**Structural Similarity of SARS-CoV2 M^{pro} and HCV NS3/4A Proteases
Suggests New Approaches for Identifying Existing Drugs Useful
as COVID-19 Therapeutics**

Khushboo Bafna¹, Robert M. Krug^{2,#}, and Gaetano T. Montelione^{1,#}

¹Department of Chemistry and Chemical Biology, Center for Biotechnology and
Interdisciplinary Sciences, Rensselaer Polytechnic Institute, Troy, New York, 12180

²Department of Molecular Biosciences, John Ring LaMontagne Center for Infectious
Disease, Institute for Cellular and Molecular Biology, University of Texas at Austin,
Austin, Texas 78712 USA.

Running title: HCV Protease Inhibitors for SARS-CoV2 M^{pro} Protease

Key words: COVID-19, Severe Acute Respiratory Syndrome (SARS), Drug Discovery,
Hepatitis C Virus (HCV) NS3/4A protease inhibitors, Human Immunodeficiency Virus 1
(HIV-1) protease inhibitors, structural bioinformatics.

[#]Corresponding authors: Robert M. Krug, Ph.D., Department of Molecular Biosciences,
John Ring LaMontagne Center for Infectious Disease, Institute for Cellular and
Molecular Biology, University of Texas at Austin, 2500 Speedway, MBB 2.122, Austin,
Texas 78712, e-mail: rkrug@austin.utexas.edu; and
Gaetano T. Montelione, Ph.D. Center for Biotechnology and Interdisciplinary Sciences,
Rensselaer Polytechnic Institute, 1623 15th St, Troy, NY 12180. e-mail:
monteg3@rpi.edu

Conflict of Interest Statement: G.T.M. is a founder of Nexomics Biosciences, Inc.

Abstract

During the current COVID-19 pandemic more than 160,000 people have died worldwide as of mid-April 2020, and the global economy has been crippled. Effective control of the SARS-CoV2 virus that causes the COVID-19 pandemic requires both vaccines and antivirals. Antivirals are particularly crucial to treat infected people during the period of time that an effective vaccine is being developed and deployed. Because the development of specific antiviral drugs can take a considerable length of time, an important approach is to identify existing drugs already approved for use in humans which could be repurposed as COVID-19 therapeutics. Here we focus on antivirals directed against the SARS-CoV2 M^{pro} protease, which is required for virus replication. A structural similarity search showed that the Hepatitis C virus (HCV) NS3/4A protease has a striking three-dimensional structural similarity to the SARS-CoV2 M^{pro} protease, particularly in the arrangement of key active site residues. We used virtual docking predictions to assess the hypothesis that existing drugs already approved for human use or clinical testing that are directed at the HCV NS3/4A protease might fit well into the active-site cleft of the SARS-CoV2 protease (M^{pro}). *AutoDock* docking scores for 12 HCV protease inhibitors and 9 HIV-1 protease inhibitors were determined and compared to the docking scores for an α -ketoamide inhibitor of M^{pro}, which has recently been shown to inhibit SARS-CoV2 virus replication in cell culture. We identified eight HCV protease inhibitors that bound to the M^{pro} active site with higher docking scores than the α -ketoamide inhibitor, suggesting that these protease inhibitors may effectively bind to the M^{pro} active site. These results provide the rationale for us to test the identified HCV protease inhibitors as inhibitors of the SARS-CoV2 protease, and as inhibitors of SARS-CoV2 virus replication. Subsequently these repurposed drugs could be evaluated as COVID-19 therapeutics.

Introduction

Coronaviruses (CoVs) cause human respiratory diseases. While several human coronaviruses cause relatively mild respiratory infections, three coronaviruses cause severe respiratory diseases in humans: Severe Acute Respiratory Syndrome virus (SARS), Middle East Respiratory Syndrome virus (MERS), and Corona Virus Infectious Disease 2019 virus (COVID-19) (1-3). The presumed etiologic cause of COVID-19 disease is the SARS-CoV2 virus (2, 3). In the current COVID-19 pandemic more than 2.3 million people have been infected worldwide (as of mid-April 2020), and more than 160,000 have died. No vaccine is available, and no antiviral drugs have yet proven to be effective.

Coronaviruses, including SARS-CoV2, are enveloped viruses. Their genome is comprised of a single, large (27-34 kilobase) positive-sense single-stranded RNA, which is directly translated by host cells. The SARS-CoV2 genome encodes 4 structural proteins, 16 non-structural proteins (NSPs) which carry out crucial intracellular functions, and 9 accessory proteins (3, 4). Many of these proteins, and their host binding partners (4), are potential targets for development of antiviral therapeutics for COVID-19. Translation of the viral RNA results in the synthesis of two polyproteins that are processed by two virally encoded cysteine proteases, the papain-like protease (PL^{pro}) and a 3C-like protease (3CL^{pro}), which is also referred to Non-Structural Protein 5 (NSP5), or as the main protease (M^{pro}). Both PL^{pro} and M^{pro} proteases are required for virus replication and are targets for antiviral development. Here we focus on the potential of using existing antiviral drugs developed for hepatitis C viral infection as inhibitors of the SARS-CoV2 M^{pro} protease.

1 Considering the urgency for identifying effective antiviral drugs, and the usually
2 lengthy process involved in approving candidate drugs for safe human use, an
3 important approach is to identify existing drugs already approved for use in humans
4 which could be repurposed as COVID-19 antiviral drugs (4-8). Viral proteases have
5 been successfully targeted for the development of antiviral drugs against human
6 immunodeficiency virus-1(HIV-1) and hepatitis C virus (HCV) (9-12), and it is has been
7 suggested that some of these drugs may also be inhibitors of COVID-19 proteases and
8 virus replication (6-8).

9
10 The main protease of SARS-CoV2 (M^{pro}) is a 67.6 kDa homodimeric cysteine
11 protease. It has about 97% sequence identity with the corresponding M^{pro} protease of
12 the SARS-CoV virus responsible for the 2003 SARS pandemic. Not surprisingly, the
13 recent 1.75 Å X-ray crystal structure of SARS-CoV2 M^{pro} protease (13, 14)
14 demonstrates its structure is very similar to this SARS-CoV M^{pro} protease (15, 16). Both
15 of these M^{pro} proteases contain three similar domains. Domains I and II adopt a double
16 β -barrel fold, with the substrate binding site located in a shallow cleft between two
17 antiparallel β -barrels of Domain I and II (Figure 1A). Both of these M^{pro} proteases also
18 have an additional C terminal helical-bundle domain, Domain III, which apparently
19 stabilizes their active homodimer forms.

20
21 The fold architectures of Domains I and II of CoV M^{pro} proteases are similar to
22 those of chymotrypsin-like proteases and the 3C family of viral proteases. Anand et al.
23 (17) previously described the striking structural similarities between the active sites of
24 transmissible gastroenteritis coronavirus (TGEV) M^{pro} protease and the 3C protease of
25 hepatitis A virus. Our three-dimensional structural similarity search of the Protein Data

Bank using the *DALI* program (18, 19), with domains I and II (excluding domain III) of the SARS-CoV2 M^{pro} protease as the three-dimensional structural query, identified several 3C-like viral proteases, including the HCV NS3/4A protease, as structurally-similar. These two structures have a structural similarity Z score = +8.4, and overall backbone root-mean-squared deviation for structurally-similar regions of ~ 3.1 Å. Like M^{pro}, the HCV NS3/4A protease also has a double β-barrel fold, with relative orientations similar to those of the SARS-CoV M^{pro} proteases, and a substrate binding site located in a shallow cleft between its two six- to eight- stranded antiparallel β-barrels (Figure 1B). Superimposition of the backbone structures of these two proteases results also in superimposition of their substrate binding pockets and their active-site catalytic residues, His41 / Cys145 and His57 / Ser139 of SARS-CoV2 M^{pro} and HCV NS3/4A proteases, respectively (Figure 2).

Fold topology, overall fold, locations of substrate binding sites, and common positioning of active-site residues can result from homologous (divergent) evolutionary relationships between proteins. To our knowledge, there is no phylogenetic evidence for common ancestors of HCV NS3/4A protease and SARS-CoV2 M^{pro}, or of HCV and SARS-CoV viruses. Indeed, the HCV NS3/4A protease is a serine-protease, with catalytic triad His57, Asp81, and Ser139, while the SARS-CoV2 M^{pro} protease is a cysteine protease, with catalytic dyad residues His41, Cys145 (13, 14). While the two structures have similar structural architectures (i.e. organization of secondary structure elements (20)), their secondary structure topologies (i.e. how the secondary structure elements are linked together) are very different (Figure 3A), and do not indicate a divergent evolutionary relationship. However, a structure-based sequence alignment, generated from the superimposed three-dimensional structures, aligns catalytic

residues His41 / His57 and Cys145 / Ser139 (Figure 3B), where each is aligned pair is found in a similar secondary structure element. Hence, it would appear that the similarities in the overall structures and relative orientations of domains I and II, the location of the active site between these domains, and the relative positions of catalytic residues in these enzymes, is a result of convergent evolution to create a similar protease active site. The HIV-1 protease is a homodimeric beta-fold protein, with the active site located between the two protomers of the dimer (12, 21) (Figure 1C). However, it does not have identifiable structural similarity with SARS-CoV2 M^{pro}.

Work over the past > 15 years on the SARS-CoV M^{pro} protease has developed an extensive understanding of structure-activity relationships of lead molecules suitable for drug discovery efforts (4, 13-16, 22-24). Recently, an α -ketoamide inhibitor of the SARS-CoV2 virus was developed based on the three-dimensional crystal structure of its M^{pro} protease (13). This inhibitor, denoted 13b, binds into the active site of M^{pro}, making hydrogen-bonded interactions with backbone amides of key residues in the oxyanion hole, and forming a “reversible” thiohemiketal covalent bond with the S ^{γ} atom of catalytic residue Cys145 (13). It inhibits the purified recombinant protease with an IC₅₀ of 0.67 $\pm$ 0.18 μ M, and inhibits SARS-CoV2 virus replication in Calu3 human lung cells with an EC₅₀ of 4-5 μ M. Virus replication was inhibited ~100-fold with ~20 μ M of this compound.

Drug discovery and development is a long and expensive process. Accordingly, we (and others) are exploring the question of whether existing FDA-approved drugs, or drugs already in human clinical trials, can bind into the active site of M^{pro} protease to inhibit its proteolytic activity. These readily available drugs might inhibit SARS-CoV2 virus replication and/or be suitable as starting points for rational drug design to develop

1 optimal inhibitors of the M^{pro} protease and SARS-CoV2 virus replication. Considering
2 the structural similarities between the HCV NS3/4A and SARS-CoV2 proteases
3 described above, we hypothesize that one or more of the several successful, and
4 currently available drugs for hepatitis C disease (HepC) that target the essential HCV
5 NS3/4A protease might also function as antivirals against the SARS-CoV2 virus that
6 causes COVID-19 disease.

7
8 Accordingly, we carried out virtual docking studies of HCV protease inhibitor
9 drugs into the active site of the SARS-CoV2 M^{pro} protease using the program *AutoDock*
10 (25). We used as a benchmark state the structure of the complex formed between the
11 SARS-CoV2 protease and the α -ketoamide inhibitor 13b described above. In order to
12 compare our studies with drugs already in clinical trials for COVID-19 disease, we also
13 carried out similar virtual docking studies using protease inhibitor drugs directed to HIV-
14 1 protease. Comparing *AutoDock* binding scores and the structural poses of the
15 enzyme-bound drugs in these predicted ligand-protein complexes, we conclude that
16 there is significant merit to the hypothesis that one or more of the HCV drugs should be
17 prioritized for further experimental studies of SARS-CoV2 M^{pro} protease enzyme
18 inhibition, their ability to inhibit viral replication in cell culture, and ultimately their
19 potential as COVID-19 therapeutics..

21 **Computational Methods**

22 Molecular docking is a well-established approach for predicting binding
23 conformations of ligands in active/binding sites of relevant protein targets. Although it is
24 well known that the most favorable binding pose is not necessarily the one observed
25 experimentally, the experimentally-determined pose is often found among the highest

1 ranked poses (26, 27). These predicted conformations and relative binding affinities
2 provide meaningful insights into protein-ligand interaction profiles and free energies,
3 and can be crucial in the drug discovery process. A number of commercial and open
4 source computational tools are available to predict protein-ligand interactions.

5
6 In this study we used the free open source *Autodock* suite (25). *AutoDockTools*
7 (28) was used for coordinate preparation, docking, and analysis of results. The
8 computational docking program *AutoDock v4.2.6* is based on an empirical free-energy
9 force field and uses a search method based on Lamarckian genetic algorithm (29).
10 Target protein coordinates were obtained from SARS-CoV2 M^{pro} X-ray crystal structure
11 (PDB id 6Y2G) (13), and structural water was removed. Three-dimensional coordinates
12 for ligand molecules were obtained from PDB (<http://www.rcsb.org/>) or from chemical
13 structure databases ChemSpider (<http://www.chemspider.com/>) and DrugBank
14 (<https://www.drugbank.ca/>). Protein and ligand coordinates were then prepared using
15 *AutoDockTools*; polar hydrogens were added to protein structures, and Gasteiger-
16 Marsili empirical atomic partial charges were added to ligands. Torsional degrees of
17 freedom were identified for each ligand respectively. In these studies, all protein
18 torsions were kept rigid. These data and parameters for each protein and ligand were
19 saved as individual PDBQT files.

20
21 The program *Autogrid* (28) was used to prepare affinity maps for all atom types in
22 the receptor and ligands. A grid of 56, 40 and 48 points in x, y and z direction, with a
23 grid spacing of 0.375 Å was used to compute electrostatic maps. The grid center was
24 placed on the center of the α -ketoamide inhibitor 13b molecule, in its complex with M^{pro}
25 (PDB id 6Y2G) (Figure 4A). The Lamarckian genetic algorithm (LGA) method was used

for sampling ligand binding conformation (29), with the following LGA parameters: 150 individuals in population; 2,500,000 energy evaluations; 27,000 maximum number of generations; and with mutation and crossover rates of 0.02 and 0.08, respectively. A maximum of 300 iterations per local search was used. An initial set of 10 docking simulations were performed for each ligand. For the top scoring drug ligands, the calculations were repeated using 100 docking simulations. For a comparative analysis, docking simulations of the α -ketoamide inhibitor 13b were also performed using the same protocol used for docking the protease inhibitor drugs.

All docking simulations were analyzed using the *AutoDockTools*. Atomic coordinates for best scoring conformation obtained in each docking simulation, for each drug-protein complex, were saved in PDB format for analysis. These protein - ligand complexes were analyzed in detailed using open source *PyMol* molecular visualization tool (30) and fully automated *Protein-Ligand Interaction Profiler* (31) (<https://projects.biotec.tu-dresden.de/plip-web/plip>).

Results

Docking simulations with α -ketoamide inhibitor 13b

As a control experiment, we used *AutoDock* to reproduce the bound conformation of the α -ketoamide inhibitor 13b into the binding site of SARS-CoV2 M^{pro}, as observed in the X-ray crystal structure (13). For these docking simulations we made the inhibitor 13b rigid by disabling the torsion angle sampling protocol of *AutoDock*. All other LGA parameters were same as those used for the other inhibitor docking simulations. This protocol was used in 100 docking simulations, and resulted in a lowest-energy binding mode for 13b that is essentially identical to the crystal structure

1 conformation (Figure 4B). Comparative analysis of the docked inhibitor with the crystal
2 structure pose shows that all of the key protein ligand interactions are reproduced. In
3 this pose, the amide oxygen of inhibitor 13b forms hydrogen bonds with the backbone N
4 atoms of Gly143, Ser144 and Cys145, which form the canonical “oxyanion hole” of the
5 cysteine protease (13). As in the crystal structure of the complex, sidechain N^ε of His41
6 stabilizes the hydroxyl group of the thiohemiketal by forming a hydrogen bond. In
7 addition, the Cys143 S^γ – inhibitor C distance of 2.3 Å is close to that of the X-ray crystal
8 structure, 1.8 Å, where a “reversible” thiohemiketal covalent bond is formed between the
9 protein and this C atom of 13b. Residues Phe140 (carbonyl O), His163 (N^ε) and Glu166
10 (sidechain carboxylic O and backbone N) also form hydrogen bonds with the inhibitor in
11 the S1 pocket. Hydrophobic interactions with Pro168 and Gln189 sidechains observed
12 in the X-ray crystal structure are also reproduced. The *AutoDock* binding energy for this
13 binding mode is -7.19 kcal/mol (Table 1). This value reflects the non-covalent
14 energetics of the ligand conformation and the ligand-protein interactions in this pose
15 within the active site. It is a useful guide for comparative analysis of relative binding
16 energies for other ligands in our study. However, this binding energy does not include
17 covalent bond energies that may result from formation of the “reversible” thiohemiketal
18 S--C bond, which may contribute to further stabilization of the complex observed in the
19 crystal structure.

20
21 Next we calculated the binding of inhibitor 13b to M^{pro} allowing dihedral bond
22 rotations, as was done in the docking simulations of the other ligands (described below).
23 The best scoring docked conformation of 13b (Figure 4C) has an *AutoDock* binding
24 energy of -9.17 kcal/mol, much lower than the crystal structure conformation. Although
25 the bound conformation of the inhibitor is now significantly different from its

1 conformation in the crystal structure, key ligand-protein interactions are reproduced.
2 The sidechain of catalytic residue His41 (N^ϵ) again forms a hydrogen bond with the
3 hydroxyl group of the thiohemiketal. The backbone N atoms of Gly143 and Cys145 of
4 the oxyanion hole also again form hydrogen bonds with the amide oxygen of inhibitor
5 13b. The hydrophobic interactions with the sidechain of Met165 are also reproduced.
6 However, in this lower-energy pose, the short $S^\gamma - C$ distance consistent with formation
7 of the thiohemiketal covalent bond, is no longer observed. The distance between the
8 thiol S^γ of Cys145 and α -keto C atom of the inhibitor is 3.5 Å, compared to 1.8 Å in the
9 crystal structure. In the crystal structure, sidechain carboxylic O atom of Glu166 forms a
10 hydrogen bond with the P1 moiety. Due to the flip in conformation of the P1 and P1'
11 groups of inhibitor 13b (labeled in Fig. 4), Glu166 now interacts with P1' moiety
12 hydrophobically. His41, His164 and Phe181 also form other hydrophobic interactions
13 with the inhibitor that were not observed in the crystal structure. This analysis of the
14 lowest energy *AutoDock* pose with the crystal structure shows that the *AutoDock*
15 protocol used in this study is able to reproduce key interactions of the known inhibitor. It
16 appears, however, that in the absence of the $S^\gamma - C$ covalent bond between the ligand
17 and the protein, the most preferred docking mode is somewhat different from the
18 docking mode observed in the crystal structure.

19
20 We also assessed other low energy poses that have fewer differences from the
21 pose observed in the crystal structure that are included among these low energy flexible
22 docking poses. One such pose is shown in Figure 4C. Here, most of the key
23 interactions observed in the crystal structure are preserved, including a short $S^\gamma - C$
24 interatomic distance between catalytic residue Cys143 and the ligand of 3.0 Å. The
25 *AutoDock* binding energy of this pose is -9.03 kcal/mol (Table 1), only slightly higher in

energy than the lowest energy binding pose described above. Hence, binding poses very similar to that observed in the crystal structure are indeed included among the low energy poses generated by *AutoDock*, even without inclusion of the energetics of covalent bond formation.

Docking simulations with HCV NS3/4a protease inhibitors

Having validated the use of *AutoDock* for this system, docking simulation calculations were used to predict bound-state poses for several HCV NS3/4A protease inhibitor drugs. These molecules were compiled from public databases (Table 1). All of these drugs have been approved for at least Phase 1 clinical trials; some are FDA approved prescription drugs useful in treating HepC. *AutoDock* inhibitor-docking calculations were carried out for each ligand, using the same protocol tested with α -ketoamide inhibitor 13b.

The *AutoDock* score of the *best scoring pose* (i.e., lowest *AutoDock* binding energy) for each of 12 HCV NSP3/4A protease inhibitors are summarized in Table 1. These are compared with the *AutoDock* score of the X-ray crystal structure PDB id 6Y2G, the complex of SARS-CoV2 M^{pro} with α -ketoamide inhibitor 13b, and with 13b poses generated with the same docking protocols used for docking these drugs. Eight of the 12 molecules studied have more favorable binding scores than the α -ketoamide inhibitor 13b: *AutoDock* score -7.19 kcal/mol for crystal structure, and -9.17 kcal/mol for best scoring pose. For these 8 inhibitors the bound-state pose with best *AutoDock* score also fits well in the active site of SARS-CoV2 M^{pro}, and recapitulates many of the key ligand-protein interactions observed in the complex with α -ketoamide inhibitor 13b.

1 The lowest energy poses for each of the eight best scoring ligands are each
2 compared to that of the X-ray crystal structure of the α -ketoamide inhibitor 13b bound to
3 SARS-CoV2 M^{pro} protease in Figure 5-8, and details of intermolecular interactions are
4 also shown. A summary of key intermolecular contacts between the inhibitor and M^{pro}
5 in the “lowest energy” poses, for each complex, are presented in Table 2. In this
6 analysis, we paid particular attention to key details of the “lowest energy” pose,
7 including interactions with the sidechains of catalytic dyad residues His41 and Cys145,
8 and hydrogen-bonded interactions with the backbone amides of Gly143, Ser144 and
9 Cys145, which form the oxyanion hole of this cysteine protease (13). As far as we are
10 aware, none of these drugs form covalent complexes with M^{pro}. Some relevant details
11 about non-covalent ligand-protein interactions in the “lowest energy” poses obtained for
12 each of the 8 top-scoring drugs are outlined in the following paragraphs.

13
14 Docking simulations with **Sovaprevir** (ligand id SOV) provided the pose with the
15 lowest predicted binding energy, -11.01 kcal/mol. In this pose (Figure 5 – left side),
16 SOV adopts a conformation similar to inhibitor 13b in the binding cavity, with many
17 analogous interactions with M^{pro}. The ligand occupies the oxyanion hole of the
18 protease, forming hydrogen bonds to the backbone amides of Gly143, Ser144, and
19 Cys145. The N^ε atom of His41 also forms direct hydrogen bonds with the ligand. The
20 thiol sulfur of the catalytic residue Cys145 does not directly contact with the ligand;
21 however, Cys145 is occluded by the bound ligand which would hinder access to the
22 thiol by a substrate. Residues of the S1 substrate pocket Phe140, Met165 and Glu166
23 also interact with this ligand.

Vaniprevir (ligand id VAN) docking provided a best scoring pose with predicted binding energy of -10.95 kcal/mol. In this pose (Figure 5 – right side), VAN occupies a much a larger portion of the binding cavity than inhibitor 13b. This could be attributed to the large size and macrocyclic nature of VAN. However, VAN forms fewer hydrogen bonds with the protein than 13b. The His41 sidechain N^ε, Asn142 sidechain O^δ, Gly143 backbone amide N, and Glu166 backbone amide N atoms form hydrogen bonds with various moieties of the ligand. No atom of the ligand is found within 4 Å of the catalytic Cys145 thiol S; however, Cys145 is occluded by the binding pose. This VAN binding mode includes extensive hydrophobic interactions with M^{pro}, involving residues Thr25, Phe140, Asn142, Glu166 and Pro168 that contribute in hydrophobic stabilization of the complex.

The predicted binding energy of the best scoring complex of **Simeprevir** (ligand id SIM) is -10.75 kcal/mol. In this pose, SIM sits in the binding cavity differently than inhibitor 13b (Figure 6 – left side). The key atoms involved in hydrogen bonding with the ligand are Thr25 sidechain hydroxyl OH, His45 backbone carbonyl O, Cys44 backbone N, and Glu166 backbone N. In addition, sidechains of residues Thr24, Met165, Glu166, and Gln189 form hydrophobic interactions with SIM. This ligand does not occupy the binding site near the catalytic residue Cys145, but would occlude its access to the substrate.

Paritaprevir (ligand id PAR) is a macrocyclic compound like vaniprevir, and adopts a similar conformation when docked into SARS-CoV M^{pro} (Fig. 6 – right side). The predicted binding energy of this best scoring pose for PAR is -10.71 kcal/mol. In this pose PAR occupies the oxyanion hole of the protease, and is stabilized by

hydrogen bonds involving Asn142 sidechain O^δ, Gly143 backbone amide N, Ser144 backbone amide N, Cys145 backbone amide N, as well as His163 sidechain N^δ and Glu166 backbone N and carbonyl O, in the binding site of M^{pro}. The pose also appears to be stabilized by hydrophobic interactions involving residues Leu27, Gln189 and Gln192.

The best scoring binding pose for **Danoprevir** (ligand id DAN) in the active site of M^{pro} has a binding energy of -9.99 kcal/mol. DAN also binds to a more extensive surface than the α -ketoamide inhibitor 13b, but overlaps the binding mode of 13b in the oxyanion hole and nearby to the catalytic dyad (Fig. 7 – left side). However, the catalytic residues His41 and Cys145 do not directly interact with DAN with either hydrogen bonds or hydrophobic interactions. It appears that due to its large size and macrocyclic nature, DAN cannot access the deeper catalytic residues for direct interactions. DAN occupies the oxyanion hole of the protease, forming hydrogen bonds with Asn142 sidechain O^δ and Gly143 backbone amide N atom. In addition, backbone amide N atoms of Thr26 and Glu 166, and sidechain N^ε of His163 also form hydrogen bonds with DAN. The pose is also stabilized by hydrophobic interactions involving residues Thr25, Leu27, Asn142 and Glu166.

The binding energy of another macrocyclic compound **Grazoprevir** (GRZ) to M^{pro} is -9.71 kcal/mol. Similar to DAN, the catalytic residue His41 does not directly interact with GRZ with either hydrogen bonds or hydrophobic interactions. GRZ forms hydrogen bonds with backbone amide N atoms of Gly143, Ser144, catalytic Cys145 and Glu166 residues (Figure 7 – right side). Hydrophobic interactions with residues Thr25,

1 Leu27, Met165, Pro168 and Gln189 also contribute to the high binding score of this
2 ligand.

3
4 The binding energy of docked **Glecaprevir** (GLE) is -9.51 kcal/mol. Similar to
5 other macrocyclic HCV inhibitors, the catalytic Cys145 deeper in the binding pocket of
6 M^{pro} is not accessible to GLE (Figure 8 – left side). The ligand forms hydrogen bonds
7 with the sidechain N^ε atom of catalytic His41 along with backbone N of residues
8 Asn142, Gly143 and Glu166. Residues that form hydrophobic interactions are Thr25,
9 Leu141, Met165, Glu166.

10
11 **Boceprevir** binds in the active site of M^{pro} in a completely different conformation
12 than the other drugs (Figure 8 – right side), and has a *AutoDock* score of -9.44 kcal/mol.
13 It interacts with M^{pro} through hydrogen bonding with carbonyl oxygen of main chain
14 Phe140, backbone N of Gly143, Ser144, catalytic Cys145 and Glu166, and sidechain N^ε
15 atom of His163. It also forms hydrophobic contacts with residues His41, Leu141,
16 His164, Met165, Glu166 and Asp187.

17 18 **Docking simulations with HIV-1 protease inhibitors**

19 Approved therapeutics for AIDS, some of which inhibit the activity of the HIV-1
20 protease, are already being explored as therapeutics for COVID-19 in several clinical
21 trials (32-35). To compare their potential as antivirals directed at the M^{pro} protease of
22 SARS-CoV2 with the HCV drugs described above, we also carried out *AutoDock*
23 docking simulation calculations to predict and rank bound-state poses for nine of these
24 AIDS drugs. These HIV-1 inhibitors were compiled from public databases (Table 3).
25 Most of these drugs are (or have been) in clinical use as therapeutics for AIDS.

AutoDock inhibitor-docking calculations were carried out for each ligand, using the same protocol tested with the α -ketoamide inhibitor 13b and HCV NS3/4A protease inhibitors. For these simulations, 10 docking trajectories were carried out for each drug. The *AutoDock* scores of the *best scoring poses* (lowest binding energy), for each of 9 HIV-1 protease inhibitors are summarized in Table 3. The *AutoDock* binding energy scores for these molecules, ranging from -8.38 to -6.58 kcal/mol, were generally lower-ranking than those of the best scoring HCV inhibitors (cf. Tables 1 and 3). Nonetheless, some of these complexes have reasonable stability, and we cannot exclude the potential of using these HIV-1 protease inhibitors as therapeutics for COVID-19. For example, the inhibitor Ritonavir (ligand id RTV), together with the inhibitor Lopinivir (ligand id AB1), is part of the formulation marketed as Kaletra (AbbVie) that is currently in clinical trials as a COVID-19 antiviral (33). Ritonavir ranks relatively high on the HIV-1 inhibitor list, with a best predicted binding energy (-7.92 kcal/mol), comparable to the α -ketoamide inhibitor 13b.

Discussion

The global health, economic, and social impact of the COVID-19 pandemic is enormous. Although social distancing policies can slow the spread of the virus, effective control of the disease will require both vaccines and antivirals. Antivirals are crucial for the present COVID-19 pandemic to treat infected people during the period of time that an effective vaccine is being developed and deployed. It would be best to treat patients at the onset of infection, i.e., at, or near the time of the first positive COVID-19 test. As a result of such treatment, virus replication would be inhibited at the onset, thereby reducing the severity and duration of the disease, reducing virus spread, reducing the need for hospitalization, and reducing deaths. Antiviral drugs also have the

1 potential to be used prophylactically, to protect a population in danger of being exposed
2 to the virus. Effective antivirals would also be needed in the future. For example, it is not
3 known whether a new vaccine will provide complete and lasting immunity for everyone,
4 or whether, after enough of the human population gains protective antibodies against
5 the current virus, a mutant SARS-CoV2 virus will emerge that is no longer effectively
6 combated by a current vaccine. However, antivirals are not as effective in ameliorating
7 the disease of patients at later times of a disease after many rounds of virus replication
8 have already occurred, i.e., when patients are hospitalized with severe symptoms.
9 Nevertheless, particularly under the current emergency situation, antivirals should be
10 administered to hospitalized patients because even a relatively small amelioration could
11 save lives.

12
13 The discovery and development of antivirals is a slow and expensive process.
14 While inhibitors of enzymes like the SARS-CoV2 protease can be important lead
15 molecules for drug discovery, challenging issues of safety, delivery, pharmacokinetics,
16 formulation, stability, and manufacturing cost all need to be successfully addressed.
17 While several antiviral targets of the SARS-CoV2 life cycle have been identified,
18 including the Spike protein – ACE-2 receptor interaction, the viral RNA-dependent RNA
19 polymerase functions, and the papain-like and main (M^{pro}) viral proteases, the M^{pro}
20 protease is an especially attractive target because effective antivirals directed against
21 proteases of other viruses have been developed and used successfully (9-12).
22 Although the optimum inhibitors of the SARS-CoV2 M^{pro} protease for COVID-19 disease
23 are yet to be developed, the potential of using existing drugs which have been approved
24 for at least clinical testing is an extremely attractive first step in dealing with the very
25 urgent need for COVID-19 antivirals.

The structural similarity of the active sites of HCV NS3/4A and SARS-CoV2 M^{pro} proteases is striking and appears to be the result of convergent evolution from different fold topologies to create a similar binding pocket. Koonin and co-workers (36) previously proposed a divergent evolutionary relationship between the cysteine proteases of positive strand RNA viruses (e.g., the 3C proteases of picornaviruses) and chymotrypsin-like serine proteases. Subsequently, based on conserved sequence motifs characteristic of the active-site regions of 3C proteases, they also predicted that coronaviruses contain a protein (later identified as the main protease, M^{pro}) similar to the 3C proteases of picornaviruses (37). Anand et al (17) described the structural similarity of the active site of TGEV coronavirus M^{pro} protease to the active site of the 3C protease of the positive strand RNA virus, hepatitis A virus. The relative positions of the sidechains of catalytic residues Cys144 and His41 superimpose on the corresponding Cys172 and His44 sidechains in the three-dimensional structure of hepatitis A virus 3C^{pro} protease, and also on the sidechains of residues Ser195 and His57 in the three-dimensional structure of chymotrypsin. At that time, antiviral drugs for HCV were not yet available. Our recent DALI search of structures in the Protein Data Bank identified HCV NS3/4A protease (structural similarity Z score = + 8.4), as well as several other viral 3C proteases (Z scores ranging from + 7.4 to + 14.1), as structurally similar to SARS-CoV2 M^{pro} protease, including (among others) proteases of flaviviruses, noroviruses, and enteroviruses; e.g. coxsackievirus, hepatitis A, Norwalk, Zika, Dengue, and West Nile viruses. These results are relevant to considering existing drugs targeted to these proteases as potential COVID-19 antivirals. However, none of these results support a convergent evolutionary relationship between coronavirus M^{pro} and the HCV

1 NS3/4A protease, considering that they have very different structural topologies (Figure
2 3A).

3
4 Molecular docking is a widely used tool for modern structure-based drug
5 discovery. It is used not only to explore the binding conformation of lead molecules in
6 the active site of drug targets, but also to estimate the strength of interaction between
7 the ligand and target. The *AutoDock* program used in our study offers a variety of
8 search algorithms to recursively evaluate ligand conformations, and uses a force-field-
9 based scoring function to rank the binding poses. The accuracy of the program has
10 been tested with a diverse set of protein–ligand complexes of biological and medicinal
11 interest (38). Although predicted *AutoDock* binding energies may not be highly
12 accurate, the relative affinities within a series of ligands can generally be reliably
13 determined. Here, we docked the M^{pro} inhibitor 13b to validate the docking protocol that
14 we used. Using a rigid conformation of 13b, this protocol reproduced the X-ray crystal
15 structure as the lowest energy pose. Allowing flexible conformations of 13b in the
16 docking simulation, bound-state poses nearly identical to the crystal structure are also
17 observed among the low-energy poses in the active site. Generally, while the best-
18 scored *AutoDock* complex does not always match the experimentally determined
19 structure, the experimentally determined structure is generally among the best scoring
20 poses (26, 27). Accordingly, the best-ranked predictions illustrated in Figures 5-8
21 should capture key features of the ligand – protein interaction, but they might not be the
22 dominant pose observed in future experimental studies.

23
24 One limitation of the *AutoDock* protocol used here is the inability to model the
25 conformational flexibility of the protein target. This problem is typically approached

1 through the generation of multiple conformations of the protein by molecular dynamics
2 before docking, which is an important future direction for this work. In addition,
3 *AutoDock* scores do not take into account covalent bond formation which may occur
4 between an inhibitor and the protein. Hence, for molecules like inhibitor 13b which can
5 form such a covalent bond, the actual binding affinity may be higher than estimated
6 from these calculations. Our *AutoDock* calculations do, however, provide information
7 about the energetics of the pose in the binding site, and the proclivity of potential
8 inhibitors to occupy the active site. Even without considering the energetics of
9 thiohemiketal bond formation, essentially the same bound-state pose observed in the
10 crystal structure is also observed among the low energy poses of inhibitor 13b in the
11 active site. Interestingly, a recent study, Jin et al. (14) found that their best inhibitor of
12 the SARS-CoV2 M^{pro} did not form a covalent bond with M^{pro}, demonstrating that
13 formation of such a covalent bond is not required for efficient inhibition.

14
15 In the course of our study, a few other studies describing relevant high-
16 throughput virtual screening and/or docking studies of potential inhibitors of the SARS-
17 CoV2 M^{pro} protease have been posted in preprint archives. Beck et al. (7) reported
18 results using deep learning to identify commercially available drugs that might bind to
19 any of several SARS-CoV2 proteins. They reported that the HIV-1 protease inhibitors
20 atazanavir and ritonavir are potential therapeutics targeting M^{pro}. They also reported that
21 the HCV NS3/4A protease inhibitors asunaprevir and simeprevir are high-scoring
22 binders to the SARS-CoV2 M^{pro} protease, although, in contrast to our results, these
23 HCV inhibitors had lower affinities than the two HIV-1 protease inhibitors. Nguyen et al.
24 (8) applied mathematical representations in machine learning, and assessed the binding
25 of 1465 FDA-approved drugs in DrugBank to various SARS-CoV2 proteins. This study

1 identified several FDA-approved drugs as potential inhibitors of M^{pro}, including all of the
2 HIV-1 protease inhibitors listed in Table 3 of our manuscript, along with one HCV
3 protease inhibitor, boceprevir. Of these HIV-1 and HCV protease inhibitors, they
4 reported that the HCV NS3/4A protease inhibitor boceprevir had the highest predicted
5 binding affinity, consistent with our results, whereas we found other HCV protease
6 inhibitors that have even higher predicted binding affinity to M^{pro}. Generally speaking,
7 the evolving literature on virtual screening and computational docking supports the
8 hypothesis that available drugs targeting the active site of the HCV NS3/4A protease
9 (and perhaps also of the HIV-1 protease) have potential as quickly mobilized antivirals
10 for COVID-19.

11
12 There have already been a few clinical trials of HIV-1 and HCV protease
13 inhibitors for treatment of COVID-19 disease. The FDA-approved drug Kaletra (AbbVie),
14 which contains the HIV-1 protease inhibitors ritonavir and lopinavir, is currently being
15 assessed by the World Health Organization (WHO) for efficacy in COVID-19 infections
16 (32-35). Initial clinical trials were not encouraging (32), but involved subjects treated 10-
17 15 days after diagnosis who may have suffered from more advanced disease. Only one
18 of the antivirals listed in Table 1 that target the HCV NS3/4A protease, danoprevir,
19 marketed under the trade name Ganovo (Ascleptis Pharma), has been studied in a small
20 clinical trial for COVID-19 treatment (39). The results of this preliminary clinical trial
21 were promising. Clearly, well-designed clinical trials of HCV and HIV-1 protease
22 inhibitors are needed for patients who have recently been found to be COVID-19
23 positive, as well as for hospitalized patients with advanced disease.

This computational study was carried out to evaluate the hypothesis that existing drugs, already FDA-approved or approved for testing in clinical trials, have the potential to bind to the active site of the M^{pro} protease of the SARS-CoV2 virus. We focused primarily on existing drugs that bind to and inhibit the NS3/4A protease of HCV because we found that the active sites of the HCV and SARS-CoV2 proteases are structurally similar, and because there has been success in using protease inhibitors as therapeutics for HepC disease (9, 10). Our computational docking used as a benchmark state the structure of the complex formed between the SARS-CoV2 protease and a known inhibitor of this protease, an α -ketoamide molecule denoted as 13b (13). The results of our computational docking of HCV protease inhibitors into the active site of the M^{pro} protease of SARS-CoV2 support the hypothesis that several of these HCV inhibitors may effectively bind to and inhibit this protease. However, we do not know yet if any of these drugs are effective for treating COVID-19 disease. Hence, these results provide the rationale for us to test these identified HCV protease inhibitors as inhibitors of the SARS-CoV2 protease activity, and as inhibitors of SARS-CoV2 virus replication. Subsequently these repurposed drugs could be evaluated as COVID-19 therapeutics.

Acknowledgements

We thank Drs. T. Acton, G. Chalmers, A. Gibbs, Y.P. Huang, G. Liu, L. Ma, C. Sander, S. Shukla, G.V.T Swapna, and R. Xiao for helpful discussions, suggestions, and comments on the manuscript. This research was supported by the National Institutes of Health grants to RMK (R01-AI11772) and to GTM (R01-GM120574).

References

1. De Wit, E., Van Doremalen, N., Falzarano, D. and Munster, V. J. SARS and MERS: Recent insights into emerging coronaviruses. *Nat Rev Microbiol* **2016**, 14: 523-534.
<https://doi.org/10.1038/nrmicro.2016.81>
2. Zhou, P., Yang, X. L., Wang, X. G., Hu, B., Zhang, L., Zhang, W., Si, H. R., Zhu, Y., Li, B., Huang, C. L., et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* **2020**, 579: 270-273. <https://doi.org/10.1038/s41586-020-2012-7>
3. Wu, F., Zhao, S., Yu, B., Chen, Y. M., Wang, W., Song, Z. G., Hu, Y., Tao, Z. W., Tian, J. H., Pei, Y. Y., et al. A new coronavirus associated with human respiratory disease in China. *Nature* **2020**, 579: 265-269. <https://doi.org/10.1038/s41586-020-2008-3>
4. Gordon, D. E., Jang, G. M., Bouhaddou, M., Xu, J., Obernier, K., O'meara, M. J., Guo, J. Z., Swaney, D. L., Tummino, T. A., Huettenhain, R., et al. A SARS-CoV-2-human protein-protein interaction map reveals drug targets and potential drug-repurposing. *bioRxiv* **2020**.
<https://doi.org/10.1101/2020.03.22.002386>
5. Liu, X. and Wang, X. J. Potential inhibitors against 2019-nCoV coronavirus M protease from clinically approved medicines. *J Genet Genomics* **2020**, 47: 119-121.
<https://doi.org/10.1016/j.jgg.2020.02.001>
6. Martinez, C. M. and Lezama, A. R. Identification of potential inhibitors of SARS-CoV-2 Main protease via a rapid in-silico drug repurposing approach. *ChemRxiv* **2020**.
<https://doi.org/10.26434/chemrxiv.12085083.v1>
7. Beck, B. R., Shin, B., Choi, Y., Park, S. and Kang, K. Predicting commercially available antiviral drugs that may act on the novel coronavirus (2019-nCoV), Wuhan, China through a drug-target interaction deep learning model. *bioRxiv* **2020**: 2020.2001.2031.929547.
<https://doi.org/10.1101/2020.01.31.929547>
8. Nguyen, D. D., Gao, K., Chen, J., Wang, R. and Wei, G.-W. Potentially highly potent drugs for 2019-nCoV. *bioRxiv* **2020**. <https://doi.org/10.1101/2020.02.05.936013>
9. Kwo, P. Y. and Vinayek, R. The therapeutic approaches for hepatitis C virus: protease inhibitors and polymerase inhibitors. *Gut Liver* **2011**, 5: 406-417.
<https://doi.org/10.5009/gnl.2011.5.4.406>
10. Mcgovern, D. R., Masaki, T., Lovell, W., Hamlett, C., Saalau-Bethell, S. and Graham, B. Protease inhibitors block multiple functions of the NS3/4A protease-helicase during the hepatitis C virus life cycle. *J Virol* **2015**, 89: 5362-5370.
<https://doi.org/10.1128/JVI.03188-14>
11. Ghosh, A. K., Osswald, H. L. and Prato, G. Recent progress in the development of HIV-1 protease inhibitors for the treatment of HIV/AIDS. *J Med Chem* **2016**, 59: 5172-5208.
<https://doi.org/10.1021/acs.jmedchem.5b01697>

12. Wlodawer, A. and Vondrasek, J. Inhibitors of HIV-1 protease: a major success of structure-assisted drug design. *Annu Rev Biophys Biomol Struct* **1998**, 27: 249-284.
<https://doi.org/10.1146/annurev.biophys.27.1.249>
13. Zhang, L., Lin, D., Sun, X., Curth, U., Drosten, C., Sauerhering, L., Becker, S., Rox, K. and Hilgenfeld, R. Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved alpha-ketoamide inhibitors. *Science* **2020**.
<https://doi.org/10.1126/science.abb3405>
14. Jin, Z., Du, X., Xu, Y., Deng, Y., Liu, M., Zhao, Y., Zhang, B., Li, X., Zhang, L., Peng, C., et al. Structure of M(pro) from COVID-19 virus and discovery of its inhibitors. *Nature* **2020**.
<https://doi.org/10.1038/s41586-020-2223-y>
15. Anand, K., Ziebuhr, J., Wadhwani, P., Mesters, J. R. and Hilgenfeld, R. Coronavirus main proteinase (3CLpro) structure: basis for design of anti-SARS drugs. *Science* **2003**, 300: 1763-1767. <https://doi.org/10.1126/science.1085658>
16. Yang, H., Yang, M., Ding, Y., Liu, Y., Lou, Z., Zhou, Z., Sun, L., Mo, L., Ye, S., Pang, H., et al. The crystal structures of severe acute respiratory syndrome virus main protease and its complex with an inhibitor. *Proc Natl Acad Sci U S A* **2003**, 100: 13190-13195.
<https://doi.org/10.1073/pnas.1835675100>
17. Anand, K., Palm, G. J., Mesters, J. R., Siddell, S. G., Ziebuhr, J. and Hilgenfeld, R. Structure of coronavirus main proteinase reveals combination of a chymotrypsin fold with an extra alpha-helical domain. *EMBO J* **2002**, 21: 3213-3224.
<https://doi.org/10.1093/emboj/cdf327>
18. Holm, L. and Sander, C. Protein structure comparison by alignment of distance matrices. *J Mol Biol* **1993**, 233: 123-138. <https://doi.org/10.1006/jmbi.1993.1489>
19. Holm, L. and Sander, C. Protein folds and families: sequence and structure alignments. *Nucleic Acids Res* **1999**, 27: 244-247. <https://doi.org/10.1093/nar/27.1.244>
20. Orengo, C. A., Michie, A. D., Jones, S., Jones, D. T., Swindells, M. B. and Thornton, J. M. CATH--a hierarchic classification of protein domain structures. *Structure* **1997**, 5: 1093-1108. [https://doi.org/10.1016/s0969-2126\(97\)00260-8](https://doi.org/10.1016/s0969-2126(97)00260-8)
21. Kozisek, M., Lepsik, M., Grantz Saskova, K., Brynda, J., Konvalinka, J. and Rezacova, P. Thermodynamic and structural analysis of HIV protease resistance to darunavir - analysis of heavily mutated patient-derived HIV-1 proteases. *FEBS J* **2014**, 281: 1834-1847.
<https://doi.org/10.1111/febs.12743>
22. Rut, W., Groborz, K., Zhang, L., Modrzycka, S., Poreba, M., Hilgenfeld, R. and Drag, M. Profiling of flaviviral NS2B-NS3 protease specificity provides a structural basis for the development of selective chemical tools that differentiate Dengue from Zika and West Nile viruses. *Antiviral Res* **2020**, 175: 104731.
<https://doi.org/10.1016/j.antiviral.2020.104731>

23. Zhang, L., Lin, D., Kusov, Y., Nian, Y., Ma, Q., Wang, J., Von Brunn, A., Leyssen, P., Lanko, K., Neyts, J., et al. alpha-Ketoamides as broad-spectrum inhibitors of Coronavirus and Enterovirus replication: Structure-based design, synthesis, and activity assessment. *J Med Chem* **2020**. <https://doi.org/10.1021/acs.jmedchem.9b01828>
24. Hilgenfeld, R. and Peiris, M. From SARS to MERS: 10 years of research on highly pathogenic human coronaviruses. *Antiviral Res* **2013**, 100: 286-295. <https://doi.org/10.1016/j.antiviral.2013.08.015>
25. Morris, G. M., Huey, R., Lindstrom, W., Sanner, M. F., Belew, R. K., Goodsell, D. S. and Olson, A. J. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J Comput Chem* **2009**, 30: 2785-2791. <https://doi.org/10.1002/jcc.21256>
26. Kolb, P., Ferreira, R. S., Irwin, J. J. and Shoichet, B. K. Docking and chemoinformatic screens for new ligands and targets. *Curr Opin Biotechnol* **2009**, 20: 429-436. <https://doi.org/10.1016/j.copbio.2009.08.003>
27. Kolb, P. and Irwin, J. J. Docking screens: right for the right reasons? *Curr Top Med Chem* **2009**, 9: 755-770. <https://doi.org/10.2174/156802609789207091>
28. Sanner, M. F. Python: a programming language for software integration and development. *J Mol Graph Model* **1999**, 17: 57-61.
29. Morris, G. M., Goodsell, D. S., Halliday, R. S., Huey, R., Hart, W. E., Belew, R. K. and Olson, A. J. Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *J Comput Chem* **1998**, 19: 1639-1662. [https://doi.org/10.1002/\(SICI\)1096-987X\(19981115\)19:14<1639::AID-JCC10>3.0.CO;2-B](https://doi.org/10.1002/(SICI)1096-987X(19981115)19:14<1639::AID-JCC10>3.0.CO;2-B)
30. Delano, W. L. The PyMOL Molecular Graphics System, Version 1.2r3pre, Schrödinger, LLC. **2009**.
31. Salentin, S., Schreiber, S., Haupt, V. J., Adasme, M. F. and Schroeder, M. PLIP: fully automated protein-ligand interaction profiler. *Nucleic Acids Res* **2015**, 43: W443-447. <https://doi.org/10.1093/nar/gkv315>
32. Cao, B., Wang, Y., Wen, D., Liu, W., Wang, J., Fan, G., Ruan, L., Song, B., Cai, Y., Wei, M., et al. A trial of lopinavir-ritonavir in adults hospitalized with severe COVID-19. *N Engl J Med* **2020**. <https://doi.org/10.1056/NEJMoa2001282>
33. Kupferschmidt, K. and Cohen, J. Race to find COVID-19 treatments accelerates. *Science* **2020**, 367: 1412-1413. <https://doi.org/10.1126/science.367.6485.1412>
34. Baden, L. R. and Rubin, E. J. COVID-19 - The search for effective therapy. *N Engl J Med* **2020**. <https://doi.org/10.1056/NEJMe2005477>
35. Young, B. E., Ong, S. W. X., Kalimuddin, S., Low, J. G., Tan, S. Y., Loh, J., Ng, O. T., Marimuthu, K., Ang, L. W., Mak, T. M., et al. Epidemiologic features and clinical course of patients infected with SARS-CoV-2 in Singapore. *JAMA* **2020**. <https://doi.org/10.1001/jama.2020.3204>

- 1
2 36. Gorbalenya, A. E., Donchenko, A. P., Blinov, V. M. and Koonin, E. V. Cysteine proteases
3 of positive strand RNA viruses and chymotrypsin-like serine proteases. A distinct protein
4 superfamily with a common structural fold. *FEBS Lett* **1989**, 243: 103-114.
5 [https://doi.org/10.1016/0014-5793\(89\)80109-7](https://doi.org/10.1016/0014-5793(89)80109-7)
6
7 37. Gorbalenya, A. E., Koonin, E. V., Donchenko, A. P. and Blinov, V. M. Coronavirus
8 genome: prediction of putative functional domains in the non-structural polyprotein by
9 comparative amino acid sequence analysis. *Nucleic Acids Res* **1989**, 17: 4847-4861.
10 <https://doi.org/10.1093/nar/17.12.4847>
11
12 38. Forli, S., Huey, R., Pique, M. E., Sanner, M. F.,Goodsell, D. S. and Olson, A. J.
13 Computational protein–ligand docking and virtual drug screening with the AutoDock suite.
14 *Nature Protocols* **2016**, 11: 905. <https://doi.org/10.1038/nprot.2016.051>
15
16 39. Chen, H., Zhang, Z., Wang, L., Huang, Z., Gong, F., Li, X., Chen, Y. and Wu, J. J. First clinical
17 study using HCV protease inhibitor danoprevir to treat naïve and experienced COVID-19
18 patients. *MedRciv* **2020**. <https://doi.org/10.1101/2020.03.22.20034041>
19
20
21

1 **Table 1. SARS-CoV2 M^{pro} and HepC 3C/4A Protease Inhibitors**

Inhibitor (Trade Name)	Identifier of Protease Inhibitor	Database ID of Protease Inhibitor Structure	AutoDock Score (kcal/mol) Lowest “Energy”	Drug Status
<u>SARS-CoV2 M^{pro} Inhibitor</u>				
α-ketoamide inhibitor 13b (no dihedral rotations)	06K	6Y2G ^a	-7.19	Not Applicable
α-ketoamide inhibitor 13b lowest “energy” pose: pose most similar to X-ray structure:	06K	6Y2G ^a	-9.17 -9.03	Not Applicable
<u>HCV NSP3/4A Protease Inhibitor Drugs</u>				
Sovaprevir (ACH-1625; Achillion)	SOV	28529313 ^b	-11.01	Investigational
Vaniprevir (MK-7009; Merck)	VAN	3SU3 ^c	-10.95	Investigational
Simeprevir (Olysio; Medivir / Janssen)	SIM	3KEE ^c	-10.75	Prescription Drug
Paritaprevir (Veruprevir / ABT-450; Abbot)	PAR	32700634 ^b	-10.71	Prescription Drug
Danoprevir (Ganovo; Array / Pfizer, Roche / Ascletis)	DAN	3M5L ^c	-9.99	Investigational
Grazoprevir (Zepatier; Merck)	GRZ	3SUD ^c	-9.71	Prescription Drug
Glecaprevir (Mavyret ^d / Maviret ^d ; AbbVie / Enanta)	GLE	35013015 ^b	-9.51	Prescription Drug
Boceprevir (Victrelis; Merck)	BOC	DB08873 ^e	-9.44	Prescription Drug
Asunaprevir (Sunvepra; Bristol-Myers Squibb)	ASU	4WF8 ^c	-8.19	Investigational
Telaprevir (Incivek / Incivo; Vertex / J&J)	TEL	3SV6 ^c	-7.56	Prescription Drug
Narlaprevir (Arlansa; Merck / R-Pharm)	NAR	3LON ^c	-7.09	Prescription Drug
Faldaprevir (Fadaprevir, Boehringer-Ingelheim)	FAL	3P8N ^c	-6.18	Investigational

2 (a) Atomic coordinates for the inhibitor taken from 6Y2G. (b) Atomic coordinates for the inhibitor were taken from the
3 ChemSpider database. (c) Atomic coordinates for the inhibitor were taken from the PDB coordinates of the
4 corresponding complex of the inhibitor bound to HCV 3C/4A protease. (d) Mavyret (or Maviret) is a multidrug
5 formulation including glecaprevir and pibrentasvir. (e) Atomic coordinates for the inhibitor were taken from the DrugBank
6 database.

1
2**Table 2. Key Interactions within the SARS-CoV2 M^{pro} Protease Active Site**

Inhibitor (Trade Name ; Manufacturer)	Identifier of Protease Inhibitor	Atoms Forming Key Hydrogen Bonds	Residues Forming Key Hydrophobic Interactions
<u>SARS-CoV2 M^{pro} Inhibitor</u>			
α-ketoamide inhibitor 13b (X-ray crystal structure PDB id 6Y2G)	06K	His41(N ^ε), Gly143(N), Ser144(N), Cys145(N), His163(N ^ε), His164(O), Glu166(O ^ε)	Asn142, Met165, Asp187, Gln189
α-ketoamide inhibitor 13b (no dihedral rotations)	06K	His41(N ^ε), Gly143(N), Ser144(N, O ^γ), Cys145(N), His163(N ^ε), Glu166(N, O ^ε)	Asn142, Met165, Pro168, Gln189
α-ketoamide inhibitor 13b (lowest “energy” pose)	06K	Thr26(N), His41(N ^ε), Gly143(N), Cys145(N)	His41, His164, Met165, Glu166, Phe181
α-ketoamide inhibitor 13b (pose most similar to X-ray structure)	06K	His41(N ^ε), Gly143(N), Glu166(N, O)	Thr25, Phe140, Met165, Leu167, Pro168, Gln189, Gln192
<u>HCV NSP3/4A Protease Inhibitor Drugs</u>			
Sovaprevir (ACH-1625; Achillion)	SOV	Thr26(O), His41(N ^ε), Gly143(N), Ser144(N), Cys145(N), Glu166(N)	Leu27, Phe140, Met165, Gln189
Vaniprevir (MK-7009; Merck)	VAN	His41(N ^ε), Asn142(O ^δ), Gly143(N), Glu166(N)	Thr25, Phe140, Asn142, Glu166, Pro168
Simeprevir (Olysio; Medivir / Janssen)	SIM	Thr25 (O ^γ), His41(O), Cys44(N), Glu166 (C ^β)	Thr24, Met165, Glu166, Gln189
Paritaprevir (ABT-450; Abbot)	PAR	Asn142(O ^δ), Gly143(N), Ser144(N), Cys145(N), His163 (N ^δ), Glu166(O,N)	Leu27, Gln189, Gln192
Danoprevir (Ganovo; Array / Pfizer, Roche / Ascleptis)	DAN	Thr26(N), Asn142(O ^δ), Gly143(N), His163(N ^ε), Glu166(N)	Thr25, Leu27, Asn142, Glu166
Grazoprevir (Zepatier; Merck)	GRZ	Gly143(N), Ser144(N), Cys145(N), Glu166(N)	Thr25, Leu27, Met165, Pro168, Gln189
Glecaprevir (Mavyret / Maviret; AbbVie)	GLE	His41(N ^ε), Asn142 (N), Gly143(N), Glu166(N)	Thr25, Leu141, Met165, Glu166
Boceprevir (Victrelis; Merck)	BOC	Phe140(O), Gly143(N), Ser144(N), Cys145(N), His163(N ^ε), Glu166(N)	His41, Leu141, His164, Met165, Glu166, Asp187

3

Table 3. HIV-1 Protease Inhibitor Drugs

Inhibitor (Trade Name ; Manufacturer)	Identifier of Protease Inhibitor	Database ID of Protease Inhibitor Structure ^a	<i>AutoDock</i> Score (kcal/mol) Lowest “Energy”	Drug Status
<u>HIV-1 Protease Inhibitor Drugs</u>				
Nelfinavir (Viracept; Agouron / Lilly / Roche / ViiV Healthcare)	NFV	3EKX	-8.83	Prescription Drug
Saquinavir (Invirase / Fortovase; Roche)	SQV	3OXC	-8.67	Prescription Drug
Indinavir (Crixivan; Merck)	IDV	2R5P	-8.39	Investigational
Ritonavir (Norvir / Kaletra ^b ; AbbVie)	RTV	2B60	-7.92	Prescription Drug
Darunavir (Prezista; Tibotec / Janssen)	DRV	4LL3	-7.47	Prescription Drug
Tipranavir (Aptivus; Boehringer Ingelheim)	TPV	6DIF	-7.34	Prescription Drug
Atazanavir Reyataz / Evotaz; Bristol-Myers Squibb	ATA	2AQU	-7.04	Prescription Drug
Lopinavir (Kaletra ^b ; AbbVie)	LOP	1MUI	-6.71	Prescription Drug
Amprenavir (Agenerase, GlaxoSmithKline)	478	3EKV	-6.58	Prescription Drug

(a) Atomic coordinates for the inhibitor were taken from the PDB coordinates of the corresponding complex of the inhibitor bound to HCV 3C/4A protease. (b) Kaletra is a formulation including both lopinavir and ritonavir protease inhibitors.

Fig. 1. X-ray crystal structures of viral proteases. A) SARS-CoV2 M^{pro} (PDB id 6Y2G), B) HCV protease NS3/4A (PDB id 2P59), and C) HIV-1 protease (PDB id 4LL3). HCV NS3/4A protease and SARS-CoV2 M^{pro} both have a double β -barrel fold architecture, with a substrate binding site located in a shallow cleft between two antiparallel β -barrels (shown in cyan and blue). Only one protomer is shown for A and B. The α -helical domain III of M^{pro} is shown in green. The HIV protease has a different fold architecture, and the substrate binding site is located in between the two protomers. For the HIV-1 protease both protomers are shown. The structures of bound inhibitors in these crystal structures are illustrated as magenta sticks.

Fig. 2. Superimposition of viral proteases. The backbone structure of the SARS-CoV2 M^{pro}, PDB 6Y2G (green) is superimposed on the backbone structure of hepatitis C virus protease NS3/4A, PDB 2P59 (cyan). The regions identified by *DALI* (18) as structurally-analogous are shown in color (green and cyan), and the regions that are not structurally-analogous are shown in gray. This superimposition of backbone atoms results in superimposition of the catalytic residues Cys145 and His41 of the SARS-CoV2 M^{pro} with Ser139 and His57 of HCV protease. Residue Asp81 of the HCV protease catalytic triad is also shown.

B

```

DSSP  -----lLLLEeeellllhhhhhhhhhhllllllllLLEEEEEEllLLEEEEEEllLLEEEEE
HCV   -----aPITAYaqqtrgllgciitsltgrdknqveGEVQIVSTATQTFLATCINGVCWTV  55
                                     |   |   |   |   |
CoV2  sgfrkmAFPSPG-----kveGCMVQVTCGTTTLNGLWLDDVVYCP  39
DSSP  llllllLllLH-----hhhLLEEEEEEllLLEEEEEEllLLEEEEE

DSSP  HHHHL-----lLLLL-----llllLllLLEEEllLLEEEEE--lLLLL
HCV   YHGAG-----tRTIA-----spkgPVIQMYTNvdQLVGWP--aPQGS  91
      |               ||
CoV2  RHVICtsedmlnpnyedllirKSNHnflvqagnvqlRVIGHSMQ--NCVLKLKvdtANPK  97
DSSP  HHHHLllllllllllhhhhhhllLHHHlleeeelleeeLLEEEEEE--LLEEEEEEellllLL

DSSP  L-lLLLLLLL-lLEEEELL---LLEEEEEEllLLEEEEEEehhHHLLllLLEEEEL
HCV   R-sLTPCTCG-sSDLYLVTR---HADVIPVRRRGdsRGSLLsprpisYLKGSGGPLLC  145
                                     |   |   |   |
CoV2  TpkYKFVRIQpgQTFSVLACyngsPSGVYQCAMPNfTIKG-----sFLNGSGSVGFN  151
DSSP  LleEEELLllLlLLEEEEEEelleEEEEEEEEllLllLllL---lLLLLllLLEEEE

DSSP  LL---LLEEEEEEELLlleeEEEEEEHHHH-----hhhhhll
HCV   PA---GHAVGLFRAAVCTrgvaKAVDFIPVENL-----ettmrs  181
      |               |
CoV2  IDydcVSFCYMHMELPT---GVHAGTDLEGNfygpfvdrqtaqaagtdtti  200
DSSP  EEllleEEEEEEEEELL---LLEEEELLllllllllllllllllllllllllllllll

```

Fig. 3. Comparison of fold topology diagrams and structure-based sequence alignments of SARS-CoV2 M^{pro} and HCV NS3/4A proteases. A) Fold topology diagrams demonstrate similar fold architectures but different topologies. B) The DALI server (18) provides a structure-based sequence alignment of HCV NS3/4A (HCV) and SARS-CoV2 M^{pro} (CoV2). Catalytic residues of HCV NS3/4A (His57, Asp81 and Ser139) and SARS-CoV2 M^{pro} (His41 and Cys145) are highlighted in bold red. Three-state secondary structure definitions by DSSP (19) (H=helix, E=sheet, L=coil) are shown above each amino acid sequence. Structurally equivalent residues are in uppercase, structurally non-equivalent residues (e.g. in loops) are in lowercase. Amino acid identities are marked by vertical bars. The structure-based alignment results in alignment of key catalytic residues His41 and Cys145 of the SARS-CoV2 M^{pro} with His57 and Ser139 of HCV NS3/4A protease.

A

α -ketoamide inhibitor 13b

13b Crystal structure conformation

13b Docked conformation

Fig. 4. Docking of α -ketoamide protease inhibitor 13b (13) to SARS-CoV2 M^{pro}. A) Grid box placed around reference ligand 13b to define the binding site. B) Lowest energy *AutoDock* pose using a rigid conformation of α -ketoamide inhibitor 13b, compared with the bound-state ligand conformation observed in the X-ray crystal structure of the complex (score = -7.19 kcal/mol). C) Lowest energy *AutoDock* pose observed among 100 docking simulations (score = -9.17 kcal/mol). D) Lowest energy *AutoDock* pose that is most similar to the conformation seen in the crystal structure (score = -9.03 kcal/mol). In each of panels B-D, the M^{pro} protein is shown in surface representation (gray), together with the X-ray crystal structure of α -ketoamide inhibitor 13b bound in the active site of M^{pro} (green sticks), and the predicted *AutoDock* conformation (magenta sticks).

Sovaprevir

Vaniprevir

Fig. 5. Docking of HCV protease NS3/4A inhibitor drugs to SARS CoV2 M^{pro}. Top panels - Molecular structures of two HCV protease inhibitor drugs. Middle panels – Lowest energy *AutoDock* pose of these HCV protease inhibitors (magenta sticks) in the SARS CoV2 M^{pro} active site, compared to the pose observed in the crystal structure (PDB id 6Y2G) available for the SARS-CoV2 M^{pro} α -ketoamide inhibitor 13b (green sticks). Bottom panels – Details of atomic interactions in the lowest energy *AutoDock* poses of these HCV protease inhibitors. Hydrogen bonds and hydrophobic interactions between the drug and the enzyme are shown with yellow solid lines and black dashed lines, respectively. Sidechains of catalytic residues His41 and Cys145 are labeled, along with other protein residues that form key interactions with these drugs.

Simeprevir

Paritaprevir

Fig. 6. Docking of HCV protease NS3/4A inhibitor drugs to SARS CoV2 M^{pro}. Top panels - Molecular structures of two HCV protease inhibitor drugs. Middle panels – Lowest energy *AutoDock* pose of these HCV protease inhibitors (magenta sticks) in the SARS CoV2 M^{pro} active site, compared to the pose observed in the crystal structure (PDB id 6Y2G) available for the SARS-CoV2 M^{pro} α -ketoamide inhibitor 13b (green sticks). Bottom panels – Details of atomic interactions in the lowest energy *AutoDock* poses of these HCV protease inhibitors. Hydrogen bonds and hydrophobic interactions between the drug and the enzyme are shown with yellow solid lines and black dashed lines, respectively. Sidechains of catalytic residues His41 and Cys145 are labeled, along with other protein residues that form key interactions with these drugs.

Danoprevir

Grazoprevir

Fig. 7. Docking of HCV protease NS3/4A inhibitor drugs to SARS CoV2 M^{pro}. Top panels - Molecular structures of two HCV protease inhibitor drugs. Middle panels – Lowest energy *AutoDock* pose of these HCV protease inhibitors (magenta sticks) in the SARS CoV2 M^{pro} active site, compared to the pose observed in the crystal structure (PDB id 6Y2G) available for the SARS-CoV2 M^{pro} α -ketoamide inhibitor 13b (green sticks). Bottom panels – Details of atomic interactions in the lowest energy *AutoDock* poses of these HCV protease inhibitors. Hydrogen bonds and hydrophobic interactions between the drug and the enzyme are shown with yellow solid lines and black dashed lines, respectively. Sidechains of catalytic residues His41 and Cys145 are labeled, along with other protein residues that form key interactions with these drugs.

Glecaprevir

Boceprevir

Fig. 8. Docking of HCV protease NS3/4A inhibitor drugs to SARS CoV2 M^{pro}. Top panels - Molecular structures of two HCV protease inhibitor drugs. Middle panels – Lowest energy *AutoDock* pose of these HCV protease inhibitors (magenta sticks) in the SARS CoV2 M^{pro} active site, compared to the pose observed in the crystal structure (PDB id 6Y2G) available for the SARS-CoV2 M^{pro} α -ketoamide inhibitor 13b (green sticks). Bottom panels – Details of atomic interactions in the lowest energy *AutoDock* poses of these HCV protease inhibitors. Hydrogen bonds and hydrophobic interactions between the drug and the enzyme are shown with yellow solid lines and black dashed lines, respectively. Sidechains of catalytic residues His41 and Cys145 are labeled, along with other protein residues that form key interactions with these drugs.