UCSF

Title

Artificial Intelligence for COVID-19 Drug Discovery and Vaccine Development.

Permalink

https://escholarship.org/uc/item/5nr1c41q

Authors

Keshavarzi Arshadi, Arash
Webb, Julia
Salem, Milad
et al.

Publication Date

2020

DOI

10.3389/frai.2020.00065

Peer reviewed

Artificial Intelligence for COVID-19 Drug Discovery and Vaccine Development

Arash Keshavarzi Arshadi[1], Julia Webb[1], Milad Salem[2], Emmanuel Cruz[3],
Stacie Calad-Thomson[4], Niloofar Ghadirian[5], Jennifer Collins[1], Elena Diez-Cecilia[3],
Brendan Kelly[3], Hani Goodarzi[6] and Jiann Shiun Yuan[2]*

[1] Burnett School of Biomedical Sciences, University of Central Florida, Orlando, FL, United States, [2] Department of Electrical and Computer Engineering, University of Central Florida, Orlando, FL, United States, [3] A2A Pharmaceuticals, Cambridge, MA, United States, [4] Atomwise Inc., San Francisco, CA, United States, [5] Department of Chemistry and Biochemistry, University of Arizona, Tucson, AZ, United States, [6] Department of Biochemistry and Biophysics, Helen Diller Family Comprehensive Cancer Center, University of California, San Francisco, San Francisco, CA, United States

SARS-COV-2 has roused the scientific community with a call to action to combat the growing pandemic. At the time of this writing, there are as yet no novel antiviral agents or approved vaccines available for deployment as a frontline defense. Understanding the pathobiology of COVID-19 could aid scientists in their discovery of potent antivirals by elucidating unexplored viral pathways. One method for accomplishing this is the leveraging of computational methods to discover new candidate drugs and vaccines *in silico*. In the last decade, machine learning-based models, trained on specific biomolecules, have offered inexpensive and rapid implementation methods for the discovery of effective viral therapies. Given a target biomolecule, these models are capable of predicting inhibitor candidates in a structural-based manner. If enough data are presented to a model, it can aid the search for a drug or vaccine candidate by identifying patterns within the data. In this review, we focus on the recent advances of COVID-19 drug and vaccine development using artificial intelligence and the potential of intelligent training for the discovery of COVID-19 therapeutics. To facilitate applications of deep learning for SARS-COV-2, we highlight multiple molecular targets of COVID-19, inhibition of which may increase patient survival. Moreover, we present CoronaDB-AI, a dataset of compounds, peptides, and epitopes discovered either *in silico* or *in vitro* that can be potentially used for training models in order to extract COVID-19 treatment. The information and datasets provided in this review can be used to train deep learning-based models and accelerate the discovery of effective viral therapies.

Keywords: COVID-19, SARS-COV-2, drug, vaccine, artificial intelligence, deep learning

INTRODUCTION

Coronaviridae is a viral family responsible for causing pneumonia-like symptoms that has been a global threat since its first outbreak in 2002 (Jabeer Khan et al., 2020). Severe Acute Respiratory Disease (SARS) and Middle Eastern Respiratory Syndrome (MERS), emerging in 2002 and 2013, respectively, caused diseases marked by both gastrointestinal and pulmonary dysfunction (Hilgenfeld and Peiris, 2013). In 2019, SARS-COV-2 was the causative agent of a third Coronavirus outbreak and has been identified as the virus responsible for COVID-19, the symptoms of which range from those of the common cold to more severe respiratory failure (Kong W.-H. et al., 2020). Despite its having been declared a pandemic by the World Health Organization (WHO), COVID-19 has continued to spread and has infected at least 20 million individuals, reaching a death toll of over half a million at the time of this review (Worldometer, 2020).

While hospitals are resorting to trial and error tactics for COVID-19 drug discovery, Virtual Screening (VS) has emerged as a popular method for discovering potent compounds due to the inefficiency of lab-based high throughput screening (HTS) (Jin et al., 2020; Kandeel and Al-Nazawi, 2020). VS for rational drug discovery is essentially an approach that involves computationally targeting a specific biomolecule (e.g., DNA, protein, RNA, lipid) of a cell to inhibit its growth and/or activation (Shoichet, 2004; Lionta et al., 2014). Structure-based and ligand-based drug discovery and design are two important subgroups of this type of screening (Lionta et al., 2014; Yu and Mackerell, 2017; Arshadi et al., 2020; Broom et al., 2020). Given our access to computationally and experimentally determined viral protein structures (Senior et al., 2020; Zhang L. et al., 2020), VS provides a rapid and cost-effective strategy for identifying antiviral candidates.

Additionally, conventional vaccine discovery methods have been costly, and it may take many years to develop an appropriate vaccine against a specified pathogen. In the early 1990s, the introduction of a genome-based vaccine design approach dubbed "Reverse Vaccinology" (RV) (Rappuoli, 2000; Bullock et al., 2020), revolutionized the field to a more efficient status, due in part to the fact that bacterial culturing was no longer required for identifying vaccine targets (Bruno et al., 2015; Heinson et al., 2015; Soria-Guerra et al., 2015). Moreover, all of the putative target protein antigens can be identified, rather than identification being limited to those isolated from bacterial cultures (Xiang and He, 2009; Bowman et al., 2011). All of these advantages taken together led scientists to generate RV prediction programs.

Over the past decade, artificial intelligence (AI)-based models have revolutionized drug discovery in general (Zhong et al., 2018; Duan et al., 2019; Lavecchia, 2019). AI has also led to the creation of many RV virtual frameworks, which are generally classified as rule-based filtering models (Naz et al., 2019; Ong et al., 2020a). Machine learning (ML) enables the creation of models that learn and generalize the patterns within the available data and can make inferences from previously unseen data. With the advent of deep learning (DL), the learning procedure can also include automatic feature extraction from raw data (Lecun et al., 2015). Moreover, it has recently been found that deep learning's feature extraction can result in superior performance compared to other computer-aided models (Ma et al., 2015; Chen et al., 2018; Zhavoronkov et al., 2019).

In this review, we provide a survey of AI-based models for COVID-19 drug discovery and vaccine development. Moreover, we identify and evaluate the best candidate targets for future treatment development. We propose that a concerted effort should be made to leverage the knowledge from pre-existing data by using machine learning approaches. To that end, we present a wide-ranging collection of small molecules, peptides, and epitopes for therapy discovery that could also direct AI-based models, screening, or generation, in an intelligent manner.

BACKGROUND OF MACHINE LEARNING METHODS FOR THERAPY DISCOVERY

In recent years, machine learning has revolutionized many fields of science and engineering. It has largely transformed our daily lives, from speech and face recognition (Alaghband et al., 2020; Grover and Toghi, 2020; Sun et al., 2020) to customized targeted advertisements (Zhai et al., 2016). The power of automatic abstract feature learning, combined with a massive volume of data, has immensely contributed to the successful application of ML (Lecun et al., 2015). Two of the most impactful areas affected are drug and vaccine discovery (Chen et al., 2018), in which ML has offered compound property prediction (Ma et al., 2015), activity prediction (Zhavoronkov et al., 2019), reaction prediction (Fooshee et al., 2018), and ligand–protein interaction.

On the prediction front, Graph Convolutional Neural Networks (GCNN) have been the favorite tool for drug discovery applications (Duvenaud et al., 2015; Kearnes et al., 2016). These networks are able to handle graphs and extract features via encoding the adjacency information within the features. Successful representation learning from molecules using GCNNs has been demonstrated in drug property prediction (Heskett et al., 2018; Bazgir et al., 2019; Liu et al., 2019), protein interface estimation (Fout et al., 2017), reactivity prediction (Coley et al., 2019), and drug–target interactions (Torng and Altman, 2019; Wang et al., 2020). Sequence-based models such as genomics, proteomics, and transcriptomics have also gained some attention in recent years due to the advancements made in the natural language processing domain. The more recent generation of context-based models are transformers that use attention mechanisms and self-supervision to extract representations from sequences (Vaswani et al., 2017; Devlin et al., 2018). Transformers have demonstrated the capacity to predict drug–target interactions (Shin et al., 2019), model protein sequences (Choromanski et al., 2020), and predict retrosynthetic reactions. These models learn to extract features from sequences on the location, context, and order of the input tokens (Belinkov and Glass, 2018). Recurrent neural networks (RNNs) and long short-term memory (LSTM) networks have successfully demonstrated the ability to perform when trained on molecules or protein sequences to predict secondary structure (Pollastri

et al., 2002), quantitative structure–activity relationship (QSAR) modeling (Chakravarti et al., 2019), and function prediction (Liu, 2017).

On the lead generation front, *de novo* design has benefitted the most from the application of deep learning. This subfield has drastically evolved from its traditional usage of ligand-based models and creating molecules from sub-blocks (Acharya et al., 2010). The current approach involves the use of state-of-the-art deep learning models such as Generative Adversarial Networks (GANs) to create data-oriented molecules (Guimaraes et al., 2017). Traditional *de novo* design fails to fully implement this exploration by constraining the generation of molecules with ligand or fragment libraries. More recent approaches utilize deep learning generative models such as variational autoencoders (VAE) (De Cao and Kipf, 2018) in order to create sequences of atoms. This approach lifts the constraints of ligand-based designs and allows the generation of unique molecules with greater diversity (Guimaraes et al., 2017; De Cao and Kipf, 2018; Jin et al., 2018; Liu et al., 2018; Simonovsky and Komodakis, 2018).

Machine learning has also improved the field of vaccine design over the past two decades. VaxiJen was the first implementation of ML in RV approaches and has shown promising results for antigen prediction (Doytchinova and Flower, 2007; Heinson et al., 2017). In addition, the recent development of Vaxign-ML, a web-based RV program leveraging machine learning approaches for bacterial antigen prediction, is a testament to the success of exercising mathematical ML-based in RV (He et al., 2010a; Heinson et al., 2017). In essence, these pipelines consist of feature extraction, feature selection, data augmentation, and cross-validation implemented to predict vaccine candidates against various bacterial and viral pathogens known to cause infectious disease. The use of biological, structural, and physiochemical features is prevalent among the approaches in this domain, as seen in reverse vaccinology and immunoinformatic methods such as IEDB and BlastP, which are feature extractors for AI-based models like RNN in the study of different pathogenic viruses (Flower et al., 2010; He and Zhu, 2015; Abbasi, 2020). More recently, graph-based features have also shown the ability to represent the antibodies instead of an expert-designed feature; Magar et al. showed that graph featurization is followed by mean pooling, and then classification is implemented using shallow and deep models (Magar et al., 2020). Deep Learning approaches have also revolutionized the field of cancer vaccinology through the improved prediction of neoantigens and their HLA binding affinity (Sher et al., 2017; Tran et al., 2019; Wu et al., 2019). Autoencoders of deep learning have shown promising improvement in extracting characteristics of human Leukocyte Antigen (HLA-A), which could be utilized in both transplantations and vaccine discovery (Miyake et al., 2018).

Key aspects of therapy discovery are safety and reliability. The Vaccine Adverse Event Reporting System (VAERS) and Vaccine Safety Databank (VSD) have been among the most popular immunization registries for tracking, recording, and predicting vaccine safety. In prior decades, implementations of computational simulation and mathematical modeling have significantly improved the tradeoff between the assessment of safety and efficacy by using the aforementioned resources (He

et al., 2010b; Vaishnav et al., 2015). Zheng et al. implemented Natural language Processing (NLP) for the identification of adverse events related to Tdap vaccines (Zheng et al., 2019).

In drug development cases, the final drug candidate produced in the process of drug discovery needs to be safe for human consumption. This requires an observation of the drug's side effects as well as confirmation that the drug is non-toxic. To accomplish this, the Toxicology in the 21st Century program (Tox-21) has screened ∼10,000 compounds from 70 screening assays, creating a database that can be used to facilitate toxicity modeling. Furthermore, the project has also expanded to contain 700 assays with nearly 1,800 molecules in the ToxCast dataset. On the side-effect prevention front, the off-target interactions are predicted and minimized *in silico*. In doing so, potential drug candidates are chosen, with consideration given to their off-target polypharmacological profiles (Zhou H. et al., 2015). In a different approach, AI-based studies were implemented to detect the potential prolongation of QT intervals and cardiotoxicity of a candidate drug, hydroxychloroquine, using ECG data from smartwatches (Li J. et al., 2020)[1].

In summary, artificial intelligence has been applied to many subfields of drug discovery and vaccine development. This improvement is crucial for the current situation and immediate SARS-COV-2 therapy discovery for several key reasons. Firstly, the automatic feature extraction ability of deep learning can support models with better accuracy and deliver more reliable results. Secondly, the generative ability demonstrated by deep learning models can be utilized to create more druggable molecules and better epitope prediction, lowering the chance of failure in the trial pipeline. Lastly, the novelty of the virus causes the data around its possible therapies to be scarce, which is a suitable scenario for transfer learning and leveraging the learned knowledge from previous tasks (e.g., TranscreenTM) (Salem et al., 2020). Transfer learning has been shown to alleviate this problem through the transferring of learned knowledge and parameters from a secondary task with big data available to the task at hand (Weiss et al., 2016). Therefore, the use of deep learning in therapy discovery for SARS-COV-2 is essential in order to make a timely and accurate response to the virus.

COVID-19 MOLECULAR MECHANISM AND TARGET SELECTION

Coronaviruses are enveloped viruses with a positive-sense single-stranded RNA genome (Fehr and Perlman, 2015). They are known to infect both humans and other eukaryotes (Andersen et al., 2020; Hoffmann et al., 2020). The novel coronavirus manages to bind to the host receptor with a higher affinity than SARS due to the increased modification of its viral spike, among other structural proteins, resulting in enhanced transmission (Zhou Y. et al., 2020).

---

[1]AI study launched to monitor cardiac safety of COVID-19 patients receiving hydroxychloroquine. Available online at: https://cardiacrhythmnews.com/ai-study-launched-to-monitor-cardiac-safety-of-covid-19-patients-receiving-hydroxychloroquine/ (accessed July 04, 2020).

SARS-CoV-2 interaction with host cells begins with attachment via the viral spike (S) protein to the host ACE2 receptor (Hoffmann et al., 2020; Zhou P. et al., 2020). ACE2 binding induces the host surface serine protease, TMPRSS2, to prime the S protein via cleavage at its S1/S2 border, facilitating viral fusion with the cell membrane (Hoffmann et al., 2020). Once inside the cell, the viral RNA genome is released into the cytosol, where it is translated by host ribosome machinery, producing two polyproteins: pp1a and pp1ab, which are then cleaved by viral 3CL protease (main protease) and PL protease. This gives rise to several non-structural proteins (nsps) as the foundation of RNA-dependent RNA polymerase (RdRP); this RdRP then transcribes a template strand of the genomic RNA, from which it then transcribes subgenomic mRNA products to be translated. These products encode the structural proteins S, E, M, and N, as well as additional accessory nsps (**Figure 1**) (Lai and Cavanagh, 1997; Kim D. et al., 2020).

The severity of the host response depends on an innate response to viral recognition, involving the expression of type-1 IFNs and pro-inflammatory cytokines (Pazhouhandeh et al., 2018; Prompetchara et al., 2020). If the antiviral response is delayed or inhibited, viral proliferation can lead to the large-scale recruitment of neutrophils and monocyte-macrophages to the lungs, creating a hyperinflammatory environment (Prompetchara et al., 2020). Overactive release of pro-inflammatory cytokines, i.e., cytokine storm (CS), has been found in COVID-19 patients and can lead to severe complications like acute respiratory distress syndrome (ARDS) (Moore and June, 2020). It has been found that levels of IL-1B, IL-1RA, IL-8, IL-10, IFNγ, IP10, MCP1, and MIP1s are higher in COVID-19 patients than in healthy adults (Huang et al., 2020). IL-6, in particular, has been highly implicated in CRS and COVID-19 severity, and inhibition of IL-6/IL-6R activity may lead to improved patient outcome, increasing its desirability as a target (**Figure 1**) (Scheller et al., 2014; Tanaka et al., 2016; Zhang C. et al., 2020).

Throughout the process of viral entry, replication, and dissemination, there are several proteins that can serve as suitable targets for therapeutic intervention. The S protein is one of the candidates receiving the most focus, as it is necessary for viral entry into host cells and is highly specific to the virus itself. The host receptor ACE2 is another possible target, but the presence of ACE2 in non-lung tissues such as heart, kidney, and intestine (Hamming et al., 2004) could complicate its inhibition. Another host protein, the TMPRSS2 protease, is essential for viral entry into the cell, making it an additional viable target (Hoffmann et al., 2020).

COVID-19 DRUG DISCOVERY

Protein-Based

The recent applications of Artificial Intelligence for COVID-19 include the virtual screening of both repurposed drug candidates and new chemical entities. For repurposed drugs, the goal has been to rapidly predict and exploit interconnected biological pathways or the off-target biology of existing medicines that are proven safe and can thus be readily tested in new clinical trials.

In one of the early attempts, Gordon et al. paved the way for the repurposing of candidate drugs by experimentally identifying 66 human proteins linked with 26 SARS-CoV-2 proteins (Gordon et al., 2020). In addition to wet-lab approaches, network-based model simulation has been the main computational approach for analyzing the virus–host interactome (Messina et al., 2020). Li et al. identified 30 drugs for repurposing by analyzing the genome sequence of three main viral family members of the coronavirus and then relating them to the human disease-based pathways (Li X. et al., 2020). In a different approach, Zhou et al. offered a combination of network-based methodologies for repurposed drug combination (Zhou Y. et al., 2020).

UK-based BenevolentAI leveraged its AI-derived knowledge graph, which integrates biomedical data from structured and unstructured sources (Richardson et al., 2020). It targeted the inhibition of host protein AAK1 and identified Baricitinib, an approved drug for the treatment of rheumatoid arthritis (Stebbing et al., 2020). Similarly, Beck et al. published an application of their DL-based drug–target interaction model that predicted commercially available antiviral drugs that may target the SARS-COV-2-related protease and helicase (Beck et al., 2020a). Atomwise has also focused on targeting several SARS-CoV-2 protein binding sites that are highly conserved across multiple coronavirus species in an effort to develop new broad-spectrum antivirals. Using its *AtomNet*® deep convolutional neural network technology (Wallach et al., 2020), Atomwise is screening millions of virtual compounds against these diverse targets alongside 15 different partnerships with academic researchers that will test the predicted compounds in their *in vitro* assays[2].

There have been several other applications of multi-task deep learning models for identifying existing drugs that can target the main viral proteins, especially the main protease (3CL^{pro}) and spike protein (Hu et al., 2020; Kadioglu et al., 2020; Kim J. et al., 2020; Redka et al., 2020). One impressive example is Cyclica's creation and mining of PolypharmDB, a platform of known drugs and their predicted binding to human protein targets that uncovered off-target applications of 30 existing drugs against the viral protein 3CL^{pro} and the ACE2 binding site as two examples (Redka et al., 2020). At least two other applications of DL-based virtual screening for the SARS-CoV-2 main protease have been published and include the open sharing of newly predicted chemical structures (Bung et al., 2020; Zhang H. et al., 2020).

ML-aided molecular docking has been one of the most prevalent approaches for virtual screening. This process normally requires the following: (1) Dataset of Druglike or Approved Molecules, (2) Crystal Structure or Homology Model of the target, (3) Molecular Docking Program, and (4) Compute Resources (Ewing et al., 2001; Pagadala et al., 2017). Through docking, many molecules have been reported to fit the binding site of various SARS-CoV-2 proteins essential for viral replication

---

[2]Atomwise Partners with Global Research Teams to Pursue Broad-Spectrum Treatments Against COVID-19 and Future Coronavirus Outbreaks | Business Wire. Available online at: https://www.businesswire.com/news/home/20200521005238/en/Atomwise-Partners-Global-Research-Teams-Pursue-Broad-Spectrum (accessed June 28, 2020).
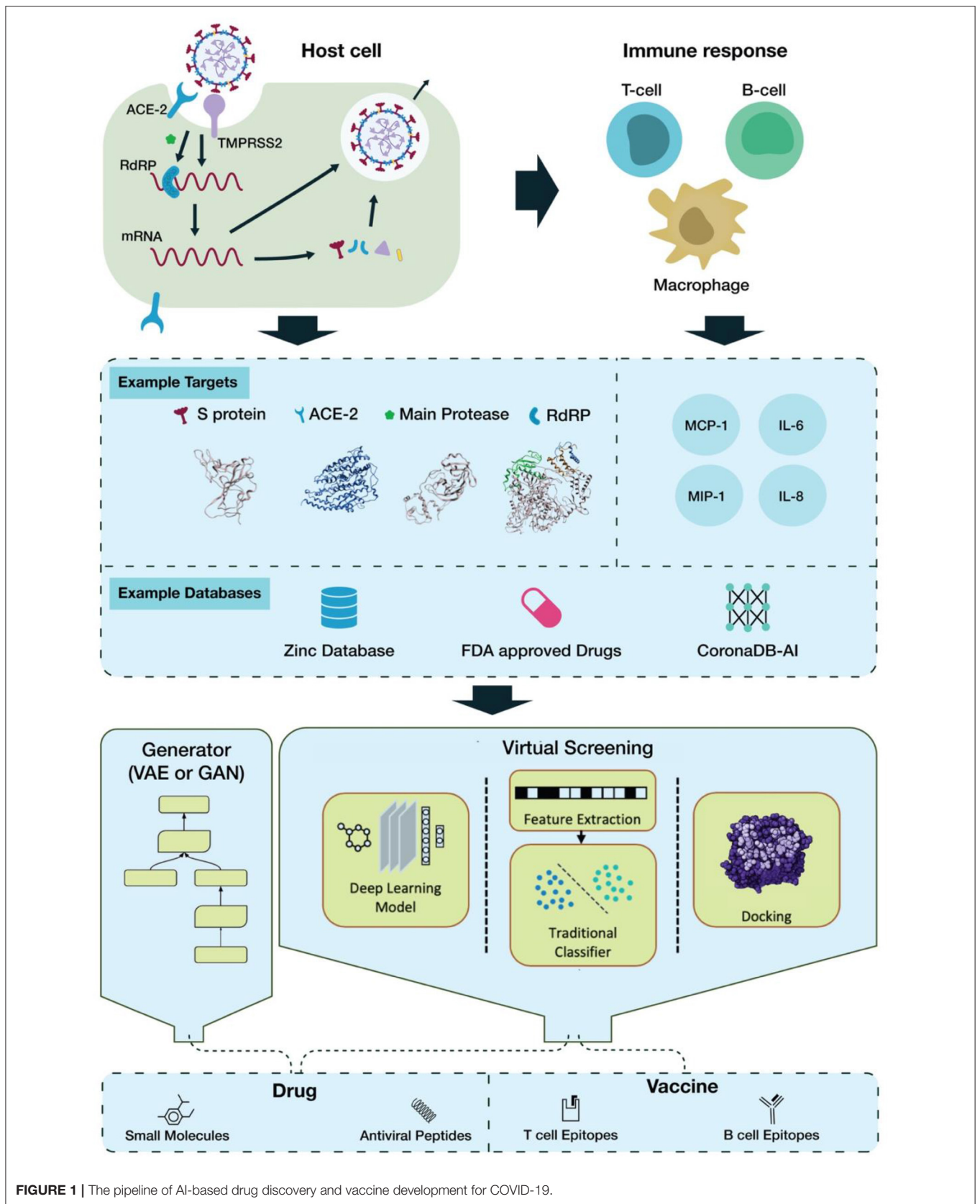
FIGURE 1 | The pipeline of AI-based drug discovery and vaccine development for COVID-19.

and infection. 3CL[pro], Spike Protein, RdRP, and PL[pro] are among those screened, as well as the host ACE2 receptor and TMPRSS2 protease (Chen et al., 2020; Choudhary et al., 2020; Kong R. et al., 2020; Smith and Smith, 2020; Wu et al., 2020). As an example, Ton et al. identified at least 1000 protease inhibitors by creating and utilizing the Deep Docking (DD) network technology approach. However, as they used the QSAR for training their model, no novel docking score was provided (Ton et al., 2020).

It is clear that 3CL[pro] is the most popular target for virtual screening (**Figure 1**). The main reason for this is its pivotal role in viral replication and transcription and its well-defined structural information. Viral protease inhibitors have been extensively studied as treatments for other viruses. In addition, deep learning-aided approaches have been the main focus of research, as their automatic feature extraction accelerates discovery. The datasets cited often rely on the ZINC database (Wu et al., 2020), while other screened datasets include the FDA-approved LOPAC library (Choudhary et al., 2020), SWEETLEAD library (Smith and Smith, 2020), or all purchasable drugs (Drugs-lib) (Chen et al., 2020). Moreover this review sampled a variety of publications witch used different computational resources. It can be carried out on a small scale on a MacOS Mojave Workstation with an 8 core Zeon E5 processor or on a large scale as with the world's strongest supercomputer, SUMMIT, for enhanced parallelization (Choudhary et al., 2020; Smith and Smith, 2020).

RNA-Based

Conserved structured elements have already been shown to play critical functional roles in the life cycles of Coronaviruses (Yang and Leibowitz, 2015). Through direct interactions with host RNA-binding proteins and helicases, structural elements add a layer of complexity to the regulatory information that is encoded in the viral RNA. Targeted disruption of the regulatory functions of these structural elements provides a largely unexplored strategy that can limit viral loads with minimal impact on the biology of normal cells (Park et al., 2011). While this idea would have been farfetched a mere 5 years ago, advances in AI-driven computational modeling and high-throughput experimental RNA shape analyses have all but overcome the critical barriers (Alipanahi et al., 2015).

Highly conserved RNA structural elements have been identified in a number of viral families, many of which have been functionally validated (Jaafar and Kieft, 2019). Some of these stem loops in SARS-CoV-2′s 5′UTRs structural elements are conserved across beta coronaviruses and are known to impact viral replication (Yang and Leibowitz, 2015). There are many functional RNA structural elements that fall within the coding sequence and the 3′UTR as well (Plant and Dinman, 2008; Stammler et al., 2011). Rangan et al. identified 106 structurally conserved regions that would be suitable biotargets for unexplored antiviral agents (Rangan et al., 2020). Moreover, they predicted at least 59 unstructured regions that are conserved within SARS-CoV-2. Park et al. identified an RNA Pseudoknot-Binding molecule against SARS-CoV-1 in target-based virtual screening (Park et al., 2011; Nakagawa et al., 2016).

Studying the changes in RNA information also allows for the identification of new and evolved targets. In a different approach, Wu et al. showed that a recently FDA-approved drug named Remdesivir could bind to the RNA-binding channel of the novel coronavirus. They discovered other candidate drugs via analyzing the proteins critical to RNA processing and pathways (Wu et al., 2020). It seems that viral genome, RdRP, and processed mRNA would make promising targets for drug repurposing.

Generative Approaches

Molecule generation has been one of the fields of drug discovery that have been most revolutionized by the implementation of artificial intelligence over the last decade. As mentioned, VAE is a generator model for enhancing the diversity of generated data. Autoencoders instruct molecules into a vector that captures properties such as bond order, element, and functional group (Bjerrum and Sattarov, 2018). Chenthamarakshan et al., together with IBM Research, demonstrated a VAE that captures molecules in a latent space. Once captured, variations are made on the original molecule vectors based on desired properties. These can then be decoded back into novel molecules (Chenthamarakshan et al., 2020). To optimize the structures, QED, Synthetic Accessibility, and LogP regressors were used to improve the latent space variations.

In a different approach, Tang et al. overcame many of the issues with traditional generative models by developing a novel advanced deep Q-learning network with fragment-based drug design (ADQN-FBDD). This allowed for the enhanced exploration of space by assembling SARS-CoV-2 molecules one fragment at a time rather than relying on latent space adjustments. After making connections and rewarding molecules with the most druglike connections, a pharmacophore and descriptor filter was used to refine the set. They demonstrated a robust method for designing novel, high-binding compounds refined to the structure of SARS-CoV-2 3CL[pro] (Tang et al., 2020). To design a drug-generative network, the following is necessary: (1) collection of Druglike Molecules, (2) a representation of these molecules *in silico* (i.e., Fingerprints, Tokenizers), (3) a method of altering molecules to increase diversity, and (4) screening and modification of the altered molecules. Pursuing GAN-related models, Insilico Medicine used three of its previously validated generative chemistry approaches to target the main protease, namely, crystal-derived pocked-based generation, homology modeling-based generation, and ligand-based generation (Zhavoronkov et al., 2020). Similar to target-based virtual screening, the main protease has been the main object of interest for scientists for *de novo* drug discovery.

COVID-19 VACCINE DISCOVERY

Identification of the best possible targets for the development of a vaccine is crucial in order to counteract a virus's high infection rate (Choudhary et al., 2020). A host immune system fights virus-infected cells either through the production of antibodies by B cells or through the direct attack of T cells (Amanat and Krammer, 2020). The HLA gene encodes MCH-I and MCH-II

proteins, which present epitopes as antigenic determinants. These proteins assist B-cell and T-cell antibodies in their ability to bind and attack invaders (Dangi et al., 2018; Gupta et al., 2020; Smith and Smith, 2020). Machine learning approaches, including Random Forest (RF), Support Vector Machine (SVM), and Recursive Feature Selection (RFE), have been basic tools for identifying antigens from protein sequences (Bowick et al., 2010; Rahman et al., 2019). However, due to their low sensitivity in the prediction of locally clustered interactions in some cases, Deep Convolutional Neural Networks (DCNN) have been a more valid alternative for the binding prediction of MHC and peptides (Han and Kim, 2017).

Since the outbreak of this first coronavirus, different AI-based approaches have been used to predict potential epitopes so as to design vaccines (Park et al., 2011; Yang and Leibowitz, 2015; Ton et al., 2020). Fast and Chen used MARIA (Chen et al., 2019) and NetMHCPan4 (Jurtz et al., 2017), two supervised neural network-driven tools, to discover potential T-cell epitopes for SARS-CoV-2 close to the 2019-nCoV spike receptor-binding domain (RBD) (Fast and Chen, 2020). The Long Short-Term Memory (LSTM) network has also shown some promising results. Abbasi et al. used this type of RNN to predict epitopes for Spike (Abbasi, 2020). Using a similar tactic, Crossman et al. employed deep-learning RNN and provided simulated sequences of Spike to identify possible targets for vaccine design (Crossman, 2020). RNN provided the sequences for a protein of interest with high sequence identity to the BLAST match.

Using a separate method, Feng et al. leveraged the iNeo tool to design a vaccine containing both B-cell and T-cell epitopes. This multi-peptide vaccine could provide a new strategy against SARS-CoV-2. Additionally, they discovered 17 vaccine peptides involving both immune cells (Nakagawa et al., 2016; Rangan et al., 2020). Ong et al. used Vaxign-RV to prioritize non-structural proteins as vaccine candidates for SARS-CoV-2 (Ong et al., 2020b). Nsp3, the largest non-structural protein of the coronavirus family, was identified as the most promising potential target for vaccine development after Spike (Ong et al., 2020b). Malone et al. also studied the entire SARS-CoV-2 proteome beyond Spike and provided a comprehensive vaccine design blueprint for SARS-CoV-2 using NEC Immune Profiler, IEDB, and BepiPred tools to create an epitope map for different HLA alleles (Malone et al., 2020).

Natural language processing models, specifically language modeling techniques, have also made an impact in the domain of COVID-19 vaccine discovery. Pre-trained transformers were used to predict protein interaction (Nambiar et al., 2020) and model molecular reactions in carbohydrate chemistry (Pesciullesi et al., 2020), which can be utilized in the process of vaccine development. Chen et al. discussed the use-case of an LSTM-based seq-2-seq model for predicting the secondary structure of certain SARS-COV-2 proteins (Karpov et al., 2019)[3]. Also, Beck et al. used transformers to repurpose commercially available drugs by predicting their interactions with viral proteins of SARS-COV-2 (Beck et al., 2020b).

---

[3]OSF Preprints. ZeroFold-Understanding Mutations of SARS-CoV-2 Spike Protein base on Secondary Structure Event Extracting for guiding Vaccine development. Available online at: https://osf.io/3vkuw/ (accessed Jul. 01, 2020).

Taking this work together, it is clear that spike protein has been the most popular candidate for virtual vaccine discovery (Oany et al., 2014). As the spike protein of SARS-COV-2 is crucial for viral entry, specific neutralizing antibodies against the receptor-binding domain of Spike can interrupt the attachment and fusion of viral proteins (Wan et al., 2019). This method could provide simulated sequences that can serve as a guide for further vaccine discovery against COVID-19 and possibly new zoonosis that may arise in the future.

DATA COLLECTION

Data-driven solutions rely on patterns embedded in the data in order to extract mathematical models. That being said, a data collection campaign will face a plethora of challenges in the case of any recently emerged virus, primarily due to the existence of bias and imbalance in the limited data available. Therefore, even the most sophisticated of modeling approaches will be ineffective when trained on such datasets. In order to overcome this issue, we compiled a multifaceted and comprehensive investigation of the existing literature, datasets, and online resources to provide potential small molecules, peptides, and epitopes. Such elements can be beneficial in the process of discovering or designing novel drugs to treat COVID-19 when used with both conventional and data-driven AI-based approaches.

We choose to focus on both potential antiviral agents and host biotarget inhibitors. The provided data entitled CoronaDB-AI in **Table 1** includes the small molecules and peptides proposed by both in-silico and in-vitro approaches. In addition to candidate scaffolds against the coronavirus's structural proteins, the potential inhibition of other respiratory tract viruses is taken into consideration to increase the therapeutic potential. Antimicrobial peptides have been validated as potent antivirals that disrupt either the viral membrane or an additional molecular mechanism of the virus (Akaji et al., 2011; Han and Kraí, 2020; Xia et al., 2020). As described before, the cytokine storm and an elevated immune response of the host plays a vital role in disease complication, so candidate immunosuppressants were also added as host-targeted agents. In addition to the potency of a candidate drug, it is crucial that the drug have high selectivity and low toxicity. Therefore, we also gathered a complete toxicity dataset from distinct databases, including ToxCast and Tox21. Finally, we gathered a comprehensive epitope-based dataset that could also guide deep learning-based models for improved vaccine development and epitope generation.

DISCUSSION

SARS-COV-2 rapidly transformed into a global challenge, costing thousands of lives, overwhelming healthcare systems, and threatening the economy all around the world. As we demonstrated above, it can be extremely challenging to experimentally perform a comprehensive potency evaluation of all drug and vaccine candidates in a timely fashion. We believe that leveraging computational models capable of filtering and generating reliable therapies can significantly speed up these discovery efforts. Employing artificial neural networks

TABLE 1 | CoronaDB-AI is a collection of small molecules, peptides, and epitopes for the purpose of COVID-19 therapy discovery.

Data provided	Discovery	Type	Mechanism of action	References
ANTIVIRAL DATA				
Total of 59,107		**Small molecules and peptides**		
50,000	*In-silico*	Small molecule	Antiviral	1
3,000	*In-silico*	Small molecule	Anti SARS2 protein	Chenthamarakshan et al., 2020
1,000	*In-silico*	Small molecule	Anti-protease	Ton et al., 2020
406	*In-vitro*	Small molecule	Inhibiting autophagy	2
802	*In-vitro*	Small molecule	Activating autophagy	2
393	*In-vitro*	Small molecule	Biotargets of coronaviruses	3
110	*In-vitro*	Peptide and small molecule	Coronavirus and respiratory disease	Pillaiyar et al., 2020
1,000	*In-silico*	Small molecule	3C protease inhibitor	Zhavoronkov et al., 2020
11	*In-silico*	Small molecule	Main protease inhibitor	Fischer et al., 2020
20	*In-vitro*	Antimicrobial peptide	Anti-SARS/MERS	Mustafa et al., 2018
7	*In-silico*	Antimicrobial peptide	Anti-MERS	Mustafa et al., 2019
277	*In-vitro*	Antimicrobial peptide	Antiviral	Wang et al., 2015
4	*In-silico*	Antimicrobial peptide	Anti-spike of sars-Cov-2	Han and Kraí, 2020
379	*In-vitro*	Small molecule	Anti-respiratory syncytial virus	Plant et al., 2015
13	*In-vitro*	Small molecule	Anti-recurrent respiratory papillomatosis by HPV-6	Alkhilaiwi et al., 2019
1,280	*In-vitro*	Small molecule	Anti-respiratory syncytial virus	Rasmussen et al., 2011
16	*In-silico*	Small molecules	Anti-SARS-COV-2	Zhou Y. et al., 2020
77	*In-silico*	Small molecules	Anti-S Protein of SARS-COV-2	Smith and Smith, 2020
10	*In-silico*	Small molecules	Anti-SARS-COV2	Hu et al., 2020
25	*In-silico*	Small molecules	Anti SARS2 Proteins	Kim J. et al., 2020
10	*In-silico*	Small molecules	ACE2 and Spike inhibitors	Choudhary et al., 2020
78	*In-silico*	Small molecules	All SARS2 proteins	Wu et al., 2020
47	*In-silico*	Small molecules	3cl protease and M pro	Tang et al., 2020
16	*In-silico*	Small molecules	3cl protease inhibitor	Chen et al., 2020
36	*In-vitro*	Small molecules	Anti- Coronavirus-OC43	Shen et al., 2019
90	*In-vitro*	Small molecules	Anti- SARS-COV-2	Touret et al., 2020
ANTI-HOST PROTEINS				
Total of 677		**Small molecules and peptides**		
6	*In-vitro*	Small molecules	Anti-IL-1β and TNFα	Laufer et al., 2002
182	*In-vitro*	Peptides	Cytokine Signaling Inhibitors	4
269	*In-silico*	Small molecules	Anti-IL-6	Shukla et al., 2019
121	*In-vitro*	Small molecules	Severe acute respiratory	5
69	*In-silico*	Small molecules	Anti-protein-protein interaction of virus-host	Gordon et al., 2020
30	*In-silico*	Small molecules	Anti-host & virus interaction	Redka et al., 2020
TOXICITY DATA				
Total of 25,333		**Small molecules**		
11,800	*In-vitro*	Small molecules	Tox21 and ToxCast	Toxicology, EPA's National Center for Computational, 2018
13,533	*In-vitro*	Small molecules	Toxic for HepG2 Cell Line	Gamo et al., 2010
VACCINE DATA				
Total of 517		**Epitopes and vaccines**		
162	*In-silico*	Epitopes	Anti-SARS-COV-2	Ahmed et al., 2020
174	*In-silico*	Epitope	Anti-SARS-COV-2	Prachar et al., 2020
2	*In-silico*	Epitope	Anti-SARS-COV-2	Fast and Chen, 2020
30	*In-silico*	Vaccine candidate	Anti-SARS-COV-2	Feng et al., 2020
7	*In-silico*	Epitope	Anti-SARS-COV-2	Lon et al., 2020
12	*In-silico*	Epitope	Anti-SARS-COV-2	Tilocca et al., 2020
59	*In-silico*	Epitope	Anti-SARS-COV-2	Sarkar et al., 2020
71	*In-silico*	Epitope	Anti-SARS-COV-2	Bhattacharya et al., 2020

[1]*Download CAS COVID-19 Antiviral Candidate Compounds Dataset | CAS. Available online at: https://www.cas.org/covid-19-antiviral-compounds-dataset (accessed April 27, 2020).*
[2]*Novel Coronavirus Information Center. Available online at: https://www.elsevier.com/connect/coronavirus-information-center (accessed April 27, 2020).*
[3]*https://www.elsevier.com/__data/assets/pdf_file/0004/978745/Copy-of-RMC-substances-coronovirus-targets-pX6.pdf (accessed April 27, 2020).*
[4]*Cytokines Inhibitor library|Targetmol|96-well. Available online at: https://www.targetmol.com/compound-library/Cytokines-inhibitors-Library (accessed April 27, 2020).*
[5]*https://www.elsevier.com/__data/assets/pdf_file/0007/977173/ResNet-Data_Coronavirus.pdf (accessed April 27, 2020).*

and supervised learning methods has proven to be a vital game-changer when used for the purpose of virtual filtering and *de novo* design. However, in order to achieve the desired performance in such intelligent methods, one requires the knowledge to recognize the most relevant biotargets in addition to a large-scale training dataset. This fact motivated us to perform a survey of biotargets that have been employed in the virtual drug and vaccine discovery literature. We observed that the viral spike protein and the main protease have been the most prevalent choices for vaccine development and drug discovery, respectively, due to their importance. Furthermore, we gathered a list of datasets titled "CoronaDB-AI" that can be used for our particular application. Having access to these key elements removes the burden of collecting training data and the required knowledge for both computer scientists and bioinformaticians and consequently enhances research outcomes.

AUTHOR CONTRIBUTIONS

AK organized and wrote most of article and gathered all the data. JW contributed to the molecular part. MS contributed to the background for AI-based methods. EC, ED-C, and BK from A2A and SC-T from Atomwise contributed to the COVID19 drug discovery. NG and JC contributed to the vaccine discovery. HG contributed to the RNA-based and molecular sections. JY provided guidance in the opportunities of deep learning in a multidiscipline collaboration. All authors contributed to the article and approved the submitted version.

ACKNOWLEDGMENTS

REFERENCES

Abbasi, B. A. (2020). Identification_of_vaccine_targets_and_design_of_vaccine_against_SARS. *OSF Preprints*. doi: 10.31219/osf.io/f8zyw

Acharya, C., Coop, A., Polli, J. E., and MacKerell, A. D. (2010). Recent advances in ligand-based drug design: relevance and utility of the conformationally sampled pharmacophore approach. *Curr. Comput. Aided-Drug Des.* 7, 10–22. doi: 10.2174/157340911793743547

Ahmed, S. F., Quadeer, A. A., and McKay, M. R. (2020). Preliminary identification of potential vaccine targets for the COVID-19 Coronavirus (SARS-CoV-2) Based on SARS-CoV immunological studies. *Viruses* 12:253. doi: 10.3390/v12030254

Akaji, K., Konno, H., Mitsui, H., Teruya, K., Shimamoto, Y., Hattori, Y., et al. (2011). Structure-based design, synthesis, and evaluation of peptide-mimetic SARS 3CL protease inhibitors. *J. Med. Chem.* 54, 7962–7973. doi: 10.1021/jm200870n

Alaghband, M., Yousefi, N., and Garibay, I. (2020). FePh: an annotated facial expression dataset for the RWTH-PHOENIX-weather 2014 Dataset. *arXiv: 2003.08759v1*. Available online at: https://arxiv.org/pdf/2003.08759.pdf

Alipanahi, B., Delong, A., Weirauch, M. T., and Frey, B. J. (2015). Predicting the sequence specificities of DNA- and RNA-binding proteins by deep learning. *Nat. Biotechnol.* 33, 831–838. doi: 10.1038/nbt.3300

Alkhilaiwi, F., Paul, S., Zhou, D., Zhang, X., Wang, F., Palechor-Ceron, N., et al. (2019). High-throughput screening identifies candidate drugs for the treatment of recurrent respiratory papillomatosis. *Papillomavirus Res.* 8:100181. doi: 10.1016/j.pvr.2019.100181

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

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

Arshadi, A. K., Salem, M., Collins, J., Yuan, J. S., and Chakrabarti, D. (2020). Deepmalaria: artificial intelligence driven discovery of potent antiplasmodials. *Front. Pharmacol.* 10:1526. doi: 10.3389/fphar.2019.01526

Bazgir, O., Zhang, R., Rahman Dhruba, S., Rahman, R., Ghosh, S., and Pal, R. (2019). REFINED (REpresentation of Features as Images With NEighborhood Dependencies): a novel feature representation for convolutional neural networks. *arXiv [Preprint] arXiv:1912.05687 (2019)*.

Beck, B. R., Shin, B., Choi, Y., Park, S., and Kang, K. (2020a). Predicting commercially available antiviral drugs that may act on the novel coronavirus (2019-nCoV), Wuhan, China through a drug-target interaction deep learning model. *bioRxiv [Preprint]*. doi: 10.1101/2020.01.31.929547

Beck, B. R., Shin, B., Choi, Y., Park, S., and Kang, K. (2020b). Predicting commercially available antiviral drugs that may act on the novel coronavirus (SARS-CoV-2) through a drug-target interaction deep learning model. *Comput. Struct. Biotechnol. J.* 18, 784–790. doi: 10.1016/j.csbj.2020.03.025

Belinkov, Y., and Glass, J. (2018). Analysis methods in neural language processing: a survey. *Trans. Assoc. Comput. Linguist.* 7, 49–72. doi: 10.1162/tacl_a_00254

Bhattacharya, M., Sharma, A. R., Patra, P., Ghosh, P., Sharma, G., Patra, B. C., et al. (2020). Development of epitope-based peptide vaccine against novel coronavirus (2019). (SARS-COV-2): immunoinformatics approach. *J. Med. Virol.* 92, 618–631. doi: 10.1002/jmv.25736

Bjerrum, E. J., and Sattarov, B. (2018). Improving chemical autoencoder latent space and molecular de novo generation diversity with heteroencoders. *Biomolecules* 8:131. doi: 10.3390/biom8040131

Bowick, G. C., and Barrett, A. D., T. (2010). Comparative pathogenesis and systems biology for biodefense virus vaccine development. *J. Biomed. Biotechnol.* (2010) 2010:236528. doi: 10.1155/2010/236528

Bowman, B. N., McAdam, P. R., Vivona, S., Zhang, J. X., Luong, T., Belew, R. K., et al. (2011). Improving reverse vaccinology with a machine learning approach. *Vaccine* 29, 8156–8164. doi: 10.1016/j.vaccine.2011.07.142

Broom, A., Rakotoharisoa, R. V., Thompson, M. C., Zarifi, N., Nguyen, E., Mukhametzhanov, N., et al. (2020). Evolution of an enzyme conformational ensemble guides design of an efficient biocatalyst. *bioRxiv [Preprint]*. doi: 10.1101/2020.03.19.999235

Bruno, L., Cortese, M., Rappuoli, R., and Merola, M. (2015). Lessons from Reverse Vaccinology for viral vaccine design. *Curr. Opin. Virol.* 11, 89–97. doi: 10.1016/j.coviro.2015.03.001

Bullock, J., Alexandra, L., Pham, K. H., Lam, C. S. N., and Luengo-Oroz, M. (2020). Mapping the landscape of artificial intelligence applications against COVID-19. *arXiv [Preprint] arXiv:2003.11336 (2020)*.

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. *ChemRxiv [Preprint]*. doi: 10.26434/chemrxiv.11998347.v2

Chakravarti, S. K., and Alla, S. R., M. (2019). Descriptor free QSAR modeling using deep learning with long short-term memory neural networks. *Front. Artif. Intell.* 2:17. doi: 10.3389/frai.2019.00017

Chen, B., Khodadoust, M. S., Olsson, N., Wagar, L. E., Fast, E., Liu, C. L., et al. (2019). Predicting HLA class II antigen presentation through integrated deep learning. *Nat. Biotechnol.* 37, 1332–1343. doi: 10.1038/s41587-019-0280-2

Chen, H., Engkvist, O., Wang, Y., Olivecrona, M., and Blaschke, T. (2018). The rise of deep learning in drug discovery. *Drug Discov. Today* 23, 1241–1250. doi: 10.1016/j.drudis.2018.01.039

Chen, Y. W., Yiu, C.-P. B., and Wong, K.-Y. (2020). Prediction of the SARS-CoV-2 (2019-nCoV) 3C-like protease (3CLpro) structure: virtual screening reveals velpatasvir, ledipasvir, and other drug repurposing candidates. *F1000Research* 9:129. doi: 10.12688/f1000research.22457.2

Chenthamarakshan, V., Das, P., Padhi, I., Strobelt, H., Lim, K. W., Hoover, B., et al. (2020). *Target-Specific and Selective Drug Design for COVID-19 Using Deep Generative Models.* Available: http://arxiv.org/abs/2004.01215 (accessed April 19, 2020).

Choromanski, K., Likhosherstov, V., Dohan, D., Song, X., Davis, J., Sarlos, T., et al. (2020). *Masked Language Modeling for Proteins via Linearly Scalable Long-Context Transformers.* Available online at: http://arxiv.org/abs/2006.03555 (accessed July 01, 2020).

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

Coley, C. W., Jin, W., Rogers, L., Jamison, T. F., Jaakkola, T. S., Green, W. H., et al. (2019). A graph-convolutional neural network model for the prediction of chemical reactivity. *Chem. Sci.* 10, 370–377. doi: 10.1039/C8SC04228D

Crossman, L. C. (2020). Leverging deep learning to simulate coronavirus spike proteins has the potential to predict future Zoonotic sequences. *bioRxiv [Preprint].* doi: 10.1101/2020.04.20.046920

Dangi, M., Kumari, R., Singh, B., and Chhillar, A. K. (2018). Advanced *in silico* tools for designing of antigenic epitope as potential vaccine candidates against coronavirus. *Bioinforma. Seq. Struct. Phylogeny.* 329–357. doi: 10.1007/978-981-13-1562-6_15

De Cao, N., and Kipf, T. (2018). MolGAN: An implicit generative model for small molecular graphs. Available online at: http://arxiv.org/abs/1805.11973 (accessed April 26, 2020).

Devlin, J., Chang, W.-M., Lee, K., Google, K. T., and Language, A. I. (2018). BERT: pre-Training of deep bidirectional transformers for language understanding. *arXiv [preprint] arXiv:1810.04805 (2018).*

Doytchinova, I. A., and Flower, D. R. (2007). VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. *BMC Bioinformatics* 8:4. doi: 10.1186/1471-2105-8-4

Duan, Y., Edwards, J. S., and Dwivedi, Y. K. (2019). Artificial intelligence for decision making in the era of Big Data – evolution, challenges and research agenda. *Int. J. Inf. Manage.* 48, 63–71. doi: 10.1016/j.ijinfomgt.2019.01.021

Duvenaud, D., Maclaurin, D., Aguilera-Iparraguirre, J., Gómez-Bombarelli, R., Hirzel, T., Aspuru-Guzik, A., et al. (2015). Convolutional networks on graphs for learning molecular fingerprints. *arXiv:1509.09292.*

Ewing, T. J. A., Makino, S., Skillman, A. G., and Kuntz, I. D. (2001). DOCK 4.0: Search strategies for automated molecular docking of flexible molecule databases. *J. Comput. Aided. Mol. Des.* 15, 411–428. doi: 10.1023/A:1011115820450

Fast, E., and Chen, B. (2020). Potential T-cell and B-cell Epitopes of 2019-nCoV. *bioRxiv [Preprint].* doi: 10.1101/2020.02.19.955484

Fehr, A. R., and Perlman, S. (2015). "Coronaviruses: an overview of their replication and pathogenesis," in *Coronaviruses: Methods and Protocols 1282.* New York, NY: Springer, 1–23. doi: 10.1007/978-1-4939-2438-7_1

Feng, Y., Qiu, M., Zou, S., Li, Y., Luo, K., Chen, R., et al. (2020). Multi-epitope vaccine design using an immunoinformatics approach for 2019 novel coronavirus in China (SARS-CoV-2). *bioRxiv [Preprint].* doi: 10.1101/2020.03.03.962332

Fischer, A., Sellner, M., Neranjan, S., Lill, M. A., and Smieško, M. (2020). Inhibitors for novel coronavirus protease identified by virtual screening of 687 million compounds. *ChemRxiv [Preprint].* doi: 10.26434/chemrxiv.11923239.v1

Flower, D. R., MacDonald, I. K., Ramakrishnan, K., Davies, M. N., and Doytchinova, I. A. (2010). Computer aided selection of candidate vaccine antigens. *Immunome Res.* 6(Suppl. 2), 1–16. doi: 10.1186/1745-7580-6-S2-S1

Fooshee, D., Mood, A., Gutman, E., Tavakoli, M., Urban, G., Liu, F., et al. (2018). Deep learning for chemical reaction prediction. *Mol. Syst. Des. Eng.* 3, 442–452. doi: 10.1039/C7ME00107J

Fout, A., Byrd, J., Shariat, B., and Ben-Hur, A. (2017). "Protein interface prediction using graph convolutional networks," in *Advances in Neural Information Processing Systems* (Long Beach, CA), 6530–6539.

Gamo, F.-J., Sanz, L. M., Vidal, J., de Cozar, C., Alvarez, E., Lavandera, J.-L., et al. (2010). Thousands of chemical starting points for antimalarial lead identification. *Nature* 465, 305–310. doi: 10.1038/nature09107

Gordon, D. E., Jang, G. M., Bouhaddou, M., Xu, J., Obernier, K., O'Meara, M. J., et al. (2020). A SARS-CoV-2-human protein-protein interaction map reveals drug targets and potential drug-repurposing. *bioRxiv [Preprint].* doi: 10.1101/2020.03.22.002386

Grover, D., and Toghi, B. (2020). MNIST dataset classification utilizing k-NN classifier with modified sliding-window metric. *Adv. Intel. Syst. Comp.* 944, 583–591. doi: 10.1007/978-3-030-17798-0_47

Guimaraes, G. L., Sanchez-Lengeling, B., Outeiral, C., Farias, L. C., and Aspuru-Guzik, A. (2017). *Objective-Reinforced Generative Adversarial Networks (ORGAN) for Sequence Generation Models.* Available online at: http://arxiv.org/abs/1705.10843 (accessed April 26, 2020).

Gupta, E., Mishra, R. K., and Niraj, R. R. K. (2020). Identification of potential vaccine candidates against *SARS-CoV-2*, a step forward to fight novel coronavirus 2019-nCoV: a reverse vaccinology approach. *bioRxiv [Preprint].* doi: 10.1101/2020.04.13.039198

Hamming, I., Timens, W., Bulthuis, M. L. C., Lely, A. T., Navis, G. J., and van Goor, H. (2004). Tissue distribution of ACE2 protein, the functional receptor for SARS coronavirus. A first step in understanding SARS pathogenesis. *J. Pathol.* 203, 631–637. doi: 10.1002/path.1570

Han, Y., and Kim, D. (2017). Deep convolutional neural networks for pan-specific peptide-MHC class I binding prediction. *BMC Bioinformat.* 18:585. doi: 10.1186/s12859-017-1997-x

Han, Y., and Kraí, P. (2020) Computational design of ACE2-based peptide inhibitors of SARS-CoV-2. *ACS Nano* 14, 5143–5147. doi: 10.1021/acsnano.0c02857

He, L., and Zhu, J. (2015). Computational tools for epitope vaccine design and evaluation. *Curr. Opin. Virol.* 11, 103–112. doi: 10.1016/j.coviro.2015.03.013

He, Y., Rappuoli, R., De Groot, A. S., and Chen, R. T. (2010b). Emerging vaccine informatics. *J. Biomed. Biotechnol.* doi: 10.1155/2010/218590

He, Y., Xiang, Z., and Mobley, H. L. T. (2010a). Vaxign: the first web-based vaccine design program for reverse vaccinology and applications for vaccine development. *J. Biomed. Biotechnol.* 2010:297505. doi: 10.1155/2010/297505

Heinson, A. I., Gunawardana, Y., Moesker, B., Denman Hume, C. C., Vataga, E., Hall, Y., et al. (2017). Enhancing the biological relevance of machine learning classifiers for reverse vaccinology. *Int. J. Mol. Sci.* 18:312. doi: 10.3390/ijms18020312

Heinson, A. I., Woelk, C. H., and Newell, M. L. (2015). The promise of reverse vaccinology. *Int. Health* 7, 85–89. doi: 10.1093/inthealth/ihv002

Heskett, C., Faircloth, B., Roper, S., and Clay, M. (2018). *Executive Insights Artificial Intelligence in Life Sciences: The Formula for Pharma Success Across the Drug Lifecycle.* Available online at: https://www.lek.com/sites/default/files/insights/pdf-attachments/2060-AI-in-Life-Sciences.pdf (accessed June 18, 2019).

Hilgenfeld, R., and Peiris, M. (2013). From SARS to MERS: 10 years of research on highly pathogenic human coronaviruses. *Antivir. Res.* 100, 286–295. doi: 10.1016/j.antiviral.2013.08.015

Hoffmann, M., Kleine-Weber, H., Schroeder, S., Kruger, N., Herrler, T., Erichsen, S., et al. (2020). SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. *Cell* 181, 271-280.e8. doi: 10.1016/j.cell.2020.02.052

Hu, F., Jiang, J., and Yin, P. (2020). *Prediction of Potential Commercially Inhibitors Against SARS-CoV-2 by Multi-Task Deep Model.* Available online at: https://arxiv.org/ftp/arxiv/papers/2003/2003.00728.pdf (accessed April 22, 2020).

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. doi: 10.1016/S0140-6736(20)30183-5

Jaafar, Z. A., and Kieft, J. S. (2019). Viral RNA structure-based strategies to manipulate translation. *Nat. Rev. Microbiol.* 17, 110–123. doi: 10.1038/s41579-018-0117-x

Jabeer Khan, R., Kumar Jha, R., Muluneh Amera, G., Jain, M., Singh, E., Pathak, A., et al. (2020). Targeting novel coronavirus 2019: a systematic drug repurposing approach to identify promising inhibitors against 3C-like proteinase and 2'-O-ribose methyltransferase: a systematic drug repurposing approach to identify promising inhibitors against 3C-like proteinase and 2'-O-ribose methyltransferase. *ChemRxiv [Preprint].* doi: 10.26434/chemrxiv.11888730.v1

Jin, W., Barzilay, R., and Jaakkola, T. (2018). Junction tree variational autoencoder for molecular graph generation. *arXiv [Preprint]. arXiv:1802.04364.*

Jin, Z., Du, X., Xu, Y., Deng, Y., Liu, M., Zhao, Y., et al. (2020). Structure of Mpro from COVID-19 virus and discovery of its inhibitors. *bioRxiv [Preprint].* doi: 10.1101/2020.02.26.964882

Jurtz, V., Paul, S., Andreatta, M., Marcatili, P., Peters, B., and Nielsen, M. (2017). NetMHCpan-4.0: improved peptide–mhc class i interaction predictions

integrating eluted ligand and peptide binding affinity data. *J. Immunol.* 199, 3360–3368. doi: 10.4049/jimmunol.1700893

Kadioglu, O., Saeed, M., Greten, H. J., and Efferth, Y. (2020). Identification of novel compounds against three targets of SARS CoV2 coronavirus by combined virtual screening and supervised machine learning. *Bull World Heal. Organ.* doi: 10.2471/BLT.20.255943

Kandeel, M., and Al-Nazawi, M. (2020). Virtual screening and repurposing of FDA approved drugs against COVID-19 main protease. *Life Sci.* 251:117627. doi: 10.1016/j.lfs.2020.117627

Karpov, P., Godin, G., and Tetko, I. V. (2019). "A transformer model for retrosynthesis," in *Artificial Neural Networks and Machine Learning – ICANN 2019: Workshop and Special Sessions. ICANN 2019. Lecture Notes in Computer Science*, Vol. 11731, eds I. Tetko, V. Kurková, P. Karpov, and Theis F (Cham: Springer). doi: 10.1007/978-3-030-30493-5_78

Kearnes, S., McCloskey, K., Berndl, M., Pande, V., and Riley, P. (2016). Molecular graph convolutions: moving beyond fingerprints. *J. Comput. Aided. Mol. Des.* 30, 595–608. doi: 10.1007/s10822-016-9938-8

Kim, D., Lee, J.-Y., Yang, J.-S., Kim, J. W., Kim, V. N., and Chang, H. (2020). The architecture of SARS-CoV-2 transcriptome. *Cell.* 181, 914–921. doi: 10.1016/j.cell.2020.04.011

Kim, J., Zhang, J., Cha, Y., Kolitz, S., Funt, J., Escalante Chong, R., et al. (2020). Advanced bioinformatics rapidly identifies existing therapeutics for patients with coronavirus disease–2019 (COVID-19). *ChemRxiv [Preprint].* doi: 10.26434/chemrxiv.12037416

Kong, R., Yang, G., Xue, R., Liu, M., Wang, F., Hu, J., et al. (2020). *COVID-19 Docking Server: An Interactive Server for Docking Small Molecules, Peptides and Antibodies Against Potential Targets of COVID-19.* Available online at: https://arxiv.org/abs/2003.00163 (accessed April 29, 2020). doi: 10.1093/bioinformatics/btaa645

Kong, W.-H., Li, Y., Peng, M.-W., Kong, D.-G., Yang, X.-B., Wang, L., et al. (2020). SARS-CoV-2 detection in patients with influenza-like illness. *Nat. Microbiol.* 5, 675–678. doi: 10.1038/s41564-020-0713-1

Lai, M. M., and Cavanagh, D. (1997). The molecular biology of coronaviruses. *Adv. Virus Res.* 48, 1–100. doi: 10.1016/S0065-3527(08)60286-9

Laufer, S., Greim, C., and Bertsche, T. (2002). An in-vitro screening assay for the detection of inhibitors of proinflammatory cytokine synthesis: A useful tool for the development of new antiarthritic and disease modifying drugs. *Osteoarthr. Cartil.* 10, 961–967. doi: 10.1053/joca.2002.0851

Lavecchia, A. (2019). Deep learning in drug discovery: opportunities, challenges and future prospects. *Drug Discovery Today* 24, 2017–2032. doi: 10.1016/j.drudis.2019.07.006

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

Li, J., Shao, J., Wang, C., and Li, W. (2020). The epidemiology and therapeutic options for the COVID-19. *Precis. Clin. Med.* 3, 71–84. doi: 10.1093/pcmedi/pbaa017

Li, X., Yu, J., Zhang, Z., Ren, J., Peluffo, A. E., Zhang, W., et al. (2020). Network bioinformatics analysis provides insight into drug repurposing for COVID-2019. *Preprints* 1–15. doi: 10.20944/preprints202003.0286.v1

Lionta, E., Spyrou, G., Vassilatis, D., and Cournia, Z. (2014). Structure-based virtual screening for drug discovery: principles, applications and recent advances. *Curr. Top. Med. Chem.* 14, 1923–1938. doi: 10.2174/1568026614666140929124445

Liu, K., Sun, X., Jia, L., Ma, J., Xing, H., Wu, J., et al. (2019). Chemi-net: A molecular graph convolutional network for accurate drug property prediction. *Int. J. Mol. Sci.* 20:3389. doi: 10.3390/ijms20143389

Liu, Q., Allamanis, M., Brockschmidt, M., and Gaunt, A. L. (2018). "Constrained graph variational autoencoders for molecule design," in *Advances in Neural Information Processing Systems* (Montreal, QC), 7795–7804.

Liu, X. (2017). *Deep Recurrent Neural Network for Protein Function Prediction from Sequence.* Available online at: https://arxiv.org/abs/1701.08318 (accessed April 26, 2020). doi: 10.1101/103994

Lon, J. R., Bai, Y., Zhong, B., Cai, F., and Du, H. (2020). Prediction and evolution of B cell epitopes of surface protein in SARS-CoV-2. *bioRxiv [Preprint].* doi: 10.1101/2020.04.03.022723

Ma, J., Sheridan, R. P., Liaw, A., Dahl, G. E., and Svetnik, V. (2015). Deep neural nets as a method for quantitative structure-activity relationships. *J. Chem. Inf. Model.* 55, 263–274. doi: 10.1021/ci500747n

Magar, R., Yadav, P., and Farimani, A. B. (2020). *Potential Neutralizing Antibodies Discovered for Novel Corona Virus Using Machine Learning.* Available onlin at: http://arxiv.org/abs/2003.08447 (accessed April 30, 2020). doi: 10.1101/2020.03.14.992156

Malone, B., Simovski, B., Moliné, C., Cheng, J., Fontenelle, H., Vardaxis, I., et al. (2020). Artificial intelligence predicts the immunogenic landscape of SARS-CoV-2: toward universal blueprints for vaccine designs. *bioRxiv [Preprint].* doi: 10.1101/2020.04.21.052084

Messina, F., Giombini, E., Agrati, C., Vairo, F., Ascoli Bartoli, T., Al Moghazi, S., et al. (2020). COVID-19: viral-host interactome analyzed by network based-approach model to study pathogenesis of SARS-CoV-2 infection. *J. Transl. Med.* 18:233. doi: 10.1186/s12967-020-02405-w

Miyake, J., Kaneshita, Y., Asatani, S., Tagawa, S., Niioka, H., and Hirano, T. (2018). Graphical classification of DNA sequences of HLA alleles by deep learning. *Hum. Cell* 31, 102–105. doi: 10.1007/s13577-017-0194-6

Moore, B. J. B., and June, C. H. (2020). Cytokine release syndrome in severe COVID-19. *Science.* 368, 473–474. doi: 10.1126/science.abb8925

Mustafa, S., Balkhy, H., and Gabere, M. (2019). *Peptide-Protein Interaction Studies of Antimicrobial Peptides Targeting Middle East Respiratory Syndrome Coronavirus Spike Protein: An In Silico Approach.* London: Hindawi. doi: 10.1155/2019/6815105

Mustafa, S., Balkhy, H., and Gabere, M. N. (2018). Current treatment options and the role of peptides as potential therapeutic components for Middle East Respiratory Syndrome (MERS): a review. *J. Infect. Public Health* 11, 9–17. doi: 10.1016/j.jiph.2017.08.009

Nakagawa, K., Lokugamage, K. G., and Makino, S. (2016). "Viral and cellular mRNA translation in coronavirus-infected cells," in *Advances in Virus Research*, Vol. 96 (Cambridge, MA: Academic Press Inc.), 165–192. doi: 10.1016/bs.aivir.2016.08.001165

Nambiar, A., Heflin, M. E., Liu, S., Maslov, S., Hopkins, M., and Ritz, A. (2020). Transforming the Language of Life: Transformer Neural Networks for Protein Prediction Tasks. *bioRxiv.* 06.15.153643, (2020). doi: 10.1101/2020.06.15.153643

Naz, K., Naz, A., Ashraf, S. T., Rizwan, M., Ahmad, J., Baumbach, J., et al. (2019). PanRV: pangenome-reverse vaccinology approach for identifications of potential vaccine candidates in microbial pangenome. *BMC Bioinformatics* 20, 1–10. doi: 10.1186/s12859-019-2713-9

Oany, A. R., Al Emran, A., and Jyoti, T. (2014). Design of an epitope-based peptide vaccine against spike protein of human coronavirus: an *in silico* approach. *Drug Des. Devel. Ther.* 8, 1139–1149. doi: 10.2147/DDDT.S67861

Ong, E., Wang, H., Wong, M. U., Seetharaman, M., Valdez, N., and He, Y. (2020a). Vaxign-ML: supervised machine learning reverse vaccinology model for improved prediction of bacterial protective antigens. *Bioinformatics* 36, 1–7. doi: 10.1093/bioinformatics/btaa119

Ong, E., Wong, M. U., Huffman, A., and He, Y. (2020b). COVID-19 coronavirus vaccine design using reverse vaccinology and machine learning. *bioRxiv [Preprint].* doi: 10.1101/2020.03.20.000141

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

Park, S. J., Kim, Y. G., and Park, H. J. (2011). Identification of rna pseudoknot-binding ligand that inhibits the - 1 ribosomal frameshifting of SARS-coronavirus by structure-based virtual screening. *J. Am. Chem. Soc.* 133, 10094–10100. doi: 10.1021/ja1098325

Pazhouhandeh, M., M.-Sahraian, A., Siadat, S. D., Fateh, A., Vaziri, F., Tabrizi, F., et al. (2018). A systems medicine approach reveals disordered immune system and lipid metabolism in multiple sclerosis patients. *Clin. Exp. Immunol.* 192, 18–32. doi: 10.1111/cei.13087

Pesciullesi, G., Schwaller, P., Laino, T., and and, J.-Reymond, L. (2020). *Carbohydrate transformer: predicting regio- and stereoselective reactions using transfer learning. ChemRxiv [Preprint].* doi: 10.26434/chemrxiv.11935635

Pillaiyar, T., Meenakshisundaram, S., and Manickam, M. (2020). Recent discovery and development of inhibitors targeting coronaviruses. *Drug Discovery Today.* 5, 668–688. doi: 10.1016/j.drudis.2020.01.015

Plant, E. P., and Dinman, J. D. (2008). The role of programmed-1 ribosomal frameshifting in coronavirus propagation. *Front. Biosci.* 13, 4873–4881. doi: 10.2741/3046

Plant, H., Stacey, C., Tiong-Yip, C. L., Walsh, J., Yu, Q., and Rich, K. (2015). High-throughput hit screening cascade to identify respiratory syncytial virus (RSV) inhibitors. *J. Biomol. Screen.* 20, 597–605. doi: 10.1177/1087057115569428

Pollastri, G., Przybylski, D., Rost, B., and Baldi, P. (2002). Improving the prediction of protein secondary structure in three and eight classes using recurrent neural networks and profiles. *Proteins Struct. Funct. Genet.* 47, 228–235. doi: 10.1002/prot.10082

Prachar, M., Justesen, S., Steen-Jensen, D. B., Winther, O., and Bagger, F. O. (2020). COVID-19 vaccine candidates: prediction and validation of 174 SARS-CoV-2 epitopes. *bioRxiv [Preprint]*. doi: 10.1101/2020.03.20.000794

Prompetchara, E., Ketloy, C., and Palaga, T. (2020). Immune responses in COVID-19 and potential vaccines: Lessons learned from SARS and MERS epidemic. *Asian Pacific J. Allergy Immunol.* 38, 1–9. doi: 10.12932/AP-200220-0772

Rahman, M. S., Rahman, M. K., Saha, S., Kaykobad, M., and Rahman, M. S. (2019). Antigenic: an improved prediction model of protective antigens. *Artif. Intell. Med.* 94, 28–41. doi: 10.1016/j.artmed.2018.12.010

Rangan, R., Zheludev, I. N., and Das, R. (2020). RNA genome conservation and secondary structure in SARS-CoV-2 and SARS-related viruses. *bioRxiv [Preprint]*. doi: 10.1101/2020.03.27.012906

Rappuoli, R. (2000). Reverse vaccinology rino rappuoli. *Curr. Opin. Microbiol.* 3, 445–450. doi: 10.1016/S1369-5274(00)00119-3

Rasmussen, L., Maddox, C., Moore, B. P., Severson, W., and White, E. L. (2011). A high-throughput screening strategy to overcome virus instability. *Assay Drug Dev Technol.* 9, 184–190. doi: 10.1089/adt.2010.0298

Redka, D. S., MacKinnon, S. S., Landon, M., Windemuth, A., Kurji, N., and Shahani, V. (2020). PolypharmDB, a Deep Learning-Based Resource, Quickly Identifies Repurposed Drug Candidates for COVID-19. *ChemRxiv [Preprint]* doi: 10.26434/chemrxiv.12071271.v1

Richardson, P., Griffin, I., Tucker, C., Smith, D., Oechsle, O., Phelan, A., et al. (2020). Baricitinib as potential treatment for 2019-nCoV acute respiratory disease. *Lancet* 395, e30–e31, 15. doi: 10.1016/S0140-6736(20)30304-4

Salem, M., Khormali, A., Arshadi, A. K., Webb, J., and Yuan, S.-J. (2020). Transcreen: transfer learning on graph-based anti-cancer virtual screening model. *Big Data Cogn. Comput.* 4:16. doi: 10.3390/bdcc4030016

Sarkar, B., Ullah, M. A., Johora, F. T., Taniya, M. A., and Araf, Y. (2020). The essential facts of wuhan novel coronavirus outbreak in china and epitope-based vaccine designing against COVID-19. *bioRxiv [Preprint]*. doi: 10.1101/2020.02.05.935072

Scheller, J., Garbers, C., and Rose-John, S. (2014). Interleukin-6: From basic biology to selective blockade of pro-inflammatory activities. *Sem. Immunol.* 26, 2–12. doi: 10.1016/j.smim.2013.11.002

Senior, A. W., Evans, R., Jumper, J., Kirkpatrick, J., Sifre, L., Green, T., et al. (2020). Improved protein structure prediction using potentials from deep learning. *Nature* 577, 706–710. doi: 10.1038/s41586-019-1923-7

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

Sher, G., Zhi, D., and Zhang, S. (2017). DRREP: deep ridge regressed epitope predictor. *BMC Genomics* 18:676. doi: 10.1186/s12864-017-4024-8

Shin, B., Park, S., Kang, K., and Ho, J. C. (2019). Self-attention based molecule representation for predicting drug-target interaction. *arXiv [Preprint] arXiv*:1908.06760.

Shoichet, B. K. (2004). Virtual screening of chemical libraries. *Nature* 432, 862–865. doi: 10.1038/nature03197

Shukla, P., Khandelwal, R., Sharma, D., Dhar, A., Nayarisseri, A., and Singh, S. K. (2019). Virtual screening of IL-6 inhibitors for idiopathic arthritis. *Bioinformation* 15, 121–130. doi: 10.6026/97320630015121

Simonovsky, M., and Komodakis, N. (2018). "GraphVAE: towards generation of small graphs using variational autoencoders," in *International Conference on Artificial Neural Networks* (Cham: Springer), 412–422. doi: 10.1007/978-3-030-01418-6_41

Smith, M., and Smith, J. C. (2020). Repurposing therapeutics for COVID-19: supercomputer-based docking to the SARS-CoV-2 viral spike protein and viral spike protein-human ACE2 interface. *ChemRxiv [Preprint]*. doi: 10.26434/chemrxiv.11871402.v4

Soria-Guerra, R. E., Nieto-Gomez, R., Govea-Alonso, D. O., and Rosales-Mendoza, S. (2015). An overview of bioinformatics tools for epitope prediction:

Implications on vaccine development. *J. Biomed. Inform.* 53, 405–414. doi: 10.1016/j.jbi.2014.11.003

Stammler, S. N., Cao, S., Chen, S. J., and Giedroc, D. (2011). A conserved RNA pseudoknot in a putative molecular switch domain of the 3′-untranslated region of coronaviruses is only marginally stable. *RNA* 17, 1747–1759. doi: 10.1261/rna.2816711

Stebbing, J., Phelan, A., Griffin, I., Tucker, C., Oechsle, O., Smith, D., et al. (2020). COVID-19: combining antiviral and anti-inflammatory treatments. *The Lancet Infectious Diseases* 20, 400–402. doi: 10.1016/S1473-3099(20)30132-8

Sun, Y., Liang, D., Wang, X., and Tang, X. (2020). *DeepID3: Face Recognition with Very Deep Neural Networks.* Available online at: http://arxiv.org/abs/1502.00873 (accessed April 26, 2020).

Tanaka, T., Narazaki, M., and Kishimoto, T. (2016). Immunotherapeutic implications of IL-6 blockade for cytokine storm. *Immunotherapy* 8, 959–970. doi: 10.2217/imt-2016-0020

Tang, B., He, F., Liu, D., Fang, M., Wu, Z., and Xu, D. (2020). AI-aided design of novel targeted covalent inhibitors against SARS-CoV-2. *bioRxiv [Preprint]*. doi: 10.1101/2020.03.03.972133

Tilocca, B., Soggiu, A., Sanguinetti, M., Musella, V., Britti, D., Bonizzi, L., et al. (2020). Comparative computational analysis of SARS-CoV-2 nucleocapsid protein epitopes in taxonomically related coronaviruses. *Microbes Infect.* 22, 188–194. doi: 10.1016/j.micinf.2020.04.002

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

Torng, W., and Altman, R. B. (2019). Graph convolutional neural networks for predicting drug-target interactions. *J. Chem. Inf. Model.* 59, 4131–4149. doi: 10.1021/acs.jcim.9b00628

Touret, F., Gilles, M., Barral, K., Nougairède, A., Decroly, E., de Lamballerie, X., et al. (2020). In vitro screening of a FDA approved chemical library reveals potential inhibitors of SARS-CoV-2 replication. *bioRxiv [Preprint]*. doi: 10.1101/2020.04.03.023846

Toxicology, EPA's National Center for Computational (2018). *ToxCast Database (invitroDB). The United States Environmental Protection Agency's Center for Computational Toxicology and Exposure. Dataset.* doi: 10.23645/epacomptox.6062623.v5

Tran, N. H., Qiao, R., Xin, L., Chen, X., Shan, B., and Li, M. (2019). Personalized deep learning of individual immunopeptidomes to identify neoantigens for cancer vaccines. *bioRxiv [Preprint]*. doi: 10.1101/620468

Vaishnav, N., Gupta, A., Paul, S., and John, G. J. (2015). Overview of computational vaccinology: vaccine development through information technology. *J. Appl. Genet.* 56, 381–391. doi: 10.1007/s13353-014-0265-2

Vaswani, A., Brain, G., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., et al. (2017). "Attention is all you need," in *31st Conference Neural Infection Processing System (NIPS 2017)*.

Wallach, I., Dzamba, M., and Heifets, A. (2020). *AtomNet: A Deep Convolutional Neural Network for Bioactivity Prediction in Structure-based Drug Discovery.* Available onlion at: http://arxiv.org/abs/1510.02855 (accessed April 22, 2020).

Wan, Y., Shang, J., Sun, S., Tai, W., Chen, J., Geng, Q., et al. (2019). Molecular mechanism for antibody-dependent enhancement of coronavirus entry. *J. Virol.* 94, 1–15. doi: 10.1128/JVI.02015-19

Wang, D., Liu, W., Shen, Z., Jiang, L., Wang, J., Li, S., et al. (2020). Deep learning based drug metabolites prediction. *Front. Pharmacol.* 10:1586. doi: 10.3389/fphar.2019.01586

Wang, G., Li, X., and Wang, Z. (2015). APD3: the antimicrobial peptide database as a tool for research and education. *Nucleic Acids Res.* 44, 1087–1093. doi: 10.1093/nar/gkv1278

Weiss, K., Khoshgoftaar, T. M., and Wang, D. D. (2016). A survey of transfer learning. *Big Data J.* 3:9. doi: 10.1186/s40537-016-0043-6

Worldometer (2020). *Coronavirus Cases. Worldometer.* Available online at: https://www.worldometers.info/coronavirus/coronavirus-cases/#daily-cases (accessed April 27, 2020).

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

Wu, J., Wang, W., Zhang, J., Zhou, B., Zhao, W., Su, Z., et al. (2019). DeepHLApan: a deep learning approach for neoantigen prediction considering both HLA-peptide binding and immunogenicity. *Front. Immunol.* 10:2559. doi: 10.3389/fimmu.2019.02559

Xia, S., Xu, W., Wang, Q., Wang, C., Hua, C., Li, W., et al. (2020). *Peptide-Based Membrane Fusion Inhibitors Targeting HCoV-229E Spike Protein HR1 and HR2 Domains. mdpi.com.* Available online at: https://www.mdpi.com/1422-0067/19/2/487 (accessed April 28, 2020).

Xiang, Z., and He, Y. (2009). Vaxign: a web-based vaccine target design program for reverse vaccinology. *Procedia Vaccinol.* 1, 23–29. doi: 10.1016/j.provac.2009.07.005

Yang, D., and Leibowitz, J. L. (2015). The structure and functions of coronavirus genomic 3' and 5' ends. *Virus Research* 206, 120–133. doi: 10.1016/j.virusres.2015.02.025

Yu, W., and Mackerell, A. D. (2017). "Computer-aided drug design methods," in *Methods in Molecular Biology, Vol. 1520,* ed P. Sass (New York, NY: Humana Press Inc.), 85–106. doi: 10.1007/978-1-4939-6634-9_5

Zhai, S., Chang, K., Zhang, R., and Zhang, Z. (2016). DeepIntent: Learning attentions for online advertising with recurrent neural networks KDD'16. in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (New York, NY: Association for Computing Machinery), 1295–1304.

Zhang, C., Wu, Z., Li, J.-W., Zhao, H., and Wang, G.-Q. (2020). The cytokine release syndrome (CRS) of severe COVID-19 and Interleukin-6 receptor (IL-6R) antagonist Tocilizumab may be the key to reduce the mortality. *Int. J. Antimicrob. Agents* 55:105954. doi: 10.1016/j.ijantimicag.2020.105954

Zhang, H., Saravanan, K. M., Yang, Y., and Hossain, T. (2020). Deep learning based drug screening for novel coronavirus 2019-nCov. *Prepr* 19, 1–17. doi: 10.20944/preprints202002.0061.v1

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:eabb3405. doi: 10.1126/science.abb3405

Zhavoronkov, A., Aladinskiy, V., Zhebrak, A., Zagribelnyy, B., Terentiev, V., Bezrukov, D. S., et al. (2020). Potential 2019-nCoV 3C-like protease inhibitors designed using generative deep learning approaches Potential COVID-19 3C-like protease inhibitors designed using generative deep learning approaches.

Insilico Med. Hong Kong Ltd A 307:E1. doi: 10.26434/chemrxiv.829102.v1

Zhavoronkov, A., Ivanenkov, Y. A., Aliper, A., Veselov, M. S., Aladinskiy, V. A., Aladinskaya, A. V., et al. (2019). Deep learning enables rapid identification of potent DDR1 kinase inhibitors. *Nat. Biotechnol.* 37, 1038–1040. doi: 10.1038/s41587-019-0224-x

Zheng, C., Yu, W., Xie, F., Chen, W., Mercado, C., Sy, L. S., et al. (2019). The use of natural language processing to identify Tdap-related local reactions at five health care systems in the Vaccine Safety Datalink. *Int. J. Med. Inform.* 127, 27–34. doi: 10.1016/j.ijmedinf.2019.04.009

Zhong, F., Xing, J., Li, X., Liu, X., Fu, Z., Xiong, Z., et al. (2018). Artificial intelligence in drug design. *Sci. China Life Sci.* 61, 1191–1204. doi: 10.1007/s11427-018-9342-2

Zhou, H., Gao, M., and Skolnick, J. (2015). Comprehensive prediction of drug-protein interactions and side effects for the human proteome. *Sci. Rep.* 5:11090. doi: 10.1038/srep11090

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

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