Computing and Artificial Intelligence 2024, 2(2), 1279.
https://doi.org/ 10.59400/cai.v2i2.1279

Review

Applications of reinforcement learning, machine learning, and virtual screening in SARS-CoV-2-related proteins

Yasunari Matsuzaka[1,2,*], Ryu Yashiro[2,3]

[1] Division of Molecular and Medical Genetics, Center for Gene and Cell Therapy, The Institute of Medical Science, The University of Tokyo, Minato-ku, Tokyo 108-8639, Japan
[2] Administrative Section of Radiation Protection, National Institute of Neuroscience, National Center of Neurology and Psychiatry, Kodaira, Tokyo 187-8551, Japan
[3] Department of Mycobacteriology, Leprosy Research Center, National Institute of Infectious Diseases, Tokyo 162-8640, Japan
*** Corresponding author:** Yasunari Matsuzaka, yasunari80808@ims.u-tokyo.ac.jp

CITATION

Matsuzaka Y, Yashiro R. Applications of reinforcement learning, machine learning, and virtual screening in SARS-CoV-2-related proteins. Computing and Artificial Intelligence. 2024; 2(2): 1279.
https://doi.org/10.59400/cai.v2i2.1279

ARTICLE INFO

Received: 9 June 2024
Accepted: 27 August 2024
Available online: 10 September 2024

Abstract: Similarly, to all coronaviruses, SARS-CoV-2 uses the S glycoprotein to enter host cells, which contains two functional domains: S1 and S2 receptor binding domain (RBD). Angiotensin-converting enzyme 2 (ACE2) is recognizable by the S proteins on the surface of the SARS-CoV-2 virus. The SARS-CoV-2 virus causes SARS, but some mutations in the RBD of the S protein markedly enhance their binding affinity to ACE2. Searching for new compounds in COVID-19 is an important initial step in drug discovery and materials design. Still, the problem is that this search requires trial-and-error experiments, which are costly and time-consuming. In the automatic molecular design method based on deep reinforcement learning, it is possible to design molecules with optimized physical properties by combining a newly devised coarse-grained representation of molecules with deep reinforcement learning. Also, structured-based virtual screening uses protein 3D structure information to evaluate the binding affinity between proteins and compounds based on physicochemical interactions such as van der Waals forces, Coulomb forces, and hydrogen bonds, and select drug candidate compounds. In addition, AlphaFold can predict 3D protein structures, given the amino acid sequence, and the protein building blocks. Ensemble docking, in which multiple protein structures are generated using the molecular dynamics method and docking calculations are performed for each, is often performed independently of docking calculations. In the future, the AlphaFold algorithm can be used to predict various protein structures related to COVID-19.

Keywords: angiotensin-converting enzyme 2; AlphaFold; Deep Q Network; molecular dynamics; SARS-CoV-2; reinforcement learning; virtual screening

1. Introduction

The novel coronavirus, severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) emerged as a human pathogen in Wuhan, China at the end of 2019 and has since spread around the world, resulting in a pandemic [1]. Symptoms appear about four to five days after being infected with the virus but can take as long as two weeks. On the other hand, asymptomatic infections have also been reported [2]. The main symptoms include fever, cough, difficulty breathing, body malaise, chills, muscle pain, headache, sore throat, and loss of smell and taste. Elderly people and people with underlying health conditions such as heart disease or diabetes are more likely to develop severe pneumonia [3]. Respiratory symptoms, high fever, diarrhea, and taste disorders have also been reported in other generations. When infected during childhood, the symptoms are mild or asymptomatic, but viral infection itself occurs,

and transmission to the elderly due to asymptomatic infection has also been reported. The host range of SARS-CoV-2 is wide, and this virus infects not only humans and wild animals, but also livestock, pets, laboratory animals, and many other animals, causing various diseases. Genetic sequence analysis has shown that this virus is like the coronavirus found in bats and pangolins [4], and it has been pointed out that these viruses may have undergone genetic recombination. Understanding the structure and function of this virus is essential to developing vaccines and treatments for coronavirus infectious disease, which emerged in 2019 (COVID-19).

Artificial Intelligence (AI) methods are being increasingly utilized to predict various aspects related to SARS-CoV-2. Here are some key findings from the search results:

I) Prediction of COVID-19 severity based on blood protein profiling:
- A study aimed to classify COVID-19 patients into mild, severe, critical, and control groups based on blood protein profiling using deep learning, random forest, and gradient-boosted trees [5].
- The ensemble classifier GBTs produced the highest accuracy in predicting disease severity (96.98%) [5].
- This approach identified specific proteins associated with COVID-19 severity, highlighting the potential for early diagnosis and treatment strategies [5].

II) Prediction of SARS-CoV-2 epitopes:
- Machine learning technologies have been used to predict target human proteins of the SARS-CoV-2 virus based on protein sequences and amino acid composition [6].
- Studies have focused on epitope prediction for SARS-CoV-2 S protein using machine learning models and immunological data from SARS-CoV [7].
- The aim is to identify nonallergenic, highly antigenic, and nontoxic epitopes that can be used in vaccine design against SARS-CoV-2 [7].

III) AI-based mutation prediction in SARS-CoV-2:
- Research is ongoing to develop AI models that predict the next variants of the SARS-CoV-2 virus based on genomic data.

These studies demonstrated the potential of AI-based methods in predicting COVID-19 severity, identifying epitopes for vaccine design, and forecasting mutations in the SARS-CoV-2 virus.

In this review, we focused our attention on the relevant new fields, such as the prediction of the SARS-CoV-2-related protein with AI, such as reinforcement learning and AlphaFold.

2. Classification and structure of SARS-CoV-2

Coronaviruses that infect birds and mammals belong to the order Nidovirales, family Coronaviridae, subfamily Orthocoronaviridae, which includes four genera: alphacoronavirus, betacoronavirus, gammacoronavirus, and deltacoronavirus. Currently, seven types of coronaviruses are known to infect humans; HCoV-229E, HCoV-NL63, HCoV-OC43, and HCoV-HKU1, which are human coronaviruses (HCoV) that routinely infect humans, SARS coronavirus (SARS-CoV-1), which

caused Severe Acute Respiratory Syndrome (SARS) in 2003, Middle East Respiratory Syndrome (MERS) coronavirus (MERS-CoV), which emerged in 2012, and the new coronaviruses (SARS-CoV-2) that is currently causing a pandemic [8]. Among the seven viruses mentioned above, HCoV-229E and HCoV-NL63 belong to the alphacoronavirus genus, and the remaining five viruses (HCoV-OC43, HCoV-HKU1, SARS-CoV-1, MERS-CoV, and SARS-CoV-2) is classified into the beta coronavirus genus, which is divided into four lineages (A, B, C, and D lineages) (**Figure 1**) [9]. Phylogenetic analysis indicates that all the coronaviruses that infect humans are derived from wild animals including bats and rodents. It is thought that coronaviruses originally carried by natural hosts including bats and rodents first infected intermediate hosts, and then eventually infected humans, causing disease. Regarding SARS-CoV-2, the sequence of a coronavirus closely related to this virus has been found in bats, so, likely, the natural host of SARS-CoV-2 is also a bat. Additionally, a coronavirus closely related to SARS-CoV-2 has been detected in Malayan pangolins, so there is a theory that Malayan pangolins are an intermediate host, but the details are unknown.

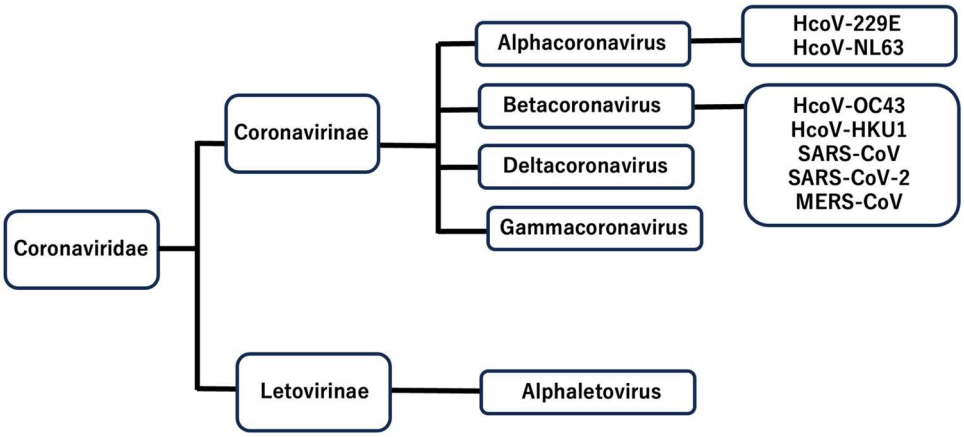

Figure 1. Taxonomy of coronaviridae family [10].

The homology of SARS-CoV-2 genomic RNA and viral proteins with SARS-CoV-1 is 79.0% for genomic RNA, 76.2% for S protein, 94.7% for E protein, 90.1% for M protein, and 90.3% for N protein [11]. Betacornavirus lineage B, which is included by SARS-CoV-1 and SARS-CoV-2, and an enveloped, single-stranded RNA virus characterized by spikes protruding from its surface and an unusually large RNA genome whose size is approximately 27 to 32 kb that is the largest among currently known RNA viruses [12]. The SARS-CoV-2 genome, whose size is approximately 30 kb, encodes four structural proteins; spike (S) protein, nucleocapsid (N) protein, membrane (M) protein, and envelope protein, each of which is essential for constructing the virus particle (**Figure 2**) [13]. The genomic RNA has a cap structure and a poly (A) sequence, at the 5’ end at the 3’ end, respectively, so it can infect host cells and function directly as mRNA. There are two open reading frames (ORF1a and ORF1b) in approximately 20 kb at the 5’ end of the viral RNA, and the start codon of ORF1b is located slightly upstream of the stop codon of ORF1b. Two proteins are translated from ORF1a and ORF1b:1a and 1a + 1b, which is synthesized by frameshifting of ribosomes. These proteins are cleaved by their proteases into more

than a dozen types of nonstructural proteins, including RNA-dependent RNA polymerase.

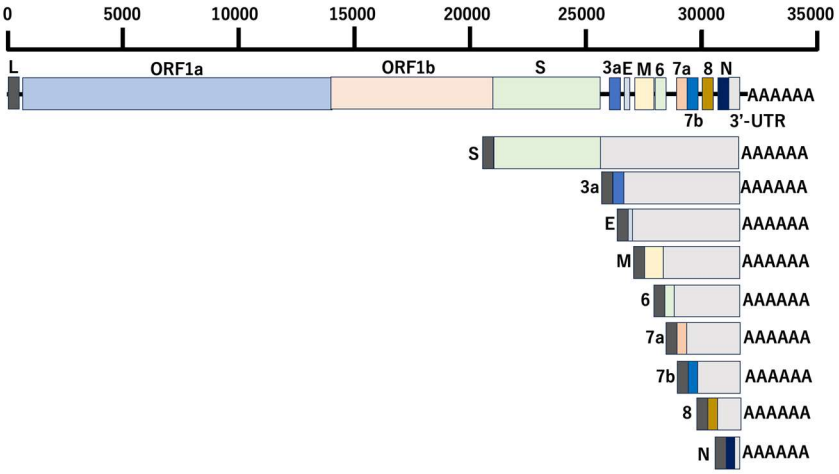

Figure 2. SARS-CoV-2 genome organization and the canonical subgenomic mRNAs. The full-length genomic RNA (29,903 nt) which also serves as an mRNA, ORF1a and ORF1b are translated [14].

3. Interaction of SARS-CoV-2 S protein with a receptor on the cell

Angiotensin-converting enzyme 2 (ACE2) is recognizable by the S protein on the surface of the SARS-CoV-2 or SARS-CoV virus [15]. On the other hand, it has also been shown that the S protein of SARS-CoV-2 does not recognize Dipeptidyl peptidase 4 (DPP4), the receptor for MERS-CoV, and APV, the receptor for HCoV-229E [16]. ACE2 and S protein combine like a lock and key, allowing the virus to enter human cells [17]. The SARS-CoV-2 virus is very similar to the SARS-CoV virus that causes severe acute respiratory syndrome (SARS), but the receptor binding region of the S protein significantly enhances the binding affinity of the SARS-CoV-2 virus to ACE2 via several mutations [18]. Like all coronaviruses, SARS-CoV-2 uses the S glycoprotein to enter host cells, which contains two functional domains: S1 and S2 RBDs [19]. The two subunits, S1 and S2 are cleaved from the S protein by host cell proteases [20]. S1 plays a role in receptor binding via RBD, and S2 plays a role in membrane fusion between the viral envelope and the cell [21].

The SARS-CoV-2 S protein first binds to the host cell's ACE2 receptor, which is a membrane protein with an enzyme domain in the cell membrane of human cells, via the S1 RBD [22]. The receptor specificity of the S protein is a major factor determining the host range and tissue tropism which is the ability to selectively infect specific tissues or organs of coronavirus [23]. It has been identified that ACE2 is the receptor for SARS-CoV-1, DPP4 is the receptor for MERS-CoV, aminopeptidase N (APN) is the receptor for HCoV-229E, and 9-O-acetylated sialic acid is a receptor for HCoV-OC43 and HCoV-HKU1, respectively [24]. The cell entry mechanisms of coronaviruses have two routes after binding to the receptor, 1) entry into the cell from the cell surface, and 2) entry into the cell via endosomes after the virus particle is taken into the cell by endocytosis [25]. When an enveloped virus invades a cell, the viral envelope needs to fuse with the cell's lipid bilayer membrane [26]. In the case of

coronaviruses, the S protein subunit S2 contains a fusion peptide, which plays an important role in membrane fusion. In SARS-CoV-1, the second route is the main pathway, in which the viral S protein taken up by endocytosis is activated by host proteases and causes membrane fusion between the endosomes and the viral envelope [27].

Host proteases that can activate the S protein of SARS-CoV-1 include cathepsin, trypsin, elastase, and TMPRSS2 [28]. Additionally, the S protein of MERS-CoV is cleaved into S1 and S2 by Furin [29]. One of the major differences between SARS-CoV-1 and SARS-CoV-2 is that SARS-CoV-2 has a characteristic sequence of consecutive basic amino acids (RRAR) in the S1/S2 cleavage site of the S protein, called "Furin cleavage site" which is absent in SARS-CoV-1, but is present in the S protein of MERS-CoV and HCoV-OC43, and efficiently cleaved by Furin and other proteases. During the virus replication cycle, the S protein is cleaved into S1 and S2, but the location and timing of cleavage differs depending on the types of coronaviruses, that is 1) S protein is synthesized in infected cells and then cleaved by host protease, and 2) when a virus invades a target cell, the S protein binds to a receptor and is then cleaved by host protease [30].

The mechanism in SARS-CoV-1 is the latter, so the S protein exists in an uncleaved state on the surface of the virus particle, and when the virus invades cells, it is cleaved by host proteases (trypsin, elastase, cathepsin, TMPRSS2) [31]. In contrast, in the case of SARS-CoV-2, cleavage occurs within the cell after S protein synthesis due to the first mechanism [32]. Experiments using pseudotyped viruses have shown that the S protein of virus particles exists as cleaved forms of S1 and S2 [33]. In addition, it has been suggested that the S protein cleavage site with the Furin needs SARS-CoV-2 to efficiently infect the human respiratory tract and that the S protein activation by TMPRSS2 is important [34]. The RBD in the SARS-CoV-1 S protein is composed of a core structure and a receptor binding motif (RBM), and the RBM directly binds to the ACE2 surface [35]. The six amino acids Y442, L472, N479, D480, T487, and Y491 in the RBM of SARS-CoV-1 are critical for binding to ACE2 and are involved in determining the host range of SARS-related coronaviruses [36]. In SARS-CoV-2, those corresponding to these six amino acids are L455, F486, Q493, S494, N501, and Y505, but except for Y505 (Y491 in SARS-CoV-1), different from amino acids [37]. Regarding the binding affinity between the RBM of SARS-CoV-2 and ACE2 in various animal species, such as humans using computer analysis of the protein structure, the RBM of SARS-CoV-2 has a high binding affinity for ACE2 in humans, civets, pigs, ferrets, cats, orangutans, monkeys (green monkeys), and bats (acetonide), and the high binding affinity of mouse and rat ACE2 was predicted to be low [38].

After membrane fusion, the virus unsheathes and the virus genome is released into the cell, whereupon virus replication begins within the cytoplasm [39]. Since coronaviruses positive-strand genomic RNA can function as mRNA, it binds to host cell ribosomes and synthesizes RNA-dependent RNA polymerase and other substances necessary for virus replication [40]. Using the positive-strand genomic RNA as a template, the mRNA encoding each viral protein is transcribed based on the synthesized complementary negative-strand RNA, and the viral protein is produced

[41]. Replication of positive-strand genomic RNA for progeny viruses also takes place. Newly synthesized viral structural proteins (S, E, and M proteins) are transported to the endoplasmic reticulum-Golgi apparatus intermediate (ERGIC) [42]. The nucleocapsid formed by the N protein and viral RNA, together with other structural proteins, forms the virus particle and buds into the ERGIC [43]. M and E proteins play important roles in the virus budding step [44]. Progeny viruses budded within ERGIC are released outside the cell by exocytosis [45].

ACE is an enzyme that catalyzes the conversion of the peptide hormone angiotensin I (Ang I) to angiotensin II (Ang II) and is well-known as a vasoconstrictor that promotes muscle contraction of blood vessel walls and narrows the lumen of blood vessels [46]. ACE2, a viral receptor, also plays a role as a vasodilator. This is because it balances ACE and relaxes blood vessel walls [47]. Both ACE and ACE2 play pivotal roles in the renin-angiotensin system (RAS), which regulates blood pressure and blood flow in multiple organs, including the lung, heart, and kidney, and conjugates a complex network of enzymes, peptide hormones, and receptors [48]. Angiotensinogen, a precursor of Ang secreted by the liver, is cleaved by the kidney enzyme renin to produce Ang I, which is converted to Ang II, an eight amino acid hormone peptide by ACE [49]. Ang II binds to the type 1 angiotensin receptor (AT1R) on the surface of microvascular muscle cells, causing vasoconstriction and promoting salt reabsorption in the kidney [50]. Vasoconstriction and salt reabsorption both contribute to increased blood pressure [51]. Therefore, when ACE activity becomes abnormally high, the amount of Ang II increases, causing hypertension.

On the other hand, ACE2 catalyzes the eight amino acid peptide of Ang II to a seven amino acid peptide (Ang 1–7) [52]. Though its action on a different receptor the Mas-1 receptor (MasR), it has the opposite effect on Ang II [53]. Although the detailed role of Ang 1–7 in blood pressure regulation is not completely understood, it is believed that Ang 1–7 decreases blood pressure and induces vasodilation [18]. Further, ACE2 splits Ang I into Ang 1–9, thereby balancing the effects of ACE by removing the substrate [54]. By converting Ang II to Ang (1–7) and Ang I to Ang 1–9, ACE2 plays an important role in maintaining the balance between vasoconstriction and vasodilation to sustain blood pressure within an appropriate range [55].

4. Reinforcement learning

4.1. Q-learning in a finite Markov decision process

Reinforcement learning is a type of machine learning that is a “mechanism for AI to automatically learn” and is a technology for machines to automatically identify and predict based on learned data [56]. It refers to a technology in which the system learns appropriate control methods through repeated trials and error. Its main feature is that it can analyze data without human intervention. In conventional machine learning, humans had to extract and adjust “feature values”, which are indices for learning the data to be analyzed [57]. However, deep learning does not require human intervention to extract feature values, so machine learning can be easily performed [58]. Machine learning is mainly composed of the following three types: supervised learning with correct data, unsupervised learning with no correct data, and

reinforcement learning. The machine learns by recognizing many images as correct data. This method is called "supervised learning" [59].

On the other hand, "unsupervised learning "is a method of learning without giving correct data [60]. Machines analyze the characteristics of data, making it possible to identify and classify data. In reinforcement learning, an agent placed in a certain environment act on the environment and seeks a policy that maximizes the reward obtained [61]. The learning progresses through a cycle in which the agent acts on the environment, the environment updates the state evaluates the action, and informs the agent of the state and reward. The action-value function and policy are optimized through learning so that the total reward obtained is maximized. The reinforcement learning repeats the following steps: 1) the agent acts on the environment, 2) the environment gives agents updated states and rewards, and 3) the agent modifies its behavioral strategy based on the reward and returns to 1).

Q-learning, a kind of reinforcement learning, is one of the policy-off Temporal Difference (TD) learning of machine learning methods [62]. *Q*-learning converges to the optimal evaluation value when it tries an infinite number of episodes in which all states are sufficiently sampled in a finite Markov decision process [63]. In *Q*-learning, each rule to be executed has a value called *Q*-value, which indicates the effectiveness of the rule, which is a pair of a state and an agent's possible actions under that state, and the value is updated each time the agent acts. For example, assume that the agent's current state is St, and there are four possible actions *a, b, c,* and *d* in this state. At this time, the agent decides the action to take based on the four *Q* values, *Q (St, a), Q (St, b), Q (St, c)*, and *Q (St, d*).

Theoretically, the *Q* value convergence has been proven even if the trial is performed an infinite number of times. Still, to expedite the convergence, actions with a large *Q* value are selected with a high probability. As a selection method, select randomly with a small probability e, otherwise select the action with the maximum *Q* value, e-Greedy method, and roulette selection used in genetic algorithm, Boltzmann distribution as below softmax Equation (1) is used.

$$\pi (St, a) = \exp (Q (st, a)/T)/\Sigma p \in A\exp (Q (St, p)/T) \quad (1)$$

where *T* is a positive constant and *A* is the set of possible actions of the agent in state St. If the action is decided, then update the state and the *Q* value of the action. As an example, the state St agent chooses action a and the state transitions to st + 1. The updated formula for the action-value function, *Q* function in *Q*-learning (2) is as follows.

$$Q (st, at) = Q (st, at) + \alpha [rt + 1 + \gamma maxat + 1\varepsilon A (st + 1) Q (st + 1, at + 1) - Q (st, at)] \quad (2)$$

here, alpha is called a learning rate, which is a numerical value that satisfies the conditions described later, and gamma is called a discount rate, which is a constant between 0 and 1 inclusive. Also, rt + 1 is the reward the agent got when it transitioned to St + 1. The above update formula means that when the current state moves to the next state, the *Q* value is brought closer to the value of the state with the highest *Q* value in the next state. This means that if a state has a high reward, that reward will propagate to states that can reach that state with each update. As a result, optimal state transition learning is performed. When the learning rate satisfies the following conditions, in *Q*-learning, all *Q* values converge to the optimal value with probability.

$$\sum_{t=0}^{\infty} \alpha(t) \rightarrow \infty \tag{3}$$

$$\sum_{t=0}^{\infty} \alpha(t)\text{^}2 < \infty \tag{4}$$

Due to this good convergence, many studies have been done on *Q*-learning, but some problems have been pointed out.

In *Q*-learning, the *Q*-function was updated by updating the number of states s× the number of actions a. However, as the number of states grows, it becomes impractical to represent the *Q* function with a table. To solve this problem, the Deep Q Network (DQN) takes the approach of expressing the *Q* function with a convolutional neural network and devises ways to converge learning [64]. However, simply replacing the *Q* function with a convolutional neural network (CNN) does not result in successful learning, so efforts have been made to converge the learning. Deep reinforcement learning is a combination of reinforcement learning and deep learning methods, and the representative method is this DQN, which is an approximation that replaces the action value function, *Q* function in *Q* learning with a CNN.

In the automatic molecular design method based on deep reinforcement learning, it is possible to design molecules with optimized physical properties by combining a newly devised coarse-grained representation of molecules with deep reinforcement learning.

Reinforcement learning and virtual screening in drug discovery and materials design.

Reinforcement learning is a type of machine learning that uses two factors: agent and environment. The agent acts and learns the feedback (reward) from the environment regarding that action, thereby deriving a behavioral guideline (strategy) to maximize the reward. The main feature is that it is less dependent on datasets. In reinforcement learning, based on feedback from the environment, it does not require a static dataset unlike unsupervised and supervised learning because the agent learns from the experience it collects. In other words, there is no need for data collection, preprocessing, or labeling before learning. The reinforcement learning workflow generally is as follows; (i) creating the environment: define the environment in which the agent operates, including the interface between the agent and the environment—introducing simulation from the standpoint of safety and experiment ability. (ii) Definition of remuneration: define rewards for goals and decide how to calculate rewards. Rewards guide the agent's behavioral choices. (iii) Creating an agent: define an agent consisting of a policy and a reinforcement learning algorithm. Specifically, the selection of the method of representing the policy: neural networks, lookup tables, etc. Choosing an appropriate learning algorithm: neural networks are commonly used because they are well suited for learning in large state and action spaces. (iv) Agent learning and verification: set conditions for learning, such as stopping conditions, and perform agent learning. After learning, verify the policy derived by the agent. Reconsider the design of reward signals and strategies, and rerun learning if necessary. Reinforcement learning is sample-inefficient, especially for model-free, on-policy algorithms, and can take minutes to days to train. Therefore, learning is often

parallelized on multiple central processing *units* (CPUs), Graphics Processing Units (GPUs), or clusters. (v) Development of measures: Investigate the learned strategies. Based on this result, the process may return to the initial stage of the workflow. Specifically, if the learning process and policy derivation do not converge within the calculation time, the following items need to be updated before relearning.

Searching for new compounds is an important initial step in drug discovery and materials design, but the problem is that this search requires trial-and-error experiments, which are costly and time-consuming. On the other hand, in recent years, in silico drug discovery and materials search, in which chemical compounds are searched for in a computer, have been attracting attention. However, it is generally difficult to search the space of discrete chemical structures of compounds, and an efficient method is required. Therefore, in recent years, new search methods using deep neural networks (DNNs), such as methods using generative models such as variational autoencoders (VAE), have been proposed [65]. These techniques attempt to circumvent this problem by learning the mapping between the discrete compound space and the continuous latent space by a generative model approximated by DNNs, and by allowing compound optimization to be performed in this continuous latent space.

However, this method had some problems. For example, there is no objective metric to evaluate whether the learned mapping is suitable for efficient optimization. In addition, the learning process of the generative model is separated from the optimization process of the compound concerning the score function of the optimization target molecule. On the other hand, in methods based on reinforcement learning; by thinking of molecular design as a Markov decision process, the agent learns the optimal policy through the rewards provided by the surrounding environment.

Virtual screening (VS) in drug discovery is a method of selecting drug candidate compounds from many compounds using computers. Naturally, it cannot be used as a drug unless it shows medicinal efficacy, so VS mainly focuses on medical efficacy and evaluates the presence or absence of activity against drug target proteins. Such VS can be broadly divided into methods based on known active compounds (ligand-based VS; LBVS) and methods based on protein three-dimensional (3D) structure (structured-based VS; SBVS). LBVS is a method that mainly uses similarity evaluation of compounds and machine learning and uses known experimental results to construct regression prediction models and classification prediction models and uses these to select compounds [66]. While drug-candidate compounds can be selected with relatively high precision. Because it learns based on the few compounds that have been tested against the target protein, it is difficult to develop guidelines for how to optimize the chemical structure of selected drug candidate compounds due to the lack of novelty in the chemical structure of predicted active compounds and the lack of 3D structural knowledge, having been pointed out as major problems.

On the other hand, SBVS uses protein 3D structure information to evaluate the binding affinity between proteins and compounds based on physicochemical interactions such as van der Waals forces, Coulomb forces, and hydrogen bonds, and select drug candidate compounds [67]. Although this method is less accurate than

LBVS because it does not use known experimental information about the target protein, it can discover highly novel drug candidates. Furthermore, the estimated binding mode between the protein and the compound can be obtained, which can provide guidelines for subsequent structural optimization of the compound. Due to the above two advantages, it is attracting a lot of attention just like LBVS. SBVS and LBVS are used together, and compounds commonly selected by both are sent to in vitro experiments. By introducing LBVS methods such as machine learning using information on known active compounds, drug candidate compounds are often narrowed down to a certain extent before the SBVS process is performed. On the other hand, many methods combine SBVS and machine learning, such as machine learning methods for protein 3D structures and for predicted binding structures obtained by docking calculations [68]. These methods hardly expose the weakness of LBVS, which is that the predicted active compound has little novelty in its chemical structure, due to the use of the structures of various proteins or the binding structures of various proteins and compounds rather than simply using knowledge of known compounds for specific proteins.

In the investigation and selection of target proteins for drug discovery, first, the target protein is selected from among the proteins involved in SARS-CoV-2, which is the target of drug discovery [69]. In addition to selecting target proteins simply based on known infection mechanisms, protein selection is performed using bioinformatics methods, such as selecting target proteins using omics analysis. However, inhibiting proteins that play an important role in the human body can lead to side effects, so it must be avoided as much as possible. If there is an essential protein that has the same function as the target protein, even if it is not a perfect match, it is necessary to show selectivity despite slight structural differences, which increases the difficulty of drug discovery. It is also important to be able to conduct experiments using gene knockout rats and mice during non-clinical trials. In addition to these conditions, to perform SBVS, it is also necessary that the 3D structure of the protein is known, or that a reliable 3D structure can be estimated by homology modeling, etc. Furthermore, the final target protein should be determined by considering the difficulty of the drug binding site.

In SBVS, even changes inside chains can greatly affect the results of docking calculations, so it is necessary to carefully prepare the protein 3D structure [70]. Various protein 3D structure is registered in the Protein Data Bank (PDB), but if the complex structure with a compound is known, the local structure is likely to allow the compound to bind easily, and highly accurate compound selection can be expected [71]. However, the required resolution is strict, and docking calculations require a resolution of at least 2.2A to 2.5A. On the other hand, if the protein 3D structure is unknown, it is necessary to predict the 3D structure. In SBVS, the ab initio method is rarely used due to the resolution mentioned above, and homology modeling is used to predict the 3D structure using homologous proteins whose structures are known. Examples of homology modeling tools include MODELLER and SWISS-MODEL, etc., [72].

However, the situation in protein 3D structure prediction changed significantly with the release of AlphaFold2 [73]. Furthermore, the ionization state of some protein

residues changes depending on the environment. Since interactions due to Coulomb forces are stronger than der Waals forces and hydrogen bonds, it is extremely important to consider the ionization state. However, changes in the ionization state cannot occur during docking calculations or molecular dynamics (MD) simulations. Therefore, it is necessary to generate an ionized state in advance, and PROPKA is most widely used for this purpose. In most cases, the human body has a nearly neutral environment, so an ionized state of pH 7.0 is often generated and used for docking calculations. However, it should be noted that proteins present in the stomach, for example, must produce an ionization state under acidic conditions.

In drug binding site prediction and selection, identifying protein surface sites (druggable sites) where drugs can be expected to bind is essential for estimating more detailed binding structures [74]. The conditions for a druggable site include having a concave region called a "pocket" when a compound binds, the concave region being of appropriate size and deep enough and having a hydrophobic surface [75]. Among these features, widely used methods include POVME, which predicts drug binding sites based on the protein surface shape, Fpocket and SiteMap, which make estimations by considering the properties of the protein surface, and FTMap, which locates small probe molecules and finds energetically stable spaces [76]. When a clear active site exists, such as in an enzyme, drug design is often aimed at that active site, binding site estimation methods are especially important if a clear concave region appears only after a compound bind [77].

Examples of such cases include when a protein binds to a compound while changing its structure (induced fit), and when designed inhibitors of protein-protein interactions [78]. Another aspect of considering whether a site is a druggable site is the degree of conservation of the amino acid residues that make up the binding site. Significant differences in target protein sequences between experimental animals such as rats and mice and humans can lead to differences in drug efficacy, leading to the suspension of drug development during clinical trials. In addition, with antiviral drugs, it is possible to suppress the acquisition of drug resistance by designing drugs that target highly conserved sites that are essential for protein function [79]. For drug binding site prediction, binding site prediction using 3D convolutional neural networks (3D-CNN) has been actively proposed, such as DeepSite, Kalasanty, and DeepSurf, and a method for predicting peptide binding sites rather than compound binding sites [80].

Evaluation of compounds based on protein 3D structure usually involves computational difficulties. Furthermore, even if it is possible to estimate a drug candidate compound that promotes or inhibits protein function, many compounds are unsuitable as drugs due to problems such as compound solubility and side effects. Based on the above, compound filtering is performed from various perspectives. The most widely used rule for designed oral drugs is Lipinski's rule of five. This is a rule that Lipinski et al. summarize the chemical properties of drugs approved for oral use. It lists four conditions: molecular weight of 500 or less, hydrogen bond accepting groups of ten or less, hydrogen bond donating groups of five or less, and water-octanol partition coefficient logP of five or less (It's called the rule of five because everything is a multiple of five). QED (quantitative estimate of drug-likeness) is also widely used

as a method to evaluate this "oral drug-likeness" using real numbers [81]. Additionally, indicators related to side effects and toxicity have been proposed, such as PAINS, which summarizes the characteristics of compound substructures that frequently cause off-target effects that bind to and inhibit or activate other proteins. In addition, LBVS-like methods are often used to select compounds to reduce the amount of docking calculations. However, this should not be done too much, as the result approaches "the discovery rate of binding compounds is high, but the novelty of the compounds is low."

Like proteins, the ionization state of compounds also changes depending on the environment. Compounds often have a range of ionization states, and an ionization state of approximately pH 7.0 ± 2.0 is generated and used for docking calculations [82]. Tools that generate the ionization state of compounds include Schrodinger's Epik, ChemAxon's JChem Protonation Plugin, and the open-source software Dimorphite-DL [83]. Additionally, some compounds may have tautomers or optical isomers may not be separated and may be grouped in one compound entry. Such isomers often have significant effects, such as changes in the interaction mode with proteins and the occurrence of collisions with proteins due to changes in the 3D structure of the compound [84]. Therefore, it is necessary to generate each isomer for these as well. Regarding this, there are JChem Protonation Plugin from ChemAxon, LigPrep from Schrodinger, and open-source software Gypsum-DL.

4.2. Protein-compound docking calculation

Docking calculation is a method for predicting the binding affinity and binding mode of a certain compound to a drug-binding site of a protein. DUD-E is a benchmark data set for SBVS, and the enrichment factor (EF) is a ratio that indicates how much the proportion of active compounds has increased after selection compared to before selection [85]. For example, the EF (EFx%) when selecting the top x% is calculated as follows.

$$EFx\% = (Posx\%/Allx\%)/(Pos100\%/All100\%)$$

The denominator is the proportion of active compounds included in the benchmark data set, and the numerator is the proportion of active compounds after selection, which based on docking calculations often narrows down the evaluation target to 1/100 or less, so it is often set to a small value such as EF1%. Commercial software such as Glid and Surflex have high prediction accuracy, whereas open-source software AutoDock and AutoDock Vina tend to have lower prediction accuracy. Also, docking calculation takes about ten seconds per compound in Glide SP mode when using one CPU core. In addition, GPU-based docking software such as Quantum. Ligand. Dock and BUDE have been developed, and AutoDock has been implemented with GPU, achieving 250 times faster speed than one CPU core when using NVIDIA Titan V [86].

While docking calculations consider structural changes in compounds, structural changes in proteins are generally not considered. The structure of a protein changes to a greater or lesser degree due to the binding of a compound (induced fit), so taking protein structural changes into account is important for improving prediction accuracy [87]. However, although there are methods that consider structural changes in protein

side chains during docking calculations, they have not become common due to the computational complexity problem [88]. Ensemble docking, in which multiple protein structures are generated using the molecular dynamics (MD) method and docking calculations are performed for each, is often performed independently of docking calculations [89]. However, since the amount of calculation is doubled by the number of protein structures used in the docking calculation, the number of applications is limited to a small number of cases.

In protein-compound docking calculations, a reranking method has been proposed that outputs multiple predicted bond structures in the docking calculation and predicts the interaction mode or interaction energy of the bond structures [90]. Therefore, accuracy is improved compared to ranking based on scores obtained by docking calculations. 3D-CNNs that use the connection structure as an input are being proposed for these as well, but interestingly, there is no significant performance difference between methods that use the interaction mode as a feature and deep learning methods [91]. This suggests a lack of data for deep learning and sufficient maturity of domain knowledge regarding interactions. An example of the application of SBVS to COVID-19 is that SBVS was performed on approximately 2100 approved drugs and active compounds with IC50 < 10 microns were identified [92]. As a result of binding energy estimation using the MM-PBSA method for each compound, since they showed good binding energies of −8.73 kcal/mol or less, in vitro assays were performed on all of them, and a good hit compound with IC50 < 10 microns was obtained.

4.3. Compound selection using the MD method

MD methods, which simulate the temporal changes in the coordinates of each atom in environments where solutions such as proteins and solvents such as water exist, are used in a variety of analyses [93]. Programs that perform MD simulations include AMBER, GROMACS, NAMD, CHARMM, and Desmond. In addition, in MD simulation, the speedup rate by using accelerators such as GPU is extremely high. From the perspective of SBVS, MD simulation makes it possible to evaluate the binding strength between a protein and a compound while explicitly considering protein structural changes, solvation, entropic effects, etc., making it possible to select compounds with higher precision [94]. In MD calculations for SBVS, simulations are performed using the predicted binding structure from docking calculations as the starting points. For example, several methods of conducting multiple short-term simulations and evaluating how stable the predicted bond structure is and of highly accurate estimation for binding energy using MM-GBSA, MM-PBSA, or MP-CAFEE, etc., have been proposed. Since the orientation of even a single side chain is important for protein structures in drug discovery, there is a possibility that attention will be focused on estimating protein structures to which compounds can easily bind using MD simulations [95].

5. Prevention of asymptomatic infections of COVID-19

To control the SARS-CoV-2 pandemic, many countries have placed restrictions on non-essential travel, and have subsequently implemented travel restrictions using a

combination of the following four strategies to lift restrictions: whitelist, unrestricted travel permission; gray list, travelers providing proof of a negative PCR and reverse transcription before arrival; red list, travelers quarantined on arrival; blacklist, ban on non-essential travel. Decisions about which list to assign to this vary by country and are often based on publicly available population-level epidemiological indicators: cases per capita, deaths per population, and positivity rate [96]. However, it has been pointed out that these indicators are incomplete, with problems such as underreporting, bias in symptomatic populations, and reporting delays.

To address these issues, it will be possible to derive optimal border policies by using real-time estimates of COVID-19 prevalence and estimating the number of asymptomatic infected people with high accuracy. Unlike normal restriction protocols, allocations can be made from limited information, based on demographic information and past test results of the incoming population. This system estimates the prevalence of COVID-19 based on test results used in the past; i) Adaptively extract a minimum set of traveler types based on demographic characteristics, country, region, age, and gender, using the least absolute shrinkage and selection operator (LASSO) regression from high-dimensional statistic [97]. ii) Estimate the prevalence of each type using the empirical Bayes method, deriving prior probabilities from previous experience. This system environment is such that the prevalence of COVID-19 is low, two in 1,000 people and arrival rates vary widely by country. As a result, testing data is unbalanced (few cases among those eligible for testing) and sparse (few arrivals from specific countries). These data characteristics are sequentially processed using the empirical Bayes method to perform appropriate processing. Utilizing the prevalence estimates described above, a subset of travelers for PCR testing is derived based solely on traveler type. This allocation of tests is done by adjusting the exploration-exploitation trade-off between the two objectives. Specifically, i) maximize the number of infected asymptomatic travelers based on current information (exploitation), and ii) assign tests based on experience to travelers for whom there is no accurate estimate, and accurately understand and update the epidemic status (exploration). For a greedy allocation to this tradeoff, allocating tests to concentrate on types with high prevalence, test data for the types with the highest number of patients and moderate prevalence will not be extracted. As the prevalence of COVID-19 is rapidly increasing in some cases, it is necessary to understand as much as possible of moderate symptoms to carry out appropriate learning. These challenges can be viewed as multi-armed bandit problems in reinforcement learning especially batch bandit problems with non-stationary, contextual, delayed feedback, and constraints. Information from pipeline tests, that are not returning results, must be considered. To solve this exploration-exploitation trade-off, the algorithm is built based on the Gittins index. Each type introduces a deterministic index representing a risk score, incorporating both estimated prevalence and uncertainty, according to which allocations are made.

Reinforcement learning, machine learning, and VS have been utilized in the search for inhibitors against SARS-CoV-2-related proteins. machine learning techniques are commonly employed to identify potential compounds for drug development quickly and accurately. In a study focusing on the SARS-CoV-2 main protease (3CLpro), machine learning-based virtual screening was used to predict new

inhibitors. Algorithms such as K-nearest neighbor (KNN), support vector machine (SVM), and Random Forest (RF) were employed, with RF showing the best performance in classifying phytochemicals as potential inhibitors.

The use of VS combined with molecular docking and molecular dynamics simulations has led to the identification of high-potential therapeutic compounds that could inhibit SARS-CoV-2 pathogenesis. These advanced computational approaches have helped narrow down a list of over 4000 compounds to 26 promising candidates [98].

In another study, deep reinforcement learning was employed after an initial virtual screening to design dual-target inhibitors against SARS-CoV-2 main protease (Mpro) and papain-like protease (PLpro) [99]. Additionally, graph generative models have been explored for designing novel drug candidates targeting SARS-CoV-2 viral proteins. Addressing minor issues, it is important to note that while virtual screening is a powerful tool for drug discovery, it can yield a high proportion of false positive hits. To mitigate this, machine learning-based approaches are increasingly being integrated into virtual screening workflows to enhance accuracy and efficiency.

6. SARS-CoV-2 protein structure prediction by AlphaFold algorithm

With the increasing number of COVID-19 cases, the AlphaFold algorithm, a deep-learning algorithm developed by DeepMind, was utilized to predict various protein structures related to COVID-19 [100]. Given the amino acid sequence, and the building blocks of a protein, AlphaFold can predict 3D protein structures. The analysis of amino acid sequences into 3D structures is typically a long-term and intensive process, involving visualization techniques for a variety of protein and structural analyses, including nuclear magnetic resonance, cryo-electron microscopy, and X-ray crystallography, and is costly. However, AlphaFold, which is an AI system predicting the 3D structure of proteins from amino acid sequence information and won the CASP13 (Critical Assessment of Structure Prediction) competition, an international competition for protein 3D structure prediction, eschews these techniques and uses a DNN that predicts distances and angles between amino acids scored with gradient descent, resulted in achieving a dramatic high score [72]. Proteins have a variety of functions due to the folding of linear chains of amino acids linked by peptide bonds to form 3D structures. By elucidating this structure, it will be possible to elucidate the proteins involved in most diseases involving proteins, especially those related to SARS-CoV-2. However, the method by which proteins fold into their final 3D structure remains a black box. Because the theoretical number is astronomical, it has been pointed out that enumerating all possible configurations of a typical protein by brute force calculations takes a long time and is known as the “protein folding problem.” By using free modeling, AlphaFold can ignore similar structures in predictions, which is particularly useful for COVID-19.

AlphaFold consists of three different layers of DNNs [101]. The first layer consists of a variational autoencoder stacked with an attention model to generate realistic fragments based on a single amino acid sequence. In the second layer, it is divided into two sublayers. The first sublayer uses a 1D Convolutional Neural

Network (CNN) on the contact map to optimize inter-residue distances. This is a 2D amino acid residue distance representation by projection of the contact map into one dimension for input into the CNN. In the second sublayer, it optimizes the scoring network and the degree to the generated substructures observed like proteins using CNN with 3D structure. After normalization, a third neural network layer is added that scores the generated proteins against the actual model. AlphaFold's structure module takes as input the features of the amino acid sequence corresponding to the input sequence and the pair representation features of the MSA (Multiple Sequence Alignment) extracted by the Evoformer part and outputs the coordinates of all atoms and the prediction reliability score pIDDT for each residue [70]. AlphaFold2 consists of four modules. i) Data preparation module: using the amino acid sequence (input sequence) of the predicted 3D structure as a query, create MSA from the database and search template 3D structure (template structure) from the database using bioinformatics tools. However, the use of a template 3D structure is optional. ii) Embedding module: the creation of an MSA representation that links raw MSA with target sequence information and a pair representation that records the relative positional relationship between residues [102]. Dense vector transformation with embedding, which fully connected layer without activation for sparse input values. iii) Evoformer (Transformer for molecular evolution) module: feature extraction from MSA and pair representation [103]. Information exchange between MSA and pair representation. Axial attention and triangular attention are performed keeping in mind the characteristics of MSA and the physical constraints of spatial graphs (proteins). iv) Structure module: integration of MSA representation, residue pair representation, and current 3D structure using IPA (Invariant point attention) module. Prediction of relative movement instructions for each residue (= (3, 3) rotation matrix and (x, y, z) translation vector for the number of residues) and side chain torsion angle. The structure module consists of eight layers with shared weights [104]. Each layer updates the features S of the amino acid sequence and the 3D representation plotted in the coordinate system Ti (corresponding to object coordinates) defined for each residue. Ti is a pair of a rotation matrix Ri, which represents a rotation that superimposes the coordinate system defined for each residue on the global coordinate system, and a vector ti, which represents a translation.

$$Ti = (Ri, ti) \tag{5}$$

This model was trained on Protein Data Bank, a freely accessible database containing 3D structures of larger biomolecules, including proteins and nucleic acids. The output is a distribution map containing the secondary structure and accessible surfaces predicted. After cross-validation of the results for the COVID-19 spike protein using the experimentally determined structure, they submitted predictions for proteins whose structure is not readily determined. These proteins have membrane proteins, proteins 3a, nsp2, nsp4, nsp6, and C-terminal domains such as papain. The structures of these proteins may represent docking sites for new drugs and therapeutics and could aid drug development in efforts to contain COVID-19. Utilizing a protein structure prediction AI program for the unique structure of the "mutant strain" of SARS-CoV-2 has the potential to change the way research is done in the field of

biology, allowing researchers to search for potential targets for new treatments before samples physically arrive.

AlphaFold2 has been released, making highly accurate protein structure prediction results available [105]. The premise of SBVS is that there is a reliable 3D protein structure, and as the 3D structures of proteins have been known so far, the targets for SBVS are naturally limited. In contrast, with the advent of AlphaFold, it has become possible to perform SBVS on proteins whose structures are unknown.

AlphaFold has significantly advanced the prediction of protein structures, including those of SARS-CoV-2, the virus responsible for COVID-19. The algorithm predicts the three-dimensional structures of proteins from their amino acid sequences, a process that traditionally requires extensive experimental techniques such as cryo-electron microscopy, nuclear magnetic resonance, and X-ray crystallography.

I. S protein

The SARS-CoV-2 spike (S) glycoprotein, which is the main target of antibodies, has been a primary focus for AlphaFold predictions. These predictions have helped elucidate the structural features of the spike protein, including its interaction with the angiotensin-converting enzyme 2 (ACE2) receptor, which is critical for the virus's entry into human cells [106]. AlphaFold's predictions have also been used to study the structural changes in different variants of the spike protein, such as those in the Omicron variant, to understand their impact on vaccine efficacy and viral transmission [107].

II. Other SARS-CoV-2 proteins

AlphaFold has also been used to predict the structures of several other SARS-CoV-2 proteins that are less well-studied but are essential for the virus's lifecycle. These include the membrane protein, Nsp2, Nsp4, Nsp6, and the papain-like proteinase (C-terminal domain) [106].

III. Methodology and Validation

AlphaFold employs a neural network architecture that integrates evolutionary, physical, and geometric constraints of protein structures. The algorithm uses multi-sequence alignments and a deep neural network to predict distances and angles between amino acids, achieving high accuracy even when no homologous structures are available [108]. The accuracy of AlphaFold's predictions has been validated by comparing them with experimentally determined structures, showing close agreement in many cases [106]. Thus, AlphaFold has revolutionized the field of protein structure prediction, particularly for SARS-CoV-2, by providing high-accuracy models that facilitate drug development and enhance our understanding of viral biology.

7. The application of AI in SARS-CoV-2-related proteins

AI has been extensively applied in various aspects related to SARS-CoV-2 proteins, particularly in COVID-19 drug discovery and vaccine design. Here are some examples of AI applications in this field:

I) In prediction of vaccine candidates, AI tools like XGBoost have been used to predict vaccine candidates from non-structural proteins of SARS-CoV-2 [109].

II) In prediction of HLA-binding peptides, feed-forward neural networks have been employed to predict HLA-binding peptides from the SARS-CoV-2 virus based on binding stability [109]

III) As for the design of multiple-epitope vaccines, deep neural networks have been utilized for the prediction and design of multi-epitope vaccines that can manage the mutation of the virus [109].

These applications demonstrate how AI and machine learning play a crucial role in accelerating the discovery of effective drugs, vaccines, and treatment strategies for combating COVID-19 by leveraging the understanding of SARS-CoV-2-related proteins.

The computational techniques have been instrumental in various aspects related to SARS-CoV-2 research. These techniques have been applied in computational protein design for COVID-19 research, including the rapid design of peptides for detecting SARS-CoV-2 proteins [110]. In silico methods, computational tools, and bioinformatics resources have been utilized to annotate SARS-CoV-2 genomes and understand viral proteins [110]. Additionally, a computational study focused on cooperative binding to multiple SARS-CoV-2 proteins has been conducted, aiming to identify compounds with potential therapeutic effects through systems computational analysis [111]. These computational approaches play a crucial role in advancing our understanding of the virus and developing strategies for diagnosis and treatment.

The application of reinforcement learning, machine learning, and virtual screening in SARS-CoV-2-related proteins has shown promising results in identifying potential inhibitors for the virus. Studies have utilized machine learning-based virtual screening, molecular docking, and molecular dynamics simulations to identify novel compounds with the potential to inhibit key proteins like the main protease (Mpro) and papain-like protease (PLpro) of SARS-CoV-2 [97]. These approaches have led to discovering inhibitors effectively targeting these proteins, offering new avenues for developing antiviral agents. Additionally, in silico reinforcement learning has been employed to design spike/ACE2 inhibitory macrocycles, showcasing the use of AI in drug discovery for COVID-19. The combination of deep reinforcement learning, and virtual screening has been instrumental in optimizing hit molecules and developing effective non-covalent inhibitors for SARS-CoV-2 proteins. Furthermore, a novel protein design framework using reinforcement learning has been proposed to design a variant of the human ACE2 that binds more tightly to the SARS-CoV-2 S protein, potentially aiding in developing therapeutic solutions for COVID-19.

The application of AI in the study of SARS-CoV-2-related proteins has been a significant area of research, particularly in the context of the COVID-19 pandemic [112]. AI has been utilized in various domains including drug repurposing, structural biology, diagnostics, and vaccine development [113]. AI has played a crucial role in determining the structure of SARS-CoV-2 proteins. By predicting the structures of viral proteins, AI helps researchers understand the virus's mechanisms and identify potential targets for drug development.

I. Protein structure prediction and analysis

AI techniques have been used to predict and analyze the structure of SARS-CoV-2 proteins, which is crucial for understanding the virus and developing targeted

therapies. For example, deep learning models have been applied to predict protein structures and interactions [72].

II. Epitope prediction for vaccine design

AI algorithms have been employed to identify potential epitopes on SARS-CoV-2 proteins that could be targets for vaccine development. One study used computational analysis to compare SARS-CoV-2 nucleocapsid protein epitopes with those of related coronaviruses [114].

III. Drug target identification

AI-powered approaches have been used to identify potential drug targets among SARS-CoV-2 proteins. For instance, graph convolutional neural networks have been applied to predict drug-target interactions [114].

IV. Vaccine candidate ranking

Machine learning tools like Vaxign-ML have been developed to rank non-structural proteins as potential SARS-CoV-2 vaccine candidates using network-based algorithms [115].

V. Inhibitor discovery

AI has been utilized to rapidly screen large compound libraries to identify potential inhibitors of SARS-CoV-2 proteins. One study used deep docking to screen 1.3 billion compounds for potential inhibitors of the SARS-CoV-2 main protease [114].

VI. Antigenicity prediction

AI models have been used to predict the protective antigenicity of SARS-CoV-2 proteins. The spike (S) protein was found to have the highest protective antigenicity score [116].

VII. Immunogenic landscape prediction

AI techniques have been applied to predict the immunogenic landscape of SARS-CoV-2, which can guide universal vaccine design strategies [116]. These applications demonstrate how AI is being leveraged to accelerate research on SARS-CoV-2 proteins, potentially leading to faster development of effective drugs and vaccines against COVID-19. The integration of AI with biological and structural data has enabled researchers to rapidly analyze vast amounts of information and generate insights that can guide experimental work in the fight against the pandemic.

8. Conclusions

The SARS-CoV-2, like other coronaviruses, utilizes the S glycoprotein with S1 and S2 domains to enter host cells by binding to ACE2 receptors. Mutations in the S protein's RBD can enhance its affinity to ACE2. Searching for new compounds in COVID-19 involves trial-and-error experiments, but methods like deep reinforcement learning and structure-based virtual screening aid in drug discovery. AlphaFold, an AI system by DeepMind, predicts protein structures accurately by combining physical and biological approaches. It uses deep learning to predict 3D protein structures from amino acid sequences, achieving atomic accuracy even without homologous structures available.

The SARS-CoV-2 virus is very similar to the SARS-CoV virus that causes SARS, but several mutations in the RBR of the S protein greatly enhance the binding affinity of the SARS-CoV-2 virus to ACE2. The SARS-CoV-2 uses the S glycoprotein to enter

host cells, which has two functional domains: S1 and S2 RBD. New search methods using DNNs, such as methods using generative models such as VAE can learn the mapping between the discrete compound space and the continuous latent space by a generative model approximated by DNNs, and by allowing compound optimization to be performed in this continuous latent space. The learning process in the generative model is separated from the optimization process of the compound concerning the score function of the optimization target molecule. On the other hand, in methods based on reinforcement learning; by thinking of molecular design as a Markov decision process, the agent learns the optimal policy based on the rewards provided by the surrounding environment. By utilizing the AI program of a protein structure prediction for the unique structure of the "mutant strain" of SARS-CoV-2, it has the potential to search for potential targets for new drugs for SARS-CoV-2.

Conflicts of interest: The authors declare no conflict of interest.

References

1. Yashavantha Rao HC, Jayabaskaran C. The emergence of a novel coronavirus (SARS-CoV-2) disease and their neuroinvasive propensity may affect in COVID-19 patients. Journal of Medical Virology. 2020; 92(7): 786-790. doi: 10.1002/jmv.25918
2. Ma Y, Deng J, Liu Q, et al. Long-Term Consequences of Asymptomatic SARS-CoV-2 Infection: A Systematic Review and Meta-Analysis. International Journal of Environmental Research and Public Health. 2023; 20(2): 1613. doi: 10.3390/ijerph20021613
3. Bongiovanni M, De Lauretis A, Manes G, et al. Clinical characteristics and outcome of COVID-19 pneumonia in elderly subjects. Journal of Infection. 2021; 82(2): e33-e34. doi: 10.1016/j.jinf.2020.08.023
4. Gupta SK, Minocha R, Thapa PJ, et al. Role of the Pangolin in Origin of SARS-CoV-2: An Evolutionary Perspective. International Journal of Molecular Sciences. 2022; 23(16): 9115. doi: 10.3390/ijms23169115
5. Yaşar Ş, Çolak C, Yoloğlu S. Artificial Intelligence-Based Prediction of Covid-19 Severity on the Results of Protein Profiling. Computer Methods and Programs in Biomedicine. 2021; 202: 105996. doi: 10.1016/j.cmpb.2021.105996
6. Dey L, Chakraborty S, Mukhopadhyay A. Machine learning techniques for sequence-based prediction of viral–host interactions between SARS-CoV-2 and human proteins. Biomedical Journal. 2020; 43(5): 438-450. doi: 10.1016/j.bj.2020.08.003
7. Cihan P, Ozger ZB. A new approach for determining SARS-CoV-2 epitopes using machine learning-based in silico methods. Computational Biology and Chemistry. 2022; 98: 107688. doi: 10.1016/j.compbiolchem.2022.107688
8. Alluwaimi AM, Alshubaith IH, Al-Ali AM, et al. The Coronaviruses of Animals and Birds: Their Zoonosis, Vaccines, and Models for SARS-CoV and SARS-CoV2. Frontiers in Veterinary Science. 2020; 7. doi: 10.3389/fvets.2020.582287
9. Kesheh MM, Hosseini P, Soltani S, et al. An overview on the seven pathogenic human coronaviruses. Reviews in Medical Virology. 2021; 32(2). doi: 10.1002/rmv.2282
10. Alexandersen S, Chamings A, Bhatta TR. SARS-CoV-2 genomic and subgenomic RNAs in diagnostic samples are not an indicator of active replication. Nature Communications. 2020; 11(1). doi: 10.1038/s41467-020-19883-7
11. Naqvi AAT, Fatima K, Mohammad T, et al. Insights into SARS-CoV-2 genome, structure, evolution, pathogenesis and therapies: Structural genomics approach. Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease. 2020; 1866(10): 165878. doi: 10.1016/j.bbadis.2020.165878

12. Chateau A, Van der Verren SE, Remaut H, et al. The Bacillus anthracis Cell Envelope: Composition, Physiological Role, and Clinical Relevance. Microorganisms. 2020; 8(12): 1864. doi: 10.3390/microorganisms8121864
13. Bai C, Zhong Q, Gao GF. Overview of SARS-CoV-2 genome-encoded proteins. Science China Life Sciences. 2021; 65(2): 280-294. doi: 10.1007/s11427-021-1964-4
14. Dërmaku-Sopjani M, Sopjani M. Interactions between ACE2 and SARS-CoV-2 S Protein: Peptide Inhibitors for Potential Drug Developments Against COVID-19. Current Protein & Peptide Science. 2021; 22(10): 729-744. doi: 10.2174/1389203722666210916141924
15. Wartecki A, Rzymski P. On the Coronaviruses and Their Associations with the Aquatic Environment and Wastewater. Water. 2020; 12(6): 1598. doi: 10.3390/w12061598
16. Roy AN, Gupta AM, Banerjee D, et al. Unraveling DPP4 Receptor Interactions with SARS-CoV-2 Variants and MERS-CoV: Insights into Pulmonary Disorders via Immunoinformatics and Molecular Dynamics. Viruses. 2023; 15(10): 2056. doi: 10.3390/v15102056
17. Scialo F, Daniele A, Amato F, et al. ACE2: The Major Cell Entry Receptor for SARS-CoV-2. Lung. 2020; 198(6): 867-877. doi: 10.1007/s00408-020-00408-4
18. Shirbhate E, Pandey J, Patel VK, et al. Understanding the role of ACE-2 receptor in pathogenesis of COVID-19 disease: a potential approach for therapeutic intervention. Pharmacological Reports. 2021; 73(6): 1539-1550. doi: 10.1007/s43440-021-00303-6
19. Lan J, Ge J, Yu J, et al. Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. Nature. 2020; 581(7807): 215-220. doi: 10.1038/s41586-020-2180-5
20. Huang Y, Yang C, Xu X feng, et al. Structural and functional properties of SARS-CoV-2 spike protein: potential antivirus drug development for COVID-19. Acta Pharmacologica Sinica. 2020; 41(9): 1141-1149. doi: 10.1038/s41401-020-0485-4
21. Li X, Yuan H, Li X, et al. Spike protein mediated membrane fusion during SARS-CoV-2 infection. Journal of Medical Virology. 2022; 95(1). doi: 10.1002/jmv.28212
22. Raghuvamsi PV, Tulsian NK, Samsudin F, et al. SARS-CoV-2 S protein: ACE2 interaction reveals novel allosteric targets. eLife. 2021; 10. doi: 10.7554/elife.63646
23. Belouzard S, Millet JK, Licitra BN, et al. Mechanisms of Coronavirus Cell Entry Mediated by the Viral Spike Protein. Viruses. 2012; 4(6): 1011-1033. doi: 10.3390/v4061011
24. Bosch BJ, Smits SL, Haagmans BL. Membrane ectopeptidases targeted by human coronaviruses. Current Opinion in Virology. 2014; 6: 55-60. doi: 10.1016/j.coviro.2014.03.011
25. Shang J, Wan Y, Luo C, et al. Cell entry mechanisms of SARS-CoV-2. Proceedings of the National Academy of Sciences. 2020; 117(21): 11727-11734. doi: 10.1073/pnas.2003138117
26. Harrison SC. Mechanism of membrane fusion by viral envelope proteins. Adv Virus Res. 2005; 64: 231-261. doi: 10.1016/S0065-3527(05)64007-9. PMID: 16139596.
27. Koppisetti RK, Fulcher YG, Van Doren SR. Fusion Peptide of SARS-CoV-2 Spike Rearranges into a Wedge Inserted in Bilayered Micelles. Journal of the American Chemical Society. 2021; 143(33): 13205-13211. doi: 10.1021/jacs.1c05435
28. Simmons G, Zmora P, Gierer S, et al. Proteolytic activation of the SARS-coronavirus spike protein: Cutting enzymes at the cutting edge of antiviral research. Antiviral Research. 2013; 100(3): 605-614. doi: 10.1016/j.antiviral.2013.09.028
29. Millet JK, Whittaker GR. Host cell entry of Middle East respiratory syndrome coronavirus after two-step, furin-mediated activation of the spike protein. Proceedings of the National Academy of Sciences. 2014; 111(42): 15214-15219. doi: 10.1073/pnas.1407087111
30. Takeda M. Proteolytic activation of SARS-CoV-2 spike protein. Microbiology and Immunology. 2021; 66(1): 15-23. doi: 10.1111/1348-0421.12945
31. Bertram S, Glowacka I, Müller MA, et al. Cleavage and Activation of the Severe Acute Respiratory Syndrome Coronavirus Spike Protein by Human Airway Trypsin-Like Protease. Journal of Virology. 2011; 85(24): 13363-13372. doi: 10.1128/jvi.05300-11
32. Chan YA, Zhan SH. The Emergence of the Spike Furin Cleavage Site in SARS-CoV-2. Kumar S, ed. Molecular Biology and Evolution. 2021; 39(1). doi: 10.1093/molbev/msab327
33. Hoffmann M, Kleine-Weber H, Pöhlmann S. A Multibasic Cleavage Site in the Spike Protein of SARS-CoV-2 Is Essential for Infection of Human Lung Cells. Molecular Cell. 2020; 78(4): 779-784. doi: 10.1016/j.molcel.2020.04.022

34. Jawad B, Adhikari P, Podgornik R, et al. Key Interacting Residues between RBD of SARS-CoV-2 and ACE2 Receptor: Combination of Molecular Dynamics Simulation and Density Functional Calculation. Journal of Chemical Information and Modeling. 2021; 61(9): 4425-4441. doi: 10.1021/acs.jcim.1c00560
35. Carvalho PPD, Alves NA. Featuring ACE2 binding SARS-CoV and SARS-CoV-2 through a conserved evolutionary pattern of amino acid residues. Journal of Biomolecular Structure and Dynamics. 2021; 40(22): 11719-11728. doi: 10.1080/07391102.2021.1965028
36. Yerukala Sathipati S, Shukla SK, Ho SY. Tracking the amino acid changes of spike proteins across diverse host species of severe acute respiratory syndrome coronavirus 2. iScience. 2022; 25(1): 103560. doi: 10.1016/j.isci.2021.103560
37. Zhai X, Sun J, Yan Z, et al. Comparison of Severe Acute Respiratory Syndrome Coronavirus 2 Spike Protein Binding to ACE2 Receptors from Human, Pets, Farm Animals, and Putative Intermediate Hosts. Gallagher T, ed. Journal of Virology. 2020; 94(15). doi: 10.1128/jvi.00831-20
38. Nour AM, Li Y, Wolenski J, et al. Viral Membrane Fusion and Nucleocapsid Delivery into the Cytoplasm are Distinct Events in Some Flaviviruses. Pierson TC, ed. PLoS Pathogens. 2013; 9(9): e1003585. doi: 10.1371/journal.ppat.1003585
39. V'kovski P, Kratzel A, Steiner S, et al. Coronavirus biology and replication: implications for SARS-CoV-2. Nature Reviews Microbiology. 2020; 19(3): 155-170. doi: 10.1038/s41579-020-00468-6
40. Ahlquist P, Noueiry AO, Lee WM, et al. Host Factors in Positive-Strand RNA Virus Genome Replication. Journal of Virology. 2003; 77(15): 8181-8186. doi: 10.1128/jvi.77.15.8181-8186.2003
41. Upadhyay M, Gupta S. Endoplasmic reticulum secretory pathway: Potential target against SARS-CoV-2. Virus Research. 2022; 320: 198897. doi: 10.1016/j.virusres.2022.198897
42. Scherer KM, Mascheroni L, Carnell GW, et al. SARS-CoV-2 nucleocapsid protein adheres to replication organelles before viral assembly at the Golgi/ERGIC and lysosome-mediated egress. Science Advances. 2022; 8(1). doi: 10.1126/sciadv.abl4895
43. Siu YL, Teoh KT, Lo J, et al. The M, E, and N Structural Proteins of the Severe Acute Respiratory Syndrome Coronavirus Are Required for Efficient Assembly, Trafficking, and Release of Virus-Like Particles. Journal of Virology. 2008; 82(22): 11318-11330. doi: 10.1128/jvi.01052-08
44. Villanueva RA, Rouillé Y, Dubuisson J. Interactions between virus proteins and host cell membranes during the viral life cycle. Int Rev Cytol. 2005; 245: 171-244. doi: 10.1016/S0074-7696(05)45006-8. PMID: 16125548
45. Seltzer S. Linking ACE2 and angiotensin II to pulmonary immunovascular dysregulation in SARS-CoV-2 infection. International Journal of Infectious Diseases. 2020; 101: 42-45. doi: 10.1016/j.ijid.2020.09.041
46. Burrell LM, Johnston CI, Tikellis C, et al. ACE2, a new regulator of the renin–angiotensin system. Trends in Endocrinology & Metabolism. 2004; 15(4): 166-169. doi: 10.1016/j.tem.2004.03.001
47. Silhol F, Sarlon G, Deharo JC, et al. Downregulation of ACE2 induces overstimulation of the renin–angiotensin system in COVID-19: should we block the renin–angiotensin system? Hypertension Research. 2020; 43(8): 854-856. doi: 10.1038/s41440-020-0476-3
48. Chappell MC. Biochemical evaluation of the renin-angiotensin system: the good, bad, and absolute? American Journal of Physiology-Heart and Circulatory Physiology. 2016; 310(2): H137-H152. doi: 10.1152/ajpheart.00618.2015
49. Tamura K, Wakui H, Azushima K, et al. Angiotensin II Type 1 Receptor Binding Molecule ATRAP as a Possible Modulator of Renal Sodium Handling and Blood Pressure in Pathophysiology. Current Medicinal Chemistry. 2015; 22(28): 3210-3216. doi: 10.2174/0929867322666150821095036
50. Blaustein MP, Leenen FHH, Chen L, et al. How NaCl raises blood pressure: a new paradigm for the pathogenesis of salt-dependent hypertension. American Journal of Physiology-Heart and Circulatory Physiology. 2012; 302(5): H1031-H1049. doi: 10.1152/ajpheart.00899.2011
51. Pratiwi A, Hakim TR, Abidin MZ, et al. Angiotensin-converting enzyme inhibitor activity of peptides derived from Kacang goat skin collagen through thermolysin hydrolysis. January-2021. 2021; 14(1): 161-167. doi: 10.14202/vetworld.2021.161-167
52. Karnik SS, Singh KD, Tirupula K, et al. Significance of angiotensin 1–7 coupling with MAS1 receptor and other GPCRs to the renin-angiotensin system: IUPHAR Review 22. British Journal of Pharmacology. 2017; 174(9): 737-753. doi: 10.1111/bph.13742
53. Santos RA. Angiotensin-(1–7). Hypertension. 2014; 63(6): 1138-1147. doi: 10.1161/hypertensionaha.113.01274

54. Bosso M, Thanaraj TA, Abu-Farha M, et al. The Two Faces of ACE2: The Role of ACE2 Receptor and Its Polymorphisms in Hypertension and COVID-19. Molecular Therapy - Methods & Clinical Development. 2020; 18: 321-327. doi: 10.1016/j.omtm.2020.06.017
55. Valente J, António J, Mora C, et al. Developments in Image Processing Using Deep Learning and Reinforcement Learning. Journal of Imaging. 2023; 9(10): 207. doi: 10.3390/jimaging9100207
56. Pudjihartono N, Fadason T, Kempa-Liehr AW, et al. A Review of Feature Selection Methods for Machine Learning-Based Disease Risk Prediction. Frontiers in Bioinformatics. 2022; 2. doi: 10.3389/fbinf.2022.927312
57. De Teyou GK, Tarabalka Y, Manighetti I, et al. Deep Neural Networks for automatic extraction of features in time series satellite images. Available online: https://arxiv.org/abs/2008.08432 (accessed on 17 May 2024).
58. Sodhani S, Faramarzi M, Mehta SV, et al. An Introduction to Lifelong Supervised Learning. Available online: https://arxiv.org/abs/2207.04354 (accessed on 17 May 2024).
59. Yang R. Unsupervised machine learning for physical concepts. Available online: https://arxiv.org/abs/2205.05279 (accessed on 17 May 2024).
60. Goel D, Neumann A, Neumann F, et al. Evolving Reinforcement Learning Environment to Minimize Learner's Achievable Reward: An Application on Hardening Active Directory Systems. Available online: https://arxiv.org/abs/2304.03998 (accessed on 17 May 2024).
61. Chitnis R, Xu Y, Hashemi B, et al. IQL-TD-MPC: Implicit Q-Learning for Hierarchical Model Predictive Control. Available online: https://arxiv.org/abs/2306.00867 (accessed on 17 May 2024).
62. Neufeld A, Sester J. Robust Q-learning Algorithm for Markov Decision Processes under Wasserstein Uncertainty. Available online: https://arxiv.org/abs/2210.00898 (accessed on 17 May 2024)
63. Ronecker MP, Zhu Y. Deep Q-Network Based Decision Making for Autonomous Driving. Available online: https://arxiv.org/abs/2303.11634 (accessed on 17 May 2024).
64. Kadurin A, Nikolenko S, Khrabrov K, et al. druGAN: An Advanced Generative Adversarial Autoencoder Model for de Novo Generation of New Molecules with Desired Molecular Properties in Silico. Molecular Pharmaceutics. 2017; 14(9): 3098-3104. doi: 10.1021/acs.molpharmaceut.7b00346
65. Dai W, Guo D. A Ligand-Based Virtual Screening Method Using Direct Quantification of Generalization Ability. Molecules. 2019; 24(13): 2414. doi: 10.3390/molecules24132414
66. Maia EHB, Assis LC, de Oliveira TA, et al. Structure-Based Virtual Screening: From Classical to Artificial Intelligence. Frontiers in Chemistry. 2020; 8. doi: 10.3389/fchem.2020.00343
67. Tran-Nguyen VK, Junaid M, Simeon S, et al. A practical guide to machine-learning scoring for structure-based virtual screening. Nature Protocols. 2023; 18(11): 3460-3511. doi: 10.1038/s41596-023-00885-w
68. Wu C, Liu Y, Yang Y, et al. Analysis of therapeutic targets for SARS-CoV-2 and discovery of potential drugs by computational methods. Acta Pharmaceutica Sinica B. 2020; 10(5): 766-788. doi: 10.1016/j.apsb.2020.02.008
69. Fassihi A, Hatami S, Sirous H, et al. Preparing a database of corrected protein structures important in cell signaling pathways. Research in Pharmaceutical Sciences. 2023; 18(1): 67. doi: 10.4103/1735-5362.363597
70. Revillo Imbernon J, Chiesa L, Kellenberger E. Mining the Protein Data Bank to inspire fragment library design. Frontiers in Chemistry. 2023; 11. doi: 10.3389/fchem.2023.1089714
71. Junk P, Kiel C. HOMELETTE: a unified interface to homology modelling software. Valencia A, ed. Bioinformatics. 2021; 38(6): 1749-1751. doi: 10.1093/bioinformatics/btab866
72. Yang Z, Zeng X, Zhao Y, et al. AlphaFold2 and its applications in the fields of biology and medicine. Signal Transduction and Targeted Therapy. 2023; 8(1). doi: 10.1038/s41392-023-01381-z
73. Konc J, Janežič D. Protein binding sites for drug design. Biophysical Reviews. 2022; 14(6): 1413-1421. doi: 10.1007/s12551-022-01028-3
74. Alzyoud L, Bryce RA, Al Sorkhy M, et al. Structure-based assessment and druggability classification of protein–protein interaction sites. Scientific Reports. 2022; 12(1). doi: 10.1038/s41598-022-12105-8
75. Piazza I, Beaton N, Bruderer R, et al. A machine learning-based chemoproteomic approach to identify drug targets and binding sites in complex proteomes. Nature Communications. 2020; 11(1). doi: 10.1038/s41467-020-18071-x
76. Rufer AC. Drug discovery for enzymes. Drug Discovery Today. 2021; 26(4): 875-886. doi: 10.1016/j.drudis.2021.01.006
77. Farooq Q ul A, Shaukat Z, Aiman S, et al. Protein-protein interactions: Methods, databases, and applications in virus-host study. World Journal of Virology. 2021; 10(6): 288-300. doi: 10.5501/wjv.v10.i6.288

78. Matthew AN, Leidner F, Lockbaum GJ, et al. Drug Design Strategies to Avoid Resistance in Direct-Acting Antivirals and Beyond. Chemical Reviews. 2021; 121(6): 3238-3270. doi: 10.1021/acs.chemrev.0c00648
79. Wang Y, Wei Z, Xi L. Sfcnn: a novel scoring function based on 3D convolutional neural network for accurate and stable protein–ligand affinity prediction. BMC Bioinformatics. 2022; 23(1). doi: 10.1186/s12859-022-04762-3
80. Kosugi T, Ohue M. Quantitative Estimate Index for Early-Stage Screening of Compounds Targeting Protein-Protein Interactions. International Journal of Molecular Sciences. 2021; 22(20): 10925. doi: 10.3390/ijms222010925
81. Eberhardt J, Santos-Martins D, Tillack AF, et al. AutoDock Vina 1.2.0: New Docking Methods, Expanded Force Field, and Python Bindings. Journal of Chemical Information and Modeling. 2021; 61(8): 3891-3898. doi: 10.1021/acs.jcim.1c00203
82. Ropp PJ, Kaminsky JC, Yablonski S, et al. Dimorphite-DL: an open-source program for enumerating the ionization states of drug-like small molecules. Journal of Cheminformatics. 2019; 11(1). doi: 10.1186/s13321-019-0336-9
83. Nash S, Vachet RW. Gas-Phase Unfolding of Protein Complexes Distinguishes Conformational Isomers. Journal of the American Chemical Society. 2022; 144(48): 22128-22139. doi: 10.1021/jacs.2c09573
84. Cleves AE, Jain AN. Structure- and Ligand-Based Virtual Screening on DUD-E+: Performance Dependence on Approximations to the Binding Pocket. Journal of Chemical Information and Modeling. 2020; 60(9): 4296-4310. doi: 10.1021/acs.jcim.0c00115
85. Tang S, Chen R, Lin M, et al. Accelerating AutoDock Vina with GPUs. Molecules. 2022; 27(9): 3041. doi: 10.3390/molecules27093041
86. Jumper J, Hassabis D. Protein structure predictions to atomic accuracy with AlphaFold. Nature Methods. 2022; 19(1): 11-12. doi: 10.1038/s41592-021-01362-6
87. Chen T, Shu X, Zhou H, et al. Algorithm selection for protein–ligand docking: strategies and analysis on ACE. Scientific Reports. 2023; 13(1). doi: 10.1038/s41598-023-35132-5
88. Mohammadi S, Narimani Z, Ashouri M, et al. Ensemble learning from ensemble docking: revisiting the optimum ensemble size problem. Scientific Reports. 2022; 12(1). doi: 10.1038/s41598-021-04448-5
89. Verburgt J, Kihara D. Benchmarking of structure refinement methods for protein complex models. Proteins: Structure, Function, and Bioinformatics. 2021; 90(1): 83-95. doi: 10.1002/prot.26188
90. Peivaste I, Ramezani S, Alahyarizadeh G, et al. Rapid and accurate predictions of perfect and defective material properties in atomistic simulation using the power of 3D CNN-based trained artificial neural networks. Scientific Reports. 2024; 14(1). doi: 10.1038/s41598-023-50893-9
91. Aziz S, Waqas M, Mohanta TK, et al. Identifying non-nucleoside inhibitors of RNA-dependent RNA-polymerase of SARS-CoV-2 through per-residue energy decomposition-based pharmacophore modeling, molecular docking, and molecular dynamics simulation. Journal of Infection and Public Health. 2023; 16(4): 501-519. doi: 10.1016/j.jiph.2023.02.009
92. Gazi R, Maity S, Jana M. Conformational Features and Hydration Dynamics of Proteins in Cosolvents: A Perspective from Computational Approaches. ACS Omega. 2023; 8(3): 2832-2843. doi: 10.1021/acsomega.2c08009
93. Wang X, Chong B, Sun Z, et al. More is simpler: Decomposition of ligand-binding affinity for proteins being disordered. Protein Science. 2022; 31(7). doi: 10.1002/pro.4375
94. Kurniawan J, Ishida T. Protein Model Quality Estimation Using Molecular Dynamics Simulation. ACS Omega. 2022; 7(28): 24274-24281. doi: 10.1021/acsomega.2c01475
95. Li Y, Hou S, Zhang Y, et al. Effect of Travel Restrictions of Wuhan City Against COVID-19: A Modified SEIR Model Analysis. Disaster Medicine and Public Health Preparedness. 2021; 16(4): 1431-1437. doi: 10.1017/dmp.2021.5
96. Chakraborty M, Shakir Mahmud M, Gates TJ, et al. Analysis and Prediction of Human Mobility in the United States during the Early Stages of the COVID-19 Pandemic using Regularized Linear Models. Transportation Research Record: Journal of the Transportation Research Board. 2022; 2677(4): 380-395. doi: 10.1177/03611981211067794
97. Samad A, Ajmal A, Mahmood A, et al. Identification of novel inhibitors for SARS-CoV-2 as therapeutic options using machine learning-based virtual screening, molecular docking and MD simulation. Frontiers in Molecular Biosciences. 2023; 10. doi: 10.3389/fmolb.2023.1060076
98. Zhang L, Zhao H, Liu J, et al. Design of SARS-CoV-2 Mpro, PLpro Dual-Target Inhibitors Based on Deep Reinforcement Learning and Virtual Screening. Future Medicinal Chemistry. 2022; 14(6): 393-405. doi: 10.4155/fmc-2021-0269
99. Higgins MK. Can We AlphaFold Our Way Out of the Next Pandemic? Journal of Molecular Biology. 2021; 433(20): 167093. doi: 10.1016/j.jmb.2021.167093

100. Ismi DP, Pulungan R, Afiahayati. Deep learning for protein secondary structure prediction: Pre and post-AlphaFold. Computational and Structural Biotechnology Journal. 2022; 20: 6271-6286. doi: 10.1016/j.csbj.2022.11.012
101. Marcu ȘB, Tăbîrcă S, Tangney M. An Overview of Alphafold's Breakthrough. Frontiers in Artificial Intelligence. 2022; 5. doi: 10.3389/frai.2022.875587
102. Bertoline LMF, Lima AN, Krieger JE, et al. Before and after AlphaFold2: An overview of protein structure prediction. Frontiers in Bioinformatics. 2023; 3. doi: 10.3389/fbinf.2023.1120370
103. DeBenedictis EA, Chory EJ, Gretton DW, et al. Systematic molecular evolution enables robust biomolecule discovery. Nature Methods. 2021; 19(1): 55-64. doi: 10.1038/s41592-021-01348-4
104. Xu YC, ShangGuan TJ, Ding XM, et al. Accurate prediction of protein torsion angles using evolutionary signatures and recurrent neural network. Scientific Reports. 2021; 11(1). doi: 10.1038/s41598-021-00477-2
105. Kilim O, Mentes A, Pál B, et al. SARS-CoV-2 receptor-binding domain deep mutational AlphaFold2 structures. Scientific Data. 2023; 10(1). doi: 10.1038/s41597-023-02035-z
106. Gutnik D, Evseev P, Miroshnikov K, Shneider M. Using AlphaFold Predictions in Viral Research. Curr Issues Mol Biol. 2023; 45: 3705-3732. doi: 10.3390/cimb45040240
107. Ali MA, Caetano-Anollés G. AlphaFold2 Reveals Structural Patterns of Seasonal Haplotype Diversification in SARS-CoV-2 Spike Protein Variants. Biology. 2024; 13(3): 134. doi: 10.3390/biology13030134
108. Jumper J, Evans R, Pritzel A, et al. Highly accurate protein structure prediction with AlphaFold. Nature. 2021; 596(7873): 583-589. doi: 10.1038/s41586-021-03819-2
109. Lv H, Shi L, Berkenpas JW, et al. Application of artificial intelligence and machine learning for COVID-19 drug discovery and vaccine design. Brief Bioinform. 2021; 22: 320. doi: 10.1093/bib/bbab320
110. Kalita P, Tripathi T, Padhi AK. Computational Protein Design for COVID-19 Research and Emerging Therapeutics. ACS Central Science. 2023; 9(4): 602-613. doi: 10.1021/acscentsci.2c01513
111. Li J, McKay KT, Remington JM, et al. A computational study of cooperative binding to multiple SARS-CoV-2 proteins. Scientific Reports. 2021; 11(1). doi: 10.1038/s41598-021-95826-6
112. Ashique S, Mishra N, Mohanto S, et al. Application of artificial intelligence (AI) to control COVID-19 pandemic: Current status and future prospects. Heliyon. 2024; 10(4): e25754. doi: 10.1016/j.heliyon.2024.e25754
113. Prasad K, Kumar V. Artificial intelligence-driven drug repurposing and structural biology for SARS-CoV-2. Current Research in Pharmacology and Drug Discovery. 2021; 2: 100042. doi: 10.1016/j.crphar.2021.100042
114. Keshavarzi Arshadi A, Webb J, Salem M, et al. Artificial Intelligence for COVID-19 Drug Discovery and Vaccine Development. Frontiers in Artificial Intelligence. 2020; 3. doi: 10.3389/frai.2020.00065
115. Ghosh A, Larrondo-Petrie MM, Pavlovic M. Revolutionizing Vaccine Development for COVID-19: A Review of AI-Based Approaches. Information. 2023; 14(12): 665. doi: 10.3390/info14120665
116. Wang L, Zhang Y, Wang D, et al. Artificial Intelligence for COVID-19: A Systematic Review. Frontiers in Medicine. 2021; 8. doi: 10.3389/fmed.2021.704256