Briefings in Bioinformatics, 00(00), 2021, 1–15

doi: 10.1093/bib/bbab113
Review

Structure-based drug repurposing against COVID-19 and emerging infectious diseases: methods, resources and discoveries

Yosef Masoudi-Sobhanzadeh†, Aysan Salemi†, Mohammad M. Pourseif†, Behzad Jafari, Yadollah Omidi and Ali Masoudi-Nejad

Corresponding authors: Ali Masoudi-Nejad, Laboratory of Systems Biology and Bioinformatics (LBB), Institute of Biochemistry and Biophysics, University of Tehran, Tehran, Iran. E-mail: amasoudin@ut.ac.ir; Yadollah Omidi, College of Pharmacy, Nova Southeastern University, Florida, USA. E-mail: yomidi@nova.edu
†These authors contributed equally to this work.

Abstract

To attain promising pharmacotherapies, researchers have applied drug repurposing (DR) techniques to discover the candidate medicines to combat the coronavirus disease 2019 (COVID-19) outbreak. Although many DR approaches have been introduced for treating different diseases, only structure-based DR (SBDR) methods can be employed as the first therapeutic option against the COVID-19 pandemic because they rely on the rudimentary information about the diseases such as the sequence of the severe acute respiratory syndrome coronavirus 2 genome. Hence, to try out new treatments for the disease, the first attempts have been made based on the SBDR methods which seem to be among the proper choices for discovering the potential medications against the emerging and re-emerging infectious diseases. Given the importance of SBDR approaches, in the present review, well-known SBDR methods are summarized, and their merits are investigated. Then, the databases and software applications, utilized for repurposing the drugs against COVID-19, are introduced. Besides, the identified drugs are categorized based on their targets. Finally, a comparison is made between the SBDR approaches and other DR methods, and some possible future directions are proposed.

Key words: COVID-19; drug repurposing; emerging diseases; SARS-CoV-2; structure-based methods

Yosef Masoudi-Sobhanzadeh (PhD) is an assistant professor at the Tabriz University of Medical Sciences (TUOMS) Research Center for Pharmaceutical Nanotechnology (RCPN), Tabriz, Iran. His expertise is in the development of algorithms and data resources for computational drug repurposing.
Aysan Salemi (PharmD) is a research assistant at TUOMS RCPN, Tabriz, Iran. She is currently working on 'systems vaccinology for identification of new antigen(s) for designing a broad-spectrum epitope-based vaccine against *Staphylococcus aureus*'.
Mohammad M. Pourseif (PhD) is an assistant professor at the TUOMS RCPN, Tabriz, Iran. His main research field is vaccine design and development against infectious diseases and cancers through the bioinformatics and systems biology approaches and *in vivo*/*in vitro* experiments.
Behzad Jafari is a full academic in the Medicinal Chemistry Department at the School of Pharmacy, Urmia University of Medical Sciences. He has worked on phage display technology, protein engineering and molecular modeling.
Yadollah Omidi (PharmD, PhD) is a full professor at the Nova Southeastern University College of Pharmacy, Florida, USA. His main research activities include the development of multifunctional nanosystems against detrimental diseases. He has published over 250 papers in the field of pharmaceutical sciences (particularly in drug delivery and targeting) with over 7700 citations and an h-index of 48.
Ali Masoudi-Nejad is now an associate professor of systems biology and bioinformatics. His research interests are computational systems biology, cancer modeling and network biology and bioinformatics.
Submitted: 10 October 2020; **Received (in revised form):** 15 February 2021

INTRODUCTION

In the late 2019, the emergence of the animal-origin coronavirus, named severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), made the world face the critical challenges. The World Health Organization (WHO) named the newly virus-induced disease, which spread rapidly throughout the globe and infected humans at a rate of knots, the coronavirus disease 2019 (COVID-19) [1]. SARS-CoV-2 was shown to be an enveloped virus with different proteins on its surface, which entraps a single-stranded ribonucleic acid (ssRNA) with its about 30 000 base pairs [2]. The long viral genome of the virus is packed inside a helical capsid formed by the nucleocapsid proteins. The SARS-CoV-2 subgenomic RNAs encodes four types of structural proteins, including envelope (E), nucleocapsid (N), spike (S) and membrane (M) proteins. Besides the genes which encode the structural proteins, there exist some specific genomic regions that encode the viral proteins required for its replication. The functions of the SARS-CoV-2's proteins and those of the host cell, interacting with them, are explained in Supplementary Tables S1–S3 (see Supplementary Data available online at http://bib.oxfordjournals.org/). A comprehensive scheme of the virus's life cycle is also presented in Figure 1.

The existing similarities between the SARS-CoV-2 and some other viruses, including SARS-CoV, MERS-CoV and human immunodeficiency viruses (HIV), help understand the SARS-CoV-2's mechanism of action and help develop the potential therapeutic and prophylactic approaches [3]. Herein, to accelerate discovering anti-COVID-19 candidate therapeutics, various techniques have been introduced, such as repurposing the antivirals and anti-inflammatory drugs, detecting antibodies, immune modulators (e.g. adjuvants) and vaccine-based targeted delivery systems. From among the introduced approaches, drug repurposing (DR) methods, which seek the hidden benefits of the existing drugs in treating different diseases, have produced promising outcomes. Given the capabilities of the computational approaches in reducing the spent time and cost of the drug discovery projects [4], many studies have been conducted to identify the candidate medicines for treating COVID-19. In this review, based on the utilized computational DR methods, the studies (at the time of conducting this review) have been divided into four classes, including (i) machine learning (ML), (ii) network-based (NB), (iii) structure-based DR (SBDR) and (iv) hybrid approaches.

Considering the fact that most of the computational DR methods rely on a sufficient volume of data, the SBDR approaches can handle the issue properly because of their reliance on the rudimentary information. For example, researchers compare the first genome of SARS-CoV-2 with the other ones and model the 3D structure of its proteins based on the knowledge obtained from some previously introduced techniques [5, 6]. Then, after applying the screening techniques to discover candidate medicines against COVID-19, the chemical/natural compounds are scored using the docking methods in a short time. Therefore, the SBDR approaches might prove to be an early and suitable drug discovery plan to cure not only COVID-19 but also the other emerging diseases. To perform the SBDR techniques, currently, there exist multiple docking software tools, which have been developed based on the different algorithms, which can be categorized based on their operational nature into several main groups as follows:

(i) Heuristic and metaheuristic-based algorithms: Detecting the conformational shape of a ligand and a protein, in which the components can bind together properly, is a nondeterministic polynomial (NP)-hard problem. To deal with such the problem, some heuristic (e.g. fast Fourier transform algorithm for protein–protein docking [7]) and metaheuristic algorithms [e.g. genetic algorithm (GA) for ligand–protein docking [8]] have been proposed. Overall, metaheuristic algorithms can yield better outcomes compared to the heuristic ones, and the probability that they fall into the local optima is lower than the heuristic algorithms [9, 10].

(ii) Grid-based and blind docking algorithms: Based on the prior knowledge about the possible binding sites of a protein, biological experts specify a box which confines the search space of an algorithm. However, in case such a knowledge does not exist, an exhaustive search strategy, named blind docking, must be carried out [11]. Although the existing toolboxes usually support two types of the grid-based and blind docking techniques to enhance the accuracy of the outcomes, it is essential to utilize efficient algorithms when the search space of the problem is large [12].

(iii) Flexible and rigid structure-based algorithms: Since the flexible docking algorithms consider the atoms' 3D geometry properties such as torsion of hyperplanes, the rigid docking algorithms ignore them. As a result, the flexible docking algorithms produce outcomes which are most likely to be observed in nature [13]. However, the molecular dynamics (MD) simulations can investigate the limitation of the rigid docking algorithms and validate or deny their outcomes [14]. This class of techniques can also be categorized based on the energy and shape complementary-based scoring algorithms, which are related to the flexible and rigid docking methods, respectively.

The main contributions of this review are listed as follows:

(i) To investigate the early SBDR-based methods against COVID-19.
(ii) To categorize the detected drugs based on their targets.
(iii) To discuss the capabilities of the SBDR methods in finding the candidate medicines to tackle not only COVID-19 but also other emerging infectious diseases.
(vi) To introduce and describe bioinformatics software suites and databases used in the SBDR approaches.

A CLASSIFICATION OF THE COMPUTATIONAL-BASED DR METHODS AGAINST COVID-19

Up to now, various computational-based DR approaches have been proposed for controlling different diseases [15]. In this section, DR methods, introduced for managing COVID-19, are categorized into some groups based on their computational techniques:

(i) ML-based methods: ML approaches are usually involved in predicting the drug–target (e.g. drug–protein, drug–pathway and drug–gene) interactions (DTIs) [16]. To this end, based on the experimentally validated DTIs, a model is generated to predict whether a drug and a target of interest have a significant interaction or not [17]. The obtained model is then applied to detect the potential medications that can inhibit the main targets of SARS-CoV-2, including S protein, M^{pro} and RNA-dependent RNA polymerase (RdRp). The ML techniques, which are usually used for screening the drugs against COVID-19, produce more accurate predictions than

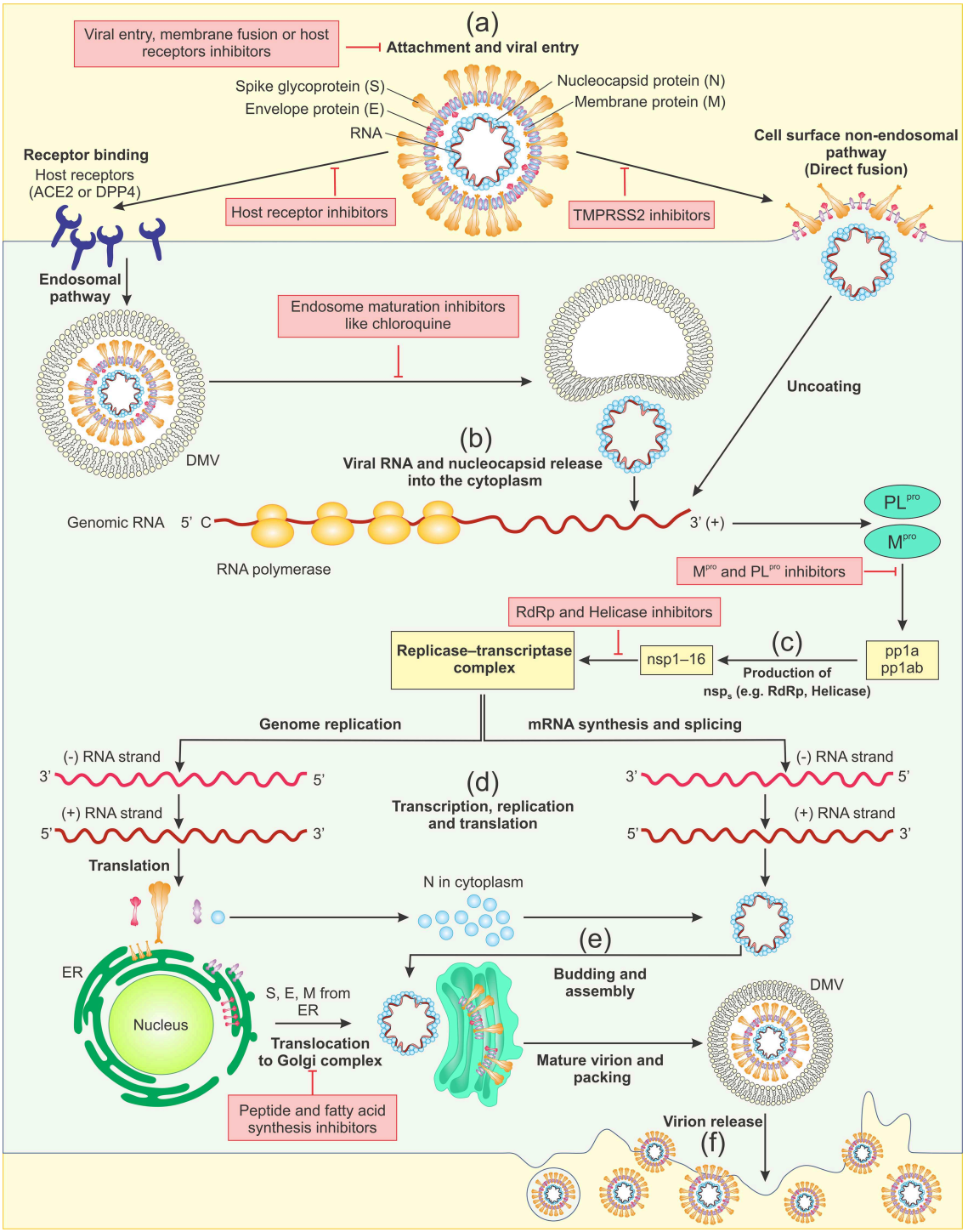

Figure 1. The schematic presentation of SARS-CoV-2's life cycle. This study has divided the life cycle of the virus into six main processes, consisting of (**A**) viral entry, (**B**) viral RNA and nucleocapsid release into the cytoplasm, (**C**) production of non-structural proteins, (**D**) replication, transcription and translation, (**E**) viral assembly and (**F**) virion release. DMV, double membrane vesicle; DPP4, dipeptidyl peptidase 4; mRNA, messenger RNA; E, envelope; ER, endoplasmic reticulum; M, membrane; N, nucleocapsid; nsps, non-structural proteins; pp1a, polyprotein 1a; pp1ab, polyprotein 1ab; S, spike.

the other technical methods [18]. Nevertheless, compared to other approaches, they are more dependent on the amount of the collected data. The predicted results also require to be validated by the secondary approaches such as *in vivo, in vitro,* and/or *in-silico* experiments [19].

(ii) NB methods: Considering the connections between different biological elements, such as genes, proteins and microRNAs, it seems that the NB approaches are the most promising techniques for managing diseases [20]. To combat COVID-19, the existing studies determined the protein–protein interactions (PPIs) between the proteins of SARS-CoV-2 and those of the host cells [21], and then, utilized various graph theory algorithms to analyze the network. For example, in a study, after the key nodes of the

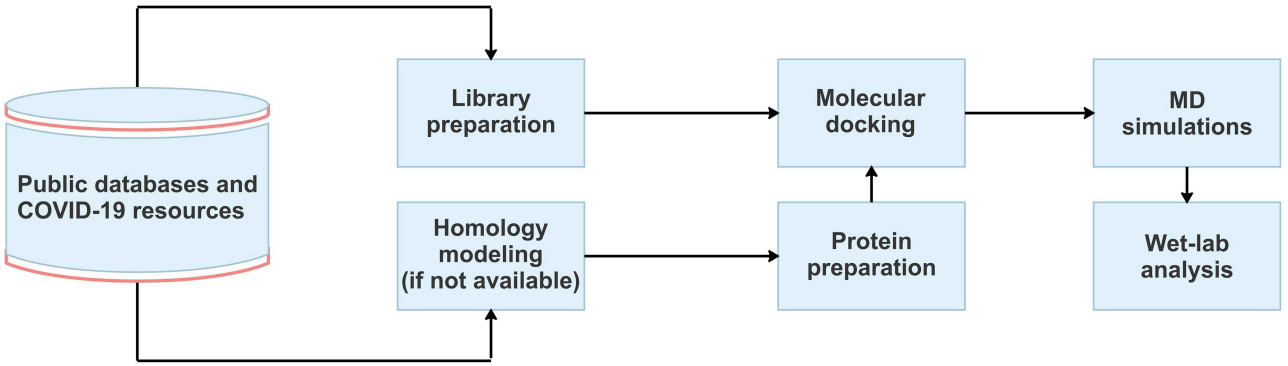

Figure 2. The general framework of the employed SBDR techniques in the different studies. These methods obtain the data of interest from the public databases and process them before the docking experiment. For the proteins which their 3D structure is not specified, homology modeling approaches are utilized. Also, to validate the outcomes further (the predicted DTIs), MD simulations and wet-lab analysis are carried out.

generated network were specified, some potential inhibitor drugs were introduced [22]. Complementary exposure components patterns were also used to find two drugs, A and B, which can be more effective than a single therapy in curing the disease [23]. Although the NB methods may yield better treatment plans, many aspects of an emerged disease are unknown [24]. Therefore, these approaches may not be applicable in the early stages of an emerging disease such as COVID-19.

(iii) SBDR methods: The first attempts to search for the candidate drugs against COVID-19 were made based on the 3D structures of SARS-CoV-2 proteins specified using *in vivo* or *in vitro* experiments. Such useful information along with a huge volume of various biological data resulted in the discovery of candidate medications against the COVID-19 infection [25]. The SBDR methods usually follow a similar framework to repurpose the drugs. For this purpose, first, drugs information is extracted from the common drug databases such as DrugBank [26], KEGG [27], DrugR+ [28, 29], TTD [30] and others. Then, after the obtained drugs information is screened against the targets of interest [e.g. S protein, angiotensin converting enzyme 2 (ACE2), M^{pro} and RdRp], the results are ranked based on the connection stability criterion (the docking score). Next, some further experiments such as MD simulation are performed to examine the behavior of the connection. The present review investigates SBDR methods which have been introduced for managing the COVID-19 outbreak, which have the ability to find the candidate medications in a short time and are capable of yielding a practical outcome. Hence, the SBDR might be a suitable approach in discovering the potential drugs for not only controlling the COVID-19 outbreak but also for treating an emerged disease.

(iv) Hybrid methods: To acquire accurate and practical results, some studies have combined the abovementioned methods in different manners and utilized their advantages. For instance, in a study, to discover the probable interactions between the existing medications and SARS-CoV-2 proteins, a model was generated to predict drug–protein interactions using the ML methods. Then, after identifying the candidate drugs, a molecular docking analysis was carried out to validate the predicted outcomes. To discover the potential anti-SARS-CoV-2 medicines, another study combined a number of bioinformatics tools that have utilized the concepts of PPI networks and molecular dockings [31]. The results of the mentioned studies showed that the Cold-Damp Plague Formula is effective in treating the COVID-19 infection. Hybrid methods, like the other DR methods, may face some limitations and reduce the accuracy of the predictions due to some flaws in bioinformatics suites and methods (as a synergic anomaly).

Some other categories such as the text-mining methods can be considered besides the above classes. At the time of conducting this review, the aforementioned methods were available as preprints. Hence, these methods along with some other novel ones are introduced and explained as future directions in the Discussion section.

A SUMMARY OF THE SBDR APPROACHES AGAINST COVID-19: RESOURCES, METHODS AND DISCOVERIES

This section examines the ways employed by the SBDR methods in combating COVID-19. The existing researches have followed a similar framework which is presented in Figure 2.

Figure 2, which can be applied to not only COVID-19 but also to other emerging diseases, consists of several main steps which are listed as follows:

(i) Collecting data: Advances in bioinformatics have yielded helpful software packages and databases, which are playing an essential role in combating COVID-19. In different studies, the information about ligand and SARS-CoV-2's proteins (such as the 3D structure of the proteins, amino acid sequences and chemical properties of drugs) is extracted from the common databases and those which have been developed especially for COVID-19.

(ii) Preparing ligands: A number of public databases (such as ZINC and PubChem) curated the existing information on the chemical compounds and facilitate the drug screening process. In case there are no data available in the data resources, various preprocessing phases (e.g. adding hydrogen atoms and removing unnecessary molecules) will be followed to optimize the structure of a ligand and make it suitable for the docking operations.

(iii) Modeling the 3D structure of the SARS-CoV-2 proteins (homology modeling): Homology modeling techniques are utilized when the 3D structure of a desired protein does not exist. In that case, the amino acid sequence of the protein is compared with that of the proteins whose 3D structure is available, and then, the protein's 3D structure is simulated.

For example, in the early stages of the pandemic, the 3D geometry information of the RdRp was simulated using the homology modeling methods.

(iv) Preparing the protein: After acquiring the 3D structure of the protein, its structure is processed and prepared for docking operations. Although this step is similar to the second one, it usually contains a series of actions (i.e. iteration steps) compared to the second step, such as generating low energy conformations and neutralizing the charged groups.

(v) Performing the docking algorithms: Various types of software packages, which commonly utilize the heuristic algorithms, have been developed to investigate the connectivity ability of the biological elements like protein–protein, ligand–protein and ligand–ligand. Since these software tools usually produce practical outcomes, they are used in many drug discovery projects to disrupt the SARS-CoV-2's life cycle.

(vi) Applying the MD simulation techniques: The flexible-based docking algorithms consider various criteria (such as the length of bonds, the type of bonds, torsions of atoms, etc.) related to the binding sites of a ligand and a protein. In contrast, the rigid-based docking algorithms do not take the mentioned criteria into account and usually require applying the secondary validation techniques like MD simulations. These simulation techniques analyze the physical movement of atoms and molecules and calculate the potential energy of a connection based on the force field's different concepts. From a difference perspective, the MD simulation methods may also be helpful after applying flexible-based docking algorithms due to the fact that unlike the flexible docking methods, they take the effects of temperature, solvation, ion bonds, etc. into consideration.

(vii) Validating using wet-lab experiments: Since the SBDR approaches usually produce the results which are most likely to be observed in a biological system, they have been converted to a popular technique in the drug discovery process against COVID-19, which examines whether there is a probabilistic interaction between ligands and SARS-CoV-2's proteins or not. However, insomuch the mentioned interaction may be affected by some parameters and elements of a biological system, the predicted drug-SARS-CoV-2 protein complex must be investigated further using the *in vivo* or *in vitro* experiment, called the wet-lab trials.

The approaches, which screen the chemical and natural compounds, target the key proteins or enzymes [e.g. RdRp, M^{pro}, PL^{pro}, NSP15, helicase, ACE2, transmembrane protease serine protease 2 (TMPRSS2), M, N, E and S proteins]. The targets are categorized into three classes, including the structural and non-structural proteins of SARS-CoV-2 and the host cells's proteins that interact with the proteins of the virus.

In the following subsections, the studies have been classified based on the SARS-CoV-2's proteins. A summary, which consists of the main aims of the work, employed databases, utilized methods and obtained results, has been also provided for every study.

RdRp-related studies

To investigate the effects of the antiviral medications against COVID-19, Elfiky carried out a similarity-based method which consists of three main steps. First, the available sequence of the SARS-CoV-2 genome was acquired from the NCBI nucleotide database. Second, the SWISS-MODEL web server was employed to model the structure of RdRp. For this purpose, the structure of SARS HCoV was obtained from the protein data bank (PDB) and was used as the simulation template because it bears a striking (about 97.8%) similarity to SRAS-CoV-2's RdRp. Besides, the outcomes of the multiple sequence alignment algorithm, obtained by the Clustal omega web server, showed that 90% of SARS' residues are identical to those of the virus. Molprobity and SAVES web servers, which comprise a variety of popular and efficient software tools, such as Verify 3D, PROCHECK, PROVE and ERRAT, were utilized to validate the model. Third, the SCIGRESS software suite was used for minimizing the generated model and for performing the molecular docking experiments. In this step, some compounds, including the FDA-approved and clinical trials drug-like compounds, were screened and examined in inhibiting SARS-CoV-2's RdRp. The results suggested that sofosbuvir, IDX-184, ribavirin and remdesivir may manage the disease better than the other chemical compounds. They also indicated that cinnamaldehyde and thymoquinone play a negative role in acting against SARS-CoV-2's RdRp [32].

NSP15-related studies

Up to now, different researchers have examined SARS-CoV-2 from various aspects and have revealed its many unknown properties. Recently, it has been reported that the NSP15 (an endoribonuclease enzyme) plays a key role in replicating the virus. To identify the potential antiviral compounds against the replication of SARS-CoV-2, Krishnan *et al*. carried out a SBDR method that consists of three main phases. In the first step, the 3D structure of NSP15 was downloaded from PDB and optimized using the Schrodinger software application. Besides, the amino acid sequence of the enzyme was compared with the other amino acid sequences using the ClustalW multiple sequence aligner algorithm. In the second step, first, 3978 compounds, having antiviral properties, were retrieved from the Enamine database. Then, the geometry and energy of the obtained compounds were optimized using the Ligprep module of Schrodinger. In the third step, the molecular docking experiments were carried out based on the standard precision (SP), extra precision (XP) and induced fit docking (IFD). For this purpose, the docking and glide energy pairs for SP, XP and IFD were set as (−5, −40), (−5.5, −40.5) and (−6, −50), respectively. As a result, eight compounds (including Z595015370, Z2228348553, Z1302426228, Z1343129850, Z16215674, Z2760938911, Z2239044677 and Z56786758) were proposed for inhibiting NSP15, which forms a significant binding energy with the enzyme [33].

M^{pro}-related studies

M^{pro}/$3CL^{pro}$ of SARS-CoV has a critical role in its life cycle. Since the genomes of SARS-CoV and SARS-CoV-2 are similar, COVID-19 may be controlled by inhibiting the M^{pro} of SARS-CoV-2. To validate the hypothesis, Qamar *et al*. conducted a computational experiment in which they examined how much plant-produced chemical compounds may bind to the enzyme and inhibit it. The authors obtained the genome of SARS-CoV-2 from the GISAID database, and then, extracted the sequence of $3CL^{pro}$. This sequence was later converted to a protein sequence using the ExPASy software suite. To determine the conserved residues and similar sequences, the multiple-sequence alignment algorithm was performed using the T-Coffee web server, and to apply the homology modeling of the enzyme, the PSI-Basic Local Alignment Search Tool (BLAST) algorithm was utilized to

calculate the similarity score between the $3CL^{pro}$ and the well-known proteins obtained from the PDB which in turn resulted in specifying the similar sequences. Then, the 3D structure of the enzyme was simulated, and its quality was checked using the Chimera and PyMOL software applications. In the next step, 32 297 phytochemical and traditional Chines medicine compounds were gathered from different repositories and screened against $3CL^{pro}$ using the molecular operating environment (MOE) software package. In the final step, to investigate the binding behavior and stability of the connections, the molecular dynamic simulation was carried out using the GROMACS software package. The findings of the above procedure included the nine plant-produced chemical compounds such as isoflavone, myricitrin and methyl rosmarinate. Based on the docking scores, it can be concluded that the mentioned herbal compounds might prove to be more effective than the other chemical compounds such as nelfinavir, prulifloxacin and colistin [34].

As described before, in the early stages of the COVID-19 appearance, the 3D structures of the SARS-CoV-2's proteins were not available. Nevertheless, the different wet-lab studies were carried out and some unknown information (such as the 3D structure of M^{pro}) was provided. In a study, Knadeel and Al-Nazawi screened the FDA-approved medicines against the detected 3D structure of SARS-CoV-2's M^{pro}. The authors extracted the compounds from the SelleckChem data resource and imported them into the Ligprep module of the Schrodinger software suite to simulate and optimize their structures. Moreover, the energies of the structures were minimized using the OPLS2005 force field and prepared for docking using the glide module of the Schrodinger software application. To acquire more accurate results, the outcomes were ranked based on the SP score, and then, 20 drugs were introduced as candidates for curing COVID-19. The results included some medicines from different drug groups such as antiviral, anti-cancer and vitamin classes. The authors also made a comparison between the sequences of SARS-CoV-2, MERS and SARS and showed that the sequence of SARS-CoV-2 is more similar to that of SARS than the sequence of MERS. For inhibiting M^{pro} of SARS, curcumin was developed and proved to have powerful clinical effects on patients suffering from SARS. However, Knadeel and Al-Nazawi indicated that their predicted compounds form better hydrogen bond interactions with the M^{pro} than curcumin [35].

In another SBDR-based research, Shamsi *et al*. screened the FDA-approved candidates against the SARS-CoV-2's M^{pro} to treat any possible indications of COVID-19. In this regard, several crystal structures of SARS-CoV-2 M^{pro} were analyzed after being downloaded from the PDB database. As a result, one of them was chosen (PDB ID: 6M03) for virtual screening, and its energy was then minimized. Next, using AutoDock Vina, a blind docking was performed based on the 2388 FDA-approved drugs, and candidate medicines with the highest binding scores were selected. The interactions mode of the selected compounds was analyzed, and those, which specifically interact with the active-site residues of the protein, were filtered. The outcomes indicated that glecaprevir and maraviroc (MVC) show the same pattern of binding with the SARS-CoV-2's M^{pro}, where the co-crystalized inhibitor was observed. The identified structures were shown to have antiviral activity that was used in the treatment of chronic hepatitis C virus (HCV) (glecaprevir) and HIV infection (MVC) [36].

Mittal *et al*. also targeted the SARS-CoV-2's M^{pro} to identify potential molecules with the best binding free energy based on the interaction behaviors (MD simulation). In the mentioned study, the crystal structure of M^{pro} in APO and HOLO forms (as a complex with covalent inhibitors, N3 and 13b) were obtained from the PDB database, and a library of FDA/repurposed ligand structures was collected from the DrugBank, SelleckChem and Repurposing hub databases. To screen the pool of the ligand structures, a three-step virtual screening approach was proposed, as follows: (i) high-throughput screening, (ii) SP and (iii) XP. The final potential compounds were selected based on the estimation of the relative binding free energies using the MM-GBSA (the molecular mechanics energies combined with the generalized Born and surface area continuum solvation) method. Finally, six potential compounds (leupeptin hemisulfate, pepstatin A, nelfinavir, birinapant, lypression and octreotide) with the best ligand-binding affinity (ΔG_{bind}) were identified [37].

Besides the described studies, Lorane *et al*. used a combined set of software packages to computationally screen and propose the most potent chemical inhibitor(s) against the M^{pro}. The researchers utilized the two compounds datasets for the virtual screening of medications against M^{pro}. One of these libraries contained 1017 previously targeted ligands (named SARS-CoV-2 target library) and the other one included 1577 compounds, which were predicted based on the ML methods such as artificial neural networks and Bayesian statistic techniques. Based on the outcomes, two lists of 100 hit compounds with the highest binding affinity score were obtained for M^{pro}. Besides, the ADME/Toxicity analysis was performed to predict the pharmacokinetic and toxicity properties of the chemical compounds by comparing them with the currently tested anti-COVID-19 drugs. Regarding the ADMET analysis, 10 molecules were selected as follows: (i) m494 from the SARS-CoV-ML dataset and (ii) nine molecules from the SARS-CoV-2-Target dataset (m57, m74, m113, m135, m152, m351, m603, m808 and m824). Next, the intermolecular interactions between the drug candidates and M^{pro} were calculated using a GA-based docking approach. The m57, m113, m135 and m808 molecules showed the highest number of molecular interactions with M^{pro} residues. Therefore, they were selected for the structural similarity analysis using the BindingDB web server. The results indicated that apixaban, a known anticoagulant drug, has the highest similarity score (42%) and might be a promising drug for curing patients infected with SARS-CoV-2 [38].

Helicase-/nsp13-related studies

Helicase or nsp13 is an enzyme that plays an essential role in the replication of SARS-CoV-2. The alignment algorithms showed that the SARS-CoV-2's helicase, which shares about 99.8% sequence similarity with the SARS-CoV's nsp13, has remained highly conserved. Therefore, it can be concluded that the enzyme is critical to the virus's life cycle. However, various SBDR studies have ignored its effect probably because of the unavailability of the enzyme's 3D structure. To address the issue, Mark *et al*. performed a SBDR experiment and screened about 970 000 chemical compounds (obtained from the ZINC database and Enamine library) against apo- and ATP/RNA-bound conformations of SARS-CoV-2's helicase. To generate the 3D structure of the enzyme, the crystal structures of SARS/MERS's nsp13 were considered as the simulation templates, and its 3D structure was then optimized using the public software packages such as NAMD. In the next step, Mark *et al*. applied the rigid algorithm-based docking software suites to find the potential drugs which may interact with the conformational shapes of the enzyme, and then, the researchers validated their

final outcomes using the MD simulations methods. According to the results, lumacaftor and cepharanthine showed a higher score with the conformational states of the enzyme, which led the researchers to conclude that they might prove to be effective in preventing the process of replicating the virus [39].

Multi-target-related studies

The conducted studies, which apply the SBDR methods to the complex formed by the interaction between the SARS-CoV-2 and the host cell's proteins such as the S protein and ACE2, are summarized in this section.

According to Indian ethnomedicines, some plants [nilavembu kudineer chooranam (NKC), seenthilkodi (*Tinospora cordifolia*) and sathakuppai (*Anethum sowa*)] are generally utilized to treat the cold- and respiratory-related disorders. Therefore, the active ingredients of the mentioned medicinal plants, whose antiviral properties have previously been reported, may be useful in curing the disease. Alagu *et al*. explored the therapeutic efficacy of 10 Indian origin medicinal plants for curing COVID-19. The researchers selected 47 active components from the mentioned medicinal plants and screened them against the S and M^{pro} of SARS-CoV-2 as well as the ACE2 enzyme of the host cell. The 3D structures of the bioactive ligands were extracted from the PubChem database, and then, their formats (.sdf) were converted into the pdf-format file using the online SMILES translator program. Additionally, the crystal structures of S protein, M^{pro} and ACE2 were retrieved from the PDB database. To optimize the 3D structures, they were refined using the ModRefiner algorithm, and their hydrogen atoms were determined using the MMFF94 force field in the Autodock suite. A molecular docking test was also performed using Autodock, and the binding affinities were determined based on kcal/mol. Next, the top 10 bioactive ligands (cucurbitacin E, orientin, bis-andrographolide, cucurbitacin B, isocucurbitacin B, vitexin, berberine, bryonolic acid, piperine and magnoflorine) were identified based on the calculated scores. It was shown that cucurbitacin E and orientin, from among the predicted outcomes, have the strongest interaction with the most critical residues of the targets. Hence, the MD simulation was performed to further investigate the interactions using the GROMACS 5.1.4 software package. The MD simulation outcomes demonstrated that the interactions between the two bioactive compounds and targets are stable, and therefore, they may be a proper option for treating COVID-19 [40].

Recently, different studies have identified the various properties of SARS-CoV-2's proteins and those which existed in and interacted with the host cells. Based on the produced data, in a study, the reported targets were systematically analyzed and compared with the other coronaviruses' proteins. As a result, 21 targets were identified, including Nsp1, Nsp3b, Nsp3c, PLpro, Nsp3e, 3CLpro, Nsp7_Nsp8 complex, Nsp9, Nsp10, Nsp12/RdRp, Nsp13/helicase, Nsp14, Nsp15, Nsp16, ORF7a, Spike, NNRBD, NCRBD, E-channel, ACE2 and TMPRSS2 were identified. To specify the 3D structures of the targets, the authors employed the homology modeling software tools, such as the TMHMM online server, that act based on the information existing in the nucleotide sequence of similar targets. These data were collected from the NCBI nucleotide database, and their relationships with the other proteins were determined using the ClustalW multiple alignment algorithm. In the next step, after the authors extracted a compound library from the ZINC drug database as well as a dataset of the natural products, they screened them against the 21 targets. After applying the homology modeling techniques, a molecular docking experiment was done, and some candidate medicines were proposed for treating COVID-19. It was found that the outcomes consist of some drugs from different groups such as antiviral, antibacterial and anti-asthmatic classes [41].

Structural proteins-related studies

Since many pieces of researches have focused on the non-structural proteins of SARS-CoV-2, some other studies have applied the SBDR approaches to inhibit the structural proteins of the virus and investigated their roles in preventing COVID-19. In a study, the structural proteins of SARS-CoV-2 (E, M and N proteins) were targeted, and the effect of the 548 natural and synthetic compounds, having antiviral properties, were examined using the docking experiments. For this purpose, the sequence of the proteins and the 3D structure of N protein were retrieved from the GenBank and PDB databases, and the chemical information of the compounds was extracted from the PubChem and SelleckChem databases. A homology modeling web tool, named I-TASSER, was employed for both M and E proteins whose 3D structure did not exist. Then, the common software suites such as YASARA, CASTp, Molsoft and Autodock vina were utilized for minimizing the energy, predicting the active site residues, determining the pharmacokinetics traits and calculating the binding affinity score, respectively. The results indicated that some combinations such as doxycycline and rutin, caffeic and ferulic acids and simeprevir and grazoprevir may inhibit the E, M and N proteins, respectively [42].

Besides the research studies, some review works investigated the usefulness of the repurposed drugs, which were previously obtained for the other diseases using the SBDR methods, against COVID-19. For example, a manuscript summarized the beneficial effects of nitazoxianide (NTZ), an FDA-approved antiparasitic agent, on patients infected with COVID-19. It was shown that NTZ has a broad antiviral activity in different viral infections such as coronavirus, influenza, hepatitis B and C viruses (HBV and HCV) and other viruses. Moreover, it revealed that NTZ has a bronchodilation activity in the contracted airways presenting the medication as a promising candidate to conduct the clinical trials. Nitazoxianide was approved as an antiparasitic therapy for diarrhea and enteritis caused by *Cryptosporidium* spp. and *Giardia lamblia* in the USA in 2002. The antiviral effects of NTZ were discovered serendipitously while treating *Cryptosporidium* diarrhea in patients suffering from HIV and HBV or HCV. The article also reported on the *in vitro* studies conducted on the application of NTZ and tizoxanide (TIZ, the active metabolite of NTZ) to combat different viruses, including various coronaviruses such as canine coronavirus S-378, bovine coronavirus (L9), murine coronavirus, human enteric coronavirus (4408) and SARS-CoV-2 [43].

RESOURCES

Computational-based DR methods strongly depend on the existing data resources whose correctness and volume play a critical role in discovering the useful medications for curing COVID-19. Although most SBDR approaches utilize the software packages, other DR techniques are usually created based on the novel algorithms. Given the importance of data and software resources, this section summarizes the resources used in the

Table 1. The public data resources employed in the SBDR methods against COVID-19.

Database	Description	Ref
Asinex	Provides screening libraries of lead-like molecules, macrocycles and fragments as well as research reagents and building blocks	–
BindingDB	A public, web-accessible database of measured binding affinities, focusing mainly on the interactions of protein	[44]
ChEMBL	A manually curated database of bioactive molecules with drug-like properties. It brings together chemical, bioactivity and genomic data to aid the translation of genomic information into effective new drugs	[45]
DrugBank	A comprehensive database which includes different drug information such as the affected targets as well as their sequences	[46]
Enamine	Provides suitable services for searching and offering different types of small molecules	[47]
NCBI	A comprehensive set of sequences collected from different databases such as GenBank and PDB	[48]
PDB	Various information about the 3D structure of proteins, nucleic acids and other biological complexes	[49]
PubChem	A part of NCBI which includes a large collection of freely accessible chemical information such as molecular formula, structures, chemical and physical properties, biological activities and safety and toxicity information	[50]
Repurposing hub	A curated and annotated collection of FDA-approved and clinical trial drugs along with their new benefits	[51]
Selleckchem	Consisting of curated data related to compound libraries and signaling inhibitors	–
UniProt	A comprehensive, high-quality and freely accessible resource of protein sequences and functional information	[52]
ZINC15	A free database of commercially available compounds for virtual screening over 230 million purchasable compounds with ready-to-dock 3D formats	[53]

*Press CTRL and click on the database name for redirecting to the relevant page.

SBDR approaches against COVID-19 and categorizes them based on their applications.

Table 1 lists the databases employed in the SBDR methods and describes their data contents. It also provides an access link to the databases along with their references. Every database contains different information, and PDB, DrugBank and NCBI from the databases are the most frequently used databases in the DR methods against COVID-19.

Table 2 lists and explains the resource especially developed to maintain the data corresponding with SARS-CoV-2 and viral disease (COVID-19).

After collecting the data of interest, they are usually analyzed using the public software applications to find their relationships with the other data. Since in the early stages of the advent of SARS-CoV-2, many aspects of its key proteins (e.g. the 3D structure of the proteins) were unknown, the homology modeling techniques were applied to determine these aspects. Table 3 lists and describes the utilized software packages employed in the SBDR methods for homology modeling and structural analyses.

Screening the drugs against SARS-CoV-2 and proposing the candidate medicines are the two main steps of the SBDR approaches. Several popular drug screening, docking software tools and MD simulations applications are listed in Table 4. Considering the total number of usages, it seems that Autodock, GROMACS and Schrödinger suites are more practical than the other packages.

A CLASSIFICATION OF THE DETECTED DTIS

As shown in the Methods section, the SBDR approaches focus on the 11 targets having different functionality. The targets are divided into three groups, including structural, non-structural and host cells' proteins, which interact with the SARS-CoV-2's proteins. The medicines which have been suggested for inhibiting the related targets are shown in this section. In other words, the repurposed drugs which may be helpful in controlling COVID-19 are listed based on the mentioned categories. The discovered medications fit into the various classes of drugs such as antiviral, anti-cancer and anti-hepatitis. Figure 3 depicts some popular proposed drugs against COVID-19 as well as their related targets. Comprehensive lists of DTIs, categorized based on the target types, are also available in Supplementary Tables S4–S6 (see Supplementary Data available online at http://bib.oxfordjournals.org/).

DISCUSSION AND FUTURE DIRECTIONS

Due to an urgent need to discover and develop cost-effective treatments with high efficacy and low side effects against COVID-19, researchers have focused on the DR methods. These approaches can not only reduce the spent time and cost of drug discovery projects but also find the candidate medicines for curing the disease. From among the reported outcomes, two repurposed drugs, remdesivir and favipiravir, suggested by the SBDR methods, have shown powerful clinical effects on treating the diseases. Besides, several studies detected antiviral medications such as ribavirin which may be useful in recovering the patients' health and mitigating their pains. Although a number of DR approaches had applicability to repurpose the drugs against COVID-19, most of them were not applicable in the preliminary stages of its advent because of the unavailability of sufficient data and knowledge about the virus. In contrast, the SBDR methods were capable of being applied in the early stages of the COVID-19's advent due to their reliance on the basic data of the virus, such as its genome sequence. Hence, SBDR techniques were quickly expanded, resulting in some rapid solutions for emerging diseases such as COVID-19. For the unknown protein structures of the virus, the SBDR approaches utilize the simulation software applications and predict their 3D structures based on the information existing in the sequences.

Table 2. The data resources developed or extended to maintain the information related to the SARS-CoV-2.

Resource	Data type	Description	Ref
CORD-19	Literature	A free resource incorporating more than 280 000 articles. The resource has provided an opportunity to develop artificial intelligence-based tools such as the Covidex	[54]
CoV3D	3D structure of the virus proteins	A weekly updated database incorporating the 3D structure of SARS-CoV-2's proteins and their complexes with other structures	[55]
COVID-19 data hub	Statistical data, policy measures, geographic information and external identifiers	An integrated data resource which has been formed by a systematic programming method based on the R programming language	[56]
COVID-evidence	Reports and literature	A frequently updated database holding trial pieces of evidence, which discusses their advantages and disadvantages	–
COVID-KOP	Drugs, phenotypes and pathways	A deductive system that integrates the ROBOKOP graph database with various data related to COVID-19	[57]
COVIEdb	Epitopes of the virus	Including the predicted B-cell and T-cell epitopes for different types of coronaviruses, which can accelerate the drug discovery process for combating COVID-19 and design an effective vaccine against it	[58]
DBCOVP	3D structure of coronavirus glycoproteins	A curated web resource containing different structural properties of the virus, helpful information to better understand the immune response, and the predicted T-cell and B-cell epitopes of the proteins	[59]
DockCoV2	Drug	A database containing binding affinity between the FDA approved and Taiwan NHI drugs and seven proteins of SARS-CoV-2 (the protein targets have been introduced in this review)	[60]
DrugRepV	The repurposed drugs	Collecting data correspond with the medicines which have been repurposed against 23 pandemic and epidemic viruses	[61]
GDSL	Gene, drug and molecular mechanism	A library including gene and drug sets which may have a role in curing COVID-19	[62]
GESS	Results of analyzing the genome of SARS-CoV-2	A web-based resource that enables the users to search, obtain and analyze the genome of the virus based on the geographical zones	[63]
GISAID	Genetic sequences and epidemiological data	Providing a rapid manner to share data related to human and animal viruses	[64]
HIT-COVID	Interventions	A database prepared by more than 200 volunteers and consisted of various information corresponding to the virus and its caused disease	[65]
HVIDB	Human-virus PPI	A database covering 35 types of virus families and their interactions with the human proteins	[66]
LitCovid	Literature	A frequently updated database which has been developed based on the PubMed data	[67]
Ma'ayan laboratory	Drugs and gene sets	Collecting predicted drugs, which may affect the virus, as well as gene sets from various resources and comparing the *in silico* predicted outcomes with wet-lab results	[68]
Nextstarin	Various mined information on the virus	By analyzing different data gathered from all around the world, this web platform, consisting of evolution information on the virus, has been organized	–
OTAVAchemicals	The virus proteins	Preparing different libraries which include chemical compounds and the predicted active sites of the virus proteins	–
SARS-CoV-2 3D	Simulated 3D structures	A precisely curated SARS-CoV-2 proteome database that has been extended based on the oligomeric modeling, binding prediction, mutation analysis and docking hypothesis	[69]
Tracker	Antibodies	Including antibodies and their related information gathered by volunteers all around the world and used for curing COVID-19	[70]
ViralZone	Molecular information	A web-based resource containing information correspond to all viral genes as well as useful information such as graphical structures of viruses	[71]
Virus-CKB	Chemical molecules, genes and proteins	Describing how the chemical molecules, genes and proteins involved in regulating a viral disease such as COVID-19	[72]

*Press CTRL and click on the resource name for redirecting to the relevant page.

The scoring function, which represents the binding affinity between a target and a ligand, is one of the key criteria to determine and differentiate between the docking software packages. The force field, empirical, ML and knowledge-based scoring functions are the four categories to calculate the binding affinity between a protein and a ligand. The suggested affinity score, which should not be misinterpreted as an activation or inhibition activity of the ligand on a target, explains only the possible molecular interaction properties such as van der Waals, hydrophobic, electrostatic and solvation forces. The docking

Table 3. Software packages employed in the SBDR techniques.

Software	Description	Ref
3Drefine	An interactive web server for refining proteins' structure and providing various statistical measurements	[73]
BIOVIA	A comprehensive set of software applications which can be used in molecular simulations, ADMET prediction, QSAR, structure-based drug design and pharmacophore modeling	–
CASTp	An online service for locating, delineating and measuring the geometric and topological properties of protein structures	[74]
CLC Genomics Workbench	A package for analyzing, comparing and visualizing next-generation sequencing (NGS) data like multiple sequence alignment and phylogenetic tree analysis	–
CoV-GLUE	A web tool maintaining information on the variations of the SARS-CoV-2's amino acid sequence (replacements, insertions and deletions)	[75, 76]
ExPASy	A tool that allows the translation of a nucleotide (DNA/RNA) sequence to a protein sequence and computes various chemical parameters of a given protein	[77]
HISNAPI	A software tool for predicting binding sites in the protein and nucleic acid interactions	[78]
iDMer	A module for screening effective compounds against viral infectious	[79]
JPred	A protein secondary structure prediction server	[80]
LIGPLOT	A program for plotting protein–ligand interactions based on a given PDB file	[81]
MODELLER	A software for homology or comparative modeling of a protein's 3D structure	[82]
MolAICal	An artificial intelligence-based soft tool for generating 3D structure of drugs in 3D pocket of a protein	[83]
Molprobity	A structure-validation web service that acts based on both proteins and nucleic acids information	[84]
ProSA	An interactive web service for the recognition of errors in 3D structures of proteins	[85]
PyRx	A structure-based virtual screening software to screen a list of compounds against targets of interest	[86]
SAVES	A server containing different packages, such as the Verify 3D tool, that are used for checking and analyzing the structure of proteins and their quality	–
SWISS-MODEL	A fully automated protein structure homology-modeling server	[87]
T-Coffee	A multiple sequence alignment package which has been developed based on a progressive algorithm	[88]
UCSF Chimera	A program for the interactive visualization and analysis of molecular structures and related data along with an extensible molecular modeling system	[89]

Note: ADMET, absorption distribution metabolism elimination and toxicity; QSAR, quantitative structure-activity relationship; DNA, deoxyribonucleic acid.
*Press CTRL and click on the software name for redirecting to the relevant page.

Table 4. Software suites employed for the docking and MD simulations.

Software	Description	Ref
Autodock	A suite that specifies how a biological element such as a small molecule binds to a receptor of known 3D structure	[90]
CaFE	A tool for predicting binding affinity based on free energy calculation methods	[91]
CHARMM	A program consisting of different tools with a comprehensive set of energy functions that support QM/MM, MM/CG and some solvent models	[92]
GOLD	It is applied to calculate the interaction score between a protein and a ligand (docking)	–
GROMACS	A comprehensive package to carry out MD experiments	[93]
MOE	A multi-objective software suite that provides various capabilities such as visualization, simulation, modeling and docking	[94]
NAMD	Software for MD simulations	[95]
Schrödinger	A chemistry-based computational tool for measuring binding free energy between biological elements such as proteins and ligands	[96]
SCIGRESS	A versatile platform for molecular design, modeling and docking	–
YASARA	A software tool for graphical representation of molecules, dynamics simulations and energy minimization	[97]

*Press CTRL and CLICK on the software name for redirecting to the relevant page.

concept is based on the searching conformational space that is not thoroughly covered by the currently available computer resources. Even by applying flexibility to the ligand and the residues of the binding pocket (semi-flexible), which is provided by some docking software tools, the procedure can just offer a snapshot of the interaction. Therefore, docking is a static procedure neglecting entropic effects. These deficiencies limit the potential of docking in discovering candidate medications. To tackle the limitations of the docking approaches, the MD simulations, as a helpful alternative approach, may yield a better conformational analysis of a target–ligand interaction from energetic and mechanistic viewpoints. Although the process

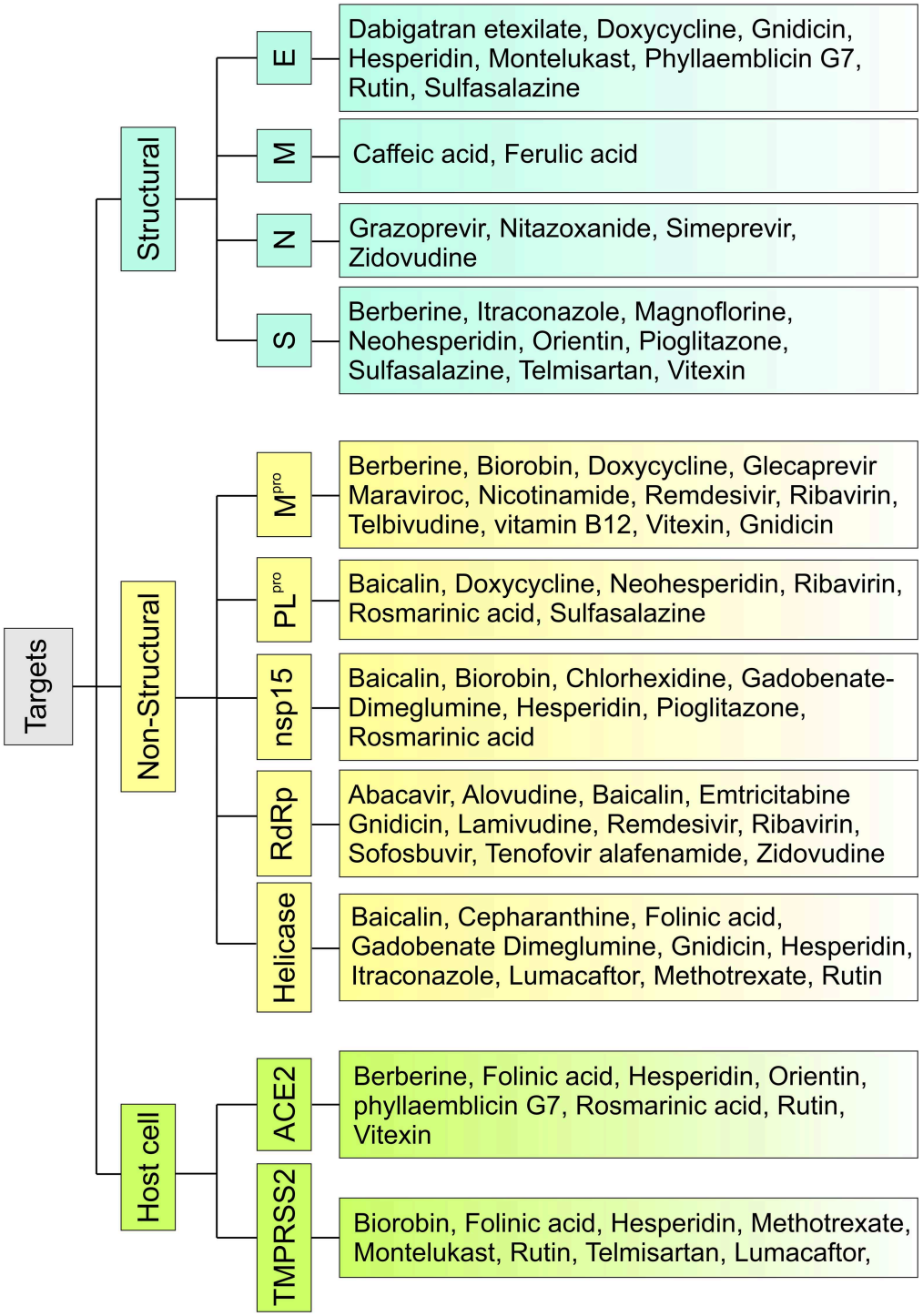

Figure 3. The proposed drugs and theirs affected targets, collected from different studies.

of running time in the MD simulations may prove too costly, the rising of high-performance computers, computer clusters and recently cloud computing can facilitate and even automate the SBDR steps.

Currently, researchers have provided a huge volume of useful and curated data in the form of repositories and databases. These data have provided opportunities for the other DR approaches such as text mining. For instance, a possible solution is to screen the sentences of the published articles and then determine their entities such as drugs and targets. Next, some features can be extracted from the validated entities and selected using the feature selection approaches. Based on the identified features, a deep neural network can be formed and trained using the efficient techniques. The obtained model can then be applied to predict the interaction of medicines and proteins of SARS-CoV-2. Finally, the outcomes can be validated

using the docking and molecular dynamic techniques. Further, the usefulness of the complementary exposure component pattern, which may play an important role in curing different diseases, can be investigated in treating COVID-19. To this end, a host–pathogen PPI network can be formed to discover a pair of drugs as a promising combination therapy. This synthetic is chosen in such a way that the drugs do not share the targets and cover the hub nodes of the network. In the process of inhibiting the hub nodes, the SBDR approaches can be utilized based on the framework introduced in the Methods section. In this regard, according to the output of this review, it seems that a combination of ribavirin and simeprevir can show a synergic result and reduce the disease complications better than the other therapeutic suggestions. Based on the different studies, these medicines, which can properly bind to RdRp and M^{pro} targets, may significantly decrease the viral loading process. A comprehensive list of possible combinations, which have been sorted by their targets and may show a synergic effect on curing COVID-19, is available in Supplementary Tables S7 and S8 (see Supplementary Data available online at http://bib.oxfordjournals.org/).

CONCLUSION

In the current review, different SBDR approaches were investigated and described for combating COVID-19. These techniques appear to be the most suitable *in silico* tools to explore *de novo* drugs against the emerging infectious diseases because they can be carried out based on the rudimentary information (e.g. whole genome of a virus) on an emerging disease. For an emerging infectious disease, the SBDR methods can be employed in five main steps, including (i) multiple sequence alignment to interpret the evolution process of a virus, (ii) homology modeling to simulate the 3D structure of proteins or enzymes if not available, (iii) drug screening to specify the potential medications to treat a designated disease, (iv) molecular docking to calculate the interaction scores between the discovered drugs and targets of interest and (v) MD simulation experiments to further investigate the nature of interactions. In addition to the methods, this review classified the database and software resources utilized in the COVID-19 projects and explained their roles. These databases and software platforms can be used to unravel the hidden information and introduce the potential drugs for curing infectious disease.

Authors' contributions

Y.M.-S., A.S. and M.M.P. have designed the work and wrote the manuscript. B.J. has written the manuscript. Y.O. and A.M.-N. have supervised the work.

Key Points

- SBDR methods are proper options for finding candidate medications in the early stages of an emerging infectious disease such as COVID-19.
- Efficient structure-based approaches have been summarized and described.
- Databases and software packages have been classified and explained.
- The repurposed drugs and their protein targets have been categorized.
- The advantages and limitations of structure-based approaches in finding potential drugs for curing emerging infectious diseases have been discussed.

Supplementary Data

Supplementary data are available online at *Briefings in Bioinformatics*.

Acknowledgements

Not applicable.

Funding

Not applicable.

References

1. Pourseif MM, Parvizpour S, Jafari B, *et al*. A domain-based vaccine construct against SARS-CoV-2, the causative agent of COVID-19 pandemic: development of self-amplifying mRNA and peptide vaccines. *Bioimpacts* 2020;**11**:65–84.
2. Huang Y, Yang C, Xu X-F, *et al*. Structural and functional properties of SARS-CoV-2 spike protein: potential antivirus drug development for COVID-19. *Acta Pharmacol Sin* 2020;**41**:1141–9.
3. Chan JF, Lau SK, To KK, *et al*. Middle East respiratory syndrome coronavirus: another zoonotic betacoronavirus causing SARS-like disease. *Clinical microbiology reviews* 2015;**28**(2):465–522.
4. Beck KR, Kaserer T, Schuster D, *et al*. Virtual screening applications in short-chain dehydrogenase/reductase research. *J Steroid Biochem Mol Biol* 2017;**171**:157–77.
5. Oppermann UC, Persson B, Filling C, *et al*. Structure-function relationships of SDR hydroxysteroid dehydrogenases. *Adv Exp Med Biol* 1997;**414**:403–15.
6. Duax W, Ghosh D, Pletnev V. Structure, mechanism of action and inhibition of dehydrogenase enzymes. In: *Structure-Based Drug Design*. Springer, 1998, 175–94. https://doi.org/10.1007/978-94-015-9028-0_16.
7. Ding X, Wu Y, Wang Y, *et al*. Accelerated CDOCKER with GPUs, parallel simulated annealing and fast Fourier transforms. *J Chem Theory Comput* 2020;**16**:3910–9.
8. Li Z, Gu J, Zhuang H, *et al*. Adaptive molecular docking method based on information entropy genetic algorithm. *Appl Soft Comput* 2015;**26**:299–302.
9. Masoudi-Sobhanzadeh Y, Motieghader H, Masoudi-Nejad A. FeatureSelect: a software for feature selection based on machine learning approaches. *BMC Bioinformatics* 2019;**20**:170.
10. Masoudi-Sobhanzadeh Y, Motieghader H, Omidi Y, *et al*. A machine learning method based on the genetic and world competitive contests algorithms for selecting genes or features in biological applications. *Sci Rep* 2021;**11**:1–19.
11. Wu G, Robertson DH, Brooks IIICL, *et al*. Detailed analysis of grid-based molecular docking: a case study of CDOCKER—A CHARMm-based MD docking algorithm. *J Comput Chem* 2003;**24**:1549–62.
12. García-Godoy MJ, López-Camacho E, García-Nieto J, *et al*. Bio-inspired optimization for the molecular docking problem:

state of the art, recent results and perspectives. *Appl Soft Comput* 2019;**79**:30–45.
13. Alogheli H, Olanders G, Schaal W, *et al*. Docking of macrocycles: comparing rigid and flexible docking in glide. *J Chem Inf Model* 2017;**57**:190–202.
14. Mishra CB, Pandey P, Sharma RD, *et al*. Identifying the natural polyphenol catechin as a multi-targeted agent against SARS-CoV-2 for the plausible therapy of COVID-19: an integrated computational approach. *Brief Bioinform* 2021;**22**(2):1346–60.
15. Masoudi-Sobhanzadeh Y, Omidi Y, Amanlou M, *et al*. Trader as a new optimization algorithm predicts drug-target interactions efficiently. *Sci Rep* 2019;**9**:1–14.
16. Lalmuanawma S, Hussain J, Chhakchhuak L. Applications of machine learning and artificial intelligence for Covid-19 (SARS-CoV-2) pandemic: a review. *Chaos Soliton Fract* 2020;**139**:110059.
17. Beck BR, Shin B, Choi Y, *et al*. 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* 2020;**18**:784–90.
18. Batra R, Chan H, Kamath G, *et al*. Screening of therapeutic agents for COVID-19 using machine learning and ensemble docking studies. *J Phys Chem Lett* 2020;**11**:7058–65.
19. Masoudi-Sobhanzadeh Y, Masoudi-Nejad A. Synthetic repurposing of drugs against hypertension: a datamining method based on association rules and a novel discrete algorithm. *BMC Bioinformatics* 2020;**21**:1–21.
20. Cheng F, Hong H, Yang S, *et al*. Individualized network-based drug repositioning infrastructure for precision oncology in the panomics era. *Brief Bioinform* 2017;**18**:682–97.
21. Zhou Y, Hou Y, Shen J, *et al*. Network-based drug repurposing for novel coronavirus 2019-nCoV/SARS-CoV-2. *Cell discovery* 2020;**6**:1–18.
22. Cheng F, Desai RJ, Handy DE, *et al*. Network-based approach to prediction and population-based validation of in silico drug repurposing. *Nat Commun* 2018;**9**:1–12.
23. Cheng F, Kovács IA, Barabási A-L. Network-based prediction of drug combinations. *Nat Commun* 2019;**10**:1–11.
24. Alaimo S, Pulvirenti A. Network-based drug repositioning: approaches, resources, and research directions. In: *Computational Methods for Drug Repurposing*. Springer, 2019, 97–113. doi: 10.1007/978-1-4939-8955-3_6.
25. Masoudi-Sobhanzadeh Y. Computational-based drug repurposing methods in COVID-19. *Bioimpacts* 2020;**10**:205.
26. Wishart DS, Knox C, Guo AC, *et al*. DrugBank: a knowledgebase for drugs, drug actions and drug targets. *Nucleic Acids Res* 2008;**36**:D901–6.
27. Kanehisa M, Goto S. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* 2000;**28**:27–30.
28. Masoudi-Sobhanzadeh Y, Omidi Y, Amanlou M, *et al*. DrugR+: a comprehensive relational database for drug repurposing, combination therapy, and replacement therapy. *Comput Biol Med* 2019;**109**:254–62.
29. Masoudi-Sobhanzadeh Y, Omidi Y, Amanlou M, *et al*. Drug databases and their contributions to drug repurposing. *Genomics* 2020;**112**:1087–95.
30. Chen X, Ji ZL, Chen YZ. TTD: therapeutic target database. *Nucleic Acids Res* 2002;**30**:412–5.
31. Han L, Wei X-X, Zheng Y-J, *et al*. Potential mechanism prediction of cold-damp plague formula against COVID-19 via network pharmacology analysis and molecular docking. *Chinas Med* 2020;**15**:1–16.
32. Elfiky AA. Anti-HCV, nucleotide inhibitors, repurposing against COVID-19. *Life Sci* 2020;**248**:117477.
33. Krishnan DA, Sangeetha G, Vajravijayan S, *et al*. Structure-based drug designing towards the identification of potential anti-viral for COVID-19 by targeting endoribonuclease NSP15. *Inform Med Unlocked* 2020;**20**:100392.
34. ul Qamar MT, Alqahtani SM, Alamri MA, *et al*. Structural basis of SARS-CoV-2 3CLpro and anti-COVID-19 drug discovery from medicinal plants. *J Pharm Anal* 2020;**10**(4):313–19.
35. Kandeel M, Al-Nazawi M. Virtual screening and repurposing of FDA approved drugs against COVID-19 main protease. *Life Sci* 2020;**251**:117627.
36. Shamsi A, Mohammad T, Anwar S, *et al*. Glecaprevir and Maraviroc are high-affinity inhibitors of SARS-CoV-2 main protease: possible implication in COVID-19 therapy. *Biosci Rep* 2020;**40**:1–16.
37. Mittal L, Kumari A, Srivastava M, *et al*. Identification of potential molecules against COVID-19 main protease through structure-guided virtual screening approach. *J Biomol Struct Dyn* 2020;1–26.
38. da Silva Hage-Melim LI, Federico LB, de Oliveira NKS, *et al*. Virtual screening, ADME/Tox predictions and the drug repurposing concept for future use of old drugs against the COVID-19. *Life Sci* 2020;**256**:117963.
39. White MA, Lin W, Cheng X. Discovery of COVID-19 inhibitors targeting the SARS-CoV-2 Nsp13 helicase. *J Phys Chem Lett* 2020;**11**:9144–51.
40. Alagu Lakshmi S, Shafreen RMB, Priya A, *et al*. Ethnomedicines of Indian origin for combating COVID-19 infection by hampering the viral replication: using structure-based drug discovery approach. *J Biomol Struct Dyn* 2020; 1–16.
41. Wu C, Liu Y, Yang Y, *et al*. Analysis of therapeutic targets for SARS-CoV-2 and discovery of potential drugs by computational methods. *Acta Pharm Sin B* 2020;**10**: 766–88.
42. Bhowmik D, Nandi R, Jagadeesan R, *et al*. Identification of potential inhibitors against SARS-CoV-2 by targeting proteins responsible for envelope formation and virion assembly using docking based virtual screening, and pharmacokinetics approaches. *Infect Genet Evol* 2020;**84**:104451.
43. Mahmoud DB, Shitu Z, Mostafa A. Drug repurposing of nitazoxanide: can it be an effective therapy for COVID-19? *J Genet Eng Biotechnol* 2020;**18**:1–10.
44. Liu T, Lin Y, Wen X, *et al*. BindingDB: a web-accessible database of experimentally determined protein–ligand binding affinities. *Nucleic Acids Res* 2007;**35**:D198–201.
45. Mendez D, Gaulton A, Bento AP, *et al*. ChEMBL: towards direct deposition of bioassay data. *Nucleic Acids Res* 2019;**47**: D930–40.
46. Wishart DS, Feunang YD, Guo AC, *et al*. DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Res* 2018;**46**:D1074–82.
47. Shivanyuk A, Ryabukhin S, Tolmachev A, *et al*. Enamine real database: making chemical diversity real. *Chemistry Today* 2007;**25**:58–9.
48. Federhen S. The NCBI taxonomy database. *Nucleic Acids Res* 2012;**40**:D136–43.
49. Berman HM, Battistuz T, Bhat TN, *et al*. The protein data bank. *Acta Crystallogr Sect D Biol Crystallogr* 2002;**58**:899–907.
50. Kim S, Chen J, Cheng T, *et al*. PubChem 2019 update: improved access to chemical data. *Nucleic Acids Res* 2019;**47**:D1102–9.
51. Corsello SM, Bittker JA, Liu Z, *et al*. The drug repurposing hub: a next-generation drug library and information resource. *Nat Med* 2017;**23**:405–8.

52. Consortium U. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res* 2019;**47**:D506–15.
53. Sterling T, Irwin JJ. ZINC 15–ligand discovery for everyone. *J Chem Inf Model* 2015;**55**:2324–37.
54. Wang LL, Lo K, Chandrasekhar Y, *et al*. CORD-19: the Covid-19 open research dataset. *ArXiv* 2020.
55. Gowthaman R, Guest JD, Yin R, *et al*. CoV3D: a database of high resolution coronavirus protein structures. *Nucleic Acids Res* 2021;**49**:D282–7.
56. Guidotti E, Ardia D. COVID-19 data hub. *J Open Source Softw* 2020;**5**:2376.
57. Korn D, Bobrowski T, Li M, *et al*. COVID-KOP: integrating emerging COVID-19 data with the ROBOKOP database. *Bioinformatics* 2020.
58. Wu J, Chen W, Zhou J, *et al*. COVIEdb: a database for potential immune epitopes of coronaviruses. *Front Pharmacol* 2020;**11**:1401.
59. Sahoo S, Mahapatra SR, Parida BK, *et al*. DBCOVP: a database of coronavirus virulent glycoproteins. *Comput Biol Med* 2020;**129**:104131.
60. Chen T-F, Chang Y-C, Hsiao Y, *et al*. DockCoV2: a drug database against SARS-CoV-2. *Nucleic Acids Res* 2021;**49**:D1152–9.
61. Rajput A, Kumar A, Megha K, *et al*. DrugRepV: a compendium of repurposed drugs and chemicals targeting epidemic and pandemic viruses. *Brief Bioinform* 2021;**22**(2):1076–84.
62. Kuleshov MV, Clarke DJ, Kropiwnicki E, *et al*. The COVID-19 gene and drug set library. *Research Square* 2020.
63. Fang S, Li K, Shen J, *et al*. GESS: a database of global evaluation of SARS-CoV-2/hCoV-19 sequences. *Nucleic Acids Res* 2021;**49**:D706–14.
64. Shu Y, McCauley J. GISAID: global initiative on sharing all influenza data–from vision to reality. *Eurosurveillance* 2017;**22**:30494.
65. Zheng Q, Jones FK, Leavitt SV, *et al*. HIT-COVID, a global database tracking public health interventions to COVID-19. *Sci Data* 2020;**7**:1–8.
66. Yang X, Lian X, Fu C, *et al*. HVIDB: a comprehensive database for human-virus protein-protein interactions. *Brief Bioinform* 2021;**22**(2):832–44.
67. Chen Q, Allot A, Lu Z. LitCovid: an open database of COVID-19 literature. *Nucleic Acids Res* 2021;**49**:D1534–40.
68. Kuleshov MV, Stein DJ, Clarke DJ, *et al*. The COVID-19 drug and gene set library. *Patterns* 2020;**1**:100090.
69. Alsulami AF, Thomas SE, Jamasb AR, *et al*. SARS-CoV-2 3D database: understanding the coronavirus proteome and evaluating possible drug targets. *Brief Bioinform* 2021;**22**(2):769–80.
70. Yang L, Liu W, Yu X, *et al*. COVID-19 antibody therapeutics tracker: a global online database of antibody therapeutics for the prevention and treatment of COVID-19. *Antib Ther* 2020;**3**:205–12.
71. Hulo C, De Castro E, Masson P, *et al*. ViralZone: a knowledge resource to understand virus diversity. *Nucleic Acids Res* 2011;**39**:D576–82.
72. Feng Z, Chen M, Liang T, *et al*. Virus-CKB: an integrated bioinformatics platform and analysis resource for COVID-19 research. *Brief Bioinform* 2021;**22**(2):882–95.
73. Bhattacharya D, Nowotny J, Cao R, *et al*. 3Drefine: an interactive web server for efficient protein structure refinement. *Nucleic Acids Res* 2016;**44**:W406–9.
74. Tian W, Chen C, Lei X, *et al*. CASTp 3.0: computed atlas of surface topography of proteins. *Nucleic Acids Res* 2018;**46**:W363–7.
75. Bernasconi A, Canakoglu A, Masseroli M, *et al*. A review on viral data sources and search systems for perspective mitigation of COVID-19. *Brief Bioinform* 2021;**22**(2):664–75.
76. Singer J, Gifford R, Cotten M, *et al*. CoV-GLUE: a web application for tracking SARS-CoV-2 genomic variation. 2020.
77. Artimo P, Jonnalagedda M, Arnold K, *et al*. ExPASy: SIB bioinformatics resource portal. *Nucleic Acids Res* 2012;**40**:W597–603.
78. Mei L-C, Wang Y-L, Wu F-X, *et al*. HISNAPI: a bioinformatic tool for dynamic hot spot analysis in nucleic acid–protein interface with a case study. *Brief Bioinform* 2021.
79. Wei Z, Gao Y, Meng F, *et al*. iDMer: an integrative and mechanism-driven response system for identifying compound interventions for sudden virus outbreak. *Brief Bioinform* 2020.
80. Drozdetskiy A, Cole C, Procter J, *et al*. JPred4: a protein secondary structure prediction server. *Nucleic Acids Res* 2015;**43**:W389–94.
81. Wallace AC, Laskowski RA, Thornton JM. LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. *Protein Eng* 1995;**8**:127–34.
82. Eswar N, Webb B, Marti-Renom MA, *et al*. Comparative protein structure modeling using Modeller. *Curr Protoc Bioinformatics* 2006;**15**:5.6. 1–30.
83. Bai Q, Tan S, Xu T, *et al*. MolAICal: a soft tool for 3D drug design of protein targets by artificial intelligence and classical algorithm. *Brief Bioinform* 2020.
84. Williams CJ, Headd JJ, Moriarty NW, *et al*. MolProbity: more and better reference data for improved all-atom structure validation. *Protein Sci* 2018;**27**:293–315.
85. Wiederstein M, Sippl MJ. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Res* 2007;**35**:W407–10.
86. Dallakyan S, Olson AJ. Small-molecule library screening by docking with PyRx. *Methods Mol Biol* 2015;**1263**:243–50.
87. Waterhouse A, Bertoni M, Bienert S, *et al*. SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res* 2018;**46**:W296–303.
88. Notredame C, Higgins DG, Heringa J. T-Coffee: a novel method for fast and accurate multiple sequence alignment. *J Mol Biol* 2000;**302**:205–17.
89. Pettersen EF, Goddard TD, Huang CC, *et al*. UCSF Chimera—a visualization system for exploratory research and analysis. *J Comput Chem* 2004;**25**:1605–12.
90. Morris GM, Huey R, Lindstrom W, *et al*. AutoDock4 and AutoDockTools4: automated docking with selective receptor flexibility. *J Comput Chem* 2009;**30**:2785–91.
91. Liu H, Hou T. CaFE: a tool for binding affinity prediction using end-point free energy methods. *Bioinformatics* 2016;**32**:2216–8.
92. Brooks BR, Brooks CL, 3rd, Mackerell AD, Jr, *et al*. CHARMM: the biomolecular simulation program. *J Comput Chem* 2009;**30**:1545–614.
93. Abraham MJ, Murtola T, Schulz R, *et al*. GROMACS: high performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* 2015;**1**:19–25.
94. Vilar S, Cozza G, Moro S. Medicinal chemistry and the molecular operating environment (MOE): application of QSAR and

molecular docking to drug discovery. *Curr Top Med Chem* 2008;**8**:1555–72.
95. Phillips JC, Braun R, Wang W, *et al*. Scalable molecular dynamics with NAMD. *J Comput Chem* 2005;**26**:1781–802.
96. Bhachoo J, Beuming T. Investigating protein–peptide interactions using the Schrödinger computational suite. In: *Modeling Peptide-Protein Interactions*. Springer, 2017, 235–54. doi: 10.1007/978-1-4939-6798-8_14.
97. Land H, Humble MS. YASARA: a tool to obtain structural guidance in biocatalytic investigations. In: *Protein Engineering*. Springer, 2018, 43–67. doi: 10.1007/978-1-4939-7366-8_4.