Reactivity-guided de novo molecular design and high throughput virtual screening of a targeted library of peptidomimetic compounds reveals charge-based structure activity relationship of potential covalent inhibitors of the main protease of SARS-CoV-2

Stephanie Sun1, Kavya Anand2, Ishani Ashok3, Bhavesh Ashok4, Ayush Bajaj5, Varsha Beldona6, Kushal Chattopadhyay3, Audrey Kwan6, Karankumar Mageswaran3, Anvi Surapaneni5, Atri Surapaneni5, Pranjal Verma2, Allen Chen3, Ria Kolala3, Andrew Liang7, Ayeeshi Poosarla3, Krithikaa Premnath2, Karthikha Sri Indran3, Jeslyn Wu3, Aishwarya Yuvaraj8, Harsha Raj9, Tanish Sathish10, Aashi Shah4, Sarah Su11, Kara Tran6, Edward Njoo12*

1 BASIS Independent Silicon Valley, San Jose, CA, 95126
2 Dougherty Valley High School, San Ramon, CA, 94582
3 Mission San Jose High School, Fremont, CA, 94539
4 Amador Valley High School, Pleasanton, CA, 94566
5 The Quarry Lane School, Dublin, CA, 94568
6 Dublin High School, Dublin, CA, 94568
7 The College Preparatory School, Oakland, CA, 94618
8 American High School, Fremont, CA, 94536
9 Fremont Christian High School, Fremont, CA, 94536
10 Irvington High School, Fremont, CA, 94538
11 Los Altos High School, Los Altos, CA, 94022
12 Department of Chemistry, Biochemistry, and Physical Science; Aspiring Scholars Directed Research Program, Fremont, CA, 94539
*Corresponding Author: stephaniesun02@gmail.com

ABSTRACT

In December of 2019, a novel coronavirus was first identified in Wuhan, China, and has since spread around the world, leaving a largely unsolved biomedical problem in its wake. Upon entry into host cells, the main protease is essential for the replication of viral RNA, which is what allows the virus to replicate inside humans. Inhibition of the main protease has been investigated as a potential strategy for inhibition of the viral replication cycle. Here, we designed a combinatorial library of small molecules and performed high-throughput virtual screening to identify a series of hit compounds that may serve as potential inhibitors of the main protease. In our design of covalent inhibitors of the coronavirus protease, we modeled a library of 361 peptidomimetic Michael acceptor small molecules, which are designed to engage the nucleophilic cysteine residue in the active site of the protease in an irreversible 1,4-conjugate addition. We then employed a variety of computational tools to determine the binding affinity of our designed compounds when bound to the protease active site, where we determined that cationic side chains are potentially beneficial for inhibition of SARS-CoV-2.
Introduction

The novel coronavirus, Severe Acute Respiratory Coronavirus 2 (SARS-CoV-2), was first identified in late 2019 in Wuhan, China, and the associated disease, later termed COVID-19, was discovered to cause respiratory infections and in more severe cases, pneumonia and death in humans [1]. SARS-CoV-2 has caused a global pandemic that has thrown countries worldwide into a state of disorder and has presented a formidable biomedical, economic, and societal problem for billions worldwide. As of June 21, 2020, over 8.8 million COVID-19 cases have been reported, with the death toll surpassing 450,000. The COVID-19 pandemic presents the need for urgent identification of inhibitors and vaccines [2].

SARS-CoV-2 belongs to a family of $\beta$-coronaviruses, which are characterized by an enveloped, single-stranded positive-sense RNA genome [3]. The novel coronavirus was determined to be similar to two zoonotic coronaviruses that emerged in the 20th century, SARS-CoV and MERS-CoV, both of which also caused respiratory infections in humans [4]. Upon entry into the human host cell, the genome of the coronavirus is translated into two polyproteins, which are processed by the main protease (M$^{\text{pro}}$) and papain-like proteases into non-structural proteins (nsps). These nsps allow for the production of RNA that encode four main structural proteins (envelope (E), membrane (M), spike (S), nucleocapsid (N)) and other accessory proteins (Figure 1) [5, 6]. Because the main protease is essential to the development and replication of SARS-CoV-2, and there are no similar proteases in the human body, the M$^{\text{pro}}$ is an ideal target for antiviral therapies [7].

Figure 1. Mechanism of action of SARS-CoV-2. The spike glycoprotein of the virus mediates entry into human cells through ACE2 receptors on human lung cells, allowing the virus to use the cell’s resources to replicate.

The main protease of SARS-CoV-2, which is well conserved across all coronaviruses, is a dimer consisting of two monomers, each of which has three domains, and a catalytic dyad, made up of His41 and Cys145, between domains I and II (residues 10-99 and 100-182, respectively), which also includes the substrate binding site (Figure 2) [8]. The dimer that is formed by the SARS-CoV-2 main protease has a contact interface between domain II of monomer A and the NH$_2$-terminal residues (“N finger”) of monomer B, which is necessary for catalytic activity because the interaction between the two monomers allows for the proper orientation of the substrate binding pocket [9].
As of now, there are no approved antiviral drugs against SARS-CoV-2, though there have been a series of recent attempts to target proteins essential to coronavirus entry or replication by performing high-throughput screenings on large libraries of molecules with the aim of identifying a few possible lead compounds [10]. We hypothesized that this approach could be applied in the design of novel chemical entities towards inhibition of the main protease of the coronavirus.

Previous efforts towards the design of a covalent inhibitor of SARS-CoV-2 have targeted the cysteine residue in the substrate binding pocket of Mpro, where an α-ketoamide inhibitor was designed based on the nucleophilic addition of Cys145 onto the α-keto group of the inhibitor and screened computationally [11]. Michael additions for irreversible binding of inhibitors to the main protease of viruses have been previously studied, where a peptidic α, β-unsaturated ester served as the Michael acceptor and demonstrated antiviral activity in cell cultures [12].

Here, we report the rational design and high-throughput virtual screening of a library of targeted compounds towards the inhibition of SARS-CoV-2. We modeled nearly four hundred designed inhibitors, screened them computationally to identify the structure-activity relationship (SAR) between side chain structures on our designed inhibitors, and predicted binding affinities to the Mpro active sites from molecular docking.

In the substrate-binding pocket of Mpro, the thiol side chain of Cys145 (Figure 2) is able to covalently bind to inhibitors, which prevents catalytic activity. The design of our inhibitors is inspired by a Michael Addition reaction: a 1,4 conjugate addition of a nucleophile (such as the primary thiol in cysteine) to an α,β-unsaturated carbonyl or analogous functional group (Figure 3).

Michael additions with dehydroalanine have been reported with thiols and have also been reported to inhibit enzyme activity through irreversible formation of a covalent bond to the active site nucleophilic residue [13, 14].
Our library of inhibitors were biologically inspired by a tripeptide (1), with structural modifications shown in Figure 4. In our design of inhibitors, the central residue was made into a dehydroalanine, which we envisioned might act as the Michael acceptor warhead. The amides of each peptide bond were methylated to minimize the possibility of proteolytic hydrolysis in biological contexts [15]. The addition of an acyl group to the N-terminus and a benzyl ester to the C-terminus served to improve the pharmacokinetic properties of the inhibitors. The other two residues (R₁ and R₃) embodied various L-amino acid side chains; cysteine was excluded from this screen at both R₁ and R₃ positions to avoid the possibility of an intramolecular macrocyclization, thus allowing for the design of 361 analog compounds. We envisioned irreversible covalent binding of an inhibitor to M^{pro} might be achieved with a binding conformation wherein the central dehydroalanine warhead is poised within reacting distance of the thiol of Cys145, as shown in Figure 3.

![Figure 4](https://example.com/figure4.png)

**Figure 4.** Modifications made to a tripeptide for the design of SARS-CoV-2 M^{pro} inhibitors.

## Results

Each structure was evaluated for predicted binding affinity to the M^{pro} via molecular docking. A docking screen was conducted using Swissdock on 361 compounds with various L-amino acid side chains as R₁ and R₃. Results were quantified by the free energy of binding (ΔG in kcal/mol) of the highest scoring docking pose. A heat map of the lowest ΔG values returned for each compound (Figure 5) is color coded by increasing binding affinity. Squares that are in grey failed to dock because they either resulted in a topology error or took too much memory on the SwissDock server.
Figure 5. Heat map of designed inhibitors docked to the main protease of SARS-CoV-2 (PDB:6Y84). The binding affinity of the hit compound is shown in yellow. Each square is color coded by increasing binding affinity.

Results from the docking study suggest that cationic amino acid side chains as either R1 or R3 give the highest binding affinities, with ΔG values around -10.00 kcal/mol, with a few exceptions. It seems that strong electrostatic interactions between the cationic side chains and a proximal Glu166 residue in the binding pocket (Figure 6) is operative in creating the high binding affinity observed. This is clearly evidenced by the fact that side chains containing a free amine (with a lysine residue) and/or guanidinium (with an arginine residue) result in a strikingly higher binding affinity to the Mpro active site in comparison to compounds with anionic or uncharged side chains. Moreover, it is found that this SAR is not highly dependent on the positioning of the substituent positions bearing the cationic side chain - substitution at R1 or R3 with a lysine or arginine seem to be equally effective in giving a high binding affinity.

Figure 6. Designed inhibitors with arginine and lysine in both R1 and R3. The structure shown in blue has lysine as R1 (ΔG = -10.26 kcal/mol). The structure in purple has lysine as R3 (ΔG = -9.82 kcal/mol).
When presented with both lysine and arginine side chains in the same compound, interactions between the Glu166 residue of the target and primary amine of the ligand’s lysine side chain are preferred (Figure 6) regardless of the order of lysine and arginine side chains in the compound. We attribute this to a greater localization of cationic character at the primary amine center over the guanidinium.

Of the 361 compounds docked, one hit compound, with two arginine side chains, shown in Figure 7, was identified. The most thermodynamically stable binding pose of this compound had a binding affinity of the -11.51 kcal/mol to the active site of the main protease. Electrostatic attractions between the cationic guanidinium side chain at R3 and a proximal anionic Glu166 residue, along with hydrogen-bond interactions between the guanidinium side chain at R1 with Thr26 are largely responsible for the ligand’s high binding affinity. In the most thermodynamically stable binding pose of the hit compound, the distance between the Glu166 residue and the cationic side chain is 3.00 Å, an indication of the strength of the electrostatic attraction.

**Figure 7.** Hit compound, with two arginine side chains, docked to the main protease. The binding pose of the hit compound with the best binding affinity to the main protease (ΔG = -11.51 kcal/mol).

**Discussion**

In a campaign to identify potential lead compounds that might serve as covalent inhibitors of the SARS-CoV-2 main protease, we screened over three hundred peptidomimetic structures with varying side chains at two positions. From this screen, it was determined that structures with at least one cationic side chain gave the highest binding affinity to the active site of the main protease. Moreover, this screen identified a top hit structure, with two arginine-like side chains, as having the highest predicted binding affinity. The hit structure described can be synthesized in under ten steps from commercially available starting materials.

Based on this initial hit structure, studies on optimized second-generation ligands and on potential pro-drug strategies to deliver dicationic ligands are currently underway. One issue with some cationic compounds is low cell membrane permeability, as evidenced by the negative clogP value (-1.38 from ChemAxon MarvinSketch) of our hit compound. This might be ameliorated by pro-drug approaches that could improve the pharmacokinetic properties of the compound. Moreover, the present study has been limited to conventional L-amino acid side chains in the structures of the compounds screened. The current work has not fully explored the possibility of D-amino acid side chains, which are less susceptible to proteolytic hydrolysis, and reversal of one or both of the stereogenic centers of our hit compound are one possible avenue of study for second-generation ligands [16]. While the hit structure reported here and other compounds from our library have yet to be evaluated in vitro or in vivo, the SAR presented provides structural insight into the chemical and biophysical factors at play in the future design and development of small molecule protease inhibitors of SARS-CoV-2.
Experimental Procedures

Molecular Mechanics Pre-Optimization

All compounds were modeled through Avogadro, an open source molecular builder and visualization tool [17]. The local pH in the protease enzyme was assumed to be equal to the bulk pH of 7.4, and the appropriate protonation states were determined. An initial molecular mechanics geometry pre-optimization with an Merck Molecular Forcefield (MMFF94) at 10,000 steps was applied to all molecules, and input files for quantum mechanical structural optimization were generated through Avogadro.

Density Functional Theory (DFT)

Quantum mechanical structural optimizations, which are necessary for accurate prediction of thermodynamically minimized geometries of each compound, was completed using ORCA, an ab initio quantum molecular modeling software [18]. Density functional theory (DFT) geometry optimizations were completed with a B3LYP functional, def2-SVP basis set, and an implicit conductor-like polarizability model (CPCM) solvation model of water (dielectric= 80.4). Density functional theory calculations were performed on a Dell PowerEdge 710 server with a 24 core Intel Xeon X5660 processor @ 2.80GHz and 32GB RAM.

Molecular Docking

Docking studies on the protease inhibitors were conducted with SwissDock, a web server developed by the Swiss Institute of Bioinformatics [19]. Swissdock is based on the docking software EADock DSS, and the algorithm starts by generating binding modes within a grid box and estimating CHARMM energies. The unliganded crystal structure of the main protease of SARS-CoV-2 (PDB: 6Y84 [20]) was used as the receptor protein. The sulfur atom of residue cysteine 145 was set as the center of the grid box. Additionally, side chains on the protein with up to 5Å of flexibility from the ligand in the predicted binding mode were co-optimized during docking. Predicted binding modes were scored by the free energy of binding (ΔG) in kcal/mol. The resulting binding poses were visualized using UCSF Chimera [21].

Acknowledgments

Raw data files for all compounds screened, including output files from DFT structural optimization and from docking experiments, are available upon request. All authors contributed to performing molecular mechanics, DFT optimization, and molecular docking experiments of the structures reported, and are listed alphabetically and by seniority. EN conceived the project. SS, AB, and AC wrote the manuscript. The authors declare no competing conflicts of interests in the work presented. The authors gratefully acknowledge Prof. Robert Downing from the Department of Computer Science & Engineering at ASDRP for his guidance with initial setup and remote access to the server.
References


[18] Neese, F. (2011). The ORCA program system. *WIREs Computational Molecular Science, 2*(1), 73–78. [https://doi.org/10.1002/wcms.81](https://doi.org/10.1002/wcms.81)

