PORTLAND PRESS

Review Article

Knowing and combating the enemy: a brief review on SARS-CoV-2 and computational approaches applied to the discovery of drug candidates

Mateus S.M. Serafim[1], Jadson C. Gertrudes[2], Débora M.A. Costa[3], Patricia R. Oliveira[4], Vinicius G. Maltarollo[3] and 🆔 Kathia M. Honorio[4,5]

[1]Department of Microbiology, Biological Sciences Institute, Federal University of Minas Gerais (UFMG), Belo Horizonte, MG, Brazil; [2]Department of Computer Science, Federal University of Ouro Preto (UFOP), Ouro Preto, MG, Brazil; [3]Department of Pharmaceutical Products, Faculty of Pharmacy, Federal University of Minas Gerais (UFMG), Belo Horizonte, MG, Brazil; [4]School of Arts, Sciences and Humanities, University of São Paulo (USP), 03828-000, São Paulo, SP, Brazil; [5]Center for Natural and Human Sciences, Federal University of ABC (UFABC), Santo Andre, SP, Brazil

Correspondence: Kathia M. Honorio (kmhonorio@usp.br)

OPEN ACCESS

Since the emergence of the new severe acute respiratory syndrome-related coronavirus 2 (SARS-CoV-2) at the end of December 2019 in China, and with the urge of the coronavirus disease 2019 (COVID-19) pandemic, there have been huge efforts of many research teams and governmental institutions worldwide to mitigate the current scenario. Reaching more than 1,377,000 deaths in the world and still with a growing number of infections, SARS-CoV-2 remains a critical issue for global health and economic systems, with an urgency for available therapeutic options. In this scenario, as drug repurposing and discovery remains a challenge, computer-aided drug design (CADD) approaches, including machine learning (ML) techniques, can be useful tools to the design and discovery of novel potential antiviral inhibitors against SARS-CoV-2. In this work, we describe and review the current knowledge on this virus and the pandemic, the latest strategies and computational approaches applied to search for treatment options, as well as the challenges to overcome COVID-19.

Introduction

In late December 2019 at Wuhan (China), an unknown acute respiratory disease was reported. At the first week of January 2020, a new virus called the severe acute respiratory syndrome-related coronavirus 2 (SARS-CoV-2) was identified as the etiological agent of the related cases, which would be later named as the coronavirus disease 2019 (COVID-19) [1,2]. Rapidly progressing from a local outbreak to a pandemic scenario, by the end of November 2020, the SARS-CoV-2 infection had been diagnosed in more than 57.8 million people, with almost 50% of cases in the Americas and over 11,789,000 of them in the United States of America (U.S.A.). Until 24 November 2020, over 1,377,000 deaths happened worldwide, with 252,460 (18.3%) in the U.S.A., due to the rapid SARS-CoV-2 spread and the severity of COVID-19 [2,3].

Figure 1 presents an overview of the main events occurring during the year of 2020 and related to SARS-CoV-2, until the submission of this document. It comprises several key events, especially the number of deaths and the efforts made by the World Health Organization (WHO).

The coronaviruses, such as SARS-CoV-2, are from a large family (*Coronaviridae*) of spherical and enveloped virus, with 120–160 nm in diameter, and a single-stranded positive sense RNA (ssRNA+) genome, containing approximately 26–32 kb. These viruses were named due to the solar crown ('corona', in latin) aspect that the viral particles exhibit under electron microscopy due to the glycoproteins at its surface [4,5].
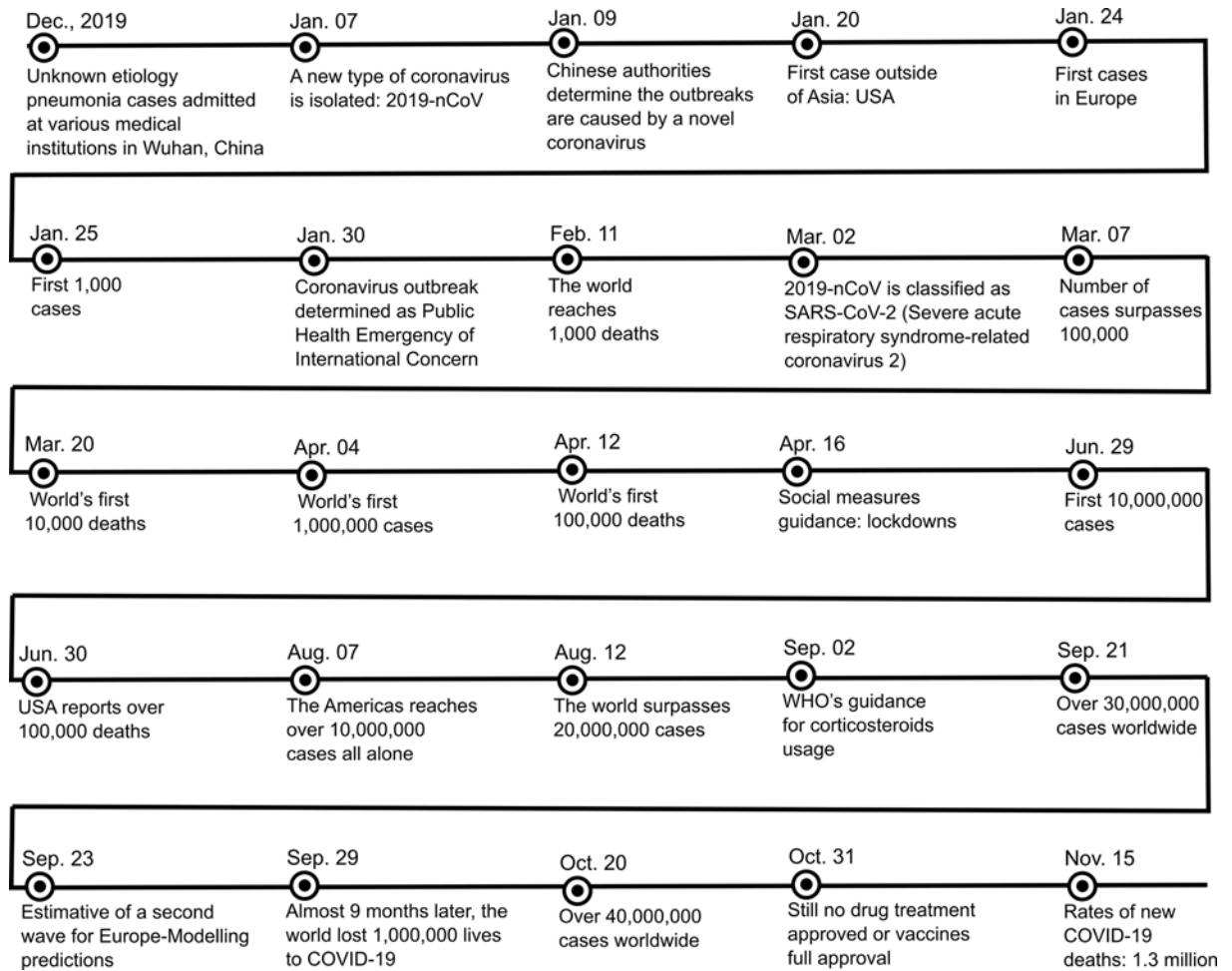
Figure 1. Main events related to SARS-CoV-2 during the year of 2020 (until 24 November)

Transmission of coronaviruses usually occurs by air, oral–fecal routes or fomites, associated with gastrointestinal and respiratory infections, due to its tropism for epithelial cells [6–8]. Generally, these infections are asymptomatic and mild, but some species of coronavirus cause serious diseases, such as hepatitis, neurological disorders, kidney failure, and severe acute respiratory syndrome (SARS), which can lead to death [6,9].

There are six known human coronaviruses (HCoV) species, such as HCoV-229E and HCoV-NL63 (*Alphacoronavirus* genus), as well as HCoV-OC43 and HKU1 (*Betacoronavirus* genus), which mainly cause common colds, but can progress to severe infections of the lower respiratory tract, especially in children and the elderly. In addition to these betacoronaviruses, two other species are highly virulent: the Middle East respiratory syndrome coronavirus (MERS-CoV) and the severe acute respiratory syndrome-related coronavirus (such as SARS-CoV and SARS-CoV-2) [1,10,11].

The transmission of SARS-CoV-2 occurs through airways and from direct contact, in addition to the contact with contaminated objects and surfaces. Incubation periods may last up to 14 days (average of 5 days), and the most common symptoms are fever, cough, fatigue, headache, and breathing difficulties (dyspnoea), as well as smell and taste loss that can last over 2 weeks. In addition to these, sore throat, myalgia, diarrhea, vomiting, and nasal congestion can also occur [12–15].

Similar to other ssRNA+ viruses, after infection of host cells, progeny occurs in the cytoplasm. During its infection, SARS-CoV-2 particles bind to receptors at the cell surface by interaction of the receptor binding domain (RBD) of the spike protein (S) with the cell receptor, the angiotensin II converting enzyme (ACE-II) [4,10]. After the virus entry to the cells, during the expression of the viral replication complex, RNA is translated into two polyproteins (PP1a and PP1ab), which encode 16 non-structural proteins (NSPs), such as the main protease (M^{pro}) and the RNA-dependent

RNA polymerase (RdRp). The remainder of the genome encodes accessory and the structural proteins (spike, membrane (M), nucleocapsid (N), and envelope (E)), followed by the assembly and release of viral particles [4,16].

Regarding the infection control measures, vaccines are considered a viable and important alternative, in particular as the first line of prevention, even more in a pandemic scenario [17,18]. Some studies on different vaccine candidates to prevent SARS-CoV-2 infections have moved into Phases II and III trials, such as Pfizer-BioNTech [19], AstraZeneca-Oxford [20], Moderna [21], Gamaleya [22], as well as Sinovac [23], some of which have successfully completed Phase III clinical trials and/or licensed for early use in late November 2020. However, over the long-term efficacy, safety or global production in large-scale and short period of time may still be a challenge to overcome, and an immunization failure or insufficient coverage are not discarded, as observed with vaccines against other coronaviruses, such as SARS-CoV and MERS-CoV [16,24].

On the other hand, drug repurposing strategies could be promising in the fight against COVID-19. Approved or licensed drugs, as previously assessed by studies that evaluated them in preclinical and clinical trials, show that these approaches could potentially reduce time and costs for making new therapies available [25]. Considering this scenario and the possibility of novel outbreaks or pandemics, an approved drug may be also used to treat diseases caused by other coronaviruses or even future mutations of SARS-CoV-2 [26–28]. Different repurposing strategies and drug combinations have been proposed (e.g. remdesivir) but showed lack of inhibitory activity or inconclusive clinical results, as well as having to take the account of side effects, thus leaving a significant opportunity for the design and development of efficient drugs to face the challenges of SARS-CoV-2 and COVID-19 [29,30].

It is interesting to mention that *in silico* virtual screening approaches associated with structural and biophysical techniques can help the design of specific inhibitors to SARS-CoV-2, and significantly enhance the quality of compounds selected for *in vitro* and *in vivo* bioassays, increasing the success of drug discovery [31–34]. For instance, structure-based approaches have shown some successful outcomes in the past, for example, the design and discovery of boceprevir, an approved hepatitis C virus (HCV) protease inhibitor, as well as oseltamivir and zanamivir, both anti-influenza drugs [35].

In the past few months, several small molecules have been described as possible inhibitors of different molecular targets for SARS-CoV-2 [36]. However, it is important to note that many of these studies are still in the initial *in silico* analyses, which only provide a preliminary theoretical view on the ligand–protein interactions and hence requiring experimental validation of the molecular targets [33].

Among the molecular targets of SARS-CoV-2, main protease or 3-chymotrypsin-like protease (M^pro/3CL^pro/nsp5) [37], papain-like protease (PL^pro/nsp3) [38], RNA-dependent RNA polymerase (RdRp/nsp12) [39], and helicase/NTPase (nsp13) [40] could be cited, which are highly conserved and essential to the viral cycle [36,41–45], as illustrated in Figure 2. Since the main viral protease is extensively studied for the design of new drug candidates to treat coronaviruses diseases, and most of the studies identified this enzyme as a valid target for broad spectrum inhibitors, we will next focus on the discussion of this macromolecule in more details [33,46,47].

Computer-aided drug design strategies as useful tools against SARS-CoV-2 macromolecules: targeting M^pro

Computer-Aided Drug Design (CADD) involves widely employed computational approaches to discover and/or design new bioactive compounds. As examples of CADD techniques we can cite molecular docking, molecular dynamics (MD) simulations, pharmacophore modeling, similarity analysis, quantitative structure–activity relationship (QSAR) analysis, and machine learning (ML) techniques [48–51] that will be discussed later.

Some recent studies have shown the feasibility of employing *in silico* methods such as molecular docking and MD simulations to perform virtual screening of molecules against the SARS-CoV-2 macromolecules and, consequently, obtaining potential repositioning drugs and selective inhibitors of these enzymes [52–54]. Figure 3 summarizes the molecular structures of the potential inhibitors and drug candidates discussed in this review.

M^pro (nsp5) is one of the most attractive viral targets for the antiviral drug discovery against SARS-CoV-2, since it plays a key role in the viral transcription and replication, and no human proteases are known with the same substrate specificity [37,46,47,55]. Furthermore, the substrate-binding pocket of this enzyme is highly conserved among all coronaviruses, suggesting that an antiviral drug targeting this active site may be effective against a broader spectrum of these viruses [37,56,57]. Nonetheless, mutations leading to changes in some amino acid residues of M^pro may provide probable drug resistance phenotypes, in particular considering the enzyme loop and the possibility of a protein folding [58]. In addition, protease resistance was also observed to MERS-CoV and other viruses, such as the human immunodeficiency virus (HIV) and HCV [59].
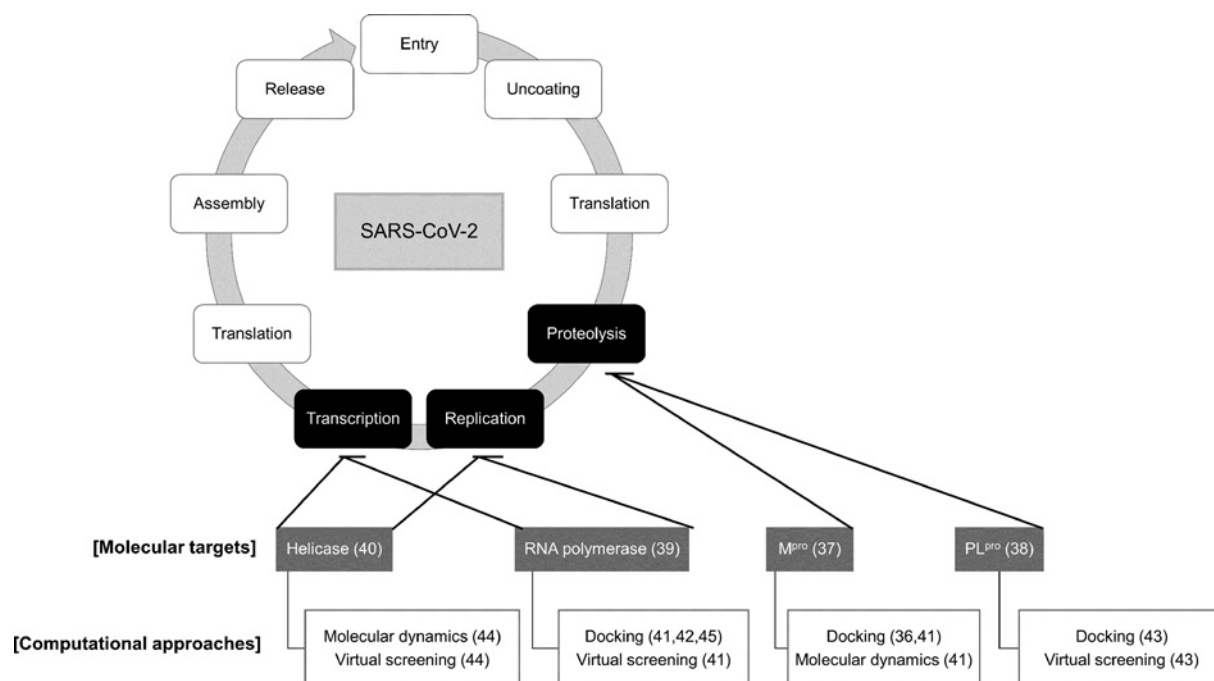
Figure 2. Interventions of computational approaches on to the design and discovery of potential inhibitors against molecular targets involved in the SARS-CoV-2 viral cycle

The diagram displays available methods that can be employed to both the M^pro and PL^pro proteases (e.g. docking), as well as the RNA polymerase and helicase (e.g. virtual screening).

M^pro is a cysteine protease with a catalytic Cys^{145} and His^{41} dyad at its active site [60,61], which cleaves the polyproteins in at least 11 conserved sites, starting with its autolytic cleavage between nsp4 and nsp6 [62,63]. The M^pro structure is composed of three domains; the catalytic dyad is located in the cleft between domains I and II [37,64,65], and the domain III is responsible for the enzyme dimerization, enabling the active form of the macromolecule [66,67]. Figure 4 illustrates X-ray crystal structures of SARS-CoV-2 M^pro in complex with some inhibitors.

Several inhibitors of SARS-CoV M^pro have been identified [46,47]. For instance, N3 (Figure 4A,B), a Michael acceptor-based inhibitor, can specifically inhibit M^pro of different coronaviruses, including SARS-CoV and MERS-CoV [56,64,69,70]. It has also showed antiviral activity in cell culture against SARS-CoV-2 and the ability to bind to the substrate-binding pocket of SARS-CoV-2 M^pro, as proven by X-ray crystallography (Figure 4A,B) [37].

Docking approaches can predict inhibitory activity and help drug design or virtual screenings, which have resulted in potential repurposing drugs against SARS-CoV-2 M^pro such as lopinavir and ritonavir [71]. The combination of docking and MD simulations, for example, allows a refinement of docking results, specifically evaluating the frequency of main interactions between residues and the drug candidates, as well as binding energy of the bioactive substances at the target, number of contacts, occupancy and stability of the target–ligand complex [72–74].

These strategies can also expand the number of compounds to be assessed in virtual screening (VS) campaigns, allowing thousands to millions of compounds to be screened from virtual chemical libraries. For instance, Jiménez-Alberto et al. (2020) showed bromocriptine, simeprevir and other FDA approved drugs to have promising inhibitory activity against the SARS-CoV-2 main protease, as result of VS and MD simulations [54]. Similarly, Kumar et al. (2020) also screened antivirals as drug repurposing therapies for COVID-19, and these studies highlighted lopinavir and ritonavir as potential treatments to be further evaluated in clinical trials against M^pro [53].

Another study investigated a combination of structure-based drug design, virtual and *in vitro* high-throughput screening of a library with more than 10,000 compounds, identifying disulfiram, carmofur, ebselen, shikonin, tideglusib, and PX-12 as SARS-CoV-2 M^pro inhibitors. Among these substances, ebselen exhibited the strongest antiviral effects in SARS-CoV-2-infected Vero cells in the low micromolar range [37]. The X-ray crystal structure of SARS-CoV-2 M^pro in complex with carmofur, an approved antineoplastic agent, also revealed that its carbonyl reactive group reacts irreversibly to bind to the catalytic Cys^{145} (Figure 4C) [75].
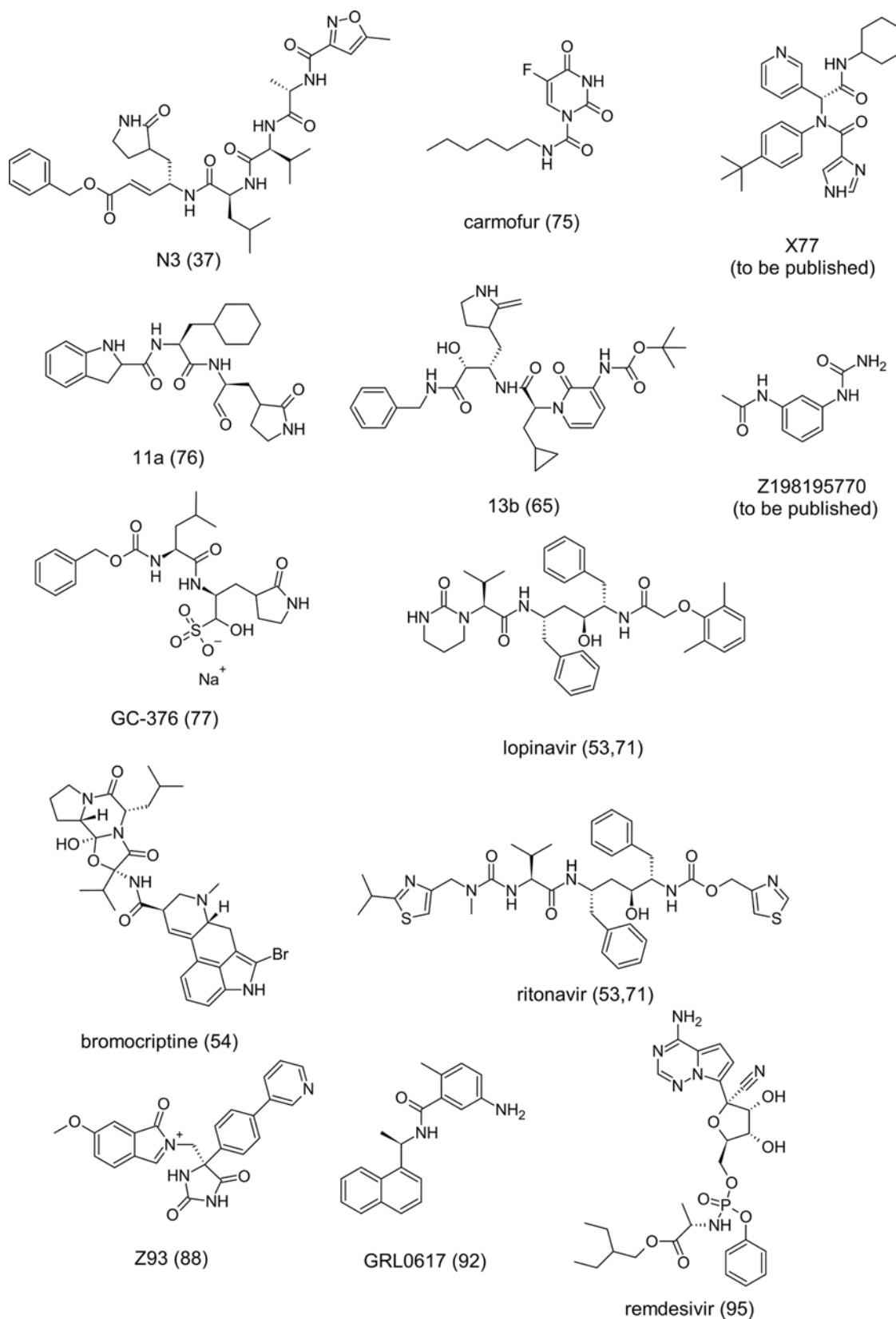
Figure 3. Two-dimensional (2D) chemical structures of the potential inhibitors and drug candidates presented in the present study, along with the respective references
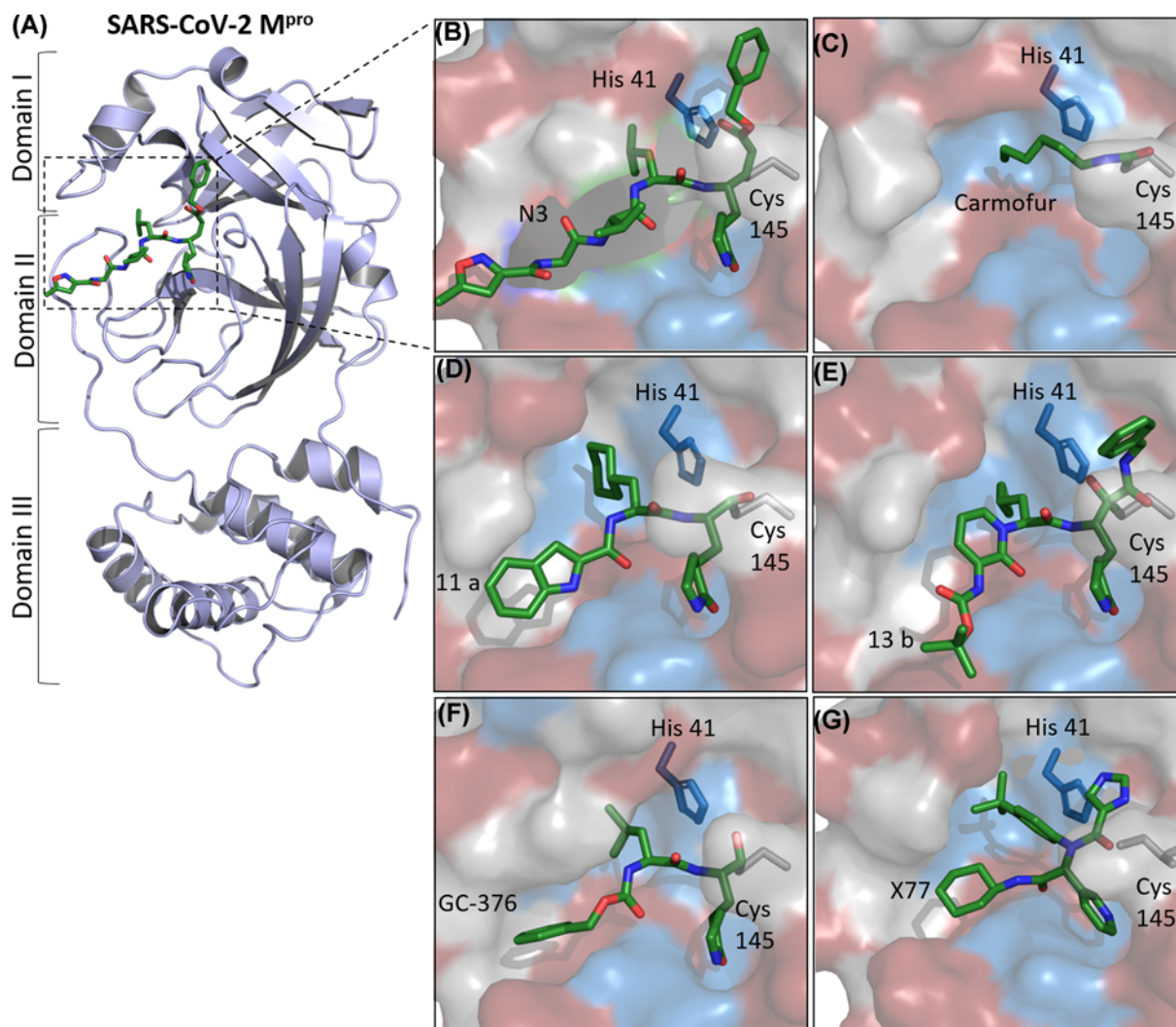
Figure 4. X-ray crystal structures of SARS-CoV-2 M^pro in complex with inhibitors

(**A**) Cartoon representation of one protomer of the dimeric M^{pro} in complex with N3 (PDB ID: 6L7U). (**B–G**) An enlarged view of the substrate-binding pocket in a surface representation (red indicating hydrophobic residues, blue indicating charged residues, and gray indicating polar ones) in complex with different compounds. The catalytic Cys^{145} and His^{41} dyad are highlighted. (B) N3, PDB ID: 6LU7. (C) Carmofur, PDB ID: 7BUY. (D) 11a, PDB ID: 6LZE. (E) 13b, PDB ID: 6Y2F. (F) GC-376, PDB ID: 6WTT. (G) X77, PDB ID: 6W63. Images were generated with PyMOL 0.99 [68].

Moreover, some compounds designed and synthesized by analyzing the substrate-binding pocket of M^{pro} revealed anti-SARS-CoV-2 activity in Vero cell cultures. For example, the crystal structure of the complexes SARS-CoV-2 M^{pro}-11a (Figure 4D) and M^{pro}-11b indicated the presence of these substances inside the substrate-binding pocket and a similar inhibitory mechanism in which occurs the C–S covalent bond formation between Cys^{145} and these compounds [76].

Zhang et al. also developed an optimized α-ketamide inhibitor of SARS-CoV-2 M^{pro}. The X-ray crystal structure of α-ketoamide (compound 13b) in complex with SARSCoV-2 M^{pro} shows the compound at the catalytic site of each protomer, between the domains I and II (Figure 4E). Compound 13b effectively prevented the viral replication in cell-based assays and exhibits a favorable pharmacokinetic *in vivo* profile [65]. Other work investigated boceprevir, GC-376, and calpain inhibitors effects on SARS-CoV-2 M^{pro} in enzymatic assays in cell cultures. The crystal structure of SARS-CoV-2 M^{pro} in complex with GC-376 also revealed molecular details on the GC-376 inhibition of the molecular target (Figure 4F) [77].

In addition to the studies above, several other structures of SARS-CoV-2 M[pro] in complex with inhibitors have been deposited at the Protein Data Bank (PDB), such as a non-covalent inhibitor X77 (PDB ID: 6W63) (Figure 4G), narlaprevir (PDB ID: 7D10), boceprevir (PDB ID: 7COM), GRL-2420 (PDB ID: 7JKV) and UAW246 (PDB ID: 6XBG). However, most of these works have not so far been published in peer-reviewed papers.

Although additional research has also identified several inhibitors as promising drugs against SARS-COV-2 M[pro], further biochemical and structural analyses, as well as *in vitro* and *in vivo* bioassays are still required. For example, several investigations predicted various small-molecules, natural compounds and approved drugs [78–80] from VS of ZINC and DrugBank databases [81,82], in conjunction with the combination of molecular docking and MD studies [53]. Focusing on drug repurposing, we can also highlight a study that apply molecular docking associated with the SCAR (steric-clashes alleviating receptors) protocol, which can help to discovery of covalent and non-covalent inhibitors in a docking model, solving, for example, steric conflicts between specific residues and reactive atoms in a screening [83]. In addition, it can even be efficient in drug repurposing [84], another approach of interest facing the urgency of the COVID-19 pandemic scenario.

It is important to mention that other strategies can evaluate and predict important characteristics of potential candidates and inhibitors of SARS-CoV-2 targets, for example, pharmacokinetics (absorption, distribution, metabolism, and excretion—ADME) and toxicity or simply ADME-Tox properties. Hage-Melim et al. (2020) screened for potential M[pro] inhibitors and the top 100 hits were evaluated by bidimensional structural similarity in order to assess their ADME-Tox properties, resulting in ten compounds, including potential repurposing drug candidates such as lopinavir, ritonavir, and remdesivir [5]. Nonetheless, in a drug repurposing approach, this kind of evaluation may not be necessary, and a given candidate could proceed quickly to clinical trials, due to the previously assessed steps of an approved or licensed drug [85].

# The importance of multitargets to CADD approaches: PL[pro] and RdRp

Another essential protease for the cleavage of the viral polyproteins is PL[pro], a cysteine protease with a classical Cys-His-Asp catalytic triad (Cys[112], His[273], Asp[287]), which cleaves the viral polyprotein releasing nsp1, nsp2 and nsp3 [86,87]. This enzyme also recognizes the consensus cleavage sequence identified by cellular deubiquitinating enzymes [87]. Therefore, substrate-derived inhibitors of PL[pro] would be expected to inhibit host cell deubiquitinases [41,76,87].

Computational approaches have also been used to predict potential SARS-CoV-2 PL[pro] inhibitors. From integrated *in silico* efforts, Mirza et al. pointed out a human ubiquitin carboxyl-terminal hydrolase-2 (USP2) inhibitor (compound Z93) as a potential lead compound against SARS-CoV-2 PL[pro] [88]. A pharmacophore model of functional centers of PL[pro] inhibitor-binding pocket and docking studies also identified 147 FDA-approved drugs, including HIV, hepatitis C, and cytomegalovirus (CMV) inhibitors, as well as drugs that have demonstrated some activity in MERS and SARS-CoV as potential opportunities for the treatment of COVID-19 [89].

Virtual screenings using ZINC and Chinese natural product databases [81], as well as FDA-approved drugs [90], have also found potential PL[pro] inhibitors. Nonetheless, alternative approaches, such as assessing inhibitors activity in Vero cell cultures [91], as well as *in vitro* protease and structural assays [92], can also be mentioned. The crystal structure of GRL0617 in complex with SARS-CoV-2 PL[pro] demonstrated that the inhibitor occupies the active site of the enzyme (Figure 5A,B) [92]. In addition, inhibition of the viral cycle has been demonstrated by GRL0617 against SARS-CoV [93].

The RdRp (nsp12) has also been described as an important coronavirus target for drug design [41,44,45]. It has an essential role in the viral cycle of coronaviruses, since it is responsible for the replication of the viral genome, with the assistance of nsp7 and nsp8 viral proteins in a polymerase complex [39,94,95]. In this sense, RdRp is considered as a primary target for nucleotide analog antiviral inhibitors such as remdesivir [96,97], which has been investigated in clinical trials against COVID-19. One study indicated that this drug seems to be capable of reducing the recovery time of severe hospitalized patients [98]. Regarding studies on the molecular target validation for this drug, the structure of the complex nsp12–nsp7–nsp8 associated with template-primer RNA and remdesivir shows that it is covalently incorporated at the first replicated base pair into the primer strand, blocking the RNA chain elongation (Figure 5C–E) [95].

Other studies applied a molecular docking approach and identified galidesivir, remdesivir, ribavirin, sofosbuvir, and tenofovir as potential drug candidates against SARS-CoV-2 RdRp [99]. Elfiky (2020) used sequence analysis, modeling and docking, identifying sofosbuvir, IDX-184, ribavirin, and remdesivir as potential therapies [100]. Beclabuvir, an HCV RdRp inhibitor, has also been predicted to bind SARS-CoV-2 RdRp [101].
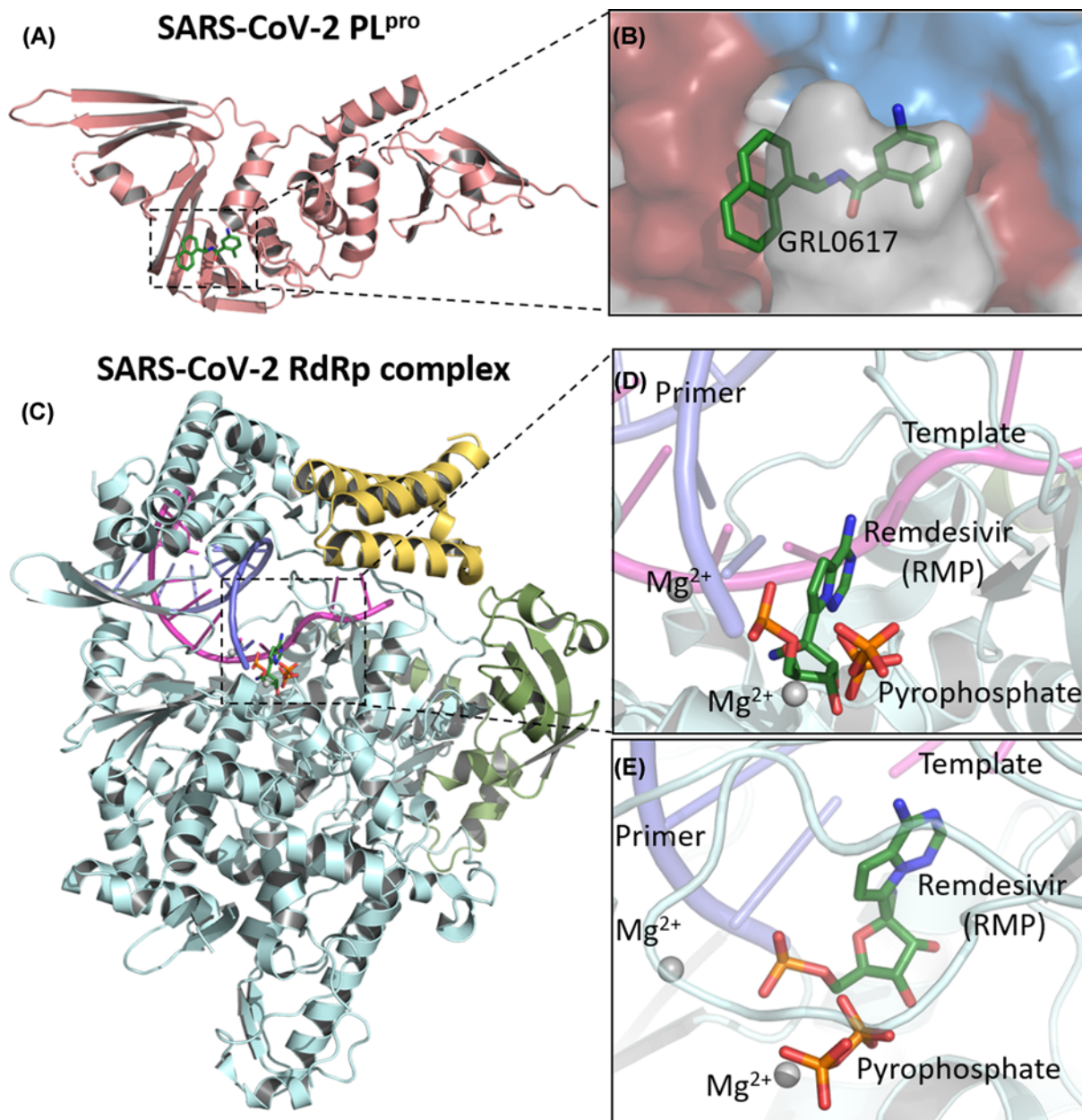
Figure 5. Crystal structure of SARS-CoV-2 PL^pro in complex with inhibitor GFL0617 and cryo-electron microscopy structure of the SARS-CoV-2 remdesivir and RNA bound RdRp complex

(**A**) Cartoon representation of papain-like protease, PDB ID: 7CMD. (**B**) An enlarged view of PL^pro substrate-binding pocket with GFL0617. (**C**) Cartoon representation of the nsp12–nsp7–nsp8 RdRp complex with a template-primer RNA and remdesivir. The structure is colored by the elements: nsp12 in pale cyan, nsp7 in pale yellow, nsp8 in bright orange, primer RNA in purple and template RNA in light magenta, PDB ID: 7BV2. (**D**) Close-up view on the RdRp active site showing the covalently bound remdesivir in its monophosphate form—RMP, pyrophosphate and magnesium ions represented by gray spheres. (**E**) RdRp binding-pocket in a different view. Images were generated with PyMOL 0.99 [68].

Helicase/NTPase (nsp13) has also been cited as a molecular target of SARS-CoV-2. This macromolecule is responsible for unwinding DNA and RNA, separating them into two single-stranded nucleic acids in the coronaviruses viral cycle [102,103]. Molecular docking analyses and structure modeling approaches have also suggested drugs and natural products as potential SARS-CoV-2 helicase inhibitors [80,104].
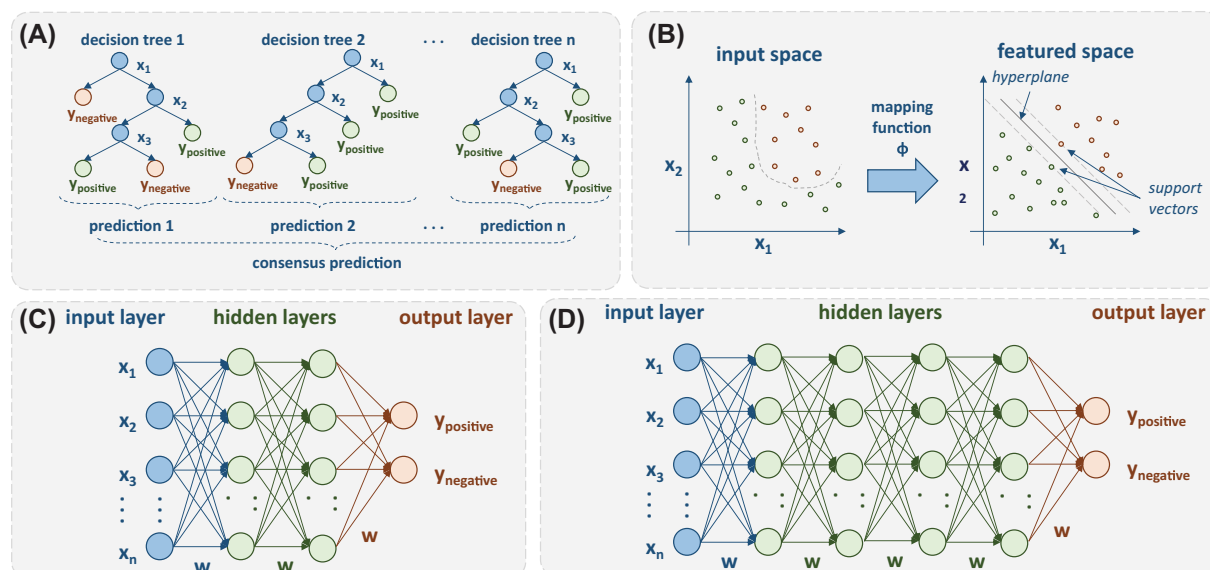
Figure 6. ML approaches

(**A**) Random Forest (RF); (**B**) SVM; (**C**) Multilayer Perceptron (MLP) and (**D**) Deep Neural Network (DNN).

In addition, a cryo electron microscopy (cryo-EM) structure of the complex nsp13:holo-RdRp:RNA complex suggests a possible role of helicase (nsp13) in the viral replication/transcription process, which needs to be confirmed through *in vitro* or *in vivo* studies [40]. Crystal structure of helicase in complex with Z198195770 (PDB ID: 5RL6), as well as other different compounds, has been determined, which can increase the recognition of this enzyme as a druggable target against SARS-CoV-2. However, *in vitro* and *in vivo* studies are necessary to confirm this macromolecule as an effective molecular target for the development of drugs against COVID-19.

ML during a pandemic: actual state and future perspectives against SARS-CoV-2

Machine learning (ML) techniques are a valuable new tool for drug discovery against SARS-CoV-2, since they can be applied to build predictive models based on previous experience (e.g. by using as training instances molecules already tested against other coronaviruses targets). ML applications in drug design commonly involves regression or classification methods for assessing the activity of compounds against a target before clinical trials. In this paper, we present the main aspects of ML approaches that may accelerate COVID-19 drug discovery, such as Ensemble Learning, Support Vector Machines (SVM), Artificial Neural Networks (ANN), and Deep Learning (DL) models [105–107]. In addition, integration of Artificial Intelligence (AI) and mechanistic modeling of signal transduction circuits, with ML algorithms, are also some recent approaches to drug repurposing models for COVID-19 [108,109]. Figure 6 presents an overview of some ML techniques.

Ensemble learning is a simple and effective learning model. The basic idea behind this approach is to build multiple models and combine their decisions in some manner [110]. Random forest (RF), illustrated in Figure 6A, is an ensemble of decision trees that uses the concept of bagging, which consists of randomly select samples to create the individual trees. Such random behavior leads to a more diverse set of decision trees, thus producing combined predictions that are more accurate than any of the individual ones. Similar approaches to RF include extra trees [111] and XGboost [112].

Aiming at discovering potential candidates to treat COVID-19, Rodrigues et al. (2020) applied computational tools to investigate the potential of diterpenes in inhibiting M^pro [113]. The authors collected a set of molecules from ChEMBL and Sistematx [114] databases to construct QSAR models using RF and multidescriptor read-across (MuDRA) models [115]. The models were used to perform a ligand-based VS, combined with a structure-based VS using Molegro Virtual Docker v6.0.1 [116]. Although no antiviral activity assays were performed, four diterpenes were selected as potential active inhibitors against the six different species of HCoV.

Similarly, Alves et al. (2020) presented a detailed study employing both structure and ligand-based computational approaches to select a set of compounds with the potential to inhibit M^pro [117]. In this work, RF was used to construct

binary QSAR models with a set of molecules obtained from the ChEMBL and PBD databases. Different descriptors were used to build the RF models, resulting in a set of 42 potential hits from the DrugBank database, 11 of which were also available in the National Center for Advancing Translational Sciences (NCATS) pharmaceutical collection. The authors' predictions found three compounds to be active and comparable to the cytopathic effect (CPE) assay data of NCATS.

Another important method used in QSAR studies is SVM (Figure 6B), which is a supervised approach based on the statistical learning theory [118]. The SVM algorithm incorporates the principles of structural risk minimization in its learning process and the reduction in complexity of mathematical functions used by the classifier. In summary, SVM constructs a separating hyperplane that maximizes the distance between the classifier and the nearest sample of each class, defined as margin separation [105]. Support vectors are the data points that lie closest to the decision surface (or hyperplane).

Although SVM presents an elegant mathematical formulation and good predictive performance, it also has some limitations, including the importance of making the right choice of hyperparameters and difficult interpretation of models. Fortunately, different strategies for tuning SVM hyperparameters can be found in literature [119,120].

Kowalewski and Ray (2020) developed a ML drug discovery pipeline to identify drug candidates for COVID-19. Initially, they collected assay data for 65 human targets with known interaction with SARS-CoV-2 proteins. SVM (for classification and regression) and RF were applied to predict the inhibitory activity and to screen FDA registered compounds and approved drugs. The predictions were filtered according to the estimated mammalian toxicity and vapor pressure with the aim at identifying volatile candidates and other inhibitors against multiple targets [121].

Inspired by the biological neural system, ANN are also interesting ML approaches. Multilayer perceptron (MLP) (Figure 6C) is a feed-forward neural network that has been widely applied in regression and classification tasks [110]. Its architecture consists of an input layer of neurons, an arbitrary number of hidden layers, and an output layer. Since the numbers of hidden layers and hidden neurons strongly influence the performance of the model, it is crucial to employ the right strategy to select hyperparameters in order to avoid overfitting and increase the generalization power of the network.

In comparison to MLP, DL neural networks architectures consist of a more complex set of hidden layers (Figure 6D) [122]. Such models contain hundreds to millions of units and require a great amount of training data to learn parameters, which is a computationally intensive process. One of the main advantages of DL models is their ability to handle complex data (i.e. text and images) by accomplishing automatic feature extraction from raw data, also called feature learning. Another frequently mentioned benefit is related to the scalability of such networks, referring to their ability to adjust the trade-offs between response time and accuracy [123].

There are a number of DL architectures that carry out specialized tasks such as (i) convolutional neural networks, which perform object detection; (ii) recurrent neural networks, commonly used for time series analysis; and (iii) adversarial neural networks, that can learn about the input data and attempt to reconstruct it as faithful as possible by only using its underlying patterns [107]. In Figure 7, we present two DL architectures recently applied in studies involving SARS-CoV-2. Figure 7A presents an example of convolutional neural networks, while Figure 7B shows an example of adversarial neural networks.

DL models have also been applied in studies related to COVID-19 identification, starting from detecting some of the disease patterns in lung X-ray images [124]. Khandelwal et al. (2020) also selected a set of molecules as drug candidates to treat COVID-19 using DL. The authors undertook a careful study on 31 drug candidates against M^{pro} (PDB ID: 6LU7), where DL architectures were employed to generate shape-based molecules starting from a 3D shape of their seed compound's pharmacophoric features. In this work, a convolutional neural network was trained as a variational autoencoder to learn the shape of PubChem's chemical structures. Recurrent neural networks were applied to sample and generate new molecules that were evaluated according to physicochemical characterization and drug-likeness. Using this procedure, they obtained a set of molecules (remdesivir, valrubicin, aprepitant, fulvestrant, and a novel therapeutic compound named nCorv-EMBS) as potential inhibitors and possible therapeutics for COVID-19, after evaluation in antiviral activity assays [125].

Bung et al. (2020) combined a generative DL model, transfer learning, and reinforcement learning to design molecules capable of inhibiting M^{pro}. The authors trained a generative model using a dataset with approximately 1.6 million drug-like molecules from ChEMBL. Transfer learning was employed to retrain the model with over 2,500 various protease inhibitor molecules, and reinforcement learning was applied to modulate the generative model aiming at producing molecules with desired properties. The trained model sampled 50,000 molecules from the chemical space, which were filtered based on many physicochemical properties. This process was followed by virtual screening and docking, resulting in a list of 31 compounds as potential hits which could be optimized into new therapies for
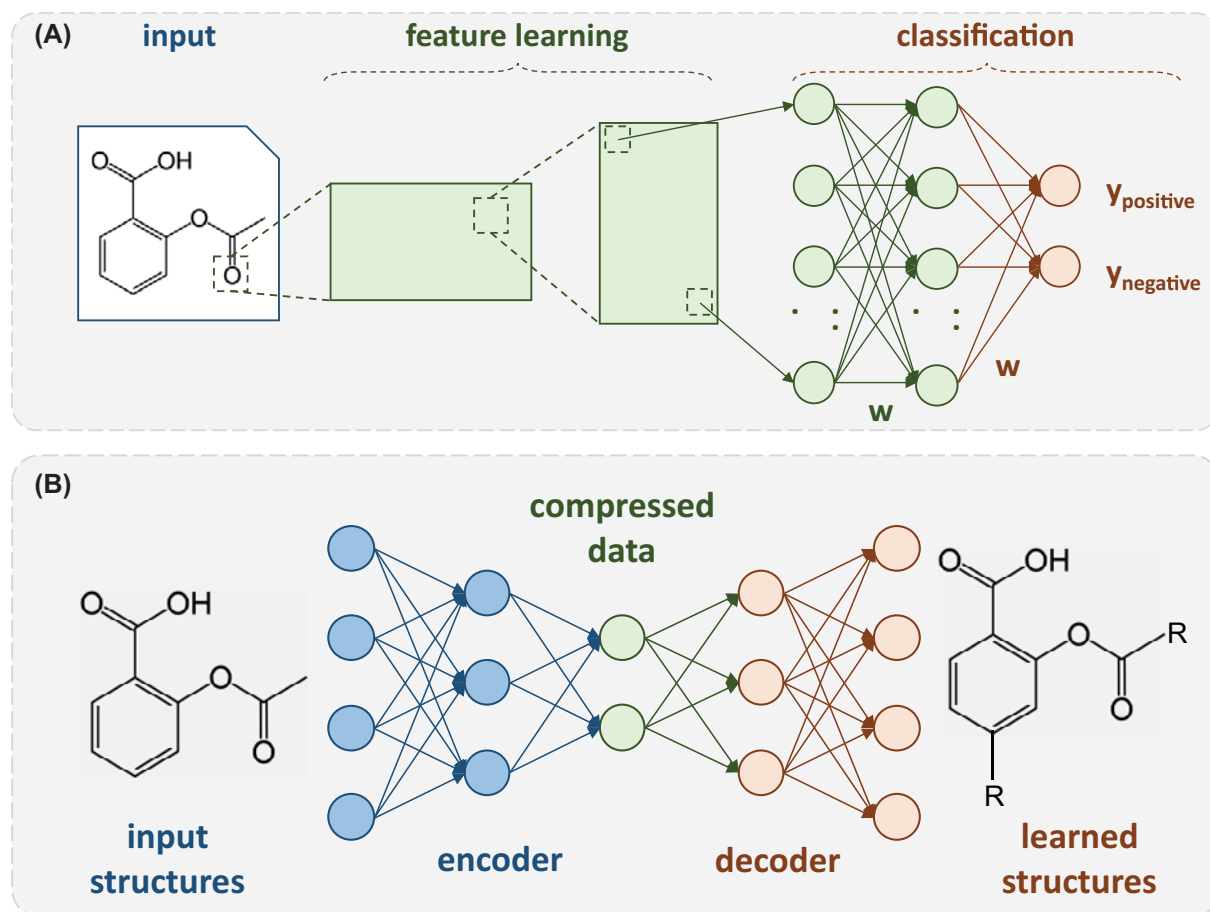
Figure 7. Special architectures of DL models
(**A**) Convolutional Neural Networks (CNN) and (**B**) Autoencoders.

COVID-19 [126]. Moreover, some important details of the docking methodology, such as definition of binding site, ligand and receptor preparation, and stereochemistry, have not also been described.

Gawriljuk et al. (2020) provided a detailed study of molecules as potential candidates against HeLa-ACE2. For this purpose, they compared different ML algorithms, including RF, support vector classification (SVC), k-nearest neighbors (kNN), and DL models, using a training dataset composed of 63 molecules and a test set containing 30 molecules. The results showed that RF and SVC presented the best prediction performance. It is worth mentioning that the DL models showed the poorest performance due to the lack of sufficient training data. The best models were employed in a virtual screening for selecting promising molecules, which were submitted to *in vitro* assays. Two of these compounds presented antiviral activity, with IC_{50} values of 8.4 µM and 540 nM, these representing useful potential starting points for COVID-19 focused drug discovery programs [106].

Conclusions

Everyday, as more lives are lost, a common effort from science, government, and population tries to gradually tackle COVID-19, aiming to mitigate its continuously growing number of disease sequels, morbidities and deaths, with the urge and rapid progression of the current pandemic scenario. Taking into account that there are not approved or licensed drugs to treat COVID-19 so far, the race to find potential drug candidates benefits from computational strategies, which have proven to be a powerful tool, with the potential to obtain a successful combined approach in the arduous process of drug design and discovery. The urgency to obtain inhibitors and potential drug candidates remains as a major objective to mitigate the disease outcomes, including the death toll. CADD and ML techniques have been employed in many protocols targeting SARS-CoV-2 macromolecules and are one the feasible options to speed up the drug design and discovery processes, leading to novel inhibitors and repurposing drugs. In this review,

we explored the potential of these many different approaches to find out or repurpose SARS-CoV-2 inhibitors as antivirals to treat COVID-19, facing, against the time, this pandemic scenario with an unflagging scientific effort.

Competing Interests

The authors declare that there are no competing interests associated with the manuscript.

Author Contribution

M.S.M.S., J.C.G., D.M.A.C., P.R.O., V.G.M., and K.M.H. conceived, designed and wrote the manuscript. P.R.O., V.G.M., and K.M.H. guided and revised all of the parts of the study. M.S.M.S., J.C.G., D.M.A.C., and V.G.M. also prepared the figures of the article. The final version of the manuscript was reviewed and approved by all authors.

Acknowledgements

Abbreviations

ADME, absorption, distribution, metabolism, and excretion; ANN, artificial neural networks; CADD, computer-aided drug design; COVID-19, coronavirus disease 2019; DL, deep learning; MD, molecular dynamics; MERS-CoV, Middle East respiratory syndrome coronavirus; ML, machine learning; MLP, multilayer perceptron; M^{pro}, main protease; NCATS, National Center for Advancing Translational Sciences; NSP, non-structural protein; PDB, Protein Data Bank; PL^{pro}, papain-like protease; QSAR, quantitative structure–activity relationship; RdRp, RNA-dependent RNA polymerase; RF, random forest; SARS-CoV-2, severe acute respiratory syndrome-related coronavirus 2; ssRNA+, single-stranded positive sense RNA; SVC, support vector classification; SVM, support vector machine.

References

1 Gorbalenya, A.E., Baker, S.C., Baric, R.S., de Groot, R.J., Drosten, C., Gulyaeva, A.A. et al. (2020) The species severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. *Nat. Microbiol.* **5**, 536–544

2 Xu, X., Chen, P., Wang, J., Feng, J., Zhou, H., Li, X. et al. (2020) Evolution of the novel coronavirus from the ongoing Wuhan outbreak and modeling of its spike protein for risk of human transmission. *Sci. China Life Sci.* **63**, 457–460, https://doi.org/10.1007/s11427-020-1637-5

3 World Health Organization (WHO) (2020) Emergency Response WHO Team (2020). Emergency Situational Updates, 24 November 2020. Weekly epidemiological update - 24 November 2020. https://www.who.int/publications/m/item/weekly-epidemiological-update---24-november-2020

4 Fehr, A.R. and Perlman, S. (2015) Coronaviruses: an overview of their replication and pathogenesis. *Coronaviruses* **1282**, 1–23, https://doi.org/10.1007/978-1-4939-2438-7˙1

5 Hage-Melim LI da, S., Federico, L.B., de Oliveira, N.K.S., Francisco, V.C.C., Correia, L.C., de Lima, H.B. et al. (2020) Virtual screening, ADME/Tox predictions and the drug repurposing concept for future use of old drugs against the COVID-19. *Life Sci.* **256**, 117963, https://doi.org/10.1016/j.lfs.2020.117963

6 de Wit, E., van Doremalen, N., Falzarano, D. and Munster, V.J. (2016) SARS and MERS: recent insights into emerging coronaviruses. *Nat. Rev. Microbiol.* **14**, 523–534, https://doi.org/10.1038/nrmicro.2016.81

7 Ong, S.W.X., Tan, Y.K., Chia, P.Y., Lee, T.H., Ng, O.T., Wong, M.S.Y. et al. (2020) Air, surface environmental, and personal protective equipment contamination by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) from a symptomatic patient. *JAMA* **323**, 1610–1612, https://doi.org/10.1001/jama.2020.3227

8 Ferretti, L., Wymant, C., Kendall, M., Zhao, L., Nurtay, A., Abeler-Dörner, L. et al. (2020) Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing. *Science* **368**, eabb6936, https://doi.org/10.1126/science.abb6936

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

10 de Wit, E., van Doremalen, N., Falzarano, D. and Munster, V.J. (2016) SARS and MERS: recent insights into emerging coronaviruses. *Nat. Rev. Microbiol.* **14**, 523–534, https://doi.org/10.1038/nrmicro.2016.81

11 Cui, J., Li, F. and Shi, Z.-L. (2019) Origin and evolution of pathogenic coronaviruses. *Nat. Rev. Microbiol.* **17**, 181–192, https://doi.org/10.1038/s41579-018-0118-9

12 Chen, N., Zhou, M., Dong, X., Qu, J., Gong, F., Han, Y. et al. (2020) Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. *Lancet Lond. Engl.* **395**, 507–513, https://doi.org/10.1016/S0140-6736(20)30211-7

13 Huang, C., Wang, Y., Li, X., Ren, L., Zhao, J., Hu, Y. et al. (2020) Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet* **395**, 497–506, https://doi.org/10.1016/S0140-6736(20)30183-5

14 Long, Q.-X., Tang, X.-J., Shi, Q.-L., Li, Q., Deng, H.-J., Yuan, J. et al. (2020) Clinical and immunological assessment of asymptomatic SARS-CoV-2 infections. *Nat. Med.* **26**, 1200–1204, https://doi.org/10.1038/s41591-020-0965-6

15 Tay, M.Z., Poh, C.M., Rénia, L., MacAry, P.A. and Ng, L.F.P. (2020) The trinity of COVID-19: immunity, inflammation and intervention. *Nat. Rev. Immunol.* **20**, 363–374, https://doi.org/10.1038/s41577-020-0311-8

16 Amanat, F. and Krammer, F. (2020) SARS-CoV-2 vaccines: status report. *Immunity* **52**, 583–589, https://doi.org/10.1016/j.immuni.2020.03.007

17 McCloskey, B., Dar, O., Zumla, A. and Heymann, D.L. (2014) Emerging infectious diseases and pandemic potential: status quo and reducing risk of global spread. *Lancet Infect. Dis.* **14**, 1001–1010, https://doi.org/10.1016/S1473-3099(14)70846-1

18 Rauch, S., Jasny, E., Schmidt, K.E. and Petsch, B. (2018) New vaccine technologies to combat outbreak situations. *Front. Immunol.* **9**, 1963, https://doi.org/10.3389/fimmu.2018.01963

19 Mulligan, M.J., Lyke, K.E., Kitchin, N., Absalon, J., Gurtman, A., Lockhart, S. et al. (2020) Phase I/II study of COVID-19 RNA vaccine BNT162b1 in adults. *Nature* **586**, 589–593, https://doi.org/10.1038/s41586-020-2639-4

20 Folegatti, P.M., Ewer, K.J., Aley, P.K., Angus, B., Becker, S., Belij-Rammerstorfer, S. et al. (2020) Safety and immunogenicity of the ChAdOx1 nCoV-19 vaccine against SARS-CoV-2: a preliminary report of a phase 1/2, single-blind, randomised controlled trial. *Lancet North Am. Ed.* **396**, 467–478, https://doi.org/10.1016/S0140-6736(20)31604-4

21 Jackson, L.A., Anderson, E.J., Rouphael, N.G., Roberts, P.C., Makhene, M., Coler, R.N. et al. (2020) An mRNA vaccine against SARS-CoV-2 — preliminary report. *N. Engl. J. Med.* **383**, 1920–1931, https://doi.org/10.1056/NEJMoa2022483

22 Logunov, D.Y., Dolzhikova, I.V., Zubkova, O.V., Tukhvatullin, A.I., Shcheblyakov, D.V., Dzharullaeva, A.S. et al. (2020) Safety and immunogenicity of an rAd26 and rAd5 vector-based heterologous prime-boost COVID-19 vaccine in two formulations: two open, non-randomised phase 1/2 studies from Russia. *Lancet North Am. Ed.* **396**, 887–897, https://doi.org/10.1016/S0140-6736(20)31866-3

23 Xia, S., Zhang, Y., Wang, Y., Wang, H., Yang, Y., Gao, G.F. et al. (2020) Safety and immunogenicity of an inactivated SARS-CoV-2 vaccine, BBIBP-CorV: a randomised, double-blind, placebo-controlled, phase 1/2 trial. *Lancet Infect. Dis.* **21**, 39–51, https://doi.org/10.1016/S1473-3099(20)30831-8

24 Trogen, B., Oshinsky, D. and Caplan, A. (2020) Adverse consequences of rushing a SARS-CoV-2 vaccine: implications for public trust. *JAMA* **323**, 2460, https://doi.org/10.1001/jama.2020.8917

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

26 Santos I de, A., Grosche, V.R., Bergamini, F.R.G., Sabino-Silva, R. and Jardim, A.C.G. (2020) Antivirals against coronaviruses: candidate drugs for SARS-CoV-2 treatment? *Front. Microbiol.* **11**, 1818, https://doi.org/10.3389/fmicb.2020.01818

27 Young, B.E., Fong, S.-W., Chan, Y.-H., Mak, T.-M., Ang, L.W., Anderson, D.E. et al. (2020) Effects of a major deletion in the SARS-CoV-2 genome on the severity of infection and the inflammatory response: an observational cohort study. *Lancet* **396**, 603–611, https://doi.org/10.1016/S0140-6736(20)31757-8

28 Morens, D.M. and Fauci, A.S. (2020) Emerging pandemic diseases: how we got to COVID-19. *Cell* **182**, 1077–1092, https://doi.org/10.1016/j.cell.2020.08.021

29 Alpern, J.D. and Gertner, E. (2020) Off-label therapies for COVID-19—are we all in this together? *Clin. Pharmacol. Ther.* **108**, 182–184, https://doi.org/10.1002/cpt.1862

30 Kalil, A.C. (2020) Treating COVID-19—off-label drug use, compassionate use, and randomized clinical trials during pandemics. *JAMA* **323**, 1897, https://doi.org/10.1001/jama.2020.4742

31 Ciulli, A. (2013) Biophysical screening for the discovery of small-molecule ligands. *Methods Mol. Biol.* **1008**, 357–388, https://doi.org/10.1007/978-1-62703-398-5´13

32 Holdgate, G., Embrey, K., Milbradt, A. and Davies, G. (2019) Biophysical methods in early drug discovery. *ADMET DMPK* **7**, 222–241, https://doi.org/10.5599/admet.733

33 Ionescu, M.I. (2020) An overview of the crystallized structures of the SARS-CoV-2. *Protein J.* **39**, 600–618, https://doi.org/10.1007/s10930-020-09933-w

34 Parvathaneni, V. and Gupta, V. (2020) Utilizing drug repurposing against COVID-19 - efficacy, limitations, and challenges. *Life Sci.* **259**, 118275, https://doi.org/10.1016/j.lfs.2020.118275

35 Talele, T.T., Khedkar, S.A. and Rigby, A.C. (2010) Successful applications of computer aided drug discovery: moving drugs from concept to the clinic. *Curr. Top. Med. Chem.* **10**, 127–141, https://doi.org/10.2174/156802610790232251

36 Beeraka, N.M., Sadhu, S.P., Madhunapantula, S.V., Rao Pragada, R., Svistunov, A.A., Nikolenko, V.N. et al. (2020) Strategies for targeting SARS CoV-2: small molecule inhibitors-the current status. *Front. Immunol.* **11**, 552925, https://doi.org/10.3389/fimmu.2020.552925

37 Jin, Z., Du, X., Xu, Y., Deng, Y., Liu, M., Zhao, Y. et al. (2020) Structure of Mpro from SARS-CoV-2 and discovery of its inhibitors. *Nature* **582**, 289–293, https://doi.org/10.1038/s41586-020-2223-y

38 Osipiuk, J., Azizi, S.-A., Dvorkin, S., Endres, M., Jedrzejczak, R., Jones, K.A. et al. (2021) Structure of papain-like protease from SARS-CoV-2 and its complexes with non-covalent inhibitors. *Nat. Commun.* **12**, 743, https://doi.org/10.1038/s41467-021-21060-3

39 Peng, Q., Peng, R., Yuan, B., Zhao, J., Wang, M., Wang, X. et al. (2020) Structural and biochemical characterization of the nsp12-nsp7-nsp8 core polymerase complex from SARS-CoV-2. *Cell Rep.* **31**, 107774, https://doi.org/10.1016/j.celrep.2020.107774

40 Chen, J., Malone, B., Llewellyn, E., Grasso, M., Shelton, P.M.M., Olinares, P.D.B. et al. (2020) Structural basis for helicase-polymerase coupling in the SARS-CoV-2 replication-transcription complex. *Cell* **182**, 1560.e13–1573.e13, https://doi.org/10.1016/j.cell.2020.07.033

41 null, F., Kumar, B.K., Sekhar, K.V.G.C., Kunjiappan, S., Jamalis, J., Balaña-Fouce, R. et al. (2020) Druggable targets of SARS-CoV-2 and treatment opportunities for COVID-19. *Bioorg. Chem.* **104**, 104269

42 Goyal, R.K., Majeed, J., Tonk, R., Dhobi, M., Patel, B., Sharma, K. et al. (2020) Current targets and drug candidates for prevention and treatment of SARS-CoV-2 (COVID-19) infection. *Rev. Cardiovasc. Med.* **21**, 365–384, https://doi.org/10.31083/j.rcm.2020.03.118

43 Prajapat, M., Sarma, P., Shekhar, N., Avti, P., Sinha, S., Kaur, H. et al. (2020) Drug targets for corona virus: a systematic review. *Indian J. Pharmacol.* **52**, 56–65

44 Saxena, A. (2020) Drug targets for COVID-19 therapeutics: ongoing global efforts. *J. Biosci.* **45**, 87, https://doi.org/10.1007/s12038-020-00067-w

45 Suganya, S., Divya, S. and Parani, M. (2020) Severe acute respiratory syndrome-coronavirus-2: Current advances in therapeutic targets and drug development. *Rev. Med. Virol.*, https://doi.org/10.1002/rmv.2174

46 Ullrich, S. and Nitsche, C. (2020) The SARS-CoV-2 main protease as drug target. *Bioorg. Med. Chem. Lett.* **30**, 127377, https://doi.org/10.1016/j.bmcl.2020.127377

47 Pillaiyar, T., Manickam, M., Namasivayam, V., Hayashi, Y. and Jung, S.-H. (2016) An overview of severe acute respiratory syndrome-coronavirus (SARS-CoV) 3CL protease inhibitors: peptidomimetics and small molecule chemotherapy. *J. Med. Chem.* **59**, 6595–6628, https://doi.org/10.1021/acs.jmedchem.5b01461

48 Maltarollo, V.G., Gertrudes, J.C., Oliveira, P.R. and Honorio, K.M. (2015) Applying machine learning techniques for ADME-Tox prediction: a review. *Expert Opin. Drug Metab. Toxicol.* **11**, 259–271, https://doi.org/10.1517/17425255.2015.980814

49 Yu, W. and MacKerell, A.D. (2017) Computer-aided drug design methods. *Methods Mol. Biol.* **1520**, 85–106, https://doi.org/10.1007/978-1-4939-6634-9_5

50 Maltarollo, V.G., Kronenberger, T., Espinoza, G.Z., Oliveira, P.R. and Honorio, K.M. (2019) Advances with support vector machines for novel drug discovery. *Expert Opin. Drug. Discov.* **14**, 23–33, https://doi.org/10.1080/17460441.2019.1549033

51 Onawole, A.T., Sulaiman, K.O., Kolapo, T.U., Akinde, F.O. and Adegoke, R.O. (2020) COVID-19: CADD to the rescue. *Virus Res.* **285**, 198022, https://doi.org/10.1016/j.virusres.2020.198022

52 Ruan, Z., Liu, C., Guo, Y., He, Z., Huang, X., Jia, X. et al. (2020) SARS-CoV-2 and SARS-CoV: virtual screening of potential inhibitors targeting RNA-dependent RNA polymerase activity (NSP12). *J. Med. Virol.*, https://doi.org/10.1002/jmv.26222

53 Kumar, Y., Singh, H. and Patel, C.N. (2020) In silico prediction of potential inhibitors for the main protease of SARS-CoV-2 using molecular docking and dynamics simulation based drug-repurposing. *J. Infect. Public Health* **13**, 1210–1223, https://doi.org/10.1016/j.jiph.2020.06.016

54 Jiménez-Alberto, A., Ribas-Aparicio, R.M., Aparicio-Ozores, G. and Castelán-Vega, J.A. (2020) Virtual screening of approved drugs as potential SARS-CoV-2 main protease inhibitors. *Comput. Biol. Chem.* **88**, 107325, https://doi.org/10.1016/j.compbiolchem.2020.107325

55 Hegyi, A. and Ziebuhr, J. (2002) Conservation of substrate specificities among coronavirus main proteases. *J. Gen. Virol.* **83**, 595–599, https://doi.org/10.1099/0022-1317-83-3-595

56 Yang, H., Xie, W., Xue, X., Yang, K., Ma, J., Liang, W. et al. (2005) Design of wide-spectrum inhibitors targeting coronavirus main proteases. *PLoS Biol.* **3**, e324, https://doi.org/10.1371/journal.pbio.0030324

57 Kim, Y., Liu, H., Galasiti Kankanamalage, A.C., Weerasekara, S., Hua, D.H., Groutas, W.C. et al. (2016) Reversal of the progression of fatal coronavirus infection in cats by a broad-spectrum coronavirus protease inhibitor. *PLoS Pathog.* **12**, e1005531, https://doi.org/10.1371/journal.ppat.1005531

58 Bzówka, M., Mitusińska, K., Raczyńska, A., Samol, A., Tuszyński, J.A. and Góra, A. (2020) Structural and evolutionary analysis indicate that the SARS-CoV-2 Mpro is a challenging target for small-molecule inhibitor design. *Int. J. Mol. Sci.* **21**, 3099, https://doi.org/10.3390/ijms21093099

59 Deng, X., StJohn, S.E., Osswald, H.L., O'Brien, A., Banach, B.S., Sleeman, K. et al. (2014) Coronaviruses resistant to a 3C-like protease inhibitor are attenuated for replication and pathogenesis, revealing a low genetic barrier but high fitness cost of resistance. *J. Virol.* **88**, 11886–11898, https://doi.org/10.1128/JVI.01528-14

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

61 Kneller, D.W., Phillips, G., O'Neill, H.M., Jedrzejczak, R., Stols, L., Langan, P. et al. (2020) Structural plasticity of SARS-CoV-2 3CL M pro active site cavity revealed by room temperature X-ray crystallography. *Nat. Commun.* **11**, 3202, https://doi.org/10.1038/s41467-020-16954-7

62 Xia, B. and Kang, X. (2011) Activation and maturation of SARS-CoV main protease. *Protein Cell* **2**, 282–290, https://doi.org/10.1007/s13238-011-1034-1

63 Muramatsu, T., Kim, Y.-T., Nishii, W., Terada, T., Shirouzu, M. and Yokoyama, S. (2013) Autoprocessing mechanism of severe acute respiratory syndrome coronavirus 3C-like protease (SARS-CoV 3CLpro) from its polyproteins. *FEBS J.* **280**, 2002–2013, https://doi.org/10.1111/febs.12222

64 Xue, X., Yu, H., Yang, H., Xue, F., Wu, Z., Shen, W. et al. (2008) Structures of two coronavirus main proteases: implications for substrate binding and antiviral drug design. *J. Virol.* **82**, 2515–2527, https://doi.org/10.1128/JVI.02114-07

65 Zhang, L., Lin, D., Sun, X., Curth, U., Drosten, C., Sauerhering, L. et al. (2020) Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved α-ketoamide inhibitors. *Science* **368**, 409–412, https://doi.org/10.1126/science.abb3405

66 Hsu, W.-C., Chang, H.-C., Chou, C.-Y., Tsai, P.-J., Lin, P.-I. and Chang, G.-G. (2005) Critical assessment of important regions in the subunit association and catalytic action of the severe acute respiratory syndrome coronavirus main protease. *J. Biol. Chem.* **280**, 22741–22748, https://doi.org/10.1074/jbc.M502556200

67 Tsai, M.-Y., Chang, W.-H., Liang, J.-Y., Lin, L.-L., Chang, G.-G. and Chang, H.-P. (2010) Essential covalent linkage between the chymotrypsin-like domain and the extra domain of the SARS-CoV main protease. *J. Biochem. (Tokyo)* **148**, 349–358, https://doi.org/10.1093/jb/mvq071

68 DeLano, W.L. (2002) The PyMOL molecular graphics system. DeLano Scientific, San Carlos, CA., http://www.pymol.org

69 Ren, Z., Yan, L., Zhang, N., Guo, Y., Yang, C., Lou, Z. et al. (2013) The newly emerged SARS-like coronavirus HCoV-EMC also has an "Achilles' heel": current effective inhibitor targeting a 3C-like protease. *Protein Cell* **4**, 248–250, https://doi.org/10.1007/s13238-013-2841-3

70 Wang, F., Chen, C., Tan, W., Yang, K. and Yang, H. (2016) Structure of main protease from human coronavirus NL63: insights for wide spectrum anti-coronavirus drug design. *Sci. Rep.* **6**, 22677, https://doi.org/10.1038/srep22677

71 Mahanta, S., Chowdhury, P., Gogoi, N., Goswami, N., Borah, D., Kumar, R. et al. (2020) Potential anti-viral activity of approved repurposed drug against main protease of SARS-CoV-2: an *in silico* based approach. *J. Biomol. Struct. Dyn.* 1–10, https://doi.org/10.1080/07391102.2020.1768902

72 Liang, J., Pitsillou, E., Karagiannis, C., Darmawan, K.K., Ng, K., Hung, A. et al. (2020) Interaction of the prototypical α-ketoamide inhibitor with the SARS-CoV-2 main protease active site in silico: molecular dynamic simulations highlight the stability of the ligand-protein complex. *Comput. Biol. Chem.* **87**, 107292, https://doi.org/10.1016/j.compbiolchem.2020.107292

73 Pant, S., Singh, M., Ravichandiran, V., Murty, U.S.N. and Srivastava, H.K. (2020) Peptide-like and small-molecule inhibitors against Covid-19. *J. Biomol. Struct. Dyn.* 1–10, https://doi.org/10.1080/07391102.2020.1757510

74 Sasidharan, S., Selvaraj, C., Singh, S.K., Dubey, V.K., Kumar, S., Fialho, A.M. et al. (2020) Bacterial protein azurin and derived peptides as potential anti-SARS-CoV-2 agents: insights from molecular docking and molecular dynamics simulations. *J. Biomol. Struct. Dyn.* 1–16, https://doi.org/10.1080/07391102.2020.1787864

75 Jin, Z., Zhao, Y., Sun, Y., Zhang, B., Wang, H., Wu, Y. et al. (2020) Structural basis for the inhibition of SARS-CoV-2 main protease by antineoplastic drug carmofur. *Nat. Struct. Mol. Biol.* **27**, 529–532, https://doi.org/10.1038/s41594-020-0440-6

76 Dai, W., Zhang, B., Jiang, X.-M., Su, H., Li, J., Zhao, Y. et al. (2020) Structure-based design of antiviral drug candidates targeting the SARS-CoV-2 main protease. *Science* **368**, 1331–1335, https://doi.org/10.1126/science.abb4489

77 Ma, C., Sacco, M.D., Hurst, B., Townsend, J.A., Hu, Y., Szeto, T. et al. (2020) Boceprevir, GC-376, and calpain inhibitors II, XII inhibit SARS-CoV-2 viral replication by targeting the viral main protease. *Cell Res.* **30**, 678–692, https://doi.org/10.1038/s41422-020-0356-z

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

79 Khaerunnisa, S., Kurniawan, H., Awaluddin, R., Suhartati, S. and Soetjipto, S. (2020) Potential inhibitor of COVID-19 main protease (Mpro) from several medicinal plant compounds by molecular docking study. *Preprints*, https://doi.org/10.20944/preprints202003.0226

80 Xu, Z., Peng, C., Shi, Y., Zhu, Z., Mu, K., Wang, X. et al. (2020) Nelfinavir was predicted to be a potential inhibitor of 2019 nCov main protease by an integrative approach combining homology modelling, molecular docking and binding free energy calculation. https://doi.org/10.1101/2020.01.27.921627

81 Kumar, D., Kumari, K., Vishvakarma, V.K., Jayaraj, A., Kumar, D., Ramappa, V.K. et al. (2020) Promising inhibitors of main protease of novel corona virus to prevent the spread of COVID-19 using docking and molecular dynamics simulation. *J. Biomol. Struct. Dyn.* 1–15, https://doi.org/10.1080/07391102.2020.1779131

82 Liu, X. and Wang, X.-J. (2020) Potential inhibitors against 2019-nCoV coronavirus M protease from clinically approved medicines. *J. Genet. Genomics* **47**, 119–121, https://doi.org/10.1016/j.jgg.2020.02.001

83 Liu, S., Zheng, Q. and Wang, Z. (2020) Potential covalent drugs targeting the main protease of the SARS-CoV-2 coronavirus. *Bioinformatics* **36**, 3295–3298, https://doi.org/10.1093/bioinformatics/btaa224

84 Zhang, Y., Zheng, Q., Zhou, Y. and Liu, S. (2020) Repurposing clinical drugs as AdoMetDC inhibitors using the SCAR strategy. *Front. Pharmacol.* **11**, 248, eCollection 2020, https://doi.org/10.3389/fphar.2020.00248

85 Trezza, A., Iovinelli, D., Santucci, A., Prischi, F. and Spiga, O. (2020) An integrated drug repurposing strategy for the rapid identification of potential SARS-CoV-2 viral inhibitors. *Sci. Rep.* **10**, 13866, https://doi.org/10.1038/s41598-020-70863-9

86 Harcourt, B.H., Jukneliene, D., Kanjanahaluethai, A., Bechill, J., Severson, K.M., Smith, C.M. et al. (2004) Identification of severe acute respiratory syndrome coronavirus replicase products and characterization of papain-like protease activity. *J. Virol.* **78**, 13600–13612, https://doi.org/10.1128/JVI.78.24.13600-13612.2004

87 Barretto, N., Jukneliene, D., Ratia, K., Chen, Z., Mesecar, A.D. and Baker, S.C. (2005) The papain-like protease of severe acute respiratory syndrome coronavirus has deubiquitinating activity. *J. Virol.* **79**, 15189–15198, https://doi.org/10.1128/JVI.79.24.15189-15198.2005

88 Mirza, M.U., Ahmad, S., Abdullah, I. and Froeyen, M. (2020) Identification of novel human USP2 inhibitor and its putative role in treatment of COVID-19 by inhibiting SARS-CoV-2 papain-like (PLpro) protease. *Comput. Biol. Chem.* **89**, 107376, https://doi.org/10.1016/j.compbiolchem.2020.107376

89 Kouznetsova, V.L., Zhang, A., Tatineni, M., Miller, M.A. and Tsigelny, I.F. (2020) Potential COVID-19 papain-like protease PLpro inhibitors: repurposing FDA-approved drugs. *Peer J.* **8**, e9965, https://doi.org/10.7717/peerj.9965

90 Arya, R., Das, A., Prashar, V. and Kumar, M. (2020) Potential inhibitors against papain-like protease of novel coronavirus (SARS-CoV-2) from FDA approved drugs. *Peer J.* **8**, e9965, https://doi.org/10.7717/peerj.9965

91 Welker, A., Kersten, C., Müller, C., Madhugiri, R., Zimmer, C., Müller, P. et al. (2020) Structure-Activity Relationships of novel benzamides and isoindolines, designed as SARS-CoV protease inhibitors - effective against SARS-CoV-2. *Chem. Med. Chem.* **16**, 340–354, https://doi.org/10.1002/cmdc.202000548

92 Gao, X., Qin, B., Chen, P., Zhu, K., Hou, P., Wojdyla, J.A. et al. (2020) Crystal structure of SARS-CoV-2 papain-like protease. *Acta Pharm. Sin. B* **11**, 237–245, https://doi.org/10.1016/j.apsb.2020.08.014

93 Ratia, K., Pegan, S., Takayama, J., Sleeman, K., Coughlin, M., Baliji, S. et al. (2008) A noncovalent class of papain-like protease/deubiquitinase inhibitors blocks SARS virus replication. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 16119–16124, https://doi.org/10.1073/pnas.0805240105

94 Subissi, L., Posthuma, C.C., Collet, A., Zevenhoven-Dobbe, J.C., Gorbalenya, A.E., Decroly, E. et al. (2014) One severe acute respiratory syndrome coronavirus protein complex integrates processive RNA polymerase and exonuclease activities. *Proc. Natl. Acad. Sci. U. S. A.* **111**, E3900–E3909, https://doi.org/10.1073/pnas.1323705111

95 Yin, W., Mao, C., Luan, X., Shen, D.-D., Shen, Q., Su, H. et al. (2020) Structural basis for inhibition of the RNA-dependent RNA polymerase from SARS-CoV-2 by remdesivir. *Science* **368**, 1499–1504, https://doi.org/10.1126/science.abc1560

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

97 Holshue, M.L., DeBolt, C., Lindquist, S., Lofy, K.H., Wiesman, J., Bruce, H. et al. (2020) First case of 2019 novel coronavirus in the United States. *N. Engl. J. Med.* **382**, 929–936, https://doi.org/10.1056/NEJMoa2001191

98 Beigel, J.H., Tomashek, K.M., Dodd, L.E., Mehta, A.K., Zingman, B.S., Kalil, A.C. et al. (2020) Remdesivir for the treatment of Covid-19 — final report. *N. Engl. J. Med.* **383**, 1813–1826, https://doi.org/10.1056/NEJMoa2007764

99 Elfiky, A.A. (2020) Ribavirin, Remdesivir, Sofosbuvir, Galidesivir, and Tenofovir against SARS-CoV-2 RNA dependent RNA polymerase (RdRp): a molecular docking study. *Life Sci.* **253**, 117592, https://doi.org/10.1016/j.lfs.2020.117592

100 Elfiky, A.A. (2020) Anti-HCV, nucleotide inhibitors, repurposing against COVID-19. *Life Sci.* **248**, 117477, https://doi.org/10.1016/j.lfs.2020.117477

101 Dutta, K. et al. (2020) Beclabuvir can inhibit the RNA-dependent RNA polymerase of newly emerged novel coronavirus (SARS-CoV-2). *Preprints*, https://doi.org/10.20944/preprints202003.0395

102 Ivanov, K.A., Thiel, V., Dobbe, J.C., van der Meer, Y., Snijder, E.J. and Ziebuhr, J. (2004) Multiple enzymatic activities associated with severe acute respiratory syndrome coronavirus helicase. *J. Virol.* **78**, 5619–5632, https://doi.org/10.1128/JVI.78.11.5619-5632.2004

103 Adedeji, A.O. and Lazarus, H. (2016) Biochemical characterization of Middle East Respiratory Syndrome coronavirus helicase. *mSphere* **1**, e00235–16, https://doi.org/10.1128/mSphere.00235-16

104 Borgio, J.F., Alsuwat, H.S., Al Otaibi, W.M., Ibrahim, A.M., Almandil, N.B., Al Asoom, L.I. et al. (2020) State-of-the-art tools unveil potent drug targets amongst clinically approved drugs to inhibit helicase in SARS-CoV-2. *Arch. Med. Sci.* **16**, 508–518, https://doi.org/10.5114/aoms.2020.94567

105 Gertrudes, J.C., Maltarollo, V.G., Silva, R.A., Oliveira, P.R., Honório, K.M. and da Silva, A.B.F. (2012) Machine learning techniques and drug design. *Curr. Med. Chem.* **19**, 4289–4297, https://doi.org/10.2174/092986712802884259

106 Gawriljuk, V.O., Kyaw Zin, P.P., Foil, D.H., Bernatchez, J., Beck, S., Beutler, N. et al. (2020) Machine learning models identify inhibitors of SARS-CoV-2. *Pharmacol. Toxicol.*, https://doi.org/10.1101/2020.06.16.154765

107 Nikolaev, G.I., Shuldov, N.A., Anischenko, A.I., Tuzikov, A.V. and Andrianov, A.M. (2020) Development of a neural network-based approach for prediction of potential HIV-1 entry inhibitors using deep learning and molecular modeling methods. In *Bioinformatics Research and Applications* (Cai, Z., Mandoiu, I., Narasimhan, G., Skums, P. and Guo, X., eds), pp. 304–311, Springer International Publishing, Cham, (Lecture Notes in Computer Science)

108 Zhou, Y., Wang, F., Tang, J., Nussinov, R. and Cheng, F. (2020) Artificial intelligence in COVID-19 drug repurposing. *Lancet Digit. Health* **2**, e667–e676, https://doi.org/10.1016/S2589-7500(20)30192-8

109 Loucera, C., Esteban-Medina, M., Rian, K., Falco, M.M., Dopazo, J. and Peña-Chilet, M. (2020) Drug repurposing for COVID-19 using machine learning and mechanistic models of signal transduction circuits related to SARS-CoV-2 infection. *Signal Transduct. Target Ther.* **5**, 1–3, https://doi.org/10.1038/s41392-020-00417-y

110 Poole, D.L. and Mackworth, A.K. (2010) *Artificial Intelligence: Foundations of Computational Agents*, Cambridge University Press, Cambridge

111 Geurts, P., Ernst, D. and Wehenkel, L. (2006) Extremely randomized trees. *Mach. Learn.* **63**, 3–42, https://doi.org/10.1007/s10994-006-6226-1

112 Chen, T. and Guestrin, C. (2016) XGBoost: A Scalable Tree Boosting System. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 785–794, Association for Computing Machinery, New York, NY, U.S.A., (KDD'16)

113 Rodrigues, G.C.S., Maia, M.D.S., de Menezes, R.P.B., Cavalcanti, A.B.S., de Sousa, N.F., de Moura, ÉP et al. (2020) Ligand and structure-based virtual screening of lamiaceae diterpenes with potential activity against a novel coronavirus (2019-NCOV). *Curr. Top. Med. Chem.* **20**, 2126–2145, https://doi.org/10.2174/1568026620666200716114546

114 Scotti, M.T., Herrera-Acevedo, C., Oliveira, T.B., Costa, R.P.O., Santos SYK de, O., Rodrigues, R.P. et al. (2018) SistematX, an online web-based cheminformatics tool for data management of secondary metabolites. *Molecule (Basel)* **23**, 103, https://doi.org/10.3390/molecules23010103

115 Alves, V.M., Golbraikh, A., Capuzzi, S.J., Liu, K., Lam, W.I., Korn, D.R. et al. (2018) Multi-Descriptor Read Across (MuDRA): a simple and transparent approach for developing accurate quantitative structure-activity relationship models. *J. Chem. Inf. Model.* **58**, 1214–1223, https://doi.org/10.1021/acs.jcim.8b00124

116 Thomsen, R. and Christensen, M.H. (2006) MolDock: a new technique for high-accuracy molecular docking. *J. Med. Chem.* **49**, 3315–3321, https://doi.org/10.1021/jm051197e

117 Alves, V., Bobrowski, T., Melo-Filho, C., Korn, D., Auerbach, S., Schmitt, C. et al. (2020) QSAR Modeling of SARS-CoV M pro inhibitors identifies Sufugolix, Cenicriviroc, Proglumetacin, and other drugs as candidates for repurposing against SARS-CoV-2. *Mol. Inform.* **40**, 2000113, https://doi.org/10.1002/minf.20200011

118 Pedrycz, W. (1999) *Advances in Kernel Methods. Support Vector Learning* (Scholkopf, B., Burges, C.J.C. and Smola, A.J., eds), p. 376, MIT Press, Cambridge, ISBN 0-262-19416-3. Neurocomputing. 2002

119 Mantovani, R., Rossi, A., Vanschoren, J. and Carvalho, A. (2015) Meta-learning recommendation of default hyper-parameter values for SVMs in classifications tasks. *CEUR Workshop Proceedings* **1455**

120 Mantovani, R., Rossi, A., Alcobaça, E., Castro Gertrudes, J., Barbon, S. and de Carvalho, A. (2020) Rethinking defaults values: a low cost and efficient strategy to define hyperparameters. pp. 1–36, Cornell University, arXiv:2008.00025v2

121 Kowalewski, J. and Ray, A. (2020) Predicting novel drugs for SARS-CoV-2 using machine learning from a >10 million chemical space. *Heliyon* **6**, e04639

122 LeCun, Y., Bengio, Y. and Hinton, G. (2015) Deep learning. *Nature* **521**, 436–444, https://doi.org/10.1038/nature14539

123 Zhang, L., Tan, Z., Song, J., Chen, J., Bao, C. and Ma, K. (2019) SCAN: a scalable neural networks framework towards compact and efficient models. pp. 1–10, Cornell University., arXiv:1906.03951v1

124 Luz, E., Silva, P., Pedrosa Silva, R., Silva, L., Moreira, G. and Menotti, D. (2020) Towards an effective and efficient deep learning model for COVID-19 patterns detection in X-ray images. pp. 1–31, Cornell University., arXiv:2004.05717v4

125 Nayarisseri, A., Khandelwal, R., Madhavi, M., Selvaraj, C., Panwar, U., Sharma, K. et al. (2020) Shape-based machine learning models for the potential novel COVID-19 protease inhibitors assisted by molecular dynamics simulation. *Curr. Top. Med. Chem.* **20**, 2146–2167, https://doi.org/10.2174/1568026620666200704135327

126 Bung, N., Krishnan, S.R., Bulusu, G. and Roy, A. (2020) *De novo* design of new chemical entities (NCEs) for SARS-CoV-2 using artificial intelligence. *Fut. Med. Chem.*, https://doi.org/10.4155/fmc-2020-0262