

## Review Article

# An Updated Review of Computer-Aided Drug Design and Its Application to COVID-19

**Arun Bahadur Gurung** <sup>1</sup>, **Mohammad Ajmal Ali** <sup>2</sup>, **Joongku Lee** <sup>3</sup>,  
**Mohammad Abul Farah** <sup>4</sup> and **Khalid Mashay Al-Anazi** <sup>4</sup>

<sup>1</sup>Department of Basic Sciences and Social Sciences, North-Eastern Hill University, Shillong, 793022 Meghalaya, India

<sup>2</sup>Department of Botany and Microbiology, College of Science, King Saud University, Riyadh 11451, Saudi Arabia

<sup>3</sup>Department of Environment and Forest Resources, Chungnam National University, 99 Daehak-ro, Yuseong-gu, Daejeon 34134, Republic of Korea

<sup>4</sup>Department of Zoology, College of Science, King Saud University, Riyadh 11451, Saudi Arabia

Correspondence should be addressed to Arun Bahadur Gurung; arunbgurung@gmail.com

Arun Bahadur Gurung and Mohammad Ajmal Ali contributed equally to this work.

Received 8 September 2020; Revised 31 May 2021; Accepted 11 June 2021; Published 29 June 2021

Academic Editor: Bilal Alatas

Copyright © 2021 Arun Bahadur Gurung et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The recent outbreak of the deadly coronavirus disease 19 (COVID-19) pandemic poses serious health concerns around the world. The lack of approved drugs or vaccines continues to be a challenge and further necessitates the discovery of new therapeutic molecules. Computer-aided drug design has helped to expedite the drug discovery and development process by minimizing the cost and time. In this review article, we highlight two important categories of computer-aided drug design (CADD), viz., the ligand-based as well as structure-based drug discovery. Various molecular modeling techniques involved in structure-based drug design are molecular docking and molecular dynamic simulation, whereas ligand-based drug design includes pharmacophore modeling, quantitative structure-activity relationship (QSARs), and artificial intelligence (AI). We have briefly discussed the significance of computer-aided drug design in the context of COVID-19 and how the researchers continue to rely on these computational techniques in the rapid identification of promising drug candidate molecules against various drug targets implicated in the pathogenesis of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). The structural elucidation of pharmacological drug targets and the discovery of preclinical drug candidate molecules have accelerated both structure-based as well as ligand-based drug design. This review article will help the clinicians and researchers to exploit the immense potential of computer-aided drug design in designing and identification of drug molecules and thereby helping in the management of fatal disease.

## 1. Introduction

Drug discovery is a lengthy process that takes around 10-15 years [1] and costs up to 2.558 billion USD for a drug to reach the market [2]. It is a multistep process that begins with the identification of suitable drug target, validation of drug target, hit to lead discovery, optimization of lead molecules, and preclinical and clinical studies [3]. Despite the high investments and time incurred for the discovery of new drugs, the success rate through clinical trials is only 13% with a relatively high drug attrition rate [4]. In the majority of the

cases (40-60%), the drug failure at a later stage has been reported due to lack of optimum pharmacokinetic properties on absorption, distribution, metabolism, excretion, and toxicity (ADME/Tox) [5]. The use of computer-aided drug discovery (CADD) techniques in preliminary studies by leading pharmaceutical companies and research groups has helped to expedite the drug discovery and development process minimizing the costs and failures in the final stage [6]. The application of rational drug design as an integral part of CADD provides useful insights into the understanding of the binding affinity and molecular interaction between target protein

and ligand. Additionally, lead identification in pharmaceutical research has been facilitated by the availability of super-computing facility, parallel processing, and advanced programs, algorithms, and tools [7]. Furthermore, recent advancements in artificial intelligence (AI) and machine learning methods have greatly aided in analyzing, learning, and explaining the pharmaceutical-related big data in the drug discovery process [8]. Different methods employed in the identification of new inhibitors from chemical databases include pharmacophore modeling, quantitative structure-activity relationship (QSAR), molecular docking, quantum mechanics, and statistical learning methods. CADD can be broadly divided into structure-based and ligand-based drug design approaches, both have been widely used in the drug discovery process in the identification of suitable lead molecules. While the structure-based drug design relies on the three-dimensional structure of the target receptor and its active sites to understand the molecular interaction between the receptor and ligand, the ligand based-drug design depends on the knowledge of ligands interacting with the given target receptor [9]. Computer-aided drug design has a large number of success stories and continues to play a vital role in the drug discovery process [10]. In this regard, the approach has been utilized in proposing drug candidates against coronavirus disease 2019 (COVID-19). COVID-19 is caused by a novel coronavirus known as severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) which taxonomically belongs to the Betacoronavirus genre and possesses high nucleotide sequence similarity with severe acute respiratory syndrome coronavirus (SARS-CoV) and Middle East respiratory syndrome coronavirus (MERS-CoV). The epidemiology, genome composition, pathogenesis, animal models, diagnostics, and vaccine development with references to various computational biology approaches for MERS-CoV infections have been comprehensively reviewed by Skariyachan et al. (2019) [11]. SARS-CoV-2 is a positive-sense single-stranded enveloped RNA virus approximately 30,000 bp in length which utilizes host cellular machinery to execute various pathogenic processes such as viral entry, genomic replication, and protein synthesis [12].

Like SARS and MERS, the genome of SARS-CoV-2 encodes sixteen nonstructural proteins (nsps) such as main protease ( $M^{pro}$ ), papain-like protease, RNA-dependent RNA polymerase (RdRp), helicase etc., four structural proteins (envelope, membrane, spike, and nucleocapsid), and other accessory proteins. While the spike glycoprotein is essential for the interaction of the virus with the host cell receptor, the nsps play a major role during the virus life cycle by engaging in the production of subgenomic RNAs [13, 14]. The nonstructural and structural proteins, therefore, offer promising targets for the design and development of antiviral agents against COVID-19 [13]. The lack of effective vaccines or drugs for the treatment of COVID-19 and the high mortality rate necessitates the rapid discovery of novel drugs [15], and computer-aided drug design is believed to be an important tool to achieve the identification of novel therapeutics. There is a possibility of the development of effective lead molecules against COVID-19 by utilizing natural lead molecules obtained through virtual screening and pharmaco-

kinetic prediction [16]. To speed up the discovery of a potential treatment for SARS-CoV-2 infection in humans, repurposing of broad-spectrum antiviral drugs is a promising strategy due to the availability of the pharmacokinetic and pharmacodynamic data of these drugs [17]. The availability of complete genome sequence of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and the elucidation of the viral protein structures through X-ray crystallography, nuclear magnetic resonance (NMR), electron microscopy, and homology modelling approach have allowed the identification of inhibitor drugs against the essential therapeutic drug targets of COVID-19. This review article provides useful insights into some of the common in silico methods used in CADD and how these methods have been currently used and can be of help in the drug discovery process of COVID-19.

## 2. Structure-Based Drug Design

The availability of the three-dimensional structure of the therapeutic target proteins and exploration of the binding site cavity forms the basis of structure-based drug design (SBDD) [18]. This approach is specific and effectively fast in the identification of lead molecules and their optimization which has helped to understand disease at a molecular level [19]. Some of the common methods employed in SBDD include structure-based virtual screening (SBVS), molecular docking, and molecular dynamics (MD) simulations. These methods find numerous applications such as assessment of binding energetics, protein-ligand interactions, and conformational changes in the receptor upon binding with a ligand [20]. Being used by many pharmaceutical industries and medicinal chemists, SBDD as a computational technique has greatly helped in the discovery of several drugs available in the market. For example, the discovery of amprenavir as a potential inhibitor of the human immunodeficiency virus (HIV) protease using protein modeling and MD simulations [21, 22], thymidylate synthase inhibitor, raltitrexed against HIV using SBDD approach [23], identification of topoisomerase II and IV inhibitor, norfloxacin which is an antibiotic commonly used against urinary tract infection using SBVS [18], the discovery of dorzolamide, a carbonic anhydrase inhibitor used against glaucoma, cystoid macular oedema using fragment-based screening [24], antituberculosis drug, isoniazid which is an enoyl-acyl-ACP reductase (InhA) inhibitor discovered through structure-based virtual screening and pharmacophore modeling [25], and flurbiprofen, a nonsteroidal anti-inflammatory drug (NSAID) used against rheumatoid arthritis, osteoarthritis etc. which targets cyclooxygenase-2 (COX-2) discovered through molecular docking approach etc. [26, 27]. The basic steps involved in SBDD consist of the preparation of target structure, identification of the ligand binding site, compound library preparation, molecular docking and scoring functions, molecular dynamic simulation, and binding free energy calculation (Figure 1).

*2.1. Preparation of the Target Structure.* With the rapid advancement in structural elucidation techniques such as X-ray and NMR, the structures deposited and available in protein data bank (PDB) have increased over the last few

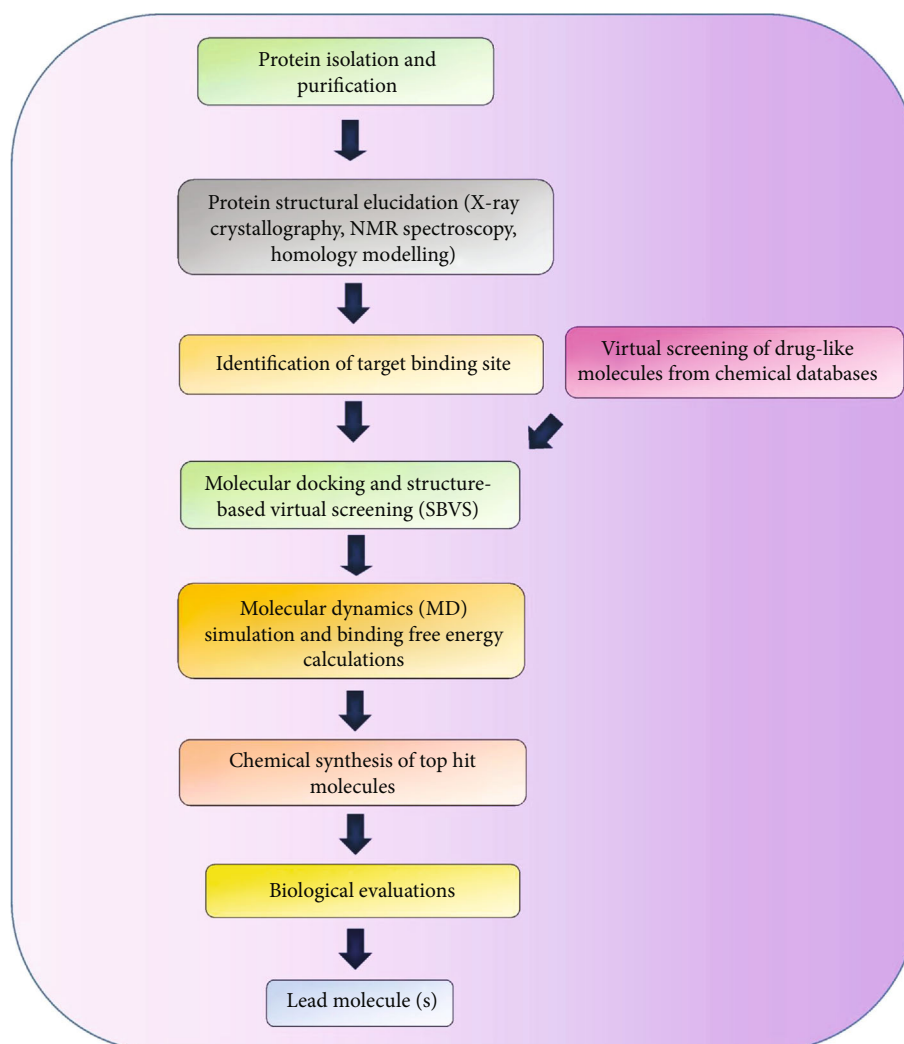


FIGURE 1: Basic steps involved in the structure-based drug design approach.

decades. Owing to the limitations of experimental techniques, many target protein structures have not been solved to date [28]. Computational technique such as comparative homology modeling [29], threading [30], and ab initio modeling [31] has been quite successful in deciphering the structures of the proteins from their sequences. Homology modeling is a widely used computational method for accurately determining the three-dimensional structure of a protein from its amino acid sequence using a suitable template structure [32]. It is a multistep process comprising of the following steps: (a) identification of template, (b) sequence alignments, (c) model building of the target (d) model refinement, and (e) model validation [29]. Protein threading is another method for protein structure prediction which is often used when (1) the target protein shares low sequence similarity with other proteins in the PDB (<25% sequence identity), and (2) the target protein shows structural similarity with some proteins in the PDB. Unlike homology modeling, which only takes into account the sequence similarity between the target and the template, protein threading considers the structural information (secondary structure, sol-

vent accessibility and pairwise interactions) encoded in the template to enhance prediction accuracy [33]. The ab initio modeling is another computational technique which is preferably used if the target protein does not have any template structures in the existing biological databases [31]. It considers a global optimization problem to find the dihedral angle values for a given protein structure which contribute to the structure's stability (possessing the global or near global minimum potential energy) [34].

**2.2. Identification of the Ligand Binding Site.** The information about the ligand-binding site is a prerequisite for carrying out specific docking. The knowledge of the binding sites can be extracted from the site-directed mutagenesis study or X-ray crystallographic structures of proteins cocrystallized with substrates or inhibitors [35]. While the experimental information about the binding site of many proteins is not available, there is plenty of software and webservers such as CASTp [36], DoGSite Scorer [37], NSiteMatch [38], DEPTH [39], MSPocket [40], MetaPocket [41], and Q-SiteFinder [42] which allows us to predict the putative binding sites of the

TABLE 1: Molecular docking tools for protein-ligand interaction studies.

Tools	Key features	Reference
AutoDock	The methods available for conformational searching in AutoDock are Lamarckian genetic algorithm, simulated annealing search, and a traditional genetic algorithm search. The prediction of binding free energies of small molecules to protein targets is based on a semiempirical free energy force field	[62]
AutoDock Vina	AutoDock Vina calculations rely on a sophisticated gradient optimization method and achieve approximately two orders of magnitude improvement in speed and better accuracy of predicting binding modes compared to AutoDock	[63]
GOLD	GOLD (genetic optimization for ligand docking) is an automated ligand docking program that allows full ligand conformational flexibility with partial flexibility of the protein and explores the binding conformations using a genetic algorithm	[64]
CDOCKER	CDOCKER (CHARMm-based DOCKER) is an automated MD docking program that uses the CHARMm19 family of force fields and offers full flexibility of ligand and CHARMm engine with reduced computation time	[65]
FlexX	FLEXX is a full automated docking tool for flexible ligands which produces reliable results with good accuracy. The FlexX method is dependent on the selection and placement of base fragments of ligand and placement and the assumption that the best base fragments interacting with the active site give a good score	[66]
Surflex	Surflex is a docking program that uses a combination of combined Hammerhead's empirical scoring function and molecular similarity method to produce putative poses of ligand fragments	[67]
GLIDE	Glide (grid-based ligand docking with energetics) performs an exhaustive search of the positional, orientational, and conformational space of a ligand binding to a receptor with reasonable computational speed. The scoring of the binding conformations is based on the ChemScore function.	[68]
DOCK6	DOCK 6 is a docking program that evaluates the conformational sampling of small molecules based on the anchor-and-grow search algorithm	[69]
SwissDock	SwissDock is a web server that allows the docking of small molecules to target proteins that are based on the EADock DSS engine	[70]

target proteins. The bulky compounds which do not fit well within the binding site pocket are rejected during the lead identification procedure.

**2.3. Compound Library Preparation.** Chemical compounds can be selected from chemical databases such as ZINC (N=230 million purchasable compounds) [43], PubChem (N=111 million pure and characterized chemical compounds) [44], MCULE (N=122 million synthetically accessible compounds) (<https://mcule.com/>), ChEMBL (>1.6 million distinct compounds) [45], DrugBank (N=14528 drug molecules) [46], and ChemSpider (N=25 million unique chemical compounds) [47]. Molecular docking is performed with drug-like compounds which are filtered using Lipinski's rule of five and ADMET (absorption, distribution, metabolism, excretion, and toxicity) parameters and other risk parameters such as acute rat toxicity, carcinogenicity, serum glutamic oxaloacetic transaminase elevation, hepatotoxicity, and inhibition of 3A4 oxidation of midazolam [28]. According to Lipinski's rule of five, a compound is considered to be orally bioactive if its physicochemical properties lies within the permissible limits such as molecular weight (MW)  $\leq 500$ , partition coefficient between n - octanol and water ( $\log P$ )  $\leq 5$ , number of hydrogen bond donor (HBD)  $\leq 5$ , and number of hydrogen bond acceptor (HBA)  $\leq 10$  [48]. Some commonly used ADMET properties include human gastrointestinal absorption (HIA), blood-brain barrier (BBB) permeation, P-glycoprotein (P-gp) inhibition, cytochromes P450 (CYP) inhibition, and plasma protein binding [49]. Besides the pharmaco-

kinetic properties, drug, and safety, the synthetic accessibility of these compounds should also be taken into account.

**2.4. Molecular Docking and Scoring Functions.** Molecular docking is a computational technique to study the interaction between a target receptor and ligand at the molecular level and allows ranking of the ligands by assessing their binding affinity towards the receptor using various scoring functions [50]. The favorable binding poses of the ligands with a target active site rely on two factors: (a) wide conformational space taking into consideration different binding poses and (b) explicit prediction of binding affinity of ligands corresponding to each binding pose [51]. A list of frequently used molecular docking programs is enumerated in Table 1. Molecular docking can be classified into two types: flexible-ligand search docking and flexible-protein docking. The ligand flexibility in the case of the flexible-ligand search docking method most commonly uses three algorithms such as systematic method, stochastic method, and simulation method [52], whereas flexible-protein docking usually relies on Monte Carlo (MC) and molecular dynamic (MD) methods [53, 54].

**2.5. Molecular Dynamic (MD) Simulation.** The MD simulation of a protein was first performed in the late 1970s [55]. This powerful physical technique is used to predict the positions of each atom in a molecular system with respect to time which is based on Newton's laws of motions governing interatomic interactions [56]. The forces between interacting



TABLE 2: A summary of commonly used molecular dynamic (MD) simulation software.

Software	Key features	Simulation system	Reference
GROMACS	GROMACS (Groningen MACHine for chemical simulation) is an efficient and versatile MD program with source code that is suited for the simulation of biological (macro) molecules in aqueous and membrane environments. The program can be run on single processors or parallel computer systems and is compatible with various force fields such as GROMOS, OPLS, AMBER, and ENCAD force fields.	Proteins, lipids, carbohydrate, nucleic acids	[71]
AMBER	Amber is an extensively used biomolecular simulation program with an assembly of codes that are designed to work together. It is a collection of codes that are designed to work together and principally divided into three major step-system preparation (antechamber, LEaP programs), simulation (sander), and trajectory analysis (ptraj analysis program).	Proteins, nucleic acids, carbohydrates	[72]
CHARMM	CHARMM (chemistry at HARvard molecular mechanics) is a widely used molecular simulation program that is primarily designed to study biological molecules such as proteins, peptides, lipids, nucleic acids, carbohydrates, and small molecule ligands. The calculations are based on different energy functions (quantum mechanical-molecular mechanical force fields, all-atom classical potential energy functions) and models such as explicit solvent, implicit solvent, and membrane models.	Proteins, lipids, carbohydrates, nucleic acids	[73]
NAMD	NAMD is a high-performance biomolecular simulation program that employs the prioritized message-driven execution capabilities of the charm++/converse parallel runtime system compatible with parallel supercomputers and workstation clusters.	Proteins, lipids, carbohydrates, nucleic acids,	[74]
Desmond	Desmond is a powerful molecular dynamic simulation program designed by D. E. Shaw with considerable speed, accuracy, and scalability. It supports explicit solvent simulations with periodic boundary conditions and can be used to model explicit membrane systems under various conditions.	Proteins, lipids	<a href="https://www.schrodinger.com/desmond">https://www.schrodinger.com/desmond</a>
Tinker	Tinker is a molecular modeling and dynamic package written primarily in a standard Fortran 95 with OpenMP extensions. It supports a wide variety of classical molecular simulations particularly biomolecular calculations and offers various force fields including the modern polarizable atomic multipole-based AMOEBA model.	Proteins, nucleic acids	[75]
LAMMPS	LAMMPS (large-scale atomic/molecular massively parallel simulator) is a classical molecular dynamic code for materials modeling. It has potentials for soft matter (biomolecules, polymers), solid-state materials (metals, semiconductors), and coarse-grained or mesoscopic systems.	Proteins, lipids, carbohydrates, nucleic acids	<a href="https://lammmps.sandia.gov/">https://lammmps.sandia.gov/</a>
DL_POLY	DL_POLY is a general purpose molecular dynamic simulation package, which allows the study of liquids of large complexity. The code is developed using the replicated data (RD) parallelization strategy.	Membranes, proteins	[76]

atoms are estimated using a suitable force field which is used to determine the overall energy of the system [57]. MD simulations have been widely used for several reasons. The position and motion of every atom of the system are captured at every point in time, which is quite tough using any experimental technique. The simulation conditions are exactly known and can be carefully modulated [58]. MD simulations have been extensively used in the structure-based drug discovery process as this technique helps to unravel many atomistic details such as binding, unbinding, and conformational changes in the receptor at a fine resolution which normally cannot be obtained from experimental studies [59, 60]. Further, using MD simulation it is possible to explore the dynamics of receptor-ligand interactions (association and dissociation) and quantify the thermodynamics, kinetics,

and free energy landscape [61]. Some examples of MD simulation programs include GROMACS, AMBER, CHARMM, NAMD, and Desmond (Table 2).

### 3. Ligand-Based Drug Design

Ligand-based drug design is another widely used approach used in computer-aided drug design and is employed when the three-dimensional structure of the target receptor is not available. The information derived from a set of active compounds against a specific target receptor can be used in the identification of physicochemical and structural properties responsible for the given biological activity which is based on the fact that structural similarities correspond to similar biological functions [77]. Some of the common techniques

used in the ligand-based virtual screening approach include pharmacophore modeling, quantitative structure-activity relationships (QSARs), and artificial intelligence (AI).

**3.1. Pharmacophore Modeling.** A pharmacophore model elucidates the spatial arrangement of chemical features in ligands that are required for interaction with the target receptor [78]. Some of the chemical features used in pharmacophore modeling include hydrogen bond donors, hydrogen bond acceptors, aromatic ring systems, hydrophobic areas, positively charged ionizable groups, and negatively charged ionizable groups [79]. Ligands having different scaffolds but the similar spatial arrangement of key interacting functional moieties can be identified using pharmacophore-based virtual screening. The bioactive conformation of the molecules within the target binding site can be incorporated into the pharmacophore model. The pharmacophore model is also often used in QSAR studies in the molecular alignment stage [80]. Some frequently used programs which allow automatic construction of the pharmacophore model include Catalyst, PHASE, LigandScout, GALAHAD, and PharmMapper (Table 3). A good pharmacophore model also incorporates spatial constraints in regions occupied by inactive molecules and often optimized further to make the model less restrictive. All the pharmacophoric features which are not consistently detected in active molecules are either made optional or removed from the final model [7]. The pharmacophore model generated should have optimum sensitivity and specificity to minimize the chances of false negative and false positive results and must be validated using an independent external test set [81]. If the information about the 3D structure of a receptor and a set of known active compounds are lacking, then a sequence-derived 3D pharmacophore model is quite useful. For example, Pharma<sup>3D</sup> utilizes knowledge of the 3D crystal structures and homology models to derive the common sequence motif important for receptor-ligand biomolecular interactions in protein families [81, 82].

**3.2. Quantitative Structure-Activity Relationships (QSARs).** QSAR studies are based on the principle that variations in the bioactivity of the compounds can be correlated with changes in the molecular structures. They are widely used in the drug discovery process in the hit to lead identification or lead optimization. A statistical model is constructed using these correlation studies, and the final model can be used to predict the biological activity of new molecules [80]. The key requirements for the generation of a reliable QSAR model are (a) a sufficient number of data sets with biological activities obtained from common experimental protocols, (b) the training and test set compounds must be appropriately selected, (c) no autocorrelation among the physicochemical properties of the ligands that may cause overfitting of the data, and (d) the applicability and predictivity of the final model must be checked using internal and external validation methods [96]. Based on how the descriptors are derived, QSAR can be classified into six different types: (a) 1D-QSAR which studies the correlation between global molecular properties such as logP and pKa with biological activities, (b) 2D-QSAR wherein biological activities are correlated

with the structural patterns such as 2D-pharmacophores and connectivity indices, (c) 3D-QSAR which studies how the biological activities correlated with noncovalent interaction fields surrounding the ligands, (d) 4D-QSAR which is an extension of 3D-QSAR with the addition of an ensemble of ligand configurations, (e) 5D-QSAR which incorporates various induced-fit models in 4D-QSAR, and (f) 6D-QSAR further extends 5D-QSAR by including different solvation models [97]. Some examples of 3D QSAR programs include the HypoGen module of Catalyst [98], PHASE [89], comparative molecular field analysis (CoMFA) [99], and comparative similarity indices analysis (CoMSIA) [100]. A list of tools for the calculation of molecular descriptors is enumerated in Table 4. QSAR technique can be classified into two types: linear and nonlinear based on chemometric methods. The linear method includes linear regression (LR), multiple linear regression (MLR), partial least squares (PLS), principal component analysis (PCA), and principal component regression (PCR). The examples of nonlinear QSAR methods are *k*-nearest neighbours (kNN), artificial neural networks (ANN), and Bayesian neural nets [97].

**3.3. Artificial Intelligence and Drug Discovery.** Artificial intelligence (AI) is a type of machine intelligence that relies on the ability of computers to learn from existing data. AI has been used in various computational modeling methods to predict the biological activities and toxicities of drug molecules [97]. Further, AI has wide applications in drug discovery such as prediction of protein folding, protein-protein interaction, virtual screening, QSAR, evaluation of ADMET properties, and de novo drug design [103]. There are two powerful methods widely used in rational drug design which include machine learning (ML) and deep learning (DL) [104]. ML algorithms that have been extensively used in drug discovery include support vector machine (SVM) [105], Random Forest (RF) [106], and Naive Bayesian (NB) [107]. Few examples of the deep learning methods are convolutional neural network (CNN), deep neural network (DNN), recurrent neural network (RNN), autoencoder, and restricted Boltzmann machine (RBM) [4]. The conventional QSAR methods can efficiently predict simple physicochemical properties such as logP and solubility. However, the QSAR prediction of complex biological properties such as drug efficacy and side effects is often not optimal as the methods use small training sets [108] and has coverage of limited chemical space [109]. The big data generated using high-throughput screening (HTS) techniques are huge challenges to traditional QSAR methods and machine learning techniques [40]. AI methods have been developed to deal with this big data of high volume and multidimensional nature to efficiently predict drug efficacy and side effects in animals or humans. The most promising approach in the present big data world is deep learning which was first used in the drug discovery process in 2012 QSAR machine learning challenge backed by Merck [110]. The results showed that deep learning models were true which can accurately predict the ADMET properties compared to traditional machine learning methods. Although, AI is an impressive method in identification of preclinical candidates in more cost and time-efficient manner, and the

TABLE 3: A list of pharmacophore modelling tools.

Tools	Description	Reference
Catalyst	Catalyst program is based on an algorithm that identifies three-dimensional configurations of chemical features common to a set of ligands, wherein each configuration is scored based both on estimated rarity and the level to which it is common to the input set.	[83]
LigandScout	LigandScout is a fully automated tool for generating pharmacophore models which detect and classifies protein-ligand interactions (hydrogen bond interactions, charge transfers, and lipophilic regions) which form the basis of the pharmacophore model used for high throughput virtual screening.	[84]
DISCO	DISCO is an automated pharmacophore method which examines the data to find all pharmacophore hypothesis that fit and serves as a complement to 3D QSAR.	[85]
PharmaGist	PharmaGist is a freely available web server used for generating ligand-based pharmacophore models, wherein the input is a set of drug-like molecules (maximum limit of 32 drug-like molecules) that have a binding affinity to the target protein.	[86]
PharmMapper	PharmMapper server is a freely available web server that is commonly used for the identification of potential target receptors for a given small molecule using the pharmacophore mapping approach.	[87]
Pharmer	Pharmer is a pharmacophore search program that organizes molecular data using the Pharmer KDB-tree and bloom fingerprints which allow rapid screening of millions of molecules in a reasonable time.	[88]
PHASE	PHASE is an advanced pharmacophore-based tool that comprehensively maps the common spatial arrangement of functional groups in a set of bioactive ligands using a novel tree-based partitioning algorithm.	[89]
ZINCPharmer	ZINCPharmer is an online web server for the screening of small molecules from the ZINC database using the Pharmer pharmacophore search program. An initial pharmacophore hypothesis can be derived either from PDB structures or by importing pharmacophore models developed using other third-party tools.	[90]
e-Pharmacophore	The e-pharmacophore method generates energetically optimized, structure-based pharmacophore models which can be used for screening of millions of compounds. The method uses the glide XP scoring function to score protein-ligand interactions and has good database screening enrichments.	<a href="https://www.schrodinger.com/e-pharmacophores">https://www.schrodinger.com/e-pharmacophores</a>
GASP	GASP program uses a genetic algorithm (GA) for the superimposition of a set of flexible ligands where the ligand possessing the lowest number of chemical features are chosen as a template, onto which other molecules are fitted.	[91]
Shape4	Shape4 is a structure-based pharmacophore program developed is to enhance increase the efficiency of database searching by considering the topographical constraints of the target binding site and consequently helps to help minimize the false positive rate.	[92]
Snooker	Snooker is a structure-based pharmacophore tool that generates pharmacophore hypotheses from homology models with critical residues for ligand binding identified through the study of Shannon entropies of structurally conserved positions in multiple sequence alignments and does not rely on the prior information of ligand structure or interactions.	[93]
Pocketv.2	Pocket v.2 is an automated program to generate a pharmacophore model from a given protein-ligand complex structure and has been designed using the pocket module of LigBuilder.	[94]
GALAHAD	GALAHAD is a pharmacophore program developed to perform flexible alignment of small molecules that bind to a target protein and share similar interaction patterns and shapes.	[95]

accurate prediction of binding affinity between a drug molecule and a receptor using AI remains challenging for quite a several reasons. Firstly, AI is a data mining method whose

performance heavily relies on the amount and quality of the available data [4, 111]. Variability in the source of data especially those derived from different biological assays and

TABLE 4: List of programs available for calculating molecular descriptors for building QSAR models.

Programs	Molecular descriptors	Reference
ADMET Predictor	Predicts over 140 properties such as solubility, logP, pKa, sites of CYP metabolism, and Ames mutagenicity.	<a href="https://www.simulations-plus.com/software/admetpredictor/">https://www.simulations-plus.com/software/admetpredictor/</a>
ChemAxon	It provides a wide range of chemical calculations such as molecular weight, elemental composition, LogP, pKa, LogD, LogS, hydrogen bond donor/acceptor (HBDA) count, and various 2D topological descriptors and 3D geometrical descriptors.	<a href="https://chemaxon.com/products/calculators-and-predictors">https://chemaxon.com/products/calculators-and-predictors</a>
PaDEL-Descriptor	PaDEL-descriptor is a standalone software for calculating molecular descriptors and fingerprints including 797 descriptors (663 1D, 2D descriptors, and 134 3D descriptors) and 10 types of fingerprints.	[101]
E-DRAGON	E-DRAGON is the electronic remote version of the noted software DRAGON, which is an application for the calculation of molecular descriptors (>1,600 molecular descriptors) used for evaluating molecular structure-activity or structure-property relationships.	<a href="http://www.vcclab.org/lab/edragon/">http://www.vcclab.org/lab/edragon/</a>
DRAGON 7.0	Dragon 7.0 calculates 5,270 molecular descriptors including the simplest atom types, functional groups and fragment counts, topological and geometrical descriptors, and three-dimensional descriptors which are organized into thirty logical blocks.	<a href="https://chm.kode-solutions.net/products_dragon.php">https://chm.kode-solutions.net/products_dragon.php</a>
CODESSA PRO	CODESSA PRO (comprehensive descriptors for structural and statistical analysis) is an extensive program for studying quantitative structure-activity/property relationships (QSAR/QSPR) by facilitating the calculation of a wide range of molecular descriptors derived from the 2D/3D geometrical structure and/or quantum-chemical wave function of small molecules.	[102]
Pre-ADMET	PreADMET is a web-based program that can be used for the calculation of drug-like physicochemical descriptors such as lipophilicity (logP), molecular weight, polar surface area, and water solubility.	<a href="https://preadmet.bmdrc.kr/">https://preadmet.bmdrc.kr/</a>
QikProp	QikProp program allows the prediction of various pharmacologically important descriptors of chemical compounds such as octanol/water and water/gas log Ps, log S, log BB, overall CNS activity, Caco-2, and MDCK cell permeabilities, log K <sub>HSA</sub> for human serum albumin binding, etc.	<a href="https://www.schrodinger.com/qikprop">https://www.schrodinger.com/qikprop</a>
ACD/labs	ACD/labs is a multipurpose tool that can be used for the calculation of a variety of physicochemical descriptors such as aqueous solubility, boiling point/vapour pressure, logD, logP, and pKa and ADME properties.	<a href="https://www.acdlabs.com/">https://www.acdlabs.com/</a>
Corina Symphony	CORINA symphony is a cheminformatic tool for structure representation and calculation of molecular descriptors of six types: global molecular descriptors, shape descriptors, quantum-mechanical properties, 2D property-weighted autocorrelation, 3D property-weighted autocorrelation, property-weighted radial distribution functions (RDF), and autocorrelation of surface properties.	<a href="https://www.mn-am.com/products/corinasymphony">https://www.mn-am.com/products/corinasymphony</a>
MOPAC	MOPAC (molecular orbital PACKage) is a semiempirical molecular orbital package for studying solid-state and molecular structures and reactions and offers calculations of various descriptors of molecules, radicals, ions, and polymers such as the vibrational spectra, thermodynamic parameters, isotopic substitution effects, and force constants.	<a href="http://openmopac.net/">http://openmopac.net/</a>

lack of high-quality data from public databases presents difficulty in efficient AI learning [112, 113].

#### 4. Case Study of COVID-19

Both ligand-based and structure-based drug design approaches have been widely used in the drug discovery process against coronavirus disease-19 (COVID-19), an infectious viral disease caused by SARS-CoV-2. To date, only a few drug-candidate molecules have undergone clinical trials, and these molecules are mostly repurposed approved drugs (Figure 2).

The lack of approved drugs and vaccines for COVID-19 and the high mortality rate of the pandemic necessitate identification of effective therapeutics. With the availability of the complete genome sequence of SARS-CoV-2 [114] and structural elucidation of the viral proteins through X-ray crystallography, NMR spectroscopy, electron microscopy and homology modeling, COVID-19 research has been rapidly pursued. Some of the important drug targets of SARS-CoV-2 are the structural protein-spike (S) protein, envelope (E) protein, membrane (M) protein, and the nucleocapsid (N) protein (Figure 3); nonstructural proteins (Nsps) (Figure 4) such as the main protease which is also known as 3C-like protease 3CL<sup>pro</sup> (nsp5), papain-like protease (PL<sup>pro</sup>, nsp3),



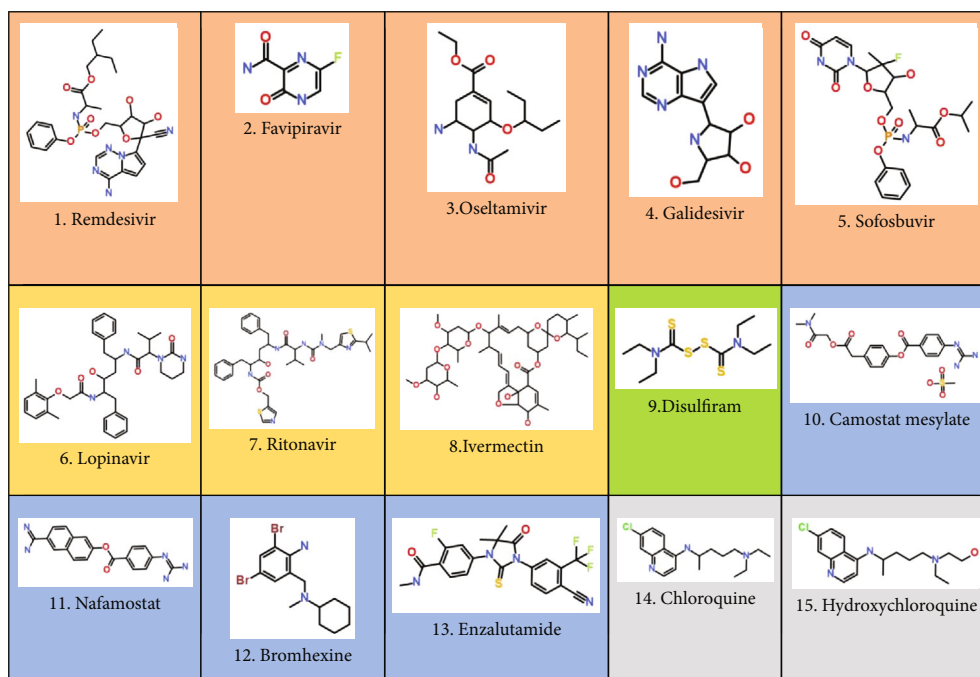


FIGURE 2: Molecules currently investigated in clinical trials where molecules 1-5 in the orange box are RNA polymerase inhibitors, molecules 6-8 in the yellow box are 3C-like protease inhibitors, molecule 9 in the green box is a papain-like protease inhibitor, molecules 10-13 in the blue box are TMPRSS2 inhibitors, and molecules 14-15 in the grey box are inhibitors of endosomal acidification.

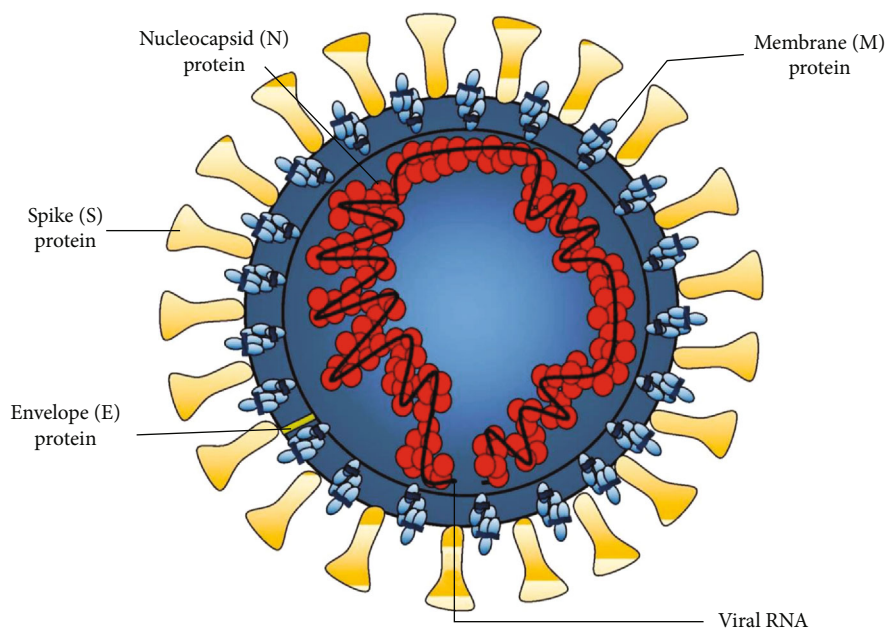


FIGURE 3: The structural proteins of SARS-CoV-2.

RNA-dependent RNA polymerase (RdRp, nsp12), nsp15 endoribonuclease, nsp16 2'-O-methyltransferase, nsp13 helicase, and host-based pharmacological targets are angiotensin-converting enzyme 2 (ACE2), transmembrane protease serine 2 (TMPRSS2), furin, and cathepsin [115]. The details of the nonstructural proteins are briefly discussed here. The main protease is a cysteine protease with a catalytic dyad (cysteine and histidine) in its active pocket [116]. The

action of the catalytic activity of  $M^{Pro}$  on polyproteins results in the release of the vital proteins required for viral replication by cleaving at least 11 sites around the C-terminal and the central regions of the viral polyproteins with sequence consensus  $X-(L/F/M)-Q↓(G/A/S)-X$  [117, 118]. Papain-like protease (PLpro) is the second SARS-CoV-2 proteases potentially targetable with small molecules which cleave three sites, with recognition sequence consensus "LXGG↓XX" [118]. It

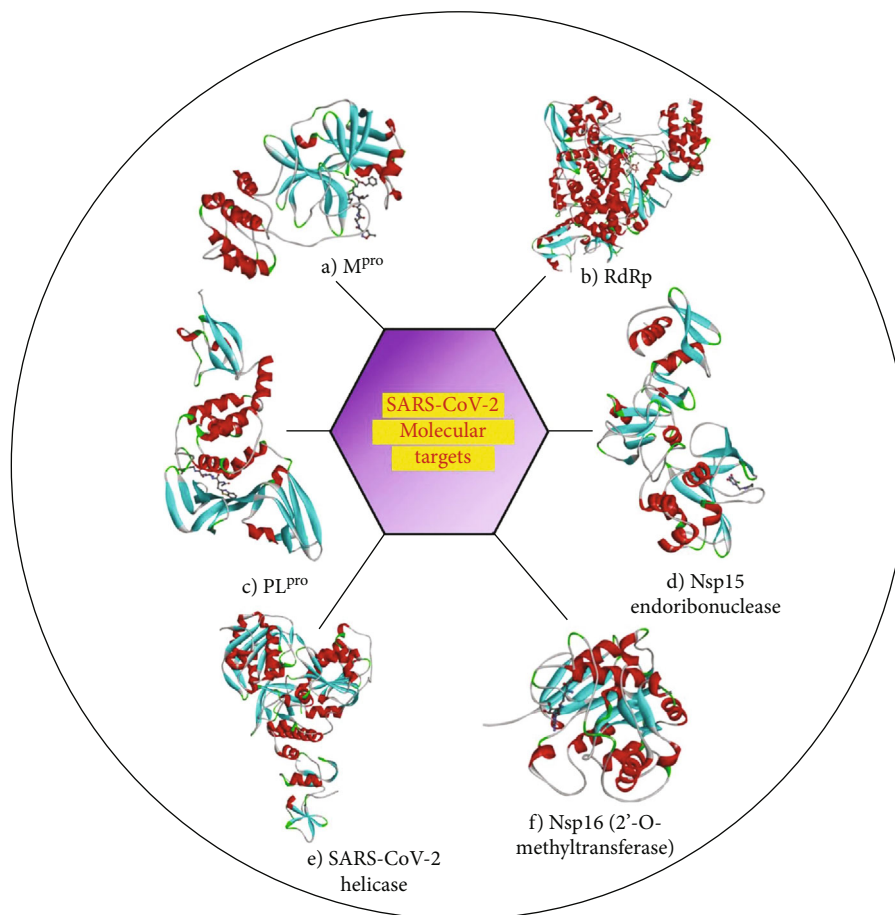


FIGURE 4: Macromolecular target structures of SARS-CoV-2. (a) X-ray crystal structure of the SARS-CoV-2 main protease in complex with an inhibitor N3 (PDB ID: 7BQY). (b) Crystal structure of Nsp12 (RdRp) bound to triphosphate form of remdesivir (PDB ID: 7BV2). (c) Crystal structure of the SARS-CoV-2 papain-like protease in complex with peptide inhibitor VIR250 (PDB ID: 6WUU). (d) Crystal structure of Nsp15 endoribonuclease from SARS-CoV-2 in complex with potential repurposing drug tipiracil (PDB ID: 6WXC). (e) Crystal structure of the SARS-CoV-2 helicase (PDB ID: 6ZSL). (f) Crystal structure of Nsp16 (2'-O-methyltransferase) from SARS-CoV-2 in complex with sinefungin. The secondary structure elements—helices, sheets, and loops—are colored in red, cyan, and grey, respectively, and the bound inhibitors are rendered as a ball-and-stick model.

is an attractive drug target because of its essential role in not only the cleavage and maturation of viral polyproteins and assembly of the replicase-transcriptase complex but also disruption of host immune responses [119]. RNA-dependent RNA polymerase (RdRp) is the cleavage product of the polyproteins 1a and 1ab from ORF1a and ORF1ab and is involved in the replication and transcription of the SARS-CoV-2 genome [120]. The catalytic core of the enzyme resembles the human right hand with differentiated palm, fingers, and thumb domains. Targeting this enzyme to halt the viral replication seems an effective therapeutic approach since the active site of the RdRp is a highly conserved and accessible region [121]. Nsp15 is a uridine-specific endoribonuclease involved in RNA processing and widely distributed in all kingdoms of life. Its catalytic C-terminal domain exhibits sequence similarity and functionality of the EndoU family enzymes [122]. The active 234-kDa hexameric enzyme cleaves both single- and double-stranded RNA at uridine sites generating 2',3'-cyclic phosphodiester and 5'-hydroxyl termini [123]. The SARS CoV-2 2'-o-methyltrans-

ferase (nsp16) is another important enzyme target essential for viral multiplication. The enzyme precisely protects the viral RNA from the cellular innate immunity by participating in the formation of a specific arrangement known as RNA cap, a structure which contributes to viral RNA stability and effective process of translation [124]. SARS-Cov-2 Nsp13 helicase is one of the critical enzyme among the 16 known CoV Nsp proteins which shows the highest sequence conservation across the CoV family, indicating their importance for viral multiplication. The enzyme possesses the NTPase and RNA helicase functions that can hydrolyze all types of NTPs and unwind RNA helix in an ATP-dependent process [125]. The transmembrane protease serine 2 (TMPRSS2) is a major host factor which regulates virus-host cell membrane fusion and cell entry by priming of the virus spike (S) protein via cleavage of the S proteins at the S1/S2 and S2 sites [126]. Furin is a type of proprotein convertases (PCs) found in the *trans*-Golgi complex and gets activated by acidic pH. The enzyme recognizes and hydrolyzes the unique "RRAR" motif in SARS-CoV-2-spike

protein [127]. Cathepsin L is a lysosomal cysteine protease belonging to a family of proteases involved in proteolysis of protein antigens produced by pathogen endocytosis. The protease cleaves the S1 subunit of the coronavirus spike glycoprotein which is required for the virus entry into human host cells, virus, and host cell endosome membrane fusion [128]. These structures solved through experimental techniques or computational homology modeling techniques can be used for structure-based virtual screening for identification of specific inhibitors of the target proteins.

The CADD methods have been successfully used in the COVID-19 drug discovery process. Selvaraj et al. (2020) solved the three-dimensional structure of SARS-CoV-2 guanine-N7 methyltransferase (nsp14) using the homology modeling method and further proposed five TCM database compounds—TCM 57025, TCM 3495, TCM 5376, TCM 20111, and TCM 31007 as potential antiviral phytochemicals based on molecular docking and simulation studies [129]. Gao et al. (2021) characterized the physicochemical property, subcellular localization, and homology model of the SARS-CoV-2 nucleocapsid protein and further explored its biological function using mass spectrometry analysis and flow cytometry [130]. Beck et al. (2020) used a pretrained deep learning-based drug-target interaction model called molecule transformer drug and identified a few Food and Drug Administration (FDA) approved antiviral drugs such as atazanavir, remdesivir, efavirenz, ritonavir, and dolutegravir showing inhibitory potential against SARS-CoV-2 3C-like proteinase [131]. Elfiky (2020) used homology modeling, molecular dynamic simulations, and molecular docking approaches to target the SARS-CoV-2 RdRp enzyme and reported the suitability of sofosbuvir, ribavirin, galidesivir, remdesivir, favipiravir, cefuroxime, tenofovir, and hydroxychloroquine as candidate drugs for clinical trials [132]. Elmezayen et al. (2020) used a structure-based virtual screening method to identify lead molecules against main proteases and human TMPRSS2. Four potential inhibitors against M<sup>Pro</sup> enzyme identified were talampicillin, lurasidone, ZINC000000702323, and ZINC000012481889, whereas promising inhibitors identified against TMPRSS2 include rubitecan, loprazolam, ZINC000015988935, and ZINC000103558522 [133]. Das et al. (2020) used a molecular docking approach to identify potential inhibitors against SARS-CoV-2 main protease by screening a set of natural products, antivirals, antifungal, antinematodes, and antiprotozoal. The inhibitors identified from the study include rutin (a natural compound), ritonavir (control drug), emetine (antiprotozoal), hesperidin (a natural compound), lopinavir (control drug), and indinavir (antiviral drug) [134]. Gurung et al. (2020) used a molecular docking approach and identified three antiviral phytochemicals: bonducellpin D, 5,7-dimethoxyflavanone-4'-O-β-d-glucopyranoside and caesalmin B as potential inhibitors of SARS-CoV-2 M<sup>Pro</sup>, SARS-CoV M<sup>Pro</sup>, and Middle East respiratory syndrome-coronavirus (MERS-CoV) M<sup>Pro</sup> [135]. Joshi et al. (2020) identified natural molecules such as δ-viniferin, myricitrin, taiwanhomoflavone A, lactucopicrin 15-oxalate, nympholide A, afzelin, biorobin, hesperidin, and phyllaemblicin B as potential inhibitors of SARS-CoV-2 M<sup>Pro</sup> using molecular docking approach [136]. Wahedi et al. (2021) explored stilbe-

noid analogues as potential anti-COVID-19 drug candidates using molecular docking and molecular dynamic simulation studies and identified piceatannol and resveratrol as important lead molecules for disrupting SARS-CoV-2 and ACE-2 complex formation [137]. Khan et al. (2020) attempted to target chymotrypsin-like protease (3CL<sup>Pro</sup>) with small molecules using molecular docking and molecular dynamic simulation approach, and the study revealed three FDA approved drugs (remdesivir, saquinavir, and darunavir) and two natural compounds (flavone and coumarin derivatives) as promising inhibitors of the target enzyme [138]. Further, the potentiality of many dietary flavonols as antiviral drugs targeting the SARS-CoV-2 enzymes and proteins (3CL<sup>Pro</sup>, PL<sup>Pro</sup>, S protein and RdRp) has been discussed comprehensively by Mouffouk et al. (2021) [139]. Umesh et al. (2021) screened chemical species from Indian spices using a computational approach (molecular docking and molecular dynamic simulation) and identified carnosol, arjunglucoside-I, and rosmanol as potent inhibitors of the novel coronavirus main protease (SARS-CoV-2 M<sup>Pro</sup>) [140]. Abdelli et al. (2021) explored essential oil from antiviral and antimicrobial plant *Ammoides verticillata* (Desf.) Briq. that blocks the function of the SARS-CoV-2 angiotensin-converting enzyme 2 (ACE2) receptor using in silico approach (molecular docking, pharmacophore mapping, and MD simulation) and identified isothymol as a promising functional inhibitor of ACE2 receptor [141]. Al-Khafaji et al. (2020) employed a covalent docking screening procedure coupled with the MD simulation technique to identify molecules that can form a covalent bond with Cys145 within the binding pocket of SARS-CoV-2 main protease and identified FDA approved drugs: saquinavir, ritonavir, and remdesivir as top three molecules [142]. Peele et al. (2020) screened FDA approved antiviral drugs, antimalarial drugs, and plant-derived natural drugs with antiviral activity through molecular docking and identified lopinavir, amodiaquine, and theaflavin digallate as promising inhibitors against SARS-CoV-2 main protease and confirmed their stability in the binding pocket of the target enzyme using molecular dynamics simulation [143]. Wang (2020) identified three potential inhibitors of SARS-CoV-2 main protease: carfilzomib, eravacycline, valrubicin, lopinavir, and elbasvir using virtual docking screening of approved drugs and drug candidates in clinical trials followed by MD simulation and binding free energy calculation [144]. Mittal et al. (2021) used antiprotease molecules for drug repurposing against COVID-19 and identified six potential inhibitors of main protease enzyme-leupeptin, hemisulphate, pepstatin A, nelfinavir, birinapant, lypressin, and octreotide using virtual screening and molecular dynamic simulation approach [145]. Using molecular modeling and virtual screening approach, Kandeel and Al-Nazami (2020) identified ribavirin and telbivudine as potential inhibitors of SARS-CoV-2 main protease enzyme from a set of FDA approved drugs [146]. ul Qamar et al. (2020) used the homology structure model of SARS-CoV-2 3CL<sup>Pro</sup> for the screening of antiviral phytochemicals and identified three lead compounds 5,7,3',4'-tetrahydroxy-2'-(3,3-dimethylallyl) isoflavone, myricitrin, and methyl rosmarinate as potential inhibitors of the target enzyme through molecular docking and molecular dynamic simulation approach [147]. Islam et al. (2020) used molecular

docking and MD simulation technique and identified five antiviral phytochemicals, viz., hypericin, cyanidin 3-glucoside, baicalin, glabridin, and  $\alpha$ -ketoamide-11r which showed a good binding affinity with SARS-CoV-2 main protease enzyme [148]. Beura and Chetti (2021) studied few derivatives of chloroquine using pharmacophore modeling, molecular docking, binding free energy calculation, and ADME property analysis and discovered molecule CQD15 as a promising inhibitor of SARS-CoV-2 main protease which shows better interactions with the target enzyme as compared to chloroquine and hydroxychloroquine [149]. Mahanta et al. (2021) screened FDA approved antimicrobial drugs using a combined approach of molecular docking and molecular dynamic simulation and proposed viomycin as a potential inhibitor of the main protease of SARS-CoV-2 [150]. Enmozhi et al. (2021) explored the potentiality of antiviral phytocompound from *Andrographis paniculata* as an SARS-CoV-2 main protease ( $M^{Pro}$ ) inhibitor using molecular docking and ADME prediction [151]. Kumar et al. (2020) screened hydroxyethylamine (HEA-) based library of chemical compounds using molecular docking where HEA is a pharmacophore derived from indinavir. They identified compound 16 as a promising inhibitor of SARS-CoV-2 3CL $^{Pro}$  which shows drug-like properties and stable binding within the binding pocket of the target enzyme throughout MD simulation studies [152]. Arun et al. (2020) used the crystal structure of SARS-CoV-2 in complex with an imidazole carboxamide inhibitor and generated an E-pharmacophore hypothesis for the repurposing of drugs. They identified two drugs binifibrate and bamifylline which bind strongly to the enzyme active site pocket as revealed from molecular docking, binding free energy calculation, and molecular dynamic simulation [153]. Gentile et al. (2020) screened marine natural product (MNP) library using hyphenated pharmacophore model, molecular docking, and molecular dynamic simulation approach and identified a total of 17 compounds (belong to the class phlorotannins isolated from the brown alga, *Sargassum spinuligerum*) as potential SARS-CoV-2  $M^{Pro}$  inhibitors [154]. Amin et al. (2020) constructed a Monte Carlo optimization-based QSAR model and used it for the virtual screening of some inhouse chemicals. The thirteen identified compounds showed good drug-likeness from SwissADME in silico study, and molecular docking studies further show their favorable interactions with target SARS-CoV-2 PL $^{Pro}$ , thereby suggesting their potentiality as a seed for drug design and optimization against SARS-CoV-2 PL $^{Pro}$  [155]. Similarly, Ghosh et al. (2020) used the Monte Carlo optimization-based QSAR model for screening a library of nature product hits. Fragment analysis of the active molecules suggests that novel potential SARS-CoV-2  $M^{Pro}$  enzyme inhibitors may be synthesized by joining fragments/features together or attaching with other scaffolds [156]. Pharmacophore modeling is a fast and effective approach in the identification of interesting lead molecules for drug discovery against COVID-19. A ligand-based pharmacophore model was generated by Law et al. (2020) using established antiviral drugs, and the model was used to estimate the antiviral activity of twenty vanillin derivatives as  $M^{Pro}$  inhibitors of SARS-CoV-2. Further, the structure-based pharmacophore model suggests that vanillin derivatives (1-20) exhibited promising

results, and these compounds were suggested to be potent COVID-19 antiviral compounds [157]. Using the X-ray crystallographic structure of COVID-19 main protease ( $M^{Pro}$ ), Daoud et al. (2020) constructed a pharmacophore model and further conducted a molecular docking study to identify antiviral drugs as potential COVID-19 main protease inhibitors. Five FDA-approved antiviral drugs (lopinavir, remdesivir, ritonavir, saquinavir, and raltegravir) were successfully captured by the pharmacophore model, and docking studies revealed that these compounds exhibit many specific binding interactions comparable to that of the cocrystallized inhibitor (X77) [158]. Skariyachan et al. (2020) explored the binding potentiality of six approved drugs (chloroquine, hydroxychloroquine, favipiravir, lopinavir, remdesivir, and ritonavir) against fifteen potential drug targets of SARS-CoV-2 (spike glycoprotein, RNA dependent RNA polymerase, nsp7, nsp8, papain-like protease, main protease, nucleocapsid protein, heptad repeat of domain 2, ADP ribose phosphatase, nsp9 RNA binding protein, endoribonuclease, orf7a, nsp10, and nsp1) using molecular docking and molecular dynamic simulation approach and concluded that out of all the six drugs, ritonavir and lopinavir showed better binding with the prioritized drug targets [159]. Furthermore, the antiviral mechanisms of these drugs (remdesivir, lopinavir/ritonavir, and chloroquine/hydroxychloroquine) against SARS-CoV-2 have been thoroughly reviewed by Uzunova et al. (2020) [160]. Singh et al. (2021) performed docking-based virtual screening from a library of 1764 antiviral drugs against SARS-CoV-2 NSP12 (RNA polymerase) and identified five compounds, viz., paritaprevir, glecaprevir, velpatasvir, remdesivir, and ribavirin which exhibited high-binding affinity with the drug target [161]. Gowrishankar et al. (2021) screened a total of 57 phytochemicals from three most commonly used Indian herbs (*Justicia adhatoda*, *Eucalyptus globulus*, and *Vitex negundo*) used in “steam inhalation therapy” against four structural protein targets of SARS-CoV-2 viz. 3CL $^{Pro}$ , ACE2, spike glycoprotein, and RdRp using molecular docking approach, and the best five lead molecules identified were apigenin-o-7-glucuronide, ellagic acid, eudesmol, viridiflorene, vasicolinone, and anisotine [162]. Ibrahim et al. (2021) explored the potentialities of eighteen repurposed drugs in clinical development against SARS-CoV-2  $M^{Pro}$  using combined molecular docking and molecular dynamic (MD) techniques and identified TMC-310911 and ritonavir as promising drugs for the treatment of COVID-19 [163].

## 5. Strengths and Challenges of CADD in COVID-19 Research

With the steady rise in the number of confirmed positive and death cases from SARS-CoV-2 infection, computer-aided drug design (CADD) emerges as a fast and reliable technique in pharmaceutical and medicinal research since it not only saves time but also helps to cut costs of designing therapeutic agents [164]. Further, realizing the severity of COVID-19 and the lack of approved therapeutic agents warrants the need for finding potent drugs in less time, and the CADD method makes this possible by facilitating the discovery of new drugs or repurposing FDA-approved drugs whose safety



and adverse effects are already known [165]. Since the inherent mutability of the SARS-CoV-2 genome may hinder disease prevention and treatment, CADD can be used efficiently to predict the effects of mutation on drug binding with the molecular receptors [166]. Therefore, CADD can greatly help in accelerating the drug discovery and development process. However, CADD methods have some limitations such as lead molecules derived from the virtual screening process that still need validation through preclinical and clinical assessments before market approval [167]. The fact that the molecular mechanism studies underlying the disease pathogenesis of COVID-19 are still underway, and the existence of bias and imbalance in the limited data available can have a major impact on the prediction accuracy of CADD methods such as artificial intelligence [168].

## 6. Conclusions

Structure-based and ligand-based drug design form two branches of the computer-aided drug discovery process which plays a significant role in the design and identification of drug molecules in reduced time and cost. The increase in the number of positive cases and deaths from COVID-19 and the lack of approved drugs and vaccines continue to be a matter of global health concern which necessitates the urgent discovery of drugs for the prevention and cure of the disease. The structural elucidation of pharmacological targets of SARS-CoV-2 has helped the researchers in the structure-based virtual identification of inhibitors, and the discovery of few lead molecules against COVID-19 has led to the use of scaffolds that can be optimized through ligand-based drug design. Realizing the possible mutability of this RNA virus and the emergence of drug resistance problems, it is, therefore, necessary to take a step further and consider targeting multiple drug targets that will be more effective and might help in overcoming drug resistance barriers.

## Data Availability

The data is not required.

## Conflicts of Interest

The authors report no conflicts of interest in this work.

## Acknowledgments

The authors would like to extend their sincere appreciation to the Deanship of Scientific Research at King Saud University for its funding of the research through the research group project #RG-1438-015. J. Lee thank the Chungnam National University, Daejeon, Republic of Korea, for the funding support.

## References

- [1] C. M. Song, S. J. Lim, and J. C. Tong, "Recent advances in computer-aided drug design," *Briefings in Bioinformatics*, vol. 10, no. 5, pp. 579–591, 2009.
- [2] J. A. DiMasi, H. G. Grabowski, and R. W. Hansen, "Innovation in the pharmaceutical industry: New estimates of R&D costs," *Journal of Health Economics*, vol. 47, pp. 20–33, 2016.
- [3] D. Vohora and G. Singh, *Pharmaceutical Medicine and Translational Clinical Research*, Academic Press, 2018.
- [4] F. Zhong, J. Xing, X. Li et al., "Artificial intelligence in drug design," *Science China Life Sciences*, vol. 61, no. 10, pp. 1191–1204, 2018.
- [5] T. Hou and X. Xu, "Recent development and application of virtual screening in drug discovery: an overview," *Current Pharmaceutical Design*, vol. 10, no. 9, pp. 1011–1033, 2004.
- [6] W. Yu and A. D. Mac Kerell, "Computer-aided drug design methods," in *Antibiotics*, Humana Press, New York, NY, 2017.
- [7] S. J. Y. Macalino, V. Gosu, S. Hong, and S. Choi, "Role of computer-aided drug design in modern drug discovery," *Archives of Pharmacal Research*, vol. 38, no. 9, pp. 1686–1701, 2015.
- [8] W. Duch, K. Swaminathan, and J. Meller, "Artificial intelligence approaches for rational drug design and discovery," *Current Pharmaceutical Design*, vol. 13, no. 14, pp. 1497–1508, 2007.
- [9] H.-J. Huang, H. W. Yu, C.-Y. Chen et al., "Current developments of computer-aided drug design," *Journal of the Taiwan Institute of Chemical Engineers*, vol. 41, no. 6, pp. 623–635, 2010.
- [10] M. Hassan Baig, K. Ahmad, S. Roy et al., "Computer aided drug design: success and limitations," *Current Pharmaceutical Design*, vol. 22, no. 5, pp. 572–581, 2016.
- [11] S. Skariyachan, S. B. Challapilli, S. Packirisamy, S. T. Kumargowda, and V. S. Sridhar, "Recent aspects on the pathogenesis mechanism, animal models and novel therapeutic interventions for Middle East respiratory syndrome coronavirus infections," *Frontiers in Microbiology*, vol. 10, p. 569, 2019.
- [12] S. A. Amin and T. Jha, "Fight against novel coronavirus: a perspective of medicinal chemists," *European Journal of Medicinal Chemistry*, vol. 201, article 112559, 2020.
- [13] B. Goyal and D. Goyal, "Targeting the dimerization of the main protease of coronaviruses: A potential broad-spectrum therapeutic strategy," *ACS Combinatorial Science*, vol. 22, no. 6, pp. 297–305, 2020.
- [14] W. Dai, B. Zhang, X.-M. Jiang et al., "Structure-based design of antiviral drug candidates targeting the SARS-CoV-2 main protease," *Science*, vol. 368, no. 6497, pp. 1331–1335, 2020.
- [15] Y.-F. Tu, C.-S. Chien, A. A. Yarmishyn et al., "A review of SARS-CoV-2 and the ongoing clinical trials," *International Journal of Molecular Sciences*, vol. 21, no. 7, p. 2657, 2020.
- [16] D. Gopal and S. Skariyachan, *Recent Perspectives on COVID-19 and Computer-Aided Virtual Screening of Natural Compounds for the Development of Therapeutic Agents Towards SARS-CoV-2*, Methods in Pharmacology and Toxicology, Springer, 2020.
- [17] T. Pillaiyar, S. Meenakshisundaram, and M. Manickam, "Recent discovery and development of inhibitors targeting coronaviruses," *Drug Discovery Today*, vol. 25, no. 4, pp. 668–688, 2020.
- [18] M. Batool, B. Ahmad, and S. Choi, "A structure-based drug discovery paradigm," *International Journal of Molecular Sciences*, vol. 20, no. 11, p. 2783, 2019.
- [19] E. Lionta, G. Spyrou, D. K. Vassilatis, and Z. Cournia, "Structure-based virtual screening for drug discovery: principles,

- applications and recent advances," *Current Topics in Medicinal Chemistry*, vol. 14, no. 16, pp. 1923–1938, 2014.
- [20] S. Kalyaanamoorthy and Y.-P. P. Chen, "Structure-based drug design to augment hit discovery," *Drug Discovery Today*, vol. 16, no. 17–18, pp. 831–839, 2011.
- [21] A. Wlodawer and J. Vondrasek, "Inhibitors of HIV-1 protease: a major success of structure-assisted drug design," *Annual Review of Biophysics and Biomolecular Structure*, vol. 27, no. 1, pp. 249–284, 1998.
- [22] D. E. Clark, "What has computer-aided molecular design ever done for drug discovery?," *Expert Opinion on Drug Discovery*, vol. 1, no. 2, pp. 103–110, 2006.
- [23] A. C. Anderson, "The process of structure-based drug design," *Chemistry & Biology*, vol. 10, no. 9, pp. 787–797, 2003.
- [24] S. Grover, M. A. Apushkin, and G. A. Fishman, "Topical dorzolamide for the treatment of cystoid macular edema in patients with retinitis pigmentosa," *American Journal of Ophthalmology*, vol. 141, no. 5, pp. 850–858, 2006.
- [25] H. Marrakchi, G. Lan  elle, and A. Qu  emard, "InhA, a target of the antituberculous drug isoniazid, is involved in a mycobacterial fatty acid elongation system, FAS-II," *Microbiology*, vol. 146, no. 2, pp. 289–296, 2000.
- [26] S. Dadashpour, T. Tuylu Kucukkilinc, O. Unsal Tan, K. Ozadali, H. Irannejad, and S. Emami, "Design, synthesis and in vitro study of 5, 6-dDiaryl-1, 2, 4-triazine-3-ylthioacetate derivatives as COX-2 and  $\beta$ -amyloid aggregation inhibitors," *Archiv der Pharmazie*, vol. 348, no. 3, pp. 179–187, 2015.
- [27] Z. Miller, K.-S. Kim, D.-M. Lee et al., "Proteasome inhibitors with pyrazole scaffolds from structure-based virtual screening," *Journal of Medicinal Chemistry*, vol. 58, no. 4, pp. 2036–2041, 2015.
- [28] X. Wang, K. Song, L. Li, and L. Chen, "Structure-based drug design strategies and challenges," *Current Topics in Medicinal Chemistry*, vol. 18, no. 12, pp. 998–1006, 2018.
- [29] V. K. Vyas, R. D. Ukawala, C. Chintha, and M. Ghate, "Homology modeling a fast tool for drug discovery: current perspectives," *Indian Journal of Pharmaceutical Sciences*, vol. 74, no. 1, pp. 1–17, 2012.
- [30] C. M.-R. Lemer, M. J. Rooman, and S. J. Wodak, "Protein structure prediction by threading methods: evaluation of current techniques," *Proteins: Structure, Function, and Genetics*, vol. 23, no. 3, pp. 337–355, 1995.
- [31] J. Lee, P. L. Freddolino, and Y. Zhang, "Ab initio protein structure prediction," in *From protein structure to function with bioinformatics*, Springer, 2017.
- [32] M. T. Muhammed and E. Aki-Yalcin, "Homology modeling in drug discovery: Overview, current applications, and future perspectives," *Chemical Biology & Drug Design*, vol. 93, no. 1, pp. 12–20, 2019.
- [33] J. Xu, F. Jiao, and L. Yu, "Protein structure prediction using threading," in *Protein structure prediction*, Springer, 2008.
- [34] M. Yousef, T. Abdelkader, and K. El-Bahnasy, "Performance comparison of ab initio protein structure prediction methods," *Ain Shams Engineering Journal*, vol. 10, no. 4, pp. 713–719, 2019.
- [35] L. Pan, C. L. Gardner, F. A. Pagliai, C. F. Gonzalez, and G. L. Lorca, "Identification of the tolfenamic acid binding pocket in PrBP from *Liberibacter asiaticus*," *Frontiers in Microbiology*, vol. 8, 2017.
- [36] T. A. Binkowski, S. Naghibzadeh, and J. Liang, "CASTp: computed atlas of surface topography of proteins," *Nucleic Acids Research*, vol. 31, no. 13, pp. 3352–3355, 2003.
- [37] A. Volkamer, D. Kuhn, F. Rippmann, and M. Rarey, "DoGSiteScorer: a web server for automatic binding site prediction, analysis and druggability assessment," *Bioinformatics*, vol. 28, no. 15, pp. 2074–2075, 2012.
- [38] J. Sun and K. Chen, "NSiteMatch: prediction of binding sites of nucleotides by identifying the structure similarity of local surface patches," *Computational and Mathematical Methods in Medicine*, vol. 2017, Article ID 5471607, 16 pages, 2017.
- [39] K. P. Tan, R. Varadarajan, and M. S. Madhusudhan, "DEPTH: a web server to compute depth and predict small-molecule binding cavities in proteins," *Nucleic Acids Research*, vol. 39, suppl\_2, pp. W242–W248, 2011.
- [40] H. Zhu and M. T. Pisabarro, "MSPocket: an orientation-independent algorithm for the detection of ligand binding pockets," *Bioinformatics*, vol. 27, no. 3, pp. 351–358, 2011.
- [41] B. Huang, "MetaPocket: a meta approach to improve protein ligand binding site prediction," *OMICS: A Journal of Integrative Biology*, vol. 13, no. 4, pp. 325–330, 2009.
- [42] A. T. R. Laurie and R. M. Jackson, "Q-SiteFinder: an energy-based method for the prediction of protein-ligand binding sites," *Bioinformatics*, vol. 21, no. 9, pp. 1908–1916, 2005.
- [43] J. J. Irwin and B. K. Shoichet, "ZINC – a free database of commercially available compounds for virtual screening," *Journal of Chemical Information and Modeling*, vol. 45, no. 1, pp. 177–182, 2005.
- [44] S. Kim, P. A. Thiessen, E. E. Bolton et al., "PubChem Substance and Compound databases," *Nucleic Acids Research*, vol. 44, no. D1, pp. D1202–D1213, 2016.
- [45] A. Gaulton, L. J. Bellis, A. P. Bento et al., "ChEMBL: a large-scale bioactivity database for drug discovery," *Nucleic Acids Research*, vol. 40, no. D1, pp. D1100–D1107, 2012.
- [46] D. S. Wishart, C. Knox, A. C. Guo et al., "DrugBank: a knowledgebase for drugs, drug actions and drug targets," *Nucleic Acids Research*, vol. 36, suppl\_1, pp. D901–D906, 2008.
- [47] H. E. Pence and A. Williams, *Chem Spider: An Online Chemical Information Resource*, ACS Publications, 2010.
- [48] C. A. Lipinski, "Lead- and drug-like compounds: the rule-of-five revolution," *Drug Discovery Today: Technologies*, vol. 1, no. 4, pp. 337–341, 2004.
- [49] A. Daina, O. Michielin, and V. Zoete, "SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules," *Scientific Reports*, vol. 7, no. 1, 2017.
- [50] S.-Y. Huang and X. Zou, "Advances and challenges in protein-ligand docking," *International Journal of Molecular Sciences*, vol. 11, no. 8, pp. 3016–3034, 2010.
- [51] I. M. Kapetanovic, "Computer-aided drug discovery and development (CADDD): in silico-chemico-biological approach," *Chemico-Biological Interactions*, vol. 171, no. 2, pp. 165–176, 2008.
- [52] S. F. Sousa, P. A. Fernandes, and M. J. Ramos, "Protein-ligand docking: current status and future challenges," *Proteins: Structure, Function, and Bioinformatics*, vol. 65, no. 1, pp. 15–26, 2006.
- [53] C. M. Oshiro, I. D. Kuntz, and J. S. Dixon, "Flexible ligand docking using a genetic algorithm," *Journal of Computer-Aided Molecular Design*, vol. 9, no. 2, pp. 113–130, 1995.

- [54] T. N. Hart and R. J. Read, "A multiple-start Monte Carlo docking method," *Proteins: Structure, Function, and Genetics*, vol. 13, no. 3, pp. 206–222, 1992.
- [55] J. A. McCammon, B. R. Gelin, and M. Karplus, "Dynamics of folded proteins," *Nature*, vol. 267, no. 5612, pp. 585–590, 1977.
- [56] M. Karplus and J. A. McCammon, "Molecular dynamics simulations of biomolecules," *Nature Structural Biology*, vol. 9, no. 9, pp. 646–652, 2002.
- [57] M. De Vivo, M. Masetti, G. Bottegoni, and A. Cavalli, "Role of molecular dynamics and related methods in drug discovery," *Journal of Medicinal Chemistry*, vol. 59, no. 9, pp. 4035–4061, 2016.
- [58] S. A. Hollingsworth and R. O. Dror, "Molecular dynamics simulation for all," *Neuron*, vol. 99, no. 6, pp. 1129–1143, 2018.
- [59] J. D. Durrant and J. A. McCammon, "Molecular dynamics simulations and drug discovery," *BMC Biology*, vol. 9, no. 1, pp. 1–9, 2011.
- [60] H. Zhao and A. Caflisch, "Molecular dynamics in drug design," *European Journal of Medicinal Chemistry*, vol. 91, pp. 4–14, 2015.
- [61] D. Huang and A. Caflisch, "The free energy landscape of small molecule unbinding," *PLoS Computational Biology*, vol. 7, no. 2, article e1002002, 2011.
- [62] G. M. Morris, R. Huey, W. Lindstrom et al., "AutoDock4 and AutoDockTools4: automated docking with selective receptor flexibility," *Journal of Computational Chemistry*, vol. 30, no. 16, pp. 2785–2791, 2009.
- [63] O. Trott and A. J. Olson, "AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading," *Journal of Computational Chemistry*, vol. 31, no. 2, pp. 455–461, 2009.
- [64] G. Jones, P. Willett, R. C. Glen, A. R. Leach, and R. Taylor, "Development and validation of a genetic algorithm for flexible docking<sup>1</sup>," *Journal of Molecular Biology*, vol. 267, no. 3, pp. 727–748, 1997.
- [65] G. Wu, D. H. Robertson, C. L. Brooks III, and M. Vieth, "Detailed analysis of grid-based molecular docking: a case study of CDOCKER?A CHARMM-based MD docking algorithm," *Journal of Computational Chemistry*, vol. 24, no. 13, pp. 1549–1562, 2003.
- [66] B. Kramer, M. Rarey, and T. Lengauer, "Evaluation of the FLEXX incremental construction algorithm for protein-ligand docking," *Proteins: Structure, Function, and Genetics*, vol. 37, no. 2, pp. 228–241, 1999.
- [67] A. N. Jain, "Surflex: fully automatic Flexible molecular docking using a molecular similarity-based search engine," *Journal of Medicinal Chemistry*, vol. 46, no. 4, pp. 499–511, 2003.
- [68] R. A. Friesner, J. L. Banks, R. B. Murphy et al., "Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy," *Journal of Medicinal Chemistry*, vol. 47, no. 7, pp. 1739–1749, 2004.
- [69] W. J. Allen, T. E. Balius, S. Mukherjee et al., "DOCK 6: impact of new features and current docking performance," *Journal of Computational Chemistry*, vol. 36, no. 15, pp. 1132–1156, 2015.
- [70] A. Grosdidier, V. Zoete, and O. Michielin, "SwissDock, a protein-small molecule docking web service based on EADock DSS," *Nucleic Acids Research*, vol. 39, suppl, pp. W270–W277, 2011.
- [71] D. Van Der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark, and H. J. C. Berendsen, "GROMACS: fast, flexible, and free," *Journal of Computational Chemistry*, vol. 26, no. 16, pp. 1701–1718, 2005.
- [72] D. A. Case, T. E. Cheatham 3rd, T. Darden et al., "The Amber biomolecular simulation programs," *Journal of Computational Chemistry*, vol. 26, no. 16, pp. 1668–1688, 2005.
- [73] B. R. Brooks, C. L. Brooks III, A. D. Mackerell Jr. et al., "CHARMM: the biomolecular simulation program," *Journal of Computational Chemistry*, vol. 30, no. 10, pp. 1545–1614, 2009.
- [74] J. C. Phillips, G. Zheng, S. Kumar, and L. V. Kalé, "NAMD: Biomolecular simulation on thousands of processors," in *ACM/IEEE SC 2002 Conference (SC'02)*, p. 36, Baltimore, MD, USA, 2002.
- [75] J. A. Rackers, Z. Wang, C. Lu et al., "Tinker 8: software tools for molecular design," *Journal of Chemical Theory and Computation*, vol. 14, no. 10, pp. 5273–5289, 2018.
- [76] W. Smith, C. W. Yong, and P. M. Rodger, "DL\_POLY: application to molecular simulation," *Molecular Simulation*, vol. 28, no. 5, pp. 385–471, 2002.
- [77] P. Prathipati, A. Dixit, and A. K. Saxena, "Computer-aided drug design: integration of structure-based and ligand-based approaches in drug design," *Current Computer-Aided Drug Design*, vol. 3, no. 2, pp. 133–148, 2007.
- [78] D. Schaller, D. Šribar, T. Noonan et al., "Next generation 3D pharmacophore modeling," *Wiley Interdisciplinary Reviews: Computational Molecular Science*, vol. 10, no. 4, article e1468, 2020.
- [79] J. H. Van Drie, "Generation of three-dimensional pharmacophore models," *Wiley Interdisciplinary Reviews: Computational Molecular Science*, vol. 3, no. 5, pp. 449–464, 2013.
- [80] C. C. Melo-Filho, R. C. Braga, and C. H. Andrade, "3D-QSAR approaches in drug design: perspectives to generate reliable CoMFA models," *Current Computer-Aided Drug Design*, vol. 10, no. 2, pp. 148–159, 2014.
- [81] A. Vuorinen and D. Schuster, "Methods for generating and applying pharmacophore models as virtual screening filters and for bioactivity profiling," *Methods*, vol. 71, pp. 113–134, 2015.
- [82] T. Klabunde, C. Giegerich, and A. Evers, "Sequence-derived three-dimensional pharmacophore models for G-protein-coupled receptors and their application in virtual screening," *Journal of Medicinal Chemistry*, vol. 52, no. 9, pp. 2923–2932, 2009.
- [83] D. Barnum, J. Greene, A. Smellie, and P. Sprague, "Identification of common functional configurations among molecules," *Journal of Chemical Information and Computer Sciences*, vol. 36, no. 3, pp. 563–571, 1996.
- [84] G. Wolber and T. Langer, "LigandScout: 3-D pharmacophores derived from protein-bound ligands and their use as virtual screening filters," *Journal of Chemical Information and Modeling*, vol. 45, no. 1, pp. 160–169, 2005.
- [85] Y. C. Martin, *Distance comparisons: a new strategy for examining three-dimensional structure—activity relationships*, ACS Publications, 1995.
- [86] D. Schneidman-Duhovny, O. Dror, Y. Inbar, R. Nussinov, and H. J. Wolfson, "PharmaGist: a webserver for ligand-based pharmacophore detection," *Nucleic Acids Research*, vol. 36, no. Web Server, pp. W223–W228, 2008.
- [87] X. Liu, S. Ouyang, B. Yu et al., "PharmMapper server: a web server for potential drug target identification using



- pharmacophore mapping approach,” *Nucleic Acids Research*, vol. 38, suppl\_2, pp. W609–W614, 2010.
- [88] D. R. Koes and C. J. Camacho, “Pharmer: efficient and exact pharmacophore search,” *Journal of Chemical Information and Modeling*, vol. 51, no. 6, pp. 1307–1314, 2011.
- [89] S. L. Dixon, A. M. Smondyrev, and S. N. Rao, “PHASE: a novel approach to pharmacophore modeling and 3D database searching,” *Chemical Biology & Drug Design*, vol. 67, no. 5, pp. 370–372, 2006.
- [90] D. R. Koes and C. J. Camacho, “ZINCPharmer: pharmacophore search of the ZINC database,” *Nucleic Acids Research*, vol. 40, no. W1, pp. W409–W414, 2012.
- [91] G. Jones, P. Willett, and R. C. Glen, “A genetic algorithm for flexible molecular overlay and pharmacophore elucidation,” *Journal of Computer-Aided Molecular Design*, vol. 9, no. 6, pp. 532–549, 1995.
- [92] J. O. Ebalunode, Z. Ouyang, J. Liang, and W. Zheng, “Novel approach to structure-based pharmacophore search using computational geometry and shape matching techniques,” *Journal of Chemical Information and Modeling*, vol. 48, no. 4, pp. 889–901, 2008.
- [93] M. P. A. Sanders, S. Verhoeven, C. de Graaf et al., “Snooker: a structure-based pharmacophore generation tool applied to class a GPCRs,” *Journal of Chemical Information and Modeling*, vol. 51, no. 9, pp. 2277–2292, 2011.
- [94] J. Chen and L. Lai, “Pocket v.2: Further Developments on receptor-based pharmacophore modeling,” *Journal of Chemical Information and Modeling*, vol. 46, no. 6, pp. 2684–2691, 2006.
- [95] N. J. Richmond, C. A. Abrams, P. R. N. Wolohan, E. Abrahamian, P. Willett, and R. D. Clark, “GALAHAD: 1. Pharmacophore identification by hypermolecular alignment of ligands in 3D,” *Journal of Computer-Aided Molecular Design*, vol. 20, no. 9, pp. 567–587, 2006.
- [96] A. Cherkasov, E. N. Muratov, D. Fourches et al., “QSAR modeling: where have you been? Where are you going to?,” *Journal of Medicinal Chemistry*, vol. 57, no. 12, pp. 4977–5010, 2014.
- [97] H. M. Patel, M. N. Noolvi, P. Sharma et al., “Quantitative structure–activity relationship (QSAR) studies as strategic approach in drug discovery,” *Medicinal Chemistry Research*, vol. 23, no. 12, pp. 4991–5007, 2014.
- [98] H. Li, J. Sutter, and R. Hoffmann, “Hypo Gen: an automated system for generating 3D predictive pharmacophore models,” *Pharmacophore Perception, Development, and Use in Drug Design*, vol. 2, p. 171, 2000.
- [99] R. D. Cramer, D. E. Patterson, and J. D. Bunce, “Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins,” *Journal of the American Chemical Society*, vol. 110, no. 18, pp. 5959–5967, 1988.
- [100] G. Klebe, U. Abraham, and T. Mietzner, “Molecular similarity indices in a comparative analysis (CoMSIA) of drug molecules to correlate and predict their biological activity,” *Journal of Medicinal Chemistry*, vol. 37, no. 24, pp. 4130–4146, 1994.
- [101] C. W. Yap, “PaDEL-descriptor: an open source software to calculate molecular descriptors and fingerprints,” *Journal of Computational Chemistry*, vol. 32, no. 7, pp. 1466–1474, 2011.
- [102] A. R. Katritzky, R. Petrukhin, H. Yang, and M. Karelson, *CODESSA PRO*, User’s manual. University of Florida, Florida, 2002.
- [103] L. Wang, J. Ding, L. Pan, D. Cao, H. Jiang, and X. Ding, “Artificial intelligence facilitates drug design in the big data era,” *Chemometrics and Intelligent Laboratory Systems*, vol. 194, article 103850, 2019.
- [104] L. Patel, T. Shukla, X. Huang, D. W. Ussery, and S. Wang, “Machine learning methods in drug discovery,” *Molecules*, vol. 25, no. 22, p. 5277, 2020.
- [105] C. Cortes and V. Vapnik, “Support-vector networks,” *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995.
- [106] L. Breiman, “Random forests,” *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.
- [107] C. Sammut and G. I. Webb, *Naïve Bayes BT-Encyclopedia of Machine Learning*, Springer Science & Business Media, 2010.
- [108] P. P. Roy, J. T. Leonard, and K. Roy, “Exploring the impact of size of training sets for the development of predictive QSAR models,” *Chemometrics and Intelligent Laboratory Systems*, vol. 90, no. 1, pp. 31–42, 2008.
- [109] T. R. Stouch, J. R. Kenyon, S. R. Johnson, X.-Q. Chen, A. Doweiko, and Y. Li, “In silico ADME/Tox: why models fail,” *Journal of Computer-Aided Molecular Design*, vol. 17, no. 2/4, pp. 83–92, 2003.
- [110] J. Ma, R. P. Sheridan, A. Liaw, G. E. Dahl, and V. Svetnik, “Deep neural nets as a method for quantitative structure–activity relationships,” *Journal of Chemical Information and Modeling*, vol. 55, no. 2, pp. 263–274, 2015.
- [111] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.
- [112] M. Davies, M. Nowotka, G. Papadatos et al., “ChEMBL web services: streamlining access to drug discovery data and utilities,” *Nucleic Acids Research*, vol. 43, no. W1, pp. W612–W620, 2015.
- [113] V. Rotemberg, A. Halpern, S. Dusza, and N. C. Codella, “The role of public challenges and data sets towards algorithm development, trust, and use in clinical practice,” *Seminars in Cutaneous Medicine and Surgery*, vol. 38, no. 1, pp. E38–E42, 2019.
- [114] N. Zhu, D. Zhang, W. Wang et al., “A novel coronavirus from patients with pneumonia in China, 2019,” *New England Journal of Medicine*, vol. 382, no. 8, pp. 727–733, 2020.
- [115] C. Gil, T. Ginex, I. Maestro et al., “COVID-19: Drug targets and potential treatments,” *Journal of Medicinal Chemistry*, vol. 63, no. 21, pp. 12359–12386, 2020.
- [116] S. Ullrich and C. Nitsche, “The SARS-CoV-2 main protease as drug target,” *Bioorganic & Medicinal Chemistry Letters*, vol. 30, no. 17, article 127377, 2020.
- [117] L. Zhang, D. Lin, X. Sun et al., “Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved  $\alpha$ -ketoamide inhibitors,” *Science*, vol. 368, no. 6489, pp. 409–412, 2020.
- [118] J. Osipiuk, S.-A. Azizi, S. Dvorkin et al., “Structure of papain-like protease from SARS-CoV-2 and its complexes with non-covalent inhibitors,” *Nature Communications*, vol. 12, no. 1, pp. 1–9, 2021.
- [119] D. Shin, R. Mukherjee, D. Grewe et al., “Papain-like protease regulates SARS-CoV-2 viral spread and innate immunity,” *Nature*, vol. 587, no. 7835, pp. 657–662, 2020.
- [120] J. Ahmad, S. Ikram, F. Ahmad, I. U. Rehman, and M. Mushtaq, “SARS-CoV-2 RNA Dependent RNA polymerase (RdRp) - A drug repurposing study,” *Heliyon*, vol. 6, no. 7, article e04502, 2020.



- [121] Y. Wang, V. Anirudhan, R. Du, Q. Cui, and L. Rong, "RNA-dependent RNA polymerase of SARS-CoV-2 as a therapeutic target," *Journal of Medical Virology*, vol. 93, no. 1, pp. 300–310, 2021.
- [122] R. Ulferts and J. Ziebuhr, "Nidovirus ribonucleases: structures and functions in viral replication," *RNA Biology*, vol. 8, no. 2, pp. 295–304, 2011.
- [123] Y. Kim, H. Y. Hwang, E. S. Ji, J. Y. Kim, J. S. Yoo, and H. J. Kwon, "Activation of mitochondrial TUFM ameliorates metabolic dysregulation through coordinating autophagy induction," *Communications Biology*, vol. 4, no. 1, pp. 1–11, 2021.
- [124] M. A. El Hassab, T. M. Ibrahim, S. T. Al-Rashood, A. Alharbi, R. O. Eskandrani, and W. M. Eldehna, "In silico identification of novel SARS-CoV-2 2'-O-methyltransferase (nsp16) inhibitors: structure-based virtual screening, molecular dynamics simulation and MM-PBSA approaches," *Journal of Enzyme Inhibition and Medicinal Chemistry*, vol. 36, no. 1, pp. 727–736, 2021.
- [125] T. Shu, M. Huang, D. Wu et al., "SARS-Coronavirus-2 Nsp 13 possesses NTPase and RNA helicase activities that can be inhibited by bismuth salts," *Virologica Sinica*, vol. 35, no. 3, pp. 321–329, 2020.
- [126] G. Ragia and V. G. Manolopoulos, "Inhibition of SARS-CoV-2 entry through the ACE2/TMPRSS2 pathway: a promising approach for uncovering early COVID-19 drug therapies," *European Journal of Clinical Pharmacology*, vol. 76, no. 12, pp. 1623–1630, 2020.
- [127] C. Wu, M. Zheng, Y. Yang et al., "Furin: a potential therapeutic target for COVID-19," *Science*, vol. 23, no. 10, article 101642, 2020.
- [128] T. Liu, S. Luo, P. Libby, and G.-P. Shi, "Cathepsin L-selective inhibitors: a potentially promising treatment for COVID-19 patients," *Pharmacology & Therapeutics*, vol. 213, article 107587, 2020.
- [129] C. Selvaraj, D. C. Dinesh, U. Panwar, R. Abhirami, E. Boura, and S. K. Singh, "Structure-based virtual screening and molecular dynamics simulation of SARS-CoV-2 guanine-N7 methyltransferase (nsp 14) for identifying antiviral inhibitors against COVID-19," *Journal of Biomolecular Structure and Dynamics*, pp. 1–12, 2021.
- [130] T. Gao, Y. Gao, X. Liu et al., "Identification and functional analysis of the SARS-CoV-2 nucleocapsid protein," *BMC Microbiology*, vol. 21, no. 1, p. 58, 2021.
- [131] B. R. Beck, B. Shin, Y. Choi, S. Park, and K. Kang, "Predicting commercially available antiviral drugs that may act on the novel coronavirus (SARS-CoV-2) through a drug-target interaction deep learning model," *Computational and Structural Biotechnology Journal*, vol. 18, pp. 784–790, 2020.
- [132] A. A. Elfiky, "SARS-CoV-2 RNA dependent RNA polymerase (RdRp) targeting: an in silico perspective," *Journal of Biomolecular Structure and Dynamics*, vol. 39, no. 9, pp. 3204–3212, 2021.
- [133] A. D. Elmezayen and K. Yelekçi, "Homology modeling and in silico design of novel and potential dual-acting inhibitors of human histone deacetylases HDAC5 and HDAC9 isozymes," *Journal of Biomolecular Structure and Dynamics*, pp. 1–19, 2020.
- [134] S. Das, S. Sarmah, S. Lyndem, and A. Singha Roy, "An investigation into the identification of potential inhibitors of SARS-CoV-2 main protease using molecular docking study," *Journal of Biomolecular Structure and Dynamics*, pp. 1–11, 2020.
- [135] A. B. Gurung, M. A. Ali, J. Lee, M. A. Farah, and K. M. Al-Anazi, "Unravelling lead antiviral phytochemicals for the inhibition of SARS-CoV-2 M<sup>pro</sup> enzyme through in silico approach," *Life Sciences*, vol. 255, article 117831, 2020.
- [136] R. S. Joshi, S. S. Jagdale, S. B. Bansode et al., "Discovery of potential multi-target-directed ligands by targeting host-specific SARS-CoV-2 structurally conserved main protease," *Journal of Biomolecular Structure and Dynamics*, vol. 39, no. 9, pp. 1–16, 2020.
- [137] H. M. Wahedi, S. Ahmad, and S. W. Abbasi, "Stilbene-based natural compounds as promising drug candidates against COVID-19," *Journal of Biomolecular Structure and Dynamics*, vol. 39, no. 9, pp. 3225–3234, 2021.
- [138] S. A. Khan, K. Zia, S. Ashraf, R. Uddin, and Z. Ul-Haq, "Identification of chymotrypsin-like protease inhibitors of SARS-CoV-2 via integrated computational approach," *Journal of Biomolecular Structure and Dynamics*, vol. 39, no. 7, pp. 2607–2616, 2020.
- [139] C. Mouffouk, S. Mouffouk, S. Mouffouk, L. Hambaba, and H. Haba, "Flavonols as potential antiviral drugs targeting SARS-CoV-2 proteases (3CL<sup>pro</sup> and PL<sup>pro</sup>), spike protein, RNA-dependent RNA polymerase (RdRp) and angiotensin-converting enzyme II receptor (ACE2)," *European Journal of Pharmacology*, vol. 891, article 173759, 2021.
- [140] D. Umesh, C. Kundu, C. Selvaraj, S. K. Singh, and V. K. Dubey, "Identification of new anti-nCoV drug chemical compounds from Indian spices exploiting SARS-CoV-2 main protease as target," *Journal of Biomolecular Structure and Dynamics*, vol. 39, no. 9, pp. 3428–3434, 2021.
- [141] I. Abdelli, F. Hassani, S. Bekkel Brikci, and S. Ghalem, "In silico study the inhibition of angiotensin converting enzyme 2 receptor of COVID-19 by *Ammoides verticillata* components harvested from western Algeria," *Journal of Biomolecular Structure and Dynamics*, vol. 39, no. 9, pp. 3263–3276, 2021.
- [142] K. Al-Khafaji, D. Al-Duhaidahawi, and T. Taskin Tok, "Using integrated computational approaches to identify safe and rapid treatment for SARS-CoV-2," *Journal of Biomolecular Structure and Dynamics*, vol. 39, no. 9, pp. 1–9, 2020.
- [143] K. A. Peele, C. Potla Durthi, T. Srihansa et al., "Molecular docking and dynamic simulations for antiviral compounds against SARS-CoV-2: a computational study," *Informatics in Medicine Unlocked*, vol. 19, article 100345, 2020.
- [144] J. Wang, "Fast identification of possible drug treatment of coronavirus disease-19 (COVID-19) through computational drug repurposing study," *Journal of Chemical Information and Modeling*, vol. 60, no. 6, pp. 3277–3286, 2020.
- [145] L. Mittal, A. Kumari, M. Srivastava, M. Singh, and S. Asthana, "Identification of potential molecules against COVID-19 main protease through structure-guided virtual screening approach," *Journal of Biomolecular Structure and Dynamics*, vol. 39, no. 10, pp. 3662–3680, 2021.
- [146] M. Kandeel and M. Al-Nazawi, "Virtual screening and repurposing of FDA approved drugs against COVID-19 main protease," *Life Sciences*, vol. 251, article 117627, 2020.
- [147] M. Tahir ul Qamar, S. M. Alqahtani, M. A. Alamri, and L. L. Chen, "Structural basis of SARS-CoV-2 3CL<sup>pro</sup> and anti-COVID-19 drug discovery from medicinal plants," *Journal of Pharmaceutical Analysis*, vol. 10, no. 4, pp. 313–319, 2020.
- [148] R. Islam, R. Parves, A. S. Paul et al., "A molecular modeling approach to identify effective antiviral phytochemicals against the main protease of SARS-CoV-2," *Journal of*

- Biomolecular Structure and Dynamics*, vol. 39, no. 9, pp. 1–12, 2020.
- [149] S. Beura and P. Chetti, “In-silico strategies for probing chloroquine based inhibitors against SARS-CoV-2,” *Journal of Biomolecular Structure and Dynamics*, vol. 39, no. 10, pp. 3747–3759, 2021.
- [150] S. Mahanta, P. Chowdhury, N. Gogoi et al., “Potential antiviral activity of approved repurposed drug against main protease of SARS-CoV-2: an in silico based approach,” *Journal of Biomolecular Structure and Dynamics*, vol. 39, no. 10, pp. 3802–3811, 2021.
- [151] S. K. Enmozhi, K. Raja, I. Sebastine, and J. Joseph, “Andrographolide as a potential inhibitor of SARS-CoV-2 main protease: an in silico approach,” *Journal of Biomolecular Structure and Dynamics*, vol. 39, no. 9, pp. 3092–3098, 2021.
- [152] S. Kumar, P. P. Sharma, U. Shankar et al., “Discovery of new hydroxyethylamine analogs against 3CLproProtein target of SARS-CoV-2: molecular docking, molecular dynamics Simulation, and structure-activity relationship studies,” *Journal of Chemical Information and Modeling*, vol. 60, no. 12, pp. 5754–5770, 2020.
- [153] K. G. Arun, C. S. Sharanya, J. Abhithaj, D. Francis, and C. Sadasivan, “Drug repurposing against SARS-CoV-2 using E-pharmacophore based virtual screening, molecular docking and molecular dynamics with main protease as the target,” *Journal of Biomolecular Structure and Dynamics*, pp. 1–12, 2020.
- [154] D. Gentile, V. Patamia, A. Scala, M. T. Sciortino, A. Piperno, and A. Rescifina, “Putative Inhibitors of SARS-CoV-2 Main Protease from A Library of Marine Natural Products: A Virtual Screening and Molecular Modeling Study,” *Marine Drugs*, vol. 18, no. 4, p. 225, 2020.
- [155] S. A. Amin, K. Ghosh, S. Gayen, and T. Jha, “Chemical-informatics approach to COVID-19 drug discovery: Monte Carlo based QSAR, virtual screening and molecular docking study of some in-house molecules as papain-like protease (PLpro) inhibitors,” *Journal of Biomolecular Structure and Dynamics*, pp. 1–10, 2020.
- [156] K. Ghosh, S. A. Amin, S. Gayen, and T. Jha, “Chemical-informatics approach to COVID-19 drug discovery: exploration of important fragments and data mining based prediction of some hits from natural origins as main protease (Mpro) inhibitors,” *Journal of Molecular Structure*, vol. 1224, article 129026, 2020.
- [157] W. Y. Law, M. R. Asaruddin, S. A. Bhawani, and S. Mohamad, “Pharmacophore modelling of vanillin derivatives, favipiravir, chloroquine, hydroxychloroquine, monolaurin and tetrodotoxin as M pro inhibitors of severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2),” *BMC Research Notes*, vol. 13, no. 1, pp. 1–8, 2020.
- [158] S. Daoud, S. J. Alabed, and L. A. Dahabiyeh, “Identification of potential COVID-19 main protease inhibitors using structure-based pharmacophore approach, molecular docking and repurposing studies,” *Acta Pharmaceutica*, vol. 71, no. 2, pp. 163–174, 2020.
- [159] S. Skariyachan, D. Gopal, S. Chakrabarti et al., “Structural and molecular basis of the interaction mechanism of selected drugs towards multiple targets of SARS-CoV-2 by molecular docking and dynamic simulation studies- deciphering the scope of repurposed drugs,” *Computers in Biology and Medicine*, vol. 126, article 104054, 2020.
- [160] K. Uzunova, E. Filipova, V. Pavlova, and T. Vekov, “Insights into antiviral mechanisms of remdesivir, lopinavir/ritonavir and chloroquine/hydroxychloroquine affecting the new SARS-CoV-2,” *Biomedicine & Pharmacotherapy*, vol. 131, article 110668, 2020.
- [161] S. K. Singh, A. K. Upadhyay, and M. S. Reddy, “Screening of potent drug inhibitors against SARS-CoV-2 RNA polymerase: an in silico approach,” *3 Biotech*, vol. 11, no. 2, pp. 1–13, 2021.
- [162] S. Gowrishankar, S. Muthumanickam, A. Kamaladevi et al., “Promising phytochemicals of traditional Indian herbal steam inhalation therapy to combat COVID-19 - an in silico study,” *Food and Chemical Toxicology*, vol. 148, p. 111966, 2021.
- [163] M. A. A. Ibrahim, A. H. M. Abdelrahman, K. S. Allemailem, A. Almatroudi, M. F. Moustafa, and M.-E. F. Hegazy, “In silico evaluation of prospective anti-COVID-19 drug candidates as potential SARS-CoV-2 main protease inhibitors,” *The Protein Journal*, pp. 1–14, 2021.
- [164] A. T. Onawole, K. O. Sulaiman, T. U. Kolapo, F. O. Akinde, and R. O. Adegoke, “COVID-19: CADD to the rescue,” *Virus Research*, vol. 285, article 198022, 2020.
- [165] S. C. Basak and L. B. Kier, “COVID-19 pandemic: how can computer-assisted methods help to rein in this global menace?,” *Current Computer-Aided Drug Design*, vol. 17, no. 1, p. 1, 2021.
- [166] T. Sharma, M. Abohashrh, M. H. Baig et al., “Screening of drug databank against WT and mutant main protease of SARS-CoV-2: towards finding potential compound for repurposing against COVID-19,” *Saudi Journal of Biological Sciences*, vol. 28, no. 5, pp. 3152–3159, 2021.
- [167] P. K. Ojha, S. Kar, J. G. Krishna, K. Roy, and J. Leszczynski, “Therapeutics for COVID-19: from computation to practices—where we are, where we are heading to,” *Molecular Diversity*, vol. 25, pp. 625–659, 2020.
- [168] A. Keshavarzi Arshadi, J. Webb, M. Salem et al., “Artificial intelligence for COVID-19 drug discovery and vaccine development,” *Frontiers in Artificial Intelligence*, vol. 3, p. 65, 2020.