Molecular modelling reveals eight novel druggable binding sites in SARS-CoV-2's spike protein

Ilke Ugur*, Antoine Marion*

Department of Chemistry, Middle East Technical University, 06800, Ankara, Turkey

e-mail: ilkeugur@metu.edu.tr; amarion@metu.edu.tr

Spike glycoprotein (S), one of the signature proteins of the SARS-CoV-2, initiates the membrane fusion and virus entry to the host cell. The S protein's key role in virus viability makes it an attractive candidate for drug design studies. Besides the recent structural characterization of the S protein, information fundamental to drug design such as possible binding sites or molecular fragments with high affinity towards the protein is unknown. We explored the druggability of this protein, focusing on its S1 and S2 domains. We performed virtual screening studies on both closed and open forms of the protein, using both cryo-EM structures and geometries obtained from molecular dynamic simulations. We targeted 20 distinct ligand binding centres with a set of about 9,000 molecules. Our docking calculations followed by molecular mechanics-based refinement of ligand/protein complexes led us to detect eight binding sites that were so far undocumented. By further focusing on a subset of approximately 1,000 approved and marketed drugs, we aimed at suggesting a new direction for drug repurposing strategies that were not considered so far. Within this approved set, our best hits include a number of antibacterial and antiviral drugs (e.g., Streptomycin, Nelfinavir), which were not yet investigated clinically in treating COVID-19. We also identified some molecules (e.g., folic acid, Famotidine) that were already suggested to be effective towards SARS-CoV-2, yet without molecular explanation. Our results also indicate a great affinity of SARS-CoV-2's S protein towards nucleoside analogues, either approved or experimental.

Introduction

The coronavirus disease 2019 (COVID-19) pandemic has struck the planet for a year and is dangerously impacting all sectors from public health to economics of nations worldwide. The response of the scientific community has naturally been immediate and commensurate with the severity of the threat caused by the spread and transmission of the virus responsible for COVID-19, i.e., the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). Teixeira da Silva *et al.* indeed reported more than 21,500 unique documents published on the topic in major online databases between January 1 and June 30, 2020.[1] As early as four months after the outbreak, promising drug candidates had already emerged from clinical trials, although no clear-cut conclusions could be drawn regarding their effectivity at the time.[2] Among other, repurposed antiviral drugs Hydroxychloroquine and the combination Lopinavir/Ritonavir, or the newly FDA approved Remdisivir were considered as promising treatments until recently. Although not peer-reviewed at the time we wrote the present report, large-scale WHO Solidarity trial results released October 15, 2020, indeed broke early hopes, showing little to no effects for these candidates.[3]

In the current context and given the recent developments regarding clinical trial results, the urge to identify new drugs for the effective treatment of COVID-19 is particularly strong, and computational drug repurposing appears a relevant route.[4] In this direction, a number of virtual screening studies have emerged in peer-reviewed and renowned online archives, aiming mostly at two major targets: i.e., the active site of the main protease[5-15] and the receptor-binding domain (RBD) of the spike (S) protein of SARS-CoV-2.[16-44] The 3-chymotrypsin-like protease (3CL-PR) and the RNA-dependent-RNA polymerase (RdRp) were also suggested to be druggable targets.[45] Inhibiting the main viral protease is a strategy that has been explored extensively to develop treatments for other viruses such as hepatitis C virus (HCV) and human immunodeficient virus (HIV)[46] and that could yield promising treatments against COVID-19 as well.[47] The general mechanism of action of viral proteases is fairly well-understood[48] and that of SARS-CoV-2's main protease has been uncovered recently via multiscale molecular modelling,[49] which could boost the development of specific therapeutics significantly in the near future.

Targeting the RBD of SARS-CoV-2's S protein is part of a different strategy that aims at blocking the fusion of the virus into the host cells. The fusion mechanism of coronaviruses (CoVs) involves sophisticated machinery recently summarised by Tang *et al.* in a comprehensive review.[50] The S protein of CoVs is a trimeric membrane-anchored protein, with each monomer made of two major domains, S1 and S2 described in detail in references [51] and [52]. Both domains have a specific role in the fusion mechanism. S1 forms the upper and outer part of the S protein and is made of two main subdomains, the N-terminal domain (NTD) and the receptor-binding domain (RBD). S1 is involved in the early stage of the fusion mechanism, i.e., binding of RBD to angiotensin-converting enzyme 2 (ACE2) presented on the host cell's surface. In the pre-fusion native state, the RBD of each monomer exists in two conformations, closed and open, the open one allowing exposure of the binding site of the S protein's RBD for anchoring to the cell's surface ACE2. It was suggested that the higher infection rate observed for SARS-CoV-2 compared to other CoVs might be due to a greater propensity of the RBD of SARS-CoV-2 to be found in the open state.[52] After anchoring, host's proteases cleave the S protein of CoVs at the S1/S2 junction, leaving only S2 attached to the virus' membrane in a so-called pre-fusion metastable state. S2 contains the machinery necessary for the entry of the virus in the host cell. The main sub-domains of S2 are, from N-terminal to C-terminal: fusion peptide (FP), heptad repeat 1 (HR1), central helix (HC), connector domain (CD), heptad repeat 2 (HR2), transmembrane domain (TM), and cytoplasmic tail (CT). After S1/S2 cleavage, the subdomains of S2 undergo a substantial conformational rearrangement, which aims at presenting the FP as the head of a spear made of HR1 reorganized as a straight and long helix, yielding the pre-hairpin intermediate form of the S protein. The fusion then proceeds in a few additional steps.

The complexity of the fusion mechanism makes the S protein of CoVs an appealing but challenging target for new treatments. Any drug binding strong enough to interfere with its conformational changes and general mechanism of action indeed holds the potential to block the fusion and prevent infection. This strategy has recently been explored by two other studies conducted in parallel to the present work. Toezler *et al.* reported the cryo-EM structure of SARS-CoV-2's S protein in complex with linoleic acid (LA) bound in a previously unknown hydrophobic pocket located at the interface between two RBDs and conserved over other CoVs.[53] The authors observed that the LA-bound structure of the S protein is mostly held in its closed state, which in turns prevents anchoring to ACE2 and eventually might prevent infection at an early stage of the fusion mechanism. Recently, Romeo *et al.* have conducted a virtual screening study to find suitable repurposed drugs that could bind to the central cavity of S protein of SARS-CoV-2.[54] By analogy with previous reports pointing out a druggable central cavity in the human respiratory syncytial virus' F-protein, Romeo *et al.* explicitly targeted the cavity at the bottom of SARS-CoV-2's S protein and identified drugs such as phthalocyanine and hypericin as potential candidates to the fusion at the stage posterior to the S1/S2 cleavage.

Besides identifying potential drug candidates that could bind effectively to the S protein of SARS-CoV-2, a significant challenge in the above-described strategy is to discover their most likely binding sites. In this direction, we have performed an extensive blind virtual screening in four structures of the S protein, sampling systematically 20 distinct binding centres with a set of 8,773 molecules containing 951 approved drugs. As a result, we report eight previously unidentified binding sites in SARS-CoV-2's S protein and point out several drug candidates that hold the potential to limit or prevent the transmission of the virus.

Methodology

Ligands library

The set of ligands used for virtual screening consisted of the entire library of 8773 molecules contained in the DrugBank database v. 5.1.5.[55] In this version of the set, 951 drugs were labelled as approved, either by the United States Food and Drug Administration (FDA) or by the European Medicines Agency (EMA). Some molecules have changed status in the newer version of the DrugBank database, such as Remdesivir that has been FDA approved since October 10, 2020. In the following, entries from the approved set are referred to as "drugs" while the terms "compounds" and "molecules" may refer to the entire set.

Each compound was extracted from the database and subject to a series of preparation steps to generate i) pdbqt files for docking and ii) molecular mechanics (MM) parameters for refinement of geometry and binding energy (see Docking and refinement). The protonation state of the molecules was assigned automatically with the open babel program[56] setting a pH of 7.4. Structures were exported as pdbqt and mol2 formats, the former being readily used in docking calculations. MM parameters were assigned via the antechamber program of AmberTools19[57], according to the General Amber Force Field (GAFF).[58] A small number of molecules could not be treated automatically by antechamber and were dealt with manually. These include silicon-containing compounds for which the bonded and Lennard-Johns parameters were adapted from data shared in Amber's archives (archive.ambermd.org/200804/0353.html). Notice that the choice of these parameters has little to no impact on the evaluation of binding energy. Atomic point charges play a more substantial role in the interaction between a ligand and its receptor and would require elaborate approaches such as restrained electrostatic potential atomic partial charges (RESP)[59] to

yield an accurate representation of quantum mechanical charge distributions. This approach is, however, hardly applicable to such a broad set of molecules, and the more approximate Gasteiger charges,[60] as assigned by the open babel program, were chosen instead for the present study. Other missing bonded parameters were generated with the parmchk2 program of AmberTools19 and saved as Amber's frcmod file format. The procedure eventually yielded for each of the 8,773 molecules a pdbqt file for docking and a mol2 file containing GAFF atom types and Gasteiger charges with a corresponding frcmod providing all necessary parameters for MM-based calculations. All files for the entire dataset may be made available upon request to the authors.

Receptors

The receptors were prepared based on the cryo-EM structures of closed and open states of SARS-CoV-2's spike protein with respective PDB-IDs 6vxx and 6vyb. Missing loops were reconstructed using the Modeller program v 9.23.[61] Each monomer contains 29 cysteine residues, all of them except C110 being identified by the DISULFIND web server [62] as involved in a disulphide bridge. All MM-based calculations used ff14SB parameters for the protein residues.[63] After manual inspection, all histidine residues were modelled in a neutral form with a proton on the epsilon nitrogen of their imidazole moiety. Amino acids with ionizable side chains were found to be in their physiologically standard protonation state by the PROPKA program v 3.1.[64-65] The structures were subject to a series of gas-phase energy minimizations to relax the geometry of added fragments and hydrogen atoms with position restraints set on backbone atoms and distance restraints to form disulphide bridges in remodelled regions. These final structures closely resemble the cryo-EM structures of the closed and open states and are further referred to as CS_00ns and OS_00ns, respectively.

CS_00ns and OS_00ns structures were solvated in a box of TIP4Pew[66] water molecules with sodium and chloride ions in a concentration of approximately 0.15 M and a buffer distance from protein atoms to the box's edges of 25 Å. After gradual temperature increase to 300 K and subsequent equilibration, the solvated structures were subject to a 25 ns-long molecular dynamics (MD) simulation in the NPT ensemble with restraints set on the connector domain (CT) of each monomer. MD simulations were carried out with the CUDA GPU-accelerated version of pmemd as part of the Amber18 distribution.[67] The last geometry of the closed and open states MDs were extracted and are further referred to as CS_25ns and OS_25ns, respectively.

Docking and refinement

All molecules from the set (8773) were docked in the closed and open state of the S protein as obtained from the cryo-EM structures (i.e., CS_00ns and OS_00ns). The drugs forming the approved set (951) were additionally docked in two structures extracted from MD simulations (i.e., CS_25ns and OS_25ns) to improve their sampling and increase the chances of identifying relevant drug candidates.

We selected twenty docking grid centres on the S protein (Figure 1a). Twelve of them were chosen to scan the druggability of the channel formed by the central helix of each monomer (orange-coloured beads). From the top of the central channel to its bottom, we regularly distributed these grid centres along the S protein's principal axis, separated from each other by 8 Å. The AutoLigand module of ADT[68] was applied to the S protein's closed state, which led us to identify six additional sites (green-coloured beads). We selected the last two grid centres by manual inspection of the protein to target the interface between RBD and NTD domains (blue-coloured beads). The coordinates of the docking grid centres are available upon request to the authors.

Virtual screening was performed by docking each molecule in the data set to each of the 20 above-described grid centres using AutoDock Vina.[69] The exhaustiveness parameter was set to its maximum value (i.e., 8). The search space was built as a cube of 18 Å edge length. Docking of all molecules (8773) in 20 grids of two receptors (i.e., CS_00ns and OS_00ns) and considering only the best pose for each attempt generated a total of 350,920 complexes. Similarly, the more extensive docking of 951 approved drugs in four receptors led to 76,080 ligand-protein complexes.

Despite the excellent performance of AutoDock Vina in predicting correct binding poses, it is recognized that binding energies are substantially less accurate.[70] Furthermore, as depicted in Figure 1b, nearly all the modelled complexes (~350,000) are found with a score ranging from 0.00 to -12.00 kcal mol^{-1}. The range of docking scores obtained for this large number of complexes is, therefore, too narrow to discriminate relevant from irrelevant binders objectively. In an attempt to increase the reliability of the predicted poses, a selection of complexes was subsequently subject to minimization and interaction energy calculation with a molecular mechanics (MM) force filed.

MM-based calculations were performed using the sander module of AmberTools19 with parameters defined as described above in the Ligand library and Receptors sections. The size of the systems was reduced to a sphere of 30 Å radius centred on the ligand's centre of mass to speed up the calculations and allow us to consider a significant number of complexes. Residues were excluded based on the distance from their centre of mass to the centre of the sphere. Capping acetyl and N-methyl groups or glycine residues were added to complement the truncated amino acid sequences. The outer layer of atoms located 10 Å away from the centre of the sphere was restrained during minimization steps with a force constant of 60.0 kcal mol^{-1} Å^{-2} (i.e., 30.0 in the input of sander). Minimizations were conducted in the gas-phase with non-bonded cut-offs set to 12 Å, in three sequential steps by relaxing: i) only the hydrogen atoms (5,000 optimization cycles), ii) hydrogen atoms and flexible protein residues (10,000 optimization cycles), and iii) hydrogen atoms, flexible protein residues and the ligand (10,000 optimization cycles). The interaction energy of each selected complex was evaluated by single-point energy calculations on the ligand, receptor, and complex in the optimized geometry of the complex. Solvation energy was included using the latest version of the generalized Born/surface area (GB/SA) implicit solvent model as implemented Amber (i.e., igb = 8). The interaction energy was calculated for each pose as the difference between the energy of the complex and the summed energies of the ligand and receptor, leading to the most favourable complexes having the most negative interaction energy.

Each of the ~76,000 complexes obtained by AutoDock Vina for the set of approved drugs was subject to the MM refinement protocol described above. For the full set of molecules, MM minimization of nearly 390,000 was deemed too computationally demanding (i.e., about 3 minutes for each complex on one CPU core), which led us to filter out a number of poses after the docking step. To this end, we identified the maximum value of the distribution of binding energies obtained from AutoDock Vina as -6.8 kcal mol^{-1} and systematically selected all complexes with a score better or equal to that value (i.e., ~145,000 complexes). Additionally, we randomly extracted 40,000 complexes with docking scores ranging from -6.8 to 0.0 kcal mol^{-1}, leading to a total of ~185,000 complexes that were subject to MM refinement for the complete set of molecules considered in this work. Complexes for which MM minimisations led to errors, due to too much of overlap between the ligand and complex atoms on the structure obtained from docking, were systematically disregarded. We then identified the maximum of the distribution of MM interaction energies at -22.0 kcal mol^{-1} (Figure 1c) and selected those complexes with an energy lower than that value for further analysis. After this last filtering step, we only kept the best pose for each molecule present in the final set of complexes successfully optimised (i.e., the best combination of receptor and docking grid centre). Overall, we report here the best binding pose for 4,135 molecules that

our procedure identified as good binders from the entire set as well as the best pose for 951 approved drugs. AutoDock Vina scores and MM interaction energies are tabulated and available upon request to the authors for the interested reader.

Binding Site Identification

The binding site identification procedure followed several structural clustering steps performed with the algorithms available in the scikit-learn python package.[71] The set of ~100,000 non-hydrogen atoms obtained from combining the structures of all 4,135 molecules was reduced to 2,000 representative atomic centres by applying the k-means clustering algorithms with all parameters set to their default, as implemented in scikit-learn. Four main agglomerates of atomic centres were identified by applying the dbscan algorithm with a gross epsilon value of 7.5 Å. Further breakdown of the main agglomerates with a finer epsilon value of 2.5 Å led to 15 clusters of representative atomic centres. In each of the main agglomerates, the atomic centres that were not assigned to a cluster were further subject to dbscan clustering with a larger epsilon value (i.e., 6-7 Å), yielding additional 13 clusters. We visually analysed these 28 clusters and reduced the number of distnincts binding site to 8.

Results & Discussion

Binding Sites

The DrugBank database contains 8773 molecules that form a versatile set of ligands in terms of structure (i.e., small cycles, macrocycles, linear chains), molecular weight, chemical fragments, and others properties (e.g., total charge, lipophilicity, hydrogen bonding character). We used this versatile set to address the druggability of SARS-CoV-2's S protein and to identify new binding sites that were so far unnoticed. We proceeded by scanning the protein's inner surface interior using 20 docking grid centres considering several conformations of the receptor and applying molecular mechanics refinement to a selected number of complexes. Out of ~350,000 complexes resulting from virtual screening, we reduced the dimensionality of our data in several steps and selected only the *better binders* of our molecular set, forming in a *better* complex with one of the receptors and bound to their corresponding *best* binding pocket. We finally selected the molecules from the entire set binding to the protein with an MM interaction energy lower or equal to -22.0 kcal mol^{-1} (see Figure 1c and Methodology). We ended up with a total of 4135 *good* binders.

The coordinates of the 4,135 molecules identified as good binders were used to define and locate the new binding sites that we discovered in SARS-CoV-2's S protein. Structural clustering analysis of the docked poses resulted in the identification of eight new binding sites (BS), which we labelled as BS A-H, and that are depicted in Figure 1d. The poses of the 951 complexes corresponding to the approved drugs were then assigned to a specific BS by a distance-based mapping of their non-hydrogen atoms to representative atomic centres. The coordinates of the representative atomic centres that define each BS are available upon request to the authors.
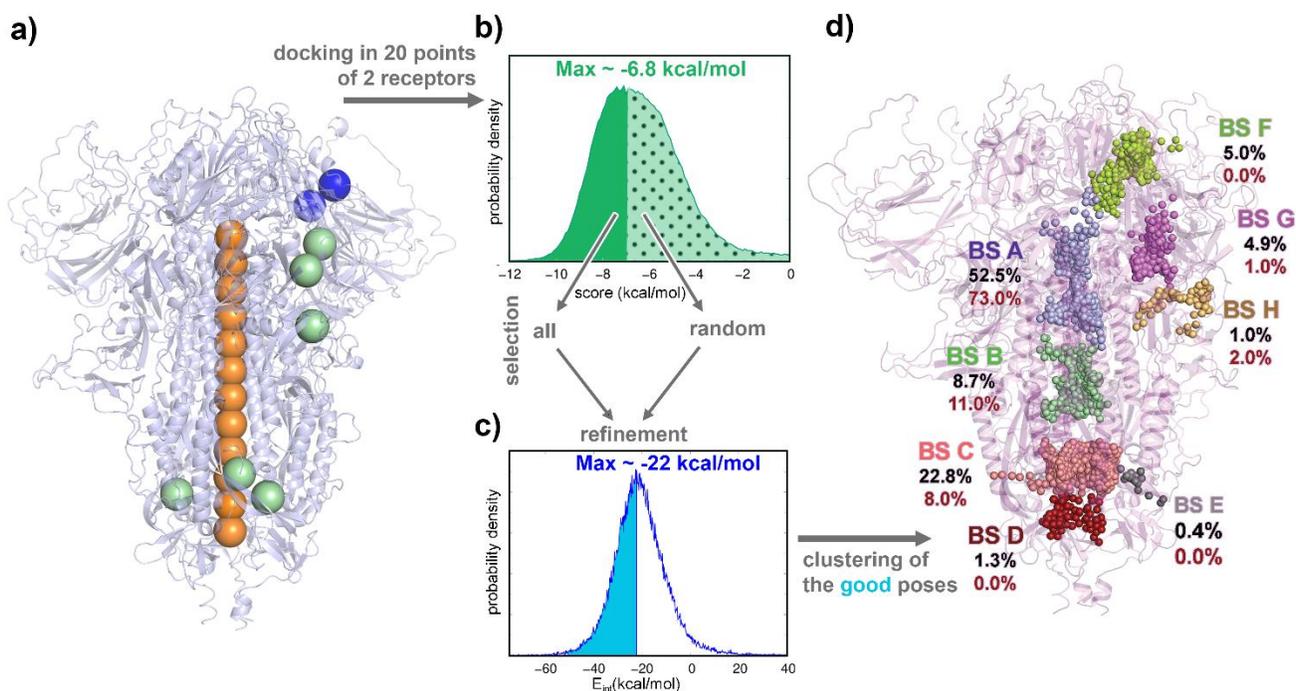
Figure 1: Summary of the virtual screening procedure and identification of new binding sites in SARS-CoV-2's spike protein. a) representation of the 20 docking grid centres in the closed conformation of SARS-CoV-2's S protein. Grid centres coloured in orange scan the interior channel of the spike protein formed by the central helix of the three monomers. Green-coloured beads represent centres identified by AutoLigand, while the blue-coloured ones were manually defined to target the interface between the N-terminal and the receptor-binding domains (NTD and RBD, respectively). b) Distribution of scores as obtained from the docking of 8,773 molecules in 20 grids for two conformations of the spike protein. c) Distribution of interaction energies after molecular mechanics-based refinement of ~145,000 ligand/protein complexes. d) Representation of the eight binding sites (BS A-H) determined by structural clustering of the 4,135 molecules identified as good binders to SARS-CoV-2's spike protein. The percent population of molecules assigned to each BS is given considering either all 4,135 binders or only the top 100 hits as coloured in black or in red, respectively.

Binding Site Characterization

Among the eight binding sites that we identified in the S protein, BS A and B are located in the upper part of the channel of the protein formed by the central helix of each monomer. BS C corresponds to the central cavity targeted by Romeo *et al.* in their recent virtual screening.[54] BS D and E can be regarded as extensions of BS C and are located in the lower part of the S protein. BSs A-E are found in the S2 domain of the S protein. Molecules binding in these BSs interact with different sub-domains of S2 and hold the potential to perturb the elaborate machinery involved in the later stage of the fusion mechanism. BS F is at the interface between the RBD and the NTD. Molecules binding in that location might fulfil a similar role as that of linoleic acid, as identified by Toezler *et al.*,[53] by increasing the interaction between the two domains and favouring the closed conformation of the protein's prefusion state. BSs G and H are mostly located in the S1 domain, below the RBD and NTD. Some of the molecules binding to BSs G and H interact with sub-domains of S2, thereby bridging between S1 and S2. Finally, we notice that BS H is located in the vicinity of the S1/S2 cleavage site, but a discussion regarding the potential impact of molecules binding in this location on the cleavage by host proteases would only be speculative. Hereafter,

7

we discuss and analyse a selection of ligands in binding in BSs A and B while the geometries and binding energies of all 4,135 good binders are available on request to the authors.
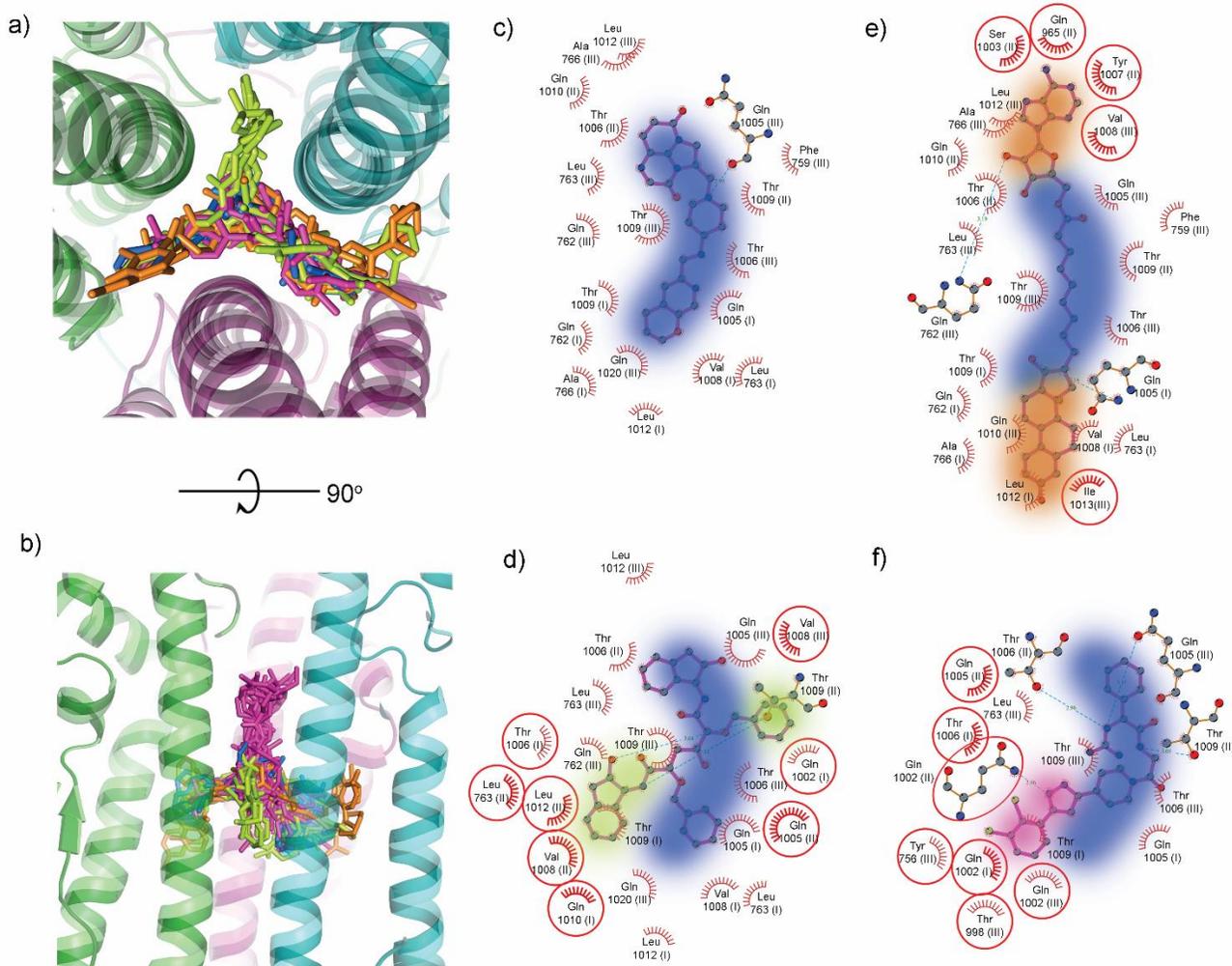

Figure 2: Illustration of the primary binding modes identified in the most populated binding site, BS A. a) The four groups of binding poses identified within BS A shown within the spike protein (cartoon representation with each colour of the cartoon showing one monomer). The view is set such that one looks from the bottom to the top of the protein along the axis of the central channel. b) Same as a), rotated by 90 degrees. c-f) Representation of ligand/receptor interactions for selected molecules from each group of binders, i.e., blue, green, orange, and pink in c, d, e, and f, respectively. Roman numbers (I, II, III) are used to label the different monomers. Ligands and hydrogen bond donor/acceptor residues are shown with ball and stick representation (C=grey, O=red, N=blue, Cl=green). Hydrophobic contacts are given by an arc with radiating spokes. Red circles around residues show the main differences compared to the reference, blue binding mode representative. Figures a) and b) were produced with Pymol; the rest was done with Ligplot$^+$.

Binding site A is the most populated with a hit rate of 53% over all the *good* complexes (4,135) and counts 73 molecules within the *best* 100 complexes (Figure 1d, blue-coloured beads). The average interaction energy of the latter 73 molecules is -60.27 kcal mol^{-1}, with ligands EM-1745 and dihydro-acarbose ranking first with -81.34 and -81.16 kcal mol^{-1}, respectively. This binding site is located approximately

8

in the middle of the spike trimer involving residues from the S2 domain. It consists of several hydrophobic residues (Leu, Ala and Val), accompanied by many polar Thr residues and charged Gln's. We visually analysed the positioning of the 30 best hits within this BS A and found four main binding modes (Figure 2a and 2b). We further examined their interactions by selecting one representative structure from each group of these main binding modes. Molecules within the blue-coloured group (Figure 2c) can be considered a scaffold binding mode common to all other groups, while the green-, orange- and pink-coloured groups are extensions of this mode. The addition of aromatic groups from the middle of the blue scaffold yields the green-coloured group, which adds new hydrophobic interactions with the receptor. The additional interactions, as compared to the blue scaffold, are circled in Figure 2d. Elongating the molecules in the blue group by adding an extra, mostly aromatic group to each end yields the orange binding mode (Figure 2e), which also brings a few new hydrophobic interactions with residue numbers Ser 1003, Tyr 1007, and Ile 1013. The fourth, pink-coloured group of binders (Figure 2f) consists of an extension of the blue scaffold with aromatic groups placed along the principal axis of the channel, which adds new interactions with the upper part of BS A. Although these extensions bring additional interactions with the protein, we cannot argue that molecules in yellow, orange or pink groups bind better than those in the blue one. Each molecule within each group indeed has its own interaction potential, and these brief analyses mostly aim at giving information about the three-dimensional shape of the binding site.
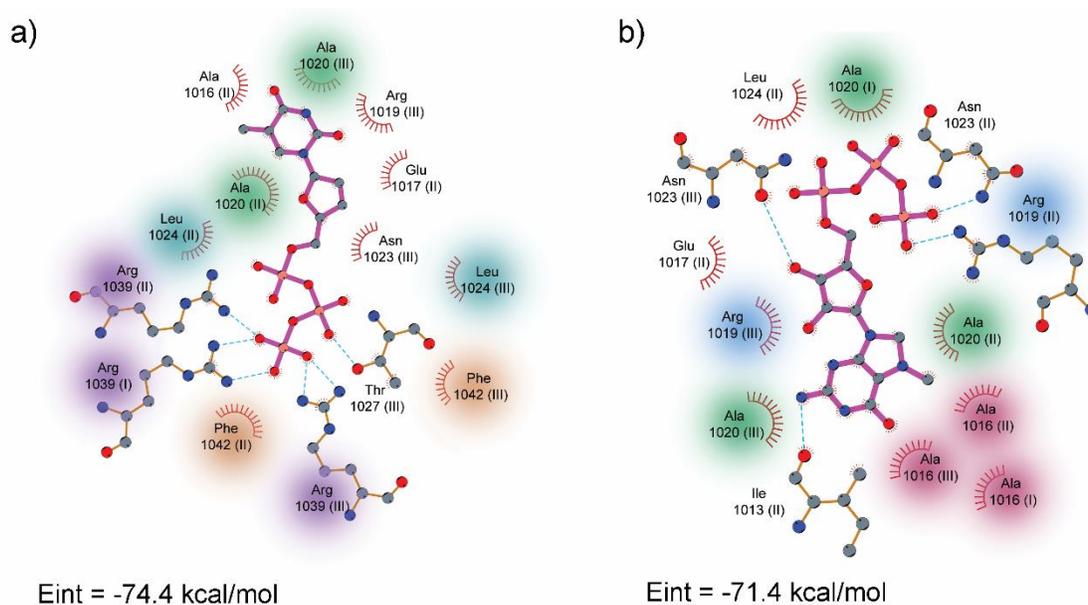

Eint = -74.4 kcal/mol

Eint = -71.4 kcal/mol

Figure 3: Protein-ligand interactions of two selected nucleotide analogue in binding site B: a) 2',3'-Dehydro-2',3'-Deoxy-Thymidine 5'-Triphosphate, and b) 7-Methyl-7,8-dihydroguanosine 5'-(tetrahydrogen triphosphate). Ligands and hydrogen bond donor/acceptor residues are shown with ball and stick representation (C=grey, O=red, N=blue, P=orange). Hydrophobic contacts are given by an arc with radiating spokes. Roman numbers (I, II, III) are used to label the different monomers. Identical residues coming from different monomers (I, II, III) are highlighted with the same colour.

Binding site B is the second most populated site considering the best 100 hits with 11 molecules (Figure 1d) and the third most populated with 8.7% of the entire set of molecules. The average interaction energy of molecules binding to BS B among the best 100 is -67.90 kcal mol^{-1}, with adenosine-5'-pentaphosphate being the best hit with -120.14 kcal mol^{-1}. The binding modes of the molecules within BS B are quite diverse and difficult to rationalize the same way as we did for BS A. However, a particular class of molecules shows a significant interaction within this binding site and caught our attention: i.e., nucleotide analogues that involve sugar-like and nucleobase-like fragments accompanied by a varying number of phosphate groups. Among the top 50 hits, we count 15 nucleotide analogues with 7 of them found in BS B. Nucleotide analogues bind to BS B with two main modes involving the symmetrical contribution of particular amino acids from two or three monomers as shown with two representative ligands in Figure 3. A shared feature of these two binding modes is that nucleobase-like and sugar-like fragments are stabilized by Ala 1020 from two or three monomers (green circles). The negatively charged phosphate groups find complementary interactions with arginine residues to form salt bridges, either in the lower part of BS B (i.e., Arg 1039 shown in purple in Figure 3a) or in the upper part (i.e., Arg 1019 shown in blue in Figure 3b). These salt bridges are formed with the contribution of two or three monomers of the S protein. Phe 1042, Leu 1024 and Ala 1016 also provide dimeric or trimeric symmetrical interactions to stabilize these molecules within the cavity. Similarly, focusing on the nucleotide analogues in BS A and BS C, we observed that phosphate groups tend to form intermolecular interactions preferentially with Gln 1002 and Lys 1038, respectively.
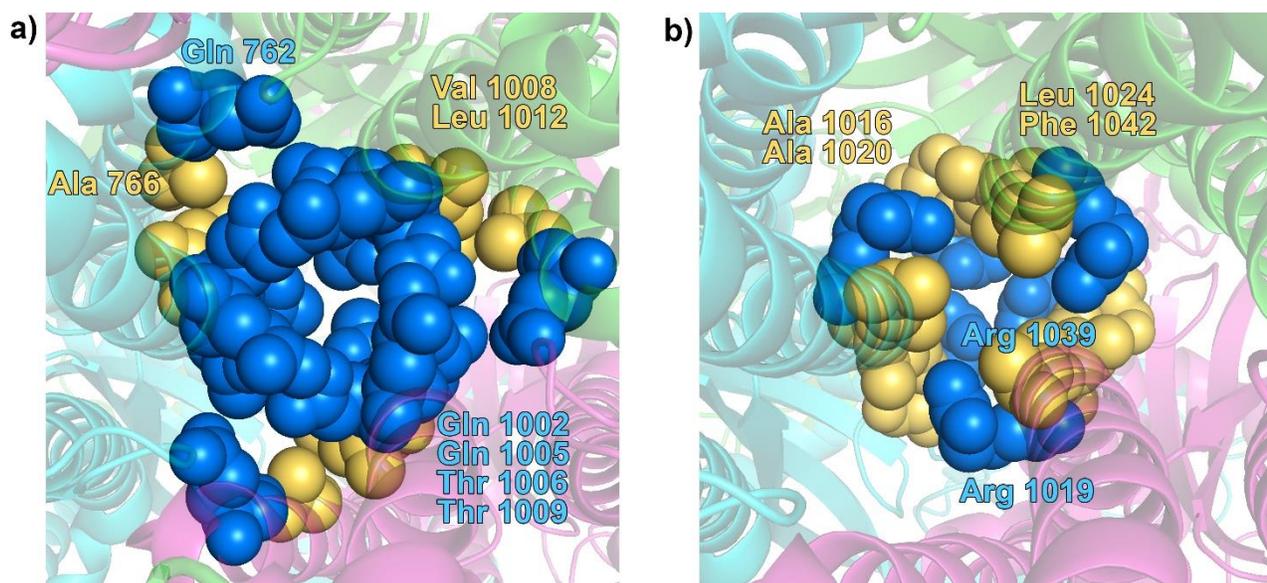

Figure 4: Representation of selected amino acids identified as the fingerprint of a) BS A and b) BS B. The monomers of the spike protein are shown with cartoon representation in three different colours. Important amino acids are represented with spheres and coloured according to their polarity as blue and pink for polar and nonpolar, respectively. The view of the picture is set such that one looks from the top to the bottom of the protein along the principal axis of the central channel.

We identified several interactions between ligands and amino acids within BSs A and B that seem relevant for effective binding. Among those, a few appear to be more crucial and to engage in interactions with many molecules, thus representing the fingerprint of these BSs. In BS A, four polar groups Thr 1006, Thr 1099, Gln 1002 and Gln 1005 are aligned on the central helix of each one of the three monomers (Figure 4a). These groups contribute to both hydrogen bonding and hydrophobic interactions. On the same helix, Val 1008 and Leu 1012 further strengthen the hydrophobic interactions in this region. Two other crucial residues, i.e., the nonpolar Ala 766 and the polar Gln 762, lie slightly away from the central channel on the HR1 domain of each monomer. Regarding BS B, we identified four nonpolar amino acids lying in the middle of the binding site (i.e., Ala 1016, 1020, Leu 1024, Phe 1042) and two charged arginine residues (i.e., Arg 1019 and Arg 1039), each one located at one end of the binding site. The intermolecular interactions in BS A and B involve amino acids from all three monomers, and molecules binding at these locations hold the potential to hinder the separation of the monomers in the later stage of the fusion mechanism.

Binding site C counts eight molecules among the top 100 of our ranking with average interaction energy of -63.96 kcal mol^{-1} and represents the second most populate BS when considering the entire set with a population of 23%. The BS is quite wide, and we could not identify any specific binding pattern worth discussing. Among the other binding sites, BSs G and H respectively count one and two molecules among the top 100 of our ranking while BSs D-F have no-hit in top of our list. Nevertheless, BSs D-H have a population of 0.4-5.0% over the whole set of molecules considered here. While we do not discuss specific interaction patterns for these BSs, we make the coordinates of the representative atomic centres that define them available to the interested reader together with the coordinates of each molecule in our final set.

Drug repurposing

About 1,000 molecules in the DrugBank database correspond to approved and marketed drugs. We performed a more extensive molecular modelling experiment for these particular molecules to identify potential candidates for drug repurposing. Each drug was docked in each of the 20 docking grid centres defined in Figure 1a, within four different geometries of the receptor S protein (i.e., two closed and two open conformations). The best pose for each docking attempt was then energy minimized and interaction energy including solvent contribution were evaluated with molecular mechanics. For each drug in the set, we kept the best combination of docking grid centre and receptor geometry. The 42 drugs having interaction energy with the S protein lower than -50.00 kcal mol^{-1} are presented in Table 1, including their DrugBank identification number, preferred binding site, calculated interaction energy, and therapeutic usage. The entirety of our results is available upon request to the authors, together with the molecular structure of the different poses within the protein. Hereafter, we discuss some of the drugs in this list that caught our attention, and a representation of the interaction patterns are given for a selection of them in Figure 5.

Several antibiotics yielded significant interaction with the S protein. In particular, Streptomycin (Figure 5a) is the best hit among the approved drugs that we investigated with an interaction energy of -93.97 kcal mol^{-1} towards BS A. It forms a number of hydrogen bonds with the fingerprint residues of BS A: i.e., Thr 1006, Thr 1009, Gln 1002, and Gln 1005. Cefpiramide (-56.11 kcal mol^{-1}; BS C), Dalfopristin (-55.20 kcal mol^{-1}; BS A), and Clindamycin (-50.64 kcal mol^{-1}; BS A) also showed good affinity towards the S protein's central regions. To the best of our knowledge, none of these antibiotics have been tested in the treatment of COVID-19 so far.

Three antiviral drugs hit the top of our ranking: Nelfinavir, Remdesivir, and Sofosbuvir. Nelfinavir (Figure 5b) binds to BS A with an interaction energy of -71.39 kcal mol^{-1}. It is used as an HIV protease inhibitor and was found to be effective towards SARS-CoV.[72] Nelfinavir's great potential to be used as a treatment of COVID-19 was recently discussed, since it was shown to block SARS-CoV-2's replication at a very low concentration (EC_{50} = 1.13 µM),[73] and suggested to be an agent to inhibit the cell fusion initiated by S protein.[74] As depicted in Figure 5b, it is mainly stabilized by hydrophobic residues within BS A. The characteristic polar residues identified in BS A and discussed in the previous section (i.e., Thr 1006, Thr 1009, Gln 1002, Gln 1005, Gln 1010) also contribute to both polar and nonpolar interactions. At the time at which we conducted this study, Remdesivir was not classified among the approved drugs and was therefore not subject to this more extensive virtual screening. Nevertheless, we added it afterward and tabulated it with the other drugs in Table 1, as it ranks quite high in our virtual screening with an interaction energy of -53.90 kcal mol^{-1} towards BS A. It was regarded as a treatment with high potential until a recent clinical trial ruled out its effectivity against COVID-19.[3] Sofosbuvir binds to BS F with an affinity of -50.08 kcal mol^{-1}. It is a nucleotide analogue that was suggested as a possible treatment of COVID-19.[75] We notice, however, that it was shown *in vitro* and clinically to target preferentially SARS-CoV-2's RNA-dependent RNA polymerase (RdRp).[76]

Three folate derivates, including folic acid itself, made the top of our interaction energy ranking: i.e., Leucovorin (-79.12 kcal mol^{-1} in BS C), folic acid (-69.71 kcal mol^{-1} in BS C), and Methotrexate (-57.58 kcal mol^{-1} in BS C). Leucovorin (Figure 5c) is mainly stabilized in BS C by a symmetrical contribution of residues from the three monomers of the S protein: i.e., the carboxylic part of the drug interacts strongly with Arg 1039. It is noteworthy to mention that folic acid has recently been examined in COVID-19 treatment as a potential inhibitor of the furin enzyme.[77]

Chlorhexidine (Figure 5d) is another noteworthy binder that shows a strong binding affinity towards BS A with an interaction energy of -75.56 kcal mol^{-1}. Gln 1002, Gln 1005 and Thr 1009 form hydrogen bonds with the drug, and the rest of the molecule is stabilized by the contribution of the nonpolar amino acids in this BS. It was not in our original set of approved drugs because this drug was labelled as withdrawn in the version of the DrugBank database that we used. Since it appeared to us as a promising candidate, we added it afterwards to the current list in Table 1. Chlorhexidine is a disinfectant and antiseptic that is also used as a mouthwash to reduce the mouth's bacteria population. It is most effective towards gram-positive bacteria, but also gram-negative bacteria, fungi and envelop viruses. If further work, including *in vitro* assays would confirm any inhibition of the S protein's activity, Chlorhexidine could represent an exciting agent in COVID-19 protection.

Another interesting hit in our list is Famotidine (Figure 5e) with an interaction energy of -50.78 kcal mol^{-1}. This histamine-2-blocker, antiulcer medicine, was recently suggested to be effective in COVID-19 treatment, and its usage is currently under investigation.[78] Famotidine also binds to BS A, however, in the upper region of the pocket. Half of the molecule is stabilized by both polar and nonpolar contributions of residues Gln 1002 and Gln 1005 coming from all three monomers.

Several iodine-based radiographic contrast agents were found to bind strongly to the various binding sites of the S protein: i.e., Iodipamide (-58.07 kcal mol^{-1} in BS F), Iopamidol (-51.89 kcal mol^{-1} in BS A), Iohexol (-51.37 kcal mol^{-1} in BS F), Ioxilan (-50.89 kcal mol^{-1} in BS A), and Ioversol (-50.88 kcal mol^{-1} in BS C). Radiographic imaging is a widely used diagnostic technique in COVID-19. Although the sole information that we provide here is the strong interaction between these agents and SARS-CoV-2's S protein, we believe that further investigation on these complexes might bring relevant insights into COVID-19 diagnostics.

Among all the drugs in the approved set, 857 bind to the S protein with interaction energy lower than -20.00 kcal mol^{-1}, and 204 with energy lower -40.00 kcal mol^{-1}. This suggests that many other relevant candidates might be of interest as a potential treatment, although not listed in Table 1. For that reason, we make our entire data available upon request to the authors for the interested reader to find information regarding specific drugs. Notably, Nafamostat binds to BS A with interaction energy of -47.58 kcal mol^{-1}. It is not an approved drug, but it is currently under investigation. We added this drug to our approved drugs modelling set since it was recently shown to be adequate to block SARS-CoV-2's activation before cell entry.[79] As shown in Figure 5f, a symmetrical contribution from Thr 998 and Thr 1009 of each monomer also significantly stabilizes this molecule within BS A.

Table 1: List of the approved drugs forming complexes with SARS-CoV-2's spike protein with molecular mechanics interaction energy lower than -50.00 kcal mol^{-1}. The DrugBank identification number (DB ID), the preferred binding site (BS; see Figure 1d), the generic name, the interaction energy (E_{int} in kcal mol^{-1}), and the primary usage is given for each drug.

| DB ID	BS	Generic name	E_{int} (kcal mol^{-1})	Usage
DB01082	A	Streptomycin	-93.97	antibiotic
DB00224	A	Indinavir	-82.04	HIV antiretroviral
DB00650	C	Leucovorin	-79.12	nutraceutical
DB00878	A	Chlorhexidine	-75.56	disinfectant
DB00220	A	Nelfinavir	-71.39	antiviral
DB00158	C	Folic acid	-69.71	nutraceutical
DB01337	A	Pancuronium	-66.75	steroid
DB00796	A	Candesartan cilexetil	-66.57	angiotensin-receptor blocker (ARB)
DB06154	C	Pentaerythritol tetranitrate	-65.03	coronary vasodilator
DB06809	A	Plerixafor	-63.51	stem cell mobilizer
DB00354	A	Buclizine	-61.78	antihistamine
DB08897	A	Aclidinium	-60.73	anticholinergic
DB00248	A	Cabergoline	-60.43	dopamine
DB09083	A	Ivabradine	-60.33	heart rate lowering agent
DB00562	A	Benzthiazide	-59.51	antihypertensive
DB00341	A	Cetirizine	-58.74	histamine H1 antagonist
DB04711	F	Iodipamide	-58.07	contrast agent
DB00563	C	Methotrexate	-57.58	folate derivative
DB00737	A	Meclizine	-57.55	histamine H1 antagonist
DB13711	A	Tritoqualine	-57.33	antihistamine
DB08932	A	Macitentan	-56.53	antagonist/blocker of endothelin receptors
DB06708	A	Lumefantrine	-56.46	antimalarial agent
DB00471	B	Montelukast	-56.44	leukotriene receptor antagonist
DB01430	A	Almitrine	-56.21	respiratory stimulant
DB00430	C	Cefpiramide	-56.11	antibiotic
DB06702	A	Fesoterodine	-55.33	antimuscarinic prodrug
DB01764	A	Dalfopristin	-55.20	antibiotic
DB01194	A	Brinzolamide	-55.10	carbonic anhydrase inhibitor
DB01180	A	Rescinnamine	-54.86	antihypertensive
DB01089	A	Deserpidine	-54.41	antipsychotic and antihypertensive
DB14761	A	Remdesivir	-53.90	antiviral
DB11262	G	Bisoctrizole	-52.56	UV-filter
DB08947	A	Iopamidol	-51.89	contrast agent
DB06207	A	Silodosin	-51.52	α1-adrenoceptor antagonist
DB01362	F	Iohexol	-51.37	contrast agent
DB11275	A	Epicriptine	-51.10	antiparkinsonian agent
DB01319	B	Fosamprenavir	-51.00	prodrug of HIV protease inhibitor
DB09135	A	Ioxilan	-50.89	contrast agent
DB09134	C	Ioversol	-50.88	contrast agent
DB00927	A	Famotidine	-50.78	histamine-2 blocker
DB01190	A	Clindamycin	-50.64	antibiotic
DB00782	A	Propantheline	-50.41	antispasmodic
DB08934	E	Sofosbuvir	-50.08	antiviral
DB00598	A	Labetalol	-50.07	decrease high blood pressure
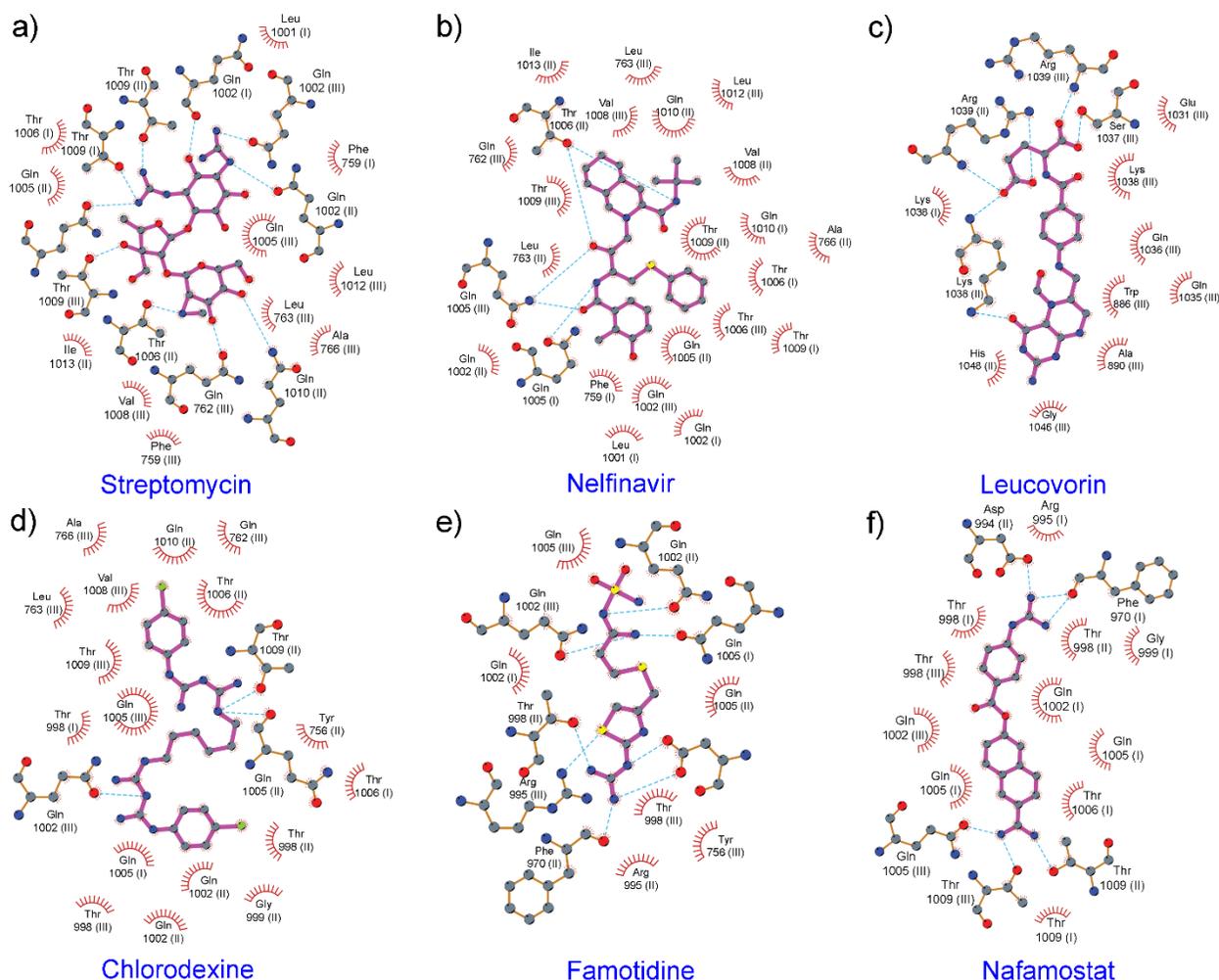
Figure 5: Representation of drug/protein interactions for selected approved drugs: a) Streptomycin, b) Nelfinavir, c) Leucovorin, d) Chlorodexine, e) Famotidine, and f) Nafamostat. Ligands and hydrogen bond donor/acceptor residues are shown with ball and stick representation (C=grey, O=red, N=blue, P=orange). Hydrophobic contacts are given by an arc with radiating spokes. Roman numbers (I, II, III) are used to label the different monomers. Identical residues coming from different monomers (I, II, III) are highlighted with the same colour.

Comments regarding the docking strategy

General difficulties encountered in virtual screening studies result from the high dimensionality of the data sets under consideration, which prevents using the most elaborate docking strategies. Energy information obtained with approximate scoring functions used in docking programs should be refined and resorted to reach accurate and reliable results.[80] Most importantly, the rigidity of the receptor may lead the docking software to miss potentially good binders. In that direction, programs that let the docking algorithm generate poses with a significant overlap between ligand and receptor atoms might overcome this problem. This would naturally lead to a particularly bad value of the scoring function, and the corresponding poses would be disregarded if the strategy stopped at that stage. The power of such protocol resides in its second step, which consists of relaxing the complexes with successive molecular dynamics-based simulations to lead to a geometry of the receptor that adapts to that of the ligand. While

14

refining all docking poses with molecular dynamics is out of consideration for the number of poses that we produced in this work, we borrowed from this idea and systematically relaxed the complexes generated by AutoDock Vina with molecular mechanics-based energy minimisations. For the set of approved drugs, all poses generated were refined, while for the whole set, we needed to make a selection to keep the computational expense to a reasonable level. In the latter case, we analysed the distribution of docking scores generated by AutoDock Vina and kept all poses systematically located in the better half of the distribution (see Methodology and Figure 1b). In line with the idea of letting the docking algorithm generate poses with a bad score and overlaps between ligand and receptors, we also randomly selected a number of complexes from the worst part of the distribution of scores. All these good and bad poses were then energy minimised and rescored with molecular mechanics. We observed that about 10% of the complexes taken from the worst part of the docking score distribution made it eventually the better part of the ranking after MM-based refinement. While a more in-depth analysis of the improvement of the poses is not in the scope of the present work, we found some specific examples that are worth discussing to illustrate the relevance of our strategy.

Considering only the set of approved drugs, we find that AutoDock Vina scored some docked poses with a particularly high, positive value of binding energy. Among many others, this is the case of Streptomycin and Indinavir, the two best binders identified in this work. The poses for Streptomycin and Indinavir that led to the best complex after MM-based refinement were indeed scored by AutoDock Vina with positive binding energies of 45.70 and 38.10 kcal mol^{-1}, respectively, while most molecules were found with a score in the range -12.00 to 0.00 kcal mol^{-1} (Figure 1b). As shown in Table 1, the interaction energy for these two complexes after MM refinement reached values as good as -93.97 and -82.04 kcal mol^{-1}, respectively. Relying only on the docking score would thus have led us to disregard these two excellent binders. Lastly, it is worth noting that the bad score assigned by AutoDock Vina is not a shortcoming of the docking program but a limitation of the setup, i.e., performing rigid docking instead of allowing some flexibility in the receptor. Instead, we want to point out that despite this limitation, AutoDock Vina appears to be powerful enough to generate poses that are highly relevant as an initial step, in order to further yield great complex geometries after a computationally affordable MM-based refinement. We trust that the strategy designed for the present blind virtual screening can be successfully adapted to other studies in drug discovery.

Conclusions

We have conducted an extensive blind virtual screening investigation on the druggability of SARS-CoV-2's spike (S) protein. Considering 20 potential binding centers and a library of about 9,000 molecules within an elaborate docking strategy, we have identified eight so far undocumented binding sites in the S protein. The most populated binding site is located in the center of the S protein, at the interface between the central helix of the three monomers. Molecules binding at that location hold the potential to prevent or slow down the virus's fusion mechanism into the host cell. Our identification and characterisation of new binding sites in the S protein are expected to lead future research in *de novo* drug design and drug repurposing.

We took a step forward in the direction of identifying drugs that could be used to prevent infection by SARS-CoV-2 by focusing on a subset of nearly 1,000 approved and marketed drugs. This second step revealed a number of compounds having a high affinity towards the S protein, with 42 drugs showing molecular mechanics-based interaction energies ranging from -50 to -94 kcal mol^{-1}. Among the top binders, we identified several antibiotics (e.g., Streptomycin -94 kcal mol^{-1}) and antivirals (e.g., Nelfinavir -71 kcal mol^{-1}) that were not considered so far in clinical trials. Nelfinavir was shown in recent in vitro

studies to prevent the virus from entering the host cell. Its high affinity towards the S protein's interior, as revealed by the present study, suggests that Nelfinavir may be involved in hindering the complex fusion mechanism of SARS-CoV-2. Other molecules such as Famotidine or folic acid were already considered by other studies as potential treatments against COVID-19 and showed up at the top of our ranking. Some radiographic contrast agents, as well as a disinfectant (i.e., Chlorhexidine -76 kcal mol^{-1}), also showed great affinity towards the S protein, opening new perspectives for the search of potent diagnostic and preventive therapeutics in the fight against the virus.

Although drug repurposing is a great hope that would serve the immediate therapeutic need we are facing today, experiences with similar infections indicate that we will be in need of novel drug research to prevent and treat COVID-19. We suggest that the nucleoside derivatives investigated in the present work hold the great potential to initiate multidisciplinary rational drug design research against SARS-CoV-2 and other viruses of the CoV family.

Acknowledgements

References

1. da Silva, J. A. T.; Tsigaris, P.; Erfanmanesh, M., Publishing volumes in major databases related to Covid-19. *Scientometrics* **2020**, 1-12.
2. Sanders, J. M.; Monogue, M. L.; Jodlowski, T. Z.; Cutrell, J. B., Pharmacologic treatments for coronavirus disease 2019 (COVID-19): a review. *Jama* **2020**, *323* (18), 1824-1836.
3. Pan, H.; Peto, R.; Karim, Q. A.; Alejandria, M.; Restrepo, A. M. H.; García, C. H.; Kieny, M. P.; Malekzadeh, R.; Murthy, S.; Preziosi, M.-P., Repurposed antiviral drugs for COVID-19; interim WHO SOLIDARITY trial results. *medRxiv* **2020**.
4. Ciliberto, G.; Cardone, L., Boosting the arsenal against COVID-19 through computational drug repurposing. *Drug Discovery Today* **2020**.
5. Shah, B.; Modi, P.; Sagar, S. R., In silico studies on therapeutic agents for COVID-19: Drug repurposing approach. *Life Sciences* **2020**, 117652.
6. Kumar, Y.; Singh, H.; Patel, C. N., In silico prediction of potential inhibitors for the Main protease of SARS-CoV-2 using molecular docking and dynamics simulation based drug-repurposing. *Journal of infection and public health* **2020**, *13* (9), 1210-1223.
7. Arun, K.; Sharanya, C.; Abhithaj, J.; Francis, D.; Sadasivan, C., Drug repurposing against SARS-CoV-2 using E-pharmacophore based virtual screening, molecular docking and molecular dynamics with main protease as the target. *Journal of Biomolecular Structure and Dynamics* **2020**, 1-12.
8. Jiménez-Alberto, A.; Ribas-Aparicio, R. M.; Aparicio-Ozores, G.; Castelán-Vega, J. A., Virtual screening of approved drugs as potential SARS-CoV-2 main protease inhibitors. *Computational biology and chemistry* **2020**, *88*, 107325.
9. Tsuji, M., Potential anti-SARS-CoV-2 drug candidates identified through virtual screening of the ChEMBL database for compounds that target the main coronavirus protease. *FEBS Open Bio* **2020**.
10. Fischer, A.; Sellner, M.; Neranjan, S.; Smieško, M.; Lill, M. A., Potential Inhibitors for Novel Coronavirus Protease Identified by Virtual Screening of 606 Million Compounds. *International Journal of Molecular Sciences* **2020**, *21* (10), 3626.
11. Baby, K.; Maity, S.; Mehta, C.; Suresh, A.; Nayak, U. Y.; Nayak, Y., Targeting SARS-CoV-2 Main Protease: A Computational Drug Repurposing Study. *Archives of medical research* **2020**.
12. Ton, A. T.; Gentile, F.; Hsing, M.; Ban, F.; Cherkasov, A., Rapid identification of potential inhibitors of SARS-CoV-2 main protease by deep docking of 1.3 billion compounds. *Molecular informatics* **2020**.
13. Gentile, D.; Patamia, V.; Scala, A.; Sciortino, M. T.; Piperno, A.; Rescifina, A., Putative inhibitors of SARS-CoV-2 main protease from a library of marine natural products: A virtual screening and molecular modeling study. *Marine drugs* **2020**, *18* (4), 225.
14. Li, Z.; Li, X.; Huang, Y.-Y.; Wu, Y.; Liu, R.; Zhou, L.; Lin, Y.; Wu, D.; Zhang, L.; Liu, H., Identify potent SARS-CoV-2 main protease inhibitors via accelerated free energy perturbation-based virtual screening of existing drugs. *Proceedings of the National Academy of Sciences* **2020**.
15. Gahlawat, A.; Kumar, N.; Kumar, R.; Sandhu, H.; Singh, I. P.; Singh, S.; Sjöstedt, A.; Garg, P., Structure-based virtual screening to discover potential lead molecules for the SARS-CoV-2 main protease. *Journal of chemical information and modeling* **2020**.
16. Prajapat, M.; Shekhar, N.; Sarma, P.; Avti, P.; Singh, S.; Kaur, H.; Bhattacharyya, A.; Kumar, S.; Sharma, S.; Prakash, A., Virtual screening and molecular dynamics study of approved drugs as inhibitors of spike protein S1 domain and ACE2 interaction in SARS-CoV-2. *Journal of Molecular Graphics and Modelling* **2020**, *101*, 107716.
17. Choudhary, S.; Malik, Y. S.; Tomar, S., Identification of SARS-CoV-2 cell entry inhibitors by drug repurposing using in silico structure-based virtual screening approach. *Frontiers in immunology* **2020**, *11*, 1664.

18.   Senathilake, K.; Samarakoon, S.; Tennekoon, K., Virtual screening of inhibitors against spike glycoprotein of 2019 novel corona virus: A drug repurposing approach. **2020**.

19.   Senathilake, K.; Samarakoon, S.; Tennekoon, K., Virtual screening of inhibitors against spike glycoprotein of SARS-CoV-2: A drug repurposing approach. **2020**.

20.   Awad, I. E.; Abu-Saleh, A. A.-A. A.; Sharma, S.; Yadav, A.; Poirier, R. A., High-throughput virtual screening of drug databanks for potential inhibitors of SARS-CoV-2 spike glycoprotein. *Journal of Biomolecular Structure and Dynamics* **2020**, 1-14.

21.   Alexpandi, R.; De Mesquita, J. F.; Pandian, S. K.; Ravi, A. V., Quinolines-Based SARS-CoV-2 3CLpro and RdRp Inhibitors and Spike-RBD-ACE2 Inhibitor for Drug-Repurposing Against COVID-19: An in silico Analysis. *Frontiers in microbiology* **2020,** *11*, 1796.

22.   Wu, Y.-C.; Chen, G.-Y.; Yao, T.-Y.; Ahmed, A.; Pan, Y.-C.; Yang, J.-C., The discovery of potential natural products for targeting SARS-CoV-2 spike protein by virtual screening. *bioRxiv* **2020**.

23.   Maffucci, I.; Contini, A., In Silico Drug Repurposing for SARS-CoV-2 Main Proteinase and Spike Proteins. *Journal of Proteome Research* **2020**.

24.   Patil, S.; Hofer, J.; Ballester, P. J.; Fattakhova, E.; DiFlumeri, J.; Campbell, A.; Oravic, M., Drug repurposing for Covid-19: Discovery of potential small-molecule inhibitors of spike protein-ACE2 receptor interaction through virtual screening and consensus scoring. **2020**.

25.   Padhi, S.; Massi, M.; Chourasia, R.; Rajashekar, Y.; Rai, A. K.; Evidente, A., ADMET profile and virtual screening of plant and microbial natural metabolites as SARS-CoV-2 S1 glycoprotein receptor binding domain and main protease inhibitors. *European journal of pharmacology* **2020**, 173648.

26.   Patel, A.; Rajendran, M.; Pakala, S. B.; Shah, A.; Patel, H.; Karyala, P., Virtual Screening of Curcumin and Its Analogs Against the Spike Surface Glycoprotein of SARS-CoV-2 and SARS-CoV. **2020**.

27.   Shehroz, M.; Zaheer, T.; Hussain, T., Computer-aided drug design against spike glycoprotein of SARS-CoV-2 to aid COVID-19 treatment. *Heliyon* **2020**, e05278.

28.   Mathew, S. M.; Benslimane, F.; Althani, A. A.; Yassine, H. M., Virtual screening of anti-viral drugs and natural compounds for potential inhibition of the novel SARS-CoV-2 spike receptor-binding domain. **2020**.

29.   Benítez-Cardoza, C. G.; Vique-Sánchez, J. L., Potential inhibitors of the interaction between ACE2 and SARS-CoV-2 (RBD), to develop a drug. *Life sciences* **2020,** *256*, 117970.

30.   Alagumuthu, M.; Rajpoot, S.; Baig, M. S., Structure-Based Design of Novel Peptidomimetics Targeting the SARS-CoV-2 Spike Protein. *Cellular and Molecular Bioengineering* **2020**, 1-9.

31.   Bharath, B.; Damle, H.; Ganju, S.; Damle, L., In silico screening of known small molecules to bind ACE2 specific RBD on Spike glycoprotein of SARS-CoV-2 for repurposing against COVID-19. *F1000Research* **2020,** *9*.

32.   Otazu, K.; Chenet-Zuta, M. E.; Ropon-Palacios, G.; Olivos-Ramirez, G. E.; Jimenez-Avalos, G. M.; Osorio-Mogollon, C.; Sosa-Amay, F.; Vargas-Rodriguez, R.; Nina-Larico, T. P.; Concu, R., Targeting Receptor Binding Domain and Cryptic Pocket of Spike glycoprotein from SARS-CoV-2 by biomolecular modeling. *arXiv preprint arXiv:2006.06452* **2020**.

33.   Deganutti, G.; Prischi, F.; Reynolds, C. A., Supervised molecular dynamics for exploring the druggability of the SARS-CoV-2 spike protein. *Journal of computer-aided molecular design* **2020**, 1-13.

34.   Sorzano, C. Ó. S.; Crisman, E.; Carazo, J. M.; León, R., Multitarget virtual screening for drug repurposing in COVID19. **2020**.

35.   Kalhor, H.; Sadeghi, S.; Abolhasani, H.; Kalhor, R.; Rahimi, H., Repurposing of the approved small molecule drugs in order to inhibit SARS-CoV-2 S protein and human ACE2 interaction through virtual screening approaches. *Journal of Biomolecular Structure and Dynamics* **2020**, 1-16.

36.   de Oliveira, O. V.; Rocha, G. B.; Paluch, A. S.; Costa, L. T., Repurposing approved drugs as inhibitors of SARS-CoV-2 S-protein from molecular modeling and virtual screening. *Journal of Biomolecular Structure and Dynamics* **2020,** (just-accepted), 1-14.

37.   Trezza, A.; Iovinelli, D.; Santucci, A.; Prischi, F.; Spiga, O., An integrated drug repurposing strategy for the rapid identification of potential SARS-CoV-2 viral inhibitors. *Scientific reports* **2020,** *10* (1), 1-8.

38.   Mahdian, S.; Ebrahim-Habibi, A.; Zarrabi, M., Drug repurposing using computational methods to identify therapeutic options for COVID-19. *Journal of Diabetes and Metabolic Disorders* **2020**, 1.

39.   Boyarsky, A., Binding Ligands that Straddle an Important Contact Site on the RBD of the Covid-19 Spike Protein. *bioRxiv* **2020**.

40.   Smieszek, S.; Przychodzen, B.; Polymeropoulos, V.; Polymeropoulos, C.; Polymeropoulos, M., Direct ACE2-Spike RBD Binding Disruption with Small Molecules: A Strategy for COVID-19 Treatment. **2020**.

41.   Unni, S.; Aouti, S.; Thiyagarajan, S.; Padmanabhan, B., Identification of a repurposed drug as an inhibitor of Spike protein of human coronavirus SARS-CoV-2 by computational methods. *Journal of Biosciences* **2020,** *45* (1), 1-20.

42.   Prashantha, C.; Gouthami, K.; Lavanya, L.; Bhavanam, S.; Jain, A.; Shakthiraju, R.; Suraj, V.; Sahana, K.; Sujana, H.; Guruprasad, N., Molecular screening of antimalarial, antiviral, anti-inflammatory and HIV protease inhibitors against spike glycoprotein of Coronavirus. *Journal of Molecular Graphics & Modelling* **2020**.

43.   Alazmi, M.; Motwalli, O., In silico virtual screening, characterization, docking and molecular dynamics studies of crucial SARS-CoV-2 proteins. *Journal of Biomolecular Structure and Dynamics* **2020**, 1-11.

44.   Somadi, G.; Sivan, S. K., Identification of therapeutic target in S2 domain of SARS nCov-2 Spike glycoprotein: Key to design and discover drug candidates for inhibition of viral entry into host cell. *Journal of Theoretical and Computational Chemistry* **2020,** *19* (07), 2050028.

45.   Artese, A.; Svicher, V.; Costa, G.; Salpini, R.; Di Maio, V. C.; Alkhatib, M.; Ambrosio, F. A.; Santoro, M. M.; Assaraf, Y. G.; Alcaro, S., Current status of antivirals and druggable targets of SARS CoV-2 and other human pathogenic coronaviruses. *Drug Resistance Updates* **2020**, 100721.

46.   Skoreński, M.; Grzywa, R.; Sieńczyk, M., Why should we target viral serine proteases when developing antiviral agents? Future Medicine: 2016.

47.   Hung, H.-C.; Ke, Y.-Y.; Huang, S. Y.; Huang, P.-N.; Kung, Y.-A.; Chang, T.-Y.; Yen, K.-J.; Peng, T.-T.; Chang, S.-E.; Huang, C.-T., Discovery of M Protease inhibitors encoded by SARS-CoV-2. *Antimicrobial agents and chemotherapy* **2020,** *64* (9).

48.   Tong, L., Viral proteases. *Chemical Reviews* **2002,** *102* (12), 4609-4626.

49.   Ramos-Guzmán, C. A.; Ruiz-Pernía, J. J.; Tuñón, I., A Microscopic Description of SARS-CoV-2 Main Protease Inhibition with Michael Acceptors. Strategies for Improving Inhibitors Design. **2020**.

50.   Tang, T.; Bidon, M.; Jaimes, J. A.; Whittaker, G. R.; Daniel, S., Coronavirus membrane fusion mechanism offers as a potential target for antiviral development. *Antiviral research* **2020**, 104792.

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

52.   Walls, A. C.; Park, Y.-J.; Tortorici, M. A.; Wall, A.; McGuire, A. T.; Veesler, D., Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein. *Cell* **2020**.

53.   Toelzer, C.; Gupta, K.; Yadav, S. K.; Borucu, U.; Davidson, A. D.; Williamson, M. K.; Shoemark, D. K.; Garzoni, F.; Staufer, O.; Milligan, R., Free fatty acid binding pocket in the locked structure of SARS-CoV-2 spike protein. *Science* **2020,** *370* (6517), 725-730.

54. Romeo, A.; Iacovelli, F.; Falconi, M., Targeting the SARS-CoV-2 spike glycoprotein prefusion conformation: virtual screening and molecular dynamics simulations applied to the identification of potential fusion inhibitors. *Virus research* **2020,** *286*, 198068.

55. Wishart, D. S.; Feunang, Y. D.; Guo, A. C.; Lo, E. J.; Marcu, A.; Grant, J. R.; Sajed, T.; Johnson, D.; Li, C.; Sayeeda, Z., DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic acids research* **2018,** *46* (D1), D1074-D1082.

56. O'Boyle, N. M.; Banck, M.; James, C. A.; Morley, C.; Vandermeersch, T.; Hutchison, G. R., Open Babel: An open chemical toolbox. *Journal of cheminformatics* **2011,** *3* (1), 33.

57. Macke, T. J.; Svrcek-Seiler, W.; Brown, R. A.; Kolossváry, I.; Bomble, Y. J.; Case, D. A.; Zhang, W.; Hou, T.; Schafmeister, C.; Ross, W. S., AmberTools Users' Manual. Ver: 2010.

58. Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A., Development and testing of a general amber force field. *Journal of computational chemistry* **2004,** *25* (9), 1157-1174.

59. Bayly, C. I.; Cieplak, P.; Cornell, W.; Kollman, P. A., A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model. *The Journal of Physical Chemistry* **1993,** *97* (40), 10269-10280.

60. Gasteiger, J., A new model for calculating atomic charges in molecules. **1978**.

61. Eswar, N.; Eramian, D.; Webb, B.; Shen, M.-Y.; Sali, A., Protein structure modeling with MODELLER. In *Structural proteomics*, Springer: 2008; pp 145-159.

62. Ceroni, A.; Passerini, A.; Vullo, A.; Frasconi, P., DISULFIND: a disulfide bonding state and cysteine connectivity prediction server. *Nucleic acids research* **2006,** *34* (suppl_2), W177-W181.

63. Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C., ff14SB: improving the accuracy of protein side chain and backbone parameters from ff99SB. *Journal of chemical theory and computation* **2015,** *11* (8), 3696-3713.

64. Søndergaard, C. R.; Olsson, M. H.; Rostkowski, M.; Jensen, J. H., Improved treatment of ligands and coupling effects in empirical calculation and rationalization of p K a values. *Journal of chemical theory and computation* **2011,** *7* (7), 2284-2295.

65. Olsson, M. H.; Søndergaard, C. R.; Rostkowski, M.; Jensen, J. H., PROPKA3: consistent treatment of internal and surface residues in empirical p K a predictions. *Journal of chemical theory and computation* **2011,** *7* (2), 525-537.

66. Horn, H. W.; Swope, W. C.; Pitera, J. W.; Madura, J. D.; Dick, T. J.; Hura, G. L.; Head-Gordon, T., Development of an improved four-site water model for biomolecular simulations: TIP4P-Ew. *The Journal of chemical physics* **2004,** *120* (20), 9665-9678.

67. Duke, R. E.; Pedersen, L. G., PMEMD 3. *University of North Carolina-Chapel Hill, Chapel Hill, NC* **2003**.

68. Harris, R.; Olson, A. J.; Goodsell, D. S., Automated prediction of ligand-binding sites in proteins. *Proteins: structure, function, and bioinformatics* **2008,** *70* (4), 1506-1517.

69. Trott, O.; Olson, A. J., AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of computational chemistry* **2010,** *31* (2), 455-461.

70. Nguyen, N. T.; Nguyen, T. H.; Pham, T. N. H.; Huy, N. T.; Bay, M. V.; Pham, M. Q.; Nam, P. C.; Vu, V. V.; Ngo, S. T., Autodock vina adopts more accurate binding poses but autodock4 forms better binding affinity. *Journal of Chemical Information and Modeling* **2019,** *60* (1), 204-211.

71. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V., Scikit-learn: Machine learning in Python. *the Journal of machine Learning research* **2011,** *12*, 2825-2830.

72. Yamamoto, N.; Yang, R.; Yoshinaka, Y.; Amari, S.; Nakano, T.; Cinatl, J.; Rabenau, H.; Doerr, H. W.; Hunsmann, G.; Otaka, A., HIV protease inhibitor nelfinavir inhibits replication of SARS-associated coronavirus. *Biochemical and biophysical research communications* **2004,** *318* (3), 719-725.

73.   Yamamoto, N.; Matsuyama, S.; Hoshino, T.; Yamamoto, N., Nelfinavir inhibits replication of severe acute respiratory syndrome coronavirus 2 in vitro. *BioRxiv* **2020**.

74.   Ianevski, A.; Yao, R.; Fenstad, M. H.; Biza, S.; Zusinaite, E.; Reisberg, T.; Lysvand, H.; Løseth, K.; Landsem, V. M.; Malmring, J. F., Potential antiviral options against SARS-CoV-2 infection. *Viruses* **2020**, *12* (6), 642.

75.   Jácome, R.; Campillo-Balderas, J. A.; de León, S. P.; Becerra, A.; Lazcano, A., Sofosbuvir as a potential alternative to treat the SARS-CoV-2 epidemic. *Scientific Reports* **2020**, *10* (1), 1-5.

76.   Wang, Y.; Anirudhan, V.; Du, R.; Cui, Q.; Rong, L., RNA-dependent RNA polymerase of SARS-CoV-2 as a therapeutic target. *Journal of Medical Virology* **2020**.

77.   Sheybani, Z.; Dokoohaki, M. H.; Negahdaripour, M.; Dehdashti, M.; Zolghadr, H.; Moghadami, M.; Masoompour, S. M.; Zolghadr, A. R., The role of folic acid in the management of respiratory disease caused by COVID-19. **2020**.

78.   Janowitz, T.; Gablenz, E.; Pattinson, D.; Wang, T. C.; Conigliaro, J.; Tracey, K.; Tuveson, D., Famotidine use and quantitative symptom tracking for COVID-19 in non-hospitalised patients: a case series. *Gut* **2020**.

79.   Hoffmann, M.; Schroeder, S.; Kleine-Weber, H.; Müller, M. A.; Drosten, C.; Pöhlmann, S., Nafamostat mesylate blocks activation of SARS-CoV-2: New treatment option for COVID-19. *Antimicrobial Agents and Chemotherapy* **2020**.

80.   Rastelli, G.; Pinzi, L., Refinement and rescoring of virtual screening results. *Frontiers in chemistry* **2019**, *7*, 498.