Identification of Potential Ligands of the Main Protease of Coronavirus SARS-CoV-2 (2019-nCoV) Using Multimodal Generative Neural-Networks.

Syeda Rehana Zia*

Department of Chemistry, University of Karachi, Karachi, Pakistan

srehana.zia@gmail.com

Keywords: *Coronavirus, COVID-19, SARS-CoV-2, 2019-nCoV, Main protease (Mpro/3CLpro), Neural-networks.*

The recent outbreak of coronavirus disease 2019 (COVID-19) is posing a global threat to human population. The pandemic caused by novel coronavirus (2019-nCoV), also called as severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2); first emerged in Wuhan city, Hubei province of China in December 2019. The rapid human to human transmission has caused the contagion to spread world-wide affecting 244,385,444 (244.4 million) people globally causing 4,961,489 (5 million) fatalities dated by 27 October 2021. At present, 6,697,607,393 (6.7 billion) vaccine doses have been administered dated by 27 October 2021, for the prevention of COVID-19 infections. Even so, this critical and threatening situation of pandemic and due to various variants' emergence, the pandemic control has become challenging; this calls for gigantic efforts to find new potent drug candidates and effective therapeutic approaches against the virulent respiratory disease of COVID-19. In the respiratory morbidities of COVID-19, the functionally crucial drug target for the antiviral treatment could be the main protease/3-chymotrypsin protease (Mpro/3CLpro) enzyme that is primarily involved in viral maturation and replication. In view of this, in the current study I have designed a library of small molecules against the main protease (Mpro) of coronavirus SARS-CoV-2 (2019-nCoV) by using multimodal generative neural-networks. The scaffold-based molecular docking of the series of compounds at the active site of the protein was performed; binding poses of the molecules were evaluated and protein-ligand interaction studies followed by the binding affinity calculations validated the findings. I have identified a number of small promising lead compounds that could serve as potential inhibitors of the main protease (Mpro) enzyme of coronavirus SARS-CoV-2 (2019-nCoV). This study would serve as a step forward in the development of effective antiviral therapeutic agents against the COVID-19.

Introduction

The recent global outbreak of novel coronavirus disease 2019 (COVID-19) is 10 times more severe than the swine flu pandemic that emerged in 2009-2010 causing 1.6 million confirmed cases with the death toll of 18,449; as stated by the World Health Organization (WHO) [1,2]. COVID-19 has been found even more deadly than expected. The coronavirus disease is caused by novel coronavirus (2019-nCoV), which is called as severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2) by the International Committee on Taxonomy of Viruses [3]. SARS-CoV-2 is a new strain of coronavirus [3]; among the four genera (ranks) of coronavirus such as alpha, beta, gamma and delta, it belongs to beta coronavirus genus [4,5]. Coronaviruses (CoVs) have been known to infect a variety of vertebrates such as avian, swine and humans [6,7]. The alpha and beta CoVs cause infections only in mammals; whereas gamma and delta CoVs mostly infect avian (birds) but few of them can infect mammals too [8]. In humans and wild animals, CoVs cause respiratory and intestinal infections [8]. So far, six CoVs species have been found to cause infection in human hosts; they are grouped together and named as human coronaviruses (HCoVs), which cause several respiratory diseases such as pneumonia, bronchiolitis and common cold, and can also infect neurological, hepatic and enteric systems [3,6,7]. These six HCoVs are HCoV-OC43, HCoV-NL63, HCoV-229E, HCoV-HKU1, Middle East respiratory syndrome coronavirus (MERS-CoV) and severe acute respiratory syndrome coronavirus (SARS-CoV) [6,7]. Among them, four species of HCoVs (HCoV-OC43, HCoV-NL63, HCoV-229E and HCoV-HKU1) are commonly distributed in the human population across the globe and cause around one-third human infections related to common cold [6,7]. However, in case of adversity, these four species of HCoVs can cause severe respiratory illness in children, elderly persons and immunocompromised patients [6,7].

Prior to the onset of COVID-19, the two human pathogens (HCoVs) SARS-CoV and MERS-CoV have previously caused emergence of viral respiratory illness SARS in 2002-2003 in China and Hong Kong, and MERS in 2012 in Saudi Arabia [3,6,7,9]. SARS-CoV and MERS-CoV were transmitted to human hosts from intermediate hosts: palm civets and dromedary (Arabian) camels respectively; the primary origin of both CoVs were likely to be bats [6–8]. The outbreak of SARS affected more than 8000 people in 29 countries with a mortality rate of 10% [6]. The MERS emergence caused infections in more than 2000 individuals with a fatality rate of ~35% [3,6,7]. The SARS and MERS outbreaks dragged the attention of the research community towards HCoVs, which had not been considered previously as highly pathogenic to humans [8]. It is commonly assumed that viruses have already prevailed in their natural reservoirs for long times [6,8]. It

turns out to be a great challenge for researchers to find out the rationality for the rapid evolution of human coronaviruses (HCoV) with frequent genomic nucleotide mutations and recombinations [6,8]. There is a growing consensus that human activities comprising urbanization, modern agricultural practices and poultry farming are the factors that have largely caused a steady spill-over of viruses from their natural hosts to other animals and to humans [6,8,10]. Recurrent mixing of species have caused severe repercussions of genetic recombination of viruses and thus allowing them to cross the species barrier [6,8]. In view of the high risks of pandemics, the concept of "One World - One Health" has already been introduced globally in 2004; which highlights the linkage between mankind, animals and environment, and encapsulates the regimes related to the health of all three: humans, animals and ecosystem [8,11,12]. In the same line, it has already been emphasized that efforts should be made to maintain preventive barriers between natural reservoirs and human community; in order to control the evolution of genetically diverse pathogens and to possibly prevent the potential damage to the mankind due to life-threatening pathogenic emergence and viral zoonosis [6,8].

The SARS-CoV-2 belongs to Coronaviridae lineage, is a large spherical enveloped virus with a diameter ranging from 50 to 200 nm; containing positive-sense, non-segmented, and a long single-stranded RNA genome [3,6,7,13]. There are mainly four structural proteins of SARS-CoV-2: spike (S), membrane (M), envelope (E), and nucleocapsid (N) [13]. The nucleocapsid (N) of the virion is symmetrically helical structure that enfolds exceptionally large-sized RNA genome of 26-32 kilobases (kb) [6]. The name coronavirus is given for the reason that under an electron microscope, coronavirus seems to have a crown-like appearance due to the presence of club-shape spike glycoproteins present on its surface; also resembles it with solar corona [3,9,14]. The spike (S) protein allows the virus to anchor the cell membrane of the host [13,15,16]. The angiotensin converting enzyme 2 (ACE2) receptors present on the host's cell membrane is the target of the viral spike protein [13,15,16].

SARS-CoV-2 was first surfaced in Wuhan city, Hubei province of China in December 2019; the epicenter of viral infection was tied to exotic organisms and seafood wholesale marketplaces in the city [3,9,15,16]. The plausibility of SARS-CoV-2 to be a laboratory-construct or purposefully engineered virus is unlikely as no clues have been found to support the hypothesis, from the comparative genomic data analysis about SARS-CoV-2 origin [17,18]. However, recent studies have supported the animal origin of SARS-CoV-2, considering bats and pangolins as natural reservoirs of the virus [18,19]. The SARS-CoV-2 has shown 91.02% whole-genome sequence identity to Pangolin-CoV and 96.2% identity to Bat-CoV

(RaTG13) [18,19]. The disease was found to be highly contagious, caused rapid transmission from human to human; due to which the WHO declared Public Health Emergency of International Concern (PHEIC) on 30 January 2020 [15,16]. By 27 October 2021, the COVID-19 epidemic has affected 244,385,444 (244.4 million) people worldwide with 4,961,489 (5 million) fatalities. At present, 6,697,607,393 (6.7 billion) vaccine doses have been administered for the prevention of COVID-19 infections [20].

The viral genome of SARS-CoV-2 constituted of ~30,000 nucleotides [21]. The RNA genome of the virus behaves as a messenger RNA (mRNA) once invaded and subsequently infected the host cell [22]. Thereby, the viral replicase gene directs the encoding of two long polyproteins namely pp1a and pplab, required for replication/transcription of the virus [21]. Proteolytic processes of the cutting of polyproteins into polypeptides and other functional machinery are performed by the proteases [22]. The main protease (Mpro) also referred to as 3-chymotrypsin protease (3CLpro) is the key viral enzyme of coronavirus SARS-CoV-2 (2019-nCoV) that plays an essential role in viral replication/transcription and maturation; thus turns it out to be a fascinating drug target [21,23].

In the current prolonged situation of global crises due to coronavirus disease (covid-19) pandemic, a sense of strong and urgent need has come out to explore novel drug discovery approaches and to discover new effective drugs, therapeutics and specialized tools for treatment [24–26]. On the other hand, a huge and significant amount of drug research has already been conducted in the last two years and continuous efforts are still under way [26,27]. In the same line, different covid-19 vaccines to prevent against SARS-CoV-2 virus have been developed and approved by WHO, namely: Pfizer-BioNTech, Moderna, Johnson & Johnson's Janssen, Oxford–AstraZeneca, Sinopharm BIBP, Covaxin, CoronaVac, and Novavax [28,29]. Vaccines have changed the contagion and their role is gratifying, but having said that there is still a strong urge for drugs which can rapidly and effectively treat covid-19 [30]. The immunity against coronavirus gained by vaccines gradually wanes after a particular time of administration; as the pandemic prevails and due to surfacing of variants (especially Omicron variant), booster shots are recommended to reduce the chances of becoming infected by the viral transmission [31]. Besides vaccines, other alternative therapeutic options are antibodies and small antiviral drug molecules [27].

Drug discovery and development approaches could be based on either conventional/traditional strategies or existing-drug repurposing/repositioning [32]. Conventional drug discovery processes are expensive and time-consuming [28,32,33]. Repurposing/repositioning of old drugs, in which

existing drugs or compounds are reused to explore new therapeutic activities is on the contrary safe, speedy and cost-effective [28,32]. However, the drug repositioning process could be hampered due to finding the unique disease-drug relationship [32]. Therefore, assorted alternative approaches including artificial intelligence (AI) based computational approaches have been developed to address the limitations of drug discovery processes [32,34].

A great effort has been extensively carried out to discover efficacious antiviral drugs and to analyze existing drugs against covid-19 by various experimental and computational approaches [24,30,35–48]. It is pertinent to mention the attempts to design small molecule antivirals by Pfizer named as PF-07321332, Nirmatrelvir [49–52]. It is a 3CLpro/Mpro inhibitor of SARS-CoV-2 [49,50]. Food and Drug Administration (FDA) is currently considering to authorize the emergency use of two oral COVID-19 antiviral pills [49]. One, Paxlovid by Pfizer (a combination of Nirmatrelvir and Ritonavir) which has been authorized by FDA and the other Molnupiravir by Merck & Co. [49]. Molnupiravir is a RdRp (viral RNA-dependent RNA polymerase) inhibitor; RdRp is a viral enzyme that synthesizes RNA [49]. However, emergency use of RdRp inhibitor Remdesivir has already been approved by FDA in October 2020 which is intravenously administered to patients [49,53]. Remdesivir and Molnupiravir are anti-Ebola virus (EBV) and anti-Venezuelan equine encephalitis virus (VEEV) repurposed drugs respectively, while Paxlovid is SARS-CoV-2 optimized drug candidate [49]. Several Mpro and RdRp antivirals are still currently under development phases [49].

Artificial Intelligence is now increasingly applied in medical and health care [32,34,54]. One of the most common sub-field of AI is machine learning [32]. Machine learning is basically a statistical approach of fitting models to data sets and to learn from training models over data [34]. Neural network is the complex form of machine learning [34]. One of the machine learning methods involving artificial deep neural networks is termed as deep learning [32,34]. AI and machine learning are cutting-edge techniques that offer solutions and could support the discovery process and optimization of novel antivirals against SARS-CoV-2 [32,55]. Several robust AI based approaches including neural network have been successfully applied during the pandemic to identify and develop anti-covid19 drugs [25–28,30,32,42,55–62].

Recently a deep-learning model LiGANN (ligand generative adversarial network) has been developed that uses generative adversarial network (GAN) for structure-based *de novo* ligand design and generating novel small molecules based on target protein structural information, 3D pocket representation, protein shapes and chemical properties [24,63,64]. Initially the protein binding pocket is voxelized

and then GAN generates 3D ligand shapes which are then converted into SMILES (simplified molecular-input line-entry system) representation of ligand's chemical structure [63]. A recent study using a combination of neural network based deep learning technique of LiGANN, lead optimization and docking has been performed [65]. Macchiagodena et al. have applied LiGANN-Autodock4 protocol to identify the possible lead compounds against Mpro of SARS-CoV-2 [65]. A series of 93 molecules were generated, among them 5 lead compounds were identified; compound-27 was identified as the best binder to the protease enzyme [65]. In that work, the authors have mentioned that 3CLpro from SARS-CoV-2 (PDB: 6LU7) share high structural similarity with 3CLpro from SARS-CoV (PDB: 1UK4); structural alignment of substrate-binding pockets of the two main proteases showed the RMSD of 0.99 A [65]. Therefore, based on that the two proteases possess similar binding modes of inhibitors [65]. The authors have thus compared and found good agreement of binding free energies of the lead compounds to the experimental value of the most potent inhibitor (ML188) of 3CLpro of SARS-CoV [65,66].

In the current research, LiGANN-SkeleDock-KDeep protocol is utilized to discover the lead compounds against the main protease of SARS-CoV-2 [63,67,68]. LiGANN, the deep-learning based multimodel generative neural network is used to generate *de novo* ligand design [63]. The three-dimensional ligand shapes complementary to the shape and chemical properties of the protease pocket of coronavirus were generated. Later, these shapes were decoded into SMILES strings which correspond to the correct molecular structures. The small molecules were docked; and the molecular interaction studies followed by the binding affinity calculations validated the findings. By this approach, potential binders have been designed which are novel for the Mpro protein target. A library of 91 compounds were generated**, among them 5 promising lead compounds were identified; compound "prot_mol00065" was identified as the best binder of the main protease. The binding free energies of the compounds range in between -5 to -10 kcal/mol. The predicted binding free energy of the "prot_mol00065" is -7.91 kcal/mol, which is in good correlation with the experimental value (ΔG = -7.98 kcal/mol) of the most potent inhibitor (ML188) of 3CLpro of SARS-CoV [66]. The results of the proposed study are also in good agreement with the work of Macchiagodena and co-workers [65].

Materials and Methods

All Tables and Figures are presented on Supplementary Information.

The co-crystallized structure of the main protease (Mpro) (viral protein) of coronavirus SARS-CoV-2 (2019-nCoV) with a peptide-like inhibitor (N3) was obtained from RCSB Protein

Data Bank (PDB ID: 6LU7) [21] – having 2.16 Å resolution. There are 306 amino acid residues (SER1 to GLN306) in PDB file of the crystal structure constituting chain A. However, another chain C comprising of a sequence length of 6 residues (02J, ALA, VAL, LEU, PJE and 010) together constitute a synthetic peptide-like inhibitor. These 6 residue names of chain C in the PDB file were manually changed to residue name LIG. The residues present in chain C were by default recognized as ATOM; therefore all residues of the chain C were manually changed from ATOM to HETATM.

The protein structure was then protonated and optimized by using ProteinPrepare web application of PlayMolecule from playmolecule.org [69,70]. ProteinPrepare utilizes an empirical method to compute protonation states of titratable/ionizable amino acid residues by using PROPKA 3.1 [71]. The PDB2PQR 2.1 [72] software used by ProteinPrepare adds missing atoms to the protein structure and performs the hydrogen bond optimization. The input PDB (PDB ID: 6LU7) was uploaded and pKa calculations were performed at default pH of 7.4. The crystallographic waters were retained and heteroatoms were also included during calculations. The computations were performed selecting all chains and the option to mutate non-standard residues was not availed thus leaving the non-standard residues unchanged. The protonated PDB file was further utilized for a structure-based *de novo* drug design tool LiGANN [63] built on multimodel generative neural-networks.

The molecular discovery web application tool LiGANN by PlayMolecule [63,69] was used to produce a library of shape-complementary molecules for binding pocket of the target protein. The protonated protein PDB file prepared in the previous section was uploaded. The three-dimensional box was automatedly defined over the location of protein where the first non-protein (ligand) molecule was present. The 3D-box can be placed at any position where the user is intended to generate potential ligands by adjusting x, y and z coordinates. In my case, I had fragments of 6 residues of peptide-like inhibitor/chain C; by default the box was centered over residue 1 of the peptide inhibitor. In order to avoid this issue, the centroid coordinates (x, y, z) of the whole peptide ligand (LIG) were provided to correctly position the 3D-box centered over the whole peptide-like inhibitor. The geometric center (centroid) of the ligand was computed by CHIMERA [73] by the following command:

define centroid mass false :LIG

The x, y, and z centroid coordinates obtained were -10.712, 12.411, and 68.831 respectively. After that, the default value of 10 was selected for both parameters: (i) ligand shape generations and (ii) decoding per shapes. After processing, a library of 91 compounds was generated in the SMILES [74] format. The output

also includes Gaussian cube files (a total of 50 files) of each chemical property such as acceptor, aromatic, donor, hydrophobic and occupancy for each shape (1-10).

The generated molecules in the SMILES format were docked in the protein's binding pocket by another tool of PlayMolecule named SkeleDock [67,69]. SkeleDock [67] performs scaffold-based molecular docking of the congeneric compounds series. SkeleDock requires a protonated protein PDB file, a template ligand PDB file, and a congeneric series in SMILES format as query molecules in csv two-column file format with unique "code" of each molecule and "SMILES" string. It works by finding a match between the template and the query and according to that correspondence it tethers/restraints the atoms of the query to acquire positions similar to the template. The protonated protein, template ligand N3 (LIG) and list of query molecules were inputted. The option to optimize with RDock [75] was selected to perform further refinement of the tethered poses obtained from SkeleDock method. The docking simulation was performed without scaffold-hopping to avoid local mismatches in the template-query alignment process. Tethering force was kept at default value of 1.0, to tether atoms of query molecules to their corresponding template atoms. Default value of 6 for probe radius was used, which defines the docking grid. Docking results include SkeleDock poses without RDock refinement and final poses which have undergone through RDock optimization; the docked poses of ligands obtained by both methods were in SDF formats. A total of 88 molecules were obtained after docking; 3 molecules were discarded. The output has two types of scores: one is SkeleDock score while the other is RDock score. The SkeleDock is the score assigned by SkeleDock, which evaluates the alignment of query molecule on to the template molecule. In SkeleDock, the score is only a positive integer; a larger number is relatively a better score, which demonstrates that query molecule aligns well with the template. Whereas, the RDock docking score is an estimate of binding energy of the final pose. More negative numerals are better which represent stronger binding. The criteria of both of the docking scores are different because one is based on molecular structure alignment while the other is an estimate of the binding energy; therefore limited correlation in docking scores obtained by the two criteria is quite rational and they thus require to be considered autonomously.

The docked poses of all ligands in SDF formats were visualized by PYMOL [76] software. The OPEN BABEL [77,78] software was used to convert compounds from SDF format to PDB format. The PDB files were visualized by VMD [79] and CHIMERA [73] softwares. The following single command was used to convert all SDF files into PDB files:

obabel *.sdf -opdb -m

The protein-ligand interactions were analyzed and two-dimensional diagram was generated by using PlexView module of PlayMolecule web application [69]. PlexView requires a protonated protein PDB file and a ligand mol2 file as input. The output includes protein-ligand interactions such as hydrogen bond and pi-pi stacking. However, besides PlexView; polar-contacts, hydrogen-bonds and all possible protein-ligand interactions were determined by using CHIMERA and PYMOL softwares.

The binding affinity of the set of ligands were also predicted by using KDeep [68] module of PlayMolecule which uses deep convolutional neural networks (DCNNs). Each ligand structure is voxelized into pharmacophoric properties namely aromatic, hydrophobic, hydrogen bond acceptor and donor, total excluded volume, metallic, negative and positive ionizable features. The input is used for DCNN model, which is already trained by employing PDB bind database. The input required for computing protein-ligand affinity is a pre-processed protonated protein PDB file obtained from ProteinPrepare and a single SDF file for all ligands. The following command was used to merge all 88 docked ligands (present in SDF format generated by SkeleDock) in a single output file (all_ligands.sdf):

obabel *.sdf -O all_ligands.sdf

The output generated in tabular form by KDeep comprising the molecular weight of each ligand in (g/mol), dissociation constant (pK_d) and free energy of binding ΔG in (kcal/mol).

Finally, the pharmacokinetic properties of the compounds were calculated by the pkCSM – pharmacokinetics web-server (http://biosig.unimelb.edu.au/) [80]. The IUPAC nomenclature of the five best compounds were generated via online webserver Convert-molecule file format conversion (via ChemAxon JChem version 19.3.0) by giving SMILES string as an input format [81]. The names generated were verified by using another web interface OPSIN: Open Parser for Systematic IUPAC nomenclature [82].

Results and Discussion

The protein structure of the main protease (Mpro) of coronavirus SARS-CoV-2 (2019-nCoV) from PDB ID: 6LU7 was protonated; and hydrogen-bond optimization was performed by ProteinPrepare web application of PlayMolecule. None of the titratable residues having pKa close to the pH 7.4 were found. The output files of ProteinPrepare PlayMolecule such as protonation table and protonation diagram are represented as **Table 1**** and **Figure 1**** respectively.

The protonated PDB file was used for the drug design tool LiGANN. LiGANN generated shape-complementary small molecule library of 92 compounds for binding pocket of the target protein as shown in the **Figure 2****.

The SkeleDock [67] tool performed scaffold-based molecular docking of the library of 91 congeneric compounds obtained by LiGANN [63], in the binding pocket of the main protease (Mpro) of coronavirus SARS-CoV-2 (2019-nCoV). The **Figure 3**** represents 88 docked compounds at the active site of the protein. The 88 docked poses of ligands were obtained separately by both SkeleDock with and without RDock refinement. The docked poses obtained by SkeleDock with RDock refinement were considered as final poses and the best ligand selection was based on RDock score. I selected five top-ranked ligands having high RDock scores and considered them as promising candidates. The **Figure 4**** shows the three-dimensional structures and atom identifiers of the top five ligands: prot_mol00065, prot_mol00012, prot_mol00002, prot_mol00037 and prot_mol00075.

The compounds' IUPAC nomenclature of the five best binders were generated via online webserver Convert- molecule file format conversion (via ChemAxon JChem version 19.3.0) by giving SMILES string as an input format of the compounds [81]. The names generated were verified by using another web interface OPSIN: Open Parser for Systematic IUPAC nomenclature [82]. Following are the chemical (IUPAC) names of these five compounds:

compound 65 (prot_mol00065):
N-butyl-N-ethyl-5-(1H-pyrazol-4-yl)-4-[(thiophen-2-yl)methyl]-4H-1,2,4-triazol-3-amine

compound 12 (prot_mol00012):
N-methyl-N-(pentan-2-yl)-5-(1H-pyrazol-3-yl)-4-[(thiophen-2-yl)methyl]-4H-1,2,4-triazol-3-amine

compound 2 (prot_mol00002):
1-{3-[(4-methyl-1H-indol-3-yl)methyl]-1,2,4-oxadiazol-5-yl}pentan-1-amine

compound 37 (prot_mol00037):
N,1-dimethyl-N-(2-{3-[(pyridin-4-yl)methyl]-1,2,4-oxadiazol-5-yl}phenyl)piperidin-4-amine

compound 75 (prot_mol00075):
4-{3-[(5-methyl-1H-imidazol-4-yl)methyl]-1,2,4-oxadiazol-5-yl}-N-[(1-propylcyclopropyl)methyl]cyclohexa-1,3-dien-1-amine.

The ligand code "prot_mol00065" was top-ranked with RDock score of -19.28. The other four ligands amongst the top five ligands were prot_mol00012, prot_mol00002, prot_mol00037 and prot_mol00075 with RDock scores of -15.68, -13.81, -13.30 and -13.04 respectively. The potential interactions between the protein and top-ranked ligands are represented in the **Table 2**** obtained by PlexView application tool of PlayMolecule; however, in order to get a clear view of polar-contacts and hydrogen-bonds, all possible protein-ligand interactions were investigated by using CHIMERA and PYMOL softwares as discussed in the following section. The two-

dimensional protein-ligand interaction diagrams generated by PlexView application tool of PlayMolecule are represented in **Figure 10**[**].

The compound "prot_mol00065" interacts with the pocket residues of the main protease substrate binding site: HIS41, MET49, PHE140, LEU141, ASN142, GLY143, SER144, CYS145, HIS163, MET165, GLU166, HIS172, ASP187, ARG188 and GLN189 as reported in the literature [65]. The ligand "prot_mol00065" formed strong hydrogen bond interaction with backbone oxygen of PHE140 via its nitrogen (N14), with a distance of 2.98 Å between the heavy atoms. Another potential hydrogen bond interaction was observed between nitrogen (N13) of the ligand and side chain oxygen OG of SER144; found at a distance of 2.82 Å. The nitrogen (N9) atom of the ligand "prot_mol00065" was found at distances 3.4 Å and 3.8 Å from backbone nitrogen of GLU166 and MET165 respectively. The nitrogen atom (N14) of the ligand is at 4 Å from the side chain OE2 atom of GLU166. On the other hand, nitrogen (N14) atom of the ligand might form a hydrogen bond interaction with backbone oxygen of LEU141; both are 3.6 Å apart. The backbone nitrogen atom of LEU141 is 3.4 Å far from nitrogen (N13) of the ligand. The nitrogen (N13) of the ligand in turn found at a distance of 3.9 Å from the side chain ND1 atom of HID163. These potential interactions of the ligand "prot_mol00065" are in good agreement with the existing data [65], might be responsible for the strong binding of the ligand at the active/catalytic site of the protein. The binding mode of ligand "prot_mol00065" in the catalytic site of protein is shown in the **Figure 5**[**].

The **Table 3**[**] depicts protein-ligand affinities calculated by KDeep module of PlayMolecule. In the **Table 4**[**] SkeleDock and RDock scores of 88 docked ligands of the congeneric series have been reported. **Table 5** shows the SMILES and codes along with the pharmacokinetic properties of all the compounds. The predicted binding affinity (ΔG) of the top-ranked ligand code "prot_mol00065" (RDock score: -19.28) was found to be -7.91 kcal/mol and pK_d of 5.86. This compound has a molecular weight of 330.16 g/mol. The poses generated by SkeleDock having RDock score greater than 0 were not considered further for KDeep scoring. This is because the more negative numerals in RDock score are better representing the best-fit and the stronger binding; in the reverse scenario if the RDock score is positive, the peculiarity of best-fitting of the ligand inside the pocket is compromised. As for instance, the binding affinity predicted by KDeep of prot_mol00071 is -10.14 kcal/mol, which is the most negative ΔG value in the series, but since its RDock score is a positive value, therefore it was ruled out.

The predicted binding affinity of the other four ligands: prot_mol00012, prot_mol00002, prot_mol00037 and prot_mol00075 are -7.69 kcal/mol with pK_d of 5.70, -7.95 kcal/mol with

pK_d of 5.89, -8.16 kcal/mol with pK_d of 6.04, and -8.26 kcal/mol with pK_d of 6.12.

The binding pattern of the ligand "prot_mol00012" in the active site of the protein is illustrated in the **Figure 6****. The nitrogen N14 of the ligand "prot_mol00012" is found at a distance of 3.55 Å from the backbone oxygen of ARG188.

The ligand "prot_mol00002" with the active site residues is depicted in **Figure 7****. The nitrogen N5 of the ligand "prot_mol00002" is found at a distance of 2.7 Å from the backbone oxygen of HIE164 forming hydrogen bond interaction. The N5 nitrogen of the ligand is in turn 2.9 Å far from the side chain nitrogen NE2 of HID41. The nitrogen N12 of the ligand is 2.8 Å away from the backbone oxygen of PHE140. On the other hand, the nitrogen N12 of the ligand is present at a distance of 3.2 Å from the side chain oxygen OE2 of GLU166. The backbone nitrogen (N) atom of GLU166 is found at distances of 3.4 and 3.6 Å from nitrogen N20 and oxygen O21 of the ligand respectively.

The **Figure 8**** is the pictorial representation of the binding conformation of ligand "prot_mol00037" inside the pocket. The ligand "prot_mol00037" formed strong hydrogen bond interactions with CYS145. The backbone nitrogen (N) of CYS145 is at 2.7 Å and 2.8 Å distances from oxygen O24 and nitrogen N23 of the ligand "prot_mol00037" respectively. The oxygen O24 is 2.9 Å and 4.0 Å away from the backbone nitrogen (N) and side chain oxygen OG of SER144. The backbone nitrogen (N) atom of GLY143 is found at a distance of 3.1 Å from the oxygen O24 of the ligand. The backbone nitrogen (N) of GLY143 is in turn 3.1 Å and 3.9 Å away from N23 and N14 nitrogen atoms of the ligand. The nitrogen N23 of the ligand "prot_mol00037" is at a distance of 3.1 Å from the backbone nitrogen of SER144.

The binding mode of ligand **"**prot_mol00075" inside the active site of the protein is exhibited in the **Figure 9****. The nitrogen N5 of the ligand "prot_mol00075" formed strong hydrogen bond interaction with backbone oxygen (O) of ARG188; both heavy atoms were found at a distance of 3.3 Å. The backbone nitrogen (N) atom of MET165 formed polar contacts with nitrogen N20 and oxygen O21 of the ligand at 3.8 Å and 3.4 Å distances. The nitrogen N15 of the ligand is found at 3.7 Å and 3.9 Å distances away from the side chain oxygen OG of SER144 and backbone nitrogen (N) of Leu141. The backbone oxygen (O) of PHE140 formed close contact with nitrogen N17 of the ligand at a distance of 3.3 Å. The nitrogen N17 of the ligand, in turn, formed polar contacts with OE1 and OE2 of GLU166 residue at 4 Å and 3 Å distances respectively. The nitrogen N20 of the ligand formed a hydrogen bond interaction with SG donor atom of CYS145; both heavy atoms were 2.9 Å distance apart.

The high docking scores, well-estimated binding affinities and key interactions of the top-five compounds: "prot_mol00065",

"prot_mol00012", "prot_mol00002", "prot_mol00037" and "prot_mol00075", describe the viability of the best-fit and the strong binding of these ligand at the active site of the protein; consequently revealing the potential of these compounds to serve as promising inhibitors of main protease (Mpro) enzyme of coronavirus SARS-CoV-2 (2019-nCoV).

In the current work, LiGANN-SkeleDock-KDeep protocol is used to optimize possible lead compounds against the Mpro enzyme of SARS-CoV-2. LiGANN generated *de novo* ligand design. The 3D ligand shapes complementary to the protease pocket of coronavirus were decoded into SMILES strings of the correct molecular structures. Molecular docking of small molecules and analysis of binding modes is followed by the binding affinity estimations. In this way, a library of 91 small compounds against the target main protease (Mpro) enzyme of SARS-CoV-2 was generated. Among the designed molecular series, 5 potential candidates were identified; compound "prot_mol00065" was identified as the best binder of the main protease. The binding free energies of the compounds range in between -5 to -10 kcal/mol. The predicted binding free energy of the compound "prot_mol00065" is -7.91 kcal/mol, which is in good agreement with the experimental value (delta ΔG = -7.98 kcal/mol) of the most potent inhibitor (ML188) of 3CLpro of SARS-CoV reported in literature [66]. The results of the proposed study are also in good correlation with the work of Macchiagodena and co-workers [65].

Conclusion

In conclusion, I have generated a library of compounds for substrate-binding pocket of the main protease (Mpro) of coronavirus SARS-CoV-2 (2019-nCoV) by using structure-based drug design tool LiGANN by PlayMolecule which is built upon multimodal generative neural-networks. The scaffold-based docking of the generated compound series at the protein's active site was performed by using SkeleDock by PlayMolecule. The binding poses were evaluated on the bases of docking scores given by RDock. The binding affinity of the protein and docked ligands were estimated by KDeep predictor from PlayMolecule based on deep convolutional neural network. I inferred that the top-ranked compounds could have the potential to inhibit the main protease (Mpro) enzyme of coronavirus SARS-CoV-2 (2019-nCoV), so as to control viral replication and maturation. This study would be helpful in developing effective antiviral agents against the COVID-19 especially when the pandemic control is challenging due to various variants' emergence.

** *All Tables and Figures are presented on Supplementary Information*

References

[1] WHO says Covid-19 is 10 times more deadly than swine flu [Internet]. Available from: https://www.france24.com/en/20200413-who-says-covid-19-is-10-times-more-deadly-than-swine-flu

[2] Coronavirus 10 times more deadly than swine flu, says WHO | The Independent [Internet]. Available from: https://www.independent.co.uk/news/health/coronavirus-deaths-cases-latest-swine-flu-who-a9462896.html

[3] Hassan S, Sheikh F, Jamal S, Ezeh J, Akhtar A. Coronavirus (COVID-19): A Review of Clinical Features, Diagnosis, and Treatment. Cureus 2020;12(3):e7355.

[4] Novel coronavirus: What we know so far [Internet]. Available from: https://www.medicalnewstoday.com/articles/novel-coronavirus-your-questions-answered

[5] Severe acute respiratory syndrome coronavirus 2. Wikipedia. Available from: https://en.wikipedia.org/wiki/Severe_acute_respiratory_syndrome_coronavirus_2

[6] Lim Y, Ng Y, Tam J, Liu D. Human Coronaviruses: A Review of Virus–Host Interactions. Diseases 2016;4(4):26.

[7] Fung T, Liu D. Human Coronavirus: Host-Pathogen Interaction. Annual Review of Microbiology 2019;73(1):529-557.

[8] Cui J, Li F, Shi Z. Origin and evolution of pathogenic coronaviruses. Nature Reviews Microbiology 2018;17(3):181-192.

[9] Singhal T. A Review of Coronavirus Disease-2019 (COVID-19). The Indian Journal of Pediatrics 2020;87(4):281-286.

[10] Jones B, Grace D, Kock R, Alonso S, Rushton J, Said M, McKeever D, Mutua F, Young J, McDermott J, Pfeiffer D. Zoonosis emergence linked to agricultural intensification and environmental change. Proceedings of the National Academy of Sciences 2013;110(21):8399-8404.

[11] Destoumieux-Garzón D, Mavingui P, Boetsch G, Boissier J, Darriet F, Duboz P, Fritsch C, Giraudoux P, Le Roux F, Morand S, Paillard C, Pontier D, Sueur C, Voituron Y. The One Health Concept: 10 Years Old and a Long Road Ahead. Frontiers in Veterinary Science 2018;5.

[12] One Health [Internet]. Available from: https://www.who.int/news-room/q-a-detail/one-health

[13] Ramphul K, Mejias S. Coronavirus Disease: A Review of a New Threat to Public Health. Cureus 2020;12(3):e7276.

[14] Fehr A, Perlman S. Coronaviruses: An Overview of Their Replication and Pathogenesis. Coronaviruses 2015:1-23.

[15] Wrapp D, Wang N, Corbett K, Goldsmith J, Hsieh C, Abiona O, Graham B, McLellan J. Cryo-EM Structure of the 2019-nCoV Spike in the Prefusion Conformation. bioRxiv Prepr Serv Biol 2020.

[16] Wrapp D, Wang N, Corbett K, Goldsmith J, Hsieh C, Abiona O, Graham B, McLellan J. Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. Science 2020;367(6483):1260-1263.

[17] Hao P, Zhong W, Song S, Fan S, Li X. Is SARS-CoV-2 originated from laboratory? A rebuttal to the claim of formation via laboratory recombination. Emerging Microbes & Infections 2020;9(1):545-547.

[18] Andersen K, Rambaut A, Lipkin W, Holmes E, Garry R. The proximal origin of SARS-CoV-2. Nature Medicine 2020;26(4):450-452.

[19] Zhang T, Wu Q, Zhang Z. Probable Pangolin Origin of SARS-CoV-2 Associated with the COVID-19 Outbreak. Current Biology 2020;30(8):1578.

[20] Coronavirus Disease (COVID-19) Situation Reports [Internet]. Available from: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports

[21] Jin Z, Du X, Xu Y, Deng Y, Liu M, Zhao Y, Zhang B, Li X, Zhang L, Peng C, Duan Y, Yu J, Wang L, Yang K, Liu F, Jiang R, Yang X, You T, Liu X, Yang X, Bai F, Liu H, Liu X, Guddat L, Xu W, Xiao G, Qin C, Shi Z, Jiang H, Rao Z, Yang H. Structure of Mpro from SARS-CoV-2 and discovery of its inhibitors. Nature 2020;582(7811):289-293.

[22] Ashour H, Elkhatib W, Rahman M, Elshabrawy H. Insights into the Recent 2019 Novel Coronavirus (SARS-CoV-2) in Light of Past Human Coronavirus Outbreaks. Pathogens 2020;9(3):186.

[23] Kumar A, Choudhir G, Shukla S, Sharma M, Tyagi P, Bhushan A, Rathore M. Identification of phytochemical inhibitors against main protease of COVID-19 using molecular modeling approaches. Journal of Biomolecular Structure and Dynamics 2020;39(10):3760-3770.

[24] Banerjee R, Perera L, Tillekeratne L. Potential SARS-CoV-2 main protease inhibitors. Drug Discovery Today 2021;26(3):804-816.

[25] Hofmarcher M, Mayr A, Rumetshofer E, Ruch P, Renz P, Schimunek J, Seidl P, Vall A, Widrich M, Hochreiter S, Klambauer G. Large-Scale Ligand-Based Virtual Screening for SARS-CoV-2 Inhibitors Using Deep Neural Networks. SSRN Electronic Journal 2020.

[26] Yassine R, Makrem M, Farhat F. Active Learning and the Potential of Neural Networks Accelerate Molecular Screening for the Design of a New Molecule Effective against SARS-CoV-2. BioMed Research International 2021;2021:1-14.

[27] Jacobs S, Moon T, McLoughlin K, Jones D, Hysom D, Ahn D, Gyllenhaal J, Watson P, Lightstone F, Allen J, Karlin I, Van Essen B. Enabling rapid COVID-19 small molecule drug design through scalable deep learning of generative models. The International Journal of High Performance Computing Applications 2021;35(5):469-482.

[28] Hsieh K, Wang Y, Chen L, Zhao Z, Savitz S, Jiang X, Tang J, Kim Y. Drug repurposing for COVID-19 using graph neural network and harmonizing multiple evidence. Scientific Reports 2021;11(1):1-13.

[29] FDA Approves First COVID-19 Vaccine | FDA 2021 [Internet]. Available from: https://www.fda.gov/news-events/press-announcements/fda-approves-first-covid-19-vaccine

[30] Wang J, Zhang Y, Nie W, Luo Y, Deng L. Computational anti-COVID-19 drug design: progress and challenges. Briefings in Bioinformatics 2021;23(1):bbab484.

[31] Dolgin E. COVID vaccine immunity is waning — how much does that matter?. Nature 2021;597(7878):606-607.

[32] Alsharif MH, Alsharif YH, Albreem MA, Jahid A, Solyman AAA, Yahya K, et al. Application of machine intelligence technology in the detection of vaccines and medicines for SARS-CoV-2. Eur Rev Med Pharmacol Sci. 2020;24(22):11977–11981.

[33] Mollica L, Decherchi S, Zia S, Gaspari R, Cavalli A, Rocchia W. Kinetics of protein-ligand unbinding via smoothed potential molecular dynamics simulations. Scientific Reports 2015;5(1):11539.

[34] Davenport T, Kalakota R. The potential for artificial intelligence in healthcare. Future Healthcare Journal 2019;6(2):94-98.

[35] Singh R, Bhardwaj V, Das P, Purohit R. A computational approach for rational discovery of inhibitors for non-structural protein 1 of SARS-CoV-2. Computers in Biology and Medicine 2021;135:104555.

[36] Bhardwaj V, Singh R, Sharma J, Rajendran V, Purohit R, Kumar S. Bioactive Molecules of Tea as Potential Inhibitors for RNA-Dependent RNA Polymerase of SARS-CoV-2. Frontiers in Medicine 2021;8:684020.

[37] Zhai T, Zhang F, Haider S, Kraut D, Huang Z. An Integrated Computational and Experimental Approach to Identifying Inhibitors for SARS-CoV-2 3CL Protease. Frontiers in Molecular Biosciences 2021;8:267.

[38] de Vries M, Mohamed A, Prescott R, Valero-Jimenez A, Desvignes L, O'Connor R, Steppan C, Devlin J, Ivanova E, Herrera A, Schinlever A, Loose P, Ruggles K, Koralov S, Anderson A, Binder J, Dittmann M. A Comparative Analysis of SARS-CoV-2 Antivirals Characterizes 3CL pro Inhibitor PF-00835231 as a Potential New Treatment

for COVID-19. Journal of Virology 2021;95(10):e01819–20.

[39] Khan A, Ali S, Khan M, Saleem S, Ali A, Suleman M, Babar Z, Shafiq A, Khan M, Wei D. Combined drug repurposing and virtual screening strategies with molecular dynamics simulation identified potent inhibitors for SARS-CoV-2 main protease (3CLpro). Journal of Biomolecular Structure and Dynamics 2020;39(13):4659-4670.

[40] Khan A, Heng W, Wang Y, Qiu J, Wei X, Peng S, Saleem S, Khan M, Ali S, Wei D. In silico and in vitro evaluation of kaempferol as a potential inhibitor of the SARS-CoV-2 main protease (3CLpro). Phytotherapy Research 2021;35(6):2841-2845.

[41] Sharma P, Bansal M, Sethi A, Poonam , Pena L, Goel V, Grishina M, Chaturvedi S, Kumar D, Rathi B. Computational methods directed towards drug repurposing for COVID-19: advantages and limitations. RSC Advances 2021;11(57):36181-36198.

[42] Muratov E, Amaro R, Andrade C, Brown N, Ekins S, Fourches D, Isayev O, Kozakov D, Medina-Franco J, Merz K, Oprea T, Poroikov V, Schneider G, Todd M, Varnek A, Winkler D, Zakharov A, Cherkasov A, Tropsha A. A critical overview of computational approaches employed for COVID-19 drug discovery. Chemical Society Reviews 2021;50(16):9121-9151.

[43] Hassanzadeganroudsari M, Ahmadi A, Rashidi N, Hossain M, Habib A, Apostolopoulos V. Computational Chemistry to Repurposing Drugs for the Control of COVID-19. Biologics 2021;1(2):111-128.

[44] Gurung A, Ali M, Lee J, Farah M, Al-Anazi K. An Updated Review of Computer-Aided Drug Design and Its Application to COVID-19. BioMed Research International 2021;2021:1-18.

[45] Singh R, Bhardwaj V, Sharma J, Purohit R, Kumar S. In-silico evaluation of bioactive compounds from tea as potential SARS-CoV-2 nonstructural protein 16 inhibitors. Journal of Traditional and Complementary Medicine 2022;12(1):35-43.

[46] Sharma J, Kumar Bhardwaj V, Singh R, Rajendran V, Purohit R, Kumar S. An in-silico evaluation of different bioactive molecules of tea for their inhibition potency against non structural protein-15 of SARS-CoV-2. Food Chemistry 2021;346:128933.

[47] Bhardwaj V, Singh R, Das P, Purohit R. Evaluation of acridinedione analogs as potential SARS-CoV-2 main protease inhibitors and their comparison with repurposed anti-viral drugs. Computers in Biology and Medicine 2021;128:104117.

[48] Bhardwaj V, Singh R, Sharma J, Rajendran V, Purohit R, Kumar S. Identification of bioactive molecules from tea plant as SARS-CoV-2 main protease inhibitors. Journal of Biomolecular Structure and Dynamics 2020;39(10):3449-3458.

[49] Cully M. A tale of two antiviral targets — and the COVID-19 drugs that bind them. Nature Reviews Drug Discovery 2021;21(1):3-5.

[50] Macchiagodena M, Pagliai M, Procacci P. Characterization of the non-covalent interaction between the PF-07321332 inhibitor and the SARS-CoV-2 main protease. Journal of Molecular Graphics and Modelling 2022;110:108042.

[51] Owen D, Allerton C, Anderson A, Aschenbrenner L, Avery M, Berritt S, Boras B, Cardin R, Carlo A, Coffman K, Dantonio A, Di L, Eng H, Ferre R, Gajiwala K, Gibson S, Greasley S, Hurst B, Kadar E, Kalgutkar A, Lee J, Lee J, Liu W, Mason S, Noell S, Novak J, Obach R, Ogilvie K, Patel N, Pettersson M, Rai D, Reese M, Sammons M, Sathish J, Singh R, Steppan C, Stewart A, Tuttle J, Updyke L, Verhoest P, Wei L, Yang Q, Zhu Y. An oral SARS-CoV-2 Mpro inhibitor clinical candidate for the treatment of COVID-19. Science 2021;374(6575):1586-1593.

[52] Halford B. How Pfizer scientists transformed an old drug lead into a COVID-19 antiviral. Chemical & Engineering News 2022;100(3):16-18.

[53] Beigel J, Tomashek K, Dodd L, Mehta A, Zingman B, Kalil A, Hohmann E, Chu H, Luetkemeyer A, Kline S,

Lopez de Castilla D, Finberg R, Dierberg K, Tapson V, Hsieh L, Patterson T, Paredes R, Sweeney D, Short W, Touloumi G, Lye D, Ohmagari N, Oh M, Ruiz-Palacios G, Benfield T, Fätkenheuer G, Kortepeter M, Atmar R, Creech C, Lundgren J, Babiker A, Pett S, Neaton J, Burgess T, Bonnett T, Green M, Makowski M, Osinusi A, Nayak S, Lane H. Remdesivir for the Treatment of Covid-19 — Final Report. New England Journal of Medicine 2020;383(19):1813-1826.

[54] Popova M, Isayev O, Tropsha A. Deep reinforcement learning for de novo drug design. Science Advances 2018;4(7):aap7885.

[55] Santana M, Silva-Jr F. De novo design and bioactivity prediction of SARS-CoV-2 main protease inhibitors using recurrent neural network-based transfer learning. BMC Chemistry 2021;15(1):1-20

[56] Jin W, Stokes J, Eastman R, Itkin Z, Zakharov A, Collins J, Jaakkola T, Barzilay R. Deep learning identifies synergistic drug combinations for treating COVID-19. Proceedings of the National Academy of Sciences 2021;118(39):e2105070118.

[57] Wang L, Zhang Y, Wang D, Tong X, Liu T, Zhang S, Huang J, Zhang L, Chen L, Fan H, Clarke M. Artificial Intelligence for COVID-19: A Systematic Review. Frontiers in Medicine 2021;8:1457.

[58] Xu Z, Su C, Xiao Y, Wang F. Artificial intelligence for COVID-19: battling the pandemic with computational intelligence. Intelligent Medicine 2022;2(1):13-29.

[59] Amilpur S, Bhukya R. Predicting novel drug candidates against Covid-19 using generative deep neural networks. Journal of Molecular Graphics and Modelling 2022;110:108045.

[60] Wang S, Sun Q, Xu Y, Pei J, Lai L. A transferable deep learning approach to fast screen potential antiviral drugs against SARS-CoV-2. Briefings in Bioinformatics 2021;22(6):bbab211.

[61] Chenthamarakshan V, Das P, Hoffman SC, Strobelt H, Padhi I, Lim KW, et al. CogMol: Target-Specific and Selective Drug Design for COVID-19 Using Deep Generative Models. Adv Neural Inf Process Syst 2020.

[62] Abdel-Basset M, Hawash H, Elhoseny M, Chakrabortty R, Ryan M. DeepH-DTA: Deep Learning for Predicting Drug-Target Interactions: A Case Study of COVID-19 Drug Repurposing. IEEE Access 2020;8:170433-170451.

[63] Skalic M, Sabbadin D, Sattarov B, Sciabola S, De Fabritiis G. From Target to Drug: Generative Modeling for the Multimodal Structure-Based Ligand Design. Molecular Pharmaceutics 2019;16(10):4282-4291.

[64] Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A, Bengio Y. Generative adversarial networks. Communications of the ACM 2020;63(11):139-144.

[65] Macchiagodena M, Pagliai M, Procacci P. Identification of potential binders of the main protease 3CLpro of the COVID-19 via structure-based ligand design and molecular modeling. Chemical Physics Letters 2020;750:137489.

[66] Jacobs J, Grum-Tokars V, Zhou Y, Turlington M, Saldanha S, Chase P, Eggler A, Dawson E, Baez-Santos Y, Tomar S, Mielech A, Baker S, Lindsley C, Hodder P, Mesecar A, Stauffer S. Discovery, Synthesis, And Structure-Based Optimization of a Series of N-(tert-Butyl)-2-(N-arylamido)-2-(pyridin-3-yl) Acetamides (ML188) as Potent Noncovalent Small Molecule Inhibitors of the Severe Acute Respiratory Syndrome Coronavirus (SARS-CoV) 3CL Protease. Journal of Medicinal Chemistry 2013;56(2):534-546.

[67] Varela-Rial A, Majewski M, Cuzzolin A, Martínez-Rosell G, De Fabritiis G. SkeleDock: A Web Application for Scaffold Docking in PlayMolecule. Journal of Chemical Information and Modeling 2020;60(6):2673-2677.

[68] Jiménez J, Škalič M, Martínez-Rosell G, De Fabritiis G. KDEEP: Protein–Ligand Absolute Binding Affinity Prediction via 3D-Convolutional Neural Networks. Journal of Chemical Information and Modeling 2018;58(2):287-296.

[69] PlayMolecule | Acellera [Internet]. Available from: https://www.acellera.com/products/playmolecule

[70] Martínez-Rosell G, Giorgino T, De Fabritiis G. PlayMolecule ProteinPrepare: A Web Application for Protein Preparation for Molecular Dynamics Simulations. Journal of Chemical Information and Modeling 2017;57(7):1511-1516.

[71] Olsson M, Søndergaard C, Rostkowski M, Jensen J. PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pKa Predictions. Journal of Chemical Theory and Computation 2011;7(2):525-537.

[72] Dolinsky T, Czodrowski P, Li H, Nielsen J, Jensen J, Klebe G, Baker N. PDB2PQR: expanding and upgrading automated preparation of biomolecular structures for molecular simulations. Nucleic Acids Research 2007;35(Web Server):W522-W525

[73] Pettersen E, Goddard T, Huang C, Couch G, Greenblatt D, Meng E, Ferrin T. UCSF Chimera?A visualization system for exploratory research and analysis. Journal of Computational Chemistry 2004;25(13):1605-1612.

[74] Weininger D. SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. Journal of Chemical Information and Modeling 1988;28(1):31-36.

[75] Ruiz-Carmona S, Alvarez-Garcia D, Foloppe N, Garmendia-Doval A, Juhos S, Schmidtke P, Barril X, Hubbard R, Morley S. rDock: A Fast, Versatile and Open Source Program for Docking Ligands to Proteins and Nucleic Acids. PLoS Computational Biology 2014;10(4):e1003571.

[76] PyMOL | pymol.org. Schrödinger L.L.C. (2015) The PyMOL Molecular Graphics System, Version 1.8. [Internet]. Available from: https://pymol.org/2

[77] O'Boyle N, Banck M, James C, Morley C, Vandermeersch T, Hutchison G. Open Babel: An open chemical toolbox. Journal of Cheminformatics 2011;3(1):33.

[78] Openbabel | PyPI [Internet]. Available from: https://pypi.org/project/openbabel

[79] Humphrey W, Dalke A, Schulten K. VMD: Visual molecular dynamics. Journal of Molecular Graphics 1996;14(1):33-38.

[80] Pires D, Blundell T, Ascher D. pkCSM: Predicting Small-Molecule Pharmacokinetic and Toxicity Properties Using Graph-Based Signatures. Journal of Medicinal Chemistry 2015;58(9):4066-4072.

[81] Convert [Internet]. Available from: https://datascience.unm.edu/tomcat/biocomp/convert

[82] Lowe D, Corbett P, Murray-Rust P, Glen R. Chemical Name to Structure: OPSIN, an Open Source Solution. Journal of Chemical Information and Modeling 2011;51(3):739-753.