frontiers | Research Topics

Drug repurposing and polypharmacology: A synergistic approach in multi-target based drug discovery

Edited by
Mithun Rudrapal, Rudra Pangeni and Keshav Raj Paudel

Published in
Frontiers in Pharmacology

ISSN 1664-8714
ISBN 978-2-83251-262-3
DOI 10.3389/978-2-83251-262-3

About Frontiers

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

Frontiers journal series

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

Dedication to quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews. Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the *Frontiers journals series*: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area.

Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers editorial office: frontiersin.org/about/contact

Drug repurposing and polypharmacology: A synergistic approach in multi-target based drug discovery

Topic editors

Mithun Rudrapal — Rasiklal M. Dhariwal Institute of Pharmaceutical Education and Research, India

Rudra Pangeni — Virginia Commonwealth University, United States

Keshav Raj Paudel — University of Technology Sydney, Australia

Citation

Rudrapal, M., Pangeni, R., Paudel, K. R., eds. (2023). *Drug repurposing and polypharmacology: A synergistic approach in multi-target based drug discovery*. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-83251-262-3

Table of contents

frontiers | Frontiers in Pharmacology

TYPE Editorial
PUBLISHED 20 December 2022
DOI 10.3389/fphar.2022.1101007

OPEN ACCESS

EDITED AND REVIEWED BY
Brian Godman,
University of Strathclyde,
United Kingdom

*CORRESPONDENCE
Mithun Rudrapal,
rsmrpal@gmail.com

SPECIALTY SECTION
This article was submitted to Drugs Outcomes Research and Policies, a section of the journal Frontiers in Pharmacology

RECEIVED 17 November 2022
ACCEPTED 13 December 2022
PUBLISHED 20 December 2022

CITATION
Rudrapal M, Paudel KR and Pangeni R (2022), Editorial: Drug repurposing and polypharmacology: A synergistic approach in multi-target based drug discovery.
Front. Pharmacol. 13:1101007.
doi: 10.3389/fphar.2022.1101007

Editorial: Drug repurposing and polypharmacology: A synergistic approach in multi-target based drug discovery

Mithun Rudrapal[1]*, Keshav Raj Paudel[2] and Rudra Pangeni[3]

[1]Department of Pharmaceutical Chemistry, Rasiklal M. Dhariwal Institute of Pharmaceutical Education & Research, Pune, India, [2]Department of Oriental Medicine Resources, Mokpo National University, Muan-gun, South Korea, [3]Virginia Commonwealth University, Richmond, VA, United States

KEYWORDS

drug repurposing, polypharmacology, multi-targeting, drug discovery, cancer, COVID-19

Editorial on the Research Topic
Drug repurposing and polypharmacology: A synergistic approach in multi-target based drug discovery

Drug repurposing (also called drug repositioning) is a process of identifying new therapeutic uses for approved and/or existing drugs for treating common, difficult-to-treat and rare diseases (Paul et al., 2022; Rudrapal et al., 2022). On the other hand, polypharmacology (or multi-targeting approach) involves the interactions of drug molecules with multiple targets of different therapeutic indications/diseases (Jamir et al., 2022).

Drug repurposing is increasingly becoming an attractive strategy worldwide as it involves lower risk, potentially reduced expenditure and shorter development timelines as compared to *de novo* drug discovery (Rudrapal et al., 2020). Rising scenarios of deadly diseases (cancer, cardiovascular illness, diabetes, infectious diseases, COVID-19) largely affect the lives of millions of people, and thereby it impose a heavy economic burden globally (Singh et al., 2020). Currently available (or FDA approved) drugs are inadequate to manage a majority of such diseases, and, therefore, there is an urgent need for new drug candidates and/or drug therapy. Drugs with multi-targeting (polypharmacology approach) potential are immensely interesting in repurposing, because this dual synergistic strategy could offer better therapeutic alternative and useful clinical candidates (Pinzi et al., 2021).

The Research Topic *"Drug repurposing and polypharmacology: A synergistic approach in multi-target based drug discovery"* was aimed to compile latest research ideas, directions, developments and advances focusing on the theme of the topic within the scope of the journal. The topic was led by three Guest Editors listed above who are experts in the subject and oversaw the entire editorial process for the submitted papers. A total of ten articles were published, including seven original research and three review articles.

In a study, Xie et al. reported that lenvatinib when combined with the PD-1 inhibitor could effectively treat patients with advanced intrahepatic cholangiocarcinoma (ICC). They concluded that this combination therapy could be a safe and better alternative option for the treatment of advanced ICC.

A review article by Liu et al. highlighted the development of novel antiviral compounds targeting the S protein of SARS-CoV-2 through screening of natural products and drug repurposing approaches. This study provided insights into the discovery of promising drug candidates from natural sources as possible anti-SARS-CoV-2 agents.

Another study reported by Yang et al. utilized a molecular docking protocol to screen out potential inhibitors targeting the main protease (M^{pro}) of SARS-CoV-2. This study resulted in five compounds (namely, N-1H-Indazol-5-yl-2-(6-methylpyridin-2-yl)quinazolin-4- amine, ergotamine, antrafenine, dihydroergotamine and phthalocyanine) as potential drug candidates to be developed for clinical trials. Further, molecular dynamics (MD) simulations confirmed that potential inhibitory effect of the five identified compounds against SARS-CoV-2 M^{pro}.

Mangione et al. investigated upon Computational Analysis of Novel Drug Opportunities (CANDO) platform to identify small molecule inhibitors against COVID-19 on the basis of multiscale therapeutic, repurposing and design approaches. Interestingly, 51 of their 276 predictions demonstrated anti-SARS-CoV-2 potential according to published reports (clinical and experimental), suggesting the ability of CANDO platform in multi-target based drug discovery.

In another study, Khan et al. investigated molecular targets and pathways of nitazoxanide as novel approaches for the treatment of hepatocellular carcinoma (HCC) by using molecular docking and network pharmacology approaches. Authors proposed that distinct therapeutic effect for nitazoxanide is possible in treating HCC, with well-defined pharmacological targets and molecular pathways.

Sun et al. represented a bibliometric analysis of publications on drug repurposing for 10 years (2010–2020), which included 2,978 of publications. Their findings reported that the United States leads in drug repurposing research, followed by China, the United Kingdom, and India. From keyword analysis, they also reported that the hotspots have been changed in recent years, with COVID-19/SARS-CoV-2/coronavirus being the most prominent topic(s) in the domain of drug discovery.

A study by Kusuma et al. proposed an approach to implement bipartite graph search optimization using the branch and bound algorithm to identify the combination or composition of Jamu formulas. In addition, the proposed method comprising one to four selected plant species for the T2DM Jamu formula was suggested by the researchers.

Cao et al. predicted the mechanism of action of licorice in the treatment of COVID-19 through an extensive computational analysis using bioinformatics tools and molecular dynamics simulation. Authors reported that phytochemicals (phaseol, glycyrol, glyasperin F) present in licorice could act against COVID-19 through the inhibition of STAT3, IL2RA, MMP1 and CXCL8.

Wang et al. investigated the target-specific compound selectivity for multi-target drug discovery and repurposing by experimental studies. Authors represented several case studies exhibiting target-specific selectivity, which could facilitate the repurposing drugs by multi-targeting approach.

A review article by Kakoti et al. summarized therapeutic drug repositioning approaches for neurodegenerative diseases with recent threats and issues. They demonstrated the neuroprotective effect of kinase inhibitors, which, however, were originally developed for oncological indications. Authors also highlighted several opportunities and challenges of drug repurposing approaches in the way of drug discovery despites many technological advancements.

In conclusion, this Research Topic has provided in-depth insights into newer research findings (experimental, computational and review reports) and latest updates including technological advancements and challenges) related to ongoing repurposing strategies and drug discovery research in various therapeutic areas of current interest. Though drug repurposing strategies have several potentials as already indicated above, it has many challenges in the process of drug discovery, whether from a scientific or regulatory perspectives. Critical evaluations of pre-clinical, clinical and observational data/evidences are required to investigate the therapeutic efficacy and safety/toxicity of a candidate drug for potential repurposing.

Author contributions

All authors worked as a team of editors for the Research Topic, and approved this editorial for publication.

Acknowledgments

Authors would like to thank all the authors and reviewers for their invaluable and timely contributions to this Research Topic.

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

References

Jamir, E., Sarma, H., Priyadarsinee, L., Nagamani, S., Kiewhuo, K., Gaur, A. S., et al. (2022). Applying polypharmacology approach for drug repurposing for SARS-CoV2. *J. Chem. Sci.* 134, 57–24. doi:10.1007/s12039-022-02046-0

Paul, A., Kumar, M., Das, P., Guha, N., Rudrapal, M., and Zaman, M. K. (2022). Drug repurposing–A search for novel therapy for the treatment of diabetic neuropathy. *Biomed. Pharmacother.* 156, 113846. doi:10.1016/j.biopha.2022.113846

Pinzi, L., Tinivella, A., Caporuscio, F., and Rastelli, G. (2021). Drug repurposing and polypharmacology to fight SARS-CoV-2 through inhibition of the main protease. *Front. Pharmacol.* 12, 636989. doi:10.3389/fphar.2021.636989

Rudrapal, M., Gogoi, N., Chetia, D., Khan, J., Banwas, S., Alshehri, B., et al. (2022). Repurposing of phytomedicine-derived bioactive compounds with promising anti-SARS-CoV-2 potential: Molecular docking, MD simulation and drug-likeness/ADMET studies. *Saudi J. Biol. Sci.* 29, 2432–2446. doi:10.1016/j.sjbs.2021.12.018

Rudrapal, M., Khairnar, S. J., and Jadhav, A. G. (2020). "Drug repurposing (DR): An emerging approach in drug discovery," in *Drug repurposing – hypothesis, molecular aspects and therapeutic applications*. Editor A. Dekebo (London: IntechOpen), 1–20.

Singh, T. U., Parida, S., Lingaraju, M. C., Kesavan, M., Kumar, D., and Singh, R. K. (2020). Drug repurposing approach to fight COVID-19. *Pharmacol. Rep.* 72, 1479–1508. doi:10.1007/s43440-020-00155-6

frontiers | Frontiers in Pharmacology

ORIGINAL RESEARCH
published: 01 June 2022
doi: 10.3389/fphar.2022.894407

Lenvatinib Combined With a PD-1 Inhibitor as Effective Therapy for Advanced Intrahepatic Cholangiocarcinoma

Lulu Xie[1†], Jingzheng Huang[1†], Linling Wang[1], Wenrui Ren[1], Hao Tian[1], Anhong Hu[1], Jun Liang[1], Yuqing Jiao[1], Yali Li[1], Qunfang Zhou[1] and Wenjing Zhang[2*]*

[1]Department of Minimally Invasive Interventional Radiology and Department of Radiology, the Second Affiliated Hospital of Guangzhou Medical University, Guangzhou, China, [2]Department of Ultrasound, the Second Affiliated Hospital of Guangzhou Medical University, Guangzhou, China

Background: Lenvatinib combined with a PD-1 inhibitor has obtained a satisfactory antitumor effect in several solid tumors. However, the efficacy and tumor response of lenvatinib with a PD-1 inhibitor in advanced intrahepatic cholangiocarcinoma still need further exploration.

Methods: This is a single-arm study for the assessment of the efficacy and tolerability of lenvatinib with a PD-1 inhibitor in intrahepatic cholangiocarcinoma patients who had chemotherapy failure. Efficacy was evaluated based on the Response Evaluation Criteria in Solid Tumors RECIST Version 1.1 (RECIST 1.1).

Results: A total of 40 patients with advanced intrahepatic cholangiocarcinoma were enrolled after the chemorefractory effect. The median progression-free survival was 5.83 ± 0.76 months. The 3-month and 6-month progression-free survival rates were 80.0% and 32.5%, respectively. The median overall survival was 14.30 ± 1.30 months. The 12-month and 18-month overall survival rates were 61.4% and 34.7%. The 3-month RECIST 1.1 evaluation was that seven patients (17.5%) showed partial response, 23 patients (57.5%) had stable disease, and 10 patients (25.0%) had progressive disease. The objective response rate was 17.5%, and the disease control rate was 75.0%. All the recorded any-grade adverse events inducing treatment termination were controllable, and there were no AE-related deaths.

Conclusion: Our study showed that a combination of lenvatinib with the PD-1 inhibitor could be an effective treatment for advanced intrahepatic cholangiocarcinoma after the chemorefractory effect.

Keywords: advanced intrahepatic cholangiocarcinoma, lenvatinib, PD-1 inhibitor, combination therapy, chemotherapy failure

OPEN ACCESS

Edited by:
Mithun Rudrapal, Rasiklal M. Dhariwal Institute of Pharmaceutical Education and Research, India

Reviewed by:
André Mauricio De Oliveira, Federal Center for Technological Education of Minas Gerais, Brazil
Prashanta Kumar Deb, Birla Institute of Technology, Mesra, India

***Correspondence:**
Qunfang Zhou
zhouqun988509@163.com
Wenjing Zhang
ven0929@163.com

†These authors have contributed equally to this work

Specialty section:
This article was submitted to Drugs Outcomes Research and Policies, a section of the journal Frontiers in Pharmacology

Received: *11 March 2022*
Accepted: *19 April 2022*
Published: *01 June 2022*

Citation:
Xie L, Huang J, Wang L, Ren W, Tian H, Hu A, Liang J, Jiao Y, Li Y, Zhou Q and Zhang W (2022) Lenvatinib Combined With a PD-1 Inhibitor as Effective Therapy for Advanced Intrahepatic Cholangiocarcinoma. Front. Pharmacol. 13:894407. doi: 10.3389/fphar.2022.894407

INTRODUCTION

Intrahepatic cholangiocarcinoma (ICC) is a hepatobiliary tumor with a high death rate, which presents an unsatisfied prognosis with 10% of 5-year overall survival (OS) and a median OS of approximately 24 months (Kelley et al., 2020). ICC ranks second and accounts for approximately 10% of primary liver malignancy (Chun and Javle, 2017). The symptoms of ICC are insidious and nonspecific which include abdominal discomfort, weight loss, indigestion, or asymptomatic elevation of liver functions on routine laboratory testing, and only a minority of patients are diagnosed at an early stage with the tumor removed by surgery (Esnaola et al., 2016). Therefore, most patients present in the advanced stage will require effective systemic therapy. Currently, the established first-line treatment was gemcitabine and cisplatin, and the second-line treatment was FOLFOX systemic chemotherapy (Rizvi et al., 2018). However, the efficacy of these approaches is still unsatisfactory, and patients easily develop the chemorefractory effect (Moeini et al., 2016). There is now no uniform therapy for advanced ICC after chemotherapy failure. The shortage of available therapeutic regimens has plagued the oncologists exploring new strategies.

Lenvatinib is an oral tyrosine kinase inhibitor that restrains the vascular endothelial growth factor receptor (VEGFR) 1–3, fibroblast growth factor receptors (FGFR) 1–4, and platelet-derived growth factor receptor (Kudo et al., 2018). Due to the advantage of inhibiting tumors with multiple pathways, this multitargeted tyrosine kinase inhibitor is being used for the treatment of many tumors (Hao and Wang, 2020). Immunotherapy has emerged as a major tool in cancer treatment with the recent success of trials with PD-1/PD-L1 axis blockade (Balar and Weber, 2017). Programmed death-1 (PD-1) is a checkpoint molecule on T cells, which plays a vital role in controlling tumor progression through immune responses (Balar and Weber, 2017). Studies have proved that a combination of therapies involving lenvatinib and the PD-1 inhibitor could produce a synergetic effect, and lenvatinib with the PD-1 inhibitor has an augment antitumor effect than alone (Kimura et al., 2018). This combination of lenvatinib and the PD-1 inhibitor now has been applied for the treatment of many cancers including hepatocellular carcinoma, renal cell carcinoma, thyroid cancer, and endometrial carcinoma (Motzer et al., 2015; Schlumberger et al., 2015; Finn et al., 2020; Makker et al., 2020). The combination of lenvatinib and a PD-1 inhibitor is efficacious and promising, and the combination is considered to be a good pair of active drugs in malignancy therapy (Wang et al., 2019).

Given these factors, this combination could be an effective treatment for advanced ICC and prolong the survival of patients. Lin J et al. reported that lenvatinib with pembrolizumab was promising in alternative patients with refractory bile tract carcinoma, and the therapeutic outcomes were delightful as a non–first-line treatment (Lin et al., 2020). Ding Y et al. reported that chemotherapy, tislelizumab, and lenvatinib could be an effective therapeutic regimen for preoperative advanced intrahepatic ICC conversion therapy (Ding et al., 2021). These studies have inspired clinical investigations of applying the regimen in patients with advanced ICC. However, studies that reported lenvatinib with a PD-1 inhibitor on advanced ICC are few, so there is still a necessity for clinical evidence to further obtain knowledge of this combination therapy. In this report, we focused on lenvatinib combined with a PD-1 inhibitor in patients with advanced ICC after chemotherapy failure.

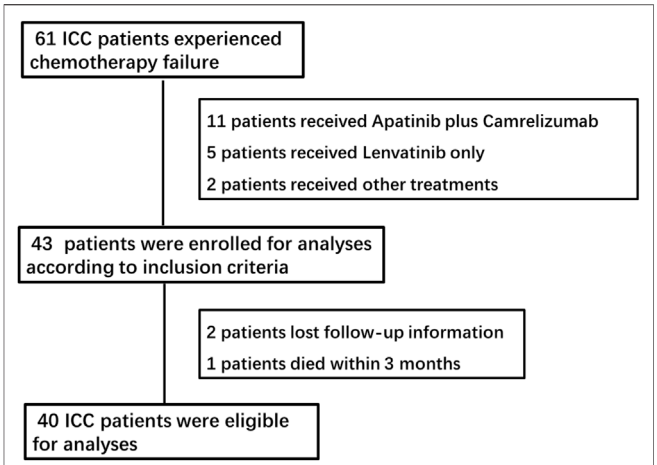


FIGURE 1 | Flow chart of the study. Efficacy and safety of the treatment.

MATERIALS AND METHODS

This study was conducted in accordance with the principles of the Declaration of Helsinki (World Medical Association, 2013), and the study protocol was approved by the Ethics Committee of the Second Affiliated Hospital of Guangzhou Medical University (no.2022-hg-ks-12).

Study Population

We retrospectively reviewed the medical records of patients who received a diagnosis of advanced ICC from June 2018 to June 2020 at the Second Affiliated Hospital of Guangzhou Medical University. Patients who met the following criteria were included in this study: 1) histologically confirmed ICC; 2) all patients experienced disease progression or could not tolerate systematic therapy; 3) at least one measurable tumor lesion according to the RECIST 1.1 criteria; 4) Eastern Cooperative Oncology Group (ECOG) performance status of 0–1; 5) patients who had adequate liver function (i.e., Child–Pugh class A or B liver function); 6) had adequate renal coagulation function; and 7) age 18–75 years. The exclusion criteria were as follows: 1) patient intolerance to lenvatinib or the PD-1 inhibitor; 2) death or missed the follow-up within 3 months; 3) inadequate liver or kidney function; and 4) patients who received other tyrosine kinase inhibitor with or without PD-1 inhibitor.

The tumor stage was assessed by systemic imaging (either enhanced computed tomography (CT) of the chest or bone scan, contrast-enhanced CT or magnetic resonance imaging (MRI) of the abdomen or brain, or positron emission tomography/computed tomography (PET/CT). Baseline levels of liver function and blood tests were collected. The albumin–bilirubin (ALBI) grade for each patient was calculated using the formula: ALBI score = (log10 bilirubin × 0.66) + (albumin × −0.085). The ALBI grade is used to identify different mortality risk subsets of patients as follows: grade 1 (lowest mortality risk) for ALBI score ≤ −2.60, grade 2 (intermediate

TABLE 1 | Baseline characteristics of patients in the entire cohort.

Parameter	Total
Age, years (median, IQR)	53.0 (43.0–58.8)
Gender, n [%] Female Male	9 [22.5] 31 [77.5]
Differentiation Moderate Poor	11 [27.5] 29 [72.5]
ECOG status 0 1	12 [30.0] 28 [70.0]
ALBI grade 1 2	31 [77.5] 9 [22.5]
Hepatitis	16 [40.0]
Metastasis, n [%] Intrahepatic Lymph nodes Lungs Bone	28 [70.0] 26 [65.0] 17 [42.5] 8 [20.0]
Previous therapy Surgery Systemic chemotherapy Thermal ablation Transarterial chemoembolization	13 [32.5] 40 [100.0] 8 [20.0] 14 [35.0]
Macrovascular tumor thrombus	10 [25.0]
Tumor size, cm, (median, IQR)	6.7 [4.9–8.2]
CA-199, U/ml, (median, IQR)	13.2 [13.2–299.7]

TABLE 2 | Therapeutic efficacy of the response and survival outcome of patients treated with lenvatinib with the PD-1 inhibitor.

Therapeutic response assessment	Evaluation of patients (*n* = 40)
Objective response rate (ORR, %)	7 (17.5%)
Disease control rate (DCR, %)	30 (75.0%)
Complete response (CR, %)	0
Partial response (PR, %)	7 (17.5%)
Stable disease (SD, %)	23 (57.5%)
Progressive disease (PD, %)	10 (25.0%)
Clinical benefit rate (%)	15 (32.5%)
Progression-free survival (median, 95% CI, months)	4.83 ± 0.68 (3.49–6.18)
Overall survival (median, 95% CI, months)	14.30 ± 1.30 (11.76–16.84)

mortality risk) for ALBI score > −2.60 and ≤ −1.39), and grade 3 (highest mortality risk) for ALBI score > −1.39 (Hiraoka et al., 2019).

Treatment and Assessment of the Response

All patients accepted contrast material–enhanced CT or MRI within 2 weeks before lenvatinib administration. Information regarding the

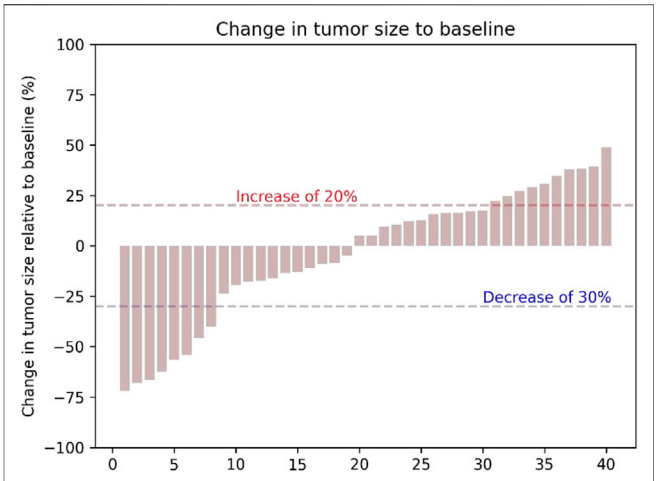


FIGURE 2 | Tumor percentage changes from the baseline in terms of the target lesion sizes.

information of initiation, completion of treatment, initial dose, dose modifications, and adverse events (AEs) during treatment was systematically collected. The prescription dosage of lenvatinib was 12 mg (for patients with a bodyweight ≥60 kg) or 8 mg (for patients with a bodyweight <60 kg) orally once a day. For the PD-1 inhibitor, the PD-1 inhibitor (tislelizumab) dose was applied according to the drug instructions.

Follow-Up

The follow-up period for this study was terminated on 30 June 2021. Laboratory tests including CA-125, albumin, bilirubin, aspartate transaminase (AST), alanine transaminase (ALT), and prothrombin time (PT) were performed to evaluate the treatment response and liver function every six weeks after treatment. Patients were evaluated at least once every six weeks after treatment. Each follow-up visit involved performing screening abdominal imaging (e.g., abdominal, chest, bone, brain CT, and/or MRI). Target tumors were selected to a maximum of two lesions per organ and five lesions in total. The minimum size for measurability is greater than 1 cm. The tumor imaging response was evaluated according to the Response Evaluation Criteria in Solid Tumors version 1.1 (Schwartz et al., 2016). In brief, the complete response (CR) was defined as the disappearance of arterial enhancement in the tumor. Partial response (PR) was defined as ≥30% shrinking in the diameter of the targeted tumors. Progressive disease (PD) was defined as at least a 20% increase in the sum of the diameter of the targeted tumors or the appearance of a new lesion. Stable disease (SD) neither met the CR nor PR and PD. The primary endpoint for the study was overall survival (OS), and the secondary endpoint was progression-free survival (PFS). OS was defined as the time from accepting lenvatinib and the PD-1 inhibitor to death or the last follow-up, and the PFS was defined as the time from the date of accepting lenvatinib and the PD-1 inhibitor to tumor progression or the last follow-up.

Statistical Analysis

The data were presented as a summary of the baseline characteristics, therapeutic efficacies, and AEs. The 3- and 6-

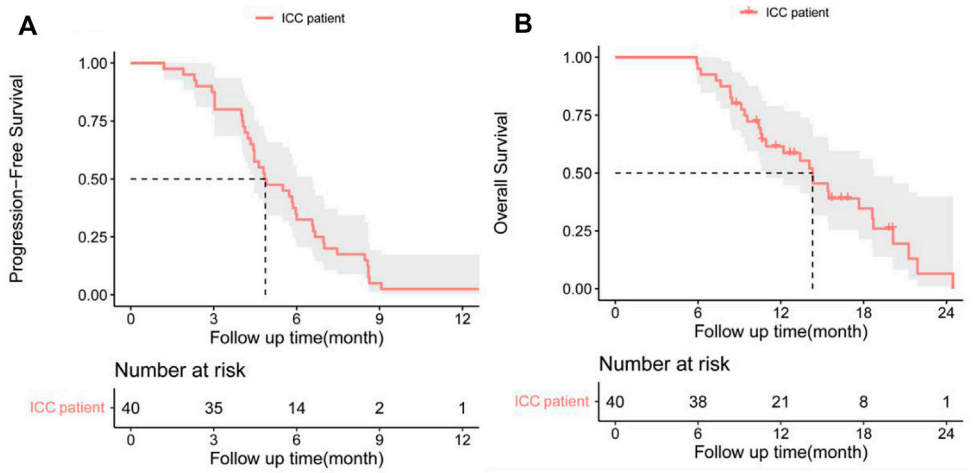


FIGURE 3 | (A) Progression-free survival (PFS) and **(B)** overall survival (OS) in patients with advanced intrahepatic cholangiocarcinoma after the chemorefractory effect.

TABLE 3 | Most common treatment-related adverse events in patients receiving lenvatinib and the PD-1 inhibitor.

Adverse events	Grades 1 and 2	Grades 3 and 4
Decreased appetite	19 (47.5)	0
Hypertension	18 (45.0)	3 (7.5)
Fatigue	13 (32.5)	1 (2.5)
Diarrhea	11 (27.5)	0
Increased ALT/AST	9 (22.5)	1 (2.5)
Proteinuria	9 (22.5)	2 (5.0)
Hypothyroidism	7 (17.5)	0
Rash or desquamation	7 (17.5)	0
Weight decreased	5 (12.5)	0

month PFS and 6-, 12-, and 18-month OS were all estimated by the Kaplan–Meier method. The hazard ratio (HR) of each clinical factor was estimated by Cox proportional hazard modeling.

RESULTS

Patient Characteristics

A total of 61 patients were enrolled for drug administration, and 40 patients were included for analyses (**Figure 1**). The median patient age was 53.0 years (range, 43.0–58.8), and 31 patients (77.5%) were males, and nine patients (22.5%) were females. In total, 28 (70.0%) patients had an ECOG performance status of 1, and 16 patients (40.0%) had HBV infection. All patients with hepatitis received regular antiviral therapy during lenvatinib and PD-1 inhibitor treatment. A total of 30 (72.5%) patients had poor tumor differentiation, and 32 (80.0%) patients had metastases, including intrahepatic, lymph nodes, lung, and bone metastases. A total of 25 patients (62.5%) received local therapy or surgery before lenvatinib and PD-1 inhibitor treatment (**Table 1**).

In the cohort, all patients had a regular follow-up, and the clinical responses were assessed. Overall, 19 of the 40 (47.5%) patients exhibited a decrease in the tumor size from the baseline (**Figure 2**). The median progression-free survival (PFS) was 5.83 ± 0.76 (95% CI, 4.34–7.33) months (**Table 2**). The 3-month and 6-month PFS rates were 80% and 32.5% (**Figure 3A**), respectively. In total, 11 patients were still alive during the follow-up period. The median overall survival (OS) was 14.30 ± 1.30 (95%: 11.76–16.84) months (**Table 2**). The 12-month and 18-month OS rates were 61.4% and 34.7% (**Figure 3B**). The 3-month RECIST 1.1 evaluation was that seven (17.5%) patients showed partial response (PR), 23 (57.5%) had stable disease (SD), and 10 (25.0%) had progressive disease (PD). The objective response rate (ORR) was 17.5%, and the disease control rate (DCR) was 75.0% (**Table 2**). We further determined the clinical benefit rate (CBR, PFS≥ 6 months) in all assessment-available patients. The CBR was 32.5% (**Table 2**).

All the recorded any-grade adverse events (AEs) inducing treatment termination were controllable, and there were no AE-related deaths. The most common AEs (any grade) were decreased appetite, hypertension, fatigue, and diarrhea (**Table 3**). For AE grade ≥3, the most common were hypertension and proteinuria (**Table 3**). Most AEs occurring during treatment were safe and tolerated.

PFS and OS Analysis

Cox-regression analysis results regarding the prognostic factors of PFS or OS were further analyzed. Univariate analysis revealed that intrahepatic metastases, lymph nodes metastases, lung metastases, bone metastases, and ALBI grade 2 were correlated with poorer PFS. Multivariate analysis illustrated that intrahepatic metastasis (HR = 3.08, 95% CI: 1.23–7.67, and p = 0.016) and ALBI grade 2 (HR = 3.84, 95% CI: 1.42–10.03, and p = 0.005) were related to poorer PFS (**Table 4**). Univariate analysis revealed that intrahepatic metastases, lymph nodes metastases, lung metastases, and bone metastases were correlated with poorer OS. Multivariate analysis illustrated that

TABLE 4 | Univariate and multivariate analyses of prognostic factors on progression-free survival (PFS) in 40 patients with advanced ICC after chemotherapy failure.

Variable	Univariate and multivariate analyses				
	Comparison	HR (95% CI)	*P*	HR (95% CI)	*P*
Tumor differentiation	Moderate vs. poor	1.09 (0.53–2.25)	0.816		
Macrovascular invasion	No vs. yes	2.34 (0.42–1.82)	0.723		
Intrahepatic metastases	**No vs. yes**	**2.34 (1.15–4.7)**	**0.019**	**3.08 (1.23–7.67)**	**0.016**
Lymph nodes metastases	**No vs. yes**	**2.72 (1.35–5.48)**	**0.005**	1.95 (0.86–4.40)	0.109
Lung metastases	**No vs. yes**	**5.03 (2.13–11.84)**	**< 0.001**	1.97 (0.76–5.15)	0.165
Bone metastases	**No vs. yes**	**2.49 (1.06–5.85)**	**0.037**	1.47 (0.61–3.50)	0.390
ALBI grade	**1 vs. 2**	**3.94 (1.71–9.07)**	**0.001**	**3.84 (1.42–10.03)**	**0.005**
ECOG status	0 vs. 1	0.74 (0.37–1.47)	0.387		
Hepatitis	No vs. yes	1.42 (0.74–2.74)	0.289		
Surgery	Yes vs. no	1.35 (0.68–2.65)	0.388		
TACE	Yes vs. no	0.92 (0.48–1.79)	0.811		
Ablation	Yes vs. no	0.92 (0.42–2.02)	0.842		
Sex	Female vs. male	1.36 (0.64–2.89)	0.432		
Smoking	No vs. yes	1.22 (0.63–2.36)	0.564		

Bold values means the P<0.05.

TABLE 5 | Univariate and multivariate analyses of prognostic factors on overall survival (OS) in 40 patients with advanced ICC after chemotherapy failure.

Variable	Univariate and multivariate analyses				
	Comparison	HR (95% CI)	*P*	HR (95% CI)	*P*
Tumor differentiation	Moderate vs. poor	1.42 (0.62–3.25)	0.404		
Macrovascular invasion	No vs. yes	0.92 (0.40–2.11)	0.847		
Intrahepatic metastases	**No vs. yes**	**5.32 (2.01–14.15)**	**0.001**	**9.02 (1.80–45.07)**	**0.007**
Lymph node metastases	**No vs. yes**	**2.66 (1.11–6.38)**	**0.029**	1.67 (0.67–4.29)	0.269
Lung metastases	**No vs. yes**	**3.31 (1.40–7.83)**	**0.006**	1.90 (0.75–5.13)	0.173
ALBI grade	1 vs. 2	0.49 (0.22–1.11)	0.088		
ECOG status	0 vs. 1	0.94 (0.40–2.18)	0.876		
Bone metastases	No vs. yes	1.32 (0.49–3.58)	0.579		
Hepatitis	No vs. yes	1.98 (0.88–4.49)	0.101		
Surgery	Yes vs. no	1.46 (0.62–3.40)	0.386		
TACE	Yes vs. no	1.50 (0.70–3.22)	0.294		
Ablation	Yes vs. no	0.67 (0.26–1.72)	0.407		
Sex	Female vs. male	1.55 (0.72–3.31)	0.261		
Smoking	No vs. yes	0.92 (0.49–1.72)	0.784		

Bold values means the P<0.05.

intrahepatic metastasis (HR = 9.02, 95% CI: 1.80–45.07, and $p = 0.007$) was related to poorer OS (**Table 5**).

DISCUSSION

The therapeutic strategy for advanced ICC is challenging worldwide as the ICC usually indicates a poor prognosis. Surgical resection is the only potentially curative treatment for ICC; however, the 5-year OS rate was 15–40% (Weber et al., 2015). For advanced or recurrent ICC, the first-line treatment was chemotherapy. However, patients usually developed refractory; then, the second-line therapy was varied and disappointed (Sirica et al., 2019). PD-1 inhibitor-based immune therapy has obtained significant improvement in several tumors, including melanoma, lung cancer, and head and neck malignancies (Sui et al., 2018; Gavrielatou et al., 2020; Guo et al., 2020). However, immune monotherapy faces many challenges in biliary cancer. Studies showed that the efficacy of the PD-1 inhibitor alone in biliary cancer remains unsatisfactory (Ueno et al., 2019). In the Makoto et al. study, combined therapy of biliary tract cancer (nivolumab PD-1 inhibitor and chemotherapy) achieved obvious better benefits than the PD-1 inhibitor alone (Ueno et al., 2019).

Combining the strategy with antiangiogenic molecular target drugs or chemotherapy could improve the efficacy of immunotherapies and has shown promising clinical results (Rizvi et al., 2018; Wang et al., 2019). Mei K et al. have reported that camrelizumab combined with apatinib has achieved promising results in the treatment of advanced ICC. The medium PFS and OS were 1.9 and 13.4 months (Mei et al., 2021). These results were superior to the previously reported efficacy of apatinib alone in ICC (Hu et al., 2020). Lin J et al. reported that lenvatinib with pembrolizumab was considered a non-first-line therapy in treating refractory bile tract carcinoma, and this study obtained ORR which was 25%, and the DCR was 78.1%. The median PFS was 4.9 months, and the 6-month PFS rate was 33.7%. The median OS was 11.0 months, and the 1-year OS rate was 39.4% (Lin et al., 2020).

In our study, the ORR was 17.5%, and the DCR was 75.0%. The median PFS and 6-month PFS rates were 5.8 months and 32.5%. The median OS was 14.3 months, and the 12-month and 18-month rates were 61.4%. Our results were in accordance with the research of Lin et al. (2020).

Ueno M et al. demonstrated lenvatinib as monotherapy for advanced biliary tract cancer, and the ORR was 11.5%. The median PFS was 3.19 months, and the median OS was 7.35 months (Ueno et al., 2020). Our study was better than Ueno M's results. Lenvatinib combined with PD-1 inhibitors have provided new ideas for advanced ICC. Previous studies have proved that lenvatinib could enhance the antitumor efficacy of PD-1 inhibitors by restraining angiogenesis (Shigeta et al., 2020). Thus, the combination of these two agents is promising and satisfactory when they are used in patients with refractory ICC. Compared with the previous reports, our study focused on ICC with chemotherapy failure, and patients were in a more advanced stage than the published literature. We found that the combination therapy of lenvatinib and the PD-1 inhibitor was effective and competent for ICC patients in a more advanced stage. A well-designed prospective trial with other second-line treatments is needed to determine the precise efficacy and safety of this combination therapy, or a further trial of this combined regimen with chemotherapy as first-line therapy is promising.

However, there were some limitations to our study; first, it was a retrospective study, so there is a need to develop a prospective trial to evaluate this combination therapy as second-line therapy in advanced ICC. Second, the small sample of our study limits more information on factors related to the prognosis; further reports with more patients and a multicenter study are needed to get more comprehensive results. Third, this study was the real-world application of lenvatinib and the PD-1 inhibitor, and it is impossible to exclude interference from the doctor and patients in terms of treatment.

In summary, our research provides evidence that a combination of lenvatinib with the PD-1 inhibitor could be an effective treatment for ICC after the chemorefractory effect. This combination could achieve controllable safety and good efficacy, thereby providing a new treatment option for advanced ICC.

DATA AVAILABILITY STATEMENT

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

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Ethics Committee of the Second Affiliated Hospital of Guangzhou Medical University. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

Conceptualization: LX and JH; data curation: QZ, LW, JH, and WZ; formal analysis: QZ, WR, and HT; data analysis: LX and QZ. Funding acquisition: QZ; investigation: LX, JH, and AH; methodology: JL, YJ, and YL; project administration: QZ and WZ; resources: QZ, JH, and WZ; original draft: LX and JH; writing—review and editing: LX, QZ, and WZ.

FUNDING

This research was funded by the National Natural Science Foundation of China (82102082).

REFERENCES

Balar, A. V., and Weber, J. S. (2017). PD-1 and PD-L1 Antibodies in Cancer: Current Status and Future Directions. *Cancer Immunol. Immunother.* 66, 551–564. doi:10.1007/s00262-017-1954-6

Chun, Y. S., and Javle, M. (2017). Systemic and Adjuvant Therapies for Intrahepatic Cholangiocarcinoma. *Cancer Control* 24, 1073274817729241. doi:10.1177/1073274817729241

Ding, Y., Han, X., Sun, Z., Tang, J., Wu, Y., and Wang, W. (2021). Systemic Sequential Therapy of CisGem, Tislelizumab, and Lenvatinib for Advanced Intrahepatic Cholangiocarcinoma Conversion Therapy. *Front. Oncol.* 11, 691380. doi:10.3389/fonc.2021.691380

Esnaola, N. F., Meyer, J. E., Karachristos, A., Maranki, J. L., Camp, E. R., and Denlinger, C. S. (2016). Evaluation and Management of Intrahepatic and Extrahepatic Cholangiocarcinoma. *Cancer* 122, 1349–1369. doi:10.1002/cncr.29692

Finn, R. S., Ikeda, M., Zhu, A. X., Sung, M. W., Baron, A. D., Kudo, M., et al. (2020). Phase Ib Study of Lenvatinib Plus Pembrolizumab in Patients with Unresectable Hepatocellular Carcinoma. *J. Clin. Oncol.* 38, 2960–2970. doi:10.1200/JCO.20.00808

Gavrielatou, N., Doumas, S., Economopoulou, P., Foukas, P. G., and Psyrri, A. (2020). Biomarkers for Immunotherapy Response in Head and Neck Cancer. *Cancer Treat. Rev.* 84, 101977. doi:10.1016/j.ctrv.2020.101977

Guo, W., Ma, J., Guo, S., Wang, H., Wang, S., Shi, Q., et al. (2020). A20 Regulates the Therapeutic Effect of Anti-PD-1 Immunotherapy in Melanoma. *J. Immunother. Cancer* 8, 8. doi:10.1136/jitc-2020-001866

Hao, Z., and Wang, P. (2020). Lenvatinib in Management of Solid Tumors. *Oncologist* 25, e302–e310. doi:10.1634/theoncologist.2019-0407

Hiraoka, A., Kumada, T., Michitaka, K., and Kudo, M. (2019). Newly Proposed ALBI Grade and ALBI-T Score as Tools for Assessment of Hepatic Function and Prognosis in Hepatocellular Carcinoma Patients. *Liver Cancer* 8, 312–325. doi:10.1159/000494844

Hu, Y., Lin, H., Hao, M., Zhou, Y., Chen, Q., and Chen, Z. (2020). Efficacy and Safety of Apatinib in Treatment of Unresectable Intrahepatic Cholangiocarcinoma: An Observational Study. *Cancer Manag. Res.* 12, 5345–5351. doi:10.2147/CMAR.S254955

Kelley, R. K., Bridgewater, J., Gores, G. J., and Zhu, A. X. (2020). Systemic Therapies for Intrahepatic Cholangiocarcinoma. *J. Hepatol.* 72, 353–363. doi:10.1016/j.jhep.2019.10.009

Kimura, T., Kato, Y., Ozawa, Y., Kodama, K., Ito, J., Ichikawa, K., et al. (2018). Immunomodulatory Activity of Lenvatinib Contributes to Antitumor Activity in the Hepa1-6 Hepatocellular Carcinoma Model. *Cancer Sci.* 109, 3993–4002. doi:10.1111/cas.13806

Kudo, M., Finn, R. S., Qin, S., Han, K. H., Ikeda, K., Piscaglia, F., et al. (2018). Lenvatinib versus Sorafenib in First-Line Treatment of Patients with Unresectable Hepatocellular Carcinoma: A Randomised Phase 3 Non-inferiority Trial. *Lancet* 391, 1163–1173. doi:10.1016/S0140-6736(18)30207-1

Lin, J., Yang, X., Long, J., Zhao, S., Mao, J., Wang, D., et al. (2020). Pembrolizumab Combined with Lenvatinib as Non-first-line Therapy in Patients with Refractory Biliary Tract Carcinoma. *Hepatobiliary Surg. Nutr.* 9, 414–424. doi:10.21037/hbsn-20-338

Makker, V., Taylor, M. H., Aghajanian, C., Oaknin, A., Mier, J., Cohn, A. L., et al. (2020). Lenvatinib Plus Pembrolizumab in Patients with Advanced Endometrial Cancer. *J. Clin. Oncol.* 38, 2981–2992. doi:10.1200/JCO.19.02627

Mei, K., Qin, S., Chen, Z., Liu, Y., Wang, L., and Zou, J. (2021). Camrelizumab in Combination with Apatinib in Second-Line or above Therapy for Advanced Primary Liver Cancer: Cohort A Report in a Multicenter Phase Ib/II Trial. *J. Immunother. Cancer* 9, e002191. doi:10.1136/jitc-2020-002191

Moeini, A., Sia, D., Bardeesy, N., Mazzaferro, V., and Llovet, J. M. (2016). Molecular Pathogenesis and Targeted Therapies for Intrahepatic Cholangiocarcinoma. *Clin. Cancer Res.* 22, 291–300. doi:10.1158/1078-0432.CCR-14-3296

Motzer, R. J., Hutson, T. E., Glen, H., Michaelson, M. D., Molina, A., Eisen, T., et al. (2015). Lenvatinib, Everolimus, and the Combination in Patients with Metastatic Renal Cell Carcinoma: A Randomised, Phase 2, Open-Label, Multicentre Trial. *Lancet Oncol.* 16, 1473–1482. doi:10.1016/S1470-2045(15)00290-9

Rizvi, S., Khan, S. A., Hallemeier, C. L., Kelley, R. K., and Gores, G. J. (2018). Cholangiocarcinoma - Evolving Concepts and Therapeutic Strategies. *Nat. Rev. Clin. Oncol.* 15, 95–111. doi:10.1038/nrclinonc.2017.157

Schlumberger, M., Tahara, M., Wirth, L. J., Robinson, B., Brose, M. S., Elisei, R., et al. (2015). Lenvatinib versus Placebo in Radioiodine-Refractory Thyroid Cancer. *N. Engl. J. Med.* 372, 621–630. doi:10.1056/NEJMoa1406470

Schwartz, L. H., Litière, S., de Vries, E., Ford, R., Gwyther, S., Mandrekar, S., et al. (2016). RECIST 1.1-Update and Clarification: From the RECIST Committee. *Eur. J. Cancer* 62, 132–137. doi:10.1016/j.ejca.2016.03.081

Shigeta, K., Datta, M., Hato, T., Kitahara, S., Chen, I. X., Matsui, A., et al. (2020). Dual Programmed Death Receptor-1 and Vascular Endothelial Growth Factor Receptor-2 Blockade Promotes Vascular Normalization and Enhances Antitumor Immune Responses in Hepatocellular Carcinoma. *Hepatology* 71, 1247–1261. doi:10.1002/hep.30889

Sirica, A. E., Gores, G. J., Groopman, J. D., Selaru, F. M., Strazzabosco, M., Wei Wang, X., et al. (2019). Intrahepatic Cholangiocarcinoma: Continuing Challenges and Translational Advances. *Hepatology* 69, 1803–1815. doi:10.1002/hep.30289

Sui, H., Ma, N., Wang, Y., Li, H., Liu, X., Su, Y., et al. (2018). Anti-PD-1/PD-L1 Therapy for Non-small-cell Lung Cancer: Toward Personalized Medicine and Combination Strategies. *J. Immunol. Res.* 2018, 6984948. doi:10.1155/2018/6984948

Ueno, M., Ikeda, M., Morizane, C., Kobayashi, S., Ohno, I., Kondo, S., et al. (2019). Nivolumab Alone or in Combination with Cisplatin Plus Gemcitabine in Japanese Patients with Unresectable or Recurrent Biliary Tract Cancer: A Non-randomised, Multicentre, Open-Label, Phase 1 Study. *Lancet Gastroenterol. Hepatol.* 4, 611–621. doi:10.1016/S2468-1253(19)30086-X

Ueno, M., Ikeda, M., Sasaki, T., Nagashima, F., Mizuno, N., Shimizu, S., et al. (2020). Phase 2 Study of Lenvatinib Monotherapy as Second-Line Treatment in Unresectable Biliary Tract Cancer: Primary Analysis Results. *BMC Cancer* 20, 1105. doi:10.1186/s12885-020-07365-4

Wang, D., Lin, J., Yang, X., Long, J., Bai, Y., Yang, X., et al. (2019). Combination Regimens with PD-1/PD-L1 Immune Checkpoint Inhibitors for Gastrointestinal Malignancies. *J. Hematol. Oncol.* 12, 42. doi:10.1186/s13045-019-0730-9

Weber, S. M., Ribero, D., O'Reilly, E. M., Kokudo, N., Miyazaki, M., and Pawlik, T. M. (2015). Intrahepatic Cholangiocarcinoma: Expert Consensus Statement. *HPB (Oxford)* 17, 669–680. doi:10.1111/hpb.12441

World Medical Association (2015). World Medical Association Declaration of Helsinki: Ethical Principles for Medical Research Involving Human Subjects. *JAMA* 310, 2191–2194. doi:10.1001/jama.2013.281053

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.

frontiers | Frontiers in Pharmacology

TYPE Original Research
PUBLISHED 25 July 2022
DOI 10.3389/fphar.2022.968148

OPEN ACCESS

EDITED BY
Mithun Rudrapal,
Rasiklal M. Dhariwal Institute of Pharmaceutical Education and Research, India

REVIEWED BY
Samah Elaidy,
Suez Canal University, Egypt
Johra Khan,
Majmaah University, Saudi Arabia

*CORRESPONDENCE
Shakeel Ahmad Khan,
shakilahmad56@gmail.com
Terence Kin Wah Lee,
terence.kw.lee@polyu.edu.hk

SPECIALTY SECTION
This article was submitted to Gastrointestinal and Hepatic Pharmacology, a section of the journal Frontiers in Pharmacology

RECEIVED 16 June 2022
ACCEPTED 06 July 2022
PUBLISHED 25 July 2022

CITATION
Khan SA and Lee TKW (2022) Investigations of nitazoxanide molecular targets and pathways for the treatment of hepatocellular carcinoma using network pharmacology and molecular docking.
Front. Pharmacol. 13:968148.
doi: 10.3389/fphar.2022.968148

Investigations of nitazoxanide molecular targets and pathways for the treatment of hepatocellular carcinoma using network pharmacology and molecular docking

Shakeel Ahmad Khan[1]* and Terence Kin Wah Lee[1,2]*

[1]Department of Applied Biology and Chemical Technology, The Hong Kong Polytechnic University, Kowloon, Hong Kong SAR, China, [2]State Key Laboratory of Chemical Biology and Drug Discovery, The Hong Kong Polytechnic University, Kowloon, Hong Kong SAR, China

Nitazoxanide has been investigated for colorectal cancer and breast cancer. However, its molecular targets and pathways have not yet been explored for hepatocellular carcinoma (HCC) treatment. Utilizing a network pharmacology approach, nitazoxanide's potential targets and molecular pathways for HCC treatment were investigated. HCC targets were extracted from the GeneCards database. Potential targets of nitazoxanide were predicted using Swiss Target Prediction and Super Pred. Intersecting targets were analyzed with VENNY online tool. Using Cytoscape, a protein-protein interaction (PPI), cluster, and core targets-pathways networks were constructed. Using the Database for Annotation, Visualization and Integrated Discovery (DAVID), gene ontology (GO), and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analyses were conducted. The nitazoxanide was molecularly docked with anti-HCC core targets by employing Auto Dock Vina. A total of 168 potential targets of nitazoxanide, 13,415 HCC-related targets, and 153 intersecting targets were identified. The top eight anti-HCC core targets were identified: SRC, EGFR, CASP3, MMP9, mTOR, HIF1A, ERBB2, and PPARG. GO enrichment analysis showed that nitazoxanide might have anti-HCC effects by affecting gene targets involved in multiple biological processes (BP) (protein phosphorylation, transmembrane receptor protein tyrosine kinase (RTKs) signaling pathway, positive regulation of MAP kinase activity, etc.). KEGG pathways and core targets-pathways network analysis indicated that pathways in cancer and proteoglycans in cancer are two key pathways that significantly contribute to the anti-HCC effects of nitazoxanide. Results of molecular docking demonstrated the potential for active interaction between the top eight anti-HCC core targets and nitazoxanide. Our research offers a theoretical basis for the notion that nitazoxanide may have distinct therapeutic effects in HCC, and the identified pharmacological targets and pathways might function as biomarkers for HCC therapy.

KEYWORDS

nitazoxanide, network, pharmacology, molecular docking, hepatocellular carcinoma

Introduction

HCC is a kind of cancer that often affects people who have a history of hepatitis or cirrhosis. Owing to increasing malignancy and morbidity, it is the second leading cause of global cancer-related demise (Khan and Lee, 2022; Yang et al., 2022). Several efficient therapeutic approaches, including biological therapy, interventional radiology, chemotherapy, resection, tumor ablation, transcatheter arterial chemical embolization (TACE), liver transplantation, etc., have been extensively employed in treating HCC in recent decades (Stehlin et al., 1988; El-Serag, 2011; Balogh et al., 2016; Hilmi et al., 2020). HCC detection in patients at an early stage is critical and has significant importance because it is strongly linked to a patient's prognosis since interventional therapy delivered at a preliminary phase of HCC may significantly improve patient outcomes. Regrettably, patients are often identified with HCC at an intermediate or advanced stage, precluding resection and transplantation. Blood vessels' active invasion, resulting in extrahepatic and intrahepatic metastases, is linked to a poor prognosis after surgical or medical treatment because of a high recurrence rate (Dutta and Mahato, 2017; Mohs et al., 2017; Daher et al., 2018).

Moreover, chemotherapy treatment with sorafenib (tyrosine multi-kinase inhibitor) has also been identified as a promising therapeutic for an advanced stage of HCC. However, its treatment can only increase overall survival by about 3 months (Llovet et al., 2008; Yuan et al., 2022). On the contrary, numerous tyrosine multi-kinase inhibitors have been shown to be dangerous or to have no effect on patient survival (Cainap et al., 2015). Efforts have been undertaken to develop promising pharmacological solutions against HCC in the lack of adequate preventative or treatment methods in order to provide patients with alternative treatment choices and enhance patient life expectancies (Jindal et al., 2019).

In this instance, scientists are trying very hard to find effective therapeutics for the treatment of HCC by adopting a drug repurposing strategy compared to traditional drug designing and development owing to its several limitations, including failure in expensive late-stage clinical trials, high attrition rates, takes a long time, high cost, etc. (Pfab et al., 2021). Repurposing currently utilized drugs offers several advantages over developing an entirely new therapeutic (Zamboni et al., 2012; Gupta et al., 2013). In a drug repurposing strategy, the failure risk is lower in terms of safety, toxicity, and formulation information already available, therefore drastically reducing the costs necessary to get the medications to patients. In fact, bringing a repurposed drug to market is ten times less expensive than bringing a unique chemical molecule to market. Since data from clinical trials about pharmacokinetics, bioavailability, etc., is already available, the drug repurposing strategy also reduces the time to make a new drug (Zamboni et al., 2012; Gupta et al., 2013; Nosengo, 2016; Pushpakom et al., 2019; Pfab et al., 2021).

Nitazoxanide was first developed as an anthelmintic agent and is an antimicrobial agent authorized by the FDA (Sisson et al., 2002; Pfab et al., 2021). Müller et al. repurposed nitazoxanide for colorectal cancer (CRC), and it has shown anticancer activity by inhibiting apoptosis, DNA fragmentation, nuclear condensation, and cell proliferation. It specifically targeted glutathione-S-transferase P1 (GSTP1), activating the AMPK pathway while suppressing c-Myc, mTOR, and WNT signaling in CRC (Müller et al., 2008; Senkowski et al., 2015). Moreover, nitazoxanide has been reported to suppress c-Myc expression, leading to tumor growth suppression and apoptosis induction in breast cancer (Fan-Minogue et al., 2013). Nitazoxanide has not been explored for HCC treatment and could be expected to down regulate the overexpressed proteins implicated in the proliferation of HCC malignancy.

Network pharmacology, multi-omics data, molecular docking, and public medical databases have enabled an alternative computational drug discovery approach and are extensively used to design and develop therapeutic drugs for many cancer types (Khan and Lee, 2022; Yuan et al., 2022). Computational drug repurposing is particularly intriguing since it allows for quicker screening of candidate drugs than traditional drug design and development (Luo et al., 2021). Computational drug repurposing develops interactions between proteins, diseases, genes, and therapeutic candidates based on open-access databases. It suggests viable therapeutic candidates, assuming they target the same proteins in treating ailments (Keenan et al., 2018). Currently, several researchers are using them to design and develop therapeutic drug candidates and explore the molecular pathways of natural products implicated in the therapy of various ailments. Therefore, we have utilized different bioinformatics tools in this research, including network pharmacology and molecular docking, to computationally repurpose and identify nitazoxanide's targets and molecular pathways that could be involved in treating HCC. The flow chart of this research is presented in Figure 1.

Materials and methods

Targets prediction of nitazoxanide

The targets of nitazoxanide were predicted using Swiss Target Prediction (http://www.swisstargetprediction.ch/, accessed on 12 May 2022) and SuperPred (https://prediction.charite.de/, accessed on 12 May 2022) web servers with limitations to "Homo sapiens" (Nickel et al., 2014; Daina et al., 2019). The targets predicted with Swiss Target Prediction and Super Pred, which have a probability greater than zero and 50%, respectively, were selected as potential targets for nitazoxanide.

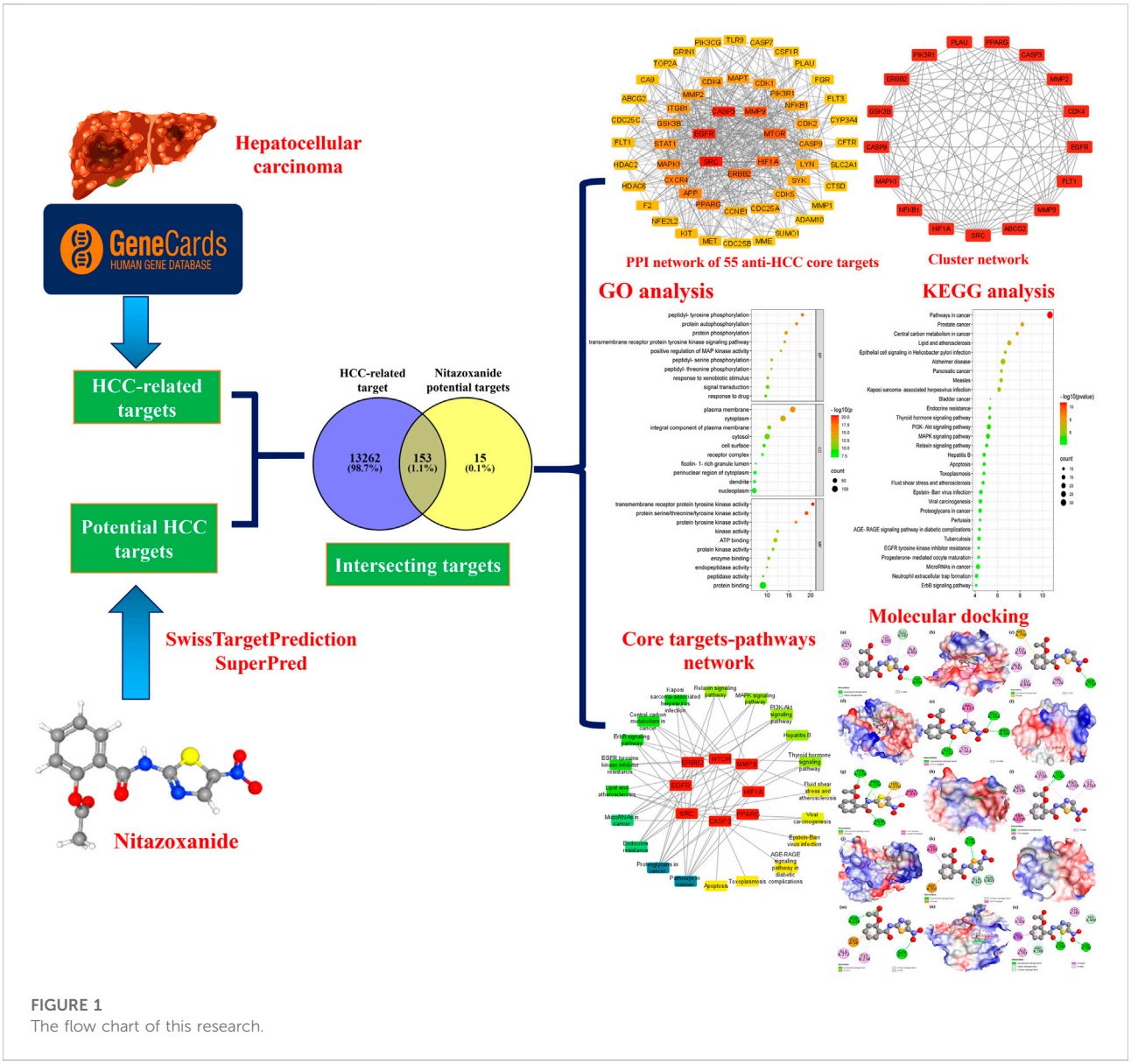


FIGURE 1
The flow chart of this research.

HCC-related targets determination

The HCC-related targets were identified by exploring the Gene Cards (https://www.genecards.org/, accessed on 12 May 2022) database for the terms "hepatic cancer, hepatic carcinoma, hepatocellular carcinoma, and hepatoma" (Rebhan et al., 1997).

Intersecting targets of HCC-related and potential targets of nitazoxanide

The intersecting targets between HCC-related and potential targets of nitazoxanide were identified using the VENNY 2.1 online tool (https://bioinfogp.cnb.csic.es/tools/venny/, accessed on 12 May 2022) (Venny, 2022). These identified intersected targets were screened for further analysis.

Protein-protein interaction analysis

The PPI analysis was performed on identified intersected targets by employing the STRING (https://string-db.org/, version 11.5, accessed on 12 May 2022) database with limitations to "Homo sapiens" and at a medium confidence score of 0.400 (von Mering et al., 2003). Moreover, their results were further explored by uploading them to Cytoscape software (version 3.9.0, Boston, MA, United States, accessed on 12 May 2022) to find the potential targets and anti-HCC core targets based on their

degrees in the network (Lopes et al., 2010). Moreover, cluster network analysis was carried out using the Molecular Complex Detection (MCODE) plugin of Cytoscape (version 3.9.0) by setting the parameters as; find clusters = in the whole network, degree cutoff = 2, node score cutoff = 0.2, K-score = 0.2, and max depth = 100.

GO and KEGG enrichment analysis

GO, and KEGG enrichment analyses were performed on identified intersected targets (determined in section 2.3) using the database for annotation, visualization, and integrated discovery (DAVID; version 6.8) (https://david.ncifcrf.gov/, accessed on 13 May 2022) (DAVID, 2022). Both analyses were performed by keeping the parameters: species, Homo sapiens; identifier, official gene symbol; gene list, list type; and remaining parameters, default values (Li et al., 2022). The results of GO enrichment analyses are comprised of three terms, including molecular functions (MF), cellular component (CC), and biological process (BP). The top 10 GO data (MF, BP, and CC) and 30 KEGG pathways were uploaded to the Bioinformatics platform (http://www.bioinformatics.com.cn/, accessed on 13 May 2022), and the results are displayed in the form of a bubble plot (Weishengxin, 2022). The enrichment of GO and pathways was deemed substantial if $p \leq 0.05$.

Network construction between anti-HCC core targets and pathways

The network was established between anti-HCC core targets and molecular pathways using Cytoscape software (version 3.9.0, Boston, MA, United States; accessed on 13 May 2022) to determine the intricate relationship between them in the treatment of HCC with nitazoxanide (Lopes et al., 2010).

Expression of anti-HCC core targets

The GEPIA database (http://gepia.cancer-pku.cn/, accessed on 14 May 2022) was used to examine the expression of the top eight anti-HCC core targets (determined in section 2.4.) in liver hepatocellular carcinoma (LIHC) (GEPIA, 2022 (Gene Expression Profiling Interactive Analysis)).

Molecular docking

The 2D chemical structure of nitazoxanide was retrieved from NCBI Pub Chem in Spatial Data File (SDF) (National Center for Biotechnology Information, 2022). 3D structure of nitazoxanide was constructed with BIOVIA Discovery Studio Visualizer 2021 and saved in PDB format (BIOVIA DS, 2016). The protein crystal structures of eight anti-HCC core targets were retrieved from Protein Data Bank (RCSB PDB: https://www.rcsb.org/search; RCSB PDB, 2022). The water molecules and ligands from protein crystal structures were extracted BIOVIA Discovery Studio Visualizer 2021. Moreover, this was employed to prepare the grid and add polar hydrogens to proteins. Each protein in PDB format was uploaded to AutoDock Vina (version 1.2.0.) and added the Kollman and Gasteiger partial charges. The PDB file of the 3D structure of nitazoxanide was then uploaded to AutoDock Vina. Proteins and nitazoxanide files were converted into pdbqt format using AutoDock Vina, and then they were utilized to write scripts for molecular docking (Trott and Olson, 2010). The docked complexes of proteins and nitazoxanide were obtained and further analyzed to determine their molecular interactions by employing BIOVIA Discovery Studio Visualizer 2021 (BIOVIA DS). Binding energy less than zero suggests that the ligand molecule may readily bind to the pockets of the targeted proteins. It is generally accepted that a lower binding energy value for a docked complex of ligand and receptor implies a stronger binding (Trott and Olson, 2010).

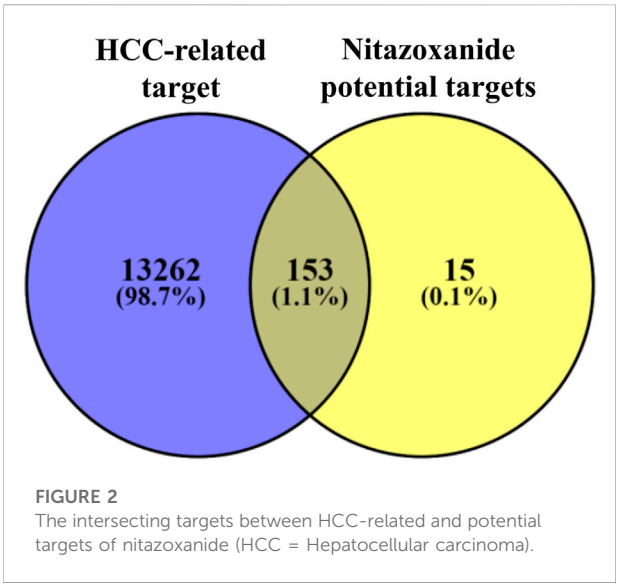


FIGURE 2
The intersecting targets between HCC-related and potential targets of nitazoxanide (HCC = Hepatocellular carcinoma).

Results

Potential targets of nitazoxanide

The nitazoxanide targets were predicted using the Swiss Target Prediction and Super Pred web servers with "Homo sapiens" limitations (Nickel et al., 2014; Daina et al., 2019). A

total of 168 potential targets were retrieved with a probability greater than zero and 50%.

HCC-related targets

By exploring the GeneCards database for the terms "hepatic cancer, hepatic carcinoma, hepatocellular carcinoma, and hepatoma," 13,415 HCC-related targets were retrieved (Rebhan et al., 1997).

Identification of intersecting targets

A total of 153 intersecting targets were identified between HCC-related and potential targets of nitazoxanide using the VENNY 2.1 online tool (Venny, 2022) (Figure 2). These identified intersecting targets were deemed as potential anti-HCC targets implicated in the treatment of HCC with nitazoxanide.

PPI network analysis

Intersecting targets (153) were uploaded to the STRING database with limitations to the species "Homo sapiens" (von Mering et al., 2003). A PPI network was obtained, which consisted of 153 nodes and 939 edges (Figure 3A). The average node degree in the network was 12.3. Moreover, the PPI network presented a 0.462 average local clustering coefficient and 406 expected number of edges. The STRING results of the PPI analysis were further imported to Cytoscape software (version 3.9.0) for better understanding and visualization of the network (Lopes et al., 2010).

The results demonstrated that the PPI network consists of 148 nodes (with the elimination of five disconnected nodes) and 939 edges (Figure 3B). The elimination of disconnected nodes from the network by the Cytoscape software (version 3.9.0) was also reported by Liu et al. (Liu et al., 2020). Moreover, network centralization, heterogeneity, density, diameter, and radius were 0.330, 0.937, 0.089, 6, and 3, respectively. The clustering coefficient, characteristics path

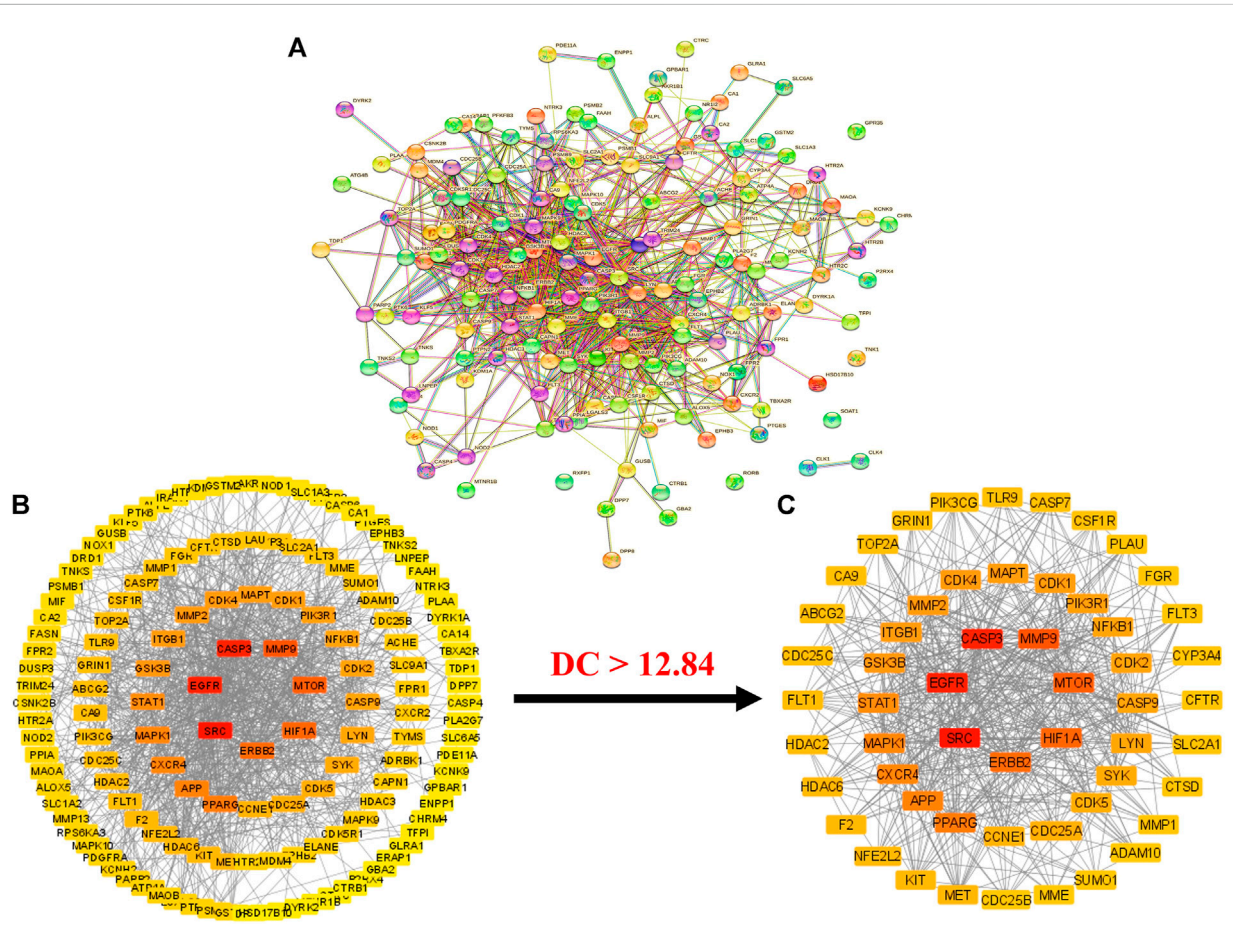


FIGURE 3
(A) STRIN PPI network **(B)** PPI network of intersecting targets **(C)** A hub network of 55 anti-HCC core targets. In both networks **(B,C)**, the transition from red to yellow represents the degree shift from highest to lowest for each node (PPI = Protein-protein interaction).

length, and the average number of neighbors were 0.409, 2.439, and 12.849, respectively.

Further, a hub network of targets with degrees greater than the average DC (12.84) was extracted and identified 55 potential HCC targets, which were classified as anti-HCC core targets (Figure 3C). The identified 55 anti-HCC core targets are presented in a bar graph based on their degree in the network, as shown in Figure 4. The top eight anti-HCC core targets are SRC (degree 60), EGFR (degree 58), CASP3 (degree 57), MMP9 (degree 47), mTOR (degree 45), HIF1A (degree 43), ERBB2 (degree 42), and PPARG (degree 38). These eight anti-HCC core targets were further investigated for molecular docking analysis with nitazoxanide.

Clusters network analysis

The cluster network analysis was further carried out on the constructed PPI network (Figure 3B) using the Molecular Complex Detection (MCODE) plugin of Cytoscape software (version 3.9.0). The PPI network was clustered into six clusters, as shown in Figure 5. Clusters 1, 2, 3, and four have 17 nodes and 97 edges, 16 nodes and 42 edges, 15 nodes and 25 edges, and five nodes and seven edges, respectively. While clusters five and six have three nodes and three edges. All of these clusters indicated the existence of potential HCC-targets for the nitazoxanide drug's therapeutic effects. Moreover, clusters one and two show the presence of the top eight anti-HCC core targets

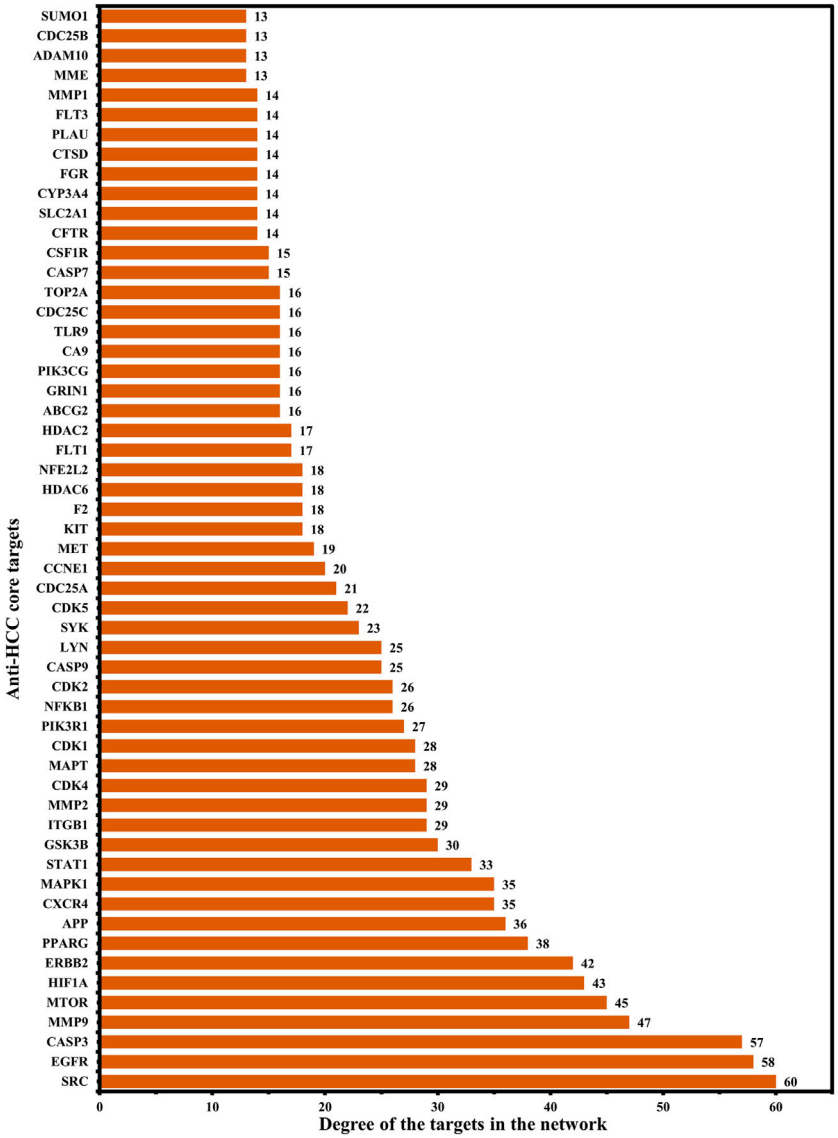


FIGURE 4
55 anti-HCC core targets in hub network ranked by DC > 12.84 (DC = Degree centrality).

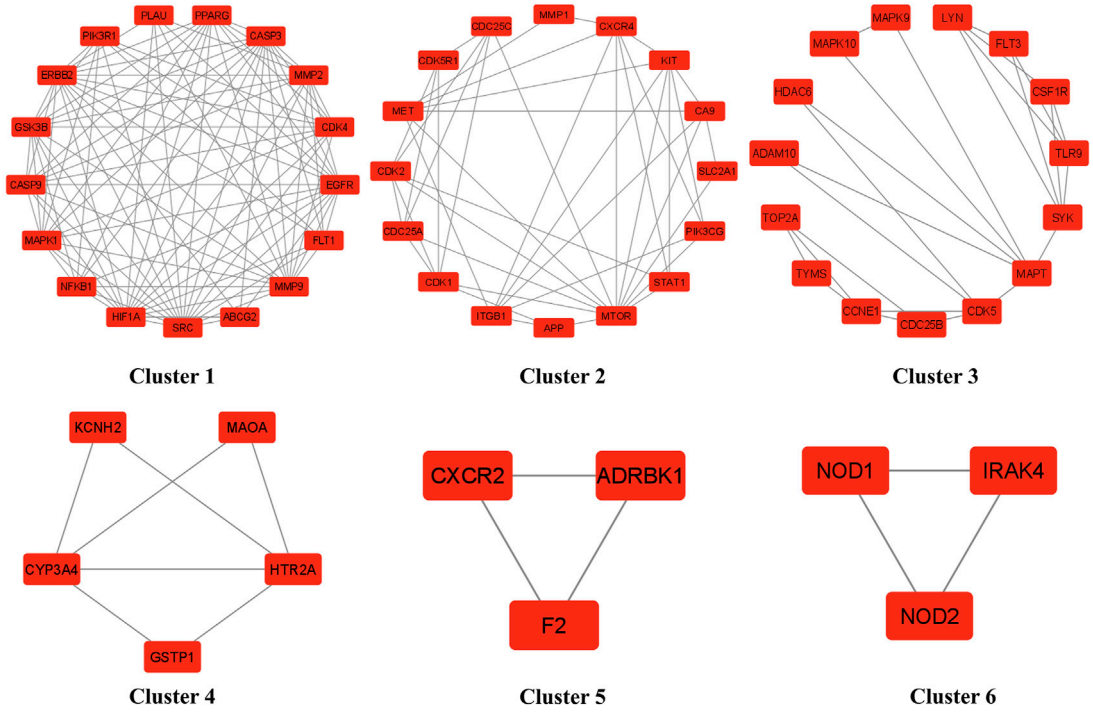


FIGURE 5
Six clusters of PPI network (PPI = Protein-protein interaction).

identified in section 3.4 (Figures 3C, 4). Cluster 1 has seven out of eight anti-HCC core targets: SRC, EGFR, CASP3, MMP9, HIF1A, ERBB2, and PPARG. On the other hand, cluster 2 has one out of eight anti-HCC core targets such as mTOR. Hence, cluster network analysis corroborated the findings of the hub network.

Expression of anti-HCC core targets in LIHC

The expression of the top eight anti-HCC core targets (SRC, EGFR, CASP3, MMP9, mTOR, HIF1A, ERBB2, and PPARG) in LIHC and normal samples were analyzed using the GEPIA database. The analysis results demonstrated that anti-HCC core targets were differentially expressed in LIHC and normal samples (Figure 6). These results corroborated that these eight anti-HCC core targets are strongly correlated to the development and progression of LIHC.

GO enrichment analysis

Anti-HCC effects of nitazoxanide drug were further investigated by performing GO enrichment analysis on 153 intersecting targets. The top 10 enriched GO terms (BP, MF, and CC) were identified. The results are presented in Figure 7. The targets attributed to the anti-HCC effects of nitazoxanide drug are implicated in multiple BP, which include peptidyl-tyrosine phosphorylation, protein autophosphorylation, protein phosphorylation, transmembrane receptor protein tyrosine kinase signaling pathway, positive regulation of MAP kinase activity, peptidyl-serine phosphorylation, etc. On the other hand, the targets implicated in the treatment of HCC with nitazoxanide drug are involved in multiple CC, including the plasma membrane, cytoplasm, integral components of the plasma membrane, cytosol, cell surface, etc. Moreover, results demonstrated that targets by which nitazoxanide treats HCC are implicated in multiple MF such as transmembrane receptor protein tyrosine kinase activity, proteins serine/threonine/tyrosine kinase activity, protein tyrosine kinase activity, kinase activity, ATP binding, etc.

KEGG enrichment analysis

The molecular mechanisms involved in the anti-HCC effects of nitazoxanide drug were further investigated by performing a KEGG pathway enrichment analysis on 153 intersecting targets. A total of 78 enriched KEGG pathways were identified at $p \leq 0.05$. The top 30 KEGG pathways were presented in Figure 8 in a bubble plot form. The molecular mechanism attributed to the anti-HCC effects of

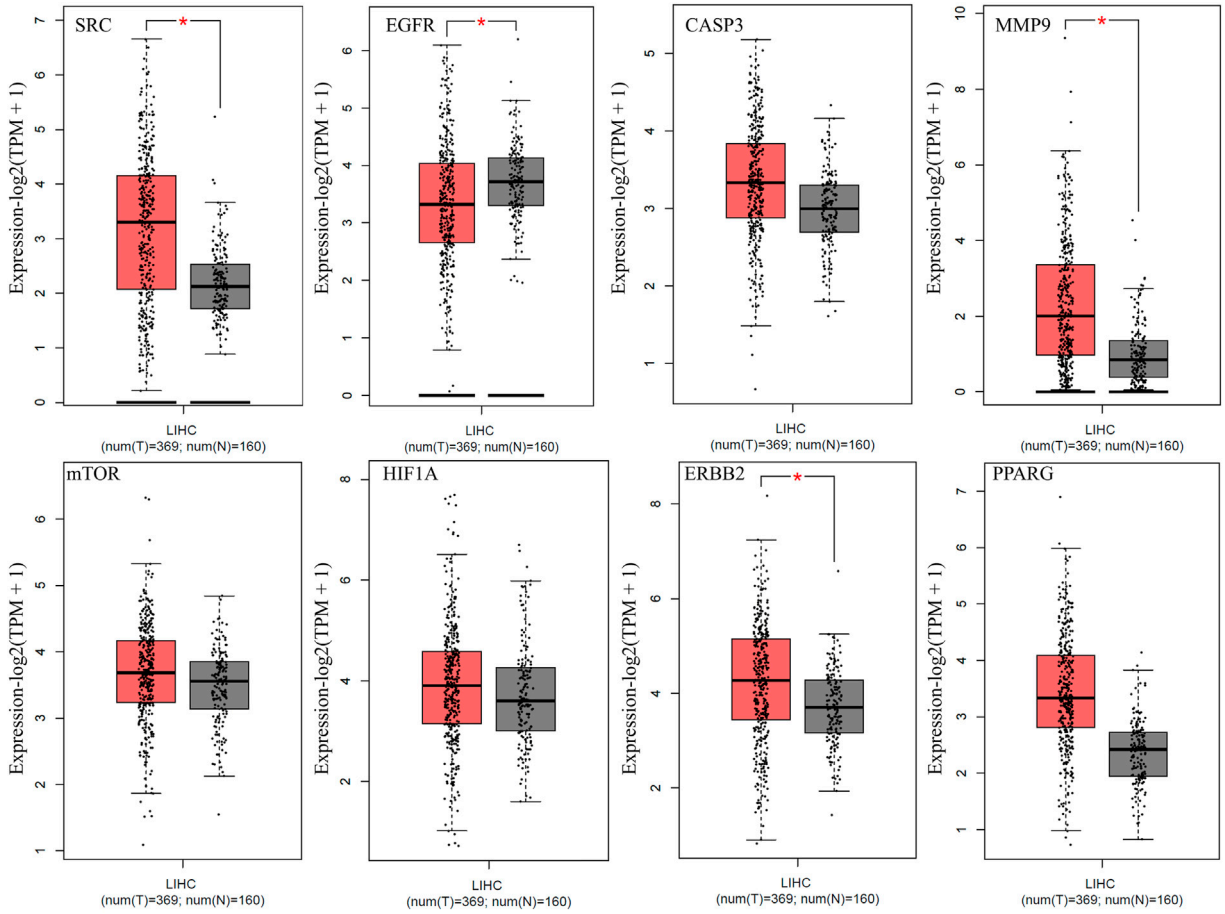


FIGURE 6
Expression of top eight anti-HCC core targets in LIHC (Red and grey colored boxes represent tumor and normal cells, respectively) (LIHC = Liver hepatocellular carcinoma).

nitazoxanide drug might be involved in pathways in cancer, PI3K-Akt signaling pathway, MAPK signaling pathway, proteoglycans in cancer, EGFR tyrosine kinase inhibitor resistance, apoptosis, hepatitis B, ErbB signaling pathway, microRNAs in cancer, etc. These findings suggest that all of these mechanisms may be implicated in a synergistic manner in the modulation of HCC by the nitazoxanide drug.

Network between anti-HCC core targets and pathways

To identify the major pathways involved in the anti-HCC effects of the nitazoxanide drug, a network between the top eight anti-HCC core targets and their corresponding pathways were constructed. The network results demonstrated that seven anti-HCC core targets (CASP3, EGFR, ERBB2, mTOR, MMP9, HIF1A, and PPARG) followed the pathways in cancer (degree 7). On the other hand, SRC, CASP3, EGFR, ERBB2, mTOR, MMP9, and HIF1A followed the proteoglycans in cancer (degree 7) (Figure 9).

Moreover, the pathways were further ranked by DC greater than the average DC (3.35) in the network to find the major pathways. A total of nine major pathways were identified and presented in a bar graph, as shown in Figure 10. Thus, these nine pathways may significantly contribute to the anti-HCC effects of the nitazoxanide drug by modulating the expression of anti-HCC core targets.

Molecular docking

The nitazoxanide drug was molecularly docked with the top eight anti-HCC core targets (SRC, EGFR, CASP3, MMP9, mTOR, HIF1A, ERBB2, and PPARG), and the findings are

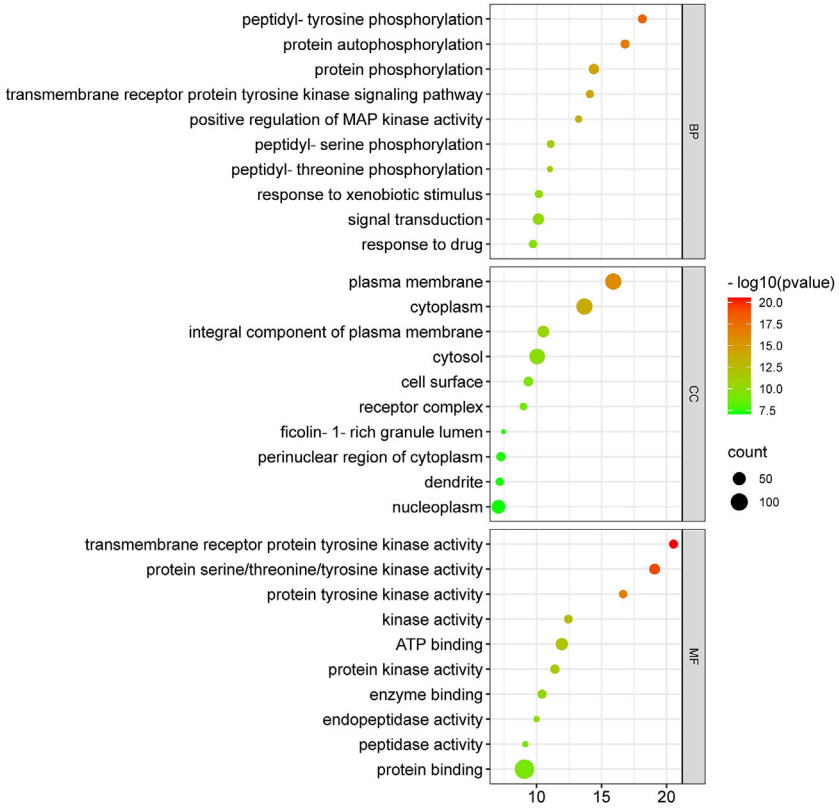


FIGURE 7
Top 10 GO enriched analysis of 153 intersecting targets involved in anti-HCC effects of nitazoxanide drug (GO = Gene ontology).

shown in Table 1. Figures 11A–O depicts docked complexes of the nitazoxanide drug and anti-HCC core targets. The results revealed that the nitazoxanide drug had a high affinity for all anti-HCC core targets. However, nitazoxanide drug had a greater binding affinity with three anti-HCC core targets (SRC, MMP9, and PPARG) and had an energy score ≥ −7.0. On the other hand, the nitazoxanide drug had a high binding affinity for mTOR, EGFR, and CASP3. Furthermore, the nitazoxanide drug had a modest binding affinity for HIF1A and ERBB2, yielding an energy score of −5.1.

Discussion

HCC often afflicts individuals with a history of hepatitis or cirrhosis. Owing to increasing malignancy and morbidity, it is the second leading cause of global cancer-related demise (Khan and Lee, 2022; Yang et al., 2022). In the lack of viable HCC preventive or therapeutic interventions, emerging trends have shifted toward drug repurposing instead of conventional drug discovery and development due to the latter's many constraints (Pfab et al., 2021). The emergence of the big data era and the growth of bioinformatics approaches provide tremendous assistance for drug discovery *via* network pharmacology (Vetrivel et al., 2021). The core principle of network pharmacology is that prospective targets may be predicted by looking at their biological pathways from a network perspective. This may aid in discovering novel active medications from medicinal compounds (Zhu et al., 2019; Vetrivel et al., 2021). In this study, a network pharmacology approach was utilized to evaluate the therapeutic mechanism of nitazoxanide as a treatment for HCC. For the first time, nitazoxanide's pharmacological effects on HCC have been examined utilizing network pharmacology and molecular docking simulations. Therapeutic targets of nitazoxanide against HCC were predicted using online databases. A total of 168 potential therapeutic targets were identified. A total of 13,415 HCC-related targets were retrieved from the online database. Furthermore, 153 intersecting targets were identified between the potential targets of nitazoxanide and HCC-related targets.

PPI and cluster network analysis of intersecting targets displayed that multiple genes such as SRC, EGFR, CASP3, MMP9, mTOR, HIF1A, ERBB2, and PPARG are implicated in the anti-HCC effects of nitazoxanide. The report demonstrates that elevated expression of SRC leads to the pathogenesis of HCC and subsequent metastasis

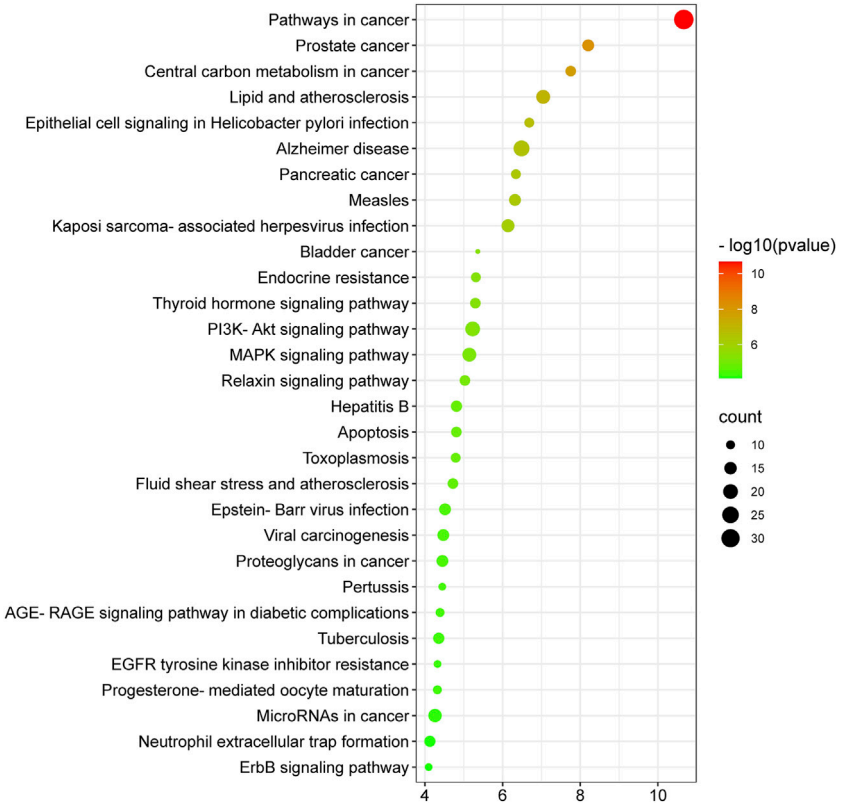


FIGURE 8
Top thirty KEGG enriched pathways of 153 intersecting targets involved in anti-HCC effects of nitazoxanide drug.

(Zhao et al., 2015). Overexpression of EGFR has been implicated in HCC pathogenesis, and activation of this receptor contributes to HCC cells' primary resistance to sorafenib (Sueangoen et al., 2020). Persad et al. reported the implication of CASP3 overexpression in the pathogenesis of HCC (Persad et al., 2004). Previous reports demonstrate that MMP9 is an oncogene implicated in HCC progression (Yan et al., 2013; Lu et al., 2015). Its higher expression in HCC tissues was also reported by Liu et al. (Liu B. et al., 2021). The mTOR signaling is involved in numerous cancer hallmarks such as cell growth, apoptosis suppression, etc. In HCC tissue samples, the mTOR pathway is more highly expressed than in liver cirrhotic tissue in the general vicinity (Ferrín et al., 2020). Reports show that HIF1A protein levels are considerably higher in human HCC samples and are linked with a poorer prognosis (Chen and Lou, 2017). In the last 3 decades, research indicated that ERbb2 expression is seldom associated with the development of HCC (Xian et al., 2005; Shi et al., 2019). However, alternative studies also revealed that ERBB2 expression was higher in 30–40 percent of HCC (Heinze et al., 1999; Shi et al., 2019). Moreover, GEPIA database analysis shows that all eight anti-HCC core targets were differentially expressed in LIHC and normal samples (Figure 6). Hence, based on the literature and the GEPIA database, all eight of these anti-HCC core targets played a crucial role in the progression of HCC and may be promising therapeutic targets for treating HCC with nitazoxanide.

The GO enrichment analysis demonstrated that nitazoxanide might be displayed anti-HCC effects by affecting gene targets implicated in multiple BP (peptidyl-tyrosine phosphorylation, protein autophosphorylation, protein phosphorylation, transmembrane receptor protein tyrosine kinase (RTKs) signaling pathway, positive regulation of MAP kinase activity, peptidyl-serine phosphorylation, etc.). Protein phosphorylation is vital for performing various activities such as biological processes, cellular localization, etc.; however, its aberrant regulation contributes to several conditions such as HCC, etc. (Liu Y. et al., 2021). MAP kinase regulates various cellular functions (apoptosis, proliferation, differentiation, etc.). MAP kinase activity is upregulated in the majority of malignancies with a high incidence rate, such as pancreatic cancer, non-small cell lung cancer, and particularly HCC (Li et al., 2021). RTKs are membrane-bound receptors necessary for cell function. By phosphorylating intracellular substrate proteins, they promote communication between cells. They govern cell proliferation, differentiation, metabolism, migration, etc., to maintain cellular homeostasis and are at the

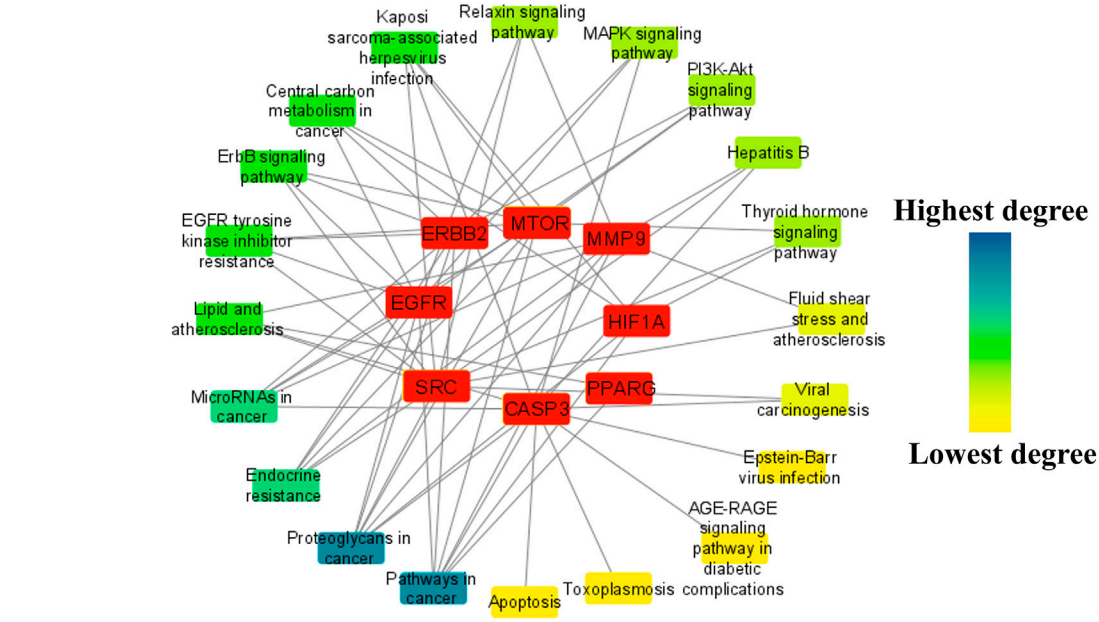


FIGURE 9
The network between the top eight anti-HCC core targets and their corresponding molecular pathways.

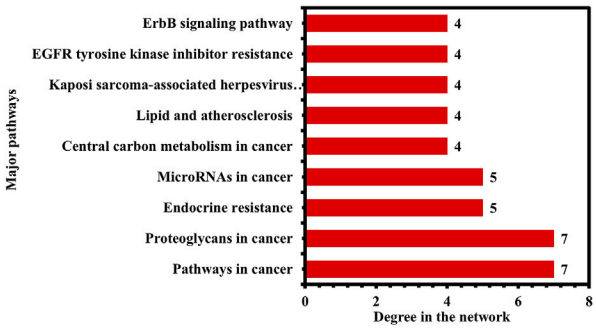


FIGURE 10
Nine pathways in the network ranked by DC > 3.35 (DC = Degree centrality).

hub of intricate signaling networks. RTK mutations or aberrant activation are common causes of the development of cancers, including HCC (Sudhesh Dev et al., 2021). In addition, the targets implicated in the treatment of HCC with the nitazoxanide drug are involved in multiple CC, including the plasma membrane, cytoplasm, integral components of the plasma membrane, cytosol, cell surface, etc. Moreover, the results demonstrated that the targets by which nitazoxanide treats HCC are implicated in multiple MF, including transmembrane receptor protein tyrosine kinase activity, protein serine/threonine/tyrosine kinase activity, protein tyrosine kinase activity, kinase activity, ATP binding, etc.

The KEGG enrichment analysis revealed that the molecular pathways underlying the anti-HCC effects of the nitazoxanide drug might involve pathways in cancer, PI3K-Akt signaling pathway, MAPK signaling pathway, proteoglycans in cancer, EGFR tyrosine kinase inhibitor resistance, apoptosis, hepatitis B, ErbB signaling pathway, microRNAs in cancer, etc. In spite of the fact that the PI3K–AKT signaling pathway regulates a wide range of cellular processes, its abnormal activation promotes the development of HCC (Rahmani et al., 2020). Consequently, PI3K-AKT suppression may be an alternative HCC therapeutic modality. RTKs trigger activation of the MAPK signaling pathway. However, its inappropriate modulation leads to abnormal cellular activity,

TABLE 1 Molecular docking of nitazoxanide with the top eight anti-HCC core targets.

Drug	Binding affinity (kcal/mol)							
	SRC	EGFR	CASP3	MMP9	mTOR	HIF1A	ERBB2	PPARG
Nitazoxanide	−7.0	−5.8	−5.5	−7.9	−6.9	−5.1	−5.1	−7.4

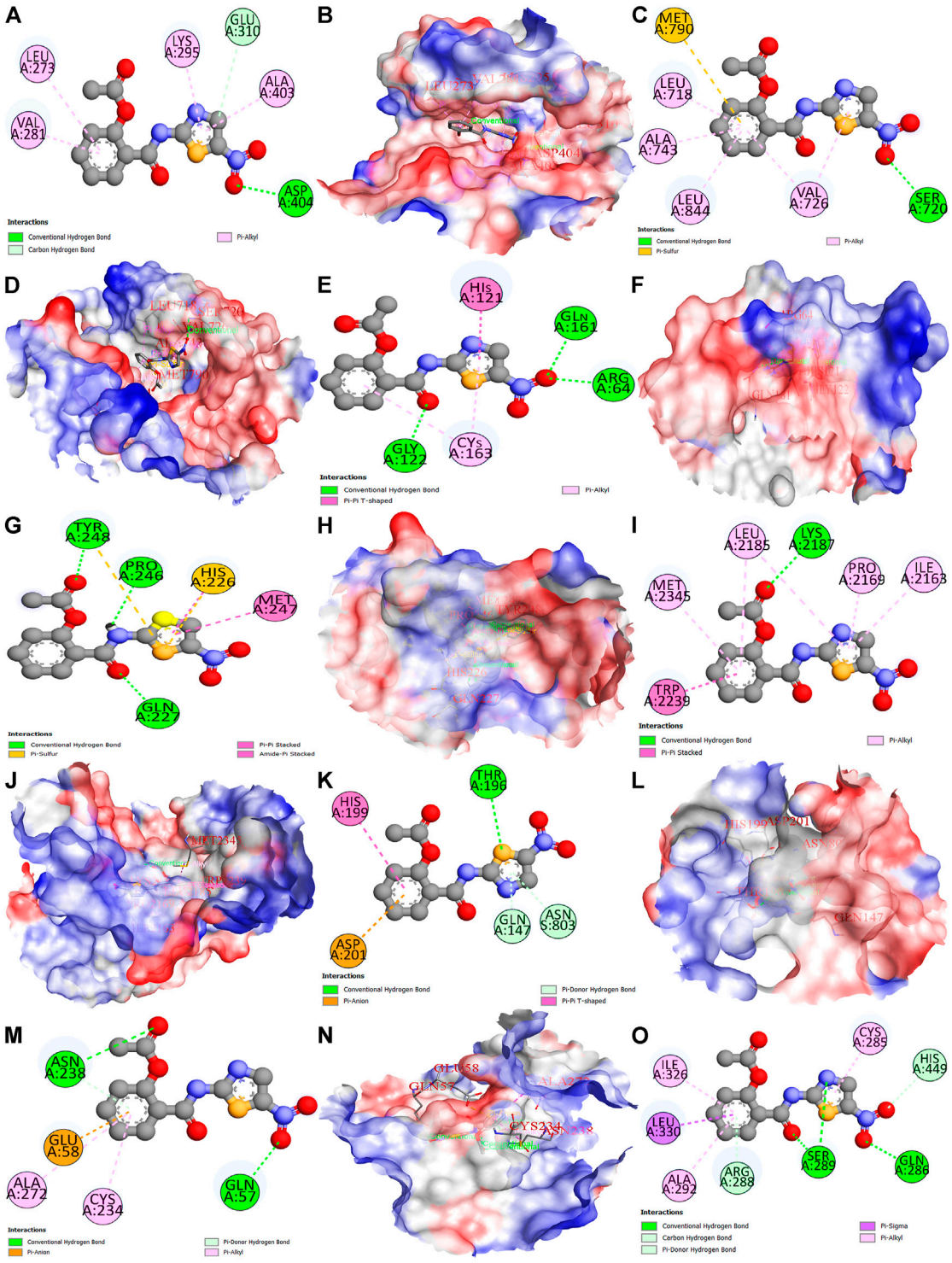


FIGURE 11
Molecular docking results of top eight anti-HCC core targets with nitazoxanide drug. Nitazoxanide drug binds with SRC (2D and 3D) **(A,B)**, EGFR (2D and 3D) **(C,D)**, CASP3 (2D and 3D) **(E,F)**, MMP9 (2D and 3D) **(G,H)**, mTOR (2D and 3D) **(I,J)**, HIF1A (2D and 3D) **(K,L)**, ERBB2 (2D and 3D) **(M,N)**, and PPARG (2D) **(O)**.

including enhanced cell growth and proliferation, dedifferentiation, and survival, which are all implicated in the etiology of malignancies, including HCC (Delire and Stärkel, 2015; Moon and Ro, 2021). The upregulation of proteoglycans such as glypican-3 leads to the development of melanoma. However, HCC patients had the highest proportion of positive instances (Ahrens et al., 2020). Therefore, inhibiting proteoglycans may be a potential treatment option for treating HCC. Moreover, the network results of core targets and pathways demonstrated nine major pathways (Figure 10). Among nine major pathways, the top two pathways were pathways in cancer (degree 7) and proteoglycans in cancer (degree 7). Seven anti-HCC core targets (CASP3, EGFR, ERBB2, mTOR, MMP9, HIF1A, and PPARG) followed the pathways in cancer. On the other hand, SRC, CASP3, EGFR, ERBB2, mTOR, MMP9, and HIF1A followed the proteoglycans in cancer (Figure 9). Thus, these two pathways may significantly contribute to the anti-HCC effects of the nitazoxanide drug by modulating the expression of anti-HCC core targets.

To further validate the results, nitazoxanide's impact on the eight anti-HCC core targets (SRC, EGFR, CASP3, MMP9, mTOR, HIF1A, ERBB2, and PPARG) was investigated in silico. The targets' molecular interactions with nitazoxanide demonstrated efficient binding in docking studies. Together, our findings show that patients with HCC have elevated levels of transcriptional expression of the expected targets and that therapy with nitazoxanide may suppress the translational expression of those targets.

Conclusion

In this study, we have successfully identified the anti-HCC core targets and their biological functions, molecular pathways, and the effect of nitazoxanide on HCC. The constructed network pharmacology revealed the significant interaction among the predicted targets that identified eight anti-HCC core targets (SRC, EGFR, CASP3, MMP9, mTOR, HIF1A, ERBB2, and PPARG) as the active bio targets of nitazoxanide in HCC. Furthermore, we found that nine key pathways are likely to be involved: pathways in cancer, proteoglycans in cancer, MicroRNAs in cancer, central carbon metabolism in cancer, lipid and atherosclerosis, Kaposi sarcoma-associated herpesvirus infection, ErbB signaling pathway, and EGFR tyrosine kinase inhibitor resistance, by which nitazoxanide treats HCC. Our study validates the notion that nitazoxanide's anti-HCC effects may emerge from synergistic interactions across several targets and pathways, and our data offer evidence to support this notion. A molecular docking simulation demonstrated the potential for active interaction between the anti-HCC core targets and nitazoxanide. Our study provides a theoretical foundation for the idea that nitazoxanide may have unique therapeutic benefits in HCC, and the pharmacological targets that have been identified may be potential biomarkers in the treatment of HCC.

Data availability statement

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

Author contributions

"Conceptualization, SK and TL; methodology SK; software, SK; validation, SK and TL; formal analysis, SK; investigation, SK and TL; data curation, SK; writing—original draft preparation, SK and TL; writing—review and editing, SK and TL; visualization, SK; supervision, TL; project administration, TL; funding acquisition, TL All authors have read and agreed to the published version of the manuscript."

Funding

This study was supported by the RGC General Research Fund (15102020), Collaborative Research Fund (C7026-18G), and Research Impact Fund (R5050-18F and R7022-20).

Acknowledgments

The authors would like to express their appreciation to the PolyU Distinguished Postdoctoral Fellowship Scheme at The Hong Kong Polytechnic University, Hong Kong, and the aforementioned grant for funding 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

References

Ahrens, T. D., Bang-Christensen, S. R., Jørgensen, A. M., Løppke, C., Spliid, C. B., Sand, N. T., et al. (2020). The role of proteoglycans in cancer metastasis and circulating tumor cell analysis. *Front. Cell Dev. Biol.* 8, 749. doi:10.3389/fcell.2020.00749

Balogh, J., Victor, D., Asham, E. H., Burroughs, S. G., Boktour, M., Saharia, A., et al. (2016). Hepatocellular carcinoma: A review. *J. Hepatocell. Carcinoma* 3, 41–53. doi:10.2147/JHC.S61146

BIOVIA DS (2016). *BIOVIA discovery studio visualizer*, 20, 779.

Cainap, C., Qin, S., Huang, W. T., Chung, I. J., Pan, H., Cheng, Y., et al. (2015). Linifanib versus sorafenib in patients with advanced hepatocellular carcinoma: Results of a randomized phase III trial. *J. Clin. Oncol.* 33, 172–179. doi:10.1200/JCO.2013.54.3298

Chen, C., and Lou, T. (2017). Hypoxia inducible factors in hepatocellular carcinoma. *Oncotarget* 8, 46691–46703. doi:10.18632/ONCOTARGET.17358

Daher, S., Massarwa, M., Benson, A. A., and Khoury, T. (2018). Current and future treatment of hepatocellular carcinoma: An updated comprehensive review. *J. Clin. Transl. Hepatol.* 6, 69–78. doi:10.14218/JCTH.2017.00031

Daina, A., Michielin, O., and Zoete, V. (2019). SwissTargetPrediction: Updated data and new features for efficient prediction of protein targets of small molecules. *Nucleic Acids Res.* 47, W357–W364. doi:10.1093/NAR/GKZ382

DAVID (2022). DAVID functional annotation bioinformatics microarray analysis. Available at: https://david.ncifcrf.gov/ (Accessed May 13, 2022).

Delire, B., and Stärkel, P. (2015). The ras/MAPK pathway and hepatocarcinoma: Pathogenesis and therapeutic implications. *Eur. J. Clin. Invest.* 45, 609–623. doi:10.1111/ECI.12441

Dutta, R., and Mahato, R. I. (2017). Recent advances in hepatocellular carcinoma therapy. *Pharmacol. Ther.* 173, 106–117. doi:10.1016/J.PHARMTHERA.2017.02.010

El-Serag, H. B. (2011). Hepatocellular carcinoma. *N. Engl. J. Med.* 365, 1118–1127. doi:10.1056/NEJMRA1001683

Fan-Minogue, H., Bodapati, S., Solow-Cordero, D., Fan, A., Paulmurugan, R., Massoud, T. F., et al. (2013). A c-Myc activation sensor-based high throughput drug screening identifies an anti-neoplastic effect of Nitazoxanide. *Mol. Cancer Ther.* 12, 1896–1905. doi:10.1158/1535-7163.MCT-12-1243

Ferrín, G., Guerrero, M., Amado, V., Rodríguez-Perálvarez, M., and de la Mata, M. (2020). Activation of mTOR signaling pathway in hepatocellular carcinoma. *Int. J. Mol. Sci.* 21, E1266. doi:10.3390/IJMS21041266

GEPIA (2022). Gene expression profiling interactive analysis. Available at: http://gepia.cancer-pku.cn/ (Accessed May 14, 2022).

Gupta, S. C., Sung, B., Prasad, S., Webb, L. J., and Aggarwal, B. B. (2013). Cancer drug discovery by repurposing: Teaching new tricks to old dogs. *Trends Pharmacol. Sci.* 34, 508–517. doi:10.1016/J.TIPS.2013.06.005

Heinze, T., Jonas, S., Karsten, A., and Neuhaus, P. (1999). Determination of the oncogenes p53 and C-erb B2 in the tumour cytosols of advanced hepatocellular carcinoma (HCC) and correlation to survival time. *Anticancer Res.* 19, 2501–2503. Available at: https://europepmc.org/article/med/10470182.

Hilmi, M., Vienot, A., Rousseau, B., and Neuzillet, C. (2020). Immune therapy for liver cancers. *Cancers (Basel)* 12, E77. doi:10.3390/CANCERS12010077

Jindal, A., Thadi, A., and Shailubhai, K. (2019). Hepatocellular carcinoma: Etiology and current and future drugs. *J. Clin. Exp. Hepatol.* 9, 221–232. doi:10.1016/J.JCEH.2019.01.004

Keenan, A. B., Jenkins, S. L., Jagodnik, K. M., Koplev, S., He, E., Torre, D., et al. (2018). The library of integrated network-based cellular signatures NIH program: System-level cataloging of human cells response to perturbations. *Cell Syst.* 6, 13–24. doi:10.1016/J.CELS.2017.11.001

Khan, S. A., and Lee, T. K. W. (2022). Network-pharmacology-based study on active phytochemicals and molecular mechanism of cnidium monnieri in treating hepatocellular carcinoma. *Int. J. Mol. Sci.* 23, 5400. doi:10.3390/ijms23105400

Li, F., Fu, X., Liu, L., and Wei, X. (2022). The anticancer mechanisms of scutellaria barbata against lung squamous cell carcinoma. *J. Oncol.* 2022, 1–12. doi:10.1155/2022/7529923

Li, H., Han, G., Li, X., Li, B., Wu, B., Jin, H., et al. (2021). MAPK-RAP1A signaling enriched in hepatocellular carcinoma is associated with favorable tumor-infiltrating immune cells and clinical prognosis. *Front. Oncol.* 11, 1988. doi:10.3389/fonc.2021.649980

Liu, B., Tian, Y., Chen, M., Shen, H., Xia, J., Nan, J., et al. (2021). CircUBAP2 promotes MMP9-mediated oncogenic effect via sponging miR-194-3p in hepatocellular carcinoma. *Front. Cell Dev. Biol.* 9, 1634. doi:10.3389/fcell.2021.675043

Liu, Y., Zhao, Q., Xu, F., Wang, K., Zhao, Y., Chen, H., et al. (2021). Dysregulation of phosphoproteins in hepatocellular carcinoma revealed via quantitative analysis of the phosphoproteome. *Oncol. Lett.* 21, 117. doi:10.3892/ol.2020.12378

Liu, Z. W., Luo, Z. H., Meng, Q. Q., Zhong, P. C., Hu, Y. J., Shen, X. L., et al. (2020). Network pharmacology-based investigation on the mechanisms of action of Morinda officinalis How. in the treatment of osteoporosis. *Comput. Biol. Med.* 127, 104074. doi:10.1016/J.COMPBIOMED.2020.104074

Llovet, J. M., Ricci, S., Mazzaferro, V., Hilgard, P., Gane, E., Blanc, J.-F., et al. (2008). Sorafenib in advanced hepatocellular carcinoma. *N. Engl. J. Med.* 359, 378–390. doi:10.1056/NEJMoa0708857

Lopes, C. T., Franz, M., Kazi, F., Donaldson, S. L., Morris, Q., Bader, G. D., et al. (2010). Cytoscape web: An interactive web-based network browser. *Bioinformatics* 26, 2347–2348. doi:10.1093/BIOINFORMATICS/BTQ430

Lu, L., Zhang, Q., Wu, K., Chen, X., Zheng, Y., Zhu, C., et al. (2015). Hepatitis C virus NS3 protein enhances cancer cell invasion by activating matrix metalloproteinase-9 and cyclooxygenase-2 through ERK/p38/NF-κB signal cascade. *Cancer Lett.* 356, 470–478. doi:10.1016/J.CANLET.2014.09.027

Luo, H., Li, M., Yang, M., Wu, F. X., Li, Y., Wang, J., et al. (2021). Biomedical data and computational models for drug repositioning: A comprehensive review. *Brief. Bioinform.* 22, 1604–1619. doi:10.1093/BIB/BBZ176

Mohs, A., Kuttkat, N., Reißing, J., Zimmermann, H. W., Sonntag, R., Proudfoot, A., et al. (2017). Functional role of CCL5/RANTES for HCC progression during chronic liver disease. *J. Hepatol.* 66, 743–753. doi:10.1016/J.JHEP.2016.12.011

Moon, H., and Ro, S. W. (2021). MAPK/ERK signaling pathway in hepatocellular carcinoma. *Cancers* 13 (12), 3026. doi:10.3390/CANCERS13123026

Müller, J., Sidler, D., Nachbur, U., Wastling, J., Brunner, T., Hemphill, A., et al. (2008). Thiazolides inhibit growth and induce glutathione-S-transferase Pi (GSTP1)-dependent cell death in human colon cancer cells. *Int. J. Cancer* 123, 1797–1806. doi:10.1002/IJC.23755

National Center for Biotechnology Information (2022). PubChem compound summary for CID 41684, nitazoxanide. Available at: https://pubchem.ncbi.nlm.nih.gov/compound/Nitazoxanide (Accessed May 15, 2022).

Nickel, J., Gohlke, B. O., Erehman, J., Banerjee, P., Rong, W. W., Goede, A., et al. (2014). SuperPred: Update on drug classification and target prediction. *Nucleic Acids Res.* 42, W26–W31. doi:10.1093/NAR/GKU477

Nosengo, N. (2016). Can you teach old drugs new tricks? *Nature* 534, 314–316. doi:10.1038/534314A

Persad, R., Liu, C., Wu, T. T., Houlihan, P. S., Hamilton, S. R., Diehl, A. M., et al. (20042004). Overexpression of caspase-3 in hepatocellular carcinomas. *Mod. Pathol.* 17 (7 17), 861–867. doi:10.1038/modpathol.3800146

Pfab, C., Schnobrich, L., Eldnasoury, S., Gessner, A., and El-Najjar, N. (2021). Repurposing of antimicrobial agents for cancer therapy: What do we know? *Cancers (Basel)* 13, 3193. doi:10.3390/cancers13133193

Pushpakom, S., Iorio, F., Eyers, P. A., Escott, K. J., Hopper, S., Wells, A., et al. (2019). Drug repurposing: Progress, challenges and recommendations. *Nat. Rev. Drug Discov.* 18, 41–58. doi:10.1038/NRD.2018.168

Rahmani, F., Ziaeemehr, A., Shahidsales, S., Gharib, M., Khazaei, M., Ferns, G. A., et al. (2020). Role of regulatory miRNAs of the PI3K/AKT/mTOR signaling in the pathogenesis of hepatocellular carcinoma. *J. Cell. Physiol.* 235, 4146–4152. doi:10.1002/JCP.29333

RCSB PDB (2022). RCSB PDB. Available at: https://www.rcsb.org (Accessed May 15, 2022).

Rebhan, M., Chalifa-Caspi, V., Prilusky, J., and Lancet, D. (1997). GeneCards: Integrating information about genes, proteins and diseases. *Trends Genet.* 13, 163. doi:10.1016/S0168-9525(97)01103-7

Senkowski, W., Zhang, X., Olofsson, M. H., Isacson, R., Höglund, U., Gustafsson, M., et al. (2015). Three-dimensional cell culture-based screening identifies the anthelmintic drug nitazoxanide as a candidate for treatment of colorectal cancer. *Mol. Cancer Ther.* 14, 1504–1516. doi:10.1158/1535-7163.MCT-14-0792

Shi, J. H., Guo, W. Z., Jin, Y., Zhang, H. P., Pang, C., Li, J., et al. (2019). Recognition of HER2 expression in hepatocellular carcinoma and its significance in postoperative tumor recurrence. *Cancer Med.* 8, 1269–1278. doi:10.1002/CAM4.2006

Sisson, G., Goodwin, A., Raudonikiene, A., Hughes, N. J., Mukhopadhyay, A. K., Berg, D. E., et al. (2002). Enzymes associated with reductive activation and action of nitazoxanide, nitrofurans, and metronidazole in *Helicobacter pylori*. *Antimicrob. Agents Chemother.* 46, 2116–2123. doi:10.1128/AAC.46.7.2116-2123.2002

Stehlin, J. S., de Ipolyi, P. D., Greeff, P. J., McGaff, C. J., Davis, B. R., McNary, L., et al. (1988). Treatment of cancer of the liver. Twenty years' experience with

infusion and resection in 414 patients. *Ann. Surg.* 208, 23–35. doi:10.1097/00000658-198807000-00004

Sudhesh Dev, S., Zainal Abidin, S. A., Farghadani, R., Othman, I., and Naidu, R. (2021). Receptor tyrosine kinases and their signaling pathways as therapeutic targets of curcumin in cancer. *Front. Pharmacol.* 12, 3266. doi:10.3389/fphar.2021.772510

Sueangoen, N., Tantiwetrueangdet, A., and Panvichian, R. (2020). HCC-derived EGFR mutants are functioning, EGF-dependent, and erlotinib-resistant. *Cell Biosci.* 10, 41. doi:10.1186/s13578-020-00407-1

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

Venny (2022). *Venny 2.1.0*. Available at: https://bioinfogp.cnb.csic.es/tools/venny/ (Accessed May 12, 2022).

Vetrivel, P., Murugesan, R., Bhosale, P. B., Ha, S. E., Kim, H. H., Heo, J. D., et al. (2021). A network pharmacological approach to reveal the pharmacological targets and its associated biological mechanisms of prunetin-5-o-glucoside against gastric cancer. *Cancers (Basel)* 13, 1918. doi:10.3390/cancers13081918

von Mering, C., Huynen, M., Jaeggi, D., Schmidt, S., Bork, P., Snel, B., et al. (2003). String: A database of predicted functional associations between proteins. *Nucleic Acids Res.* 31, 258–261. doi:10.1093/NAR/GKG034

Weishengxin (2022). Weishengxin - data analysis and visualization experts around you. Available at: http://www.bioinformatics.com.cn/ (Accessed May 13, 2022).

Xian, Z. H., Zhang, S. H., Cong, W. M., Wu, W. Q., and Wu, M. C. (2005). Overexpression/amplification of HER-2/neu is uncommon in hepatocellular carcinoma. *J. Clin. Pathol.* 58, 500–503. doi:10.1136/JCP.2004.023556

Yan, X. L., Jia, Y. L., Chen, L., Zeng, Q., Zhou, J. N., Fu, C. J., et al. (2013). Hepatocellular carcinoma-associated mesenchymal stem cells promote hepatocarcinoma progression: Role of the S100A4-miR155-SOCS1-MMP9 axis. *Hepatology* 57, 2274–2286. doi:10.1002/HEP.26257

Yang, S., Cai, C., Wang, H., Ma, X., Shao, A., Sheng, J., et al. (2022). Drug delivery strategy in hepatocellular carcinoma therapy. *Cell Commun. Signal.* 20, 26. doi:10.1186/s12964-021-00796-x

Yuan, M., Shong, K., Li, X., Ashraf, S., Shi, M., Kim, W., et al. (2022). A gene Co-expression network-based drug repositioning approach identifies candidates for treatment of hepatocellular carcinoma. *Cancers (Basel)* 14, 1573. doi:10.3390/cancers14061573

Zamboni, W. C., Torchilin, V., Patri, A. K., Hrkach, J., Stern, S., Lee, R., et al. (2012). *IN-CANCER-NANOTECHNOLOGY*, 18, 3229–3241. doi:10.1158/1078-0432.CCR-11-2938/88462/P/BEST-PRACTICES-Best practices in cancer nanotechnology: Perspective from NCI nanotechnology alliance*Clin. Cancer Res.*

Zhao, R., Wu, Y., Wang, T., Zhang, Y., Kong, D., Zhang, L., et al. (2015). Elevated Src expression associated with hepatocellular carcinoma metastasis in northern Chinese patients. *Oncol. Lett.* 10, 3026–3034. doi:10.3892/OL.2015.3706

Zhu, J., Li, B., Ji, Y., Zhu, L., Zhu, Y., Zhao, H., et al. (2019). β elemene inhibits the generation of peritoneum effusion in pancreatic cancer via suppression of the HIF1A VEGFA pathway based on network pharmacology. *Oncol. Rep.* 42, 2561–2571. doi:10.3892/or.2019.7360

frontiers | Frontiers in Pharmacology

TYPE Review
PUBLISHED 09 August 2022
DOI 10.3389/fphar.2022.955648

OPEN ACCESS

EDITED BY
Keshav Raj Paudel,
University of Technology Sydney,
Australia

REVIEWED BY
Arnab Mukherjee,
Indian Institute of Science Education and Research, Pune, India
Giuseppe Felice Mangiatordi,
National Research Council (CNR), Italy

*CORRESPONDENCE
Tao Wang,
tao-wang@zju.edu.cn

SPECIALTY SECTION
This article was submitted to Drugs Outcomes Research and Policies, a section of the journal Frontiers in Pharmacology

RECEIVED 29 May 2022
ACCEPTED 13 July 2022
PUBLISHED 09 August 2022

CITATION
Liu X-H, Cheng T, Liu B-Y, Chi J, Shu T and Wang T (2022), Structures of the SARS-CoV-2 spike glycoprotein and applications for novel drug development.
Front. Pharmacol. 13:955648.
doi: 10.3389/fphar.2022.955648

Structures of the SARS-CoV-2 spike glycoprotein and applications for novel drug development

Xiao-Huan Liu, Ting Cheng, Bao-Yu Liu, Jia Chi, Ting Shu and Tao Wang*

School of Biological Science, Jining Medical University, Jining, China

COVID-19 caused by SARS-CoV-2 has raised a health crisis worldwide. The high morbidity and mortality associated with COVID-19 and the lack of effective drugs or vaccines for SARS-CoV-2 emphasize the urgent need for standard treatment and prophylaxis of COVID-19. The receptor-binding domain (RBD) of the glycosylated spike protein (S protein) is capable of binding to human angiotensin-converting enzyme 2 (hACE2) and initiating membrane fusion and virus entry. Hence, it is rational to inhibit the RBD activity of the S protein by blocking the RBD interaction with hACE2, which makes the glycosylated S protein a potential target for designing and developing antiviral agents. In this study, the molecular features of the S protein of SARS-CoV-2 are highlighted, such as the structures, functions, and interactions of the S protein and ACE2. Additionally, computational tools developed for the treatment of COVID-19 are provided, for example, algorithms, databases, and relevant programs. Finally, recent advances in the novel development of antivirals against the S protein are summarized, including screening of natural products, drug repurposing and rational design. This study is expected to provide novel insights for the efficient discovery of promising drug candidates against the S protein and contribute to the development of broad-spectrum anti-coronavirus drugs to fight against SARS-CoV-2.

KEYWORDS

COVID-19, spike glycoprotein, small-molecule inhibitors, drug development, computer-aided drug development

1 Introduction

The 2019 novel coronavirus disease (COVID-19) caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has rapidly spread to more than 210 countries and has become a serious threat to global public health (Jiang et al., 2020; Lai et al., 2020). To date, the globe is still struggling with COVID-19. Coronaviruses (CoVs) can infect humans and animals and cause a variety of diseases, such as fever, severe respiratory illness and pneumonia, threatening human health and public safety. CoVs are mainly divided into four genera, α-CoV, beta-CoV,

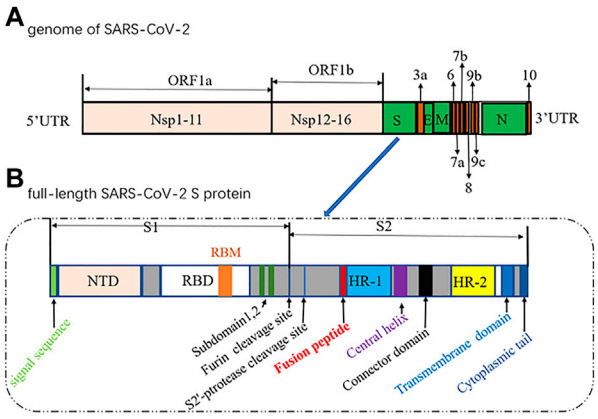


FIGURE 1
Diagrams of the genome of SARS-CoV-2 (1) and the full-length SARS-CoV-2 S protein (2).(NTD, N-terminal domain; RBM, receptor-binding motif; RBD, receptor-binding domain; HR, heptad repeat).

gamma-CoV, and delta-CoV (Ou et al., 2020). During the past 2decades, β-CoVs have caused three severe zoonotic outbreaks: severe acute respiratory syndrome-CoV (SARS-CoV) in 2003, Middle East respiratory syndrome-CoV (MERS-CoV) in 2012, and newly emerged SARS-CoV-2 in late 2019. To date, several promising antiviral medicines, such as remdesivir (Schooley et al., 2021), molnupiravir (Holman et al., 2021), and paxlovid (ritonavir/PF-07321332) (Zhao et al., 2021), have been developed or approved for marketing; unfortunately, no specific medicine or standard treatment has been developed yet to combat COVID-19.

Although the physiology-based approach is a traditional and proven drug discovery paradigm for the development of novel drugs, emerging computer-aided drug development has become a promising alternative to accelerate the modern discovery process (Yadav et al., 2020; Gentile et al., 2021). These *in silico* strategies could very effectively identify novel active scaffolds for a validated target. Therefore, target identification (i.e., one or more targets) has become a key starting point for a successful drug discovery project, which is also true for the development of pancoronavirus (HCoV) antiviral drugs (Liu et al., 2021).

The genome of SARS-CoV-2 contains two large overlapping open reading frames (Figure 1A, ORF1a and ORF1b) encoding 16 non-structural proteins (Nsp1 to 16), along with open reading frames encoding four structural proteins (spike (S), membrane (M), envelope (E), and nucleocapsid (N)) and nine accessory proteins (Chan et al., 2020; Pillay, 2020). The trimeric S protein (~180 kDa, Figure 1B), consisting of the S1 and S2 subunit, is crucial for the virus to enter the cell. In particular, S1 contains a receptor-binding domain (RBD) that binds to angiotensin-converting enzyme 2 (ACE2) to initiate the entry of the virus into cells (Wang et al., 2020). Considering that the SARS-CoV-2 viral life cycle starts with the binding of the S-RBD to the host ACE2 receptor, the S protein, especially the S-RBD, is considered a key molecular target for the development of vaccines, therapeutic agents, and diagnostic methods against COVID-19 (Pandey et al., 2021; Souza et al., 2021; Tan et al., 2021; Zahradník et al., 2021; Gyebi et al., 2022).

Given the importance of the S protein in the context of the COVID-19 pandemic, in this review, the molecular features of the S protein of SARS-CoV-2 are highlighted. Additionally, computational tools available for the treatment of COVID-19 were also provided. Finally, recent advances in the novel development of antivirals against the S protein are summarized. Taken together, this study provides an essential foundation for the design and development of efficient antiviral agents based on the SARS-CoV-2 S protein.

2 Structures and functions of the SARS-CoV-2 spike glycoprotein

2.1 S protein: A key target for antivirals

The surface transmembrane spike glycoprotein S is a typical class I viral fusion protein that is responsible for viral attachment to host cells, subsequent virus–cell membrane fusion and humoral and cell-mediated response induction (Hatmal et al., 2020). The overall structure of the SARS-CoV-2 S protein is quite similar to that of SARS-CoV S; the main conformational difference lies in the position of the receptor-binding domain (RBD) (Wrapp et al., 2020). ACE2 can bind more tightly to the SARS-CoV-2 S protein (with ~15 nM affinity) than to the SARS-CoV S protein. This may help explain the enhanced pathogenicity of COVID-19 compared with that of SARS-CoV. The amino acid sequence identity of the S proteins of SARS-CoV-2 and SARS-CoV is approximately 77%, also indicating that they are closely related phylogenetically (Zhou et al., 2020).

Usually, when the S protein is processed and hydrolysed by one or multiple host proteases [e.g., furin and transmembrane protease serine protease-2 (TMPRSS-2)], it will lead to the formation of active and fusion-competent S protein (Ou et al., 2020). For binding to the host cell receptor, the RBD undergoes a transient hinge-like conformational change from the "down" conformation (receptor inaccessible) to the "up" conformation (receptor accessible) (Peng et al., 2020; Meirson et al., 2021). In coronaviruses, the fusion-competent S protein (Figure 2A)) usually forms a trimer carrying the receptor-binding subunit S1 (700 amino acids) and the membrane-fusion subunit S2 (600 amino

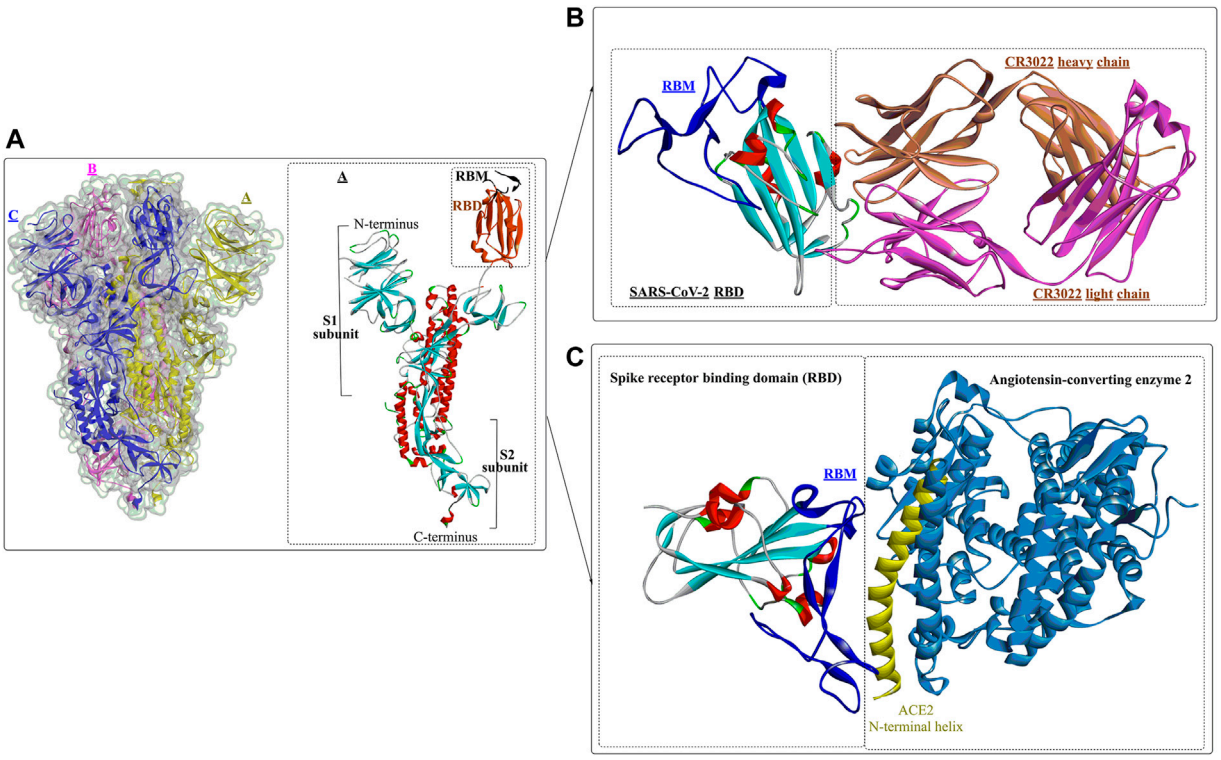


FIGURE 2
Crystal structure of prefusion SARS-CoV-2 spike glycoprotein (1, PDB ID: 6VSB), antibody CR3022 binding to the SARS-CoV-2 RBD (2, PDB ID: 6W41), and the SARS-CoV-2 spike receptor-binding domain bound with ACE2 (3, PDB ID: 6M0J).

acids). It is noted that an insertion of four amino acid residues at the junction of S1 and S2 of SARS-CoV-2 will generate a polybasic cleavage site (RRAR), which would greatly facilitate effective cleavage (Andersen et al., 2020; Coutard et al., 2020). For SARS-CoV and SARS-CoV-2 (Lan et al., 2020), angiotensin-converting enzyme 2 (ACE2) is required for binding to target cells (Figure 2C), while dipeptidyl peptidase 4 (DPP4) is the necessary cellular receptor of MERS-CoV (Raj et al., 2013).

The RBDs of SARS-CoV-2, MERS-CoV, and SARS-CoV are located in the S1 subunit (Du et al., 2017) and are composed of a core subdomain and a receptor-binding motif (RBM) mediating viral attachment to host cells. Differences in the RBM domains would lead to the use of different receptors in varying hosts (Lu et al., 2013; Wang et al., 2013). Upon binding to the receptor, the S1 subunit dissociates from the trimeric S protein and is then exposed to the other subunit, S2 (Benton et al., 2020), (Cai et al., 2020). The S2 subunit contains several important structural elements (Walls et al., 2016), including an N-terminal fusion peptide (FP), heptad repeat 1 (HR1), the central helix (CH), the connector domain (CD), heptad repeat 2 (HR2), the transmembrane region (TM), and the cytoplasmic tail (CT). FP can bind to the target cell membrane and, once bound, will induce S2 into a prehairpin state to connect the viral and cellular membranes. Then, 3 HR1 regions self-assemble into a trimeric coiled coil, and 3 HR2 regions fold into the interface of the HR1 inner core, forming a six-helix bundle (6-HB) structure (Yuan et al., 2017).

Wang et al. determined the crystal structure of the RBD of SARS-CoV-2 bound to the cell receptor ACE2 (Figure 2C), and the results revealed that the interaction modes resemble those of the SARS-CoV RBD (Lan et al., 2020). The binding site is composed of five-stranded antiparallel β sheets, several short connecting helices, and loops. Among these secondary structures, four pairs of disulfide bonds formed by eight cysteine residues were also identified, which are used for stabilizing the β sheets. Analysis of the critical residues associated with RBD binding revealed that a total of 16 residues in the RBD (Figure 2C in dark blue) might interact with the N-terminal helix of ACE2. Among these critical residues, hydrophilic interactions (13 hydrogen bonds and 3 salt bridges) are formed during binding to the ACE2 receptor. It is worth noting that in this study, no interactions between the N-acetyl-β-glucosaminide (NAG) glycans and SARS-CoV-2 RBD were found, although the glycan-RBD interaction is believed to be associated with the binding of the SARS-CoV RBD to ACE2 (Li et al., 2005).

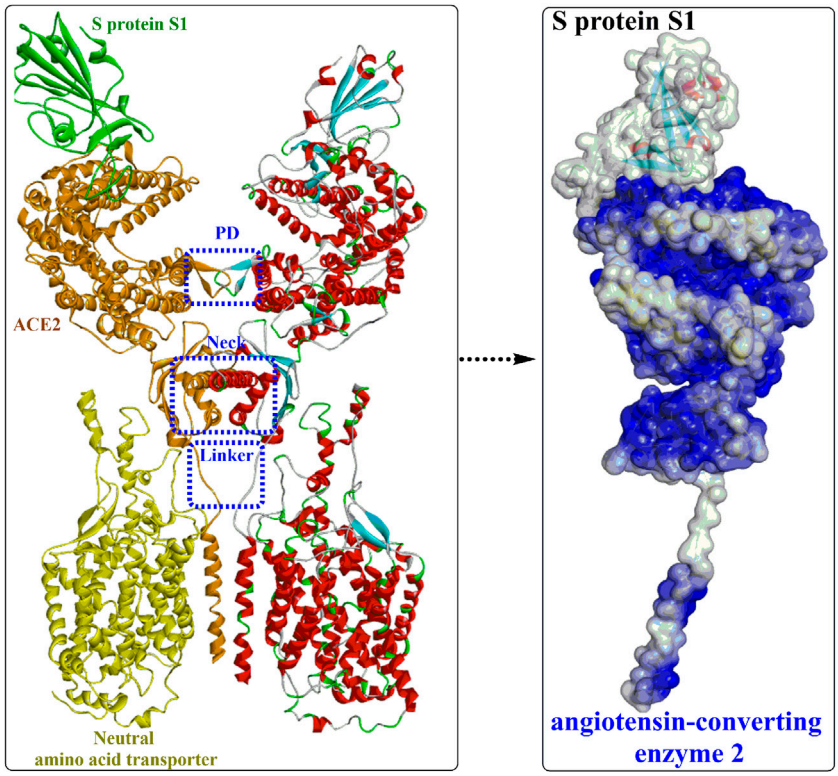


FIGURE 3
Overall structure of the RBD-ACE2-neutral amino acid transporter complex (PDB ID: 6M17).

In a recent study (Yuan et al., 2020), a highly conserved cryptic epitope in the RBD of SARS-CoV-2 and SARS-CoV was discovered, which could be recognized by the neutralizing antibody CR3022 (Figure 2B). The presence of a glycan would induce CR3022 to bind more tightly to SARS-CoV than SARS-CoV-2. Although this special domain is distal from the traditional RBD, it makes cross-reactive binding between SARS-CoV-2 and SARS-CoV possible. In particular, it was found that the binding epitope could be accessed by CR3022 only when at least two of the three RBDs on the trimeric S protein were in the "up" conformation.

2.2 Interactions between the S protein and ACE2

Yan et al. (Yan et al., 2020) released the cryo-electron microscopy structures of full-length human ACE2 in the presence of the neutral amino acid transporter B0AT1 and the RBD of the S protein of SARS-CoV-2 (Figure 3). The ACE2-B0AT1 complex is assembled as a dimer of heterodimers (Figure 3A) with two critical functional domains, including an N-terminal peptidase domain (PD, residues 19–615) and a C-terminal collectrin-like domain (CLD) of ACE2. The RBD is recognized and directly binds to the PD of ACE2 mainly through polar interactions, and the homodimerization process is mediated by CLD (Song et al., 2018).

Interactions between S1-RBD and ACE2 mainly occur in the region constructed by residues F486 to V503 (*β**, coloured yellow). The two ends of the β* region strongly interact with the N- and C-termini of the α1 helix and certain areas on the α2 helix and *β*1 sheet. Moreover, the interaction can be further stabilized by interactions through several polar residues in the middle of α1 (Figure 4). At the N-terminus of α1, P499, T500, and N501 of the RBD form a network of H-bonds with Y41, Q42, K353, V503, G354, D355, and R357 from ACE2 (Figure 4B). In the middle of β*, Lys417 and Tyr453 of the RBD interact with Asp30 and His34 of ACE2, respectively (shown in green, Figure 4B). At the C-terminus of α1, Q493, C488 of the RBD is H-bonded to K31 and T27 of ACE2, respectively (shown in green, Figure 4B), whereas F486 of the RBD interacts with M82 and Y83 of ACE2 through van der Waals forces (shown in purple, Figure 4B). It is clear that those identified residues associated with the interactions with ACE2 would certainly make potential targets for inhibitors against virus replication. In addition, it was found that the T470-F490 loop (activated in

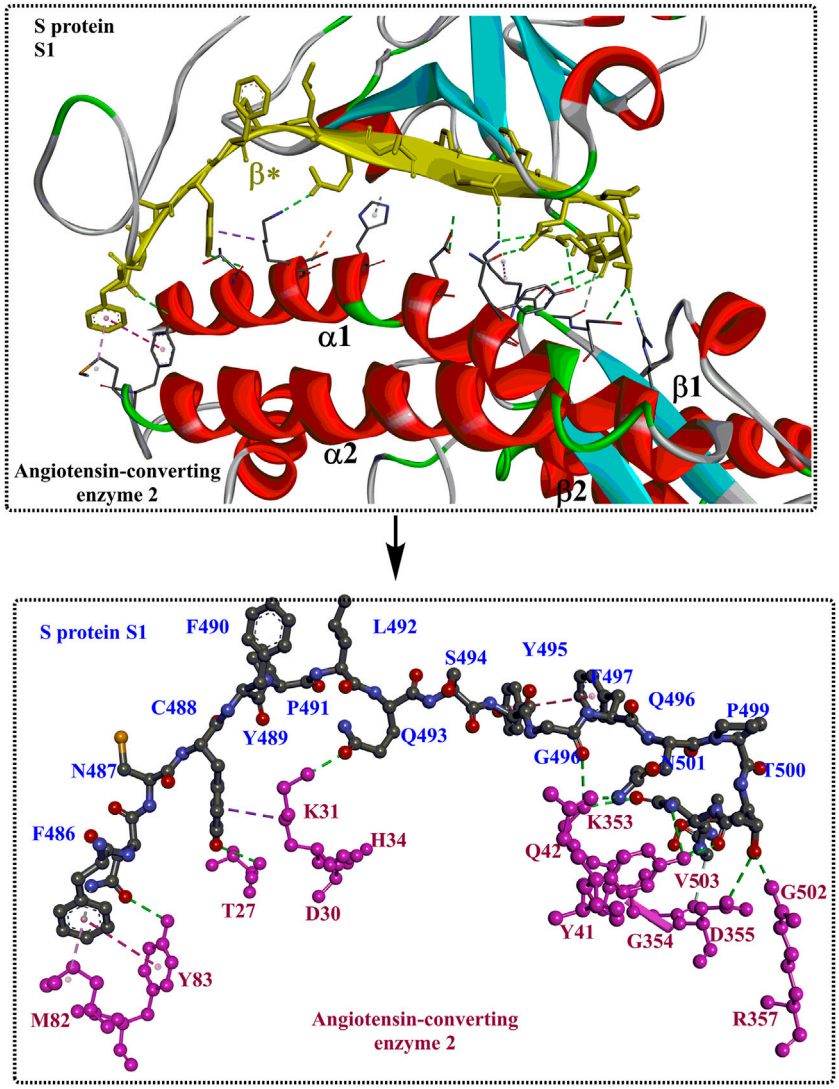


FIGURE 4
The interaction modes of S1 and ACE2 (PDB ID: 6M17).

the open state) and Q498-Y505 residues within the RBD domain of SARS-CoV-2 S act as viral determinants for the specific recognition of SARS-CoV-2 RBD by ACE2 (Xu et al., 2021).

2.3 Mutations in the spike protein and related interactions

Owing to the crucial role of the S protein in entering the cell for the virus, key mutations of the S protein might alter the virus' infectivity, virulence, and antigenicity, leading to the reduced effectiveness of therapeutic antibodies and vaccines (Wang et al., 2021a; Wang et al., 2021b; Starr et al., 2021). Results showed that mutants (N501Y, E484K, and K417 N/T) with high mutation frequencies might have become the main genotypes for the spread of SARS-CoV-2 (Yi et al., 2021). Recent quantitative analysis of the stability of the ACE2-RBD complex for the Omicron variant (SARS-CoV-2 B.1.1.529) showed that its RBD could bind more strongly to the target human ACE2 protein than the original strain through increased hydrogen bonding interactions and a more buried solvent-accessible surface region (Lupala et al., 2022). This might help explain why 85% of previously characterized neutralization antibodies lost their efficacy against the new variant Omicron (Cao et al., 2022). Omicron (B.1.1.529) exhibits more than thirty amino acid mutations in the receptor-binding motif of the spike protein, and the increases in transmissibility and immune evasion have caused a challenging and threatening situation worldwide (Kannan et al., 2021; Meo et al., 2021).

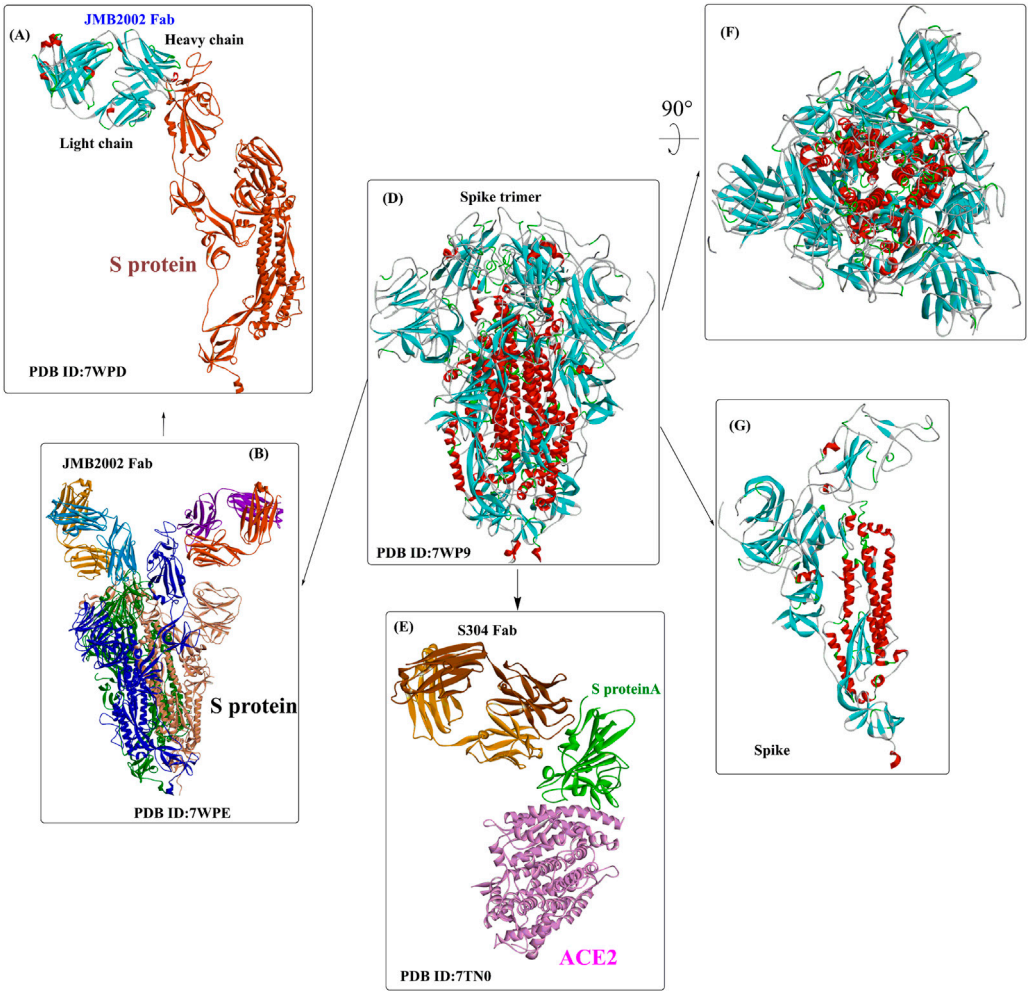


FIGURE 5
The interaction modes of the Omicron spike protein and ACE2.

Recently, several crystal structures of the Omicron spike trimer in complex with angiotensin-converting enzyme 2 (ACE2) or the therapeutic antibody (JMB 2002) (Han et al., 2022; Hong et al., 2022; McCallum et al., 2022; Yin et al., 2022) have been released. With 15 mutated residues, the overall structure of the Omicron ACE2-RBD complex is similar to that of the wild-type ACE2-RBD complex (Figure 5). Most Omicron mutations are located on the surface of the spike protein and change binding epitopes to many current antibodies. In the ACE2-binding site (Yin et al., 2022), compensating mutations strengthen RBD binding to ACE2, forming additional interactions with ACE2, including interactions from the RBD mutations N477 (hydrogen bonds), R493, Q496, R498 (hydrogen bonds), and Y501 (packing interactions) to ACE2 (Figure 6A). Moreover, RBD-RBD interactions from one of the two down RBDs to the up RBD were found, which might be capable of stabilizing the up conformation of the RBD. In contrast, the mutant residues (Omicron residues L371, P373, and F375) located at the entrance to the fatty acid–binding pocket could probably distort the pocket and destabilize the RBDs in the all closed-down conformation. All the mentioned interactions could further contribute to the higher affinity of Omicron. Similar findings also revealed new salt bridges, hydrogen bonds and π-stacking interactions formed by mutated residues R493, S496, R498 and Y501 in the RBD with ACE2 (Figure 6D) (Mannar et al., 2022). It was also found that the mutant residues N471 (H bond), H505 (van der Waals interactions), and R498 (salt bridge) also play important roles in hACE2 binding (Mannar et al., 2022).

2.4 N-terminal domain of the S protein of SARS-CoV-2

It has been revealed that S1 consisted of the NTD and the RBD plays a critical role in the lifecycle of SARS-CoV-2.

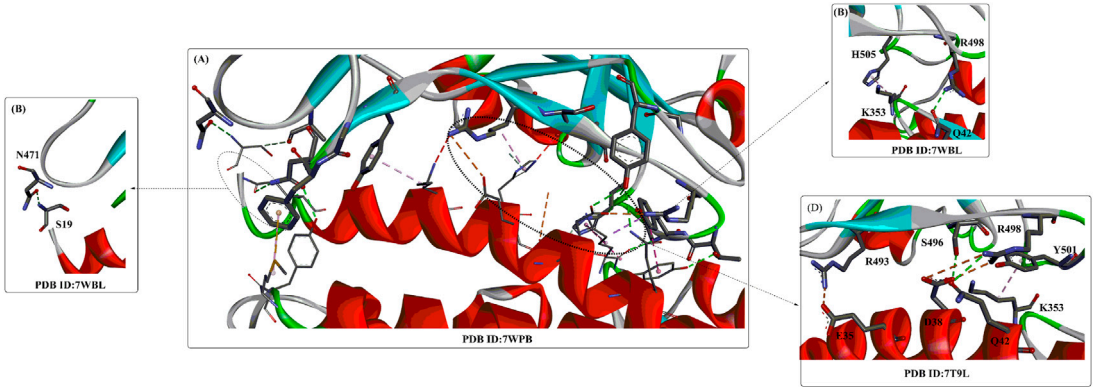


FIGURE 6
The interaction modes of the Omicron spike protein and ACE2.

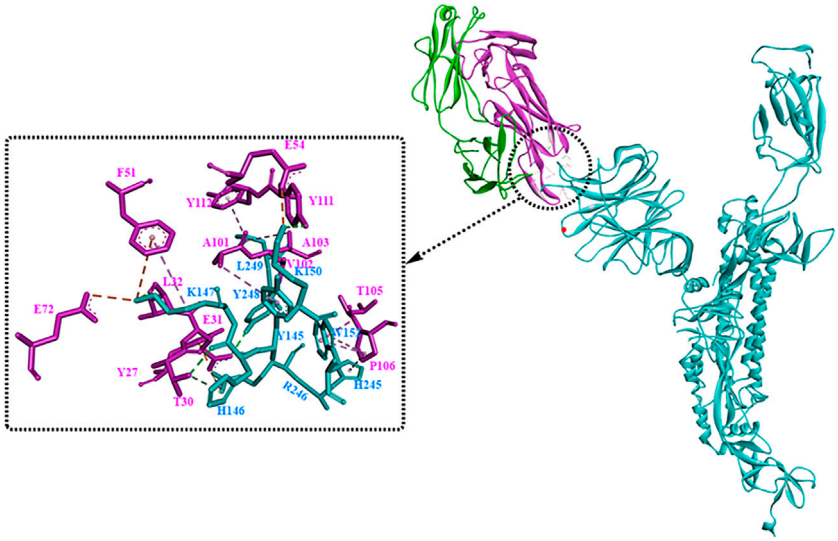


FIGURE 7
The interaction modes of the 4A8 and S-NTD complex (PDB:7C2L).

Though the detailed functions of NTD have not been well investigated, several studies showed that drug development against the NTD, especially the NTD-directed antibodies, might be another promising strategy (Ciuffreda et al., 2021; Di Gaetano et al., 2021; Schuurs et al., 2021). In a study by Chi et al. (2020), a neutralizing human antibody binding to the NTD of the S protein was developed and investigated. The biological results showed that the neutralizing capacity of 4A8 with EC_{50} of 0.61 mg/ml, moreover it could protect the ACE2-293T cells with an EC_{50} of 49 mg/ml. From the structure of the complex between 4A8 and S-NTD, it shows that the heavy chain of 4A8 mainly participates in binding to the NTD, on the contrary the light chain is away from the RBD. On this basis, it was estimated that 4A8 might play important functions in restraining the conformational changes of the S protein. Also, at the surface area of the 4A8-NTD interface critical residues including K147, Y248, K150, H146, R246, H245, W152, L129, N149, and so on could interact with 4A8 by means of H-bounds, salt bridges and hydrophilic interactions (Figure 7). In another study (Cerutti et al., 2021), structural analysis of seven potent NTD-directed neutralizing antibodies revealed a common highly

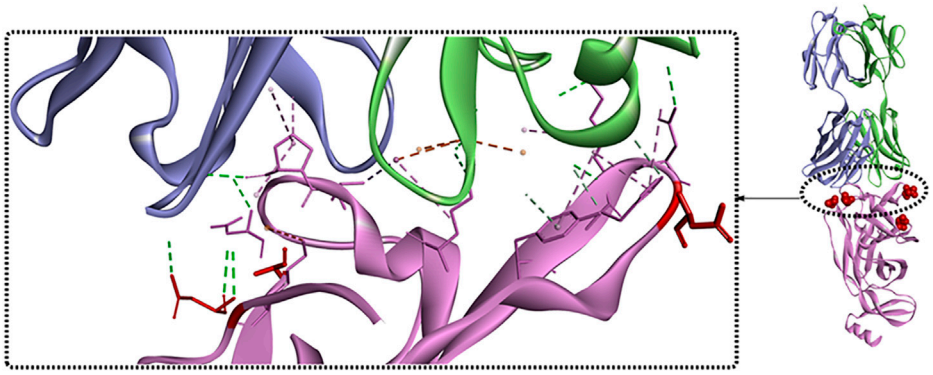


FIGURE 8
Crystal structure analysis of S-NTD bound to S2M28 Fab (PDB:7LY3).

electropositive binding site, which is formed by a mobile β-hairpin and several flexible loops including the critical residues glycans N17, N74, N122, and N149. Similarly, Matthew et al. also identified a supersite (site I, Figure 8), which could be recognized by all known NTD-specific neutralizing antibodies (McCallum et al., 2021). These studies indicate that potent NTD-directed neutralizing antibodies might probably target the single supersite.

3 Targets for novel drug development

Given that viral entry is mainly mediated by the trimeric spike protein, the S protein is considered a major therapeutic target for the treatment of SARS-CoV-2 infections. To interfere with the S protein-hACE2 interaction, neutralizing antibodies are usually the most traditional and functional strategies. However, inhibitors of the protein–protein interaction (PPI) between the S protein and hACE2 have recently drawn increasing attention for the development of potential antiviral agents to prevent viral attachment and cellular entry to control the ongoing COVID-19 pandemic (Tai et al., 2020; Bojadzic et al., 2021). In addition, although small molecular weight inhibitors (SMIs) are usually not considered potential candidates for PPI modulation, an increasing number of studies have revealed that SMIs could also be effective against certain PPIs (Scott et al., 2016; Risner et al., 2020; Pan et al., 2021).

Considering that anti-SARS-CoV-2 neutralizing antibodies have been extensively investigated (Ju et al., 2020; Cho et al., 2021; Wang et al., 2021c; Lucas et al., 2021), in this section, the recent drug development of novel small-molecule inhibitors (nonpeptide small molecules) that can interfere with viral entry or viral propagation is highlighted, including the discovery of natural products, drug repurposing, and novel drug development.

3.1 Computational tools developed for COVID-19 treatment

In addition to the traditional computational tools frequently used for rational drug development, such as Discovery Studio, Gold and AutoDock, several types of computational resources, tools and databases have recently been developed to investigate the ever-growing available data against COVID-19 and its related diseases.

(1) The D3Targets-SARS-CoV-2 web server (https://www.d3pharma.com/D3Targets-SARS-CoV-2/index.php.) is a webserver capable of predicting potential drug targets and identifying lead compounds against specific or multiple targets via structure-based virtual screening against COVID-19 (Shi et al., 2020). It provides two strategies for target prediction and virtual screening: the structure-based method (D3Pockets) and the ligand-based method (D3Similarity). The potential ligand-binding pockets is predicted by D3Pockets, and the docking process is performed with AutoDock Vina. By the end of 27-05-2021, 56 potential proteins (constructed by homology modelling or *de novo* prediction) involved in the whole process of virus life have been included.
(2) D3Similarity is a ligand-based method developed based on the molecular similarity evaluation between the submitted molecule(s) and the active compounds in the database (604 molecules) (Yang et al., 2021a). The 2D molecular similarity is evaluated by using Open Babel based on the Tanimoto coefficient (Tc) values between the SMILES of the

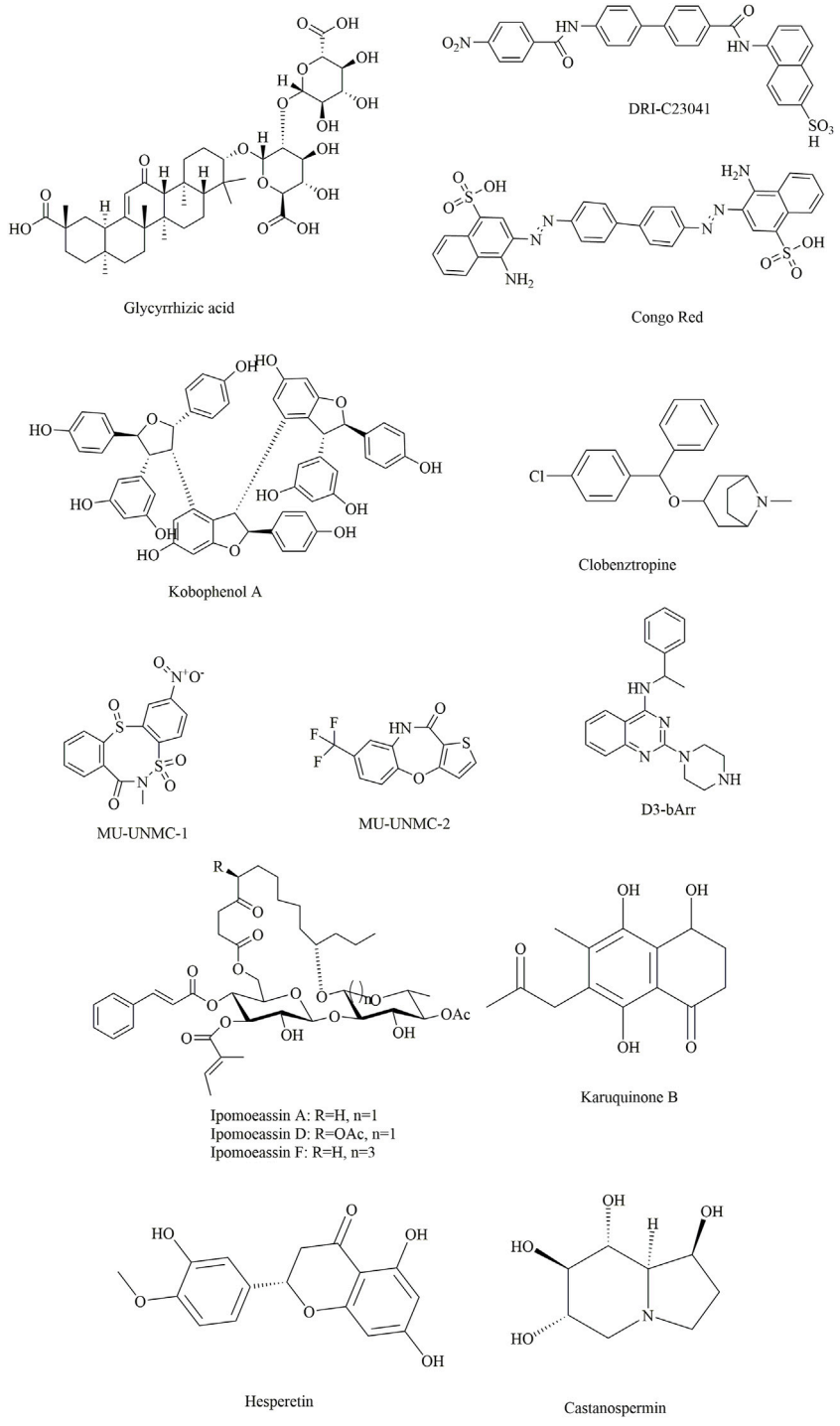


FIGURE 9
Several potential natural products showing inhibitory effects against SARS-CoV-2.

input structure and the molecules in the database. The 3D molecular similarity was evaluated by using MolShaCS (Molecular Shape and Charge Similarity). D3Pockets is a web server developed for systematically exploring protein pocket dynamics based on either molecular dynamic simulation trajectories or conformational ensembles with large-scale conformational changes (Chen et al., 2019). Based on D3Pockets, the stability, continuity, and

correlation of protein pockets could be investigated, and the results could also be visualized with PyMOL.

(3) CovidExpress (https://stjudecab.github.io/covidexpress) is an open-access database and interactive visualization tool for intuitive investigation of SARS-CoV-2-related transcriptomes, and we collected approximately 1,500 human bulk RNA-seq datasets from publicly available resources (Djekidel et al., 2021). It can be used to examine the relative gene expression levels in different tissues, cell lines, and especially the response to SARS-CoV-2. Based on this database, a series of commonly regulated genes (~345 genes, 280 upregulated and 65 downregulated) in SARS-CoV-2-infected lung and nasal cells were identified, such as the interferon response genes *OASL TNF*, *IL1A*, and *CXCL10*.

(4) The COVID-19 Docking Server (http://ncov.schanglab.org.cn) is a web server that can be used for the prediction of the binding modes between COVID-19 targets and ligands (Kong et al., 2020). It provides a free and interactive tool for the prediction of COVID-19 target-ligand interactions and subsequent drug development. A total of 27 targets (e.g., spike protein, nucleocapsid protein, main protease, papain-like protease, and RNA-dependent RNA polymerase) involved in the virus life cycle were collected or constructed based on homology modelling and prepared for docking on the website. For different ligands, the implementation methods are different. For small molecular weight ligands, Open Babel is applied for format transformation and 3D coordinate generation, and AutoDock Vina is used for molecular docking. However, for macromolecular drugs (e.g., peptides and antibodies), CoDockPP is used as a docking engine with a multistage fast Fourier transform (FFT)-based strategy for both global docking and site-specific docking.

(5) MolAICal (https://molaical.github.io) was developed for the rational design of potential 3D drug structures in the 3D pockets of specific targets achieved by using a deep learning model and classical algorithms (Bai et al., 2021). It contains two main modules: one can employ the genetic algorithm, deep learning and the Vinardo score for rational drug design. The second module can use a deep learning generative model and molecular docking (achieved by AutoDock Vina) for virtual screening. The models used in this tool have been fully trained by different databases and methods. Several user-defined rules (e.g., Lipinski's rule of five, synthetic accessibility) are also introduced for filtering out undesired hits.

(6) COVID19 db (http://hpcc.siat.ac.cn/covid19db or http://www.biomedicalweb.com/covid19db) is a user-friendly and open-access platform that integrates 95 COVID-19-related human transcriptomic datasets of 4,127 human samples across 13 body sites associated with exposure to 33 microbes and 33 drugs/agents in GEO and 39, 930 drug–target–pathway interactions among 2,037 drugs, 1,116 targets, and 207 pathways in DrugCentral and KEGG (Zhang et al., 2022). In addition, 14 different analytical applications (included in the differential expression and coexpression modules) and a web service tool are designed and integrated to analyse the integrated data or the obtained human transcriptomic data. Moreover, a drug discovery tool is provided for the identification of potential drugs and targets of COVID-19 and its related diseases at the whole transcriptomic level.

In addition to those mentioned above, some other computational tools have been developed to meet the urgent demand of the COVID-19 outbreak, which are listed in Table 1.

3.2 Small-molecule inhibitors against the S protein

In attempts to discover and identify small-molecule inhibitors against the S protein, an efficient drug screening strategy is extremely important for structure-based, fragment-based, mechanism-based, and computer-guided drug discovery (Figure 9).

3.2.1 Discovery of natural products for use as inhibitors against the S protein

Compared with neutralizing antibodies, small-molecule inhibitors might be more challenging for the blockade of RBD-hACE2 interactions due to the lack of well-defined binding pockets. However, they might offer alternatives that are more broadly active, more patient-friendly, less immunogenic, and more controllable than antibodies due to improved pharmacokinetics, stability, and dosage logistics (Song and Buchwald, 2015; Bojadzic and Buchwald, 2018; Xiu et al., 2020). Traditionally, natural products are important sources for novel drug development due to their rich sources, chemical diversity, large chemical space diversity, and biological activities. Therefore, natural products could make good starting points for modern drug design for the treatment of COVID-19 (Hu et al., 2021a; Mahmudpour et al., 2021). Together with the combination of computer-aided drug design and biological verification, drug development based on natural products is believed to be an efficient strategy for modern drug discovery.

Yu et al. applied this method to discover bioactive monomers from the active ingredient of licorice (*Glycyrrhiza uralensis* Fisch) for broad-spectrum anti-coronavirus candidates (Yu et al., 2021). In addition, surface plasmon resonance (SPR) assays, NanoBit assays and MTT assays were used simultaneously to determine the binding activities, inhibitory activities, and cell toxicities of selected compounds. The results showed that glycyrrhizic acid (ZZY-44) was an efficient (IC_{50}: 22 μM) and broad-spectrum

TABLE 1 Other computational tools developed for the analysis of COVID-19-related data.

Name	Functions	Website	Ref
Virus-CKB	(1) Target prediction; (2) platform of viral-associated computing resources; (3) drug development	https://www.cbligand.org/g/virus-ckb	Feng et al. (2021)
DINC-COVID	Ensemble docking with flexible SARS-CoV-2 proteins	http://dinc-covid.kavrakilab.org/	Hall-Swan et al. (2021)
DeepR2cov	Discovery of potential agents for treating the excessive inflammatory response in COVID-19 patients by a deep representation on heterogeneous drug networks	https://github.com/pengsl-lab/DeepR2cov.git	Wang et al. (2021d)
CoV-AbDab	A coronavirus antibody database containing over 1400 published/patented antibodies and nanobodies	http://opig.stats.ox.ac.uk/webapps/coronavirus	Raybould et al. (2021)
SARS-CoV-2 3D	Supply and analysis of possible experimentally solved and created 3D structures of SARS-CoV-2	https://sars3d.com/	Alsulami et al. (2021)
CORDITE	Combination of state-of-the-art data on potential drugs against the SARS-CoV-2	https://cordite.mathematik.uni-marburg.de	Martin et al. (2020)
CoV3D	Resource for up-to-date coronavirus protein structures	https://cov3d.ibbr.umd.edu	Gowthaman et al. (2021)
DockCoV2	Prediction of the binding affinities of FDA-approved and taiwan national health insurance (NHI) drugs against specific targets	https://covirus.cc/drugs/	Chen et al. (2021)

anti-coronavirus molecule with low toxicity (CC_{50} > 100 μM) *in vitro*, which could disrupt the interaction between the RBD and ACE2 (K_D = 0.87 μM).

Damir Bojadzic et al. (2021) identified several promising candidates by screening a compound library of organic dyes. Among them, Congo red, direct violet 1, Evans blue and novel drug-like compounds (DRI-C23041, DRI-C91005) showed inhibitory effects against the interaction of hACE2 with the spike protein with low micromolar activity (IC_{50}: 0.2–3.0 μM). Notably, the results revealed that the inhibitors identified could bind the SARS-CoV-2-S protein but not hACE2, which provides great significance for the development of small-molecule inhibitors of PPIs critical for SARS-CoV-2 attachment/entry.

Suresh Gangadevi et al. discovered that kobophenol A is a potential inhibitor capable of blocking the interaction between the ACE2 receptor and S1-RBD by virtual screening of a library of natural compounds (Gangadevi et al., 2021). In this study, a computer-aided drug design strategy was applied to screen natural compounds, determine conformational changes and predict potential binding sites, including molecular docking and molecular Dynamic studies. The results showed that kobophenol A from *Caragana sinica* extract could disrupt the interaction between ACE2 and the SARS-CoV-2 S protein (IC_{50}: 1.81 ± 0.04 μM and EC_{50}: 71.6 μM). In addition, two potential binding sites for Kobophenol A were predicted, including the ACE2 hydrophobic pocket and the spike1/ACE2 interface.

Structure-based drug development is a proven method for high-throughput screening of specific compounds. In a study, through molecular Dynamic simulations and molecular docking, the hydrophobic pocket at the FP domain was first investigated, revealing the key binding regions (especially the FP hinge loop) and interactions. Then, a pharmacophore model was generated based on the predicted binding interaction. After that, nearly 200,000 drug-like compounds in the NCATS inhouse library were screened according to pharmacophore- and 3D-shape-based searches. Then, the 2,000 top-scoring compounds from docking were selected, clustered and visually inspected. Ultimately, 120 compounds were prioritized for further evaluation. This led to the discovery of two novel chemotypes of entry inhibitors (clobenztropine and D3-βArr), which displayed single-digit micromolar inhibition against SARS-CoV-2 (IC_{50}: 12.6 and 15.8 μM) as well as SARS-CoV-1 and MERS (Hu et al., 2021b). It is interesting that although the two inhibitors are structurally distinct, they showed a similar binding mode at the fusion peptide (FP) domain, including an H-bond formed between Asp867 and the N atom of the polar headgroup and the π–π stacking interaction with Phe833. This further demonstrates the importance of the FP-binding site as a promising target for the structure-based development of novel inhibitors as drug candidates for treating COVID-19.

Similarly, two compounds (MU-UNMC-1 and MU-UNMC-2) were identified as being capable of blocking both SARS-CoV-2 replication at submicromolar IC_{50} values in human bronchial epithelial cells (0.67 and 1.72 μM) and Vero cells (5.35 and 1.63 μM) and the replication of rapidly transmitting variants of concern, including South African variant B.1.351 (IC_{50} = 9.27 and 3.00 mM) and Scotland variant B.1.222 (IC_{50} = 2.64 and 1.39 mM) (Acharya et al., 2021). In particular, MU-UNMC-2 could function synergistically with remdesivir (RDV), indicating that RDV and MU-UNMC-2 might be developed as a combination therapy to fight SARS-CoV-2.

Ipomoeassins A-E, as a new family of glycoresins, were isolated from the leaves of *Ipomoea squamosa* found in the Suriname rainforest in 2005 (Cao et al., 2005). They were shown to inhibit the proliferation of A2780 human ovarian

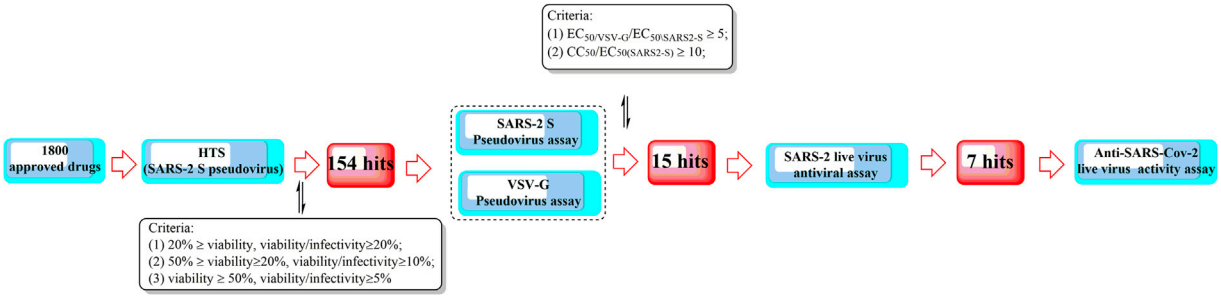


FIGURE 10
A high-throughput screening method for the efficient discovery of SARS-CoV-2 entry inhibitors using a SARS2-S pseudotyped virus.

cancer cells; among them, Ipomoeassin F (Ipom-F) is a potent natural cytotoxin that inhibits the growth of many tumour cell lines as a selective inhibitor of Sec61-mediated protein translocation at the ER membrane (Zong et al., 2019). However, in a recent study, it was found that Ipomoeassin-F could also inhibit the *in vitro* biogenesis of the SARS-CoV-2 spike protein (O'Keefe et al., 2021). It was also revealed that integration of the viral S protein and ACE2 into the endoplasmic reticulum membrane was significantly reduced by Ipom-F, while several other viral membrane proteins were unaffected.

In a study by Mathew (Al-Sehemi et al., 2020), 31,000 natural compounds of the natural product activity and species source (NPASS) library were screened for the discovery of special hits capable of interfering with the SARS-CoV-2 spike protein. The results showed that Castanospermine from a culture extract of *Fusarium solani* and Karuquinone B from different plant species (e.g., *Cassine glaucawere*) were identified and selected based on their binding affinity and pharmacokinetic data. However, no information is available regarding the antiviral activities of kuquinone B, and castanospermine was determined to show antiviral effects (Chang et al., 2013) against various viruses *in vitro* and *in vivo*, such as Ebola (Dowall et al., 2016) and Zika (Bhushan et al., 2020). Similar to Anamika et al. (Basu et al., 2020), the natural products hesperidin, emodin and chrysin were found to be capable of inhibiting SARS-CoV-2. In particular, hesperidin from *Citrus aurantium* could interfere with the interactions between ACE2 and the spike protein. In addition, its interaction was predicted to be located in the middle shallow part of the surface of RBD of Spike, in which the dihydroflavone part was parallel with the β-6 sheet of RBD and the sugar part was inserted into the binding site in the direction away from ACE2 (Wu et al., 2020).

3.2.2 Drug repurposing for inhibitors against the S protein

Recently, Yang et al. proposed a high-throughput screening method for the efficient discovery of SARS-CoV-2 virus entry inhibitors using SARS2-S pseudotyped virus (Figure 10) (Yang et al., 2021b). The results showed that 7 drugs could significantly inhibit SARS2 replication and reduce supernatant viral RNA load with a promising level of activity. Among them, trimeprazine, azelastine hydrochloride, and clemastine, classified as histamine receptor antagonists with clemastine, were determined to show the strongest anti-SARS2 activity. In addition, clemastine is capable of targeting the sigma 1 and sigma 2 receptors (Gordon et al., 2020). Sigma-1 and sigma-2 are endoplasmic binding sites, and sigma-1 is usually considered as a pluripotent chaperone for regulating Ca^{++} fluxes (Ortega-Roldan et al., 2013) and the K^{+} channels (Abraham et al., 2019). In addition, studies also showed that neuroprotection, neuroregulation, and modulation of the proliferative status of cells might also be associated with the functions of sigma 1 (Abate et al., 2020). On the contrary, sigma-2 might play roles in regulating cell death (Pati et al., 2017). Therefore, attention should be given to antihistamine drugs for the development of antiviral agents.

In a study by Smith et al. (2020), SUMMIT, the world's most powerful supercomputer, was applied for the identification of approved drugs that could bind to either the S-protein receptor recognition region or the S protein-human ACE2 interface. Additionally, in this study, an ensemble virtual high-throughput screening docking strategy in combination with restrained temperature replica-exchange molecular dynamic (restrained T-REMD) simulations was used. The results showed that 77 hits (24 having official approval) from over 8,000 drugs, metabolites, and natural products were found to be capable of binding efficiently to the target. Among them, the three top-scoring ligands (cepharanthine, ergoloid, and hypericin) have ZINC15 annotations; however, no experimental testing has been reported regarding the effectiveness of the identified drugs.

In addition to computer-aided drug discovery, other novel and efficient tools for the rapid screening of SARS-CoV-2 inhibitors must also be established. To achieve this, an electrochemical impedance spectroscopy (EIS)-based biosensor was designed and characterized (Kiew et al., 2021). In this system, the core sensing element mainly consisted of a recombinant ACE2 protein-coated palladium nanothin-film (ACE2-Pd-NTF) electrode, which could be

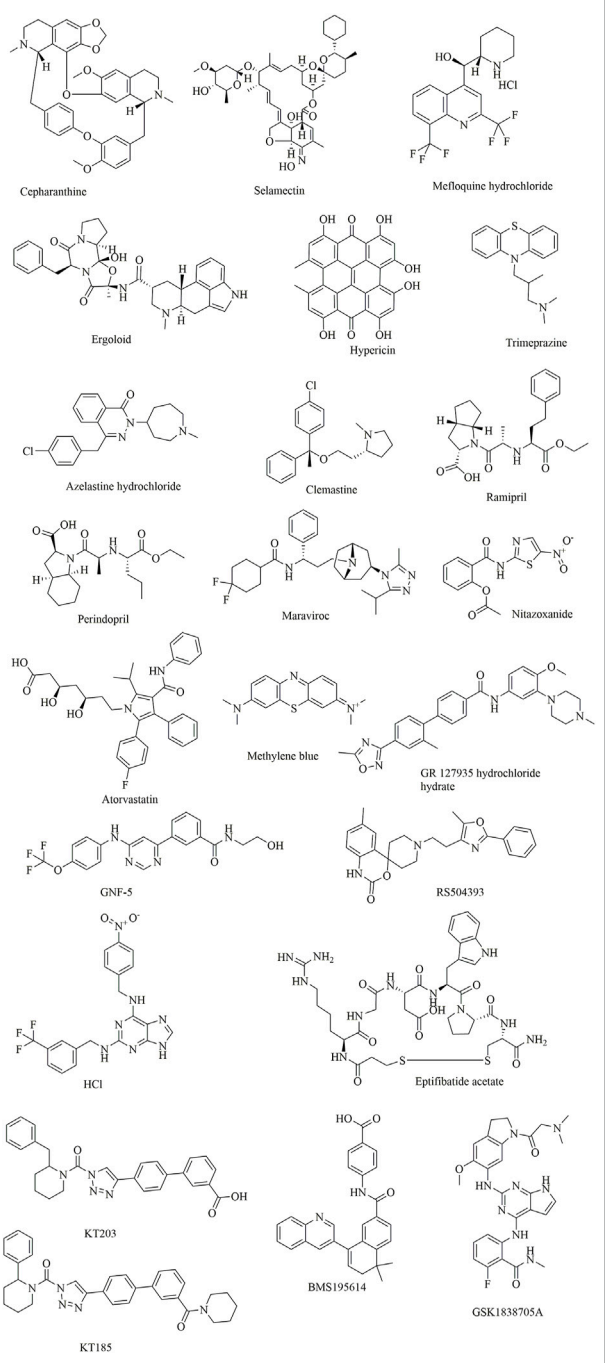


FIGURE 11
Clinically approved drugs repurposed for the treatment of COVID-19.

used to detect alterations occurring in the binding of S-protein to ACE2 when exposed to the test molecules. With this method, several potential pharmacological leads that could disturb SARS-CoV-2-ACE2 binding were successfully identified, such as ramipril and perindopril. Although it is yet to be fully explored at present, this approach has good potential for becoming a mainstream approach for efficient, timesaving, and cost-effective drug discovery and repurposing in the future.

In another study, it was found that maraviroc, FTY720, nitazoxanide and atorvastatin could inhibit SARS-CoV-2 replication in cell culture by screening 19 small molecules and 3 biologics (Risner et al., 2020). However, confocal microscopy with overexpressed S protein revealed that maraviroc reduced the extent of S protein-mediated cell fusion.

Considering the huge chemical space of organic dyes, it is believed that small-molecule inhibitors for PPIs would be more likely to be discovered in such compounds (Downing et al., 2017; Bojadzic and Buchwald, 2018). In a recent study, methylene blue (MB) was found to be capable of inhibiting the interaction between the S protein and ACE2 in a concentration-dependent manner (IC_{50} = 3–3.5 μM), even in the absence of light (Bojadzic et al., 2020). In another study, nonphotoactivated MB showed *in vitro* activity at a very low micromolar range with an EC_{50} of 0.30 ± 0.03 μM and an EC_{90} of 0.75 ± 0.21 μM at a multiplicity of infection of 0.25 against SARS-CoV-2 (strain IHUMI-3) (Gendrot et al., 2020). As a tricyclic phenothiazine compound, it was approved by the FDA for the treatment of methemoglobinemia. However, MeBlu shows dose-dependent toxicity, with symptoms including nausea, vomiting, and haemolysis, when used at doses >2 mg/kg (Dabholkar et al., 2021). In addition to disturbing the direct interaction between SARS-CoV-2 spike protein and ACE2, MB was reported to play various biological roles in blocking the entry of SARS-CoV-2 into the cells, such as preventing the endocytosis of virions into the cells by increasing endosomal and lysosomal intracellular pH and inhibiting the intermediate stages of endocytosis; blocking the formation of the NLRP3 complex to prevent the cytokine storm (van den Berg and Te Velde, 2020); inhibiting nitric oxide synthase and promoting saturation of oxygen to terminate the effects of bradykinin (Ghahestani et al., 2020; Karamyan, 2021).

Shweta et al. investigated FDA-approved LOPAC library drugs against both the RBD of the spike protein and the ACE2 host cell receptor with a high-throughput virtual screening approach and molecular simulations (Choudhary et al., 2020). The results showed that GR 127935 hydrochloride hydrate, GNF-5, RS504393, TNP, and eptifibatide acetate were capable of binding to the ACE2 receptor. In addition, KT203, BMS195614, KT185, RS504393, and GSK1838705A could bind to the RBD of the spike protein.

In a recent study, repurposing clinically approved drugs for the treatment of COVID-19 in a 2019-nCoV-related coronavirus model was achieved (Fan et al., 2020). The results showed that cepharanthine (CEP), selamectin, and mefloquine hydrochloride exhibited complete inhibition of cytopathic effects in cell culture at 10 mmol/L. In particular, CEP displayed the most potent inhibition of GX_P2V infection (EC_{50} = 0.98 mmol/L). In another study using Calu-3 cells, CEP was also determined to show an

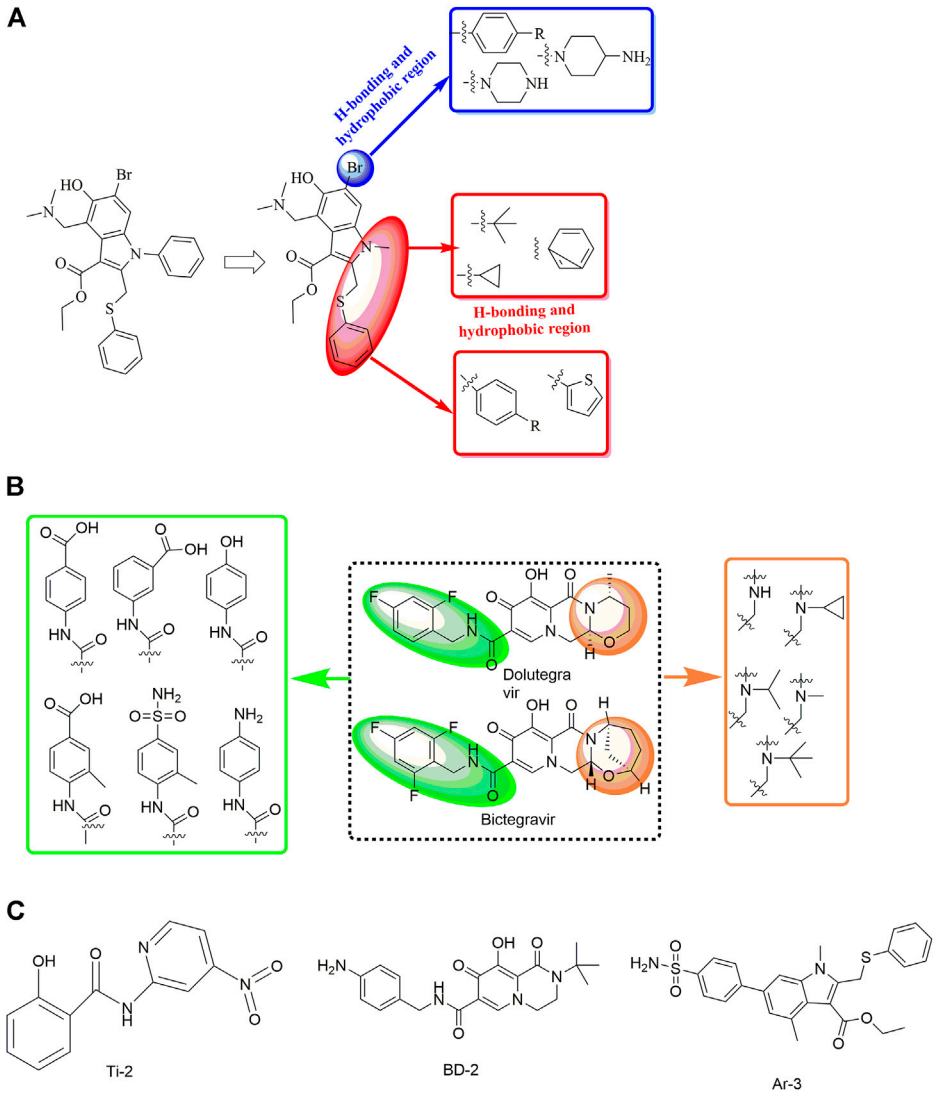


FIGURE 12
Strategies for the further structural optimization of arbidol **(A)**, dolutegravir and bictegravir **(B)** and the structures of BD-2, Ti-2 and Ar-3.

inhibitory effect against SARS-CoV-2, with an IC_{50} of 30 μM (as opposed to an IC_{50} of 4.47 μM in Vero cells) (Jeon et al., 2020; Ko et al., 2020). CEP, a Japanese-approved alopecia drug, is an alkaloid used frequently to treat radiation-induced leukopenia, exudative middle ear catarrh, and viper bite. Transcriptome analysis indicated that CEP could efficiently reverse most dysregulated genes and pathways in infected cells, such as the ER stress/unfolded protein response and HSF1-mediated heat shock response (Li et al., 2021).

3.2.3 Novel drug development for inhibitors against the S protein

In a study by Sun et al. (2021), the discovery and rational design of small-molecule inhibitors of the SARS-CoV-2 S protein was achieved (Figure 12). First, molecular docking with the Lamarckian genetic algorithm was applied for the screening of 14 antiviral molecules by analysing the binding energy and interactions between the ligands and the receptor, the SARS-CoV-2 S protein. This approach led to the discovery of tizoxanide, dolutegravir, bictegravir, and arbidol, which have high binding energies and are capable of binding to the S1 and S2 subunits. Then, structure-based rational design was performed using the molecular connection method and a bioisosterism strategy by introducing specific functional groups to enhance the binding energies and interactions with the S protein. In this way, Ti-2, BD-2, and Ar-3 were identified with much stronger binding ability to the S protein. Although no experimental data about the antiviral activities have been

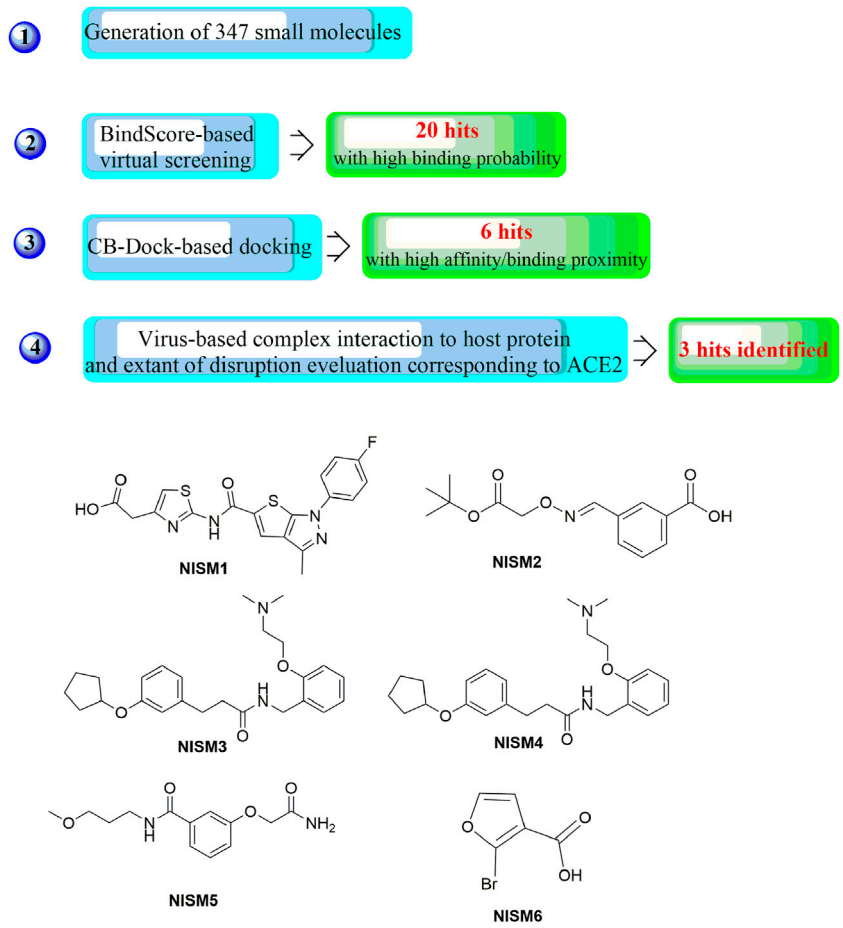


FIGURE 13
Strategies for the further structural optimization of arbidol **(A)**, dolutegravir and bictegravir **(B)** and the structures of BD-2, Ti-2 and Ar-3.

reported to date, the strategy used in this study might be valuable in the rational design of novel anti-SARS-CoV-2 drugs.

In a recent study, another strategy was proposed for the rational design of novel inhibitors against the binding of SARS-CoV-2 to the ACE2 receptor (Figure 13) (Mishra and Nandi, 2021). First, favourable and accessible binding pockets of the RBD were established using the deep convolution neural network (DCNN) model based on parameters such as the hydrogen bond acceptor/donor, hydrophobicity, and ionization energy. Based on the four different binding pockets obtained, decoded *de novo* drug molecules were then generated with a shape-captioning network, leading to the formation of a total of 347 ligands in the simplified molecular input line entry system (SMILES) strings. BindScope based on a DCNN was then used for virtual high-throughput screening according to the simulated binding probabilities, which resulted in the identification of the top 20 molecules with high probabilities of binding. The obtained 20 hits were then subjected to further detailed screening using the CB-Dock server with AutoDock Vina, and 6 compounds showing the best potential to inhibit the spike protein–ACE2 interaction were ultimately identified. Although the newly designed compounds lack antiviral activities, they possess excellent drug scores and are nontoxic and nonmutagenic compared with the several existing antiviral drugs available on the market. Similarly, Rituparno et al. developed an atomistic *de novo* inhibitor generation-guided drug repurposing strategy based on the free-energy validation by well-tempered metadynamics (Chowdhury et al., 2021). It contains three main steps including: generation of new molecules, structural similarity mapping and validation of the binding abilities, well-tempered metadynamics free energy calculations.

In one study, three small-molecule fusion inhibitors with potent inhibitory activity against MERS-CoV were identified (Figures 14A–C); these inhibitors could bind to the surface of HR1, interfering with HR2 recognition of HR1 (Kandeel et al., 2020). These compounds are considered the first generation of MERS-CoV small-molecule fusion inhibitors. In another study

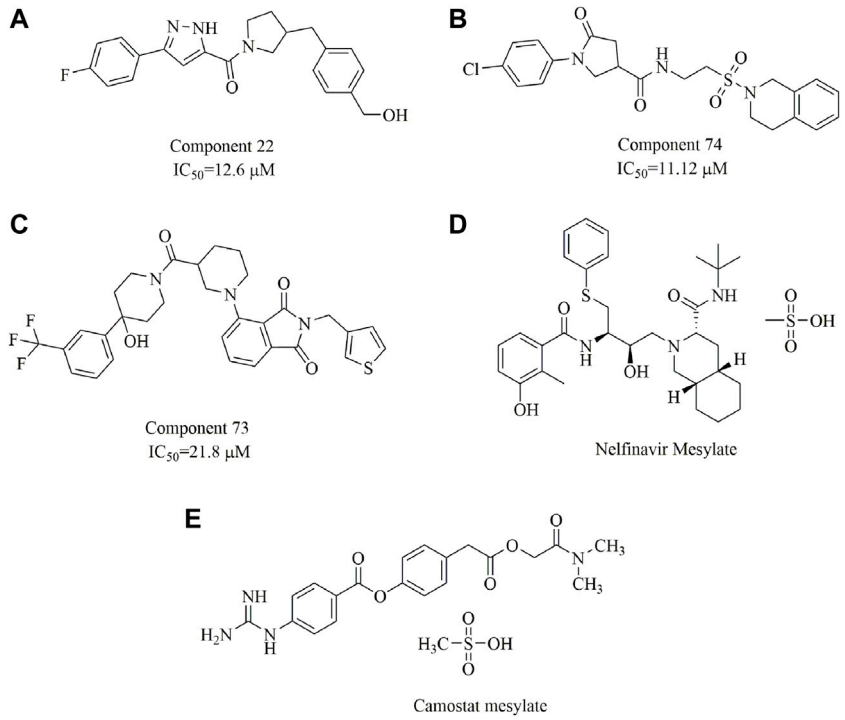


FIGURE 14
Chemical structures of the identified small-molecule fusion inhibitors against MERS-CoV and their inhibitory properties.

(Musarrat et al., 2020), nelfinavir mesylate (Viracept, an anti-HIV drug, Figure 14D) was found to be a potent inhibitor of cell fusion caused by the SARS-CoV-2 S protein, with complete inhibition even at 10 μM. Markus et al. discovered that the S protein of SARS-CoV-2 is cleaved by the serine protease TMPRSS2 and that cell entry could be inhibited by the clinically proven protease inhibitor camostat mesylate (Figure 14E (Hoffmann et al., 2020). Additionally, ligands have been suggested to bind to the interfaces of the trimeric structure of the SARS-CoV-2 S protein and may destabilize the quaternary S protein structure, thereby interfering with the SARS-CoV-2 life cycle (Bongini et al., 2020). In our opinion, this is of importance for the discovery of promising drug candidates but requires evidence-based support.

3.2.4 Application of machine learning for COVID-19 drug discovery

Machine learning (ML) and deep learning (DL) algorithms as two of the most widely used artificial intelligence technology could also be applied to predict drug-target interactions and then validate the predicted drugs in terms of chemical, biological, and physical characteristics based on the various predictive models (Patel et al., 2020; Crampon et al., 2022). To date, two main types of ML algorithms have been developed including: supervised learning (from training samples with known labels) and unsupervised learning (from training samples without known labels) (Rifaioglu et al., 2019). In a study by Batra et al. (2020), a rational screening strategy was developed combining machine learning-based models and high-fidelity ensemble docking studies. Firstly, viable targets for drug discovery wase determined as SARS-CoV-2 S-protein at its host receptor region or the S-protein: human ACE2 interface. Then, the Vina scores were estimated by random forest (RF) regression models for the construction of molecular descriptors, which were applied to represent the molecules for the development of the ML models. The validated ML models were then used for virtual screening ligands from drug and biomolecule data sets. Top scoring 187 hits (75 FDA-approved) were further validated by all atom docking studies, and important molecular descriptors and promising chemical fragments are identified to guide future experiments. Coveney et al. designed a novel *in silico* method for drug design by coupling ML with physics-based (PB) simulations (Bhati et al., 2021). The accurate PB simulations would make the drug design process smarter by calculating the binding free energies of obtained hits from the output of a deep learning (DL) algorithm, which will then fed back to the DL algorithm to improve its predictive performance. Recently, a machine-learning method was proposed capable of identifying drug mechanism of actions based on the cell image features (Han et al., 2021). In this method, the supervised information theoretic

metric-learning (ITML) algorithm was used for converting the characteristics of drugs with similar mechanism of actions clustered by affinity propagation algorithm. Therefore, this method would be more useful in the development of candidates with similar action mechanisms.

Undoubtedly, these results clearly demonstrate the power and efficiency of the ML-based screening. However, in such studies, development of accurate and reliable ML models is the key part of a successful ML-based strategy (Lv et al., 2021), including the data quality and algorithm design. Therefore, it is important to train and validate the models over established data sets.

4 Conclusion and perspectives

Since 2019, the outbreak of SARS-CoV-2 has posed a global health emergency. The high morbidity and mortality associated with COVID-19, especially the lack of an approved efficient drug or vaccine for SARS-CoV-2, presents the urgent need for developing standard antiviral therapies. Drug development to counter COVID-19 could be streamlined by targeting different viral proteins, especially the S protein, which is an effective choice to interfere with viral entry into host cells. Computer-aided modern drug development provides a time- and effort-saving alternative for hit identification, lead optimization and rational drug design. However, just as every coin has two sides, computer-aided drug development also has its own disadvantages. Virtual screening aided by structure-based docking has inherent deficiencies caused by various factors, such as the lack of crystal structures of target proteins and the influence of various conformations and pockets, which could lead to false-negative results. Hence, improving the performance and the accuracy of the computational resources to streamline the workflow used is still needed. In addition, to maximum full play the function of the computational resources, deeper insights into the Spike protein structure, function, and interactions with ACE2 is still essential. Especially, it has demonstrated that specific mutations in the S protein will greatly influence its infectivity, transmissibility, virulence. Therefore, more intensively studies on the adaptive evolutionary mechanisms will further help develop proper and more efficient strategies to fight SARS-CoV-2.

To date, most studies focused on novel drug development for the prevention and treatment of COVID-19 are usually at the *in vitro* experimental stage in the absence of actual *in vivo* data. This might also result in the discovery of nonfunctional ligands in animal or *in vivo* experiments. Therefore, a more efficient strategy should also be investigated by integrating computational resources with *in vivo* experiments. In this way, it would to the greatest extent avoid the false positives and thereby maximize the odds of success in following development process.

Recently, multitarget drugs (MTDs) have attracted great attention due to their advantages in the treatment of complex diseases such as Alzheimer's disease. Since ACE2 is a multifunctional protein, MTDs targeting several sub-pathologies simultaneously might present a better approach for the treatment of COVID-19. While it is conceivable that rational design of MTDs with excellent performance against SARS-CoV-2 is a huge challenge for medicinal chemists; it demonstrates great potential and provides a promising method for treating complex diseases, including COVID-19.

Significantly, it has been reported that phospholipidosis was a common mechanism underlying the antiviral activity of many repurposed drugs (Tummino et al., 2021). Therefore, one the one hand as mentioned above, adequate, and timely *in vitro* tests would be more important for the detection of phospholipidosis and elimination of the identified false positives in early drug discovery. On the other hand, to avoid phospholipidosis, drug discovery or screening of antiviral drugs should be focused more on the target-directed mechanism as highlighted in this review.

In summary, this review is expected to provide a potential framework for designing and developing promising anti-SARS-CoV-2 therapeutics.

Author contributions

TW conceived and designed the research; TC, B-YL, JC, and TS searched, collected, and synthesized the related literatures; X-HL and TW wrote the manuscript. All authors read and approved the manuscript.

Funding

The authors are supported by the Research Fund for Academician Lin He New Medicine (JYHL2021MS23), Supporting Fund for Teachers' research of Jining Medical University (JYFC2019KJ012), and the Natural Science Foundation of Rizhao (RZ2021ZR21). The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from the one mentioned above.

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

References

Abate, C., Niso, M., Abatematteo, F. S., Contino, M., Colabufo, N. A., and Berardi, F. (2020). PB28, the sigma-1 and sigma-2 receptors modulator with potent anti-SARS-CoV-2 activity: A review about its pharmacological properties and structure affinity relationships. *Front. Pharmacol.* 11, 589810. doi:10.3389/fphar.2020.589810

Abraham, M. J., Fleming, K. L., Raymond, S., Wong, A. Y. C., and Bergeron, R. (2019). The sigma-1 receptor behaves as an atypical auxiliary subunit to modulate the functional characteristics of Kv1.2 channels expressed in HEK293 cells. *Physiol. Rep.* 7, e14147. doi:10.14814/phy2.14147

Acharya, A., Pandey, K., Thurman, M., Klug, E., Trivedi, J., Sharma, K., et al. (2021). Discovery and evaluation of entry inhibitors for SARS-CoV-2 and its emerging variants. *J. Virol.* 95, e0143721. doi:10.1128/JVI.01437-21

Al-Sehemi, A. G., Olotu, F. A., Dev, S., Pannipara, M., Soliman, M. E., Carradori, S., et al. (2020). Natural products database screening for the discovery of naturally occurring SARS-cov-2 spike glycoprotein blockers. *ChemistrySelect* 5, 13309–13317. doi:10.1002/slct.202003349

Alsulami, A. F., Thomas, S. E., Jamasb, A. R., Beaudoin, C. A., Moghul, I., Bannerman, B., et al. (2021). SARS-CoV-2 3D database: Understanding the coronavirus proteome and evaluating possible drug targets. *Brief. Bioinform.* 22, 769–780. doi:10.1093/bib/bbaa404

Andersen, K. G., Rambaut, A., Lipkin, W. I., Holmes, E. C., and Garry, R. F. (2020). The proximal origin of SARS-CoV-2. *Nat. Med.* 26, 450–452. doi:10.1038/s41591-020-0820-9

Bai, Q., Tan, S., Xu, T., Liu, H., Huang, J., and Yao, X. (2021). MolAICal: A soft tool for 3D drug design of protein targets by artificial intelligence and classical algorithm. *Brief. Bioinform.* 22, bbaa161. doi:10.1093/bib/bbaa161

Basu, A., Sarkar, A., and Maulik, U. (2020). Molecular docking study of potential phytochemicals and their effects on the complex of SARS-CoV2 spike protein and human ACE2. *Sci. Rep.* 10, 17699. doi:10.1038/s41598-020-74715-4

Batra, R., Chan, H., Kamath, G., Ramprasad, R., Cherukara, M. J., and Sankaranarayanan, S. K. R. S. (2020). Screening of therapeutic agents for COVID-19 using machine learning and ensemble docking studies. *J. Phys. Chem. Lett.* 11, 7058–7065. doi:10.1021/acs.jpclett.0c02278

Benton, D. J., Wrobel, A. G., Xu, P., Roustan, C., Martin, S. R., Rosenthal, P. B., et al. (2020). Receptor binding and priming of the spike protein of SARS-CoV-2 for membrane fusion. *Nature* 588, 327–330. doi:10.1038/s41586-020-2772-0

Bhati, A. P., Wan, S., Alfe, D., Clyde, A. R., Bode, M., Tan, L., et al. (2021). Pandemic drugs at pandemic speed: Infrastructure for accelerating COVID-19 drug discovery with hybrid machine learning- and physics-based simulations on high-performance computers. *Interface focus* 11, 20210018. doi:10.1098/rsfs.2021.0018

Bhushan, G., Lim, L., Bird, I., Chothe, S. K., Nissly, R. H., and Kuchipudi, S. V. (2020). Iminosugars with endoplasmic reticulum α-glucosidase inhibitor activity inhibit ZIKV replication and reverse cytopathogenicity *in vitro*. *Front. Microbiol.* 11, 531. doi:10.3389/fmicb.2020.00531

Bojadzic, D., Alcazar, O., and Buchwald, P. (2020). Methylene blue inhibits the SARS-CoV-2 spike-ACE2 protein-protein interaction-a mechanism that can contribute to its antiviral activity against COVID-19. *Front. Pharmacol.* 11, 600372. doi:10.3389/fphar.2020.600372

Bojadzic, D., Alcazar, O., Chen, J., Chuang, S. T., Condor Capcha, J. M., Shehadeh, L. A., et al. (2021). Small-molecule inhibitors of the coronavirus spike: ACE2 protein-protein interaction as blockers of viral attachment and entry for SARS-CoV-2. *ACS Infect. Dis.* 7, 1519–1534. doi:10.1021/acsinfecdis.1c00070

Bojadzic, D., and Buchwald, P. (2018). Toward small-molecule inhibition of protein-protein interactions: General aspects and recent progress in targeting costimulatory and coinhibitory (immune checkpoint) interactions. *Curr. Top. Med. Chem.* 18, 674–699. doi:10.2174/1568026618666180531092503

Bongini, P., Trezza, A., Bianchini, M., Spiga, O., and Niccolai, N. (2020). A possible strategy to fight COVID-19: Interfering with spike glycoprotein trimerization. *Biochem. Biophys. Res. Commun.* 528, 35–38. doi:10.1016/j.bbrc.2020.04.007

Cai, Y., Zhang, J., Xiao, T., Peng, H., Sterling, S. M., Walsh, R. M., et al. (2020). Distinct conformational states of SARS-CoV-2 spike protein. *Sci. (New York, N.Y.)* 369, 1586–1592. doi:10.1126/science.abd4251

Cao, S., Ipomoeassins, A-E., Guza, R. C., Wisse, J. H., Miller, J. S., Evans, R., et al. (2005). Ipomoeassins A-E, cytotoxic macrocyclic glycoresins from the leaves of Ipomoea squamosa from the Suriname rainforest. *J. Nat. Prod.* 68, 487–492. doi:10.1021/np049629w

Cao, Y., Wang, J., Jian, F., Xiao, T., Song, W., Yisimayi, A., et al. (2022). Omicron escapes the majority of existing SARS-CoV-2 neutralizing antibodies. *Nature* 602, 657–663. doi:10.1038/s41586-021-04385-3

Cerutti, G., Guo, Y., Zhou, T., Gorman, J., Lee, M., Rapp, M., et al. (2021). Potent SARS-CoV-2 neutralizing antibodies directed against spike N-terminal domain target a single supersite. *Cell Host Microbe* 29, 819–833.e7. doi:10.1016/j.chom.2021.03.005.e817

Chan, J. F., Kok, K. H., Zhu, Z., Chu, H., To, K. K. W., Yuan, S., et al. (2020). Genomic characterization of the 2019 novel human-pathogenic coronavirus isolated from a patient with atypical pneumonia after visiting Wuhan. *Emerg. Microbes Infect.* 9, 221–236. doi:10.1080/22221751.2020.1719902

Chang, J., Block, T. M., and Guo, J. T. (2013). Antiviral therapies targeting host ER alpha-glucosidases: Current status and future directions. *Antivir. Res.* 99, 251–260. doi:10.1016/j.antiviral.2013.06.011

Chen, T. F., Chang, Y. C., Hsiao, Y., Lee, K. H., Hsiao, Y. C., Lin, Y. H., et al. (2021). DockCoV2: A drug database against SARS-CoV-2. *Nucleic Acids Res.* 49, D1152–d1159. doi:10.1093/nar/gkaa861

Chen, Z., Zhang, X., Peng, C., Wang, J., Xu, Z., Chen, K., et al. (2019). D3Pockets: A method and web server for systematic analysis of protein pocket dynamics. *J. Chem. Inf. Model.* 59, 3353–3358. doi:10.1021/acs.jcim.9b00332

Chi, X., Yan, R., Zhang, J., Zhang, G., Zhang, Y., Hao, M., et al. (2020). A neutralizing human antibody binds to the N-terminal domain of the Spike protein of SARS-CoV-2. *Science* 369, 650–655. doi:10.1126/science.abc6952

Cho, A., Muecksch, F., Schaefer-Babajew, D., Wang, Z., Finkin, S., Gaebler, C., et al. (2021). Anti-SARS-CoV-2 receptor-binding domain antibody evolution after mRNA vaccination. *Nature* 600, 517–522. doi:10.1038/s41586-021-04060-7

Choudhary, S., Malik, Y. S., and Tomar, S. (2020). Identification of SARS-CoV-2 cell entry inhibitors by drug repurposing using *in silico* structure-based virtual screening approach. *Front. Immunol.* 11, 1664. doi:10.3389/fimmu.2020.01664

Chowdhury, R., Sai Sreyas Adury, V., Vijay, A., Singh, R. K., and Mukherjee, A. (2021). Atomistic de-novo inhibitor generation-guided drug repurposing for SARS-CoV-2 spike protein with free-energy validation by well-tempered metadynamics. *Chem. Asian J.* 16, 1634–1642. doi:10.1002/asia.202100268

Ciuffreda, L., Lorenzo-Salazar, J. M., Alcoba-Florez, J., Rodriguez-Perez, H., Gil-Campesino, H., Inigo-Campos, A., et al. (2021). Longitudinal study of a SARS-CoV-2 infection in an immunocompromised patient with X-linked agammaglobulinemia. *J. Infect.* 83, 607–635. doi:10.1016/j.jinf.2021.07.028

Coutard, B., Valle, C., de Lamballerie, X., Canard, B., Seidah, N. G., and DEcroly, E. (2020). The spike glycoprotein of the new coronavirus 2019-nCoV contains a furin-like cleavage site absent in CoV of the same clade. *Antivir. Res.* 176, 104742. doi:10.1016/j.antiviral.2020.104742

Crampon, K., Giorkallos, A., Deldossi, M., Baud, S., and Steffenel, L. A. (2022). Machine-learning methods for ligand-protein molecular docking. *Drug Discov. Today* 27, 151–164. doi:10.1016/j.drudis.2021.09.007

Dabholkar, N., Gorantla, S., Dubey, S. K., Alexander, A., Taliyan, R., and Singhvi, G. (2021). Repurposing methylene blue in the management of COVID-19: Mechanistic aspects and clinical investigations. *Biomed. Pharmacother. = Biomedecine Pharmacother.* 142, 112023. doi:10.1016/j.biopha.2021.112023

Di Gaetano, S., Capasso, D., Delre, P., Pirone, L., Saviano, M., Pedone, E., et al. (2021). More is always better than one: The N-terminal domain of the spike protein as another emerging target for hampering the SARS-CoV-2 attachment to host cells. *Int. J. Mol. Sci.* 22, 6462. doi:10.3390/ijms22126462

Djekidel, M. N., Rosikiewicz, W., Peng, J. C., Kanneganti, T.-D., Hui, Y., Jin, H., et al. CovidExpress: An interactive portal for intuitive investigation on SARS-CoV-2 related transcriptomes. bioRxiv: the preprint server for biology (2021).

Dowall, S. D., Bewley, K., Watson, R. J., Vasan, S. S., Ghosh, C., Konai, M. M., et al. (2016). Antiviral screening of multiple compounds against Ebola virus. *Viruses* 8, E277. doi:10.3390/v8110277

Downing, N. S., Shah, N. D., Aminawung, J. A., Pease, A. M., Zeitoun, J. D., Krumholz, H. M., et al. (2017). Postmarket safety events among novel therapeutics approved by the US food and drug administration between 2001 and 2010. *Jama* 317, 1854–1863. doi:10.1001/jama.2017.5150

Du, L., Yang, Y., Zhou, Y., Lu, L., Li, F., and Jiang, S. (2017). MERS-CoV spike protein: A key target for antivirals. *Expert Opin. Ther. Targets* 21, 131–143. doi:10.1080/14728222.2017.1271415

Fan, H. H., Wang, L. Q., Liu, W. L., An, X. P., Liu, Z. D., He, X. Q., et al. (2020). Repurposing of clinically approved drugs for treatment of coronavirus disease 2019 in a 2019-novel coronavirus-related coronavirus model. *Chin. Med. J.* 133, 1051–1056. doi:10.1097/CM9.0000000000000797

Feng, Z., Chen, M., Liang, T., Shen, M., Chen, H., and Xie, X. Q. (2021). VirusCKB: An integrated bioinformatics platform and analysis resource for COVID-19 research. *Brief. Bioinform.* 22, 882–895. doi:10.1093/bib/bbaa155

Gangadevi, S., Badavath, V. N., Thakur, A., Yin, N., De Jonghe, S., Acevedo, O., et al. (2021). Kobophenol A inhibits binding of host ACE2 receptor with spike RBD domain of SARS-CoV-2, a lead compound for blocking COVID-19. *J. Phys. Chem. Lett.* 12, 1793–1802. doi:10.1021/acs.jpclett.0c03119

Gendrot, M., Andreani, J., Duflot, I., Boxberger, M., Le Bideau, M., Mosnier, J., et al. (2020). Methylene blue inhibits replication of SARS-CoV-2 *in vitro*. *Int. J. Antimicrob. Agents* 56, 106202. doi:10.1016/j.ijantimicag.2020.106202

Gentile, F., Fernandez, M., Ban, F., Ton, A. T., Mslati, H., Perez, C. F., et al. (2021). Automated discovery of noncovalent inhibitors of SARS-CoV-2 main protease by consensus Deep Docking of 40 billion small molecules. *Chem. Sci.* 12, 15960–15974. doi:10.1039/d1sc05579h

Ghahestani, S. M., Shahab, E., Karimi, S., and Madani, M. H. (2020). Methylene blue may have a role in the treatment of COVID-19. *Med. Hypotheses* 144, 110163. doi:10.1016/j.mehy.2020.110163

Gordon, D. E., Jang, G. M., Bouhaddou, M., Xu, J., Obernier, K., White, K. M., et al. (2020). A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. *Nature* 583, 459–468. doi:10.1038/s41586-020-2286-9

Gowthaman, R., Guest, J. D., Yin, R., Adolf-Bryfogle, J., Schief, W. R., and Pierce, B. G. (2021). CoV3D: A database of high resolution coronavirus protein structures. *Nucleic Acids Res.* 49, D282–d287. doi:10.1093/nar/gkaa731

Gyebi, G. A., Adegunloye, A. P., Ibrahim, I. M., Ogunyemi, O. M., Afolabi, S. O., and Ogunro, O. B. (2022). Prevention of SARS-CoV-2 cell entry: Insight from *in silico* interaction of drug-like alkaloids with spike glycoprotein, human ACE2, and TMPRSS2. *J. Biomol. Struct. Dyn.* 40, 2121–2145. doi:10.1080/07391102.2020.1835726

Hall-Swan, S., Devaurs, D., Rigo, M. M., Antunes, D. A., Kavraki, L. E., and Zanatta, G. (2021). DINC-COVID: A webserver for ensemble docking with flexible SARS-CoV-2 proteins. *Comput. Biol. Med.* 139, 104943. doi:10.1016/j.compbiomed.2021.104943

Han, L., Shan, G., Chu, B., Wang, H., Wang, Z., Gao, S., et al. (2021). Accelerating drug repurposing for COVID-19 treatment by modeling mechanisms of action using cell image features and machine learning. *Cogn. neurodynamics*, 1–9. doi:10.1007/s11571-021-09727-5

Han, P., Li, L., Liu, S., Wang, Q., Zhang, D., Xu, Z., et al. (2022). Receptor binding and complex structures of human ACE2 to spike RBD from omicron and delta SARS-CoV-2. *Cell* 185, 630–640.e10. doi:10.1016/j.cell.2022.01.001.630

Hatmal, M. M., Alshaer, W., Al-Hatamleh, M. A. I., Hatmal, M., Smadi, O., Taha, M. O., et al. (2020). Comprehensive structural and molecular comparison of spike proteins of SARS-CoV-2, SARS-CoV and MERS-CoV, and their interactions with ACE2. *Cells* 9, E2638. doi:10.3390/cells9122638

Hoffmann, M., Kleine-Weber, H., Schroeder, S., Kruger, N., Herrler, T., Erichsen, S., et al. (2020). SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. *Cell* 181, 271–280. doi:10.1016/j.cell.2020.02.052.e278

Holman, W., Holman, S., McIntosh, S., Painter, W., Painter, G., Bush, J., et al. (2021). Accelerated first-in-human clinical trial of EIDD-2801/MK-4482 (molnupiravir), a ribonucleoside analog with potent antiviral activity against SARS-CoV-2. *Trials* 22, 561. doi:10.1186/s13063-021-05538-5

Hong, Q., Han, W., Li, J., Xu, S., Wang, Y., Xu, C., et al. (2022). Molecular basis of receptor binding and antibody neutralization of Omicron. *Nature* 604, 546–552. doi:10.1038/s41586-022-04581-9

Hu, X., Chen, C. Z., Xu, M., Hu, Z., Guo, H., Itkin, Z., et al. (2021). Discovery of small molecule entry inhibitors targeting the fusion peptide of SARS-CoV-2 spike protein. *ACS Med. Chem. Lett.* 12, 1267–1274. doi:10.1021/acsmedchemlett.1c00263

Hu, X., Shrimp, J. H., Guo, H., Xu, M., Chen, C. Z., Zhu, W., et al. (2021). Discovery of TMPRSS2 inhibitors from virtual screening as a potential treatment of COVID-19. *ACS Pharmacol. Transl. Sci.* 4, 1124–1135. doi:10.1021/acsptsci.0c00221

Jeon, S., Ko, M., Lee, J., Choi, I., Byun, S. Y., Park, S., et al. (2020). Identification of antiviral drug candidates against SARS-CoV-2 from FDA-approved drugs. *Antimicrob. Agents Chemother.* 64, e00819–20. doi:10.1128/AAC.00819-20

Jiang, S., Du, L., and Shi, Z. (2020). An emerging coronavirus causing pneumonia outbreak in wuhan, China: Calling for developing therapeutic and prophylactic strategies. *Emerg. Microbes Infect.* 9, 275–277. doi:10.1080/22221751.2020.1723441

Ju, B., Zhang, Q., Ge, J., Wang, R., Sun, J., Ge, X., et al. (2020). Human neutralizing antibodies elicited by SARS-CoV-2 infection. *Nature* 584, 115–119. doi:10.1038/s41586-020-2380-z

Kandeel, M., Yamamoto, M., Al-Taher, A., Watanabe, A., Oh-Hashi, K., Park, B. K., et al. (2020). Small molecule inhibitors of Middle East respiratory syndrome coronavirus fusion by targeting cavities on heptad repeat trimers. *Biomol. Ther.* 28, 311–319. doi:10.4062/biomolther.2019.202

Kannan, S., Shaik Syed Ali, P., and Sheeza, A. (2021). Omicron (B.1.1.529) - variant of concern - molecular profile and epidemiology: A mini review. *Eur. Rev. Med. Pharmacol. Sci.* 25, 8019–8022. doi:10.26355/eurrev_202112_27653

Karamyan, V. T. (2021). Between two storms, vasoactive peptides or bradykinin underlie severity of COVID-19? *Physiol. Rep.* 9, e14796. doi:10.14814/phy2.14796

Kiew, L. V., Chang, C. Y., Huang, S. Y., Wang, P. W., Heh, C. H., Liu, C. T., et al. (2021). Development of flexible electrochemical impedance spectroscopy-based biosensing platform for rapid screening of SARS-CoV-2 inhibitors. *Biosens. Bioelectron.* 183, 113213. doi:10.1016/j.bios.2021.113213

Ko, M., Jeon, S., Ryu, W.-S., and Kim, S. Comparative analysis of antiviral efficacy of FDA-approved drugs against SARS-CoV-2 in human lung cells: Nafamostat is the most potent antiviral drug candidate. bioRxiv : the preprint server for biology, 2020.2005.2012.090035 (2020).

Kong, R., Yang, G., Xue, R., Liu, M., Wang, F., Hu, J., et al. (2020). COVID-19 docking server: A meta server for docking small molecules, peptides and antibodies against potential targets of COVID-19. *Bioinforma. Oxf. Engl.* 36, 5109–5111. doi:10.1093/bioinformatics/btaa645

Lai, C. C., Shih, T. P., Ko, W. C., Tang, H. J., and Hsueh, P. R. (2020). Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and coronavirus disease-2019 (COVID-19): The epidemic and the challenges. *Int. J. Antimicrob. Agents* 55, 105924. doi:10.1016/j.ijantimicag.2020.105924

Lan, J., Ge, J., Yu, J., Shan, S., Zhou, H., Fan, S., et al. (2020). Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *Nature* 581, 215–220. doi:10.1038/s41586-020-2180-5

Li, F., Li, W., Farzan, M., and Harrison, S. C. (2005). Structure of SARS coronavirus spike receptor-binding domain complexed with receptor. *Science* 309, 1864–1868. doi:10.1126/science.1116480

Li, S., Liu, W., Chen, Y., Wang, L., An, W., An, X., et al. (2021). Transcriptome analysis of cepharanthine against a SARS-CoV-2-related coronavirus. *Brief. Bioinform.* 22, 1378–1386. doi:10.1093/bib/bbaa387

Liu, X. H., Zhang, X., Lu, Z. H., Zhu, Y. S., and Wang, T. (2021). Potential molecular targets of nonstructural proteins for the development of antiviral drugs against SARS-CoV-2 infection. *Biomed. Pharmacother.* 133, 111035. doi:10.1016/j.biopha.2020.111035

Lu, G., Hu, Y., Wang, Q., Qi, J., Gao, F., Li, Y., et al. (2013). Molecular basis of binding between novel human coronavirus MERS-CoV and its receptor CD26. *Nature* 500, 227–231. doi:10.1038/nature12328

Lucas, C., Vogels, C. B. F., Yildirim, I., Rothman, J. E., Lu, P., Monteiro, V., et al. (2021). Impact of circulating SARS-CoV-2 variants on mRNA vaccine-induced immunity. *Nature* 600, 523–529. doi:10.1038/s41586-021-04085-y

Lupala, C. S., Ye, Y., Chen, H., Su, X. D., and Liu, H. (2022). Mutations on RBD of SARS-CoV-2 Omicron variant result in stronger binding to human ACE2 receptor. *Biochem. Biophys. Res. Commun.* 590, 34–41. doi:10.1016/j.bbrc.2021.12.079

Lv, H., Shi, L., Berkenpas, J. W., Dao, F. Y., Zulfiqar, H., Ding, H., et al. (2021). Application of artificial intelligence and machine learning for COVID-19 drug discovery and vaccine design. *Brief. Bioinform.* 22, bbab320. doi:10.1093/bib/bbab320

Mahmudpour, M., Nabipour, I., Keshavarz, M., and Farrokhnia, M. (2021). Virtual screening on marine natural products for discovering TMPRSS2 inhibitors. *Front. Chem.* 9, 722633. doi:10.3389/fchem.2021.722633

Mannar, D., Saville, J. W., Zhu, X., Srivastava, S. S., Berezuk, A. M., Tuttle, K. S., et al. (2022). SARS-CoV-2 Omicron variant: Antibody evasion and cryo-EM structure of spike protein-ACE2 complex. *Science* 375, 760–764. doi:10.1126/science.abn7760

Martin, R., Lochel, H. F., Welzel, M., Hattab, G., Hauschild, A. C., and Heider, D. (2020). CORDITE: The curated CORona drug InTERactions database for SARS-CoV-2. *iScience* 23, 101297. doi:10.1016/j.isci.2020.101297

McCallum, M., Czudnochowski, N., Rosen, L. E., Zepeda, S. K., Bowen, J. E., Walls, A. C., et al. (2022). Structural basis of SARS-CoV-2 Omicron immune

evasion and receptor engagement. *Science* 375, 864–868. doi:10.1126/science.abn8652

McCallum, M., De Marco, A., Lempp, F. A., Tortorici, M. A., Pinto, D., Walls, A. C., et al. (2021). N-terminal domain antigenic mapping reveals a site of vulnerability for SARS-CoV-2. *Cell* 184, 2332–2347.e16. doi:10.1016/j.cell.2021.03.028.2332

Meirson, T., Bomze, D., and Markel, G. (2021). Structural basis of SARS-CoV-2 spike protein induced by ACE2. *Bioinforma. Oxf. Engl.* 37, 929–936. doi:10.1093/bioinformatics/btaa744

Meo, S. A., Meo, A. S., Al-Jassir, F. F., and Klonoff, D. C. (2021). Omicron SARS-CoV-2 new variant: Global prevalence and biological and clinical characteristics. *Eur. Rev. Med. Pharmacol. Sci.* 25, 8012–8018. doi:10.26355/eurrev_202112_27652

Mishra, P. M., and Nandi, C. K. (2021). Structural decoding of a small molecular inhibitor on the binding of SARS-CoV-2 to the ACE 2 receptor. *J. Phys. Chem. B* 125, 8395–8405. doi:10.1021/acs.jpcb.1c03294

Musarrat, F., Chouljenko, V., Dahal, A., Nabi, R., Chouljenko, T., Jois, S. D, et al. (2020). The anti-HIV drug nelfinavir mesylate (Viracept) is a potent inhibitor of cell fusion caused by the SARSCoV-2 spike (S) glycoprotein warranting further evaluation as an antiviral against COVID-19 infections. *J. Med. Virol.*

O'Keefe, S., Roboti, P., Duah, K. B., Zong, G., Schneider, H., Shi, W. Q., et al. (2021). Ipomoeassin-F inhibits the *in vitro* biogenesis of the SARS-CoV-2 spike protein and its host cell membrane receptor. *J. Cell Sci.* 134, jcs257758. doi:10.1242/jcs.257758

Ortega-Roldan, J. L., Ossa, F., and Schnell, J. R. (2013). Characterization of the human sigma-1 receptor chaperone domain structure and binding immunoglobulin protein (BiP) interactions. *J. Biol. Chem.* 288, 21448–21457. doi:10.1074/jbc.M113.450379

Ou, X., Liu, Y., Lei, X., Li, P., Mi, D., Ren, L., et al. (2020). Characterization of spike glycoprotein of SARS-CoV-2 on virus entry and its immune cross-reactivity with SARS-CoV. *Nat. Commun.* 11, 1620. doi:10.1038/s41467-020-15562-9

Pan, C., Yang, H., Lu, Y., Hu, S., Wu, Y., He, Q., et al. (2021). Recent advance of peptide-based molecules and nonpeptidic small-molecules modulating PD-1/PD-L1 protein-protein interaction or targeting PD-L1 protein degradation. *Eur. J. Med. Chem.* 213, 113170. doi:10.1016/j.ejmech.2021.113170

Pandey, P., Rane, J. S., Chatterjee, A., Kumar, A., Khan, R., Prakash, A., et al. (2021). Targeting SARS-CoV-2 spike protein of COVID-19 with naturally occurring phytochemicals: An *in silico* study for drug development. *J. Biomol. Struct. Dyn.* 39, 6306–6316. doi:10.1080/07391102.2020.1796811

Patel, L., Shukla, T., Huang, X., Ussery, D. W., and Wang, S. (2020). Machine Learning Methods in Drug Discovery, *Mol. (Basel, Switz.*, 25. 5277. doi:10.3390/molecules25225277

Pati, M. L., Hornick, J. R., Niso, M., Berardi, F., Spitzer, D., Abate, C., et al. (2017). Sigma-2 receptor agonist derivatives of 1-Cyclohexyl-4-[3-(5-methoxy-1, 2, 3, 4-tetrahydronaphthalen-1-yl)propyl]piperazine (PB28) induce cell death via mitochondrial superoxide production and caspase activation in pancreatic cancer. *BMC cancer* 17, 51. doi:10.1186/s12885-016-3040-4

Peng, C., Zhu, Z., Shi, Y., Wang, X., Mu, K., Yang, Y., et al. (2020). Computational insights into the conformational accessibility and binding strength of SARS-CoV-2 spike protein to human angiotensin-converting enzyme 2. *J. Phys. Chem. Lett.* 11, 10482–10488. doi:10.1021/acs.jpclett.0c02958

Pillay, T. S. (2020). Gene of the month: The 2019-nCoV/SARS-CoV-2 novel coronavirus spike protein. *J. Clin. Pathol.* 73, 366–369. doi:10.1136/jclinpath-2020-206658

Raj, V. S., Mou, H., Smits, S. L., Dekkers, D. H. W., Muller, M. A., Dijkman, R., et al. (2013). Dipeptidyl peptidase 4 is a functional receptor for the emerging human coronavirus-EMC. *Nature* 495, 251–254. doi:10.1038/nature12005

Raybould, M. I. J., Kovaltsuk, A., Marks, C., and Deane, C. M. (2021). CoV-AbDab: The coronavirus antibody database. *Bioinforma. Oxf. Engl.* 37, 734–735. doi:10.1093/bioinformatics/btaa739

Rifaioglu, A. S., Atas, H., Martin, M. J., Cetin-Atalay, R., Atalay, V., and Dogan, T. (2019). Recent applications of deep learning and machine intelligence on *in silico* drug discovery: Methods, tools and databases. *Brief. Bioinform.* 20, 1878–1912. doi:10.1093/bib/bby061

Risner, K. H., Tieu, K. V., Wang, Y., Bakovic, A., Alem, F., Bhalla, S., et al. Maraviroc inhibits SARS-CoV-2 multiplication and s-protein mediated cell fusion in cell culture. *bioRxiv : the preprint server for biology* (2020).

Schooley, R. T., Carlin, A. F., Schooley, J. R., Valiaeva, N., Zhang, X. Q., Clark, A. E., et al. Rethinking remdesivir: Synthesis, antiviral activity and pharmacokinetics of oral lipid prodrugs. *bioRxiv : the preprint server for biology* (2021).

Schuurs, Z. P., Hammond, E., Elli, S., Rudd, T. R., Mycroft-West, C. J., Lima, M. A., et al. (2021). Evidence of a putative glycosaminoglycan binding site on the glycosylated SARS-CoV-2 spike protein N-terminal domain. *Comput. Struct. Biotechnol. J.* 19, 2806–2818. doi:10.1016/j.csbj.2021.05.002

Scott, D. E., Bayly, A. R., Abell, C., and Skidmore, J. (2016). Small molecules, big targets: Drug discovery faces the protein-protein interaction challenge. *Nat. Rev. Drug Discov.* 15, 533–550. doi:10.1038/nrd.2016.29

Shi, Y., Zhang, X., Mu, K., Peng, C., Zhu, Z., Wang, X., et al. (2020). D3Targets-2019-nCoV: A webserver for predicting drug targets and for multi-target and multi-site based virtual screening against COVID-19. *Acta Pharm. Sin. B* 10, 1239–1248. doi:10.1016/j.apsb.2020.04.006

Smith, M., and Smith, J. C. (2020). *Repurposing therapeutics for the wuhan coronavirus nCov-2019: Supercomputer-based docking to the viral S protein and human ACE2 interface*. ChemRxiv. Cambridge: Cambridge Open Engage.

Song, W., Gui, M., Wang, X., and Xiang, Y. (2018). Cryo-EM structure of the SARS coronavirus spike glycoprotein in complex with its host cell receptor ACE2. *PLoS Pathog.* 14, e1007236. doi:10.1371/journal.ppat.1007236

Song, Y., and Buchwald, P. (2015). TNF superfamily protein-protein interactions: Feasibility of small- molecule modulation. *Curr. Drug Targets* 16, 393–408. doi:10.2174/1389450116666150223115628

Souza, P. F. N., Mesquita, F. P., Amaral, J. L., Landim, P. G. C., Lima, K. R. P., Costa, M. B., et al. (2021). The human pandemic coronaviruses on the show: The spike glycoprotein as the main actor in the coronaviruses play. *Int. J. Biol. Macromol.* 179, 1–19. doi:10.1016/j.ijbiomac.2021.02.203

Starr, T. N., Greaney, A. J., Dingens, A. S., and Bloom, J. D. (2021). Complete map of SARS-CoV-2 RBD mutations that escape the monoclonal antibody LY-CoV555 and its cocktail with LY-CoV016. *Cell Rep. Med.* 2, 100255. doi:10.1016/j.xcrm.2021.100255

Sun, C., Zhang, J., Wei, J., Zheng, X., Zhao, X., Fang, Z., et al. (2021). Screening, simulation, and optimization design of small molecule inhibitors of the SARS-CoV-2 spike glycoprotein. *PLoS One* 16, e0245975. doi:10.1371/journal.pone.0245975

Tai, W., He, L., Zhang, X., Pu, J., Voronin, D., Jiang, S., et al. (2020). Characterization of the receptor-binding domain (RBD) of 2019 novel coronavirus: Implication for development of RBD protein as a viral attachment inhibitor and vaccine. *Cell. Mol. Immunol.* 17, 613–620. doi:10.1038/s41423-020-0400-4

Tan, T. K., Rijal, P., Rahikainen, R., Keeble, A. H., Schimanski, L., Hussain, S., et al. (2021). A COVID-19 vaccine candidate using SpyCatcher multimerization of the SARS-CoV-2 spike protein receptor-binding domain induces potent neutralising antibody responses. *Nat. Commun.* 12, 542. doi:10.1038/s41467-020-20654-7

Tummino, T. A., Rezelj, V. V., Fischer, B., Fischer, A., O'Meara, M. J., Monel, B., et al. (2021). Drug-induced phospholipidosis confounds drug repurposing for SARS-CoV-2. *Science* 373, 541–547. doi:10.1126/science.abi4708

van den Berg, D. F., and Te Velde, A. A. (2020). Severe COVID-19: NLRP3 inflammasome dysregulated. *Front. Immunol.* 11, 1580. doi:10.3389/fimmu.2020.01580

Walls, A. C., Tortorici, M. A., Bosch, B. J., Frenz, B., Rottier, P. J. M., DiMaio, F., et al. (2016). Cryo-electron microscopy structure of a coronavirus spike glycoprotein trimer. *Nature* 531, 114–117. doi:10.1038/nature16988

Wang, N., Shi, X., Jiang, L., Zhang, S., Wang, D., Tong, P., et al. (2013). Structure of MERS-CoV spike receptor-binding domain complexed with human receptor DPP4. *Cell Res.* 23, 986–993. doi:10.1038/cr.2013.92

Wang, P., Casner, R. G., Nair, M. S., Wang, M., Yu, J., Cerutti, G., et al. (2021). Increased resistance of SARS-CoV-2 variant P.1 to antibody neutralization. *Cell Host Microbe* 29, 747–751.e4. doi:10.1016/j.chom.2021.04.007.e744

Wang, P., Nair, M. S., Liu, L., Iketani, S., Luo, Y., Guo, Y., et al. (2021). Antibody resistance of SARS-CoV-2 variants B.1.351 and B.1.1.7. *Nature* 593, 130–135. doi:10.1038/s41586-021-03398-2

Wang, Q., Zhang, Y., Wu, L., Niu, S., Song, C., Zhang, Z., et al. (2020). Structural and functional basis of SARS-CoV-2 entry by using human ACE2. *Cell* 181, 894–904. doi:10.1016/j.cell.2020.03.045.e899

Wang, X., Xin, B., Tan, W., Xu, Z., Li, K., Li, F., et al. (2021). DeepR2cov: Deep representation learning on heterogeneous drug networks to discover anti-inflammatory agents for COVID-19. *Brief. Bioinform.* 22, bbab226. doi:10.1093/bib/bbab226

Wang, Z., Schmidt, F., Weisblum, Y., Muecksch, F., Barnes, C. O., Finkin, S., et al. (2021). mRNA vaccine-elicited antibodies to SARS-CoV-2 and circulating variants. *Nature* 592, 616–622. doi:10.1038/s41586-021-03324-6

Wrapp, D., Wang, N., Corbett, K. S., Goldsmith, J. A., Hsieh, C. L., Abiona, O., et al. (2020). Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Sci. (New York, N.Y.)* 367, 1260–1263. doi:10.1126/science.abb2507

Wu, C., Liu, Y., Yang, Y., Zhang, P., Zhong, W., Wang, Y., et al. (2020). Analysis of therapeutic targets for SARS-CoV-2 and discovery of potential drugs by computational methods. *Acta Pharm. Sin. B* 10, 766–788. doi:10.1016/j.apsb.2020.02.008

Xiu, S., Dick, A., Ju, H., Mirzaie, S., Abdi, F., Cocklin, S., et al. (2020). Inhibitors of SARS-CoV-2 entry: Current and future opportunities. *J. Med. Chem.* 63, 12256–12274. doi:10.1021/acs.jmedchem.0c00502

Xu, C., Wang, Y., Liu, C., Zhang, C., Han, W., Hong, X., et al. (2021). Conformational dynamics of SARS-CoV-2 trimeric spike glycoprotein in complex with receptor ACE2 revealed by cryo-EM. *Sci. Adv.* 7, eabe5575. doi:10.1126/sciadv.abe5575

Yadav, M., Dhagat, S., and Eswari, J. S. (2020). Emerging strategies on *in silico* drug development against COVID-19: Challenges and opportunities. *Eur. J. Pharm. Sci.* 155, 105522. doi:10.1016/j.ejps.2020.105522

Yan, R., Zhang, Y., Li, Y., Xia, L., Guo, Y., and Zhou, Q. (2020). Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2. *Sci. (New York, N.Y.)* 367, 1444–1448. doi:10.1126/science.abb2762

Yang, L., Pei, R. J., Li, H., Ma, X. N., Zhou, Y., Zhu, F. H., et al. (2021b). Identification of SARS-CoV-2 entry inhibitors among already approved drugs. *Acta Pharmacol. Sin.* 42, 1347–1353. doi:10.1038/s41401-020-00556-6

Yang, Y., Zhu, Z., Wang, X., Zhang, X., Mu, K., Shi, Y., et al. (2021a). Ligand-based approach for predicting drug targets and for virtual screening against COVID-19. *Brief. Bioinform.* 22, 1053–1064. doi:10.1093/bib/bbaa422

Yi, H., Wang, J., Wang, J., Lu, Y., Zhang, Y., Peng, R., et al. (2021). The emergence and spread of novel SARS-CoV-2 variants. *Front. Public Health* 9, 696664. doi:10.3389/fpubh.2021.696664

Yin, W., Xu, Y., Xu, P., Cao, X., Wu, C., Gu, C., et al. (2022). Structures of the Omicron spike trimer with ACE2 and an anti-Omicron antibody. *Science* 375, 1048–1053. doi:10.1126/science.abn8863

Yu, S., Zhu, Y., Xu, J., Yao, G., Zhang, P., Wang, M., et al. (2021). Glycyrrhizic acid exerts inhibitory activity against the spike protein of SARS-CoV-2. *Phytomedicine.* 85, 153364. doi:10.1016/j.phymed.2020.153364

Yuan, M., Wu, N. C., Zhu, X., Lee, C. C. D., So, R. T. Y., Lv, H., et al. (2020). A highly conserved cryptic epitope in the receptor binding domains of SARS-CoV-2 and SARS-CoV. *Science* 368, 630–633. doi:10.1126/science.abb7269

Yuan, Y., Cao, D., Zhang, Y., Ma, J., Qi, J., Wang, Q., et al. (2017). Cryo-EM structures of MERS-CoV and SARS-CoV spike glycoproteins reveal the dynamic receptor binding domains. *Nat. Commun.* 8, 15092. doi:10.1038/ncomms15092

Zahradník, J., Marciano, S., Shemesh, M., Zoler, E., Harari, D., Chiaravalli, J., et al. (2021). SARS-CoV-2 variant prediction and antiviral drug design are enabled by RBD *in vitro* evolution. *Nat. Microbiol.* 6, 1188–1198. doi:10.1038/s41564-021-00954-4

Zhang, W., Zhang, Y., Min, Z., Mo, J., Ju, Z., Guan, W., et al. (2022). COVID19db: A comprehensive database platform to discover potential drugs and targets of COVID-19 at whole transcriptomic scale. *Nucleic Acids Res.* 50, D747–d757. doi:10.1093/nar/gkab850

Zhao, Y., Fang, C., Zhang, Q., Zhang, R., Zhao, X., Duan, Y., et al. (2021). Crystal structure of SARS-CoV-2 main protease in complex with protease inhibitor PF-07321332. *Protein & Cell* 13 (9), 689–693.

Zhou, P., Yang, X. L., Wang, X. G., Hu, B., Zhang, L., Zhang, W., et al. (2020). A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* 579, 270–273. doi:10.1038/s41586-020-2012-7

Zong, G., Hu, Z., O'Keefe, S., Tranter, D., Iannotti, M. J., Baron, L., et al. (2019). Ipomoeassin F binds Sec61α to inhibit protein translocation. *J. Am. Chem. Soc.* 141, 8450–8461. doi:10.1021/jacs.8b13506

frontiers | Frontiers in Pharmacology

TYPE Original Research
PUBLISHED 11 August 2022
DOI 10.3389/fphar.2022.978741

OPEN ACCESS

EDITED BY
Mithun Rudrapal,
Rasiklal M. Dhariwal Institute of Pharmaceutical Education and Research, India

REVIEWED BY
Rohini Karunakaran,
AIMST University, Malaysia
Tung Bui Thanh,
Vietnam National University, Vietnam

*CORRESPONDENCE
Wisnu Ananta Kusuma,
ananta@apps.ipb.ac.id

SPECIALTY SECTION
This article was submitted to Drugs Outcomes Research and Policies, a section of the journal Frontiers in Pharmacology

RECEIVED 26 June 2022
ACCEPTED 14 July 2022
PUBLISHED 11 August 2022

CITATION
Kusuma WA, Habibi ZI, Amir MF, Fadli A, Khotimah H, Dewanto V and Heryanto R (2022), Bipartite graph search optimization for type II diabetes mellitus Jamu formulation using branch and bound algorithm.
Front. Pharmacol. 13:978741.
doi: 10.3389/fphar.2022.978741

Bipartite graph search optimization for type II diabetes mellitus Jamu formulation using branch and bound algorithm

Wisnu Ananta Kusuma[1,2]*, Zulfahmi Ibnu Habibi[1], Muhammad Fahmi Amir[1], Aulia Fadli[1], Husnul Khotimah[1], Vektor Dewanto[1] and Rudi Heryanto[2,3]

[1]Department of Computer Science, Faculty of Mathematics and Natural Sciences, IPB University, Bogor, Indonesia, [2]Tropical Biopharmaca Research Center, IPB University, Bogor, Indonesia, [3]Department of Chemistry, Faculty of Mathematics and Natural Sciences, IPB University, Bogor, Indonesia

Jamu is an Indonesian traditional herbal medicine that has been practiced for generations. Jamu is made from various medicinal plants. Each plant has several compounds directly related to the target protein that are directly associated with a disease. A pharmacological graph can form relationships between plants, compounds, and target proteins. Research related to the prediction of Jamu formulas for some diseases has been carried out, but there are problems in finding combinations or compositions of Jamu formulas because of the increase in search space size. Some studies adopted the drug–target interaction (DTI) implemented using machine learning or deep learning to predict the DTI for discovering the Jamu formula. However, this approach raises important issues, such as imbalanced and high-dimensional dataset, overfitting, and the need for more procedures to trace compounds to their plants. This study proposes an alternative approach by implementing bipartite graph search optimization using the branch and bound algorithm to discover the combination or composition of Jamu formulas by optimizing the search on a plant–protein bipartite graph. The branch and bound technique is implemented using the search strategy of breadth first search (BrFS), Depth First Search, and Best First Search. To show the performance of the proposed method, we compared our method with a complete search algorithm, searching all nodes in the tree without pruning. In this study, we specialize in applying the proposed method to search for the Jamu formula for type II diabetes mellitus (T2DM). The result shows that the bipartite graph search with the branch and bound algorithm reduces computation time up to 40 times faster than the complete search strategy to search for a composition of plants. The binary branching strategy is the best choice, whereas the BrFS strategy is the best option in this research. In addition, the the proposed method can suggest the composition of one to four plants for the T2DM Jamu formula. For a combination of four plants, we obtain *Angelica Sinensis*, *Citrus aurantium*, *Glycyrrhiza uralensis*, and *Mangifera indica*. This approach is expected to be an alternative way to discover the Jamu formula more accurately.

KEYWORDS
branch and bound, diabetes mellitus, drug–target interaction, graph traversing, jamu

Introduction

Jamu, known as Indonesian herbal medicine, is local wisdom that must be preserved because it has been practiced for generations (Elfahmi et al., 2014). The 2010 Basic Health Research results show that more than 50% of Indonesians use herbal medicine (Purwaningsih, 2013). Jamu is made from various plants that are considered to have healing properties based on practical experience. Zuhud et al. (2001) identified approximately 1,845 forest plant species in Indonesia that have the potential as medicinal plants. National Agency of Drug and Food Control in Indonesia noted that approximately 283 plant species were officially registered and used for treatment. Thus, Jamu has the potential to be developed. Because of the vast biodiversity of Indonesia's indigenous medicinal plants, herbal medicine has the potential for economic development (Elfahmi et al., 2014). However, this herbal medicine has not been widely used because the discovery of herbal formulas has not been supported by its scientific basis (Noor et al., 2022).

Various efforts to make herbal medicine have a computational-based scientific basis have been carried out. Research on herbal medicine by Afendi et al. (2010) put forward the hypothesis that at least one Jamu formula has a composition of four herbal plants. One main plant directly affects disease, and the other three are supporting plants that have analgesic, antimicrobial, and anti-inflammatory properties. The 3138 herbal formulas taken from 465 plants were classified into nine properties (Afendi et al., 2010). Afendi et al. (2013) continued their research by looking for the relationship between plant composition and herbal medicine efficacy using a statistical approach to classify nine properties of the 3138 Jamu formulas derived from these 465 plants. Classification based on partial least squares discriminant analysis produced an accuracy of 71.6%. Fitriawan et al. (2013) conducted a similar study with a machine learning approach using the support vector machine (SVM) method, resulting in an accuracy of 71%. Puspita et al. (2016) conducted another study that reported the study of feature selection using clustering techniques to reduce the number of irrelevant features before training using SVM.

Prediction of herbal composition based on plant composition still does not obtain high accuracy. In addition, formula predictions based on plant composition cannot describe the interaction mechanism between compounds contained in plants and target proteins that represent certain diseases. The network pharmacology approach, first presented by Hopkins (2008), provides an opportunity to investigate the molecular complexity of herbal formulas and the correlation between herbal formulas and disease complexes (Wu et al., 2013; Du et al., 2014). It has been shown to work in various herbal compositions used in traditional medicine (Emig et al., 2013; Lotfi Shahreza et al., 2018). Furthermore, in this big data era, we can repurpose traditional medicines by analyzing the combinatorial properties of herbal formulas and their mechanism of action (Newman et al., 2008; Huffman and Shenvi, 2019). With the rapid advances in bioinformatics and systems biology, network-based drug discovery is seen as a promising approach to more cost-effective drug discovery (Keith et al., 2005; Jia et al., 2009; Schadt et al., 2009; Zhang et al., 2019; Chaudhari et al., 2020; Noor et al., 2022).

One of the representations of network-based drug discovery is drug–target interaction (Li et al., 2009). Many studies predicted interactions between compounds and target proteins, such as using machine learning techniques, classification algorithms, learning to rank algorithms, and deep learning algorithms (Xu et al., 2021). Yamanishi et al. laid the basis for drug–target interaction (DTI) prediction research. Their systematic study employed a bipartite local model based on an SVM Yamanishi et al. (2008). Yamanishi et al. (2010) used a distance learning algorithm as a classifier. Other studies have used a binary classification approach with machine learning techniques, such as SVM and random forest (RF), to predict drug or compound interactions with target proteins (Nasution et al., 2019; Shi et al., 2019; Erlina et al., 2020; Wijaya et al., 2021). In this binary classification approach, the features that represent DTI are obtained from the compound fingerprint and the descriptor of the protein. For example, in Erlina et al. (2020), the PubChem fingerprint was used for its compound consisting of 881 features and a dipeptide descriptor consisting of 400 features. The total number of features is 1281. Therefore, Erlina et al. (2020) reported that the binary classification model for this DTI faces high-dimensionality problems that affect the model's accuracy.

Several studies used deep learning to predict this DTI (Fitriawan et al., 2016; Lee et al., 2019; Mei and Zhang, 2019; Sulistiawan et al., 2020; Sajadi et al., 2021). Lee et al. (2019) proposed a deep learning-based prediction model capturing local residue patterns of proteins participating in DTIs. This was motivated by reports about conventional learning-based prediction models being not informative in predicting accurate DTIs. Sajadi et al. (2021) proposed a method based on deep unsupervised learning for drug–target interaction prediction called AutoDTI++ to solve the sparsity problem of the interaction matrix. Sulistiawan et al. (2020) used stacked autoencoder (SAE) as pretraining for initializing weights on the deep neural network (DNN) to prevent learning from stopping too quickly. SAE for DNN pretraining can prevent the layer outputs from vanishing during the training process (Boulila et al., 2021) and help to achieve better generalization in prediction results (Bahi and Batouche, 2018). However, there is a drawback to using binary classification for predicting DTI. It simplifies the DTI issue by modeling high-dimensional compound–protein

and their complex associations into a binary classification model without considering the relationship between compounds or proteins (Mei and Zhang, 2019).

Thus, Pliakos et al. (2019) and Fadli et al. (2021) used multilabel classification based on deep learning to generate a prediction model for DTI. Multilabel classification can be used to solve binary classification problems. In multilabel classification, the training process produces a model that maps input vectors to one or more classes. The prediction of the target is only determined based on the pattern of the existing compound structure. Utilizing proteins as class labels can reduce the input dimensions because it does not require feature extraction of the protein. In addition, from a machine learning perspective, apart from being able to predict several interactions at once, the multilabel classification model can identify possible correlations between class labels (proteins) to increase the performance of DTI predictions (Pliakos et al., 2019).

The approach to machine learning for DTI, whether using conventional machine learning methods, ensemble methods, or deep learning, raises an important issue. Its application to predict the formula of herbal compounds for certain diseases often leads to different results. Erlina et al. (2020) reported the results of different herbal compounds using multilayer perceptron, SVM, and RF. To conclude which compounds and target proteins have the most potential, Erlina et al. (2020) analyzed the overlapping predictions of herbal compounds across all methods. Likewise, Fadli et al. (2021) have built four different models based on compound features. The four models produce several different predictive compounds, so it is necessary to perform an overlapping analysis of the predicted results. Another limitation of the machine learning approach in predicting the Jamu formula is that we cannot immediately know what plants contain these candidate compounds. To get plant information, we have to do a literature study or look for it in databases, such as KNApSAcK (Afendi et al., 2012) and IJAH Analytics (http://ijah.apps.cs.ipb.ac.id).

This research proposes a new approach using the graph traversing technique to overcome the limitation of machine learning approaches. In this study, a tree representing DTI was built, and the unknown interactions were determined based on similarity measurements among compounds and proteins that meet a specific threshold value. Furthermore, the bipartite network representing DTI was expanded into a bipartite plant–protein network. Moreover, we applied a graph traversing algorithm with the branch and bound technique to perform tree searches to find medicinal plants for certain diseases. Morrison et al. (2016) stated that the branch and bound algorithm has been used successfully to find exact solutions for a wide area of optimization problems. Zhang et al. (2012) used branch and bound with a bipartite graph to solve the single vehicle routing problem with a toll-by-weight scheme. The results showed that branch and bound outperforms the best-known exact algorithms at that time for the unweighted minimum latency problem and was able to find the optimal solution. Wang et al. (2019) used the branch and bound technique to traverse the bipartite graph of resource allocation problems in radio broadcast scheduling, and the results showed that the algorithm greatly reduces the searching space and execution time. In bioinformatics, Sridhar et al. (2008) demonstrated branch and bound usage to explore the metabolic networks and find the target for known successful drugs. The algorithm can accurately identify the target enzymes that interacted with the drugs and reduce the total search time compared with the exhaustive search. Zhou et al. (2016) used a modified branch and bound algorithm to find the global minimum energy conformation in structure-based computational protein design. The algorithm is able to exploit the structure of residue–residue interaction graph to significantly accelerate the process. Thus, we proposed to use the branch and bound technique because of its ability to find the solution optimally while being able to reduce search time and space. To show the performance of the proposed method, we compared our method with the complete search algorithm, which searched all nodes in the tree without pruning. In this study, we specifically apply the proposed method to search for the Jamu formula for type II diabetes mellitus (T2DM) disease. T2DM is a disease characterized by carbohydrate, fat, and protein metabolism disorders and a lack of work and insulin secretion (Fatimah, 2015). We hope that the proposed method will become an alternative method for predicting interactions in drug–target and for searching the Jamu formula for T2DM.

Materials and methods

Data acquisition

Data acquisition was done using web crawling techniques on several databases and related research results. These data were used to build three pharmacological networks as follows.

1) Network A represents a plant–compound–protein network. The plant data are taken from the KNApSAcK database (Afendi et al., 2012). The compound data are taken from PubChem (Kim et al., 2019) and KNApSAcK database. The protein target data are taken from PubChem BioAssay (Wang et al., 2017).
2) Network B is an extension of network A by adding 10 compounds obtained from searching over compounds in the ChemMine-Tools database (Backman et al., 2011). These 10 compounds have similarity scores of at least 0.9 to each compound in network A.
3) Network C represents the relationship between the T2DM target proteins in Uniprot and the compounds in PubChem BioAssay.

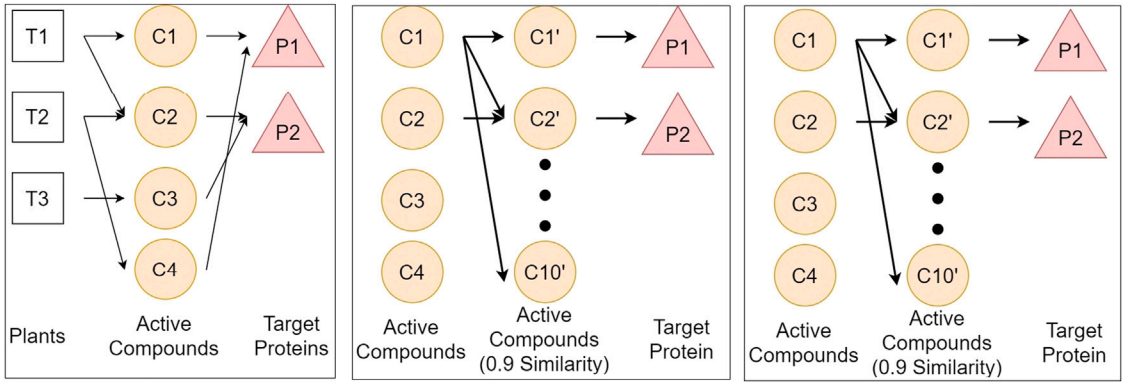


FIGURE 1
Illustration of three networks of plants, compounds, and proteins, respectively represented by T, C, and P. We define three networks from different databases to get the relationship between plants and protein. Network A connects plants from KNApSAcK, compounds from KNApSAcK and PubChem, and proteins from PubChem BioAssay. Network B connects compounds in network A, compounds in network B taken from ChemmineTools, and the target protein is the same as that in Network A. Network C connects proteins from Usman et al. (2020), proteins from Uniprot, and compounds from PubChem BioAssay.

TABLE 1 Results of data acquisition from the various databases.

Network	Data	Data resources	Results
A	Plants	KNApSAcK	- 711 plants
	Compound	KNApSAcK	- obtained 4926 compounds from 711 plants with 7725 interactions of plant–compound
			- only 581 plants have at least one compound
	Compound	PubChem	- only 2780 of 4926 compounds have CID and are categorized as compound
			- only 541 plants have at least one compound
	Target protein	PubChem BioAssay	- obtained 2308 target proteins with 131.798 interactions of compound–protein
			- only 1063 compounds have at least one target protein
B	Compound	ChemmineTools	- obtained 9647 compounds from the expansion of network A
	Target protein	PubChem BioAssay	- obtained 2465 target protein from 9647 compound
C	Target protein	Usman et al. (2020)	- 21 target proteins associated with T2DM
			- The score of betweenness centrality (BC) and closeness centrality (CC)
	Target protein	UniProt	- MGI to GI id conversion for each target protein
	Compound	Pubchem BioAssay	- obtained 803 compounds have interaction with 14 target proteins of T2DM

Figure 1 illustrates the three above-mentioned networks. The details of data and source databases for each network are provided in Table 1.

Each compound has a CID and a CAS ID. The CAS ID is used to find a CID that corresponds to the compound in the PubChem database.

Data preprocess

The data was preprocessed on the target protein of T2DM because of Usman et al. (2020). There are 21 proteins, each has the betweenness centrality (BC) and closeness centrality (CC) values. The two values are averaged and then normalized to the range of 0–1. This value becomes the weight of a protein. Table 2 shows the normalization results. Genes in Table 2 are attributes that indicate the gene name of the T2DM protein. BC and CC are the BC and CC, respectively, whereas AVG is the average value of BC and CC. The normalization results are shown in the NORM column. From the data acquisition results, only 14 T2DM proteins could be targeted by at least one compound. Therefore, we carried out analysis and experiment with those 14 T2DM target proteins.

TABLE 2 Protein weight normalization results.

Gene	BC	CC	AVG	NORM
INS	0.3211	0.6250	0.4731	1.000
AKT1	0.2435	0.5128	0.3782	0.799
TCF7L2	0.2003	0.5714	0.3859	0.816
KCNJ11	0.1342	0.5000	0.3171	0.670
UBC	0.1097	0.4878	0.2987	0.632
PPARG	0.0952	0.5128	0.3040	0.643
GCGR	0.0780	0.4762	0.2771	0.586
INSR	0.0775	0.5000	0.2888	0.610
IAPP	0.0526	0.4348	0.2437	0.515
SOCS3	0.0518	0.4348	0.2433	0.514
EP300	0.0443	0.4167	0.2305	0.487
PPARA	0.0311	0.4082	0.2197	0.464
WFS1	0.0186	0.4444	0.2315	0.489
APOE	0.0163	0.3846	0.2004	0.424
FOXO1	0.0096	0.3704	0.1900	0.402
STAT3	0.0066	0.3509	0.1787	0.378
PTH	0.0044	0.3509	0.1776	0.375
CTLA4	0.0000	0.3448	0.1724	0.364
MTNR1B	0.0000	0.3922	0.1961	0.414
PRKACA	0.0000	0.3390	0.1695	0.358
SOD3	0.0000	0.3448	0.1724	0.364

Measurement of compound similarity

Measurement of the similarity of two compounds was carried out using the Tanimoto coefficient. The Tanimoto coefficient is used to measure the degree of similarity with the formula. The more similar the two compounds are, the higher their Tanimoto coefficient. To be specific, the Tanimoto coefficient approaches 1 when two compounds have more similarities. By contrast, the Tanimoto coefficient approaches 0 when two compounds have more dissimilarities. For this reason, the compound structure is encoded into binary bits representing the compound's molecular structure. We utilized the Klekota–Roth fingerprint, which has 4860 binary features. For each compound, the fingerprint algorithm encodes 1 in a bit if there is a corresponding molecular structure, and 0, otherwise. Eq. 1 shows the Tanimoto coefficient formula.

$$coef = \frac{c}{(a + b - c)} \quad (1)$$

where a denotes the number of bits 1 in the first compound.b denotes the number of bits 1 in the second compound.c denotes the number of 1 bit in both compounds.

In this study, we used the fingerprint algorithm to calculate the similarity score between the compounds in networks A and C.

Connecting networks A, B, and C

The three main networks from this research are networks A, B, and C, as shown in Figure 1. The vertex or node of the network is a component, which can be a plant, a compound, a protein, or a disease, whereas the edge represents the connection between components. We used an adjacency list data structure to store the networks. An adjacency list is a data structure that stores graphs like a neighbor list. By using this data structure, we can speed up the tracing process because enumerating a vertex's neighbors can be done in O(k), where k is the number of neighbors of a vertex (Blandford et al., 2003).

Networks A and B connection

Networks A and B are connected because network B is an extension of network A. The expansion is through the similarity between compounds in network A (denoted as C_a) and compounds in network B (denoted as C_b).

Networks A and C connection

Network A (denoted as N_a) and C (denoted as N_c) are connected through two pathways, namely, compound similarities and protein similarities. The compound similarity was formed by calculating the similarity of each compound in network C (denoted as C_c) with the compound in network A (C_a) using the Klekota–Roth fingerprinting and Tanimoto coefficient. For each compound in network C (C_c), we record all compounds with the highest similarity score and create a new edge between C_a and C_c. The pseudocode is provided in Supplementary Figure S1.

The protein similarity pathway was formed by looking at the proteins in networks A and C that are the same. For each exact protein pair, a new edge is created between the two proteins. The pseudocode is provided in Supplementary Figure S2.

Networks B and C connection

Networks B and C are linked by protein similarity. A new edge is created between every protein in network B that is the same as the protein in network C. The pseudocode is provided in Supplementary Figure S3.

Graph traversing for constructing a weighted bipartite network plant–protein

Graph traversing from T2DM proteins to plants aims to determine which plant can target T2DM proteins. For this process to be efficient, it is necessary to trace the T2DM protein to the compound in network A. Then, we stored the interaction information between proteins and compounds in network A. Next, for each compound, we traced it back to the plant containing the compound in network A. Any components (plants, compounds, or proteins) that cannot be traced from the

T2DM protein was removed from the network to form a simpler network so that the search process became more efficient.

Next, we conducted graph traversing to construct a bipartite network of plant–compound–protein as follows:

1) Traverse from network C to network B

A complete search was started by searching from network C to network B. At first, the compound in network C was removed. Then, a search was carried out from the T2DM proteins (denoted as P_c) to each compound in network B (C_b). First, look for P_c and P_b that are the same, where P_b denotes proteins in network B. Next, store information on which target protein (P_t) traced each compound C_b. Again, a search was carried out from the T2DM proteins (P_c) to each compound C_b and recorded any target protein (P_t) connected with compound C_a whose similarity weight to compound B (C_b) is at least 0.9. The pseudocode is provided in Supplementary Figure S4.

2) Traverse from network C to network A.

First, we removed the compounds in network C. A search was carried out from the T2DM proteins P_c to the compound C_a in network A. Then, we stored any target protein that is connected to all compounds C_a. If there was a stored target protein (P_t) connected to the compound C_a, the previously recorded weight is updated to 1. Similar to the previous step, we traced the T2DM proteins (P_c) but retained all compounds in network C. For each compound C_a, record the T2DM protein (P_c) as protein target P_t that traced it (if the protein has not been recorded previously), and update the edge weights if the similarity score of C_a and C_c traversed is greater than the previous edge weight. Last, backtracking was carried out to all compounds up to plants in network A. For each plant, if the weight of the target protein is greater than the weight of the previous target protein (if it has been recorded), then update the weight and record all the target proteins that had interaction with compounds traceable from the plant. The pseudocode is provided in Supplementary Figure S5.

Graph traversing over networks A, B, and C finally produces a relationship between plants and proteins. This relationship is represented as a weighted bipartite graph between plants and proteins. From the graph search results, some compounds and plants cannot target any T2DM protein. These components were eliminated from the network, leaving 1467 compounds and 460 plants in network A. Each compound and plant pair in network A has information in the form of any T2DM protein that can be traced, along with edge weights found during tracing the protein.

Examples of the search results and the stored information:
′73399′: ({′60391226′: [′Akt1′, 0.7993787198, 0.9]}, set([′60391226′])).

Information:

1) 73399 is CID a compound.
2) 60391226 is a GI of T2DM protein that can be traced from this compound.
3) Akt1 is the symbol gene for the T2DM protein.
4) 0.7993787198 is the weight of the T2DM proteins.
5) 0.9 is the edge weight that is passed when tracing the T2DM protein.
6) set([′60391226′]) is a set data structure to prevent double counting.

If the node is a plant, the information that changes is only the CID of the compound in the Latin name of the plant in question. For example, "*Schisandra chinensis Baill.*": ({′60391226′: [′Akt1′, 0.7993787198, 0.9]}, set([′60391226′])). The difference with the previous result, namely, "*Schisandra chinensis Baill.*" is the plant's Latin name.

Composition of k plants as a candidate for herbal formula

Each plant has a relationship with one or more target proteins of TD2M. Each relationship has a different value. The greater the value of the relationship between a plant and protein indicates that the plant is associated with the target protein. In addition to the correlation value, each significant protein in T2DM has its weight. The known correlation value and protein weight will be used as a benchmark in calculating the herbal formula score using Eq. 2.

$$\boldsymbol{Formula\_score} = \sum \boldsymbol{P_i W_i} \tag{2}$$

where P_i denotes protein weight *i*th and W_i denotes edge weight of protein *i*th.

The higher the score of a formula, the better the formula will be in treating T2DM. If T plants and k unique plants are selected, there will be C(T, k) possible herbal formula candidates with k constituent plants. C(n, r) is a function that returns the value of the number of combinations of r objects from n objects. It shows that the memory and time complexity in finding the combination of k plants that make up the herbal formula is $O(T^k)$. However, memory usage optimization can be done by limiting the number of candidate herbal formulas. For example, if we only want an F for herbal formulas with the highest score, we can use a priority queue data structure to store the candidate herbal formulas. A priority queue is a data structure in the form of a (binary) heap. The (binary) heap itself is a complete binary tree. This data structure has characteristics: for each subtree with root X, the left and right child subtrees are smaller (or equal to) X. The complexity of inserting and popping data in the priority queue is O(log n), where n is the number of data stored in the priority queue, whereas the top process has a complexity of O(1).

The top is retrieving data with the maximum value in the priority queue, whereas the pop is removing data with the maximum value from the priority queue. This characteristic can be used to store the score of herbal formula candidates. Hence, the data stored in the priority queue is the herbal candidate with the highest score' the score will be stored in the form of −score. This is because if the priority queue already accommodates F candidate herbal formulas, then the pop process will issue the herbal formula with the lowest score (−highest score).

With this technique and data structure, the memory complexity becomes O(F), where F is the number of candidates with the highest score. If this process is paralleled with t threads, then it takes t priority queues, each of which accommodates F herbal candidates. The t priority queues will then be combined into a priority queue. Memory complexity becomes O(t*F).

By contrast, the time complexity is still $O(T^k)$ because it must produce all combinations of k plants as candidate herbal formulas. However, optimization can be done by reducing T. Because the herbal formula to be sought has the highest score, plants that do not have the maximum edge weight for a T2DM protein can be eliminated. Then, search for plants that are not a subset (smaller) of T2DM protein than other plants, i.e., if we choose a plant X whose T2DM protein is a subset of another plant Y with more T2DM protein, then it is more optimal if we choose plant Y.

Graph traversing using branch and bound technique

According to Morrison et al. (2016), branch and bound is a fundamental methodology and is widely used in solving exact solutions for NP-hard optimization problems. Branch and bound implicitly generate all possible solutions to the problem by storing partial solutions called subproblems in the tree structure. Unexplored nodes in the tree generate branches and partition the solution space into smaller regions that can be solved recursively (branching), and a pruning rule is used to reduce the search space size that proves to be nonoptimal (pruning). In the branch and bound algorithm, three components are not explicitly explained but can significantly influence the algorithm's performance. These components are search strategy, branching strategy, and pruning rules.

Searching strategy

In this study, the branch and bound technique will be implemented using the search strategy of breadth-first search (BrFS) (Bundy and Wallen, 1984), depth-first search (DFS) (Morrison et al., 2016), and best-first search (BFS) (Morrison et al., 2016). The difference between these three search strategies is the order in which the nodes are searched. BrFS performs a search by searching for the nearest neighbor, or in this case, the nearest neighbor is a child of that node. A search by visiting the nearest neighbor will make the search comprehensive. Next, the second strategy is DFS. A search on DFS will perform a search focused on one of the paths until it encounters a leaf node. After the leaf nodes are traced, backtracking is carried out and traced again on other paths. The last search strategy is BFS. In the BFS search strategy, the node visited first is the node with the most optimal partial solution. In the case of this herbal formula, of course, the most optimal solution is to get the most significant profit. The difference in search of the three search strategies is shown in Figure 2. The value in the circle is the profit of each node, whereas the value outside the node is the order of the node search.

Branching strategy

In addition to determining the search strategy, the branching strategy also considerably influences its use. In this study, the branching strategy used is binary branching (Devroye, 1998; Morrison et al., 2016)) and wide branching (Morrison et al., 2016). The binary branching makes each node form two children, namely, the selected plant condition and the unselected plant condition. In the wide branching strategy, nodes will form as many children as N-level nodes as many plants are added. The different forms of the two branching strategies is shown in Figure 3.

Pruning rule

The final aspect of the branch and bound algorithm is the pruning rule. In this study, the pruning rule used is the lower bounds to ensure the result is the most optimal solution. The lower bounds pruning rule starts by sorting the data from the most significant profit. Each node will calculate the maximum profit that may be obtained. The node will not be traversed if the profit is not greater than the maximum profit of the temporary solution.

The problem of finding Jamu formulas can be approached as a 1–0 KNApSAcK problem, so it can be solved using the branch and bound technique (Ezugwu et al., 2019). Suppose that there are data that have weight (w_i) and profit (p_i) stored in an array. Furthermore, the data are sorted by the highest p_i/w_i value to find the maximum profit that can be obtained with the maximum limit (W) allowed. Figure 4 shows the data that have weight and profit.

Figure 5 shows the process of finding the maximum profit from Figure 4 using the lower bounds pruning rule. The letter X in each node represents an item that was added (1), not added

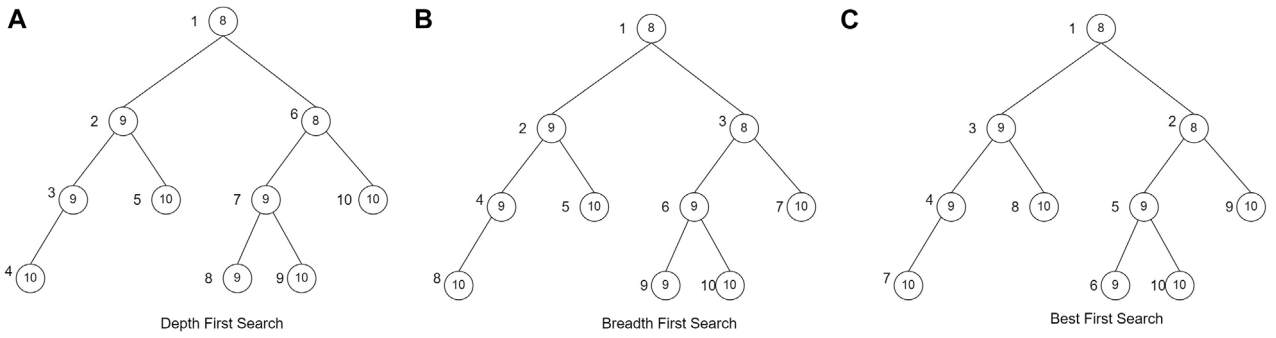


FIGURE 2
Searching strategy (Morrison et al., 2016)

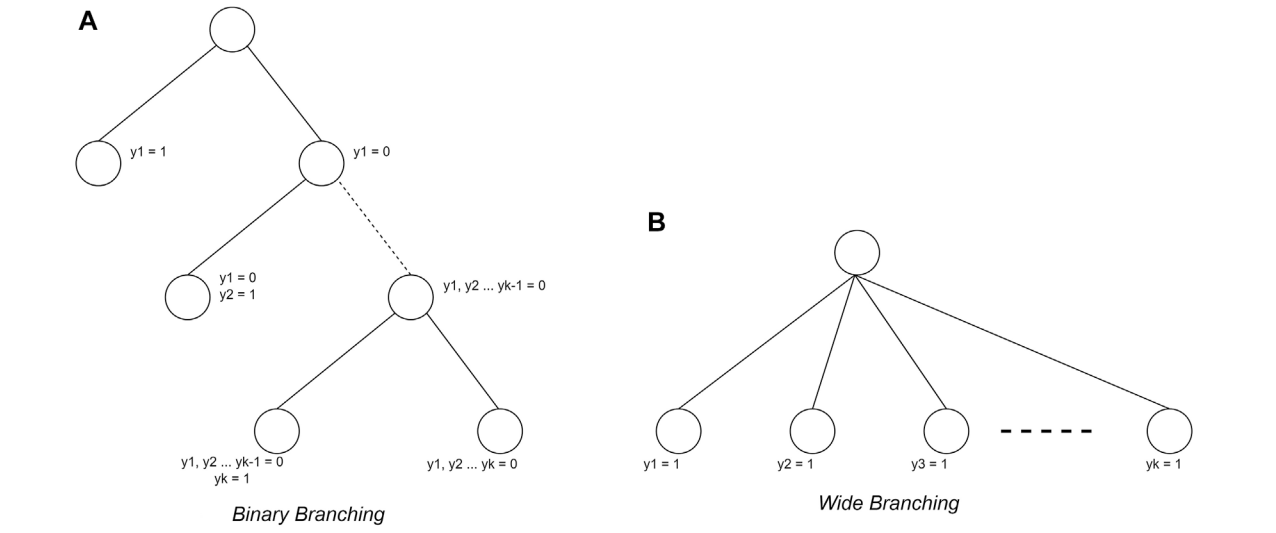


FIGURE 3
Branching strategy (Morrison et al., 2016)

(−), or not added (0) at that node. The letter B on each node is the maximum profit value obtained if the node is traced. Each node has its weight (w) based on the weight of the added items.

Implementation of branch and bound technique

The first step is to store plant data using a struct (a collection of variable definitions wrapped in a specific name). The plant data are represented as a bipartite plant–protein graph. Each significant protein in the data has a weight taken from two centrality values, namely, BC and CC. Edges that connect plants and proteins have different weights that are taken when tracing graphs.

Examples of data stored using a struct:
′*Leucaena glauca*′: ({′3041727′: [′Ppara′, 0.4643250105, 0.9]},
{′60391226′: [′Akt1′, 0.7993787198, 0.9]},
{′13432234′: [′Pparg′, 0.6426582082, 0.9]}).

The set of plant data is stored in an array of structs. The plant struct consists of the name of the plant (name), the weight of the edge of the 14th plant protein (value), plant weight (weight), and total plant profit (totalValue) calculated from Eq. 2. After being saved, the data are sorted by the totalValue parameter. Storage using arrays allows accessing plant struct data based on array index.

i	0	1	2	3
wi	11	12	8	7
pi	23	24	15	13
pi/wi	23/11	24/12	15/8	13/7

FIGURE 4
Data with weight and profit.

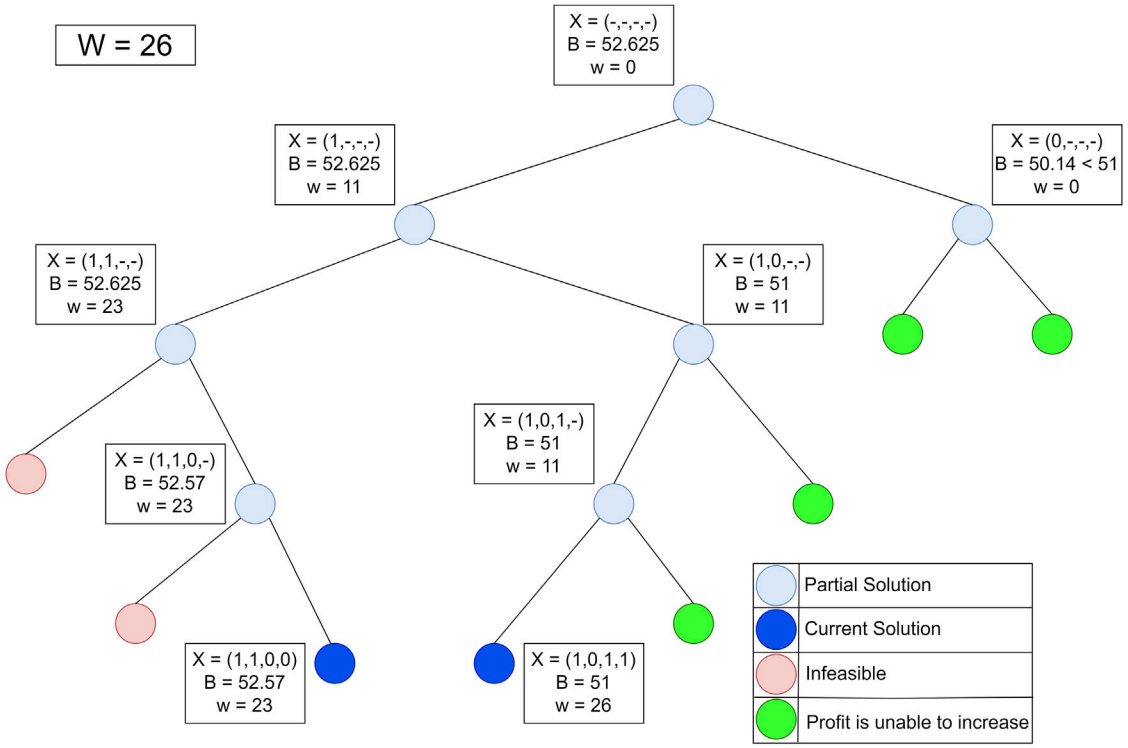


FIGURE 5
Lower bounds pruning rule.

Implementation of the branch and bound using breadth first search

The BrFS strategy was first tried using the queue data structure. The queue data structure is used because it has FIFO (first in, first out) properties, which follow the BrFS strategy. The BrFS strategy is implemented using the binary branching strategy and the lower bounds pruning rule. Figure 6 illustrates the use of the queue data structure in tree tracing. The node value in Figure 6 is the order of browsing in the tree. When the third node has been accessed and left the queue, the two children of that node will enter the queue. Each

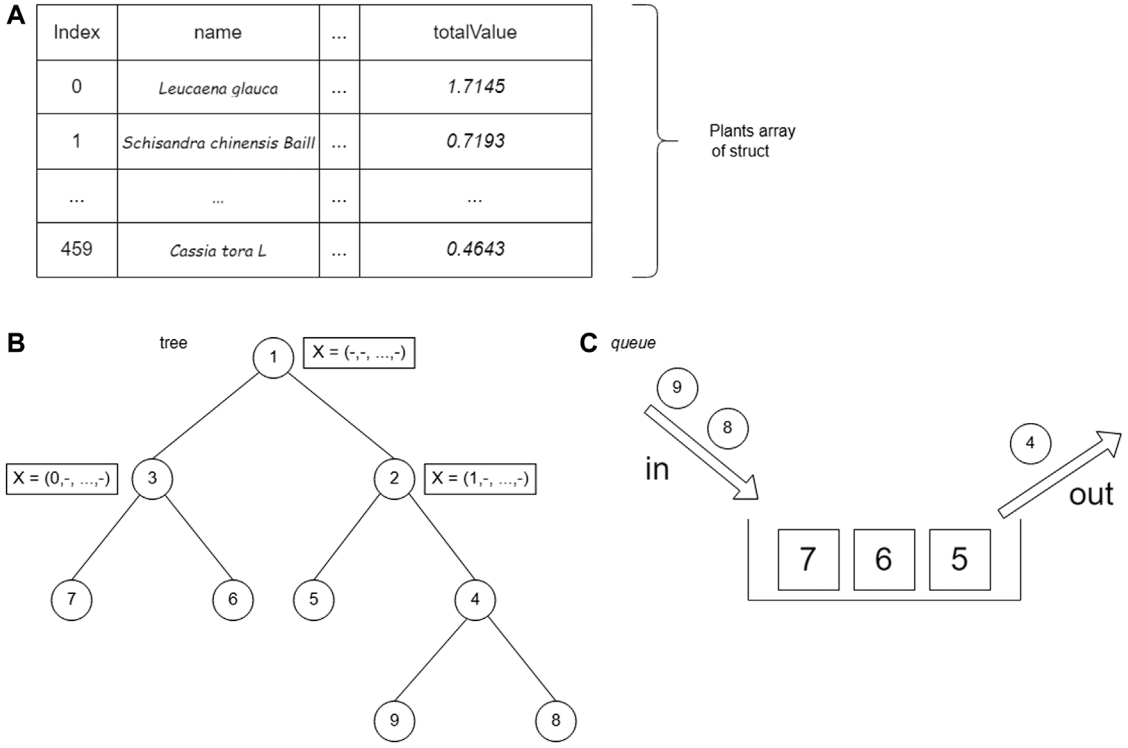


FIGURE 6
Use of the queue data structure in tree tracing.

node in the tree will store a list of plant indexes and the totalValue of the sum of each stored plant. The X symbol in Figure 6 shows the plants stored in that node.

The branch and bound algorithm starts by generating the root node. The root node is then stored in the queue data structure. After generating the root node, the next step is to enter a loop that will stop when there are no nodes in the queue. Based on the nature of the FIFO queue data structure, data from the queue is fetched (FRONT) and removed from the queue (POP).

The next step generates the child of the node. There are two children, namely, the condition of a plant being added and that not being added. At this stage, there is a bound function call, which is a function that calculates the upper bound of a node shown in Supplementary Figure S6. After getting the bound value of the child node, the value is compared with the temporary maxProfit. If the bound value of the child node is greater than maxProfit, then the child will be stored in the queue. In addition, if the profit on the node is greater than maxProfit, then the maxProfit value will be replaced with the node's profit.

Implementation of the branch and bound using depth-first search

The next step is implementing a DFS lookup strategy using a stack data structure. The stack data structure is used because it has first in, last out properties, which are in accordance with the DFS search strategy. The DFS strategy is implemented using a binary branching strategy and lower bounds pruning rules. The stack data structure is used in tree tracing. The value of the node is the order of tracing in the tree. The search is carried out by always prioritizing accessing the right child until it reaches the leaf node. After reaching the leaf node, backtracking is performed and traces the left child if the right child has been traced. Figure 7 shows the use of the stack data structure on the tree.

Implementation of the branch and bound using best-first search

The last search strategy is BFS by using the priority queue data structure. The priority queue data structure is a data structure in the form of a (binary) heap tree. (Binary) heap tree is a data structure in the form of a complete binary tree and has the characteristic that the value of each left child and right child of a node will not be greater than its parent. The structure of the (binary) heap tree can be seen in Figure 8, where the value for each node is the total value of the nodes in the tree. The complexity of deleting data and adding data to the priority queue is O (log n), for n is the amount of data that has been stored in the priority queue, whereas the process of accessing

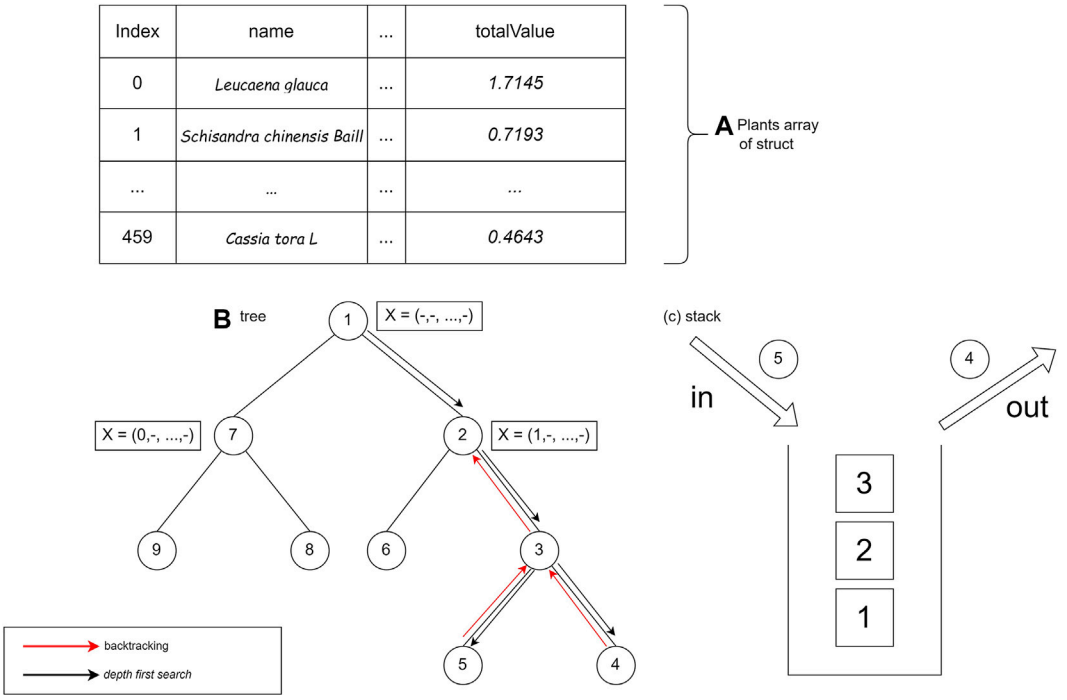


FIGURE 7
Use of the stack data structure on the tree.

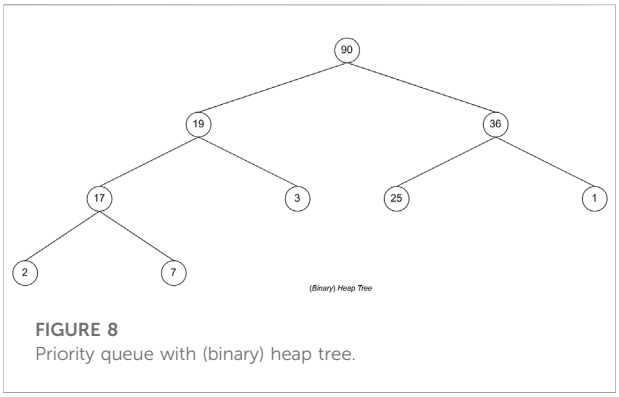


FIGURE 8
Priority queue with (binary) heap tree.

leading data has a complexity of O (1). A priority queue follows the BFS search strategy, which will execute nodes based on the most optimal solution.

Using a wide branching strategy

After implementing the binary branching strategy, a wide search strategy was implemented. In the wide branching search strategy, the BrFS strategy is used because it has the shortest computation time compared with the other two search strategies in the previous experiment. In contrast to binary branching, where each internal node must make two children for the condition of the plant being added or not, in the wide branching strategy, each node will create a different number of children depending on the plant index added last to that node. Figure 9 shows a wide branching strategy. The computation time of the wide branching strategy using BrFS is much longer than the binary branching strategy; therefore, wide branching experiments with DFS and BFS search strategies were not carried out. The X symbol in Figure 9 shows the selected crop index at each node.

Results and discussion

Comparison of computing time and search space

Differences in the use of search strategies and branching strategies will affect the computational time and search space. In order to get the best strategy, each computation time and search space of the strategy will be compared. Each strategy, search and branching, will get the same input data, namely, the plant–protein bipartite graph. The comparison of the search space will be seen from the number of candidate solutions generated in each strategy.

Three search strategies combined with two branching strategies resulted in four different branch and bound strategies: BrFS with binary branching, DFS with binary branching, BFS with binary branching, and BrFS with wide branching. The four strategies in the branch and bound

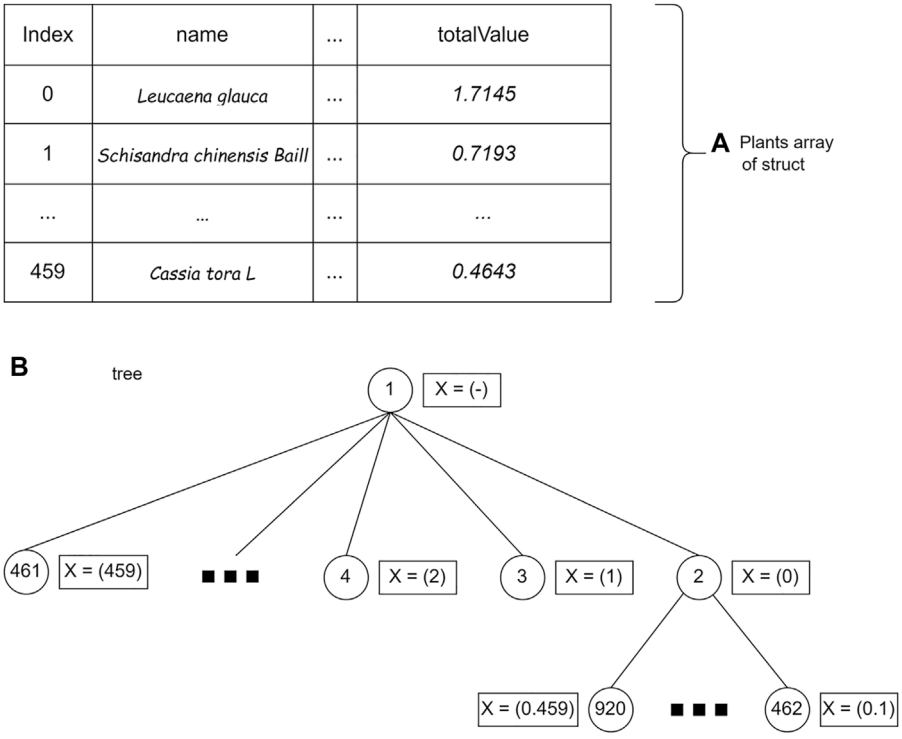


FIGURE 9
Wide branching strategy.

TABLE 3 Complete data on the computational time for each strategy.

The number of plant (k)	BrFS	DFS	BFS	Wide branching	Complete search
2	0.25	0.256	0.31	1.14	0.98
3	11.40	11.64	18.70	48.25	169.12
4	476.21	483.75	1070.58	2106.60	20285.02

algorithm have different computational times. Table 3 shows the complete data on the computational time for each strategy in this study: The experiment was conducted using PC with Intel Core i3 1.8 GHz processor, 6 GB RAM, SSD Sandisk 120 GB, and Linux Ubuntu 16.04 Operating System.

The binary search strategy produced optimum computational time, especially in the BrFS and DFS search strategies. In the combination of two plants, the longest time is the wide branching strategy, but for the combination of three and four plants, the use of complete search, as done in previous studies, requires a very long computational time. In addition to the computation time, the search space size can also be seen by calculating the number of solutions generated for each strategy. Figure 10 shows the difference in the search space size using the calculation of the number of solutions generated for each strategy in log(n) units.

In Figure 10, it can be seen that the search space of complete search is far above other strategies. The wide branching strategy has a broad search over the binary branching strategy. The binary branching strategy with either BrFS, DFS, or BFS search strategies has almost the same search area.

Pruning process

The branch and bound and complete search algorithms have the same worst-case complexity $O(T^k)$, i.e., when no nodes are

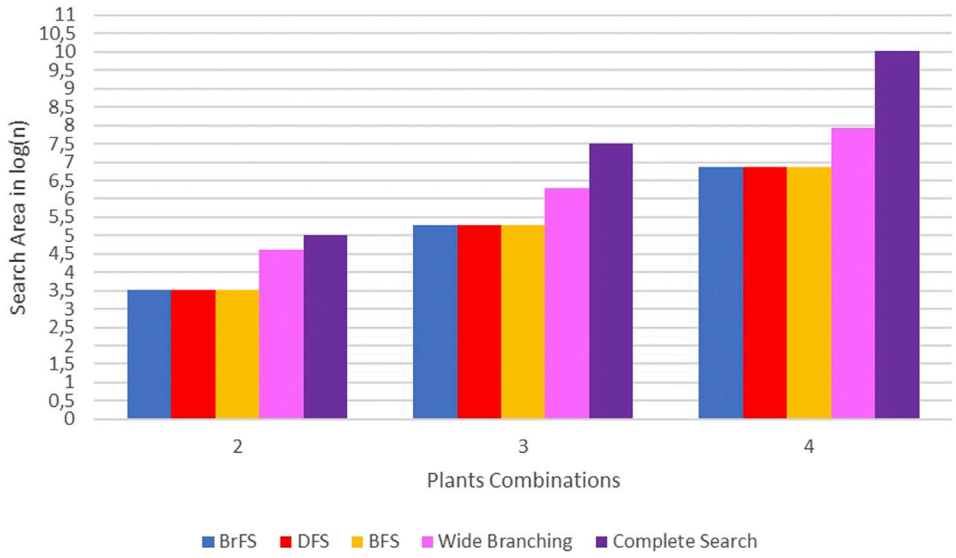


FIGURE 10
Comparison of search space area in log(n) units.

pruned. In this study, the branch and bound algorithm has a better computational time than the complete search strategy. It proves that the pruning process was successfully carried out in this study.

Comparison of branching strategies

The search space and computational time of the wide branching and binary branching strategies differ quite a lot. The wide branching strategy has a longer computation time and a larger search space than the binary branching strategy because of its inability of the pruning process. Each node in the wide branching forms a very large number of child nodes and takes time and space for each level of the tree to be formed. At node level 1 will raise to 460 child nodes and will be more and more for the next node. The number of nodes at each level results in the length of the process to generate nodes with $w_i = W$, where the greater the W requested, the higher the target level.

Search strategy comparison

In the search strategy, BFS is not better than BrFS or DFS. BFS performs a search based on the most optimum node on the node to be searched. Searching based on the most optimum node is expected to be able to cut the search space better, although it has a greater complexity when the process of deleting and adding nodes to the list. After testing, the search space of BFS with BrFS and DFS is not much different. It caused the computation time of BFS to be longer than BrFS and DFS. The best search strategy in this study is BrFS, which has a slightly better computational time difference than DFS. Searching using the BrFS strategy can reduce search space better than other strategies. The BrFS strategy of tracing nodes with the difference that the children on that node are plants are added and not added. The BrFS strategy traces nodes from the root, which nodes access high-scoring plants so that when high-scoring nodes are not added, it speeds up the process of pruning those nodes.

Composition of k plants

From the previous search results, 460 plants had at least one target protein in T2DM. Of the 460 plants, up to four combinations will be used to create a candidate for herbal formula. For every k combination of plants, 10 candidate herbal formulas with the highest score will be taken. The higher the formula score, the more traceable T2DM target protein and the more remarkable the edge weights traced that protein.

If $k = 1$, *Mangifera indica* got the highest score with 9 out of 21 (42.8%) traceable T2DM proteins, or if seen from the formula score, 4.39 out of a maximum score of 11.3 (38.8%). The summary of the 10 candidates' Jamu formulas for the combination of one plant with the highest score can be seen in Supplementary Table S1. Moreover, *Punica granatum* could only target seven T2DM proteins, but the formula score was higher than *Argemone mexicana*, *Salvia miltiorrhiza*, and *Daucus carota*. It shows that the edge or protein weight targeted by *Punica granatum* is greater than the three plants. A comparison of target proteins and edge weights between *Punica granatum* and *Argemone mexicana* can be seen in Supplementary Table S2.

When we compare edge weights, *Punica granatum* is relatively always higher than *Argemone mexicana*. It makes the Punica

TABLE 4 Execution time for composition k plants.

Combination/Composition *k*	Time to k (sec)	Time to- (k+1)/(k) (sec)
1	0.005	68
2	0.34	744.11
3	253	131.28
4	33214	—

TABLE 5 Best results of the composition of plants for the Jamu formula.

Composition of plant	Latin name	Formula score
2	*Citrus aurantium, Mangifera indica*	5.26512
3	*Angelica sinensis, Citrus aurantium, Mangifera indica*	5.77630
4	*Angelica sinensis, Citrus aurantium, Glycyrrhiza uralensis, Mangifera indica*	6.13136

granatum formula score higher than *Argemone mexicana*, although Punica granatum cannot target *GCGR* protein. If $k = 2$, the composition of *Mangifera indica* and *Citrus aurantium* obtained the highest score, 5.26 (46.5%), and 11 T2DM proteins (52.3%) could be traced. The summary of the 10 candidate herbal formulas for the combination of two plants with the highest scores can be seen in Supplementary Table S3.

Referring to the composition of one plant, *Mangifera indica* got the highest score and could target 9 T2DM proteins. From Supplementary Table S3, it can be seen that *Mangifera indica* mostly appears in every candidate's Jamu formula. However, the number of T2DM proteins is only approximately 10 or 11. It indicates that the second plant paired with *Mangifera indica* only added approximately two new target proteins. However, it is also possible that the edge weight of the second plant is higher than the edge weight of *Mangifera indica*. The candidates with the highest scores are *Mangifera indica* and *Citrus aurantium*. The comparison of the protein weights of the two can be seen in Supplementary Table S4. The contribution of *Citrus aurantium* is in the *KCNJ11* protein, in which the edge weight value of *Citrus aurantium* is greater than that of *Mangifera indica*. In addition, two T2DM proteins cannot be targeted by *Mangifera indica*, namely, *MTNR1B* and *EP300* proteins.

If $k = 3$, the composition of *Angelica sinensis*, *Citrus aurantium*, and *Mangifera indica* had the highest score, 5.7763 (51.1%), and there were 12 T2DM proteins (57.1%) that could be traced. A summary of the 10 best Jamu formula candidates can be seen in Supplementary Table S5. Supplementary Table S5 shows several Jamu formulas that have the same score. The plant compositions target the same T2DM protein and have the same edge weights. For a composition of three plants, the Jamu formula scores are approximately 50% of the maximum score, and all of them targeted 12 T2DM proteins.

If $k = 4$, the herbal formula candidates with the highest score are *Angelica sinensis*, *Citrus aurantium*, *Glycyrrhiza uralensis*, and *Mangifera indica*, with a score of 6.13 (54.2%), and there are 13 T2DM proteins (61.9%) that can be targeted. The 10 best candidates can be seen in Supplementary Table S6. The highest score for the composition of the three plants was the combination of Angelica sinensis, Citrus aurantium, and Mangifera indica, which can target 12 T2DM proteins. For $k = 4$, the composition reappeared as a candidate for herbal medicine with the highest score, plus the plant Glycyrrhiza uralensis. It shows that one new protein can be targeted by Glycyrrhiza uralensis but cannot be targeted by the other three plants. Glycyrrhiza uralensis consistently appeared in all 10 candidate herbs, meaning that of the four plant combinations, only Glycyrrhiza uralensis targeted a T2DM protein that neither did the other three. If traced back from the T2DM protein to the information stored in each plant, it was seen that the TCF7l2 protein was only targeted by Glycyrrhiza uralensis.

The limitation of this study is that it can only be used up to a composition of four plants. Doing a combination of five plants without reducing the number of plants will take much time. When the combination is one plant, it only takes 0.005 s for the program to finish. The combination of the two plants takes 0.34 s. Combinations of three and four plants take 253 s and 33,214 s, respectively. The comparison of the increase in program execution time is comprehensively shown in Table 4. This problem can be overcome by using parallel computing, which is beyond the scope of this study.

Best composition of Jamu formula

The experiment results showed that the plant combinations were obtained from two plant combinations to four plant combinations. The best of each composition of plants can be seen in Table 5. The composition of the two plants consists of

Citrus aurantium and *Mangifera indica* with a total formula score of 5.26512. In formulas for three and four plants, *Citrus aurantium* and *Mangifera indica* plants also existed. It can be concluded that *Citrus aurantium* and *Mangifera indica* plants dominate the Jamu formulas for two, three, and four plants.

The best results from the four plant compositions are *Angelica sinensis*, *Citrus aurantium*, *Glycyrrhiza uralensis*, and *Mangifera indica*. From the literature study, all the mentioned plants had the potential to be used as T2DM treatments. Li and Chen (2007) and Li et al. (2007) research analyzed the effects of Angelica sinensis polysaccharides on diabetic rats. The results showed that polysaccharides contained in *Angelica sinensis* could not only significantly reduce blood glucose levels but also improve the clinical symptoms of T2DM in the rats. Jia et al. (2015), which conducted research on the effects of Citrus aurantium in diabetic mice, reported that neohesperidin derived from Citrus aurantium helped increase oral glucose tolerance and insulin sensitivity as well as decrease insulin resistance in the diabetic mice. Moreover, aromatherapy produced from Citrus aurantium extracts also helped to relieve anxiety and fatigue in T2DM patients (Abdollahi and Mobadery 2020). Glycyrrhiza uralensis can also be used for T2DM treatment and prevention because of its flavonoids containing α-glycosidase and PTP1B inhibitory activities. Both inhibitions have been suggested as potential therapeutic targets for drug discovery for T2DM patients (Guo et al., 2015). As for Mangifera indica, Ngo et al. (2019) showed that its leaves extract contained potential hypoglycemic and antioxidant properties, which could be beneficial for T2DM patients, by inhibiting a starch digestive enzyme, possessing glucose uptake capacity and adsorption, and suppressing the production of nitric oxide, which its high level could cause diabetes complications. Further research is needed to verify and determine the potential efficacy of Jamu composition using *Angelica sinensis*, *Citrus aurantium*, *Glycyrrhiza uralensis*, and *Mangifera indica* plants.

Conclusion

Bipartite graph search optimization with branch and bound algorithms for predicting Jamu formulas can reduce computation time. The complete search strategy has the worst-case and best-case complexities of $O(T^k)$, where T is the number of plant data, and k is the number of plant combinations. The branch and bound algorithm has the worst-case complexity of $O(T^k)$ and the best-case of O(T). Although the worst case is the same, the branch and bound algorithm achieves faster computation time. In this study, we found that the best branching strategy is the binary strategy t and the best search strategy are BrFS and DFS.

The proposed method suggests that the potential plant composition for the type II diabetes mellitus Jamu formula comprises *Angelica sinensis*, *Citrus aurantium*, *Glycyrrhiza uralensis*, and *Mangifera indica*. We note that this composition requires experimental validation, which is beyond our current scope. In addition, *Citrus aurantium* and *Mangifera indica* plants dominate the three- and four-plant composition for Jamu formulas. This approach is expected to be an alternative way to discover the Jamu formula more accurately.

Data availability statement

Publicly available datasets were analyzed in this study. These data can be found here: knapsackfamily.com; https://pubchem.ncbi.nlm.nih.gov/; https://pubchem.ncbi.nlm.nih.gov/bioassay/; https://chemminetools.ucr.edu/; https://www.uniprot.org/.

Author contributions

Conceptualization, WK; methodology, WK, ZH, MA, HK, and VD; software, ZH and MA; validation, WK, RH, HK, and VD; analysis, WK, ZH, MA, and AF; data acquisition, MA, ZH, and AF; writing—original draft preparation, WK, ZH, MA, and AF; writing—review and editing, WK, AF, VD, and RH; and visualization, ZH, MA, and AF. All authors have read and agreed to the published version of the manuscript.

Acknowledgments

The author thanks Tropical Biopharmaca Research Center, IPB University, for its invaluable support 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

Supplementary material

The supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fphar.2022.978741/full#supplementary-material.

References

Abdollahi, F., and Mobadery, T. (2020). The effect of aromatherapy with bitter orange (Citrus aurantium) extract on anxiety and fatigue in type 2 diabetic patients. *Adv. Integr. Med.* 7, 3–7. doi:10.1016/J.AIMED.2019.01.002

Afendi, F. M., Darusman, L. K., Hirai, A., Altaf-Ul-Amin, M., Takahashi, H., Nakamura, K., et al. (2010). System biology approach for elucidating the relationship between Indonesian herbal plants and the efficacy of jamu. *Proc. - IEEE Int. Conf. Data Min. ICDM* 9 (1), 661–668. doi:10.1109/ICDMW.2010.105

Afendi, F. M., Darusman, L. K., Hirai Morita, A., Altaf-Ul-Amin, M., Takahashi, H., Nakamura, K., et al. (2013). Efficacy prediction of jamu formulations by PLS modeling. *Curr. Comput. Aided. Drug Des.* 9, 46–59. doi:10.2174/1573409911309010005

Afendi, F. M., Okada, T., Yamazaki, M., Hirai-Morita, A., Nakamura, Y., Nakamura, K., et al. (2012). KNApSAcK family databases: Integrated metabolite-plant species databases for multifaceted plant research. *Plant Cell Physiol.* 53, e1. doi:10.1093/PCP/PCR165

Backman, T. W. H., Cao, Y., and Girke, T. (2011). ChemMine tools: An online service for analyzing and clustering small molecules. *Nucleic Acids Res.* 39, W486–W491. doi:10.1093/NAR/GKR320

Bahi, M., and Batouche, M. (2018). Drug-target interaction prediction in drug repositioning based on deep semi-supervised learning. *IFIP Adv. Inf. Commun. Technol.* 522, 302–313. doi:10.1007/978-3-319-89743-1_27

Blandford, D. K., Blelloch, G. E., and Kash, I. A. (2003). "Compact representations of separable graphs," in Proceedings of the Fourteenth Annual ACM-SIAM Symposium on Discrete Algorithms SODA '03, Baltimore Maryland, January 12 - 14, 2003 (USA: Society for Industrial and Applied Mathematics), 679–688.

Boulila, W., Driss, M., Al-Sarem, M., Saeed, F., and Krichen, M. (2021). Weight initialization techniques for deep learning algorithms in remote sensing: Recent trends and future perspectives. arXiv 2021, arXiv:2102.07004.

Bundy, A., and Wallen, L. (1984). "Breadth-first search," in *Catalogue of artificial intelligence tools*. Symbolic computation. Editors A. Bundy and L. Wallen (Berlin, Heidelberg: Springer). doi:10.1007/978-3-642-96868-6_25

Chaudhari, R., Fong, L. W., Tan, Z., Huang, B., and Zhang, S. (2020). An up-to-date overview of computational polypharmacology in modern drug discovery. *Expert Opin. Drug Discov.* 15, 1025–1044. doi:10.1080/17460441.2020.1767063

Devroye, L. (1998). "Branching processes and their applications in the analysis of tree structures and tree algorithms," in *Probabilistic methods for algorithmic discrete mathematics*. Algorithms and combinatorics. *16*. Editors M. Habib, C. McDiarmid, J. Ramirez-Alfonsin, and B. Reed (Berlin, Heidelberg: Springer). doi:10.1007/978-3-662-12788-9_7

Du, H., Zhao, X., and Zhang, A. (2014). Identifying potential therapeutic targets of a natural product Jujuboside B for insomnia through network pharmacology. *Plant Sci. Today* 1, 69–79. doi:10.14719/PST.2014.1.2.26

Elfahmi, E., Woerdenbag, H. J., and Kayser, O. (2014). Jamu: Indonesian traditional herbal medicine towards rational phytopharmacological use. *J. Herb. Med.* 4, 51–73. doi:10.1016/J.HERMED.2014.01.002

Emig, D., Ivliev, A., Pustovalova, O., Lancashire, L., Bureeva, S., Nikolsky, Y., et al. (2013). Drug target prediction and repositioning using an integrated network-based approach. *PLoS One* 8, e60618. doi:10.1371/JOURNAL.PONE.0060618

Erlina, L., Paramita, R. I., Kusuma, W. A., Fadilah, F., Tedjo, A., Pratomo, I. P., et al. (2020). Virtual screening on Indonesian herbal compounds as COVID-19 SupportiveTherapy: Machine learning and pharmacophore modeling approaches. doi:10.21203/rs.3.rs-29119/v1

Ezugwu, A. E., Pillay, V., Hirasen, D., Sivanarain, K., and Govender, M. (2019). A comparative study of meta-heuristic optimization algorithms for 0 – 1 Knapsack problem: Some initial results. *IEEE Access* 7, 43979–44001. doi:10.1109/ACCESS.2019.2908489

Fadli, A., Kusuma, W. A., Annisa, A., Batubara, I., and Heryanto, R. (2021). Screening of potential Indonesia herbal compounds based on multi-label classification for 2019 coronavirus disease. *Big Data Cogn. Comput.* 5, 75. doi:10.3390/bdcc5040075

Fatimah, R. N. (2015). Diabetes melitus tipe 2. *J. Major.* 4 (5), 93–101.

Fitriawan, A., Kusuma, W. A., and Heryanto, R. (2013). "A classification system for Jamu efficacy based on formula using Support Vector Machine," in 2013 International Conference on Advanced Computer Science and Information Systems (ICACSIS), Sanur Bali, Indonesia, 28-29 September 2013, 291–295. doi:10.1109/ICACSIS.2013.6761591

Fitriawan, A., Wasito, I., Syafiandini, A. F., Amien, M., and Yanuar, A. (2016). "Multi-label classification using deep belief networks for virtual screening of multi-target drug," in Proceedings of the 2016 International Conference on Computer, Control, Informatics and its Applications (IC3INA): Recent Progress in Computer, Control, and Informatics for Data Science, Tangerang, Indonesia, 3–5 October 2016, 102–107.

Guo, Z., Niu, X., Xiao, T., Lu, J., Li, W., Zhao, Y., et al. (2015). Chemical profile and inhibition of α-glycosidase and protein tyrosine phosphatase 1B (PTP1B) activities by flavonoids from licorice (Glycyrrhiza uralensis Fisch). *J. Funct. Foods* 14, 324–336. doi:10.1016/J.JFF.2014.12.003

Hopkins, A. (2008). Network pharmacology: The next paradigm in drug discovery. *Nat. Chem. Biol.* 4, 682–690. doi:10.1038/nchembio.118

Huffman, B. J., and Shenvi, R. A. (2019). Natural products in the "marketplace": Interfacing synthesis and biology. *J. Am. Chem. Soc.* 141, 3332–3346. doi:10.1021/JACS.8B11297/SUPPL_FILE/JA8B11297_SI_002.XLSX

Jia, J., Zhu, F., Ma, X., Cao, Z., Cao, Z. W., Li, Y., et al. (2009). Mechanisms of drug combinations: Interaction and network perspectives. *Nat. Rev. Drug Discov.* 8 (2), 111–128. doi:10.1038/nrd2683

Jia, S., Hu, Y., Zhang, W., Zhao, X., Chen, Y., Sun, C., et al. (2015). Hypoglycemic and hypolipidemic effects of neohesperidin derived from Citrus aurantium L. in diabetic KK-Ay mice. *Food Funct.* 6, 878–886. doi:10.1039/C4FO00993B

Keith, C. T., Borisy, A. A., and Stockwell, B. R. (2005). Multicomponent therapeutics for networked systems. *Nat. Rev. Drug Discov.* 4 (1), 71–78. doi:10.1038/nrd1609

Kim, S., Chen, J., Cheng, T., Gindulyte, A., He, J., He, S., et al. (2019). PubChem in 2021: New data content and improved web interfaces. *Nucleic Acids Res.* 49 (D1), D1388–D1395. doi:10.1093/nar/gkaa971

Kotlyar, M., Fortney, K., and Jurisica, I. (2012). Network-based characterization of drug-regulated genes, drug targets, and toxicity. *Methods* 57, 499–507. doi:10.1016/J.YMETH.2012.06.003

Lee, I. I., Keum, J., and Nam, H. I. (2019). DeepConv-DTI: Prediction of drug-target interactions *via* deep learning with convolution on protein sequences. *PLOS Comput. Biol.* 15, e1007129. doi:10.1371/JOURNAL.PCBI.1007129

Li, C. J., and Chen, P. (2007). Lowering sugar effect and mechanism of Angelica polysaccharide to Wister diabetic rats induced by STZ. *J. Qiqihar Med. Coll.* 28, 1158–1161.

Li, C. J., Zhang, Y. Z., and Meng, W. F. (2007). The hypoglycemic mechanism of Angelica polysaccharide in type-2 diabetic rats. *J. Qiqihar Med. Coll.* 28, 1422–1424.

Li, J., Zhu, X., and Chen, J. Y. (2009). Building disease-specific drug-protein connectivity maps from molecular interaction networks and PubMed abstracts. *PLoS Comput. Biol.* 5 (7), e1000450. doi:10.1371/journal.pcbi.1000450

Lotfi Shahreza, M., Ghadiri, N., Mousavi, S. R., Varshosaz, J., and Green, J. R. (2018). A review of network-based approaches to drug repositioning. *Brief. Bioinform.* 19, 878–892. doi:10.1093/BIB/BBX017

Mei, S., and Zhang, K. (2019). A multi-label learning framework for drug repurposing. *Pharmaceutics* 11, E466. doi:10.3390/pharmaceutics11090466

Morrison, D. R., Jacobson, S. H., Sauppe, J. J., and Sewell, E. C. (2016). Branch-and-bound algorithms: A survey of recent advances in searching, branching, and pruning. *Discrete Optim.* 19, 79–102. doi:10.1016/J.DISOPT.2016.01.005

Nasution, A. K., Wijaya, S. H., and Kusuma, W. A. (2019). "Prediction of drug-target interaction on Jamu formulas using machine learning approaches," in 2019 International Conference on Advanced Computer Science and information Systems (ICACSIS), Bali, Indonesia, 12-13 October 2019, 169–174. doi:10.1109/ICACSIS47736.2019.8979795

Newman, D. J., Cragg, G. M., and Kingston, D. G. I. (2008). Natural products as pharmaceuticals and sources for lead structures. *Pract. Med. Chem.*, 159–186. doi:10.1016/B978-0-12-374194-3.00008-1

Ngo, D. H., Ngo, D. N., Vo, T. T. N., and Vo, T. S. (2019). Mechanism of action of Mangifera indica leaves for anti-diabetic activity. *Sci. Pharm.* 87, 13. doi:10.3390/SCIPHARM87020013

Noor, F., Tahir ul Qamar, M., Ashfaq, U. A., Albutti, A., Alwashmi, A. S. S., Aljasir, M. A., et al. (2022). Network pharmacology approach for medicinal plants: Review and assessment. *Pharmaceuticals* 15, 572. doi:10.3390/PH15050572

Pliakos, K., Vens, C., and Tsoumakas, G. (2019). Predicting drug-target interactions with multi-label classification and label partitioning. *IEEE/ACM Trans. Comput. Biol. Bioinform.* 18, 1596–1607. doi:10.1109/tcbb.2019.2951378

Purwaningsih, E. H. (2013). Jamu, Obat Tradisional Asli Indonesia Pasang Surut Pemanfaatannya di Indonesia. *eJKI.* 1 (2), 85–89. doi:10.23886/ejki.1.2065.85-89

Puspita, M. N., Kusuma, W. A., Kustiyo, A., and Heryanto, R. (2016). "A classification system for jamu efficacy based on formula using support vector machine and k-means algorithm as a feature selection," in 2015 International Conference on Advanced Computer Science and Information Systems (ICACSIS),

Depok, Indonesia, 10-11 October 2015, 215–220. doi:10.1109/ICACSIS.2015.7415176

Sajadi, S. Z., Zare Chahooki, M. A., Gharaghani, S., and Abbasi, K. (2021). AutoDTI++: Deep unsupervised learning for DTI prediction by autoencoders. *BMC Bioinforma.* 22, 204. doi:10.1186/s12859-021-04127-2

Schadt, E. E., Friend, S. H., and Shaywitz, D. A. (2009). A network view of disease and compound screening. *Nat. Rev. Drug Discov.* 8 (4), 286–295. doi:10.1038/nrd2826

Shi, H., Liu, S., Chen, J., Li, X., Ma, Q., Yu, B., et al. (2019). Predicting drug-target interactions using Lasso with random forest based on evolutionary information and chemical structure. *Genomics* 111, 1839–1852. doi:10.1016/j.ygeno.2018.12.007

Sridhar, P., Song, B., Kahveci, T., and Ranka, S. (2008). Mining metabolic networks for optimal drug targets. *Pac. Symp. Biocomput.* 13, 291–302. doi:10.1142/9789812776136_0029

Sulistiawan, F., Kusuma, W. A., Ramadhanti, N. S., and Tedjo, A. (2020). "Drug-target interaction prediction in coronavirus disease 2019 case using deep semi-supervised learning model," in 2020 International Conference on Advanced Computer Science and Information Systems (ICACSIS), Depok, Indonesia, 17-18 October 2020, 83–88. doi:10.1109/ICACSIS51025.2020.9263241

Usman, M. S., Kusuma, W. A., Afendi, M. F., and Heryanto, R. (2020). Identification of significant proteins associated with diabetes mellitus by using network analysis of protein-protein interactions. *Comput. Eng. Appl.* 8 (1), 41. doi:10.18495/comengapp.v8i1.283

Wang, S., Wu, T., Yao, Y., Bu, D., and Cai, S. (2019). Constrained maximum weighted bipartite matching: A novel approach to radio broadcast scheduling. *Sci. China Inf. Sci.* 62 (7), 72102. doi:10.1007/s11432-017-9324-0

Wang, Y., Bryant, S. H., Cheng, T., Wang, J., Gindulyte, A., Shoemaker, B. A., et al. (2017). PubChem BioAssay: 2017 update. *Nucleic Acids Res.* 45 (D1), D955–D963. doi:10.1093/nar/gkw1118

Wijaya, S. H., Afendi, F. M., Batubara, I., Huang, M., Ono, N., Kanaya, S., et al. (2021). Identification of targeted proteins by jamu formulas for different efficacies using machine learning approach. *Life* 11, 866. doi:10.3390/LIFE11080866

Wu, L., Wang, Y., Nie, J., Fan, X., and Cheng, Y. (2013). A network pharmacology approach to evaluating the efficacy of Chinese medicine using genome-wide transcriptional expression data. *Evid. Based. Complement. Altern. Med.* 2013, 915343. doi:10.1155/2013/915343

Xu, L., Ru, X., and Song, R. (2021). Application of machine learning for drug–target interaction prediction. *Front. Genet.* 12, 680117. doi:10.3389/FGENE.2021.680117

Yamanishi, Y., Araki, M., Gutteridge, A., Honda, W., and Kanehisa, M. (2008). Network pharmacology databases for traditional Chinese medicine: Review and assessment. *Bioinformatics* 24, i232–i240. doi:10.1093/BIOINFORMATICS/BTN162

Yamanishi, Y., Kotera, M., Kanehisa, M., and Goto, S. (2010). Drug-target interaction prediction from chemical, genomic and pharmacological data in an integrated framework. *Bioinformatics* 26 (12), i246–54. doi:10.1093/bioinformatics/btq176

Zhang, R., Zhu, X., Bai, H., and Ning, K. (2019). Network pharmacology databases for traditional Chinese medicine: Review and assessment. *Front. Pharmacol.* 10, 123. doi:10.3389/fphar.2019.00123

Zhang, Z., Qin, H., Zhu, W., and Lim, A. (2012). The single vehicle routing problem with toll-by-weight scheme: A branch-and-bound approach. *Eur. J. Operational Res.* 220 (2), 295–304. doi:10.1016/j.ejor.2012.01.035

Zhou, Y., Wu, Y., and Zeng, J. (2016). Computational protein design using AND/OR branch-and-bound search. *J. Comput. Biol.* 23 (6), 439–451. doi:10.1089/cmb.2015.0212

Zuhud, E. A. M., Aziz, S., Ghulamahdi, M., Andarwulan, N., and Darusman, L. K. (2001). "Dukungan teknologi pengembangang obat asli Indonesiadari segi budaya, pelestarian dan pasca panen," in *Workshop on agribusiness development based on biopharmaca* (Jakarta: Departemen Pertanian).

frontiers | Frontiers in Pharmacology

TYPE Original Research
PUBLISHED 25 August 2022
DOI 10.3389/fphar.2022.970494

OPEN ACCESS

EDITED BY
Mithun Rudrapal,
Rasiklal M. Dhariwal Institute of Pharmaceutical Education and Research, India

REVIEWED BY
Harun Patel,
R. C. Patel Institute of Pharmaceutical Education and Research, India
Abdul Issahaku,
University of KwaZulu-Natal, South Africa

*CORRESPONDENCE
Ram Samudrala,
ram@compbio.org

SPECIALTY SECTION
This article was submitted to Drugs Outcomes Research and Policies, a section of the journal Frontiers in Pharmacology

RECEIVED 16 June 2022
ACCEPTED 07 July 2022
PUBLISHED 25 August 2022

CITATION
Mangione W, Falls Z and Samudrala R (2022), Optimal COVID-19 therapeutic candidate discovery using the CANDO platform.
Front. Pharmacol. 13:970494.
doi: 10.3389/fphar.2022.970494

Optimal COVID-19 therapeutic candidate discovery using the CANDO platform

William Mangione, Zackary Falls and Ram Samudrala*

Department of Biomedical Informatics, Jacobs School of Medicine and Biomedical Sciences, University at Buffalo, Buffalo, NY, United States

The worldwide outbreak of SARS-CoV-2 in early 2020 caused numerous deaths and unprecedented measures to control its spread. We employed our Computational Analysis of Novel Drug Opportunities (CANDO) multiscale therapeutic discovery, repurposing, and design platform to identify small molecule inhibitors of the virus to treat its resulting indication, COVID-19. Initially, few experimental studies existed on SARS-CoV-2, so we optimized our drug candidate prediction pipelines using results from two independent high-throughput screens against prevalent human coronaviruses. Ranked lists of candidate drugs were generated using our open source cando.py software based on viral protein inhibition and proteomic interaction similarity. For the former viral protein inhibition pipeline, we computed interaction scores between all compounds in the corresponding candidate library and eighteen SARS-CoV proteins using an interaction scoring protocol with extensive parameter optimization which was then applied to the SARS-CoV-2 proteome for prediction. For the latter similarity based pipeline, we computed interaction scores between all compounds and human protein structures in our libraries then used a consensus scoring approach to identify candidates with highly similar proteomic interaction signatures to multiple known anti-coronavirus actives. We published our ranked candidate lists at the very beginning of the COVID-19 pandemic. Since then, 51 of our 276 predictions have demonstrated anti-SARS-CoV-2 activity in published clinical and experimental studies. These results illustrate the ability of our platform to rapidly respond to emergent pathogens and provide greater evidence that treating compounds in a multitarget context more accurately describes their behavior in biological systems.

KEYWORDS

COVID-19, SARS-CoV-2, drug discovery, multitargeting, computational drug repurposing, computational biology

1 Introduction

The severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and the disease caused by its infection, COVID-19, was first documented in Wuhan, China in December 2019. It spread rapidly and was declared a pandemic by the World Health Organization in March 2020, causing over 5.9 million deaths across the world as of February 2022

(Organization, 2022). The scientific community immediately began employing various tools and methods to identify medical interventions that would reduce the threat posed by this novel coronavirus. Numerous institutions conducted clinical trials evaluating the ability of therapeutics to decrease COVID-19 lethality, often reporting conflicting results for the same drug (e.g. chloroquine and remdesivir) (Wang Y. et al., 2020; Chowdhury et al., 2020; Spinner et al., 2020). Few clearly conclusive success stories were reported in the months immediately following the outbreak with the most notable being dexamethasone, an anti-inflammatory corticosteroid that reduced death rates in patients suffering from a hyperactive immune system response known as a cytokine storm (Group, 2021). Further, it took nearly two years for a direct antiviral therapeutic indisputably capable of significantly preventing death from COVID-19 to be approved by the FDA, specifically both molnupiravir and the nirmatrelvir/ritonavir combination drugs in December of 2021 (Mahase, 2021; Hammond et al., 2022), which speaks to the complexity of this disease and the urgent need for innovative technologies that rapidly and effectively identify promising therapies. Such technologies will not only be useful in the present but also to combat any new emerging pathogens.

Significant advances made in the field of computational drug discovery were deployed in the context of COVID-19 with the goal of uncovering viable solutions (Mohamed et al., 2021). For example, multiple studies utilized virtual docking methods to identify compounds with strong affinity to SARS-CoV-2 proteins (Vijayan et al., 2020; Wang, 2020; Baby et al., 2021). Others used network-based bioinformatics methods to suggest drug repurposing candidates or better understand SARS-CoV-2 pathology, taking advantage of large scale human and virus protein-protein interaction knowledge (Zhou et al., 2020; Ghandikota et al., 2021; Gysi et al., 2021). On the clinical side, applications of traditional and deep machine learning methods have been utilized to identify high-risk patients, such as convolutional neural networks that analyze CT and X-ray images (Ardakani et al., 2020; Ozturk et al., 2020). Deep learning approaches have also been directly applied to identify drug candidates for treating COVID-19 (Liu et al., 2021; Pham et al., 2021).

In this study we describe and evaluate the performance of our Computational Analysis of Novel Drug Opportunities (CANDO) multiscale therapeutic drug discovery, repurposing, and design platform for identifying small molecules that show potential in inhibiting the SARS-CoV-2 virus and treating COVID-19. CANDO was originally designed as a shotgun repurposing platform for exactly this type of epidemic/pandemic scenario utilizing multiscale modeling techniques and adhering to multitarget drug theory, but has since been enhanced to carry out novel drug discovery against all indications (Jenwitheesuk and Samudrala, 2003b, 2005; Jenwitheesuk et al., 2008; Horst et al., 2012; Minie et al., 2014; Sethi et al., 2015; Chopra et al., 2016; Chopra and Samudrala, 2016; Falls et al., 2019; Fine et al., 2019; Mangione and Samudrala, 2019; Schuler et al., 2019; Schuler and Samudrala, 2019; Mangione et al., 2020b; Hudson and Samudrala, 2021; Schuler et al., 2021) as well as novel drug design (Overhoff et al., 2021). The relatively recent introduction of higher order biological data such as protein pathways, protein-protein interactions, drug side effects, and protein-disease associations has further augmented our ability to describe compound behavior holistically, with subsequent improved performance (Moukheiber et al., 2021; Schuler et al., 2021; Mangione, 2022; Mangione et al., 2022). Our platform is freely available to the scientific community and a detailed description of the software implementation has been published (Mangione et al., 2020a).

We employed two separate predictive pipelines within CANDO to suggest putative drug candidates for COVID-19: one first optimized our compound-protein interaction protocol against SARS-CoV and then applied it to SARS-CoV-2, and the other searched for compounds that were similar to those known to possess anti-coronavirus activity based on interactions computed with all human proteins. We originally published three different ranked lists of putative drug candidates in March and May of 2020 using the CANDO platform (Mangione et al., 2020b; Group, 2020). In May 2020, we published an assortment of drug candidates that were highly ranked by CANDO and were at the time being investigated in clinical trials to treat COVID-19. Since then several of our top scoring compounds have been validated by us and by others which we analyze in detail here. The significant number of top-ranked therapeutics successfully validated in this study, our previous work with the Ebola Virus Disease outbreak in West Africa in 2014 (Chopra et al., 2016), as well as our earlier validation studies and analyses (Jenwitheesuk and Samudrala, 2003b,a, 2005; Jenwitheesuk et al., 2008; Costin et al., 2010; Nicholson et al., 2011; Michael et al., 2011a,b), all suggest that CANDO is an effective tool to combat newly emerging epidemics and pandemics.

2 Results and discussion

Figure 1 illustrates the pipelines and protocols used within the CANDO platform to produce the three lists of drug candidates; a detailed description follows below.

2.1 Compound-protein interaction protocol parameter optimization

We initially assessed the robustness of predictions made by the CANDO platform by inspecting the recapture rate of small molecules identified to be active against SARS-CoV, MERS-CoV, and other coronavirus species from two high-

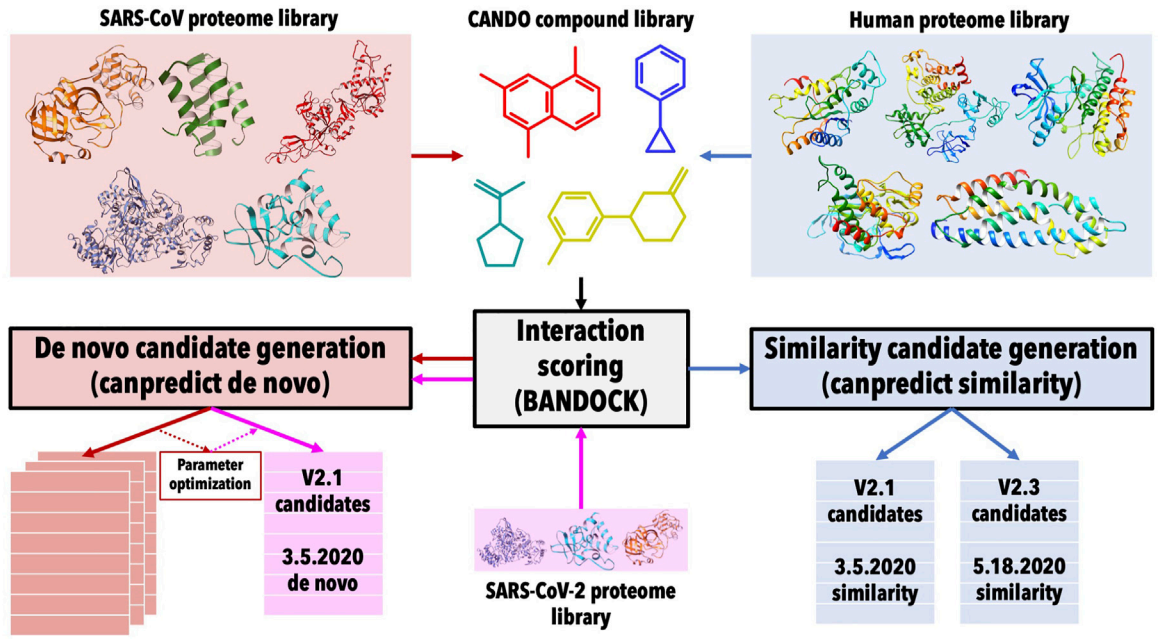


FIGURE 1
Overview of COVID-19 drug candidate prediction pipelines within the CANDO platform. Drug/compound structure libraries were curated from DrugBank (Wishart et al., 2018) and protein structure libraries comprising both the human and SARS-CoV proteomes were extracted from the Protein Data Bank (Burley et al., 2019). Interaction scores between every protein and compound in the corresponding libraries were calculated using our bioanalytic docking (BANDOCK) protocol (Mangione et al., 2020a; Schuler et al., 2021). The interaction scores with the SARS-CoV proteins were used for the *de novo* candidate generation pipeline (red) that identified compounds with the highest binding scores to multiple viral proteins, while the interaction scores with the human proteins were used for a similarity based candidate generation pipeline (blue) that identified candidates based on the similarity of their proteomic interaction signatures to drugs/compounds known to be effective against SARS-CoV *in vitro*. The interaction scoring protocol parameters were optimized against SARS-CoV and then applied to modeled protein structures from the SARS-CoV-2 proteome in the *de novo* candidate generation pipeline to produce the 3.5.20 *de novo* candidate list. Two distinct signature similarity drug candidate lists were generated using the version 2.1 CANDO compound library initially followed by an enhanced v2.3 compound library denoted as 3.5.20 similarity and 5.18.20 similarity, respectively. The predictions in these three lists were validated using evidence from published clinical and experimental studies to not only verify our platform but to determine optimal candidates that are safe and effective at treating COVID-19 downstream.

throughput screens by Shen et al. and Dyall et al. (Dyall et al., 2014; Shen et al., 2019).

We parameterized our compound-protein interaction scoring protocol *via* the discounted cumulative gain metric after generating many matrices using various criteria (see Section 3.4). Figure 2 depicts how well each parameter set ranked the actives present in the three separate screens. Among the top four competitive parameter sets, two did not have any screens ranked within the top 10 and were discarded. The parameter set we chose to apply to SARS-CoV-2 ranked 25th for SARS-CoV, 3rd for HCoV-NL63, and 10th for HCoV-OC43. We selected this over the other competitive parameter set because omacetaxine mepesuccinate, one of the strongest actives identified in the Dyall screen, was ranked 2nd versus being ranked 14th in the discarded set. The final interaction scoring protocol and corresponding de novo candidate generation pipeline parameters included the integer based Extended-connectivity fingerprint (ECFP) with a diameter of 10, dCxP scoring protocol, and a compound-protein interaction score cutoff of 0.9 (see Section 3.3 and Section 3.4).

2.2 Generation and validation of drug candidates

We generated three lists of drug candidates from corresponding pipelines that mixed and matched the protocols and data sources as described in the methods: 1) Using the parameters identified in the previous step, we generated a list of 155 approved drug candidates with strong interaction scores to SARS-CoV-2 proteins where the top scoring compounds all had interaction scores greater than or equal to 0.9 to one or both of the main (Mpro) or papain-like (PLpro) proteases (identified as 3.5.20 *de novo*). 2) The nonredundant synthesis of the 18 actives from the Shen study and 21 actives from the Dyall study as well as 2 promising manually added candidates oseltamivir and remdesivir served as input to the interaction signature similarity pipeline since it does not require EC50 values. These 38 compounds were then used to generate 45 approved drug candidates using the signature similarity pipeline (3.5.20 similarity). 3) We later

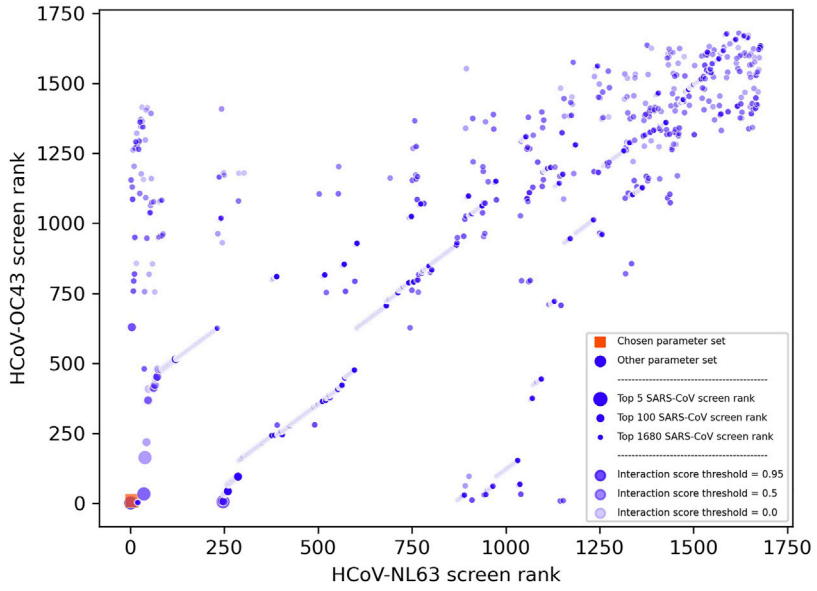


FIGURE 2
Visualization of parameter optimization set ranks across three coronavirus screens. This scatter plot depicts the ranks of each set of parameters for the interaction scoring protocol and *de novo* candidate generation pipeline within CANDO after using the discounted cumulative gain metric to score how well each corresponding pipeline ranked sets of active compounds against three separate coronavirus species: HCoV-NL63, HCoV-OC43, and SARS-CoV. The ranks for the HCoV-NL63 and HCoV-OC43 screens are depicted along the horizontal and vertical axes, respectively, while the size of the points depicts if the screen against SARS-CoV ranked within the top 5, 100, or 1,680 for each parameter set. The shade reflects the interaction score threshold that was used by the *de novo* pipeline to filter the candidates, scaled continuously from 0.0 (lightest) to 0.95 (darkest). The chosen parameter set (orange box) was the second ranked among all three screens with ranks of 3, 10, and 25 for HCoV-NL63, HCoV-OC43, and SARS-CoV, respectively, and used an ECFP10 integer based fingerprint, dCxP scoring protocol, and 0.9 compound-protein interaction score cutoff. The strong and consistent performance of this parameter set across three different coronavirus species justified our selection and warranted its use in generating drug candidates to inhibit SARS-CoV-2.

TABLE 1 Summary details of drug candidates generated by the CANDO platform. For each candidate list, the total number of candidates that were initially generated by our prediction modules, the number of viable candidates after manual filtering (removing ions and dyes) prior to validation, the number of approved compounds, the number of candidates that were matched *via* literature search using the CoronaCentral and GHDDI resources ("Checked"), the number of candidates with EHR evidence or *in vitro* activity less than 100 μM ("Validated"), the hit rate percentage, the Pearson correlation coefficient ("CC") between the full virus validation ranks and their EC50 scores (including the combined and nonredundant lists), and the discounted cumulative gain ("DCG") score are given. Overall, we obtained hit rates ranging from 13.5 to 29.9% using the CANDO platform, with the signature similarity pipelines yielding the highest success rates and the direct viral inhibition *de novo* pipeline accurately ranking the best, most potent, candidates.

	Total	Viable	Approved	Checked	Validated	Hit rate	CC	DCG
3.5.20 *de novo*	225	224	155	48	21	13.5%	0.41	0.96
3.5.20 similarity	115	114	45	17	11	24.4%	0.63	0.24
5.18.20 similarity	100	97	97	48	29	29.9%	0.35	0.22
Combined	440	435	297	113	61	20.5%	0.30	—
Nonredundant	419	414	275	102	51	18.5%	0.37	—

repeated the similarity pipeline with a sublibrary of 85 anti-SARS-CoV-2 actives and an enhanced CANDO compound library (v2.3) to generate a list of 97 approved drug candidates (5.18.20 similarity).

We scoured the literature to see if other studies validated our candidates from our three lists against SARS-CoV-2, primarily utilizing two different resources that collate detailed information on therapeutic interventions against COVID-19: CoronaCentral and the Targeting COVID-19 Portal from the Global Health Drug Discovery Institute (GHDDI) (see Section 3.6). Table 1 gives a summary of the number of predicted candidates and validations, along with correlation coefficients and discounted

TABLE 2 Complete list of validated candidates generated by the CANDO platform. The names of the 51 compounds, their ranks in the 3.5.20 *de novo*, 3.5.20 similarity, and 5.18.20 similarity lists, the full virus EC50s, main protease IC50s, and EHR-based evidence are given. Only the lowest full virus EC50 for each candidate is shown. The *de novo* pipeline identified better, more potent, full virus inhibition candidates, while the signature similarity pipeline identified a greater fraction of validated candidates accurately.

Compound	3.5.20 *de novo*	3.5.20 similarity	5.18.20 similarity	SARS-CoV-2 EC50 (μM)	Mpro IC50 (μM)	Other
Omacetaxine mepesuccinate	1	—	—	0.03	—	—
Chlorpromazine		3	11	3.14	—	—
Clomipramine		4	—	5.63	—	—
Entrectinib	—	—	4		—	58.4 μM IC50 Spike protein binding ACE2
Mycophenolate mofetil	7	—	—	0.87	—	—
Imipramine	127	8	—	10.0	—	—
Toremifene	—	—	8	2.5	—	—
Tamsulosin	100	14	38		—	18% relative risk reduction (death)
Bepridil	15	—	—	0.86	72	—
Azelastine	—	—	15	2.24	—	—
Zuclopenthixol	—	28	18	1.35	—	—
Masitinib	—	20	50	3.2	—	—
Erythromycin	—	—	20	—	—	70% reduction SARS-2 infection at 100ug/ml
Chloroquine	—	21	96	7.28	—	—
Ritonavir	—	—	21		13.7	—
Hydroxychloroquine	—	22		4.14	—	—
Cobicistat	—	—	22		6.7	—
Amodiaquine	—	23	40	0.13	—	—
Nilotinib	—	26	—	1.88	—	4.21 μM IC50 Spike protein binding ACE2
Pimozide	—	—	26		42	—
Diphenhydramine	28	—	—	17.4	—	—
Clomifene	—	29	84	9.73	—	—
Remdesivir	30	—	—	0.76	—	—
Butenafine	—	—	35	—	5.4	—
Moxifloxacin	—	44	—	239.7	—	—
Clarithromycin	—	—	47	—	—	78% reduction in severe respiratory failure versus chloroquine
Saquinavir	—	—	54	—	9.92	—
Simeprevir	—	—	55	2.3	48.2	—
Ouabain	—	—	56	0.024	—	—
Azithromycin	—	—	57	2.12	—	—
Tranylcypromine	57	—	—	—	8.64	—
Almitrine	—	—	68	1.42	—	—
Tamoxifen	—	—	74	8.98	—	—
Colistimethate	—	—	75	—	—	Mpro 17% bound (50 μM)
Lopinavir	—	—	76	9.12	—	—
Terconazole	144	—	78	11.92	—	—
Silodosin	81	—	—	—	—	18% relative risk reduction (death)
Atazanavir	—	—	82	0.22	60.7	—
Triamterene	86	—	—	—	—	23.5 μM IC50 Spike protein binding ACE2
Hydroxyzine	90	—	—	15.3	—	0.42 hazard ratio (death)
Itraconazole	—	—	90	0.39	—	—

(Continued on following page)

TABLE 2 (*Continued*) Complete list of validated candidates generated by the CANDO platform. The names of the 51 compounds, their ranks in the 3.5.20 *de novo*, 3.5.20 similarity, and 5.18.20 similarity lists, the full virus EC50s, main protease IC50s, and EHR-based evidence are given. Only the lowest full virus EC50 for each candidate is shown. The *de novo* pipeline identified better, more potent, full virus inhibition candidates, while the signature similarity pipeline identified a greater fraction of validated candidates accurately.

Compound	3.5.20 *de novo*	3.5.20 similarity	5.18.20 similarity	SARS-CoV-2 EC50 (μM)	Mpro IC50 (μM)	Other
Ebastine	—	—	92	0.5	57	—
Avatrombopag	—	—	95	5.71	—	—
Trimipramine	99	—	—	1.5	—	—
Flunarizine	105	—	—	19.05	—	—
Tadalafil	108	—	—	—	—	100 μM IC50 preventing Spike protein binding to ACE2
Thalidomide	109	—	—	—	—	11 versus 23 median days SARS-CoV-2 negative conversion from admission, 18.5 *vs.* 30 days length hospital stay
Paroxetine	111	—	—	—	—	0.52 hazard ratio (death or intubation)
Ifenprodil	117	—	—	—	46.86	Mpro 39% bound (50 μM)
Nebivolol	123	—	—	2.72	—	—
Doxazosin	133	—	—	—	—	74% relative risk reduction (death)
Levofloxacin	145	—	—	418.6	—	—
Teniposide	149	—	—	—	—	46.3 μM IC50 Spike protein binding ACE2

cumulative gain scores. Table 2 gives a full breakdown of the validations from each list as well as two drugs with weak EC50s not counted as validated: moxifloxacin and levofloxacin. This includes full virus, main protease, other miscellaneous *in vitro* (for example, inhibition of SARS-CoV-2 spike protein binding to the human ACE2 receptor), and electronic health record (EHR) studies. The studies demonstrating the activities are provided in Supplementary Table S1 while the energetic stability of the designated hits are provided in Supplementary Table S2. Figure 3 uses a Sankey diagram to illustrate the validation of all candidates with EC50s less than 10μM, which includes 31 drugs that were found to be effective against SARS-CoV-2 in full virus inhibition studies. Overall, a total of 51 drugs showed efficacy against SARS-CoV-2 out of 275 nonredundant candidates for a hit rate of 18.5%.

In addition to these validations gathered from the literature, 30 candidates were evaluated by our collaborator, Ennaid Therapeutics, of which 11 displayed *in vitro* efficacy; a patent has been filed for their use (Samudrala et al., 2020).

Aside from moxifloxacin and diphenhydramine, all validations of candidates ranked in the top 50 of their respective lists have full virus EC50 values less than 10 μM. The same is true for those in the top 100 with the exception of hydroxyzine and terconazole. The second strongest reported EC50 (0.03 μM) was obtained using omacetaxine mepesuccinate, the top ranked candidate from the 3.5.20 *de novo* list, which is only slightly weaker than the best EC50 belonging to ouabain (0.024 μM), ranked 56 in the 5.18.20 similarity list. Figure 4 illustrates the proposed mechanism of omacetaxine mepesuccinate inhibiting SARS-CoV-2 *via* strong predicted interactions to the main and papain-like proteases. Two other drugs known to inhibit both SARS-CoV-2 proper as well as its main protease, bepridil and ebastine, were present in the 3.5.20 *de novo* and 5.18.20 similarity lists respectively, with the latter having a relatively weak interaction score to the main protease of 0.82 while the former received a score of 0.98. However, the protease inhibition activity of ebastine is supported by it being the third most similar compound to nelfinavir, a known human immunodeficiency virus protease inhibitor, based on their proteomic interaction signature similarity, suggesting the CANDO platform is capable of recognizing/predicting mechanistic behavior in multiple ways.

We also investigated why moxifloxacin was deemed a candidate despite its low reported efficacy (Figure 5). Moxifloxacin was predicted by the 3.5.20 similarity pipeline and received a score of two meaning it was in the top 25 most similar compounds to two coronavirus actives (average rank 19.5). Moxifloxacin was the 18th most similar compound to mefloquine and the 21st most similar to emetine; the former is a treatment for malaria, similar to many other anti-malarials with moderate activity (~4–15 μM) against coronaviruses *in vitro* (Dyall et al., 2014; Ellinger et al., 2021), and the latter is an experimental treatment for amoebiasis with demonstrated activity against not only SARS-CoV-2 (EC50 0.46 μM) (Choy et al., 2020), but many other coronavirus species (Dyall et al., 2014; Shen et al., 2019). Moxifloxacin having similarity to one strong and one moderate anti-coronavirus compound would suggest a stronger EC50 than 239.7 μM; we attribute this result to a progressive decrease in behavioral/functional similarity signal strength/relevance as the distance between their proteomic

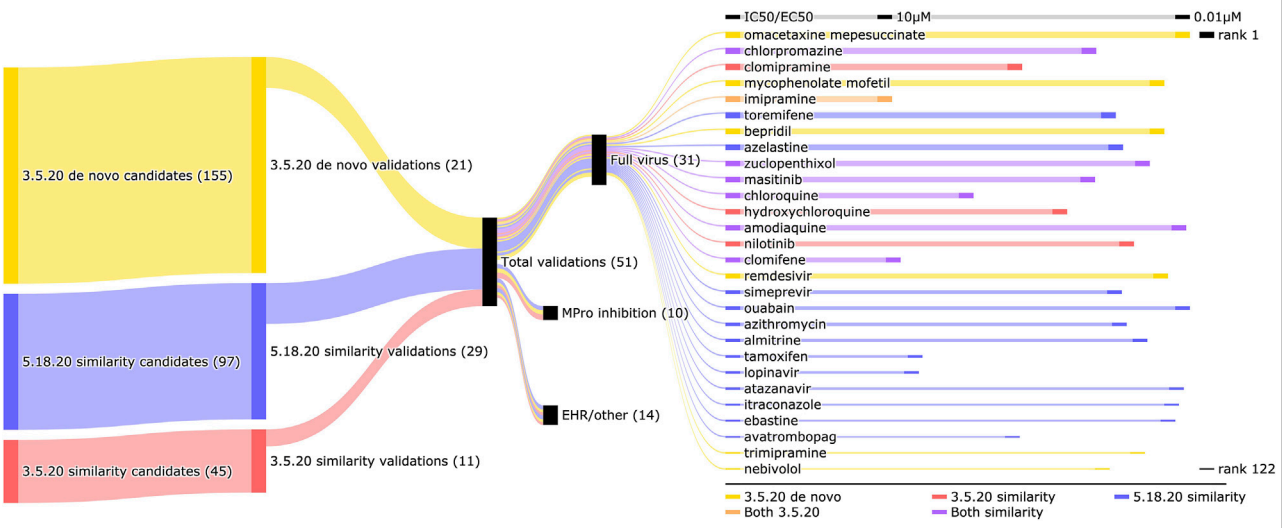


FIGURE 3
Validation of SARS-CoV-2 full virus inhibition candidates generated by the CANDO platform. The flow of validations among the three candidate lists generated using the CANDO platform are depicted from left to right using a Sankey diagram. The 155 candidates from 3.5.20 *de novo* (yellow) yielded 21 validations, while the similarity counterpart from the same date (red) produced 11 validations from 45 candidates. The 5.18.20 similarity list (blue) of 97 approved drugs resulted in 29 validations, resembling the hit rate of 3.5.20 similarity list and more than twice that of the lone *de novo* list. The 51 total validations were comprised of 31 full virus studies, 10 main protease (Mpro) inhibition studies, and 14 EHR or other inhibition based studies. The compounds that were validated *via* a full virus inhibition of less than 10 μM are shown prioritized by their rank in the list (or best rank if in multiple lists) corresponding to the thickness of their bars (ranging from rank 1–122). All but six drugs were in a single list, five drugs were in both similarity based lists (purple) and one was in both the 3.5.20 similarity and *de novo* lists (orange). The length of the horizontal bar next to the names of the compounds indicates the lowest reported EC50 or IC50 from published experimental studies progressing on a linear scale. The second strongest reported EC50 (0.03 μM) belongs to omacetaxine mepesuccinate, which is the top candidate from 3.5.20 *de novo*. The correlation between rank and strength of inhibition is sub-moderate (0.3718), and this is possibly due to the variation in assay design among different studies (viral replication reduction, viral entry inhibition, viral induced cytopathic effect reduction, etc). Overall, the CANDO platform was able to identify several candidates with potent anti-SARS-CoV-2 activity using two different predictive pipelines, verifying its potential to rapidly and efficiently respond to emerging threats to global health.

interaction signatures relative to those of known coronavirus actives increases. In other words, the signal disappears as we move further down the ranks as depicted in Figure 5.

The second to last validation in the 3.5.20 similarity list is clomifene, an infertility treatment in women, at rank 29 with a score of 2 and EC50 of 9.73 μM; it is similar to the coronavirus active compounds tamoxifen (rank 2) and toremifene (rank 11), constituting an average rank of 6.5. Additionally, all other validations from the same list have an average rank of less than or equal to 6.5 regardless of the score, which ranges from two to six. This implies setting the cutoff rank for the canpredict module to a lower value will produce stronger candidates and is further supported by the higher hit rate observed in the 5.18.20 similarity list (29.9 *vs.* 24.4% for 3.5.20 similarity) which was produced with a cutoff of ten. However the candidates predicted in the 5.18.20 similarity list benefited from using anti-SARS-CoV-2 drugs specifically, as opposed to actives against other coronavirus species, and had over double the number of active compounds when compared to the actives used to generate the 3.5.20 similarity list.

The candidates generated using the human proteome interaction signature similarity pipeline had higher validation rates relative to the direct compound-protein inhibition *de novo* pipeline; yet some of the candidates generated by the latter demonstrated stronger *in vitro* efficacy. The increase in hit rate is due to the similarity pipeline utilizing the structural knowledge embedded in the results of countless coronavirus studies, whereas the *de novo* pipeline relies exclusively on the fidelity of the compound-protein interactions computed using our interaction scoring protocols, which are prone to inaccuracies. The *de novo* pipeline was better tuned to correctly rank the strong inhibitors as interaction scoring parameters were first optimized for SARS-CoV using the discounted cumulative gain metric, which prioritizes ranking the strongest active compounds near the top of the list. This suggests that weighting the active compounds based on their available EC50 values for the full proteome interaction similarity pipeline may produce more potent candidates.

Our observed hit rate of 18.5% is likely conservative as not all of the compounds from the three candidate lists have been validated for efficacy against SARS-CoV-2 in published clinical and experimental studies. Conversely, the fraction of these 51 validations analyzed in this study that will result in

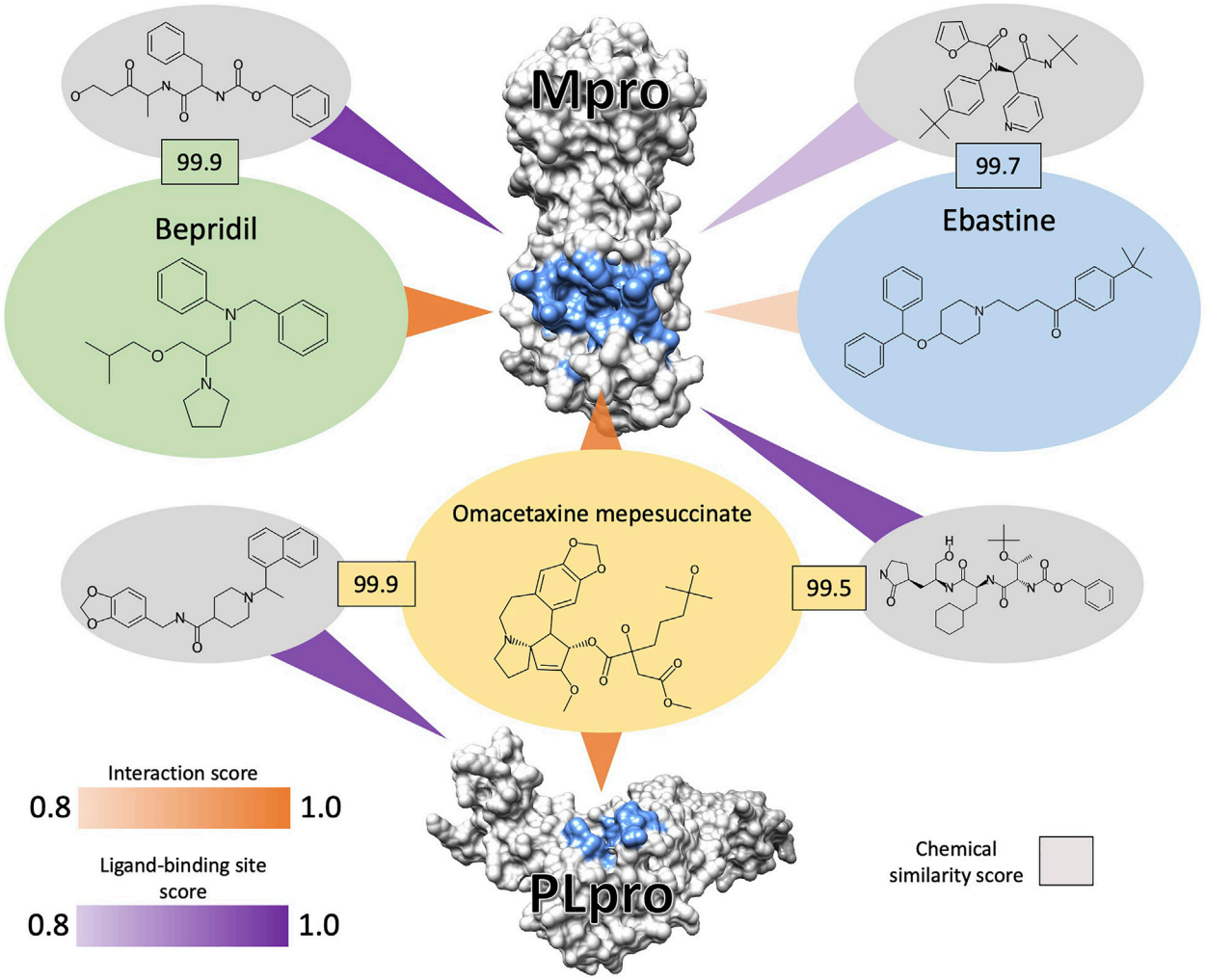


FIGURE 4
Analysis of selected interactions between SARS-CoV-2 proteases and top ranked CANDO-generated drug candidates. The main (Mpro, top) and the papain-like (PLpro, bottom) proteases are depicted in grey with the binding site residues colored blue. Bepridil (green) and omacetaxine mepesuccinate (orange) ranked at 15 and 1 in the 3.5.20 *de novo* list, and ebastine (blue) ranked at 92 in the 5.18.20 similarity list, are shown bound to one or both proteases. These constitute example interactions of when CANDO made a successful prediction as well as illustrate why candidate generation is not perfect from a mechanistic multiscale perspective. The interaction score (orange triangles) between the compounds and the proteases were generated using the bioanalytic docking protocol BANDOCK, with higher scores (maximum 1.0) predicting a higher likelihood of interaction. The ligand associated with the binding site predictions by the COACH algorithm and chosen as the template for BANDOCK are depicted in grey ellipses (full names available in the Supplementary Material), all of which are strong coronavirus protease inhibitors. These ligands are compared to the query drug using the ECFP10 chemical fingerprint *via* RDKit and a similarity score is assessed based on the Sorenson-Dice coefficient. The percentile of the similarity (black outlined boxes) from the corresponding distribution of all similarities between the query compounds and all ligands in the binding site library is multiplied by the confidence score associated with the binding site prediction from COACH (purple triangles) to serve as the final score. Bepridil inhibits the full SARS-CoV-2 virus and Mpro *in vitro* with EC50s of 0.86 and 72 μM, which was successfully assigned a strong interaction score of 0.98. On the other hand, ebastine also inhibits the full virus and Mpro with EC50s of 0.5 and 57 μM, yet was assigned a lower interaction score of 0.82. Despite the strong percentile similarity score between ebastine and its template ligand (99.7), the confidence score for this binding site prediction was 0.82, significantly lowering the final interaction score. However, ebastine is the 3rd most similar compound to nelfinavir, a known human immunodeficiency virus protease inhibitor with activity against SARS-CoV-2, based on interaction similarity to a library of 5,317 human proteins, suggesting its putative mechanism as a protease inhibitor. Omacetaxine mepesuccinate, the second strongest full virus inhibitor predicted by CANDO with an EC50 of 0.03 μM, was the top candidate from the *de novo* list and has interaction scores of 0.960 and 0.964 with Mpro and PLpro, respectively, and has not yet been validated in terms of target specificity. Based on the high interaction scores, we propose this as its mechanism not only for SARS-CoV-2, but for all other coronavirus species against which it has activity. In this manner, the mechanistic understanding of drug candidate behavior is readily deciphered in a multiscale manner, from the atomic-level fingerprints between the novel drug candidates and the interacting ligands to the evolutionary information embedded at the protein and proteome scales, and exemplifies the ability of the CANDO platform to accurately identify novel drug candidates and their mechanisms *via* a multi-pronged approach.

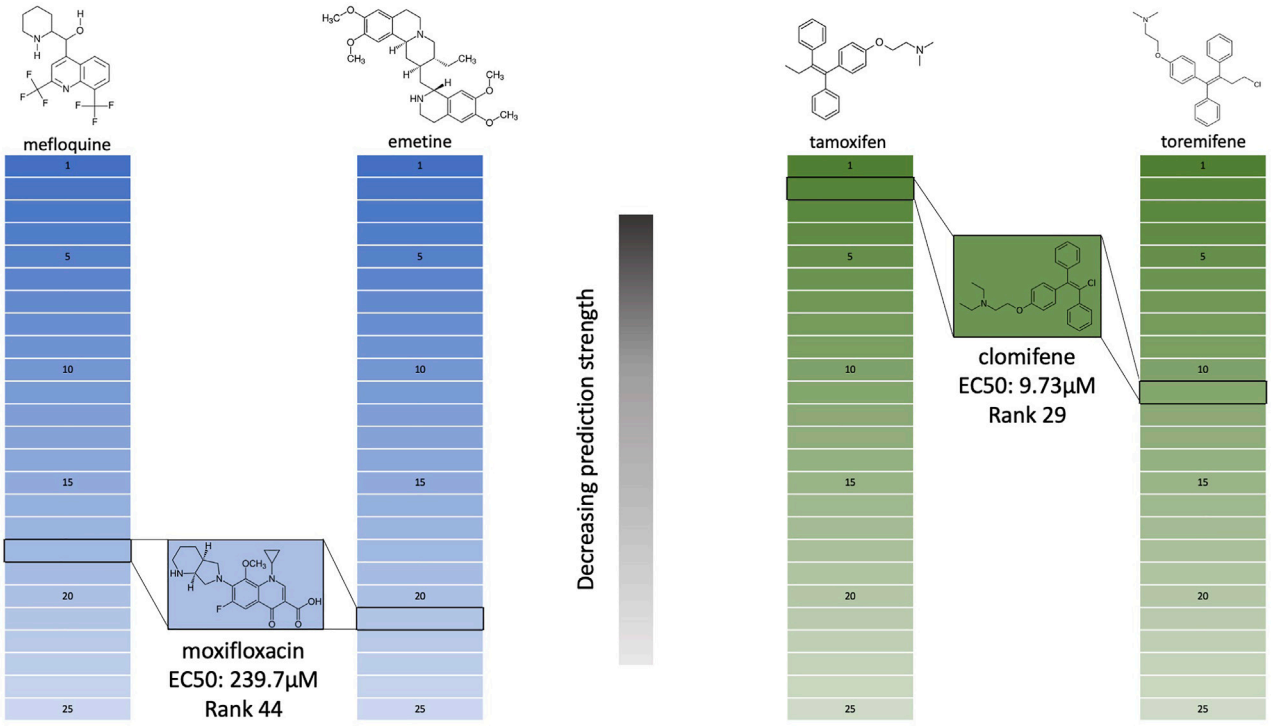


FIGURE 5
Analysis of the efficacy of two SARS-CoV-2 inhibitors with respect to proteomic interaction signature similarities predicted using the CANDO platform. The structures of two validated compounds from the 3.5.20 similarity list, moxifloxacin (blue) and clomifene (green), are shown with EC50 values of 239.7 and 9.73 μM, respectively. The EC50 values are based on the full virus *in vitro* inhibition of SARS-CoV-2. Their ranks in the list of the top 25 most similar compounds to two different coronavirus actives are outlined in black; moxifloxacin is at rank 18 and 21 in comparison to mefloquine and emetine, and clomifene is at rank 2 and 11 in comparison to tamoxifen and toremifene, respectively. These ranks are determined by the similarity coefficient (Sorenson-Dice) of the proteomic interaction signatures between the query compound and all others in the CANDO library. The proteomic signatures are vectors of interaction scores between a compound and a library of 5,317 human proteins computed using our in-house docking protocol BANDOCK (see Section 3.3). The fundamental hypothesis underlying the CANDO platform is that similar drugs will have similar behavior in biological systems as measured by their proteomic interaction signatures. Despite the relatively high rank (44) of moxifloxacin in the 3.5.20 similarity list, its measured EC50 was poor; this is explained by its lower interaction signature similarity to the two coronavirus actives depicted suggesting behavioral signal strength inversely correlates with rank. On the other hand, clomifene, the next highest prediction from the 3.5.20 similarity list at rank 29, has a stronger EC50 and ranks higher in the similarity lists to two coronavirus active compounds. However the reported EC50 values of mefloquine and emetine are strong at 4–15 and 0.46 μM, respectively, which implies that behavioral similarity signal is preserved for highly ranked compounds and that using lower rank cutoff thresholds produces stronger candidates.

clinical utility is limited due to a variety of factors such as pharmacokinetics, pharmacodynamics, safety, and cost. Multiple candidates that we listed as validations, specifically chloroquine, hydroxychloroquine, and azithromycin, have had conflicting reports of clinical benefit (Wang Y. et al., 2020; Chowdhury et al., 2020; Spinner et al., 2020; Echeverría-Esnal et al., 2021); regardless, we consider them a successful prediction of the CANDO platform due to the extensive number of *in vitro* studies reporting their SARS-CoV-2 inhibition, which is what the compound-proteome interaction analytics pipelines present in CANDO optimize for at present. Furthermore, even if CANDO fails to accurately score a known interaction with our bioanalytic docking protocol (BANDOCK) for a compound with reported activity, as in the case of ebastine and the SARS-CoV-2 main protease, its therapeutic mechanism may still be elucidated by inspecting the behavior of highly similar compounds based on their proteomic interaction signatures. Consequently, we are actively implementing methods to further refine the feasibility of our candidates based on the aforementioned factors.

3 Methods

3.1 Compound structure library and known actives curation

The CANDO v2.1 compound library consisted of 8,696 drug and drug-like small molecule three-dimensional structures, including 1,979 approved for human use, and was extracted from DrugBank (Wishart et al., 2018); this library was used for the initial predictions. We later updated the CANDO compound library to v2.3 that included 13,194 compounds from DrugBank

consisting of 2,449 approved drugs and 2,519 small molecule metabolites, with the remaining classified as experimental/investigational. Biologic therapeutics were not included in our analyses.

Initially, compounds were considered as a coronavirus active if they were identified in one of two high-throughput screens by Shen et al. and Dyall et al. (Dyall et al., 2014; Shen et al., 2019). The former screened a library of 290 compounds against SARS-CoV and Middle East respiratory syndrome coronavirus (MERS-CoV). The latter screened a 2,000 compound library against four different coronavirus strains: human coronavirus OC43 (HCoV-OC43), human coronavirus NL63 (HCoV-NL63), MERS-CoV, and murine coronavirus (MHV-A59; also known as mouse hepatitis virus). Out of 60 successful hits from both studies, 18 compounds from the Shen study along with their EC50s against HCoV-OC43 and HCoV-NL63, as well as 12 compounds from the Dyall study and their EC50s against SARS-CoV were mapped to our compound library. These three actives sublibraries were used for the compound-protein interaction scoring protocol parameter optimization (see Section 3.4).

The nonredundant combination of actives in the Shen and Dyall studies were used for the signature similarity candidate generation pipeline (see Section 3.5). We also added oseltamivir and remdesivir as at that time (February 2020) evidence suggested that they may inhibit SARS-CoV-2 or related coronaviruses (Wang M. et al., 2020; Coenen et al., 2020), resulting in an actives library of 38 compounds.

As more data became available regarding *in vitro* efficacy values for compounds against SARS-CoV-2, a second sublibrary of 85 actives with reported EC50 values less than or equal to 10 μM was extracted on May 7, 2020 from the Targeting COVID-19 Portal from GHDDI (Leng, 2020), which contained 17/38 compounds from the previous list. The updated CANDO compound library along with the new GHDDI actives sublibrary were used for the enhanced signature similarity candidate generation pipeline (see Section 3.5).

3.2 Protein structure library curation

The available SARS-CoV x-ray diffraction protein structures were obtained from the Protein Data Bank (PDB) (Burley et al., 2019) and initially served as our representative coronavirus proteome, comprising eighteen total structures. These eighteen SARS-CoV proteins were used for the compound-protein interaction protocol optimization (see Section 3.3).

A SARS-CoV-2 protein library of 24 structures was modeled from sequence using the I-TASSER v5.1 suite (Yang et al., 2015) and comprised the proteome used for the remaining analyses. We prioritized 18/24 proteins that were modeled by I-TASSER using homology to known coronavirus structures. These 18 SARS-CoV-2 proteins were used for the *de novo* pipeline, while both iterations of the signature similarity based pipeline (see Section 3.5) used a library of 5,317 human protein x-ray diffraction structures extracted from the PDB. The former piepline is implemented using the canpredict *de novo* module, and the latter is implemented using the canpredict similarity module, in the cando.py Python package (Mangione et al., 2020a; Mangione and Falls, 2022)).

3.3 Compound-protein interaction calculation

We utilized our in-house bioinformatic analytics-based docking protocol BANDOCK to generate interaction scores between every compound and every protein structure; these scores serve as a proxy for binding strength/probability (Minie et al., 2014; Sethi et al., 2015; Falls et al., 2019; Hudson and Samudrala, 2021). The COACH algorithm from the I-TASSER suite (Yang et al., 2013) was used to predict binding sites for each protein. COACH outputs an associated score and binding ligand for every binding site in a protein and is the primary data used by BANDOCK to generate interaction scores. For a given compound and protein pair, every interacting ligand predicted by COACH is compared to the query compound by computing the similarity coefficient of their chemical fingerprints generated *via* RDKit (Landrum, 2013). The maximum resulting coefficient (i.e. the strongest match) and its associated binding site score are then used to compute the final interaction score for the compound-protein pair, depending on the scoring protocol parameters. This is repeated iteratively for each protein in a given library (e.g. SARS-CoV, SARS-CoV-2, human, nonredundant PDB), resulting in a proteomic interaction signature for every drug/compound, represented an $N \times M$ matrix, where N is the number of drugs/compounds and M is the number of proteins.

Interaction scoring (BANDOCK) parameters were systematically varied to identify those optimal for assessing anti-coronavirus activity. These were 1) the chemical fingerprinting method: ECFP or functional-class fingerprint (FCFP) with diameters of 0, 2, 4, 6, 8, and 10 and length of 2048; 2) the fingerprint style: binary vs integer based for the compounds/ligands; 3) the scoring protocol: the binding site score from COACH (Pscore), the Tanimoto or Sorenson-Dice coefficient of the binding site ligand from COACH to the query drug (Cscore) for binary or integer fingerprints, respectively, the percentile of the Cscore in the distribution of all I-TASSER ligand Cscores to the query drug (dCscore), or products of these (Pscore × Cscore, Pscore × dCscore); and 4) thresholds: Pscore and Cscore (or dCscore) thresholds so that any binding site or compound-ligand similarity coefficient (or its percentile) that does not exceed each cutoff, respectively, are ignored. A compound-protein interaction matrix was generated for each of these parameter combinations.

Computed interaction scores with the 18 SARS-CoV proteins were used for compound-protein scoring protocol parameter optimization, while the scores computed (using the parameters identified in the previous step) with the 18 SARS-CoV-2 proteins were used for the *de novo* candidate generation pipeline. The scores computed with a library of 5,317 human PDB structures were used for the similarity-based pipelines (see section 3.5). The initial parameters were an ECFP4 binary fingerprint with Tanimoto coefficients for Cscores, Pscore scoring protocol, and a dCscore threshold of 0.5 (50th percentile), which were used to generate the March 5 2020 aka 3.5.20 list of candidates. The enhanced parameters were an ECFP4 integer fingerprint with Sorenson-Dice coefficient for Cscores, Pscore × dCscore scoring protocol, and a dCscore threshold of 0.75 (75th percentile), which were used to generate the May 18, 2020 aka 3.18.20 candidate list.

3.4 Parameter optimization using coronavirus active compound recovery

We identified the best parameters for BANDOCK that optimally ranked the compounds identified *via* high throughput screens against three different coronavirus species (SARS-CoV, HCoV-NL63, and HCoV-OC43), each of which were assessed separately *via de novo* drug candidate generation. We also varied the cutoff threshold of interaction scores to consider so that the interaction scores with proteins below that threshold were not considered in the total for a given compound. The cutoffs in this study were incremented by 0.05, starting with 0.0 (no threshold) and ending with 1.0 (maximum score). The discounted cumulative gain metric (Järvelin and Kekäläinen, 2002; Dupret, 2011), often employed for search engine optimization and other early recognition problems, was used to assess how well each matrix properly ranked the active compounds in the proper order given their associated EC50/IC50 values from each of the three species separately. Our previous work has identified this metric as the optimal one for drug repurposing studies (Schuler et al., 2021). Briefly, discounted cumulative gain (DCG) rewards lists of predictions that rank the optimal known actives at the top and progressively penalizes lower ranked ones *via* the equation:

$$DCG_p = \sum_{i=1}^{p} \frac{2^{rel_i} - 1}{log_2(i+1)} \quad (1)$$

where p is the length of the list, i is the rank, and rel_i is the relevance score of the item at position/rank i which is the inverse of the EC50 values (1/EC50) for the 36 nonredundant actives.

Parameter sets utilizing any of the following criteria were discarded due to trivial candidate rankings: Pscore scoring protocol, interaction score threshold of 1.0, and Cscore threshold of 1.0. Interaction scores generated using the Pscore protocol did not utilize the chemical fingerprint similarity value between the binding site ligand and the query compound and subsequently failed to discriminate between two compounds that used the same ligand. Using an interaction score or Cscore threshold of 1.0 required the chemical fingerprint similarity score to equal 1.0, meaning identical compounds, therefore ensuring the only predicted candidates were known coronavirus inhibitors.

3.5 COVID-19 drug candidate generation

To generate drug candidates against COVID-19, we used both a *de novo* pipeline that ranked compounds based on their predicted interaction scores against proteins from SARS-CoV-2, and a similarity pipeline that searched the CANDO drug/compound library for compounds similar to those deemed as actives in terms of their interaction signatures. The former protocol summed the computed interaction scores of each compound against all viral proteins and ranked them from best to worst. Interaction scores below particular thresholds were ignored in the sums (see section 3.4). For the initial iteration of the latter similarity protocol, drug candidates were ranked by their frequency of occurrence in the top 25 most similar compounds to each of the 38 coronavirus actives, while the enhanced iteration ranked compounds by frequency of occurrence in the top 10 most similar compounds to the 85 GHDDI actives. We kept track of the number of coronavirus actives each compound was similar to within the cutoff threshold along with their average ranks (which served as a tie-breaker) to produce the final ranked list of candidates.

The outputs of our pipelines were three ranked lists of drug candidates: one using the direct viral inhibition pipeline from the initial iteration (3.5.20 *de novo*), a second using the similarity based candidate generation pipeline from the initial iteration (3.5.20 similarity), and the third using the similarity based pipeline using the enhanced actives list (5.18.20 similarity).

3.6 External validation studies curation

We analyzed GHDDI (Leng, 2020) and CoronaCentral (Lever and Altman, 2021) for up-to-date information on COVID-19 therapeutic interventions which could independently and prospectively validate our top ranked candidates. Both sources utilize deep learning or natural language processing methods to automatically extract and annotate information from SARS-CoV-2 studies to produce lists of possible actives. We manually parsed the manuscripts that were annotated with and matched the name of any candidate compounds from our three prediction lists for corresponding efficacy values (EC50, IC50, hazard ratios, etc) while eliminating studies that were purely

computational or did not investigate the candidate compound as the primary intervention.

4 Conclusion

This study highlights how CANDO may be used to rapidly generate promising leads for drug development when time is critical, provided the therapeutic intervention is possible within established dosing guidelines. Our study is an assessment of potential therapeutics for treating COVID-19 which were all generated within three months of the pandemic declaration by the WHO. Considering that it took almost one year for a vaccine (Food and Administration, 2022) and two years for a potent antiviral such as molnupiravir or nirmatrelvir (Mahase, 2021; Hammond et al., 2022) to become available, we have exemplified that computational drug discovery and repurposing platforms like ours can be strategically used to alleviate the burden of emergent pathogens ahead of time. Additional studies, ideally *via in vivo* and/or clinical studies, verifying the efficacy of these identified candidates is necessary in most cases, however for already approved drug candidates such as those explored in this study the need for trials demonstrating safety is greatly diminished. Additionally, retrospective EHR analysis may also be used to indirectly examine clinical benefits in human patients as in the case of fluoxetine (Oskotsky et al., 2021).

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 below: http://compbio.buffalo.edu/data/mc_cando_covid19/.

Author contributions

WM conceived the prediction pipelines, research design, approach and methods, conducted all experiments and analysis, implemented all pipelines, and drafted the manuscript. ZF helped with data generation, research design, approach, and methods, and editing the manuscript. RS conceived the prediction pipelines, research design, approach and methods, edited the manuscript, and supervised the overall project. All authors have read and agreed to the published version of the manuscript.

Funding

This work was supported in part by a NIH Director's Pioneer Award (DP1OD006779), a NIH Clinical and Translational Sciences Award (UL1TR001412), NIH T15 Award (T15LM012495), an NCATS ASPIRE Design Challenge Award, an NCATS ASPIRE Reduction-to-Practice Award, and startup funds from the Department of Biomedical Informatics at the University at Buffalo.

Acknowledgments

The authors would like to acknowledge the support provided by the Center for Computational Research at the University at Buffalo. We would also like to thank all members of the Samudrala Computational Biology Group.

Conflict of interest

A patent has been filled with the United States Patent and Trademark Office (USPTO Application number: 63/120,633) claiming some of the small molecule compounds identified using the approach discussed in this manuscript for the treatment of COVID-19, and were validated by Ennaid Therapeutics, LLC in a propreitary study. The compounds exclusive to the patent are not included in the list of the 51 validated actives and are not discussed further in the manuscript.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fphar.2022.970494/full#supplementary-material

References

Ardakani, A. A., Kanafi, A. R., Acharya, U. R., Khadem, N., and Mohammadi, A. (2020). Application of deep learning technique to manage Covid-19 in routine clinical practice using ct images: Results of 10 convolutional neural networks. *Comput. Biol. Med.* 121, 103795. doi:10.1016/j.compbiomed.2020.103795

Baby, K., Maity, S., Mehta, C. H., Suresh, A., Nayak, U. Y., Nayak, Y., et al. (2021). Targeting sars-cov-2 main protease: A computational drug repurposing study. *Arch. Med. Res.* 52, 38–47. doi:10.1016/j.arcmed.2020.09.013

Burley, S. K., Berman, H. M., Bhikadiya, C., Bi, C., Chen, L., Di Costanzo, L., et al. (2019). Rcsb protein Data Bank: Biological macromolecular structures enabling research and education in fundamental biology, biomedicine, biotechnology and energy. *Nucleic Acids Res.* 47, D464–D474. doi:10.1093/nar/gky1004

Chopra, G., Kaushik, S., Elkin, P. L., and Samudrala, R. (2016). Combating ebola with repurposed therapeutics using the cando platform. *Molecules* 21, 1537. doi:10.3390/molecules21121537

Chopra, G., and Samudrala, R. (2016). Exploring polypharmacology in drug discovery and repurposing using the cando platform. *Curr. Pharm. Des.* 22, 3109–3123. doi:10.2174/1381612822666160325121943

Chowdhury, M. S., Rathod, J., and Gernsheimer, J. (2020). A rapid systematic review of clinical trials utilizing chloroquine and hydroxychloroquine as a treatment for Covid-19. *Acad. Emerg. Med.* 27, 493–504. doi:10.1111/acem.14005

Choy, K.-T., Wong, A. Y.-L., Kaewpreedee, P., Sia, S. F., Chen, D., Hui, K. P. Y., et al. (2020). Remdesivir, lopinavir, emetine, and homoharringtonine inhibit sars-cov-2 replication *in vitro*. *Antivir. Res.* 178, 104786. doi:10.1016/j.antiviral.2020.104786

Coenen, S., van Der Velden, A. W., Cianci, D., Goossens, H., Bongard, E., Saville, B. R., et al. (2020). Oseltamivir for coronavirus illness: Post-hoc exploratory analysis of an open-label, pragmatic, randomised controlled trial in European primary care from 2016 to 2018. *Br. J. Gen. Pract.* 70, e444–e449. doi:10.3399/bjgp20X711941

Costin, J. M., Jenwitheesuk, E., Lok, S.-M., Hunsperger, E., Conrads, K. A., Fontaine, K. A., et al. (2010). Structural optimization and de novo design of dengue virus entry inhibitory peptides. *PLoS Negl. Trop. Dis.* 4, e721. doi:10.1371/journal.pntd.0000721

Dupret, G. (2011). Discounted cumulative gain and user decision models. International Symposium on String Processing and Information Retrieval. Springer, 2–13.

Dyall, J., Coleman, C. M., Venkataraman, T., Holbrook, M. R., Kindrachuk, J., Johnson, R. F., et al. (2014). Repurposing of clinically developed drugs for treatment of Middle East respiratory syndrome coronavirus infection. *Antimicrob. Agents Chemother.* 58, 4885–4893. doi:10.1128/AAC.03036-14

Echeverría-Esnal, D., Martin-Ontiyuelo, C., Navarrete-Rouco, M. E., De-Antonio Cusco, M., Ferrández, O., Horcajada, J. P., et al. (2021). Azithromycin in the treatment of Covid-19: A review. *Expert Rev. anti. Infect. Ther.* 19, 147–163. doi:10.1080/14787210.2020.1813024

Ellinger, B., Bojkova, D., Zaliani, A., Cinatl, J., Claussen, C., Westhaus, S., et al. (2021). A sars-cov-2 cytopathicity dataset generated by high-content screening of a large drug repurposing collection. *Sci. Data* 8, 70. doi:10.1038/s41597-021-00848-4

Falls, Z., Mangione, W., Schuler, J., and Samudrala, R. (2019). Exploration of interaction scoring criteria in the cando platform. *BMC Res. Notes* 12, 318. doi:10.1186/s13104-019-4356-3

Fine, J., Lackner, R., Samudrala, R., and Chopra, G. (2019). Computational chemoproteomics to understand the role of selected psychoactives in treating mental health indications. *Sci. Rep.* 9, 13155. doi:10.1038/s41598-019-49515-0

Food and Administration (2022). *Comirnaty and pfizer-biontech covid-19 vaccine*.

Ghandikota, S., Sharma, M., and Jegga, A. G. (2021). Secondary analysis of transcriptomes of sars-cov-2 infection models to characterize Covid-19. *Patterns* 2, 100247. doi:10.1016/j.patter.2021.100247

Group, R. C. (2021). Dexamethasone in hospitalized patients with Covid-19. *N. Engl. J. Med. Overseas. Ed.* 384, 693–704. doi:10.1056/nejmoa2021436

Group (2020). *Cando platform putative drug candidates against covid-19*.

Gysi, D. M., Do Valle, I., Zitnik, M., Ameli, A., Gan, X., Varol, O., et al. (2021). Network medicine framework for identifying drug-repurposing opportunities for Covid-19. *Proc. Natl. Acad. Sci. U. S. A.* 118, e2025581118. doi:10.1073/pnas.2025581118

Hammond, J., Leister-Tebbe, H., Gardner, A., Abreu, P., Bao, W., Wisemandle, W., et al. (2022). Oral nirmatrelvir for high-risk, nonhospitalized adults with Covid-19. *N. Engl. J. Med. Overseas. Ed.* 386, 1397–1408. doi:10.1056/nejmoa2118542

Horst, J. A., Laurenzi, A., Bernard, B., and Samudrala, R. (2012). Computational multitarget drug discovery. *Polypharmacology Drug Discov.*, 263–301. doi:10.1002/9781118098141.ch13

Hudson, M. L., and Samudrala, R. (2021). Multiscale virtual screening optimization for shotgun drug repurposing using the cando platform. *Molecules* 26, 2581. doi:10.3390/molecules26092581

Järvelin, K., and Kekäläinen, J. (2002). Cumulated gain-based evaluation of ir techniques. *ACM Trans. Inf. Syst.* 20, 422–446. doi:10.1145/582415.582418

Jenwitheesuk, E., Horst, J. A., Rivas, K. L., Van Voorhis, W. C., and Samudrala, R. (2008). Novel paradigms for drug discovery: Computational multitarget screening. *Trends Pharmacol. Sci.* 29, 62–71. doi:10.1016/j.tips.2007.11.007

Jenwitheesuk, E., and Samudrala, R. (2005). Identification of potential multitarget antimalarial drugs. *JAMA* 294, 1490–1491. doi:10.1001/jama.294.12.1490

Jenwitheesuk, E., and Samudrala, R. (2003a). Identifying inhibitors of the sars coronavirus proteinase. *Bioorg. Med. Chem. Lett.* 13, 3989–3992. doi:10.1016/j.bmcl.2003.08.066

Jenwitheesuk, E., and Samudrala, R. (2003b). Improved prediction of hiv-1 protease-inhibitor binding energies by molecular dynamics simulations. *BMC Struct. Biol.* 3, 2. doi:10.1186/1472-6807-3-2

Landrum, G. (2013). *Rdkit: A software suite for cheminformatics, computational chemistry, and predictive modeling*.

Leng, L. D. (2020). *Targeting covid-19: Ghddi info sharing portal*.

Lever, J., and Altman, R. B. (2021). Analyzing the vast coronavirus literature with CoronaCentral. *Proc. Natl. Acad. Sci. U. S. A.* 118, e2100766118. doi:10.1073/pnas.2100766118

Liu, Y., Wu, Y., Shen, X., and Xie, L. (2021). Covid-19 multi-targeted drug repurposing using few-shot learning. *Front. Bioinform.* 1, 18. doi:10.3389/fbinf.2021.693177

Mahase, E. (2021). *Covid-19: Molnupiravir reduces risk of hospital admission or death by 50% in patients at risk, msd reports*.

Mangione, W. (2022). *Comprehensive elucidation of small molecule therapeutic behavior using multitarget theory*. Ann Arbor, Michigan: Ph.D. thesis, University. at Buffalo. Copyright - Database copyright ProQuest LLC; ProQuest does not claim copyright in the individual underlying works; Last updated - 2022-03-21.

Mangione, W., and Falls, Z. (2022). cando.py.

Mangione, W., Falls, Z., Chopra, G., and Samudrala, R. (2020a). cando. py: Open source software for predictive bioanalytics of large scale drug–protein–disease data. *J. Chem. Inf. Model.* 60, 4131–4136. doi:10.1021/acs.jcim.0c00110

Mangione, W., Falls, Z., Melendy, T., Chopra, G., and Samudrala, R. (2020b). Shotgun drug repurposing biotechnology to tackle epidemics and pandemics. *Drug Discov. Today* 25, 1126–1128. doi:10.1016/j.drudis.2020.05.002

Mangione, W., Falls, Z., and Samudrala, R. (2022). Effective holistic characterization of small molecule effects using heterogeneous biological networks. bioRxiv.

Mangione, W., and Samudrala, R. (2019). Identifying protein features responsible for improved drug repurposing accuracies using the cando platform: Implications for drug design. *Molecules* 24, 167. doi:10.3390/molecules24010167

Michael, S., Isern, S., Garry, R., Costin, J., Jenwithesuk, E., and Samudrala, R. (2011a). *Optimized dengue virus entry inhibitory peptide (dn81)*.

Michael, S., Isern, S., Garry, R., Costin, J., Jenwithesuk, E., and Samudrala, R. (2011b). *Optimized dengue virus entry inhibitory peptide(1oan1)*.

Minie, M., Chopra, G., Sethi, G., Horst, J., White, G., Roy, A., et al. (2014). Cando and the infinite drug discovery frontier. *Drug Discov. Today* 19, 1353–1363. doi:10.1016/j.drudis.2014.06.018

Mohamed, K., Yazdanpanah, N., Saghazadeh, A., and Rezaei, N. (2021). Computational drug discovery and repurposing for the treatment of Covid-19: A systematic review. *Bioorg. Chem.* 106, 104490. doi:10.1016/j.bioorg.2020.104490

Moukheiber, L., Mangione, W., Maleki, S., Falls, Z., Gao, M., and Samudrala, R. (2021). Identifying protein features and pathways responsible for toxicity using machine learning, cando, and tox21 datasets: Implications for predictive toxicology. *bioRxiv*.

Nicholson, C. O., Costin, J. M., Rowe, D. K., Lin, L., Jenwitheesuk, E., Samudrala, R., et al. (2011). Viral entry inhibitors block dengue antibody-dependent enhancement *in vitro*. *Antivir. Res.* 89, 71–74. doi:10.1016/j.antiviral.2010.11.008

Organization (2022). *Who coronavirus (covid-19) dashboard*.

Oskotsky, T., Marić, I., Tang, A., Oskotsky, B., Wong, R. J., Aghaeepour, N., et al. (2021). Mortality risk among patients with Covid-19 prescribed selective serotonin reuptake inhibitor antidepressants. *JAMA Netw. Open* 4, e2133090. doi:10.1001/jamanetworkopen.2021.33090

Overhoff, B., Falls, Z., Mangione, W., and Samudrala, R. (2021). A deep-learning proteomic-scale approach for drug design. *Pharmaceuticals* 14, 1277. doi:10.3390/ph14121277

Ozturk, T., Talo, M., Yildirim, E. A., Baloglu, U. B., Yildirim, O., Acharya, U. R., et al. (2020). Automated detection of Covid-19 cases using deep neural networks with x-ray images. *Comput. Biol. Med.* 121, 103792. doi:10.1016/j.compbiomed.2020.103792

Pham, T.-H., Qiu, Y., Zeng, J., Xie, L., and Zhang, P. (2021). A deep learning framework for high-throughput mechanism-driven phenotype compound screening and its application to Covid-19 drug repurposing. *Nat. Mach. Intell.* 3, 247–257. doi:10.1038/s42256-020-00285-9

Samudrala, R., Falls, Z., and Mangione, W. (2020). *Coronavirus treatment compositions and methods*.

Schuler, J., Falls, Z., Mangione, W., Hudson, M. L., Bruggemann, L., Samudrala, R., et al. (2021). Evaluating the performance of drug-repurposing technologies. *Drug Discov. Today* 27, 49–64. doi:10.1016/j.drudis.2021.08.002

Schuler, J., Mangione, W., Samudrala, R., and Ceusters, W. (2019). Foundations for a realism-based drug repurposing ontology. In 10th Annual International Conference on Biomedical Ontology. 1–8.

Schuler, J., and Samudrala, R. (2019). Fingerprinting cando: Increased accuracy with structure-and ligand-based shotgun drug repurposing. *ACS omega* 4, 17393–17403. doi:10.1021/acsomega.9b02160

Sethi, G., Chopra, G., and Samudrala, R. (2015). Multiscale modelling of relationships between protein classes and drug behavior across all diseases using the cando platform. *Mini Rev. Med. Chem.* 15, 705–717. doi:10.2174/1389557515666150219145148

Shen, L., Niu, J., Wang, C., Huang, B., Wang, W., Zhu, N., et al. (2019). High-throughput screening and identification of potent broad-spectrum inhibitors of coronaviruses. *J. Virol.* 93, e0002319. doi:10.1128/JVI.00023-19

Spinner, C. D., Gottlieb, R. L., Criner, G. J., López, J. R. A., Cattelan, A. M., Viladomiu, A. S., et al. (2020). Effect of remdesivir vs standard care on clinical status at 11 days in patients with moderate Covid-19: A randomized clinical trial. *Jama* 324, 1048–1057. doi:10.1001/jama.2020.16349

Vijayan, V., Pant, P., Vikram, N., Kaur, P., Singh, T., Sharma, S., et al. (2020). Identification of promising drug candidates against nsp16 of sars-cov-2 through computational drug repurposing study. *J. Biomol. Struct. Dyn.* 39, 6713–6727. doi:10.1080/07391102.2020.1802349

Wang, J. (2020). Fast identification of possible drug treatment of coronavirus disease-19 (Covid-19) through computational drug repurposing study. *J. Chem. Inf. Model.* 60, 3277–3286. doi:10.1021/acs.jcim.0c00179

Wang, M., Cao, R., Zhang, L., Yang, X., Liu, J., Xu, M., et al. (2020a). Remdesivir and chloroquine effectively inhibit the recently emerged novel coronavirus (2019-ncov) *in vitro*. *Cell Res.* 30, 269–271. doi:10.1038/s41422-020-0282-0

Wang, Y., Zhang, D., Du, G., Du, R., Zhao, J., Jin, Y., et al. (2020b). Remdesivir in adults with severe Covid-19: A randomised, double-blind, placebo-controlled, multicentre trial. *Lancet* 395, 1569–1578. doi:10.1016/S0140-6736(20)31022-9

Wishart, D. S., Feunang, Y. D., Guo, A. C., Lo, E. J., Marcu, A., Grant, J. R., et al. (2018). Drugbank 5.0: A major update to the drugbank database for 2018. *Nucleic Acids Res.* 46, D1074–D1082. doi:10.1093/nar/gkx1037

Yang, J., Roy, A., and Zhang, Y. (2013). Protein–ligand binding site recognition using complementary binding-specific substructure comparison and sequence profile alignment. *Bioinformatics* 29, 2588–2595. doi:10.1093/bioinformatics/btt447

Yang, J., Yan, R., Roy, A., Xu, D., Poisson, J., Zhang, Y., et al. (2015). The i-tasser suite: Protein structure and function prediction. *Nat. Methods* 12, 7–8. doi:10.1038/nmeth.3213

Zhou, Y., Hou, Y., Shen, J., Huang, Y., Martin, W., Cheng, F., et al. (2020). Network-based drug repurposing for novel coronavirus 2019-ncov/sars-cov-2. *Cell Discov.* 6, 14. doi:10.1038/s41421-020-0153-3

frontiers | Frontiers in Pharmacology

TYPE Original Research
PUBLISHED 02 September 2022
DOI 10.3389/fphar.2022.1003310

OPEN ACCESS

EDITED BY
Mithun Rudrapal,
Rasiklal M. Dhariwal Institute of Pharmaceutical Education and Research, India

REVIEWED BY
Habibu Tijjani,
Bauchi State University, Nigeria
Johra Khan,
Majmaah University, Saudi Arabia

*CORRESPONDENCE
Lixin Zhang,
534259627@qq.com,
orcid.org/0086-135-2919-4157
Wang Zan,
33058766@qq.com,
orcid.org/0086-186-2812-0209
Xiao Zhang,
954073462@qq.com,
orcid.org/0086-130-8661-6376

SPECIALTY SECTION
This article was submitted to Drugs Outcomes Research and Policies, a section of the journal Frontiers in Pharmacology

RECEIVED 26 July 2022
ACCEPTED 11 August 2022
PUBLISHED 02 September 2022

CITATION
Cao J-F, Gong Y, Wu M, Yang X, Xiong L, Chen S, Xiao Z, Li Y, Zhang L, Zan W and Zhang X (2022), Exploring the mechanism of action of licorice in the treatment of COVID-19 through bioinformatics analysis and molecular dynamics simulation.
Front. Pharmacol. 13:1003310.
doi: 10.3389/fphar.2022.1003310

Exploring the mechanism of action of licorice in the treatment of COVID-19 through bioinformatics analysis and molecular dynamics simulation

Jun-Feng Cao[1,2], Yunli Gong[3], Mei Wu[1], Xingyu Yang[1], Li Xiong[1], Shengyan Chen[1], Zixuan Xiao[3], Yang Li[3], Lixin Zhang[4]*, Wang Zan[5]* and Xiao Zhang[2]*

[1]Clinical Medicine, Chengdu Medical College, Chengdu, China, [2]Chengdu Medical College of Basic Medical Sciences, Chengdu, China, [3]Laboratory Medicine, Chengdu Medical College, Chengdu, China, [4]Yunnan Academy of Forestry Sciences, Kunming, China, [5]Chengdu Medical College of Pharmacy, Chengdu, China

Purpose: The rapid worldwide spread of Corona Virus Disease 2019 (COVID-19) has become not only a global challenge, but also a lack of effective clinical treatments. Studies have shown that licorice can significantly improve clinical symptoms such as fever, dry cough and shortness of breath in COVID-19 patients with no significant adverse effects. However, there is still a lack of in-depth analysis of the specific active ingredients of licorice in the treatment of COVID-19 and its mechanism of action. Therefore, we used molecular docking and molecular dynamics to explore the mechanism of action of licorice in the treatment of COVID-19.

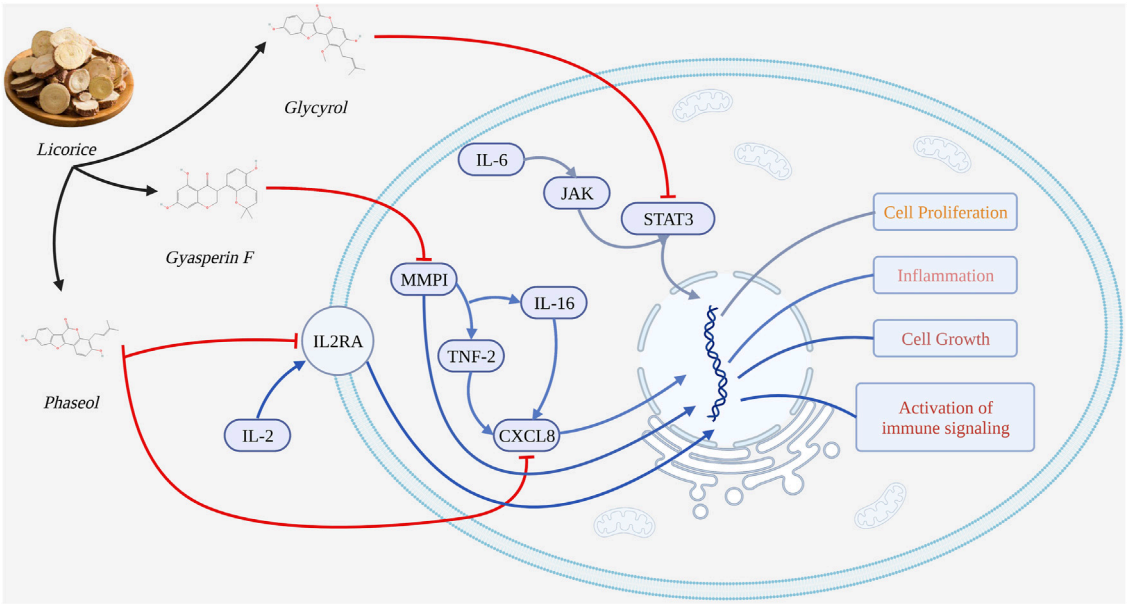
Methods: We used bioinformatics to screen active pharmaceutical ingredients and potential targets, the disease-core gene target-drug network was established and molecular docking was used for verification. Molecular dynamics simulations were carried out to verify that active ingredients were stably combined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

Results: Licorice had 255 gene targets, COVID-19 had 4,628 gene targets, the intersection gene targets were 101. Kyoto Encyclopedia of Genes and Genomes (KEGG) and Gene ontology (GO) analysis showed that licorice played an important role mainly through the signaling pathways of inflammatory factors and oxidative stress. Molecular docking showed that Glycyrol, Phaseol and Glyasperin F in licorice may playe a role in treating COVID-19 by acting on STAT3, IL2RA, MMP1, and CXCL8. Molecular dynamics were used to demonstrate and analyze the binding stability of active ingredients to protein targets.

Conclusion: This study found that Phaseol in licorice may reduce inflammatory cell activation and inflammatory response by inhibiting the activation of CXCL8 and IL2RA; Glycyrol may regulate cell proliferation and survival by acting on STAT3. Glyasperin F may regulate cell growth by inhibiting the activation of MMP1, thus reducing tissue damage and cell death caused by excessive inflammatory response and promoting the growth of new tissues. Therefore, licorice is proposed as an effective candidate for the treatment of COVID-19 through STAT3, IL2RA, MMP1, and CXCL8.

KEYWORDS

licorice, COVID-19, bioinformatics analysis, molecular docking, molecular dynamics



GRAPHICAL ABSTRACT
The mechanisms analysis of licorice in the treatment of COVID-19.

Introduction

Corona Virus Disease 2019 (COVID-19) is a respiratory disease caused by Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) (Fernandes et al., 2022). Signs and symptoms of COVID-19 disease vary from patient to patient, but the most common clinical signs include fever, fatigue, cough, anorexia, sputum production and shortness of breath (Rai et al., 2021). Less common symptoms such as sore throat, headache, confusion, hemoptysis, shortness of breath, and chest tightness, as well as mild symptoms such as nausea, vomiting, diarrhea, and gastrointestinal complications have also been reported (Majumder and Minko, 2021). SARS-CoV-2 transmission usually occurs via respiratory droplets with an average incubation period of 6.4 days. Although most patients tend to be mildly ill, a small number of patients develop severe hypoxia requiring hospitalization and mechanical ventilation (Ochani et al., 2021). In severe cases, pneumonia, severe acute respiratory syndrome, heart failure, renal failure, and even death occur (Rai et al., 2021). However, there is a lack of effective COVID-19 therapeutic agents with few side effects. Therefore, screening and investigating drugs to treat COVID-19 will contribute significantly to the global fight against the COVID-19 epidemic.

Clinical evidence suggests that herbal drugs are effective against viral infections such as influenza, SARS and SARS-CoV-2 by targeting viral cell entry, viral replication and host

antiviral immune response steps. Among the drugs and formulations recommended by Chinese authorities for COVID-19 treatment, the dried root of licorice is one of the most commonly used ingredients in formulations. Recent reports also suggest that licorice extracts may play a potential role in the fight against COVID-19 and related diseases (Li et al., 2021). According to the Chinese Pharmacopoeia, licorice is able to nourish the spleen, remove heat, prevent toxicity, remove phlegm, and relieve cough, cramps and pain, thus harmonizing the effects of other drugs (Ng et al., 2021).

Many studies have reported that active compounds isolated from licorice have antitumor, antibacterial, antiviral, anti-inflammatory, immunomodulatory and several other activities that help restore and protect the nervous, digestive, respiratory, endocrine and cardiovascular systems (Yang et al., 2015). Licorice has many pharmacological effects and is often used as a unique "guiding drug," accounting for more than half of the traditional and modern prescriptions and formulations. The modulating effects of licorice on other herbs include significant detoxification, treatment of drug and food poisoning, or suppression of adverse reactions, and this "guiding" effect has been tested in many preparations. According to available studies, the pharmacological effects of licorice and natural products such as glycyrrhizin have beneficial effects on the prevention of some immune reactions triggered by COVID-19 (Zhang et al., 2021). In addition to antiviral and anti-inflammatory properties, one of the components of licorice has a mechanism to enhance autophagy, which studies have shown to be necessary for COVID-19 treatment (Abraham and Florentine, 2021).

Numerous studies have been conducted to find many active components in licorice that can hinder SARS-COV-2 infection and alleviate the clinical symptoms of COVID-19. Gomaa and Abdel-Wadood demonstrated the antiviral activity of licorice sweeteners and licorice extracts. The most common mechanism of antiviral activity is due to disruption of viral uptake into host cells and disruption of the interaction between SARS-COV2 and the receptor binding structural domain (RBD) of ACE2 (Gomaa and Abdel-Wadood, 2021). Luo found that quercetin, the active component of licorice, has a strong docking ability with IL-6, suggesting that licorice may primarily reduce IL-6 levels in response to COVID-19 inflammatory outbreaks, which represents a prospective therapeutic strategy for moderate COVID-19 (Luo et al., 2022). Yi et al. found that the triterpenoid licorice saponin A3 (A3) and glycyrrhizic acid (GA) could effectively inhibit SARS-CoV-2 by targeting nsp7 and the stinging protein RBD, respectively (Yi et al., 2022). However, licorice as a traditional Chinese medicine contains a large number of active ingredients, and the complex drug composition seriously hinders the application of licorice in clinical COVID-19 treatment, and the specific mechanism of action of licorice for the treatment of COVID-19 is still unclear. Molecular dynamics allows a comprehensive and systematic simulation of the interaction and binding stability between small molecule monomers and protein targets with the help of powerful computational capabilities.

Molecular dynamics (MD) is based on large computer clusters (even supercomputers) and aims to computationally obtain data on the microstructure, physicochemical properties, and performance characterization parameters of materials (Collier et al., 2020). Molecular dynamics complements and digs deeper into the traditional materials discipline, which is mainly experimental. The data obtained from calculations are used to study and analyze the mechanism behind the experiments at multiple levels from micro, meso and macro scales (Nam, 2021). Molecular dynamics simulations help to discover the relationships on protein, protein-ligand, protein-protein, protein-DNA and other biomolecular interactions (Al-Shar'i and Al-Balas, 2019). Molecular dynamics simulations not only help to understand the physical processes of systems at the atomic level, but also allow the discovery of empirically undetectable hidden states. In addition, experimental measurements of thermodynamic properties in biomolecular systems are usually expensive and time-consuming (Filipe and Loura, 2022). Accurate theoretical calculations of their free energies by numerical simulations are becoming increasingly important in medical biology, where 3D structures of small molecule-protein complexes can reveal how and where a protein interacts with a drug small molecule.

In this study, we screened licorice for potential active small molecules by bioinformatics. The core intersection targets of licorice and COVID-19 were screened. Protein-protein interaction (PPI), Kyoto Encyclopedia of Genes and Genomes (KEGG) and Gene ontology (GO) were used to analyze the potential association among the core intersection targets to explore the mechanism of action and potential pathways. To further validate the relationship between active small molecules and key protein targets we performed molecular motion system simulations through a supercomputer platform. Molecular motion system simulations enable systematic study and analysis of drugs to treat diseases from the cellular level to the chemical moiety level. Molecular docking was used to determine the affinity of monomeric compounds to protein targets, and molecular dynamics was used to simulate the stability of bound complexes and to analyze the dynamics of complexes after binding.

Therefore, this study of the potential mechanism of licorice in the treatment of COVID-19 may provide new ideas and necessary theoretical basis for clinical treatment.

Material and methods

Identification and screening of active compounds of licorice

In this study, all compounds of licorice were screened and analyzed using the Traditional Chinese Medicine System Pharmacology Database (TCMSP) (Xie et al., 2021). We

evaluated the drug components in terms of absorption, distribution, metabolism and excretion and screened by two key parameters, oral bioavailability (OB) and drug similarity (DL). OB largely determines the impact of drug small molecules on disease and DL is used for early screening and refinement of candidate compounds in drug development. Active compounds of licorice were screened on the basis of OB ≥ 30% and DL ≥ 0.18.

Analysis and screening of core intersection gene targets

We used the GeneCards database and "COVID-19" and "SAR-Cov-2" were used as keywords to obtain disease gene targets. We also imported licorice into the GeneCards database to obtain drug gene targets. Drug gene targets and disease gene targets were intersected through the venny website to obtain intersecting gene targets. And the intersecting gene targets were screened to obtain core intersecting gene targets by relevance score ≥2 as a threshold, which is a comprehensive evaluation of the association of genes with the studied diseases.

Construction of protein-protein interaction network for corona virus disease 2019 interaction in licorice treatment

The STRING database was used to analyze protein-protein interactions (PPI) for licorice treatment of COVID-19. In this study, all the core intersecting targets were imported into Cytoscape 3.7.1 for analysis in order to elucidate the interactions between potential protein targets (Pan et al., 2020). The network topology parameters were analyzed by Cytoscape 3.7.1, and the hub protein targets were screened according to the criteria of nodal degree value and median centroid value greater than the mean.

Gene target enrichment analysis

Interacting gene targets were analyzed by Gene Ontology (GO) functional annotation and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment in the DAVID database. In this study, the relevant biological processes (BP), cellular components (CC) and molecular functions (MF) of the gene targets were obtained by GO enrichment. The core intersecting targets were imported into the DAVID database and the selected species was "*Homo sapiens*" (Xiong et al., 2020). We performed KEGG pathway enrichment analysis for the relevant signaling pathways involved in the disease-related targets and performed gene target screening at $p < 0.05$. The main biological processes and signaling pathways were analyzed for licorice treatment of COVID-19. The Omicshare tool platform was used to visualize the results of GO enrichment and KEGG enrichment (Cao et al., 2022).

Validation of molecular docking and docking protocols

Molecular docking was used to study the molecular affinity of the active small molecules of licorice to the COVID-19 protein target. The crystal structures of the proteins used for docking were downloaded from the PDB database and the 3D structures of the small molecules were downloaded from the PUBCHEM database. We used AutoDock Vina 1.1.2 software for the molecular docking work. Prior to docking, PyMol 2.5 was used to process all receptor proteins (Burley et al., 2017). ADFRsuite 1.0 was used to convert all processed small molecules and receptor proteins into the PDBQT format required for docking with AutoDock Vina 1.1.2. The docked conformation with the highest output score was considered to be the binding conformation for subsequent molecular dynamics simulations (Ravindranath et al., 2015). In this study, the original crystal ligand of the protein target was used as a positive reference by re-docking the original crystal ligand and the protein. The consistency of the binding pattern can indicate the correctness of the molecular docking scheme (Cao et al., 2022).

Molecule dynamics

In this study, the small molecule-protein complexes obtained by molecular docking were used as the initial structures for all-atom molecular dynamics simulations, respectively (Mithun et al., 2022). AMBER 18 software was used for the molecular dynamics simulations (Maier et al., 2015; Lee et al., 2020). The LEaP module was used to add hydrogen atoms to the system, a truncated octahedral TIP3P solvent box was added at a distance of 10 Å from the system, and Na+/Cl-was added to the system to balance the system charge. At the maintenance temperature of 298.15 K, the NVT (isothermal isomer) system simulation was performed for 500 ps to further distribute the solvent molecules uniformly in the solvent box. The equilibrium simulation of the whole system was performed for 500 ps at NPT (isothermal isobaric). Finally, two composite systems are simulated for 50 ns of NPT system under periodic boundary conditions (Larini et al., 2007).

MMGBSA binding free energy calculation

The binding free energy between the protein and ligand for all systems was calculated by the MM/GBSA method (Chen et al.,

TABLE 1 The core active compounds in licorice.

MOL ID	molecule_name	OB	MW	DL
MOL002311	Glycyrol	90.77	366.39	0.66
MOL004990	7,2′,4′-trihydroxy-5-methoxy-3-arylcoumarin	83.71	300.28	0.27
MOL004904	Licopyranocoumarin	80.36	384.41	0.65
MOL004891	Shinpterocarpin	80.29	322.38	0.72
MOL005017	Phaseol	78.76	336.36	0.57
MOL004841	Licochalcone B	76.75	286.30	0.19
MOL004810	Glyasperin F	75.83	354.38	0.53
MOL001484	Inermine	75.18	284.28	0.53
MOL000500	Vestitol	74.65	272.32	0.20
MOL005007	Glyasperins M	72.67	368.41	0.59
MOL004941	(2R)-7-hydroxy-2-(4-hydroxyphenyl)chroman-4-one	71.12	256.27	0.18

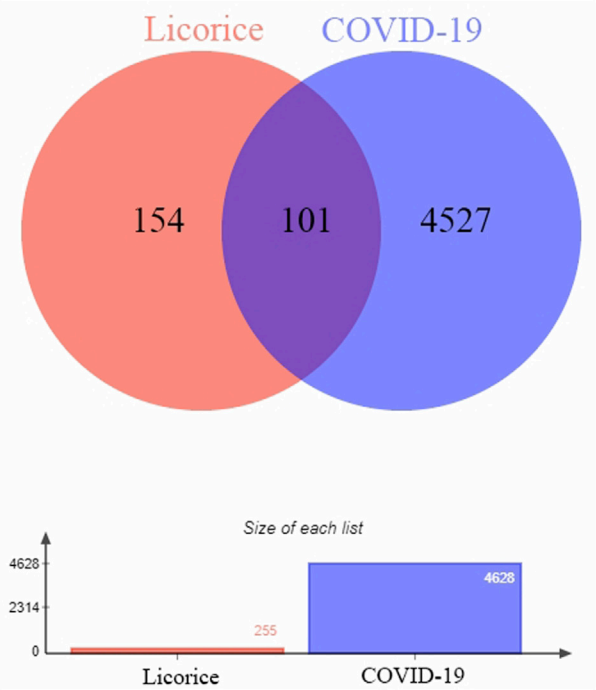


FIGURE 1
Intersection targets-active ingredient networks. Targets of the intersection of licorice and COVID-19.

2020). The MD trajectory of 50 ns was used as the calculation in this study. The calculation equations ars as follows:

$$\Delta G_{bind} = \Delta G_{complex} - \left(\Delta G_{receptor} + \Delta G_{ligand}\right)$$
$$= \Delta E_{internal} + \Delta E_{VDW} + \Delta E_{elec} + \Delta G_{GB} + \Delta G_{SA}$$

In this formula, the non-polar solvation free energy (ΔG_{GA}) was calculated based on solvent accessible surface area (SA) and the product of surface tension (γ), $\Delta G_{GA} = 0.0072 \times SASA$ (Cao et al., 2022).

Results

Identification of potentially active compounds in licorice

The identification of potentially active compounds in licorice was based on the criteria of DL ≥ 0.18 and OB ≥ 30%. 200 potential compounds in licorice were retrieved from the TCMSP database. By further improving the OB score (OB ≥ 70%), 11 core active compounds were screened from licorice, shown in Table 1.

Acquisition of intersectional target genes

In this study, 255 gene targets of licorice and 4,628 gene targets of COVID-19 were obtained. A total of 101 intersecting gene targets were processed by Venny, shown in Figure 1.

Core intersectional target screening and protein interaction network diagram construction

In this study, core intersectional gene targets were obtained from the GeneCards database based on relevance score, and relevance score ≥2 were considered as core intersectional gene targets. The STRING database was used to analyze the 27 core intersectional protein targets of COVID-19 and licorice, and a protein interaction network diagram was constructed for the treatment of COVID-19 with licorice, shown in Figure 2A. 11 key intersectional protein targets (such as: STAT3, IL2RA, CXCL8, etc.) were obtained by increasing the confidence score (confidence level ≥0.95), and the

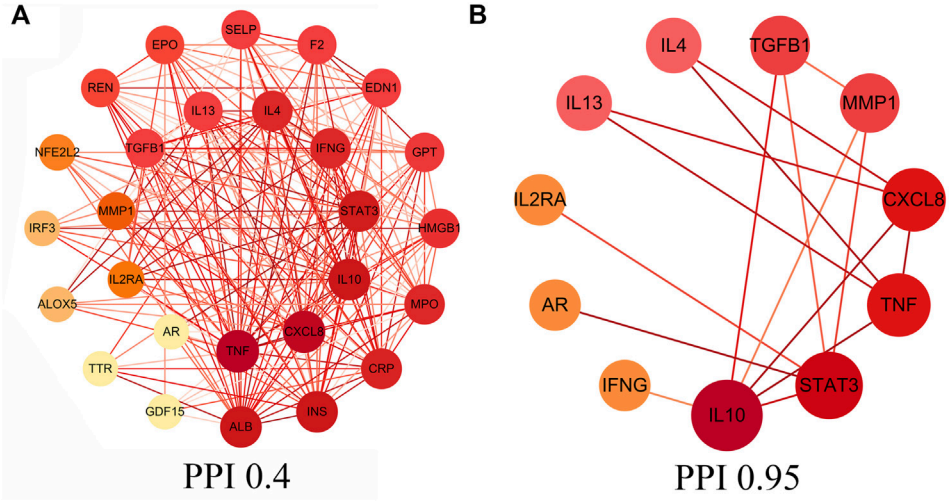


FIGURE 2
Protein-protein interaction (PPI) network. **(A)** PPI network of protein targets, **(B)** PPI network of key protein targets (confidence>0.95).

11 key intersectional protein targets were used to construct the key protein interaction network diagram, shown in Figure 2B.

Gene ontolog and kyoto encyclopedia of genes and genomes enrichment analysis

The 27 core intersectional gene targets were imported into the DAVID database for enrichment analysis. At $p < 0.05$, the GO enrichment analysis yielded 222 GO entries, including 193 BP entries, 10 CC entries and 19 MF entries. The results showed that biological processes were highly correlated with inflammation and cytokine transmission, mainly involving the positive regulation of gene expression, cytokine-mediated signaling pathway and inflammatory response. In cellular component, external side of plasma membrane, extracellular space and extracellular region account for a relatively large amount. In molecular functions, transcription regulatory region sequence-specific DNA binding, cytokine activity and growth factor activity were relatively high, shown in Figures 3A–F. KEGG pathway analysis yielded 72 pathways, and KEGG enrichment analysis showed that the enriched pathways involved multiple pathways related to immune response regulation and inflammation, mainly cytokine-cytokine receptor interaction, pathways in cancer, inflammatory bowel disease and other signaling pathways, shown in Figures 3G, H.

Disease-core gene target-drug network

The disease-core gene target-drug network was constructed to demonstrate the main signaling pathways and biological processes of licorice for the treatment of COVID-19, shown in Figure 4.

Molecular docking

The 11 key intersection protein targets were selected for molecular docking. The results indicate that the CXCL8/Phaseol complex was mainly maintained by hydrophobic interactions. The small molecule Phaseol interacted with E29 on the protein by hydrogen bonding and with V25, V27, V58, and I22 by hydrophobic interactions, shown in Figure 5A. The binding of the IL2RA/Phaseol complex was maintained mainly by hydrogen bonding and hydrophobic interactions. The small molecule Phaseol interacted with Y119, E116, R117, T14, and E9 on the protein by hydrogen bonding and with Y119, F121, F15, E9, and E116 by hydrophobic interactions. In addition, we also observed pi-pi conjugation between Phaseol and F15, shown in Figure 5B. In the MMP1/Glyasperin F complex, the small molecule Glyasperin F interacted with A84 on the protein by hydrogen bonding and with H83, V115, L81, Y140, and H118 by hydrophobic interactions, shown in Figure 5C. The binding of STAT3/Glycyrol indicated that the small molecule Glycyrol hydrogen bonds with S611, E612, and S613 on the protein, hydrophobic interaction with P629 and S613, and also cation pi conjugation with R609, shown in Figure 5D. The molecular docking results score are shown in Figure 6.

Molecular dynamics results

The root mean square deviation of the molecular dynamics simulations can reflect the motility of the complexes, and the larger RMSD and the more intense fluctuations indicate intense motility. The simulation results suggested that the RMSD fluctuations of MMP1/Glyasperin F and STAT3/Glycyrol were

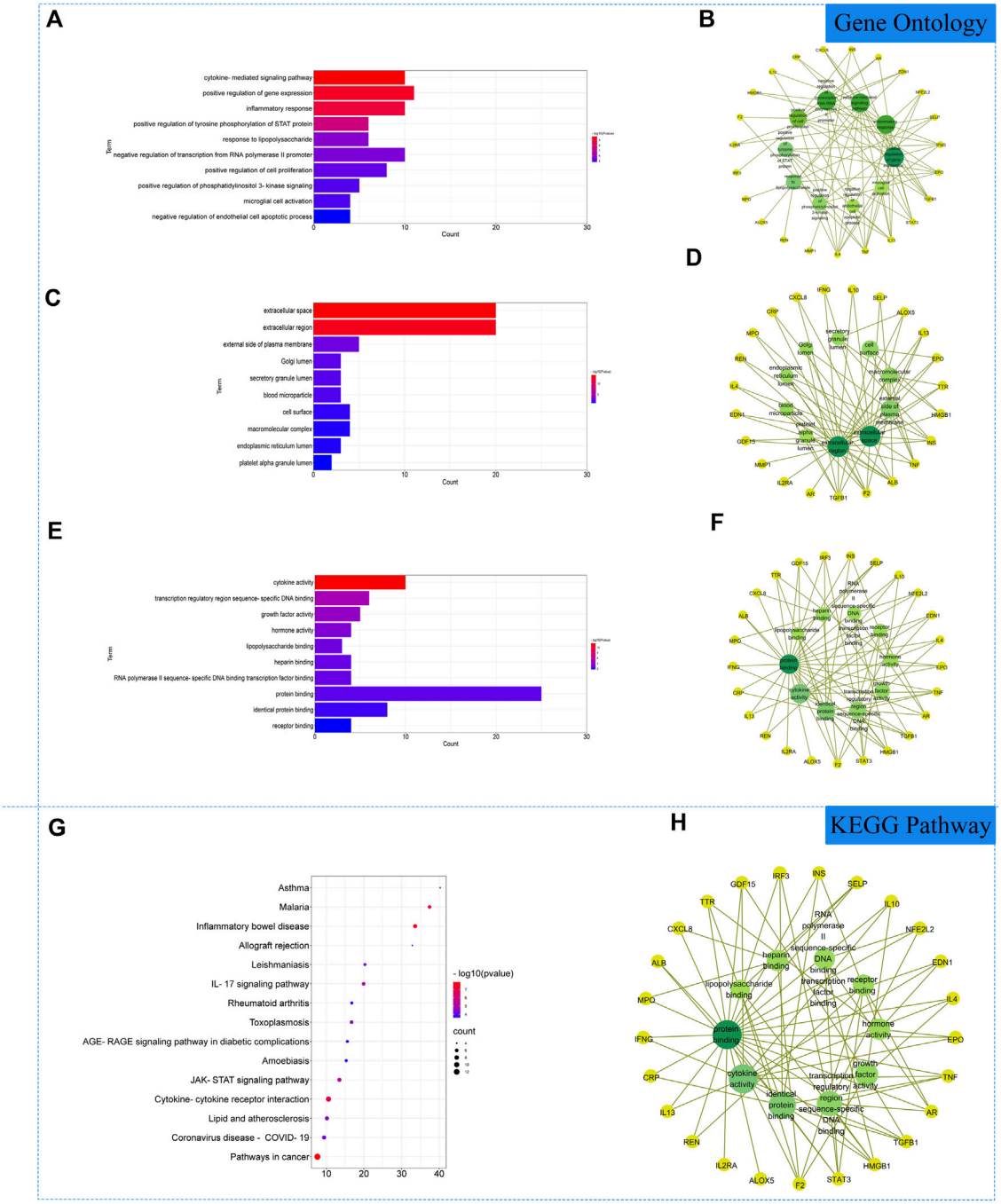


FIGURE 3
Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG). Analysis of related genes. **(A)** The top 10 terms in biological processes (BP) were greatly enriched. **(B)** The subnetwork displayed the top 10 BP terms and related genes. **(C)** The top 10 terms in cellular components (CC) were greatly enriched. **(D)** The subnetwork displayed the top 10 CC terms and related genes. **(E)** The top 10 terms in molecular function (MF) were greatly enriched. **(F)** The subnetwork displayed the top 10 MF terms and related genes. **(G)** The top 15 KEGG pathways were showed. **(H)** The subnetworks displayed the top 15 KEGG pathways.

within 4 Å, which implied that the system was less kinetic. Therefore, combining the magnitude of RMSD and stability, we can determine the stability of these complexes from strong to weak in the order of STAT3/Glycyrol, MMP1/Glyasperin F, CXCL8/Phaseol, and IL2RA/Phaseol. The results are shown in Figure 7.

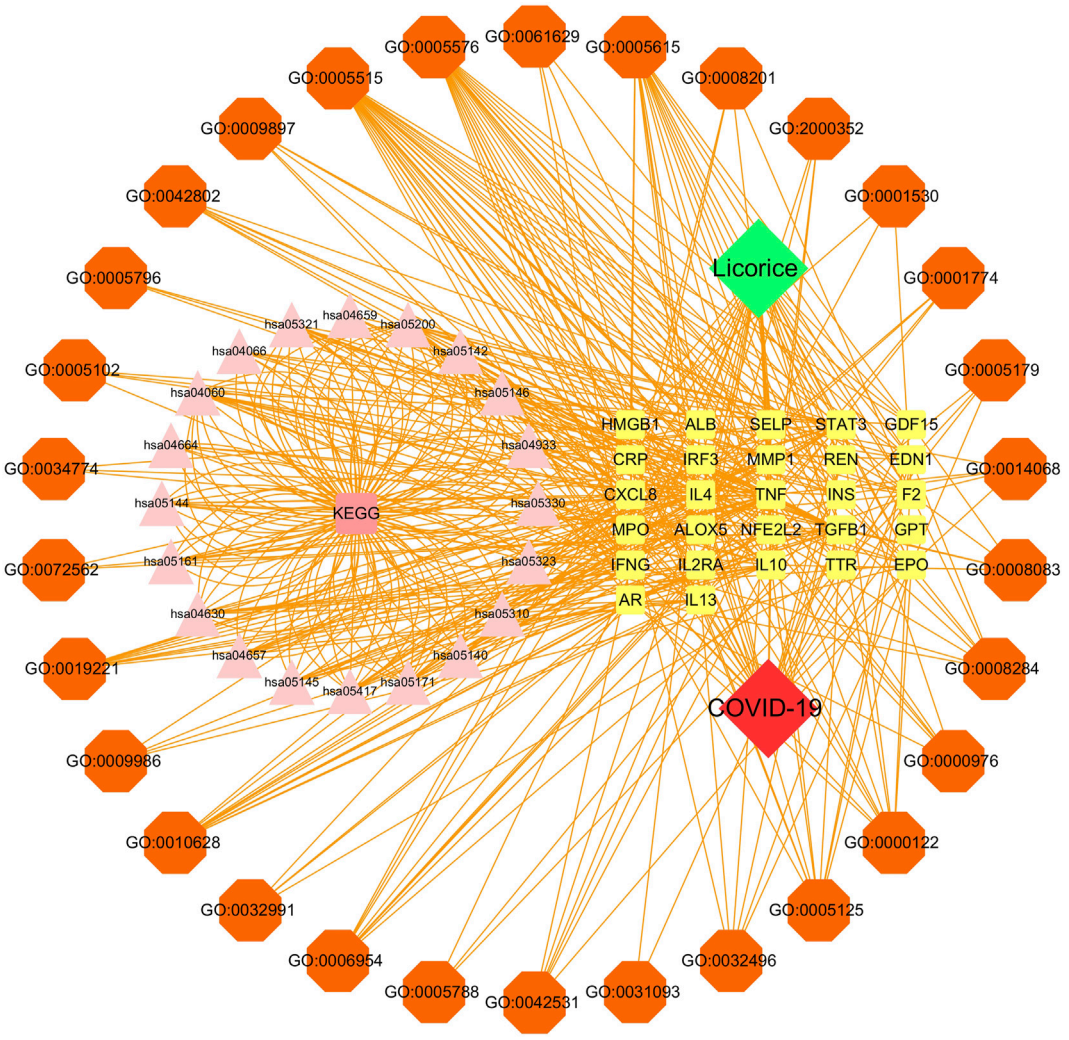


FIGURE 4
Disease-core gene target-drug network. Square nodes represent gene targets, triangular nodes represent signaling pathways (KEGG), and octagonal nodes represent gene ontology (GO) of related genes.

Combined free energy calculation results

Based on the trajectory of molecular dynamics simulations, the binding energy was calculated in this study using the MM-GBSA method. The binding energy can more accurately reflect the binding mode of small molecules and target proteins. The experimental results showed that CXCL8/Phaseol, IL2RA/Phaseol, MMP1/Glyasperin F, STAT3/Glycyrol were −39.51 ± 2.06 kcal/mol, −20.12 ± 3.38 kcal/mol, −43.70 ± 1.80 kcal/mol, −11.85 ± 1.06 kcal/mol. Negative values indicate that these two molecules have binding affinity to the target protein, and lower values indicate stronger binding. The simulation results suggested that these molecules and the corresponding binding affinities are very strong. The MMP1/Glyasperin F binding energy was the highest, with a value of 43.70 ± 1.80 kcal/mol. The binding energies of these complexes were mainly contributed by van der Waals energy as well as electrostatic energy. The experimental results are shown in Table 2.

Hydrogen bond analysis

Hydrogen bonding is one of the strongest non-covalent binding interactions, and a higher number indicates better binding. The experimental results suggested that the number of hydrogen bonds of the four complexes was basically 1-2 in the middle and late stages of the simulation. Among them, the hydrogen bonding diagram of MMP1/Glyasperin F complex showed more sparse in the late stage of simulation, implying that hydrogen bonding was not the main force for it to maintain stability. Combining the results of MGBSA and the binding pattern, we suggested that hydrophobic interaction

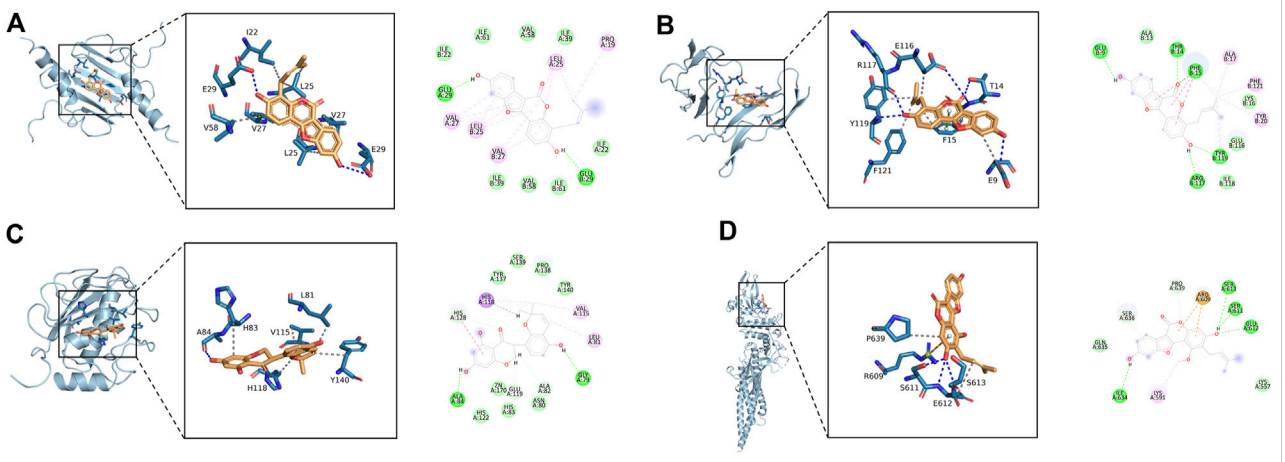


FIGURE 5
Molecular docking of active ingredients and core targets. **(A)** CXCL8/Phaseol, **(B)** IL2RA/Phaseol, **(C)** MMP1/Glyasperin F, **(D)**STAT3/Glycyrol.

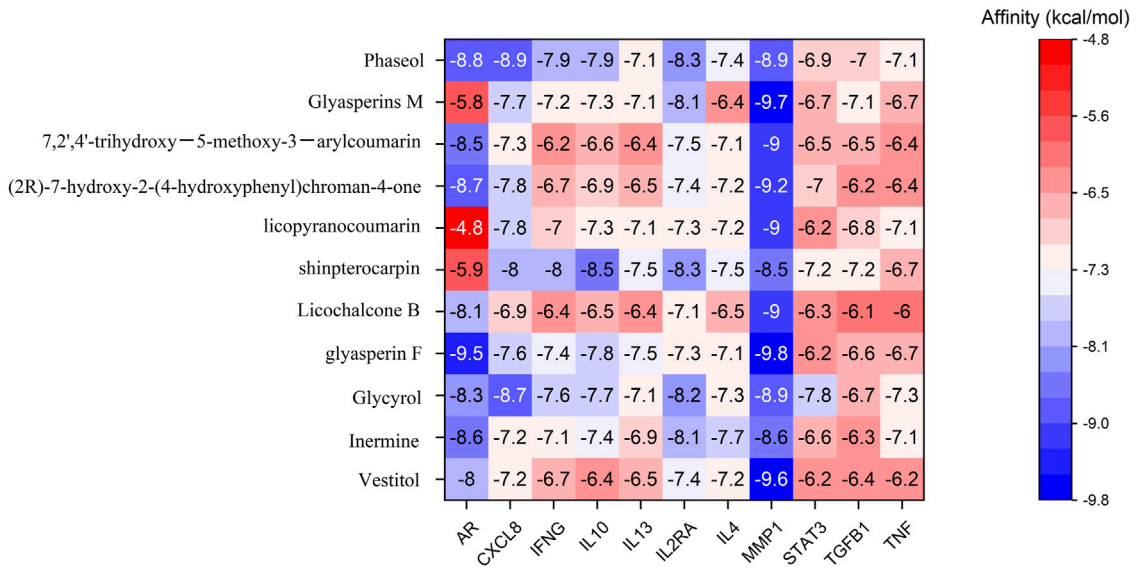


FIGURE 6
Screening docking results between ligands and receptors.

was the main force for MMP1/Glyasperin F to maintain stability. The results are shown in Figure 8.

The stability of the target protein at the residue level

RMSF can respond to the flexibility of the protein during molecular dynamics simulation. Usually the protein flexibility decreases after the drug binds to the protein, which in turn achieves the effect of stabilizing the protein while exerting the enzymatic activity. The simulation results showed that the RMSF of proteins in MMP1/Glyasperin F, STAT3/Glycyrol were low. Especially for MMP1 protein, the RMSF of most of the dashed lines was below 2 Å, implying that the complex binding was more stable. In contrast, the RMSFs of the proteins in IL2RA/Phaseol and CXCL8/Phaseol were larger, suggesting that these two proteins were more flexible. The results are shown in Figure 9.

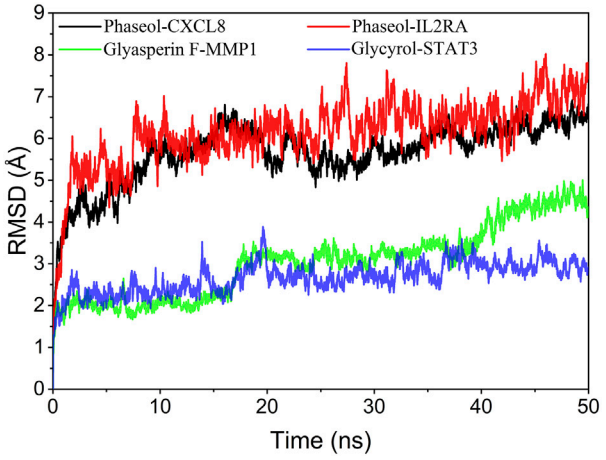


FIGURE 7
Complex root mean square deviation (RMSD) difference over time.

Analysis of the radius of gyration

The radius of gyration can reflect the degree of compactness of the complex, and the size of fluctuation can be very intuitive to determine the compactness or system convergence. The fluctuations of the radius of gyration were MMP1/Glyasperin F, STAT3/Glycyrol, CXCL8/Phaseol, IL2RA/Phaseol from the largest to the smallest, respectively. The results are shown in Figure 10.

Analysis of solvent accessible surface area

The Solvent Accessible Surface Area (SASA) is calculated as the interface surrounded by the solvent. The larger the area indicates that the complex can interact with the aqueous solution. In addition, the fluctuation of SASA reflects the exposure of the protein surface and the change of the buried area. The fluctuations of SASA suggested that MMP1/Glyasperin F, STAT3/Glycyrol, CXCL8/Phaseol fluctuated less and the SASA values were small. The results are shown in Figure 11.

Discussion

In this study, we investigated the pharmacological mechanism of action of licorice for the treatment of COVID-19 by molecular docking and molecular dynamics simulation. It was found that the important active chemical components Phaseol in licorice may reduce inflammatory cell activation and inflammatory response by inhibiting the activation of CXCL8 and IL2RA; Glycyrol may act mainly on STAT3 to regulate cell proliferation and survival; And Glyasperin F may regulate cell growth by inhibiting the activation of MMP1, thereby reducing tissue damage and cell death caused by excessive inflammatory responses and promoting the growth of new tissues. Therefore, the active small molecules Phaseol, Glycyrol and Glyasperin F in licorice may act on CXCL8, IL2RA, STAT3, and MMP1 to treat COVID-19 by reducing tissue damage and inflammatory response.

Analysis of bioinformatics results

In this study, Phaseol, Glycyrol and Glyasperin F in licorice may treat COVID-19 to reduce the inflammatory response and promote cell survival by acting on CXCL8, IL2RA, STAT3, and MMP1.

Phaseol may reduce inflammatory cell activation and inflammatory response by inhibiting the activation of CXCL8 and IL2RA. PPI analysis suggested that CXCL8 and IL2RA were closely associated with targets of inflammatory response regulation. GO analysis results suggested that CXCL8 was mainly involved in chemokine activity and interleukin eight receptor binding. KEGG pathway analysis identified IL2RA in pathways such as cellular senescence and MIF-mediated glucocorticoid regulation. The analysis showed that CXCL8 acted as a chemokine that attracts neutrophils, basophils and T cells, but not monocytes. It was also involved

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

System name	ΔE_{vdw}	ΔE_{elec}	ΔG_{GB}	ΔG_{SA}	ΔG_{bind}
Phaseol-CXCL8	−40.63 ± 1.83	−14.59 ± 1.22	21.41 ± 1.27	−5.69 ± 0.10	−39.51 ± 2.06
Phaseol-IL2RA	−28.99 ± 2.58	−10.08 ± 9.51	22.88 ± 6.68	−3.92 ± 0.27	−20.12 ± 3.38
Glyasperin F-MMP1	−39.19 ± 1.25	−12.17 ± 2.77	11.90 ± 2.37	−4.24 ± 0.08	−43.70 ± 1.80
Glycyrol-STAT3	−14.88 ± 1.19	−1.77 ± 2.41	6.56 ± 2.24	−1.75 ± 0.14	−11.85 ± 1.06

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

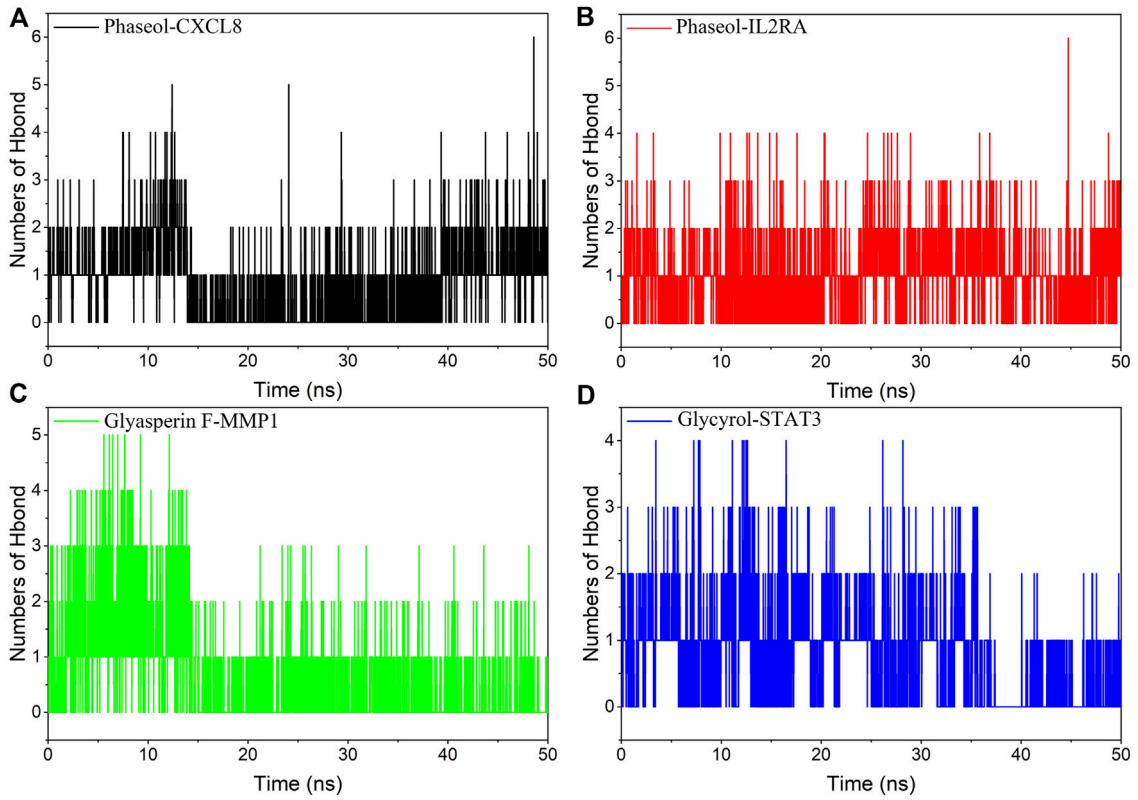


FIGURE 8
Changes in the number of hydrogen bonds between small molecule ligands and protein receptors in complex system simulations. **(A)**CXCL8/Phaseol, **(B)** IL2RA/Phaseol, **(C)** MMP1/Glyasperin F, **(D)**STAT3/Glycyrol.

in neutrophil activation. The results of GO analysis suggested that IL2RA was mainly involved in drug binding and interleukin two binding. KEGG pathway analysis revealed IL2RA in pathways such as immune cell activation and tumor microenvironment regulation. The results suggested that IL2RA was involved in the regulation of immune tolerance by controlling the activity of regulatory T cells (TREG), which could regulate the inflammatory response by suppressing the activation and expansion of self-reactive T cells.

Glycyrol may act mainly on STAT3 to regulate cell proliferation and survival. PPI analysis suggested that STAT3 was closely associated with targets that regulate cell growth and apoptosis. GO analysis results suggested that STAT3 was mainly involved in DNA-binding transcription factor activity and sequence-specific DNA binding. KEGG pathway analysis identified STAT3 in pathways such as cellular senescence. The results of the analysis suggested that IL2RA was involved in the regulation of immune tolerance through the control of regulatory T cell (TREG) activity. TREG can regulate inflammatory responses by suppressing the activation and expansion of self-reactive T cells. The analysis showed that IL6 could participate in cell cycle regulation by regulating the transcriptional activity of STAT3, and STAT3 inhibited cellular autophagy by suppressing EIF2AK2/PKR activity.

Glyasperin F may regulate cell growth by inhibiting the activation of MMP1. PPI analysis suggested that MMP1 was closely associated with targets that regulate protein hydrolysis and processing. GO analysis results suggested that MMP1 was mainly involved in calcium binding and metallopeptidase activity. KEGG pathway analysis identified MMP1 in interleukin six family signaling and other pathways. The analysis showed that MMP1 was mainly involved in extracellular matrix breakdown in normal physiological processes (such as: embryonic development, reproduction and tissue remodeling) as well as in disease processes (such as: arthritis and metastasis).

However, the results of bioinformatics analysis could only predict the potential relationship between the drug and the key target. Therefore, this study further validated the mechanism of action of licorice for COVID-19 treatment using molecular docking and molecular dynamics.

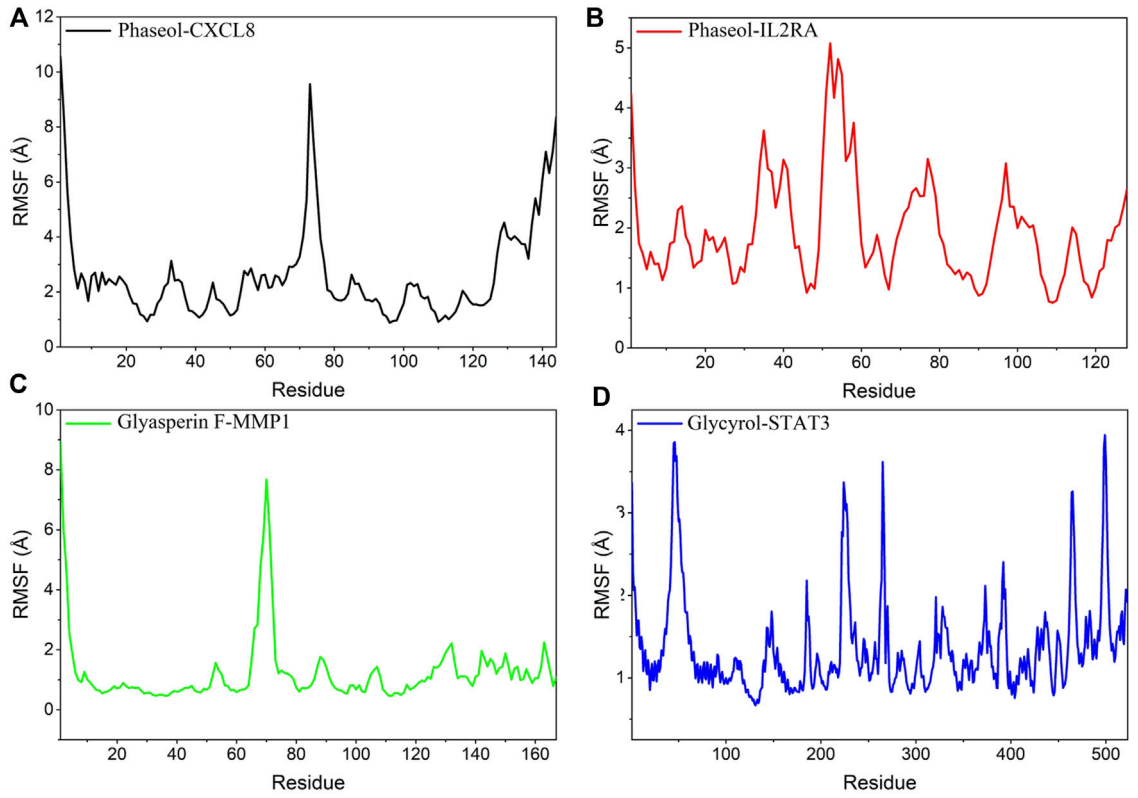


FIGURE 9
Changes in the stability of protein targets at the residue level. **(A)** CXCL8/Phaseol, **(B)** IL2RA/Phaseol, **(C)** MMP1/Glyasperin F, **(D)**STAT3/Glycyrol.

Analysis of molecular docking and molecular dynamics

Molecular docking simulations revealed strong affinity of drug active ingredients (such as: Phaseol, Glycyrol, and Glyasperin F) to protein targets (such as: CXCL8, IL2RA, STAT3, and MMP1). Molecular dynamics results suggested that the drug small molecules and protein complexes could maintain a very stable binding state and thus exert pharmacological effects in the treatment of COVID-19.

Phaseol was able to act stably on CXCL8 and IL2RA, and in particular CXCL8/Phaseol showed strong stability. Molecular docking showed that the binding energy of small molecule Phaseol to CXCL8 and IL2RA reached −8.9 and −8.3, respectively. Based on the trajectory of molecular dynamics simulations, we used the MMGBSA method to calculate the binding energy, which can more accurately reflect the binding mode of small molecules to target proteins. The binding free energy results showed −39.51 ± 2.06 kcal/mol and −20.12 ± 3.38 kcal/mol for CXCL8/Phaseol and IL2RA/Phaseol. In molecular dynamics simulations, the RMSD of both CXCL8/Phaseol and IL2RA/Phaseol gradually converged in the first 10 ns of the simulation and maintained stable fluctuations in subsequent simulations, implying increasing stability of the complex after binding. CXCL8/Phaseol binding results showed that the small molecule Phaseol interacted with E29 on the protein by hydrogen bonding and with V25, V27, V58, and I22 by hydrophobic interaction. The IL2RA/Phaseol binding results showed that the small molecule of drug interacted with Y119, E116, R117, T14, and E9 on the protein by hydrogen bonding, and with Y119, F121, F15, E9, and E116 by hydrophobic interaction, and the pi-pi conjugation occurred between Phaseol and F15.

The binding of Glycyrol to STAT3 was relatively stable and molecular docking showed that the binding energy of the small molecule to NLRP3 was −7.8. The free energy of binding results showed STAT3/Glycyrol to be −11.85 ± 1.06 kcal/mol. The RMSD fluctuations of STAT3/Glycyrol were all within 2 Å, implying a small movement of the STAT3/Glycyrol system. The STAT3/Glycyrol binding results indicated that the small molecule Glycyrol interacted with S611, E612, and S613 on the protein by hydrogen bonding, with P629 and S613 by hydrophobic interaction, and also with R609 by cation pi conjugation.

Glyasperin F bound to MMP1 could form a very stable complex. Molecular docking showed that the binding energy

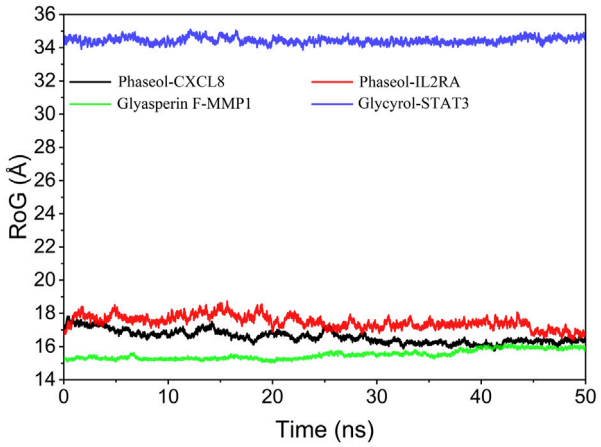


FIGURE 10
Analysis of protein folding state and overall conformation.

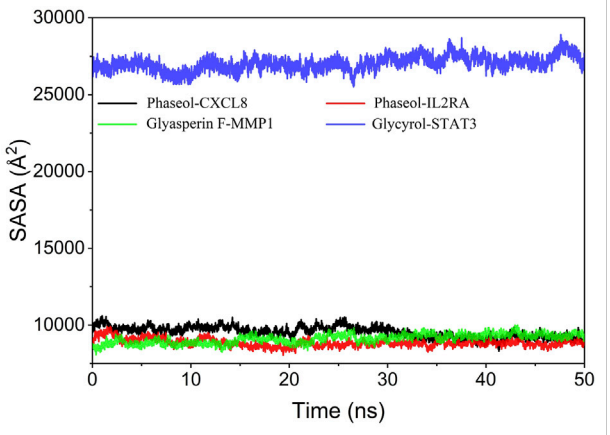


FIGURE 11
Analysis of Solvent Accessible Surface Area (SASA).

of small molecule Glyasperin F to MMP1 reached −9.8. The binding free energy results showed that MMP1/Glyasperin F was −43.70 ± 1.80 kcal/mol. The hydrogen bonding of MMP1/Glyasperin F was sparse at the late stage of molecular dynamics simulation, implying that hydrogen bonding was not the main force for its stability maintenance. The results of MMP1/Glyasperin F binding indicated that the small molecule interacted with A84 on the protein by hydrogen bonding and with H83, V115, L81, Y140, and H118 by hydrophobic interaction.

This study not only analyzed the relevant bioinformatics findings, but also used a supercomputer platform to simulate the microscopic evolution of complex systems of small molecule drugs and proteins through molecular dynamics. The computer simulations visualized the binding states of CXCL8/Phaseol, IL2RA/Phaseol, STAT3/Glycyrol and MMP1/Glyasperin F. The results of molecular dynamics simulations showed that the simulated binding of the four complexes could remain relatively stable.

Therefore, the results of this study can further explain the mechanism of action of active small molecules of licorice for the treatment of COVID-19 and related signaling pathways.

Phaseol may reduce inflammatory cell activation and inflammatory response through CXCL8 and IL2RA

Phaseol is the active component derived from licorice. Phaseol was found to be closely associated with the IL6-STAT3 signaling pathway (Lu et al., 2020), and Phaseol could alleviate the inflammatory effects in lipopolysaccharide (LPS)-induced RAW264.7 cells (Li et al., 2017).

CXCL8 (also known as CXCL8) belongs to the elastin-like recombinant (ELR) CXC chemokine family (Liu et al., 2016). CXCL8 can be secreted by different cell types, including blood monocytes, alveolar macrophages, fibroblasts, endothelial cells, and epithelial cells (Ha et al., 2017). CXCL8 acts as a chemokine by directing neutrophils to the site of infection. Moreover, CXCL8 is also involved in pro-inflammatory signaling cascades along with other cytokines and plays a role in systemic inflammatory response syndrome (SIRS).

CXCL8 is a highly selective pro-inflammatory chemokine, and local and systemic elevations of CXCL8 have been found in various inflammatory diseases as well as in SIRS and sepsis (Haas et al., 2016). CXCL8 is barely detectable in the physiological state, but can be stimulated by pro-inflammatory cytokines such as tumor necrosis factor a (TNFa) and interleukin-1b (IL-1b) and mediated by the transcription factors NF-κB and activator protein-1 (AP-1), which can lead to a 10 to 100 fold upregulation of CXCL8 expression. The function of CXCL8 is mainly dependent on its interaction with specific cell surface G protein-coupled receptors (GPCR), CXCR1 and CXCR2 (Liu et al., 2016; Ha et al., 2017). CXCL8 contributes to the pathology of angiogenesis, fibrosis, infection, atherosclerosis, and tumor growth. Clinical studies have shown that elevated plasma levels of CXCL8 and other ELR-CXC chemokines can occur with acute indications such as arthritis, chronic obstructive pulmonary disease (COPD), asthma, cystic fibrosis, atherosclerosis, inflammatory bowel disease (IBD), psoriasis, and cancer, as well as acute indications such as reperfusion injury and acute respiratory distress syndrome (ARDS) (Cheng et al., 2017). Leukocyte recruitment is critical in many acute and chronic inflammatory diseases. Chemokines are key mediators of leukocyte recruitment during the inflammatory response, and

the chemokine interleukin-8/CXCL8 is a classic neutrophil chemoattractant (Martínez-Burgo et al., 2019). CXCL8 inhibits the chemotactic response of neutrophils and suppresses the neutrophil-induced inflammatory response (Zhou et al., 2019). And CXCL8 promotes the activation and recruitment of macrophages and monocytes, which is a prerequisite for the shift from acute to chronic inflammation (Mohr et al., 2017). CXCL8 has been reported to recruit leukocytes from the blood into tissues during inflammation, and in turn, inflammation worsened by activated leukocytes can increase CXCL8 levels (Zhou et al., 2019). And it has been shown that monoammonium glycyrrhizinate (MAG) of licorice has anti-inflammatory properties. Mag inhibited the mRNA expression of TNF-α-induced chemokines (including CXCL8, CX3CL1, and CXCL16) in human dermal microvascular endothelial cell line (HMEC-1) cells in a dose-dependent manner and reduced the secretion of these chemokines (Cao et al., 2014).

IL2RA (also known as CD25) is a core component of the trimeric IL-2 receptor complex and plays a key role in mediating interleukin two immunomodulatory functions (Borysewicz-Sańczyk et al., 2020). IL2RA is a membrane protein that is involved in the regulation of immune tolerance by controlling the activity of regulatory T cells (TREG). Interleukin 2 (IL2) is a lymphocyte growth factor that plays an important role in the regulation of immune homeostasis as an essential self-tolerance regulator. It was found that cellular responsiveness to IL-2 directly depends on cellular expression of IL2RA, that IL-2 signaling increases with increased IL2RA expression, and that IL2RA directly affects binding stability in the IL-2/IL-2R complex (Buhelt et al., 2019). Plasma IL2RA levels were also found to be significantly elevated in COVID-19 patients (Galván-Peña et al., 2021; Sayah et al., 2021).

IL2RA is the receptor subunit that increases the affinity of the receptor for IL2 cytokines (Akman et al., 2021). Expression of IL2RA has been described at high levels on the surface of regulatory T cells (Tregs), a population of T cells with the ability to suppress self-reactive T cells. Further studies have shown that IL2RA plays a crucial role in sensitizing T cells to induce cell death (Borysewicz-Sańczyk et al., 2020). Changes in IL2RA expression may affect immune and inflammatory signaling cascade responses, which in turn affect $CD4^+$ T cell differentiation and TReg cell suppressive activity (Asouri et al., 2020). IL2 signaling is involved in the differentiation and homeostasis of regulatory T cells (Tregs), and IL2 signaling is involved in the induction of cell growth and effector T cell proliferation (Zeebroeck et al., 2021). Pre-activation of IL-12, IL-15, and IL-18 was shown to upregulate IL2RA (CD25) expression (Akman et al., 2021). Further studies have shown that IL2RA plays a crucial role in sensitizing T cells to induce cell death.

Therefore, we suggested that Phaseol may reduce inflammatory cell activation and inflammatory response by acting on CXCL8 and IL2RA, thereby reducing tissue damage from excessive inflammatory response and alleviating the clinical symptoms of COVID-19.

Glycyrol may affect cell proliferation and survival by regulating STAT3

Glycyrol exhibits a variety of biological effects, including antioxidant and anti-inflammatory effects and modulation of intrinsic immunity (Shin et al., 2011; Fu et al., 2014; Kim et al., 2020). It has been shown that Glycyrol-induced cell death is associated with apoptosis and autophagy, Glycyrol can bind to TOPK proteins and inhibit their kinase activity, leading to the activation of apoptotic signalling pathways (Xu and Kim, 2014; Lu et al., 2019).

STAT3 is a component of the acute phase response factor (APRF) complex activated by interleukin-6 (IL-6), and STAT3 plays a key role in many cellular processes such as cell growth and apoptosis by mediating the expression of multiple cellular stimuli (Hillmer et al., 2016). STAT3 is involved in regulating biological processes such as cell growth, differentiation and survival, inflammation and hematopoiesis (Gao et al., 2018; Liu et al., 2021; Zhao et al., 2021).

STAT3 is a latent transcription factor that mediates extracellular signals, such as cytokines and growth factors, by interacting with peptide receptors on the cell surface. STAT3 protein is transcriptionally activated mainly through tyrosine phosphorylation. Activated STAT3 dimers translocate to the nucleus and bind to sequence-specific DNA elements, thereby transcribing target genes (You et al., 2015). Recent studies have shown that STAT3 protein is expressed in CD4 T cells, T helper Th17 cells, Th1 and Th2 cells and that STAT3α isoforms may interact with proteins such as Probanin one to regulate pathological immune responses. The IL-6/JAK/STAT3 pathway is a major signaling pathway involved in regulating the inflammatory response in disease pathogenesis. JAK/STAT3 signaling promotes inflammation by regulating the development of innate lymphocytes in the immune response (Kang et al., 2021). STAT3 plays a central role in JAK/STAT signaling (You et al., 2015). IL-6 is the main stimulator of STAT3 *in vivo*, especially during inflammatory outbreaks. IL-6 signaling acts primarily through the JAK/STAT pathway, mainly through STAT3. Both of these factors can form IL-6 amplifiers that produce a cascade of amplifying effects associated with inflammation. This effect promotes various pro-inflammatory cytokines and chemokines, including IL-6, and recruits macrophages and lymphocytes, thereby enhancing the positive feedback loop formed by IL-6 and STAT3 (Luo et al., 2022). Licorice was found to reduce IL-6 levels, which is the main stimulator of STAT3 *in vivo*, especially during inflammatory outbreaks (Richard, 2021). Inhibition of STAT3 activity improved the pulmonary inflammatory response in LPS-induced acute lung injury (ALI) (Xu et al., 2020). And one

study found a clinical therapeutic effect on lung inflammation by inhibiting STAT3 pathway (Zhao et al., 2016).

Therefore, we suggested that Glycyrol may act on STAT3 to regulate cell proliferation and survival, thereby reducing cell death due to inflammatory stimuli and promoting the growth of new tissue.

Glyasperin F may regulate cell growth by affecting the activation of MMP1

Glyasperin F is an isoflavone compound, studies have found that Glyasperin F can inhibit the proliferation of lung cancer cells (Ngnintedo et al., 2016; Kuete et al., 2018).

MMP-1 is one of the most abundant enzymes in the family of matrix metalloproteinases (MMPs), which are mesenchymal collagenases secreted by a variety of cells including fibroblasts, endothelial and inflammatory cells (Gopal et al., 2016; Erdem et al., 2020). MMP-1 is capable of degrading type I, II, and III collagen, which plays a key role in extracellular matrix (ECM) remodeling in normal development and pathology (Affara et al., 2011).

MMP1 can be activated by several pro-inflammatory cytokines and growth factors and its expression is increased in alveolar epithelial cells during pulmonary fibrosis, and it inhibits mitochondrial respiration and oxidative stress, while promoting cell proliferation and migration (Lee et al., 2013). Various inflammatory factors (including CXCL8, IL-1β, and TNF-α) have been reported to contribute to the expression of MMP1 (Chen et al., 2019). MMP1 plays a clinically important role in inflammatory diseases and has been associated with many pathological processes, including wound healing, tumor metastasis and arthritis (Affara et al., 2011). Several reports suggest that MMP1 is indeed upregulated in patients suffering from diseases such as COPD and lung cancer (Carver et al., 2015). MMP-1 has been widely reported to lyse the extracellular matrix (ECM) and to promote angiogenesis. MMP1 was found to induce expression of vascular endothelial growth factor receptor 2 (VEGFR2) and endothelial cell proliferation, stimulate the serine/threonine protein kinase MARK2 and activate the transcription factor NF-κB for vascular remodeling and angiogenesis (Ng et al., 2022).

Many studies have found that licorice inhibited the high expression of matrix metalloproteinase-1 (MMP-1) and -3 (MMP-3) and down-regulated the expression of inflammatory cytokines such as IL-6, TNF-α, and IL-10. These findings strongly suggest that licorice regulates the abnormal expression of MMP-1 and MMP-3 mainly through its antioxidant and anti-inflammatory properties as well as (Kong et al., 2015; Gopal et al., 2016).

Therefore, we proposed that Glyasperin F may regulate cell growth by affecting the activation of MMP1, thereby promoting recovery of injured tissues.

The mechanisms analysis of licorice in the treatment of corona virus disease 2019

The summary of the mechanisms analysis of licorice in the treatment of COVID-19 is shown in Graphical Abstract.

Conclusion

This study explored the pharmacological mechanism of licorice for the treatment of COVID-19 by molecular docking and molecular dynamics simulations. We found that Phaseol in licorice may reduce inflammatory cell activation and inflammatory response by inhibiting the activation of CXCL8 and IL2RA; Glycyrol may regulate cell proliferation and survival by acting on STAT3. And Glyasperin F may regulate cell growth by inhibiting the activation of MMP1, thus reducing tissue damage and cell death caused by excessive inflammatory response and promoting the growth of new tissues.

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

Author contributions: J-FC, YG, MW, LX, XY, and XZ, contributed to the conception of the study; J-FC, MW, YG, XY, SC, ZX, and YL contributed significantly to analysis and manuscript preparation; J-FC, YG, MW, WZ, LZ, and SC performed the data analyses and wrote the manuscript; XZ, J-FC, WZ, and LZ helped perform the analysis with constructive discussions.

Conflict of interest

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

Publisher's note

References

Abraham, J., and Florentine, S. (2021). Licorice (Glycyrrhiza glabra) extracts-suitable pharmacological interventions for COVID-19? A review. *Plants (Basel)* 10, 2600. doi:10.3390/plants10122600

Affara, M., Dunmore, B. J., Sanders, D. A., Johnson, N., Print, C. G., and Charnock-Jones, D. S. (2011). MMP1 bimodal expression and differential response to inflammatory mediators is linked to promoter polymorphisms. *BMC Genomics* 12, 43. doi:10.1186/1471-2164-12-43

Akman, B., Hu, X., Liu, X., HatipoğluHua, Y., Chen, Y., Chan, W. C., et al. (2021). PRDM1 decreases sensitivity of human NK cells to IL2-induced cell expansion by directly repressing CD25 (IL2RA). *J. Leukoc. Biol.* 109, 901–914. doi:10.1002/JLB.2A0520-321RR

Al-Shar'i, N. A., and Al-Balas, Q. A. (2019). Molecular dynamics simulations of adenosine receptors: Advances, applications and trends. *Curr. Pharm. Des.* 25, 783–816. doi:10.2174/1381612825666190304123414

Asouri, M., Rokni, H. A., Sahraian, M. A., Fattahi, S., Motamed, N., Doosti, R., et al. (2020). Analysis of single nucleotide polymorphisms in HLA-DRA, IL2RA, and HMGB1 genes in multiple sclerosis. *Rep. Biochem. Mol. Biol.* 9, 198–208. doi:10.29252/rbmb.9.2.199

Borysewicz-Sańczyk, H., Sawicka, B., Wawrusiewicz-Kurylonek, N., Głowińska-Olszewska, B., Kadłubiska, A., Gościk, J., et al. (2020). Genetic association study of IL2RA, IFIH1, and CTLA-4 polymorphisms with autoimmune thyroid diseases and type 1 diabetes. *Front. Pediatr.* 8, 481. doi:10.3389/fped.2020.00481

Buhelt, S., Søndergaard, H. B., Oturai, A., Ullum, H., Essen, M. R., and Sellebjerg, F. (2019). Relationship between multiple sclerosis-associated IL2RA risk allele variants and circulating T cell phenotypes in healthy genotype-selected controls *Cells*, 8. Cells, E634. doi:10.3390/cells8060634

Burley, S. K., Berman, H. M., Kleywegt, G. J., Markley, J. L., Nakamura, H., and Velankar, S. (2017). Protein data bank (PDB): The single global macromolecular structure archive. *Methods Mol. Biol.* 1607, 627–641. doi:10.1007/978-1-4939-7000-1_26

Cao, J., Li, L., Xiong, L., Wang, C., Chen, Y., and Zhang, X. (2022). Research on the mechanism of berberine in the treatment of COVID-19 pneumonia pulmonary fibrosis using network pharmacology and molecular docking. *Phytomed. Plus.* 2, 100252. doi:10.1016/j.phyplu.2022.100252

Cao, N., Chen, T., Guo, Z., Qin, S., and Li, M. (2014). Monoammonium glycyrrhizate suppresses tumor necrosis factor-α induced chemokine production in HMEC-1 cells, possibly by blocking the translocation of nuclear factor-κB into the nucleus. *Can. J. Physiol. Pharmacol.* 92, 859–865. doi:10.1139/cjpp-2014-0022

Carver, P. I., Anguiano, V., D'Armiento, J. M., and Shiomi, T. (2015). Mmp1a and Mmp1b are not functional orthologs to human MMP1 in cigarette smoke induced lung disease. *Exp. Toxicol. Pathol.* 67, 153–159. doi:10.1016/j.etp.2014.11.004

Chen, Y., Peng, S., Cen, H., Lin, Y., Huang, C., Chen, Y., et al. (2019). MicroRNA hsa-miR-623 directly suppresses MMP1 and attenuates IL-8-induced metastasis in pancreatic cancer. *Int. J. Oncol.* 55, 142–156. doi:10.3892/ijo.2019.4803

Chen, Y., Zheng, Y., Fong, P., Mao, S., and Wang, Q. (2020). The application of the MM/GBSA method in the binding pose prediction of FGFR inhibitors. *Phys. Chem. Chem. Phys.* 22, 9656–9663. doi:10.1039/d0cp00831a

Cheng, H., Yu, H., Gordon, J. R., Li, F., and Cheng, J. (2017). Effects of K11R and G31P mutations on the structure and biological activities of CXCL8: Solution structure of human CXCL8 $_{(3\text{-}72)}$ K11r/G31P. *Molecules* 22, 1229. doi:10.3390/molecules22071229

Collier, T. A., Piggot, T. J., and Allison, J. R. (2020). Molecular dynamics simulation of proteins. *Methods Mol. Biol.* 2073, 311–327. doi:10.1007/978-1-4939-9869-2_17

Erdem, J. s., Arnoldussen, Y. J., Tajik, S., Ellingsen, D. G., and Zienolddiny, S. (2020). Effects of mild steel welding fume particles on pulmonary epithelial inflammation and endothelial activation. *Toxicol. Ind. Health* 36, 995–1001. doi:10.1177/0748233720962685

Fernandes, Q., Inchakalody, V. P., Merhi, M., Mestiri, S., Taib, N., Abo El-Ella, D. M., et al. (2022). Emerging COVID-19 variants and their impact on SARS-CoV-2 diagnosis, therapeutics and vaccines. *Ann. Med.* 54, 524–540. doi:10.1080/07853890.2022.2031274

Filipe, H. A. L., and Loura, L. M. S. (2022). *Molecular Dynamics Simulations: Advances and Applications*, 27.*Molecules*

Fu, Y., Zhou, H., Wang, S., and Wei, Q. (2014). Glycyrol suppresses collagen-induced arthritis by regulating autoimmune and inflammatory responses. *PLoS One* 9 (7), e98137. doi:10.1371/journal.pone.0098137

Galván-Peña, S., Leon, J., Chowdhary, K., Michelson, D. A., Vijaykumar, B., Yang, L., et al. (2021). Profound Treg perturbations correlate with COVID-19 severity. *Proc. Natl. Acad. Sci. U. S. A.* 118, e2111315118. doi:10.1073/pnas.2111315118

Gao, Y., Zhao, H., Wang, P., Wang, J., and Zou, L. (2018). The roles of SOCS3 and STAT3 in bacterial infection and inflammatory diseases. *Scand. J. Immunol.* 88, e12727. doi:10.1111/sji.12727

Gomaa, A. A., and Abdel-Wadood, Y. A. (2021). The potential of glycyrrhizin and licorice extract in combating COVID-19 and associated conditions. *Phytomed. Plus.* 1, 100043. doi:10.1016/j.phyplu.2021.100043

Gopal, S. K., Greening, D. W., Zhu, H., Simpson, R. J., and Mathias, R. A. (2016). Transformed MDCK cells secrete elevated MMP1 that generates LAMA5 fragments promoting endothelial cell angiogenesis. *Sci. Rep.* 6, 28321. doi:10.1038/srep28321

Ha, H., Debnath, B., and Neamati, N. (2017). Role of the CXCL8-CXCR1/2 Axis in cancer and inflammatory diseases. *Theranostics* 7, 1543–1588. doi:10.7150/thno.15625

Haas, M., Kaup, F. J., and Neumann, S. (2016). Canine pyometra: A model for the analysis of serum CXCL8 in inflammation. *J. Vet. Med. Sci.* 78, 375–381. doi:10.1292/jvms.15-0415

Hillmer, E. J., Zhang, H., Li, H. S., and Watowich, S. S. (2016). STAT3 signaling in immunity. *Cytokine Growth Factor Rev.* 31, 1–15. doi:10.1016/j.cytogfr.2016.05.001

Kang, D., Sp, N., En, C., Rugamba, A., Jing, X., Jing, R., et al. (2021). Non-toxic sulfur inhibits LPS-induced inflammation by regulating TLR-4 and JAK2/STAT3 through IL-6 signaling. *Mol. Med. Rep.* 24, 485. doi:10.3892/mmr.2021.12124

Kim, Y., Shrestha, R., Kim, S., Kim, J. A., Lee, J., Jeong, T. C., et al. (2020). *In vitro* characterization of Glycyrol metabolites in human liver microsomes using HR-resolution MS spectrometer coupled with tandem mass spectrometry. *Xenobiotica.* 50 (4), 380–388. doi:10.1080/00498254.2019.1636418

Kong, S., Chen, H., Yu, X., Zhang, X., Feng, X., Kang, X., et al. (2015). The protective effect of 18β-Glycyrrhetinic acid against UV irradiation induced photoaging in mice. *Exp. Gerontol.* 61, 147–155. doi:10.1016/j.exger.2014.12.008

Kuete, V., Ngnintedo, D., Fotso, G. W., Karaosmanoğlu, O., Ngadjui, B. T., Keumedjio, F., et al. (2018). Cytotoxicity of seputhecarpan D, thonningiol and 12 other phytochemicals from African flora towards human carcinoma cells. *BMC Complement. Altern. Med.* 18 (1), 36. doi:10.1186/s12906-018-2109-9

Larini, L., Mannella, R., and Leporini, D. (2007). Langevin stabilization of molecular-dynamics simulations of polymers by means of quasisymplectic algorithms. *J. Chem. Phys.* 126, 104101. doi:10.1063/1.2464095

Lee, K. M., Shim, H., Lee, G. S., Park, I. H., Lee, O. S., Lim, S. C., et al. (2013). Chitin from the extract of cuttlebone induces acute inflammation and enhances MMP1 expression. *Biomol. Ther.* 21, 246–250. doi:10.4062/biomolther.2013.036

Lee, T. S., Allen, B. K., Giese, T. J., Guo, Z., Li, P., Lin, C., et al. (2020). Alchemical binding free energy calculations in AMBER20: Advances and best practices for drug discovery. *J. Chem. Inf. Model.* 60, 5595–5623. doi:10.1021/acs.jcim.0c00613

Li, H., Yoon, J. H., Won, H. J., Ji, H. S., Yuk, H. J., Park, K. H., et al. (2017). Isotrifoliol inhibits pro-inflammatory mediators by suppression of TLR/NF-κB and TLR/MAPK signaling in LPS-induced RAW264.7 cells. *Int. Immunopharmacol.* 45, 110–119. doi:10.1016/j.intimp.2017.01.033

Li, J., Xu, D., Wang, L., Zhang, M., Zhang, G., Li, E., et al. (2021). Glycyrrhizic acid inhibits SARS-CoV-2 infection by blocking spike protein-mediated cell attachment. *Molecules* 26, 6090. doi:10.3390/molecules26206090

Liu, Q., Li, A., Tian, Y., Wu, D., Liu, Y., Li, T., et al. (2016). The CXCL8-CXCR1/2 pathways in cancer. *Cytokine Growth Factor Rev.* 31, 61–71. doi:10.1016/j.cytogfr.2016.08.002

Liu, Y., Liao, S., Bennett, S., Tang, H., Song, D., Wood, D., et al. (2021). STAT3 and its targeting inhibitors in osteosarcoma. *Cell Prolif.* 54, e12974. doi:10.1111/cpr.12974

Lu, S., Ye, L., Yin, S., Zhao, C., Yan, M., Liu, X., et al. (2019). Glycyrol exerts potent therapeutic effect on lung cancer via directly inactivating T-LAK cell-originated protein kinase. *Pharmacol. Res.* 147, 104366. doi:10.1016/j.phrs.2019.104366

Lu, X., Wu, X., Jing, L., Tao, L., Zhang, Y., Huang, R., et al. (2020). Network pharmacology analysis and experiments validation of the inhibitory effect of JianPi Fu recipe on colorectal cancer LoVo cells metastasis and growth. *Evid. Based. Complement. Altern. Med.* 2020, 4517483. doi:10.1155/2020/4517483

Luo, W., Ding, R., Guo, X., Zhan, T., Tang, T., Fan, R., et al. (2022). Clinical data mining reveals Gancao-Banxia as a potential herbal pair against moderate COVID-19 by dual binding to IL-6/STAT3. *Comput. Biol. Med.* 145, 105457. doi:10.1016/j.compbiomed.2022.105457

Maier, J. A., Martinez, C., Kasavajhala, K., Wickstrom, L., Hauser, K. E., and Simmerling, C. (2015). ff14SB: Improving the accuracy of protein side chain and backbone parameters from ff99SB. *J. Chem. Theory Comput.* 11, 3696–3713. doi:10.1021/acs.jctc.5b00255

Majumder, J., and Minko, T. (2021). Recent developments on therapeutic and diagnostic approaches for COVID-19. *Aaps J.* 23, 14. doi:10.1208/s12248-020-00532-2

Martínez-Burgo, B., Cobb, S. L., Pohl, E., Kashanin, D., Paul, T., Kirby, J. A., et al. (2019). A C-terminal CXCL8 peptide based on chemokine-glycosaminoglycan interactions reduces neutrophil adhesion and migration during inflammation. *Immunology* 157, 173–184. doi:10.1111/imm.13063

Mithun, R., Ismail, C., Johra, K., Mohammad, A. A., Mohammad, N. M., Rohitash, Y., et al. (2022). Identification of bioactive molecules from Triphala (Ayurvedic herbal formulation) as potential inhibitors of SARS-CoV-2 main protease (Mpro) through computational investigations. *J. King Saud Univ. Sci.* 34 (3), 101826. doi:10.1016/j.jksus.2022.101826

Mohr, T., Haudek-Prinz, V., Slany, A., Grillari, J., Micksche, M., and Gerner, C. (2017). Proteome profiling in IL-1β and VEGF-activated human umbilical vein endothelial cells delineates the interlink between inflammation and angiogenesis. *PLoS One* 12, e0179065. doi:10.1371/journal.pone.0179065

Nam, K. H. (2021). Room-temperature structure of xylitol-bound glucose isomerase by serial crystallography: Xylitol binding in the M1 site induces release of metal bound in the M2 site. *Int. J. Mol. Sci.* 22, 3892. doi:10.3390/ijms22083892

Ng, L., Wong, S. K., Huang, Z., Lam, C. S., Chow, A. K., Foo, D. C., et al. (2022). CD26 induces colorectal cancer angiogenesis and metastasis through CAV1/MMP1 signaling. *Int. J. Mol. Sci.* 23, 1181. doi:10.3390/ijms23031181

Ng, S. L., Khaw, K. Y., Ong, Y. S., Goh, H. P., Kifli, N., Teh, S. P., et al. (2021). Licorice: A potential herb in overcoming SARS-CoV-2 infections. *J. Evid. Based. Integr. Med.* 26, 2515690x21996662. doi:10.1177/2515690X21996662

Ngnintedo, D., Fotso, G. W., Kuete, V., Nana, F., Sandjo, L. P., Karaosmanoğlu, O., et al. (2016). Two new pterocarpans and a new pyrone derivative with cytotoxic activities from ptycholobium contortum (N.E.Br.) brummitt (leguminosae): Revised NMR assignment of mundulea lactone. *Chem. Cent. J.* 10, 58. doi:10.1186/s13065-016-0204-x

Ochani, R., Asad, A., Yasmin, F., Shaikh, S., Khalid, H., Batra, S., et al. (2021). COVID-19 pandemic: From origins to outcomes. A comprehensive review of viral pathogenesis, clinical manifestations, diagnostic evaluation, and management. *Infez. Med.* 29, 20–36.

Pan, B., Fang, S., Zhang, J., Pan, Y., Liu, H., Wang, Y., et al. (2020). Chinese herbal compounds against SARS-CoV-2: Puerarin and quercetin impair the binding of viral S-protein to ACE2 receptor. *Comput. Struct. Biotechnol. J.* 18, 3518–3527. doi:10.1016/j.csbj.2020.11.010

Rai, P., Kumar, B. K., Deekshit, V. K., Karunasagar, I., and Karunasagar, I. (2021). Detection technologies and recent developments in the diagnosis of COVID-19 infection. *Appl. Microbiol. Biotechnol.* 105, 441–455. doi:10.1007/s00253-020-11061-5

Ravindranath, P. A., Forli, S., Goodsell, D. S., Olson, A. J., and Sanner, M. F. (2015). AutoDockFR: Advances in protein-ligand docking with explicitly specified binding site flexibility. *PLoS Comput. Biol.* 11, e1004586. doi:10.1371/journal.pcbi.1004586

Richard, S. A. (2021). Exploring the pivotal immunomodulatory and anti-inflammatory potentials of glycyrrhizic and glycyrrhetinic acids. *Mediat. Inflamm.* 2021, 6699560. doi:10.1155/2021/6699560

Sayah, W., Berkane, I., Guermache, I., Sabri, M., Lakhal, F. Z., Rahali, S. Y., et al. (2021). Interleukin-6, procalcitonin and neutrophil-to-lymphocyte ratio: Potential immune-inflammatory parameters to identify severe and fatal forms of COVID-19. *Cytokine* 141, 155428. doi:10.1016/j.cyto.2021.155428

Shin, E. M., Kim, S., Merfort, I., and Kim, Y. S. (2011). Glycyrol induces apoptosis in human Jurkat T cell lymphocytes via the Fas-FasL/caspase-8 pathway. *Planta Med.* 77 (3), 242–247. doi:10.1055/s-0030-1250260

Xie, R., Lin, Z., Zhong, C., Li, S., Chen, B., Wu, Y., et al. (2021). Deciphering the potential anti-COVID-19 active ingredients in Andrographis paniculata (Burm. F.) Nees by combination of network pharmacology, molecular docking, and molecular dynamics. *RSC Adv.* 11, 36511–36517. doi:10.1039/d1ra06487h

Xiong, H., Dong, Z., Lou, G., Gan, Q., Wang, J., and Huang, Q. (2020). Analysis of the mechanism of Shufeng Jiedu capsule prevention and treatment for COVID-19 by network pharmacology tools. *Eur. J. Integr. Med.* 40, 101241. doi:10.1016/j.eujim.2020.101241

Xu, M. Y., and Kim, Y. S. (2014). Antitumor activity of Glycyrol via induction of cell cycle arrest, apoptosis and defective autophagy. *Food Chem. Toxicol.* 74, 311–319. doi:10.1016/j.fct.2014.10.023

Xu, S., Pan, X., Mao, L., Pan, H., Xu, W., Hu, Y., et al. (2020). Phospho-Tyr705 of STAT3 is a therapeutic target for sepsis through regulating inflammation and coagulation. *Cell Commun. Signal.* 18, 104. doi:10.1186/s12964-020-00603-z

Yang, R., Wang, L., Yuan, B., and Liu, Y. (2015). The pharmacological activities of licorice. *Planta Med.* 81, 1654–1669. doi:10.1055/s-0035-1557893

Yi, Y., Li, J., Lai, X., Zhang, M., Kuang, Y., Bao, Y. O., et al. (2022). Natural triterpenoids from licorice potently inhibit SARS-CoV-2 infection. *J. Adv. Res.* 36, 201–210. doi:10.1016/j.jare.2021.11.012

You, L., Wang, Z., Li, H., Shou, J., Jing, Z., Xie, J., et al. (2015). The role of STAT3 in autophagy. *Autophagy* 11, 729–739. doi:10.1080/15548627.2015.1017192

Zeebroeck, L. V., Hornero, R. A., Côrte-Real, B. F., Hamad, I., Meissner, T. B., and Kleinewietfeld, M. (2021). Fast and efficient genome editing of human FOXP3(+) regulatory T cells. *Front. Immunol.* 12, 655122. doi:10.3389/fimmu.2021.655122

Zhang, Q. H., Huang, H. Z., Qiu, M., Wu, Z. F., Xin, Z. C., Cai, X. F., et al. (2021). Traditional uses, pharmacological effects, and molecular mechanisms of licorice in potential therapy of COVID-19. *Front. Pharmacol.* 12, 719758. doi:10.3389/fphar.2021.719758

Zhao, J., Liu, X., Chen, Y., Zhang, L. S., Zhang, Y. R., Ji, D. R., et al. (2021). STAT3 promotes schistosome-induced liver injury by inflammation, oxidative stress, proliferation, and apoptosis signal pathway. *Infect. Immun.* 89, e00309-20. doi:10.1128/IAI.00309-20

Zhao, J., Yu, H., Liu, Y., Gibson, S. A., Yan, Z., Xu, X., et al. (2016). Protective effect of suppressing STAT3 activity in LPS-induced acute lung injury. *Am. J. Physiol. Lung Cell. Mol. Physiol.* 311, L868–L880. doi:10.1152/ajplung.00281.2016

Zhou, Y., Xu, W., and Zhu, H. (2019). CXCL8((3-72)) K11R/G31P protects against sepsis-induced acute kidney injury via NF-κB and JAK2/STAT3 pathway. *Biol. Res.* 52, 29. doi:10.1186/s40659-019-0236-5

frontiers | Frontiers in Pharmacology

TYPE Original Research
PUBLISHED 23 September 2022
DOI 10.3389/fphar.2022.1003480

OPEN ACCESS

EDITED BY
Mithun Rudrapal,
Rasiklal M. Dhariwal Institute of Pharmaceutical Education and Research, India

REVIEWED BY
Rakesh K. Sindhu,
Chitkara University, India
Lalit Mohan Nainwal,
GD Goenka University, India

*CORRESPONDENCE
Tero Aittokallio,
tero.aittokallio@fimm.fi

SPECIALTY SECTION
This article was submitted to Drugs Outcomes Research and Policies, a section of the journal Frontiers in Pharmacology

RECEIVED 26 July 2022
ACCEPTED 15 August 2022
PUBLISHED 23 September 2022

CITATION
Wang T, Pulkkinen OI and Aittokallio T (2022), Target-specific compound selectivity for multi-target drug discovery and repurposing. *Front. Pharmacol.* 13:1003480. doi: 10.3389/fphar.2022.1003480

Target-specific compound selectivity for multi-target drug discovery and repurposing

Tianduanyi Wang[1,2], Otto I. Pulkkinen[1,3,4] and Tero Aittokallio[1,3,4,5,6]*

[1]Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Helsinki, Finland, [2]Department of Computer Science, Aalto University, Espoo, Finland, [3]Helsinki Institute for Information Technology (HIIT), Department of Computer Science, University of Helsinki, Helsinki, Finland, [4]Department of Mathematics and Statistics and InFLAMES Research Flagship, University of Turku, Turku, Finland, [5]Institute for Cancer Research, Department of Cancer Genetics, Oslo University Hospital, Oslo, Norway, [6]Oslo Centre for Biostatistics and Epidemiology (OCBE), Faculty of Medicine, University of Oslo, Oslo, Norway

Most drug molecules modulate multiple target proteins, leading either to therapeutic effects or unwanted side effects. Such target promiscuity partly contributes to high attrition rates and leads to wasted costs and time in the current drug discovery process, and makes the assessment of compound selectivity an important factor in drug development and repurposing efforts. Traditionally, selectivity of a compound is characterized in terms of its target activity profile (wide or narrow), which can be quantified using various statistical and information theoretic metrics. Even though the existing selectivity metrics are widely used for characterizing the overall selectivity of a compound, they fall short in quantifying how selective the compound is against a particular target protein (e.g., disease target of interest). We therefore extended the concept of compound selectivity towards target-specific selectivity, defined as the potency of a compound to bind to the particular protein in comparison to the other potential targets. We decompose the target-specific selectivity into two components: 1) the compound's potency against the target of interest (absolute potency), and 2) the compound's potency against the other targets (relative potency). The maximally selective compound-target pairs are then identified as a solution of a bi-objective optimization problem that simultaneously optimizes these two potency metrics. In computational experiments carried out using large-scale kinase inhibitor dataset, which represents a wide range of polypharmacological activities, we show how the optimization-based selectivity scoring offers a systematic approach to finding both potent and selective compounds against given kinase targets. Compared to the existing selectivity metrics, we show how the target-specific selectivity provides additional insights into the target selectivity and promiscuity of multi-targeting kinase inhibitors. Even though the selectivity score is shown to be relatively robust against both missing bioactivity values and the dataset size, we further developed a permutation-based procedure to calculate empirical *p*-values to assess the statistical significance of the observed selectivity of a compound-target pair in the given bioactivity dataset. We present several case studies that show how the target-specific selectivity can distinguish between

highly selective and broadly-active kinase inhibitors, hence facilitating the discovery or repurposing of multi-targeting drugs.

KEYWORDS

drug selectivity, drug repurposing, drug discovery and development, kinase inhibition activity, polypharmacological effects

1 Introduction

Compound selectivity is a critical factor when developing new drugs or repurposing existing drugs for new uses (Bosc et al., 2017; Schipper et al., 2022). Binding affinity measurements of a compound across various target proteins enable systematic mapping of the target activity space and bioactivity spectrum of the compound. If a compound has a narrow target profile and activity spectrum, i.e., it binds effectively to a few specific targets, then the compound is considered as more selective than a compound with a wide activity spectrum and which binds to multiple targets with similar affinities. The overall selectivity of a compound can therefore be characterized in terms of how narrow or wide its bioactivity spectrum is. Compounds that potently bind to a single target protein are often easier to develop and optimize for clinical use. However, most of the currently used drugs have relatively broad polypharmacological profile, that is, their phenotypic responses are due to interactions with multiple protein targets at different degrees of binding affinity. For instance, kinases are promising therapeutic targets for various indications, including cancer, autoimmune diseases, inflammatory diseases, and cardiovascular diseases, but due to their structural similarity, it is rather challenging to develop highly selective kinase inhibitors (Davis et al., 2011). However, such polypharmacological effects of kinase inhibitors make them also potential candidates for drug repurposing, provided the compound has sufficient selectivity against the off-target proteins driving the disease progression.

A number of statistical and information theoretic metrics have been introduced to quantify compound selectivity. For example, the standard selectivity score calculates the number of targets bound by a compound above a given binding affinity threshold (Karaman et al., 2008). The Gini selectivity metric quantifies how widely the binding affinity measurements of a compound are spread across the target space (Graczyk, 2007; Ursu et al., 2020). More specifically, if there are only a few high binding affinities in the bioactivity spectrum, while the rest of the target activities remain weak, then the binding affinities are unevenly distributed, thus resulting in a high Gini coefficient, and the compound is considered selective. The selectivity entropy also estimates how the binding affinities of a compound distribute across the target space (Uitdehaag and Zaman, 2011). A high entropy indicates that the compound binds to many targets at comparable affinities, and is hence considered non-selective, while low entropy indicates a strong binding to only a few targets, thus making the compound selective (Uitdehaag et al., 2012). While the dissociation constant K_d is often used as an estimate of the binding affinity, the Partition index makes use of association constant K_a instead (Cheng et al., 2010). Partition index quantifies the compound selectivity by calculating the fraction of binding strength (as measured by K_a) to a reference target in comparison to other targets. Recently, the KInhibition Selectivity Score (KISS) was designed for percentage inhibition target activity data, with user-defined on- and off-targets as prior information (Bello and Gujral, 2018). In KISS calculation, penalties are placed on off-target effects by empirical penalty functions, so that lower penalty and higher on-target effects indicate that the compound is selective.

The existing selectivity metrics estimate certain characteristics of a compound's bioactivity spectrum from slightly different perspectives, hence leading to a variable performance in different drug discovery applications (Bosc et al., 2017; Miljković and Bajorath, 2018a; Miljković and Bajorath, 2018b). However, none of the existing metrics are designed for identifying selective compounds for a given target protein of interest. This is because the current selectivity metrics effectively estimate the narrowness of the bioactivity spectrum across the potential targets and consider a compound as highly selective if it binds to only a single target, regardless of the target identity. This makes it difficult to use these metrics for finding selective compounds against a specific target. A target-specific selectivity analysis is needed in many applications, e.g., when developing or repurposing drugs against a specific disease target, while guaranteeing that the drug should not have strong off-target activities toward other proteins which may lead to unwanted side effects (Aittokallio, 2022). To fill this gap, we introduce a target-specific compound selectivity scoring approach to facilitate identification of selective compounds against a given target protein (Figure 1). We demonstrate here the performance and use of the novel selectivity score in the context of kinase inhibitors, which are known to have a wide degree of polypharmacological activities, but the general approach is applicable also to other drug and target classes.

2 Results

2.1 Kinase target activity dataset for the selectivity scoring

To develop and test the new selectivity score, we used a published dataset of fully-measured compound-target

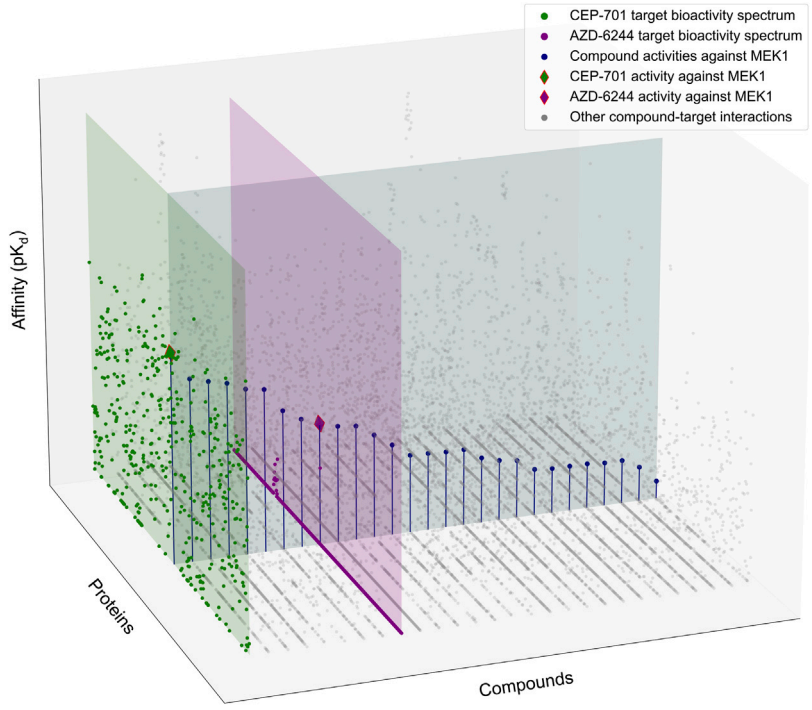


FIGURE 1
Schematic illustration of the target-specific drug selectivity concept. A subset of the Davis et al. dataset (Davis et al., 2011), where 28 randomly selected compounds and all 442 kinases were used for the illustration purposes. The gray horizontal panel shows the activity profile of the 28 kinase inhibitors against MEK1, where the compounds are ordered based on their relative potencies against MEK1. The green and purple vertical panels show the bioactivity spectrums of the compounds CEP-701 and AZD-6244, respectively, across the 442 kinase targets. Even though CEP-701 has the highest potency against MEK1 across all the compounds, it also has other high-potency targets, indicating that CEP-701 is not highly selective against MEK1. While AZD-6244 is not the most potent compound against MEK1, it has its highest potency against MEK1, and therefore AZD-6244 is considered as more selective against MEK1 than CEP-701.

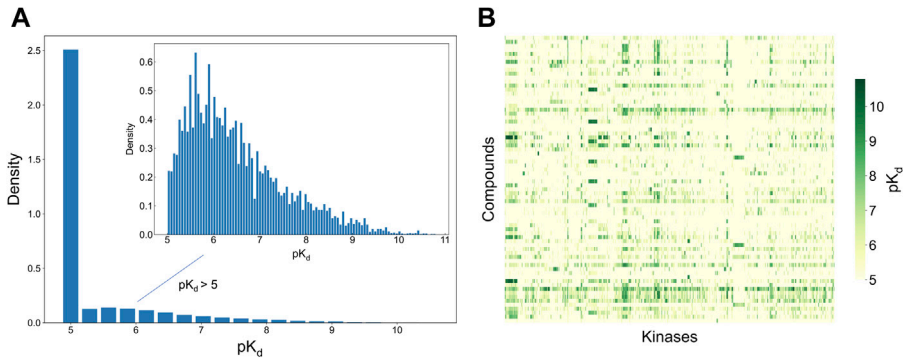


FIGURE 2
Bioactivity data (pK_d values) in the Davis dataset (Davis et al., 2011), containing 72 compounds and 442 kinases. **(A)** The bioactivity distributions, where the larger one includes all bioactivity data, and the smaller one (inset) includes only those bioactivities with $pK_d > 5$ (the pairs with $K_d = 10$ uM, i.e., $pK_d = 5$, indicate no activity in the primary screen). **(B)** Heatmap of the target activities. Higher pK_d (lower K_d) values indicate stronger compound-kinase activities.

interactions between 72 kinase inhibitors and 442 kinases (Davis et al., 2011). Figure 2 shows the distribution of the measured compound-kinase interactions in terms of pK_d. In this bioactivity data matrix, a large number of compound-kinase pairs show no activity, with $pK_d = 5$, i.e., $K_d = 10$ uM, and only a few compound-target pairs show strong potency, with $pK_d > 9$ i.e., $K_d < 1$ nM. As expected with kinase inhibitors that are known to have varied degrees of target promiscuity, many compounds have relatively strong activities against multiple kinases, and many kinases have a number of potent inhibitors. This makes the Davis et al. dataset an excellent test bench for developing and testing a new selectivity method, since it encompasses compounds and kinases with different polypharmacological activities and wide differences in their activity spectra, including both highly promiscuous compounds targeting multiple kinases at low concentrations, and highly selective compounds with narrow target activity profiles.

2.2 Decomposition of target-specific compound selectivity

Given a compound $c_i \in C$ and a target $t_j \in T$, the bioactivity spectrum of the compound c_i can be defined as $B_{c_i} = \{K_{c_i,t_j} \mid t_j \in T\}$, and the activity profile of the target t_j can be defined as $P_{t_j} = \{K_{c_i,t_j} \mid c_i \in C\}$, where K_{c_i,t_j} is the interaction strength between c_i and t_j (here, dissociation constant K_d, but in general it can be any binding affinity estimate).

The existing compound selectivity metrics try to characterize the distributional properties of B_{c_i}, essentially measuring whether a compound interacts with only a few or larger number of targets. However, such a compound-specific approach is not sufficient when a specific protein target is under investigation. When assessing the target-specific compound selectivity, two aspects of the pairwise interactions need to be considered (1): how the interaction strength of a compound is distributed across its targets, i.e., characterizing B_{c_i}; and (2) how the interaction strength of a target is distributed across the compounds, i.e., characterizing P_{t_j} (see the horizontal and vertical panels of Figure 1).

Given a set of compounds C and a set of targets T, that are explored in a target activity profiling study, the task of finding the most potent compounds and highest affinity targets among the compound and target spaces can be formulated as an optimization problem:

$$c^*(t_i) = \underset{c_i}{argmax}\, P_{t_j}, s.t.\ t_j \epsilon T$$

$$t^*(c_i) = \underset{t_i}{argmax}\, B_{c_j}, s.t.\ c_i \epsilon C$$

However, as was illustrated in Figure 1, the optimal solutions to these two objectives do not agree in general, i.e., the most potent compound $c^*(t_j)$ (e.g., CEP-701 in Figure 1) for a target t_j (e.g., MEK1) is not necessarily among the compounds (e.g., AZD-6244) that each exert their highest affinity toward t_j and are considered selective in this respect. Likewise, the selective potency of a compound (AZD-6244) for a target (MEK1) does not imply that the most potent compound (CEP-701) for the same target shows superior potency over the other targets. Therefore, the target-specific selectivity needs to be formulated as a multi-objective optimization problem that considers both B_{c_i} and P_{t_j}.

Intuitively, for a target t_j, one tries to find the compound c_i that simultaneously maximizes K_{c_i,t_j} in P_{t_j} and minimizes some statistic describing $B_{c_i} \backslash \{K_{c_i,t_j}\}$, for example, the mean of the set $B_{c_i} \backslash \{K_{c_i,t_j}\}$. In addition to the global mean, we also used a more local statistic by taking the mean of the h-nearest neighbors of K_{c_i,t_j} in B_{c_i}, i.e., $B_{c_i,\, hNN(t_j)}$, where $hNN(t_j)$ denotes the h-nearest neighbors of K_{c_i,t_j} in B_{c_i} in terms of target activity.

We formulated the above two statistics relative to K_{c_i,t_j} as below:

$$Global\ relative\ potency \quad G_{c_i,t_j} = K_{c_i,t_j} - mean\left(B_{c_i} \backslash \left\{K_{c_i,t_j}\right\}\right) \quad (1)$$

$$Local\ relative\ potency \quad L_{c_i,t_j} = K_{c_i,t_j} - mean\left(B_{c_i,\, hNN(t_j)}\right) \quad (2)$$

Additionally, K_{c_i,t_j} is termed as *absolute potency*. Based on these definitions, the target-specific selectivity can be obtained as a solution of the bi-objective optimization problem, in which one maximizes simultaneously both the absolute potency and the relative potency, which can be easily solved using the ε-constraint method (Haimes, 1971; Miettinen, 1999) (see Materials and methods for details). Here, we used the neighborhood size of $h = 5$ in the local relative potency, unless otherwise specified.

In the Davis dataset, 1,208 selective compound-kinase pairs were identified among the 31,824 total pairs between 72 compounds and 442 kinases when using the local relative potency (Figure 3A); while using the global relative potency, 660 selective pairs were identified (Figure 3B). Even if the use of the local relative potency in the optimization problem led to 1.8-fold more selective compound-target pairs, compared to using the global relative potency, there is still a relatively large overlap between the identified selective compound-target pairs (Figure 3C). Since the local and global relative potencies capture different aspects of B_{c_i}, they lead to different optimal solutions. However, selective compound-target pairs identified using both statistics can be considered together, based on the needs of the user.

2.3 The integrated target-specific compound selectivity score

When applying the bi-objective optimization to identify selective compound-target pairs, an integrated selectivity score can be calculated by combining both the local and global relative potencies to quantify the selectivity of a compound for a given target. Such integrated selectivity score S_{c_i,t_j} for the compound-target pair (c_i, t_j) is formally defined as:

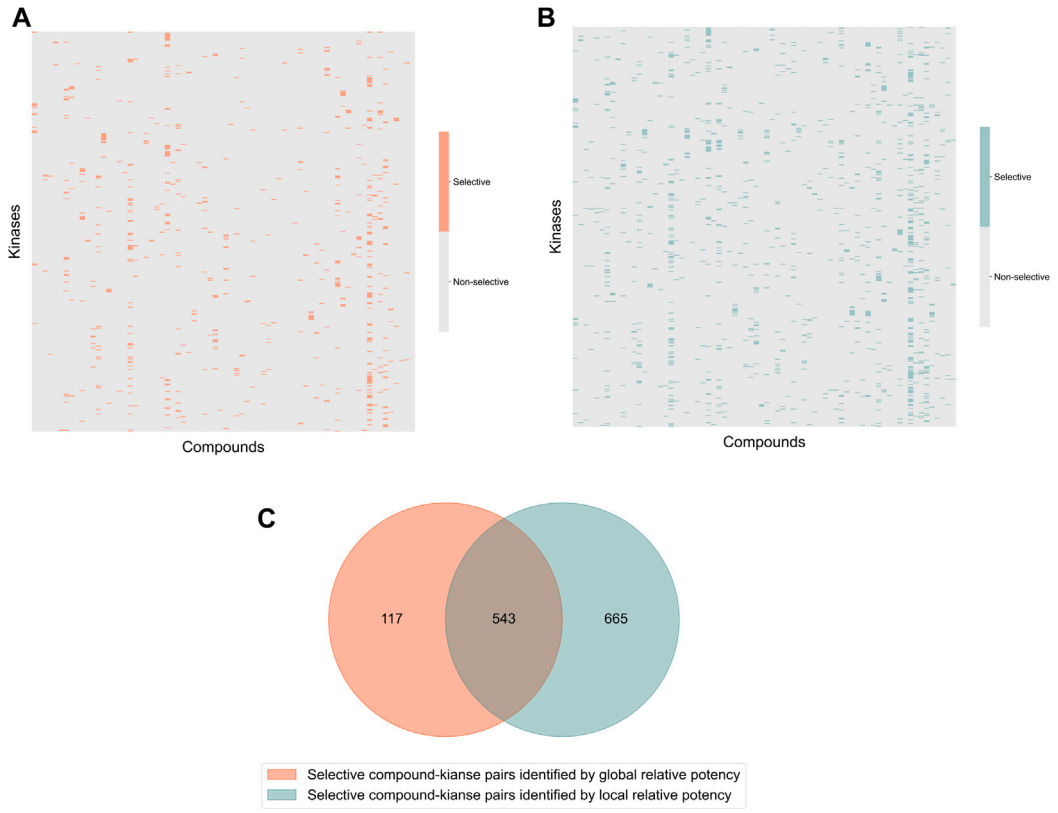


FIGURE 3
Heatmaps of the identified selective and broadly-active compound-kinase pairs among 72 compounds and 442 kinases when using **(A)** local relative potency and **(B)** global relative potency; **(C)** the overlap of the identified selective compound-kinase pairs identified using the local and global relative potency.

$$S_{c_i,t_j} = \alpha \cdot L_{c_i,t_j} + (1-\alpha) \cdot G_{c_i,t_j} \quad (3)$$

where the parameter α adjusts for the contributions of the local and global relative potency to the selectivity score.

The global relative potency G_{c_i,t_j} focuses on comparing the compound's interaction strength against a specific target, relative to the average affinity to the other targets, and it therefore reflects the general interaction strength over $B_{c_i}\backslash\{K_{c_i,t_j}\}$. Large G_{c_i,t_j} indicates that the K_{c_i,t_j} is generally high compared to the $mean(B_{c_i}\backslash\{K_{c_i,t_j}\})$, but we note that ($c_i$, t_j) is not necessarily the only pair with strong interaction. For example, its nearest neighbor in terms of target potency, K_{c_i,t_a}, may be as high as K_{c_i,t_j}, meaning that compound c_i has similar interaction strength against t_j and t_a. Therefore, the local relative potency L_{c_i,t_j} was introduced to better distinguish between K_{c_i,t_a} and K_{c_i,t_j}, since it emphasizes the local potency, relative to the average of neighbor targets, instead of all the other protein targets.

A weighted sum of the two relative potencies can be used to quantify the integrated selectivity of a compound-target pair to be optimized in Eq. 3. The weight of each relative potency term can be freely adjusted by the user. When more weight is placed on the local relative potency, then the selectivity score will focus more on distinguishing between the given target and its nearest neighbors in terms of the interaction strength, hence identifying compounds most potent against the given target in the context of the target neighborhood. As a default option, the mean of local and global relative potencies can be used (i.e., α = 0.5), if none of the terms is considered more important than the other in the particular drug discovery or repurposing application (Figure 4A).

Figure 4B shows the correlation between the integrated selectivity scores and pK_d values, colored for three example compounds discussed below. In general, and as was expected, a higher interaction strength (absolute potency measured by pK_d) corresponds to higher selectivity. However, by combining the local and global relative potencies, one can discover compounds that are selective, yet may have relatively weak interaction strengths. For example, MLN-120B has a relatively weak absolute potency of pK_d = 7.72 with IKK2, but it was identified as selective against IKK2 with a relatively high selectivity score of 2.11. In the Davis dataset, MLN-120B is the second most potent inhibitor of kinase IKK2, yet having the highest local relative potency. This example shows that with the

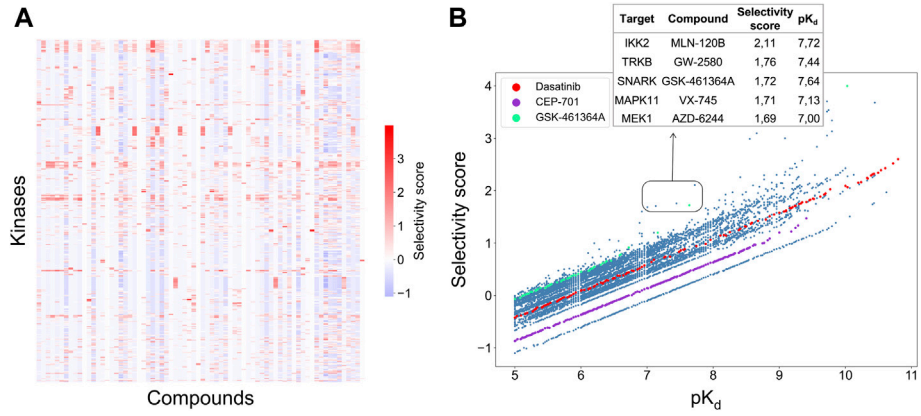


FIGURE 4
(A) Heatmap of the selectivity scores between 72 compounds and 442 kinases when using the weighting factor $\alpha = 0.5$ in Eq. 3; **(B)** Correlation between the selectivity scores and absolute potency pK_d across the 31,824 compound-kinase pairs in the Davis dataset. Higher scores indicate higher selectivity. Examples of compound-kinase pairs with relatively low interaction strengths and high selectivity scores are highlighted in the box, and details shown in the inset table.

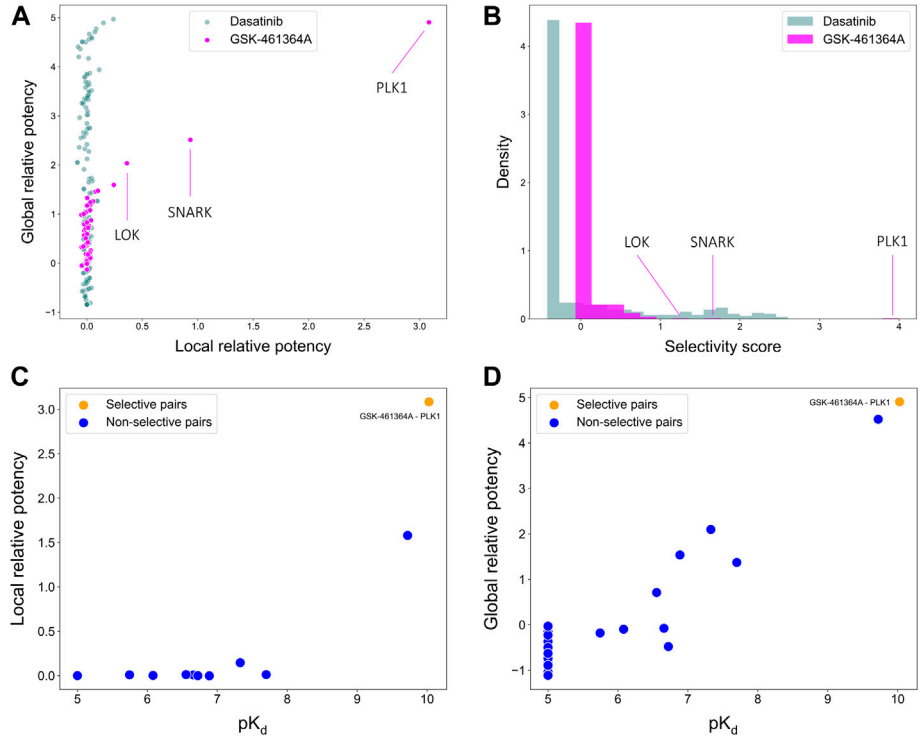


FIGURE 5
Comparisons of **(A)** local and global relative potencies and **(B)** distributions of selectivity scores for dasatinib and GSK-461364A across 442 kinase targets. GSK-461364A was identified through bi-objective optimization as selective against PLK1 using both **(C)** local relative potency and **(D)** global relative potency.

adjustable weights, it is possible to reach a balance between compound potency and selectivity, with the aim to find maximally selective and potent compounds for a particular target of interest.

2.4 The application of target-specific selectivity score to kinase inhibition

To illustrate the use of the target-specific selectivity score, Figure 5 shows the selectivity scores and relative potencies of two compounds: dasatinib and GSK-461364A. GSK-461364A is known to be highly selective against only a few kinase targets, PLK1, SNARK, and LOK, with much higher selectivity scores than for other kinases (Figure 5B). In contrast, dasatinib is a broad-spectrum multi-kinase inhibitor, and therefore many of its targets have high global relative potencies, but none of these targets have a high local relative potency (Figure 5A). A high global relative potency indicates that the compound shows overall selectivity to any target in general, since it considers the mean of $B_{c_i} \backslash \{K_{c_i,t_j}\}$. Thus, when considering dasatinib to be selective against a set of targets, more weight can be placed on the global relative potency; when searching for selective compounds against a few specific targets, more weight can be placed on the local relative potency. In this way, the target-specific compound selectivity score provides flexibility and becomes applicable to different drug discovery needs.

As shown above, GSK-461364A was identified as a highly selective PLK1 inhibitor since it has both high local and global relative potencies against PLK1 (Figures 5C,D). The bi-objective optimization also identified GSK-461364A as an optimally selective compound against SNARK and LOK, due to its high local relative potency (Supplementary Figure S1). For many kinase targets, such as PLK1, multiple highly selective compounds can be rather easily identified from the Davis dataset, but for some other targets, such as SNARK, LOK, and other targets shown in Supplementary Figure S1, compromises between the potency and selectivity need to be made through the bi-objective optimization. A *Pareto front* was generated to illustrate all the equally optimal compounds for a given target. For example, multiple compounds were identified as optimally selective for the kinase TNIK (see Supplementary Figure S1). The most selective compounds for the target can then be identified using the selectivity score (Eq. 3), along with other available information, including physicochemical properties of the compounds or their toxicity profile. In this way, the pareto optimization provides the user with additional quantitative information for the drug discovery process.

2.5 Evaluation of the stability of the target-specific selectivity score

To evaluate the stability of the target-specific selectivity score, we first studied the impact of missing bioactivity values by adding 20, 40, 60 and 80% of missing values to the full bioactivity data matrix, while keeping all the compounds and targets in the matrix. When considering all compound-target pairs, the global relative potency was in general more robust to missing data than the local relative potency (Figures 6A,B). For each kinase target, the recall value was calculated using the identified selective compounds from the full data matrix as true positives, using both local and global relative potencies (Figures 6C,D). As expected, the recall tends to decrease when increasing the missing value rates in the bioactivity data matrix. When only 20% of non-missing data are available, the recall values were distributed mostly at zero, suggesting that the identified selective pairs are not stable anymore. Based on the above results, the methodology appears reasonably consistent in bioactivity data matrices that have maximally 20% of bioactivity pairs missing.

Next, we studied the effects of various bioactivity data matrix sizes on the stability of the identifications. Data matrices of increasing sizes were subsampled from the full data matrix, with 20, 40, 60 and 80% of compounds and targets included, and the selective compound-target pairs were identified based on each subsampled matrix. For a compound-target pair, the number of times it was identified as selective in the submatrices of different sizes was considered as a measure of consistency. If a compound-target pair was identified as selective in all the data matrices, regardless of the bioactivity matrix size, it indicates that even with a very small data size, for example 20% of the compounds and targets that corresponds to 4% of the full data matrix, the method can still identify the selective pairs, and the result is consistent with that when using the larger bioactivity data matrices.

Supplementary Figure S2 shows the overall heat map counting the occurrences of selective pairs consistently identified across different sizes of submatrices. A count of 5 means a compound-target pair was identified as selective in all submatrices of different sizes, and a count of 1 means a compound-target pair was identified only once as selective. Some compound-target pairs were only present in the largest data matrix, i.e., the full data matrix, thus they can only be identified once. Similar to Figure 6, the selectivity score tends to be more stable when using global relative potency, as the identified pairs are more consistent compared to that when using the local relative potency (Supplementary Figure S2A). As expected, gradually decreasing the data matrix size leads to identification of certain targets with many selective compounds, indicating increased instability. In general, when the data size is the smallest, i.e., 4% of the full data matrix, the method starts to behave inconsistently, suggesting that larger data matrices are required.

2.6 Statistical evaluation of the relative potency using empirical *p*-values

Statistical properties of the relative potency were next studied by randomly permuting the compound-target bioactivity matrix. Local and global relative potencies were calculated based on the

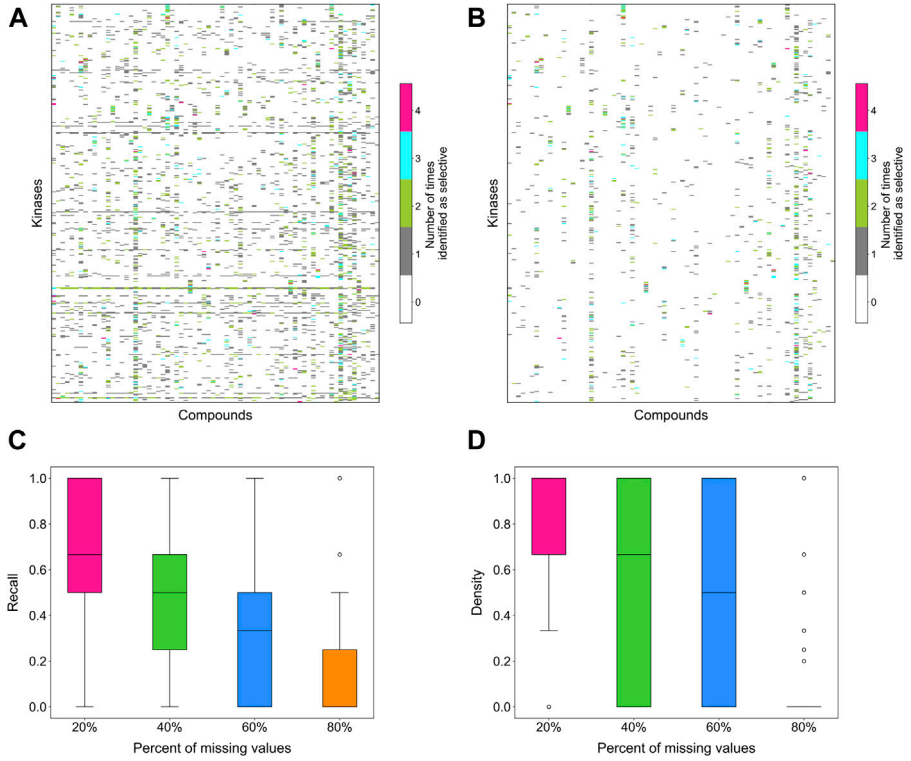


FIGURE 6
The number of times a compound was identified as selective for a target in bioactivity matrices with missing values when using **(A)** local relative potency and **(B)** global relative potency. Upper row: the heatmaps show the overall results in the matrix between 72 compounds and 442 kinases when adding 20, 40, 60 and 80% of missing values to the full bioactivity data matrix. In panel a, gray stripes correspond to kinases for which almost all compounds are identified as selective, indicating instability; Bottom row: the boxplots of the recall of identification of selective drug-kinase pairs from data matrices with missing values when using **(C)** local relative potency and **(D)** global relative potency, using the selective pairs identified in the full data matrix as ground truth.

permuted matrices to form the background distribution for null hypothesis. As expected, the background distributions were concentrated at around zero, especially for the local relative potency (Figures 7A,B). Next, the empirical *p*-values were calculated for each compound-target pair based on the background distributions (Figure 7C). The empirical *p*-values for the global relative potency were almost uniformly distributed, as would be expected for a proper statistic, but for the local relative potency, the *p*-values tend to be either very small or close to 1. The ill-distributed *p*-values of local relative potency may be due to the local neighborhood size ($h = 5$) that was used as default in its calculation. When comparing the *p*-value distributions of compound-target pairs identified as selective with those of non-selective pairs, it was observed that *p*-values for selective pairs are more concentrated around zero, i.e., indicating statistically significant target-specific selectivity (Figures 7D,E).

We note that the local relative potency is closely related to the global relative potency, since when the number of neighbors *h* is increased to all the targets, the local relative potency becomes equal to the global relative potency. Thus, we wanted to study the effect of using increasing numbers of nearest neighbors when calculating the local relative potency for the bi-objective optimization. In general, different numbers of nearest neighbors resulted in rather similar detections, which are distinct compared to using the global relative potency (Supplementary Figure S3A). When comparing the identified selective compounds per target, using the local relative potency based on different numbers of nearest neighbors, we calculated recall values using selective compounds identified by the global relative potency as true positives. The recall distributions showed that the performance of the local relative potency is again relatively consistent when the number of nearest neighbors varies (Supplementary Figure S3B). Taken together, the consistent behavior of the local relative potency calculation indicates that the two versions of the relative potency capture both unique and common properties of the compound-target interactions.

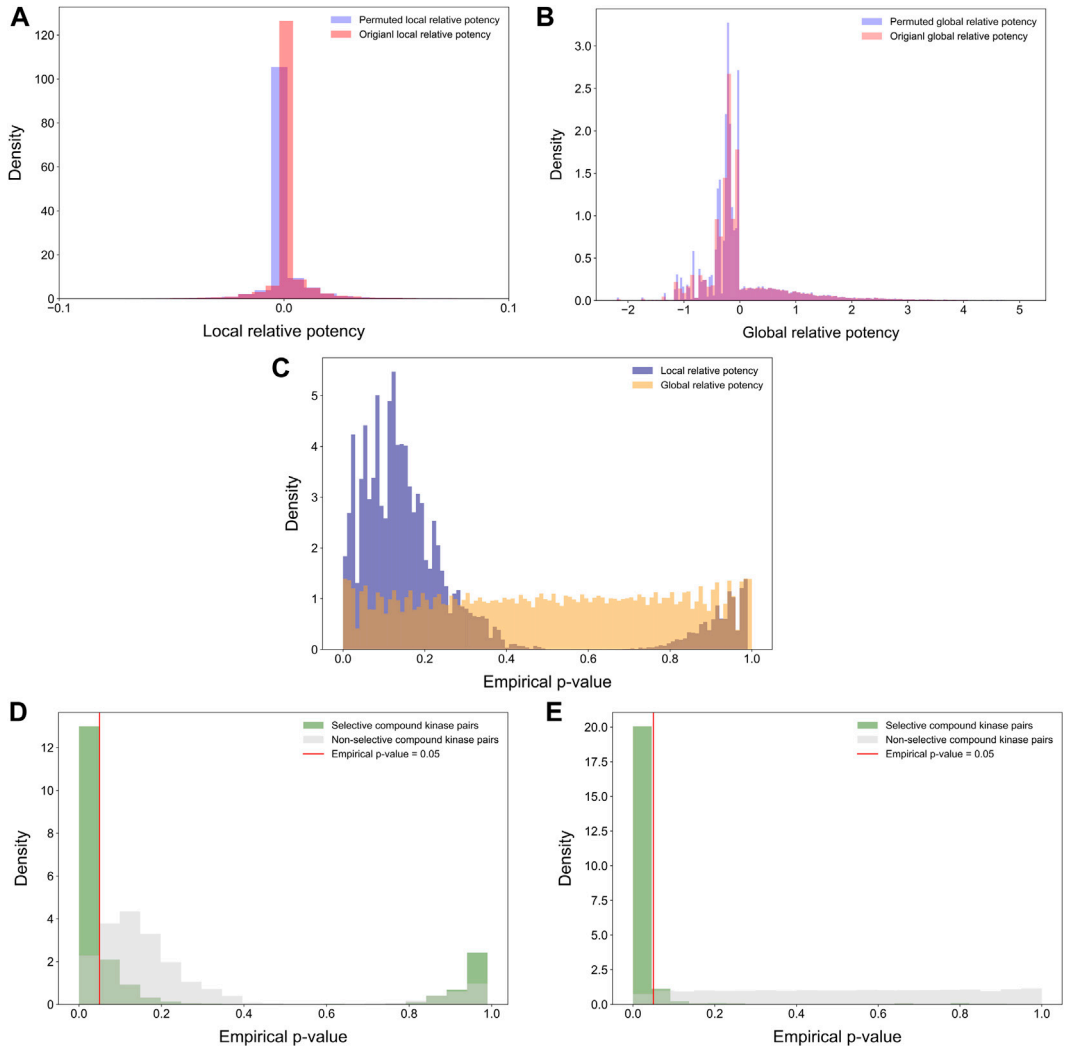


FIGURE 7
Distributions of **(A)** permuted and original local relative potencies; and **(B)** permuted and original global relative potencies; **(C)** the empirical *p*-values calculated with permutation procedure for both local and global relative potency; empirical *p*-values of **(D)** local relative potency and **(E)** global relative potency colored by whether the compound-kinase pair is identified as selective or not.

2.7 Comparison of target-specific selectivity with existing selectivity metrics

Since most of the existing compound selectivity metrics are designed only from the perspective of compound selectivity, it is not straightforward to make comparisons between those metrics and our target-specific selectivity metric. Furthermore, the metrics are also designed for different bioactivity readouts, and may have different directions and scales to indicate selectivity. To make a reasonable comparison, we z-scaled and standardized all the metrics so that the smaller the metric, the more selective the compound (see Materials and Methods). For our target-specific selectivity score, we used a summarized, target-agnostic selectivity score, calculated as the mean of selectivity scores of a compound across all available targets. We also used the number of identified selective targets for each compound as a measure of the compound's overall selectivity, regardless of the target. Supplementary Figure S4 shows that such summarized measures coincide among the selectivity metrics, since in effect, they all measure whether a compound has a strong activity against multiple or only a few targets. For example, the local and global relative potencies correlated well with the standard score using pK_d of 7 as the activity cut-off (Supplementary Figure S4).

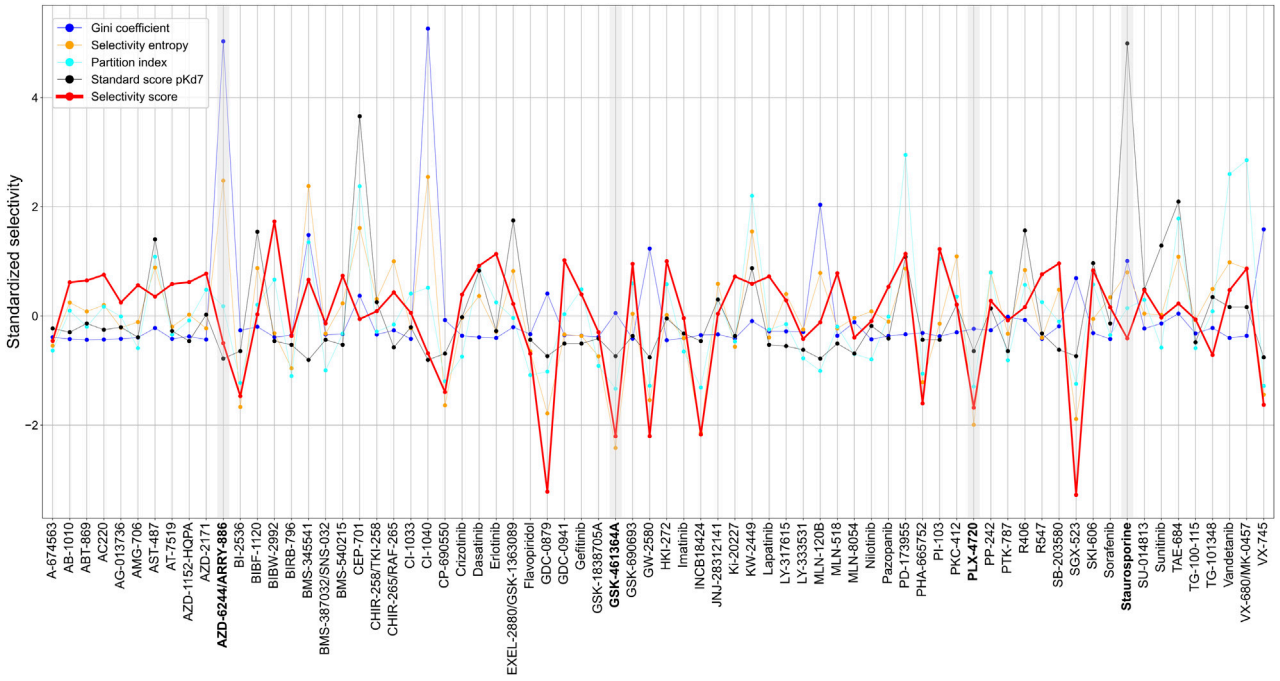


FIGURE 8
Comparison of various target-agnostic compound selectivity metrics. Local and global relative potencies are summarized along the targets of each compound in the target-specific selectivity metric, and all the metrics are z-scaled and standardized so that the smaller metric values indicate more selective compounds. The boldfaced compounds are discussed in the text.

Across the 72 kinase inhibitors in the Davis dataset, most of the target-agnostic summary metrics identified selective and broadly-active compounds rather consistently, expect for the Gini coefficient that was not highly correlated with the other metrics (Figure 8). This could be due to the different data types required by Gini coefficient, which was designed for percent inhibition values instead of K_d data. For example, dasatinib and staurosporine are two well-known broadly-active kinase inhibitors, and they were considered as non-selective by most of the metrics. Similarly, more target-specific compounds, such as GSK-461364A and PLX-4720, were identified as highly selective compounds by most of the summary metrics. As an exception, AZD-6244 was considered non-selective in terms of Gini coefficient and selectivity entropy, with relatively high scores compared to other compounds in the dataset, whereas AZD-6244 was considered relatively selective by our selectivity score and the standard score. Upon inspecting the Davis dataset, AZD-6244 has interaction strengths of pK_d >5 with 13 out of 442 kinases, which are mainly MEKs and EGFR mutants (Supplementary Figure S5).

These results demonstrate a consistent performance of our target-specific selectivity metric, when using it to measure the overall target-agnostic compound selectivity.

2.8 Comparison of target-specific selectivity with partition index

To make a more detailed, target-specific comparison, the partition index scores were calculated such that each kinase target was used separately as the reference target (see Figure 9A which shows the negative logarithm of the target-specific partition indices). The vertical stripes indicate that the partition index considers many compounds to be selective against all the targets, suggesting that the partition index is not generally capable of finding selective compound-target pairs. When comparing the partition index with our target-specific selectivity score, it was observed that the two metrics are generally well correlated, as expected, but the new selectivity metric was more distinctive in terms of identifying selective compound-target pairs (Figure 9B). Especially, when the partition index is small, between 0 and 1, the selectivity score can still distinguish between the highly selective and broadly-active kinase inhibitors better than the partition index.

Supplementary Table S1 shows several example compound-target pairs that have low partition indices, yet higher and more different target-specific selectivity scores (the black bordered points in the bottom right corner of Figure 9B).

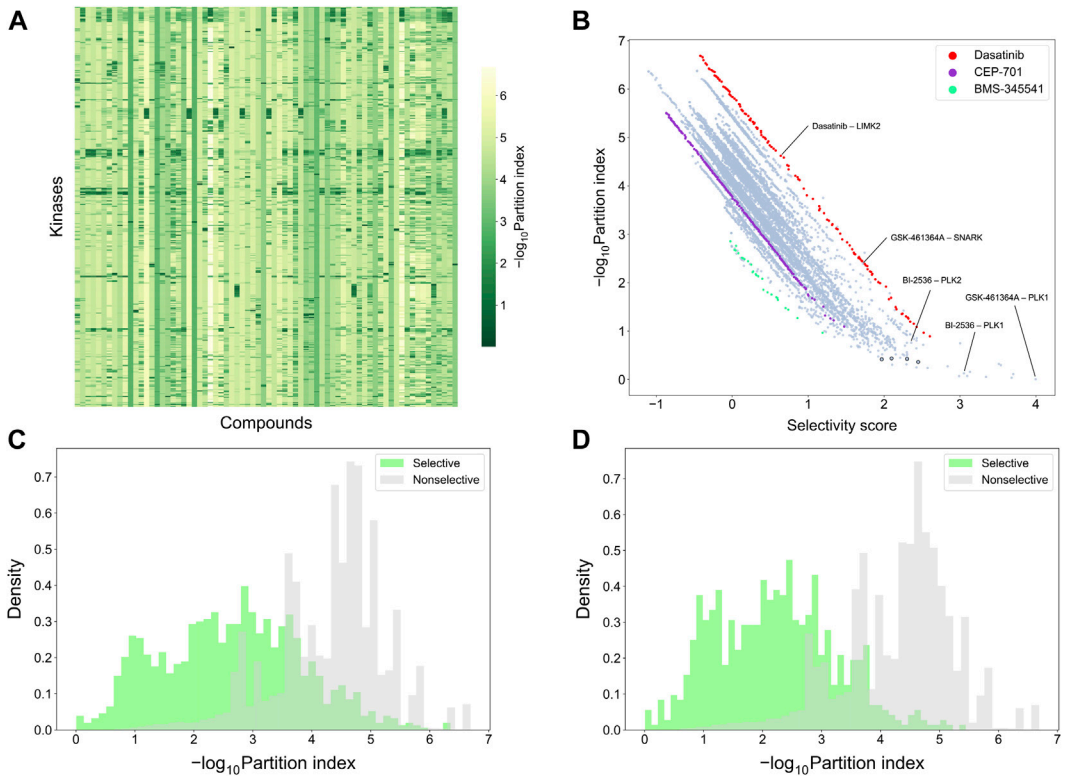


FIGURE 9
Comparison of target-specific selectivity score and partition index. Upper row: **(A)** Heatmap of -log10 (partition index) for each kinase, where smaller values indicate more selectivity; **(B)** correlation between partition index and selectivity score across the 31,824 compound-kinase pairs in the Davis dataset; Bottom row: distributions of -log10 (partition index) colored by whether the compound-kinase pair was selective when using **(C)** local relative potency and **(D)** global relative potency.

For example, the compound-target pairs (nilotinib, DDR1) and (PTK-787, KIT) have selectivity scores of 2.30 (0.11% quantile) and 1.97 (0.38%), respectively, while their partition indices are 0.43 (0.057%) and 0.42 (0.053%), suggesting that target-specific selectivity score provides slightly better separation for the pairs, as further supported by the significant *p*-values, using both local and global relative potency (Supplementary Table S1). Such observations suggest that the new selectivity score harnesses different information than the partition index, thus providing additional perspective to the target-specific discovery or repurposing of selective compounds.

To further compare the two selectivity approaches, Figures 9C,D shows the distributions of the partition index for the compound-target pairs identified as selective or non-selective by the target-specific selectivity score. Regardless of whether using the local or global relative potency, the partition indices of the selective pairs tend to have lower values than those of non-selective pairs, indicating that the two methods are generally consistent with each other. However there exists also pairs identified as selective by the target-specific score, yet having a relatively large partition index values, or vice versa, shown as the overlaps of two distributions in Figure 9C,D. For example, the pair (BI-2536, PLK1) has a very low partition index of 0.13, indicating relatively high selectivity. In the Davis dataset, GSK-461364A is the most potent inhibitor of PKL1 ($pK_d = 10.03$), with BI-2536 being the second most potent ($pK_d = 9.72$) (Supplementary Figure S5). From the compound perspective, both compounds have their highest potency against PLK1. However, GSK-461364A has a pK_d of 7.64 for its second most potent target (SNARK), while BI-2536 has a pK_d of 9.09 against PLK2. Since BI-2536 has very similar potencies against its top-2 most potent targets, it is not considered as selective against PLK1 when GSK-461364A is available in the library. These examples further demonstrated that our method provides an added value for finding target-specific selective compounds.

3 Discussion

Finding selective compounds is considered important for kinase drug discovery since many of the current kinase inhibitors

are relatively promiscuous. This is the case also with many kinase inhibitors marketed or under current development (Cohen et al., 2021), and it remains a challenging task to find more targeted and selective inhibitors that can both improve efficacy and reduce the unwanted off-target toxicity (Attwood et al., 2021). Our results show that the new target-specific selectivity score provides an added value for the discovery of multi-targeting, yet selective compounds in the case when a target of interest is pre-defined. The selectivity score derived from the relative potencies measures the target-specific compound selectivity quantitatively and provides flexibility for the user. The bi-objective optimization was capable of identifying the maximally selective compound-target pairs in the presence of a wide degree of polypharmacological effects. The flexibility comes from the user-adjustable weights for the local and global relative potencies in the selectivity score, as well as from using both the relative and absolute potency in the bi-objective optimization. Such flexibility allows wide applications, based on different user needs, for example, finding the most selective compound for a single target or group of targets. Thus, the new metric is expected to become beneficial in kinase inhibitor development, and more broadly in lead compound identification in drug discovery and for repurposing multi-targeting drugs.

The advantage of the target-specific selectivity score is that it requires only the bioactivity measurements of the compound-target pairs, without the need to provide other information of the compounds, such as their on/off targets or chemical structures. This makes our approach widely applicable to various types of bioactivity measurements. In case the available bioactivity data contains various studies of target activities using multi-dose assays, such as a mix of K_i, K_d and IC_{50} readouts, then the bioactivity readouts can be summarized and integrated using our previously developed data transformations (Wang et al., 2020). Due to its data-driven approach, the approach is not only limited to kinase inhibitors, but once sufficient amounts of similar bioactivity data become available for other target classes, such as G-protein-coupled receptors (GPCRs), the same approach is directly applicable to these data. Apart from calculating the target-specific selectivity score, the approach also provides optimal solutions of the most selective compound-target pairs based on the given bioactivity data. Finally, the target-specific selectivity enables the user to find selective compounds for the particular targets of interest. Such target-specificity provides a unique perspective to analyzing compound selectivity, and expands the application area of the current compound selectivity metrics in multi-target drug discovery and repurposing.

The limitation of any data-driven approach is the data availability and quality. Since the target-specific selectivity approach requires experimentally measured bioactivity data, we recommend that at least 80% of the compound-target pairs should have measured bioactivities to obtain a reliable performance. Such a requirement limits the approach to only compounds with sufficient amounts of target bioactivity measurements available. However, the approach can be further developed by incorporating other information of either compounds or targets, for example, compound structural similarity (Lo et al., 2019) to infer selectivity of novel compounds, even without any measured bioactivities. Alternatively, machine learning methods can be used to predict bioactivities for the compound-target pairs that have not yet been explored experimentally (Bora et al., 2016; Merget et al., 2017; Öztürk et al., 2018; Thafar et al., 2019; Vamathevan et al., 2019; Bagherian et al., 2020; Nguyen et al., 2020; Schneider et al., 2020; Cichońska et al., 2021; Ye et al., 2021), after which the target-specific compound selectivity metric can be applied to the fully predicted compound target interaction matrix to identify selective lead compounds against any target of interest. In the general method development, we did not distinguish between the on- and off-targets, or penalized targets that may lead to adverse effects in clinical applications, but such factors could be later incorporated into the general selectivity scoring approach when applied to a particular disease or cellular context, similar to the KInhibition Selectivity Score (Bello and Gujral, 2018), but this will require careful distinction between the therapeutic and toxicity-related targets.

4 Conclusion

We have developed a novel target-specific compound selectivity metric by decomposing the selectivity into absolute and relative potencies. Two statistics were used to describe the relative potency, local and global relative potencies, which characterized the target-specific compound selectivity from different aspects and can be combined using a weighted sum as the integrated selectivity score to facilitate the quantification of compound selectivity. A bi-objective optimization problem was used for maximizing both absolute and relative potencies to identify the maximally target-specific selective compounds in a given compound-target interaction dataset. The new selectivity approach is expected to contribute to finding selective compounds with improved target-specificity, as well as to enable repurposing of existing multi-targeting drugs for new disease indications that are driven by the specific disease protein.

5 Materials and methods

The workflow of the target-specific compound selectivity scoring is illustrated in Supplementary Figure S6.

5.1 Compound-target interaction data for method development

The compound target activity data used to develop and test the target-specific compound selectivity were obtained from

Davis et al. (Davis et al., 2011), hereby called the Davis dataset. In the Davis dataset, dissociation constant K_d was measured for all pairs between 72 compounds and 442 kinases. In our analyses, $pK_d = -\log_{10}(K_d)$ is used, and the larger is the pK_d the stronger the binding affinity.

5.2 Decomposition of target-specific compound selectivity

Similar to our previous work on identification of selective drug combination treatment effects (Pulkkinen et al., 2021), two aspects of compound binding properties were considered to quantify target-specific selectivity (1): the compound's potency against the target of interest, termed the *absolute potency*; and (2) the compound's potencies against other targets, termed the *relative potency*. To find a selective compound for a given target protein, we consider that the compound needs to be potent enough against the target, and simultaneously, it must have a weak or no activity against the other potential targets.

The absolute potency can be basically any multi-dose bioactivity measurement, such as K_i, K_d, IC_{50} or EC_{50}, which measures the binding affinity between the compound and target of interest. The relative potency can be quantified in different ways, for example, as the difference between the absolute potency and the mean of a compound's potencies against all the other targets, except for the target of interest. Such relative potency uses as reference the compound's overall binding affinity with all other targets, thus termed as *global relative potency*. A more focused measure of relative potency is to consider only those targets having the closest potencies to the target of interest, for example, the difference between the absolute potency and the mean of *h* nearest neighbors' potencies with the target of interest. Such calculation measures the compound's average interaction strength within the local neighborhood of the target of interest, thus termed as *local relative potency*. If the mean value is higher than the absolute potency, this indicates that the compound has similar or stronger binding activity with several targets.

5.3 Bi-objective optimization to identify target-specific selective compounds

Selectivity score provides a quantitative tool to understand and quantify target-specific compound selectivity. However, in most cases, it is difficult to find the optimally selective compound for a specific protein target. Therefore, we used bi-objective optimization to find the most selective compound-target pairs given a particular compound-target interaction dataset. Two separate bi-objective optimization problems were solved to identify target-specific selective compounds (1): maximizing both absolute potency K_{c_i,t_j} and local relative potency L_{c_i,t_j} (2); maximizing both absolute potency K_{c_i,t_j} and global relative potency G_{c_i,t_j}.

Let us denote by K_{c_i,t_j} the binding strength of a compound c_i from a set of compounds $C = \{c_i\}$ against a target protein t_j from a set of protein targets $T = \{t_j\}$. The activity spectrum of a compound c_i can then be defined as $B_{c_i} = \{K_{c_i,t_j} \mid t_j \in T\}$.

For the optimization formulation, the two relative potencies are formally defined as follows:

Local relative potency:

$$L_{c_i,t_j} = K_{c_i,t_j} - \frac{1}{n}\sum_{h=1}^{n} K_{hNN(c_i,t_j)}$$

where $K_{hNN(c_i,t_j)}$ denotes the absolute potency of hth nearest neighbor of t_j given c_i.

Global relative potency:

$$G_{c_i,t_j} = K_{c_i,t_j} - \frac{1}{|T|-1}\sum_{l=1}^{|T|} K_{c_i,t_l},\ (l \neq j)$$

The bi-optimization problem is to maximize both the absolute potency and the relative potency, which can be solved using the ε-constraint method (Haimes, 1971; Miettinen, 1999) as follows:

$$\bigcup_{\varepsilon\in\mathbb{R}} \left\{ \underset{c,\,t}{argmax}\, K_{c,t} \,\middle|\, L_{c,t} < \varepsilon,\ c \in C,\ t \in T \right\}$$

$$\bigcup_{\varepsilon\in\mathbb{R}} \left\{ \underset{c,\,t}{argmax}\, K_{c,t} \,\middle|\, G_{c,t} < \varepsilon,\ c \in C,\ t \in T \right\}$$

Here, the relative potency can be calculated either by local or global relative potency, $L_{c,t}$ or $G_{c,t}$, respectively. We used $h = 5$ as default neighborhood size in the local relative potency.

5.4 Evaluation of target-specific selectivity

We carried out several analyses to evaluate the performance and stability of the target-specific selectivity score.

5.4.1 The effect of matrix size and missing bioactivity values

We first studied the effect of compound-kinase interaction matrix sizes on the identification of selective compound-kinase pairs. Increasingly sized submatrices were sampled using 20, 40, 60, 80 and 100% of the compounds and kinases in the full matrix, respectively. In each submatrix, the same selectivity identification method was applied to generate a binary matrix with 0 indicating non-selective and 1 selective compound-kinase pairs. The matrices were aligned by the identity of compounds and kinases and added up accordingly. For example, all the five submatrices contain the first 20% of the compounds and kinases. Therefore, the sum of the binary matrices, which ranges between 1 and 5, indicates how well the method reproduces the same selectivity identification for the compound-kinase pairs present in the particular part of the matrix.

Next, the effect of missing bioactivity values was studied. For each kinase, 20, 40, 60, 80% compounds were randomly subsampled from the set of all compounds, and these were

assigned as missing, to form matrices with random artificial missing values. Such matrices were generated with 20, 40, 60, 80% missing values independently (i.e., missing completely at random). The same selectivity identification method was applied to all the matrices. The identified selective compound-kinase pairs from each subsampled matrix were compared to those identified based on the original full data matrix, without missing data, to study the effect of increasing the amount of missing data.

5.4.2 Permutation procedure to calculate empirical p-values

The original compound-kinase bioactivity matrix was randomly shuffled for 10,000 times, corresponding to a bioactivity matrix between compounds and kinases where the labels of the compounds/kinases were randomized, and the identification method was applied to each of those randomized matrices to form the background distributions for the local and global relative potencies. Then, for the observed local and global relative potencies calculated from the original matrix, empirical *p*-values were calculated as the percentage of values in the background distribution smaller or equal than the observed local and global relative potencies, respectively.

5.4.3 Relationships between h, local and global relative potency

To study the effect of the number of nearest neighbors *h* used in the calculation of the local relative potency, an increasing number of 1, 5, 20, 100 nearest neighbors were used to calculate the local relative potency. Then, for each kinase, the number of identified selective compounds was compared among the local relative potencies when using different numbers of nearest neighbors.

The local relative potency becomes equal to global relative potency when setting the number of nearest neighbors equal to all available neighbors, i.e., $h = |T| - 1$. Therefore, selectivity identified using global relative potency was considered as the ground truth, against which the selectivity identified using different local relative potencies were compared, and the recall values were calculated:

$$Recall = \frac{TP}{P}$$

Here, *TP* is the number of true positives, i.e., the overlap between the selective compound-target pairs identified both by the local relative potency, using different numbers of nearest neighbors, and by the global relative potency, considered as the ground truth. *P* is the number of positive cases, i.e., the selective compound target pairs identified by the global relative potency.

5.5 Comparison of compound selectivity metrics

5.5.1 General compound selectivity metric comparison

Since most of the existing compound selectivity metrics are not target-specific, we used the number of selective targets identified for each compound as a target-agnostic selectivity metric based on our target-specific selectivity approach to make a fair comparison with the other selectivity metrics. Different metrics may also have different ranges as well as different directions. Thus, for comparison, all the metrics were normalized to zero mean and unit standard deviation using the z-scaling:

$$z = \frac{x - \mu}{\sigma}$$

where x is the value of the original selectivity score, and μ and σ are the mean and standard deviation of the original selectivity scores, respectively. All the metrics were also normalized in direction, such that the smaller the value of the metrics, the more selective is the compound.

5.5.2 Target-specific compound selectivity comparison

As described in the original work (Cheng et al., 2010), partition index can be considered as a target-specific compound selectivity metric when choosing specific reference target. Thus, we calculated partition index for each compound-target pair separately as follows:

$$Partition\ index\ of\ \left(c_i, t_j\right) = \frac{\frac{1}{K_{c_i t_j}}}{\sum_{t_j} \frac{1}{K_{c_i t_j}}}$$

This calculation was then compared with our target-specific compound selectivity score calculated from the local and global relative potency. In negative logarithm form, the smaller the partition index, the more selective is the compound-target pair.

5.6 Software tools

Python programming language (version 3.7, https://www.python.org) was used for all the analyses. Python libraries Pandas (version 1.3.4) (McKinney and W, 2010; Reback et al., 2020) and Numpy (version 1.21.2) (Harris et al., 2020) were used for data processing and bi-objective optimization. Python libraries Matplotlib (version 3.5.1) (Hunter, 2007), Seaborn (version 0.11.0) (Waskom, 2021) and venn (0.1.3, https://pypi.org/project/venn/) were used for making the figures.

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found here: https://static-content.springer.com/esm/art%3A10.1038%2Fnbt.1990/MediaObjects/41587_2011_BFnbt1990_MOESM5_ESM.xls.

Author contributions

TW: developed and implemented the selectivity score, conducted experiments and analysis, prepared the figures and

drafted the manuscript. OP: designed the optimization approach and the selectivity score, and revised the manuscript. TA: conceived and supervised the work, drafted the manuscript, and provided funding support.

Funding

TW was supported by a salary grant from Doctoral Programme in Integrative Life Science. TA was supported by the Norwegian Cancer Society (grant 216104), Helse Sør-Øst (2020026), Radium Hospital Foundation, Finnish Cancer Foundation, and the Academy of Finland (grants 313267, 326238, 340141, 345803 and 344698). The open-access publication fees were covered by the University of Helsinki Library.

Acknowledgments

The authors thank Prof. Juho Rousu, Dr. Aik Choon Tan and Dr. Sandor Szedmak for many fruitful discussions on the drug and target selectivity analyses.

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fphar.2022.1003480/full#supplementary-material

References

Aittokallio, T., (2022). What are the current challenges for machine learning in drug discovery and repurposing? *Expert Opin. Drug Discov.* 17 (5), 423–425. doi:10.1080/17460441.2022.2050694

Attwood, M. M., Fabbro, D., Sokolov, A. V., Knapp, S., and Schiöth, H. B., (2021). Trends in kinase drug discovery: Targets, indications and inhibitor design. *Nat. Rev. Drug Discov.* 20 (11), 839–861. doi:10.1038/s41573-021-00252-y

Bagherian, M, Sabeti, E, Wang, K, Sartor, MA, Nikolovska-Coleska, Z, and Najarian, K (2020). Machine learning approaches and databases for prediction of drug–target interaction: A survey paper. *Brief. Bioinform.* 22 (1), 247–269. doi:10.1093/bib/bbz157

Bello, T, and Gujral, TS (2018). KInhibition: A kinase inhibitor selection portal. *iScience* 8, 49–53. doi:10.1016/j.isci.2018.09.009

Bora, A, Avram, S, Ciucanu, I, Raica, M, and Avram, S (2016). Predictive models for fast and effective profiling of kinase inhibitors. *J. Chem. Inf. Model.* 56 (5), 895–905. doi:10.1021/acs.jcim.5b00646

Bosc, N, Meyer, C, and Bonnet, P (2017). The use of novel selectivity metrics in kinase research. *BMC Bioinforma.* 18 (1), 17. doi:10.1186/s12859-016-1413-y

Cheng, AC, Eksterowicz, J, Geuns-Meyer, S, and Sun, Y (2010). Analysis of kinase inhibitor selectivity using a thermodynamics-based partition index. *J. Med. Chem.* 53 (11), 4502–4510. doi:10.1021/jm100301x

Cichońska, A, Ravikumar, B, Allaway, RJ, Wan, F, Park, S, Isayev, O, et al. (2021). Crowdsourced mapping of unexplored target space of kinase inhibitors. *Nat. Commun.* 12 (1), 3307. doi:10.1038/s41467-021-23165-1

Cohen, P, Cross, D, and Jänne, PA (2021). Kinase drug discovery 20 years after imatinib: Progress and future directions. *Nat. Rev. Drug Discov.* 20 (7), 551–569. doi:10.1038/s41573-021-00195-4

Davis, MI, Hunt, JP, Herrgard, S, Ciceri, P, Wodicka, LM, Pallares, G, et al. (2011). Comprehensive analysis of kinase inhibitor selectivity. *Nat. Biotechnol.* 29 (11), 1046–1051. doi:10.1038/nbt.1990

Graczyk, PP (2007). Gini coefficient: A new way to express selectivity of kinase inhibitors against a family of kinases. *J. Med. Chem.* 50 (23), 5773–5779. doi:10.1021/jm070562u

Haimes, Y (1971). On a bicriterion formulation of the problems of integrated system identification and system optimization. *IEEE Trans. Syst. man, Cybern.* 1 (3), 296–297.

Harris, CR, Millman, KJ, van der Walt, SJ, Gommers, R, Virtanen, P, Cournapeau, D, et al. (2020). Array programming with NumPy. *Nature* 585 (7825), 357–362. doi:10.1038/s41586-020-2649-2

Hunter, JD (2007). Matplotlib: A 2D graphics environment. *Comput. Sci. Eng.* 9 (3), 90–95. doi:10.1109/mcse.2007.55

Karaman, M. W., Herrgard, S., Treiber, D. K., Gallant, P., Atteridge, C. E., Campbell, B. T., et al. (2008). A quantitative analysis of kinase inhibitor selectivity. *Nat. Biotechnol.* 26 (1), 127–132. doi:10.1038/nbt1358

Lo, YC, Liu, T, Morrissey, KM, Kakiuchi-Kiyota, S, Johnson, AR, Broccatelli, F, et al. (2019). Computational analysis of kinase inhibitor selectivity using structural knowledge. *Bioinformatics* 35 (2), 235–242. doi:10.1093/bioinformatics/bty582

McKinney, W (2010). "Data structures for statistical computing in python," in Proceedings of the 9th Python in Science Conference, (June 28 - July 3, 2010: Austin, TX).

Merget, B, Turk, S, Eid, S, Rippmann, F, and Fulle, S (2017). Profiling prediction of kinase inhibitors: Toward the virtual assay. *J. Med. Chem.* 60 (1), 474–485. doi:10.1021/acs.jmedchem.6b01611

Miettinen, K (1999). *Nonlinear multiobjective optimization, vol. 12 of international series in operations research &. Management science*. Berlin, Germany: Springer.

Miljković, F, and Bajorath, J (2018). Data-driven exploration of selectivity and off-target activities of designated chemical probes. *Molecules* 23 (10), 2434. doi:10.3390/molecules23102434

Miljković, F, and Bajorath, J (2018). Exploring selectivity of multikinase inhibitors across the human kinome. *ACS Omega* 3 (1), 1147–1153. doi:10.1021/acsomega.7b01960

Nguyen, T, Le, H, Quinn, TP, Nguyen, T, Le, TD, and Venkatesh, S (2020). GraphDTA: Predicting drug–target binding affinity with graph neural networks. *Bioinformatics* 37 (8), 1140–1147. doi:10.1093/bioinformatics/btaa921

Öztürk, H, Özgür, A, and Ozkirimli, E (2018). DeepDTA: Deep drug–target binding affinity prediction. *Bioinformatics* 34 (17), i821–i829. doi:10.1093/bioinformatics/bty593

Pulkkinen, OI, Gautam, P, Mustonen, V, and Aittokallio, T (2021). Multiobjective optimization identifies cancer-selective combination therapies. *PLoS Comput. Biol.* 16 (12), e1008538. doi:10.1371/journal.pcbi.1008538

Reback, J, McKinney, W, Van Den Bossche, J, Augspurger, T, Cloud, P, Klein, A, et al. (2020). *pandas-dev/pandas*. Zenodo. Pandas 1.0. 5.

Schipper, LJ, Zeverijn, LJ, Garnett, MJ, and Voest, EE (2022). Can drug repurposing accelerate precision oncology? *Cancer Discov.* 12 (7), 1634–1641. doi:10.1158/2159-8290.CD-21-0612

Schneider, P, Walters, WP, Plowright, AT, Sieroka, N, Listgarten, J, Goodnow, RA, et al. (2020). Rethinking drug design in the artificial intelligence era. *Nat. Rev. Drug Discov.* 19 (5), 353–364. doi:10.1038/s41573-019-0050-3

Thafar, M., Raies, A. B., Albaradei, S., Essack, M., and Bajic, V. B., (2019). Comparison study of computational prediction tools for drug-target binding affinities. *Front. Chem.* 7, 782. doi:10.3389/fchem.2019.00782

Uitdehaag, J, and Zaman, GJ (2011). A theoretical entropy score as a single value to express inhibitor selectivity. *BMC Bioinforma.* 12 (1), 94–11. doi:10.1186/1471-2105-12-94

Uitdehaag, JC, Verkaar, F, Alwan, H, de Man, J, Buijsman, RC, and Zaman, GJ (2012). A guide to picking the most selective kinase inhibitor tool compounds for pharmacological validation of drug targets. *Br. J. Pharmacol.* 166 (3), 858–876. doi:10.1111/j.1476-5381.2012.01859.x

Ursu, A, Childs-Disney, JL, Angelbello, AJ, Costales, MG, Meyer, SM, and Disney, MD (2020). Gini coefficients as a single value metric to define chemical probe selectivity. *ACS Chem. Biol.* 15 (8), 2031–2040. doi:10.1021/acschembio.0c00486

Vamathevan, J, Clark, D, Czodrowski, P, Dunham, I, Ferran, E, Lee, G, et al. (2019). Applications of machine learning in drug discovery and development. *Nat. Rev. Drug Discov.* 18 (6), 463–477. doi:10.1038/s41573-019-0024-5

Wang, T, Gautam, P, Rousu, J, and Aittokallio, T (2020). Systematic mapping of cancer cell target dependencies using high-throughput drug screening in triple-negative breast cancer. *Comput. Struct. Biotechnol. J.* 18, 3819–3832. doi:10.1016/j.csbj.2020.11.001

Waskom, ML (2021). Seaborn: Statistical data visualization. *J. Open Source Softw.* 6 (60), 3021. doi:10.21105/joss.03021

Ye, Q, Hsieh, C-Y, Yang, Z, Kang, Y, Chen, J, Cao, D, et al. (2021). A unified drug–target interaction prediction framework based on knowledge graph and recommendation system. *Nat. Commun.* 12 (1), 6775. doi:10.1038/s41467-021-27137-3

frontiers | Frontiers in Pharmacology

TYPE Review
PUBLISHED 26 September 2022
DOI 10.3389/fphar.2022.974849

OPEN ACCESS

EDITED BY
Mithun Rudrapal,
Rasiklal M. Dhariwal Institute of Pharmaceutical Education and Research, India

REVIEWED BY
Luca Pinzi,
University of Modena and Reggio Emilia, Italy
Kandi Sridhar,
Agrocampus Ouest, France

*CORRESPONDENCE
Yuehua Wan,
wanyuehua@zjut.edu.cn

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

SPECIALTY SECTION
This article was submitted to Drugs Outcomes Research and Policies, a section of the journal Frontiers in Pharmacology

RECEIVED 21 June 2022
ACCEPTED 12 August 2022
PUBLISHED 26 September 2022

CITATION
Sun G, Dong D, Dong Z, Zhang Q, Fang H, Wang C, Zhang S, Wu S, Dong Y and Wan Y (2022), Drug repositioning: A bibliometric analysis.
Front. Pharmacol. 13:974849.
doi: 10.3389/fphar.2022.974849

Drug repositioning: A bibliometric analysis

Guojun Sun[1†], Dashun Dong[1†], Zuojun Dong[1†], Qian Zhang[1], Hui Fang[2], Chaojun Wang[3], Shaoya Zhang[1], Shuaijun Wu[1], Yichen Dong[4] and Yuehua Wan[2*]

[1]Institute of Pharmaceutical Preparations, Department of Pharmacy, Zhejiang University of Technology, Hangzhou, China, [2]Institute of Information Resource, Zhejiang University of Technology, Hangzhou, China, [3]Hangzhou Aeronautical Sanatorium for Special Service of Chinese Air Force, Hangzhou, China, [4]Faculty of Chinese Medicine, Macau University of Science and Technology, Macau, China

Drug repurposing has become an effective approach to drug discovery, as it offers a new way to explore drugs. Based on the Science Citation Index Expanded (SCI-E) and Social Sciences Citation Index (SSCI) databases of the Web of Science core collection, this study presents a bibliometric analysis of drug repurposing publications from 2010 to 2020. Data were cleaned, mined, and visualized using Derwent Data Analyzer (DDA) software. An overview of the history and development trend of the number of publications, major journals, major countries, major institutions, author keywords, major contributors, and major research fields is provided. There were 2,978 publications included in the study. The findings show that the United States leads in this area of research, followed by China, the United Kingdom, and India. The Chinese Academy of Science published the most research studies, and NIH ranked first on the h-index. The Icahn School of Medicine at Mt Sinai leads in the average number of citations per study. Sci Rep, Drug Discov. Today, and Brief. Bioinform. are the three most productive journals evaluated from three separate perspectives, and pharmacology and pharmacy are unquestionably the most commonly used subject categories. Cheng, FX; Mucke, HAM; and Butte, AJ are the top 20 most prolific and influential authors. Keyword analysis shows that in recent years, most research has focused on drug discovery/drug development, COVID-19/SARS-CoV-2/coronavirus, molecular docking, virtual screening, cancer, and other research areas. The hotspots have changed in recent years, with COVID-19/SARS-CoV-2/coronavirus being the most popular topic for current drug repurposing research.

KEYWORDS

drug repurposing, bibliometrics, drug development, COVID-19, virtual screening

1 Introduction

Sir James Black, a winner of the 1988 Nobel Prize, clearly recognized well before the 21st century that drug repurposing strategies would occupy an important place in the future of new drug discovery (Raju, 2000). In 2004, Ted T. Ashburn et al. (Ashburn and Thor, 2004) summarized previous research and developed a general approach to drug development using

drug repurposing, retrospectively looking for new indications for approved drugs and molecules that are waiting for approval for new pathways of action and targets. These molecules are usually safe in clinical trials but do not show sufficient efficacy for the treatment of the disease originally targeted (Southan et al., 2013). The definition of the term "drug repurposing" has been endorsed by scholars (Dudley et al., 2011) and used by them (Li et al., 2011; Cheng et al., 2012). It should be pointed out that the synonyms of "drug repurposing" often used by academics also include drug repositioning (Rosa and Santos, 2020), drug rediscovery (Simsek et al., 2018), drug redirecting (Jang et al., 2019), drug retasking (Scherman and Fetro, 2020), and therapeutic switching (Kim et al., 2019; Kurdi et al., 2019). After the research study by Ashburn et al., Allarakhia et al. expanded the starting materials for drug repositioning to include products that were discontinued for commercial reasons, expired patents, and candidates for laboratory testing (Allarakhia, 2013). In the discovery process of a completely new drug, the difficulty usually lies in its safety and efficacy, which are the main potential causes of failure of most drugs in the approval (Schuster et al., 2005) or clinical development stage (Milne, 2017). Using existing knowledge about a drug or known target (Mercorelli et al., 2018), the time, risk, and cost of developing a drug using drug repositioning are reduced (Joshua, 2011), thereby greatly increasing the efficiency and economics of drug development, providing a better risk–reward trade-off, and making it easier to win the favor of venture capital firms (Ashburn and Thor, 2004).

Since the 1990s, the repositioning of sildenafil for male erectile dysfunction (Goldstein et al., 1998) and pulmonary hypertension (Badesch et al., 2007), the development of a new efficacy of bupropion for smoking cessation (Hurt et al., 1997), new applications of thalidomide for multiple myeloma (Singhal et al., 1999; Barlogie, 2001), and chronic graft-versus-host disease (Vogelsang et al., 1992) have generated intense interest from pharmaceutical companies and academics (Kumar et al., 2019). These classic success stories rely on three traditional approaches: 1) molecular biology approaches (Pujol et al., 2010), 2) *in vivo* and *ex vivo* experimental approaches (Kuter, 2007; Swinney and Anthony, 2011), and 3) expert knowledge-based approaches (Kumar et al., 2019). Due to the unknown, complex, and information-fragmented nature of drug candidates and potential new mechanisms of action (Yella et al., 2018), this activity is dependent on multiple factors, and success is often fortuitous (Kumar et al., 2019). At the beginning of the 21st century, cheminformatics (Feng et al., 2007; Joshua, 2011), bioinformatics (Salazar et al., 2006; Feng et al., 2021), systems biology (Lv et al., 2018; Turanli et al., 2021), genomics (Zhao et al., 2016; Mirza et al., 2017), polypharmacology (Reddy and Zhang, 2013; Anighohro et al., 2014), precision medicine (Delavan et al., 2018; Tanoli et al., 2020), and other disciplines, combined with artificial intelligence (Yang et al., 2019), have developed rapidly. These rapidly growing disciplines have promoted the generation of systematic (Talevi and Bellera, 2020) computer methods to make the drug repositioning process cheaper and shorter (Vanhaelen et al., 2017; Luo et al., 2021). Computational drug repositioning is classified as "disease-centric" or "target/gene-centric" or "drug-centric" depending on the source of discovery (Li et al., 2016). This process relies on public biochemical databases such as DrugBank (Mihai et al., 2019; Mazzolari et al., 2020), ChEMBL (Mendez et al., 2019), Cmap (Lin et al., 2020), PDB (Berman et al., 2000), OMIM (Amberger et al., 2014), etc., to provide the appropriate information. In fact, to make the computational drug repurposing process, including the molecular docking and virtual screening steps, more convenient, database tools specifically developed for drug repurposing, such as EK-DRD (Zhao et al., 2019), DREIMT (Troulé et al., 2021), DrugSig (Wu H. et al., 2017), RepoDB (Malas et al., 2019), Promiscuous 2.0 (Gallo et al., 2021), etc., have been reported in the last few years. In addition, it has been found in the literature that only 10% of the research results have been carried out in the "drug-centric" pathway, which holds great prospects for future development (Parisi et al., 2020). With the help of database tools, it is now possible to perform computational screening of even a staggering number of hundreds of millions of compounds (Fischer et al., 2020). Computer methods to carry out this screening include machine learning (Napolitano et al., 2013), network modeling (Francisco, 2013; Lotfi Shahreza et al., 2018), text mining, and semantic reasoning (Christos et al., 2011; Yuan et al., 2017; Ji et al., 2020), among others. The ultimate objective of repositioning is to transfer one or two of the most relevant results to clinical applications. Therefore, validation is quite important (Li et al., 2016) and requires consideration of multiple factors, such as price, toxicity levels, bioavailability, and differences between validated and computational models (Li et al., 2016; Jarada et al., 2020). Current validation methods include experimental validation (Kang et al., 2014), electronic health records to aid validation (Xu et al., 2015), cross-validation (Wu Z. et al., 2017; Ozsoy et al., 2018), gold standard dataset evaluation (Luo et al., 2021), literature citation validation (Chopra et al., 2016), and expert consultation (Jarada et al., 2020).

Today, drug repositioning is increasingly prominent in the development of drugs for a variety of neurological diseases (Athauda and Foltynie, 2018; Kessing et al., 2019), cancer (Gupta et al., 2013; Efferth, 2017), rare diseases (Sardana et al., 2011; Southall et al., 2019), and infectious diseases (Pietschmann, 2017; Muratov et al., 2021). An increasing number of pharmaceutical companies are also establishing relevant R&D programs (Kettle and Wilson, 2016) or funding support (Tummino et al., 2021). To translate relevant research results efficiently and smoothly, national departments within the United Kingdom, the United States, and the Netherlands have (Paul and Lewis-Hall, 2013; Vanhaelen et al., 2017) launched initiatives or programs to build partnerships between pharmaceutical companies and academia and to further explore scientific and commercial opportunities (Yella et al., 2018). It is certain that drug repositioning currently presents several dilemmas, such as intellectual property challenges (Breckenridge and Jacob, 2019), data platforms, and analytical

techniques that need to be improved (Kumar et al., 2019), that financial support remains important for technology development and clinical trials (Verbaanderd et al., 2021), and that some scientists deny the practical utility of the approach (Edwards, 2020).

There have been systematic analyses of terminology in the drug repurposing literature (Langedijk et al., 2015), text mining of drug–disease combinations (Baker et al., 2018), and the progression of a particular drug (Li X. et al., 2020), but no studies have yet provided a broad overview of publications on the topic of drug repurposing research. When independent researchers or collectives (including pharmaceutical companies, academia, and government departments) seek drug repurposing partnership partners and seek to obtain a concise overview of comprehensive current research hotspots, the lack of relevant intelligence analysis to aid decision-making often makes the process convoluted and time-consuming (Frail et al., 2015). The bibliometric approach can solve the aforementioned problems relatively fairly, but at present, scholars have only studied the bibliometrics of aspirin, a drug repurposing (Li X. et al., 2020); there has not been a panoramic study of drug repurposing, and therefore, this study is necessary. Bibliometrics is a useful tool combining multiple parameters for the quantitative analysis of scholarly publications and is currently used to assess research hotspots and trends in a wide range of disciplines and industries, such as management (Vogel and Güttel, 2013; Feng et al., 2017), sociology (Rey-Martí et al., 2016; Sharifi, 2021), economics (Zhang et al., 2019), medicine (Tao et al., 2012; Powell et al., 2016), environmental engineering (Colares et al., 2020; Mao et al., 2021), and agronomy (Canas-Guerrero et al., 2013; Giraldo et al., 2019). Therefore, this study uses bibliometric methods (Leung et al., 2017) to quantitatively assess the following elements of drug repositioning publications: 1) major contributors: countries, research institutions, and authors; 2) modes of collaboration: intercountry collaborations; 3) the most productive journals; 4) the most frequently used disciplinary knowledge; and 5) research trends, judged by analyzing author keywords, Essential Science Indicators (ESI) high citations, and hot research studies.

2 Methodology and data processing

2.1 Data collection

We use the Web of Science™ core database, an authoritative academic information data service platform produced by Clarivate (version © 2021 Clarivate.). Due to its rigorous selection of journals, the Web of Science (WOS) Core Collection Database is now internationally recognized as a database for evaluating the scientific output or disciplinary development of scholars and institutions. Among the subdatabases, SCI-E mainly includes global journals in basic science research, covering basic pharmacological and medical research related to the theme of this study, "drug repositioning," while SSCI includes social science, covering ethical, nursing, psychological, and other social science research related to this study.

The data were obtained on 25 October 2021 through the WOS Core Collection Database Citation Indexes SCI-E and SSCI, using the formula "drug repurposing" OR "drug repositioning" OR "drug rediscovery" OR "therapeutic switching" OR "drug redirecting" OR "drug rediscovery" OR "drug retasking" search query, searching in the "subject" field and defining the document type as "Article" and "Review". The publication time parameters were initially limited to publications related to "drug repositioning" published between 1990 and 2020. A total of 3,009 documents were obtained, of which only 31 were published in two decades from 1990 to 2009. Of these 31 documents, except for one document that is still frequently used by scholars as a retrospective source for drug repurposing definitions in these years (Ashburn and Thor, 2004), the remaining 30 were cited by other authors during the period of 2010–2020 as shown in Figure 1. The overall level of interest in these studies shows a fluctuating downward trend as opposed to the rising citation fervor for drug repurposing, entering a stage of decline even under the less-demanding evaluation criterion of a 5-year maturation window (Jacsó, 2009). As the literature ages, its content becomes stale and obsolete in the perspective of intelligence sources, and the value of the metrics for judging current research trends is low. Therefore, we further narrowed the study to 2,978 publications published from 2010 to 2020.

2.2 Data import and deduplication

The complete records of all retrieved documents are downloaded and imported for processing into Derwent Data Analyzer (DDA) version V10, a data cleaning, multiperspective data mining, and visualization software from Clarivate that improves data analysis efficiency and reduces labor costs. After importing all records of WOS documents into DDA, they are classified and measured according to a list of fields such as keyword, country/region, institution, author, research field, journal, etc. For each item in the list fieldset, DDA has a built-in data cleaning tool for automatic data deduplication.

2.3 Data splitting or merging

After the machine has removed duplicates, the items in the set of fields still need to be manually verified for splitting or merging. It is to be noted that the regions of certain countries are presented separately, while they are usually considered as a single country internationally. Therefore, we need to perform merging, such as combining Wales, Scotland, England, and Northern

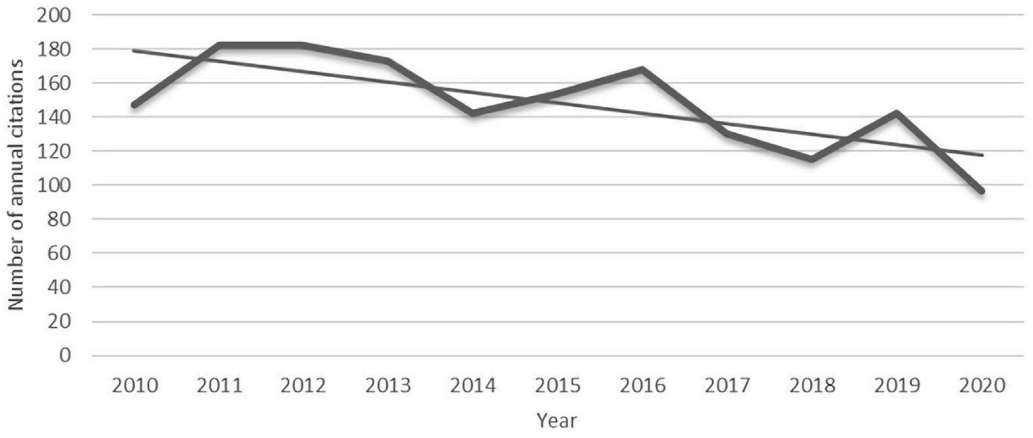


FIGURE 1
Total number of citations per year from 2010 to 2020 for 30 publications published from 1990 to 2009.

Ireland into the United Kingdom column and combining Hong Kong and Macau regions into the China column.

To address the possible problem of different authors with the same name, the following two main verification steps were performed: 1) returning to the WOS database to search for publications under that author's name under the original search formula conditions and 2) for authors whose publications provide disputed information (this also includes three cases: first, two or more authors with the same name but not the same person; second, two or more authors with the same abbreviated name, but the full names were found to be different after a search; and third, similar signatures being different variants of the same author's name), in addition to searching the ORCID-related information of the authors concerned for judgment, we checked different institutional websites as well as encyclopedias to look for changes in the study and work history of authors with the same or similar names from 2010 to 2020 to determine whether they were the same person. Based on the verification, we then split or merged the results.

2.4 Data analysis and visualization

After data cleaning and matrix analysis by DDA, various types of cluster plots and bubble plots can be obtained to reveal the useful information behind the data. The bibliometric fields of publication volumes, countries, international collaborations, institutions, research areas, journals, authors, highly cited research studies, and author keywords were analyzed in this study. It should be noted that because some studies were published online ahead of time and the study publication date was a year or two behind, for statistical purposes, the year of publication of such research studies was included as the year of online publication. (e.g., a study shown in the reference as published in 2022 may have been published online in 2020).

3 Results

3.1 Number and type of publications

Of the 2,978 papers obtained using the search criteria mentioned previously, the main ones were research studies (2248; 75.49%) and reviews (730; 24.51%). Furthermore, individual publications are not only classified by journals in the single category of research studies or reviews but also belong to other categories. These publications were also related to proceeding studies (68; 2.28%), early access (24; 0.81%), book chapters (7; 0.24%), data studies (2; 0.07%), and retracted publications (2; 0.07%). The vast majority of research studies and reviews were published in English (2967; 99.631%), with the remainder in Japanese (3; 0.101%), Chinese (2; 0.067%), Czech (1; 0.034%), French (1; 0.034%), German (1; 0.034%), Hungarian (1; 0.034%), Korean (1; 0.034%), and Portuguese (1; 0.034%). Ninety were from SSCI, and the remaining 2888 were from SCI-E. Further, 1,996 were from Open Access. An annual analysis of published research studies is shown in Figure 2. The number of publications for every year expanded from 17 in 2010 to 970 in 2020. Annual publications on the subject have increased by more than 64 times. The number of annual publications has been increasing at a relatively high rate since 2015, while in 2020, there was a spike in the number of publications and annual citations, probably due to the

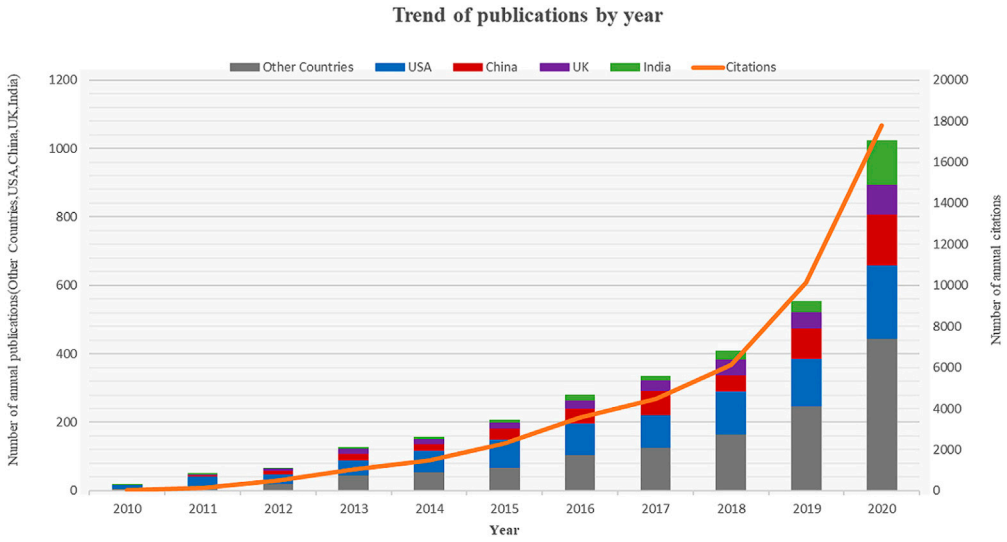


FIGURE 2
Annual trends in the number of articles published and citations related to drug repositioning.

TABLE 1 Top 20 most productive countries/regions in the field of drug repositioning.

Rank	Country	TP	TC	h-index	ACPP	nCC	SMCP (%)	Region
1	The United States	918	27,355	74	29.8	59	48.15	Anglo-America
2	P.R. China	485	11,147	49	22.98	39	36.70	Asia
3	The United Kingdom	284	8,762	43	30.85	57	69.01	Europe
4	India	247	3,203	27	12.97	37	30.77	Asia
5	Italy	232	6,024	39	25.97	40	47.41	Europe
6	Germany	171	5,213	36	30.49	50	67.25	Europe
7	South Korea	161	2221	21	13.8	24	29.20	Asia
8	Japan	146	3,037	26	20.8	22	25.34	Asia
9	Brazil	125	1911	24	15.29	29	42.20	Latin America
10	France	116	3,627	26	31.27	35	56.03	Europe
11	Canada	111	4,641	28	41.81	46	62.16	Anglo-America
12	Spain	109	2305	27	21.15	38	58.72	Europe
13	Australia	73	1816	23	24.88	36	79.45	Oceania
14	The Netherlands	73	1,559	22	21.36	37	75.34	Europe
15	Switzerland	59	2126	23	36.03	32	67.80	Europe
16	Sweden	58	1,434	19	24.72	37	86.21	Europe
17	Taiwan Region	58	1,110	17	19.14	8	36.21	Asia
18	Argentina	51	749	17	14.69	16	43.14	Latin America
19	Belgium	48	1,062	18	22.13	26	81.25	Europe
20	Mexico	47	1,162	19	24.85	15	42.55	Latin America

Notes: TP, total papers; TC, total citations; ACPP, average citations per publication; nCC, number of cooperative countries; and SMCP, share of multinational cooperation publications.

COVID-19 pandemic, a global public health emergency that prompted special attention from scientists. Among the four countries with the highest number of publications (the United States, China, the United Kingdom, and India), the United States has maintained a high growth volume since 2010, while China was the fastest in terms of average annual growth

rate in the last three years. In 2020, the number of publications in India surged and surpassed the production of the United Kingdom.

3.2 Countries and number of publications

With respect to the 2978 publications related to drug repositioning research, 89 countries contributed to the field of drug repositioning research. The number of publications and citations from the 20 most productive countries/regions is shown in Table 1. There are nine countries/regions in Europe, five in the Americas, five in Asia, and one in Oceania.

The four most productive countries/regions are, in order, the United States, China, the United Kingdom, and India. The United States is the absolute leader in this field, with 918 research studies on drug repositioning published since 2010, which is already more than the next highest number of publications in China and the United Kingdom combined. This is followed by India (247), Italy (232), Germany (171), South Korea (161), and Japan (146). Other productive countries include Brazil (125), France (116), Canada (111), Spain (109), Australia (73), the Netherlands (73), and Switzerland (59). In terms of publication impact, the United States led the Total citations (TC) rankings with 27,355, twice as many as that of China (11,147), which ranked second. We also included the average citations per publication (ACPP) in the comparison, which is calculated by dividing the TC by the TP (total papers) value and is a relative number that may better reflect the individual or collective level of attention than the individual TC and TP values. Canada ranked first in ACPP at 41.81, closely followed by the United Kingdom (30.85) and Germany (30.49). In addition, the h-index was originally proposed as a simple quantification that a researcher had at least h publications cited h times, reflecting to a certain extent the research results of the researcher as an individual (Hirsch, 2005). Later, the word "researcher" in the definition began to be replaced by collective words such as "academic group or institution (Van Raan, 2006)," "journal (Braun et al., 2006)," and "country (Csajbók et al., 2007)," becoming an indicator of the level of collective research to some extent. Undoubtedly, the h-index of the United States ranks first in this field with 74 times. Taking all parameters into account, we find that publications in the United Kingdom, the United States, and Canada perform better on average. While the number of publications in China and India is significant, they have received low levels of attention.

3.3 National/regional cooperation

It should be noted that DDA analysis software is nationally identified based on the location of each researcher's institution address provided in the publication. If a publication is coauthored by institutions from more than two countries, the publication is defined as the result of an international collaboration. Whether there is some affiliation between the various institutions of the research group that produces the multicountry collaboration is not taken into account. As shown in Table 1, among the publications of the top 20 countries and regions, the proportion of international collaborations is quite high in European countries, especially in Sweden (86.21%) and Belgium (81.25%). Asian and Latin American countries are generally underrepresented. In addition, the United States, the most active country in publishing and the country with the most collaborations—with 59 countries or regions—still has over 50% of the studies published overall.

Figure 3 depicts the academic collaboration network for the top 20 countries and regions in terms of productivity. Using DDA software, the network was mapped using a co-occurrence matrix. The size of the circles is proportional to the extent of each country's contribution, the lines between the circles represent the collaboration between countries/regions, and the thickness of the connecting lines indicates the frequency of collaboration (Bao et al., 2018). The results show that the United States cooperates most frequently with China and the United Kingdom and has the closest cooperation with them. In addition, Mexico, Belgium, Argentina, Taiwan, Japan, and Korea have slightly sparser cooperation networks among the 20 most productive countries/regions, while the remaining countries have more extensive cooperation networks among themselves.

3.4 Contributions of leading bodies

A total of 3,530 institutes were involved in drug repositioning research. The top 20 productive institutes are shown in Table 2. Eight of the top 20 institutions are from the United States, again indicating the dominance of the United States in drug repositioning research; three are from the United Kingdom; two are from China; and Brazil, France, Mexico, Canada, the Netherlands, Austria, and Sweden each have one. The Chinese Academy of Science ranked first in terms of the number of research studies, followed by Case Western Reserve University and the NIH. The Icahn School of Medicine at Mt Sinai ranked first in ACPP at 77.32. The NIH had the highest h-index value of 22. The Icahn School of Medicine at Mt Sinai was the best performer in ACPP at 77.32, followed by the University of California, San Francisco (67.82) and Johns Hopkins University (65.68), both of which are US-based research institutions. Compared with US research institutions, Chinese research institutions are at the back of the pack in terms of ACPP, and their impact needs to be improved.

The collaboration network between the 15 largest institutions in 2010–2020 is shown in Figure 4. The collaboration network provides a more visual view of the collaboration with different institutions and thus helps in the search for more beneficial collaborations. Next to the name of each institution is its total

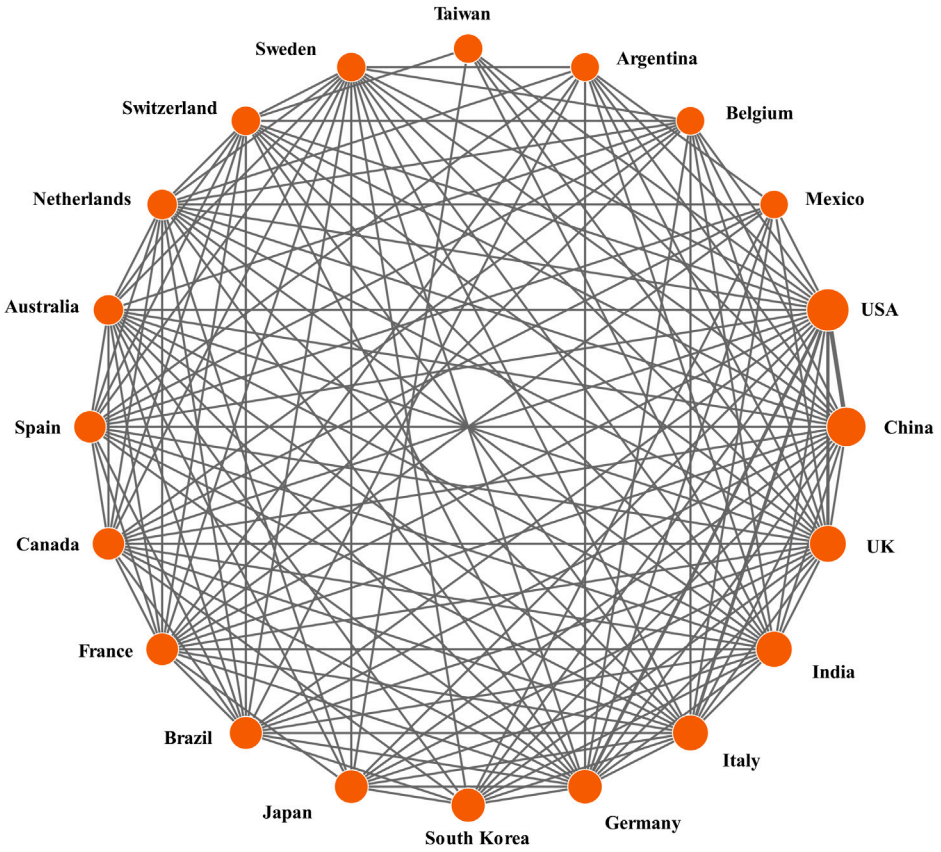


FIGURE 3
Cooperation between the top 20 most efficient countries/regions.

number of publications. At the intersections of these institutions, yellow dots indicate collaborations with the other top 10 research institutions. It should be noted that the number of yellow dots can indicate the output of cooperation and the strength of interagency cooperation. The nodal data with no crossover points represent the number of publications produced by the institute, either by its independent work or in collaboration with research institutions outside the top 15 (Bao et al., 2019). From Figure 4, we see that the University of Cambridge established the largest collaborative network, followed by the large network established by four institutions, the NIH, the Icahn School of Medicine at Mt Sinai, Karolinska Institute, and King's College London. In terms of the number of copublications with established institutions, the Chinese Academy of Science and Shanghai Jiao Tong University copublished as many as six, followed by the University of Cambridge and King's College London and the NIH and the Icahn School of Medicine at Mt Sinai. Analyzing the aforementioned three pairs of institutional combinations, King's College London has two publications that are the product of collaboration between the three research institutions. The University of Sao Paulo and Aix-Marseille University are relatively independent in this research area. Combining the ranking of multiple parameters, we found that the NIH and Icahn Sch Med Mt Sinai in the United States are the most vocal institutions in terms of academic research result perspective on the topic.

3.5 Contribution of leading research areas

An analysis of research areas gives a good indication of the scope of application of the research topic, with an unrestrained number of 74 areas covered, with the top 20 areas of research in terms of publication volume shown in Table 3. Briefly, "pharmacology and pharmacy" took first place with 962 articles, followed by "biochemistry and molecular biology", and for ACPP, the top three were science and technology-other topics (36.1)", "mathematics (32.79)", and "cell biology (29.65)".

TABLE 2 Top 20 most productive institutions in the field of drug repositioning for the period of 2010–2020.

Rank	Institution	TP	TC	ACPP	h-Index	PMCP (%)	Country/region
1	Chinese Acad Sci	54	1,286	23.81	19	98.15	China/Asia
2	Case Western Reserve Univ	38	1799	47.34	20	86.84	The United States/Anglo-America
3	NIH	37	1777	48.03	22	72.97	The United States/Anglo-America
4	Stanford Univ	35	1,401	40.03	16	80.00	The United States/Anglo-America
5	Univ Sao Paulo	34	452	13.29	13	76.47	Brazil/Latin America
6	Harvard Med Sch	33	1,078	32.67	18	84.85	The United States/Anglo-America
7	Univ Cambridge	32	788	24.63	14	90.63	The United Kingdom/Europe
8	Icahn Sch Med Mt Sinai	28	2165	77.32	15	75.00	The United States/Anglo-America
9	Kings Coll London	28	605	21.61	13	96.43	The United Kingdom/Europe
10	Aix Marseille Univ	27	1,183	43.81	15	92.59	France/Europe
11	Univ Nacl Autonoma Mexico	27	943	34.93	17	88.89	Mexico/Latin America
12	Shanghai Jiao Tong Univ	25	524	20.96	13	76.00	China/Asia
13	Univ Toronto	24	457	19.04	11	95.83	Canada/Anglo-America
14	Karolinska Inst	23	708	30.78	10	100.00	Sweden/Europe
15	Leiden Univ	23	327	14.22	11	78.26	The Netherlands/Europe
16	UCL	23	584	25.39	14	95.65	The United Kingdom/Europe
17	HM Pharma Consultancy	22	23	1.05	2	4.55	Austria/Europe
18	Johns Hopkins Univ	22	1,445	65.68	16	95.45	The United States/Anglo-America
19	NCI	22	602	27.36	14	100.00	The United States/Anglo-America
20	Univ Calif San Francisco	22	1,492	67.82	13	86.36	The United States/Anglo-America

Notes: TP, total papers; TC, total citations; ACPP, average citations per publication; and PMCP, Proportion of multi-institutional collaborative publications.

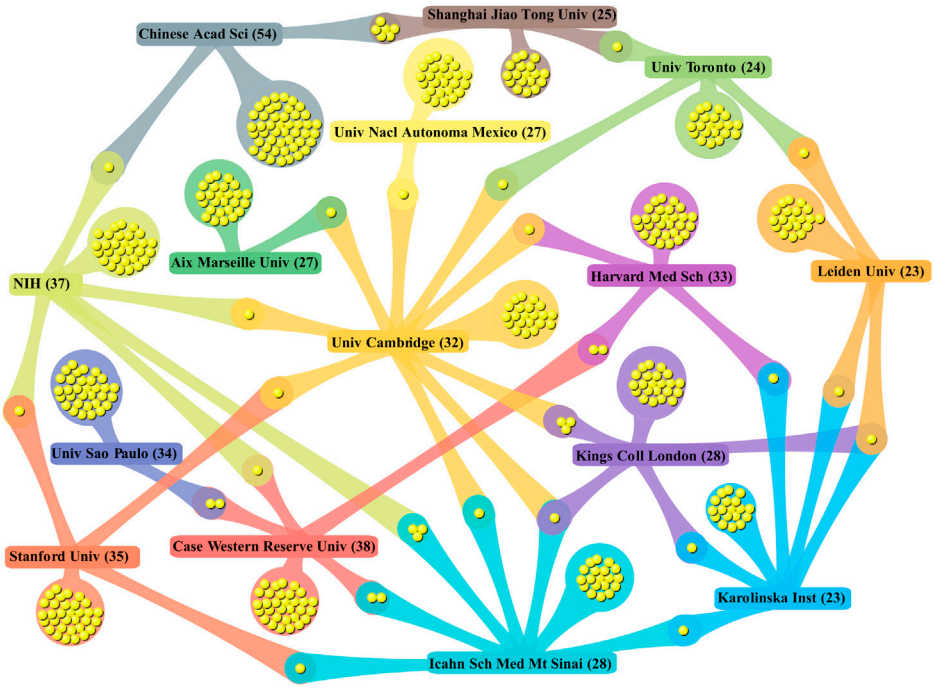


FIGURE 4
Collaboration matrix mapped between the first 15 productive bodies.

TABLE 3 Contribution of the top 20 research areas in the field of drug repositioning.

Rank	Research Area	TP	TC	ACPP	h-Index	SP%
1	Pharmacology & Pharmacy	962	25,243	26.24	67	32.3
2	Biochemistry & Molecular Biology	721	18,768	26.03	59	24.21
3	Oncology	302	7,104	23.52	40	10.14
4	Chemistry	274	5,539	20.22	33	9.2
5	Mathematical & Computational Biology	242	6,671	27.57	40	8.13
6	Science & Technology-Other Topics	234	8,448	36.1	42	7.86
7	Computer science	215	5,392	25.08	38	7.22
8	Biotechnology & Applied Microbiology	189	5,384	28.49	36	6.35
9	Cell biology	185	5,486	29.65	34	6.21
10	Research & Experimental Medicine	157	4,322	27.53	31	5.27
11	Microbiology	151	3,714	24.6	32	5.07
12	Neurosciences & Neurology	136	2513	18.48	26	4.57
13	Biophysics	114	2071	18.17	25	3.83
14	Genetics & Heredity	94	1878	19.98	23	3.16
15	Infectious diseases	93	2603	27.99	28	3.12
16	Immunology	68	1744	25.65	22	2.28
17	Mathematics	66	2164	32.79	27	2.22
18	General & Internal Medicine	64	1,299	20.3	21	2.15
19	Parasitology	56	1,079	19.27	18	1.88
20	Virology	56	1,079	19.27	18	1.88

Notes: TP, total papers; TC, total citations; ACPP, average citations per publication; and SP%, share of publications.

Figure 5 shows a bubble graph of the top 20 drug repositioning research areas. The bubble plot shows three dimensions of the data, namely, research area, year of publication, and the number of publications. The horizontal change in bubble size illustrates the growing trend of research areas over time, the vertical size of the bubble shows the most popular research areas in that year, and the number in the bubble indicates the frequency of the topic in the research area and the number of publications in that year (Chen et al., 2016). The number of research results in each relevant field is increasing year by year. Biophysics increased from five in 2019 to 77 in 2020, a more than 15-fold increase, suggesting that drug repositioning may have made a breakthrough or become widely used in this field. The field of virology was in a downturn from 2010 to 2014, with only one publication, with a gradual increase in relevant studies after 2015.

3.6 Contribution of major journals

For scholars studying drug repositioning-related topics, knowing which journals publish relevant research is important in deciding which journals to read or submit their research studies to. A total of 2,988 publications related to drug repositioning research were published in 845 journals during the period of 2010–2020. The top 25 journals in terms of a total number of studies published are shown in Table 4 Sci Rep topped the list with 75 studies published, followed by *PLoS One* (73; 2.52%), *J. Biomol. Struct. Dyn* (67; 2.45%), *Bioinformatics* (53; 2.25%), and *BMC bioinformatics* (50 articles; 1.78%). The rest of the journals had a share of less than 1.5%. In terms of total citations (TC), at present, studies in *Drug Discov.* have been cited a total of 2,119 times over the past 10 years, followed in rank by PLoS One (1800) and Bioinformatics (1,677). For ACPP, *Drug Discov. Today* still holds first place with a high frequency of 50.45 times, followed by *PLoS Comput* (33.14 times). The impact factor (IF) of a journal is calculated by dividing the total number of citations of all publications in the journal in the previous two years by the number of publications (Garfield, 2006). Thus, Table 4 shows that the ACPP of drug repurposing publications included in most journals is much higher than that of IF, which roughly verifies that the number of scholars interested in drug repurposing is relatively high. In terms of the impact factor (IF) of specific journals, except for *Oncotarget* and B*MC Syst. Biol.*, which have not been included in SCI since 2018 and 2020, *Brief. Bioinform.* has the highest value of 11.622, followed by *Drug Discov. Today* (7.851), *Bioinformatics* (6.937), *Cancers* (6.639), *Eur. J. Med. Chem* (6.514), and *Expert. Opin. Drug Discov.* (6.098). The bubble chart shows that *J. Biomol. Struct. Dyn.* featured 64 publications in 2020, compared to a combined total of only four publications in the previous ten years; the Oncotarget journal inclusion in this category peaked in 2016–2017 (Figure 6).

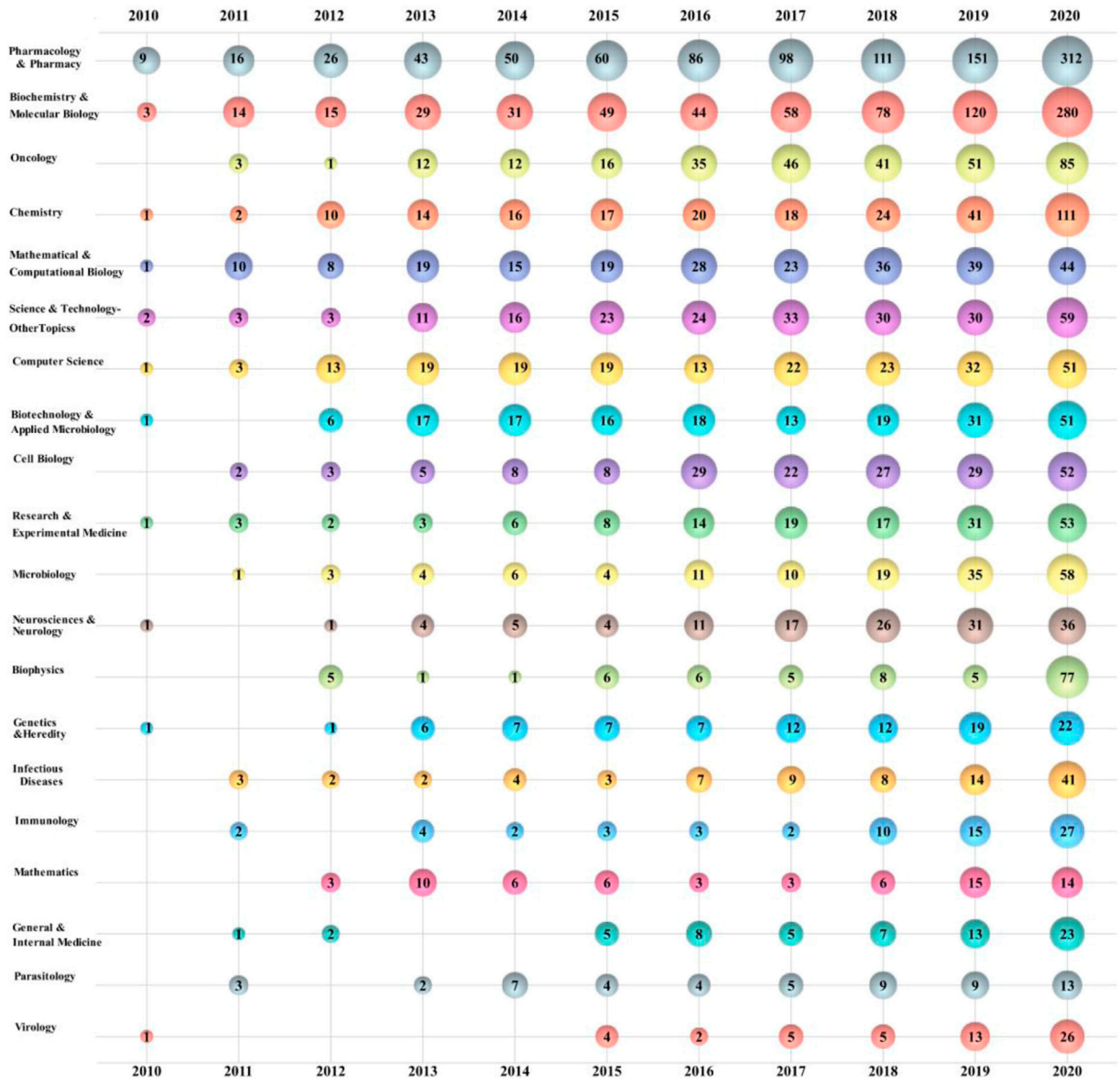


FIGURE 5
Bubble chart of the top 20 drug repositioning research areas by year.

3.7 Contribution of the lead author

For scholars interested in the topic of drug repositioning, it is useful to know how other researchers are working on the issue to facilitate communication and collaboration between scholars. A total of 15,620 authors contributed to studies within our measurement consideration, and Table 5 shows the top 20 prolific authors by a number of publications. Of these 20 highly productive authors, seven were from the United States, three were from Argentina, and two were from Germany, indicating a relatively high concentration of drug repositioning research in certain countries. In addition, the NIH (United States), Case Western Reserve University (United States), Tech University Dresden (Germany), and the National University of La Plata (Argentina) each have two of these academics.

Cheng, FX leads the list with 25 research studies, followed by Talevi, A (23) and Mucke, HAM (22). For the list of corresponding authors, the top three remain, in order, Mucke, HAM (22), Cheng, FX (17), and Talevi, A (17). In terms of ACPP ranking, Butte, AJ was ranked first with 154.33 points, followed by Cheng, FX (100.56), Tang, Y (95), and Dudley, JT (82). Cheng,

TABLE 4 Top 25 journals publishing studies in drug repositioning studies.

Rank	*Journal Title*	TP	TC	ACPP	IF (2020)
1	*Sci Rep*	75	1,081	14.41	4.38
2	*PLoS One*	73	1800	24.66	3.24
3	*J. Biomol. Struct. Dyn*	67	1,000	14.93	3.110
4	*Bioinformatics*	53	1,677	31.64	6.937
5	*BMC Bioinformatics*	50	658	13.16	3.169
6	*Front. Pharmacol*	43	1,073	24.95	5.811
7	*Drug Discov. Today*	42	2119	50.45	7.851
8	*Molecules*	40	329	8.23	4.412
9	*ASSAY DRUG DEV. TECHNOL.*	39	224	5.74	1.738
10	*Int. J. Mol. Sci.*	39	785	20.13	5.924
11	*Oncotarget*	38	861	22.66	—
12	*Antimicrob. Agents Chemother*	36	770	21.39	5.191
13	*Brief. Bioinform*	36	1,585	44.03	11.622
14	*J. Chem Inf. Model.*	35	1,134	32.4	4.956
15	*Curr. Top. Med. Chem.*	34	447	13.15	3.295
16	*Curr. Med. Chem.*	30	343	11.43	4.53
17	*Cancers*	27	185	6.85	6.639
18	*Eur. J. Med. Chem.*	26	418	16.08	6.514
19	*Int. J. Antimicrob. Agents*	22	729	33.14	5.283
20	*Expert. Opin. Drug Discov.*	21	366	17	6.098
21	*Antiviral Res.*	19	312	16.42	5.927
22	*PLoS Comput. Biol.*	19	889	46.79	4.475
23	*Biochem. Biophys. Res. Commun.*	17	218	12.82	3.575
24	*BMC Syst. Biol.*	17	407	23.94	—
25	*Curr. Pharm. Design*	17	312	18.35	3.116

Notes: TP, total papers; TC, total citations; ACPP, average citations per publication; and IF: impact factor.

FX still has the highest h-index at 21, followed by Zheng, W (17), Talevi, A (12), Xu, R (11), and Schroeder, M (11). The h-index has two drawbacks when researchers of the same topic are compared with each other (Bornmann and Daniel, 2007). One is that the scholar's h-index does not decrease over time but only grows or stays the same, and it is not possible to obtain information on whether the scholar is still in an academic career. In this study, we narrow the study to the most recent publications from 2010 to 2020, taking into account the timeliness of the h-index response information. Second, older scholars usually enter academia earlier and have an advantage in their h-indexes in comparison with those of younger scholars. Therefore, this phenomenon must be targeted for analysis or illustration. Thus, by combining the authors' educational experiences and employment relationship changes that were recorded in the WOS database and ORCID business cards, we inferred that more than half of the scholars in the top 20 in terms of the number of publications received their Ph.D. before 2008, and two scholars, Mucke, HAM and Zheng, Wei, are older. In contrast, Cheng, FX, a scholar from Case Western Reserve Univ, completed his Ph.D. without a gap in 2013 and may have a longer academic career in the future; therefore, Cheng, FX's h-index in the field of drug repositioning is likely to grow more in the future and Cheng, FX is likely to have more academic influence.

3.8 Research hotspots and trends

To reveal the focus of drug repositioning research and research trends, the author keywords and the highly cited and hot research topics of the ESI for each of the 2978 publications were analyzed, which were also derived from the core database of the WOS database (SCI-E/SSCI) (Liao et al., 2019). Highly cited studies were defined as studies in the top 1% of the citations for all studies in the same ESI discipline within the 10-year range of inclusion of ESI inclusion (Chang et al., 2020). A hot research topic of the ESI refers to a study published in two years with a citation frequency within one of the corresponding disciplines in the world in the last two months (Li L. et al., 2020).

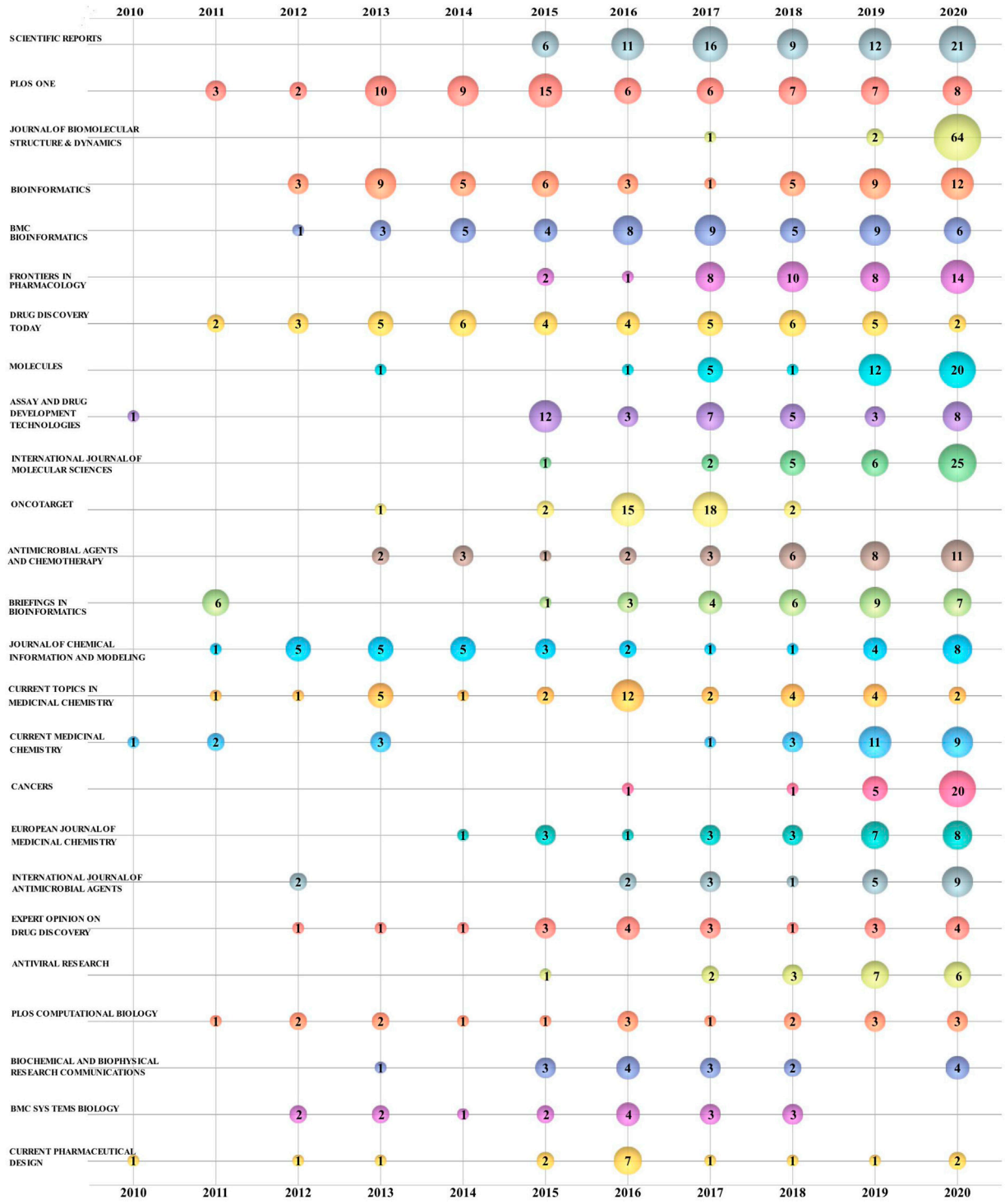


FIGURE 6
Bubble chart of the top 25 drugs repositioned by year in terms of journal production.

3.8.1 Author keyword analysis

Author keywords tend to provide more information and have thus become a widespread focus (Chen et al., 2021; Zhen et al., 2022). The data of 6,083 author keywords in the search results were merged to make keywords with the same meaning represented by a single unified word. In the end, 5,616 author

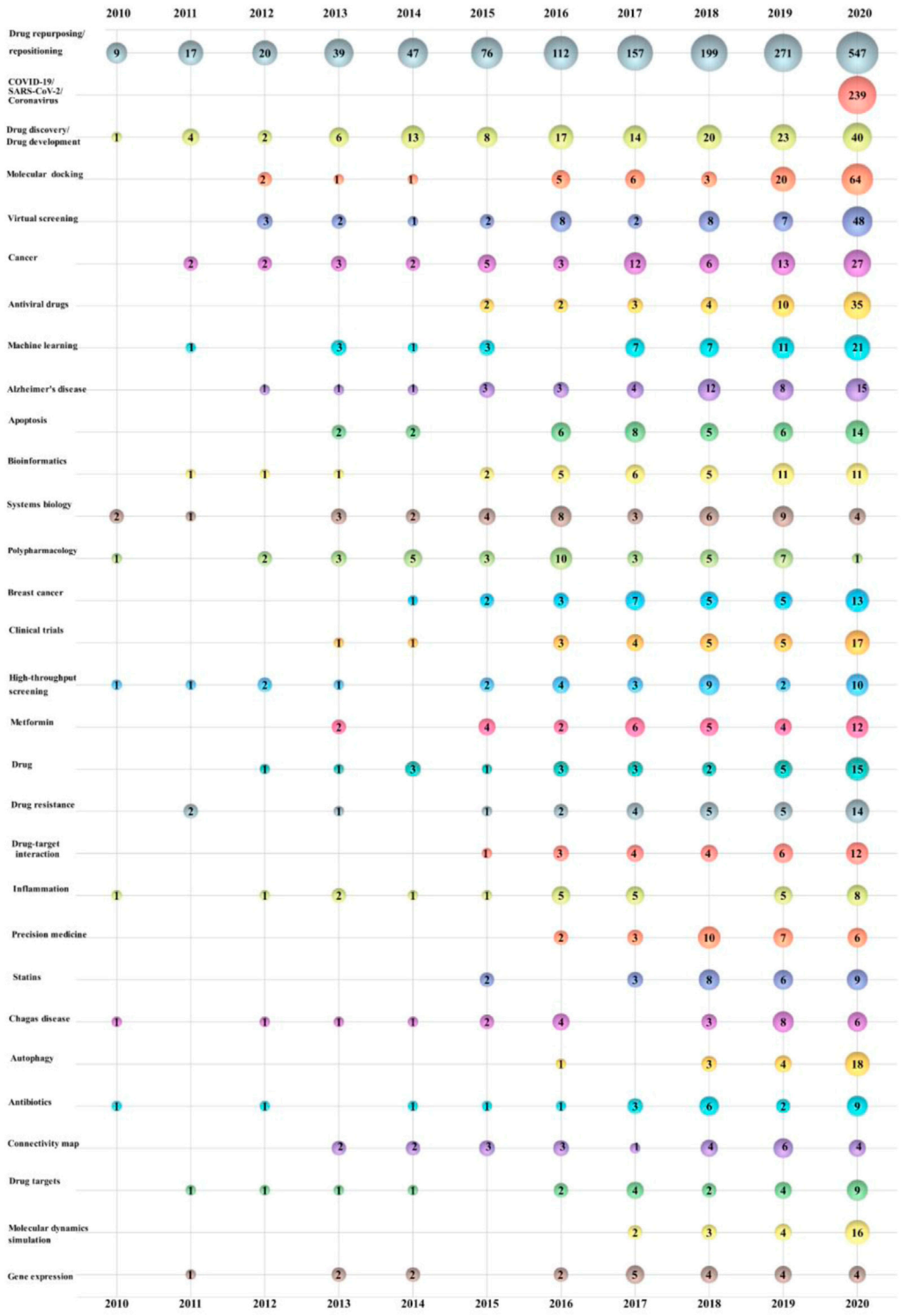


FIGURE 7
Bubble chart of the top 30 author keywords by year.

keywords were obtained. It should be specified that some publications without author keywords were excluded from the statistical analysis. Of these author keywords, 4,296 were used only once, representing 76.50% of the total. A total of 1,216 (21.65%) appeared 2–10 times, 79 (1.41%) appeared 10–20 times, 37 (0.66%) appeared 21–50 times, and the remaining eight

TABLE 5 Contribution of the top 20 authors to drug repurposing studies.

Rank	Author	TP	TC	ACPP	H-Index	TPR	Institution (Current), Country/Region
1	Cheng, FX	25	2514	100.56	21	17	Case Western Reserve Univ, USA/Anglo-America
2	Talevi, A	23	446	19.39	12	17	Natl Univ La Plata UNLP, Argentina/Latin America
3	Mucke, HAM	22	23	1.05	2	22	HM Pharma Consultancy, Austria/Oceania
4	Zheng, W	19	1,189	62.58	17	12	NIH,USA/Anglo-America
5	Xu, R	16	330	20.63	11	15	Case Western Reserve Univ, USA/Anglo-America
6	Dudley, JT	15	1,218	81.2	10	7	Icahn Sch Med Mt Sinai, USA/Anglo-America
7	Schroeder, M	15	454	30.27	11	12	Tech Univ Dresden, Germany/Europe
8	Andre, N	12	471	39.25	9	5	Aix Marseille Univ, France/Europe
9	Wang, QuanQiu	12	237	19.75	9	0	ThinTek LLC,USA/Anglo-America
10	Arga, KY	11	175	15.91	8	6	Marmara Univ, Turkey/Asia
11	Haupt, V. Joachim	11	399	36.27	8	0	Tech Univ Dresden, Germany/Europe
12	Carrillo, C	10	192	19.2	8	1	Inst Ciencias and Tecnol Cesar Milstein, Argentina/Latin America
13	Duenas-Gonzalez, A	10	326	32.6	8	9	Univ Nacl Autonoma Mexico, Mexico/Latin America
14	Bellera, Carolina L	10	192	19.2	7	0	Natl Univ La Plata, Argentina/Latin America
15	Sun, Wei	10	508	50.8	8	0	NIH,USA/Anglo-America
16	Tang, Y	10	950	95	9	6	East China Univ Sci and Technol, Peoples R China/Asia
17	Tempone, AG	10	113	11.3	7	7	Adolfo Lutz Inst, Ctr Parasitol and Mycol, Brazil/Latin America
18	Aittokallio, T	9	431	47.89	8	6	Aalto Univ, Finland/Europe
19	Bae, JS	9	39	4.33	4	9	Kyungpook Natl Univ, South Korea/Asia
20	Butte, AJ	9	1,389	154.33	9	5	Univ Calif San Francisco, USA/Anglo-America

Notes: TP, total papers; TC, total citations; ACPP, average citations per publication; and TPR, total number of publications for which they are responsible.

(0.14%) were used between 51 and 1,500 times. All keywords cumulatively appear a total of 12,400 times, while the top 30 most used author keywords appear 2,967 times alone, or approximately 23.93%, as shown in Figure 7. The comparison of keywords in recent years allows for tracking the frontiers of research and predicting hotspots and trends in drug repositioning research. The bubble plots show the three dimensions of the data, namely, the year of publication, the author's keywords, and the number of corresponding publications. The horizontal change in the size of the bubble illustrates the increasing trend of author keywords over time, the vertical size of the bubble shows the most popular keywords in that year, and the numbers in the bubble indicate the frequency of author keywords and the number of publications.

The top 30 keywords include five diseases: "COVID-19/SARS-CoV-2/Coronavirus" (239), "Cancer" (75), "Alzheimer's disease" (48), "Breast cancer (36)", and "Chagas disease" (27). Drug names appear four times, "Antiviral drugs" (56), "Metformin" (35), "Statins" (28), and "Antibiotics" (25), which reveal the diseases and applications to which drugs were often repositioned during these 11 years. There were four subject categories, "Bioinformatics" (43), "Polypharmacology" (42), "Systems biology" (42), and "Precision medicine" (28) and eight occurrences of research methods, namely, "Virtual screening" (81), "Molecular docking" (64), "Machine learning" (54, eighth), "Clinical trials" (36), "High-throughput screening" (35), "Connectivity map" (28), and "Molecular dynamics simulation" (26).

In the context of the pandemic in 2020, there was a surge in research on the subject, with "COVID-19/SARS-CoV-2/Coronavirus" topping the list of keywords as soon as they appeared that year. "Virtual screening" is a research method that appeared seven times more frequently in 2020 than in the previous year. Since "Drug repurposing/repositioning" is a subject matter and a strategy for drug discovery/drug development, it would not make much sense to analyze these two keywords. Molecular docking is one of the core steps of virtual screening, and the COVID-19 pandemic generated many opportunities for the practice of drug repositioning. Therefore, high-quality studies of the keywords "COVID-19/SARS-CoV-2/Coronavirus", "Virtual screening", and "Molecular docking" were surveyed, as shown in the bubble chart, in the past two years, reflecting the relevant research trend in recent years. Wang, F et al. developed a new free reverse docking server based on a consensus algorithm (combining several docking algorithm strategies) to address the original shortcomings of computational molecular docking in drug repositioning, such as a low success rate, cumbersome operational steps, and reliance on code writing (Wang et al., 2019). M Lapillo et al. extensively evaluated the performance assessment of docking-based target fishing methods and developed a consensus docking-based target fishing tactic

TABLE 6 All ESI hot citation studies from 2011 to 2020.

No	Author	Title	TC	Journal	Institution,Country/Region	OPC
1	Gordon, DE et al.	A SARS-CoV-2 protein interaction map reveals targets for drug repurposing	952	*Nature*	Univ Calif San Francisco, United States et al.	France; England
2	Wu, CR et al.	Analysis of therapeutic targets for SARS-CoV-2 and discovery of potential drugs by computational methods	817	*Acta Pharm. Sin. B*	Huazhong Univ Sci and Technol, Peoples R China et al.	None
3	Liu, C et al.	Research and Development on Therapeutic Agents and Vaccines for COVID-19 and Related Human Coronavirus Diseases	543	*ACS Central Sci*	CAS, United States	None
4	Elfiky, AA	Ribavirin, Remdesivir, Sofosbuvir, Galidesivir, and Tenofovir against SARS-CoV-2 RNA dependent RNA polymerase (RdRp): A molecular docking study	363	*Life Sci*	Cairo Univ, Egypt	None
5	Tu, YF et al.	A Review of SARS-CoV-2 and the Ongoing Clinical Trials	324	*Int. J. Mol. Sci*	Natl Yang Ming Univ, Taiwan	None
6	Jeon, S et al.	Identification of Antiviral Drug Candidates against SARS-CoV-2 from FDA-Approved Drugs	211	*Antimicrob. Agents Chemother*	Inst Pasteur Korea, South Korea	None
7	Wang, JM	Fast Identification of Possible Drug Treatment of Coronavirus Disease-19 (COVID-19) Through Computational Drug Repurposing Study	199	*J. Chem. Inf. Model*	Univ Pittsburgh, United States	None
8	Rosa, SGV et al.	Clinical trials on drug repositioning for COVID-19 treatment	131	*Rev. Panam. Salud Publica*	Univ Fed Fluminense, Brazil	None
9	Singh, TU et al.	Drug repurposing approach to fight COVID-19	86	*Pharmacol. Rep*	ICAR Indian Vet Res Inst, India	None
10	Rut, W et al.	Activity profiling and crystal structures of inhibitor-bound SARS-CoV-2 papain-like protease: A framework for anti-COVID-19 drug design	69	*Sci. Adv*	Wroclaw Univ Sci and Technol, Poland et al.	The United States
11	Bindu, S et al.	Non-steroidal anti-inflammatory drugs (NSAIDs) and organ damage: A current perspective	63	*Biochem. Pharmacol*	Bose Inst, India et al.	None

Notes: TC, total citations; and OPC, other partner countries.

(Lapillo et al., 2019). In a virtual screening process, Gervasoni, S. conducted a literature search for molecular binding sites for SARS-CoV-2-associated protein targets while combining pocket and docking searches to propose a new pocket mapping strategy that identifies binding cavities with significantly better performance than pocket detection alone (Gervasoni et al., 2020). Xie, L et al. screened antitoxic drugs based on the multitarget structure of the pathway center and stated that this inhibition of multiple targets in one pathway would be more effective than targeting a single protein, and the chance of drug resistance was smaller, which could be applied to other pathways (Xie and Xie, 2019). Li, Z et al. reported a virtual screening method based on accelerated free energy perturbation absolute binding free energy (FEP-ABFE) prediction and stated that the virtual screening method based on the prediction of FEP-ABFE will play a role in many other drug repositioning studies (Li Z. et al., 2020). After a series of drug repurposing computational screens and various validation activities by several scientists, it was agreed that raltegravir (Beck et al., 2020; Elfiky, 2020), clonidine (Jeon et al., 2020; Xu et al., 2020), chloroquine and hydroxychloroquine (Fantini et al., 2020) have therapeutic effects in the treatment of novel coronavirus.

In addition, from the studies on the keyword "Machine learning" over the 10-year period shown in the bubble chart, it was found that the classical machine learning algorithms of support vector machines (Kinnings et al., 2011; Pérez-Sánchez et al., 2014; Zhao and So, 2018), regularized least squares (Hao et al., 2016; Zhou et al., 2019), logistic regression (Qabaja et al., 2014; Liu et al., 2015; Xu et al., 2017), and random forests (Cao et al., 2014; Coelho et al., 2016) have been widely used in inferring drug–target and drug–disease interactions.

3.8.2 Analysis of hot research topics

While the level of influence of a study is reflected by a combination of many aspects, the number of citations remains an important indicator (Wu Y. et al., 2020). Based on the definition of highly cited and hot ESI papers in Section 3.8 of this study, a total of 108 highly cited studies were obtained, of which 11 were hot research topics. Hot research topics are shown in Table 6. It should be noted that the first-ranked author is used here as a representative, and the corresponding institution is shown. This rule is followed in Section 3.8.3 of this study. All hot research topics were published in 2020, and with the exception of an article describing the damage caused by nonsteroidal anti-inflammatory drugs (NSAIDs) to multiple organs and new information on drug repurposing (Bindu et al., 2020), the remaining studies focused on drug repositioning therapeutic target studies in novel coronavirus pneumonia (Wu C. et al., 2020; Gordon et al., 2020), screening drug studies (Elfiky, 2020; Jeon et al., 2020; Rut et al., 2020; Singh et al., 2020; Wang, 2020), reviews of clinical trials (Rosa and Santos, 2020; Tu et al., 2020),

TABLE 7 Top 20 highly cited ESI publications from 2011 to 2020.

No	Author (PY)	Title	TC	TCPY	Journal	Institution,Country/ Region	OPC
1	Wishart, DS et al. (2018)	DrugBank 5.0: a major update to the DrugBank database for 2018	1820	606.7	*Nucleic Acids Res*	Univ Alberta, Canada et al.	None
2	Pushpakom, S et al. (2019)	Drug repurposing: progress, challenges and recommendations	885	442.5	*Nat. Rev. Drug Discov*	Univ Liverpool, England et al.	None
3	Maier, L et al. (2018)	Extensive impact of non-antibiotic drugs on human gut bacteria	639	213.0	*Nature*	European Mol Biol Lab, Germany et al.	Japan
4	Zhou, YD et al. (2020); Cheng, FX et al. (2020)	Network-based drug repurposing for novel coronavirus 2019-nCoV/SARS-CoV-2	609	609.0	*Cell Discov*	Cleveland Clin, United States et al.	None
5	Anighohro, A et al. (2014)	Polypharmacology: Challenges and Opportunities in Drug Discovery	492	70.3	*J. Med. Chem*	Univ Modena and Reggio Emilia, Italy et al.	Germany
6	Cheng, FX et al. (2012)	Prediction of Drug-Target Interactions and Drug Repositioning via Network-Based Inference	491	54.6	*PLoS Comput. Biol*	E China Univ Sci and Technol, Peoples R China	None
7	Langhans, SA (2018)	Three-Dimensional *in Vitro* Cell Culture Models in Drug Discovery and Drug Repositioning	395	131.7	*Front. Pharmacol*	Alfred I DuPont Hosp Children, United States	None
8	Xu, M et al. (2016)	Identification of small-molecule inhibitors of Zika virus infection and induced neural cell death via a drug repurposing screen	389	77.8	*Nat. Med*	NIH, United States et al.	China
9	Sirota, M et al. (2011); Dudley, JT et al. (2011)	Discovery and Preclinical Validation of Drug Indications Using Compendia of Public Gene Expression Data	327	32.7	*Sci. Transl. Med*	Stanford Univ, United States	None
10	Sriram, K et al. (2018)	G Protein-Coupled Receptors as Targets for Approved Drugs: How Many Targets and How Many Drugs?	311	103.7	*Mol. Pharmacol*	Univ Calif San Diego, United States	None
11	Dudley, JT et al. (2011)	Exploiting drug-disease relationships for computational drug repositioning	282	28.2	*Brief. Bioinform*	Arizona State Univ, United States et al.	None
12	Medina-Franco, JL et al. (2013)	Shifting from the single to the multitarget paradigm in drug discovery	285	35.6	*Drug Discov. Today*	Univ Nacl Autonoma Mexico, Mexico et al.	The United States
13	Peters, JU (2013)	Polypharmacology - Foe or Friend?	275	34.4	*J. Med. Chem*	F Hoffmann La Roche Ltd., Switzerland	None
14	Yoshida, GJ et al. (2015)	Metabolic reprogramming: the emerging concept and associated therapeutic strategies	255	42.5	*J. Exp. Clin. Cancer Res*	Japan Soc Promot Sci, Japan	None
15	Skrott, Z et al. (2017)	Alcohol-abuse drug disulfiram targets cancer via p97 segregase adaptor NPL4	249	62.3	*Nature*	Palacky Univ/Czech Republic et al.	Denmark; Sweden; Switzerland; The United States; China
16	Li, J et al. (2016)	A survey of current trends in computational drug repositioning	242	48.4	*Brief. Bioinform*	Chinese Acad Med Sci, Peoples R China et al.	The United States
17	Stokes, JM et al. (2020)	A Deep Learning Approach to Antibiotic Discovery	235	235	*Cell*	MIT, United States et al.	Canada
18	Reddy, AS et al. (2013)	Polypharmacology: drug discovery for the future	228	28.5	*Expert Rev. Clin. Pharmacol*	Univ Texas Houston, United States	None
19	Menden, MP et al. (2013)	Machine Learning Prediction of Cancer Cell Sensitivity to Drugs Based on Genomic and Chemical Properties	229	28.6	*PLoS One*	Wellcome Trust Genome Campus Cambridge, England et al.	The United States
20	Beck, BR et al. (2020)	Predicting commercially available antiviral drugs that may act on the novel coronavirus (SARS-CoV-2) through a drug-target interaction deep learning model	225	225	*Comp. Struct. Biotechnol. J*	Deargen Inc., South Korea et al.	The United States

Notes: PY, publication year; TC, total citations; TCPY, total citations per year; and OPC, other partner countries.

and reports of other coronavirus therapeutic agents and vaccine studies (Liu et al., 2020). From the perspective of cooperation, most of them were completed by a country's independent agency. In terms of the countries and regions studied, four studies involved US scholars, five studies involved Asian scholars, and one contribution was from an African scholar. In addition, "A SARS-CoV-2 protein interaction map reveals targets for drug repurposing (Gordon et al., 2020)", published in *Nature* by Gordon, DE with a total of 125 scholars from the United States, the United Kingdom, and France was the most cited publication with 952 citations.

3.8.3 Analysis of the most cited studies

Eleven hot research topics were removed from the 108 highly cited ESI studies, and the top 20 most cited studies were selected from the remaining highly cited studies for analysis. In terms of year of publication, the study by Dudley, JT et al. published in *NUCLEIC ACIDS RESEARCH* in February 2011 was the earliest of these studies (Dudley et al., 2011). Five highly cited studies were published in 2013, and three studies were published as recently as 2020. Two studies were published in *Nature*, and one each was published in *Nat. Rev. Drug Discov.* and *Nat. Med. subj. of Nature E*; *J. Med. Chem. L* was next with two studies. There were 12 studies with the first author or coauthor from the United States, representing more than half of those in Table 7, followed by China (4), Canada (2), England (2), Germany (2), Japan (2), and Switzerland (2) in order of contribution of two or more studies. Nine studies were based on collaborations between different institutions in multiple countries. One of them, entitled "Alcohol-abuse drug disulfiram targets cancer via p97 segregase adapter NPL4", published in *Nature* in 2017 by Skrott, Z et al. is a collaboration between scholars from six countries: Czech Republic, the United States, Denmark, Sweden, Switzerland, and China (Skrott et al., 2017). In TC, "DrugBank 5.0: a major update to the DrugBank database for 2018" (Wishart et al., 2018) by Canadian University of Alberta scientists Wishart, DS et al. ranked first (1820 total citations). The most cited publication on an annual basis was "Network-based drug repurposing for novel coronavirus 2019-nCoV/SARS-CoV-2", published in 2020, which was authored by Zhou, YD et al. and was the highest annual average cited publication with 609 citations (Gordon et al., 2020). The scientists Cheng, FX and Dudley, JT, contributed to two of these 20 publications and are important influencers in the field.

The three studies published in 2020 focus on novel coronavirus-related drug rediscovery activities (Zhou et al., 2020) and the use of deep learning techniques (Beck et al., 2020; Stokes et al., 2020). Dudley, JT et al. (2011) and Pushpakom, S et al. (2019) provided systematic reviews of the methods and challenges of drug repositioning at that time (Dudley et al., 2011; Pushpakom et al., 2019). Initially, Sirota, M et al. (2011) explored the role of integrating genome-wide computational approaches for predicting reusable drugs (Sirota et al., 2011), while from 2013 onward, Peters, JU et al., Medina-Franco et al., JL et al., Reddy, AS et al., and Anighoro, A et al. generally recognized the importance of combining multiple points of pharmacological knowledge for drug repositioning studies (Medina-Franco et al., 2013; Peters, 2013; Reddy and Zhang, 2013; Anighohro et al., 2014). In the face of a worldwide health emergency caused by the Zika virus epidemic, Xu et al. (2016) used drug repositioning to identify lead compounds for drug development (Xu et al., 2016). Of course, techniques related to the mining of repositionable drugs through experimental high-throughput screening, a traditional experimental approach, are not without progress; for example, Langhans (2018) explored the challenges of transferring 3D cell culture technology to the use of high-throughput screening (HTS) (Langhans, 2018).

4 Discussion

In 1995, Mchugh et al. investigated the immunomodulatory action mechanism of thalidomide in humans, which was the first relevant publication on drug repositioning (Mchugh et al., 1995). The publication time can be divided into three phases: the growth period of 1995–2009, the steady growth period of 2010–2018, and the rapid rise from 2019 and beyond. The 2978 publications studied between 2010 and 2020 were completed by 15,338 authors from 3,530 research institutions in 89 countries, and at the time of this study's completion, the WOS database had surpassed more than 1,400 publications in 2021 under the same search restrictions for the topic, with more than 31,000 citations for the year, supporting further evidence that the topic is still gaining momentum worldwide.

The publication countries/regions are divided into three types: first, countries with a traditionally developed medical level, mainly developed countries in Western Europe, North America, and Oceania; second, countries with a developed pharmaceutical manufacturing industry, such as India and Japan in Asia; and third, developing countries with some research potential, such as China, Brazil, Argentina, and Mexico. In terms of national cooperation, Western European countries have shown a high degree of cooperation, with the United States, China, and the United Kingdom cooperating more frequently. This may be because Western European countries have a tradition of cooperation in the field of research, and the United States, China, and the United Kingdom are the most powerful countries in terms of drug repositioning publications and therefore cooperate more with each other. The United States accounts for half of the 20 most productive institutions, which may explain why the United States still publishes more than 50% of its studies independently, despite having the largest international collaborative network base, because it already

has the most active and high-quality producing institutions within the country for research institutions seeking collaboration. Furthermore, 19 of these 20 institutions are universities and research institutes, and one is a company, HM Pharma Consultancy, which was established in 2000 to focus on drug repositioning for the development of new drugs (Nosengo, 2016). This evidence suggests that the topic of drug repositioning is not only widely studied in academia but also has a place in the industry.

The 2978 studies are spread across 74 research areas, but pharmacology and pharmacy and biochemistry and molecular biology account for a larger proportion of the total number of studies. It is quite notable that the majority of studies reported in biophysics did not rise significantly until 2020. The reasons for this may be the following: first, there was a breakthrough in basic research in this field in 2020 and second, due to the novel coronavirus, research in this direction has increased its application for the prevention and control of the pandemic.

In terms of journals, *Sci Rep* ranked first, followed by *PLoS One* and *J. Biomol. Struct. Dyn*. In terms of lead authors, three have the most productive and influential positions: Cheng, FX is the most prolific author, based on the number of papers and h-index; Mucke, HAM is the most frequent corresponding author; and Butte, AJ is the top author in terms of ACPP ranking. Even though Latin American countries do not have an advantage in terms of national cooperation or the total number of institutional funding units, Latin American scholars have overcome many obstacles and are actively at the forefront of scholarship, contributing significantly to the field.

Through the analysis of the authors' keywords, cancer has been the main disease addressed by this method. Metformin has been found by many scientists to have a good inhibitory effect on various tumors, mainly in gynecology (Kumar et al., 2013; Xu et al., 2015; Gadducci et al., 2016; Seliger et al., 2019), and it has become a specific drug that has been most frequently mentioned in drug reuse in recent years. In terms of "antiviral drugs", scholars not only use drug repositioning to find antiviral drugs to treat diseases, such as Ebola (Kouznetsova et al., 2014; Dyall et al., 2018) and HIV (Trivedi et al., 2020), that have plagued humans for a long time but also use this method to seek treatments for infectious diseases, such as Zika virus (Xu et al., 2016; Chan et al., 2017) and novel coronaviruses that have threatened several countries and even the world. For these diseases, emergency research on drug repositioning has played an important role in reducing mortality, calming patient fears, and restoring economic production when no specific drugs or vaccines were initially available during the pandemic. The combination of precision medicine and drug repositioning studies, often used to seek treatments for rare diseases (Álvarez-Machancoses et al., 2020) and, in particular, genetically related diseases (Reay et al., 2020), is expected to be fully developed in the future. In the past 2 years, "Virtual screening", together with "Molecular docking" and "Machine learning", has become the most cutting-edge and important research methods in related technology fields, constantly improving the accuracy of drug reuse and screening. Currently, to develop more efficient and accurate research, there are two trends in the use of drug repositioning. One is the combination of various methods, such as the use of text mining and network analysis, and the creation of statistical models for predicting semantic link association to assess the relationship between pharmacological target pairings (Chen et al., 2012); text analysis combined with machine learning (Zhu et al., 2020) to develop drugs for Parkinson's disease; prediction of new DTIs using data from multiple databases (Olayan et al., 2018); and the obtained relocated anticancer drugs were verified by cross-validation, literature, and experimental verification (Cheng et al., 2021). Second, the most advanced algorithms are applied and improved, such as matrix decomposition (Xuan et al., 2019; Huang et al., 2020; Meng et al., 2021; Tang et al., 2021; Sadeghi et al., 2022) and matrix completion (Luo et al., 2018; Yan et al., 2022) and deep learning (Aliper et al., 2016; Zeng et al., 2019; Chiu et al., 2020; Stokes et al., 2020; Lee and Chen, 2021; Liu et al., 2021).

In fact, some of the studies in the list of highly cited research topics on novel coronaviruses drug repurposing studies are currently approaching 3,000 citations on Google Scholar (Gordon et al., 2020). The percentage of highly cited studies and hot research topics related to novel coronaviruses is also a good indication that the method has made an indelible contribution to the study of novel coronaviruses and similar infectious diseases. Auxiliary technology for the experimental screening of traditional drugs is also developing (Langhans, 2018), which also promotes drug repositioning or other drug development processes. Furthermore, the high-quality results of Elfiky, AA, a scientist from Cairo University, Egypt (Elfiky, 2020), suggest that relevant research in some economically underdeveloped countries may reach top levels worldwide due to the return or affiliation of some prominent scientists.

5 Conclusion

For this research, the literature on drug repositioning research published in the SCI-E and SSCI sections of WOS core journals from 2010 to 2020 was analyzed based on bibliometrics and DDA software. This area has been of interest to scientists since the end of the 20th century and entered a period of rapid growth in 2019, with the peak far from being reached. Using bibliometrics as a tool, the United States has become a world leader in terms of the number of submissions, number of high-quality studies, funding support, strength of research institutions, and number of top scholars, followed by China and the United Kingdom, where more research is being performed in this area. As a method of drug discovery, drug repurposing is closely related to the development of various biomedical disciplines, and computer-related disciplinary

methods, such as mathematical computational biology and computer science, have taken an important place in the research of this field in the last decade. The authors' keyword analysis suggests that research in the field of the novel coronavirus will remain valuable until the associated pandemic is completely contained. Virtual screening, molecular docking, machine learning, and other related technical fields still need long-term development to achieve efficient and accurate repositioning of drugs (Kumar et al., 2019). Precision medicine, combined with drug repositioning, is the most promising direction for the future. In conclusion, drug repositioning can help to treat more diseases, such as drug resistance, poor drug selectivity, and limited therapeutic options.

This study may help some scholars with an initial interest in drug repositioning-related research to gain a concise and rapid understanding of the current state of global research, as well as offer some relevant information to institutions or groups seeking collaboration.

6 Limitations

It is worth noting that this study has some biases and limitations. First, there are still some issues with the publications included in the study based on subject terms: 1) some relevant publications that do not use the search formula in this study may have been excluded from this study and 2) there may also be a small number of articles whose use of some of the aforementioned search terms deviates significantly from the general understanding; yet, such publications are included in this study. Second, some extraneous factors distort the credibility of the bibliometric statistics. 1) When analyzing the keywords of publications, some publications are excluded from the statistical analysis because they do not list author keywords (e.g., (Gordon et al., 2020)). 2) Excessive self-citation by some authors (Haghighat and Hayatdavoudi, 2021) inflates the actual level of interest in the publication. 3) for a publication, when an author submits more than one institution's address information, this publication is counted as research results by each institution. Finally, in future work, patents from the WOS database associated with the topic of drug repositioning will be analyzed to provide another perspective on the situation of the topic in terms of applications and technological innovations.

Author contributions

YW and DD contributed to the conception and design of the study. DD organized the database and performed the statistical analysis. DD and YW wrote the first draft of the manuscript. GS and QZ reviewed and edited the manuscript. GS and ZD provided financial support. All authors contributed to manuscript revision and read and approved the submitted version.

Funding

This study was supported by the Science and Technology Department of Zhejiang Province (grant no. 2022C25007), China.

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

Aliper, A., Plis, S., Artemov, A., Ulloa, A., Mamoshina, P., and Zhavoronkov, A. (2016). Deep learning applications for predicting pharmacological properties of drugs and drug repurposing using transcriptomic data. *Mol. Pharm.* 13 (7), 2524–2530. doi:10.1021/acs.molpharmaceut.6b00248

Allarakhia, M. (2013). Open-source approaches for the repurposing of existing or failed candidate drugs: Learning from and applying the lessons across diseases. *Drug Des. devel. Ther.* 7, 753–766. doi:10.2147/DDDT.S46289

Álvarez-Machancoses, Ó., Galiana, E. J. D., Cernea, A., de la Viña, J. F., and Fernández-Martínez, J. L. (2020). On the role of artificial intelligence in genomics to enhance precision medicine. *Pharmgenomics. Pers. Med.* 13, 105–119. doi:10.2147/PGPM.S205082

Amberger, J. S., Bocchini, C. A., Schiettecatte, F., Scott, A. F., and Hamosh, A. (2014). OMIM.org: Online Mendelian Inheritance in Man (OMIM), an Online catalog of human genes and genetic disorders. *Nucleic Acids Res.* 43 (1), D789–D798. doi:10.1093/nar/gku1205

Anighohro, A., Bajorath, J., and Rastelli, G. (2014). Polypharmacology: Challenges and opportunities in drug discovery. *J. Med. Chem.* 57 (19), 7874–7887. doi:10.1021/jm5006463

Ashburn, T. T., and Thor, K. B. (2004). Drug repositioning: Identifying and developing new uses for existing drugs. *Nat. Rev. Drug Discov.* 3 (8), 673–683. doi:10.1038/nrd1468

Athauda, D., and Foltynie, T. (2018). Drug repurposing in Parkinson's disease. *CNS Drugs* 32 (8), 747–761. doi:10.1007/s40263-018-0548-y

Badesch, D. B., Hill, N. S., Burgess, G., Rubin, L. J., Simonneau, G., Galie, N., et al. (2007). Sildenafil for pulmonary arterial hypertension associated with connective tissue disease. *J. Rheumatol.* 34 (12), 2417–2422.

Baker, N. C., Ekins, S., Williams, A. J., and Tropsha, A. (2018). A bibliometric review of drug repurposing. *Drug Discov. Today* 23 (3), 661–672. doi:10.1016/j.drudis.2018.01.018

Bao, G., Fang, H., Chen, L., Wan, Y., Xu, F., Yang, Q., et al. (2018). Soft robotics: Academic insights and perspectives through bibliometric analysis. *Soft Robot.* 5, 229–241. doi:10.1089/soro.2017.0135

Bao, G., Pan, L., Fang, H., Wu, X., Yu, H., Cai, S., et al. (2019). Academic review and perspectives on robotic exoskeletons. *IEEE Trans. Neural Syst. Rehabil. Eng.* 27 (11), 2294–2304. doi:10.1109/TNSRE.2019.2944655

Barlogie, B., Desikan, R., Eddlemon, P., Spencer, T., Zeldis, J., MuNshiN., et al. (2001). Extended survival in advanced and refractory multiple myeloma after single-agent thalidomide: Identification of prognostic factors in a phase 2 study of 169 patients. *Blood* 98 (2), 492–494. doi:10.1182/blood.V98.2.492

Beck, B. R., Shin, B., Choi, Y., Park, S., and Kang, K. (2020). Predicting commercially available antiviral drugs that may act on the novel coronavirus (SARS-CoV-2) through a drug-target interaction deep learning model. *Comput. Struct. Biotechnol. J.* 18, 784–790. doi:10.1016/j.csbj.2020.03.025

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., et al. (2000). The protein data bank. *Nucleic Acids Res.* 28 (1), 235–242. doi:10.1093/nar/28.1.235

Bindu, S., Mazumder, S., and Bandyopadhyay, U. (2020). Non-steroidal anti-inflammatory drugs (NSAIDs) and organ damage: A current perspective. *Biochem. Pharmacol.* 180, 114147. doi:10.1016/j.bcp.2020.114147

Bornmann, L., and Daniel, H. D. (2007). What do we know about the h index? *J. Am. Soc. Inf. Sci. Technol.* 58 (9), 1381–1385. doi:10.1002/asi.20609

Braun, T., Glänzel, W., and Schubert, A. (2006). A Hirsch-type index for journals. *Scientometrics* 69 (1), 169–173. doi:10.1007/s11192-006-0147-4

Breckenridge, A., and Jacob, R. (2019). Overcoming the legal and regulatory barriers to drug repurposing. *Nat. Rev. Drug Discov.* 18 (1), 1–2. doi:10.1038/nrd.2018.92

Canas-Guerrero, I., Mazarrón, F. R, P.-M. A., Calleja-Perucho, C., and Díaz-Rubio, G. (2013). Bibliometric analysis of research activity in the "Agronomy" category from the Web of Science, *Eur. J. Agron.* 50(1), 19–28. doi:10.1016/j.eja.201h3.05.002

Cao, D. S., Zhang, L. X., Tan, G. S., Xiang, Z., Zeng, W. B., Xu, Q. S., et al. (2014). Computational prediction of DrugTarget interactions using chemical, biological, and network features target interactions using chemical, biological, and network features. *Mol. Inf.* 33 (10), 669–681. doi:10.1002/minf.201400009

Chan, J. F.-W., Chik, K. K.-H., Yuan, S., Yip, C. C.-Y., Zhu, Z., Tee, K.-M., et al. (2017). Novel antiviral activity and mechanism of bromocriptine as a Zika virus NS2B-NS3 protease inhibitor. *Antivir. Res.* 141, 29–37. doi:10.1016/j.antiviral.2017.02.002

Chang, X., Zhang, R., Xiao, Y., Chen, X., Zhang, X., and Liu, G. (2020). Mapping of publications on asphalt pavement and bitumen materials: A bibliometric review. *Constr. Build. Mater.* 234, 117370. doi:10.1016/j.conbuildmat.2019.117370

Chen, B., Ding, Y., and Wild, D. J. (2012). Assessing drug target association using semantic linked data. *PLoS Comput. Biol.* 8 (7), e1002574. doi:10.1371/journal.pcbi.1002574

Chen, H., Wang, X., He, L., Chen, P., Wan, Y., Yang, L., et al. (2016). Chinese energy and fuels research priorities and trend: A bibliometric analysis. *Renew. Sustain. Energy Rev.* 58, 966–975. doi:10.1016/j.rser.2015.12.239

Chen, Y., Li, Y., Guo, L., Hong, J., Zhao, W., Hu, X., et al. (2021). Bibliometric analysis of the inflammasome and pyroptosis in brain. *Front. Pharmacol.* 11, 626502. doi:10.3389/fphar.2020.626502

Cheng, F., Zhou, Y., Jie, L., Li, W., Liu, G., and Yun, T. (2012). Prediction of chemical–protein interactions: Multitarget-QSAR versus computational chemogenomic methods. *Mol. Biosyst.* 8 (9), 2373–2384. doi:10.1039/c2mb25110h

Cheng, X., Zhao, W., Zhu, M., Wang, B., Wang, X., Yang, X., et al. (2021). Drug repurposing for cancer treatment through global propagation with a greedy algorithm in a multilayer network. *Cancer Biol. Med.* 18, 0. doi:10.20892/j.issn.2095-3941.2020.0218

Chiu, Y., Chen, H., Gorthi, A., Mostavi, M., Zheng, S., Huang, Y., et al. (2020). Deep learning of pharmacogenomics resources: Moving towards precision oncology. *Brief. Bioinform.* 21 (6), 2066–2083. doi:10.1093/bib/bbz144

Chopra, G., Kaushik, S., Elkin, P. L., and Samudrala, R. (2016). Combating ebola with repurposed therapeutics using the CANDO platform. *Molecules* 21 (12), 1537. doi:10.3390/molecules21121537

Christos, A., Anuj, S., Vassilis, V., Spyros, D., and Aris, P. (2011). Literature mining, ontologies and information visualization for drug repurposing. *Brief. Bioinform.* 12 (4), 357–368. doi:10.1093/bib/bbr005

Coelho, E. D., Arrais, J. P., and Oliveira, J. L. (2016). Computational discovery of putative leads for drug repositioning through drug-target interaction prediction. *PLoS Comput. Biol.* 12 (11), e1005219. doi:10.1371/journal.pcbi.1005219

Colares, G. S., Dell'Osbel, N., Wiesel, P. G., Oliveira, G. A., Lemos, P. H. Z., da Silva, F. P., et al. (2020). Floating treatment wetlands: A review and bibliometric analysis. *Sci. Total Environ.* 714, 136776. doi:10.1016/j.scitotenv.2020.136776

Csajbók, E., Berhidi, A., Vasas, L., and Schubert, A. (2007). Hirsch-index for countries based on essential science indicators data. *Scientometrics* 73 (1), 91–117. doi:10.1007/s11192-007-1859-9

Delavan, B., Roberts, R., Huang, R., Bao, W., Tong, W., and Liu, Z. (2018). Computational drug repositioning for rare diseases in the era of precision medicine. *Drug Discov. Today* 23 (2), 382–394. doi:10.1016/j.drudis.2017.10.009

Dudley, J. T., Deshpande, T., and Butte, A. J. (2011). Exploiting drug-disease relationships for computational drug repositioning. *Brief. Bioinform.* 12 (4), 303–311. doi:10.1093/bib/bbr013

Dyall, J., Johnson, J. C., Postnikova, E., Cong, Y., Zhou, H., Gerhardt, D. M., et al. (2018). *In vitro* and *in vivo* activity of amiodarone against Ebola virus. *J. Infect. Dis.* 218 (5), S592–S596. doi:10.1093/infdis/jiy345

Edwards, A. (2020). What are the odds of finding a COVID-19 drug from a lab repurposing screen? *J. Chem. Inf. Model.* 60 (12), 5727–5729. doi:10.1021/acs.jcim.0c00861

Efferth, T. (2017). From ancient herb to modern drug: Artemisia annua and artemisinin for cancer therapy. *Semin. Cancer Biol.* 46, 65–83. doi:10.1016/j.semcancer.2017.02.009

Elfiky, A. A. (2020). Ribavirin, remdesivir, sofosbuvir, galidesivir, and tenofovir against SARS-CoV-2 rna dependent rna polymerase (RdRp): A molecular docking study. *Life Sci.* 253, 117592. doi:10.1016/j.lfs.2020.117592

Fantini, J., Di Scala, C., Chahinian, H., and Yahi, N. (2020). Structural and molecular modelling studies reveal a new mechanism of action of chloroquine and hydroxychloroquine against SARS-CoV-2 infection. *Int. J. Antimicrob. Agents* 55 (5), 105960. doi:10.1016/j.ijantimicag.2020.105960

Feng, B. Y., Simeonov, A., Jadhav, A., Babaoglu, K., Inglese, J., Shoichet, B. K., et al. (2007). A high-throughput screen for aggregation-based inhibition in a large compound library. *J. Med. Chem.* 50 (10), 2385–2390. doi:10.1021/jm061317y

Feng, Y., Zhu, Q., and Lai, K.-H. (2017). Corporate social responsibility for supply chain management: A literature review and bibliometric analysis. *J. Clean. Prod.* 158, 296–307. doi:10.1016/j.jclepro.2017.05.018

Feng, Z., Chen, M., Liang, T., Shen, M., Chen, H., and Xie, X. Q. (2021). Virus-CKB: An integrated bioinformatics platform and analysis resource for COVID-19 research. *Brief. Bioinform.* 22 (2), 882–895. doi:10.1093/bib/bbaa155

Fischer, A., Sellner, M., Neranjan, S., Smieško, M., and Lill, M. A. (2020). Potential inhibitors for novel coronavirus protease identified by virtual screening of 606 million compounds. *Int. J. Mol. Sci.* 21 (10), 3626. doi:10.3390/ijms21103626

Frail, D. E., Brady, M., Escott, K. J., Holt, A., Sanganee, H. J., Pangalos, M. N., et al. (2015). Pioneering government-sponsored drug repositioning collaborations: Progress and learning. *Nat. Rev. Drug Discov.* 14 (12), 833–841. doi:10.1038/nrd4707

Francisco, A. (2013). Drug interaction networks: An introduction to translational and clinical applications. *Cardiovasc. Res.* 97 (4), 631–641. doi:10.1093/cvr/cvs289

Gadducci, A., Biglia, N., Tana, R., Cosio, S., and Gallo, M. (2016). Metformin use and gynecological cancers: A novel treatment option emerging from drug repositioning. *Crit. Rev. Oncol. Hematol.* 105, 73–83. doi:10.1016/j.critrevonc.2016.06.006

Gallo, K., Goede, A., Eckert, A., Moahamed, B., Preissner, R., and Gohlke, B.-O. (2021). Promiscuous 2.0: A resource for drug-repositioning. *Nucleic Acids Res.* 49 (1), D1373–D1380. doi:10.1093/nar/gkaa1061

Garfield, E. (2006). The history and meaning of the journal impact factor. *JAMA* 295 (1), 90–93. doi:10.1001/jama.295.1.90

Gervasoni, S., Vistoli, G., Talarico, C., Manelfi, C., Beccari, A. R., Studer, G., et al. (2020). A comprehensive mapping of the druggable cavities within the SARS-CoV-2 therapeutically relevant proteins by combining pocket and docking searches as implemented in pockets 2.0. *Int. J. Mol. Sci.* 21 (14), 5152. doi:10.3390/ijms21145152

Giraldo, P., Benavente, E., Manzano-Agugliaro, F., and Gimenez, E. (2019). Worldwide research trends on wheat and barley: A bibliometric comparative analysis. *Agronomy* 9 (7), 352. doi:10.3390/agronomy9070352

Goldstein, I., Lue, T. F., Padma-Nathan, H., Rosen, R. C., Steers, W. D., and Wicker, P. A. (1998). Oral sildenafil in the treatment of erectile dysfunction. *J. Urology* 338 (20), 1197–1203. doi:10.1016/S0022-5347(02)80386-X

Gordon, D. E., Jang, G. M., Bouhaddou, M., Xu, J., Obernier, K., White, K. M., et al. (2020). A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. *Nature* 583 (7816), 459–468. doi:10.1038/s41586-020-2286-9

Gupta, S. C., Sung, B., Prasad, S., Webb, L. J., and Aggarwal, B. B. (2013). Cancer drug discovery by repurposing: Teaching new tricks to old dogs. *Trends Pharmacol. Sci.* 34 (9), 508–517. doi:10.1016/j.tips.2013.06.005

Haghighat, M., and Hayatdavoudi, J. (2021). How hot are hot papers? The issue of prolificacy and self-citation stacking. *Scientometrics* 126 (1), 565–578. doi:10.1007/s11192-020-03749-2

Hao, M., Wang, Y., and Bryant, S. H. (2016). Improved prediction of drug-target interactions using regularized least squares integrating with kernel fusion technique. *Anal. Chim. Acta* 909, 41–50. doi:10.1016/j.aca.2016.01.014

Hirsch, J. E. (2005). An index to quantify an individual's scientific research output. *Proc. Natl. Acad. Sci. U. S. A.* 102 (46), 16569–16572. doi:10.1073/pnas.0507655102

Huang, F., Qiu, Y., Li, Q., Liu, S., and Ni, F. (2020). Predicting drug-disease associations via multi-task learning based on collective matrix factorization. *Front. Bioeng. Biotechnol.* 8, 218. doi:10.3389/fbioe.2020.00218

Hurt, R. D., Sachs, D., Glover, E. D., Offord, K. P., Johnston, J. A., Dale, L. C., et al. (1997). A comparison of sustained-release bupropion and placebo for smoking cessation. *N. Engl. J. Med.* 337 (17), 1195–1202. doi:10.1056/NEJM199710233371703

Jacsó, P. (2009). Five-year impact factor data in the Journal Citation Reports. *Online Inf. Rev.* 33 (3), 603–614. doi:10.1108/14684520910969989

Jang, H. J., Chung, I. Y., Lim, C., Chung, S., Kim, B., Kim, E. S., et al. (2019). Redirecting an anticancer to an antibacterial hit against methicillin-resistant *Staphylococcus aureus*. *Front. Microbiol.* 10, 350. doi:10.3389/fmicb.2019.00350

Jarada, T. N., Rokne, J. G., and Alhajj, R. (2020). A review of computational drug repositioning: Strategies, approaches, opportunities, challenges, and directions. *J. Cheminform.* 12 (1), 46–23. doi:10.1186/S13321-020-00450-7

Jeon, S., Ko, M., Lee, J., Choi, I., Byun, S. Y., Park, S., et al. (2020). Identification of antiviral drug candidates against SARS-CoV-2 from FDA-approved drugs. *Antimicrob. Agents Chemother.* 64 (7), e008190–00820. doi:10.1128/AAC.00819-20

Ji, X., Jin, C., Dong, X., Dixon, M. S., Williams, K. P., and Zheng, W. (2020). Literature-wide association studies (LWAS) for a rare disease: Drug repurposing for inflammatory breast cancer. *Molecules* 25 (17), 3933. doi:10.3390/molecules25173933

Joshua, S. S. (2011). Mining small-molecule screens to repurpose drugs. *Brief. Bioinform.* 12 (4), 327–335. doi:10.1093/bib/bbr028

Kang, Y., Hodges, A., Ong, E., Roberts, W., Piermarocchi, C., and Paternostro, G. (2014). Identification of drug combinations containing imatinib for treatment of BCR-ABL+ leukemias. *PLoS One* 9 (7), e102221. doi:10.1371/journal.pone.0102221

Kessing, L. V., Rytgaard, H. C., Gerds, T. A., Berk, M., Ekstrøm, C. T., and Andersen, P. K. (2019). New drug candidates for depression–a nationwide population-based study. *Acta Psychiatr. Scand.* 139 (1), 68–77. doi:10.1111/acps.12957

Kettle, J. G., and Wilson, D. M. (2016). Standing on the shoulders of giants: A retrospective analysis of kinase drug discovery at AstraZeneca. *Drug Discov. Today* 21 (10), 1596–1608. doi:10.1016/j.drudis.2016.06.007

Kim, D., Kim, W., Jeong, S., Kim, D., Yoo, J.-W., and Jung, Y. (2019). Therapeutic switching of sulpiride, an anti-psychotic and prokinetic drug, to an anti-colitic drug using colon-specific drug delivery. *Drug Deliv. Transl. Res.* 9 (1), 334–343. doi:10.1007/s13346-018-00599-7

Kinnings, S. L., Liu, N., Tonge, P. J., Jackson, R. M., Xie, L., and Bourne, P. E. (2011). A machine learning-based method to improve docking scoring functions and its application to drug repurposing. *J. Chem. Inf. Model.* 51 (2), 408–419. doi:10.1021/ci100369f

Kouznetsova, J., Sun, W., Martínez-Romero, C., Tawa, G., Shinn, P., Chen, C. Z., et al. (2014). Identification of 53 compounds that block Ebola virus-like particle entry via a repurposing screen of approved drugs. *Emerg. Microbes Infect.* 3 (1), e84–7. doi:10.1038/emi.2014.88

Kumar, R., Harilal, S., Gupta, S. V., Jose, J., Uddin, M. S., Shah, M. A., et al. (2019). Exploring the new horizons of drug repurposing: A vital tool for turning hard work into smart work. *Eur. J. Med. Chem.* 182, 111602. doi:10.1016/j.ejmech.2019.111602

Kumar, S., Meuter, A., Thapa, P., Langstraat, C., Giri, S., Chien, J., et al. (2013). Metformin intake is associated with better survival in ovarian cancer: A case-control study. *Cancer* 119 (3), 555–562. doi:10.1002/cncr.27706

Kurdi, A. I. B., Elliott, R. A., and Chen, L.-C. (2019). Clinical and economic implications of therapeutic switching of angiotensin receptor blockers to angiotensin-converting enzyme inhibitors: A population-based study. *J. Hypertens.* 37 (6), 1285–1293. doi:10.1097/HJH.0000000000002009

Kuter, D. J. (2007). New thrombopoietic growth factors. *Blood* 109 (11), 4607–4616. doi:10.1182/blood-2006-10-019315

Langedijk, J., Mantel-Teeuwisse, A. K., Slijkerman, D. S., and Schutjens, M.-H. D. B. (2015). Drug repositioning and repurposing: Terminology and definitions in literature. *Drug Discov. Today* 20 (8), 1027–1034. doi:10.1016/j.drudis.2015.05.001

Langhans, S. A. (2018). Three-dimensional *in vitro* cell culture models in drug discovery and drug repositioning. *Front. Pharmacol.* 9, 6. doi:10.3389/fphar.2018.00006

Lapillo, M., Tuccinardi, T., Martinelli, A., Macchia, M., Giordano, A., and Poli, G. (2019). Extensive reliability evaluation of docking-based target-fishing strategies. *Int. J. Mol. Sci.* 20 (5), 1023. doi:10.3390/ijms20051023

Lee, C. Y., and Chen, Y.-P. P. (2021). Prediction of drug adverse events using deep learning in pharmaceutical discovery. *Brief. Bioinform.* 22 (2), 1884–1901. doi:10.1093/bib/bbaa040

Leung, X. Y., Sun, J., and Bai, B. (2017). Bibliometrics of social media research: A co-citation and co-word analysis. *Int. J. Hosp. Manag.* 66, 35–45. doi:10.1016/j.ijhm.2017.06.012

Li, H., Liu, A., Zhao, Z., Xu, Y., Lin, J., Jou, D., et al. (2011). Fragment-based drug design and drug repositioning using multiple ligand simultaneous docking (MLSD): Identifying celecoxib and template compounds as novel inhibitors of signal transducer and activator of transcription 3 (STAT3). *J. Med. Chem.* 54 (15), 5592–5596. doi:10.1021/jm101330h

Li, J., Zheng, S., Chen, B., Butte, A. J., Swamidass, S. J., and Lu, Z. (2016). A survey of current trends in computational drug repositioning. *Brief. Bioinform.* 17 (1), 2–12. doi:10.1093/bib/bbv020

Li, L., Lu, J., Fang, H., Yin, Z., Wang, T., Wang, R., et al. (2020a). Lattice Boltzmann method for fluid-thermal systems: Status, hotspots, trends and outlook. *Ieee Access* 8, 27649–27675. doi:10.1109/ACCESS.2020.2971546

Li, X., Rousseau, J. F., Ding, Y., Song, M., and Lu, W. (2020b). Understanding drug repurposing from the perspective of biomedical entities and their evolution: Bibliographic research using aspirin. *JMIR Med. Inf.* 8 (6), e16739. doi:10.2196/16739

Li, Z., Li, X., Huang, Y.-Y., Wu, Y., Liu, R., Zhou, L., et al. (2020c). Identify potent SARS-CoV-2 main protease inhibitors via accelerated free energy perturbation-based virtual screening of existing drugs. *Proc. Natl. Acad. Sci. U. S. A.* 117 (44), 27381–27387. doi:10.1073/pnas.2010470117

Liao, H., Tang, M., Li, Z., and Lev, B. (2019). Bibliometric analysis for highly cited papers in operations research and management science from 2008 to 2017 based on Essential Science Indicators. *Omega* 88, 223–236. doi:10.1016/j.omega.2018.11.005

Lin, K., Li, L., Dai, Y., Wang, H., Teng, S., Bao, X., et al. (2020). A comprehensive evaluation of connectivity methods for L1000 data. *Brief. Bioinform.* 21 (6), 2194–2205. doi:10.1093/bib/bbz129

Liu, C., Zhou, Q., Li, Y., Garner, L. V., Watkins, S. P., Carter, L. J., et al. (2020). Research and development on therapeutic agents and vaccines for COVID-19 and related human coronavirus diseases. *ACS Cent. Sci.* 6 (3), 315–331. doi:10.1021/acscentsci.0c00272

Liu, R., Wei, L., and Zhang, P. (2021). A deep learning framework for drug repurposing via emulating clinical trials on real-world patient data. *Nat. Mach. Intell.* 3 (1), 68–75. doi:10.1038/s42256-020-00276-w

Liu, Z., Guo, F., Gu, J., Wang, Y., Li, Y., Wang, D., et al. (2015). Similarity-based prediction for Anatomical Therapeutic Chemical classification of drugs by integrating multiple data sources. *Bioinformatics* 31 (11), 1788–1795. doi:10.1093/bioinformatics/btv055

Lotfi Shahreza, M., Ghadiri, N., Mousavi, S. R., Varshosaz, J., and Green, J. R. (2018). A review of network-based approaches to drug repositioning. *Brief. Bioinform.* 19 (5), 878–892. doi:10.1093/bib/bbx017

Luo, H., Li, M., Wang, S., Liu, Q., Li, Y., and Wang, J. (2018). Computational drug repositioning using low-rank matrix approximation and randomized algorithms. *Bioinformatics* 34 (11), 1904–1912. doi:10.1093/bioinformatics/bty013

Luo, H., Li, M., Yang, M., Wu, F.-X., Li, Y., and Wang, J. (2021). Biomedical data and computational models for drug repositioning: A comprehensive review. *Brief. Bioinform.* 22 (2), 1604–1619. doi:10.1093/bib/bbz176

Lv, B.-M., Tong, X.-Y., Quan, Y., Liu, M.-Y., Zhang, Q.-Y., Song, Y.-F., et al. (2018). Drug repurposing for Japanese encephalitis virus infection by systems biology methods. *Molecules* 23 (12), 3346. doi:10.3390/molecules23123346

Malas, T. B., Vlietstra, W. J., Kudrin, R., Starikov, S., Charrout, M., Roos, M., et al. (2019). Drug prioritization using the semantic properties of a knowledge graph. *Sci. Rep.* 9 (1), 1–10. doi:10.1038/s41598-019-42806-6

Mao, G., Hu, H., Liu, X., Crittenden, J., and Huang, N. (2021). A bibliometric analysis of industrial wastewater treatments from 1998 to 2019. *Environ. Pollut.* 275, 115785. doi:10.1016/j.envpol.2020.115785

Mazzolari, A., Gervasoni, S., Pedretti, A., Fumagalli, L., Matucci, R., and Vistoli, G. (2020). Repositioning dequalinium as potent muscarinic allosteric ligand by

combining virtual screening campaigns and experimental binding assays. *Int. J. Mol. Sci.* 21 (17), 5961. doi:10.3390/ijms21175961

Mchugh, S., Rifkin, I., Deighton, J., Wilson, A., Lachmann, P., Lockwood, C., et al. (1995). The immunosuppressive drug thalidomide induces T helper cell type 2 (Th2) and concomitantly inhibits Th1 cytokine production in mitogen- and antigen-stimulated human peripheral blood mononuclear cell cultures. *Clin. Exp. Immunol.* 99 (2), 160–167. doi:10.1111/j.1365-2249.1995.tb05527.x

Medina-Franco, J. L., Giulianotti, M. A., Welmaker, G. S., and Houghten, R. A. (2013). Shifting from the single to the multitarget paradigm in drug discovery. *Drug Discov. Today* 18 (9-10), 495–501. doi:10.1016/j.drudis.2013.01.008

Mendez, D., Gaulton, A., Bento, A. P., Chambers, J., De Veij, M., Félix, E., et al. (2019). ChEMBL: Towards direct deposition of bioassay data. *Nucleic Acids Res.* 47 (1), D930–D940. doi:10.1093/nar/gky1075

Meng, Y., Jin, M., Tang, X., and Xu, J. (2021). Drug repositioning based on similarity constrained probabilistic matrix factorization: COVID-19 as a case study. *Appl. Soft Comput.* 103, 107135. doi:10.1016/j.asoc.2021.107135

Mercorelli, B., Palù, G., and Loregian, A. (2018). Drug repurposing for viral infectious diseases: How far are we? *Trends Microbiol.* 26, 865–876. doi:10.1016/j.tim.2018.04.004

Mihai, D. P., Nitulescu, G. M., Ion, G. N. D., Ciotu, C. I., Chirita, C., and Negres, S. (2019). Computational drug repurposing algorithm targeting TRPA1 calcium channel as a potential therapeutic solution for multiple sclerosis. *Pharmaceutics* 11 (9), 446. doi:10.3390/pharmaceutics11090446

Milne, T. A. (2017). Mouse models of MLL leukemia: Recapitulating the human disease. *Blood* 129 (16), 2217–2223. doi:10.1182/blood-2016-10-691428

Mirza, N., Sills, G. J., Pirmohamed, M., and Marson, A. G. (2017). Identifying new antiepileptic drugs through genomics-based drug repurposing. *Hum. Mol. Genet.* 26 (3), 527–537. doi:10.1093/hmg/ddw410

Muratov, E. N., Amaro, R., Andrade, C. H., Brown, N., Ekins, S., Fourches, D., et al. (2021). A critical overview of computational approaches employed for COVID-19 drug discovery. *Chem. Soc. Rev.* 50 (16), 9121–9151. doi:10.1039/d0cs01065k

Napolitano, F., Zhao, Y., Moreira, V. M., Tagliaferri, R., Kere, J., D'Amato, M., et al. (2013). Drug repositioning: A machine-learning approach through data integration. *J. Cheminform.* 5 (1), 30–39. doi:10.1186/1758-2946-5-30

Nosengo, N. (2016). New tricks for old drugs. *Nature* 534 (7607), 314–316. doi:10.1038/534314a

Olayan, R. S., Haitham, A., and Bajic, V. B. (2018). Ddr: Efficient computational method to predict drug–target interactions using graph mining and machine learning approaches. *Bioinformatics* 34 (7), 1164–1173. doi:10.1093/bioinformatics/btx731

Ozsoy, M. G., Özyer, T., Polat, F., and Alhajj, R. (2018). Realizing drug repositioning by adapting a recommendation system to handle the process. *BMC Bioinforma.* 19 (1), 136. doi:10.1186/s12859-018-2142-1

Parisi, D., Adasme, M. F., Sveshnikova, A., Bolz, S. N., Moreau, Y., and Schroeder, M. (2020). Drug repositioning or target repositioning: A structural perspective of drug-target-indication relationship for available repurposed drugs. *Comput. Struct. Biotechnol. J.* 18, 1043–1055. doi:10.1016/j.csbj.2020.04.004

Paul, S. M., and Lewis-Hall, F. (2013). Drugs in search of diseases. *Sci. Transl. Med.* 5 (186), 186fs18. doi:10.1126/scitranslmed.3004452

Pérez-Sánchez, H., Cano, G., and García-Rodríguez, J. (2014). Improving drug discovery using hybrid softcomputing methods. *Appl. Soft Comput.* 20, 119–126. doi:10.1016/j.asoc.2013.10.033

Peters, J. U. (2013). Polypharmacology - foe or friend? *J. Med. Chem.* 56 (22), 8955–8971. doi:10.1021/jm400856t

Pietschmann, T. (2017). Clinically approved ion channel inhibitors close gates for hepatitis C virus and open doors for drug repurposing in infectious viral diseases. *J. Virol.* 91 (2), e01914–01916. doi:10.1128/JVI.01914-16

Powell, A. G. M. T., Hughes, D. L., Wheat, J. R., and Lewis, W. G. (2016). The 100 most influential manuscripts in gastric cancer: A bibliometric analysis. *Int. J. Surg.* 28, 83–90. doi:10.1016/j.ijsu.2016.02.028

Pujol, A., Mosca, R., Farrés, J., and Aloy, P. (2010). Unveiling the role of network and systems biology in drug discovery. *Trends Pharmacol. Sci.* 31 (3), 115–123. doi:10.1016/j.tips.2009.11.006

Pushpakom, S., Iorio, F., Eyers, P. A., Escott, K. J., Hopper, S., Wells, A., et al. (2019). Drug repurposing: Progress, challenges and recommendations. *Nat. Rev. Drug Discov.* 18 (1), 41–58. doi:10.1038/nrd.2018.168

Qabaja, A., Alshalalfa, M., Alanazi, E., and Alhajj, R. (2014). Prediction of novel drug indications using network driven biological data prioritization and integration. *J. Cheminform.* 6 (1), 1–14. doi:10.1186/1758-2946-6-1

Raju, T. (2000). The Nobel chronicles. 1988: James whyte Black, (b 1924), gertrude elion (1918-99), and george H hitchings (1905-98). *Lancet* 355 (9208), 1022. doi:10.1016/S0140-6736(05)74775-9

Reay, W. R., Atkins, J. R., Carr, V. J., Green, M. J., and Cairns, M. J. (2020). Pharmacological enrichment of polygenic risk for precision medicine in complex disorders. *Sci. Rep.* 10 (1), 879. doi:10.1038/s41598-020-57795-0

Reddy, A. S., and Zhang, S. (2013). Polypharmacology: Drug discovery for the future. *Expert Rev. Clin. Pharmacol.* 6 (1), 41–47. doi:10.1586/ecp.12.74

Rey-Martí, A., Ribeiro-Soriano, D., and Palacios-Marqués, D. (2016). A bibliometric analysis of social entrepreneurship. *J. Bus. Res.* 69 (5), 1651–1655. doi:10.1016/j.jbusres.2015.10.033

Rosa, S. G. V., and Santos, W. C. (2020). Clinical trials on drug repositioning for COVID-19 treatment. *Rev. Panam. Salud Publica* 44, e40. doi:10.26633/RPSP.2020.40

Rut, W., Lv, Z., Zmudzinski, M., Patchett, S., Nayak, D., Snipas, S. J., et al. (2020). Activity profiling and crystal structures of inhibitor-bound SARS-CoV-2 papain-like protease: A framework for anti–COVID-19 drug design. *Sci. Adv.* 6 (42), eabd4596. doi:10.1126/sciadv.abd4596

Sadeghi, S., Lu, J., and Ngom, A. (2022). A network-based drug repurposing method via non-negative matrix factorization. *Bioinformatics* 38 (5), 1369–1377. doi:10.1093/bioinformatics/btab826

Salazar, C., Schütze, J., and Ebenhöh, O. (2006). Bioinformatics meets systems biology. *Genome Biol.* 7 (1), 303. doi:10.1186/gb-2006-7-1-303

Sardana, D., Zhu, C., Zhang, M., Gudivada, R. C., Yang, L., and Jegga, A. G. (2011). Drug repositioning for orphan diseases. *Brief. Bioinform.* 12 (4), 346–356. doi:10.1093/bib/bbr021

Scherman, D., and Fetro, C. (2020). Drug repositioning for rare diseases: Knowledge-based success stories. *Therapie* 75 (2), 161–167. doi:10.1016/j.therap.2020.02.007

Schuster, D., Laggner, C., and Langer, T. (2005). Why drugs fail - a study on side effects in new chemical entities. *Curr. Pharm. Des.* 11 (27), 3545–3559. doi:10.2174/138161205774414510

Seliger, C., Luber, C., Gerken, M., Schaertl, J., Proescholdt, M., Riemenschneider, M. J., et al. (2019). Use of metformin and survival of patients with high-grade glioma. *Int. J. Cancer* 144 (2), 273–280. doi:10.1002/ijc.31783

Sharifi, A. (2021). Urban sustainability assessment: An overview and bibliometric analysis. *Ecol. Indic.* 121, 107102. doi:10.1016/j.ecolind.2020.107102

Simsek, M., Meijer, B., van Bodegraven, A. A., de Boer, N. K. H., and Mulder, C. J. J. (2018). Finding hidden treasures in old drugs: The challenges and importance of licensing generics. *Drug Discov. Today* 23 (1), 17–21. doi:10.1016/j.drudis.2017.08.008

Singh, T. U., Parida, S., Lingaraju, M. C., Kesavan, M., Kumar, D., and Singh, R. K. (2020). Drug repurposing approach to fight COVID-19. *Pharmacol. Rep.* 72 (6), 1479–1508. doi:10.1007/s43440-020-00155-6

Singhal, S., Mehta, J., Desikan, R., Ayers, D., Roberson, P., Eddlemon, P., et al. (1999). Antitumor activity of thalidomide in refractory multiple myeloma. *N. Engl. J. Med.* 341 (21), 1565–1571. doi:10.1056/NEJM199911183412102

Sirota, M., Dudley, J. T., Kim, J., Chiang, A. P., Morgan, A. A., Sweet-Cordero, A., et al. (2011). Discovery and preclinical validation of drug indications using compendia of public gene expression data. *Sci. Transl. Med.* 3 (96), 96ra77. doi:10.1126/scitranslmed.3001318

Skrott, Z., Mistrik, M., Andersen, K. K., Friis, S., Majera, D., Gursky, J., et al. (2017). Alcohol-abuse drug disulfiram targets cancer via p97 segregase adaptor NPL4. *Nature* 552 (7684), 194–199. doi:10.1038/nature25016

Southall, N. T., Natarajan, M., Lau, L. P. L., Jonker, A. H., Deprez, B., Guilliams, T., et al. (2019). The use or generation of biomedical data and existing medicines to discover and establish new treatments for patients with rare diseases–recommendations of the irdirc data mining and repurposing task force. *Orphanet J. Rare Dis.* 14 (1), 225–228. doi:10.1186/s13023-019-1193-3

Southan, C., Williams, A. J., and Ekins, S. (2013). Challenges and recommendations for obtaining chemical structures of industry-provided repurposing candidates. *Drug Discov. Today* 18 (1-2), 58–70. doi:10.1016/j.drudis.2012.11.005

Stokes, J. M., Yang, K., Swanson, K., Jin, W., Cubillos-Ruiz, A., Donghia, N. M., et al. (2020). A deep learning approach to antibiotic discovery. *Cell* 180 (4), 688–702. doi:10.1016/j.cell.2020.01.021

Swinney, D. C., and Anthony, J. (2011). How were new medicines discovered? *Nat. Rev. Drug Discov.* 10 (7), 507–519. doi:10.1038/nrd3480

Talevi, A., and Bellera, C. L. (2020). Challenges and opportunities with drug repurposing: Finding strategies to find alternative uses of therapeutics. *Expert Opin. Drug Discov.* 15 (4), 397–401. doi:10.1080/17460441.2020.1704729

Tang, X., Cai, L., Meng, Y., Xu, J., Lu, C., and Yang, J. (2021). Indicator regularized non-negative matrix factorization method-based drug repurposing for COVID-19. *Front. Immunol.* 11, 603615. doi:10.3389/fimmu.2020.603615

Tanoli, Z., Alam, Z., Ianevski, A., Wennerberg, K., Vähä-Koskela, M., and Aittokallio, T. (2020). Interactive visual analysis of drug–target interaction networks using drug target profiler, with applications to precision medicine and drug repurposing. *Brief. Bioinform.* 21 (1), 211–220. doi:10.1093/bib/bby119

Tao, T., Zhao, X., Lou, J., Bo, L., Wang, F., Li, J., et al. (2012). The top cited clinical research articles on sepsis: A bibliometric analysis. *Crit. Care* 16 (3), R110–R117. doi:10.1186/cc11401

Trivedi, J., Mohan, M., and Byrareddy, S. N. (2020). Drug repurposing approaches to combating viral infections. *J. Clin. Med.* 9 (11), 3777. doi:10.3390/jcm9113777

Troulé, K., López-Fernández, H., García-Martín, S., Reboiro-Jato, M., Carretero-Puche, C., Martorell-Marugán, J., et al. (2021). Dreimt: A drug repositioning database and prioritization tool for immunomodulation. *Bioinformatics* 37 (4), 578–579. doi:10.1093/bioinformatics/btaa727

Tu, Y. F., Chien, C. S., Yarmishyn, A. A., Lin, Y. Y., Luo, Y. H., Lin, Y. T., et al. (2020). A review of SARS-CoV-2 and the ongoing clinical trials. *Int. J. Mol. Sci.* 21 (7), 2657. doi:10.3390/ijms21072657

Tummino, T. A., Rezelj, V. V., Fischer, B., Fischer, A., O'Meara, M. J., Monel, B., et al. (2021). Drug-induced phospholipidosis confounds drug repurposing for SARS-CoV-2. *Science* 373 (6554), 541–547. doi:10.1126/science.abi4708

Turanli, B., Altay, O., Borén, J., Turkez, H., Nielsen, J., Uhlen, M., et al. (2021). Systems biology based drug repositioning for development of cancer therapy. *Semin. Cancer Biol.* 68, 47–58. doi:10.1016/j.semcancer.2019.09.020

Van Raan, A. F. (2006). Comparison of the Hirsch-index with standard bibliometric indicators and with peer judgment for 147 chemistry research groups. *Scientometrics* 67 (3), 491–502. doi:10.1556/Scient.67.2006.3.10

Vanhaelen, Q., Mamoshina, P., Aliper, A. M., Artemov, A., Lezhnina, K., Ozerov, I., et al. (2017). Design of efficient computational workflows for *in silico* drug repurposing. *Drug Discov. Today* 22 (2), 210–222. doi:10.1016/j.drudis.2016.09.019

Verbaanderd, C., Rooman, I., and Huys, I. (2021). Exploring new uses for existing drugs: Innovative mechanisms to fund independent clinical research. *Trials* 22 (1), 322. doi:10.1186/s13063-021-05273-x

Vogel, R., and Güttel, W. H. (2013). The dynamic capability view in strategic management: A bibliometric review. *Int. J. Manag. Rev.* 15 (4), 426–446. doi:10.1111/ijmr.12000

Vogelsang, G. B., Farmer, E. R., Hess, A. D., Altamonte, V., Beschorner, W. E., Jabs, D. A., et al. (1992). Thalidomide for the treatment of chronic graft-versus-host disease. *N. Engl. J. Med.* 326 (16), 1055–1058. doi:10.1056/NEJM199204163261604

Wang, F., Wu, F., Li, C., Jia, C., Su, S., Hao, G., et al. (2019). Acid: A free tool for drug repurposing using consensus inverse docking strategy. *J. Cheminform.* 11 (1), 73–11. doi:10.1186/s13321-019-0394-z

Wang, J. (2020). Fast identification of possible drug treatment of coronavirus disease-19 (COVID-19) through computational drug repurposing study. *J. Chem. Inf. Model.* 60 (6), 3277–3286. doi:10.1021/acs.jcim.0c00179

Wishart, D. S., Feunang, Y. D., Guo, A. C., Lo, E. J., Marcu, A., Grant, J. R., et al. (2018). DrugBank 5.0: A major update to the DrugBank database for 2018. *Nucleic Acids Res.* 46 (D1), D1074–D1082. doi:10.1093/nar/gkx1037

Wu, C., Liu, Y., Yang, Y., Zhang, P., Zhong, W., Wang, Y., et al. (2020a). Analysis of therapeutic targets for SARS-CoV-2 and discovery of potential drugs by computational methods. *Acta Pharm. Sin. B* 10 (5), 766–788. doi:10.1016/j.apsb.2020.02.008

Wu, H., Huang, J., Zhong, Y., and Huang, Q. (2017a). DrugSig: A resource for computational drug repositioning utilizing gene expression signatures. *PLoS One* 12 (5), e0177743. doi:10.1371/journal.pone.0177743

Wu, Y., Chen, J., Fang, H., and Wan, Y. (2020b). Intimate partner violence: A bibliometric review of literature. *Int. J. Environ. Res. Public Health* 17 (15), 5607. doi:10.3390/ijerph17155607

Wu, Z., Cheng, F., Li, J., Li, W., Liu, G., and Tang, Y. (2017b). Sdtnbi: An integrated network and chemoinformatics tool for systematic prediction of drug–target interactions and drug repositioning. *Brief. Bioinform.* 18 (2), 333–347. doi:10.1093/bib/bbw012

Xie, L., and Xie, L. (2019). Pathway-centric structure-based multi-target compound screening for anti-virulence drug repurposing. *Int. J. Mol. Sci.* 20 (14), 3504. doi:10.3390/ijms20143504

Xu, D., Ham, A. G., Tivis, R. D., Caylor, M. L., Tao, A., Flynn, S. T., et al. (2017). Msbis: A multi-step biomedical informatics screening approach for identifying medications that mitigate the risks of metoclopramide-induced tardive dyskinesia. *EBioMedicine* 26, 132–137. doi:10.1016/j.ebiom.2017.11.015

Xu, H., Aldrich, M. C., Chen, Q., Liu, H., Peterson, N. B., Dai, Q., et al. (2015). Validating drug repurposing signals using electronic health records: A case study of metformin associated with reduced cancer mortality. *J. Am. Med. Inf. Assoc.* 22 (1), 179–191. doi:10.1136/amiajnl-2014-002649

Xu, J., Shi, P.-Y., Li, H., and Zhou, J. (2020). Broad spectrum antiviral agent niclosamide and its therapeutic potential. *ACS Infect. Dis.* 6 (5), 909–915. doi:10.1021/acsinfecdis.0c00052

Xu, M., Lee, E. M., Wen, Z., Cheng, Y., Huang, W. K., Qian, X., et al. (2016). Identification of small-molecule inhibitors of Zika virus infection and induced neural cell death via a drug repurposing screen. *Nat. Med.* 22 (10), 1101–1107. doi:10.1038/nm.4184

Xuan, P., Cao, Y., Zhang, T., Wang, X., Pan, S., and Shen, T. (2019). Drug repositioning through integration of prior knowledge and projections of drugs and diseases. *Bioinformatics* 35 (20), 4108–4119. doi:10.1093/bioinformatics/btz182

Yan, Y., Yang, M., Zhao, H., Duan, G., Peng, X., and Wang, J. (2022). Drug repositioning based on multi-view learning with matrix completion. *Brief. Bioinform.* 23, bbac054. doi:10.1093/bib/bbac054

Yang, X., Wang, Y., Byrne, R., Schneider, G., and Yang, S. (2019). Concepts of artificial intelligence for computer-assisted drug discovery. *Chem. Rev.* 119 (18), 10520–10594. doi:10.1021/acs.chemrev.8b00728

Yella, J. K., Yaddanapudi, S., Wang, Y., and Jegga, A. G. (2018). Changing trends in computational drug repositioning. *Pharmaceuticals* 11 (2), 57. doi:10.3390/ph11020057

Yuan, L., William, K., Thompson, T., Zeng, Z., Berendsen, M. A., Jonnalagadda, S. R., et al. (2017). Natural language processing for EHR-based pharmacovigilance: A structured review. *Drug Saf.* 40 (11), 1075–1089. doi:10.1007/s40264-017-0558-6

Zeng, X., Zhu, S., Liu, X., Zhou, Y., Nussinov, R., and Cheng, F. (2019). deepDR: a network-based deep learning approach to *in silico* drug repositioning. *Bioinformatics* 35 (24), 5191–5198. doi:10.1093/bioinformatics/btz418

Zhang, D., Zhang, Z., and Managi, S. (2019). A bibliometric analysis on green finance: Current status, development, and future directions. *Finance Res. Lett.* 29, 425–430. doi:10.1016/j.frl.2019.02.003

Zhao, C., Dai, X., Li, Y., Guo, Q., Zhang, J., Zhang, X., et al. (2019). EK-DRD: A comprehensive database for drug repositioning inspired by experimental knowledge. *J. Chem. Inf. Model.* 59 (9), 3619–3624. doi:10.1021/acs.jcim.9b00365

Zhao, J., Cheng, F., Wang, Y., Arteaga, C. L., and Zhao, Z. (2016). Systematic prioritization of druggable mutations in~ 5000 genomes across 16 cancer types using a structural genomics-based approach. *Mol. Cell. Proteomics* 15 (2), 642–656. doi:10.1074/mcp.M115.053199

Zhao, K., and So, H.-C. (2018). Drug repositioning for schizophrenia and depression/anxiety disorders: A machine learning approach leveraging expression data. *IEEE J. Biomed. Health Inf.* 23 (3), 1304–1315. doi:10.1109/jbhi.2018.2856535

Zhen, G., Yingying, L., and Jingcheng, D. (2022). Drug therapies for copd: A bibliometric review from 1980 to 2021. *Front. Pharmacol.* 13, 820086. doi:10.3389/fphar.2022.820086

Zhou, L., Li, Z., Yang, J., Tian, G., Liu, F., Wen, H., et al. (2019). Revealing drug-target interactions with computational models and algorithms. *Molecules* 24 (9), 1714. doi:10.3390/molecules24091714

Zhou, Y., Hou, Y., Shen, J., Huang, Y., Martin, W., and Cheng, F. (2020). Network-based drug repurposing for novel coronavirus 2019-nCoV/SARS-CoV-2. *Cell Discov.* 6, 14. doi:10.1038/s41421-020-0153-3

Zhu, Y., Jung, W., Wang, F., and Che, C. (2020). Drug repurposing against Parkinson's disease by text mining the scientific literature. *Libr. Hi Tech.* 38 (4), 741–750. doi:10.1108/lht-08-2019-0170

frontiers | Frontiers in Pharmacology

TYPE Review
PUBLISHED 03 October 2022
DOI 10.3389/fphar.2022.1007315

OPEN ACCESS

EDITED BY
Mithun Rudrapal,
Rasiklal M. Dhariwal Institute of Pharmaceutical Education and Research, India

REVIEWED BY
Alejandro Schcolnik-Cabrera,
Hôpital Maisonneuve-Rosemont, Canada
Praveen Pasala,
Jawaharlal Nehru Technological University Anantapur, India

*CORRESPONDENCE
Bibhuti Bhusan Kakoti,
bibhutikakoti@dibru.ac.in

SPECIALTY SECTION
This article was submitted to Drugs Outcomes Research and Policies, a section of the journal Frontiers in Pharmacology

RECEIVED 30 July 2022
ACCEPTED 12 September 2022
PUBLISHED 03 October 2022

CITATION
Kakoti BB, Bezbaruah R and Ahmed N (2022), Therapeutic drug repositioning with special emphasis on neurodegenerative diseases: Threats and issues.
Front. Pharmacol. 13:1007315.
doi: 10.3389/fphar.2022.1007315

Therapeutic drug repositioning with special emphasis on neurodegenerative diseases: Threats and issues

Bibhuti Bhusan Kakoti*, Rajashri Bezbaruah and Nasima Ahmed

Department of Pharmaceutical Sciences, Faculty of Science and Engineering, Dibrugarh University, Dibrugarh, India

Drug repositioning or repurposing is the process of discovering leading-edge indications for authorized or declined/abandoned molecules for use in different diseases. This approach revitalizes the traditional drug discovery method by revealing new therapeutic applications for existing drugs. There are numerous studies available that highlight the triumph of several drugs as repurposed therapeutics. For example, sildenafil to aspirin, thalidomide to adalimumab, and so on. Millions of people worldwide are affected by neurodegenerative diseases. According to a 2021 report, the Alzheimer's disease Association estimates that 6.2 million Americans are detected with Alzheimer's disease. By 2030, approximately 1.2 million people in the United States possibly acquire Parkinson's disease. Drugs that act on a single molecular target benefit people suffering from neurodegenerative diseases. Current pharmacological approaches, on the other hand, are constrained in their capacity to unquestionably alter the course of the disease and provide patients with inadequate and momentary benefits. Drug repositioning–based approaches appear to be very pertinent, expense- and time-reducing strategies for the enhancement of medicinal opportunities for such diseases in the current era. Kinase inhibitors, for example, which were developed for various oncology indications, demonstrated significant neuroprotective effects in neurodegenerative diseases. This review expounds on the classical and recent examples of drug repositioning at various stages of drug development, with a special focus on neurodegenerative disorders and the aspects of threats and issues viz. the regulatory, scientific, and economic aspects.

KEYWORDS

drug repurposing, alzheimer, parkinson, neurodegenarative disease, artificial intelligence

Abbreviations: AD, Alzheimer's disease; AI, artificial intelligence; ALS, amyotrophic lateral sclerosis; CMap, connectivity map; COMT, catechol-o-methyltransferase; DR, drug repurposing; EMR, electronic medical record; FDA, U.S. food and drug administration; FTD, frontotemporal dementia; HD, huntington's disease; MAO, monoamine oxidase; ML, machine learning; MS, multiple sclerosis,; NDs, neurodegenerative diseases; OMOP, observational medical outcomes partnership; PD, Parkinson's disease

1 Introduction

There has been a relentless search for the discovery of drugs in various therapeutic segments. Of late repurposing also referred to as drug repositioning has gained interest in recent years. As per the reports, various discoveries have taken place in the finding of new molecules and the development of alternative strategies using repurposing strategies. In comparison to the classical drug discovery process, the new approach of Drug repurposing (DR) has various advantages and has opened new vistas in the field of Pharmacology and Medicinal chemistry. Treatment of rare and intractable diseases, minimizing attrition rates, reducing the cost of therapy, etc. are some of the advantages of drug repurposing. Essentially it is a new way of approaching drug compounds and targets that have been abandoned during the development stages either to their risks or other issues. This review shed light on the classical and recent examples of DR at various stages of drug development, with a special focus on neurodegenerative diseases (NDs) and the aspects of threats and issues viz. the regulatory, scientific, and economic aspects.

2 Drug repurposing approaches

As stated by the U.S. Census Bureau, the world's population on 1st January 2022 was estimated to be 7.8 billion. This depicts that there is an expansion of 74 million people or a 0.9% growth rate (The Economic Times, 2021). Furthermore, there has been an escalation in the figure of geriatric people that is supplementing the world population growth. The dwellers of each country in the world are enduring build-up in both the proportion and size of elderly persons. It is envisioned that 1 in every 6 people in the globe will be in the age group of 60 years or beyond by 2030. In developed countries, life expectancy is ascending in small doses above 80 years. While there is a deviation in the assortment of a country's population towards older ages, the frequency of incidence and progression of incurable ND has heightened. Aging is the leading risk factor for nearly all ND including Alzheimer's disease (AD) and Parkinson's disease (PD) (Hou et al., 2019). The number of people being afflicted by AD is anticipated to surge up to 135 million by 2050 because AD alone can affect between one-third and one-half of people above the age of 85 years. NDs are expected to have disastrous repercussions on individuals, families, and societies unless efficient aids are discovered to minimize the progression of these diseases.

Over the past century, NDs have generated distinctive and convincing challenges to effective drug discovery. In America, AD and PD are the two uttermost prevalent NDs with 5 million Americans existing with AD as well as more than 500,000 people diagnosed with PD (Karlawish et al., 2017). Yet another group of people comprising millions more are affected with rare NDs, such as amyotrophic lateral sclerosis (ALS), multiple sclerosis (MS), Huntington's disease (HD), frontotemporal dementia (FTD), and spinal muscular atrophy (Katsnelson et al., 2016; Correale et al., 2017). The healthcare cost of contrasting dementias and AD scores for over US$200 billion, an amount presumably to escalate by 2050 if these disorders persist to be unrecoverable (Barnes, 2021). It has been unveiled that there is no cure for MS even though as many as 9 immunomodulatory compositions have reached FDA approval for MS since 2000. This is shockingly diverse from the instance that even though the number of AD patients is pondered to approximately double in the following 10 years, only four non-disease-modifying compounds were passed for AD during that equivalent period (Crismon, 1994; Cusi et al., 2007; Birks and Evans, 2015; Birks and Harvey, 2018). Besides the overwhelming load of AD and other NDs on our healthcare system touching a bothersome level and unfulfilled efficacious cure, the urgency for the well-timed creation of competent therapies has been increasing bit by bit.

Currently, treatments accessible for NDs can barely handle the symptoms or terminate the progression of the disease (Durães et al., 2018). The drug discovery process right from target identification and validation to licensed use of a drug is a daunting task that comes with a long gestation period. DR (DR) is a present-day trending strategy that overcomes several shortcomings of the denovo development of entirely new drugs. It speeds up the discovery process and is efficient, economical, riskless, and reduces the failure rates in the clinical development and testing phases (Tanoli et al., 2021). With the expanding necessity for the treatment of NDs and the commitment given by DR, it makes sense that old drugs are being used as new treatments for these diseases. Nonetheless, the foremost issue in drug repositioning is tracking down novel drug-disease relationships. To deal with this issue, there are a range of approaches and two cardinal strategies of DR, viz., on-target and off-target (refer Figure 1) (Rudrapal et al., 2020). In on-target (target-centric) DR, the pharmacological mechanism of a drug molecule that is previously established is correlated to a new therapeutic implication. In this plan of action, the biological target of the drug candidate is unaltered, but the ailment is dissimilar. It incorporates computational approaches, biological experimental approaches, and mixed approaches (Ferreira and Andricopulo, 2016). On the other way around, in the off-target (drug-centric) profile, the pharmacological mechanism of a drug candidate is unrecognized. Drugs and drug candidates respond to new targets, out of the original scope, for afresh curative indications. Consequently, the targets along with the indications are unique (Ashburn and Thor, 2004). In the sphere of DR emphasis is given to three significant stages: procreation of candidate compounds, preclinical analysis, and clinical trial. For the production of candidate compounds, it's of high priority to determine relevant drugs for potential remedial indications. Notable advances have been made in the understanding of neurodegerative disease biology. Likewise, a plethora of fresh accessible resources has simplified drug

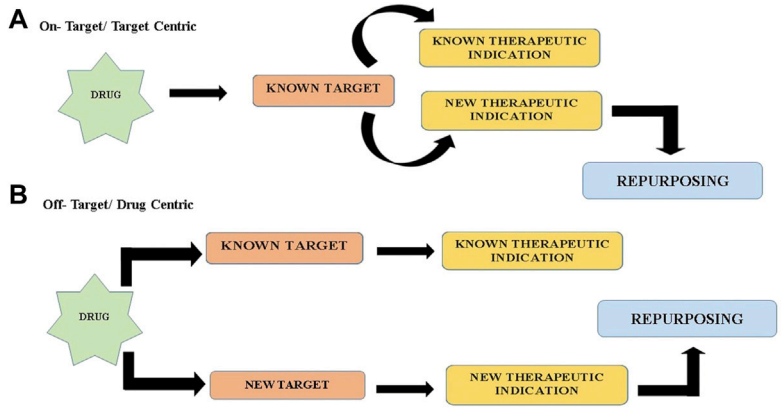


FIGURE 1
Two cardinal strategies of drug repurposing **(A)** On target/target Centric **(B)** Off target/Drug Centric.

discovery attempts through the medium of drug reprofiling. These incorporate bounteous data from clinical, mechanistic and epidemiological research, development of biomarkers, and a number of well-validated models, both cell and animal-based. Nowadays, the most prevailing drug reprofiling approaches in NDs are predominantly grounded on ad hoc clinical and epidemiological risk assessment in human testing and preclinical alterations in rodent models (Ashburn and Thor, 2004). However, for the accomplishment of superior DR in NDs, more precise and standardized approaches for both activity-based and computational methods should be put into effect. In conjunction with swift advancement in the scientific study of disease, the accessibility to contrasting sophisticated tools available in genomics and bioinformatics and assured clinical drug libraries will immeasurably hasten and promote future endeavours in neurodegenerative disease drug repositioning. For exploration of novel therapeutic liabilities for neurodegenerative disease, two alternative and complementary approaches perhaps be applied widely, one is activity-based/experiment based phenotypic screening and the other is theoretical/in silico-based/computational approaches (Rudrapal et al., 2020). DR can also be approached through a combination of both fields.

2.1 Experiment-based approaches

When it comes to the series of actions in drug discovery and drug repurposing, the experiment-based/experimental screening approaches are frequently supposed to be the fundamental step. It refers to the identification of original compounds for new pharmacological utilization entrenched on experimental assays. It necessarily blends protein target-based and cell/organism-based screens *in vitro* and/or *in vivo* disease models without necessitating the employment of every structural data of biological target proteins. In this approach, structural data of target proteins as well as the drug-induced cell/disease phenotypic information is not mandatory. The activity-based approach is also time and labor-consuming and amid the screening process, the generation of false positive hits is low. Experimental repositioning comprises a handful of approaches, essentially the cell assay approach, target screening approach, animal model approach, and clinical approach (Lionta et al., 2014; Oprea and Overington, 2015).

Affinity chromatography and mass spectrometry are two broadly operated proteomic techniques in analyzing drug candidates (Brehmer et al., 2005). In the present age, drug target analysis along with drug repositioning are entangled. DR is distinctive from drug discovery in terms of modification of drug targets. The affinity of drug ligands can be predicted using a cellular thermal stability assay which can map the contact patterns of intracellular targets (Molina et al., 2013). Utilizing this method, a considerable number of molecular on and off-targets have been divulged for numerous clinically approved drugs. New biological targets of well-known drugs are derived *via* affinity matrices chiefly observed in the area of kinases (Klaeger et al., 2016; Scott et al., 2016).

2.2 *In-silico* approaches

To accomplish effective therapies for neurodegenerative disease and get the therapies to the clinic faster, computational drug repurposing, or the in silico screening of FDA-approved compounds is advantageous. For investigating drug-target binding kinetics and drug residence times of prevailing drugs or drug candidates, using the computer as assistance for molecular docking is a notable approach (De

Benedetti and Fanelli, 2018). In silico/computational drug reprofiling, simulated screening of public databases of mountainous drug/chemical libraries is executed by adopting computational biology and bioinformatics/cheminformatics tools. In this approach, the potential bioactive molecules are identified based on the molecular interaction between the drug molecule and protein target (Talevi, 2018). This calls for structural data of target proteins and drug-induced cell/disease phenotypic data. In-silico based approach is time and labor efficient and has a higher rate of false positive hits during the screening.

For many neurodegenerative disorders, it should be considered that drugs look for satisfactory penetration into the blood-brain barrier (BBB). The two sections for curative means of accessing brain targeting are invasive and non-invasive categories (Alam et al., 2010; Gabathuler, 2010). The invasive category encompasses the transitory rise in BBB permeability, and the non-invasive category primarily engages in the transformation of drug molecules *via* a physiological, chemical, or colloidal carrier system approach. Simultaneously, these methods are also connected with computational approaches.

Lately, the amalgamation of economically feasible large-scale computational capacity with high-throughput clinical, molecular, and structural biology technologies has constructed a modernistic and favorable circumstance to logically repurpose conventional drugs by adopting computational frameworks rather than chance findings. Currently available computational approaches/strategies to DR can be branched into molecular, clinical, and structure-based (biophysical) methods.

Intending to conclude drugs that may modify disease gene marks, molecular approaches have opted which aims to match the drug-gene expression marks pre-and post-drug treatment with disease gene expression marks. It does not depend on prior recognition of the target molecule for high-throughput screening of existing compounds. Currently, resources such as CMap (Connectivity Map) and LINCS are limited in the case of neurodegenerative disease. Molecular approaches of computational drug repositioning integrate genetic, epigenetic, proteomic, transcriptomic and metabolomics evidence to determine promising and up-to-date indications for drugs. Additionally, techniques such as network integration, correlating gene expression profiles amidst a disease model and drug-treated condition, prediction of drug-protein interactions, and implementation of genotype-phenotype associations are also being practiced (Yang and Agarwal, 2011; Chen et al., 2017; Luo et al., 2017). There is an enormous demand for the generation of databases based on transcriptomic drug perturbation in CNS tissues to ascertain the drug response to inappropriate tissue and cell types for neurodegenerative disease. Recently, for AD (AD), a proteotranscriptomic-based computational drug repositioning method named Drug Repositioning Perturbation Score/Class (DRPS/C) resulted based on inverse associations between disease-induced or drug-induced gene and protein perturbation patterns (Lee et al., 2020). Another such instance in the matter of ND is the work by Zhang et al. where the National Human Genome Research Institute-European Bioinformatics Institute Genome-Wide Association Study catalog, PubMed, and the Human Metabolome database were precisely extracted to generate an assembly of proteomic, metabolomics, and genetic signatures of AD (Zhang et al., 2016; Wishart et al., 2018; Buniello et al., 2019). By commixing this multi-omics data with the Therapeutic Target database and Drug Bank drug-target databases, the authors of the study were capable of illustrating a list of 75 drug predictions in AD (Wishart et al., 2006; Li et al., 2018).

In clinical methods of drug discovery and repurposing, large-scale health data such as the electronic medical record (EMR), insurance claims data, clinical trial data, health registries, health surveys, and personal genome testing companies are engaged as a supreme asset. Mount Sinai BioMe cohort and the eMERGE network are two notable illustrations of EMR databases. Meticulous medicine approaches can be utilized with the aid of an abundant sample size. It is effortless to identify drugs that are efficacious in indications other than the primary drug use by taking the patient medication history as an asset. For instance, the latest reconsideration of human trials and Medicare pharmacy claim specifics has recommended that when compared to nonuser counterparts, statin users experience a lower incidence of AD (Geifman et al., 2017). Likewise, utilizing EMR laboratory testing data from Ajou University a group of researchers compared the 'clinical signatures' or laboratory test values of patients before drug administration and following drug administration and found two therapies for Kawasaki syndrome that is terbutaline sulfate and ursodeoxycholic acid evoked identical changes in laboratory values. Correlating the disease pairs disclosed that there is a significant extent of resemblance in clinical signatures between Kawasaki syndrome and Amyloid lateral sclerosis (ALS), advocating that terbutaline sulfate can be competent in treating ALS besides Kawasaki syndrome. One of the shortcomings of clinical methods is that before analysis clinical data must be changed into a structured database. Moreover, EMR evidence is oftentimes inadequate and cluttered. In the event of neurodegenerative disease patients are to be longitudinally outlined and for NDs with lengthy disease courses it's strenuous to track the physical and mental wellness and consequences. Also for genetic subtype-specific drug repurposing, clinical data should be paired with genetic data.

However, substantial improvement has been made in the computerized recovery of knowledge from unstructured EMR data (Ford et al., 2016; Delespierre et al., 2017). Recently, Observational Medical Outcomes Partnership (OMOP), has been simulated by the Observation Health Data Sciences and Informatics program. OMOP is a universally accepted scheme to

transform claimed information and EMR record data into a uniform and consistent data format with familiar data representations essentially terminologies, coding schemes, etc. (Hripcsak et al., 2015). As a result of mutable data coding and formatting, consecutive statistical analyses can be intended with the slightest information loss. There are alternative linkage procedures that include probabilistic matching strategies and 'fuzzy' matching techniques and these techniques take advantage of multiple field values to compare records even when no single field is an exact match (Dean et al., 2001; Malin and Sweeney, 2005).

In biophysical methods, drug-target predictions can be accomplished by taking biochemical characteristics of drugs into accounts such as binding affinity or biophysical properties like 3D conformation (Holdgate et al., 2013; March-Vila et al., 2017). These methods comprise structural, ligand-based, and molecular docking methods and possibly be principally advantageous in NDs such as HD with well-established targets (Nance, 2017). Structural methods utilize the complete advantage of 3D protein configuration data to determine structurally identical drugs that might conceal similar targets (March-Vila et al., 2017). Structural methods employ local site similarity metrics to describe protein binding sites or those that identify two protein environments that can bind the same ligand that is chemiosmotic protein environments (de Franchi et al., 2010; Jalencas and Mestres, 2013). If the hypothesis is such that two diseases share similar target proteins, then a structurally similar molecule/drug may be dynamically useful in both diseases. This can be illustrated by the fact that patients with AD and HD both have marked extra synaptic NR2B subunit-containing N-Methyl-D-aspartate receptors (NMDARs) and increased phosphorylation of NMDARs (Song et al., 2003; Hoe et al., 2009). Establishing drugs that hinder the extra synaptic NMDAR activity using addressing structurally analogous ligands or binding sites depicts a credible strategy for DR in both of these conditions (Ehrnhoefer et al., 2012).

Ligand-based methods presume that two molecules may share similar targets if they share a similar bioactivity profile. To verify innovative targets for conventional drugs/compounds, ligand-based methods pay attention to chemical and biological knowledge such as binding affinity; cellular activity; absorption, distribution, metabolism, and excretion data (Gregori-Puigjane and Mestres, 2008; March-Vila et al., 2017). Ligand-based methods entrust public bioactivity databases such as PubChem, DrugBank, and ChEMBL in opposition to structure-based methods, Docking-based methods implement molecular docking simulations either to predict promising drugs for a given target or novel targets for existing drugs (Kitchen et al., 2004). One such example of docking-based repurposing is to single out droperidol as an established drug in AD by the application of high-throughput ligand–protein inverse docking due to droperidol's high binding affinity to seven AD target proteins (Xie et al., 2016). Although biophysical methods are competent in drug repositioning, they look for prior labeling of target molecules and demand crystallographic evidence of target and drug molecules.

In recent years, several companies are developing and elaborated Artificial intelligence (AI) and machine learning (ML) based frameworks for drug discovery. These methods are exceptionally proficient at linking diverse classes of data. There has been a blooming diversion towards the evolvement of ML techniques to efficaciously dig for transcriptomic, structural, and clinical data (Mani et al., 2012; Kadurin et al., 2017; Shameer et al., 2017; Butler et al., 2018; Wang et al., 2018; Smith et al., 2019). IBM adopted AI-based text-mining approaches to constitute a semantic model of ALS-associated RNA-binding proteins that may exemplify drug targets. BM could uncover potential ALS-associated RNA-binding by application of this model to a new set of RNA-binding proteins (Bakkar et al., 2018).

2.2.1 Artificial intelligence/machine learning algorithms

In recent years, several companies are developing and elaborated Artificial intelligence (AI) and machine learning (ML) based frameworks for drug discovery. These methods are exceptionally proficient at linking diverse classes of data. There has been a blooming diversion towards the evolvement of ML techniques to efficaciously dig for transcriptomic (Wang et al., 2018; Smith et al., 2019), structural (Kadurin et al., 2017; Butler et al., 2018; Popova et al., 2018), and clinical (Shameer et al., 2017; Nemati et al., 2018). ML is one of the forms of artificial intelligence. It does facilitate vigorous interrogation of multiple datasets by using statistical techniques to determine formerly undetected associations and patterns in the data and in the recent past the approaches have been presenting promising outcomes when applied to drug repurposing of neurodegenerative diseases (Myszczynska et al., 2020). Machine learning algorithms are chiefly classified into supervised, unsupervised and reinforcement learning approaches (Bharadwaj et al., 2021). The ongoing methods most frequently applied to neurodegenerative disease-related data are the supervised machine learning algorithms. IBM adopted AI-based text-mining approaches to constitute a semantic model of ALS-associated RNA-binding proteins that may exemplify drug targets. BM could uncover potential ALS-associated RNA-binding by application of this model to a new set of RNA-binding proteins (Bakkar et al., 2018). In a study, a novel computational approach was reported to predict drug repositioning grounded on a ML algorithm and data integration. The approach in the study relied on the persistent analysis of classification mismatches as genuine reclassifications opportunities. The definiteness of the results were of high levels and were rational with several literature reports (Napolitano et al., 2013). In another

study, a novel method "PREDICT" was presented which was based on the observation that drugs that are similar can also be indicated for similar disease (Gottlieb et al., 2011). The method obtained tremendous specificity and sensitivity, more desirable than the existing methods in predicting the large-scale drug indications for both approved drugs and novel molecules. In recent years, it has been a remarkable preference to pave the way for novel computational approaches and deep learning (DL) methods is one such example which commits to intensify the capableness of drug repurposing methods. Approaches known as deep neural networks (DNNs) are adopted by DL which encompasses artificial neural networks including plentiful hidden layers between the input and output layers (Ma et al., 2015). One instance of a work that selected deep learning was in which the authors confirmed how DNN trained on gigantic transcriptional response datasets can assort different drugs to therapeutic categories solely established on their transcriptional profile (Aliper et al., 2016). Additionally, favourable outcomes were obtained by means of a deep learning-based algorithmic framework termed as DeepDTIs (Drug target interaction) which ascertained drug-target interactions using chemical structures and known interactions. (Wen et al., 2017).

2.2.2 Network-based methods

By virtue of immense present-day progress in the sphere of system biology has led to the progression in applications such as drug repurposing. Networks are clear, understandable and flexible data structures on which associations can be implied using many statistical and computational approaches. The perception of interaction network is massively engaged in biology. In network models pairwise relations between various objects is exhibited. Schematically, in such networks, nodes are represented by entities (genes, proteins, complexes, metabolite, disease), while edges represent interactions or relationship between two nodes such the relationship between drugs and known gene targets and large number of diverse connections between two nodes can be displayed concurrently (Savva et al., 2019). Despite of the potency of such approaches has been verified for considerable times with drug-target interaction prediction, these methods are afflicted by the deficiency of current knowledge on molecular interactome, leading to noisy results. Network-based drug repositioning methods can be organized into categories based on their main source of biological data: 1) gene regulatory networks, 2) metabolic networks, and 3) drug interaction networks (Approaches et al., 2019). Moreover, a fourth category, integrated approaches, using multiple data sources simultaneously, can also be supplemented.

For example, a recent work proposed a untried bidirectional drug repositioning approach that comprised of Top-down and Bottom-up approaches and eventually provided information about significant repositioning drug candidates (Rakshit et al., 2015). This method takes into account tripartite indication-drug-target network (IDTN), also considering the topological significance (choosing most potent drugs based on seven topological parameters, such as degree, betweenness, centroid, closeness, eccentricity, radiality, and stress, which are basic network measures used to analyse a network) of drugs. A separate study proposed a different approach based on a two-pass random walk with restart on the drug-disease heterogeneous network, referred to as TP-NRWRH, to predict new indications for approved drugs (Liu et al., 2016). It was applied on three different types of networks, that is, integrated drug-drug similarity, disease-disease similarity, and drug-disease networks. This method was evaluated and in case study on the AD it showed that nine of top 10 predicted drugs have been approved or are investigational for neurodegenerative diseases.

2.2.3 Genome-wide association studies-based methods

Another robust tool for drug repurposing is the utilization of genomics technologies. For the past few years, genome wide association studies (GWAS) has been another source of data which is being exploited for new information regarding the association of specific genomic variations known as single nucleotide polymorphisms (SNPs), with complex trait human diseases, such as AD, multiple sclerosis, etc. (Savva et al., 2019). GWAS can distinguish thousands of SNPs synchronously and these data are used by researchers to detect genes that are linked with a specific disease trait and to analyse how these variations affect responses to drugs. Furthermore, GWAS can be indicated to identify alternative indications for existing drugs rapidly and systematically (Hurle et al., 2013). However, objections such as inadequacy of data regarding whether an activator or inhibitor is needed to observe an effect, makes it burdensome to use GWAS information alone. While applying GWAS to initiate repurposing of drug candidates, the basic process is to analyse the catalogue of SNPs linked with the disease to determine a subgroup of genes that are speculated to be drug targets according to the drug ability of the gene's product. Thereafter, process demands to select which of these gene products, if any, are targets for the drugs that are in the pharmaceutical channels at that instant. One such illustration detected by this approach is a clinical candidate Biib-033 (Biogen Idec, Cambridge, MA, USA), which is an antibody targeting the leucine-rich repeat and immunoglobulin domain-containing 1 (LINGO-1), which was developed for multiple sclerosis. Two GWAS studies detected LINGO-1 as a target for essential tremor, which is a neurological disorder, propounding that it could be repurposed for vital tremor ailments (Gudjonsdottir et al., 2009; Clark et al., 2010). .

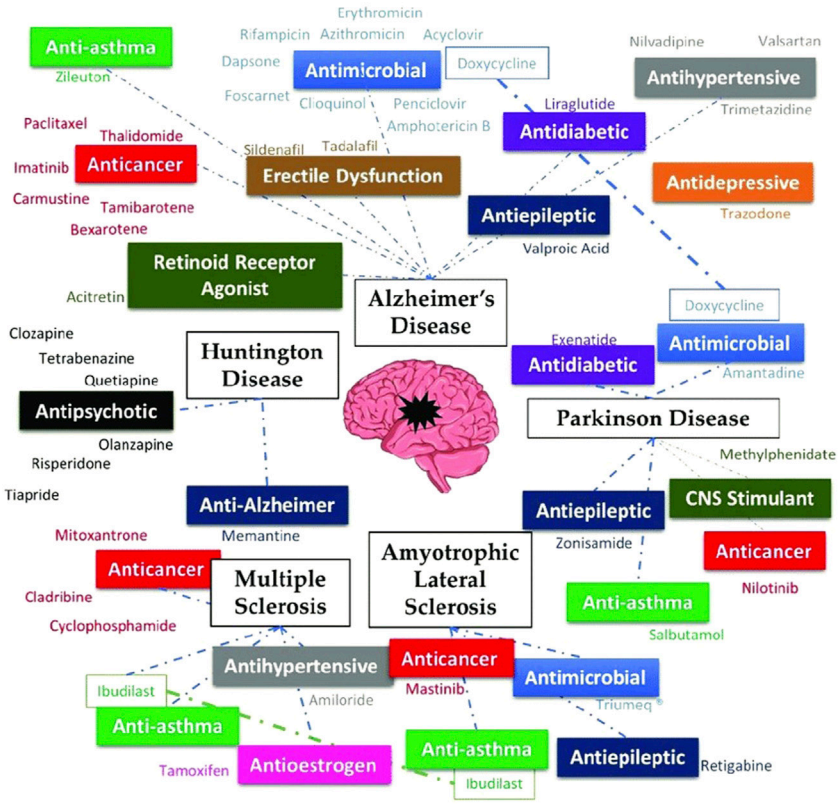


FIGURE 2
Summary of a few drugs repurposed for neurodegenerative diseases, adapted from (Durães et al., 2018) *via* CC by 4.0 license.

3 Drug repurposing for neurodegenerative diseases

Diseases that affect the central as well as the peripheral nervous system, are known as neurodegenerative diseases (NDs). More than 600 distinct neuropathological illnesses exist, which include stroke, Parkinson's disease, brain tumors, and epilepsy. Considering that the global population is growing, there are more NDs than ever before (Siuly and Zhang, 2016; Matilla-Dueñas et al., 2017; Kumar et al., 2021). In the next 20 years, neurodegenerative disorders that impact motor function will overtake cardiovascular disease as the second most common cause of mortality, according to the World Health Organization. No ND is currently curable due to its poorly understood molecular basis, and the medicines available merely treat the symptoms or slow the disease's course (Onyango et al., 2021). Since the medicine's pharmacokinetic and pharmacodynamic properties are already known, DR is the most beneficial new technique for the creation of an effective treatment for NDs. The promise of old medications for the most important NDs, like Amyotrophic lateral sclerosis, Huntington's disease, Parkinson's disease, Multiple sclerosis, and Alzheimer's disease has been the subject of numerous studies (Durães et al., 2018). Figure 2 represents a summary of drugs repurposed for some neurodegenerative diseases.

3.1 Alzheimer's disease

AD accounts for 80% of occurrences of dementia in senior persons. The gradual memory loss, the incapacity to learn, and the deterioration in behavior and function are its signs. Although the exact pathology of AD is unknown, it is thought to be related to the buildup of amyloid-β plaques in the brain, which eventually cause neuronal and synaptic degeneration (Scheltens et al., 2016). The majority of AD medications are used to address cognitive impairments or other symptoms, and they work best when started early (Appleby et al., 2013).

Commonly prescribed drugs for AD are cholinesterase inhibitors viz. Galantamine, Donepezil, Rivastigmine etc. Galantamine, an alkaloid found in *Galanthus* species, has been researched as a potential treatment for peripheral neuropathies and myopathies. It has the potential to block

TABLE 1 List of repurposed drugs for AD.

Drug name	Earlier indication	Repurposed	References
Carmustine	It is a small, lipophilic, non-ionized nitrosourea molecule that can cross the blood-brain barrier and is employed as an alkylating agent in cases of brain cancer	Carmustine, at a non-toxic dose, demonstrated a significant reduction in amyloid-β development in cells overexpressing the precursor protein to the amyloid protein	Hayes et al. (2013)
Bexarotene	A retinoid X receptor antagonist is used to treat cutaneous T-cell lymphomas	In mice overexpressing familial AD mutations, it has been demonstrated to be effective at reversing neurodegeneration, enhancing cognition, and lowering amyloid-β levels	Tousi (2015)
Tamibarotene	It is an agonist of the retinoic acid receptor and is used to treat acute promyelocytic leukemia	It can influence a variety of pathways involved in the pathogenesis of AD, including those that control the release of pro-inflammatory chemokines and cytokines by brain cells, the behavior of animals with increased senescence, and cortical acetylcholine levels	Fukasawa et al. (2012)
Paclitaxel	It is an antimitotic drug authorized for the treatment of non-small cell lung cancer as well as ovarian and breast cancer	Although paclitaxel can be a substrate for P-gp and only penetrates a small portion of the central nervous system, it is particularly helpful in treating tauopathies because it reduces tau protein phosphorylation	Brunden et al. (2011)
Thalidomide	It prevents angiogenesis, endothelial cell growth, and blood-brain barrier disruption	Through the inhibition of tumor necrosis factor-α, it can minimize the death of hippocampus neurons	Ryu and McLarnon (2008)
Azithromycin, erythromycin	Macrolide antibiotics	They prevent the production of the amyloid precursor protein, which lowers the amyloid-β levels in the brain	Appleby et al. (2013)
Tetracyclines	Antibiotic (protein synthesis inhibitors)	It has been discovered that it encourages the destruction of fibrils and inhibits the synthesis of amyloid-β	Diomede et al. (2010)
Rifampicin	Use for *Mycobacterium* infections	It has shown results in the reduction of amyloid-β fibrils in a dose-dependent manner because of reduced production and enhanced elimination of amyloid-β	Tomiyama et al. (1996)
Acyclovir, penciclovir, foscarnet	antiviral drugs	In AD cell models, decreases phosphorylated tau protein and amyloid-β	Wozniak and Itzhaki (2010)
Amphotericin B	Antifungal drug	It has been demonstrated to slow down the production of amyloid-β (but posses toxicity)	Hartsel and Biochemistry (2003)
Clioquinol	Antifungal, Antiparasitic	In transgenic mice brains, it shows a reduction in the amyloid-β plaques	Grossi et al. (2009)
Valproic acid	Antiepileptic drug	Due to its ability to alleviate memory impairments and diminish the production of amyloid-β plaques in transgenic mice, it is recommended as a neuroprotective treatment for AD.	Smith et al. (2010)
Valsartan	Antihypertensive (angiotensin receptor blocker)	Chronic adverse stress, which can increase brain angiotensin II levels, is one of the main environmental factors of AD. Because it has been shown that angiotensin II increases are linked to amyloidogenesis, using angiotensin receptor blockers may be useful in delaying the loss of cognitive processing. Additionally, valsartan reduces inflammation, vasoconstriction, and mitochondrial dysfunction while encouraging acetylcholine release	Culman et al. (2002)
Trimetazidine	Anti-ischemic drug	It can penetrate the blood-brain barrier, lower free radical production, enhance axonal regeneration, and effectively myelinate both healthy and damaged axons	Hassanzadeh et al. (2015)
Liraglutide	Anti-diabetic drug	It demonstrated brain penetration and indicated physiological changes in the brain that improved learning and reduced the development of amyloid-β and inflammation in the brain	Mcclean et al. (2011)
Ghrelin	Peptide hormone (synthesized in the alimentary tract which controls appetite)	It has been shown that ghrelin, as well as its deacylated precursor, has neuroprotective effects by preventing programmed cell death and reducing the rise of interleukins induced by amyloid-β	Wagner et al. (2017)
Acitretin	Retinoid receptor activators	It reported an increase in antioxidant regulation and amyloid- β clearing enzymes	Tippmann et al. (2009)
Zileuton	Antiasthma drug	Zileuton, which inhibits 5-lipoxygenase, is thought to offer therapeutic benefits for AD. This is due to the finding that	Di Meco et al. (2014)

(Continued on following page)

TABLE 1 (*Continued*) List of repurposed drugs for AD.

Drug name	Earlier indication	Repurposed	References
		5-lipoxygenase is more prevalent in AD, creating it an exciting target within this context. Research using zileuton in mice revealed a decrease in amyloid-β accumulation	
Sildenafil/tadalafil	Erectile dysfunction drugs (inhibitors of phosphodiesterase-5) Phosphodiesterase-5 regulates cGMP, which in turn regulates memory problems caused on by amyloid-β	In aged mouse models, sildenafil was effective in reducing amyloid-β and suppressing neuroinflammation. Furthermore, Tadalafil showed neuroprotection and an increase of cognition	García-Barroso et al. (2013), Zhang et al. (2013)
Trazodone	Antidepressant	Trazodone has demonstrated potential in suppressing signaling *via* the PERK/eIF2α-P branch of the unfolded protein response, which is overactivated in AD patients and harms regulating translation s in cells	Halliday et al. (2017)

muscle acetylcholinesterase. Galantamine's ability to improve nerve impulse transmission also makes it useful for reversing neuromuscular blockade during anesthesia. During the 1960s through the 1980s, the majority of galantamine use was confined to Italy, Bulgaria, Germany, and France under the brand name Nivalin®. Galantamine's therapeutic properties for the treatment of AD were first investigated in the 1980s, and it was only in 2000 that it was included in the arsenal of drugs used to treat AD (Mucke, 2015). The production of misfolded proteins, oxidative stress, mitochondrial dysfunction, and impaired cell metabolism are only a few of the signaling pathways that may be involved in the pathogenesis of both cancer and neurodegeneration. The goal of the subsequent research was to see whether cancer medications may also be used to treat AD. Following these, investigations have been made to see if cancer medications can also be used to treat AD (Monacelli et al., 2017). Pathogens can enter the CNS in a variety of ways, depending on the organism, which may speed up the development of AD. The first is accomplished by a damaged BBB (Orgogozo et al., 2003). Some viruses, like the herpes virus, can go dormant after the original infection and then reawaken decades later in elderly people, causing delayed harmful complications (Nagarajan and Wilde, 2005). According to a 2020 study model, immunocompromised people who were exposed to *C. pneumoniae* through their noses developed Aβ plaque and NFTs in their olfactory cortex as well as in hippocampus (Sundar et al., 2020). Thus antimicrobials such as Rifampicin, Amphotericin B, acyclovir, penciclovir, foscarnet etc. (see Table 1) have also been researched to see whether they may be used to treat AD, specially its symptoms (Iqbal et al., 2020). Antidiabetics are also used to treat AD because type 2 diabetes has been established as a risk factor for the disease. According to studies, AD sufferers' brains have become less sensitive to insulin signalling. Insulin therapy has been shown to improve memory and cognition while also protecting the brain from damage and controlling the levels of phosphorylated tau protein. Additionally, insulin can promote cell growth, repair, and activation of neural stem cells. As a result, substances that affect insulin release may potentially be beneficial for AD. Analogues of glucagon-like peptide 1, which increase insulin production, may also have an impact on a number of AD-related processes, including tau phosphorylation, amyloid-β reduction, and impaired neuronal function and cell death (Perry et al., 2003; Zhao et al., 2004). Some drugs repurposed for AD are listed in Table 1.

3.2 Parkinson's disease

PD is a multifactorial neurological condition that impairs a patient's ability to move. Dopamine neurons in the putamen and caudate areas of the brain are the main targets of Parkinson's disease. Due to mitochondrial DNA deletion, elevated ROS and RNS generation decreased antioxidant function, and dopamine inhibition, the activities of mitochondria are reduced in the substantia nigra of parkinsonian brains (Ryan et al., 2015; Reeve et al., 2018). As dopamine is oxidized by both Monoamine oxidase (MAO) A and B, the level of dopamine drops in PD (Alexander, 2004). Primary tremor, akinesia, rigidity, bradykinesia, lack of postural instability, and secondary motor symptoms including the freezing of gait, micrographia, and speech issues are the hallmarks of Parkinson's disease (PD). In PD, non-motor symptoms include sensory impairment, autonomic dysregulation, neurobehavioral abnormalities, and sleep problems are also possible. Parkinson's disease is treated with levodopa, carbidopa, amantadine, rotigotine, dopamine agonists, Catechol-O-methyltransferase (COMT) inhibitors, anticholinergics Selegiline, rasagiline, safinamide, etc (Gupta and Shukla, 2021). The most recent therapy options for PD include newer dopaminergic medications, immunotherapies, drug repurposing, medications that target non-dopaminergic neurotransmitters, regenerative treatments, and deep brain stimulation. Many medications are currently undergoing clinical trials. Several medications, including the following, are being repurposed for PD: The antibiotic doxycycline, which has been investigated for its anti-PD effects after being once

TABLE 2 List of repurposed drugs for PD.

Drug name	Earlier indication	Repurposed	References
Amantadine	Anti influenza	As a mild glutamate receptor antagonist, it is used to treat Parkinson's disease (PD), boosting dopamine and preventing its reuptake	Lee and Kim, (2016)
Nilotinib	Tyrosine kinase Abl inhibitors, used to treat chronic myeloid leukaemia	It was found that α-synuclein build-up and increased α-synuclein expression are both signs of Abl activation in neurodegeneration. Nilotinib accelerates α-synuclein breakdown by preventing Abl phosphorylation	Pagan et al. (2016)
Zonisamide	Antiepileptic drug	Increased dosages revealed a reduction in intracellular dopamine. Both motor and non-motor symptoms have responded well to this medication, but its exact mode of action is yet unknown	Fox et al. (2018)
Methylphenidate	Central nervous system stimulant used to treat attention-deficit hyperactivity disorder	This medication has been found in numerous studies to be beneficial in lowering PD-related gait problems and non-motor symptoms	Devos et al. (2013)
Exenatide	Glucagon-like peptide-1 (used for type 2 diabetes)	It has proven to be capable of neuroprotection and beneficial neuroplastic change, which can stop or reduce the progression of the disease. It can cross the blood-brain barrier and offers neuroprotection by turning on GLP-1 receptors	Jankovic, (2017)

identified as a possible anti-AD therapeutic approach (Dominguez-Meijide et al., 2021). Differences in doxycycline concentration can distinguish between an antibacterial and an anti-inflammatory effect. Smaller concentrations than the ones used to treat microorganisms with antibiotics do not influence bacterial susceptibility, according to studies, but they do exhibit anti-inflammatory activity, which is connected to their neuroprotective effects. Doxycycline's antioxidant properties and its capacity to transform early species of α-synuclein oligomers (a presynaptic neuronal protein connected to PD genetically and neuropathologically) into non-toxic and non-seeding species are two additional ways that aid neuroprotection (Dominguez-Meijide et al., 2021). Only oligomeric species of α-synuclein have been discovered to bind to doxycycline, however, the physiological monomeric forms of α-synuclein are still present. Table 2 represents repurposed drug for PD. The anti-PD activity of antiasthma medications, specifically β2-adrenoreceptor agonists, has been researched. Recent research has connected the β2-adrenoreceptor to the control of the SNCA-synuclein gene. More particular, stimulation of the β2-adrenoreceptor was demonstrated to exhibit neuroprotection. Three anti-asthmatic drugs were investigated, and salbutamol, the one with the highest blood-brain barrier permeability, demonstrated the greatest promise. The conducted analysis revealed that all three medications were capable of lowering the abundance of SNCA-mRNA and α-synuclein (Mittal et al., 2017).

3.3 Huntington's disease

HD is characterized by dementia, behavioral and mental abnormalities, and involuntary choreatic movements (McColgan and Tabrizi, 2018). The multifunctional protein huntingtin (HTT) develops a mutant form as a result of a genetic mutation, which causes toxicity and causes neuronal death and malfunction. When a mutation in the HTT gene's exon 1 on chromosome 4p16.3 results in CAG (C-cytosine, A-adenine, and G-guanine) trinucleotide DNA segment extension, repetition, and multiplicity, HD develops. In a gene, the CAG segment is typically repeated between 10 and 35 times. However, due to mutations, more than 36 CAG repeats are produced, which results in the genesis of HD (Tabrizi et al., 2020). The slow degeneration of neurons in the basal ganglia, particularly the caudate nucleus and putamen to the cerebral cortex, signals the beginning of HD (Kshirsagar et al., 2021). The symptoms of HD begin to appear in adults, and they worsen with time until they eventually result in death within years. The sole alternative is to control the symptoms since there is no known cure for this illness s (Roos, 2010).

Tetrabenazine was initially created as a result of research into the design of straightforward drugs with reserpine-like antipsychotic action. It functions as both a mild blocker of the D2 dopamine postsynaptic neurons and a highly selective, reversible inhibitor of monoamine absorption by presynaptic neurons. Research on this substance as an antipsychotic was conflicting, thus this medication was repurposed for conditions like HD that are characterized by abnormal, involuntary hyperkinetic movements. Tetrabenazine has never been shown to elicit signs of dyskinesia, making it a safer drug to use in HD than dopamine receptor blockers (Paleacu, 2007). For the treatment of HD, several medications with dopamine antagonistic action have been investigated. This is the situation with the antipsychotic drug tiapride, a D2 receptor antagonist. Selegiline, however, is a popular option for the treatment of Huntington's chorea in Europe (Roos et al., 1982). A neuroleptic medication called clozapine is used to treat schizophrenia. With little antagonistic activity toward the D2 dopaminergic receptors, it exhibits a high affinity for the D1 and D4 dopamine receptors. Although clinical trials had mixed outcomes, it was recommended as a good symptomatic medication for chorea due to its low prevalence of

TABLE 3 List of repurposed drugs for HD.

Drug name	Earlier indication	Repurposed	References
Clozapine	Neuroleptic drug	Although clinical trials had mixed outcomes, it was recommended as a good symptomatic medication for chorea due to its low prevalence of extrapyramidal side effects	Bonuccelli et al. (1994)
Tetrabenazine	Intended to have antipsychotic effects but produced conflicting success	Repurposed to treat HD symptoms, it functions as a mild blocker of D2 dopamine postsynaptic neurons and a high-affinity, reversible inhibitor of monoamine uptake by presynaptic neurons	Paleacu (2007)
Olanzapine	Antipsychotic drug	It is routinely prescribed for the treatment of HD's motor and behavioural symptoms. Although this medication has a strong affinity for serotonin receptors, it is antagonistic to dopamine D2 receptors	Paleacu et al. (2002)
Risperidone	Antipsychotic drug	It is used to treat schizophrenia and bipolar disorder as a D2 receptor antagonist and serotonin agonist, and it can also be used to treat HD chorea	Duff et al. (2008)
Memantine	Used to treat AD.	Investigation into memantine's efficacy for treating HD revealed that it could lower neurons' sensitivity to glutamate-mediated excitotoxicity	Beister et al. (2004)

TABLE 4 Some repurposed drugs for ALS and MS.

Drug name	Earlier indication	Repurposed	References
Masitinib	Tyrosine kinase inhibitor (used to treat canine cancer)	Tyrosine kinase inhibitors may be effective against the aberrant glial cells that grow in ALS, explaining their usage in the disease	Trias et al. (2016)
Triumeq® (dolutegravir + abacavir + lamivudine)	An antiretroviral Drug used in anti-HIV therapy	Based on the fact that ALS patients had reverse transcriptase blood concentrations comparable to HIV-infected patients and that a human endogenous retrovirus was found to be expressed in the brains of ALS victims, this medicine was investigated for the treatment of the disease	Clinicaltrials (2022)
Retigabine	Anti-epileptic drug (causes membrane hyperpolarization by attaching to voltage-gated potassium channels, which increases the M-current.)	Because it is believed that neurons in this condition are hyper-excited and fire more frequently than usual, ultimately leading to cell death, it can promote motor neuron survival and lower excitability, which is beneficial in the treatment of ALS.	Wainger et al. (2021)
Tamoxifen	An antioestrogen drug (authorized for use in breast cancer chemotherapy and chemoprevention)	The discovery of neurological improvements in patients and disease stability in ALS patients who had breast cancer treated with tamoxifen led to the drug's accidental repurposing for the treatment of ALS.	Chen et al. (2020)
Mitoxantrone	An anthracenedione that has been proven effective in the treatment of breast and prostate cancer, acute leukaemia, and lymphoma	Mitoxantrone has also been licensed for the treatment of MS due to its immunosuppressive properties, which are connected to variable responses of the T- and B-cells in the central nervous system to antigens, myelin degradation brought on by macrophages, and axonal lesions	Fox (2004)
Cyclophosphamide	An alkylating agent treatment of leukaemia, lymphomas, and breast carcinoma	Cyclophosphamide is used in MS because it can have an immunosuppressive and immunomodulatory effect. Additionally, cyclophosphamide has good absorption in the central nervous system and can cross the blood-brain barrier	Awad and Stue (2009)
Amiloride	A diuretic medication	Amiloride can prevent the neuronal proton-gated acid-sensing ion channel 1 (ASIC1), which is overexpressed in axons and oligodendrocytes in MS lesions, from having its neuroprotective and myeloprotective effects. A further benefit of amiloride's preventive action occurring later in the course of inflammation is that it makes it active even before inflammation begins	Arun et al. (2013)
Ibudilast	Phosphodiesterases inhibitor used for bronchial asthma and cerebrovascular disorders	Ibudilast can prevent the brain's microglia and astrocytes from releasing tumor necrosis factor, which reduces neuronal degeneration. It is also helpful in MS because it can prevent oligodendrocyte apoptosis, suppress astrocyte apoptosis, and prevent demyelination	Barkhof et al. (2010)

extrapyramidal side effects (Bonuccelli et al., 1994). Another antipsychotic medicine, olanzapine, is frequently recommended to treat HD's behavioral and motor symptoms. While antagonizing dopamine D2 receptors, this medication has a high affinity for serotonin receptors. It can be advised when irritation, sleep issues, weight loss, and chorea are present because it is safe and well tolerated (Paleacu et al., 2002). As a D2 receptor antagonist and serotonin agonist, the antipsychotic risperidone, which is used to treat schizophrenia and bipolar disorder, can also be used to treat HD chorea. It demonstrated positive results in stabilizing mental symptoms and motor deterioration (Duff et al., 2008). Quetiapine, an atypical antipsychotic, has a strong affinity for dopamine and serotonin receptors. Even though there haven't been many instances of quetiapine being used to treat HD symptoms, those have emphasized the drug's value in treating chorea, particularly when it's coupled with psychiatric symptoms (Alpay and Koroshetz, 2006). An adamantane derivative called memantine is used to treat AD. It is an inhibitor of N-methyl-D-aspartate (NMDA) that is non-competitive. A large influx of calcium enters the cell as a result of excessive NMDA receptor stimulation, which ultimately results in cell death. Memantine can therefore stop this calcium influx in neuronal cells and stop the death of brain cells. When memantine's effectiveness in treating HD was investigated, it was shown that it could lessen the susceptibility of neurons to glutamate-mediated excitotoxicity (Beister et al., 2004). Table 3 represents list of repurposed drugs for HD.

3.4 Other neurodegenerative diseases

Upper and lower motor neurons, which regulate the voluntary muscles, die as a result of the condition known as ALS. Muscles eventually weaken and shrink as a result, which causes muscular atrophy. Other signs include difficulty breathing, swallowing, speaking, and twitching or rigid muscles. Most ALS causes are aetiologically unknown, with genetic inheritance accounting for roughly 10% of cases (Kiernan et al., 2011). Only two medications, edaravone, and riluzole, are presently accessible to postpone the development of the illness, albeit they cannot reverse the symptoms once they have appeared (Zoccolella et al., 2007; Sawada, 2017). Another autoimmune condition affecting the central nervous system is MS. It is a protracted, inflammatory disorder in which the myelin and axons are partially or completely damaged. Its progression is uncertain, and its early symptoms include temporary neurological impairments that eventually turn severe. There is currently no approved treatment for MS, however, there are medications that can slow the disease's progression and symptoms (Trapp and Nave, 2008). Several drugs are currently being repurposed for the treatment of ALS as well as MS. Table 4 represents some drugs that are under clinical trial for ALS or MS.

3.5 Unsuccessful repurposed drugs for neurodegenerative diseases

Even though there have been numerous instances of pharmacological repurposing, numerous attempts at repositioning have also been unsuccessful. A drug may look promising in computational analyses or *in vitro* assays but not *in vivo*, requiring the investigation of the medicine to be stopped in favor of other activities. This was the situation with latrepirdine, an antihistamine that was repurposed for AD and HD after being licensed in Russia for the treatment of rhinitis brought on by allergies. Despite the lack of a characterized mechanism of action, it had been suggested that it might alter the activity of channels and neurotransmitters, avoiding amyloid toxicity among other things (Bezprozvanny, 2010). In actuality, phase III studies unsuccessful to find any appreciable variations in the course of the disease, despite phase II research showing improvement in AD patients related to placebo (Doody et al., 2008). There have also been attempts to employ the anti-hypercholesterolemic medications simvastatin and atorvastatin for AD. This notion was developed in response to the important finding that cardiovascular illness and AD frequently co-occur. Studies had demonstrated that statins could raise neuroprotection and reduce amyloid-β levels, among other positive benefits. However, none of them were effective in the management of AD (DL et al., 2005; Sano et al., 2011). Studies evaluating the use of selective serotonin reuptake inhibitors, commonly used as antidepressants, in the treatment of AD have also been conducted. Although nortriptyline and paroxetine originally showed an improvement in cognitive abilities, subsequent analyses revealed that there was no improvement in cognitive behavior even after these medications had addressed mood disorders (Nebes et al., 2003).

In phase II investigations for the treatment of ALS, the antibiotic ceftriaxone seemed promising, but it also failed to demonstrate clinical efficacy in phase III tests (Cudkowicz et al., 2014). Even cladribine was initially rejected as an MS treatment before it was approved (Leist and Weissert, 2011). Even though DR is encouraging in the creation of new treatments for ND, the approval procedure can be challenging and frequently leads to the failure of repurposing initiatives.

5 Opportunities and challenges of drug repurposing

Owing to its proficiency in sparing time and cost, drug repositioning has become a crucial method for exploiting new therapeutic implications of current drugs or drug candidates. Such an ingenious type of approach will undeniably accelerate the drug development process. Concurrently, in the case of neurological diseases, some restraints need to be considered during the process of drug repositioning. Repurposing drugs

experiences humongous challenges due to which there are limitations in the market for repurposed drugs. A single phase III clinical trial of a repurposed drug for AD can cost up to 300–400 million dollars (Shineman et al., 2014). This demonstrates even though repurposed drugs can deviate from the initial development stage and safety testing, they demand profound high-risk extravagant clinical trials to establish efficacy. As a result of the sluggish progression of neurodegenerative diseases, clinical trials might take a long duration. Further, apart from proving the drug penetration into the brain, many times drugs must be tested for safety issues in geriatric populations who periodically have comorbidities and undergo treatment that may interreact with the repurposed drug. DR may be difficult in neuropathological states considering its complex molecular and cellular signaling mechanisms. The reason that drugs respond to multiple targets despite affecting a single target might accelerate the risk of a range of adverse reactions (Vogt and Mestres, 2010). An all-inclusive evaluation of the assets as well as lacking these adverse effects can assist us to figure out drug repositioning from a more multifaceted perspective (Reddy and Zhang, 2013). Other challenges in repurposing drugs include limited or no patent protection or patent life, commercialization, and reimbursement challenges.

In pursuance of overcoming obstructions encountered in the course of drug repurposing, we can consider several proposals. In the first place, it is inspirited to furnish more financial support in conjunction with technical assistance for clinical trials of drugs to be repurposed. Pharmaceutical companies are exceedingly doubtful to finance human trials of approved drugs to be repositioned unless there is a viable commercial strategy. This generates a favorable circumstance where government and foundations can take the eagerness to do something. Currently, several groups are taking a large interest in funding pilot trials of repurposed drug candidates with the hopes of paying more impetus to drug repurposing. Foundations such as 'Cures Within Reach' are entirely centralized on aiding repurposing studies. The MJFF, ADDF, Cure Parkinson's Trust, Alzheimer's Society (United Kingdom), the Multiple Myeloma Research Foundation, and others have financed repurposing trials. In association with government initiatives, various academic centers are also heading the repurposing attempts. Secondly, to augment data sharing it is crucial to constitute an exhaustive data analysis platform. The enormous volume of data piled up by approved drugs or drug candidates for clinical trials can be stored in an assorted manner and can be unlocked and reanalyzed adopting Information science services and artificial intelligence. The bottleneck in the research process is that data derived from biological databases and human trials are massive and perplexed and the conventional data processing methods cannot work out with it. We can unquestionably improve our understanding of the disease from this big data and make more accurate disease-related strategies. Nevertheless, there is a considerable breach between producing biomedical data and data analysis. Expertise needs to find technical clarifications to ensure the efficiency of research with less energy and time. Finally, it is fundamental to resolve patent restrictions and take judicious surveillance in pursuance of facilitating the DR process. The utilization of drug reprofiling should be backed by a risk handling strategy and the drug's safety assurance can be established by clinical trial information or data from post-marketing surveillance.

6 Conclusion

In recent years, many repurposed drugs have found their place as potential agents to treat various neurodegenerative diseases. As already discussed many companies are developing and elaborating the strategic advantages of using Artificial intelligence (AI) and machine learning (ML) based frameworks for drug discovery in this segment. Despite these advancements, there are threats to the precise analysis of existing pre-clinical and clinical evidence concerning particularly from regulatory and scientific perspectives. Apart from focussing on the efficacy of the newly repurposed drugs, robust post-authorization studies are equally important.

Author contributions

All authors contributed to the design of the article. RB, NA, and BBK wrote the article. RB and NA edited the article and contributed to the interpretation of the included papers. All authors have read, reviewed, and approved the final paper. All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

Conflict of interest

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

Publisher's note

References

Alam, M. I., Beg, S., Samad, A., Baboota, S., Kohli, K., Ali, J., et al. (2010). Strategy for effective brain drug delivery. *Eur. J. Pharm. Sci.* 40, 385–403. doi:10.1016/j.ejps.2010.05.003

Alexander, G. E. (2004). Biology of Parkinson's disease: Pathogenesis and pathophysiology of a multisystem neurodegenerative disorder. *Dialogues Clin. Neurosci.* 6, 259–280. doi:10.31887/DCNS.2004.6.3/GALEXANDER

Aliper, A., Plis, S., Artemov, A., Ulloa, A., Mamoshina, P., Zhavoronkov, A., et al. (2016). Deep learning applications for predicting pharmacological properties of drugs and drug repurposing using transcriptomic data. *Mol. Pharm.* 13, 2524–2530. doi:10.1021/acs.molpharmaceut.6b00248

Alpay, M., and Koroshetz, W. J. (2006). Quetiapine in the treatment of behavioral disturbances in patients with Huntington's disease. *Psychosomatics* 47, 70–72. doi:10.1176/APPI.PSY.47.1.70

Appleby, B. S., Nacopoulos, D., Milano, N., Zhong, K., and Cummings, J. L. (2013). A review: Treatment of alzheimer's disease discovered in repurposed agents. *Dement. Geriatr. Cogn. Disord.* 35, 1–22. doi:10.1159/000345791

Approaches, N. D. R., Alaimo, S., and Pulvirenti, A. (2019). Approaches. *Netw. based drug Repositioning* 1903, 97–113. doi:10.1007/978-1-4939-8955-3_6

Arun, T., Tomassini, V., Sbardella, E., De Ruiter, M. B., Matthews, L., Leite, M. I., et al. (2013). Targeting ASIC1 in primary progressive multiple sclerosis: Evidence of neuroprotection with amiloride. *Brain* 136, 106–115. doi:10.1093/BRAIN/AWS325

Ashburn, T. T., and Thor, K. B. (2004). Drug repositioning: Identifying and developing new uses for existing drugs. *Nat. Rev. Drug Discov.* 3, 673–683. doi:10.1038/nrd1468

Awad, A., and Stue, O. (2009). Cyclophosphamide in multiple sclerosis: Scientific rationale, history and novel treatment paradigms. *Ther. Adv. Neurol. Disord.* 2, 50–61. doi:10.1177/1756285609344375

Bakkar, N., Kovalik, T., Lorenzini, I., Spangler, S., Lacoste, A., Sponaugle, K., et al. (2018). Artificial intelligence in neurodegenerative disease research: Use of IBM watson to identify additional RNA-binding proteins altered in amyotrophic lateral sclerosis. *Acta Neuropathol.* 135, 227–247. doi:10.1007/s00401-017-1785-8

Barkhof, F., Hulst, H. E., Drulović, J., Uitdehaag, B. M. J., Matsuda, K., Landin, R., et al. (2010). Ibudilast in relapsing-remitting multiple sclerosis. *Neurology* 74, 1033–1040. doi:10.1212/WNL.0B013E3181D7D651

Barnes, L. L. (2021). Alzheimer disease in african American individuals: Increased incidence or not enough data? *Nat. Rev. Neurol.* 181 (18), 56–62. doi:10.1038/s41582-021-00589-3

Beister, A., Kraus, P., Kuhn, W., Dose, M., Weindl, A., and Gerlach, M. (2004). The N-methyl-D-aspartate antagonist memantine retards progression of Huntington's disease. *J. Neural Transm. Suppl.*, 117–122. doi:10.1007/978-3-7091-0579-5_14

Bezprozvanny, I. (2010). The rise and fall of Dimebon. *Drug News Perspect.* 23, 518–523. doi:10.1358/DNP.2010.23.8.1500435

Bharadwaj Prakash, K. B., and Kanagachidambaresan, G. R. (2021). "Pattern recognition and machine learning," in *Programming with TensorFlow*. Cham: Springer. doi:10.1007/978-3-030-57077-4_11

Birks, J. S., and Evans, J. G. (2015). Rivastigmine for alzheimer's disease (review) rivastigmine for alzheimer's disease. *Cochrane Database Syst. Rev.*, 2015, 4–7. doi:10.1002/14651858.CD001191

Birks, J. S., and Harvey, R. J. (2018). Donepezil for dementia due to Alzheimer's disease. *Cochrane Database Syst. Rev.* 6, CD001190. doi:10.1002/14651858.CD001190.pub3

Bonuccelli, U., Ceravolo, R., Maremmani, C., Nuti, A., Rossi, G., and Muratorio, A. (1994). Clozapine in Huntington's chorea. *Neurology* 44, 821–823. doi:10.1212/WNL.44.5.821

Brehmer, D., Greff, Z., Godl, K., Blencke, S., Kurtenbach, A., Weber, M., et al. (2005). Cellular targets of gefitinib. *Cancer Res.* 65, 379–382. doi:10.1158/0008-5472.379.65.2

Brunden, K. R., Yao, Y., Potuzak, J. S., Ferrer, N. I., Ballatore, C., James, M. J., et al. (2011). The characterization of microtubule-stabilizing drugs as possible therapeutic agents for Alzheimer's disease and related tauopathies. *Pharmacol. Res.* 63, 341–351. doi:10.1016/J.PHRS.2010.12.002

Buniello, A., Macarthur, J. A. L., Cerezo, M., Harris, L. W., Hayhurst, J., Malangone, C., et al. (2019). The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res.* 47, D1005–D1012. doi:10.1093/nar/gky1120

Butler, K. T., Davies, D. W., Cartwright, H., Isayev, O., and Walsh, A. (2018). Machine learning for molecular and materials science. *Nature* 559, 547–555. doi:10.1038/s41586-018-0337-2

Chen, B., Ma, L., Paik, H., Sirota, M., Wei, W., Chua, M. S., et al. (2017). Reversal of cancer gene expression correlates with drug efficacy and reveals therapeutic targets. *Nat. Commun.* 8, 16022. doi:10.1038/ncomms16022

Chen, P. C., Hsieh, Y. C., Huang, C. C., and Hu, C. J. (2020). Tamoxifen for amyotrophic lateral sclerosis: A randomized double-blind clinical trial. *Med. Baltim.* 99, e20423. doi:10.1097/MD.0000000000020423

Clark, L. N., Park, N., Kisselev, S., Rios, E., Lee, J. H., and Louis, E. D. (2010). Replication of the LINGO1 gene association with essential tremor in a North American population. *Eur. J. Hum. Genet.* 18, 838–843. doi:10.1038/ejhg.2010.27

Clinicaltrials (2022). NCT02868580 safety and tolerability of antiretroviral (triumeq) in patients with amyotrophic lateral sclerosis (ALS). - full text view - ClinicalTrials.gov. Available at: https://clinicaltrials.gov/ct2/show/NCT02868580 (Accessed July 7, 2022).

Correale, J., Gaitán, M. I., Ysrraelit, M. C., and Fiol, M. P. (2017). Progressive multiple sclerosis: From pathogenic mechanisms to treatment. *Brain* 140, 527–546. doi:10.1093/brain/aww258

Crismon, M. L. (1994). Tacrine: First drug approved for alzheimer's disease. *Ann. Pharmacother.* 28, 744–751. doi:10.1177/106002809402800612

Cudkowicz, M. E., Titus, S., Kearney, M., Yu, H., Sherman, A., Schoenfeld, D., et al. (2014). Safety and efficacy of ceftriaxone for amyotrophic lateral sclerosis: A multi-stage, randomised, double-blind, placebo-controlled trial. *Lancet. Neurol.* 13, 1083–1091. doi:10.1016/S1474-4422(14)70222-4

Culman, J., Blume, A., Gohlke, P., and Unger, T. (2002). The renin-angiotensin system in the brain: Possible therapeutic implications for AT1-receptor blockers. *J. Hum. Hypertens.* 16, S64–S70. doi:10.1038/sj.jhh.1001442

Cusi, C., Cantisani, T. A., Celani, M. G., Incorvaia, B., Righetti, E., Candelise, L., et al. (2007). Galantamine for Alzheimer's disease and mild cognitive impairment. *Neuroepidemiology* 28, 116–117. doi:10.1159/000101510

De Benedetti, P. G., and Fanelli, F. (2018). Computational modeling approaches to quantitative structure–binding kinetics relationships in drug discovery. *Drug Discov. Today* 23, 1396–1406. doi:10.1016/j.drudis.2018.03.010

de Franchi, E., Schalon, C., Messa, M., Onofri, F., Benfenati, F., Rognan, D., et al. (2010). Binding of protein kinase inhibitors to synapsin I inferred from pair-wise binding site similarity Measurements. *PLoS One* 5, e12214. doi:10.1371/journal.pone.0012214

Dean, J. M., Vernon, D. D., Cook, L., Nechodom, P., Reading, J., and Suruda, A. (2001). Probabilistic linkage of computerized ambulance and inpatient hospital discharge records: A potential tool for evaluation of emergency medical services. *Ann. Emerg. Med.* 37, 616–626. doi:10.1067/mem.2001.115214

Delespierre, T., Denormandie, P., Bar-Hen, A., and Josseran, L. (2017). Empirical advances with text mining of electronic health records. *BMC Med. Inf. Decis. Mak.* 17, 1–15. doi:10.1186/s12911-017-0519-0

Devos, D., Moreau, C., Delval, A., Dujardin, K., Defebvre, L., and Bordet, R. (2013). Methylphenidate : A treatment for Parkinson's disease? *CNS Drugs* 27, 1–14. doi:10.1007/S40263-012-0017-Y

Di Meco, A., Lauretti, E., Vagnozzi, A. N., and Praticò, D. (2014). Zileuton restores memory impairments and reverses amyloid and tau pathology in aged Alzheimer's disease mice. *Neurobiol. Aging* 35, 2458–2464. doi:10.1016/J.NEUROBIOLAGING.2014.05.016

Diomede, L., Cassata, G., Fiordaliso, F., Salio, M., Ami, D., Natalello, A., et al. (2010). Tetracycline and its analogues protect *Caenorhabditis elegans* from β amyloid-induced toxicity by targeting oligomers. *Neurobiol. Dis.* 40, 424–431. doi:10.1016/J.NBD.2010.07.002

Dl, S., Mn, S., Dj, C., J, L., Lj, L., S, P., et al. (2005). Atorvastatin therapy lowers circulating cholesterol but not free radical activity in advance of identifiable clinical benefit in the treatment of mild-to-moderate AD. *Curr. Alzheimer Res.* 2, 343–353. doi:10.2174/1567205054367900

Dominguez-Meijide, A., Parrales, V., Vasili, E., González-Lizárraga, F., König, A., Lázaro, D. F., et al. (2021). Doxycycline inhibits α-synuclein-associated pathologies *in vitro* and *in vivo*. *Neurobiol. Dis.* 151, 105256. doi:10.1016/J.NBD.2021.105256

Doody, R. S., Gavrilova, S. I., Sano, M., Thomas, R. G., Aisen, P. S., Bachurin, S. O., et al. (2008). Effect of dimebon on cognition, activities of daily living, behaviour, and global function in patients with mild-to-moderate alzheimer's disease: A randomised, double-blind, placebo-controlled study. *Lancet (London, Engl.* 372, 207–215. doi:10.1016/S0140-6736(08)61074-0

Duff, K., Beglinger, L. J., O'Rourke, M. E., Nopoulos, P., Paulson, H. L., and Paulsen, J. S. (2008). Risperidone and the treatment of psychiatric, motor, and cognitive symptoms in Huntington's disease. *Ann. Clin. Psychiatry* 20, 1–3. doi:10.1080/10401230701844802

Durães, F., Pinto, M., and Sousa, E. (2018). Old drugs as new treatments for neurodegenerative diseases. *Pharmaceuticals* 11, E44. doi:10.3390/PH11020044

Ehrnhoefer, D. E., Wong, B. K. Y., and Hayden, M. R. (2012). Europe PMC funders group convergent pathogenic pathways in alzheimer ' s and Huntington disease. *Shar. targets drug Dev.* 10, 853–867. doi:10.1038/nrd3556.Convergent

Ferreira, L. G., and Andricopulo, A. D. (2016). *Drug repositioning approaches to parasitic diseases: A medicinal chemistry perspective*. Amsterdam, Netherlands: Elsevier. doi:10.1016/j.drudis.2016.06.021

Ford, E., Carroll, J. A., Smith, H. E., Scott, D., and Cassell, J. A. (2016). Extracting information from the text of electronic medical records to improve case detection: A systematic review. *J. Am. Med. Inf. Assoc.* 23, 1007–1015. doi:10.1093/jamia/ocv180

Fox, E. J. (2004). Mechanism of action of mitoxantrone. *Neurology* 63, S15–S18. doi:10.1212/WNL.63.12_SUPPL_6.S15

Fox, S. H., Katzenschlager, R., Lim, S. Y., Barton, B., de Bie, R. M. A., Seppi, K., et al. (2018). International Parkinson and movement disorder society evidence-based medicine review: Update on treatments for the motor symptoms of Parkinson's disease. *Mov. Disord.* 33, 1248–1266. doi:10.1002/MDS.27372

Fukasawa, H., Nakagomi, M., Yamagata, N., Katsuki, H., Kawahara, K., Kitaoka, K., et al. (2012). Tamibarotene: A candidate retinoid drug for alzheimer's disease. *Biol. Pharm. Bull.* 35, 1206–1212. doi:10.1248/bpb.b12-00314Available at: https://www.jstage.jst.go.jp/article/bpb/35/8/35_b12-00314/_article/-char/ja/(Accessed June 27, 2022)

Gabathuler, R. (2010). Approaches to transport therapeutic drugs across the blood-brain barrier to treat brain diseases. *Neurobiol. Dis.* 37, 48–57. doi:10.1016/j.nbd.2009.07.028

García-Barroso, C., Ricobaraza, A., Pascual-Lucas, M., Unceta, N., Rico, A. J., Goicolea, M. A., et al. (2013). Tadalafil crosses the blood–brain barrier and reverses cognitive dysfunction in a mouse model of AD. *Neuropharmacology* 64, 114–123. doi:10.1016/J.NEUROPHARM.2012.06.052

Geifman, N., Brinton, R. D., Kennedy, R. E., Schneider, L. S., and Butte, A. J. (2017). Evidence for benefit of statins to modify cognitive decline and risk in Alzheimer's disease. *Alzheimer's Res. Ther.* 9, 10. doi:10.1186/s13195-017-0237-y

Gottlieb, A., Stein, G. Y., Ruppin, E., and Sharan, R. (2011). Predict: A method for inferring novel drug indications with application to personalized medicine. *Mol. Syst. Biol.* 7, 496–499. doi:10.1038/msb.2011.26

Gregori-Puigjane, E., and Mestres, J. (2008). A ligand-based approach to mining the chemogenomic space of drugs. *Comb. Chem. High. Throughput Screen.* 11, 669–676. doi:10.2174/138620708785739952

Grossi, C., Francese, S., Casini, A., Rosi, M. C., Luccarini, I., Fiorentini, A., et al. (2009). Clioquinol decreases amyloid-β burden and reduces working memory impairment in a transgenic mouse model of alzheimer's disease. *J. Alzheimer's Dis.* 17, 423–440. doi:10.3233/JAD-2009-1063

Gudjonsdottir, H., Jonsdottir, G. A., Palsson, S. T., Jonsson, T., Bjornsdottir, G., Böttcher, Y., et al. (2009). Variant in the sequence of the LINGO1 gene confers risk of essential tremor. *Nat. Genet.* 41, 277–279. doi:10.1038/ng.299

Gupta, S., and Shukla, S. (2021). Non-motor symptoms in Parkinson's disease: Opening new avenues in treatment. *Curr. Res. Behav. Sci.* 2, 100049. doi:10.1016/j.crbeha.2021.100049

Halliday, M., Radford, H., Zents, K. A. M., Molloy, C., Moreno, J. A., Verity, N. C., et al. (2017). Repurposed drugs targeting eIF2α-P-mediated translational repression prevent neurodegeneration in mice. *Brain* 140, 1768–1783. doi:10.1093/BRAIN/AWX074

Hartsel, S., and Biochemistry, T. W. (2003). Amphotericin B binds to amyloid fibrils and delays their formation: A therapeutic mechanism? *Biochemistry* 42, 6228–6233. doi:10.1021/bi0270384

Hassanzadeh, G., Hosseini, A., Pasbakhsh, P., Akbari, M., Ghaffarpour, M., Takzare, N., et al. (2015). Trimetazidine prevents oxidative changes induced in a rat model of sporadic type of Alzheimer's disease. *Acta Med. Iran.* 53, 17–24. Available at: http://acta.tums.ac.ir/index.php/acta/article/view/4942 (Accessed June 27, 2022).

Hayes, C. D., Dey, D., Palavicini, J. P., Wang, H., Patkar, K. A., Minond, D., et al. (2013). Striking reduction of amyloid plaque burden in an Alzheimer's mouse model after chronic administration of carmustine. *BMC Med.* 11, 81. doi:10.1186/1741-7015-11-81

Hoe, H. S., Fu, Z., Makarova, A., Lee, J. Y., Lu, C., Feng, L., et al. (2009). The effects of amyloid precursor protein on postsynaptic composition and activity. *J. Biol. Chem.* 284, 8495–8506. doi:10.1074/jbc.M900141200

Holdgate, G., Geschwindner, S., Breeze, A., Davies, G., Colclough, N., Temesi, D., et al. (2013). Biophysical methods in drug discovery from small molecule to pharmaceutical. *Methods Mol. Biol.* 1008, 327–355. doi:10.1007/978-1-62703-398-5_12

Hou, Y., Dan, X., Babbar, M., Wei, Y., Hasselbalch, S. G., Croteau, D. L., et al. (2019). Ageing as a risk factor for neurodegenerative disease. *Nat. Rev. Neurol.* 15, 565–581. doi:10.1038/s41582-019-0244-7

Hripcsak, G., Duke, J. D., Shah, N. H., Reich, C. G., Huser, V., Schuemie, M. J., et al. (2015). Observational health data sciences and Informatics (OHDSI): Opportunities for observational researchers. *Stud. Health Technol. Inf.* 216, 574–578. doi:10.3233/978-1-61499-564-7-574

Hurle, M. R., Yang, L., Xie, Q., Rajpal, D. K., Sanseau, P., and Agarwal, P. (2013). Computational drug repositioning: From data to therapeutics. *Clin. Pharmacol. Ther.* 93, 335–341. doi:10.1038/clpt.2013.1

Iqbal, U. H., Zeng, E., and Pasinetti, G. M. (2020). The use of antimicrobial and antiviral drugs in alzheimer's disease. *Int. J. Mol. Sci.* 21, 4920. doi:10.3390/IJMS21144920

Jalencas, X., and Mestres, J. (2013). Chemoisosterism in the proteome. *J. Chem. Inf. Model.* 53, 279–292. doi:10.1021/ci3002974

Jankovic, J. (2017). Exenatide – A drug for diabetes and Parkinson disease? *Nat. Rev. Neurol.* 1311 (13), 643–644. doi:10.1038/nrneurol.2017.140

Kadurin, A., Nikolenko, S., Khrabrov, K., Aliper, A., and Zhavoronkov, A. (2017). DruGAN: An advanced generative adversarial autoencoder model for de Novo generation of new molecules with desired molecular properties *in silico*. *Mol. Pharm.* 14, 3098–3104. doi:10.1021/acs.molpharmaceut.7b00346

Karlawish, J., Jack, C. R., Rocca, W. A., Snyder, H. M., and Carrillo, M. C. (2017). Alzheimer's disease: The next frontier—special report 2017. *Alzheimer's Dement.* 13, 374–380. doi:10.1016/j.jalz.2017.02.006

Katsnelson, A., De Strooper, B., and Zoghbi, H. Y. (2016). Neurodegeneration: From cellular concepts to clinical applications. *Sci. Transl. Med.* 8, 364ps18–6. doi:10.1126/scitranslmed.aal2074

Kiernan, M. C., Vucic, S., Cheah, B. C., Turner, M. R., Eisen, A., Hardiman, O., et al. (2011). Amyotrophic lateral sclerosis. *Lancet* 377, 942–955. doi:10.1016/S0140-6736(10)61156-7

Kitchen, D. B., Decornez, H., Furr, J. R., and Bajorath, J. (2004). Docking and scoring in virtual screening for drug discovery: Methods and applications. *Nat. Rev. Drug Discov.* 3, 935–949. doi:10.1038/nrd1549

Klaeger, S., Gohlke, B., Perrin, J., Gupta, V., Heinzlmeir, S., Helm, D., et al. (2016). Chemical proteomics reveals ferrochelatase as a common off-target of kinase inhibitors. *ACS Chem. Biol.* 11, 1245–1254. doi:10.1021/acschembio.5b01063

Kshirsagar, P. B., Kanhere, H. S., Bansinge, P. C., Rathod, S. K., Khandare, V. S., and Das, R. K. (2021). Huntington's disease: Pathophysiology and therapeutic intervention. *GSC Biol. Pharm. Sci.* 15, 171–184. doi:10.30574/gscbps.2021.15.2.0140

Kumar, R., Aadil, K. R., Mondal, K., Mishra, Y. K., Oupicky, D., Ramakrishna, S., et al. (2021). Neurodegenerative disorders management: State-of-art and prospects of nano-biotechnology. doi:10.1080/07388551.2021.1993126

Lee, H. M., and Kim, Y. (2016). Drug repurposing is a new opportunity for developing drugs against neuropsychiatric disorders. *Schizophr. Res. Treat.* 2016, 6378137. doi:10.1155/2016/6378137

Lee, S. Y., Song, M. Y., Kim, D., Park, C., Park, D. K., Kim, D. G., et al. (2020). A proteotranscriptomic-based computational drug-repositioning method for Alzheimer's disease. *Front. Pharmacol.* 10, 1–11. doi:10.3389/fphar.2019.01653

Leist, T. P., and Weissert, R. (2011). Cladribine: Mode of action and implications for treatment of multiple sclerosis. *Clin. Neuropharmacol.* 34, 28–35. doi:10.1097/WNF.0B013E318204CD90

Li, Y. H., Yu, C. Y., Li, X. X., Zhang, P., Tang, J., Yang, Q., et al. (2018). Therapeutic target database update 2018: Enriched resource for facilitating bench-to-clinic research of targeted therapeutics. *Nucleic Acids Res.* 46, D1121–D1127. doi:10.1093/nar/gkx1076

Lionta, E., Spyrou, G., Vassilatis, D., and Cournia, Z. (2014). Structure-based virtual screening for drug discovery: Principles, applications and recent advances. *Curr. Top. Med. Chem.* 14, 1923–1938. doi:10.2174/1568026614666140929124445

Liu, H., Song, Y., Guan, J., Luo, L., and Zhuang, Z. (2016). Inferring new indications for approved drugs via random walk on drug-disease heterogenous networks. *BMC Bioinforma.* 17, 539. doi:10.1186/s12859-016-1336-7

Luo, Y., Zhao, X., Zhou, J., Yang, J., Zhang, Y., Kuang, W., et al. (2017). A network integration approach for drug-target interaction prediction and computational drug repositioning from heterogeneous information. *Nat. Commun.* 8, 573. doi:10.1038/s41467-017-00680-8

Ma, J., Sheridan, R. P., Liaw, A., Dahl, G. E., and Svetnik, V. (2015). Deep neural nets as a method for quantitative structure-activity relationships. *J. Chem. Inf. Model.* 55, 263–274. doi:10.1021/ci500747n

Malin, B. A., and Sweeney, L. (2005). A secure protocol to distribute unlinkable health data. *AMIA Annu. Symp. Proc.*, 2005, 485–489.

Mani, S., Chen, Y., Elasy, T., Clayton, W., and Denny, J. (2012). Type 2 diabetes risk forecasting from EMR data using machine learning. *AMIA Annu. Symp. Proc.* 2012, 606–615.

March-Vila, E., Pinzi, L., Sturm, N., Tinivella, A., Engkvist, O., Chen, H., et al. (2017). On the integration of *in silico* drug design methods for drug repurposing. *Front. Pharmacol.* 8, 298. doi:10.3389/fphar.2017.00298

Matilla-Dueñas, A., Corral-Juan, M., Rodríguez-Palmero Seuma, A., Vilas, D., Ispierto, L., Morais, S., et al. (2017). Rare neurodegenerative diseases: Clinical and genetic update. *Adv. Exp. Med. Biol.* 1031, 443–496. doi:10.1007/978-3-319-67144-4_25

Mcclean, P. L., Parthsarathy, V., Faivre, E., and Holscher, C. (2011). The diabetes drug liraglutide prevents degenerative processes in a mouse model of alzheimer's disease. *J. Neurosci.* 31, 6587–6594. doi:10.1523/JNEUROSCI.0529-11.2011

McColgan, P., and Tabrizi, S. J. (2018). Huntington's disease: A clinical review. *Eur. J. Neurol.* 25, 24–34. doi:10.1111/ENE.13413

Mittal, S., Bjørnevik, K., Im, D. S., Flierl, A., Dong, X., Locascio, J. J., et al. (2017). β2-Adrenoreceptor is a regulator of the α-synuclein gene driving risk of Parkinson's disease. *Sci. (80-.)* 357, 891–898. doi:10.1126/science.aaf3934

Molina, D. M., Jafari, R., Ignatushchenko, M., Seki, T., Larsson, E. A., Dan, C., et al. (2013). Monitoring drug target engagement in cells and tissues using the cellular thermal shift assay. *Science* 341, 84–87. doi:10.1126/science.1233606

Monacelli, F., Cea, M., Borghi, R., Odetti, P., and Nencioni, A. (2017). Do cancer drugs counteract neurodegeneration? Repurposing for alzheimer's disease. *J. Alzheimer's Dis.* 55, 1295–1306. doi:10.3233/JAD-160840

Mucke, H. A. (2015). The case of galantamine: Repurposing and late blooming of a cholinergic drug. *Future Sci. OA* 1, FSO73. doi:10.4155/fso.15.73

Myszczynska, M. A., Ojamies, P. N., Lacoste, A. M. B., Neil, D., Saffari, A., Mead, R., et al. (2020). Applications of machine learning to diagnosis and treatment of neurodegenerative diseases. *Nat. Rev. Neurol.* 16, 440–456. doi:10.1038/s41582-020-0377-8

Nagarajan, D. V., and Wilde, M. (2005). The reawakening of a sleeping little giant. *Emerg. Med. J.* 22, 384–386. doi:10.1136/EMJ.2003.008656

Nance, M. A. (2017). *Genetics of Huntington disease*. 1st ed. Amsterdam, Netherlands: Elsevier B.V. doi:10.1016/B978-0-12-801893-4.00001-8

Napolitano, F., Zhao, Y., Moreira, V. M., Tagliaferri, R., Kere, J., D'Amato, M., et al. (2013). Drug repositioning: A machine-learning approach through data integration. *J. Cheminform.* 5, 30–39. doi:10.1186/1758-2946-5-30

Nebes, R. D., Pollock, B. G., Houck, P. R., Butters, M. A., Mulsant, B. H., Zmuda, M. D., et al. (2003). Persistence of cognitive impairment in geriatric patients following antidepressant treatment: A randomized, double-blind clinical trial with nortriptyline and paroxetine. *J. Psychiatr. Res.* 37, 99–108. doi:10.1016/S0022-3956(02)00085-7

Nemati, S., Holder, A., Razmi, F., Stanley, M. D., Clifford, G. D., and Buchman, T. G. (2018). An interpretable machine learning model for accurate prediction of sepsis in the ICU. *Crit. Care Med.* 46, 547–553. doi:10.1097/CCM.0000000000002936

Onyango, I. G., Bennett, J. P., and Stokin, G. B. (2021). Regulation of neuronal bioenergetics as a therapeutic strategy in neurodegenerative diseases. *Neural Regen. Res.* 16, 1467–1482. doi:10.4103/1673-5374.303007

Oprea, T. I., and Overington, J. P. (2015). Computational and practical aspects of drug repositioning. *Assay. Drug Dev. Technol.* 13, 299–306. doi:10.1089/adt.2015.29011.tiodrrr

Orgogozo, J. M., Gilman, S., Dartigues, J. F., Laurent, B., Puel, M., Kirby, L. C., et al. (2003). Subacute meningoencephalitis in a subset of patients with AD after Abeta42 immunization. *Neurology* 61, 46–54. doi:10.1212/01.WNL.0000073623.84147.A8

Pagan, F., Hebron, M., Valadez, E. H., Torres-Yaghi, Y., Huang, X., Mills, R. R., et al. (2016). Nilotinib effects in Parkinson's disease and dementia with lewy bodies. *J. Park. Dis.* 6, 503–517. doi:10.3233/JPD-160867

Paleacu, D., Anca, M., and Giladi, N. (2002). Olanzapine in Huntington's disease. *Acta Neurol. Scand.* 105, 441–444. doi:10.1034/J.1600-0404.2002.01197.X

Paleacu, D. (2007). Tetrabenazine in the treatment of Huntington's disease. *Neuropsychiatr. Dis. Treat.* 3, 545–551. (Accessed July 6, 2022).

Perry, T. A., Lahiri, D. K., Sambamurti, K., Chen, D., Mattson, M. P., Egan, J. M., et al. (2003). Glucagon-like peptide-1 decreases endogenous amyloid-beta peptide (Abeta) levels and protects hippocampal neurons from death induced by Abeta and iron. *J. Neurosci. Res.* 72, 603–612. doi:10.1002/JNR.10611

Popova, M., Isayev, O., and Tropsha, A. (2018). Deep reinforcement learning for de novo drug design. *Sci. Adv.* 4, eaap7885–14. doi:10.1126/sciadv.aap7885

Rakshit, H., Chatterjee, P., and Roy, D. (2015). A bidirectional drug repositioning approach for Parkinson's disease through network-based inference. *Biochem. Biophys. Res. Commun.* 457, 280–287. doi:10.1016/j.bbrc.2014.12.101

Reddy, A. S., and Zhang, S. (2013). Polypharmacology: Drug discovery for the future. *Expert Rev. Clin. Pharmacol.* 6, 41–47. doi:10.1586/ecp.12.74

Reeve, A. K., Grady, J. P., Cosgrave, E. M., Bennison, E., Chen, C., Hepplewhite, P. D., et al. (2018). Mitochondrial dysfunction within the synapses of substantia nigra neurons in Parkinson's disease. *NPJ Park. Dis.* 4, 9–10. doi:10.1038/s41531-018-0044-6

Roos, R. A. C., Buruma, O. J. S., Bruyn, G. W., Kemp, B., and van der Velde, E. A. (1982). Tiapride in the treatment of Huntington's chorea. *Acta Neurol. Scand.* 65, 45–50. doi:10.1111/J.1600-0404.1982.TB03060.X

Roos, R. A. C. (2010). Huntington's disease: A clinical review. *Orphanet J. Rare Dis.* 5, 40–48. doi:10.1186/1750-1172-5-40

Rudrapal, M., Khairnar, S. J., and Jadhav, A. G. (2020). Drug repurposing (DR): An emerging approach in drug discovery. in *Drug repurposing*, London UK: IntechOpen, 1–20. doi:10.5772/intechopen.93193

Ryan, B. J., Hoek, S., Fon, E. A., and Wade-Martins, R. (2015). Mitochondrial dysfunction and mitophagy in Parkinson's: From familial to sporadic disease. *Trends biochem. Sci.* 40, 200–210. doi:10.1016/J.TIBS.2015.02.003

Ryu, J. K., and McLarnon, J. G. (2008). Thalidomide inhibition of perturbed vasculature and glial-derived tumor necrosis factor-α in an animal model of inflamed Alzheimer's disease brain. *Neurobiol. Dis.* 29, 254–266. doi:10.1016/J.NBD.2007.08.019

Sano, M., Bell, K. L., Galasko, D., Galvin, J. E., Thomas, R. G., Van Dyck, C. H., et al. (2011). A randomized, double-blind, placebo-controlled trial of simvastatin to treat Alzheimer disease. *Neurology* 77, 556–563. doi:10.1212/WNL.0B013E318228BF11

Savva, K., Zachariou, M., Oulas, A., Minadakis, G., Sokratous, K., Dietis, N., et al. (2019). *Computational drug repurposing for neurodegenerative diseases*. Amsterdam, Netherlands: Elsevier. doi:10.1016/b978-0-12-816125-8.00004-3

Sawada, H. (2017). Clinical efficacy of edaravone for the treatment of amyotrophic lateral sclerosis. *Expert Opin. Pharmacother.* 18, 735–738. doi:10.1080/14656566.2017.1319937

Scheltens, P., Blennow, K., Breteler, M. M. B., de Strooper, B., Frisoni, G. B., Salloway, S., et al. (2016). Alzheimer's disease. *Lancet* 388, 505–517. doi:10.1016/S0140-6736(15)01124-1

Scott, T., Susana, M., Smitha, K., Nettles, K., Mohammad, F., Gary, L. J., et al. (2016). Crizotinib inhibits NF2-associated schwannoma through inhibition of focal adhesion kinase 1. *Oncotarget* 7 (N), 54515–54525. doi:10.18632/oncotarget.10248

Shameer, K., Johnson, K. W., Yahi, A., Miotto, R., Li, L. I., Ricks, D., et al. (2017). Predictive modeling of hospital readmission rates using electronic medical record-wide machine learning: A case-study using mount sinai heart failure cohort. *Pac. Symp. Biocomput.* 0, 276–287. doi:10.1142/9789813207813_0027

Shineman, D. W., Alam, J., Anderson, M., Black, S. E., Carman, A. J., Cummings, J. L., et al. (2014). Overcoming obstacles to repurposing for neurodegenerative disease. *Ann. Clin. Transl. Neurol.* 1, 512–518. doi:10.1002/acn3.76

Siuly, S., and Zhang, Y. (2016). Medical big data: Neurological diseases diagnosis through medical data analysis. *Data Sci. Eng.* 1, 54–64. doi:10.1007/s41019-016-0011-3

Smith, A. M., Gibbons, H. M., and Dragunow, M. (2010). Valproic acid enhances microglial phagocytosis of amyloid-beta(1-42). *Neuroscience* 169, 505–515. doi:10.1016/J.NEUROSCIENCE.2010.04.041

Smith, A. M., Walsh, J. R., Long, J., Davis, C. B., Henstock, P., Hodge, M. R., et al. (2019). Deep learning of representations for transcriptomics-based phenotype prediction. Priprint bioRxiv, 1–38. doi:10.1101/574723

Song, C., Zhang, Y., Parsons, C. G., and Liu, Y. F. (2003). Expression of polyglutamine-expanded huntingtin induces tyrosine phosphorylation of N-methyl-D-aspartate receptors. *J. Biol. Chem.* 278, 33364–33369. doi:10.1074/jbc.M304240200

Sundar, S., Battistoni, C., McNulty, R., Morales, F., Gorky, J., Foley, H., et al. (2020). An agent-based model to investigate microbial initiation of Alzheimer's via the olfactory system. *Theor. Biol. Med. Model.* 17, 5–15. doi:10.1186/s12976-020-00123-w

Tabrizi, S. J., Flower, M. D., Ross, C. A., and Wild, E. J. (2020). Huntington disease: New insights into molecular pathogenesis and therapeutic opportunities. *Nat. Rev. Neurol.* 1610 (16), 529–546. doi:10.1038/s41582-020-0389-4

Talevi, A. (2018). Drug repositioning: Current approaches and their implications in the precision medicine era. *Expert Rev. Precis. Med. Drug Dev.* 3, 49–61. doi:10.1080/23808993.2018.1424535

Tanoli, Z., Vähä-Koskela, M., and Aittokallio, T. (2021). Artificial intelligence, machine learning, and drug repurposing in cancer. *Expert Opin. Drug Discov.* 16, 977–989. doi:10.1080/17460441.2021.1883585

The Economic Times (2021). World population growth: World population grew by 74 million over past year: US census Bureau - the economic times. Available at: https://economictimes.indiatimes.com/news/international/world-news/world-population-grew-by-74-million-over-past-year-us-census-bureau/articleshow/88611764.cms (Accessed September 2, 2022).

Tippmann, F., Hundt, J., Schneider, A., Endres, K., and Fahrenholz, F. (2009). Up-regulation of the α-secretase ADAM10 by retinoic acid receptors and acitretin. *FASEB J.* 23, 1643–1654. doi:10.1096/FJ.08-121392

Tomiyama, T., Shoji, A., Kataoka, K. I., Suwa, Y., Asano, S., Kaneko, H., et al. (1996). Inhibition of amyloid β protein aggregation and neurotoxicity by Rifampicin: Its possible function as a hydroxyl radical scavenger (*). *J. Biol. Chem.* 271, 6839–6844. doi:10.1074/JBC.271.12.6839

Tousi, B. (2015). The emerging role of bexarotene in the treatment of alzheimer's disease: Current evidence. *Neuropsychiatr. Dis. Treat.* 11, 311–315. doi:10.2147/NDT.S61309

Trapp, B. D., and Nave, K. A. (2008). Multiple sclerosis: An immune or neurodegenerative disorder?, *Annu Rev Neurosci.*, 31, 247–269. doi: doi:10.1146/annurev.neuro.30.051606.09431331

Trias, E., Ibarburu, S., Barreto-Núñez, R., Babdor, J., Maciel, T. T., Guillo, M., et al. (2016). Post-paralysis tyrosine kinase inhibition with masitinib abrogates neuroinflammation and slows disease progression in inherited amyotrophic lateral sclerosis. *J. Neuroinflammation* 13, 177. doi:10.1186/S12974-016-0620-9

Vogt, I., and Mestres, J. (2010). Drug-target networks. *Mol. Inf.* 29, 10–14. doi:10.1002/minf.200900069

Wagner, J., Vulinović, F., Grünewald, A., Unger, M. M., Möller, J. C., Klein, C., et al. (2017). Acylated and unacylated ghrelin confer neuroprotection to mesencephalic neurons. *Neuroscience* 365, 137–145. doi:10.1016/J.NEUROSCIENCE.2017.09.045

Wainger, B. J., Macklin, E. A., Vucic, S., McIlduff, C. E., Paganoni, S., Maragakis, N. J., et al. (2021). Effect of ezogabine on cortical and spinal motor neuron excitability in amyotrophic lateral sclerosis: A randomized clinical trial. *JAMA Neurol.* 78, 186–196. doi:10.1001/JAMANEUROL.2020.4300

Wang, L., Xi, Y., Sung, S., and Qiao, H. (2018). RNA-seq assistant: Machine learning based methods to identify more transcriptional regulated genes. *BMC Genomics* 19, 1–13. doi:10.1186/s12864-018-4932-2

Wen, M., Zhang, Z., Niu, S., Sha, H., Yang, R., Yun, Y., et al. (2017). Deep-learning-based drug-target interaction prediction. *J. Proteome Res.* 16, 1401–1409. doi:10.1021/acs.jproteome.6b00618

Wishart, D. S., Feunang, Y. D., Marcu, A., Guo, A. C., Liang, K., Vázquez-Fresno, R., et al. (2018). Hmdb 4.0: The human metabolome database for 2018. *Nucleic Acids Res.* 46, D608–D617. doi:10.1093/nar/gkx1089

Wishart, D. S., Knox, C., Guo, A. C., Shrivastava, S., Hassanali, M., Stothard, P., et al. (2006). DrugBank: A comprehensive resource for *in silico* drug discovery and exploration. *Nucleic Acids Res.* 34, 668–672. doi:10.1093/nar/gkj067

Wozniak, M. A., and Itzhaki, R. F. (2010). Antiviral agents in alzheimer's disease: Hope for the future? *Ther. Adv. Neurol. Disord.* 3, 141–152. doi:10.1177/1756285610370069

Xie, H., Wen, H., Qin, M., Xia, J., Zhang, D., Liu, L., et al. (2016). *In silico* drug repositioning for the treatment of Alzheimer's disease using molecular docking and gene expression data. *RSC Adv.* 6, 98080–98090. doi:10.1039/c6ra21941a

Yang, L., and Agarwal, P. (2011). Systematic drug repositioning based on clinical side-effects. *PLoS One* 6, e28025. doi:10.1371/journal.pone.0028025

Zhang, J., Guo, J., Zhao, X., Chen, Z., Wang, G., Liu, A., et al. (2013). Phosphodiesterase-5 inhibitor sildenafil prevents neuroinflammation, lowers beta-amyloid levels and improves cognitive performance in APP/PS1 transgenic mice. *Behav. Brain Res.* 250, 230–237. doi:10.1016/J.BBR.2013.05.017

Zhang, M., Schmitt-Ulms, G., Sato, C., Xi, Z., Zhang, Y., Zhou, Y., et al. (2016). Drug repositioning for Alzheimer's disease based on systematic "omics" data mining. *PLoS One* 11, e0168812–e0168815. doi:10.1371/journal.pone.0168812

Zhao, W. Q., Chen, H., Quon, M. J., and Alkon, D. L. (2004). Insulin and the insulin receptor in experimental models of learning and memory. *Eur. J. Pharmacol.* 490, 71–81. doi:10.1016/J.EJPHAR.2004.02.045

Zoccolella, S., Beghi, E., Palagano, G., Fraddosio, A., Guerra, V., Samarelli, V., et al. (2007). Riluzole and amyotrophic lateral sclerosis survival: A population-based study in southern Italy. *Eur. J. Neurol.* 14, 262–268. doi:10.1111/J.1468-1331.2006.01575.X

frontiers | Frontiers in Pharmacology

TYPE Original Research
PUBLISHED 05 October 2022
DOI 10.3389/fphar.2022.962863

OPEN ACCESS

EDITED BY
Mithun Rudrapal,
Rasiklal M. Dhariwal Institute of Pharmaceutical Education and Research, India

REVIEWED BY
Ismail Celik,
Erciyes University, Turkey
Vishma Pratap Sur,
Institute of Biotechnology (ASCR), Czechia

*CORRESPONDENCE
Xinbo Yang,
ubeihang@126.com
Yuanjie Zheng,
yjzheng@sdnu.edu.cn

SPECIALTY SECTION
This article was submitted to Drugs Outcomes Research and Policies, a section of the journal Frontiers in Pharmacology

RECEIVED 06 June 2022
ACCEPTED 12 September 2022
PUBLISHED 05 October 2022

CITATION
Yang X, Xing X, Liu Y and Zheng Y (2022), Screening of potential inhibitors targeting the main protease structure of SARS-CoV-2 via molecular docking. *Front. Pharmacol.* 13:962863. doi: 10.3389/fphar.2022.962863

Screening of potential inhibitors targeting the main protease structure of SARS-CoV-2 *via* molecular docking

Xinbo Yang[1]*, Xianrong Xing[2], Yirui Liu[3] and Yuanjie Zheng[1]*

[1]School of Information Science and Engineering, Shandong Normal University, Jinan, China, [2]Department of Pharmacy, Shandong Medical College, Jinan, China, [3]School of Clinical Medicine, Harbin Medical University, Daqing, China

The novel coronavirus disease (COVID-19) caused by SARS-CoV-2 virus spreads rapidly to become a global pandemic. Researchers have been working to develop specific drugs to treat COVID-19. The main protease (M^{pro}) of SARS-CoV-2 virus plays a pivotal role in mediating viral replication and transcription, which makes it a potential therapeutic drug target against COVID-19. In this study, a virtual drug screening method based on the M^{pro} structure (Protein Data Bank ID: 6LU7) was proposed, and 8,820 compounds collected from the DrugBank database were used for molecular docking and virtual screening. A data set containing 1,545 drug molecules, derived from compounds with a low binding free energy score in the docking experiment, was established. N-1H-Indazol-5-yl-2-(6-methylpyridin-2-yl)quinazolin-4-amine, ergotamine, antrafenine, dihydroergotamine, and phthalocyanine outperformed the other compounds in binding conformation and binding free energy over the N3 inhibitor in the crystal structure. The bioactivity and ADMET properties of these five compounds were further investigated. These experimental results for five compounds suggested that they were potential therapeutics to be developed for clinical trials. To further verify the results of molecular docking, we also carried out molecular dynamics (MD) simulations on the complexes formed by the five compounds and M^{pro}. The five complexes showed stable affinity in terms of root mean square distance (RMSD), root mean square fluctuation (RMSF), radius of gyration (Rg), and hydrogen bond. It was further confirmed that the five compounds had potential inhibitory effects on SARS-CoV-2 M^{pro}.

KEYWORDS

SARS-CoV-2, main protease, molecular docking, virtual screening, drug repurposing, molecular dynamics

1 Introduction

From December 2019, the world witnessed an outbreak of an acute respiratory disease (Han et al., 2020; Rothan and Byrareddy, 2020). In the early stages of the disease outbreak, Zhou et al. (2020) obtained the full-length genomic sequences of the virus collected from five patients. These sequences exhibited 79.6% homology with SARS-CoV. In addition, the newly found virus exhibited 96% identity to bat coronavirus at the whole-genome level. The International Committee of Taxonomy of Viruses named the virus "severe acute respiratory syndrome coronavirus 2" (SARS-CoV-2), and the World Health Organization (WHO) announced this new disease as a novel coronavirus disease 2019 (COVID-19) (Anand et al., 2020; Wang et al., 2020). According to data from the WHO, over 526 million confirmed cases and over six million deaths have been recorded by 29 May 2022.

Similar to SARS-CoV, SARS-CoV-2 also belongs to the β-coronavirus class but is more contagious and mutable (Tang B. et al., 2020; Shereen et al., 2020). Vaccination has been widely promoted as an important preventive measure against COVID-19. As on 9 September 2022, the WHO reported that there were 371 COVID-19 vaccine candidates in development, of which 172 have entered clinical trials (COVID-19 vaccine tracker and landscape, 2022, https://www.who.int/publications/m/item/draft-landscape-of-covid-19-candidate-vaccines). Among the vaccines in clinical development, the number of types ranked was protein subunit vaccines (32%), RNA vaccines (23%), viral vector (non-replicating) vaccines (13%), inactivated virus vaccines (13%), DNA vaccines (9%), and other types of vaccines. As research on protein subunit vaccines was relatively mature and was the priority vaccine development method, the number of protein subunit vaccines was the largest among COVID-19 vaccines. However, persistent mutations of the virus can affect the vaccine's preventive effect, especially Omicron, which largely evaded the antibodies elicited by the vaccine (Planas et al., 2022). SARS-CoV-2 comprises a single-stranded positive-sense RNA genome that encodes both structural and non-structural proteins. The non-structural proteins include RNA-dependent RNA polymerase, coronavirus main protease (M^{pro}, also known as 3C-like protease, $3CL^{pro}$), and papain-like protease (PL^{pro}). When the viral genome enters the host cell, the host cell protein translation mechanism translates it into a viral polyprotein, which is then cleaved into effector proteins by the viral proteases M^{pro} and PL^{pro} (Tang X. et al., 2020; Liu et al., 2020; Zhang et al., 2020). Since M^{pro} can cleave polyproteins at no less than 11 conserved sites, it plays a vital role in the replication of viral particles (ul Qamar et al., 2020). Therefore, it is an attractive target for the screening of antiviral inhibitors. The high-resolution crystal structure of SARS-CoV-2 M^{pro} was presented by the Zihe Rao and Haitao Yang's research team. They also provided a basis for drug screening and design based on the structure of the M^{pro} (Jin et al., 2020).

The research and development of a new drug is a time-consuming process that requires huge financial investment. In the current global crisis, the repositioning of existing drugs seems to be a potentially useful tool in searching for new therapeutic options (Serafin et al., 2020). Computer-assisted virtual screening provides an inexpensive and rapid alternative to high-throughput screening for drug discovery. Furthermore, virtual screening technology can optimize the selection of potential drugs (de Carvalho Gallo et al., 2018). In the past few decades, virtual screening has played an important role in the discovery of small molecule inhibitors of therapeutic targets. Various ligands and structure-based virtual screening methods have been used to identify small-molecule ligands for proteins of interest (Bharatham et al., 2017; Singh and Jana, 2017; Li et al., 2020). Virtual screening technology has revealed several compound molecules that can inhibit SARS-CoV activity (Wei et al., 2006; Niu et al., 2008; Wang et al., 2017).

In this study, we investigated potential M^{pro} inhibitors using a docking-based virtual screening approach. We used a variety of screening strategies, such as molecular docking, molecular dynamics (MD) simulations, biological activity, and ADMET prediction. The AutoDock Tools were used to prepare the M^{pro} receptor model of SARS-CoV-2. A Vina-based molecular docking program was encoded, and M^{pro} and compounds (from DrugBank, with the 3D structure) were docked. The compounds were sorted based on the combined free energy score. The potential drug compounds with inhibitory effects on M^{pro} were determined by analyzing the binding mode between the compounds with better scoring results and M^{pro}. The bioactivity and ADMET properties of the five selected compounds were further explored. Simultaneously, we performed MD simulation experiments on the complexes of five compounds and M^{pro}. The purpose of this study was to identify potential drug compounds from DrugBank by molecular docking and MD simulations. This method can rapidly predict whether a compound has inhibitory effect on the activity of M^{pro} based on the physicochemical properties of the compound and the stability of the protein–ligand complex.

2 Materials and methods

2.1 Receptor (SARS-CoV-2 M^{pro} protein) preparation

SARS-CoV-2 M^{pro} is a key CoV enzyme, which plays a pivotal role in mediating viral replication and transcription, making it an attractive drug target for treating COVID-19 (Anand et al., 2002; Yang et al., 2003; Jin et al., 2020).

The complex crystal structure of M^{pro} and the N3 inhibitor (PDB ID: 6LU7) (Jin et al., 2020) was downloaded from the

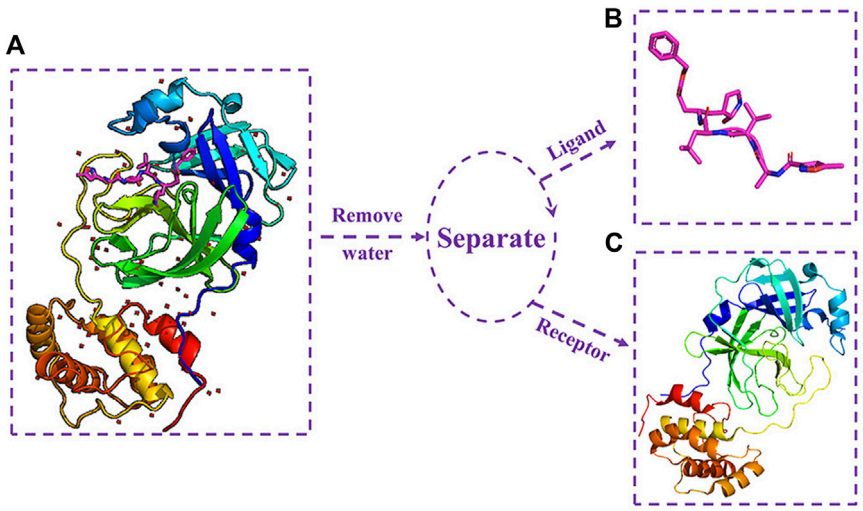


FIGURE 1
(A) Complex crystal structure of M^{pro} protein with N3; **(B)** three-dimensional structure of N3 inhibitor; and **(C)** three-dimensional structure of M^{pro} protein.

Protein Data Bank (http://www.rcsb.org). M^{pro} was isolated from the complex crystal structure using PyMOL. The separation process of M^{pro} and N3 inhibitor is shown in Figure 1. Figure 1A shows the complex crystal structure of M^{pro} protein with N3, and (c) shows the 3D structure of M^{pro}.

2.2 Ligand data set preparation

For the docking experimental ligand (which is composed of a drug molecule data set and N3 inhibitor), the N3 inhibitor was isolated from the SARS-CoV-2 main protease crystal complex. Figure 1B shows the 3D structure of N3 inhibitor.

The drug molecule data set contained 8,820 molecules with their 3D structures. They were obtained from DrugBank (https://www.drugbank.ca/) in the SDF format (Wishart et al., 2018).

2.3 Pre-processing of receptor and ligands

The docking program requires files stored in the Protein Data Bank, especially in the Partial Charge and Atom Type (PDBQT) format. M^{pro} standardization involved Gasteiger charges and the addition of polar hydrogen atoms. The conversion of the file format from the Protein Data Bank (PDB) format to PDBQT format was implemented using AutoDock Tools.

Data standardization was performed as a part of the pre-processing. The drug molecules of the ligand data set were first added to polar hydrogen atoms using Open Babel software (O'Boyle et al., 2011). Subsequently, Gasteiger charges were added using Raccoon (Forli et al., 2016) and broken down into 8,820 small molecule files in a PDBQT format.For the N3 inhibitor, Open Babel software was used to add polar hydrogen atoms and Gasteiger charges, followed by converting the format from PDB to PDBQT.

2.4 Molecular docking and screening

Molecular docking was performed using AutoDock Vina and the standardized docking data. In this study, the center of grid box was set to (−10.807, 12.541, 68.917) Å for (center_x, center_y, center_z). Meanwhile, the size of the grid box was defined as (30, 30, 30) Å for (size_x, size_y, size_z). Figure 2A shows the setting information of the grid box, and (b) shows the 3D structure of the grid box for M^{pro}. To generate as many different binding modes as possible, the num-modes was set to 20 (maximum number of binding modes to generate), and the energy range was set to 6 kcal/mol (maximum energy difference between the best binding mode and the worst one displayed [kcal/mol]). The number of CPUs was set to 20 (CPU = 20), and the explicit random seed was set to 200.

We encoded a bash script file to implement the docking process. This script file encapsulated the Vina program and the parameters required for the Vina program, including the parameters set in the previous paragraph and the input and output parameters. It could automatically execute the Vina program and perform docking experiments with each ligand molecule and the receptor and finally showed the score of each

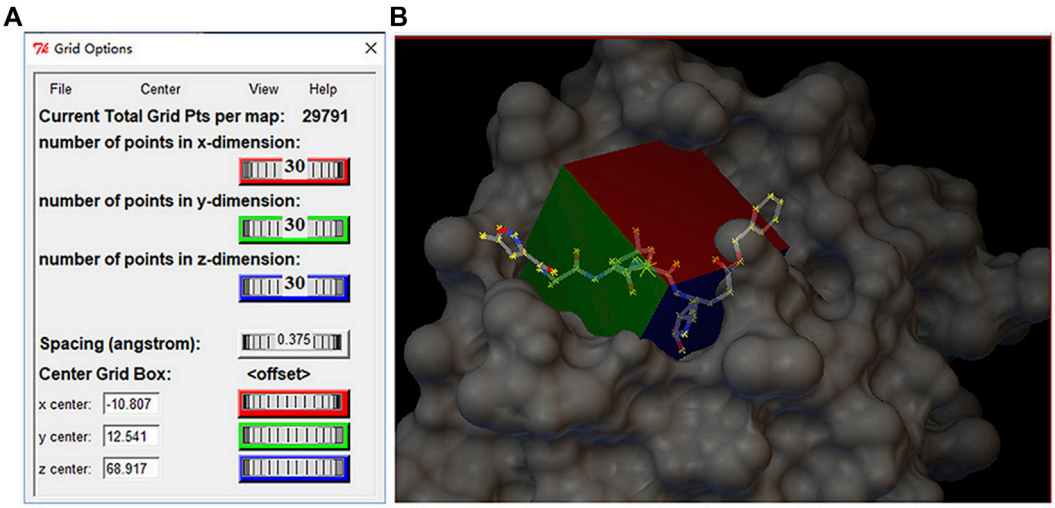


FIGURE 2
(A) Setting information of the grid box and**(B)** three-dimensional structure of gird box in the M^{pro} protein.

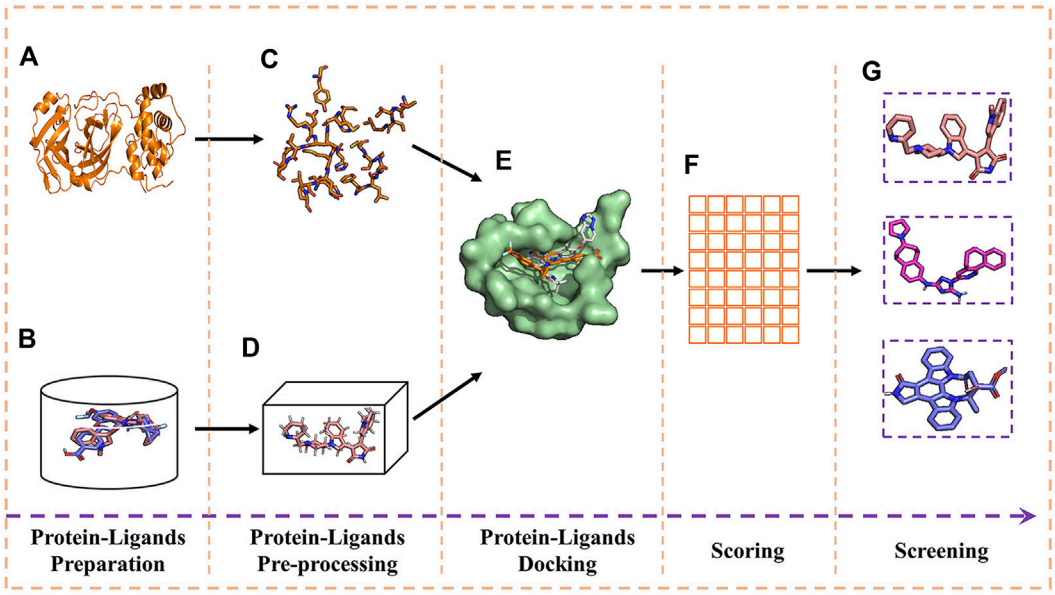


FIGURE 3
Experimental process of molecular docking and virtual screening. **(A)** Receptor; **(B)** ligands; **(C)** receptor pre-processing; **(D)** ligand pre-processing; **(E)** docking of receptor and ligands; **(F)** table of free energy score; and **(G)** virtual screening for the docking result.

ligand molecule. It was used to calculate the free energy score of M^{pro} with different conformations of each ligand.

Screening for the potential drug molecule was achieved by implementing a specific Python script program. The optimal docking score of M^{pro} with the N3 inhibitor was used as a reference standard for the analysis and screening to establish a candidate drug molecule data set. Figure 3 shows the experimental process of molecular docking and virtual screening.

Discovery Studio Visualizer was used to analyze the interactions and types of interactions between compounds and M^{pro} (docking complexes).

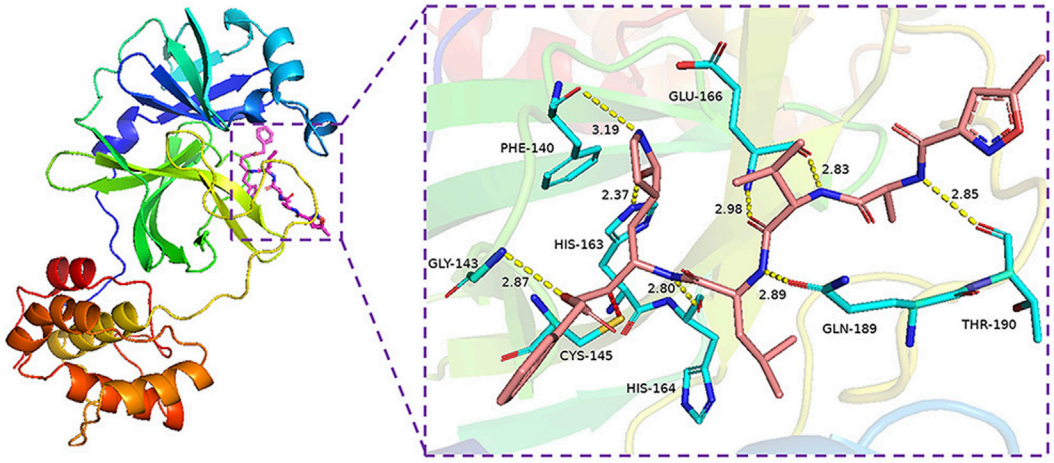


FIGURE 4
Interaction of covalent bonding and hydrogen bonding between M^{pro} protein and N3.

TABLE 1 Groups of 1,545 compounds in DrugBank.

No.	Group	Count
1	Approved	95
2	Approved; experimental	4
3	Approved; experimental; investigational	1
4	Approved; investigational	108
5	Experimental	736
6	Experimental; investigational	8
7	Investigational	543
8	Others	50

2.5 Molecular dynamics simulation

Molecular dynamics (MD) simulation is an effective method to predict the stability of protein–ligand complexes (Bharadwaj et al., 2021). In this study, we used GROMACS (version: 2022.2, https://www.gromacs.org/) for molecular dynamics simulation. The topology file for M^{pro} protease was generated using the gmx pdb2gmx command, with the addition of the gromos53a6 force field. The topologies of the five drug molecules were generated using the PRODRG (http://davapc1.bioch.dundee.ac.uk/) server, and their respective topology files with parameters set to chirality: Yes; charge: Full; and EM: NO was also generated.

The simulation system adopted a rectangular solvated box and used gmx grompp for the energy detection and minimization processes. The maximum number of minimization steps to perform was set to 50,000, the energy step size was set to 0.01, and the energy minimization algorithm adopted the steepest descent minimization. At the same time, the system was stabilized by 100 ps NVT and NPT balance. The V-rescale thermal bath coupling algorithm was used in the NVT ensemble, and the Parrinello–Rahman pressure coupling method was used in the NPT ensemble. Finally, we performed 100 ns MD simulations of the equilibrium system at a temperature of 300 K and pressure of 1 bar. The RMSD, RMSF, Rg, number of hydrogen bonds, and protein–ligand interactions of the MD simulation results were recorded and analyzed for further validation of our virtual screening results. The results of molecular dynamics simulations were visualized using qtgrace (version: V26) software.

2.6 Bioactivity and ADMET property prediction

As an alternative to clinical experiments, computer technology was a fast and efficient method to predict the pharmacodynamic properties of compounds (Zaki et al., 2022). We combined the DrugBank database and used Molinspiration Cheminformatics (https://www.molinspiration.com) and admetSAR web service (Yang et al., 2019) to predict the bioactivity and ADMET properties of the five screened compounds, respectively.

3 Results

3.1 Molecular docking and screening

In this study, AutoDock Vina was used to perform the docking of the screened molecules with modeled M^{pro}. Each ligand generated 20 conformations. These conformations were further subjected to virtual screening evaluation. From the docking search, the conformation with the lowest docked energy was selected as the

TABLE 2 Top 30 compounds from the docking results.

No.	Accession number	Chemical formula	Group	Binding energy (kcal/mol)
1	DB12983	$C_{32}H_{18}N_8$	Investigational	-10.7
2	DB12225	$C_{36}H_{45}N_5O_5S$	Investigational	-10.4
3	DB11651	$C_{30}H_{23}N_5O$	Investigational	-10.3
4	DB13050	$C_{38}H_{52}N_6O_2$	Investigational	-10.2
5	DB11913	$C_{28}H_{25}FN_6O_3$	Investigational	-10.1
6	DB14883	$C_{29}H_{24}FN_7O$	Investigational	-10.1
7	DB14894	$C_{28}H_{21}F_4NO_7$	Investigational	-10.0
8	DB00320	$C_{33}H_{37}N_5O_5$	Approved; investigational	-9.9
9	DB06486	$C_{32}H_{29}N_5O_2$	Investigational	-9.9
10	DB08450	$C_{21}H_{16}N_6$	Experimental	-9.8
11	DB12411	$C_{30}H_{34}N_8$	Investigational	-9.8
12	DB00696	$C_{33}H_{35}N_5O_5$	Approved	-9.7
13	DB04868	$C_{28}H_{22}F_3N_7O$	Approved; investigational	-9.7
14	DB12323	$C_{27}H_{21}F_3N_8O$	Investigational	-9.7
15	DB12719	$C_{25}H_{24}F_2N_2O_3$	Investigational	-9.7
16	DB11791	$C_{23}H_{17}FN_6O$	Approved; investigational	-9.6
17	DB11799	$C_{21}H_{18}F_3N_3O_5$	Approved; investigational	-9.6
18	DB11977	$C_{33}H_{37}F_2N_7O_4$	Investigational	-9.6
19	DB13648	$C_{44}H_{50}N_4O_2$	Experimental	-9.6
20	DB00820	$C_{22}H_{19}N_3O_4$	Approved; investigational	-9.5
21	DB01761	$C_{28}H_{29}F_3N_6$	Experimental	-9.5
22	DB04016	$C_{40}H_{35}N_2O_6P$	Experimental	-9.5
23	DB06888	$C_{22}H_{21}N_5O_3$	Experimental	-9.5
24	DB12200	$C_{23}H_{19}N_3O_2$	Investigational	-9.5
25	DB13109	$C_{25}H_{28}N_8O_3$	Investigational	-9.5
26	DB01200	$C_{32}H_{40}BrN_5O_5$	Approved; investigational	-9.4
27	DB01419	$C_{30}H_{26}F_6N_4O_2$	Approved	-9.4
28	DB04330	$C_{29}H_{19}Cl_2N_3O_6S$	Experimental	-9.4
29	DB06630	$C_{30}H_{25}F_{10}NO_3$	Investigational	-9.4
30	DB07020	$C_{20}H_{14}N_6O_2$	Experimental	-9.4

best conformation. The molecular docking results for AutoDock Vina are presented in Supplementary Table S1.

The binding energy of the interaction of N3 with M^{pro} was -7.8 kcal/mol. N3 is mainly stabilized by interacting with M^{pro} through the formation of covalent and hydrogen bonds. The S atom of Cys145 of M^{pro} forms a covalent bond with C20 of N3. As shown in Figure 4, N3 forms seven hydrogen bonds with Gly143, His164, Glu166, Thr190, Gln189, His163, and Phe140 of M^{pro} (Tang B. et al., 2020). These results show the active pocket position of M^{pro}. As a reference, 1,545 compounds, with energy values lower than -7.8 kcal/mol, were obtained. The groups of these compounds in DrugBank are listed in Table 1. According to the order of energy value, the top 30 compounds from the molecular docking analysis are listed in Table 2.

The top 30 compounds were distributed among four different groups of compounds. Among these, we selected the compounds with the best docking energy. For the "approved" type, we selected two compounds. Next, we analyzed the interactions of the compounds $C_{33}H_{35}N_5O_5$ (DB00696, generic name: ergotamine), $C_{30}H_{26}F_6N_4O_2$ (DB01419, generic name: antrafenine), $C_{33}H_{37}N_5O_5$ (DB00320, generic name: dihydroergotamine), $C_{21}H_{16}N_6$ (DB08450, generic name: N-1H-indazol-5-yl-2-(6-methylpyridin-2-yl) quinazolin-4-amine), and $C_{32}H_{18}N_8$ (DB12983, generic name: phthalocyanine) with M^{pro}.

The conformation diagrams of these compounds are displayed in Figure 5. The interactions of the M^{pro} protease with each of the five molecules are shown in Figures 6–10. Hydrophobic interactions were visualized using LIGPLOT (Laskowski and Swindells, 2011) Figure 7. Other interactions, including conventional hydrogen bond, carbon–hydrogen bond, pi–donor hydrogen bond, alkyl, pi–alkyl, halogen (fluorine), and pi–pi t-shaped, were visualized using Discovery Studio Visualizer. The interactions of residues with their respective ligands are shown in Supplementary Table S2.

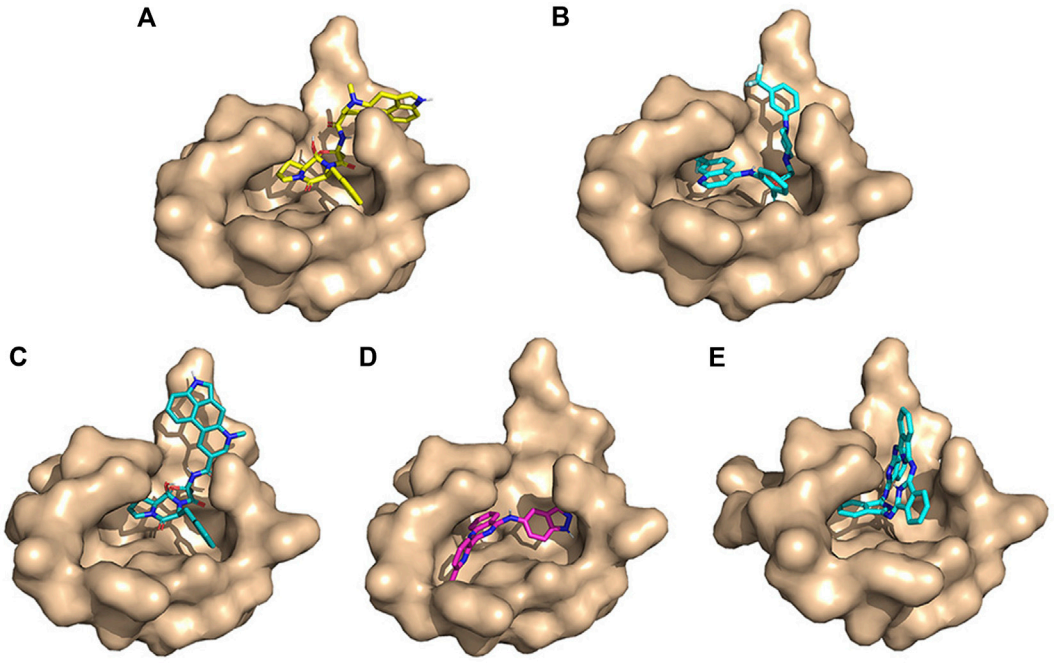


FIGURE 5
Conformation diagrams of these compounds. **(A)** Ergotamine; **(B)** antrafenine; **(C)** dihydroergotamine; **(D)** N-1H-indazol-5-yl-2-(6-methylpyridin-2-yl)quinazolin-4-amine; and **(E)** phthalocyanine.

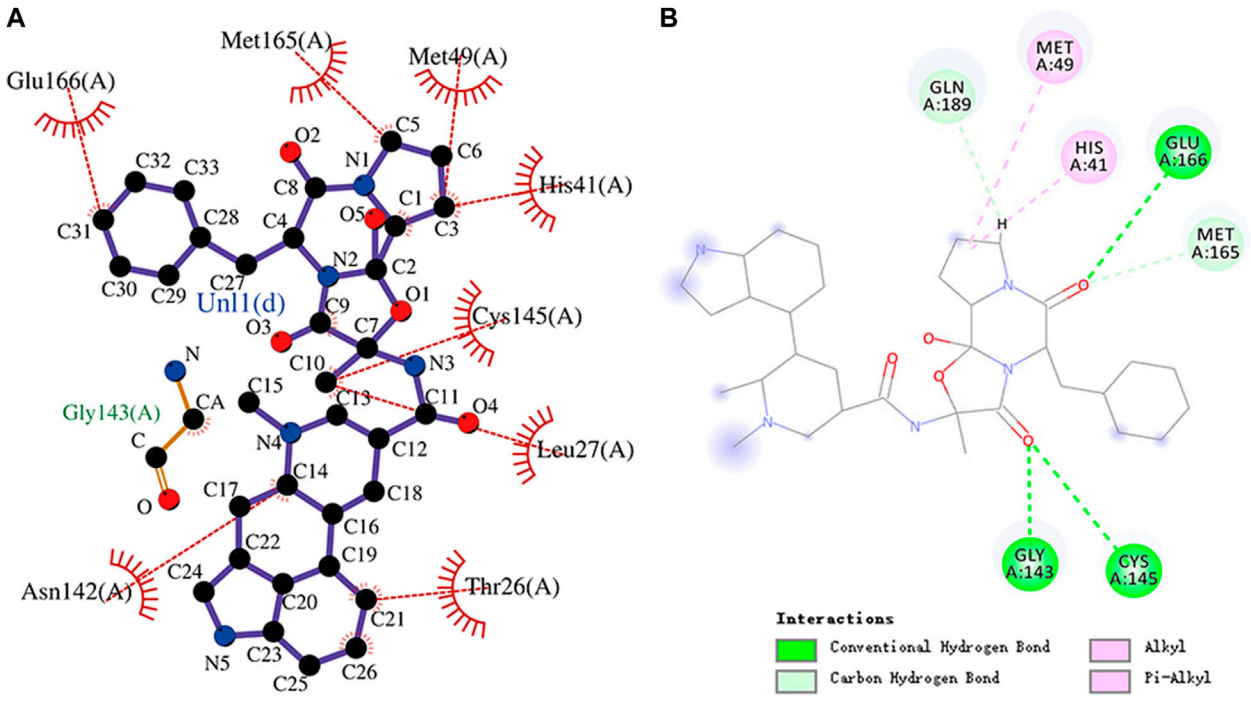


FIGURE 6
(A) Hydrophobic interaction between M^{pro} and ergotamine and **(B)** two-dimensional plot of ergotamine interaction with the amino acid residues.

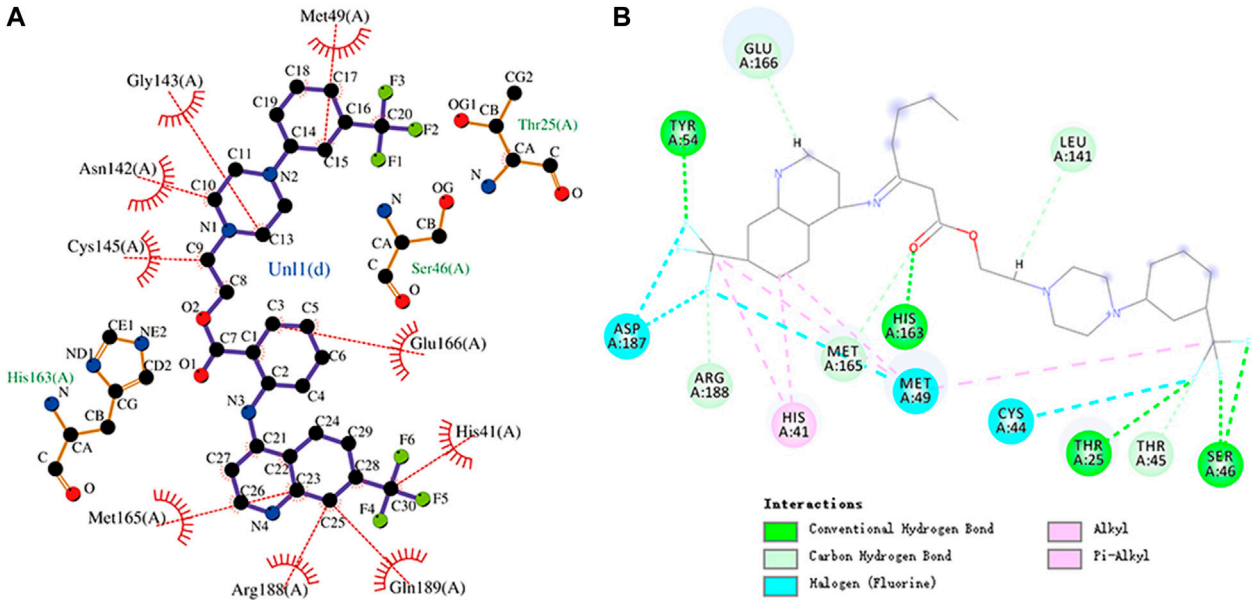


FIGURE 7
(A) Hydrophobic interaction between M^{pro} and antrafenine and **(B)** two-dimensional plot of antrafenine interaction with the amino acid residues.

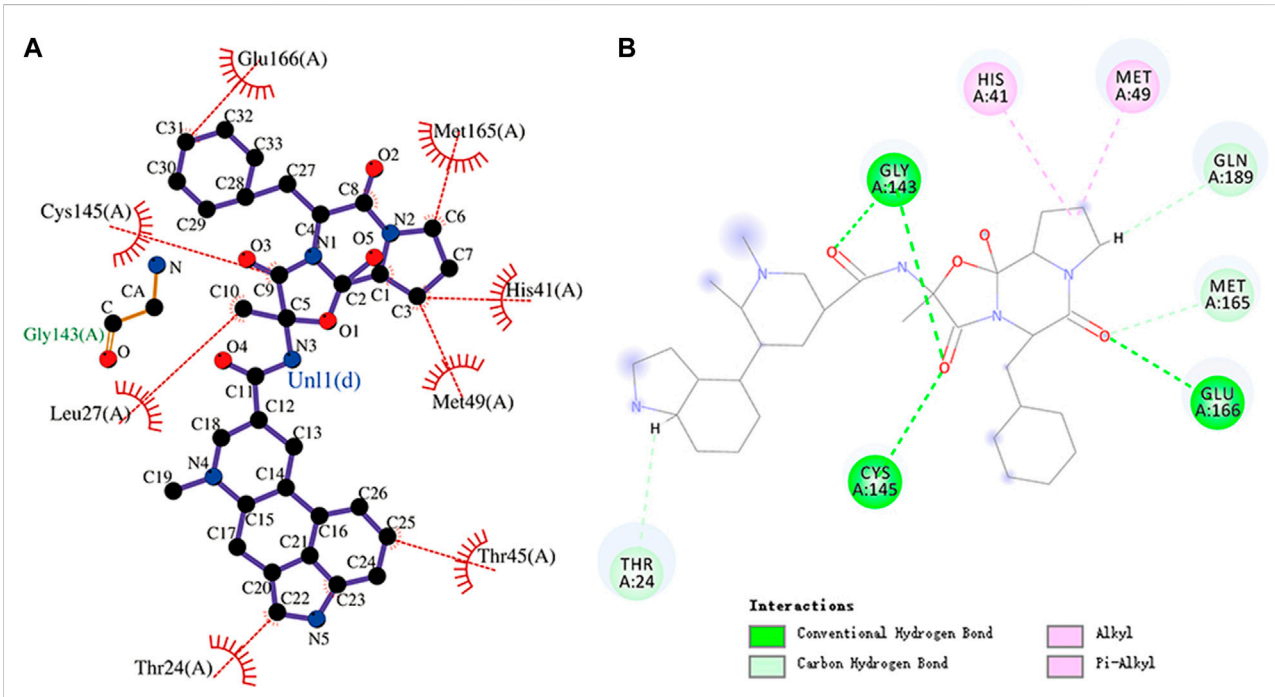


FIGURE 8
(A) Hydrophobic interaction between M^{pro} and dihydroergotamine and **(B)** two-dimensional plot of dihydroergotamine interaction with the amino acid residues.

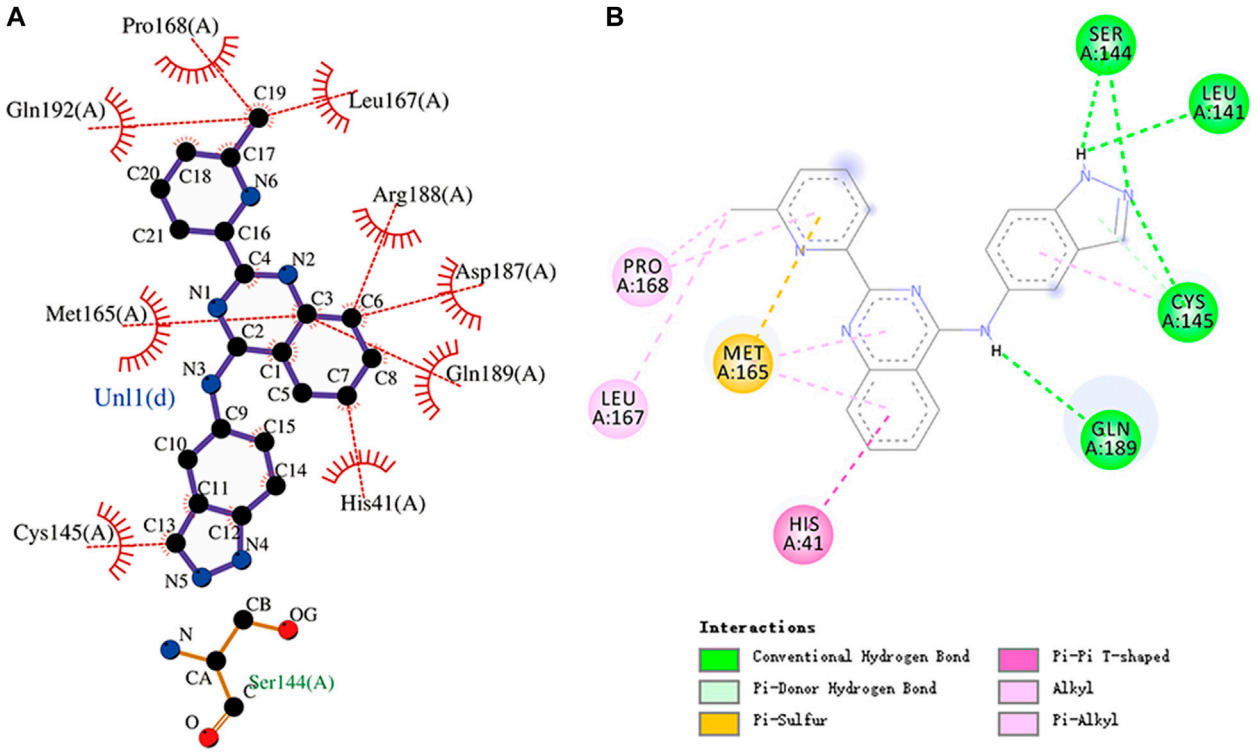


FIGURE 9
(A) Hydrophobic interaction between M^{pro} and N-1H-indazol-5-yl-2-(6-methylpyridin-2-yl)quinazolin-4-amine and **(B)** two-dimensional plot of N-1H-indazol-5-yl-2-(6-methylpyridin-2-yl)quinazolin-4-amine interaction with the amino acid residues.

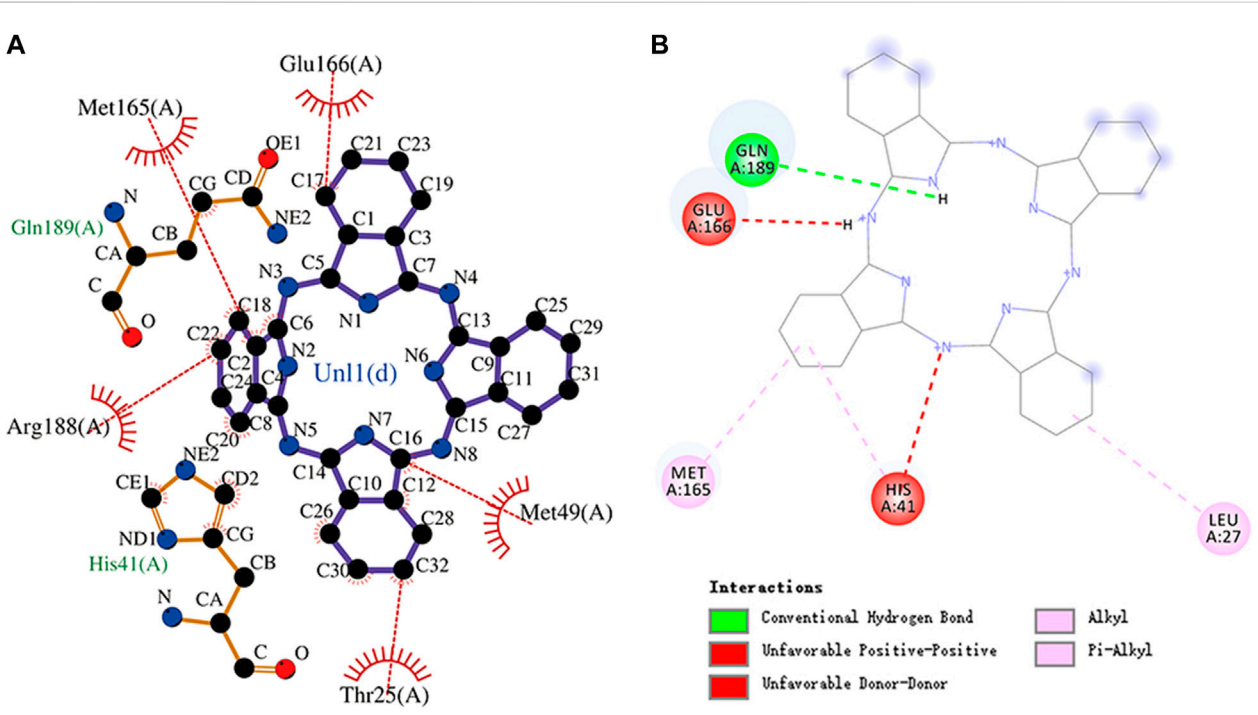


FIGURE 10
(A) Hydrophobic interaction between M^{pro} and phthalocyanine and **(B)** two-dimensional plot of phthalocyanine interaction with the amino acid residues.

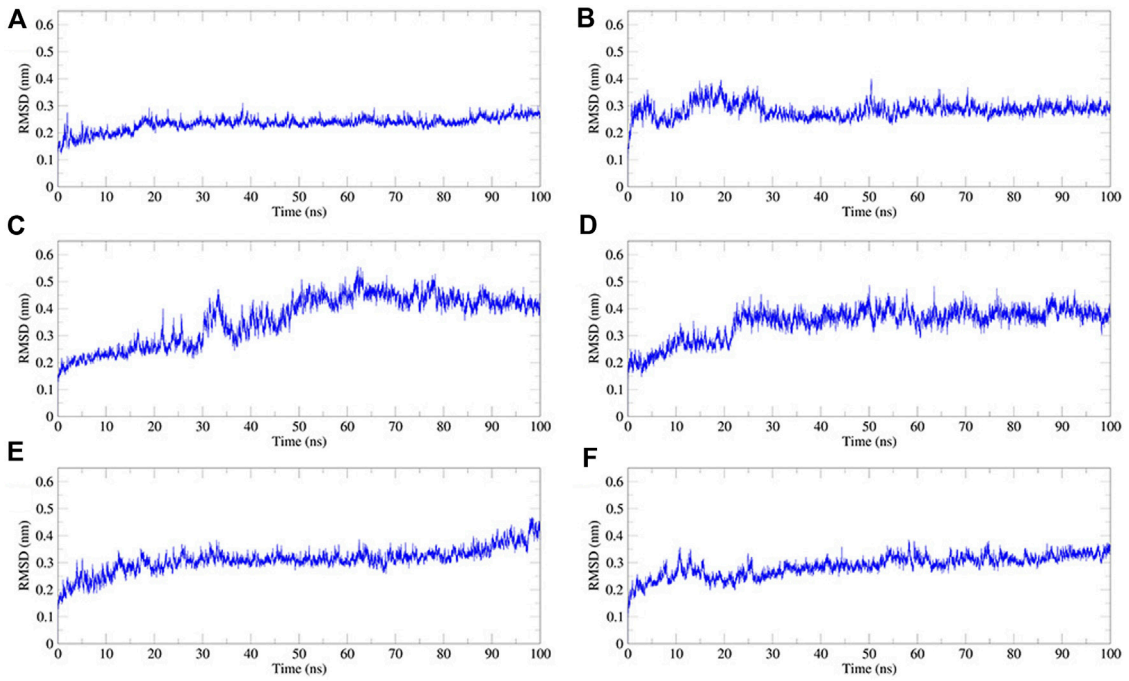


FIGURE 11
Plot of root mean square deviation (RMSD) values, during 100 ns MD simulation of compound–M^{pro} complexes. **(A)** N3–M^{pro} RMSD; **(B)** N-1H-indazol-5-yl-2-(6-methylpyridin-2-yl)quinazolin-4-amine–M^{pro} RMSD; **(C)** phthalocyanine–M^{pro} RMSD; **(D)** antrafenine–-M^{pro} RMSD **(E)** ergotamine–M^{pro} RMSD; and **(F)** dihydroergotamine–M^{pro} RMSD.

3.2 Molecular dynamics simulation

We performed 100 ns MD simulations for each of the five compounds and N3 inhibitors in complex with M^{pro}. As shown in Figure 8, complex N3–M^{pro} trajectory stabilized around 20 ns, N-1H-indazol-5-yl-2-(6-methylpyridin-2-yl) quinazolin-4-amine–M^{pro} stabilized around 25 ns, phthalocyanine–M^{pro} stabilized around 35 ns, antrafenine–M^{pro} stabilized around 25 ns, ergotamine–M^{pro} stabilized around 25 ns, and dihydroergotamine–M^{pro} stabilized around 25 ns The results showed that the RMSD of the five complexes in the MD trajectory interval (35–50 ns) fluctuated from 2.47 to 3.59 Å for dihydroergotamine Figure 9, 2.71–3.71 Å for ergotamine, 2.43–4.75 Å for phthalocyanine, 2.99–4.56 Å for antrafenine, 2.25–3.40 Å for N-1H-indazol-5-yl-2-(6-methylpyridin-2-yl) quinazolin-4-amine, and 1.91–3.33 Å for the N3 inhibitor and M^{pro} complex. The smaller the RMSD value, the smaller the fluctuation of the complex structure. Compared with N3 inhibitors, the RMSD of the five molecule–M^{pro} complexes have little difference Figure 10.

Rg is an important indicator for evaluating the compactness of the docking architecture. The smaller the cyclotron radius, the better the compactness Figure 11, and hence, the more stable the protein structure. The Rg results of the five molecules and N3 inhibitors in the complex with M^{pro} are shown in Figure 12. The average Rg of the N3–M^{pro} complex was about 22.5 Å. The average Rg of the five complexes showed little difference and was lower than that of the N3–M^{pro} complex except for ergotamine–M^{pro} complex. The average Rg of the drug dihydroergotamine–M^{pro}, phthalocyanine–M^{pro}, N-1H-indazol-5-yl-2-(6-methylpyridin-2-yl) quinazolin-4-amine–M^{pro}, and antrafenine–M^{pro} complexes were all about 22 Å. The average Rg of the ergotamine–M^{pro} complex was about the highest (22.8 Å). This shows that the results of the Rg analysis are consistent with the results of the RMSD trajectory analysis.

The RMSF can be used to observe how individual amino acids fluctuate during the simulation. It is possible to compare the effects of different small-molecule ligands on the spatial structural fluctuations of proteins by calculating the RMSF value. The smaller the value of RMSF, the smaller the disturbance of the small-molecule ligand to the protein, and therefore the stronger the stability of the complex. We calculated the RMSF value for each of the five small molecules and N3 inhibitors bound to M^{pro}. The calculated values are shown in Figure 13. The RMSF results showed that the average RMSF value of M^{pro} bound to dihydroergotamine was 1.75 Å, which indicates less fluctuation in the complex

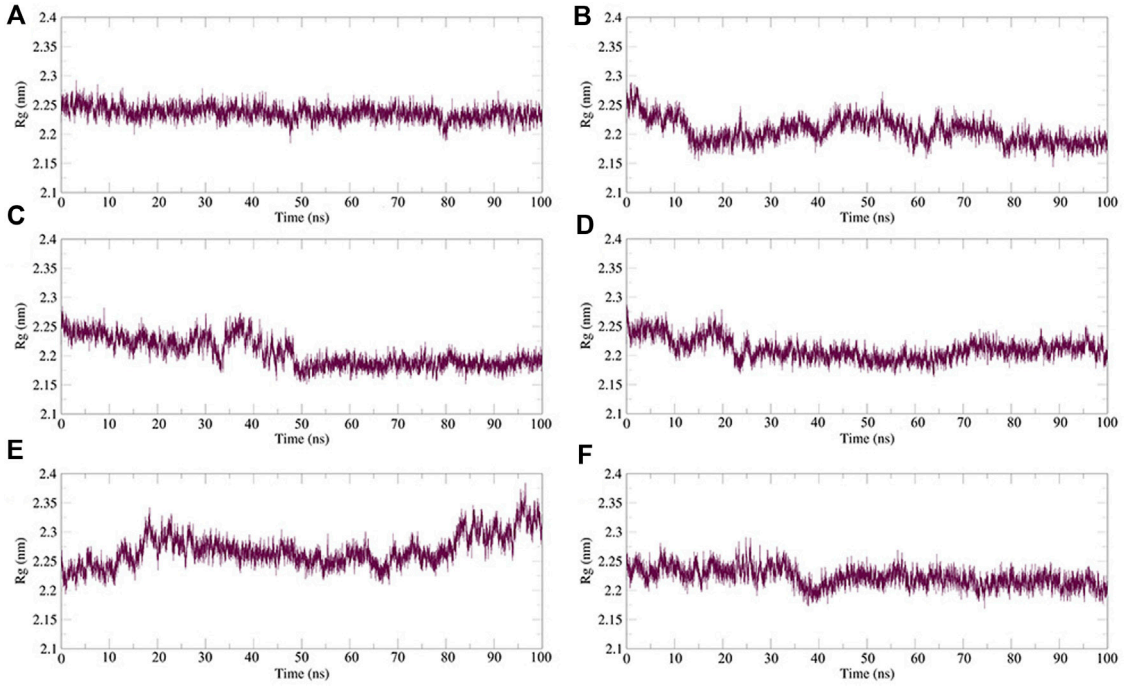


FIGURE 12
Plot of radius of gyration (Rg) values, during 100 ns MD simulation of compound–M^{pro} complexes. **(A)** N3–M^{pro} Rg; **(B)** N-1H-indazol-5-yl-2-(6-methylpyridin-2-yl)quinazolin-4-amine–M^{pro} Rg; **(C)** phthalocyanine–M^{pro} Rg; **(D)** antrafenine–M^{pro} Rg; **(E)** ergotamine–M^{pro} Rg; and **(F)** dihydroergotamine–M^{pro} Rg.

structure. However, the residues Met49 (7.07 Å), Tyr54 (3.67 Å), Arg188 (2.30 Å), and Thr24 (2.30 Å) showed a slight fluctuation in the dihydroergotamine–M^{pro} complex during the simulation. Thus, from the perspective of RMSF, the stability order of the complex formed with the main protease is dihydroergotamine, antrafenine, ergotamine, N-1H-indazol-5-yl-2-(6-methylpyridin-2-yl) quinazolin-4-amine, and phthalocyanine.

Hydrogen bonds between the ligand and key residues of the main protease were investigated using 100 ns MD simulations as shown in Figure 14. During the 100 ns simulation, there were multiple hydrogen bonds between the five compounds and M^{pro}. The results confirmed that the five compounds in the MD system had a strong inhibitory effect on M^{pro}, and there was a good binding effect between the compounds and M^{pro} in the pocket of M^{pro}.

3.3 Pharmacodynamic properties

The results of bioactivity and ADMET are shown in Supplementary Tables S3, S4, respectively.

Molinspiration Cheminformatics can predict the GPCR ligand, ion channel modulator, kinase inhibitor, nuclear receptor ligand, protease inhibitor, and enzyme inhibitor values of compounds to evaluate their biological activities. According to reports, a bioactivity score of -5.0–0.0 is considered moderately active, and a score of ≥0 is considered active (Mokhnache et al., 2019; Rahman et al., 2021). From the predicted results, it can be concluded that N-1H-indazol-5-yl-2-(6-methylpyridin-2-yl)quinazolin-4-amine (score 0.38) is an active enzyme inhibitor, and the other four compounds can be approximately considered active enzyme inhibitors.

The admetSAR can predict ADMET for pharmacodynamic studies of five compounds in the host. In Supplementary Table S4, parameters such as molecular weight, water solubility (logS), human intestinal absorption, blood–brain barrier, Caco-2 permeable, human oral bioavailability, and toxicity of the compounds were listed. The results showed that the solubility value of antrafenine was slightly lower than -4, while the values of other compounds were higher than -4. This indicated that the solubility of the five compounds was suitable (Rahman et al., 2021). None of the five compounds were carcinogenic. But it should be noted that phthalocyanine and N-1H-indazol-5-yl-2-(6-methylpyridin-2-yl)quinazolin-4-amine were shown to have AMES toxicity. Regarding drug-likeliness, only N-1H-indazol-5-yl-2-(6-methylpyridin-2-yl)quinazolin-4-amine met the requirements of the five rules, and other compounds were larger than the ideal molecular weight of 500. By comparison,

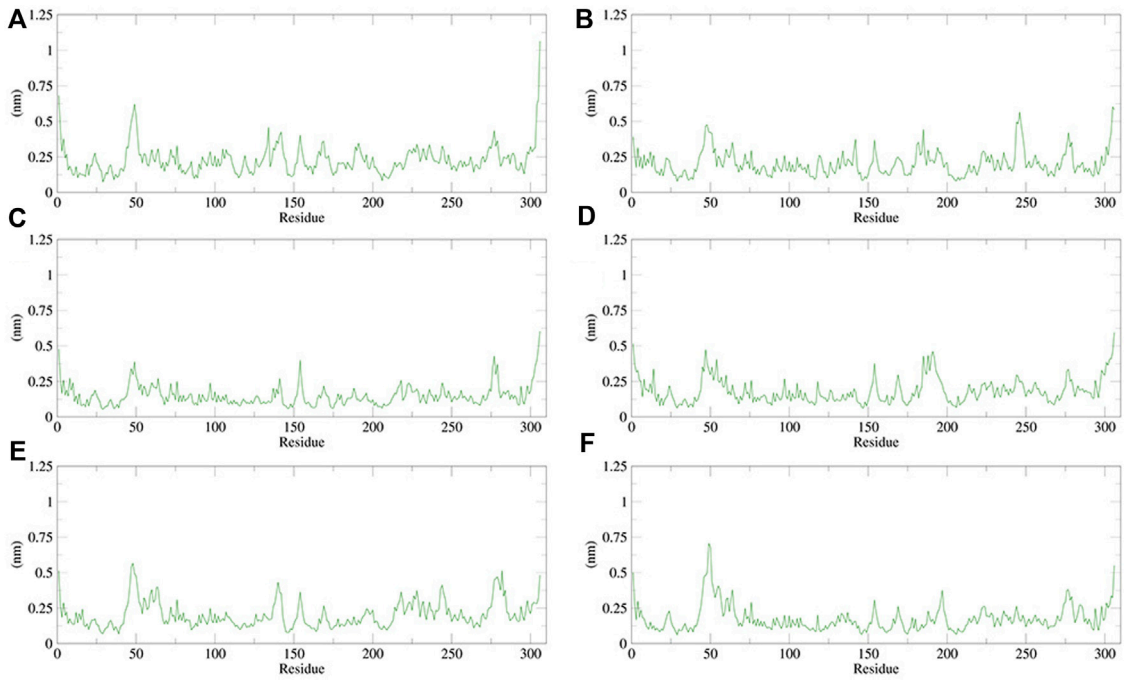


FIGURE 13
Plot of root mean square fluctuations (RMSF) values, during 100 ns MD simulation of compound–M^{pro} complexes. **(A)** N3–M^{pro} RMSF; **(B)** N-1H-indazol-5-yl-2-(6-methylpyridin-2-yl)quinazolin-4-amine–M^{pro} RMSF; **(C)** phthalocyanine–M^{pro} RMSF; **(D)** antrafenine–M^{pro} RMSF; **(E)** ergotamine–M^{pro} RMSF; and **(F)** dihydroergotamine–M^{pro} RMSF.

it was concluded that the drug candidate order of the five compounds was ergotamine, dihydroergotamine, antrafenine, N-1H-indazol-5-yl-2-(6-methylpyridin-2-yl)quinazolin-4-amine, and phthalocyanine.

4 Discussion

The docking of small-molecule compounds to receptor binding sites and the estimation of the binding affinity of the complex are important components of the structure-based drug design process. AutoDock Vina is an open-source program for drug discovery, molecular docking, and virtual screening, which significantly improves the average accuracy of the binding mode predictions (Herowati and Widodo, 2014; Xiang et al., 2015).

Ergotamine is an α-1 selective adrenergic agonist that is commonly used in the treatment of migraine disorders. The binding energy of ergotamine and M^{pro} was -9.7 kcal/mol. Ergotamine forms hydrogen bonds with residues Gly143, Cys145, and Glu166, respectively. Glu166, Met165, Met49, His41, Cys145, Leu27, Thr26, and Asn142 residues and the hydrophobic groups of ergotamine can engage through hydrophobic interactions. Ergotamine has an alkyl interaction with residue Met49 and a pi–alkyl interaction with residue His41. The molecular docking representation of ergotamine with M^{pro} is shown in Figure 6.

The interaction energy between antrafenine and M^{pro} was -9.4 kcal/mol. Antrafenine forms hydrogen bonds with residues Thr25, Ser46, Tyr54, and His163, respectively. Met49, Glu166, His41, Gln189, Arg188, Met165, Cys145, Asn142, and Gly143 residues and antrafenine can engage through hydrophobic interactions. There were also alkyl, pi–alkyl, and halogen (fluorine) interactions between antrafenine and residues. The molecular docking representation of antrafenine with M^{pro} is shown in. Moreover, antrafenine is a piperazine derivative drug, which exhibits analgesic and anti-inflammatory effects similar to naproxen.

The interaction energy between dihydroergotamine and M^{pro} was -9.9 kcal/mol. Dihydroergotamine forms hydrogen bonds with residues Gly143, Cys145, and Glu166, respectively. Moreover, dihydroergotamine is stabilized by the interaction with M^{pro} through hydrophobic interactions, involving Glu166, Leu27, Cys145, Thr24, Thr45, Met49, His41, and Met165 residues. There were also carbon–hydrogen bond, alkyl, and pi–alkyl interactions between dihydroergotamine and residues. The molecular docking representation of dihydroergotamine with M^{pro} is shown in .

N-1H-indazol-5-yl-2-(6-methylpyridin-2-yl)quinazolin-4-amine is an experimental drug molecule. The binding energy of N-1H-indazol-5-yl-2-(6-methylpyridin-2-yl)quinazolin-4-amine and M^{pro} was -9.8 kcal/mol. N-1H-indazol-5-yl-2-(6-methylpyridin -2-yl)

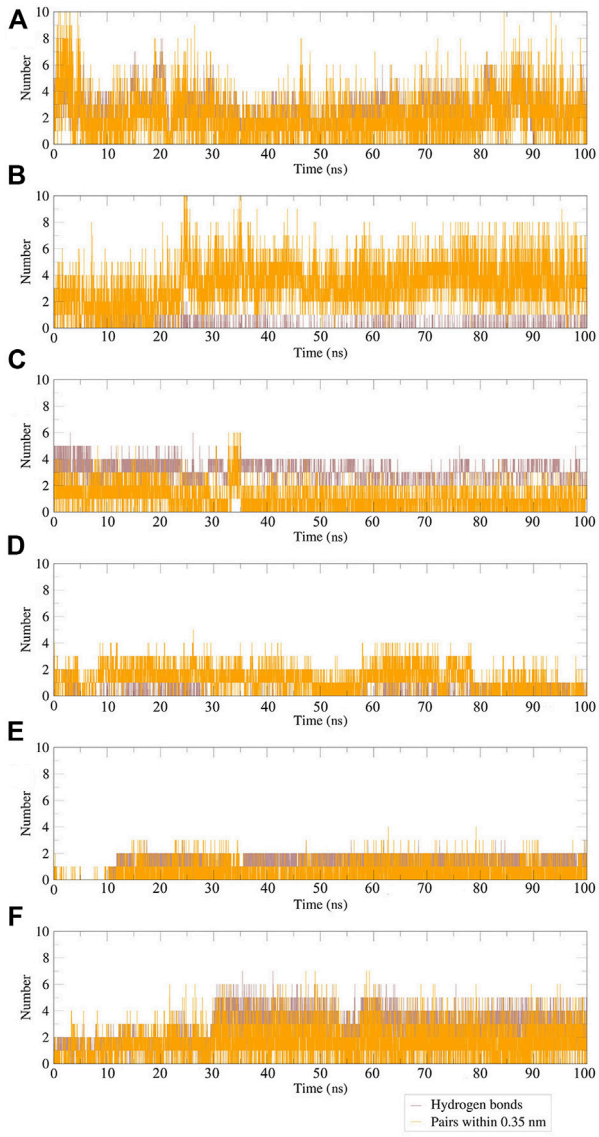


FIGURE 14
Plot of hydrogen bonds in compound–M^{pro} complexes during 100 ns MD simulation. **(A)** N3–M^{pro}; **(B)** N-1H-indazol-5-yl-2-(6-methylpyridin-2-yl)quinazolin-4-amine–M^{pro}; **(C)** antrafenine–M^{pro}; **(D)** ergotamine–M^{pro}; **(E)** dihydroergotamine–M^{pro}; **(F)** phthalocyanine–M^{pro}.

quinazolin-4-amine forms hydrogen bonds with residues Leu141, Ser144, Cys145, and Gln189, respectively. Gln192, Pro168, and Leu167 residues and the methyl of N-1H-indazol-5-yl-2-(6-methylpyridin-2-yl)quinazolin-4-amine can engage through hydrophobic interactions. Furthermore, Arg188, Asp187, Met165, Gln189, His41, and Cys145 residues can form hydrophobic interactions with N-1H-indazol-5-yl-2-(6-methylpyridin -2-yl) quinazolin-4-amine. There were also pi–donor hydrogen bond, alkyl, pi–alkyl, halogen (fluorine), and pi–pi T-shaped interactions between N-1H-indazol-5-yl-2-(6 -methylpyridin-2-yl)quinazolin-4-amine and residues. The molecular docking representation of N-1H-indazol-5-yl-2-(6-methylpyridin-2-yl)quinazolin-4-amine with M^{pro} is shown in.

Phthalocyanine is an 18-electron large conjugated system compound comprising four isoindole units. The center of the conjugated ring structure has a large cavity that can accommodate metal ions (such as iron, cobalt, and nickel). Phthalocyanine has the lowest binding energy value of -10.7 kcal/mol. Glu166, Met165, Arg188, His41, Met49, and Thr25 residues and the isoindole ring of phthalocyanine can engage through hydrophobic interactions. The carbonyl group of Gln189 (hydrogen bond acceptor) forms a hydrogen bond with the NH group of acting phthalocyanine (hydrogen bond donor). Phthalocyanine also has alkyl and pi–alkyl interactions with residues. However, it should be noted that phthalocyanine has unfavorable interactions with residues His41 and Glu166, respectively. The molecular docking representation of phthalocyanine with M^{pro} is shown in.

5 Conclusion

Molecular docking and molecular dynamics simulations have been widely used in drug screening and drug design. In this study, we present several exciting findings about SARS-CoV-2 M^{pro}. The compounds analyzed in this study can be used as potential inhibitors of SARS-CoV-2 M^{pro}: Ergotamine is an approved medication for the treatment of migraine disorders, and antrafenine is used as an anti-inflammatory and analgesic agent for the relief of mild-to-moderate pain. Furthermore, we have uncovered dihydroergotamine, N-1H-indazol-5-yl-2-(6-methylpyridin-2-yl) quinazolin-4-amine, and phthalocyanine, which may be developed as potential treatments against SARS-CoV-2 infections. Structural optimization and clinical trials are needed for these compounds to become strong drug candidates. At present, no biological experiments have been carried out in this study. However, through high-throughput molecular docking and molecular dynamics simulations, it was confirmed that these five compounds can form stable conformational structures with M^{pro} and have potential inhibitory effects on SARS-CoV-2. At the same time, this study provides research ideas and helps for drug designing and drug reusing for the treatment of SARS-CoV-2.

Data availability statement

Publicly available datasets were analyzed in this study. These data can be found at: the complex crystal structure of M^{pro} protein with an N3 inhibitor (PDB ID: 6lu7) was downloaded from the Protein Data Bank (http://www.rcsb.org). The drug molecule data set contains 8,820 molecules, with 3D structures, which is in the SDF format obtained from DrugBank (https://www.drugbank.ca/).

Author contributions

XY and YZ contributed to the conception and design of the study. YL and XY contributed to the collection and collation of data. XY and XX performed the statistical analysis. XY and XX wrote the first draft of the manuscript. All authors contributed to manuscript revision and read and approved the submitted version.

Funding

This work was supported in part by the National Natural Science Foundation of China (No. 81871508) and the Major Program of Shandong Province Natural Science Foundation (ZR2019ZD04).

Acknowledgments

The authors kindly thank all participants for their contributions to this study. The authors also thank the Protein Data Bank and DrugBank for their data support services.

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fphar.2022.962863/full#supplementary-material

References

Anand, K., Karade, S., Sen, S., and Gupta, R. (2020). Sars-cov-2: camazotz's curse. *Med. J. Armed Forces India* 76, 136–141. doi:10.1016/j.mjafi.2020.04.008

Anand, K., Palm, G. J., Mesters, J. R., Siddell, S. G., Ziebuhr, J., and Hilgenfeld, R. (2002). Structure of coronavirus main proteinase reveals combination of a chymotrypsin fold with an extra α-helical domain. *EMBO J.* 21, 3213–3224. doi:10.1093/emboj/cdf327

Bharadwaj, K. K., Sarkar, T., Ghosh, A., Baishya, D., Rabha, B., Panda, M. K., et al. (2021). Macrolactin a as a novel inhibitory agent for sars-cov-2 mpro: Bioinformatics approach. *Appl. Biochem. Biotechnol.* 193, 3371–3394. doi:10.1007/s12010-021-03608-7

Bharatham, N., Finch, K. E., Min, J., Mayasundari, A., Dyer, M. A., Guy, R. K., et al. (2017). Performance of a docking/molecular dynamics protocol for virtual screening of nutlin-class inhibitors of mdmx. *J. Mol. Graph. Model.* 74, 54–60. doi:10.1016/j.jmgm.2017.02.014

de Carvalho Gallo, J. C., de Mattos Oliveira, L., Araújo, J. S. C., Santana, I. B., and dos Santos Junior, M. C. (2018). Virtual screening to identify leishmania braziliensis n-myristoyltransferase inhibitors: Pharmacophore models, docking, and molecular dynamics. *J. Mol. Model.* 24, 260. doi:10.1007/s00894-018-3791-8

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

Han, Q., Lin, Q., Jin, S., and You, L. (2020). Coronavirus 2019-ncov: A brief perspective from the front line. *J. Infect.* 80, 373–377. doi:10.1016/j.jinf.2020.02.010

Herowati, R., and Widodo, G. P. (2014). Molecular docking studies of chemical constituents of tinospora cordifolia on glycogen phosphorylase. *Procedia Chem.* 13, 63–68. doi:10.1016/j.proche.2014.12.007

Jin, Z., Du, X., Xu, Y., Deng, Y., Liu, M., Zhao, Y., et al. (2020). Structure of mpro from sars-cov-2 and discovery of its inhibitors. *Nature* 582, 289–293. doi:10.1038/s41586-020-2223-y

Laskowski, R. A., and Swindells, M. B. (2011). Ligplot+: Multiple ligand–protein interaction diagrams for drug discovery. *J. Chem. Inf. Model* 51 (10), 2778–2786. doi:10.1021/ci200227u

Li, T., Tan, X., Yang, R., Miao, Y., Zhang, M., Xi, Y., et al. (2020). Discovery of novel glyceraldehyde-3-phosphate dehydrogenase inhibitor via docking-based virtual screening. *Bioorg. Chem.* 96, 103620. doi:10.1016/j.bioorg.2020.103620

Liu, C., Zhou, Q., Li, Y., Garner, L. V., Watkins, S. P., Carter, L. J., et al. (2020). Research and development on therapeutic agents and vaccines for Covid-19 and related human coronavirus diseases. *ACS Cent. Sci.* 6, 315–331. doi:10.1021/acscentsci.0c00272

Mokhnache, K., Madoui, S., Khither, H., and Charef, N. (2019). Drug-likeness and pharmacokinetics of a bis-phenolic ligand: Evaluations by computational methods. *Sch. J. App Med. Sci.* 1, 167–173.

Niu, C., Yin, J., Zhang, J., Vederas, J. C., and James, M. N. (2008). Molecular docking identifies the binding of 3-chloropyridine moieties specifically to the s1 pocket of sars-cov mpro. *Bioorg. Med. Chem.* 16, 293–302. doi:10.1016/j.bmc.2007.09.034

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. doi:10.1186/1758-2946-3-33

Planas, D., Saunders, N., Maes, P., Guivel-Benhassine, F., Planchais, C., Buchrieser, J., et al. (2022). Considerable escape of sars-cov-2 omicron to antibody neutralization. *Nature* 602, 671–675. doi:10.1038/s41586-021-04389-z

Rahman, F., Tabrez, S., Ali, R., Alqahtani, A. S., Ahmed, M. Z., and Rub, A. (2021). Molecular docking analysis of rutin reveals possible inhibition of sars-cov-2 vital proteins. *J. Tradit. Complement. Med.* 11, 173–179. doi:10.1016/j.jtcme.2021.01.006

Rothan, H. A., and Byrareddy, S. N. (2020). The epidemiology and pathogenesis of coronavirus disease (Covid-19) outbreak. *J. Autoimmun.* 109, 102433. doi:10.1016/j.jaut.2020.102433

Serafin, M. B., Bottega, A., Foletto, V. S., da Rosa, T. F., Hörner, A., and Hörner, R. (2020). Drug repositioning is an alternative for the treatment of coronavirus Covid-19. *Int. J. Antimicrob. Agents* 55, 105969. doi:10.1016/j.ijantimicag.2020.105969

Shereen, M. A., Khan, S., Kazmi, A., Bashir, N., and Siddique, R. (2020). COVID-19 infection: Origin, transmission, and characteristics of human coronaviruses. *J. Adv. Res.* 24, 91–98. doi:10.1016/j.jare.2020.03.005

Singh, A., and Jana, N. K. (2017). Discovery of potential zika virus rna polymerase inhibitors by docking-based virtual screening. *Comput. Biol. Chem.* 71, 144–151. doi:10.1016/j.compbiolchem.2017.10.007

Tang, B., Bragazzi, N. L., Li, Q., Tang, S., Xiao, Y., and Wu, J. (2020a). An updated estimation of the risk of transmission of the novel coronavirus (2019-ncov). *Infect. Dis. Model.* 5, 248–255. doi:10.1016/j.idm.2020.02.001

Tang, X., Wu, C., Li, X., Song, Y., Yao, X., Wu, X., et al. (2020b). On the origin and continuing evolution of sars-cov-2. *Natl. Sci. Rev.* 7, 1012–1023. doi:10.1093/nsr/nwaa036

ul Qamar, M. T., Alqahtani, S. M., Alamri, M. A., and Chen, L.-L. (2020). Structural basis of sars-cov-2 3clpro and anti-Covid-19 drug discovery from medicinal plants. *J. Pharm. Anal.* 10, 313–319. doi:10.1016/j.jpha.2020.03.009

Wang, L., Bao, B.-B., Song, G.-Q., Chen, C., Zhang, X.-M., Lu, W., et al. (2017). Discovery of unsymmetrical aromatic disulfides as novel inhibitors of sars-cov main protease: Chemical synthesis, biological evaluation, molecular docking and 3d-qsar study. *Eur. J. Med. Chem.* 137, 450–461. doi:10.1016/j.ejmech.2017.05.045

Wang, L., Wang, Y., Ye, D., and Liu, Q. (2020). Review of the 2019 novel coronavirus (sars-cov-2) based on current evidence. *Int. J. Antimicrob. Agents* 55, 105948. doi:10.1016/j.ijantimicag.2020.105948

Wei, D.-Q., Zhang, R., Du, Q.-S., Gao, W.-N., Li, Y., Gao, H., et al. (2006). Anti-sars drug screening by molecular docking. *Amino Acids* 31, 73–80. doi:10.1007/s00726-006-0361-7

Wishart, D. S., Feunang, Y. D., Guo, A. C., Lo, E. J., Marcu, A., Grant, J. R., et al. (2018). Drugbank 5.0: A major update to the drugbank database for 2018. *Nucleic Acids Res.* 46, D1074–D1082. doi:10.1093/nar/gkx1037

Xiang, L., Xu, Y., Zhang, Y., Meng, X., and Wang, P. (2015). Virtual screening studies of Chinese medicine coptidis rhizoma as alpha7 nicotinic acetylcholine receptor agonists for treatment of alzheimer's disease. *J. Mol. Struct.* 1086, 207–215. doi:10.1016/j.molstruc.2015.01.021

Yang, H., Lou, C., Sun, L., Li, J., Cai, Y., Wang, Z., et al. (2019). Admetsar 2.0: web-service for prediction and optimization of chemical admet properties. *Bioinformatics* 35, 1067–1069. doi:10.1093/bioinformatics/bty707

Yang, H., Yang, M., Ding, Y., Liu, Y., Lou, Z., Zhou, Z., et al. (2003). The crystal structures of severe acute respiratory syndrome virus main protease and its complex with an inhibitor. *Proc. Natl. Acad. Sci. U. S. A.* 100, 13190–13195. doi:10.1073/pnas.1835675100

Zaki, A. A., Ashour, A., Elhady, S. S., Darwish, K. M., and Al-Karmalawy, A. A. (2022). Calendulaglycoside a showing potential activity against sars-cov-2 main protease: Molecular docking, molecular dynamics, and sar studies. *J. Tradit. Complement. Med.* 12, 16–34. doi:10.1016/j.jtcme.2021.05.001

Zhang, L., Lin, D., Sun, X., Curth, U., Drosten, C., Sauerhering, L., et al. (2020). Crystal structure of sars-cov-2 main protease provides a basis for design of improved α-ketoamide inhibitors. *Science* 368, 409–412. doi:10.1126/science.abb3405

Zhou, P., Yang, X.-L., Wang, X.-G., Hu, B., Zhang, L., Zhang, W., et al. (2020). A pneumonia outbreak associated with a new coronavirus of probable bat origin. *nature* 579, 270–273. doi:10.1038/s41586-020-2012-7

Frontiers in Pharmacology

Explores the interactions between chemicals and living beings

The most cited journal in its field, which advances access to pharmacological discoveries to prevent and treat human disease.

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