Computational Methods in Road towards Drug Discovery against SARS-CoV2

Syed Babar Jamal, Shumaila Naz, Raees Khan, Adnan Haider, Rabail Zehra Raza, Sumra Wajid Abbasi

ABSTRACT
SARS-CoV2 has affected millions of people around the globe with hundreds of mortalities. The emergence of SARS-CoV2 is very recent, and there is no potential drug or vaccine available. In this review, we have compiled the most frequently used computational methods in drug discovery, target proteins of SARS-CoV2 as well as implementation of computational methods. Most recent literature on SARS-CoV2 has been compiled from various journal search engines including Google Scholar, Academia, PubMed, Scopus, Research Gate, and the Web of Science. The keywords chosen for the searches were COVID-19, Corona Virus, SARS-CoV2, drug development and future directions. This review has far reaching implications to both the public health and pharmaceutical industries for potential novel drug development against SARS-CoV2.

Key Words: SARS-CoV-2, COVID-19, In Silico Drug Design, Drug Repurposing.

Introduction
The novel coronavirus has affected more than 24 million people across the world including 6.7 million active cases, 61,400 critical cases and 835,679 deaths have been reported by the World Health Organization (WHO), till date. In order to track the causative agent of the infection, the public health expert's teams in the Wuhan city, Hubei province of China embarked on investigations to figure out the exact route of virus spread and transmission. Epidemiological resources and molecular based approaches were ramped up to discern the formidable nature of infection followed by, environmental assessments of the wholesale seafood market, where the pathogen spill-over happened at the early outbreak time. Later, the epidemiological and molecular genotyping confirmed that a novel coronavirus (n-CoV), closely similar to the Severe Acute Respiratory Syndrome-related coronavirus (SARS-CoV) and the Middle East Respiratory Syndrome-related coronavirus (MERS-CoV). Initially, it was considered a spill-over and that the infection was disseminated into the local pool by animal-to-human contact; however, later human-to-human transmission was confirmed in patients with n-CoV in the southern part of China. The genome of the virus is around 30 kb encoding 15 proteins. As, SARS-CoV2 has emerged recently, there is currently no drug or vaccine available to combat this deadly pandemic. The high number of deaths associated with the virus is because of lack of treatment for SARS-CoV2. Therefore, the development of a seamless, smooth and quick process for the discovery of drugs against the virus is the need of the day. The identification of novel drug compound is drug discovery. This is a complex process as it involves various fields of sciences apart from being laborious and time-consuming process. It is for this reason numerous techniques of virtual screening were introduced in order to overcome the delays that usually hampers the development of drugs. For example, using virtual screening some of the protein structures of SARS-CoV2 have been stored in the protein data bank, and further structure can be added to the data bank via computational analysis. This review article summarizes the drug design via virtual screening based on their structure, ligand interactions with their receptors and the analysis of their potential efficacy against SARS-CoV2.
Potential drug targets

Major structural proteins and non structural protein (NSPs) play an important part in the design of drugs. Up till now, the seven main druggable targets for COVID-19 have been identified including: spike (S) glycoprotein, membrane (M) protein, envelope (E) protein, nucleocapsid (N) protein, 3-chymotrypsin-like protease (3CLpro)/main protease, helices and hemagglutinin esterase. Additionally, sixteen NSPs also known as functional proteins are tested as potential candidates for the development of drugs.

Structural proteins drug targets

The structural proteins of coronavirus that make up the genome of virus are envelope protein, spike glycoprotein, membrane protein, and nucleocapside protein. While these proteins are less conserved as compared to NSPs, they still have a crucial role to play in the life cycle of the virus. S glycoprotein, with a molecular weight of approximately 141178 kDa and 1273 amino acids, aids in the entry of virus by recognizing the host receptor and plays a crucial role in the pathogenesis of virus and organism tropism. SARS-CoV2 S protein has been reported to share approximately 75% homology with the SARS-CoV2 S protein. Similarly, SARS-CoV2 subunits S1 and S2 share 70% and 99% similarity with the S1 and S2 subunits of SARS-CoV, respectively. The E protein is a small integral membranous protein consisting of N terminal domain (NTD), a hydrophobic domain, and a C terminal chain containing 76-109 amino acids and 8-12 kDa in size. Although the E protein is the smallest of all other structural proteins, it has made a significant contribution to the assembly, budding, enveloping and virulence. In coronaviruses, M protein (220-260 length) is found in large concentrations out of all proteins. Due to its membrane-bending properties, the key role of M protein has been reported in the promotion of viral assembly. The N protein of corona virus is the most abundantly expressed at the early stages of infection in host samples. The N protein of SARS-CoV-2 shows 90 percent sequence similarity to the N protein of SARS-CoV-N. It is a multi-functional protein that bundles the viral RNA genome into a ribonucleo protein complex known as nucleocapsid for genome protection.

Non-structural proteins drug targets

Like other coronaviruses, SARS-CoV-2, has 16 highly conserved NSPs with different roles, including the development of the replication–transcription complex (RTC). Literature outlining the detailed functions of most of the NSPs have been published. X-ray crystallographic structures of the important NSPs of COVID-19 are currently available in the RCSB PDB (Table 1, Figure 1).

<table>
<thead>
<tr>
<th>Sr. No</th>
<th>Important NSPs</th>
<th>PDB ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Main protease (M3pro, 3CLpro, nsp5)</td>
<td>6Y28</td>
</tr>
<tr>
<td>2</td>
<td>Papain-like protease (3CLpro/nsp3)</td>
<td>6W9C [To be published]</td>
</tr>
<tr>
<td>3</td>
<td>RNA-dependent RNA polymerase (RdRp, nsp12 in complex with cofactors nsp7 and nsp8)</td>
<td>6M71</td>
</tr>
<tr>
<td>4</td>
<td>Methyltransferase-stimulatory factor complex of nsp16 and nsp10</td>
<td>6W61 [To be published]</td>
</tr>
<tr>
<td>5</td>
<td>Complex of nspP10- nsp16</td>
<td>6W75 and 6W4H [To be published]</td>
</tr>
<tr>
<td>6</td>
<td>nsp9-binding protein</td>
<td>6W48 [To be published]</td>
</tr>
<tr>
<td>7</td>
<td>nsp15 endoribonuclease</td>
<td>6VWW</td>
</tr>
</tbody>
</table>

While any NSP, which plays a pivotal role in viral infection, might be used as a druggable target. The accessibility of the x-ray crystallographic structures of proteins along with the co-crystallized ligand, significantly enhances the probability of success.

Fig 1. Three dimensional structures of recently reported Non-structural Proteins of COVID-19 host-based drug targets

Angiotensin-converting enzyme 2, abbreviated as ACE2, mediates the entry of coronavirus into the host cell. ACE2 has also been reported as a specific host for the receptor binding domain (RBD) of SARS-CoV spike protein. Recently it has been reported
that not only SARS-CoV-2 host receptor is identical to SARS-CoV but SARS-CoV-2 spike RDB sequence is also similar to the RBD of SARS-CoV. Additionally, interactions between the key residues of RBD and AEC2 are crucial for the anti-COVID-19 drug development. 

ACE2 is the host site for the discovery of COVID-19 therapy according to latest research.

Computational methods in drug discovery

Structure-based drug design

Drug designing based on their structure are carried out using the already available targeted protein structure models. The protein structures are elucidated using nuclear magnetic resonance (NMR), X-ray diffraction (XRD) and molecular simulation. After deducing the structure of protein via different techniques, molecular modelling softwares are used for analysing the receptor's binding sites and physicochemical properties of the drug such as hydrophobicity, electrostatic interactions/field, hydrogen bonding and major residues. After collecting all the relevant information about the drug, drug designing techniques are applied or databank of small molecules is searched for marking the suitable molecules which have higher binding affinity and can easily interact or fit the binding site present on the receptor. Once data is obtained about the selected molecules via the above mentioned method, the selected molecules are synthesized and their biological efficacies are tested to develop the drug. Hence, designing of drug based on their structure plays a pivotal role during drug designing and development.

Ligand-based drug design

For developing drug using ligand-based approach, hunting of small molecules data bank is not necessary. Instead, it depends on the available literature about the molecule and its interaction (binding mechanism to the targeted site) with the macromolecule or protein. Using the existed information about the known molecules, a pharmacophore model is derived which helps in defining that a molecule must possess minimum structural characteristics for interacting and binding to the target. Which further leads to designing of new molecules that can easily bind to the target molecule. Moreover, in designing ligand-based drugs, use of quantitative structure–activity relationships (QSAR) also play a crucial role. QSAR suggests a correlation between the obtained results (experimental biological activity results) from actual experiment and calculated properties (theoretically) which in turn helps in predicting the efficacy of the newly formed analogues.

Virtual screening

Virtual screening plays a pivotal role in drug designing. This method is cost effective and has already identified lead compounds in few reported studies. The conventional method of identifying the lead compound is that the researcher physically screens chemical databanks for finding potential biological targets. The lead compound is identified via carrying out separate biochemical analysis of thousands of compounds. This consumes both time and money. To overcome the hurdles faced during the conventional methods, a new cheaper less time-consuming and less laborious method was introduced which was called high-throughput screening. This technique is highly effective in initial screening of the lead compound or new drug. It helps in selection of active compound or from the vast library of compounds. The purpose of this technique is to screen out the potentially active small molecules. The virtual screening is even more cost effective and less time consuming, as compared to the high throughput screening method in which one has to screen out each and every compound available in the data bank.

Molecular docking

In drug designing, molecular docking is an important tool for predicting the interaction between lead molecules (small molecules) and the target protein. Similarly, using molecular docking, one can also predict the interaction between two protein
molecules. In short, molecular docking helps in finding the active site (binding sites) among the interacting molecules. Molecular docking works on the principle of ligand receptor interaction which banks on the energy matching and spatial shape matching. This phenomenon is called the theory of “inducing fit”. Furthermore, molecular docking also helps in finding the most suited conformation of both ligands and receptors. Molecular docking is further classified into flexible, semi flexible and rigid docking.

**Pharmacophore modeling**
A pharmacophore describes the pivotal molecular properties of ligand for recognition by biological molecules. It also explains binding of structurally diverse ligand molecules to the active site on the receptors. A drug or ligand upon its interaction with the targeted molecules forms active conformers of the target molecules via reorientation in terms of geometry and energy. Furthermore, replacement of a single functional group can result into completely different interaction with the target molecule. That’s why Ehrlich suggested the pharmacophores concept in 1909, by referring to the molecular geometry (arrangement of atoms) having the characteristic properties. Gund in 1977 further explained pharmacophores concept by claiming that pharmacophores are molecules that recognize the receptors and form the basis of biological activity of that molecule.

**Quantitative structure–activity relationship (QSAR)**
QSAR is a quantitative study designed to identify and predict how biological macromolecules interact with the organic micro molecules. It provides a relationship between experimentally derived biological efficacy versus the calculated properties of the molecules such as distribution, absorption and metabolism of molecules in organisms. This technique is of utmost importance in terms of its accuracy and effectiveness in case of unknown receptor structures in drug designing. In drug discovery, mostly QSAR technique is implemented for identifying the structure of molecules which can have excellent inhibitory effect on selected and specified targets. Furthermore, it has less toxicity (non-specific activity).

**In silico drug repurposing and drug discovery**
The recent SARS-CoV-2 outbreak welcomes and calls for potential treatment approaches using different FDA approved drugs. The application of computer-aided drug design methods is very effective to find out promising and potent drug repurposing targets quickly, particularly after having solved the comprehensive three-dimensional structures of chief viral proteins. Much work has been done in this respect and literature is flooded with large amount of published data. For example, Jin et al., documented a data-based drug repositioning design using structure-based drug design (SBDD) and high-throughput screening approaches to design and identify potent drug targets against the SARS-CoV-2 main protease. Computational screening, along with cell-based confirmation suggested that ebselen, an anti-inflammatory, antioxidant, and cytoprotective drug, could be repurposed for the treatment of COVID-19. The screening results and structural details of key drug targets (3CLpro/main protease, S protein, RdRp and PLpro) were also reported by Canrong Wua et al. They studied every single protein encoded by the SARS-CoV2 genes, made their comparison with all the respective proteins from other corona viruses, modelled their 3D structures, and designed 19 structures which might be developed using a comparative modeling approach. A total of 21 drug targets including 2 human specific targets as well, were identified and screened against compound libraries namely ZINC drug database and an in-house database of natural products by conducting target-based virtual ligand screening. The results of this research has provided new leads and drug targets for further SARS-CoV-2 in vitro and in vivo investigations, new ideas for all drug currently undergoing clinical trials, and also notable new drug reposting approaches for treating SARS-CoV-2 infections. Using the recently reported main protease crystal structure together with a co-crystallised inhibitor (RCSB PDB ID: 6LU7), Liu et al., performed virtual screening followed by molecular docking of approved drugs as well the drug candidates undergoing clinical trials. For the top docking hits, MD simulations were performed with binding free energy calculations. A number of known drugs, including carfilzomib, eravacycline, valrubicin, lopinavir, and elbasvir, have been identified as potential main protease inhibitors. Another study reported the theoretical
structure for main protease and its mode of interaction with various plausible protease inhibitor drugs such as remdesivir, nelfinavir, lopinavir, ritonavir and aketoamide using molecular docking approach. HIV protease inhibitors have also been recommended as one of the COVID-19 therapies. Molecular docking analysis of 61 molecules of known antiviral activity was conducted, of which four protease inhibitors were found to be useful, namely: Lopinavir, Asunaprevir, Indinavir, and Ritonavir.

Similarly, Wahedi and co-workers reported using different aforesaid computational approaches that stilbene-based compounds, especially resveratrol, could be potent anti-COVID-19 drug targets acting as spike protein disruptors. Further, to identify the potent inhibitors against SARS-CoV-2 PLpro, Amin et al., also reported a Monte Carlo optimization based QSAR followed by virtual screening of some in-house virtual chemicals and molecular study. It was hypothesized that studied chemicals may emerge as potent therapeutics against COVID-19 after extensive wet lab validation.

**Conclusion**

The deadly SARS-CoV-2 is one of the seven transmissible human coronaviruses believed to have originated from the bat Coronavirus. In extreme cases, treatments are confined to supportive care and COVID-19 management. Various in silico approaches have been influential in the drug screening and designing process. For instance, multi-scale simulations studies of different drug targets and protein-inhibitors docked complexes may be useful in the classification of important binding pockets for the selected targets and in the elucidation of mode of action of drugs at atomic level. Integrated computational approaches are expected to improve the drug design, development and repurposing process and to help find successful drugs with new mode of actions that could potentially be implemented to combat the novel SARS-CoV-2.

**REFERENCES**

21. Neuman BW, Kiss G, Kunding AH, Bhella D, Baksh MF,


