure 3C). On the other hand, GSK3 β -control, and GSK3 β -GSK429286A complexes showed the least overall fluctuations.

By measuring the distribution of atoms around the axis of a protein, the radius of gyration (Rg) provides insight into the compactness profile of a complex in a biological system. The GSK3 β -control, GSK3 β -BMS-754807 and GSK3 β -GSK429286A complexes had average Rg values of 2.13, 2.15, and 2.17 nm, respectively. Rg plot showed lesser compactness in GSK3 β -control and GSK3 β -BMS-754807 than GSK3 β -GSK429286A complexes. It contingent that after binding, these compounds make GSK3 β stable, due to GSK3 β showing less Rg trajectories (Figure 3D). Among both compounds, BMS-754807 showed better stability in the catalytic pocket of GSK3 β .

The Solvent-accessible surface area (SASA) of a protein refers to the portion of its surface area that interacts with its surrounding solvent molecules. The average SASA values for GSK3 β -control, GSK3 β -BMS-754807, and GSK3 β -GSK429286A complexes were plotted during the 100 ns simulation. The SASA values for the GSK3 β -control, GSK3 β -BMS-754807, and GSK3 β -GSK429286A complexes were 170.51, 178.10, and 180.42 nm²,

TABLE 4 ADMET calculation of GSK429286 A and BMS-754807.

Property	Model name		Predicted value		Unit
			GSK429286A	BMS-754807	
Absorption	Water sol		−4.068	−2.92	log mol/L
	Caco2 per		0.699	1.256	log Papp in 10−6 cm/s
	Intestinal abs		88.306	87.281	% Absorbed
	Skin Per		−2.761	−2.735	log Kp
	P-glycoprotein (P-gp) substrate		Y	Y	
	P-gp I inhibitor		Y	N	
	P-gp II inhibitor		Y	N	
Distribution	VDss (human)		−0.148	0.847	log L/kg
	Fraction unbound		0	0.143	Fu
	Per	−1.129	−1.747	−0.528	log BB
		−2.193	−3.615	−1.665	log PS
Metabolism	substrate	N	N	N	
		Y	N	Y	
	inhibitor	N	N	N	
		Y	N	N	
		Y	N	N	
		N	N	N	
		Y	N	N	
Excretion	Total Clearance		−0.02	−0.393	log mL/min/kg
	Renal OCT2 substrate		N	N	
Toxicity	AMES toxicity		N	N	
	Max. Tolerated dose (human)		−0.205	0.738	log mg/kg/day
	inhibitor	N	N	N	
		Y	Y	N	
	LD50		2.192	2.446	mol/kg
	LOAEL		1.717	1.859	log mg/kg_bw/day
	Hepato		Y	Y	
	Skin Sensitization		N	N	
	T. Pyriformis		0.349	0.285	log mM
	MinNw		1.501	1.63	

(per. = permeability; sol. = solubility; Y=Yes; N=No).

respectively (Figure 4A). Further, GSK3 β -control and GSK3 β -BMS-754807 overlapped each other in 2D projection analysis, whereas GSK3 β -GSK429286A showed a different pattern (Figure 4B). SASA exploration indicated that upon binding of control, BMS-754807, surface exposure has been reduced and the GSK429286A compound increases the surface area of solvent accessibility. Further, hydrogen bond analysis was performed of the docked complexes. To evaluate the stability of the docked complexes, 100 ns simulations of GSK3 β -control, GSK3 β -BMS-

754807, and GSK3 β -GSK429286A were conducted in the presence of a solvent environment. The control and BMS-754807 compound showed an average 2–7 H-bond with GSK3 β protein whereas, the GSK429286A compound showed 2–6 H-bond. It inferred that the BMS-754807 compound showed more stable interaction and might work as a potential drug against the GSK3 β protein (Figure 4C-E).

Next, Mean square displacement (MSD) was determined. It found that the GSK3 β -BMS-754807 complex had a higher

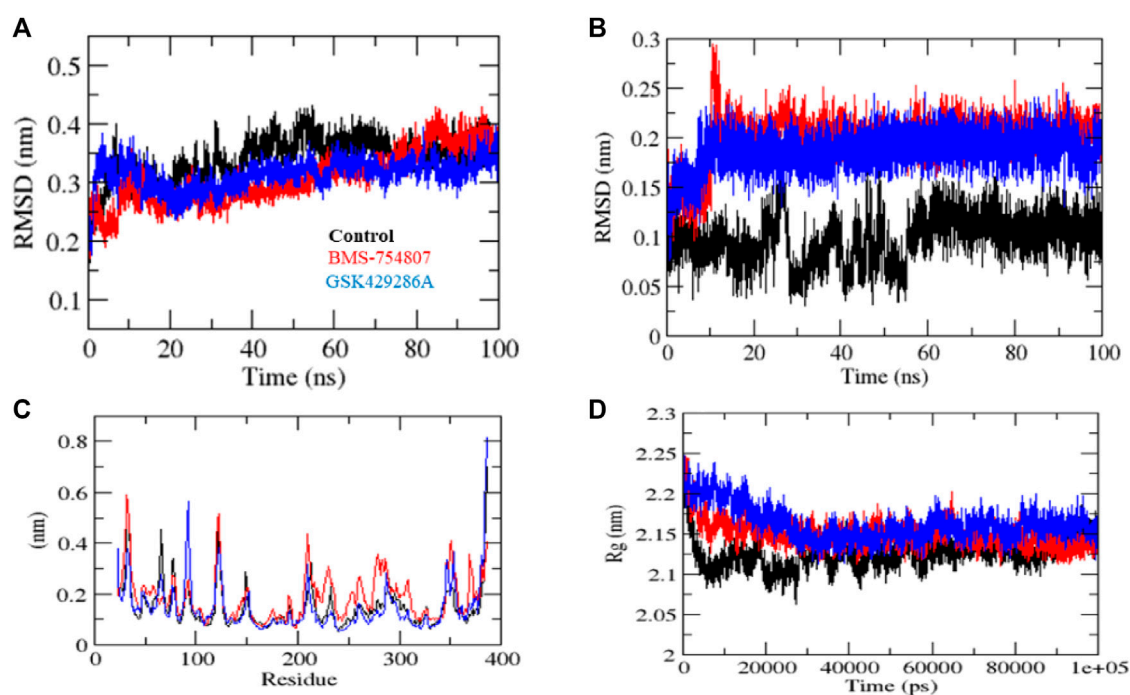


FIGURE 3
MD simulation studies of complexes. RMSD plot (A), RMSD plot of ligands (B), RMSF plot (C), and Rg plot (D) of GSK3 β protein with the ligands.

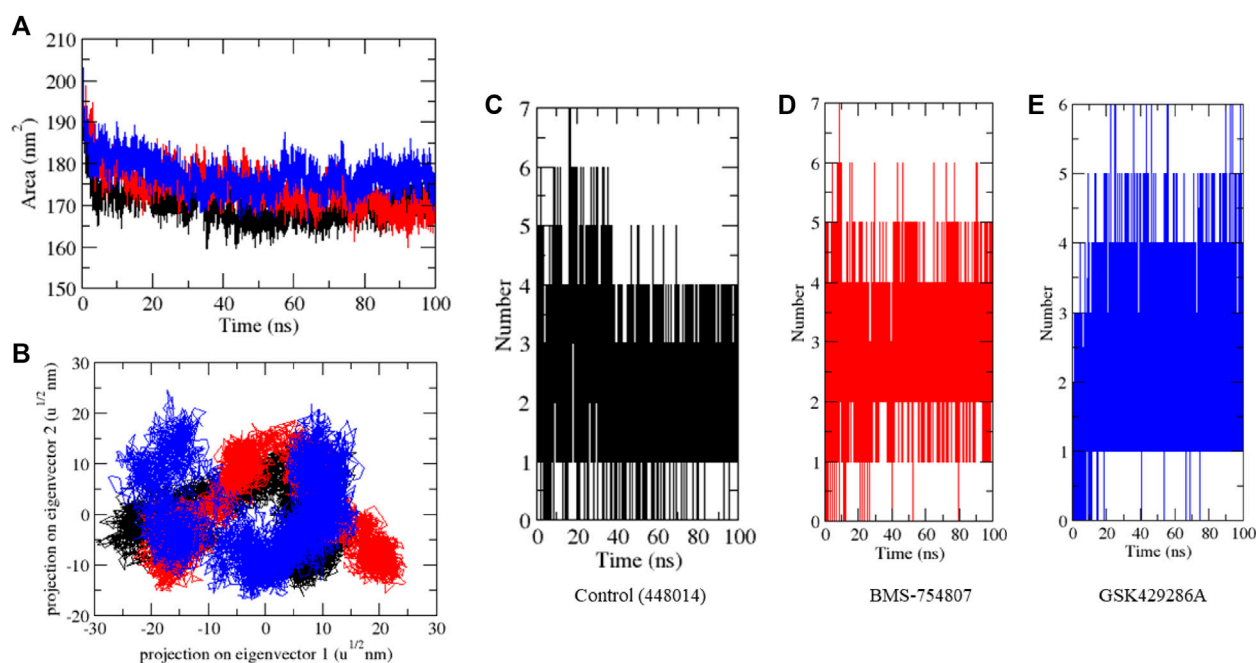


FIGURE 4
SASA plot (A), 2D projection of complexes (B), and number of H-bonds in complexes (C–E).

displacement than the control and the GSK3 β -GSK429286A complex (Figure 5A). The Gibbs' free energy (GFE) landscape was computed using GROMACS analysis modules, and the first

(PC1) and second (PC2) eigenvectors were projected to generate a Comparable GFE contour map, where darker blue shades indicate lower energy levels. During the simulations, the global minima of

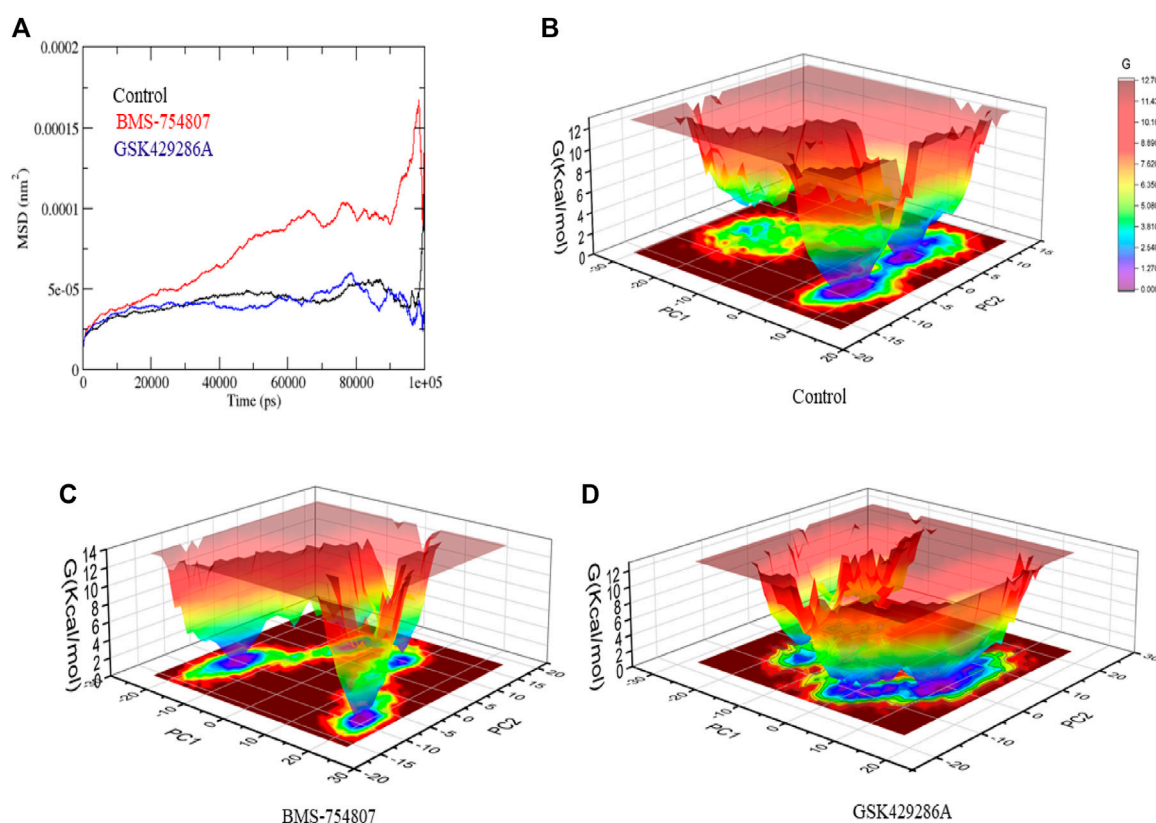


FIGURE 5
Mean square displacement plot of complexes (A), GFE landscape plot (B–D).

GSK3 β fluctuated due to the binding of the complexes to the GSK3 β protein. The GSK3 β -control and GSK3 β -GSK429286A were showing similar projections; and GSK3 β -BMS-754807 was showing dissimilar global minima, demonstrating that BMS-754807 global minima drastically changed during the simulation (Figure 5B–D). The above results suggested that the BMS-754807 compound might be used as a possible drug for the GSK3 β protein.

Several small-molecule inhibitors of GSK3 β , such as CHIR-99021, CHIR-98014, SB216763, SB415286, AR-A011418, CG701338, and CG202796, have been utilized in preclinical studies involving cell and animal models to investigate the potential involvement of GSK3 β in cancer pathogenesis (Walz et al., 2017). These compounds, however, are primarily classified as “toolkit compounds” due to a lack of adequate ADMET properties required for advancement as drug candidates to clinical trials. To date, clinical trials on GSK3 β inhibitors, including tideglusib and LY2090314, have demonstrated their tolerability, indicating that concerns about GSK3 β inhibition causing widespread metabolic toxicity were not justified. LY2090314 has a suboptimal pharmacokinetic profile, so the lack of toxicity observed could be attributed to inadequate systemic exposure (Zamek-Gliszczynski et al., 2013). The selected compounds in this study demonstrated promising drug-like properties and have been proposed to inhibit cancer progression via their interaction with GSK3 β .

4 Conclusion

In this study, an insilico screening approach was employed to investigate potential anti-cancer compounds targeting the GSK3 β protein. BMS-754807 and GSK429286A were discovered to have high binding affinity and stability to the GSK3 β protein. The favorable interactions were found to be attributed to various Van der Waals forces and H-bonding interactions. Additionally, both compounds exhibited promising drug-like properties. These findings provide a basis for further investigation *in vitro* and *in vivo* to develop potent GSK3 β inhibitors for cancer management.

Data availability statement

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

Author contributions

Conceptualization, FA and NX; formal analysis, AA, SH, and HB; methodology, LH, AS, and AA; original draft preparation, LH and FA; review and editing, AA, SA, FA, and NX. All authors read

and approved the final manuscript. All authors contributed to the article and approved the final version.

Funding

This work was supported by Discipline Collaborative Innovation Team Program of Double First-class and High-level Universities for Guangzhou University of Chinese Medicine—No. 2021XK01.

Acknowledgments

The researchers would like to acknowledge Deanship of Scientific Research, Taif University for funding this work.

References

- Banerjee, P., Eckert, A. O., Schrey, A. K., and Preissner, R. (2018). ProTox-II: A webserver for the prediction of toxicity of chemicals. *Nucleic acids Res.* 46, W257–W263. doi:10.1093/nar/gky318
- Bhat, R., Xue, Y., Berg, S., Hellberg, S., Ormo, M., Nilsson, Y., et al. (2003). Structural insights and biological effects of glycogen synthase kinase 3-specific inhibitor AR-A014418. *J. Biol. Chem.* 278, 45937–45945. doi:10.1074/jbc.M306268200
- Carboni, J. M., Wittman, M., Yang, Z., Lee, F., Greer, A., Hurlburt, W., et al. (2009). BMS-754807, a small molecule inhibitor of insulin-like growth factor-1R/IR. *Mol. Cancer Ther.* 8, 3341–3349. doi:10.1158/1535-7163.mct-09-0499
- Cohen, P., and Frame, S. (2001). The renaissance of GSK3. *Nat. Rev. Mol. Cell Biol.* 2, 769–776. doi:10.1038/35096075
- Dallakyan, S., and Olson, A. J. (2015). Small-molecule library screening by docking with PyRx. *Methods Mol. Biol.* 1263, 243–250. doi:10.1007/978-1-4939-2269-7_19
- Dickey, A., Schleicher, S., Leahy, K., Hu, R., Hallahan, D., and Thotala, D. K. (2011). GSK-3 β inhibition promotes cell death, apoptosis, and *in vivo* tumor growth delay in neuroblastoma Neuro-2A cell line. *J. Neurooncol.* 104, 145–153. doi:10.1007/s11060-010-0491-3
- Domoto, T., Uehara, M., Bolidong, D., and Minamoto, T. (2020). Glycogen synthase kinase 3 β in cancer biology and treatment. *Cells* 9, 1388. doi:10.3390/cells9061388
- Duda, P., Akula, S. M., Abrams, S. L., Steelman, L. S., Martelli, A. M., Cocco, L., et al. (2020). Targeting GSK3 and associated signaling pathways involved in cancer. *Cells* 9, 1110. doi:10.3390/cells9051110
- Gao, S., Li, S., Duan, X., Gu, Z., Ma, Z., Yuan, X., et al. (2017). Inhibition of glycogen synthase kinase 3 beta (GSK3 β) suppresses the progression of esophageal squamous cell carcinoma by modifying STAT3 activity. *Mol. Carcinog.* 56, 2301–2316. doi:10.1002/mc.22685
- Gentile, G., Merlo, G., Pozzan, A., Bernasconi, G., Bax, B., Bamborough, P., et al. (2012). 5-Aryl-4-carboxamide-1, 3-oxazoles: Potent and selective GSK-3 inhibitors. *Bioorg. Med. Chem. Lett.* 22, 1989–1994. doi:10.1016/j.bmcl.2012.01.034
- Gould, T. D., Einat, H., Bhat, R., and Manji, H. K. (2004). AR-A014418, a selective GSK-3 inhibitor, produces antidepressant-like effects in the forced swim test. *Int. J. Neuropsychopharmacol.* 7, 387–390. doi:10.1017/s1461145704004535
- Greten, F. R., and Grivennikov, S. I. (2019). Inflammation and cancer: Triggers, mechanisms, and consequences. *Immunity* 51, 27–41. doi:10.1016/j.immuni.2019.06.025
- Humphrey, W., Dalke, A., and Schulten, K. (1996). Vmd: Visual molecular dynamics. *J. Mol. Graph.* 14 (33–38), 33–38. doi:10.1016/0263-7855(96)00018-5
- Jacobs, K. M., Bhave, S. R., Ferraro, D. J., Jaboin, J. J., Hallahan, D. E., and Thotala, D. (2012). GSK-3 β : A bifunctional role in cell death pathways. *Int. J. Cell Biol.* 2012, 930710. doi:10.1155/2012/930710
- Kim, S., Kim, S. A., Han, J., and Kim, I. S. (2021). Rho-kinase as a target for cancer therapy and its immunotherapeutic potential. *Int. J. Mol. Sci.* 22, 12916. doi:10.3390/ijms222312916
- Klamer, G., Song, E., Ko, K. H., O'Brien, T. A., and Dolnikov, A. (2010). Using small molecule GSK3946; inhibitors to treat inflammation. *Curr. Med. Chem.* 17, 2873–2881. doi:10.2174/092986710792065090
- Kunnimalaiyaan, S., Schwartz, V. K., Jackson, I. A., Clark Gamblin, T., and Kunnimalaiyaan, M. (2018). Antiproliferative and apoptotic effect of LY2090314, a GSK-3 inhibitor, in neuroblastoma *in vitro*. *BMC cancer* 18, 560–568. doi:10.1186/s12885-018-4474-7
- Macalino, S. J., Gosu, V., Hong, S., and Choi, S. (2015). Role of computer-aided drug design in modern drug discovery. *Arch. Pharm. Res.* 38, 1686–1701. doi:10.1007/s12272-015-0640-5
- Mathuram, T. L., Ravikumar, V., Reece, L. M., Karthik, S., Sasikumar, C. S., and Cherian, K. M. (2016). Tideglusib induces apoptosis in human neuroblastoma IMR32 cells, provoking sub-G0/G1 accumulation and ROS generation. *Environ. Toxicol. Pharmacol.* 46, 194–205. doi:10.1016/j.etap.2016.07.013
- Medina, M., and Avila, J. (2010). Glycogen synthase kinase-3 (GSK-3) inhibitors for the treatment of Alzheimer's disease. *Curr. Pharm. Des.* 16, 2790–2798. doi:10.2174/138161210793176581
- Mishra, H., Kesharwani, R. K., Singh, D. B., Tripathi, S., Dubey, S. K., and Misra, K. (2019). Computational simulation of inhibitory effects of curcumin, retinoic acid and their conjugates on GSK-3 beta. *Netw. Model. Analysis Health Inf. Bioinforma.* 8, 3–7. doi:10.1007/s13721-018-0177-x
- Pal, K., Cao, Y., Gaisina, I. N., Bhattacharya, S., Dutta, S. K., Wang, E., et al. (2014). Inhibition of GSK-3 induces differentiation and impaired glucose metabolism in renal cancer. *Mol. Cancer Ther.* 13, 285–296. doi:10.1158/1535-7163.mct-13-0681
- Palomo, V., and Martinez, A. (2017). Glycogen synthase kinase 3 (GSK-3) inhibitors: A patent update (2014–2015). *Expert Opin. Ther. Pat.* 27, 657–666. doi:10.1080/13543776.2017.1259412
- Paul, D., Sanap, G., Shenoy, S., Kalyane, D., Kalia, K., and Tekade, R. K. (2021). Artificial intelligence in drug discovery and development. *Drug Discov. Today* 26, 80–93. doi:10.1016/j.drudis.2020.10.010
- Pires, D. E., Blundell, T. L., and Ascher, D. B. (2015). pkCSM: predicting small-molecule pharmacokinetic and toxicity properties using graph-based signatures. *J. Med. Chem.* 58, 4066–4072. doi:10.1021/acs.jmedchem.5b00104
- Pol-Fachin, L., Fernandes, C. L., and Verli, H. (2009). GROMOS96 43a1 performance on the characterization of glycoprotein conformational ensembles through molecular dynamics simulations. *Carbohydr. Res.* 344, 491–500. doi:10.1016/j.carres.2008.12.025
- Sahin, I., Eturi, A., De Souza, A., Pamarthy, S., Tavora, F., Giles, F. J., et al. (2019). Glycogen synthase kinase-3 beta inhibitors as novel cancer treatments and modulators of antitumor immune responses. *Cancer Biol. Ther.* 20, 1047–1056. doi:10.1080/15384047.2019.1595283
- Saud, S. M., Li, W., Gray, Z., Matter, M. S., Colburn, N. H., Young, M. R., et al. (2016). Diallyl disulfide (DADS), a constituent of garlic, inactivates NF- κ B and prevents colitis-induced colorectal cancer by inhibiting GSK-3 β . *Cancer Prev. Res. (Phila)* 9, 607–615. doi:10.1158/1940-6207.capr-16-0044
- Schüttelkopf, A. W., and Van Aalten, D. M. (2004). Prodrgr: A tool for high-throughput crystallography of protein-ligand complexes. *Acta Crystallogr. D. Biol. Crystallogr.* 60, 1355–1363. doi:10.1107/s0907444904011679
- Sliwoski, G., Kothiwale, S., Meiler, J., and Lowe, E. W., Jr. (2014). Computational methods in drug discovery. *Pharmacol. Rev.* 66, 334–395. doi:10.1124/pr.112.007336
- Talele, T. T., Khedkar, S. A., and Rigby, A. C. (2010). Successful applications of computer aided drug discovery: Moving drugs from concept to the clinic. *Curr. Top. Med. Chem.* 10, 127–141. doi:10.2174/156802610790232251
- Trnski, D., Sabol, M., Gojevic, A., Martinic, M., Ozretic, P., Musani, V., et al. (2015). GSK3 β and Gli3 play a role in activation of Hedgehog-Gli pathway in human colon cancer — targeting GSK3 β downregulates the signaling pathway and reduces cell proliferation. *Biochim. Biophys. Acta* 1852, 2574–2584. doi:10.1016/j.bbdis.2015.09.005

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.

- Walz, A., Ugolkov, A., Chandra, S., Kozikowski, A., Carneiro, B. A., O'halloran, T. V., et al. (2017). Molecular pathways: Revisiting glycogen synthase kinase-3 β as a target for the treatment of cancer. *Clin. Cancer Res.* 23, 1891–1897. doi:10.1158/1078-0432.ccr-15-2240
- Yuan, S., Chan, H. S., and Hu, Z. (2017). Using PyMOL as a platform for computational drug design. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* 7, e1298. doi:10.1002/wcms.1298
- Zamek-Gliszczyński, M. J., Abraham, T. L., Alberts, J. J., Kulanthaivel, P., Jackson, K. A., Chow, K. H., et al. (2013). Pharmacokinetics, metabolism, and excretion of the glycogen synthase kinase-3 inhibitor LY2090314 in rats, dogs, and humans: A case study in rapid clearance by extensive metabolism with low circulating metabolite exposure. *Drug Metab. Dispos.* 41, 714–726. doi:10.1124/dmd.112.048488
- Zeng, J., Liu, D., Qiu, Z., Huang, Y., Chen, B., Wang, L., et al. (2014). GSK3 β overexpression indicates poor prognosis and its inhibition reduces cell proliferation and survival of non-small cell lung cancer cells. *PLoS One* 9, e91231. doi:10.1371/journal.pone.0091231
- Zhang, J. S., Herreros-Villanueva, M., Koenig, A., Deng, Z., De Narvajas, A. A., Gomez, T. S., et al. (2014). Differential activity of GSK-3 isoforms regulates NF- κ B and TRAIL- or TNF α induced apoptosis in pancreatic cancer cells. *Cell Death Dis.* 5, e1142. doi:10.1038/cddis.2014.102
- Zhuang, Y. W., Wu, C. E., Zhou, J. Y., Chen, X., Wu, J., Jiang, S., et al. (2017). Solasodine inhibits human colorectal cancer cells through suppression of the AKT/glycogen synthase kinase-3 β / β -catenin pathway. *Cancer Sci.* 108, 2248–2264. doi:10.1111/cas.13354



OPEN ACCESS

EDITED BY

Khurshid Ahmad,
Yeungnam University, Republic of Korea

REVIEWED BY

Shakir Khan,
Massachusetts General Hospital and
Harvard Medical School, United States
Murali M.,
University of Mysore, India

*CORRESPONDENCE

Abdulrahim A. Alzain,
✉ abdulrahim.altoam@gmail.com

[†]These authors have contributed equally
to this work

RECEIVED 14 April 2023

ACCEPTED 22 May 2023

PUBLISHED 07 June 2023

CITATION

Mukhtar RM, Abdelmoniem N,
Elrufaie HA, Edris A, Ghaboosh H,
Mahgoub MA, Garelnabi EAE, Osman W,
Sherif AE, Ashour A, Ghazawi KF,
Samman WA, Alhaddad AA, Bafail R,
Ibrahim SRM, Mohamed GA and Alzain AA
(2023), Unlocking the potential of
approved drugs for the allosteric
inhibition of tropomyosin-receptor
kinase A using molecular docking and
molecular dynamics studies.
Front. Chem. 11:1205724.
doi: 10.3389/fchem.2023.1205724

COPYRIGHT

© 2023 Mukhtar, Abdelmoniem, Elrufaie,
Edris, Ghaboosh, Mahgoub, Garelnabi,
Osman, Sherif, Ashour, Ghazawi,
Samman, Alhaddad, Bafail, Ibrahim,
Mohamed and Alzain. 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.

Unlocking the potential of approved drugs for the allosteric inhibition of tropomyosin-receptor kinase A using molecular docking and molecular dynamics studies

Rua M. Mukhtar^{1†}, Nihal Abdelmoniem^{1†}, Hisham A. Elrufaie^{1†},
Alaa Edris¹, Hiba Ghaboosh², Mohanad A. Mahgoub¹,
Elrashied A. E. Garelnabi³, Wadah Osman^{4,5}, Asmaa E. Sherif^{4,6},
Ahmed Ashour^{4,6}, Kholoud F. Ghazawi⁷, Waad A. Samman⁸,
Aisha A. Alhaddad⁸, Rawan Bafail⁹, Sabrin R. M. Ibrahim^{10,11},
Gamal A. Mohamed¹² and Abdulrahim A. Alzain^{1*}

¹Department of Pharmaceutical Chemistry, Faculty of Pharmacy, University of Gezira, Gezira, Sudan,

²Department of Pharmaceutics, Faculty of Pharmacy, University of Gezira, Gezira, Sudan, ³Department of Pharmaceutical Chemistry, Faculty of Pharmacy, University of Khartoum, Khartoum, Sudan, ⁴Department of Pharmacognosy, Faculty of Pharmacy, Prince Sattam Bin Abdulaziz University, Al-kharj, Saudi Arabia, ⁵Department of Pharmacognosy, Faculty of Pharmacy, University of Khartoum, Khartoum, Sudan, ⁶Department of Pharmacognosy, Faculty of Pharmacy, Mansoura University, Mansoura, Egypt, ⁷Clinical Pharmacy Department, College of Pharmacy, Umm Al-Qura University, Makkah, Saudi Arabia,

⁸Department of Pharmacology and Toxicology, College of Pharmacy, Taibah University, Al-Madinah Al-Munawwarah, Saudi Arabia, ⁹Department of Pharmaceutics and Pharmaceutical Technology, College of Pharmacy, Taibah University, Medina, Saudi Arabia, ¹⁰Preparatory Year Program, Department of Chemistry, Batterjee Medical College, Jeddah, Saudi Arabia, ¹¹Department of Pharmacognosy, Faculty of Pharmacy, Assiut University, Assiut, Egypt, ¹²Department of Natural Products and Alternative Medicine, Faculty of Pharmacy, King Abdulaziz University, Jeddah, Saudi Arabia

Tropomyosin-receptor kinase A (TrkA) is the primary isoform among the tropomyosin-receptor kinases that have been associated with human cancer development, contributing to approximately 7.4% of all cancer cases. TrkA represents an attractive target for cancer treatment; however, currently available TrkA inhibitors face limitations in terms of resistance development and potential toxicity. Hence, the objective of this study was to identify new allosteric-approved inhibitors of TrkA that can overcome these challenges and be employed in cancer therapy. To achieve this goal, a screening of 9,923 drugs from the ChEMBL database was conducted to assess their repurposing potential using molecular docking. The top 49 drug candidates, exhibiting the highest docking

Abbreviations: TrkA, tropomyosin-receptor kinase A; TrkB, tropomyosin-receptor kinase B; TrkC, tropomyosin-receptor kinase C; NGF, nerve growth factor; Trks, tropomyosin-receptor kinases; RTKs, receptor tyrosine kinases; BDNF, brain-derived neurotrophic factor; NT4/5, neurotrophin-4/5; NT3, neurotrophin-3; ATP, adenosine triphosphate; JM, juxtamembrane; DFG, ASP-PHE-GLY; FFF, PHE-PHE-PHE; CNS, central nervous system; PDB, Protein Data Bank; OPLS, optimized potentials for liquid simulations; FDA, Food and Drug Administration; HTVS, high-throughput virtual screening; SP, standard precision; XP, extra-precision; MM-GBSA, molecular mechanics-generalized born surface area; MD, molecular dynamics; TIP3P, transferable interaction potential; NVT, isothermal-isochoric; NPT, isothermal-isobaric; RMSD, root mean square deviation; RMSF, root mean square fluctuation.

scores (−11.569 to −7.962 kcal/mol), underwent MM-GBSA calculations to evaluate their binding energies. Delanzomib and tibalosin, the top two drugs with docking scores of −10.643 and −10.184 kcal/mol, respectively, along with MM-GBSA dG bind values of −67.96 and −50.54 kcal/mol, were subjected to 200 ns molecular dynamic simulations, confirming their stable interactions with TrkA. Based on these findings, we recommend further experimental evaluation of delanzomib and tibalosin to determine their potential as allosteric inhibitors of TrkA. These drugs have the potential to provide more effective and less toxic therapeutic alternatives. The approach employed in this study, which involves repurposing drugs through molecular docking and molecular dynamics, serves as a valuable tool for identifying novel drug candidates with distinct therapeutic uses. This methodology can contribute to reducing the attrition rate and expediting the process of drug discovery.

KEYWORDS

cancer, tropomyosin-receptor kinase A, repurposing, molecular docking, molecular dynamics, drug discovery, health and wellbeing

1 Introduction

Tropomyosin-receptor kinases (Trks), a subfamily of the protein kinase superfamily, belong to the receptor tyrosine kinases and consist of three isoforms: TrkA, TrkB, and TrkC. These isoforms function as receptors for the neurotrophin family, which includes high-affinity growth factors such as nerve growth factor (NGF), which binds to TrkA, brain-derived neurotrophic factor (BDNF) and neurotrophin-4/5 (NT4/5), which bind to TrkB, and neurotrophin-3 (NT3), which binds to TrkC (Wang et al., 2009).

Previous experimental research has provided cumulative data indicating the involvement of Trks in the pathogenesis of a diverse range of human cancers, which has led to their recognition as promising targets for cancer treatment (Wang et al., 2009; Alam et al., 2017). TrkA, in particular, is considered oncogenic, with mounting evidence pointing to its overexpression and involvement in cancer development (Griffin et al., 2020). It is the most common isoform of Trks and is frequently associated with gene mutations or fusions, which result in the formation of oncogenes responsible for approximately 7.4% of all human cancer cases (Guo et al., 2022).

Trk inhibitors can be classified into four categories based on their binding interactions: type I, type II, type III, and type IV (Wu et al., 2015). Type I inhibitors are ATP-competitive and bind to the ATP active site. Type II inhibitors, on the other hand, are ATP non-competitive and exhibit pseudo-competitive binding kinetics by extending into a deep hydrophobic pocket within the ATP-binding site. Type III inhibitors are allosteric and bind adjacent to the ATP-binding site, while type IV inhibitors bind to regions other than the kinase domain of the protein (Yan et al., 2019). Type II inhibitors offer higher selectivity than type I inhibitors, but their large molecular size limits their druggability. However, both type I and type II inhibitors face challenges due to the emergence of secondary mutations in the ATP active site of Trks, particularly TrkA. Type III and type IV inhibitors provide isoform selectivity, although the effectiveness of type IV inhibitors as anticancer agents remains uncertain (Yan et al., 2019). Therefore, this study aims to identify allosteric TrkA selective inhibitors (type III) to overcome the existing limitations of TrkA inhibitors. These inhibitors could

potentially be used in the management of various cancers associated with TrkA activation, such as lung, breast, cervix, thyroid, and oral cavity cancers (Lagadec et al., 2009; Sasahira et al., 2013; Faulkner et al., 2018; Gao et al., 2018; Faulkner et al., 2020).

Similar to other RTKs, TrkA comprises three domains: an extracellular domain responsible for ligand binding, a transmembrane domain, and an intracellular catalytic domain (Amatu et al., 2019). The region between the transmembrane domain and the catalytic domain, known as the juxtamembrane (JM) region, consists of approximately 60 residues. Interestingly, this region exhibits approximately 36% similarity with TrkB and 40% similarity with TrkC (Su et al., 2017; Furuya et al., 2017). In the inactive state, the Asp-Phe-Gly (DFG) motif of TrkA's activation loop adopts an "out" conformation. This conformation is stabilized by edge-to-face interactions involving three phenylalanine residues: DFG motif Phe669, gatekeeper Phe589, and back pocket Phe575. Together, they form a unique FFF motif, along with the Leu564 residue in the α -C helix and the JM region. This combination generates an allosteric site adjacent to the ATP-binding site of TrkA (Bagal et al., 2019). X-ray crystallography studies have revealed that this allosteric site binds type III inhibitors, effectively maintaining TrkA in an inactive conformation (Simard et al., 2009; Heinrich et al., 2010). Therefore, this study leverages the knowledge of this allosteric site to identify type III TrkA inhibitors, utilizing its potential for modulating TrkA activity.

The process of drug discovery and development is known for its high attrition rate, involving significant time, cost, and effort, making the introduction of a new drug to the market a challenging endeavor (Mohammed et al., 2022; Gowtham et al., 2022). In the field of cancer research, the strategy of drug repositioning or repurposing has gained widespread application. This strategy involves repurposing approved or investigational drugs for new indications that were not initially intended for their use (Pushpakom et al., 2018; Gazerani, 2019; Omer et al., 2022). By leveraging existing drugs, the drug repositioning approach significantly reduces the time required for the drug discovery process by 3–5 years, lowers costs by \$0.3 billion, and reduces failure rates in the later stages of development. This is because the drugs being investigated have already demonstrated sufficient

safety profiles, enabling them to swiftly enter phases II and III of clinical trials (Fu et al., 2022; Issa et al., 2021).

Computational techniques play a vital role in drug repurposing, encompassing various approaches, such as molecular docking, genetic association, pathway mapping, data mining, and signature matching (Fu et al., 2022).

In this study, molecular docking coupled with MM-GBSA calculations and molecular dynamics (MD) simulations were employed to investigate drugs from the ChEMBL database. The aim of this study was to assess their potential for repurposing as drug candidates for cancer treatment, specifically targeting the TrkA allosteric site.

2 Materials and methods

All *in silico* studies, with the exception of the molecular dynamics (MD) simulations, were conducted using Maestro v12.8 from Schrödinger. The MD simulations were performed using Academic Desmond v6.5 by D.E. Shaw Research.

2.1 Protein and ligand preparation

The crystallographic structure of TrkA, along with the co-crystallized ligand (PDB ID: 6D20) (Bagal et al., 2019), was obtained from the Protein Data Bank (PDB) (<https://www.rcsb.org/>). To prepare the TrkA structure for subsequent calculations, a three-step processing procedure was performed using the Protein Preparation Wizard in Maestro.

In the first step, basic adjustments were made to the protein structure, including assigning bond orders, adding hydrogen atoms to those that were missing, creating zero-order bonds for metals and disulfide bonds, converting selenomethionines to methionines, filling in missing side chains and loops, removing water molecules beyond 5.00 Å from heterogroups, and generating potential ionization states of heteroatoms at a pH of 7 ± 2 .

The second step involved optimizing hydrogen bonds and assigning orientations to the crystalized water molecules. The protonation states of the residues were also determined using the PROPKA tool at a pH of 7.0.

Finally, the third step involved restrained minimization, which was performed using the OPLS4 force field (Abdelmoniem et al., 2023). This step aimed to achieve a more stable and energetically favorable conformation of the TrkA structure for subsequent calculations.

We downloaded the drugs library from the ChEMBL database at <https://www.ebi.ac.uk/chembl/>. Specifically, we focused on the category of drug molecules. Within this category, we narrowed our focus to small molecules, which encompass various types, such as FDA-approved, world-approved, and investigational compounds. In total, we selected 9,923 small molecules from this category, representing a diverse range of therapeutic classes. To prepare the library for further analysis, the LigPrep tool in Maestro was employed (Alzain and Elbadwi, 2021). LigPrep not only generated low-energy three-dimensional structures for the input compounds but also produced multiple output structures for each compound. This was achieved by considering various factors, such

as possible ionization states, tautomers, and stereoisomers. The LigPrep process was executed with the default settings, ensuring comprehensive exploration of the chemical space represented by the drug library.

2.2 Grid generation and molecular docking

The prepared protein structure underwent the receptor grid generation process, a crucial step for ligand docking. This process generated a grid file representing the site on the receptor where the ligand docking would occur. The receptor grid generation panel in Maestro was utilized to configure the grid generation job (Mohamed et al., 2022). The ligand molecule that bound to the TrkA allosteric site was identified and excluded from the grid generation process. This step helped define the position and size of the allosteric site surrounding the ligand. The van der Waals scaling and other options in the panel were kept at their default settings, and the grid generation process was initiated.

To evaluate the strength and affinity of the compounds toward the target's allosteric site, the prepared library underwent molecular docking using the ligand docking panel in the Glide tool of Maestro (Elbadwi et al., 2021; Alzain et al., 2022). Initially, the library was subjected to a high-throughput virtual screening (HTVS) mode. The top compounds were then filtered based on their docking scores and subsequently subjected to an extra-precision (XP) docking mode. This multi-step docking process enabled the identification of potential compounds that exhibited favorable binding characteristics and affinity for the TrkA allosteric site. As a reference, the co-crystallized ligand was also docked onto the allosteric site.

2.3 MM-GBSA calculations

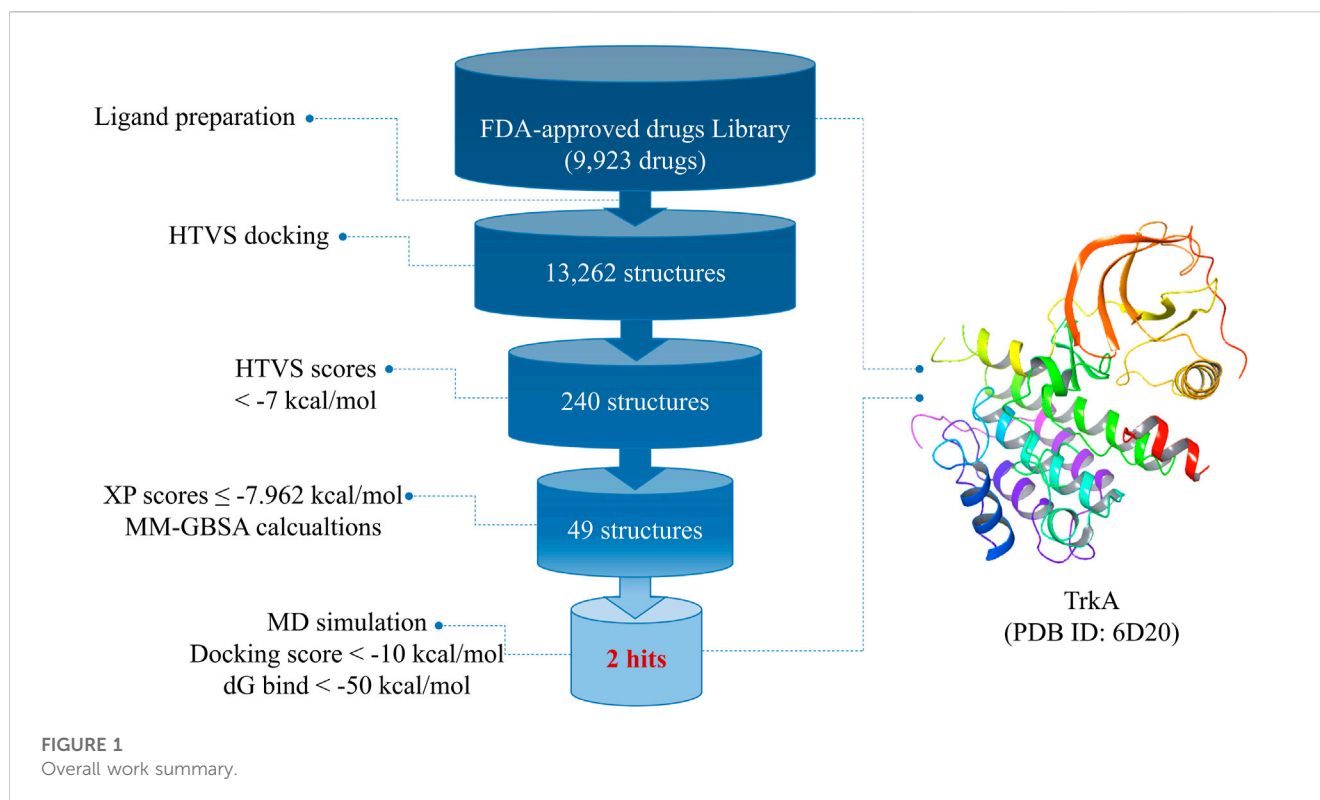
The ligand that poses with the best docking scores were selected and subjected to free-binding energy calculations using the molecular mechanics-generalized born surface area (MM-GBSA) method. These calculations were performed using the Prime tool in Maestro. The MM-GBSA method was utilized to estimate the free-binding energy of the ligand–receptor complex. The specific equation employed in these calculations to determine the free-binding energy is as follows:

$$\Delta E = E_c - E_R - E_L,$$

where ΔE is the free-binding energy, E_c is the ligand–receptor complex energy, E_R is the receptor energy, and E_L is the ligand energy (Obubeid et al., 2022). The force field and the solvent model were set to be OPLS4 and VSGB, respectively.

2.4 MD simulation

The two ligand–protein complexes with the best docking scores and free-binding energy, as well as the co-crystallized ligand, were chosen for the molecular dynamics (MD) simulation study. The MD simulations were conducted using Desmond software (Alzain et al., 2022).



Prior to the simulation process, the biological system was set up using the System Builder panel in Desmond. This involved solvating the ligand–protein complexes with 12,437 TIP3P water molecules in an orthorhombic-shaped box with dimensions of $10 \times 10 \times 10 \text{ \AA}$. Additionally, 51.167 mM of Na^+ ions (with a total charge of +35) and 51.167 mM of Cl^- ions (with a total charge of –35) were added as salt to maintain the system's electrostatic neutrality. The OPLS4 force field was employed to minimize the energy of the system. Subsequently, the system underwent equilibration in two ensembles: isothermal–isochoric (NVT) and isothermal–isobaric (NPT). During the NVT ensemble, the system's temperature was maintained at 300 K, while during the NPT ensemble, both temperature and atmospheric pressure (1 bar) were kept constant. The Nose–Hoover chain thermostat and the Martyna–Tobias–Klein barostat methods were employed to maintain the desired temperature and pressure conditions, respectively. The trajectory was recorded at a 100-ps interval, resulting in a total of 2,000 frames.

The analysis of the simulation results was performed using the Simulation Interaction Diagram tool provided by Desmond.

3 Results and discussion

Figure 1 provides an overview of the research workflow, which employed various *in silico* methods to investigate the potential discovery of drug candidates from FDA-approved drugs for the inhibition of TrkA protein kinases. In the drug discovery process, molecular docking and molecular dynamics simulations play crucial

roles in understanding ligand–receptor interactions. These computational approaches are particularly valuable for developing medications targeting new and challenging diseases such as cancer. The research also utilized virtual screening and drug repurposing strategies to identify potential drug candidates. By leveraging these *in silico* techniques, the study aimed to uncover promising candidates for TrkA inhibition.

3.1 Molecular docking

The molecular docking analysis was performed using the Glide module of Schrödinger. Glide is a powerful tool that accurately determines the positions and orientations of ligands within the active site of the receptor, providing valuable information on the compounds' affinity and activity (Alzain and Elbadwi, 2021; Meng et al., 2011). It employs various scoring functions to rank and select the best poses for further analysis (Friesner et al., 2004). Glide offers three levels of docking methodologies: high-throughput virtual screening (HTVS), standard precision (SP), and extra-precision (XP). Each methodology differs in accuracy, with HTVS being the fastest but least accurate and XP being the most accurate but time-consuming. The docking time for screening one compound ranges from 2 s (HTVS) to 2 min (XP) (Eltaib and Alzain, 2022). These methodologies can be used sequentially to efficiently filter a large number of compounds. Furthermore, the molecular docking performed by Glide sets the stage for predicting the free-binding energy using methods such as MM-GBSA calculations.

TABLE 1 Docking scores and MM-GBSA dG bind energies of the nine selected best ligand poses and the reference bound to TrkA allosteric site.

Compound name	Docking score kcal/mol	MM-GBSA dG bind kcal/mol
Delanzomib	−10.643	−67.96
Tibalosin	−10.184	−50.54
Vismodegib	−9.948	−53.56
Hexoprenaline	−9.666	−62.39
Merestininib	−9.342	−64.24
Etanterol	−9.146	−55.76
Ractopamine	−9.117	−53.23
Primidolol	−9.084	−54.35
Cliropamine	−9.022	−52.04
TrkA–ligand	−10.689	−105.51

After preparing the library of FDA-approved drugs (9,923 molecules) using the LigPrep tool, we generated a total of 23,334 conformers and tautomers. These compounds were subjected to molecular docking against the TrkA allosteric site using the high-throughput virtual screening (HTVS) mode. Among them, 230 structures with docking energies below -7 kcal/mol were identified as potential ligands. Since this number was manageable for further analysis, these 230 structures were directly subjected to molecular docking using the extra-precision (XP) mode, bypassing the standard precision (SP) level. Among the XP docking results, 49 structures were selected based on their docking scores, which ranged from -11.569 to -7.962 kcal/mol, for subsequent free-binding energy prediction.

3.2 MM-GBSA calculations

Docking results provide insights into whether ligands bind to the active site of the target protein. However, to determine if this binding is stable and capable of eliciting a response, it is crucial to assess the free-binding energy of the receptor–ligand complex (Lyne et al., 2006). Therefore, the top 49 structures from the docking results were further analyzed using the MM-GBSA method, which accounts for the solvent's influence on the ligand–protein complex binding. For comparison, the co-crystallized ligand of TrkA was also subjected to XP and MM-GBSA calculations as a reference.

Among the 49 structures, nine drugs were selected based on their docking scores (<-9) and MM-GBSA dG bind energies (<-50 kcal/mol) for further investigation (Table 1). As shown in Table 1, none of the nine chosen drugs achieved better docking scores or MM-GBSA dG bind energies than the reference compound, which had a docking score of -10.689 and MM-GBSA dG bind of -105.51 kcal/mol. However, the results are considered satisfactory since the difference in docking scores between the selected compounds and the reference is minimal, and their MM-GBSA dG bind energies are highly favorable.

Among the chosen drugs, delanzomib and tibalosin, with docking scores of -10.643 and -10.184 kcal/mol and MM-GBSA dG bind energies of -67.96 and -50.54 kcal/mol, respectively, stood out as representatives for further analysis of their interaction patterns.

3.3 Ligand–residue interaction analysis

The delanzomib/TrkA complex exhibited three hydrogen bonds with LEU486, LYS544, and GLY670 residues, along with hydrophobic contacts with LEU486, PHE521, LEU564, LEU567, ILE572, VAL573, PHE575, PHE589, LEU641, PHE646, ILE666, and PHE669 (Table 2; Figure 2A), while the tibalosin/TrkA complex formed one hydrogen bond with ASP668, one salt bridge with ASP668, and hydrophobic contacts with LEU486, PHE521, LEU564, LEU567, PHE589, ILE572, VAL573, LEU641, PHE646, ILE666, and PHE669 (Table 2; Figure 2B). On the other hand, the reference/TrkA complex exhibited six hydrogen bonds, involving GLY483, SER484, LEU486, ARG673, and ASP668 residues. Additionally, it formed one halogen bond with HIE648, one pi-cation interaction with LYS544, and hydrophobic contacts with LEU486, LEU564, LEU567, MET587, PHE589, ILE572, VAL573, PHE575, LEU641, PHE646, ILE666, and PHE669 (Table 2; Figure 2C).

The ligand–residue interaction analysis provides insights into the differences observed in the docking scores and MM-GBSA dG bind energies among the reference, delanzomib, and tibalosin complexes. The reference compound showed the highest number of hydrogen bonds (6), followed by delanzomib (3) and Tibalosin (1). This highlights the importance of hydrogen bond interactions in contributing to the binding affinity of ligands (Klebe and Böhm, 1997; Chen et al., 2016; Anandan et al., 2022).

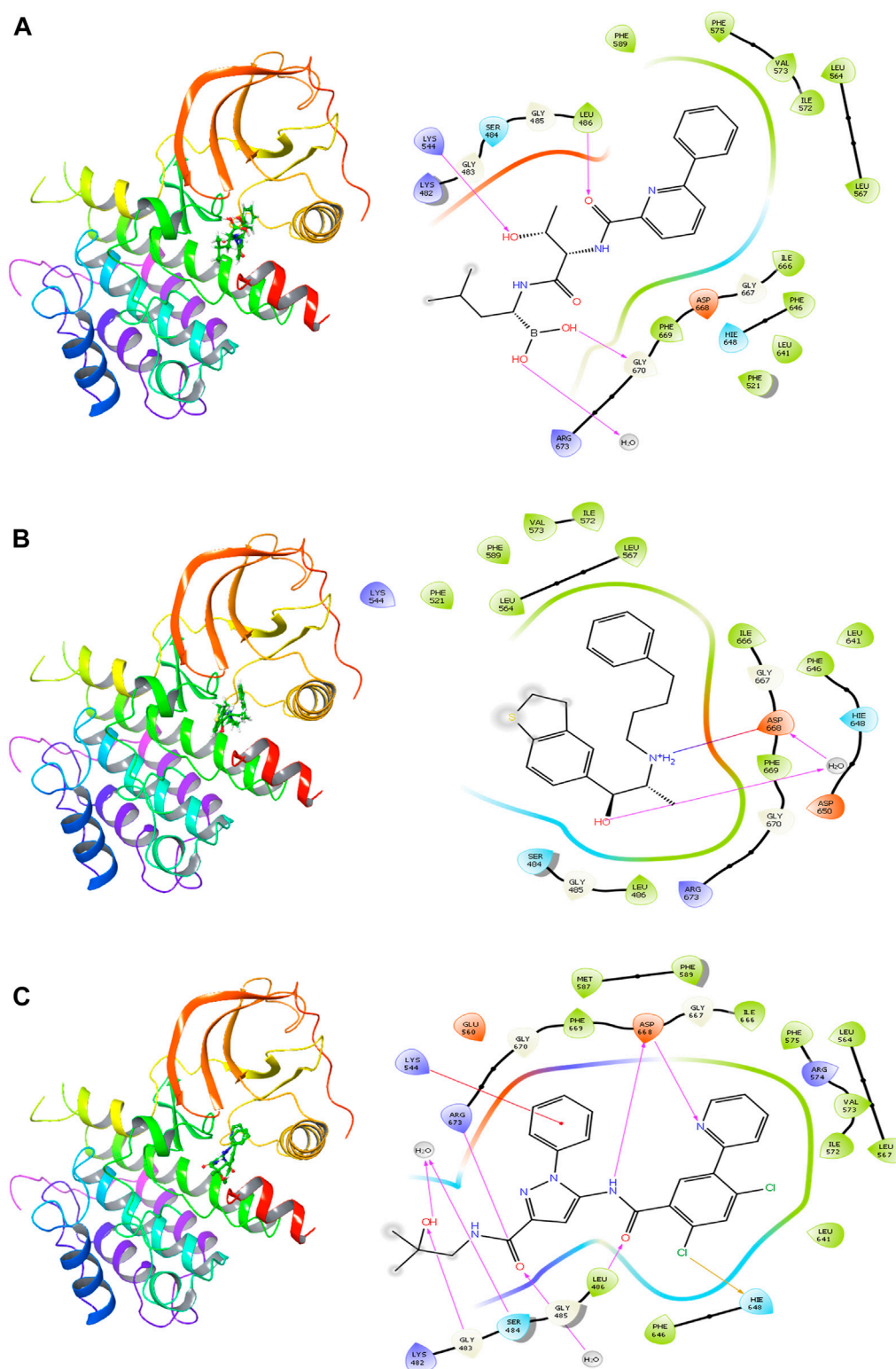
It is worth noting that the interactions observed between delanzomib, tibalosin, and specific residues in the TrkA protein align with findings from previous research articles investigating small molecules as TrkA allosteric inhibitors (Furuya et al., 2017; Su

TABLE 2 Ligand–residue interactions of delanzomib and tibalosin at the TrkA allosteric site.

Compound name	H-bond	Salt bridge	Hydrophobic interaction	Other interaction
Delanzomib	LEU486, LYS544, and GLY670	-	LEU486, PHE521, LEU564, LEU567, ILE572, VAL573, PHE575, PHE589, LEU641, PHE646, ILE666, and PHE669	Polar interaction: SER484 and HIE648
				Charged negative: ASP668
				Charged positive: LYS482, LYS544, and ARG673
Tibalosin	ASP668	ASP668	LEU486, PHE521, LEU564, LEU567, PHE589, ILE572, VAL573, LEU641, PHE646, ILE666, and PHE669	Polar interaction: SER484 and HIE648
				Charged negative: ASP668 and ASP650
				Charged positive: LYS544 and ARG673
Reference	GLY483, SER484, LEU486, ASP668, and ARG673	-	LEU486, LEU564, LEU567, MET587, PHE589, ILE572, VAL573, PHE575, LEU641, PHE646, ILE666, and PHE669	Polar interaction: SER484 and HIE648
				Charged negative: GLU560 and ASP668
				Charged positive: LYS482, LYS544, ARG574, and ARG673
				Halogen bond: HIE648
				Pi–cation: LYS544

et al., 2017; Bagal et al., 2019; Subramanian et al., 2019; Guo et al., 2022). For instance, ASP668 has been reported to form hydrogen bonds with several top inhibitors discovered by different scientific groups. In the case of tibalosin, ASP668 interacts with the hydroxyl group and the amino group of the (2R)-2-[(4-phenylbutyl)amino]propan-1-ol moiety through hydrogen bond and salt bridge interactions, respectively. Although delanzomib does not form a hydrogen bond with ASP668, it establishes hydrogen bond interactions with LEU486, LYS544, and GLY670, which have also been documented in previous studies (Furuya et al., 2017; Bagal et al., 2019; Subramanian et al., 2019; Guo et al., 2022). Furthermore, previous studies by Guo et al. (2022) and Bagal et al. (2019). have emphasized the significance of hydrophobic interactions with LYS544, LEU564, and PHE589 in TrkA

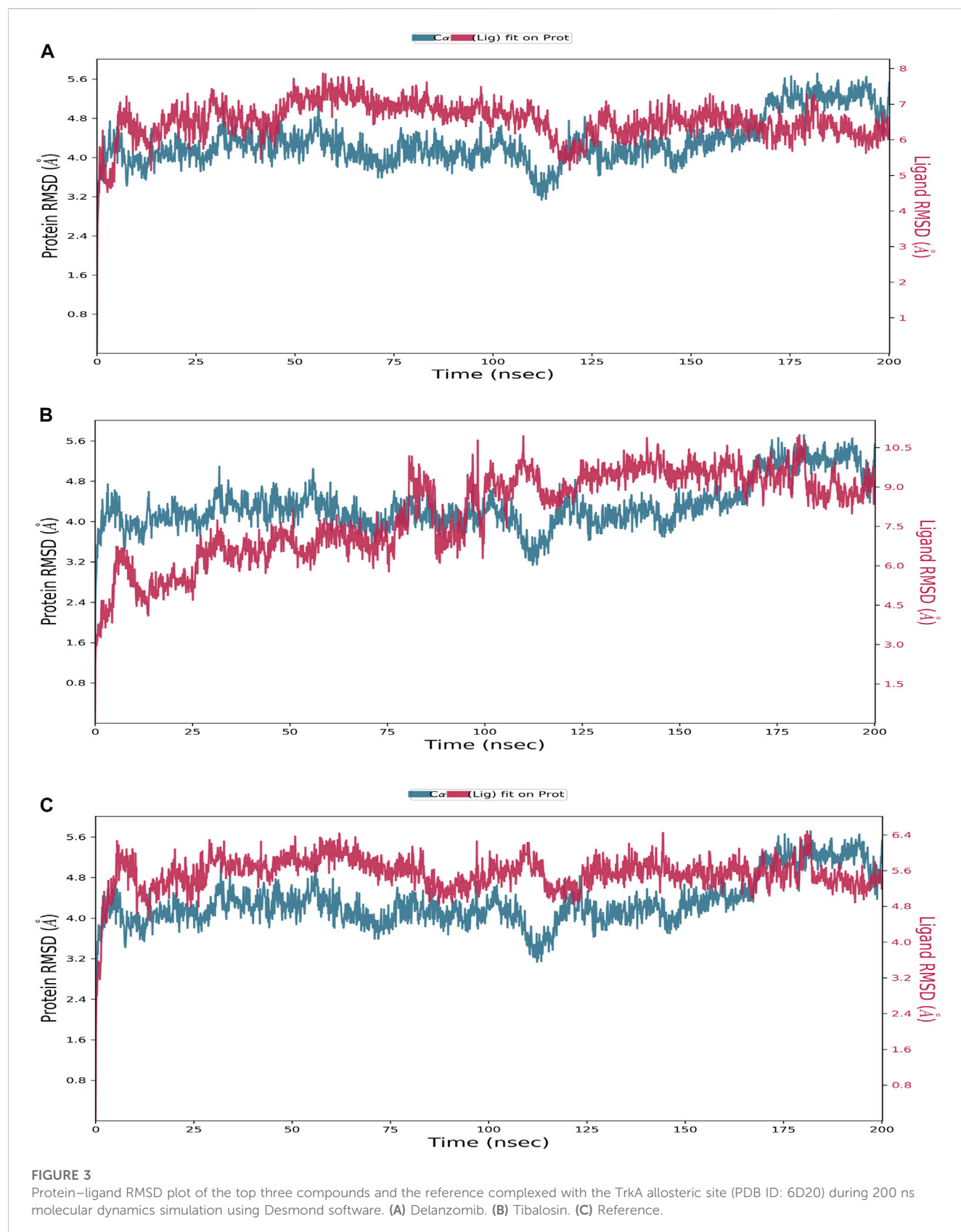
allosteric inhibitors. In the case of delanzomib and tibalosin, both compounds form hydrophobic contacts with LEU564 and PHE589 and establish a charged positive contact with LYS544. Additionally, a thorough review of the identity and previous records of delanzomib and tibalosin was conducted to determine if any documented activity or correlation with cancer treatment exists. Delanzomib is an orally active boronate-based proteasome inhibitor that specifically targets the chymotrypsin-like activity of the proteasome (Doloff, 2015). While information regarding delanzomib’s impact on bone remodeling is limited, one study has explored its effects on osteoclasts (Zangari and Suva, 2016). In a study by Mopei et al., delanzomib demonstrated promising efficacy and antimutagenic properties in human multiple

**FIGURE 2**

2D and 3D interactions of the best three hits with the TrkA allosteric site (PDB ID: 6D20) using Glide software. (A) Delanzomib. (B) Tibalotin. (C) Reference.

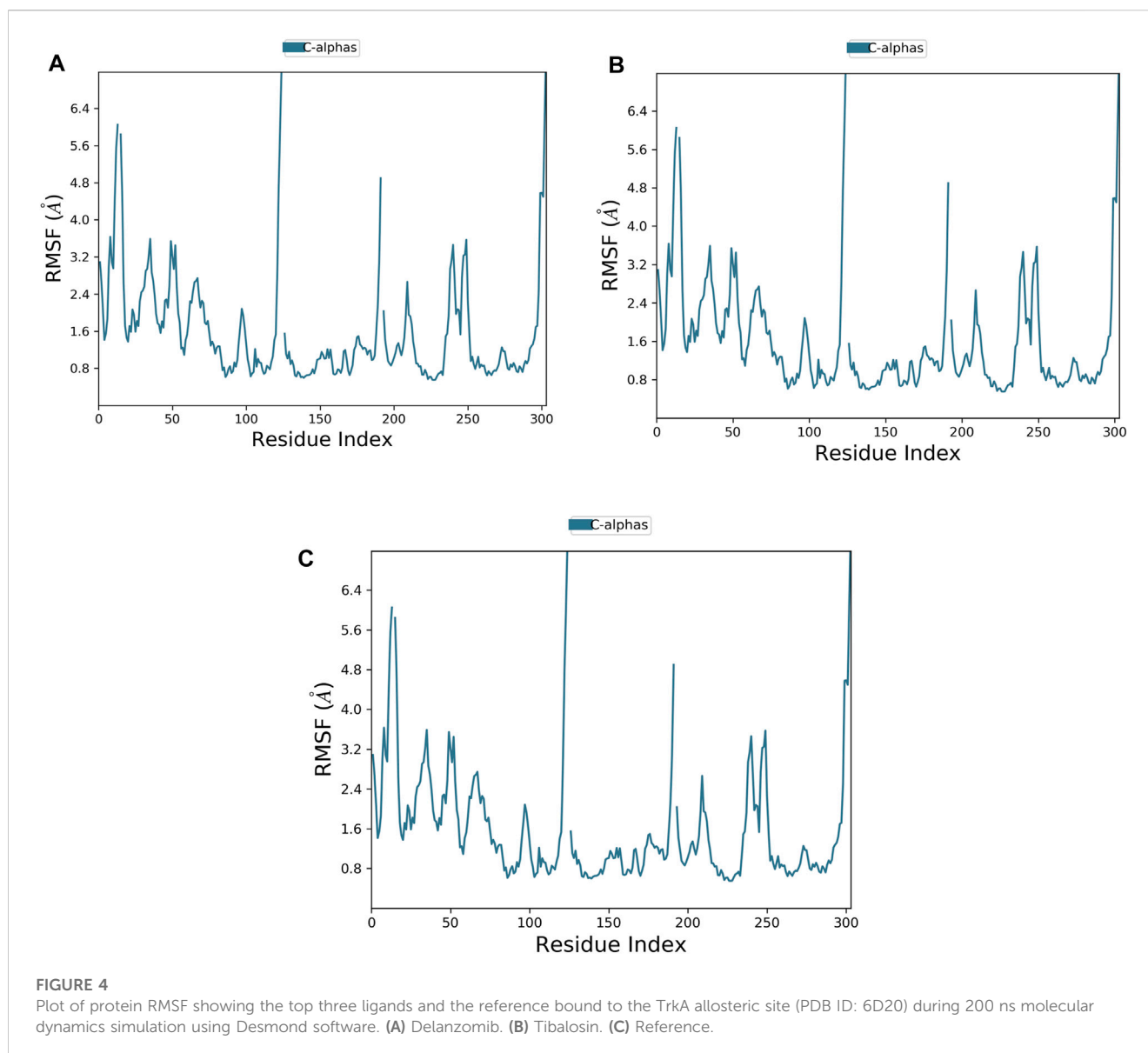
myeloma cell lines and patient-derived cells (Wang et al., 2019). Additionally, delanzomib has shown significance in the treatment of renal cell carcinoma (RCC). When combined

with ritonavir, these two drugs exhibited synergistic effects in suppressing colony formation and inhibiting the growth of renal cancer (Isono et al., 2018).



Tibalosin, a phenylethylamine derivative, has been shown to reduce arterial pressure in hypertension animal models (Staessen et al., 1983). It exhibits potent antihypertensive effects; however, side effects prevent its clinical use at a daily dose of 150 mg. Combination

therapy with a beta-adrenoceptor-blocking medication appears to be more effective in treating hypertension than thiazide therapy alone (Staessen et al., 1986). Currently, there is no available data linking tibalosin to cancer or its anticancer properties.



In conclusion, based on the docking, MM-GBSA, and interaction pattern analysis results, delanzomib and tibalosin demonstrate promise as TrkA inhibitors. Delanzomib has reported an anticancer activity, while tibalosin presented a potential anticancer activity as a TrkA inhibitor, which is being reported for the first time in this study. Furthermore, both drugs were subjected to a 200 ns molecular dynamics (MD) simulation study to further explore their behavior and interactions.

3.4 MD simulations

The previous techniques employed rigid structures for proteins and ligands, whereas molecular dynamics (MD) simulation takes into account the conformational changes in the receptor and ligand. MD simulation provides a more realistic representation of the dynamic behavior occurring under physiological conditions, allowing for a thorough investigation of the complex's stability,

flexibility, and binding interactions (Kumar et al., 2019; Jordaan et al., 2020; Aghajani et al., 2022; Gowtham et al., 2023).

In this study, MD simulations were conducted for the complexes of the two best compounds, delanzomib and tibalosin, with TrkA, as well as the reference structure (the co-crystallized ligand of 6D20). The analyzed data include the root mean square deviation (RMSD), the root mean square fluctuation (RMSF), and the protein–ligand contacts observed during the 200 ns simulation.

Starting with the RMSD analysis of the protein's C α atoms (Figure 3), it can be observed that the protein exhibited a similar pattern of deviations with an average RMSD of 4.31 Å when complexed with the two compounds and the reference. This average RMSD value is relatively compatible, as an RMSD of 1–3 Å is generally acceptable for small globular proteins (Obubeid et al., 2022). It indicates the overall stability of the TrkA–ligand complexes. Delanzomib exhibited the lowest range of RMSD values, indicating greater stability than tibalosin and the

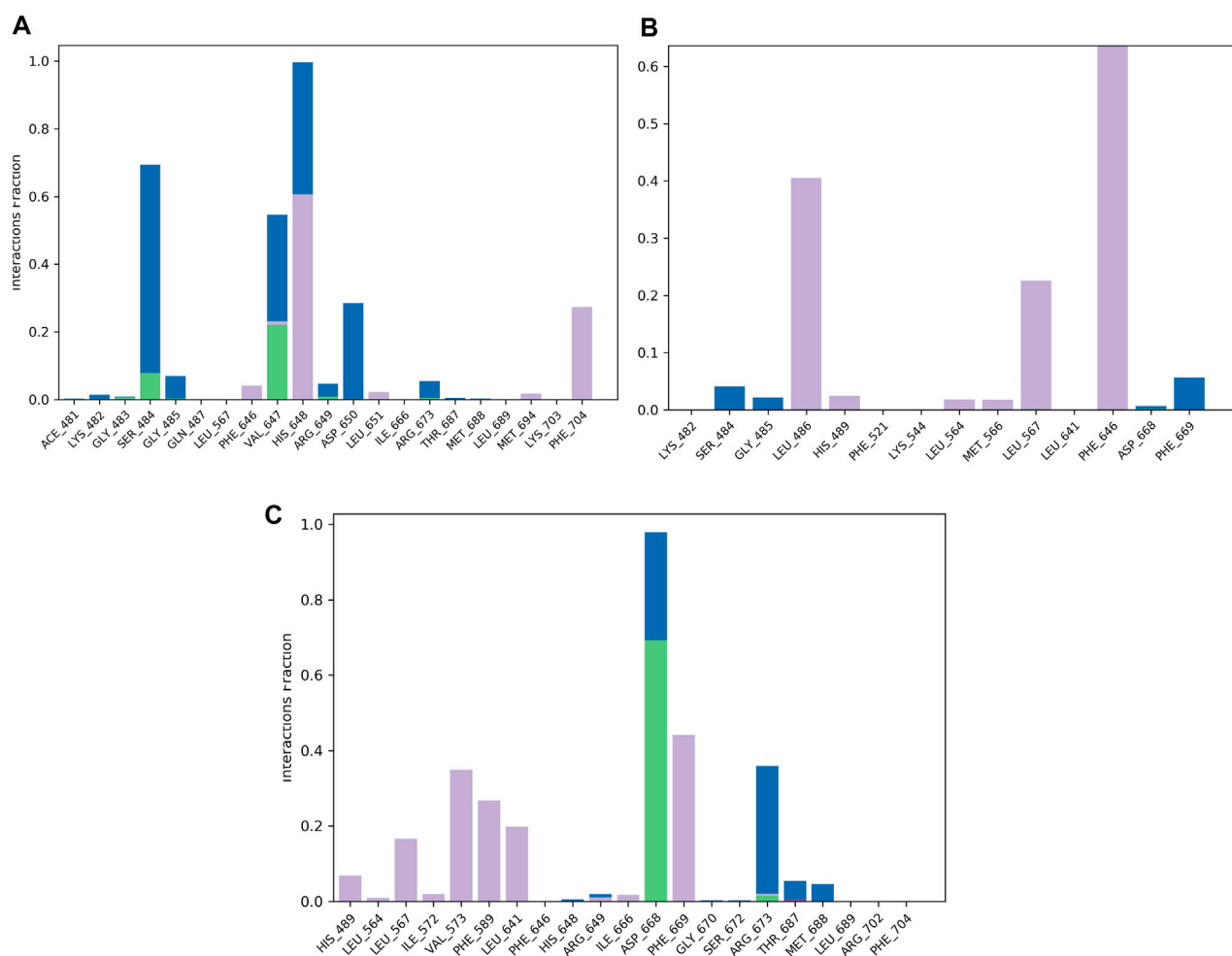


FIGURE 5

Protein–ligand contact histogram of the top three compounds and the reference complexed with the TrkA allosteric site (PDB ID: 6D20) during 200 ns molecular dynamics simulation using Desmond software. (A) Delanzomib. (B) Tibalosin. (C) Reference.

reference. Delanzomib also showed minimal fluctuations along the simulation duration, with an average ligand RMSD of 2.81 Å. It is worth noting that the behavior of delanzomib closely resembled that of the reference.

On the other hand, tibalosin initially exhibited high fluctuations during the first 100 ns of the simulation, but its behavior became more similar to the reference in the second half of the simulation. Although tibalosin showed higher fluctuations, its average ligand RMSD of 1.63 Å was the smallest among delanzomib (2.81 Å) and the reference (2.61 Å).

The flexibility of the TrkA protein and the movement of its residues were assessed by monitoring the RMSF value of the Ca atoms. A lower RMSF value indicates less flexibility and greater stability (Alzain, 2022). As shown in Figure 4, the protein exhibited similar RMSF patterns with both compounds and the reference, with an average RMSF value of 1.59 Å. The low average RMSF value, combined with the previously discussed average RMSD values, confirms the stability of the studied complexes.

The protein–ligand contact histogram (Figure 5) provided information about the binding and non-binding interactions between the protein and the two compounds, as well as the reference, during the simulation. The delanzomib–TrkA complex formed contacts with SER484 (H-bond 10% and water bridges 50%), VAL647 (H-bond 25%, hydrophobic 3%, and water bridges 22%), HIS648 (hydrophobic 60% and water bridges 40%), ASP650 (water bridges 25%), and PHE704 (hydrophobic 30%).

The tibalosin–TrkA complex interacted with LEU486 (hydrophobic 40%), LEU567 (hydrophobic 20%), PHE646 (hydrophobic 70%), and PHE669 (water bridges 5%). Considering the significant role of H-bonds in the binding of an inhibitor to a kinase (Wu et al., 2021), these results suggest that delanzomib exhibits higher inhibitory activity than tibalosin. This conclusion is supported by the protein–ligand contact histogram, which shows that delanzomib forms H-bonds with two residues, whereas tibalosin does not form any H-bonds.

The reference–TrkA complex had contacts with LEU567 (hydrophobic 15%), VAL573 (hydrophobic 35%), PHE589 (hydrophobic 30%), LEU641 (hydrophobic 20%), ASP668 (H-bond

70% and water bridge 30%), PHE669 (hydrophobic 45%), and ARG673 (H-bond 2%, hydrophobic 3%, and water bridges 37%).

In conclusion, based on the MD results, delanzomib and tibalosin were identified as type III inhibitors, as they exhibited similar effects on the protein compared to the reference molecule (a potent, selective, and allosteric type III TrkA binder named molecule 23) in terms of RMSD and RMSF plots (Bagal et al., 2019).

4 Conclusion

TrkA, the most prevalent isoform associated with a wide range of human malignancies, is a crucial target for cancer therapy. This study aimed to identify potential allosteric TrkA inhibitors for the treatment of cancer. To achieve this objective, multiple computational approaches were employed to screen a library of 9,923 approved drugs from the ChEMBL database, assessing their repurposing potential as allosteric inhibitors against the TrkA protein. Initially, the library was docked into the allosteric site of TrkA using HTVS and XP modes. This screening process yielded 49 compounds with favorable docking scores, which were further evaluated through MM-GBSA calculations to determine their free-binding energies. Among the 49 compounds, nine exhibited MM-GBSA dG bind energies below -50 and were selected for detailed analysis in this study. The interaction patterns of the top two drugs, delanzomib and tibalosin, were examined. These compounds displayed several common interactions with previously identified TrkA allosteric inhibitors, and notably, delanzomib has been reported to possess antimutagenic and anti-cancer effects. Subsequently, delanzomib and tibalosin underwent MD simulations, demonstrating good stability at the protein's allosteric site. Based on these findings, delanzomib and tibalosin are considered promising hits against TrkA. Further experimental investigations are warranted to validate their potential as inhibitors of this protein, holding significant prospects for future cancer therapies.

References

- Abdelmoniem, N., H. Abdallah, M., M. Mukhtar, R., Moutasim, F., Rafie Ahmed, A., Edris, A., et al. (2023). Identification of novel natural dual HDAC and Hsp90 inhibitors for metastatic TNBC using e-pharmacophore modeling, molecular docking, and molecular dynamics studies. *Molecules* 28, 1771. doi:10.3390/molecules28041771
- Aghajani, J., Farnia, P., Farnia, P., Ghanavi, J., and Velayati, A. A. (2022). Molecular dynamic simulations and molecular docking as a potential way for designed new inhibitor drug without resistance. *Tanaffos* 21 (1), 1–14.
- Alam, M. S., Choi, S. U., and Lee, D. U. (2017). Synthesis, anticancer, and docking studies of salicyl-hydrazone analogues: A novel series of small potent tropomyosin receptor kinase A inhibitors. *Bioorg. Med. Chem.* 25 (1), 389–396. doi:10.1016/j.bmc.2016.11.005
- Alzain, A. A., and Elbadwi, F. A. (2021). Identification of novel TMPRSS2 inhibitors for COVID-19 using e-pharmacophore modelling, molecular docking, molecular dynamics and quantum mechanics studies. *Inf. Med. Unlocked* 26, 100758. doi:10.1016/j.imu.2021.100758
- Alzain, A. A. (2022). Insights from computational studies on the potential of natural compounds as inhibitors against SARS-CoV-2 spike omicron variant. *Sar. QSAR Environ. Res.* 00 (00), 953–968. doi:10.1080/1062936X.2022.2152486
- Alzain, A. A., Ismail, A., Fadlilmola, M., A Mohamed, M., Mahjoub, M., A Makki, A., et al. (2022). De novo design of novel spike glycoprotein inhibitors using e-pharmacophore modeling, molecular hybridization, ADMET, quantum mechanics and molecular dynamics studies for COVID-19. *Pak J. Pharm. Sci.* 35, 313–321.
- Amatu, A., Sartore-Bianchi, A., Bencardino, K., Pizzutillo, E. G., Tosi, F., and Siena, S. (2019). Tropomyosin receptor kinase (TRK) biology and the role of NTRK gene fusions in cancer. *Ann. Oncol. Off. J. Eur. Soc. Med. Oncol.* 30, viii5–viii15. doi:10.1093/annonc/mdz383
- Anandan, S., Gowtham, H. G., Shivakumara, C. S., Thampy, A., Singh, S. B., Murali, M., et al. (2022). Integrated approach for studying bioactive compounds from *Cladosporium* spp. against estrogen receptor alpha as breast cancer drug target. *Sci. Rep.* 12 (1), 22446. doi:10.1038/s41598-022-22038-x
- Bagal, S. K., Omoto, K., Blakemore, D. C., Bungay, P. J., Bilsland, J. G., Clarke, P. J., et al. (2019). Discovery of allosteric, potent, subtype selective, and peripherally restricted TrkA kinase inhibitors. *J. Med. Chem.* 62 (1), 247–265. doi:10.1021/acs.jmedchem.8b00280
- Chen, D., Oezguen, N., Urvil, P., Ferguson, C., Dann, S. M., and Savidge, T. C. (2016). Regulation of protein-ligand binding affinity by hydrogen bond pairing. *Sci. Adv.* 2 (3), e1501240. doi:10.1126/sciadv.1501240
- Doloff, N. G. (2015). Emerging therapeutic strategies for overcoming proteasome inhibitor resistance. *Adv. Cancer Res.* 127 (1), 191–226. doi:10.1016/bs.acr.2015.03.002
- Elbadwi, F. A., Khairy, E. A., Alsamami, F. O., Mahadi, M. A., Abdalrahman, S. E., Ahmed, Z. A. M., et al. (2021). Identification of novel transmembrane Protease Serine

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

Supervision, conceptualization, and software: AbA; methodology: RM, NA, HE, AE, RB, WS, KG, and AiA; writing—original draft preparation: RM, NA, HE, and EG; writing—review and editing: HG, MM, AiA, WO, AS, AhA, SI, GM, WS, and KG. All authors contributed to the article and approved the submitted version.

Funding

This study was supported via funding from Prince Sattam bin Abdulaziz University (project number (PSAU/2023/R/1444).

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.

Type 2 drug candidates for COVID-19 using computational studies. *Inf. Med. Unlocked* 26, 100725. doi:10.1016/j.imu.2021.100725

Eltai, L., and Alzain, A. A. (2022). Targeting the omicron variant of SARS-CoV-2 with phytochemicals from Saudi medicinal plants: Molecular docking combined with molecular dynamics investigations. *J. Biomol. Struct. Dyn.* 0 (0), 1–13. doi:10.1080/07391102.2022.2146203

Faulkner, S., Griffin, N., Rowe, C. W., Jobling, P., Lombard, J. M., Oliveira, S. M., et al. (2020). Nerve growth factor and its receptor tyrosine kinase TrkA are overexpressed in cervical squamous cell carcinoma. *FASEB bioAdvances* 2 (7), 398–408. doi:10.1096/fba.2020-00016

Faulkner, S., Jobling, P., Rowe, C. W., Rodrigues Oliveira, S., Roselli, S., Thorne, R. F., et al. (2018). Neurotrophin receptors TrkA, p75(NTR), and sortilin are increased and targetable in thyroid cancer. *Am. J. Pathol.* 188 (1), 229–241. doi:10.1016/j.ajpath.2017.09.008

Friesner, R. A., Banks, J. L., Murphy, R. B., Halgren, T. A., Klicic, J. J., Mainz, D. T., et al. (2004). Glide: A new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J. Med. Chem.* 47 (7), 1739–1749. doi:10.1021/jm0306430

Fu, L., Jin, W., Zhang, J., Zhu, L., Lu, J., Zhen, Y., et al. (2022). Repurposing non-oncology small-molecule drugs to improve cancer therapy: Current situation and future directions. *Acta Pharm. Sin. B* 12 (2), 532–557. doi:10.1016/j.apsb.2021.09.006

Furuya, N., Momose, T., Katsuno, K., Fushimi, N., Muranaka, H., Handa, C., et al. (2017). The juxtamembrane region of TrkA kinase is critical for inhibitor selectivity. *Bioorg. Med. Chem. Lett.* 27 (5), 1233–1236. doi:10.1016/j.bmcl.2017.01.056

Gao, F., Griffin, N., Faulkner, S., Rowe, C. W., Williams, L., Roselli, S., et al. (2018). The neurotrophic tyrosine kinase receptor TrkA and its ligand NGF are increased in squamous cell carcinomas of the lung. *Sci. Rep.* 8 (1), 8135. doi:10.1038/s41598-018-26408-2

Gazerani, P. (2019). Identification of novel analgesics through a drug repurposing strategy. *Pain Manag.* 9 (4), 399–415. doi:10.2217/pmt-2018-0091

Gowtham, H. G., Ahmed, F., Anandan, S., Shivakumara, C. S., Bilagi, A., Pradeep, S., et al. (2023). *In silico* computational studies of bioactive secondary metabolites from wedelia trilobata against anti-apoptotic B-cell lymphoma-2 (Bcl-2) protein associated with cancer cell survival and resistance. *Molecules* 28, 1588. doi:10.3390/molecules28041588

Gowtham, H. G., Murali, M., Singh, S. B., Shivamallu, C., Pradeep, S., Shivakumar, C. S., et al. (2022). Phytoconstituents of Withania somnifera unveiled Ashwagandhanolide as a potential drug targeting breast cancer: Investigations through computational, molecular docking and conceptual DFT studies. *PLoS One* 17, 0275432–e275524. doi:10.1371/journal.pone.0275432

Griffin, N., Marsland, M., Roselli, S., Oldmeadow, C., Attia, J., Walker, M. M., et al. (2020). The receptor tyrosine kinase trka is increased and targetable in HER2-positive breast cancer. *Biomolecules* 10 (9), 1329–1413. doi:10.3390/biom10091329

Guo, J., Xiang, S., Wang, J., Zhou, Y., Wang, Z., Zhang, Z., et al. (2022). Discovery of novel TrkA allosteric inhibitors: Structure-based virtual screening, biological evaluation and preliminary SAR studies. *Eur. J. Med. Chem.* 228, 114022. doi:10.1016/j.ejmech.2021.114022

Heinrich, T., Grädler, U., Böttcher, H., Blaukat, A., and Shutes, A. (2010). Allosteric IGF-1R inhibitors. *ACS Med. Chem. Lett.* 1 (5), 199–203. doi:10.1021/ml100044h

Isono, M., Sato, A., Asano, T., Okubo, K., and Asano, T. (2018). Delanzomib interacts with ritonavir synergistically to cause endoplasmic reticulum stress in renal cancer cells. *Anticancer Res.* 38 (6), 3493–3500. doi:10.21873/anticancer.12620

Issa, N. T., Stathias, V., Schürer, S., and Dakshanamurthy, S. (2021). Machine and deep learning approaches for cancer drug repurposing. *Semin. Cancer Biol.* 68, 132–142. doi:10.1016/j.semcancer.2019.12.011

Jordaan, M. A., Ebenezer, O., Damoyi, N., and Shapi, M. (2020). Virtual screening, molecular docking studies and DFT calculations of FDA approved compounds similar to the non-nucleoside reverse transcriptase inhibitor (NNRTI) efavirenz. *Heliyon* 6 (8), e04642. doi:10.1016/j.heliyon.2020.e04642

Klebe, G., and Böhm, H. J. (1997). Energetic and entropic factors determining binding affinity in protein-ligand complexes. *J. Recept. Signal Transduct. Res.* 17 (1–3), 459–473. doi:10.3109/10799899709036621

Kumar, A., Rath, E., and Kini, S. G. (2019). E-pharmacophore modelling, virtual screening, molecular dynamics simulations and *in-silico* ADME analysis for identification of potential E6 inhibitors against cervical cancer. *J. Mol. Struct.* 1189, 299–306. doi:10.1016/j.molstruc.2019.04.023

Lagade, C., Meignan, S., Adriaenssens, E., Foveau, B., Vanhecke, E., Romon, R., et al. (2009). TrkA overexpression enhances growth and metastasis of breast cancer cells. *Oncogene* 28 (18), 1960–1970. doi:10.1038/ncr.2009.61

Lyne, P. D., Lamb, M. L., and Saeh, J. C. (2006). Accurate prediction of the relative potencies of members of a series of kinase inhibitors using molecular docking and MM-GBSA scoring. *J. Med. Chem.* 49 (16), 4805–4808. doi:10.1021/jm060522a

Meng, X.-Y., Zhang, H.-X., Mezei, M., and Cui, M. (2011). Molecular docking: A powerful approach for structure-based drug discovery. *Curr. Comput. Aided. Drug Des.* 7 (2), 146–157. doi:10.2174/157340911795677602

Mohamed, L. M., Eltigani, M. M., Abdallah, M. H., Ghaboosh, H., Bin Jardan, Y. A., Yusuf, O., et al. (2022). Discovery of novel natural products as dual MNK/PIM inhibitors for acute myeloid leukemia treatment: Pharmacophore modeling, molecular docking, and molecular dynamics studies. *Front. Chem.* 10, 1–15. doi:10.3389/fchem.2022.975191

Mohammed, A. O., Abo-Idrees, M. I., Makki, A. A., Ibraheem, W., and Alzain, A. A. (2022). Drug repurposing against main protease and RNA-dependent RNA polymerase of SARS-CoV-2 using molecular docking, MM-GBSA calculations and molecular dynamics. *Struct. Chem.* 33 (5), 1553–1567. doi:10.1007/s11224-022-01999-9

Obubeid, F. O., Eltigani, M. M., Mukhtar, R. M., Ibrahim, R. A., Alzain, M. A., Elbadawi, F. A., et al. (2022). Dual targeting inhibitors for HIV-1 capsid and cyclophilin A: Molecular docking, molecular dynamics, and quantum mechanics. *Mol. Simul.* 48, 1476–1489. doi:10.1080/08927022.2022.2097673

Omer, S. E., Ibrahim, T. M., Krar, O. A., Ali, A. M., Makki, A. A., Ibraheem, W., et al. (2022). Drug repurposing for SARS-CoV-2 main protease: Molecular docking and molecular dynamics investigations. *Biochem. Biophys. Rep.* 29, 101225. doi:10.1016/j.bbrep.2022.101225

Pushpakom, S., Iorio, F., Eyers, P. A., Escott, K. J., Hopper, S., Wells, A., et al. (2018). Drug repurposing: Progress, challenges and recommendations. *Nat. Rev. Drug Discov.* 18 (1), 41–58. doi:10.1038/nrd.2018.168

Sasahira, T., Ueda, N., Yamamoto, K., Bhawal, U. K., Kurihara, M., Kirita, T., et al. (2013). Trks are novel oncogenes involved in the induction of neovascularization, tumor progression, and nodal metastasis in oral squamous cell carcinoma. *Clin. Exp. Metastasis* 30 (2), 165–176. doi:10.1007/s10585-012-9525-x

Simard, J. R., Klüter, S., Grütter, C., Getlik, M., Rabiller, M., Rode, H. B., et al. (2009). A new screening assay for allosteric inhibitors of cSrc. *Nat. Chem. Biol.* 5 (6), 394–396. doi:10.1038/nchembio.162

Staessen, J., Fagard, R., Fiocchi, R., Lijnen, P., M'Buyamba-Kabangu, J. R., and Amery, A. (1986). Chronic treatment with tibolone in essential hypertension. *Arch. Int. Pharmacodyn. Ther.* 279 (1), 162–176.

Staessen, J., Fagard, R., Grauwels, R., Lijnen, P., Verschueren, L. J., and Amery, A. (1983). Effects of tibolone, a new alpha-adrenoreceptor antagonist, in essential hypertension. *Clin. Pharmacol. Ther.* 33 (5), 556–564. doi:10.1038/CLPT.1983.76

Su, H. P., Rickert, K., Burlein, C., Narayan, K., Bukhtiyarova, M., Hurzy, D. M., et al. (2017). Structural characterization of nonactive site, TrkA-selective kinase inhibitors. *Proc. Natl. Acad. Sci. U. S. A.* 114 (3), E297–E306. doi:10.1073/pnas.1611577114

Subramanian, G., Johnson, P. D., Zachary, T., Roush, N., Zhu, Y., Bowen, S. J., et al. (2019). Deciphering the allosteric binding mechanism of the human tropomyosin receptor kinase A (hTrkA) inhibitors. *ACS Chem. Biol.* 14 (6), 1205–1216. doi:10.1021/acscmbio.9b00126

Wang, M., Liang, L., Lu, J., Yu, Y., Zhao, Y., Shi, Z., et al. (2019). Delanzomib, a novel proteasome inhibitor, sensitizes breast cancer cells to doxorubicin-induced apoptosis. *Thorac. Cancer* 10, 918–929. doi:10.1111/1759-7714.13030

Wang, T., Yu, D., and Lamb, M. L. (2009). Trk kinase inhibitors as new treatments for cancer and pain. *Expert Opin. Ther. Pat.* 19 (3), 305–319. doi:10.1517/13543770902721261

Wu, P., Nielsen, T. E., and Clausen, M. H. (2015). FDA-approved small-molecule kinase inhibitors. *Trends Pharmacol. Sci.* 36 (7), 422–439. doi:10.1016/j.tips.2015.04.005

Wu, X., Li, Q., Wan, S., and Zhang, J. (2021). Molecular dynamics simulation and free energy calculation studies of the binding mechanism of allosteric inhibitors with TrkA kinase. *J. Biomol. Struct. Dyn.* 39 (1), 202–208. doi:10.1080/07391102.2019.1708798

Yan, W., Lakkaniga, N. R., Carlomagno, F., Santoro, M., McDonald, N. Q., Lv, F., et al. (2019). Insights into current tropomyosin receptor kinase (TRK) inhibitors: Development and clinical application. *J. Med. Chem.* 62 (4), 1731–1760. doi:10.1021/acs.jmedchem.8b01092

Zangari, M., and Suva, L. J. (2016). The effects of proteasome inhibitors on bone remodeling in multiple myeloma. *Bone* 86, 131–138. doi:10.1016/j.bone.2016.02.019



OPEN ACCESS

EDITED BY

Khurshid Ahmad,
Yeungnam University, Republic of Korea

REVIEWED BY

Murali M.,
University of Mysore, India
Elancheran Ramakrishnan,
Annamalai University, India
Mohammad Rizki Fadhil Pratama,
Universitas Muhammadiyah
Palangkaraya, Indonesia

*CORRESPONDENCE

Nusrat Shafiq,
✉ dr.nusratshafiq@gcwuf.edu.pk,
✉ gqumarin@gmail.com
Mohammed Bourhia,
✉ bourhiamohammed@gmail.com

RECEIVED 31 March 2023

ACCEPTED 09 June 2023

PUBLISHED 27 June 2023

CITATION

Rashid M, Maqbool A, Shafiq N,
Bin Jordan YA, Parveen S, Bourhia M,
Nafidi H-A and Khan RA (2023), The
combination of multi-approach studies
to explore the potential therapeutic
mechanisms of imidazole derivatives as
an MCF-7 inhibitor in
therapeutic strategies.
Front. Chem. 11:1197665.
doi: 10.3389/fchem.2023.1197665

COPYRIGHT

© 2023 Rashid, Maqbool, Shafiq, Bin
Jordan, Parveen, Bourhia, Nafidi and
Khan. 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.

The combination of multi-approach studies to explore the potential therapeutic mechanisms of imidazole derivatives as an MCF-7 inhibitor in therapeutic strategies

Maryam Rashid¹, Ayesha Maqbool¹, Nusrat Shafiq ^{1*},
Yousef A. Bin Jordan², Shagufta Parveen^{1,3},
Mohammed Bourhia^{4*}, Hiba-Allah Nafidi⁵ and
Rashid Ahmed Khan⁶

¹Synthetic and Natural Product Drug Discovery Laboratory, Department of Chemistry, Government College Women University Faisalabad, Faisalabad, Pakistan, ²Department of Pharmaceutics, College of Pharmacy, King Saud University, Riyadh, Saudi Arabia, ³Department of Applied Chemistry, Beijing Institute of Technology, Beijing, China, ⁴Department of Chemistry and Biochemistry, Faculty of Medicine and Pharmacy, Ibn Zohr University, Laayoune, Morocco, ⁵Department of Food Science, Faculty of Agricultural and Food Sciences, Laval University, Quebec City, QC, Canada, ⁶Nuclear Institute for Agriculture and Biology (NIAB), Faisalabad, Pakistan

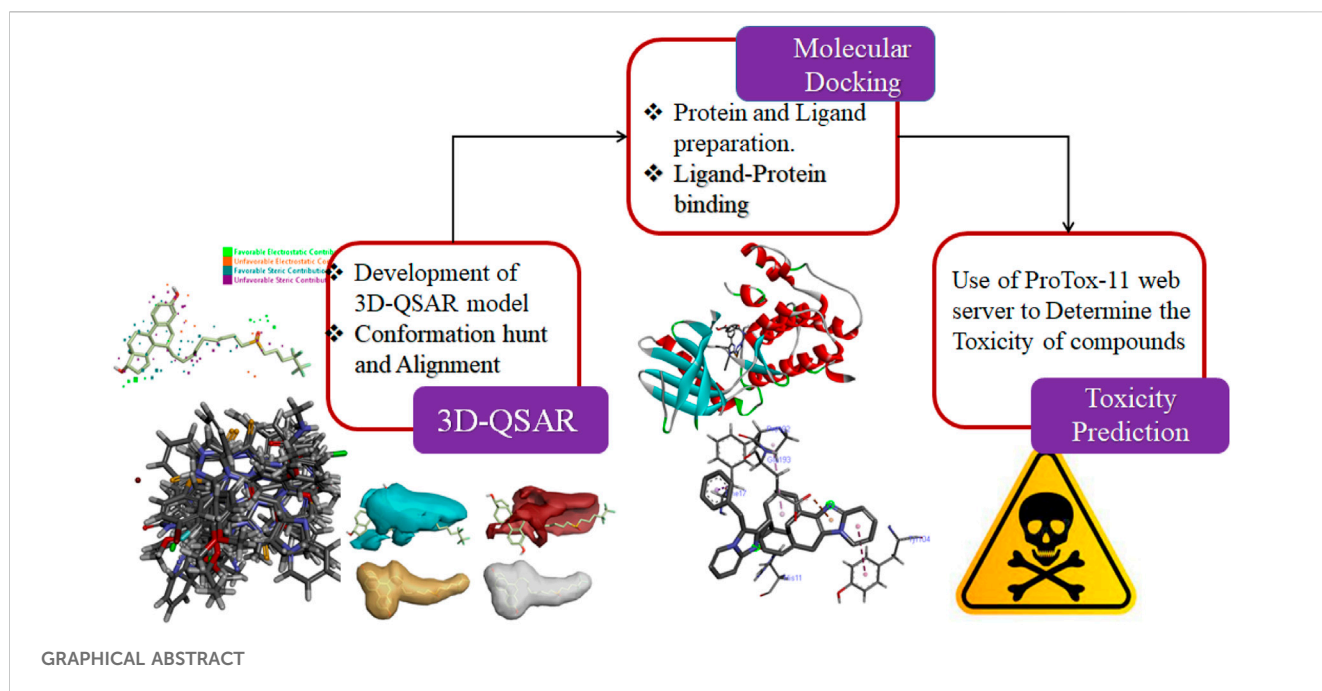
Breast cancer covers a large area of research because of its prevalence and high frequency all over the world. This study is based on drug discovery against breast cancer from a series of imidazole derivatives. A 3D-QSAR and activity atlas model was developed by exploring the dataset computationally, using the machine learning process of Flare. The dataset of compounds was divided into active and inactive compounds according to their biological and structural similarity with the reference drug. The obtained PLS regression model provided an acceptable $r^2 = 0.81$ and $q^2 = 0.51$. Protein-ligand interactions of active molecules were shown by molecular docking against six potential targets, namely, TTK, HER2, GR, NUDT5, MTHFS, and NQO2. Then, toxicity risk parameters were evaluated for hit compounds. Finally, after all these screening processes, compound **C10** was recognized as the best-hit compound. This study identified a new inhibitor C10 against cancer and provided evidence-based knowledge to discover more analogs.

KEYWORDS

breast cancer, drug discovery, imidazole, virtual screening, MCF-7

Introduction

Cancer is the abnormal and uncontrolled growth of cells that is caused by the mutation of genes. This mutation may lead to an accelerated rate of cell division, so it is the major cause of death worldwide (Ali et al., 2017). A frequently occurring cancer in women is breast cancer and approximately 1 million women are affected by it every year. Obesity, consumption of alcohol, genetics, aging, menopause, diabetes mellitus (type 2), high estrogen levels, radiation exposure,



smoking, menarche, sex, and physical activity are the major risk factors responsible for causing breast cancer (Ataollahi et al., 2015; Escala-Garcia et al., 2020; Anandan et al., 2022).

The genetic mutation causes the development and progression of breast tumors. Anomalous amplification and mutation of genes cause the initiation of tumors such as a mutation in the Breast Cancer gene (BRCA1/2), RB Transcriptional Corepressor 1 (RB1), Human epidermal growth factor receptor 2 (HER2), Fragile Histidine Triad Diadenosine Triphosphatase (FHIT), tumor protein P53, Epidermal Growth Factor Receptor (EGFR), extracellular signal-regulated kinase (ERK), Mitogen-activated protein kinase (MEK), and Rat sarcoma (Ras) genes that can lead to breast cancer (Figure 1) (Dickson, 1990; Sun et al., 2017; Lakshmithendral et al., 2019a).

Endogenous estrogen and exogenous estrogen can both cause breast cancer. Hormone replacement therapy (HRT; the process in which endogenous estrogen is administered in menopausal females) also increases the risk of breast cancer (Sun et al., 2017). Androgen hormone is converted to estrogen through an enzyme complex, aromatase. Aromatase is detected in the stromal cell component of the breast; it is also located in the breast epithelial cells. Studies have shown that the level of aromatase was higher in breast tumor cells than in normal cells (Brueggemeier et al., 2003; Pasqualini and Chetrite, 2005). Leptin is another hormone involved in breast cancer; its overexpression causes an increase in cell proliferation and thus leads to breast cancer (Jardé et al., 2011).

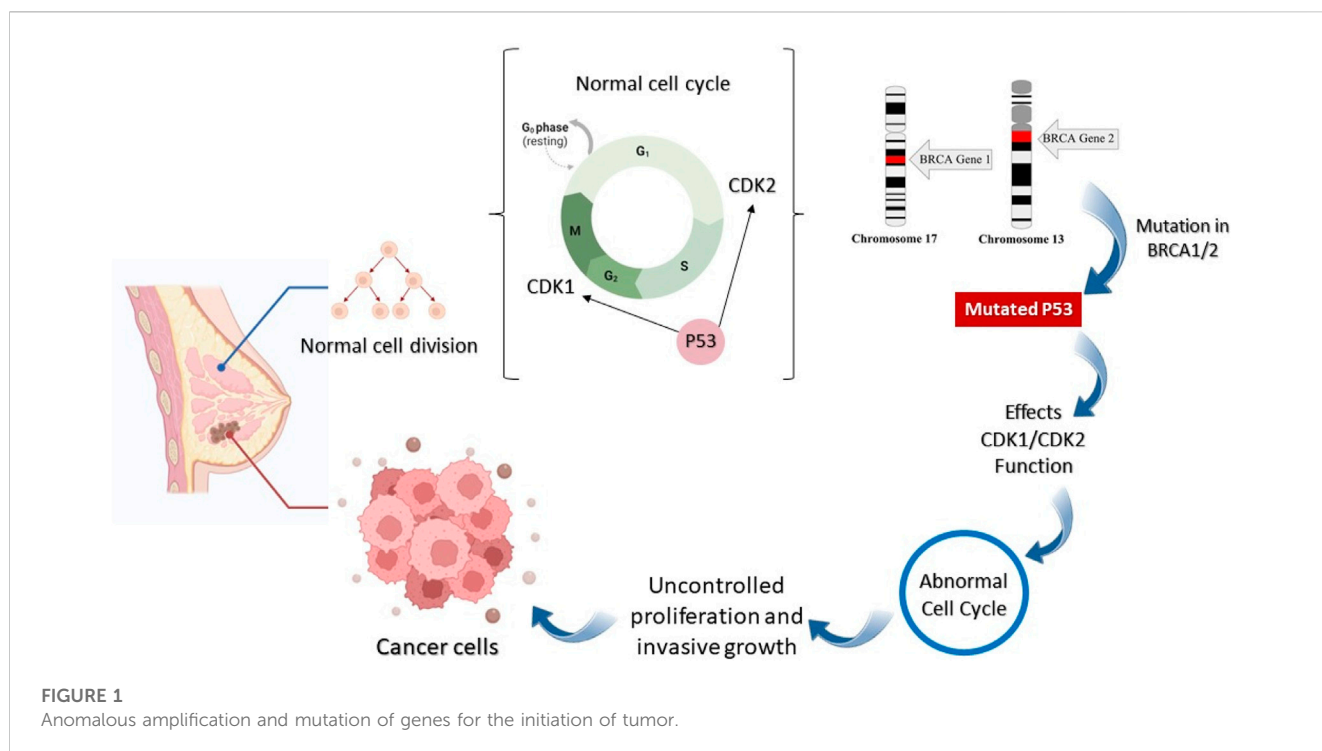
Breast cancer is a serious problem that needs to be solved. For this purpose, considerable advances have been made in breast cancer treatment (Howell et al., 2014). Till now, many drugs have been synthesized to cure this deadly disease. Breast cancer-targeted medication utilizes molecules or drugs that suppress breast cancer cell growth in various ways (Maruthanila et al., 2017). Targeted drugs either kill the cancer cells or retard their growth. For example, the expression of abnormal genes such as HER2 (which

stimulates breast cancer cell growth) can be blocked by using this medication (Masoud and Pagès, 2017; Lakshmithendral et al., 2019b).

The most commonly targeted breast cancer cell line is MCF-7 because it has been proven to be the most suitable cell line for the investigation of breast cancer all over the world (Lee et al., 2015). MCF-7 cells are universally used for experiments on ER (estrogen receptor) positive breast cancer cells. They are cultured easily, and they maintain their ER expression during treatment with a targeted drug. For this reason, they are highly suitable for anti-hormone therapy resistance studies. MCF-7 cells are very well distinguished and an excellent experience of this cell line permits researchers to utilize these cells to bring more insights into the treatment of breast cancer through viable *in vitro* assays (Comşa et al., 2015).

The use of machine learning has created a revolutionary impact on chemical sciences by quickening the use of computational chemistry methods (Keith et al., 2021). Computer-aided drug designing aims at the discovery and analysis of suitable medications and biologically active compounds by computational approaches. In structure-based drug designing (SBDD), 3D structural information of proteins is utilized to design new drugs by identifying the sites and their interactions that are useful for the biological activity of ligands. In ligand-based drug designing (LBDD), ligand information is utilized to set up an interrelation between their physiochemical characteristics and biological activities. This information is useful for designing new drugs and for the optimization of already known drugs to enhance their activity.

Drug discovery is a costly procedure and time-consuming process; therefore, we have employed computational processes for drug discovery. The advancements in computational methods and high-throughput virtual screening have developed a remarkable pharmaceutical approach that does not only reduce the time



phase but also introduces highly efficient drugs, having efficient biological activity and minimum side effects for a specific disease (Lakshmithendral et al., 2019b).

Imidazole is the core of FDA-approved drugs with acceptable activities in practice. Several compounds having imidazole core have been utilized for their medicinal uses in clinical trials for several diseases. There is an increasing trend towards imidazole-based medicinal chemistry which has added promising and potential therapeutic values of imidazole-derived compounds for treating incurable diseases. The compounds with imidazole scaffold provide electronic-rich characteristics responsible for binding with a variety of enzymes, proteins, and receptors compared to the other heterocyclic rings. In this study, the role of imidazole drugs as anti-breast cancer agents have been discussed using the computational approach (Chopra and Sahu, 2019). Heterocyclic compounds are very well-known molecules in organic chemistry because they show remarkable medicinal properties as well as anticancer properties (Ali et al., 2017). Imidazoles are very important heterocyclic compounds that are widely utilized all over the world for drug discovery processes and are compounds of interest for researchers for centuries (Gaba and Mohan, 2016). Previous studies have proved the vital role of imidazole and its derivatives in medicinal chemistry because of their efficient uses as anti-coagulant, anti-cancer, anti-parasitic, anti-helminthic, anti-fungal, antimicrobial, anti-inflammatory, antibacterial, anti-viral, anti-diabetic, anti-malarial, antihypertensive, and anti-tubercular drugs (Abbasov et al., 2012; Verma et al., 2013; Mumtaz et al., 2016; Ali et al., 2017). Some FDA-approved anticancer imidazole derivatives are shown in Figure 2.

In spite of extensive studies on imidazole derivatives and their *in vitro* potential activity, their *in-vivo* and, specifically, *in*

silico activity of breast cancer has not been carried out. The computer-based methodologies, i.e., *in silico* approaches are powerful tools for the recognition of synthetic imidazole compounds and their potential to inhibit breast cancer. Using these approaches, the new drug candidates can be evaluated in a faster way, reducing costs and accelerating drug discovery (Gowtham et al., 2023). This study aimed to investigate the *in silico* anticancer activities of imidazole derivatives (Rizzo et al., 2014).

In this study, maximum tools used for structure- and ligand-based drug designing have been used, and key regularity features governing the toxicity and anticancer activity of imidazole derivatives have been studied. By discovering and characterizing potential imidazole derivatives as anti-breast cancer agents, this research will contribute to the growing repertoire of drug candidates, expanding the possibilities for future therapeutic interventions. It has the potential to revolutionize the therapeutic landscape by providing new and effective options for breast cancer patients, deepening our understanding of the disease, and inspiring further advancements in the field. It will provide some more valuable insights into the virtual screening and drug designing process and will demonstrate the drug designing process which will lead to drug discovery containing a pharmacophore against breast cancer which is a harmful disease affecting millions of lives all over the world (Alam and Khan, 2017).

Materials and methods

In this study, the 3D-QSAR modeling has been accomplished using the Forge V6.0 software. A total of 84 compounds

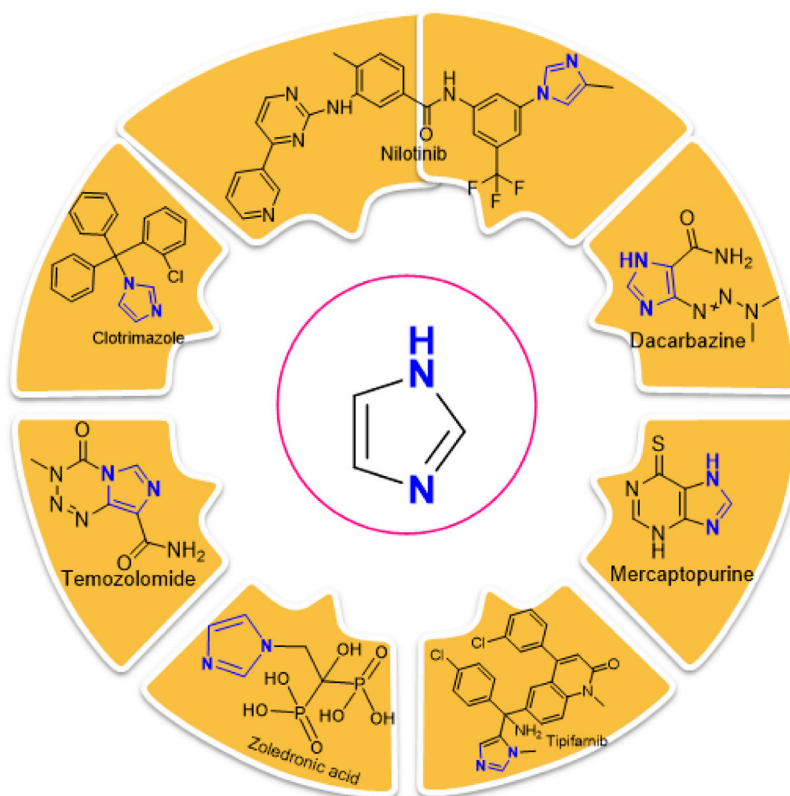


FIGURE 2
FDA-Approved imidazole derivatives.

(Supplementary Table S1) with reported anti-breast cancer activity were used while developing the model. An FDA-approved drug Fulvestrant which is a steroidal anti-estrogen used to treat hormone-receptor-positive metastatic breast cancer is used as a reference compound.

The development of the 3D-QSAR model

Data collection and structure preparation

The dataset of imidazole compounds was collected from prior reports/literature. Their structures were drawn in Chem Draw professional (Perkin Elmer), and 2-Dimensional structures were converted into 3-Dimensional structures using Chem3D Ultra (Version 19.1.0.8, Perkin Elmer). The value of enzyme inhibition (experimental activity) was expressed as (IC_{50}) for the training dataset which was then altered to its positive logarithmic scale using the formula: $pIC_{50} = -\log(IC_{50})$ and defined as a dependent variable. The database of compounds was generated in Microsoft Excel as a CSV output file (comma delimited).

Conformation hunt and pharmacophore generation

To demonstrate a hypothesis for 3D conformation, the Field Template module of Forge V6.0 software was used as no structural data was attainable for imidazole derivatives in their target-bound

state. For this purpose, the information about field and shape was utilized by the template from the library of 84 compounds. The hypothesis was developed by generating the three-dimensional field point pattern and calculating the field points of bioactive conformation.

Compound alignment and the development of the 3D-QSAR model

At the connexion point of a 3D grid, the 3D-QSAR method calculates various molecular properties as molecular descriptors. This methodology covers the complete data of aligned training set compounds. The pharmacophore template was transported into the Forge V6.0 software, followed by the alignment of compounds with the associated template. After the alignment of 84 compounds with known IC_{50} values, the 3D-QSAR model was built using the Field point-based descriptors. While building the model, the maximum distance of sample points was set to 1.0 Å, the maximal number of components was set to 20, Y scrambles were adjusted to 50, and volumetric as well as electrostatic fields were also used. For overall resemblance, 50% dice volume similarity and 50% field similarity were achieved using the Forge software. The experimental activity (IC_{50}) of compounds was changed to pIC_{50} which is equal to the negative log of IC_{50} . The set of 84 compounds was divided into the training set and test set with a ratio of 80% and 20%, respectively, and one compound was selected as the reference drug to assess the QSAR modeling using the activity stratified method.

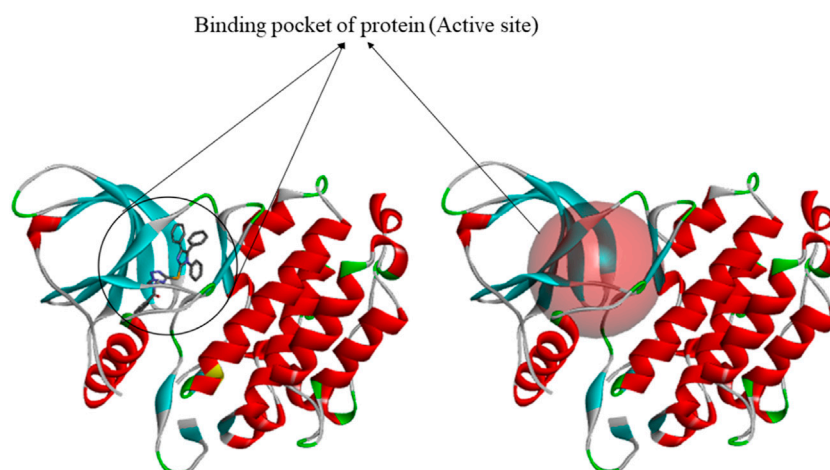


FIGURE 3
Generation of binding site by using discovery studio.

QSAR model validation

The model was verified by q^2 (cross regression coefficient), r^2 (regression coefficient), and similarity score of conformers for every ligand. LOO technique (leave-one-out) was used to assess the derived 3D-QSAR model. The LOO cross-validation technique is thought to be one of the most efficient techniques for the validation of the regression model having a small training dataset. The data size of $N-1$ was used for training and the remaining one was tested; N identifies the complete dataset. In the LOOCV technique, the process of testing and training was repeated for the N number of times, and in this way, each data was passed through the testing method. Then, the test data which is not in the training set is used to derive the 3D-QSAR model.

SAR activity-atlas models visualization

The global aspect of training data was studied quantitatively by using the Bayesian approach. The hydrophobicity, electrostatics, and shape attributes, which lie beneath the SAR of a particular set of compounds, are better understood by this approach. These 3D models were viewed to achieve valuable information. The three types of interconnected biochemical evaluated data including regions explored analysis, activity cliff summary, and average of actives were revealed by the Activity-atlas study. The regions explored analysis exhibited the areas of aligned and fully explored compounds. The details about negative and positive electrostatic sites, appropriate and inappropriate hydrophobicity, and appropriate shape of actives were provided by the activity cliff summary. On the other hand, an average of actives helped in showing the common parts in active compounds which were selected.

Target prediction analysis by molecular docking

The preparation of protein

The 3D structures of target proteins (PDB ID: 7CHM, 3PP0, 4UDD, 5NWH, 3HY3, and 4ZVM) (Supplementary Table S2) were

downloaded from the RCSB PDB database (<https://www.rcsb.org/>). The protein preparation was performed to accomplish various tasks such as identifying the active site, deleting alternate conformations, interpolating missing atoms in incomplete residues, protonating titratable residues, modeling the missing loop areas, and removing the water molecules and heteroatoms (Alam and Khan, 2019). The ligands of proteins were used to identify the active sites from the “Define and Edit Binding Site” option in Discovery Studio, and SBD_Site_Sphere was generated (Figure 3).

The preparation of active ligands

Active ligands along with reference Fulvestrant (standard drug) having known inhibitory potentials were collected from literature, and SDF files of some of the compounds were downloaded from PubChem while others were generated from the Chem3D software where they were optimized through MM2 and MMF9 force field. For ligand preparation, the “Open Babel” software was used. For this purpose, the input file was imported in the form of sdf-MDL MOL format and the output path was selected in pdb format. Then the ligand was converted into the desired form.

Protein-ligand docking studies

PDBQT files of proteins and ligands were prepared, and by using the Graphical User Interface program AutoDock Tools, grid box creation was accomplished. Fragmental volumes to the protein, polar hydrogens, united atom Kollman charges, and solvation parameters were assigned by AutoDock tools. After preparation, all files were saved as PDBQT. Preparation of the grid map was done by using a grid box via AutoGrid. The grid size was selected as $60 \times 60 \times 60$ XYZ points with a grid spacing of 0.375\AA . Default settings were used for all other parameters. Autodock was used for docking protocol and information about proteins and ligands was used along with grid box features in the configuration file. Both ligands and proteins are considered rigid when using Autodock. The results lower than

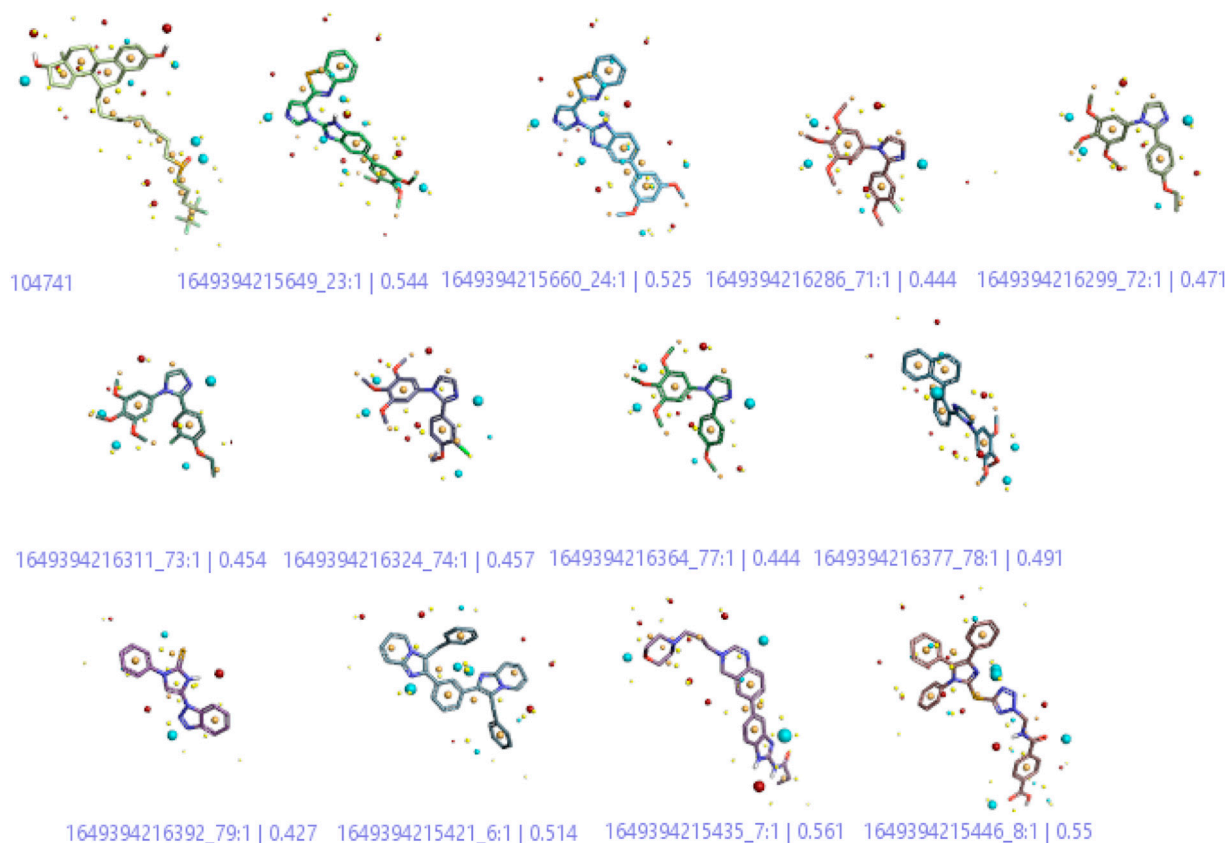


FIGURE 4

3-Dimensional Field points for QSAR model development. The red colour indicates positive electrostatic potential while negative electrostatic potential is represented by the blue colour. The orange color shows hydrophobicity and the yellow color denotes van der Waals descriptors localization.

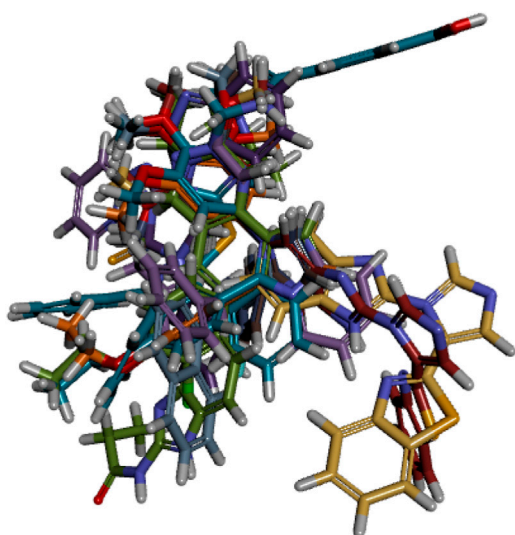


FIGURE 5

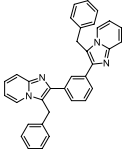
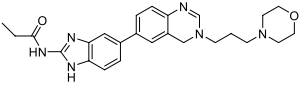
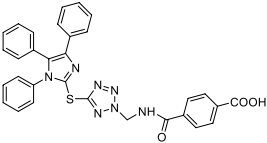
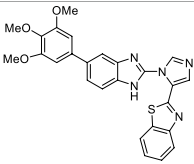
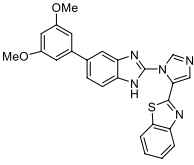
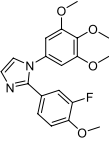
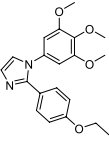
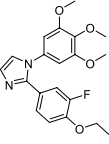
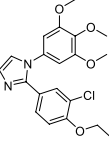
Conformational alignment of active compounds.

1.0 Å in the root-mean-square deviation were assembled and were depicted by the result with the most suitable free-binding energy while the results with the lowest binding affinity or binding energy were extracted and subjected to further analysis (Azam and Abbasi, 2013).

Molecular dynamics simulation

MD simulation was accomplished via the iMOD server (<https://imods.iqfr.csic.es/>) to assess the physical movement and stability of protein-ligand complexes (Sumera et al., 2022). The structural dynamics of the protein-ligand complexes were analyzed using iMODS and the molecular motion was also determined. The iMOD server employs Normal mode analysis (NMA) to calculate the internal coordinates of protein to evaluate its stability. In this study, the conformational fluctuations of docked complexes were demonstrated and their slow dynamics were investigated using NMA (Kirar et al., 2022).

TABLE 1 Active compounds obtained after the 3D-QSAR model development.

Sr No	Compounds	Structure	IC ₅₀ (μM)	PIC ₅₀	References
1	C6		0.30	6.5228	Meenakshisundaram et al. (2019)
2	C7		0.38 ± 0.08	6.4202	Fan et al. (2020)
3	C10		0.38 ± 0.04	6.4202	Al-Blewi et al. (2021)
4	C21		0.018 ± 0.0039	7.7447	Edukondalu et al. (2021)
5	C22		0.10 ± 0.028	7	Edukondalu et al. (2021)
6	C68		0.074 ± 0.017	7.1307	Romagnoli et al. (2016)
7	C69		0.0015	8.8239	Romagnoli et al. (2016)
8	C70		0.0034	8.4685	Romagnoli et al. (2016)
9	C71		0.0007	9.1549	Romagnoli et al. (2016)

(Continued on following page)

TABLE 1 (Continued) Active compounds obtained after the 3D-QSAR model development.

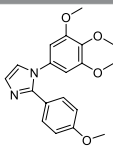
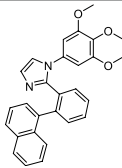
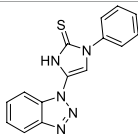
Sr No	Compounds	Structure	IC ₅₀ (μM)	PIC ₅₀	References
10	C74		0.23	6.6382	Romagnoli et al. (2016)
11	C75		0.0046	8.3372	Romagnoli et al. (2016)
12	C76		6.4 ± 0.18	8.1938	Khayyat et al. (2021)

TABLE 2 Binding energy values in kcal/mol.

Compounds	Proteins PDB IDs					
	4ZVM	5NWH	7CHM	3HY3	4UUD	3PP0
C6	−10.3	−8.7	−9.4	−8.1	−8.8	−7.6
C7	−7.2	−6.2	−8.6	−8.2	−6.4	−10.0
C10	−7.9	−7.8	−10.5	−9.0	−8.0	−8.1
C21	−8.3	−7.8	−9.2	−9.7	−9.1	−7.2
C22	−8.9	−7.5	−9.2	−8.7	−8.8	−9.0
C68	−6.6	−5.9	−8.3	−6.6	−5.4	−8.3
C69	−6.2	−6.1	−7.6	−6.7	−5.1	−7.6
C70	−6.6	−5.8	−8.1	−6.9	−5.2	−8.1
C71	−6.1	−5.9	−7.5	−6.1	−5.1	−7.5
C74	−6.1	−6.1	−7.6	−6.4	−5.3	−7.6
C75	−6.2	−6.6	−8.3	−6.1	−6.5	−8.3
C76	−7.5	−6.9	−8.9	−8.5	−6.2	−10.3
References	−6.8	−6.2	−7.9	−8.2	−6.1	−6.6

Toxicity prediction

The toxicity of the compounds was determined using ProTox-11 (Banerjee et al., 2018). With the help of this tool, the toxicity of compounds can be freely estimated by inserting the name of the compound or by simply writing its canonical smiles. The 2-Dimensional structure of the compound is used as input for this webserver. ProTox-11 is distributed in different classes depending on the toxicity such as Organ toxicity,

immunotoxicity, carcinogenicity, cytotoxicity, and mutagenicity.

Geometry optimization and reactivity determination

The DFT calculations were performed using Gauss view 06 and Gaussian. The 2D structure of the molecule was drawn by using

TABLE 3 Toxicity risk parameters.

Compounds	Hepatotoxicity	Carcinogenicity	Immunotoxicity	Mutagenicity	Cytotoxicity
C6	−0.60	−0.61	−0.99	+0.54	−0.86
C10	−0.54	−0.51	−0.99	−0.55	−0.71
C21	+0.56	+0.5	+0.85	+0.66	−0.56
C76	+0.62	+0.55	−0.99	+0.62	−0.86

Perkin-Elmer ChemDraw and then converted into a 3D structure through the Chem3D software. Geometrical optimization was done with B3LYP (an exchange-correlation function) and basic sets of 6-311G (Türker et al., 2010). The reactivity and stability of the compound were determined by calculating the energy gap between the HOMO-LUMO orbitals.

Results and discussion

3D-QSAR modeling on imidazole derivatives

Conformation hunt and pharmacophore generation

A three-dimensional structure-activity relationship (3D-QSAR) was performed to throw more light on a series of imidazoles. For this purpose, a conformational hunt was carried out on these compounds (C1–C84). (Alam and Khan, 2017). The three-dimensional pattern of field points (Figure 4) was identified by illustrating the derived conception of bioactive conformation with its calculated field points. Four distinct molecular fields were calculated, namely, negative and positive electrostatic potential, hydrophobicity, and shape/van der Waal descriptors. To draw a pharmacophore template resembling the bioactive conformation (for further virtual screening), a molecular field-based similarity approach was employed.

Figure 4 shows that the compounds with similar field points bind at the same target site. This characteristic provides a linear correlation between biological activity and structural similarity of ligands (Low et al., 2005). Based on the parameter, Figure 4 shows the active ligands, and their similarity metric was found within the range of 56% to 42%.

Alignment and development of the 3D-QSAR model

The ligand alignment in the protein context is required to use the 3D similarity metric for activity atlas model development. To ensure accurate model development, this alignment must be inspected. The compounds in the training set were aligned to ensure that the molecules being compared were in the same relative orientation (Figure 5). This alignment is necessary because molecules can adopt different conformations or spatial arrangements due to the freedom of rotation around single bonds. After it, the 3D-QSAR model was built by using the Field points-based descriptors. The activity interactive graph plot was used to represent the fitness of the derived 3D-QSAR model. This graph displays the comparison between predicted and actual activity

with cross-validation data points. Fairly good activity-descriptors' relationship accuracy of 81% was achieved by the derived 3D-QSAR model as the regression coefficient was $r^2 = 0.81$. Similarly, as mentioned by the cross-validation regression coefficient ($q^2 = 0.51$), a high activity-prediction accuracy of 51% was attained. The derived 3D-QSAR model was proved to be very reliable to predict the anticancer and cytotoxic activity of imidazole derivatives as an MCF7 cell-line inhibitor (Table 1).

Figure 5 shows that there were displayed little spaces in ligands to accommodate small conformation changes and variations of moieties present in aligned molecules. High molecules have tight alignment to restrict the substitution or replacement of any group present in them. Low-active molecules, in comparison, lack steric tightness and have the capacity to substitute any moiety in the context of an activity enhancer (Low et al., 2005).

The regulation of the SAR mechanism of imidazole derivatives by field points

The identification of field points (coefficient and variance) governing the anticancer activity

The QSAR model was also viewed in a 3D form to unveil the structure-activity relationship (SAR) mechanism of imidazole derivatives. The field points named coefficient and variance (associated with the bioactivity of training set compounds) were analyzed in a 3D structural form for the purpose. The derived model points for QSAR were contrasted with the reference compound for better comprehension of space field point localization (Alam and Khan, 2014). In a robust model, the high coefficient and variance field points were proved to be the highly essential correlating parameters. According to the results, electrostatic and steric coefficients both play a major role in modulating the anticancer activity as represented by the large size of red, cyan, green, and pink field points (Figure 6). Field points containing high steric and electrostatic variance indicated regions of high changes while the field points containing low variance represented the regions with less changes or no changes (Figure 6).

Field contribution in activity prediction

"View field contributions to predicted activity" study was done on imidazole derivatives. This evaluates the extent to which imidazole derivatives fit the derived field-based 3D-QSAR model and regions of structural field points governing the predicted activity. These field contributions (Figures 6A, B) were represented by purple, blue, and red color regions. According to the results, the orange- and purple-colored areas denote the region

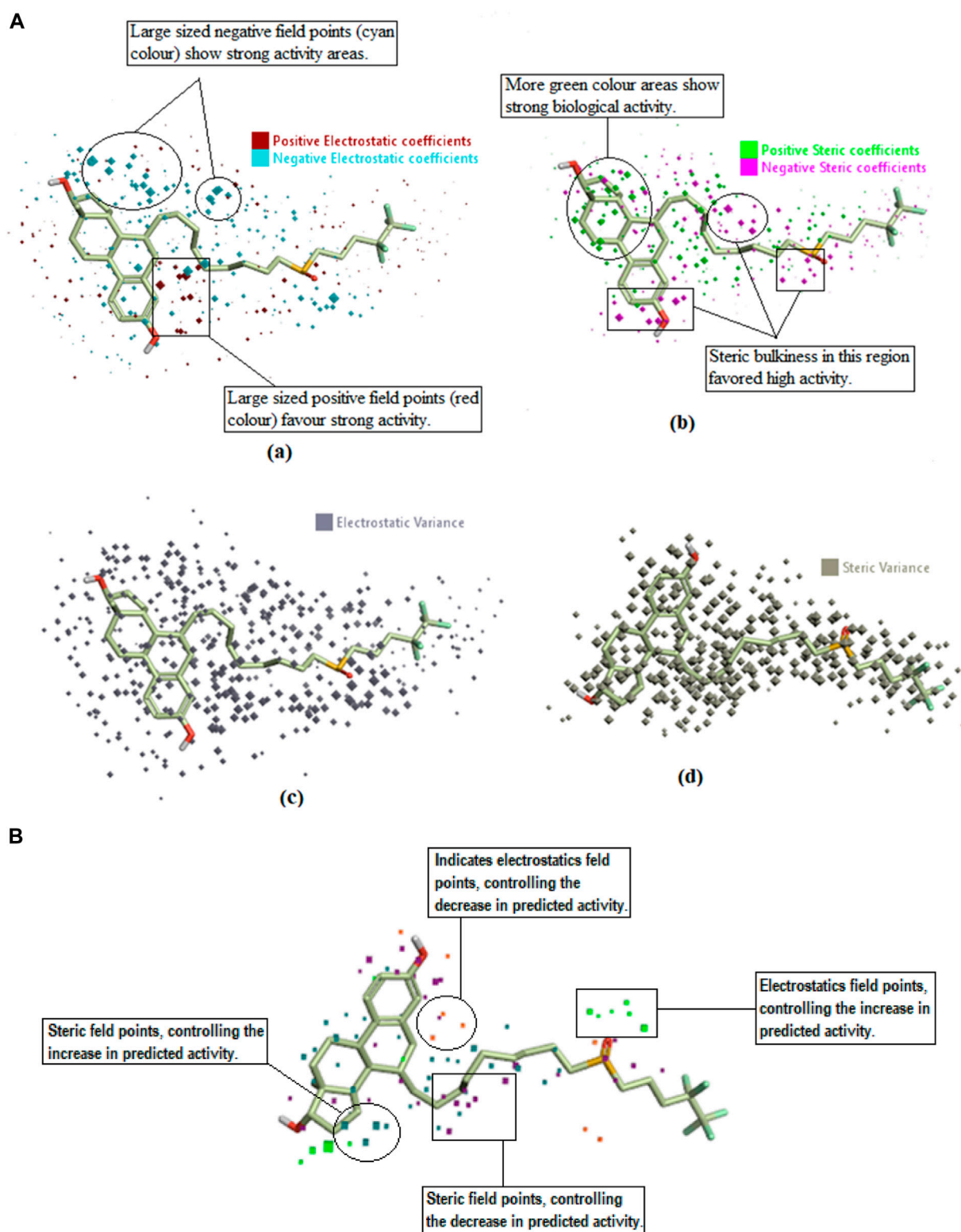


FIGURE 6

(A) Coefficients and variance field points of Fulvestrant; (a) Electrostatic coefficient; (b) Electrostatic variance; (c) Steric coefficients; (d) Steric variance; (e) Field contributions to the predicted activity. (B) Field contributions to the predicted activity.

of electrostatics and steric field points, respectively, having the negative regulation capability on predicted activity (decrease anticancer activity). Whereas the green- and zinc-colored areas

denote the regions of electrostatics and steric field points, respectively, with a positive regulation capability on the predicted activity (increased anticancer activity).

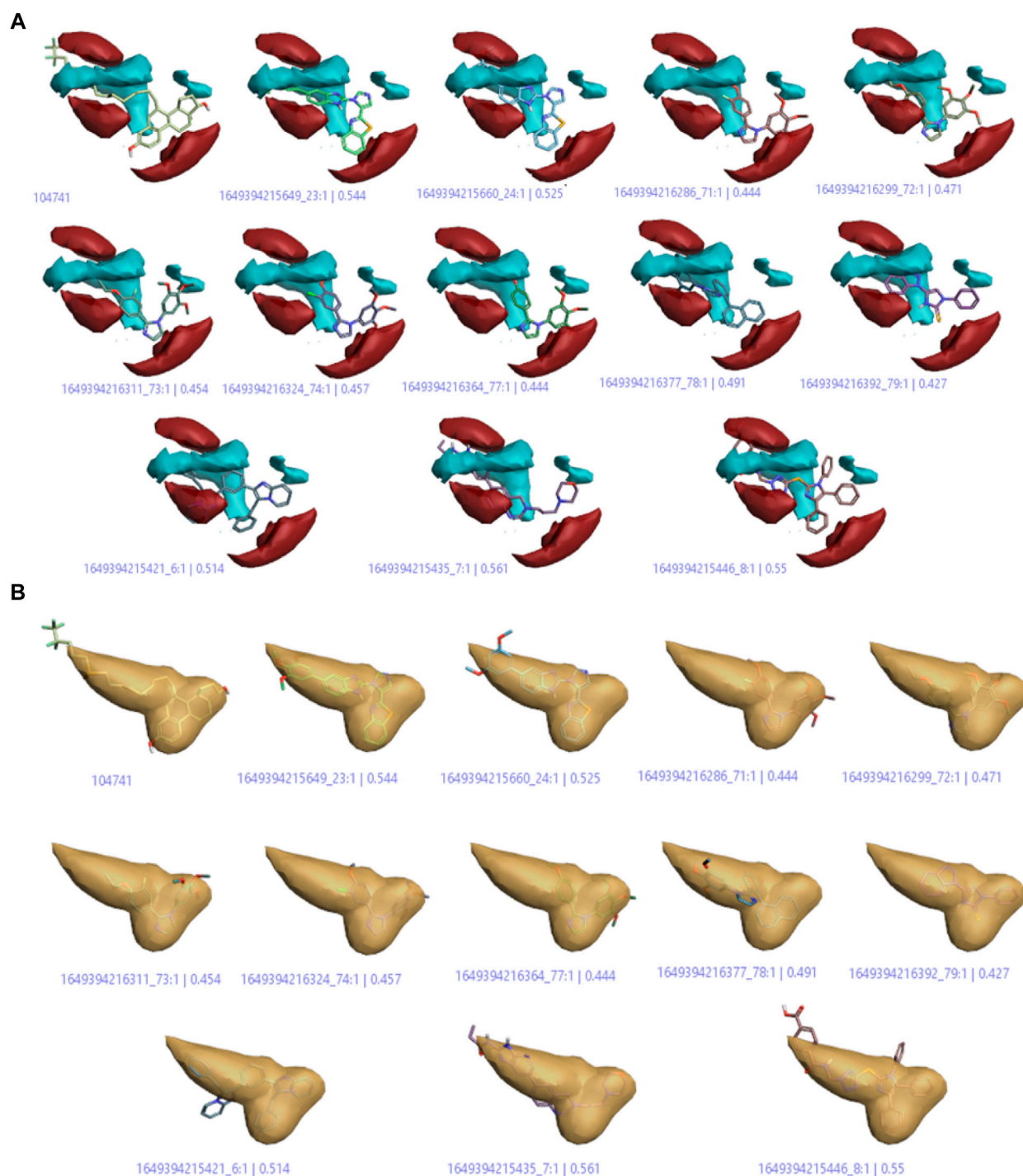


FIGURE 7

(A) Positive and negative electrostatic regions (B) The hydrophobic interaction regions of active compounds.

Activity-atlas visualization for SAR mechanism identification

SAR study was practiced through the activity-atlas visualization technique and was used to unveil the key features of imidazole, regulating the anticancer activity and designing more novel drugs. For this purpose, an activity cliffs summary

and an average of actives study were performed on imidazole derivatives.

Average of actives model

On the basis of this model, imidazole compounds having a pIC_{50} value higher than 6.4 were classified as active compounds while the

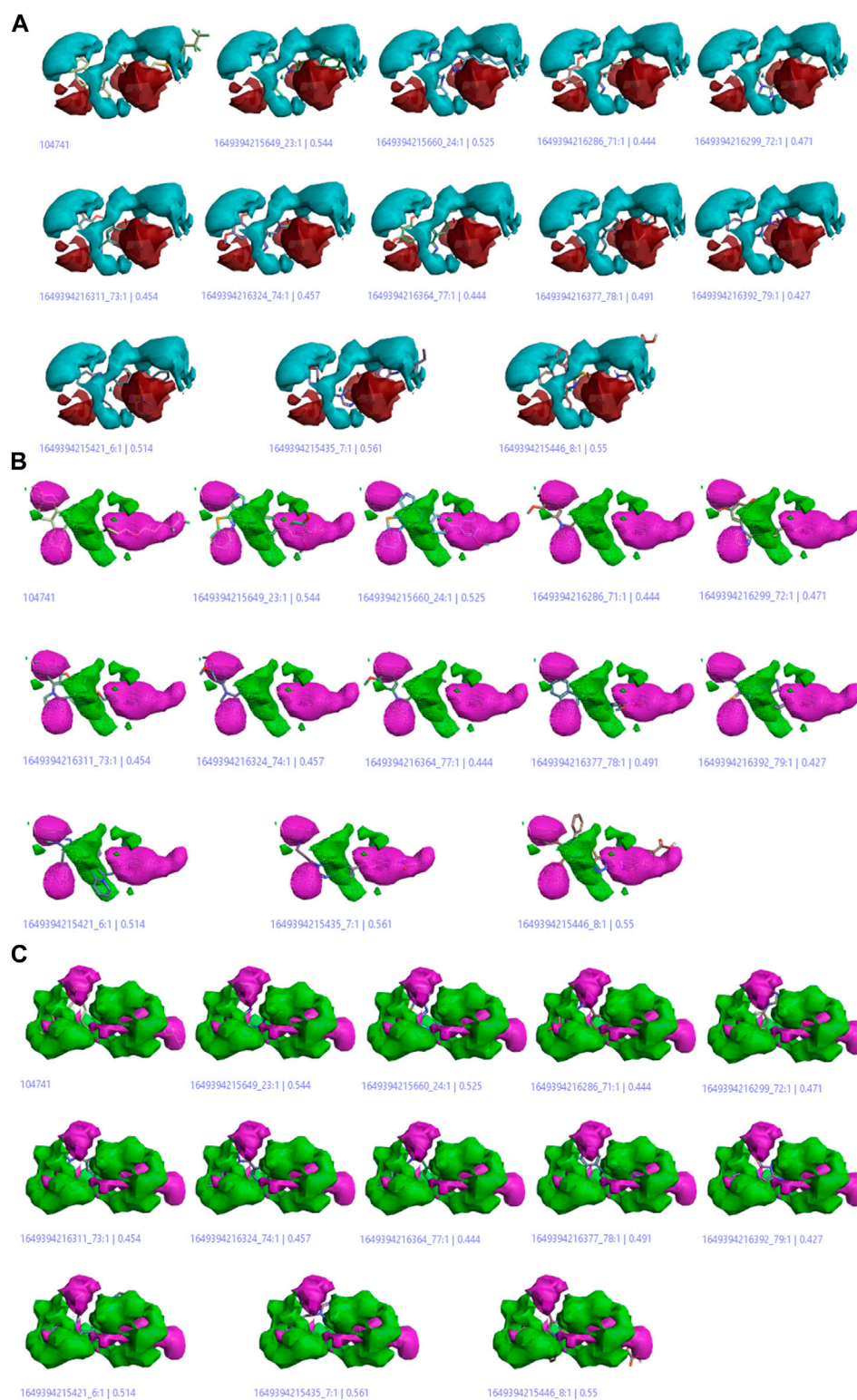


FIGURE 8

(A) Cliff summary of electrostatics. (B) The areas of favorable and unfavorable hydrophobics. (C) Favourable shape region and unfavorable shape region.

rest of the compounds were considered inactive. This model (Figure 7; Fig. S1) represents the areas of high activity that reference drugs and active ligands have in common.

In Figure 7A, the positive and negative electrostatic regions represented by the red color sites and cyan color sites, respectively, correlate with the anticancer activity, i.e., more of

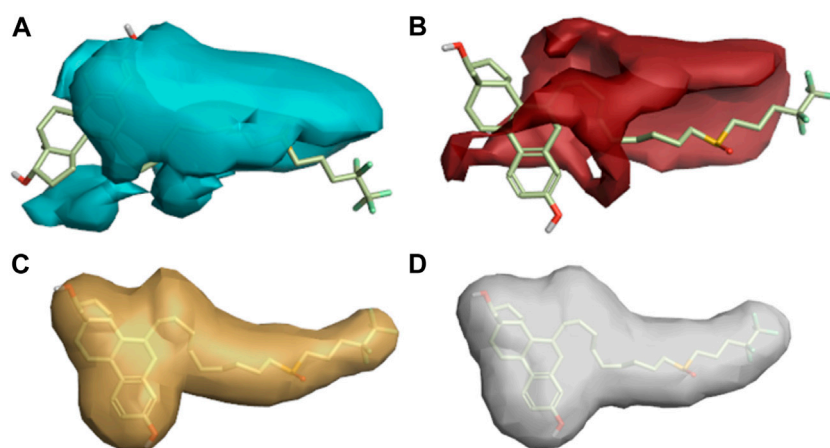


FIGURE 9

Regions explored for Fulvestrant by active molecules in the activity atlas model. (A) Regions explored in negative electrostatics, (B) Regions explored in positive electrostatics, (C) Regions explored in Hydrophobic, and (D) Shape explored.

these regions indicate more anticancer activity. Figure 7B shows the hydrophobic interaction regions of active compounds as indicated by the yellow color, and Supplementary Figure S1 shows the shape regions of active compounds as represented by the white color.

Activity cliff summary

The activity cliff summary diagrams as indicated by Figure 8A correlate with the biologically active parts of imidazole drugs with the reference drug. The cliff summary of electrostatics in Figure 8A is visualized in two colors: red and cyan. The presence of the red color indicating the positive electrostatic field and the cyan color indicating the negative electrostatic field is favorable for high anticancer activity. Figure 8B shows the areas of favorable and unfavorable hydrophobics represented by the green and purple colors, respectively. Whereas in Figure 8C, the green color shows a favorable shape region, and the purple color shows an unfavorable shape region.

Regions explored

The descriptive features of compounds were explored in this model aside from their biological activities (Attig et al., 2022). The more red and cyan colors indicating positive and negative fields, respectively, show the areas of strong SAR with the reference drug. The average regions explored also represent the areas of active compounds that would not take part in an anticancer activity (Figure 9).

Validation of the 3D-QSAR model

Molecular characteristics regulating the active compounds as anticancer agents were retrieved for further prediction of their anticancer activity based on derived SAR models. Before that, prediction performance was analyzed on the test set and training set compounds by predicting their anticancer activity. This prediction was done by means of derived models and then the distance value (error) was compared. In order to perform the comparison, predicted activity and distance to models' columns

were examined for each derived model. The important ligand fields were illustrated for each derived model through this study and after that, these characteristics were utilized for virtual screening.

Ligand-based virtual screening

To predict hits, a series of ligand-based virtual screening experiments were performed. Only high-hit compounds were selected having the value of 'excellence'. The excellence of hits was set by taking a threshold of docking score -8.7 kcal/mol to compare the biological activity. The predicted activities were expected to be reliable because most of the characteristics in compounds were the same as the training set. In contrast, compounds having poor field point similarities were excluded to evade the false positive compounds by ineffective predicted activities. Also, the derived QSAR model was used to predict the hit compounds for anticancer activity (Lakshmithendral et al., 2019b).

Structure-based virtual screening (SBVS)

SBVS of selected compounds was performed to discover new valuable drugs in order to treat breast cancer (El Aissouq et al., 2021). AutoDock Tools provided notable results with overall binding energy of all the selected compounds ranging from -5.1 to -10.5 kcal/mol. Most of the compounds have binding free energy greater than the reference compound Fulvestrant when observed with all the selected proteins as shown in Table 2.

The docking results explain that compounds C6, C10, C21, and C76 are the most hit compounds giving excellent results, so only these compounds will be subjected to further study.

Molecular Interaction and Binding Mode. The top hit compounds represented by the shaded area in Table 2 were selected to evaluate the binding site interactions between the ligand and the target protein.

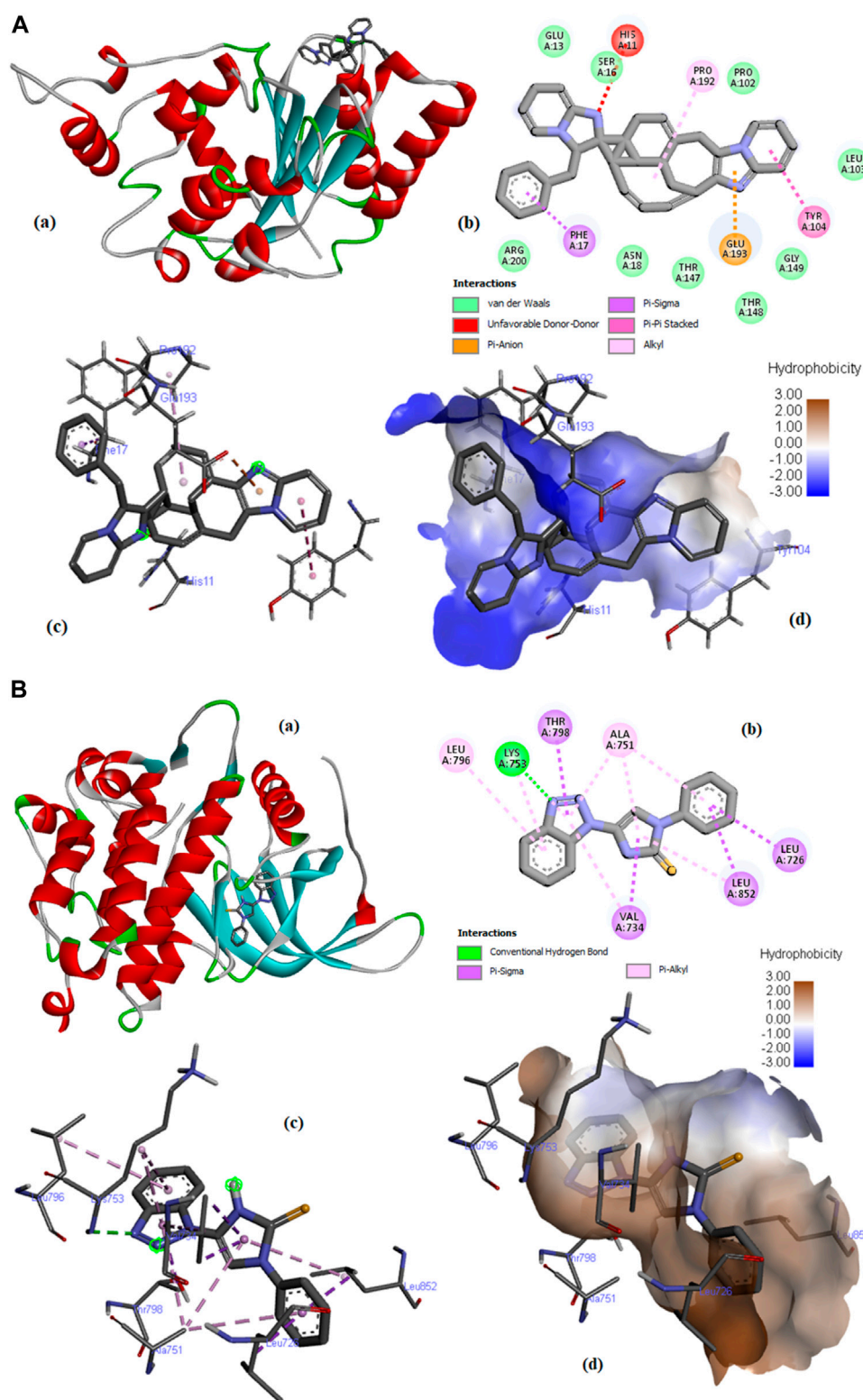
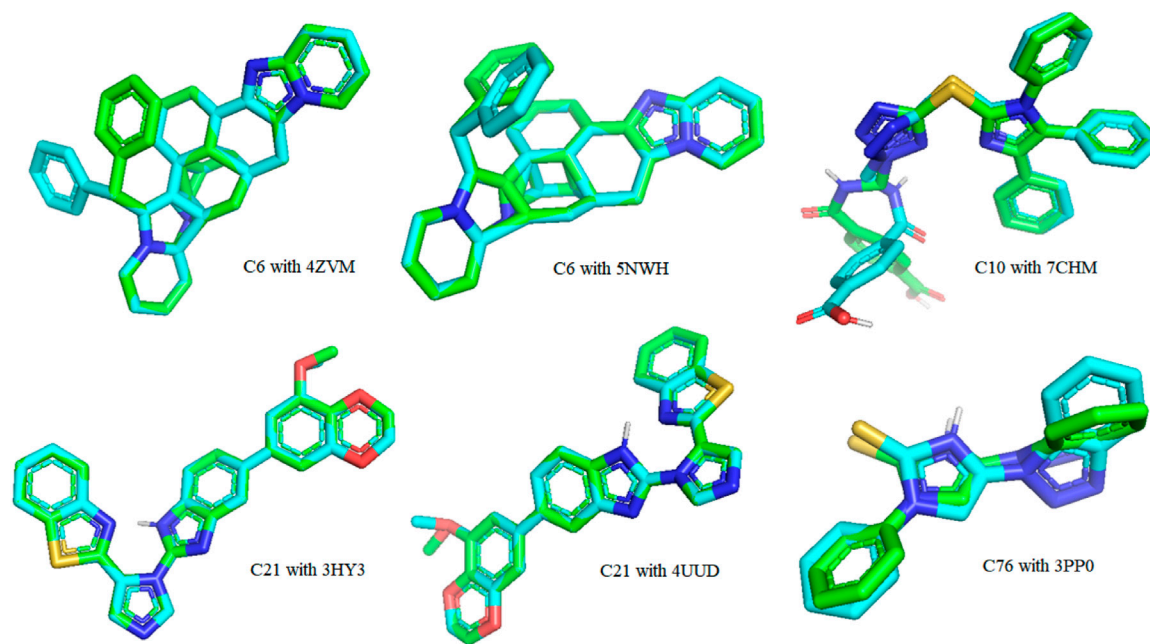
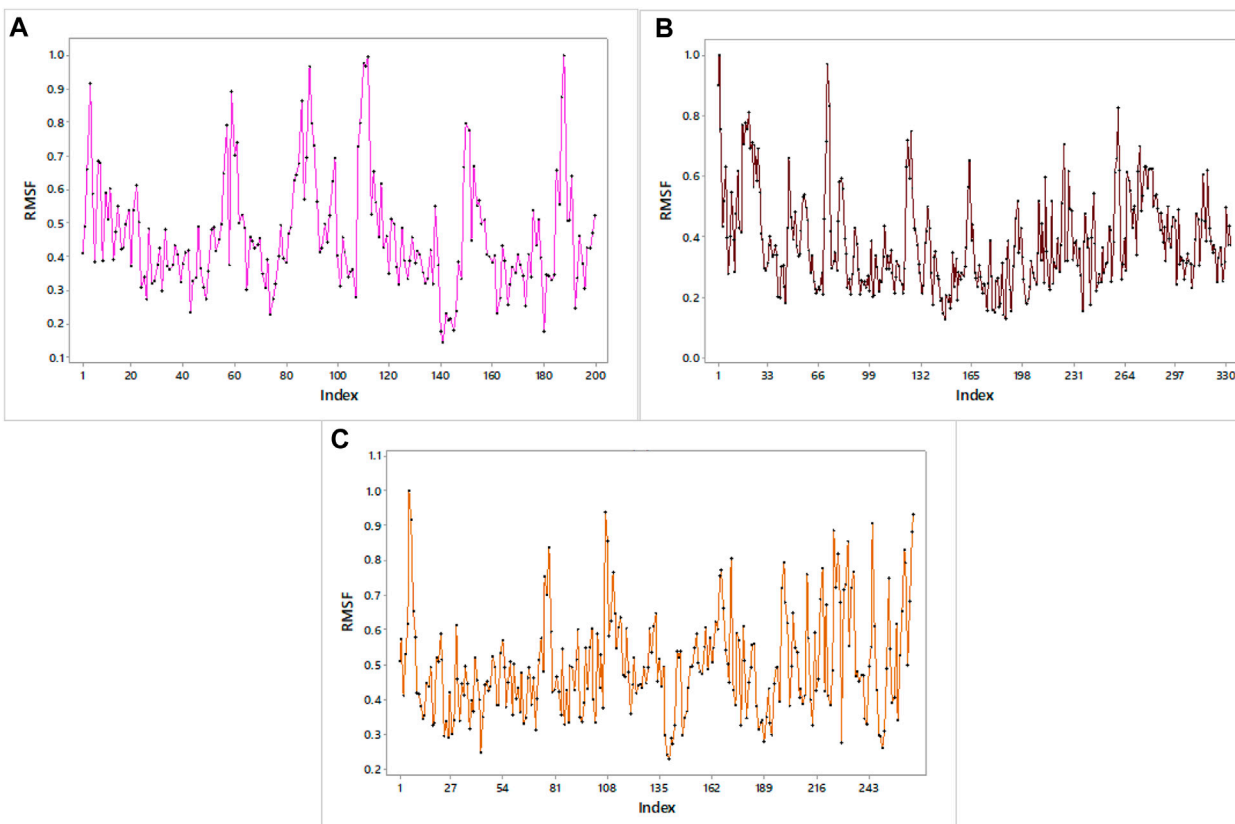


FIGURE 10

(A) Molecular docking of compound **C6** with PDB ID: 4ZMV; (a) 3D view of the best-selected conformation, (b) 2D Interactions, (c) Ligand interactions, and (d) Hydrophobicity. (B) Molecular docking of compound **C76** with PDB ID: 3PP0 (a) 3D view of the best-selected conformation, (b) 2D Interactions, (c) Ligand interactions, and (d) Hydrophobicity.

**FIGURE 11**

A 3-dimensional view of the best-docked pose of the ligand that fitted into the binding pocket of the protein receptor binding site.

**FIGURE 12**

RMSF profiles of (A) 3HY3, (B) 4UUD, and (C) 7CHM.

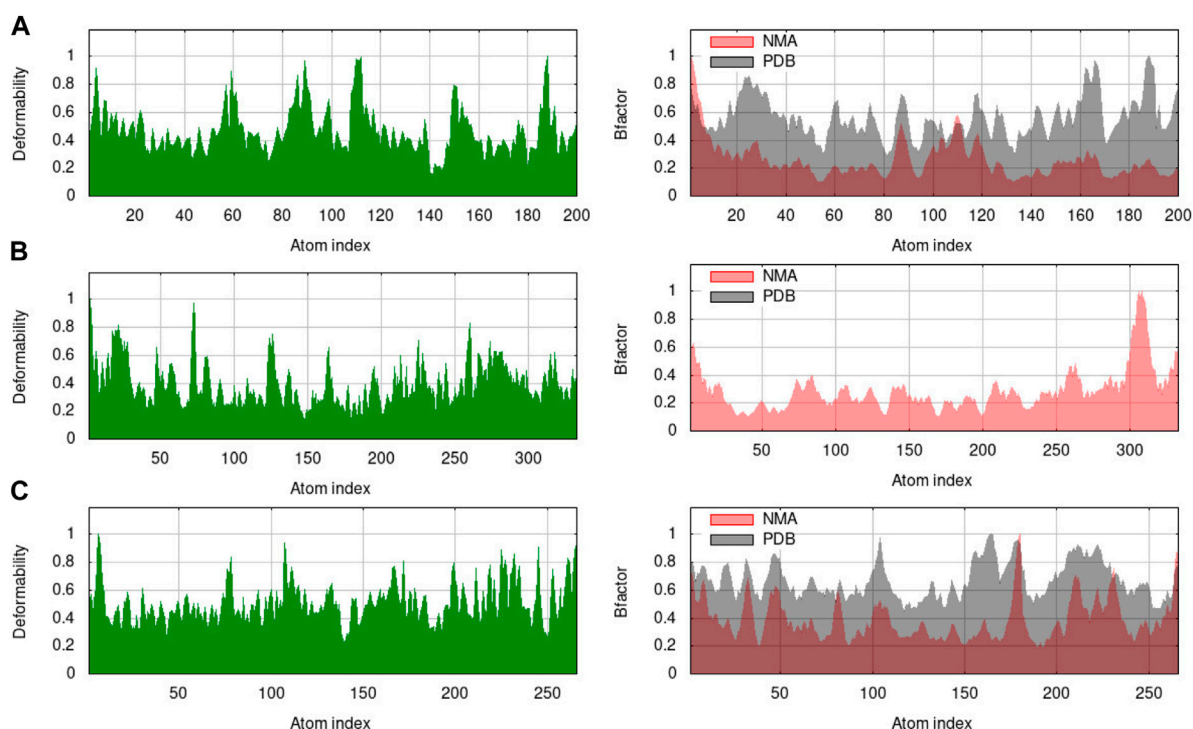


FIGURE 13
Deformability and B-factor of docked complexes. (A) 3HY3-C10, (B) 4UUD-C10, and (C) 7CHM-C10.

Compound **C6** was fixed in the binding pocket of protein (PDB ID: 4ZVM) (Figure 10A) by undergoing electrostatic interactions (Pi-Anion) with GLU193 and hydrophobic interactions (Pi-Sigma, Pi-Pi Stacked, and Alkyl) with PHE17, TYR104, and PRO192 (Supplementary Table S3).

The docking of compound **C6** with protein (PDB ID: 5NWH) is described in Supplementary Figure S2. **C6** undergoes interactions with the protein by hydrophobic interactions (Pi-Sigma and Pi-Alkyl) with VAL 49, VAL29, and PRO86, Pi-Cation electrostatic interactions with ARG196, and other interactions (Pi-Sulfur) with CYS91 (Supplementary Table S3).

Supplementary Figure S3 describes the docking of compound **C10** with the protein having PDB ID: 7CHM. The compound fits in the binding pocket of the protein through hydrogen bonding (Conventional Hydrogen Bond and Carbon Hydrogen Bond) with amino acids ASP608, SER611, GLN530, ASP674, and ILE607; hydrophobic interactions (Pi-Sigma, Pi-Pi Stacked, Alkyl, and Pi-Alkyl) with ILE531, ILE607, LEU654, ALA651, PRO673, VAL539, ILE663, and ALA551; and miscellaneous interactions (Pi-Sulfur) with MET602, CYS604, and MET671 (Supplementary Table S4).

However, the docking of compound **C21** with PDB ID: 3HY3 provided different results (Supplementary Figure S4). **C10** interacted by forming a conventional hydrogen bond with TRP109 and hydrophobic interactions (Pi-Pi Stacked, Pi-Pi T-shaped, Alkyl, and Pi-Alkyl) with TYR83, TRP109, PRO81,

MET90, TYR152, TYR153, and LYS150 (Supplementary Table S4) (Elancheran et al., 2023a).

Compound **C21** fits in the binding pocket of protein (PDB ID: 4UUD) (Supplementary Figure S5) through hydrogen bonding (conventional hydrogen bond and carbon-hydrogen bond) with GLN642, GLN738, and PRO637, hydrophobic interactions (Pi-Pi Stacked) with TRP557 and TYR735, and miscellaneous interactions (Pi-Sulfur) with MET745 (Supplementary Table S2). Docking results of Compound **C76** with PDB ID: 3PP0 are described in Figure 10B. **C76** interacted by forming conventional hydrogen bond with LYS753 and hydrophobic interactions (Pi-Sigma and Pi-Alkyl) with LEU726, VAL734, THR798, LEU852, ALA751, and LYS753 (Supplementary Table S2).

Validation of docking

The re-docking of the native ligand with the protein receptor binding site was performed to validate the docking process by using the PyMOL molecules graphic system, version 2.4.1. The crystal structures were aligned to compare their changes in conformation and displacement. The results were reported in root-mean-squared deviation (RMSD) to calculate the deviation between analogous atoms of two proteins, i.e., the docked pose and the corresponding crystal conformer. Redocking of all hit compounds with PDB IDs 4ZVM, 5NWH, 7CHM, 3HY3, 4UUD, and 3PP0 resulted in the

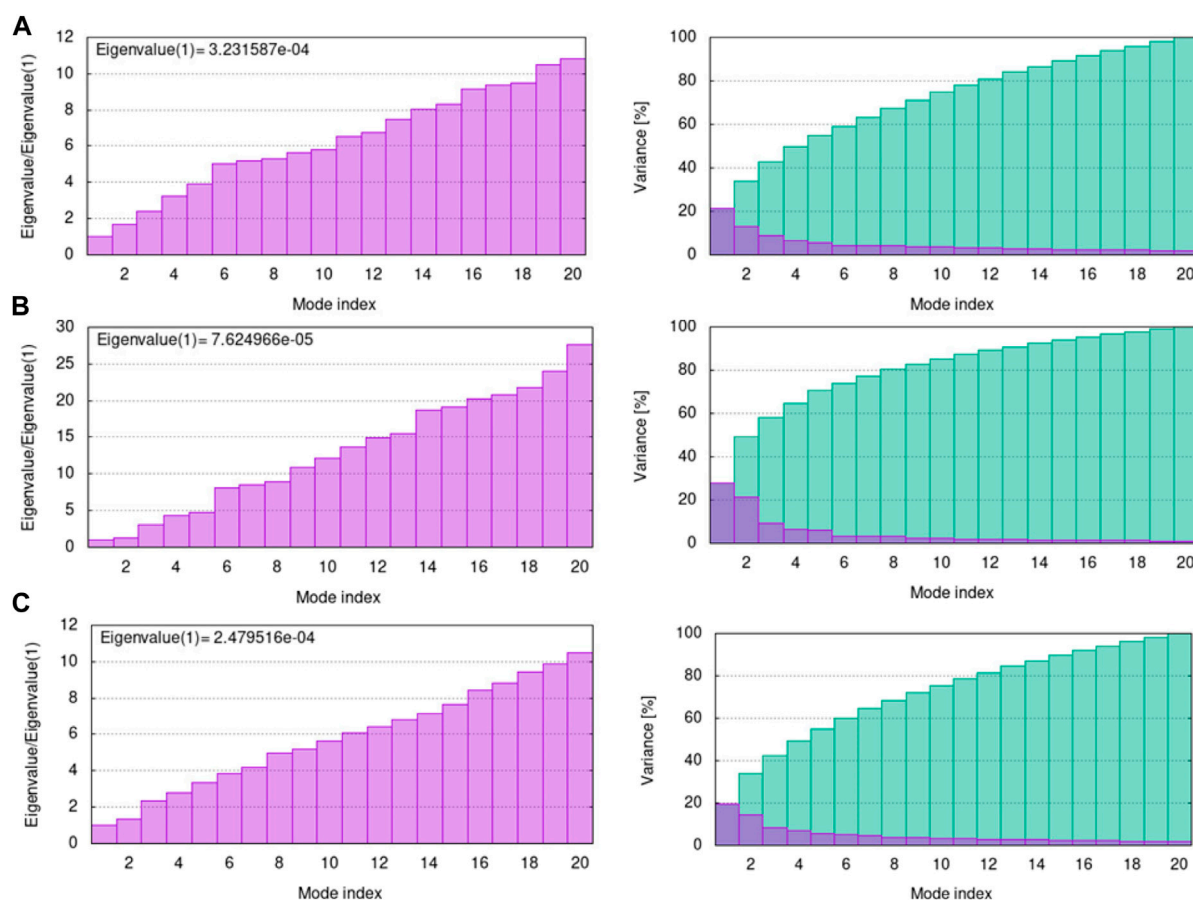


FIGURE 14
Eigenvalue and variance of docked complexes (A) 3HY3-C10, (B) 4UUD-C10, and (C) 7CHM-C10.

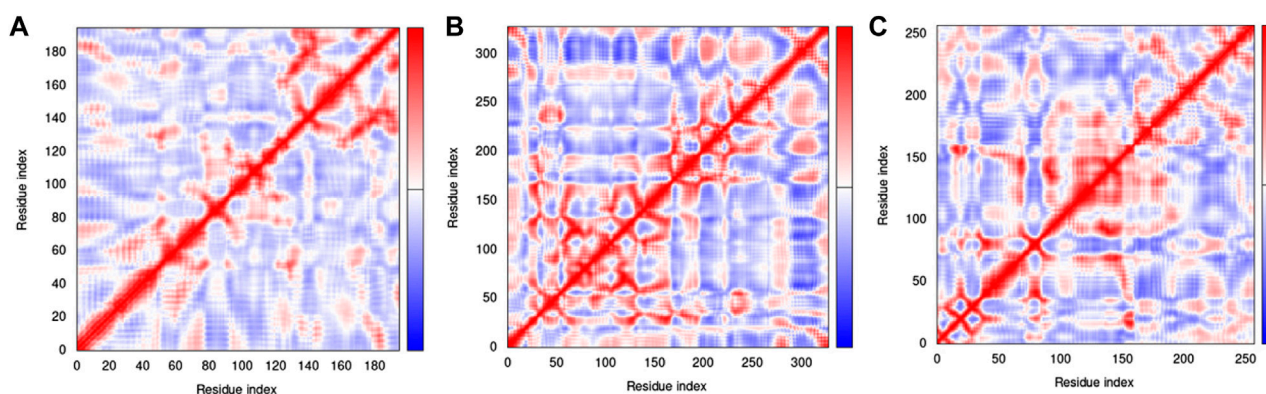


FIGURE 15
Covariance matrix of complexes (A) 3HY3-C10, (B) 4UUD-C10, and (C) 7CHM-C10.

RMSD values shown in [Supplementary Table S5](#). The lower value of RMSD revealed that the ligands were bound to target very closely to the original conformation, hence, signifying the accuracy of results. The RMSD value close to zero was considered to be ideal. A superimposed view is displayed in [Figure 11](#).

Molecular dynamics simulation

The RMSF graph provides insight into the flexibility of individual atoms or residues in the protein. It shows how much they deviate from their average positions during the simulation. The maximum value of RMSF indicates greater flexibility, while the

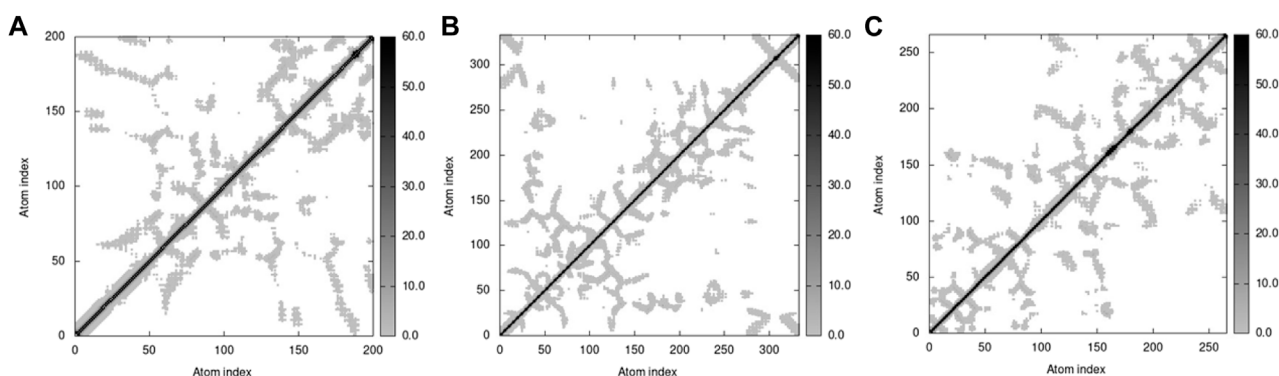


FIGURE 16
Elastic maps of docked proteins (A) 3HY3-C10, (B) 4UUD-C10, and (C) 7CHM-C10.

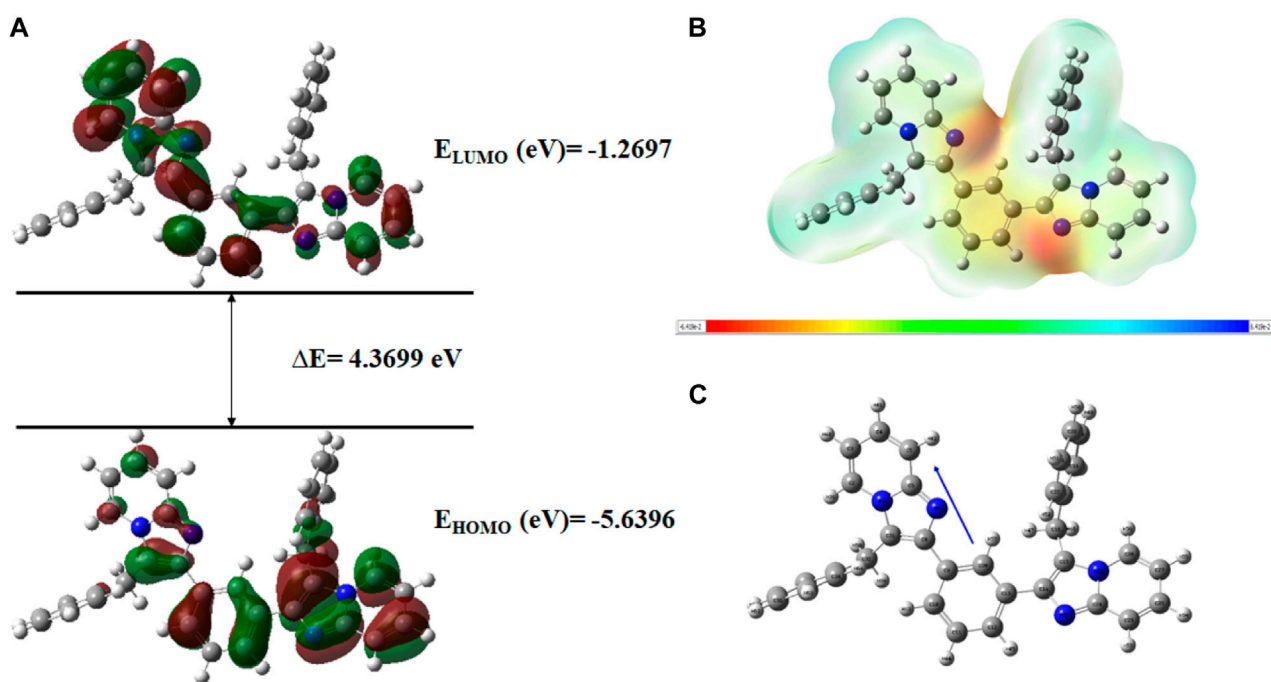
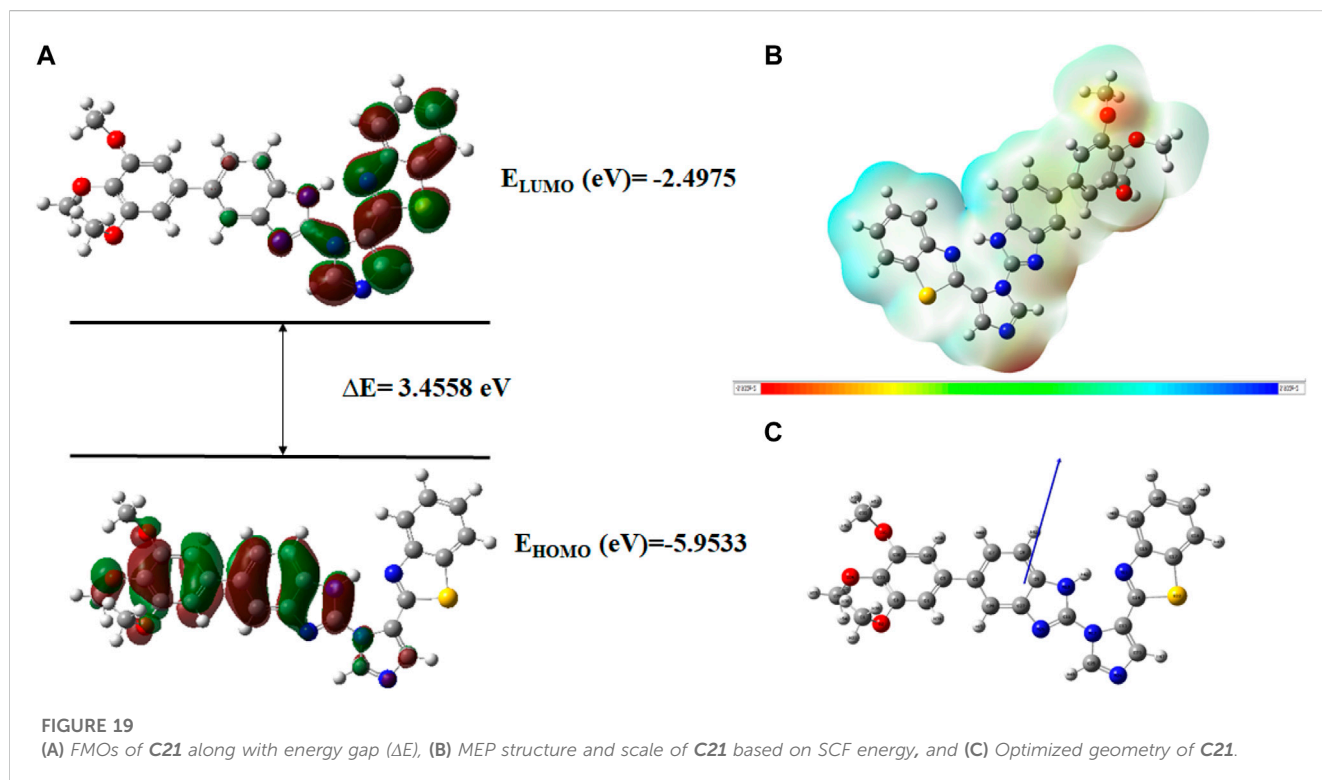
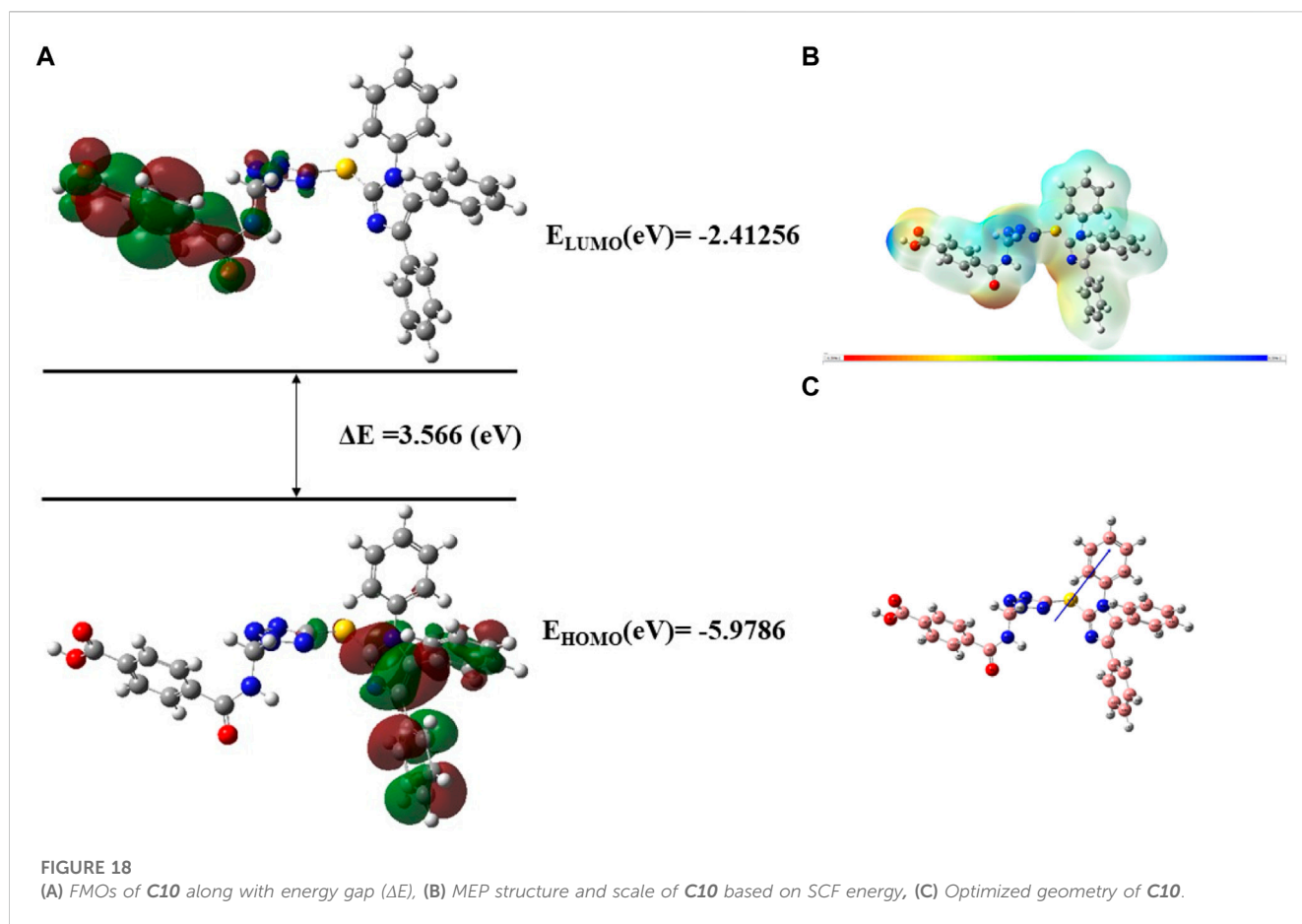


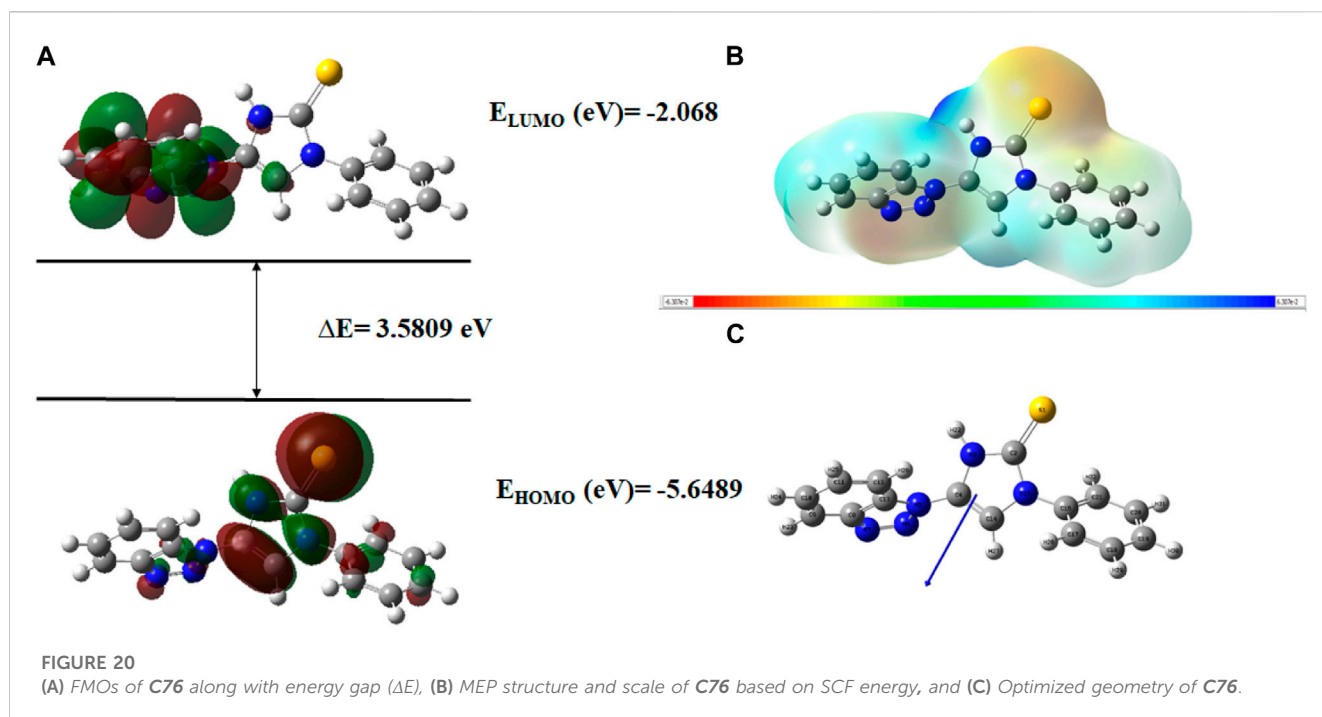
FIGURE 17
(A) FMOs of **C6** along with energy gap (ΔE), (B) MEP structure and scale of **C6** based on SCF energy, and (C) Optimized geometry of **C6**.

smallest value denotes the system's restricted motion across the simulation course. In all of our proteins, the RMSF graph showed a number of areas having high flexibility (Figure 12). The complex of C10 with H3Y3 showed a maximum number of peaks, indicating more flexible movements. All three docked complexes showed a maximum RMSF of 1.0. It indicated that the atoms or residues are, on average, deviating from their average positions by around 1 Å (Angstrom).

The B-factor, eigenvalue, deformability, covariance matrix, variance map, and elastic network model of the protein serves as

a representation of its stability. The mobility characteristics of the docked proteins are determined by the deformability and B-factor. The peaks are associated with the protein regions with deformability whereas the areas with the highest peaks are those with the greatest deformability (Santra and Maiti, 2022). In B-factor graphs, the comparison between the PDB field and NMA of the docked complexes is provided. The B-factor graphs of 3H3Y-C10 and 7CHM-C10 complexes showed that the PDB data predicted higher B-factors compared to the NMA data. It suggested that the B-factor values predicted by the





computational simulations using NMA showed lower mobility or flexibility than what was predicted by the experimentally determined B-factor values from the Protein Data Bank. **Figure 13** illustrates the deformability and B-factor of 3HY3-C10, 4UUD-C10, and 7CHM-C10, respectively.

The eigenvalue indicates motion stiffness which is directly connected to the energy needed to deform the structure. If the eigenvalue is low, it means that the complex is more easily deformable. The eigenvalues of 3H3Y, 4UUD, and 7CHM complexes with **C10** are $3.231587\text{e-}04$, $7.624966\text{e-}05$, and $2.479516\text{e-}04$, respectively. It means that all of our docked complexes showed low eigenvalues, indicating a considerable amount of deformability and, hence, good flexibility and stability of the molecular motion. The individual variance is shown by purple-shaded bars in the variance graph of **C10** with target proteins, while the bars with green shading show cumulative variance. The eigenvalue and variance graphs of protein-ligand complexes (H3Y3-C10, 4UUD-C10, and 7CHM-C10) are shown in **Figure 14**.

The covariance matrix indicates correlations among the pairs of residues in a protein-ligand complex (**Figure 15**). The red and white colors showed correlated and uncorrelated motion, respectively, while anticorrelations are represented by the blue color. Greater correlation means the formation of a better complex. The covariance matrices for H3Y3-C10, 4UUD-C10, and 7CHM-C10 complexes exhibited good correlations and minimal anticorrelations.

The elastic network model of docked proteins shows relationships between the atoms where the stiffer regions are indicated by the darker grey areas (**Figure 16**). All protein elastic maps yielded reliable results.

Toxicity prediction

Toxicity results provided valuable information related to the toxicological profile of selected compounds after molecular docking studies and thus may be useful for drug designing (to select the dosage and preferred route of administration (**Banerjee et al., 2018**)). However, all these results are preliminary and must be confirmed by experiment (**Table 3**).

As shown in **Table 3**, among all the hit compounds, **C10** was found to be inactive in all toxicity parameters and, thus, it was considered the best-hit compound.

Density functional theory

Frontier molecular orbitals

Very useful information about the compounds can be provided by Frontier molecular orbitals (FMO) such as electronegativity, stability, reactivity, and chemical hardness and softness (**Elancheran et al., 2023b**). The HOMO and LUMO parameters are used to compute the chemical reactivity descriptors and to assess the molecular reactivity (**Al-Janabi et al., 2021**). The energy values of HOMO and LUMO were determined by the DFT method as shown in **Supplementary Table S6**. Contour diagrams of FMOs are shown in **Figures 17–20**.

The geometry optimization was done to minimize the energy and to find the most stable atomic arrangement. Optimized geometries of all compounds are shown in **Figures 17–20** along with the numbering system and the vector of dipole moment.

Computational description

The DFT calculations for ionization energy, electron affinity, energy gap, electronegativity, chemical potential, electrophilicity

Parameters	C6	C10	C21	C76
E_{HOMO} (eV)	-5.6396	-5.9786	-5.9533	-5.6489
E_{LUMO} (eV)	-1.2697	-2.41256	-2.4975	-2.068
ENERGY GAP ΔE (eV)	4.3699	3.566	3.4558	3.5809
IONIZATION POTENTIAL ($I = -E_{\text{HOMO}}$)	5.6396	5.9786	5.9533	5.6489
ELECTRON AFFINITY ($A = -E_{\text{LUMO}}$)	1.2697	2.41256	2.4975	2.068
ELECTRONEGATIVITY ($\chi = \frac{(I+A)}{2}$) (eV)	3.4546	4.195	4.2254	3.8584
CHEMICAL POTENTIAL ($M = -\frac{(I+A)}{2}$) (eV)	-3.4546	-4.195	-4.2254	-3.8584
CHEMICAL HARDNESS ($H = \frac{(I-A)}{2}$) (eV)	2.1849	1.78	1.7279	1.79
CHEMICAL SOFTNESS ($S = \frac{1}{2H}$) (eV)	0.2288	0.28	0.2893	0.2793
ELECTROPHILICITY INDEX ($\Omega = \frac{\mu^2}{2\eta}$) (eV)	2.731	4.94	5.1663	4.1584
NUCLEOPHILICITY INDEX ($N = \frac{1}{\omega}$) (eV)	0.366	0.2024	0.1936	0.2405
MAXIMUM CHARGE TRANSFER INDEX ($\Delta N_{\text{MAX}} = \frac{\mu}{\eta}$) (eV)	1.581	2.3567	2.4454	2.155
DIPOLE MOMENT (DEBYE)	3.970092	7.126807	6.225672	4.797580

index, chemical softness and hardness, additional electronic charges, maximum charge transfer index, and dipole moment were performed as explained in [Supplementary Table S6](#)).

A high E_{HOMO} indicates that the molecule is a strong electron donor and can easily donate electrons to the receptor, increasing biological activity. According to this concept, the activity ranking of the hit compounds is given below with an increasing E_{HOMO} value.

$$C10 > C21 > C76 > C6$$

A low E_{LUMO} value indicates that the compound can easily accept electrons from the donor molecule, increasing biological activity ([Mu and Gao, 2022](#)). According to this criterion, an increase in the biological activity of compounds is as follows:

$$C6 > C76 > C10 > C21$$

The third parameter is the energy gap (ΔE) between HOMO and LUMO. If the energy gap is small, it indicates that the molecule is soft, biologically active, less stable, and has a high chemical reactivity. In other words, the biological activity increases with a decrease in the energy gap. The order of ranking should be

$$C21 > C10 > C76 > C6$$

A high chemical potential (CP) or Lower electronegativity (X) value indicates electron delocalization. It means that the molecule can easily form bonds and coordinate easily with the biological system. So according to our calculated DFT data, an increase in the biological activity of compounds is given below:

$$C21 > C10 > C76 > C6$$

The dipole moment also affects the biological activity of the compound. The high value of the dipole moment indicates the strong ligand-protein interaction, thus increasing the biological activity ([Sayin and Üngördü, 2018](#)). [Supplementary Table S6](#)

shows that the dipole moment of our hit compounds are in the order of

$$C10 > C21 > C76 > C6$$

Molecular electrostatic potential (MEP)

In order to determine a chemical mechanism, MEP maps and MEP contours play a very important role. Molecular electrostatic potential helps to determine the hydrogen bonding interactions and to interpret the nucleophilic as well as electrophilic reactions ([Horchani et al., 2020](#)). MEP can be used to indicate the shape of the molecule and the sizes of the negative, positive, and neutral electrostatic potential. The molecular structure of drugs along with the interaction among different physicochemical properties can be predicted by the MEP scale ([Al-Janabi et al., 2021](#)). The MEP of compounds **C6**, **C10**, **C21**, and **C76** is determined under the basis set of B3LYP/6-311G. The negative charge is indicated by red and yellow areas that represent the electrophilic attack sites. The green color indicated a neutral charge while the blue region indicating the positively charged areas represented the nucleophilic reactivity ([Bendjeddou et al., 2016](#)). The MEP structures and MEP scales of compounds are shown in [Figures 17–20](#).

Conclusion

After the virtual screening, **C10** was found to be the best imidazole derivative clearing all the filters. The 3D-QSAR models generated in this study provided valuable insights into the structural features and molecular interactions that contribute to the compounds' activity against breast cancer cells. The derived PLS regression model confirmed a fairly acceptable value of regression coefficient ($r^2 = 0.81$) and cross-validation regression coefficient ($q^2 = 0.51$). Docking results based on the binding free energy values were found to support the best-hit compounds. The DFT calculations also confirm the best alternative cancer inhibitor. These predictions aided in rationalizing

the observed biological activities and potential mechanisms of action of these compounds against breast cancer cells. MD simulation study supported the docking results of **C10** to its target proteins or receptors. This study showcased the lead compound's stability and robustness, suggesting its suitability for further preclinical and clinical evaluations. The compound's favorable binding profile, coupled with its ability to sustain its interactions over extended simulation periods, instills confidence in its potential as a promising candidate for subsequent stages of drug development. The results acquired from the present study may be utilized in the future to develop more imidazole-based therapeutics against cancer. The identification of Compound C10 as a lead compound opens up avenues for further drug development and optimization and offers valuable insights and potential directions for future research and clinical applications.

Data availability statement

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

Author contributions

MR: Drafting and evaluation of results; AM: Library preparation, drafting, 3DQSAR, and Docking; NS: Conceptualization, supervision, Funding, corrections, proofreadings, overall final editing, and setting; YB: English editing and proof editing; SP: data analysis and Literature; MB: Funding acquisition and molecular docking validation; H-AN: data analysis and validation; RK: Library data analysis, proofreading, and English check.

References

- Abbasov, V., Marzouk, A. A., Mammadov, A., Kazimova, S., and Talybov, A. (2012). Imidazole derivatives, synthesis and biological activity. *Process. Petrochem. oil-refining* 13, 52.
- Al-Blewi, F., Shaikh, S. A., Naqvi, A., Aljohani, F., Aouad, M. R., Ihmaid, S., et al. (2021). Design and synthesis of novel imidazole derivatives possessing triazole pharmacophore with potent anticancer activity, and in silico ADMET with GSK-3 β molecular docking investigations. *Int. J. Mol. Sci.* 22, 1162. doi:10.3390/ijms22031162
- Al-Janabi, A. S., Elzupir, A. O., and Yousef, T. A. (2021). Synthesis, anti-bacterial evaluation, DFT study and molecular docking as a potential 3-chymotrypsin-like protease (3CLpro) of SARS-CoV-2 inhibitors of a novel Schiff bases. *J. Mol. Struct.* 1228, 129454. doi:10.1016/j.molstruc.2020.129454
- Alam, S., and Khan, F. (2017). 3D-QSAR studies on Maslinic acid analogs for Anticancer activity against Breast Cancer cell line MCF-7. *Sci. Rep.* 7, 6019–6113. doi:10.1038/s41598-017-06131-0
- Alam, S., and Khan, F. (2019). 3D-QSAR, Docking, ADME/Tox studies on Flavone analogs reveal anticancer activity through Tankyrase inhibition. *Sci. Rep.* 9, 5414–5415. doi:10.1038/s41598-019-41984-7
- Alam, S., and Khan, F. (2014). QSAR and docking studies on xanthone derivatives for anticancer activity targeting DNA topoisomerase IIa. *Drug Des. Dev. Ther.* 8, 183–195. doi:10.2147/DDDT.S51577
- Ali, I., Lone, M. N., and Aboul-Enein, H. Y. (2017). Imidazoles as potential anticancer agents. *MedChemComm* 8, 1742–1773. doi:10.1039/c7md00067g
- Anandan, S., Gowtham, H. G., Shivakumara, C., Thamby, A., Singh, S. B., Murali, M., et al. (2022). Integrated approach for studying bioactive compounds from *Cladosporium* spp. against estrogen receptor alpha as breast cancer drug target. *Sci. Rep.* 12, 22446. doi:10.1038/s41598-022-22038-x
- Ataollahi, M., Sharifi, J., Paknahad, M., and Paknahad, A. (2015). Breast cancer and associated factors: A review. *J. Med. life* 8, 6–11.
- Attig, N., Arshad, U., Brogi, S., Shafiq, N., Imtiaz, F., Parveen, S., et al. (2022). Exploring the anti-SARS-CoV-2 main protease potential of FDA approved marine drugs using integrated machine learning templates as predictive tools. *Int. J. Biol. Macromol.* 1, 1415–1428. doi:10.1016/j.ijbiomac.2022.09.086
- 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, 63–16. doi:10.1186/1742-4682-10-63
- Banerjee, P., Eckert, A. O., Schrey, A. K., and Preissner, R. (2018). ProTox-II: A webserver for the prediction of toxicity of chemicals. *Nucleic acids Res.* 46, W257–W263. doi:10.1093/nar/gky318
- Brueggemeier, R. W., Richards, J. A., and Petrel, T. A. (2003). Aromatase and cyclooxygenases: Enzymes in breast cancer. *J. steroid Biochem. Mol. Biol.* 86, 501–507. doi:10.1016/s0960-0760(03)00380-7
- Chopra, P. N., and Sahu, J. K. (2019). Biological significance of imidazole-based analogues in new drug development. *Current Drug Discovery Technologies.* 17(5), pp.574–584. doi:10.2174/1570163816666190320123340
- Comşa, S., Cimpean, A. M., and Raica, M. (2015). The story of MCF-7 breast cancer cell line: 40 years of experience in research. *Anticancer Res.* 35, 3147–3154.
- Dickson, R. B. (1990). Stimulatory and inhibitory growth factors and breast cancer. *J. Steroid Biochem. Mol. Biol.* 37, 795–803. doi:10.1016/0960-0760(90)90422-h
- Edukondalu, P., Sireesha, R., Bandaru, C. M., Rao, M. V. B., Kala, P., and Raju, R. R. (2021). Design, Synthesis and anticancer evaluation of 2-(5-(Benzo[d]thiazol-2-yl)-1H-imidazol-1-yl)-5-aryl-1H-benzo[d]imidazole derivatives. *Chem. Data Collect.* 35, 100753. Elsevier. doi:10.1016/j.cdc.2021.100753
- El Aissouq, A., Chedadi, O., Bouachrine, M., and Ouammou, A. (2021). Identification of novel SARS-CoV-2 inhibitors: A structure-based virtual screening approach. *J. Chem.* 2021, 1–7. doi:10.1155/2021/1901484

Acknowledgments

The authors are thankful to the Higher Education Commission of Pakistan for their financial assistance to conduct this study under Grant# TDF03-172. The authors would like to extend their sincere appreciation to the Researchers Supporting Project, King Saud University, Riyadh, Saudi Arabia for funding this work through the project number (RSP2023R457).

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/fchem.2023.1197665/full#supplementary-material>

- Elancheran, R., Karthikeyan, B., Srinivasan, S., Krishnasamy, K., and Kabilan, S. (2023). Synthesis, crystal structure, DFT and Hirshfeld surface analysis of 4-fluoro-N-(1,3-dioxoisindolin-2-yl)benzamide. *Eur. J. Chem.* 14, 1–8. doi:10.5155/eurjchem.14.1.1-8.2335
- Escala-Garcia, M., Morra, A., Canisius, S., Chang-Claude, J., Kar, S., Zheng, W., et al. (2020). Breast cancer risk factors and their effects on survival: A mendelian randomisation study. *BMC Med.* 18, 327–410. doi:10.1186/s12916-020-01797-2
- Fan, C., Zhong, T., Yang, H., Yang, Y., Wang, D., Yang, X., et al. (2020). Design, synthesis, biological evaluation of 6-(2-amino-1H-benzo [d] imidazole-6-yl) quinazolin-4 (3H)-one derivatives as novel anticancer agents with Aurora kinase inhibition. *Eur. J. Med. Chem.* 190, 112108. doi:10.1016/j.ejmech.2020.112108
- Gaba, M., and Mohan, C. (2016). Development of drugs based on imidazole and benzimidazole bioactive heterocycles: Recent advances and future directions. *Med. Chem. Res.* 25, 173–210. doi:10.1007/s00044-015-1495-5
- Gowtham, H. G., Ahmed, F., Anandan, S., Shivakumara, C., Bilagi, A., Pradeep, S., et al. (2023). In silico computational studies of bioactive secondary metabolites from wedelia trilobata against anti-apoptotic B-cell lymphoma-2 (Bcl-2) protein associated with cancer cell survival and resistance. *Molecules* 28, 1588. doi:10.3390/molecules28041588
- Howell, A., Anderson, A. S., Clarke, R. B., Duffy, S. W., Evans, D. G., Garcia-Closas, M., et al. (2014). Risk determination and prevention of breast cancer. *Breast Cancer Res.* 16, 446–519. doi:10.1186/s13058-014-0446-2
- Jardé, T., Perrier, S., Vasson, M.-P., and Caldefie-Chézet, F. (2011). Molecular mechanisms of leptin and adiponectin in breast cancer. *Eur. J. cancer* 47, 33–43. doi:10.1016/j.ejca.2010.09.005
- Keith, J. A., Vassilev-Galindo, V., Cheng, B., Chmiela, S., Gastegger, M., Müller, K.-R., et al. (2021). Combining machine learning and computational chemistry for predictive insights into chemical systems. *Chem. Rev.* 121, 9816–9872. doi:10.1021/acs.chemrev.1c00107
- Khayyat, A. N., Mohamed, K. O., Malebari, A. M., and El-Malah, A. (2021). Design, synthesis, and antiproliferative activities of novel substituted imidazole-thione linked benzotriazole derivatives. *Molecules* 26, 5983. doi:10.3390/molecules26195983
- Kirar, M., Singh, H., and Sehrawat, N. (2022). Virtual screening and molecular dynamics simulation study of plant protease inhibitors against SARS-CoV-2 envelope protein. *Inf. Med. Unlocked* 30, 100909. doi:10.1016/j.imu.2022.100909
- Lakshmithendral, K., Saravanan, K., Elancheran, R., Archana, K., Manikandan, N., Arjun, H., et al. (2019). Design, synthesis and biological evaluation of 2-(phenoxyethyl)-5-phenyl-1, 3, 4-oxadiazole derivatives as anti-breast cancer agents. *Eur. J. Med. Chem.* 168, 1–10. doi:10.1016/j.ejmech.2019.02.033
- Lee, A. V., Oesterreich, S., and Davidson, N. E. (2015). MCF-7 cells—Changing the course of breast cancer research and care for 45 years. *JNCI J. Natl. Cancer Inst.* 107, djv073. doi:10.1093/jnci/djv073
- Low, C. M. R., Buck, I. M., Cooke, T., Cushnir, J. R., Kalindjian, S. B., Kotecha, A., et al. (2005). Scaffold hopping with molecular field points: Identification of a cholecystokinin-2 (CCK2) receptor pharmacophore and its use in the design of a prototypical series of pyrrole- and imidazole-based CCK2 antagonists. *J. Med. Chem.* 48, 6790–6802. doi:10.1021/jm049069y
- Maruthanila, V., Elancheran, R., Kunnumakkara, A., Kabilan, S., and Kotoky, J. (2017). Recent development of targeted approaches for the treatment of breast cancer. *Breast Cancer* 24, 191–219. doi:10.1007/s12282-016-0732-1
- Masoud, V., and Pagès, G. (2017). Targeted therapies in breast cancer: New challenges to fight against resistance. *World J. Clin. Oncol.* 8, 120. doi:10.5306/wjco.v8.i2.120
- Meenakshisundaram, S., Manickam, M., and Pillaiyar, T. (2019). Exploration of imidazole and imidazopyridine dimers as anticancer agents: Design, synthesis, and structure-activity relationship study. *Arch. Pharm.* 352, 1900011. doi:10.1002/ardp.201900011
- Mu, M., and Gao, H. (2022). DFT study on the substituent effect of anticancer picoline-diazido-Pt (IV) compounds. *Front. Oncol.* 11, 749178. doi:10.3389/fonc.2021.749178
- Mumtaz, A., Saeed, A., Fatima, N., Dawood, M., Rafique, H., and Iqbal, J. (2016). Imidazole and its derivatives as potential candidates for drug development. *Bangladesh J. Pharmacol.* 11, 756–764. doi:10.3329/bjp.v11i4.26835
- Pasqualini, J. R., and Chetrite, G. S. (2005). Recent insight on the control of enzymes involved in estrogen formation and transformation in human breast cancer. *J. steroid Biochem. Mol. Biol.* 93, 221–236. doi:10.1016/j.jsbmb.2005.02.007
- Rizzo, L., Longato, G., Ruiz, A., Tinti, S., Possenti, A., Vendramini-Costa, D., et al. (2014). In vitro, in vivo and in silico analysis of the anticancer and estrogen-like activity of guava leaf extracts. *Curr. Med. Chem.* 21, 2322–2330. doi:10.2174/0929867321666140120120031
- Romagnoli, R., Baraldi, P. G., Prencipe, F., Oliva, P., Baraldi, S., Tabrizi, M. A., et al. (2016). Design and synthesis of potent in vitro and in vivo anticancer agents based on 1-(3', 4', 5'-trimethoxyphenyl)-2-aryl-1H-imidazole. *Sci. Rep.* 6, 26602–26617. doi:10.1038/srep26602
- Santra, D., and Maiti, S. (2022). Molecular dynamic simulation suggests stronger interaction of Omicron-spike with ACE2 than wild but weaker than Delta SARS-CoV-2 can be blocked by engineered S1-RBD fraction. *Struct. Chem.* 33, 1755–1769. doi:10.1007/s11224-022-02022-x
- Sayin, K., and Üngördü, A. (2018). Investigation of anticancer properties of caffeinated complexes via computational chemistry methods. *Spectrochimica Acta Part A Mol. Biomol. Spectrosc.* 193, 147–155. doi:10.1016/j.saa.2017.12.013
- SumeraAnwer, F., Waseem, M., Fatima, A., Malik, N., Ali, A., Zahid, S., et al. (2022). Molecular docking and molecular dynamics studies reveal secretory proteins as novel targets of temozolomide in glioblastoma multiforme. *Molecules* 27, 7198. doi:10.3390/molecules27217198
- Sun, Y.-S., Zhao, Z., Yang, Z.-N., Xu, F., Lu, H.-J., Zhu, Z.-Y., et al. (2017). Risk factors and preventions of breast cancer. *Int. J. Biol. Sci.* 13, 1387–1397. doi:10.7150/ijbs.21635
- Türker, L., Gümüş, S., and Atalar, T. (2010). A DFT study on nitro derivatives of pyridine. *J. Energetic Mater.* 28, 139–171. doi:10.1080/07370650903273224
- Verma, A., Joshi, S., and Singh, D. (2013). Imidazole: Having versatile biological activities. *J. Chem.* 2013, 1–12. doi:10.1155/2013/329412

Frontiers in Chemistry

Explores all fields of chemical science across the periodic table

Advances our understanding of how atoms, ions, and molecules come together and come apart. It explores the role of chemistry in our everyday lives - from electronic devices to health and wellbeing.

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

