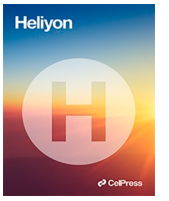
Heliyon 10 (2024) e24209

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon



Review article

Recent advances in chemometric modelling of inhibitors against SARS-CoV-2

Qianqian Wang [a], Xinyi Lu [a], Runqing Jia [b,*], Xinlong Yan [b,***], Jianhua Wang [c,****], Lijiao Zhao [a], Rugang Zhong [a], Guohui Sun [a,**]

[a] *Beijing Key Laboratory of Environmental and Viral Oncology, Faculty of Environment and Life, Beijing University of Technology, Beijing 100124, PR China*
[b] *Department of Biology, Faculty of Environment and Life, Beijing University of Technology, Beijing 100124, PR China*
[c] *Beijing Municipal Key Laboratory of Child Development and Nutriomics, Translational Medicine Laboratory, Capital Institute of Pediatrics, Beijing 100124, PR China*

A R T I C L E I N F O

Keywords:
COVID-19
Candidate drugs
3CLpro
hACE2/S protein
QSAR

A B S T R A C T

The outbreak of the novel coronavirus disease 2019 (COVID-19), caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), has caused great harm to all countries worldwide. This disease can be prevented by vaccination and managed using various treatment methods, including injections, oral medications, or aerosol therapies. However, the selection of suitable compounds for the research and development of anti-SARS-CoV-2 drugs is a daunting task because of the vast databases of available compounds. The traditional process of drug research and development is time-consuming, labour-intensive, and costly. The application of chemometrics can significantly expedite drug R&D. This is particularly necessary and important for drug development against pandemic public emergency diseases, such as COVID-19. Through various chemometric techniques, such as quantitative structure-activity relationship (QSAR) modelling, molecular docking, and molecular dynamics (MD) simulations, compounds with inhibitory activity against SARS-CoV-2 can be quickly screened, allowing researchers to focus on the few prioritised candidates. In addition, the ADMET properties of the screened candidate compounds should be further explored to promote the successful discovery of anti-SARS-CoV-2 drugs. In this case, considerable time and economic costs can be saved while minimising the need for extensive animal experiments, in line with the 3R principles. This paper focuses on recent advances in chemometric modelling studies of COVID-19-related inhibitors, highlights current limitations, and outlines potential future directions for development.

* Corresponding authors.
** Corresponding author.
*** Corresponding authors.
**** Corresponding author.
E-mail addresses: wangqq1998@emails.bjut.edu.cn (Q. Wang), xinyi_lu@163.com (X. Lu), runqingjia@bjut.edu.cn (R. Jia), yxlong@bjut.edu.cn (X. Yan), wangjianhua@shouer.com.cn (J. Wang), zhaolijiao@bjut.edu.cn (L. Zhao), lifesci@bjut.edu.cn (R. Zhong), sunguohui@bjut.edu.cn (G. Sun).

https://doi.org/10.1016/j.heliyon.2024.e24209
Received 22 August 2023; Received in revised form 2 January 2024; Accepted 4 January 2024
Available online 9 January 2024

1. Introduction

In December 2019, an unprecedented coronavirus disease appeared in Wuhan, Hubei province, China. It spread widely and quickly, and its clinical manifestations are fever, cough, dyspnoea, and symptoms of pneumonia [1]. Simultaneously, diseases with the same symptoms occur in many countries worldwide. On February 11, 2020 the World Health Organization (WHO) named the disease as coronavirus Disease 2019 (COVID-19). The International Committee on Virus Taxonomy officially announced the Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) as the causative agent of COVID-19 [2]. As of September 2023, the cumulative number of confirmed cases worldwide, as counted by the World Health Organization, exceeded 770 million, and the cumulative confirmed deaths exceeded 6.9 million [3]. Furthermore, although the development of vaccines and inoculations has aroused wide social concerns, it appears that vaccinations do not prevent the rapid transmission of SARS-CoV-2, such as the latest Omicron sub-variants BF.7, BA.5.2, BQ.1, BQ.1.1, XBB, XBB.1.5, XBD, and XBF etc [4–7].

SARS-CoV-2 belongs to a large family of coronaviruses, which are RNA viruses with an envelope and linear single-stranded positive strand genome, with a total genome length of ~30 kb [8]. According to genome sequence analysis, SARS-CoV-2 is approximately 80 % similar to SARS-CoV and has the same structural and molecular characteristics as other coronaviruses [9]. As shown in Fig. 1A, the genome of SARS-CoV-2 contains 14 open reading frames (ORFs), in which the first two ORFs, ORF1a and ORF1b (overlapping with a −1 ribosomal frame-shift), account for approximately two-thirds of the genome from the 5′-end, and are translated to the polyproteins pp1a and pp1ab, respectively [10]. The polyproteins pp1a and pp1ab are cleaved proteolytically at 11 sites by two viral proteases (papain-like protease (PLpro) and the main protease (Mpro or 3CLpro), to produce 16 non-structural proteins (Nsp1-Nsp16, although the function of Nsp11 is unknown) that are essential for viral replication and transcription [11]. Among these, Nsp3, also called PLpro, participates in the cleavage of polyproteins to generate Nsp1, Nsp2, and Nsp3 via proteolysis [12]. Nsp5 is a homodimeric cysteine protease commonly known as chymotrypsin-like protease 3CLpro or Mpro [13,14]. During the life cycle of the virus, 3CLpro plays an indispensable role via the cleavage of viral polyproteins to produce multiple Nsps (Nsp4-Nsp16), thus indirectly participating in the generation of viral sub-genomic mRNAs encoding four structural proteins and ultimately completing the reproduction and release of the progeny virus [13,14].

Considering the vital roles of 3CLpro and PLpro in the life cycle of SARS-CoV-2, they are attractive therapeutic targets for COVID-19 drug discovery [15–17]. The remaining one-third of the viral genome at the 3′-end encodes structural and accessory proteins. Accessory proteins lie between these structural genes, including ORF3a, ORF3b, ORF6, ORF7a, ORF7b, ORF8, ORF9b, ORF9c and ORF10 [18–20]. Four main structural proteins are responsible for the assembly of new virions, including spike protein (S), envelope protein (E), membrane protein (M), and nucleocapsid protein (N), which are responsible for host recognition, formation of the viral envelope, viral stability, and interaction with the RNA genome (Fig. 1B) [21]. In particular, the sequence composition of the S protein

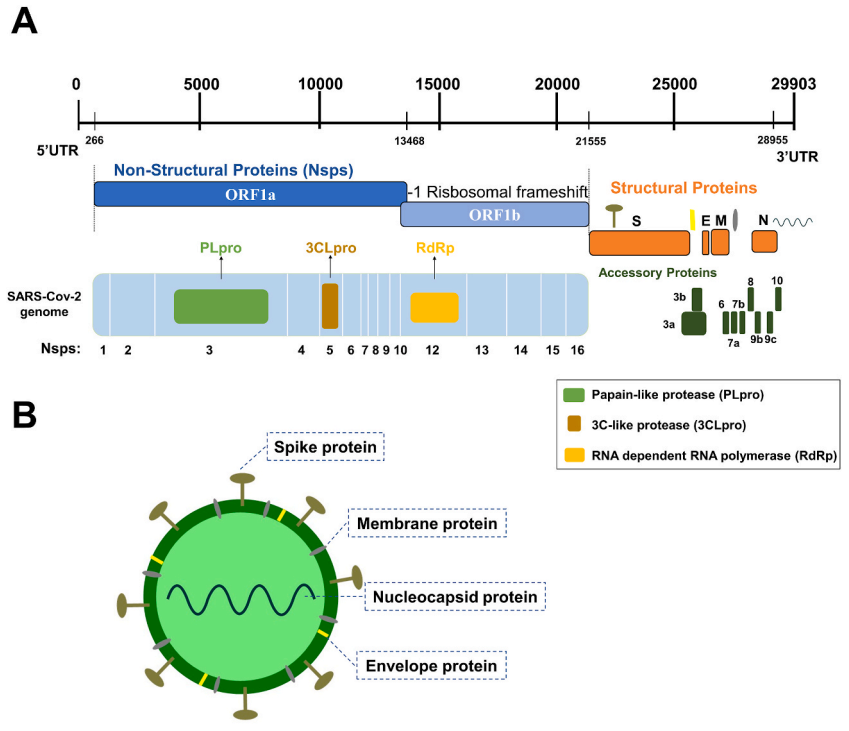


Fig. 1. (**A**) Genome organisation of the SARS-CoV-2. The viral genome encodes 16 non-structural proteins (Nsps) required for replication/transcription along with the structural proteins required for the assembly of new virions. (**B**) Structural diagram of SARS-CoV-2, including four structural proteins, S protein, E protein, M protein and N protein.

encoded by the SARS-CoV-2 genome is the most different from that of SARS and MERS (Middle East Respiratory Syndrome) [11,22]. Part of the S protein spike can extend and attach to the human angiotensin-converting enzyme (hACE2) protein, allowing viruses to enter cells and endanger human health (Fig. 2) [11,20]. Over time, SARS-CoV-2 has evolved into new mutants, such as alpha, beta, gamma, delta, Omicron, and various subvariants [4–7,23,24]. These mutations change the structure and characteristics of SARS-CoV-2 to varying degrees, bringing more challenges to the development of related inhibitor drugs [18].

Based on the physiological and biological characteristics of SARS-CoV-2, researchers from academic and medical circles have developed relevant drugs for the treatment and prevention, including vaccines for the prevention and control of COVID-19 transmission and small-molecule oral drugs for treatment [25,26]. For example, the COVID-19 therapeutic drug Paxlovid (PF-07321332/Nirmatrelvir in combination with the CYP3A4 inhibitor ritonavir), developed by Pfizer in the United States, is the result of further processing and optimisation of the PF-00835231 inhibitor developed during the SARS period [25,27–29]. It was the first approved anti-SARS-CoV-2 drug by the American Food and Drug Administration (FDA), bringing a new dawn for the treatment of COVID-19 [30,31]. On February 11, 2022 the National Medical Products Administration officially approved the import registration of Paxlovid for the treatment of adult COVID-19 patients in China [32]. However, recent studies found that after taking Paxlovid treatment, COVID-19 patients reappeared positive a few days after the disappearing of symptoms, and the virus load was much higher than that of patients who had not taken Paxlovid before, this phenomenon is called "Paxlovid rebound" [33–35]. Many mRNA vaccine

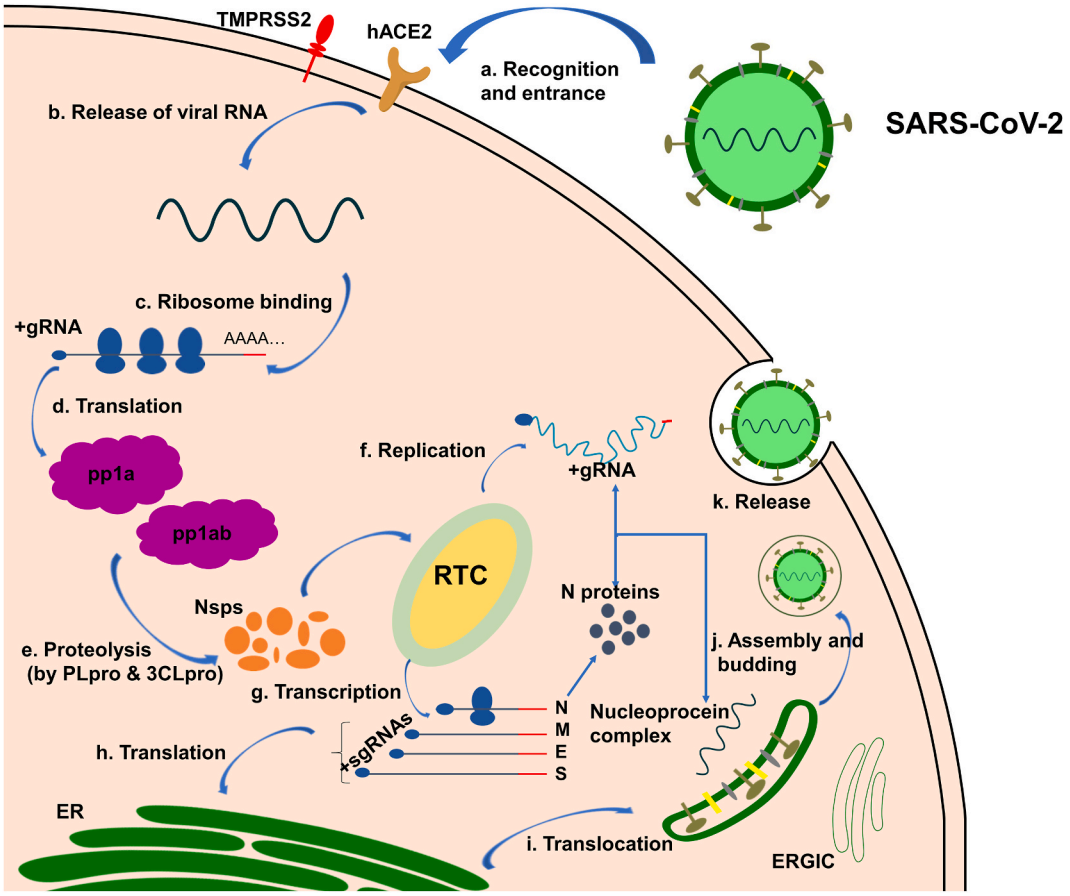


Fig. 2. The life cycle of SARS-CoV-2, including viral entry, replication and transcription, assembly and release. During the viral infection, SARS-CoV-2 injects its genome into the host cell via endosomes or direct fusion of the viral envelope with the host cell membrane that mediated by the binding of S protein to the human ACE2 at the cell surface (step **a**). After entering into the host cell, the viral RNA is translated by the host ribosomes (steps **b-d**). The translated products, polyproteins pp1a and pp1ab, are proteolytically cleaved into non-structural proteins Nsp1-16 by the viral proteases, PLpro and 3CLpro (step **e**). Several Nsps (Nsp2-16) together with other factors to form a replication-transcription complex (RTC) within the infected host cell. Nsp2-11 are supposed to play a supporting role, and Nsp12-16 provide the required enzymatic function for viral genome replication/transcription inside the RTC. The RNA (+) strand first gets replicated to the RNA (−) strand and then the negative-strand is used either for replication to the RNA (+) strand for new virion assembly (step **f**) or transcription of sub-genomic mRNAs (step **g**). These sub-genomic mRNAs are translated to the structural proteins S, M, E, N, and the accessory proteins (step **h**). The S, M, and E proteins enter the endoplasmic reticulum (ER), and the N protein attaches to the genomic RNA (+) strand to produce nucleoprotein complex. The nucleoprotein complex and the structural proteins move to the ER-Golgi intermediate compartment (ERGIC) where the virions assemble, mature, and bud off from the Golgi in the form of small vesicles (steps **i-j**). These vesicles travel to the host cell membrane where they are released into the extracellular region through exocytosis (step **k**).

recipients infected with Omicron also experienced a rebound after receiving Paxlovid treatment, and the symptoms, viral load, viral duration, and other indicators were more serious than those in the acute infection period before recovery, resulting in the impairment of COVID-19 specific immunity [36]. Therefore, there is an urgent need to develop new and efficient oral drugs for treating COVID-19, especially easily transmissible SARS-CoV-2 variants, such as Omicron. There may already be highly potent anti-COVID-19 compounds in the library of various chemicals or drugs (e.g. DrugBank, Zinc, ChEMBL, PubChem, ChemSpider, and ToxNet), but they have not been found to date. However, traditional drug development pipelines are time-consuming (~12 years on average) and costly (~2.6 billion USD) [37], and do not meet the emergency requirements of public health emergencies such as COVID-19. Thus, it is not difficult to understand the dilemma of discovering more potent compounds for COVID-19 treatment within a limited time and with available human resources. In view of this situation, rapid and efficient chemometric screening methods, such as quantitative structure-activity/toxicity relationship (QSAR/QSTR), machine learning/artificial intelligence, molecular docking and molecular dynamics (MD), are particularly important, may lead to the discovery of "drug repurposing" and "completely novel drugs" [37–41].

QSAR methodology refers to the establishment of a linear or non-linear relationship model between the structural features of a molecule and its biological and physicochemical properties using mathematical methods [42]. Compared with traditional drug development, with the aid of *in silico* screening methods such as QSAR, not only can the R&D cycle be significantly shortened, but economic costs will also be saved, as well as avoiding a large number of unnecessary animal experiments [42,43]. QSAR methodology is one of the hot spots in the field of pharmaceutical science and has been widely applied to the discovery of antiviral drugs, including anti-SARS-CoV-2 agents [40].

When a molecule is predicted to have obvious inhibitory potency against the target viral protein by the QSAR method, this does not mean that the molecule can be successfully developed into a clinical antiviral drug. For example, other factors, such as binding stability, affinity, absorption, distribution, metabolism, excretion, and toxicity (ADMET), also influence druggability. This means that after QSAR predictions, a series of subsequent validations must be carried out, such as molecular docking and MD simulations. Docking can explore the interaction mode of the protein-ligand complex, providing the binding free energy (docking scores) and key amino acid residues responsible for the inhibitory activity. MD simulations can calculate the binding free energy of the ligand compound and the target protein using various energy-contributing components such as solvent-accessible surface area (SASA), polar solvation,

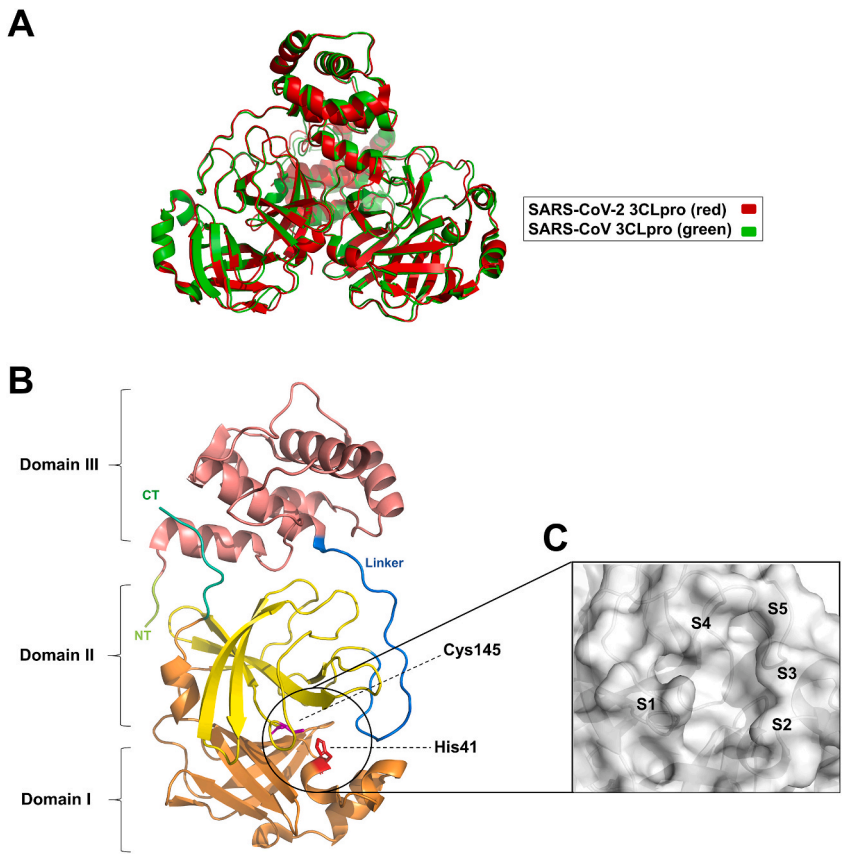


Fig. 3. (**A**) The structure of SARS-CoV-2 3CLpro (red, PDB: 6XHU) and SARS-CoV 3CLpro (green, PDB: 1UJ1); (**B**) Three structural domains (domain I: orange, domain II: yellow, domain III: pale red) of SARS-CoV-2 3CLpro protomer (PDB code: 6XHU), and two key amino acid residues His41 and Cys145. CT, Carbon terminal; NT, Nitrogen terminal; (**C**) The surface representation for the catalytic pocket (sub-pockets: S1–S5) of SARS-CoV-2 3CLpro. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

Table 1
Summary of the QSAR models on the inhibitors of 3CLpro and hACE2/Spike protein.

Target	End point	N_{total}	N_{tr}	N_{test}	Algorithm	Outliers	Parameter	AD analysis	True external set	Ref.
SARS-CoV 3CLpro	IC_{50} (μM)	34	26	8	MLR	–	$R^2 = 0.748$ $Q^2 = 0.628$ $R^2_{test} = 0.723$	No	Yes	[58]
	IC_{50} (μM)	104	78	26	PLS	0	$R^2 = 0.756$ $Q^2 = 0.708$ $Q^2_{Fn} = 0.752$	Yes	Yes	[44]
	IC_{50} (μM)	40	32	8	MLR	1	$R^2 = 0.838$ $R^2_{test} = 0.735$ $Q^2_{LOO} = 0.757$	Yes	Yes	[59]
	IC_{50} (μM)	101	85	16	CNN	–	$R^2 = 0.98$ $R^2_{test} = 0.84$	No	Yes	[62]
	IC_{50} (μM)	21	14	7	MLR	0	$R^2 = 0.91$ $Q^2_{LOO} = 0.78$ $R^2_{test} = 0.85$ $CCC_{test} = 0.92$	Yes	Yes	[63]
	IC_{50} (μM)	81	59	22	Monte Calo	0	$R^2 = 0.96$ $R^2_{test} = 0.92$	No	Yes	[64]
	Ki (nM)	62	50	12	MLR	3	$R^2 = 0.898$ $R^2_{test} = 0.799$ $Q^2_{LOM} = 0.848$	Yes	Yes	[65]
	IC_{50} (nM)	84	63	21	Monte Calo	–	$R^2 = 0.9314$ $Q^2 = 0.9271$ $R^2_{test} = 0.9243$ $Q^2_{test} = 0.8986$	No	Yes	[66]
	IC_{50} (nM)	73	51	22	MLR	1	$R^2 = 0.907$ $Q^2_{LOO} = 0.866$ $R^2_{test} = 0.517$	Yes	No	[67]
	IC_{50} (μM)	220	176	44	machine learning	–	$AUROC_{tr} = 0.998$ $AUROC_{test} = 0.914$	No	Yes	[68]
	IC_{50} (mg/L)	110	90	20	COMSIA	–	$R^2 = 0.74$ $Q^2 = 0.54$ $R^2_{test} = 0.71$	No	Yes	[69]
SARS-CoV-2 3CLpro	IC_{50} (nM)	190	144	46	PLS	1	$R^2 = 0.789$ $Q^2 = 0.741$ $Q^2_{test\text{-}F1} = 0.787$ $Q^2_{test\text{-}F2} = 0.786$	Yes	Yes	[70]
	IC_{50} (μM)	33	25	6	Bayesian classification	–	$ROC_{tr} = 0.747$ $ACC_{tr} = 0.840$ $ROC_{test} = 1.000$ $ACC_{test} = 0.833$	No	Yes	[72]
	IC_{50} (μM)	46	37	11	CoMFA-PLS	–	$R^2 = 0.97$ $Q^2_{LOO} = 0.81$ $R^2_{test} = 0.95$	No	Yes	[73]
					CoMSIA-PLS		$R^2 = 0.94$ $Q^2_{LOO} = 0.76$ $R^2_{test} = 0.91$			
	IC_{50} (μM)	69	51	18	SW -MLR	–	$R^2 = 0.703$ $R^2_{test} = 0.565$	No	Yes	[74]
					GA-MLR		$R^2 = 0.740$ $R^2_{test} = 0.654$			
					SW-ANN		$R = 0.872$ $R_{test} = 0.745$			
					GA-ANN		$R = 0.865$ $R_{test} = 0.814$			
					SW-SVM		$R = 0.839$ $R_{test} = 0.812$			
					GA-SVM		$R = 0.913$ $R_{test} = 0.814$			
					HQSAR-PLS		$R^2 = 0.867$ $R^2_{test} = 0.715$			
	IC_{50} (μM)	55	39	16	Monte Calo	–	$R^2 = 0.9203$ $Q^2 = 0.8508$	No	Yes	[75]
					MLR	1	$R^2 = 0.9288$ $Q^2_{LOO} = 0.8902$ $R^2_{test} = 0.8558$	Yes		
hACE2/S protein	Active /inactive	1747	80 %	20 %	CNN	–	$FP = 2.7$ % $FN = 9.6$ %	No	Yes	[89]

(*continued on next page*)

Table 1 (*continued*)

Target	End point	N_{total}	N_{tr}	N_{test}	Algorithm	Outliers	Parameter	AD analysis	True external set	Ref.
	Docking score	166	–	–	–	–	$Q^2 < 0.45$ $R^2 > 0.7$	–	–	[90]
	Ki (nM)	31	18	13	3D-QSAR	–	$R^2 = 0.984$ $R^2_{test} = 0.652$	No	Yes	[91]
	IC_{50} (μM)	21	17	4	CoMSIA-PLS	–	$R^2 = 0.962$ $Q^2 = 0.65$	No	No	[94]

electrostatic energy, and van der Waals energy [44]. Therefore, the prediction of pharmacokinetics and bioavailability is indispensable. For example, the international standard "Lipinski's principle of five" of an orally active drug includes [45,46]: (1) the number of hydrogen donors (the number of hydrogen atoms connected to N and O) is less than 5; (2) the number of hydrogen receptors (the number of N and O) is less than 10; (3) the relative molecular weight is less than 500; (4) the octanol–water partition coefficient (logP) is less than 5; (5) 10 or fewer rotatable bonds. This rule of thumb describes the molecular properties vital for the pharmacokinetics of a molecule in the human body, including its ADME properties [47], and has been widely used in the initial screening of compound libraries since its formulation [48]. ADMET properties can determine the efficacy and toxicity of a drug, and are a remarkable indicator for evaluating whether a compound has drug-likeness.

In fact, the above chemometric tools or methods are indispensable for successful screening of few potent candidates, followed by experimental confirmation. In this case, the time from research to market approval can be dramatically reduced, especially being very important in the background of emerging or re-emerging pandemic diseases such as SARS, MERS and COVID-19. In this review, we discuss recent advances in the chemometric modelling of inhibitors against SARS-CoV-2 (including host-targeting therapies) to provide new ideas for the development of anti-SARS-CoV-2 drugs.

2. QSAR models of SARS-CoV-2 main protease inhibitors

2.1. 3CLpro/Mpro

3CLpro (Mpro), which cleaves polyproteins at no fewer than 11 sites to generate Nsps, is a crucial target for antiviral drug development. The SARS-CoV-2 genome has over 80 % identity with SARS-CoV, and genomic sequence analysis indicated that the 3CLpro of SARS-CoV-2 (PDB code: 6XHU) and 3CLpro of SARS-CoV (PDB code: 1U1J) are highly similar (Fig. 3A) [49–51]. Owing to the high sequence similarity between SARS-CoV-2 and SARS-CoV, compounds obtained in the past that exhibited remarkable inhibitory effects on SARS-CoV may also act on SARS-CoV-2. Researchers have shifted their attention to investigating the inhibitory effects of these compounds on the activity of SARS-CoV-2 and have made great progress.

Fig. 3B shows the single-chain structure of SARS-CoV-2 3CLpro (PDB code: 6XHU), as well as the two most popular amino acid residues, His41 and Cys145, which frequently participate in the binding process of small molecules to 3CLpro [52]. As mentioned above, 3CLpro can hydrolyse the pp1a and pp1ab primitive polyproteins from at least 11 conserved cleavage sites, participate in their processing, and form replicase complex [53]. 3CLpro plays an important role in viral replication and proliferation and its deletion is fatal. The active form of 3CLpro comprises homodimers containing two protomers, and each monomer consists of three domains (Fig. 3B). Among them, domain I (Phe10-Tyr99) and domain II (Lys100-Pro182) contain chymotrypsin like skeletons, both domains have six-stranded anti-parallel β-sheet structure that form the substrate-binding site containing the catalytic dyad formed by His41 and Cys145, where the N-terminal finger (domain I, residues 1–7) inserts itself between domains II and III of the adjacent protomer, thus participates the formation of dimer and the S1 subsite pocket of substrate binding. Domain III (Thr198-Val303) contains five anti-parallel α helical structure, and connects with domain II through a long ring region (residues 183–198) (Fig. 3B) [19]. The C-terminus of domain III is also involved in dimerization via a salt-bridge interaction between Glu290 of a protomer and Arg4 from its adjacent protomer [14,19]. Existing 3CLpro inhibitors can affect the replication of viral RNA by blocking polymerisation between domains I and II in the two protomers [19,53]. As shown in Fig. 3C, the catalytic site of 3CLpro is located at the intersection of domains I and II and can be divided into five subpockets: S1, S2, S3, S4, and S5 [54]. 3CLpro-targeting inhibitors can specifically bind to this active site to interfere with protein function [54,55].

2.2. QSAR studies of the 3CLpro inhibitors

The homology between the 3CLpro of SARS-CoV and SARS-CoV-2 is as high as 96 % [56]. The 3CLpro sequence of SARS-CoV-2 is only 12 of 306 residues different from that of SARS-CoV which can be used as a homologous target for drug screening and repurposing [57]. Many compounds that inhibit the 3C-like protease of SARS-CoV also inhibit its main protease activity of SARS-CoV-2. Therefore, many scholars consider this a starting point to collect a large number of such compounds to establish predictive models, which lay a good foundation for the development and identification of new drugs for the treatment of COVID-19 [40]. A summary of the QSAR modelling of the 3CLpro inhibitors is shown in Table 1, which contains the response endpoint, data division, modelling algorithm, and other important information.

For example, Khan et al. constructed a 2D-QSAR model to predict the inhibitory activity of small-molecule carboxamides against

the SARS-CoV 3CLpro [58]. The modelling dataset contained 34 small-molecule carboxamides consisting of two series of compounds: N-(benzo [1–3] triazol-1-yl)-N-(benzyl) acetamido) phenyl) carboxamides and N-(*tert*-butyl)-2-(N-arylamido)-2-(pyridin-3-yl) acetamides. The best QSAR model ($R^2 = 0.748$, $Q^2 = 0.628$, $R^2_{test} = 0.723$) contained four molecular descriptors: SsssCH (sum of atom-type E-State: >CH-), minHBint3 (minimum E-State descriptor of strength for potential hydrogen bonds of path length 3), ETA_Eta_B_RC (a measure of branching), and CrripenLogP. The model equation ($pIC_{50} = 3.633–2.957ETA\_Eta\_B\_RC - 1.179minHBint3 – 1.635SsssCH + 0.678CrripenLogP$ (1)) showed that the former three descriptors were negatively correlated with inhibitory activity, while the latter contributed positively to inhibitory activity. Then, a total of 28 newly designed small molecules based on the scaffold of the most active molecule in the dataset were subjected to QSAR predictions, and the results showed that 25 designed molecules exhibited good predictions (Table S1).

In another study by De et al., the experimental pIC_{50} values of 104 SARS-CoV 3CLpro inhibitors were used to develop a QSAR model for virtual screening of potential anti-COVID-19 drugs [44]. The best model ($R^2 = 0.756$, $Q^2 = 0.708$, $Q^2_{Fn} = 0.752$) was developed by partial least squares (PLS) technique and with a 3:1 ratio of training to test sets. According to the coefficients of the model equation ($pIC_{50} = – 1.586 + 1.333B04[O–Cl] – 0.122F01[C–N] + 0.631B06[N–N] + 0.059ETA_{dBeta} + 0.778B05[C–N] – 0.297nRCONHR$ (2)), four of the six descriptors contributed to an increase in inhibitory activity. Conversely, the presence of the remaining two features was found to be unfavourable for enhancing the inhibitory activity. Using the developed model, predictions were made using three large databases: 8722 antivirals, 11309 peptidomimetics, and 6968 proteases obtained from Asinex (http://www.asinex.com/). Subsequently, the top 25 molecules from each database (Table S2) were identified and confirmed experimentally.

In a study performed by Chtita et al., 40 asymmetrical aromatic disulphides against the main protease of SARS-CoV were collected to establish a QSAR model using the multiple linear regression (MLR) method [59]. The performance of all the developed models was comprehensively evaluated according to the OECD principles of QSAR validation [60] and Golbraikh and Tropsha's criteria [61]. The best model showed a good overall performance ($R^2 = 0.838$, $Q^2_{LOO} = 0.757$, $R^2_{test} = 0.735$). From the model equation ($IC_{50} = 128.780–2.590E_{HOMO-1} + 4.855E_{HOMO} + 51.701S1S2 – 123.760\ S2Bnz + 5.682\ 10^{-06}$ BI (3)), it can be analysed that the descriptors Balaban index (BI), highest occupied molecular orbital energy (E_{HOMO}) and bond length between the two sulfur atoms (S1S2) increase the inhibitory activity, while molecular orbital energy below HOMO energy (E_{HOMO-1}) and bond length between sulfur atom and the benzene ring (S2Bnz) decrease the inhibitory activity. The authors used the two compounds with the highest anti-SARS-CoV activity as templates to design 20 new compounds with higher predicted activity by modifying the groups on the benzene ring. The feasibility of the 20 new compounds was further evaluated based on their drug-like properties (detailed in Section 4.3).

Kumari and Subbarao collected 101 compounds from PubChem Bioassay with defined experimental inhibition data (IC_{50}) on SARS-CoV, followed by transformation into $–logIC_{50}$ (pIC_{50}), and used these data to train a QSAR model based on a deep learning-based convolutional neural network (CNN) [62]. The statistical results revealed that the model had $R^2 = 0.98$ for the training set, and $R^2_{test} = 0.84$ for the test set. The results of the model suggested that the benzimidazole scaffold exhibited higher potency than 4-(4-phenylpiperazine-1-carbonyl)-1H-quinolin-2-one. Benzimidazole, which contains 4-methoxyphenyl and 2-methylsulfanyl groups, exhibited high inhibitory activity. The phytochemical database, natural products from the NCI Diversity Set IV, and FDA-approved drugs were screened using the CNN-QSAR model, and 264, 163, and 520 compounds with potential inhibitory activity (5.00–6.25, 5.00–6.39, 5.00–7.06, respectively) were screened from the three datasets. Further molecular docking analysis was performed to validate the predicted biological activity against 3CLpro of SARS-CoV-2, where NPACT106, NSC5159, and diosmin showed better binding affinities. It is well-known that α-ketoamide derivatives represent a class of potential inhibitors of the main protease of coronavirus [56]. Thus, Oubahmane et al. collected 21 α-ketoamide derivatives to establish the genetic algorithm-MLR (GA-MLR)--based 2D-QSAR models [63]. The best model equation ($pIC_{50} = −17.0891 + 16.1547GATS8i + 0.6723NRS + 36.4881G2p – 0.428H8s$ (4)) contained four molecular descriptors with satisfactory statistical parameters: $R^2 = 0.91$, $Q^2_{LOO} = 0.78$, $R^2_{test} = 0.85$. Subsequently, the model was used to predict 713 newly desired compounds, followed by pharmacophore fitting, molecular docking, MD simulations, and ADMET predictions.

Soleymani et al. collected 81 isatin and indole-based compounds with good inhibitory activity against SARS-CoV 3CLpro, and the reliable Monte Carlo QSAR models were subsequently built [64]. The parameters of the best model were $R^2 = 0.96$ and $R^2_{test} = 0.92$. The substructures contributing to the increase of inhibitory activity include the presence of nitrogen or oxygen with double bond, nitrogen with oxygen, at least one ring, the combination of aliphatic oxygen with double bond, presence of oxygen with double bond and branching and presence of aromatic carbon in first ring. Conversely, the presence of nitrogen with sulfur, consecutive aliphatic carbon with aliphatic nitrogen with branching, aromatic carbon with branching in fourth ring and aliphatic carbon with branching in fourth ring may reduce the inhibitory effect.

Apart from obtaining data from published literature, much available data can also be acquired from official web databases, such as the Binding database (https://www.bindingdb.org/bind/) and the CHEMBL database (https://www.ebi.ac.uk/chembl/).

Zaki et al. reported GA-MLR-based QSAR modelling using 62 compounds obtained from Binding database that showed extensive inhibitory activity against the main protease of SARS-CoV [65]. In this study, the response endpoint is the inhibition constant K_i, which was also converted into pK_i ($-\log_{10}K_i$) for modelling purpose. The developed QSAR model ($pK_i = 4.618 + 2.774fnotringCamdN3B + 0.762aroN\_sp3O\_4B + 0.035fringClipo5B + 0.962flipoplaN2B − 0.279\ com\_sp2O\_5A$ (5)) was strictly validated according to OECD guidelines, with good statistical parameters ($Q^2_{LOO} = 0.859$, $Q^2_{LMO} = 0.848$, $R^2 = 0.898$, and $R^2_{test} = 0.799$). Mechanistic interpretation indicated that non-ring carbons and nitrogens, amide nitrogen, sp2-hybridised carbons and lipophilic atoms were responsible for governing the inhibitory activity. The most active compound in the modelling dataset was selected as the original template to generate a subdataset of 360 heterocyclic variants using RDKit (http://www.rdkit.org). The authors also downloaded a sub-dataset containing 8453 food compounds from FooDB (http://foodb.ca/) and combined the two external subdatasets to generate a dataset containing 8813 molecules for QSAR-based virtual screening. Notably, the choice of food compounds is based on the fact that most of these

compounds are naturally nontoxic and have immunostimulatory effects. The top ten active compounds from the modelling dataset were obtained for further docking and MD analyses.

Ničkčović et al. also extracted 84 heterocyclic molecules in the Binding database with the definitive inhibitory activity (IC_{50}) against 3CLpro of SARS-CoV to construct the Monte Carlo-based QSAR model [66]. It should be noted that the SMILES notation and molecular graph-based descriptors were used for model development. The following statistical parameters were present in the best QSAR model: $R^2 = 0.9314$, $Q^2_{LOO} = 0.9271$, $R^2_{test} = 0.9243$, and $CCC_{test} = 0.9414$, indicating that the Monte Carlo optimisation method was capable of producing a reproducible and interpretable QSAR model with high predictability. In addition, the analysis of the selected SMILES notation-based descriptors revealed that molecular fragments like "(… (… …." (additional branching), "(… N … (…" (additional branching realised by adding aliphatic nitrogen), "(… O … (…" (additional branching realised by adding aliphatic oxygen atom), "+ + + + CL–N = = = " (the presence of chlorine and nitrogen atoms), "+ + + + N–-O = = = " (the presence of oxygen and nitrogen atoms) "C···C···C …" (propyl group), "N···C···C …" (ethyl amine group), "O···C… ….", "O···C···C …" and "C···O···C …" (methoxy, ethoxy and dimethoxy groups), "c … c … …." and "c … c … c …" (aromatic carbon atoms) that have a positive influence on the inhibitory potency against 3CLpro of SARS-CoV (Fig. 4). Based on these fragments, they successfully designed five new molecules using the most active molecule in the dataset as a template, and their predicted inhibitory effects on 3CLpro of SARS-CoV were higher than those of the template molecule. A docking study was performed to validate the binding affinities of the designed molecules to 3CLpro.

Ishola and co-workers retrieved 73 compounds with SARS-CoV 3CLpro inhibitory activity from the CHEMBL database as a modelling dataset [67]. The correlation model was generated using the GA-MLR with acceptable statistical parameters ($R^2 = 0.907$, $Q^2_{LOO} = 0.866$, $R^2_{test} = 0.609$) according to Golbraikh and Tropsha's criteria [61]. The model equation was shown as $pIC_{50} = 5.531 + 1.824ALogP2 + 2.209AATS8v – 1.324AATS3i + 1.107MATS6 – 1.227GATS8e – 1.648BCUTp\text{-}1h – 1.448ZMIC2 – 1.229VE1\_D(7)$ (6). The Y-randomisation validation indicated that the obtained model was not generated by chance. Applicability domain (AD) analysis showed that all compounds were within the model's AD, except for one training compound, indicating an acceptable predictive ability for the newly designed molecules against SARS-CoV 3CLpro. Tejera et al. collected 220 compounds with inhibitory activity against the 3CLpro of SARS-CoV from the CHEMBL database [68]. Compared to Ishola's study, more available data were included in this study. The graph convolutional network approach was used to create a binary classifier for inhibitors and non-inhibitors, in which the SMILES notation of a molecule was converted into a molecular graph that was used as an input for the graph convolutional networks. The best classification model yielded an area under the receiver operating characteristic curve (AUC) of 0.914 for the test set, suggesting good predictive performance of the model. The model was then applied to virtual screening for potential 3CLpro inhibitors of SARS-CoV using drug molecules in the DrugBank database, and the top 20 candidates were obtained for further docking and MD analysis. In 2022, Ghaleb et al. reported a 3D-QSAR modelling study based on 110 pyridine N-oxide-based antiviral compounds with inhibitory activity against SARS-CoV 3CLpro [69]. The best model was developed using the comparative molecular similarity index analysis (CoMSIA) method, in which the SEHDA model showed the best statistical parameters with $Q^2 = 0.54$ and $R^2_{test} = 0.71$. Based on the contour maps of the 3D-QSAR model, they further designed ten new pyridine N-oxide compounds and predicted their inhibitory activity, followed by molecular docking and ADMET evaluation. With the continuous efforts of researchers from all walks of life, many studies have reported molecules that inhibit the 3CLpro activity of SARS-CoV-2. These compounds can be used to establish QSAR predictive models, which will make a certain contribution to the design and development of anti-SARS-CoV-2 drugs. Kumar et al. collected 190 compounds with inhibition data against 3CLpro of SARS-CoV-2 and established a 2D-QSAR model based on the PLS method [70]. All statistical parameters of the model ($R^2 = 0.789$, $Q^2 = 0.741$, $Q^2_{F1} = 0.787$, $Q^2_{F2} = 0.786$) met the strict validation criteria for QSAR modelling. The model was used to screen five large databases containing 66495 compounds (including Asinex Antiviral database, Asinex Peptidomimetics database, CAS COVID-19 antiviral database, US FDA-approved antiviral drugs, the dataset reported by Ton et al. [71]). After combined screening using the QSAR and 3D pharmacophore models, 14738 compounds were selected for further molecular docking analyses (see 4.1 section).

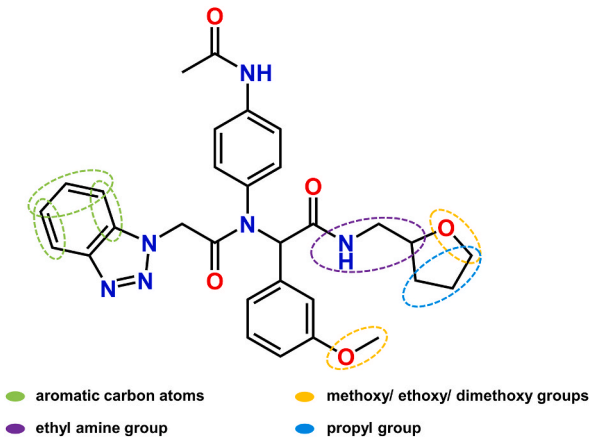


Fig. 4. Molecular fragments with positive influence on SARS-CoV 3CLpro inhibitory activity.

Amin et al. collected 33 representative baicalein derivatives from the published literature that could inhibit the activity of SARS-CoV-2 3CLpro to varying degrees [72]. They established a Bayesian classification-based QSAR model ($ROC = 0.747$, $ACC = 0.840$, $ROC_{test} = 1.000$, $ACC_{test} = 0.833$) using Discovery Studio software and determined the active molecule to have a pIC_{50} ($-logIC_{50}$) value greater than 5.0, and an inactive molecule with a value less than 5.0. After eliminating the compounds without definite inhibitory activity against SARS-CoV-2 3CLpro, the remaining 25 molecules with definite inhibitory activity values were used to construct an MLR model ($R^2 = 0.956$, $Q^2_{LOO} = 0.928$). The two predictive models were validated using the widely accepted validation criteria. Using the established model, the molecular fragment structure that plays a positive and negative role in the inhibitory effect was analysed to modify the compound structure and design new inhibitors. Bi-acetyl amine and acetamido group containing 2-oxo pyrrolidine moiety, cyclohexyl and cyclohexyl methyl groups, acetamido methylene (iso-butyl) acetamide moiety, and oxy-anion function were identified as positive influencers of 3CLpro inhibition. Subsequently, four chromone-based molecules were designed and predicted to have excellent inhibitory activity ($pIC_{50} > 7.523$), providing important candidate molecules for the development of novel drugs for COVID-19.

In addition to the derivatives of baicalein and α-ketoamide derivatives, 9,10-dihydrophenanthroline derivatives also exhibit potential to develop into a new drug for the treatment of COVID-19. In a study by Daoui et al., a 3D-QSAR modelling based on 46 9,10-dihydrophenanthrene molecules was performing by comparative molecular field analysis (CoMFA) and comparative molecular similarity index analysis (CoMSIA) approaches, in which CoMFA includes two descriptors: steric (S) and electrostatic (E), and CoMSIA includes five descriptors: steric (S), electrostatic (E), hydrophobic (H), hydrogen bond donor (D), and hydrogen bond acceptor (A) [73]. These 46 molecules showed different inhibitory effects against the main protease of SARS-CoV-2, 3CLpro. Based on the two analytical approaches, the CoMFA/SE model ($R^2 = 0.97$, $Q^2_{LOO} = 0.81$, $R^2_{test} = 0.95$) and CoMSIA/SEHDA model ($R^2 = 0.94$, $Q^2_{LOO} = 0.76$, $R^2_{test} = 0.91$) exhibited the best statistical performance. Analysis of the contributions of different fields revealed that the inhibitory activity of the 9,10-dihydrophenanthrene derivatives on SARS-CoV-2 3CLpro was mainly affected by their A, S, H, and E properties. Subsequently, 96 new molecules were designed from the template molecules in the modelling dataset, followed by activity prediction, drug-likeness screening, docking analysis, MD simulations, and ADMET prediction.

In a QSAR study performed by Adhikari and co-workers, 69 compounds with potential inhibitory effects on SARS-CoV-2 3CLpro were selected to develop three different types of QSAR models, namely, the MLR QSAR model, hologram QSAR model, and non-linear QSAR model that included an artificial neural network (ANN) model and support vector machine (SVM) model [74]. For the MLR and non-linear models, two different algorithms were used, the genetic algorithm (GA) and stepwise algorithm (SW), to screen the descriptor variables responsible for inhibitory potency. All the established models are internally robust and externally predictive, such as SW-MLR ($R^2 = 0.703$, $R^2_{test} = 0.565$), GA-MLR ($R^2 = 0.740$, $R^2_{test} = 0.654$), SW-ANN ($R = 0.872$, $R_{test} = 0.745$), GA-ANN ($R = 0.865$, $R_{test} = 0.814$), SW-SVM ($R = 0.839$, $R_{test} = 0.812$), GA-SVM ($R = 0.913$, $R_{test} = 0.814$), and HQSAR ($R^2 = 0.867$, $R^2_{test} = 0.715$), indicating they can be used to predict the inhibitory activity of untested compounds. Analysis of the structural features required for high antiviral activity revealed that structural fragments such as benzyl acetate amide, 2-oxopyrrolidinyl methyl, methylene, benzyl, and benzyloxy carboxamide play a positive role. According to the predictions of these models, curcumin, ribavirin, lurasidone, saquinavir, lopinavir, elbasvir, paritaprevir, sepimostat, and remdesivir were potent inhibitors of SARS-CoV-2 3CLpro (Table S3).

Recently, Oubahmane's team developed 2D-QSAR models using 55 dihydrophenanthrene derivatives as potent inhibitors of SARS-CoV-2 3CLpro with Monte Carlo optimisation method and the Genetic Algorithm Multi-Linear Regression (GA-MLR) method [75]. The best Monte Carlo QSAR model ($R^2 = 0.9203$ and $Q^2 = 0.8508$) revealed that feature fragments like "(…. …. …", "(…. O … (…" and "++++N—" can enhance the inhibitory activity, while "O …. …. …", "n …..... …" and "n … c …. …" can decrease the inhibitory activity. For the best GA-MLR QSAR model, the statistical metrics ($R^2 = 0.9288$, $Q^2_{LOO} = 0.8902$, $R^2_{test} = 0.8558$) are excellent, which can reliably predict the inhibitory activity of the newly designed compounds.

Currently, researchers have mainly focused on two aspects of modelling SARS-CoV-2 3CLpro inhibitors. The first is "old drugs for new use or drug repurposing". Owing to the high similarity between SARS-CoV-2 3CLpro and SARS-CoV 3CLpro, it is feasible to establish QSAR models for SARS-CoV 3CLpro and screen potential molecules that significantly inhibit the 3CLpro activity of SARS-CoV-2 and the key molecular fragments imparting the inhibitory effects. The second is to develop QSAR models using compounds that can directly target SARS-CoV-2 3CLpro. According to the established models, it can be analysed that some structural fragments play a positive role in inhibitory activity. For example, the reduction of molecular skeleton branches, shortening of the bond length of two sulfur atoms, increasing the bond length between sulfur atoms and the benzene ring, and increasing the number of aromatic nitrogen atoms can improve the inhibitory activity against 3CLpro [59,65,76]. At the same time, lipophilic compounds exhibit higher inhibitory activity compared to hydrophilic compounds [58,65]. Screening of such active structures allows for the design, optimisation, and identification of potential candidate compounds. The incorporation of these active structures into new inhibitors may result in the formation of highly potent compounds. Although the dataset of SARS-CoV-2 3CLpro inhibitors is insufficient compared to that of SARS-CoV, the experimental data were obtained by directly acting on SARS-CoV-2 3CLpro, which is more reliable and promising than indirect modelling by SARS-CoV 3CLpro inhibitors. In short, modelling both entry points has its own advantages and can provide support for the design and discovery of anti-COVID-19 drugs in the future. It can also be seen that the methods and software employed by each research group to establish QSAR predictive models are different, including MLR/PLS-QSAR, Monte Calo-QSAR, and machine-learning-based non-linear QSAR models. These models demonstrate good internal robustness and external predictive performance, meeting internationally accepted validation standards [60]. Model equations and descriptor analyses or interpretations can provide an important feature reference and theoretical basis for the design, development, and innovation of anti-SARS-CoV-2 drugs targeting 3CLpro in clinical practice.

3. QSAR models of spike protein and human receptor protein ACE2 inhibitors

3.1. S protein and hACE2

SARS-CoV-2 is primarily transmitted through the respiratory tract but can also spread through physical contact. A recent study suggested that SARS-CoV-2 may also infect individuals through aerosol exposure to the eyes [77]. SARS-CoV-2 utilises transmembrane serine proteinase 2 (TMPRSS2) and/or histone proteinases B/L (cathepsin B/L, Cat-B/L) present on the surface of host cells to facilitate activation of the viral S protein, which enables the virus to bind to hACE2 receptors on the cell surface and enter the host cell (Fig. 2) [78]. As shown in Fig. 5A, the S protein can be hydrolysed into S1 and S2 subunits by TMPRSS2 or Cat-B/L, thereby mediating receptor binding and membrane fusion [79]. Furthermore, S1 is subdivided into an N-terminal domain (NTD) and a receptor-binding domain (RBD), which directly bind to the extracellular peptidase domain (PD) of the cell surface ACE2 receptor on human respiratory epithelial cells [79,80]. The S2 subunit comprises an N-terminal fusion peptide (FP), two heptad repeats (HR1 and HR2) separated by a central helix (CH), connector domain (CD), transmembrane domain (TM), and a cytoplasmic tail (CT) [81,82]. When the virus invades the host, the S1 subunits fall off from the S protein, the S2 subunits expose the fusion peptide region, and the FP segment is inserted into the host cell membrane, finally completing membrane fusion [83,84]. SARS-CoV-2 is internalised by endocytosis; viral RNA is released, replicated, and translated within host cells, and virus particles are further assembled and released, opening a new round of infection in host cells. Fig. 5B and C shows the morphology of the complex formed by RDB and hACE2 as well as the four active amino acid residues on hACE2, namely Tyr41, Tyr83, Lys353, and Gln24, which facilitate the interaction between hACE2 and small-molecule inhibitors [14].

hACE2, also known as ACEH (angiotensin-converting enzyme Homolog), is the first human homolog of angiotensin converting enzyme (ACE), which was cloned from the complementary DNA of human heart failure and lymphoma [85,86]. The hACE2 protein is composed of 805 amino acids, and its structure includes an N-terminal signal peptide region, a catalytic domain, a transmembrane region, and a C-terminal intracellular domain [87]. Membrane-linked hACE2 distributed on the cell surface can be hydrolysed and cleaved by related proteases into soluble hACE2 that lacks the transmembrane region and C-terminal intracellular domain. The relative molecular weight of soluble hACE2 is approximately 89,000 [88].

3.2. QSAR studies of the inhibitors interfering the binding of S protein and hACE2

The spike protein (S protein) on the surface of SARS-CoV-2 is very important for binding to the hACE2 receptor and viral entry into the body. Consequently, the protein-protein interaction (PPI) between the SARS-CoV-2 RBD and hACE2 may become a therapeutic

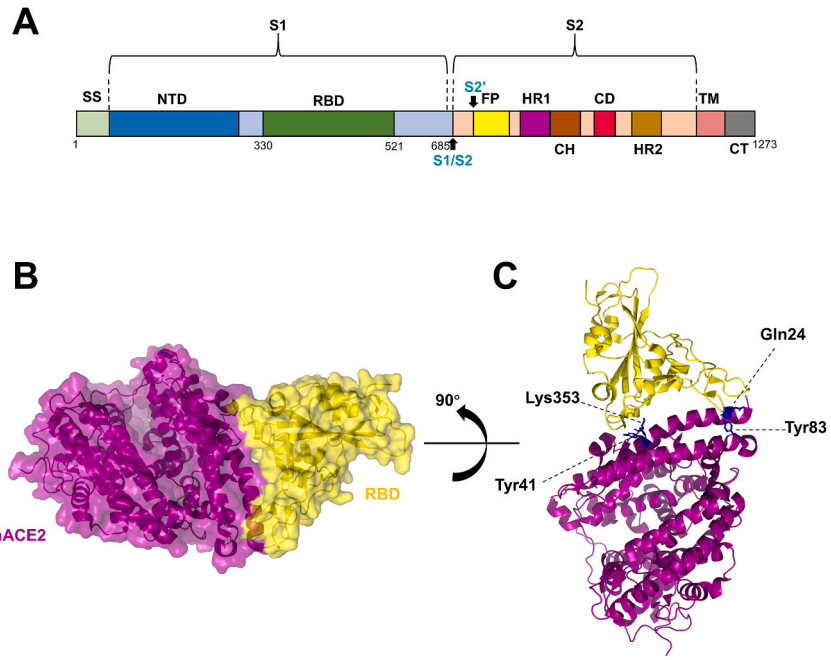


Fig. 5. (**A**) Full-length SARS-CoV-2 S protein. SS, single sequence; NTD, N-terminal domain; RBD, receptor-binding domain; S1, subdomain 1; S2, subdomain 2; S1/S2, S1/S2 protease cleavage site; S2′, S2′protease cleavage site; FP, fusion peptide; HR1, heptad repeat 1; CH, central helix; CD, connector domain; HR2, heptad repeat 2; TM, transmembrane domain; CT, cytoplasmic tail. Arrows indicate the protease cleavage site. (**B**) Cartoon representation of the complex structure of the SARS-CoV-2 RBD (yellow) bound to hACE2 (purple) (PDB: 6M0J). (**C**) Interface between hACE2 (purple) and bound RBD (yellow) in the ribbon diagram (PDB: 6M0J), four key amino acid residues were highlighted (Gln24, Tyr41, Tyr83, Lys353). (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

target for drug development to inhibit the entry of SARS-CoV-2 into host cells. Pirolli et al. collected 1747 PPI inhibitors from the iPPI-DB database to train a CNN-based QSAR model (Fig. 6) with an accuracy of 0.95, precision of 0.92, F1-score of 0.91, and area under the ROC curve of 0.99 [89]. The false positive and negative rates were 2.7 % and 9.6 %, respectively. After validation, the CNN-based model was used for virtual screening to identify potential PPI inhibitors from a drug-like library (~2 M compounds). The molecules predicted to be potential PPI inhibitors were further filtered using toxicity prediction and docking analyses.

Plonka and co-workers [90] performed a docking analysis of 166 aminothioureas that can act at the interface between the spike protein of SARS-CoV-2 and the hACE2 receptor, followed by machine learning-based QSAR modelling using the docking scores as the response endpoint. However, because of the extensive size of the PPI and the binding site differences between the studied molecules and PPI, the internal robustness of the developed QSAR models was relatively low ($Q^2 < 0.45$), although the correlation coefficient R^2 exceeded 0.7. Subsequently, ADMET predictions were performed to identify the best candidates as PPI inhibitors, and one molecule (FSoOH) (Table S3) was explored in further validation assays.

Zarezade et al. developed a 3D pharmacophore QSAR model for novel hACE2 [91]. The dataset contained 31 compounds, among which 18 compounds were classified as the training set, and the remaining compounds were classified as the test set. The training compounds were structurally different from each other and displayed 4-5 orders of magnitude difference, thus characterising the types and functions of these compounds. The best pharmacophore model based on the training set was more reliable and applicable, with R^2 of 0.984 and R^2_{test} of 0.837, respectively. Compared to the 2D-QSAR model, this model considers the three-dimensional conformation of molecules and incorporates more comprehensive molecular information; thus, it has better potential for virtual screening from a large database. The best pharmacophore model was then used to screen the inhibitory activity of 1264479 compounds from the PubChem and Zinc databases, and 492697 ligands successfully passed the best pharmacophore model. ADMET, molecular docking, and MD simulations were applied to the candidate compounds for further screening.

In fact, the hACE2 receptor can not only bind to SARS-CoV-2 but can also be identified as a functional receptor of SCV (pathogen of severe acute respiratory syndrome) and the H–CoV NL63 virus [92,93]. These viruses can also use hACE2 to enter host cells, endangering human health. Among the early QSAR studies, Torres et al. used the data of 21 compounds with inhibitory activity on hACE2 to conduct 3D-QSAR CoMSIA analysis [94]. The best CoMSIA model showed satisfactory statistical parameters, with $Q^2 = 0.652$ and $R^2 = 0.962$, indicating a good predictive performance. Although this work did not involve specific content related to SARS-CoV-2, its focus was still on hACE2; therefore, it has some guiding significance for the design and development of anti-COVID-19 drugs.

Table 1 shows the basic information on the models established for hACE2 inhibitors. At present, the models for compounds inhibiting the binding of SARS-CoV-2 spike protein to hACE2 are not as rich as those for compounds inhibiting 3CLpro, and most of the models are classification-based QSAR models. In fact, it can be found that the data set of compounds targeting PPI is still large.

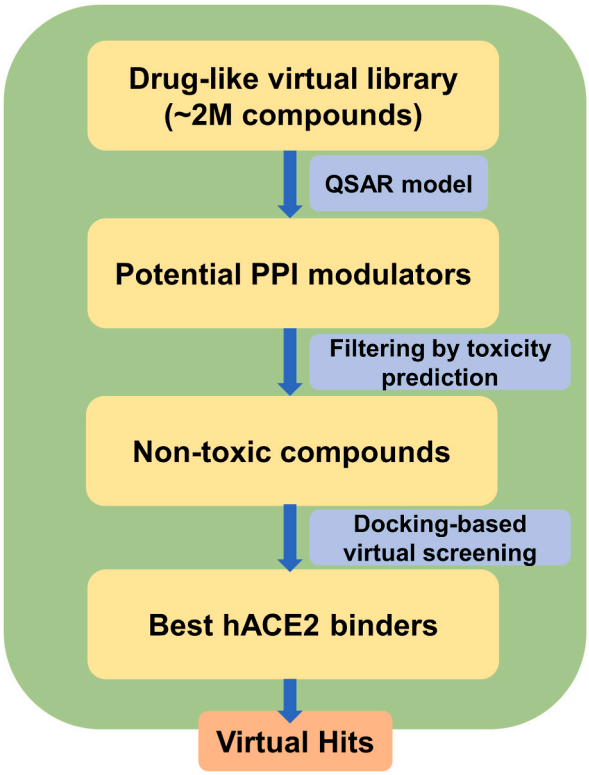


Fig. 6. Workflow of the virtual screening strategy in study by Pirolli et al. [89].

However, owing to the disunity of the binding sites between these compounds and PPI, corresponding QSAR models with excellent performance cannot be successfully established.

4. Molecular docking, molecular dynamics, drug-like properties

4.1. Molecular docking

Scientists have proposed molecular docking approaches to investigate chemical problems in biological systems at the molecular level. It is a computational technique that leverages computer technology to simulate intricate interactions between small-molecule ligands and macromolecular receptors, facilitating computer-aided virtual drug screening. By iteratively optimising the conformation and position of ligands and/or receptors, we identified the optimal conformation of the ligand-receptor complex. This allowed us to calculate the binding mode and affinity, score the results, and select the most promising ligands based on these scoring results. Subsequently, the selected ligand is used as a candidate for further exploration and development [95].

Molecular docking aims to identify the active binding sites of small molecule ligands and target proteins, enabling them to form energetically favourable conformations. In general, molecular docking includes three interconnected components: binding-site recognition, conformation search algorithms, and scoring functions. Binding-site recognition involves identifying the active sites within target protein molecules that interact with ligands. The conformational search algorithm focuses on determining the optimal position, orientation, and conformation of the ligand using an optimisation algorithm that considers the flexibility of the ligand. The scoring function is responsible for evaluating the binding conformation during the search process, providing a measure of the quality of the ligand-receptor interaction [95,96].

Currently, molecular docking is primarily categorised into three types: flexible, semiflexible, and rigid. The distinction between these categories lies in whether the conformations of the ligand and receptor can be changed. Flexible docking is employed when both the receptor and ligand conformations can be altered; semi-flexible docking is utilised when the receptor conformation remains unchanged and the ligand conformation can be adjusted; and rigid docking refers to a scenario in which neither the receptor nor ligand conformations can be modified [97,98]. Each of these three docking methods has unique advantages and characteristics. Rigid docking stands out for its simplicity and speed, which makes it suitable for handling large structural molecules. Semi-flexible docking, on the other hand, is particularly useful for studying the interactions between small and large molecules and remains the most commonly employed approach in drug research and design. Although semiflexible docking may exhibit lower docking efficiency, it requires high accuracy in molecular conformation and yields precise docking results. This feature makes it particularly suitable for the precise examination of molecular recognition processes. With ongoing advancements in computer capabilities, the variety of software options available for molecular docking continues to expand, accompanied by continual improvements in their performance. Notable examples are DOCK, AutoDock, FlexX, and GOLD. These software packages were specifically designed and developed to facilitate the molecular docking processes [99–102]. Researchers can select different search algorithms and scoring functions based on specific needs and operational preferences. These variations in molecular docking methods allow researchers to explore different aspects of protein-ligand interactions and tailor their approach to the specific requirements of their research. Each algorithm and scoring function has its own strengths and limitations, and researchers can select the most suitable based on factors such as the accuracy, speed, and nature of the molecules being studied. The availability of various options enhances the flexibility and applicability of molecular docking techniques in drug discovery and other research fields.

In the CNN-based QSAR modelling by Kumari et al., a molecular docking technique was employed on three datasets to validate the predictions from the CNN-QSAR model [62]. First, the binding energy of the top 10 compounds selected from the phytochemical database ranged from −42.5 to −29.3 kJ/mol, and seven of them belong to the flavonoid compounds. The molecule (NPACT0106) (Table S3), with the best interaction with 3CLpro of SARS-CoV-2, formed two hydrogen bonds with Thr26, one hydrogen bond with His163, and one hydrogen bond with Glu166. Secondly, for the natural products database, the binding energy of the top 10 compounds ranged from −48.4 to −30.1 kJ/mol. The most promising compound NSC5159 (Table S3) (−48.2 kJ/mol) forms eight hydrogen bonds with residues Thr26, Leu141, Gly143, Ser144, Cys145 and Arg188. For drug repurposing, the binding energy of the top 10 compounds in FDA-approved drugs ranged from −44.8 to −31.6 kJ/mol. Among these, diosmin (Table S3) exhibited the highest binding affinity (−44.8 kJ/mol) with 3CLpro, which forms three hydrogen bonds with Phe140 and His163.

After developing the QSAR model of SARS-CoV-2 3CLpro based on 21 α-ketoamide inhibitors, Oubahmane et al. employed a docking procedure to explore the interactions and conformational patterns of three selected hit compounds within the active site of SARS-CoV-2 3CLpro by AutoDock software [63]. By comparing the docking results with those of the reference compound, it was found that only one (**329**) of the three candidate compounds (Table S3) exhibited a higher binding affinity than the reference compound. From the docking results, it can be concluded that the three candidate compounds interact with the two key catalytic residues, Cys145 and His41, of 3CLpro by forming a hydrogen bond with Cys145 and/or a hydrophobic bond with His41, suggesting that they are promising inhibitors.

After completing the QSAR-based virtual screening of the anti-SARS-CoV activity of 8813 compounds by Zaki et al., flexible-rigid docking was applied to analyse the binding modes of the top ten molecules in the modelling dataset [65]. The docking results showed that the most active molecule interacted with 3CLpro through the residues His41, Met49, Tyr54, Phe140, Leu141, Asn142, Ser144, Cys145, His163, His164, Met165, Glu166, Asp187, Arg188, and Gln189. The combination of a lipophilic region and a hydrogen bond donor is very important for the interaction between the inhibitor and the S2 subpocket of 3CLpro. Similarly, molecular docking was also used by Ničkčović et al. to validate the inhibitory activity of the five newly designed molecules (from QSAR-aided design, see Table S3) on the 3CLpro of SARS-CoV. The results showed that the new molecule with the highest predictive activity in the QSAR also

possessed the best docking score [66].

Ishola et al. carried out a docking study using VINA software on 73 molecules collected from the CHEMBL database used for QSAR modelling [67] and found that seven compounds (Table S3) had significant binding affinities of −43.1 to −38.5 kJ/mol for SARS-CoV 3CLpro. Importantly, hydrophobic interactions play a crucial role in the binding of inhibitors to 3CLpro, and residues His41, Cys44, Met49, Leu287, and Tyr237 are commonly involved in this interaction. By analysing the docking results, they proposed that two individual inhibitors, one that binds to the active catalytic site (domains I and II) and the other that disrupts the dimerization interface (domain III), can be linked together to create a bifunctional inhibitor with significantly increased binding affinity and specificity [67].

After virtual screening of potential SARS-CoV 3CLpro inhibitors using the established QSAR model, Tejera et al. used the Gold software to perform molecular docking analysis to evaluate the binding affinities of the top 20 candidate compounds against SARS-CoV-2 3CLpro [68]. According to the consensus docking score in the Gold software, the top five compounds were inositol nicotinate, niceritrol, rebastinib, aleplasinin, and liothyronine (Table S3), while the worst-scoring compounds were ortataxel, isofluorophate, nicotinic acid, aluminium nicotinate, and amobarbital. However, the scoring functions employed for these calculations often overlook numerous critical factors that significantly affect the molecular recognition and stability of the complexes, prioritising computational efficiency over comprehensive analyses. Therefore, it is recommended that post-processing ligand-receptor complexes be predicted using docking tools by employing more accurate methods that rely on the analysis of MD trajectories.

In a study by Ghaleb et al., molecular docking was applied to ten newly designed pyridine N-oxide compounds and four reference compounds (chloroquine, hydroxychloroquine, compound 63, and compound 31) [69]. By analysing the top three new compounds and the reference compound 63 (Table S3), it was found that these compounds mainly bound to SARS-CoV-2 3CLpro through hydrogen bonds, and Gly143 made a very important contribution to the formation of hydrogen bonds. Furthermore, an increase in phosphoric and nitrogen dioxide substituents on the pyridine moiety in pyridine N-oxide compounds can greatly enhance their inhibitory effect.

Kumar et al. implemented molecular docking using Biovia Discovery Studio (https://www.3dsbiovia) for the screened 14738 compounds derived from QSAR predictions. The CDOCKER interaction module was used to check each receptor-ligand complex, and the top-scoring molecules with only the poses of non-covalent interactions (ionic bonds, hydrophobic interactions, hydrogen bonds, etc.) were retained for further analysis. The top six molecules (Table S3), including AsinexAntiV-BDD-26908890, CASAntiV-865453-58-3, CASAntiV-865453-40-3, CASAntiV-2043031-84-9, CASAntiV-1370259-79-2, and CASAntiV-2043031-85-0, had significant predicted activities and binding energies [70].

To further explore the inhibitory activities of the four newly designed compounds (Table S3), Amin et al. conducted docking studies using the AutoDock tool. The results showed that all four molecules could dock to the active site of SARS-CoV-2 3CLpro, and D4 had the highest binding energy [72]. Similarly, the AutoDock tool was used by Daoui et al. to predict the potential non-covalent interaction profiles between drug molecules and the active sites of SARS-CoV-2 3CLpro [73]. Based on the results from QSAR predictions, they used the newly designed 9,10-dihydrophenanthrene derivatives, template molecule T40, original covalent inhibitor N3, and reference molecule disulfiram as ligands to explore their binding affinities with the active pocket of 3CLpro and screened the potential non-covalent inhibitors of 3CLpro. The docking results showed that the non-covalent interactions were commonly hydrophobic, electrostatic, hydrogen bonding, and van der Waals interactions, and the residues involved in these interactions were mainly Cys145, His41, Asn142, and Glu166. As a result, there were 11 molecules (Table S3) exhibiting higher binding stability than T40 and disulfiram with the active pocket of 3CLpro, thereby providing a solid theoretical basis for nominating them as novel anti-COVID-19 drugs.

To discover potential PPI inhibitors [89], non-toxic molecules from a drug-like library (ZINC15 database) were docked to the hACE2/spike PPI interface for virtual screening (Fig. 5). The top four molecules (Table S3) were identified as PPI inhibitors that prevented hACE2 from binding to the spike protein of SARS-CoV-2 and viral infection. Remarkably, all four compounds exhibited consistent interaction patterns at the hACE2 binding site, indicating a common mode of binding. Additionally, our results highlighted the potential significance of the N-benzyl carbamoyl group, which was present in three of the top four compounds, suggesting that this group is a promising scaffold for the development of novel drug candidates targeting COVID-19.

Although Plonka et al. failed to develop an acceptable QSAR model for predicting the docking scores of aminothiourea with PPI (the interface between the SARS-CoV-2 S protein and the hACE2 receptor) [90], the docking results based on the Gold software provided some important structural information for optimising these PPI inhibitors. Thiadiazole rings are frequently present in top-binding compounds. Among these molecules, 5-(2-methylfuran-3-yl)-2-(2-hydroxyphenylamino)-1,3,4-thiadiazole (FSoOH) (Table S3) yielded the best docking results. The molecule bound tightly to the Leu29-His34 region of the hACE2 receptor terminal helix and, to a lesser extent, weakly bound to the Pro49-Tyr495 region of the S protein. ADMET properties are also important for the development of these molecules as effective drugs against COVID-19.

In a search for potential hACE2 inhibitors from PubChem and Zinc databases, Zarezade et al. found that 81 compounds were successfully docked to the active site of hACE2 [91]. Among these, 73 compounds had a higher predicted inhibitory activity than the most active modelling compounds. Unexpectedly, 29 of these compounds could be successfully docked to the active site of 3CLpro, implying that they can be used as dual inhibitors of hACE2 and 3CLpro. ZINC12562757 (Table S3) showed the most potential as the hACE2 inhibitor with a docking energy of −63.7 kJ/mol. Compared to other compounds, ZINC12562757 interacted with many amino acid residues of hACE2 to form five hydrogen bonds with Asn149, Pro346, Lys363, and Glu375. Compound 112,260,215 (Table S3), the best potential inhibitor of 3CLpro, exhibited a docking energy of −46.7 kJ/mol, and could interact with four key amino acid residues (His41, Cys145, Met165, and Glu166) to form numerous hydrogen bonds. Moreover, 10 molecules (Evo_1–10) were designed with compound 112,260,215 as the molecular template, and Evo_1 (Table S3), which had the highest docking energy, was further considered. The three candidate compounds were subjected to MD simulations to evaluate their reliability as potential inhibitors.

In the design of hACE2 inhibitors, Torres et al. used the FlexX software to perform a docking analysis [94]. Using this software, 100 binding poses of each ligand molecule bound to the hACE2 protein (PDB ID:1R4L) were explored, among which the best pose was

selected by ranking the consensus scores based on various scoring functions. The results showed that these ligands mainly bound to the three pockets of hACE2, and the electrostatic S1 subsite contributed to the interaction with the polar residues of the natural substrate of hACE2. Based on the docking results, the lead optimisation tool LeapFrog was used to design a series of novel potential hACE2 inhibitors, followed by QSAR, ADMET, and drug-likeness predictions.

4.2. Molecular dynamics

MD is a computational method that involves modelling a molecular system using a classical force field and simulating its movement to predict its behaviour [103]. By constantly updating and recording the molecular configuration, MD generates a set of conformations representing different states of the system. From these conformations, various macroscopic properties such as thermodynamics can be calculated and analysed. MD simulations can simulate the conformational changes in molecules by considering their thermal motion, which follows the Boltzmann distribution at a given temperature. This method is particularly useful for exploring the behaviour of proteins, including their response to mechanical forces, in a cellular environment. For example, external forces can induce conformational changes in proteins, which may affect the accessibility of ligands to their binding sites. Additionally, MD simulations allow researchers to simulate various experimental conditions such as mutations, ligand binding, temperature changes, and solvent effects. The ability to control a simulated environment provides researchers with additional flexibility and opportunities. Several parameters are commonly used to evaluate the MD simulation results. These include the root-mean-square deviation (RMSD), which measures the overall structural deviation of the simulated system compared with a reference structure. The root-mean-square fluctuation (RMSF) quantifies the local flexibility of atoms within the system. The radius of gyration (Rg) characterises the overall compactness of the system, while the number of hydrogen bonds indicates the strength and stability of the hydrogen bonding interactions [104]. Methods such as Molecular Mechanics/Generalised Born Surface Area (MM/GBSA) or Molecular Mechanics/Poisson-Boltzmann Surface area (MM/PBSA) are commonly employed to assess the binding affinities and stability of complexes. These methods provide more accurate approximations of the binding energies and stability by considering solvation effects and other factors [105]. In summary, MD simulations offer a valuable approach to study the conformational dynamics of molecules, particularly proteins, under various conditions. The ability to manipulate the simulation environment and use specific parameters and energy calculations enables researchers to gain insights into the molecular behaviour and interactions.

In the virtual screening of SARS-CoV-2 inhibitors, MD simulations and molecular docking can be used in conjugation with QSAR to further evaluate the screening results of the QSAR model and focus on the most promising inhibitors. MD simulations have an advantage over molecular docking because they directly reflect the molecular docking process without being affected by scoring functions, which is helpful in understanding the docking mode of small molecules. However, MD simulations often require a significant amount of time, which makes them unsuitable for rapid docking. It is typically used to explore the binding mode of a ligand to its corresponding receptor. Therefore, a combined approach involving molecular docking and MD simulations is commonly used. After conducting the molecular docking experiments, MD simulations were performed on a few selected molecules to achieve a higher level of screening optimisation.

For example, Kumari and Subbarao employed MD simulations to further investigate the stability of selected ligand-protein complexes, specifically inhibitors, derived from docking analysis of the SARS-CoV-2 3CLpro protein [62]. The top-ranked docked complexes were chosen to understand the reliability of the complex through various analyses such as RMSD, RMSF, Rg, and H-bond interactions. Over the course of 100 ns of MD simulation analysis, the docked complex of 3CLpro-diosmin remained stable. This suggests that diosmin (Table S3) is a promising drug candidate targeting the main protease of SARS-CoV-2.

To discover potential-ketoamide inhibitors targeting the main protease of SARS-CoV-2, Oubahmane et al. performed MD simulations after QSAR, pharmacophore fitting, and molecular docking-based virtual screening [63]. In this study, 160 ns MD simulation were conducted using the AMBER software package. Three different compounds (Table S3) were simulated to form complexes with SARS-CoV-2 3CLpro. The MM/PBSA method was employed to calculate the binding free energy by considering the van der Waals, solvation, and electrostatic energies. Based on the findings, the binding energies of complex **329** (−161.4 kJ/mol) and complex **331** (−197.1 kJ/mol) were found to exhibit the most favourable binding when compared to the other two complexes. The strong binding of these two compounds to 3CLpro can be attributed to specific protein residues, namely, Asn140, Gly141, Ser142, His41, Asn49, Met163, and Gln187. These residues enhance both the electrical and van der Waals interactions between the compounds and proteins, contributing to the favourable binding energies observed.

Zaki et al. performed MD simulations to determine the stability and convergence of 3CLpro with and without the most active molecule, **1** (Table S3), in a modelled dataset [65]. The RMSF, RMSD, and Rg values revealed that the binding of molecule **1** with 3CLpro was highly stable compared to that to apo-3CLpro. The amino residues Glu166, Asp187, Arg188, His41 and His172 participated in the formation of water bridges, H-bonding interaction, polar contacts and π-π stacking with molecule **1**. The MM/GBSA calculations conducted to evaluate the free energy change of ligand binding in the main protease Mpro of SARS-CoV-2 revealed an average binding energy of -219.6 ± 16.7 kJ/mol. The notably lower binding energy suggests a higher affinity of the molecule for 3CLpro. Consequently, this finding opens up new possibilities for the development of novel inhibitors targeting SARS-CoV-2. In MD simulations, along with MM/GBSA, MM/PBSA is another commonly used method for calculating binding free energy. Ishola et al. utilised this approach to calculate the binding free energies of five molecules (Table S3) that were initially screened using molecular docking and ADMET analyses [67]. To investigate the intricate structural and dynamic mechanisms of 3CLpro and its complexes with compounds, the MD simulation trajectories of 100 ns simulations were thoroughly analysed. They assessed the stability of the molecule-protein binding by analysing RMSF, RMSD, and Rg. The three best-performing molecules (CHEMBL194398, CHEMBL196635, and CHEMBL210097) (Table S3), based on their binding energy rankings, were selected, with binding energies of

−70.4 kJ/mol, −53.4 kJ/mol, and −47.5 kJ/mol. These findings indicate that these molecules effectively bind to the SARS-CoV-2 3CLpro enzyme.

Tejera et al. also performed an MD simulation of the top 20 3CLpro inhibitor candidates screened using the QSAR model and predicted the binding free energies of the 20 compounds using the MM-PBSA method, as implemented in the AMBER package [68]. Interestingly, levothyroxine, which was not one of the top five docked compounds, exhibited the highest binding free energy. Conversely, inositol nicotinate, which was ranked as the best compound in the docking calculations, had a relatively poor estimated free energy of binding, ranking the third worst. These findings emphasise the significance of employing free-energy simulations to refine and validate docking results when selecting the most promising 3CLpro inhibitor candidates. The top three candidates for 3CLpro repurposing were levothyroxine, amobarbital, and ABP-700 (Table S3).

In the study conducted by Kumar et al., MD simulation analysis was performed using Biovia Discovery Studio [70]. First, they screened the ADMET properties of six compounds; the top three candidate drugs (CASAntiV-865453-58-3, CASAntiV-865453-40-3, and CASAntiV-2043031-84-9) were subjected to MD simulations for 100 ns (Table S3), and various parameters were calculated from the resulting trajectories. Notably, the ligand-bound systems exhibited lower RMSD, RMSF, and Rg values, indicating their increased stability. Furthermore, the binding energies of the ligand-protein complexes were assessed using the thermodynamic MM/GBSA approach, which considers all MD trajectories. The calculated binding energies strongly indicated that two compounds (CASAntiV-865453-58-3 and CASAntiV-865453-40-3) formed strong bindings with the protein, with binding energy values of −148.1 kJ/mol and −161.3 kJ/mol, respectively. These findings suggest that these compounds have a high affinity for SARS-CoV-2 3CLpro and are promising drug candidates.

Similarly, Amin et al. performed MD simulations on four newly designed 3CLpro inhibitors (Table S3). Composite systems of the four molecules combined with 3CLpro exhibited lower RMSD, RMSF, and Rg values [72]. Surprisingly, D2 exhibited a more stable binding state than the other compounds. At the same time, MM-PBSA tool calculated the binding energy of four complexes in a 20ns MD simulation, and the binding energy of compound D4 was −238.3 ± 1.9 kJ/mol, indicating that compound D4 had a more satisfactory binding affinity with 3CLpro.

Daoui et al. evaluated the ADMET properties of 11 newly designed compounds against SARS-CoV-2 3CLpro before MD simulation; only D06 and D30 were unsuccessful (see 4.3 for detailed information). The MM/GBSA method was also used to calculate the binding free energies of these compounds, except for D06 and D30 (Table S3). The binding free energies of compound D08, D23 and D76 were better, which were −230.5 kJ/mol, −248.9 kJ/mol, and −239.5 kJ/mol, respectively. A 100ns MD simulation was applied to the three compounds. Lower RMSD and RMSF values indicated that all three compounds could stably bind to SARS-CoV-2 3CLpro. These results indicated that small molecules using the 9,10-dihydrophenanthrene fragment as a template may be candidate drugs for inhibiting the activity of SARS-CoV-2 3CLpro [73].

Zarezade et al. evaluated the stability of hACE2-ZINC12562757, 3CLpro-112,260,215, and 3CLpro-Evo_1 using a 50 ns MD simulation (Table S3) [91]. The RMSD and RMSF values were consistently low during the simulation, indicating that the hACE2-ZINC12562757, 3CLpro-112,260,215, and 3CLpro-Evo_1 complexes exhibited good stability. The binding energies of ZINC12562757, 112,260,215 and Evo_1 were −63.0 ± 27.3 kJ/mol, −92.7 ± 13.0 kJ/mol and −159.8 ± 96.4 kJ/mol, respectively.

There are many structural fragments in the literature that have been shown to positively contribute to the inhibitory activity of compounds [64,66,75]. Researchers can design new compounds on the basis of these substructures, and subsequently utilizing these compounds directly for docking and MD. For example, Saleh designed 12 new peptidomimetic fullerene-based derivatives [106], and the docking results showed that these compounds had a high binding affinity (−1175.4 kJ/mol to −607.2 kJ/mol) for 3CLpro. In MD studies, the Gly-α and Ser-α showed good stability and strong binding affinity to 3CLpro. Unfortunately, the 12 new compounds did not perfectly comply with the adoption of the Lipsky rule, then may not be developed as oral drugs.

4.3. Drug-like properties

The activity of a compound is a primary consideration in rational drug design. However, the drug properties and their impact on human health are crucial, particularly for virtual screening approaches. It is essential to identify compounds that not only exhibit high inhibitory activity against the virus but also demonstrate favourable safety profiles and minimal harm to the human body. The failure to meet these criteria can hinder the development of new drugs. In computer-aided screening for potential drugs against COVID-19, it is important to consider both the antiviral efficacy and the impact of the compound on human health. Various descriptors have been developed to characterise drug-like properties; among them, Lipinski's five rules are widely referenced. These rules provide a detailed framework for evaluating the compatibility of drugs with the human body. The Lipinski five rules, also known as the Rule of Five, assess properties such as the molecular weight (MW), lipophilicity (logP), hydrogen bond donors (HBD), hydrogen bond acceptors (HBA), and polar surface area (PSA). These properties help predict the absorption, distribution, metabolism, excretion, and toxicity (ADMET) characteristics of a compound. By adhering to these rules, researchers can assess the likelihood of a compound being orally active and having good bioavailability [107,108].

The Lipinski's five rules and other relevant drug-like properties allow researchers to prioritise compounds with a higher chance of success in terms of both activity against the virus and compatibility with human health. This comprehensive approach will promote the development of safer and more effective drugs against COVID-19.

In the experimental screening of new effective inhibitors of SARS-CoV-2, the "Lipinski five rules" and ADMET properties are widely utilised criteria to assess the potential applicability of screened compounds in human medicine. For the 20 newly designed anti-SARS-CoV compounds derived from the QSAR predictions by Chtita et al., the pkCSM online tool (http://biosig.unimelb.edu.au/pkcsm/) was employed to evaluate the physical and chemical properties of these compounds [59]. Remarkably, all 20 new compounds

conformed to Lipinski's five rules. However, in the assessment of the ADMET properties, only 13 of the 20 new compounds met these requirements. These 13 compounds (Table S3), which demonstrated favourable ADMET profiles, were subsequently selected as promising inhibitor candidates for 3CLpro of SARS-CoV. Considering the high similarity between the 3CLpro of SARS-CoV and that of SARS-CoV-2, these 13 compounds have the potential to aid in the treatment of COVID-19. Similarly, the three compounds (Table S3) screened by Oubahmane et al. had good ADMET properties, and all three compounds had better penetration scores through the blood-brain barrier (BBB) than the reference compound [63].

In addition to the "five Lipinski rules," Veber et al. proposed an additional set of rules for assessing the drug likeness of a molecule [109]. These rules provide additional criteria for evaluating the likelihood of oral bioavailability. According to "Veber Rules", a molecule is considered acceptable if it satisfies the following conditions: having 10 or fewer rotatable bonds, a polar surface area equal to or less than 140 Å2, and 12 or fewer hydrogen donors and acceptors. In the research conducted by Ničkčović et al., all five newly designed molecules (Table S3) were found to comply with both the "Lipinski five rules" and "Veber Rules" [66]. This indicates that these molecules possess desirable drug-like properties and are likely to exhibit satisfactory oral bioavailability. By meeting these criteria, the molecules are more likely to be effectively absorbed, distributed, and metabolised within the body, thereby increasing their potential as viable candidates for drug development.

Ghaleb et al. examined the ADMET properties of the top three compounds in molecular docking [69]. They found that the three newly designed compounds were easily absorbed, but did not easily cross the BBB, and only compound A5 showed high Caco-2 permeability. Compounds A9 and A10 were easily metabolised in the liver, and these three compounds showed no AMES toxicity. A comprehensive evaluation confirmed that compound A5 showed better potential as an anti-COVID-19 drug (Table S3). Daoui et al. evaluated the ADMET properties of 11 newly designed compounds against the SARS-CoV-2 3CLpro [73]. However, they used Osiris

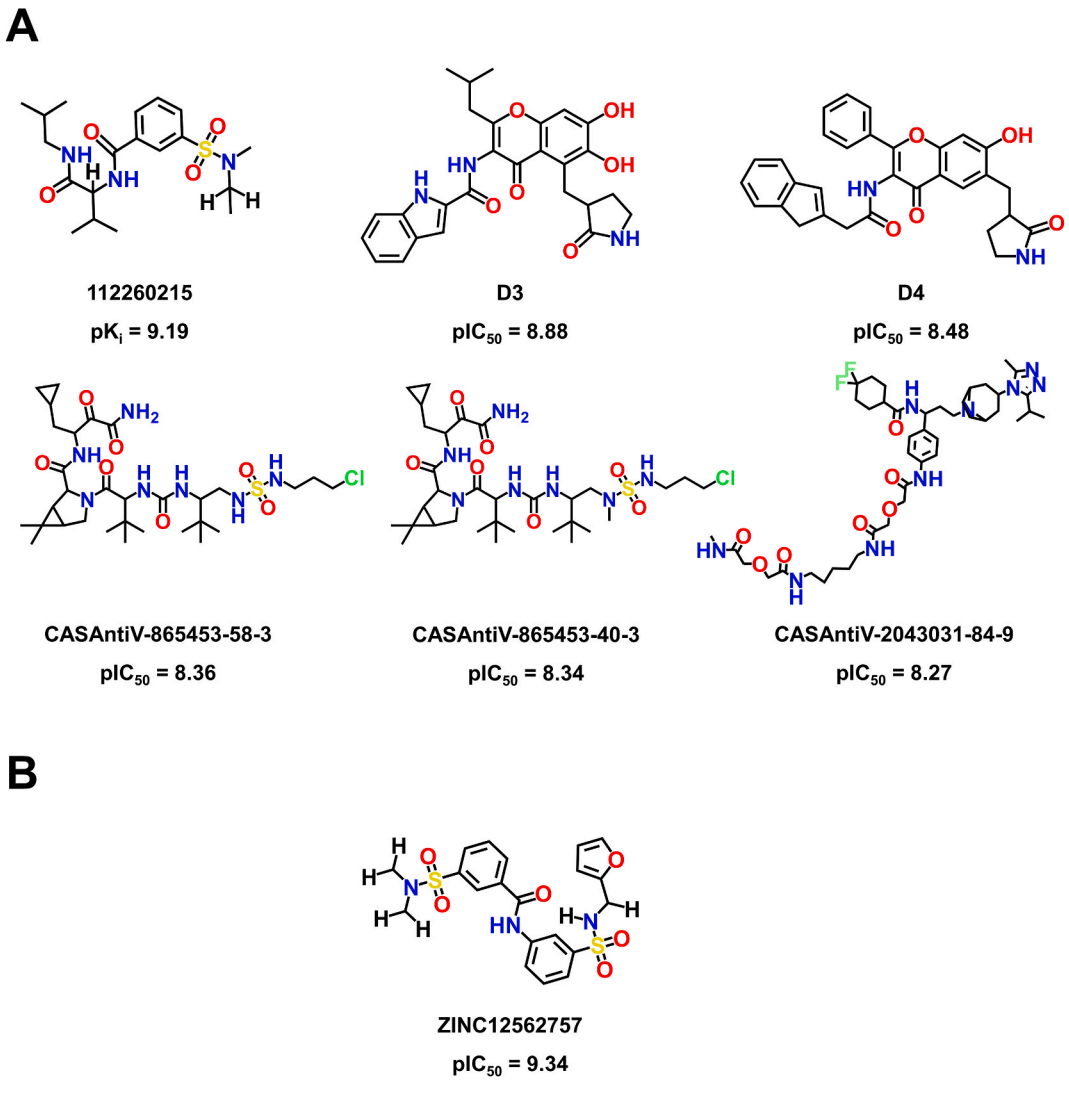


Fig. 7. The most promising inhibitors of **(A)** SARS-COV-2 3CLpro **(B)** and hACE2/S protein.

computations (https://www.organic-chemistry.org/prog/peo/) to predict the compound toxicity. Unfortunately, compounds D06 and D30 did not successfully pass the toxicity predictions because the predicted results showed that D06 may have slight toxicity as an irritant, whereas D30 is potentially mutagenic and carcinogenic. Thus, only the other nine compounds were used for MM-GBSA and MD simulations (Table S3).

In the toxicity filtering of QSAR identified PPI inhibitors by Pirolli et al., they employed the Cramer classification to categorise a total of 423565 compounds based on their toxicity grades [89]. The Cramer classification (https://www.vegahub.eu/) is a commonly used method for predicting the potential toxicity of chemical molecules. According to their *in silico* toxicological analysis, the classification results revealed the distribution of the compounds across different toxicity classes. Of the 423565 compounds analysed, 400453 were assigned to Class III, indicating that they were classified as highly toxic; 20025 compounds were categorised as Class I, indicating they were deemed non-toxic; and 3086 compounds were assigned to Class II, representing an intermediate toxicity level.

In a study by Zarezade et al., 492697 ligands were screened using the Lipinski and Veber rules, with 414631 ligands passing the standards, followed by docking at the active site of hACE2 [91]. Torres et al. evaluated 77 compounds, half of which violated Lipinski's five rules [94]. Similarly, they used Osiris computations to screen for toxicity, and compounds containing halogenated pyridine rings passed Osiris testing. Although these conclusions are specific to SARS-CoV, they can still provide support the development of COVID-19 drugs.

With advancements in computer science and the continuous improvement of new drug-screening techniques, virtual screening has emerged as a key technology in computer-aided drug design. In the search for SARS-CoV-2 inhibitors, virtual screening enables rapid narrowing of compound libraries, significantly expediting the initial stages of drug discovery. Various computational methods such as molecular docking, molecular dynamics (MD) simulations, and drug-like evaluation have been employed to predict compounds with promising activity against targets. It is important to note that identification of active compounds through virtual screening is the initial step. Further experimental investigations and validations are necessary to assess the potential of these compounds. The subsequent stages may involve additional experimental techniques, preclinical studies, and clinical trials to determine the suitability of a compound as a new anti-COVID-19 drug [110]. However, the use of chemometric models and computational tools allows for the efficient screening of large compound databases, thereby enhancing the overall process. By leveraging these models, researchers can quickly filter and prioritise potential candidates, saving time and resources in subsequent experimental and clinical phases. This accelerated screening process aids in the identification of promising compounds and expedites the drug discovery pipeline, ultimately contributing to the development of effective treatments for COVID-19 and other diseases.

Fig. 7 presents a summary of the most promising compounds identified as potential anti-COVID-19 drugs through a series of screening steps, including QSAR, molecular docking, MD simulations, and drug-like evaluations. In Fig. 7A, six representative compounds are highlighted that showed highly predictive inhibition of the 3CLpro enzyme of SARS-CoV-2. In Fig. 7B, only one compound with clear predictive inhibitory activity against hACE2 is shown. It is important to note that screening studies for inhibitors of the hACE2/S protein complex are relatively limited compared to those for 3CLpro. Therefore, the number of potential inhibitors identified for the hACE2/S protein complex was smaller than that identified for 3CLpro, and existing studies on this target may employ different endpoints for evaluation. Fig. 7 provides a visual representation of the most promising compounds that emerged from the screening process targeting specific proteins involved in SARS-CoV-2 infection. These compounds demonstrated potential inhibitory activities and are being further explored and evaluated for their effectiveness as anti-COVID-19 drugs.

5. Opportunities and challenges of computational modelling methods

The emergence of some variants poses an even greater challenge to our prevention and treatment of COVID-19. However, it has been found that the SARS-CoV-2 gene sequence is highly conserved, and the variants have a high degree of genetic similarity to the original virus, which indicates that inhibitors of the original virus may also have an inhibitory effect on the variants [111]. Although the QSAR modelling mainly relies on the original experimental data, we can further investigate the potential inhibitors by combining chemometrics, which is a combination of QSAR modelling, molecular docking and MD simulation. Molecular docking and MD simulation can directly target compounds to the protein crystals of variants, analyse the binding affinity and stability of compounds to different variant proteins, and directly assess whether the compounds can interact with the new variants [112,113].

In the process of drug design and development, computational modelling methods such as artificial intelligence (AI) have become an important part. AI such as machine learning is often used to discover new drugs or to improve the therapeutic efficacy of chemicals [114], it can accelerate the drug discovery and design process, and save the cost, especially in virtual screening, which greatly reduces the number of experiments and the waste of time and resources [115]. For huge datasets, AI can rapidly integrate multivariate information, jointly analyse the knowledge from different domains, and provide accurate and personalized predictions for special objects. However, as a theoretical drug development method, it may not be able to break through the barrier of ethics and regulations for the time being [116]. It is also constrained by limited experimental data, and the quality and quantity of which can directly affect the performance of computational modelling methods. At the same time, most modelling algorithms such as machine learning are black boxes, with no detailed explanation of the decision-making process, affecting the credibility of the results [117]. In conclusion, AI provide efficient and innovative ways for drug design and development, and provide strong support for clinical trials, but also face challenges in data quality, model interpretability, and regulations. If overcoming these challenges, computational modelling methods, such as AI, may play a greater role in drug development.

6. Conclusion and Prospect

The ongoing global concern surrounding SARS-CoV-2 necessitates the continuous research and development of effective inhibitors against the virus. The emergence of viral mutations has added to this uncertainty and poses a threat to public health. Vulnerable populations, such as the elderly, children, and individuals with underlying health conditions, are particularly at risk and may experience severe illness or even death due to SARS-CoV-2 infection. This paper provides a comprehensive review of research progress in the field of QSAR-based inhibitors targeting SARS-CoV-2 3CLpro as well as the hACE2/S protein interaction. Through the establishment and refinement of QSAR models, novel potential inhibitors were identified and subsequently subjected to molecular docking, MD simulations, and drug-like evaluations. These *in silico* screening methods hold promise for identifying compounds that may serve as effective COVID-19 specific drugs in the future.

In addition to 3CLpro and hACE2/S protein interactions, RNA-dependent RNA polymerase (RdRp), N protein N-terminal domain (N-NTD), and papain-like protease (PLpro) are popular targets for inhibitor development [118,124]. Remdesivir, a drug specific for COVID-19, exerts its inhibitory effects by targeting RdRp [119,120]. Although some QSAR models based on experimental data are available for RdRp and PLpro inhibitors, the amount of data is relatively limited compared to that for 3CLpro.

Several recently published papers have reviewed the development of potential inhibitors of SARS-CoV-2 from different perspectives (Table 2). Katre et al. reviewed the development of potential inhibitors targeting 3CLpro and focused on the antiviral activity of two inhibitors, favouring the experimental direction [121]. Similarly, Banerjee et al. reviewed potential inhibitors of 3CLpro and focused on theoretical calculations for a large number of inhibitors [122]. Gao et al. and Rolta et al. reviewed the development of potential inhibitors for many targets and both focused on the theoretical calculations (QSAR, Docking, MD, etc.) of potential inhibitors [123,124]. The difference is that Gao et al. covered a large number of potential inhibitors and included potential biological drugs (vaccines), whereas Rolta et al. described only the development of methylxanthines as a class of potential inhibitors. In addition to the 3CLpro target, PLpro is also a popular target for antiviral inhibitors; Calleja et al. and Amin et al. detailed the development of theoretical calculations for potential inhibitors of PLpro in their reviews [125,126]. Elkashlan et al. described several large databases for studying potential inhibitors of SARS-CoV-2 as well as the application and development of machine learning models [127]. Our review focuses on chemometric modelling as an important point, focusing on the protein targets 3CLpro, hACE2/S protein, and PPI, we detailed a series of studies on QSAR modelling, molecular docking, molecular dynamics, and ADMET around chemical potential inhibitors of these targets.

Owing to the collective efforts of scientists worldwide, many drugs with significant inhibitory activity against SARS-CoV-2 are currently undergoing clinical trials and some have even been approved for marketing. The continuous development of these drugs has brought both hope and pressure to fight SARS-CoV-2. Identifying the most suitable compound from a large pool of potential candidates is time-consuming and requires significant resources. In this regard, the application of chemometric methods, such as quantitative structure-activity relationship (QSAR), molecular docking, molecular dynamics (MD) simulations, and drug-like evaluation, can assist researchers in narrowing the scope of investigation and provide valuable supporting data for future studies.

Indeed, the vast number of compounds that have the potential to inhibit the activity of SARS-CoV or SARS-CoV-2 makes traditional experimental screening processes time-consuming and resource-intensive. Therefore, the use of reliable QSAR models based on computational methods is a valuable solution. By utilizing known inhibitor activity data, a QSAR model can rapidly screen large compound databases, significantly reducing the need for extensive experimental testing. The QSAR model serves as an initial filter, narrowing down the range of compounds with unknown activities and prioritising those with a higher potential for further investigation. Subsequent steps, such as molecular docking and MD simulations, were employed to optimise and validate the results obtained from the QSAR model. These techniques provide detailed insights into the interactions between compounds and their target proteins, helping to refine the selection of candidate compounds. Moreover, the QSAR approach allows the identification and analysis of molecular structural fragments that play a crucial role in the activity of compounds. This information provides valuable guidance for the design and development of new inhibitors, and facilitates the exploration of novel drug candidates for the treatment of COVID-19 and other coronaviruses.

Overall, the integration of QSAR models, molecular docking, and MD simulations not only accelerates the screening process but also offers a cost-effective approach to resource management. This enables the identification of potential compounds with desired activities, optimises their drug-like properties, and provides important insights into the development of effective drugs to combat

Table 2
Information derived from SARS-CoV-2 inhibitors-related reviews published in the last few years.

Ref.	Target	Number of Inhibitors	Inhibitor type	Method
[121]	3CLpro	2	Chemical	Experimental
[122]	3CLpro	Massive	Chemical	Theoretical (Docking/MD) Experimental
[123]	3CLpro/S protein/RdRp/PLpro	Massive	Chemical/Biological	Theoretical (QSAR/Docking/MD/Machine Learning)
[124]	3CLpro/S protein/N-NTD	7	Chemical (Methylxanthines)	Theoretical (Docking/MD/ADMET)
[125]	PLpro	14	Chemical	Theoretical (SAR)
[126]	3CLpro/PLpro	Massive	Chemical	Theoretical (QSAR)
[127]	SARS-CoV-2	Massive	Chemical	Theoretical (Machine Learning)
This Review	**3CLpro/hACE2/S protein/ PPI**	**Massive**	**Chemical**	**Theoretical (QSAR/Docking/MD/ADMET)**

coronaviruses.

CRediT authorship contribution statement

Qianqian Wang: Writing – original draft. **Xinyi Lu:** Data curation. **Runqing Jia:** Writing – original draft. **Xinlong Yan:** Writing – original draft. **Jianhua Wang:** Writing – review & editing. **Lijiao Zhao:** Data curation. **Rugang Zhong:** Writing – review & editing. **Guohui Sun:** Writing – review & editing, Supervision, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (No. 82003599) and The Project of Cultivation for Young Top-Motch Talents of Beijing Municipal Institutions (No. BPHR202203016), Science and Technology General Project of the Beijing Municipal Education Commission (No. KM202110005005), and the Beijing Natural Science Foundation (No. 7242193, 7222016).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.heliyon.2024.e24209.

References

[1] P. Zhou, X.L. Yang, X.G. Wang, B. Hu, L. Zhang, W. Zhang, H.R. Si, Y. Zhu, B. Li, C.L. Huang, H.D. Chen, J. Chen, Y. Luo, H. Guo, R.D. Jiang, M.Q. Liu, Y. Chen, X.R. Shen, X. Wang, X.S. Zheng, K. Zhao, Q.J. Chen, F. Deng, L.L. Liu, B. Yan, F.X. Zhan, Y.Y. Wang, G.F. Xiao, Z.L. Shi, A pneumonia outbreak associated with a new coronavirus of probable bat origin, Nature 579 (2020) 270–273, https://doi.org/10.1038/s41586–020–2012–7.

[2] A.E. Gorbalenya, S.C. Baker, R.S. Baric, R.J. de Groot, C. Drosten, A.A. Gulyaeva, B.L. Haagmans, C. Lauber, A.M. Leontovich, B.W. Neuman, D. Penzar, S. Perlman, L.L.M. Poon, D.V. Samborskiy, I.A. Sidorov, I. Sola, J. Ziebuhr, C.S. Grp, The species Severe acute respiratory syndrome–related coronavirus: classifying 2019–nCoV and naming it SARS–CoV–2, Nat. Microbiol. 5 (2020) 536–544, https://doi.org/10.1038/s41564-020-0695-z.

[3] WHO, WHO Coronavirus (COVID–19) Dashboard , https://covid19.who.int/data (accessed on 10 September 2023).

[4] Q. Wang, S. Iketani, Z. Li, L. Liu, Y. Guo, Y. Huang, A.D. Bowen, M. Liu, M. Wang, J. Yu, R. Valdez, A.S. Lauring, Z. Sheng, H.H. Wang, A. Gordon, L. Liu, D. D. Ho, Alarming antibody evasion properties of rising SARS–CoV–2 BQ and XBB subvariants, Cell 186 (2023) 279–286, https://doi.org/10.1016/j.cell.2022.12.018.

[5] M. Mohammed, COVID–19: What we know about new omicron variant BF .7, https://medicalxpress.com/news/2022-12-covid-omicron-variant-bf7.html.

[6] S. Chatterjee, M. Bhattacharya, S. Nag, K. Dhama, C. Chakraborty, A detailed overview of SARS–CoV–2 Omicron: its sub–variants, mutations and pathophysiology, clinical characteristics, immunological landscape, immune escape, and therapies, Viruses 15 (2023) 167, https://doi.org/10.3390/v15010167.

[7] M. Sabbatucci, A. Vitiello, S. Clemente, A. Zovi, M. Boccellino, F. Ferrara, C. Cimmino, R. Langella, A. Ponzo, P. Stefanelli, G. Rezza, Omicron variant evolution on vaccines and monoclonal antibodies, Inflammopharmacology 31 (2023) 1779–1788, https://doi.org/10.1007/s10787-023-01253-6.

[8] R. Lu, X. Zhao, J. Li, P. Niu, B. Yang, H. Wu, W. Wang, H. Song, B. Huang, N. Zhu, Y. Bi, X. Ma, F. Zhan, L. Wang, T. Hu, H. Zhou, Z. Hu, W. Zhou, L. Zhao, J. Chen, Y. Meng, J. Wang, Y. Lin, J. Yuan, Z. Xie, J. Ma, W.J. Liu, D. Wang, W. Xu, E.C. Holmes, G.F. Gao, G. Wu, W. Chen, W. Shi, W. Tan, Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding, Lancet 395 (2020) 565–574, https://doi.org/10.1016/S0140-6736(20)30251-8.

[9] F. Wu, S. Zhao, B. Yu, Y.M. Chen, W. Wang, Z.G. Song, A new coronavirus associated with human respiratory disease in China, Nature 579 (2020) 265–269, https://doi.org/10.1038/s41586-020-2008-3.

[10] Y.W. Chen, C.B. Yiu, K.Y. Wong, Prediction of the SARS–CoV–2 (2019–nCoV) 3C–like protease (3CL (pro)) structure: virtual screening reveals velpatasvir, ledipasvir, and other drug repurposing candidates 9 (2020) 129, https://doi.org/10.12688/f1000research.22457.2. F1000Res.

[11] A. Wu, Y. Peng, B. Huang, X. Ding, X. Wang, P. Niu, J. Meng, Z. Zhu, Z. Zhang, J. Wang, J. Sheng, L. Quan, Z. Xia, W. Tan, G. Cheng, T. Jiang, Genome composition and divergence of the novel coronavirus (2019–nCoV) originating in China, Cell Host Microbe 27 (2020) 325–328, https://doi.org/10.1016/j.chom.2020.02.001.

[12] J. Lei, Y. Kusov, R. Hilgenfeld, Nsp3 of coronaviruses: structures and functions of a large multi–domain protein, Antiviral Res 149 (2018) 58–74, https://doi.org/10.1016/j.antiviral.2017.11.001.

[13] K. Anand, J. Ziebuhr, P. Wadhwani, J.R. Mesters, R. Hilgenfeld, Coronavirus main proteinase (3CLpro) structure: basis for design of anti–SARS drugs, Science 300 (2003) 1763–1767, https://doi.org/10.1126/science.1085658.

[14] I.M. Jahirul, I.N. Nawal, A.M. Siddik, M. Kabir, M.A. Halim, A review on structural, non–structural, and accessory proteins of SARS–CoV–2: highlighting drug target sites, Immunobiology 228 (2023) 152302, https://doi.org/10.1016/j.imbio.2022.152302.

[15] J. Osipiuk, S.A. Azizi, S. Dvorkin, M. Endres, R. Jedrzejczak, K.A. Jones, S. Kang, R.S. Kathayat, Y. Kim, V.G. Lisnyak, S.L. Maki, V. Nicolaescu, C.A. Taylor, C. Tesar, Y.A. Zhang, Z. Zhou, G. Randall, K. Michalska, S.A. Snyder, B.C. Dickinson, A. Joachimiak, Structure of papain–like protease from SARS–CoV–2 and its complexes with non–covalent inhibitors, Nat. Commun. 12 (2021) 743, https://doi.org/10.1038/s41467-021-21060-3.

[16] D. Shin, R. Mukherjee, D. Grewe, D. Bojkova, K. Baek, A. Bhattacharya, L. Schulz, M. Widera, A.R. Mehdipour, G. Tascher, P.P. Geurink, A. Wilhelm, G.J. Van Noort, van der Heden, H. Ovaa, S. Muller, K.P. Knobeloch, K. Rajalingam, B.A. Schulman, J. Cinatl, G. Hummer, S. Ciesek, I. Dikic, Papain–like protease regulates SARS–CoV–2 viral spread and innate immunity, Nature 587 (2020) 657–662, https://doi.org/10.1038/s41586-020-2601-5.

[17] K. Amporndanai, X. Meng, W. Shang, Z. Jin, M. Rogers, Y. Zhao, Z. Rao, Z.J. Liu, H. Yang, L. Zhang, P.M. O'Neill, Hasnain S. Samar, Inhibition mechanism of SARS–CoV–2 main protease by ebselen and its derivatives, Nat. Commun. 12 (2021) 3061, https://doi.org/10.1038/s41467-021-23313-7.

[18] C.R. Wu, W.C. Yin, Y. Jiang, H.E. Xu, Structure genomics of SARS–CoV–2 and its Omicron variant: drug design templates for COVID–19, Acta Pharmacol. Sin. 43 (2022) 3021–3033, https://doi.org/10.1038/s41401-021-00851-w.

[19] R. Arya, S. Kumari, B. Pandey, H. Mistry, S.C. Bihani, A. Das, V. Prashar, G.D. Gupta, L. Panicker, M. Kumar, Structural insights into SARS–CoV–2 proteins, J. Mol. Biol. 433 (2021) 166725, https://doi.org/10.1016/j.jmb.2020.11.024.

[20] H. Yang, Z. Rao, Structural biology of SARS–CoV–2 and implications for therapeutic development, Nat. Rev. Microbiol. 19 (2021) 685–700, https://doi.org/10.1038/s41579-021-00630-8.

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

[22] F. Li, Structure, function, and evolution of coronavirus spike proteins, Annu. Rev. Virol. 3 (2016) 237–261, https://doi.org/10.1146/annurev-virology-110615-042301.

[23] X. He, C. He, W. Hong, K. Zhang, X. Wei, The challenges of COVID–19 Delta variant: prevention and vaccine development, MedComm 2 (2021) 846–854, https://doi.org/10.1002/mco2.95.

[24] S. Nasreen, H. Chung, S. He, K.A. Brown, J.B. Gubbay, S.A. Buchan, D.B. Fell, P.C. Austin, K.L. Schwartz, M.E. Sundaram, A. Calzavara, B. Chen, M. Tadrous, K. Wilson, S.E. Wilson, J.C. Kwong, Effectiveness of COVID–19 vaccines against symptomatic SARS–CoV–2 infection and severe outcomes with variants of concern in Ontario, Nat. Microbio. l 7 (2022) 379–385, https://doi.org/10.1038/s41564-021-01053-0.

[25] Y. Zhou, H. Wang, L. Yang, Q. Wang, Progress on COVID-19 chemotherapeutics discovery and novel technology, Molecules 27 (2022) 8257, https://doi.org/10.3390/molecules27238257.

[26] M. Li, H. Wang, L. Tian, Z. Pang, Q. Yang, T. Huang, J. Fan, L. Song, Y. Tong, H. Fan, COVID–19 vaccine development: milestones, lessons and prospects, Signal. Transduct. Target. Ther. 7 (2022) 146, https://doi.org/10.1038/s41392-022-00996-y.

[27] R.L. Hoffman, R.S. Kania, M.A. Brothers, J.F. Davies, R.A. Ferre, K.S. Gajiwala, M. He, R.J. Hogan, K. Kozminski, L.Y. Li, J.W. Lockner, J. Lou, M.T. Marra, LJ Jr Mitchell, B.W. Murray, J.A. Nieman, S. Noell, S.P. Planken, T. Rowe, K. Ryan, GJ Rd Smith, J.E. Solowiej, C.M. Steppan, B. Taggart, Discovery of ketone-based covalent inhibitors of coronavirus 3CL proteases for the potential therapeutic treatment of COVID-19, J. Med. Chem. 63 (2020) 12725–12747, https://doi.org/10.1021/acs.jmedchem.0c01063.

[28] D.R. Owen, C. Allerton, A.S. Anderson, L. Aschenbrenner, M. Avery, S. Berritt, B. Boras, R.D. Cardin, A. Carlo, K.J. Coffman, A. Dantonio, L. Di, H. Eng, R. Ferre, K.S. Gajiwala, S.A. Gibson, S.E. Greasley, B.L. Hurst, E.P. Kadar, A.S. Kalgutkar, J.C. Lee, J. Lee, W. Liu, S.W. Mason, S. Noell, J.J. Novak, R.S. Obach, K. Ogilvie, N.C. Patel, M. Pettersson, D.K. Rai, M.R. Reese, M.F. Sammons, J.G. Sathish, R.S.P. Singh, C.M. Steppan, A.E. Stewart, J.B. Tuttle, L. Updyke, P. R. Verhoest, L. Wei, Q. Yang, Y. Zhu, An oral SARS–CoV–2 M(pro) inhibitor clinical candidate for the treatment of COVID–19, Science 374 (2021) 1586–1593, https://doi.org/10.1126/science.abl4784.

[29] B.R. Casey, R.C. Vernick, A. Bahekar, D. Patel, A.I. Ncogo, Ranolazine toxicity secondary to Paxlovid, Cureus 15 (2023) e37153, https://doi.org/10.7759/cureus.37153.

[30] Us FDA, Coronavirus (COVID–19) Update: FDA authorizes first oral antiviral for treatment of COVID–19. https://www.fda.gov/news-events/press-announcements/coronavirus-covid-19-update-fda-authorizes-first-oral-antiviral-treatment-covid-19 (accessed on 22 December 2021).

[31] Us FDA, Coronavirus (COVID–19) Update: FDA authorizes pharmacists to prescribe Paxlovid with certain limitations. https://www.fda.gov/news-events/press-announcements/coronavirus-covid-19-update-fda-authorizes-pharmacists-prescribe-paxlovid-certain-limitations (accessed on 6 July 2022).

[32] NMPA. https://www.nmpa.gov.cn/zhuanti/yqyjxzd/yqyjxd/20220212085753142.html (accessed on 11 February 2022).

[33] E.Y. Dai, K.A. Lee, A.B. Nathanson, A.T. Leonelli, B.A. Petros, T. Brock–Fisher, S.T. Dobbins, B.L. MacInnis, A. Capone, N. Littlehale, J. Boucau, C. Marino, A. K. Barczak, P.C. Sabeti, M. Springer, K.E. Stephenson, Viral kinetics of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) omicron infection in mRNA-vaccinated individuals treated and not treated with nirmatrelvir-ritonavir (2022) 22278378, https://doi.org/10.1101/2022.08.04.22278378 medRxiv [Preprint].

[34] N. Ranganath, J.C. O'Horo, D.W. Challener, S.M. Tulledge–Scheitel, M.L. Pike, M. O'Brien, R.R. Razonable, A. Shah, Rebound phenomenon after nirmatrelvir/ritonavir treatment of coronavirus disease 019 (COVID-19) in high-risk persons, Dis 76 (2023) 537–539, https://doi.org/10.1093/cid/ciac481.

[35] M.E. Charness, K. Gupta, G. Stack, J. Strymish, E. Adams, D.C. Lindy, H. Mohri, D.D. Ho, Rebound of SARS-CoV-2 infection after nirmatrelvir-ritonavir treatment, N. Engl. J. Med. 387 (2022) 1045–1047, https://doi.org/10.1056/NEJMc2206449.

[36] J.A. Hay, S.M. Kissler, J.R. Fauver, C. Mack, C.G. Tai, R.M. Samant, S. Connolly, D.J. Anderson, G. Khullar, M. MacKay, M. Patel, S. Kelly, A. Manhertz, T. Eiter, D. Salgado, T. Baker, B. Howard, J.T. Dudley, C.E. Mason, M. Nair, Y. Huang, J. DiFiori, D.D. Ho, N.D. Grubaugh, Y.H. Grad, Quantifying the impact of immune history and variant on SARS–CoV–2 viral kinetics and infection rebound: a retrospective cohort study, medRxiv 11 (2022) e81849, https://doi.org/10.7554/eLife.81849.

[37] H. Chan, H. Shan, T. Dahoun, H. Vogel, S. Yuan, Advancing drug discovery via artificial intelligence, Trends, Pharmacol. Sci. 40 (2019) 592–604, https://doi.org/10.1016/j.tips.2019.06.004.

[38] M.N. Gomes, R.C. Braga, E.M. Grzelak, B.J. Neves, E. Muratov, R. Ma, L.L. Klein, S. Cho, G.R. Oliveira, S.G. Franzblau, C.H. Andrade, QSAR–driven design, synthesis and discovery of potent chalcone derivatives with antitubercular activity, Eur. J. Med. Chem. 137 (2017) 126–138, https://doi.org/10.1016/j.ejmech.2017.05.026.

[39] M.H. Baig, K. Ahmad, S. Roy, J.M. Ashraf, M. Adil, M.H. Siddiqui, S. Khan, M.A. Kamal, I. Provaznik, I. Choi, Computer aided drug design: success and limitations, Curr. Pharm. Des. 22 (2016) 572–581, https://doi.org/10.2174/1381612822666151125000550.

[40] A.B. Gurung, M.A. Ali, J. Lee, M.A. Farah, K.M. Al–Anazi, An updated review of computer-aided drug design and its application to COVID-19, BioMed Res. Int. 21 (2021) 8853056, https://doi.org/10.1155/2021/8853056.

[41] Y.L. Ng, C.K. Salim, J. Chu, Drug repurposing for COVID–19: approaches, challenges and promising candidates, Pharmacol. Ther. 228 (2021) 107930, https://doi.org/10.1016/j.pharmthera.2021.107930.

[42] S. Chen, F. Li, G. Sun, L. Zhao, R. Zhong, QSAR modeling and its advances in antiviral drug design and screening, Chem. Reagents 43 (2021) 895–905, https://doi.org/10.13822/j.cnki.hxsj.2021008038.

[43] A.P. Worth, A. Bassan, J. De Bruijn, S.A. Gallegos, T. Netzeva, G. Patlewicz, M. Pavan, I. Tsakovska, S. Eisenreich, The role of the European Chemicals Bureau in promoting the regulatory use of (Q)SAR methods, SAR, QSAR Environ. Res. 18 (2007) 111–125, https://doi.org/10.1080/10629360601054255.

[44] P. De, S. Bhayye, V. Kumar, K. Roy, In silico modeling for quick prediction of inhibitory activity against 3CL (pro) enzyme in SARS CoV diseases, J. Biomol. Struct. Dyn. 40 (2022) 1010–1036, https://doi.org/10.1080/07391102.2020.1821779.

[45] C.A. Lipinski, Lead– and drug–like compounds: the rule–of–five revolution, Drug Discov. Today Technol. 1 (2004) 337–341, https://doi.org/10.1016/j.ddtec.2004.11.007.

[46] C.A. Lipinski, F. Lombardo, B.W. Dominy, P.J. Feeney, Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings, Adv. Drug Deliv. Rev. 46 (2001) 3–26, https://doi.org/10.1016/s0169-409x(00)00129-0.

[47] Z. O'Brien, M.M. Fallah, Small molecule kinase inhibitors approved by the FDA from 2000 to 2011: a systematic review of preclinical ADME data, Expert. Opin. Drug Metab. Toxicol. 9 (2013) 1597–1612, https://doi.org/10.1517/17425255.2013.834046.

[48] S. Alam, S. Nasreen, A. Ahmad, M.P. Darokar, F. Khan, Detection of natural inhibitors against human liver cancer cell lines through QSAR, molecular docking and ADMET studies, Curr. Top Med. Chem. 21 (2021) 686–695, https://doi.org/10.2174/1568026620666201204155830.

[49] H. Lu, C.W. Stratton, Y.W. Tang, Outbreak of pneumonia of unknown etiology in Wuhan, China: the mystery and the miracle, J. Med. Virol. 92 (2020) 401–402, https://doi.org/10.1002/jmv.25678.

[50] N. Zhu, D. Zhang, W. Wang, X. Li, B. Yang, J. Song, X. Zhao, B. Huang, W. Shi, R. Lu, P. Niu, F. Zhan, X. Ma, D. Wang, W. Xu, G. Wu, G.F. Gao, W. Tan, A novel coronavirus from patients with pneumonia in China, 2019, N. Engl. J. Med. 382 (2020) 727–733, https://doi.org/10.1056/NEJMoa2001017.

[51] E. Petersen, M. Koopmans, U. Go, D.H. Hamer, N. Petrosillo, F. Castelli, M. Storgaard, S. Al Khalili, L. Simonsen, Comparing SARS–CoV–2 with SARS–CoV and influenza pandemics, Lancet Infect. Dis. 20 (2020) 238–244, https://doi.org/10.1016/S1473-3099(20)30484-9.

[52] J.C. Ferreira, S. Fadl, A.J. Villanueva, W.M. Rabeh, Catalytic dyad residues His41 and Cys145 impact the catalytic activity and overall conformational fold of the main SARS-CoV-2 protease 3-chymotrypsin-like protease, Front. Chem. 9 (2021) 692168, https://doi.org/10.3389/fchem.2021.692168.

[53] Z. Jin, X. Du, Y. Xu, Y. Deng, M. Liu, Y. Zhao, B. Zhang, X. Li, L. Zhang, C. Peng, Y. Duan, J. Yu, L. Wang, K. Yang, F. Liu, R. Jiang, X. Yang, T. You, X. Liu, X. Yang, F. Bai, H. Liu, X. Liu, L.W. Guddat, W. Xu, G. Xiao, C. Qin, Z. Shi, H. Jiang, Z. Rao, H. Yang, Structure of M(pro) from SARS–CoV–2 and discovery of its inhibitors, Nature 582 (2020) 289–293, https://doi.org/10.1038/s41586-020-2223-y.

[54] Q. Hu, Y. Xiong, G.H. Zhu, Y.N. Zhang, Y.W. Zhang, P. Huang, G.B. Ge, The SARS–CoV–2 main protease (M(pro)): structure, function, and emerging therapies for COVID–19, MedComm 3 (2022) e151, https://doi.org/10.1002/mco2.151.

[55] G. La Monica, A. Bono, A. Lauria, A. Martorana, Targeting SARS-CoV-2 main protease for treatment of COVID-19: covalent inhibitors structure-activity relationship in-sights and evolution perspectives, J. Med. Chem. 65 (2022) 12500–12534, https://doi.org/10.1021/acs.jmedchem.2c01005.

[56] L. Zhang, D. Lin, X. Sun, U. Curth, C. Drosten, L. Sauerhering, S. Becker, K. Rox, R. Hilgenfeld, Crystal structure of SARS–CoV–2 main protease provides a basis for design of improved alpha–ketoamide inhibitors, Science 368 (2020) 409–412, https://doi.org/10.1126/science.abb3405.

[57] Y. Chen, Q. Liu, D. Guo, Emerging coronaviruses: genome structure, replication, and pathogenesis, J. Med. Virol. 92 (2020) 418–423, https://doi.org/10.1002/jmv.25681.

[58] P.M. Khan, V. Kumar, K. Roy, In silico modeling of small molecule carboxamides as inhibitors of SARS-CoV 3CL protease: an approach towards combating COVID-19, Comb. Chem. High Throughput Screen. 24 (2021) 1281–1299, https://doi.org/10.2174/1386207323666200914094712.

[59] S. Chtita, A. Belhassan, M. Bakhouch, A.I. Taourati, A. Aouidate, S. Belaidi, M. Moutaabbid, S. Belaaouad, M. Bouachrine, T. Lakhlifi, QSAR study of unsymmetrical aromatic disulfides as potent avian SARS–CoV main protease inhibitors using quantum chemical descriptors and statistical methods, Chemometr. Intell. Lab. Syst. 210 (2021) 104266, https://doi.org/10.1016/j.chemolab.2021.104266.

[60] OECD, Guidance Document on the Validation of (Quantitative) Structure–Activity Relationship [(Q)SAR] Models. OECD Series on Testing and Assessment, No. 69. OECD Publishing, Paris https://www.oecd.org/env/guidance-document-on-the-validation-of-quantitative-structure-activity-relationship-q-sar-models-9789264085442-en.htm (accessed on 3 September 2014).

[61] A. Golbraikh, A. Tropsha, Beware of q2, J. Mol. Graph. Model. 20 (2002) 269–276, https://doi.org/10.1016/s1093-3263(01)00123-1.

[62] M. Kumari, N. Subbarao, Development of a deep learning–based quantitative structure–activity relationship model to identify potential inhibitors against the 3C–like protease of SARS–CoV–2, Future Med. Chem. 14 (2022) 1541–1559, https://doi.org/10.4155/fmc-2021-0063.

[63] M. Oubahmane, I. Hdoufane, I. Bjij, C. Jerves, D. Villemin, D. Cherqaoui, COVID–19: in silico identification of potent alpha–ketoamide inhibitors targeting the main protease of the SARS–CoV–2, J. Mol. Struct. 1244 (2021) 130897, https://doi.org/10.1016/j.molstruc.2021.130897.

[64] N. Soleymani, S. Ahmadi, F. Shiri, A. Almasirad, QSAR and molecular docking studies of isatin and indole derivatives as SARS 3CLpro inhibitors, BMC Chem 17 (2023) 32, https://doi.org/10.1186/s13065-023-00947-w.

[65] M. Zaki, S.A. Al–Hussain, V.H. Masand, S. Akasapu, S.O. Bajaj, N. El–Sayed, A. Ghosh, I. Lewaa, Identification of anti-SARS-CoV-2 compounds from food using QSAR-based virtual screening, molecular docking, and molecular dynamics simulation analysis, Pharmaceuticals 14 (2021) 354, https://doi.org/10.3390/ph14040357.

[66] V.P. Ničkčović, G.R. Nikolic, B.M. Nedeljkovic, N. Mitic, S.F. Danic, J. Mitic, Z. Marcetic, D. Sokolovic, A.M. Veselinovic, In silico approach for the development of novel antiviral compounds based on SARS–COV–2 protease inhibition, Chem. Zvesti 76 (2022) 4393–4404, https://doi.org/10.1007/s11696-022-02170-8.

[67] A.A. Ishola, O. Adedirin, T. Joshi, S. Chandra, QSAR modeling and pharmacoinformatics of SARS coronavirus 3C–like protease inhibitors, Comput. Biol. Med. 134 (2021) 104483, https://doi.org/10.1016/j.compbiomed.2021.104483.

[68] E. Tejera, C.R. Munteanu, A. Lopez–Cortes, A. Cabrera–Andrade, Y. Perez–Castillo, Drugs repurposing using QSAR, docking and molecular dynamics for possible inhibitors of the SARS-CoV-2 M(pro) protease, Molecules 25 (2020) 5172, https://doi.org/10.3390/molecules25215172.

[69] A. Ghaleb, A. Aouidate, H. Ayouchia, M. Aarjane, H. Anane, S.E. Stiriba, In silico molecular investigations of pyridine N–Oxide compounds as potential inhibitors of SARS–CoV–2: 3D QSAR, molecular docking modeling, and ADMET screening, J. Biomol. Struct. Dyn. 40 (2022) 143–153, https://doi.org/10.1080/07391102.2020.1808530.

[70] V. Kumar, S. Kar, P. De K. Roy, J. Leszczynski, Identification of potential antivirals against 3CLpro enzyme for the treatment of SARS–CoV–2: a multi–step virtual screening study, SAR, QSAR Environ. Res. 33 (2022) 357–386, https://doi.org/10.1080/1062936X.2022.2055140.

[71] A.T. Ton, F. Gentile, M. Hsing, F. Ban, A. Cherkasov, Rapid identification of potential inhibitors of SARS-CoV-2 main protease by deep docking of 1.3 billion compounds, Mol. Inform. 39 (2020) e2000028, https://doi.org/10.1002/minf.202000028.

[72] S.A. Amin, S. Banerjee, S. Singh, I.A. Qureshi, S. Gayen, T. Jha, First structure–activity relationship analysis of SARS–CoV–2 virus main protease (Mpro) inhibitors: an endeavor on COVID–19 drug discovery, Mol. Divers. 25 (2021) 1827–1838, https://doi.org/10.1007/s11030-020-10166-3.

[73] O. Daoui, S. Elkhattabi, S. Chtita, Rational identification of small molecules derived from 9,10–dihydrophenanthrene as potential inhibitors of 3CL (pro) enzyme for COVID–19 therapy: a computer–aided drug design approach, Struct. Chem. 33 (2022) 1667–1690, https://doi.org/10.1007/s11224-022-02004-z.

[74] N. Adhikari, S. Banerjee, S.K. Baidya, B. Ghosh, T. Jha, Ligand–based quantitative structural assessments of SARS–CoV–2 3CL (pro) inhibitors: an analysis in light of structure–based multi–molecular modeling evidences, J. Mol. Struct. 1251 (2022) 132041, https://doi.org/10.1016/j.molstruc.2021.132041.

[75] M. Oubahmane, I. Hdoufane, C. Delaite, A. Sayede, D. Cherqaoui, A. El Allali, Design of potent inhibitors targeting the main protease of SARS-CoV-2 using QSAR modeling, molecular docking, and molecular dynamics simulations, Pharmaceuticals 16 (2023) 608, https://doi.org/10.3390/ph16040608.

[76] V. Kumar, K. Roy, Development of a simple, interpretable and easily transferable QSAR model for quick screening antiviral databases in search of novel 3C–like protease (3CLpro) enzyme inhibitors against SARS–CoV diseases, SAR QSAR Environ. Res. 31 (2020) 511–526, https://doi.org/10.1080/1062936X.2020.1776388.

[77] G. Davis, K. Li, F.G. Thankam, D.R. Wilson, D.K. Agrawal, Ocular transmissibility of COVID–19: possibilities and perspectives, Mol. Cell. Biochem. 477 (2022) 849–864, https://doi.org/10.1007/s11010-021-04336-6.

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

[79] A.C. Walls, Y.J. Park, M.A. Tortorici, A. Wall, A.T. McGuire, D. Veesler, Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein, Cell 181 (2020) 281–292, https://doi.org/10.1016/j.cell.2020.02.058.

[80] T.T. Lam, N. Jia, Y.W. Zhang, M.H. Shum, J.F. Jiang, H.C. Zhu, Y.G. Tong, Y.X. Shi, X.B. Ni, Y.S. Liao, W.J. Li, B.G. Jiang, W. Wei, T.T. Yuan, K. Zheng, X. M. Cui, J. Li, G.Q. Pei, X. Qiang, W.Y. Cheung, L.F. Li, F.F. Sun, S. Qin, J.C. Huang, G.M. Leung, E.C. Holmes, Y.L. Hu, Y. Guan, W.C. Cao, Identifying SARS–CoV–2–related coronaviruses in Malayan pangolins, Nature 583 (2020) 282–285, https://doi.org/10.1038/s41586-020-2169-0.

[81] D. Wrapp, N. Wang, K.S. Corbett, J.A. Goldsmith, C.L. Hsieh, O. Abiona, B.S. Graham, J.S. McLellan, Cryo–EM structure of the 2019–nCoV spike in the prefusion conformation, Science 367 (2020) 1260–1263, https://doi.org/10.1126/science.abb2507.

[82] S. Xia, M. Liu, C. Wang, W. Xu, Q. Lan, S. Feng, F. Qi, L. Bao, L. Du, S. Liu, C. Qin, F. Sun, Z. Shi, Y. Zhu, S. Jiang, L. Lu, Inhibition of SARS–CoV–2 (previously 2019–nCoV) infection by a highly potent pan–coronavirus fusion inhibitor targeting its spike protein that harbors a high capacity to mediate membrane fusion, Cell Res. 30 (2020) 343–355, https://doi.org/10.1038/s41422-020-0305-x.

[83] R. Ling, Y. Dai, B. Huang, W. Huang, J. Yu, X. Lu, Y. Jiang, In silico design of antiviral peptides targeting the spike protein of SARS–CoV–2, Peptides 130 (2020) 170328, https://doi.org/10.1016/j.peptides.2020.170328.

[84] X. Wang, S. Xia, Y. Zhu, L. Lu, S. Jiang, Pan–coronavirus fusion inhibitors as the hope for today and tomorrow, Protein Cell 12 (2021) 84–88, https://doi.org/10.1007/s13238-020-00806-7.

[85] M. Donoghue, F. Hsieh, E. Baronas, K. Godbout, M. Gosselin, N. Stagliano, M. Donovan, B. Woolf, K. Robison, R. Jeyaseelan, R.E. Breitbart, S. Acton, A novel angiotensin–converting enzyme–related carboxypeptidase (ACE2) converts angiotensin I to angiotensin 1–9, Circ. Res. 87 (2000) 1–9, https://doi.org/10.1161/01.res.87.5.e1.

[86] S.R. Tipnis, N.M. Hooper, R. Hyde, E. Karran, G. Christie, A.J. Turner, A human homolog of angiotensin–converting enzyme. Cloning and functional expression as a captopril–insensitive carboxypeptidase, J. Biol. Chem. 275 (2000) 33238–33243, https://doi.org/10.1074/jbc.M002615200.

[87] A.J. Turner, N.M. Hooper, The angiotensin–converting enzyme gene family: genomics and pharmacology, Trends Pharmacol. Sci. 23 (2002) 177–183, https://doi.org/10.1016/s0165-6147(00)01994-5.

[88] C. Vickers, P. Hales, V. Kaushik, L. Dick, J. Gavin, J. Tang, K. Godbout, T. Parsons, E. Baronas, F. Hsieh, S. Acton, M. Patane, A. Nichols, P. Tummino, Hydrolysis of biological peptides by human angiotensin–converting enzyme–related carboxypeptidase, J. Biol. Chem. 277 (2002) 14838–14843, https://doi.org/10.1074/jbc.M200581200.

[89] D. Pirolli, B. Righino, M.C. De Rosa, Targeting SARS-CoV-2 spike protein/ACE2 protein-protein interactions: a computational study, Mol. Inform. 40 (2021) e2060080, https://doi.org/10.1002/minf.202060080.

[90] W. Plonka, A. Paneth, P. Paneth, Docking and QSAR of aminothioureas at the SARS-CoV-2 s-protein-human ACE2 receptor interface, Molecules 25 (2020) 4645, https://doi.org/10.3390/molecules25204645.

[91] V. Zarezade, H. Rezaei, G. Shakerinezhad, A. Safavi, Z. Nazeri, A. Veisi, O. Azadbakht, M. Hatami, M. Sabaghan, Z. Shajirat, The identification of novel inhibitors of human angiotensin–converting enzyme 2 and main protease of SARS–Cov–2: a combination of in silico methods for treatment of COVID–19, J. Mol. Struct. 1237 (2021) 130409, https://doi.org/10.1016/j.molstruc.2021.130409.

[92] W. Li, M.J. Moore, N. Vasilieva, J. Sui, S.K. Wong, M.A. Berne, M. Somasundaran, J.L. Sullivan, K. Luzuriaga, T.C. Greenough, H. Choe, M. Farzan, Angiotensin–converting enzyme 2 is a functional receptor for the SARS coronavirus, Nature 426 (2003) 450–454, https://doi.org/10.1038/nature02145.

[93] H. Hofmann, K. Pyrc, L. van der Hoek, M. Geier, B. Berkhout, S. Pohlmann, Human coronavirus NL63 employs the severe acute respiratory syndrome coronavirus receptor for cellular entry, Proc. Natl. Acad. Sci. USA 102 (2005) 7988–7993, https://doi.org/10.1073/pnas.0409465102.

[94] J.E. Torres, R. Baldiris, R. Vivas–Reyes, Design of angiotensin-converting enzyme 2 (ACE2) inhibitors by virtual lead optimization and screening, J. Chin. Chem. Soc. 59 (2012) 1394–1400, https://doi.org/10.1002/jccs.201200079.

[95] L. Pinzi, G. Rastelli, Molecular docking: shifting paradigms in drug discovery, Int. J. Mol. Sci. 20 (2019) 4331, https://doi.org/10.3390/ijms20184331.

[96] L.G. Ferreira, S.R. Dos, G. Oliva, A.D. Andricopulo, Molecular docking and structure–based drug design strategies, Molecules 20 (2015) 13384–13421, https://doi.org/10.3390/molecules200713384.

[97] V.B. Sulimov, D.C. Kutov, A.V. Sulimov, Advances in docking, Curr. Med. Chem. 26 (2019) 7555–7580, https://doi.org/10.2174/0929867325666180904115000.

[98] F. Tessaro, L. Scapozza, How 'Protein-Docking' translates into the new emerging field of docking small molecules to nucleic acids? Molecules 25 (2020) 2749, https://doi.org/10.3390/molecules25122749.

[99] N.S. Pagadala, K. Syed, J. Tuszynski, Software for molecular docking: a review, Biophys. Rev. 9 (2017) 91–102, https://doi.org/10.1007/s12551-016-0247-1.

[100] N.A. Murugan, A. Podobas, D. Gadioli, E. Vitali, G. Palermo, S. Markidis, A review on parallel virtual screening softwares for high-performance computers, Pharmaceuticals 15 (2022) 63, https://doi.org/10.3390/ph15010063.

[101] A.K. Oyedele, A.T. Ogunlana, I.D. Boyenle, A.O. Adeyemi, T.O. Rita, T.I. Adelusi, M. Abdul-Hammed, O.E. Elegbeleye, T.T. Odunitan, Docking covalent targets for drug discovery: stimulating the computer–aided drug design community of possible pitfalls and erroneous practices, Mol. Divers. 27 (2022) 1–25, https://doi.org/10.1007/s11030-022-10523-4.

[102] S. Tang, R. Chen, M. Lin, Q. Lin, Y. Zhu, J. Ding, H. Hu, M. Ling, J. Wu, Accelerating AutoDock vina with GPUs, Molecules 27 (2022) 3041, https://doi.org/10.3390/molecules27093041.

[103] R. Friedman, K. Boye, K. Flatmark, Molecular modelling and simulations in cancer research, Biochim. Biophys. Acta 1836 (2013) 1–14, https://doi.org/10.1016/j.bbcan.2013.02.001.

[104] P. Chowdhury, In silico investigation of phytoconstituents from Indian medicinal herb 'Tinospora cordifolia (giloy)' against SARS–CoV–2 (COVID–19) by molecular dynamics approach, J. Biomol. Struct. Dyn. 39 (2021) 6792–6809, https://doi.org/10.1080/07391102.2020.1803968.

[105] S. Genheden, U. Ryde, The MM/PBSA and MM/GBSA methods to estimate ligand–binding affinities, Expert. Opin. Drug. Discov. 10 (2015) 449–461, https://doi.org/10.1517/17460441.2015.1032936.

[106] N.A. Saleh, In-silico study: docking simulation and molecular dynamics of peptidomimetic fullerene-based derivatives against SARS-CoV-2 Mpro, 3 Biotech 13 (2023) 185, https://doi.org/10.1007/s13205-023-03608-w.

[107] Z.R. Guo, [Strategy of molecular drug design: activity and druggability], Acta Pharm. Sin. 45 (2010) 539–547, https://doi.org/10.16438/j.0513-4870.2010.05.016.

[108] X. Li, D. Kong, Predicting ligand druggability for drug discovery, Computers and Applied Chemistry 29 (2012) 999–1003, https://doi.org/10.16866/j.com.app.chem2012.08.023.

[109] D.F. Veber, S.R. Johnson, H.Y. Cheng, B.R. Smith, K.W. Ward, K.D. Kopple, Molecular properties that influence the oral bioavailability of drug candidates, J. Med. Chem. 45 (2002) 2615–2623, https://doi.org/10.1021/jm020017n.

[110] R.A. Al–Horani, S. Kar, K.F. Aliter, Potential anti-COVID-19 therapeutics that block the early stage of the viral life cycle: structures, mechanisms, and clinical trials, Int. J. Mol. Sci. 21 (2020) 5224, https://doi.org/10.3390/ijms21155224.

[111] K. Dhama, F. Nainu, A. Frediansyah, M.I. Yatoo, R.K. Mohapatra, S. Chakraborty, H. Zhou, M.R. Islam, S.S. Mamada, H.I. Kusuma, A.A. Rabaan, S. Alhumaid, A.A. Mutair, M. Iqhrammullah, J.A. Al-Tawfiq, M.A. Mohaini, A.J. Alsalman, H.S. Tuli, C. Chakraborty, H. Harapan, Global emerging Omicron variant of SARS-CoV-2: impacts, challenges and strategies, J. Infect. Public Health 16 (2023) 4–14, https://doi.org/10.1016/j.jiph.2022.11.024.

[112] E. Pitsillou, J.J. Liang, R.C. Beh, A. Hung, T.C. Karagiannis, Molecular dynamics simulations highlight the altered binding landscape at the spike-ACE2 interface between the Delta and Omicron variants compared to the SARS-CoV-2 original strain, Comput. Biol. Med. 149 (2022) 106035, https://doi.org/10.1016/j.compbiomed.2022.106035.

[113] A. Hossain, M.E. Rahman, M.S. Rahman, K. Nasirujjaman, M.N. Matin, M.O. Faruqe, M.F. Rabbee, Identification of medicinal plant-based phytochemicals as a potential inhibitor for SARS-CoV-2 main protease (Mpro) using molecular docking and deep learning methods, Comput. Biol. Med. 157 (2023) 106785, https://doi.org/10.1016/j.compbiomed.2023.106785.

[114] B.H. Gulumbe, Z.M. Yusuf, A.M. Hashim, Harnessing artificial intelligence in the post-COVID-19 era: a global health imperative, Trop. Doct. 53 (2023) 414–415, https://doi.org/10.1177/00494755231181155.

[115] A. Sharma, T. Virmani, V. Pathak, A. Sharma, K. Pathak, G. Kumar, D. Pathak, Artificial intelligence-based data-driven strategy to accelerate research, development, and clinical trials of COVID Vaccine, BioMed Res. Int. 6 (2022) 7205241, https://doi.org/10.1155/2022/7205241.

[116] M. Anshari, M. Hamdan, N. Ahmad, E. Ali, H. Haidi, COVID-19, artificial intelligence, ethical challenges and policy implications, AI Soc. 38 (2023) 707–720, https://doi.org/10.1007/s00146-022-01471-6.

[117] S. Huang, J. Yang, S. Fong, Q. Zhao, Artificial intelligence in the diagnosis of COVID-19: challenges and perspectives, Int. J. Biol. Sci. 17 (2021) 1581–1587, https://doi.org/10.7150/ijbs.58855.

[118] S.G. Jamalipour, S. Iravani, Potential inhibitors of SARS-CoV-2: recent advances, J. Drug Target. 29 (2021) 349–364, https://doi.org/10.1080/1061186X.2020.1853736.

[119] Y.N. Lamb, Remdesivir: first approval, Drugs 80 (2020) 1355–1363, https://doi.org/10.1007/s40265-020-01378-w.

[120] K. Rosenberg, Remdesivir in the treatment of COVID-19, Am. J. Nurs. 121 (2021) 55, https://doi.org/10.1097/01.NAJ.0000731668.01845.8c.

[121] S.G. Katre, A.J. Asnani, K. Pratyush, N.G. Sakharkar, A.G. Bhope, K.T. Sawarkar, V.S. Nimbekar, Review on development of potential inhibitors of SARS-CoV-2 main protease (M(Pro)), Futur, J. Pharm. Sci. 8 (2020) 36, https://doi.org/10.1186/s43094-022-00423-7.

[122] R. Banerjee, L. Perera, LMV Tillekeratne, Potential SARS-CoV-2 main protease inhibitors, Drug Discov. Today 26 (2021) 804–816, https://doi.org/10.1016/j.drudis.2020.12.005.

[123] K. Gao, R. Wang, J. Chen, L. Cheng, J. Frishcosy, Y. Huzumi, Y. Qiu, T. Schluckbier, X. Wei, G.W. Wei, Methodology-centered review of molecular modeling, simulation, and prediction of SARS-CoV-2, Chem. Rev. 122 (2022) 11287–11368, https://doi.org/10.1021/acs.chemrev.1c00965.

[124] R. Rolta, D. Salaria, B. Sharma, O. Awofisayo, O.A. Fadare, S. Sharma, C.N. Patel, V. Kumar, A. Sourirajan, D.J. Baumler, K. Dev, Methylxanthines as potential inhibitor of SARS-CoV-2: an in silico approach, Curr. Pharmacol. Rep. 8 (2022) 149–170, https://doi.org/10.1007/s40495-021-00276-3.

[125] D.J. Calleja, G. Lessene, D. Komander, Inhibitors of SARS-CoV-2 PLpro, Front. Chem. 10 (2022) 876212, https://doi.org/10.3389/fchem.2022.876212.
[126] S.A. Amin, S. Banerjee, K. Ghosh, S. Gayen, T. Jha, Protease targeted COVID-19 drug discovery and its challenges: insight into viral main protease (Mpro) and papain-like protease (PLpro) inhibitors, Bioorg. Med. Chem. 29 (2021) 115860, https://doi.org/10.1016/j.bmc.2020.115860.
[127] M. Elkashlan, R.M. Ahmad, M. Hajar, F. Jasmi, Al, J.M. Corchado, N.A. Nasarudin, M.S. Mohamad, A review of SARS-CoV-2 drug repurposing: databases and machine learning models, Front. Pharmacol. 14 (2023) 1182465, https://doi.org/10.3389/fphar.2023.1182465.