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 inhibitor of SARS-CoV-2

Stephanie Sun[1], Kavya Anand[2], Ishani Ashok[3], Bhavesh Ashok[4], Ayush Bajaj[5], Varsha Beldona[5], Kushal Chattopadhyay[3], Audrey Kwan[6], Karankumar Mageswaran[3], Anvi Surapaneni[4], Atri Surapaneni[5], Pranjal Verma[2], Allen Chen[3], Ria Kolala[3], Andrew Liang[7], Ayeeshi Poosarla[3], Krithikaa Premnath[2], Karthikha Sri Indran[3], Jesyln Wu[3], Aishwarya Yuvaraj[8], Harsha Raj[9], Tanish Sathish[10], Aashi Shah[4], Sarah Su[11], Kara Tran[6], *Edward Njoo*[12,*]

[1] BASIS Independent Silicon Valley, San Jose, CA; [2] Doughery Valley High School, San Ramon, CA; [3] Mission San Jose High School, Fremont, CA; [4] Amador Valley High School, Pleasanton, CA; [5] The Quarry Lane School, Dublin, CA; [6] Dublin High School, Dublin, CA; [7] The College Preparatory School, Oakland, CA; [8] American High School, Fremont, CA; [9] Fremont Christian High School, Fremont, CA; [10] Irvington High School, Fremont, CA; [11] Los Altos High School, Los Altos, CA; [12] Department of Chemistry, Biochemistry, and Physical Science; Aspiring Scholars Directed Research Program, Fremont, CA

KEYWORDS: COVID-19, SARS-COV-2, Protease Inhibitors, Computer-Guided Drug Design, Organic Chemistry, Medicinal Chemistry, Antiviral Drugs, Molecular Docking

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 8, 2020, over 7.1 million COVID-19 cases have been reported, with the death toll surpassing 400,000. The COVID-19 pandemic presents the need for urgent identification of inhibitors and vaccines (2).

SARS-CoV-2 belongs to a family of β-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 (Mpro) 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 Mpro is an ideal target for antiviral therapies (7).

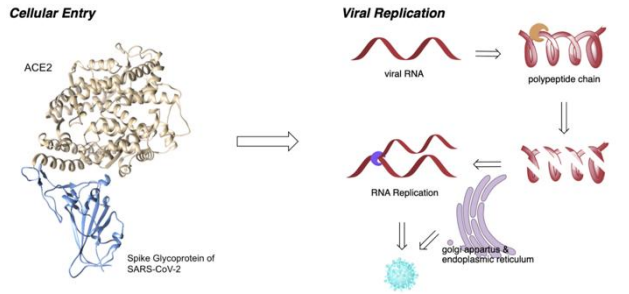

Figure 1. Mechanism of action of SARS-CoV-2. The spike glycoprotein of the virus mediates entry into human cells through ACE2 receptors on human lung cells, allowing the virus to use the cell's resources to replicate.

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 NH2- 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).

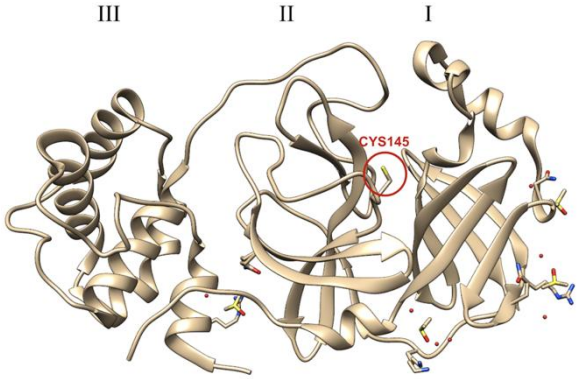

Figure 2. Crystal structure of the SARS-CoV-2 M^{pro} monomer (PDB: 6Y84), with each of the three domains labeled. Cys145 - the target in our design of inhibitors – is circled.

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).

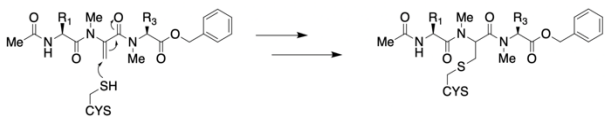

Figure 3. Reaction mechanism of the Michael Addition between the thiol of Cys145 and the dehydroalanine on our designed inhibitors, resulting in irreversible covalent binding of our inhibitors to the active site residue.

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 Fig-

ure 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 Mpro 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.

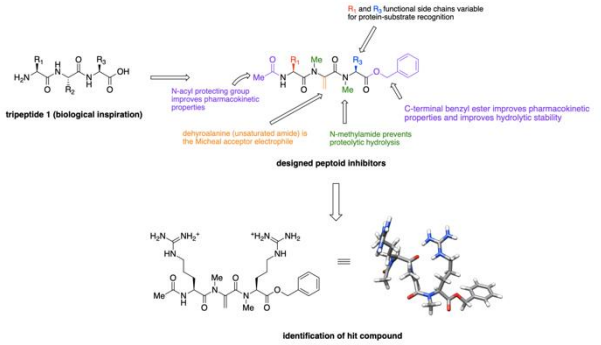

Figure 4. Modifications made to a tripeptide for the design of SARS-CoV-2 M^{pro} inhibitors.

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.

R3 R1	G	A	V	I	L	P	M	S	T	N	Q	D	E	K	R	F	Y	W	H
G	-8.19	-8.52	-8.31	-8.64	-8.76	-8.36	-9.05	-8.57	-8.64	-8.11	-9.04	-8.34	-8.5	-9.80	-9.51	-8.32	-8.87		
A	-7.86	-8.47	-8.57	-8.69	-8.24	-7.97	-8.83	-8.2	-8.35	-8.47	-9.14	-8.54	-8.26	-9.52	-10.27		-9.17		
V	-8.68	-8.16	-8.21	-9.14	-8.63	-8.76	-8.78	-8.64	-8.38	-8.47	-8.28	-8.07	-8.69	-9.49	-9.96	-8.75	-8.54		
I	-8.69	-8.72	-9.35	-8.82	-8.85	-8.15	-9.59	-8.72	-8.93	-8.86	-8.21	-8.26	-8.9	-9.27	-10.16		-8.76		
L	-8.34	-8.28	-8.23	-9.12	-8.90	-8.18	-9.22	-8.98	-9.08	-9.06	-8.79	-8.44	-8.86	-10.42	-10.13	-9.01	-8.98		
P	-8.52	-8.07	-8.49	-8.63	-8.61	-7.92	-9.16	-9.07	-7.98	-7.74	-8.67	-8.13	-7.90	-8.74	-9.43	-9.38	-9.62		
M	-8.71	-8.62	-8.68	-8.93	-8.87	-8.50	-8.99	-9.05	-8.55	-9.69	-8.76	-9.35	-8.34	-9.89	-10.37	-9.49	-8.88		
S	-8.26	-8.74	-8.42	-10.69	-8.60	-7.94	-8.81	-8.87	-8.81	-8.27	-7.18	-7.87	-8.08	-10.04	-9.92	-8.42	-9.35		
T	-8.45	-8.34	-8.76	-8.74	-8.13	-8.12	-8.69	-8.68	-8.85	-8.93	-9.05	-8.56	-8.48	-10.12	-9.91	-8.63	-8.85		
N	-8.82	-8.64	-9.02	-8.42	-8.99	-7.96	-8.26	-8.88	-8.63	-8.70	-8.96	-8.58	-8.58	-9.89	-10.06	-8.78	-9.35		
Q	-8.89	-8.45	-9.59	-9.11	-9.00	-9.90	-9.43	-8.35	-8.96	-9.10	-8.57	-8.68	-8.54	-9.85	-10.21	-8.77	-8.85		
D	-8.35	-7.67	-8.41	-8.25	-8.66	-7.73	-8.87	-8.33	-8.00	-7.69	-8.92	-7.52	-8.12	-9.17	-10.17	-8.57	-7.98		
E	-8.38	-8.51	-8.61	-8.79	-8.82	-7.71	-8.97	-7.75	-8.19	-8.94	-8.81	-8.38	-8.31	-10.34	-10.13	-8.74	-8.87		
K	-9.32	-10.43	-10.46	-9.65	-9.57	-9.50	-10.73	-9.36	-10.09	-9.94	-10.32	-10.19	-10.20	-9.52	-9.82	-9.70	-9.20	-10.45	
R	-9.94	-10.00	-10.60	-10.26	-10.70	-9.96	-10.60	-10.41	-10.12	-10.39	-10.08	-9.31	-10.16	-10.26	-11.51	-10.13	-10.57		
F	-8.26	-9.14	-8.45	-8.74	-9.08	-9.04	-9.30	-8.86	-8.84	-8.58	-8.07	-8.66	-8.53	-9.61		-8.78	-8.85		
Y	-9.61	-9.05	-8.63	-8.93	-9.03	-8.95	-8.86	-8.55	-8.44	-8.94	-7.70	-8.58	-8.87	-9.31	-10.49		-8.98		
W														-9.83	-9.19				
H																			

KEY	≤ -10.00	-9.99 to -9.50	-9.49 to -9.00	-8.99 to -8.50	-8.49 to -8.00	≥ -7.99	Error

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 M^{pro} 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.

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.

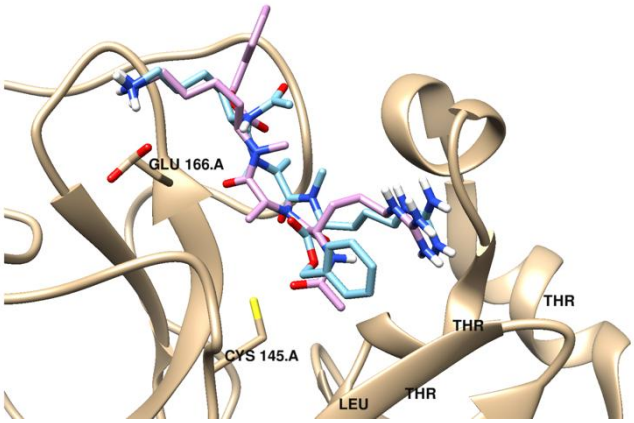

Figure 6. Designed inhibitors with arginine and lysine in both R_1 and R_3. The structure shown in the blue has lysine as R_1 (ΔG = -10.26 kcal/mol). The structure in purple has lysine as R_3 (ΔG = -9.82 kcal/mol)

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.

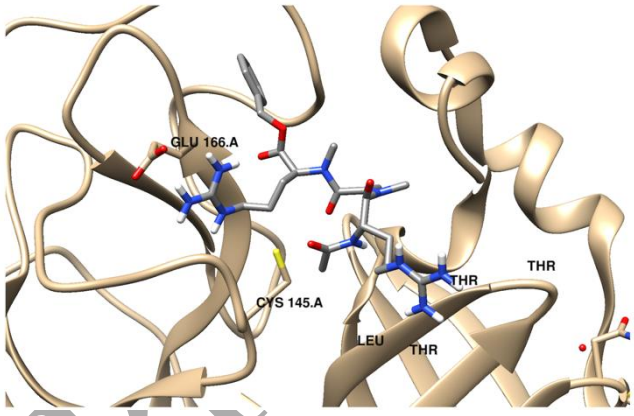

Figure 7. Hit compound, with two arginine side chains, docked to the main protease. The binding pose of the hit compound with the best binding affinity to the main protease (ΔG = -11.51 kcal/mol)

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.

MATERIALS AND METHODS

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).

ACKNOWLEDGEMENTS

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

(1) Wu, Fan, et al. "A New Coronavirus Associated with Human Respiratory Disease in China." Nature, vol. 579, no. 7798, 2020, pp. 265–269., doi:10.1038/s41586-020-2008-3.

(2) Rothan, Hussin A., and Siddappa N. Byrareddy. "The Epidemiology and Pathogenesis of Coronavirus Disease (COVID-19) Outbreak." Journal of Autoimmunity, vol. 109, 2020, p. 102433., doi:10.1016/j.jaut.2020.102433.

(3) "The Species Severe Acute Respiratory Syndrome-Related Coronavirus: Classifying 2019-NCoV and Naming It SARS-CoV-2." Nature Microbiology, vol. 5, no. 4, 2020, pp. 536–544., doi:10.1038/s41564-020-0695-z.

(4) Ren, Li-Li, et al. "Identification of a Novel Coronavirus Causing Severe Pneumonia in Human." Chinese Medical Journal, vol. 133, no. 9, 2020, pp. 1015–1024., doi:10.1097/cm9.0000000000000722.

(5) Fehr, Anthony R., and Stanley Perlman. "Coronaviruses: An Overview of Their Replication and Pathogenesis." Coronaviruses Methods in Molecular Biology, 2015, pp. 1–23., doi:10.1007/978-1-4939-2438-7_1.

(6) Ramajayam, R., et al. "Recent Development of 3C and 3CL Protease Inhibitors for Anti-Coronavirus and Anti-Picornavirus Drug Discovery." Biochemical Society Transactions, vol. 39, no. 5, 2011, pp. 1371–1375., doi:10.1042/bst0391371.

(7) Dai, Wenhao, et al. "Structure-Based Design of Antiviral Drug Candidates Targeting the SARS-CoV-2 Main Protease." Science, 2020, doi:10.1126/science.abb4489.

(8) Shi, Jiahai, and Jianxing Song. "The Catalysis of the SARS 3C-like Protease Is under Extensive Regulation by Its Extra Domain." FEBS Journal, vol. 273, no. 5, 2006, pp. 1035–1045., doi:10.1111/j.1742-4658.2006.05130.x.

(9) Goyal, Bhupesh, and Deepti Goyal. "Targeting the Dimerization of the Main Protease of Coronaviruses: A Potential Broad-Spectrum Therapeutic Strategy." ACS Combinatorial Science, 2020, doi:10.1021/acscombsci.0c00058.

(10) Ton, Anh-Tien, et al. "Rapid Identification of Potential Inhibitors of SARS-CoV-2 Main Protease by Deep Docking of 1.3 Billion Compounds." Molecular Informatics, 2020, doi:10.1002/minf.202000028.

(11) Zhang, Linlin, et al. "Crystal Structure of SARS-CoV-2 Main Protease Provides a Basis for Design of Improved α-Ketoamide Inhibitors." Science, 2020, doi:10.1126/science.abb3405.

(12) Tan, J., et al. "3C Protease of Enterovirus 68: Structure-Based Design of Michael Acceptor Inhibitors and Their Broad-Spectrum Antiviral Effects against Picornaviruses." Journal of Virology, vol. 87, no. 8, 2013, pp. 4339–4351., doi:10.1128/jvi.01123-12.

(13) Naidu, B. Narasimhulu, et al. "Michael Addition of Amines and Thiols to Dehydroalanine Amides: A Remarkable Rate Acceleration in Water." The Journal of Organic Chemistry, vol. 68, no. 26, 2003, pp. 10098–10102., doi:10.1021/jo034762z.

(14) Zhang, Yuan, et al. "Michael Addition of Dehydroalanine-Containing MAPK Peptides to Catalytic Lysine Inhibits the Activity of Phosphothreonine Lyase." FEBS Letters, vol. 589, no. 23, 2015, pp. 3648–3653., doi:10.1016/j.febslet.2015.10.025.

(15) Radzicka, Anna, and Richard Wolfenden. "Rates of Uncatalyzed Peptide Bond Hydrolysis in Neutral Solution and the Transition State Affinities of Proteases." Journal of the American Chemical Society, vol. 118, no. 26, 1996, pp. 6105–6109., doi:10.1021/ja954077c.

(16) Tugyi, R., et al. "Partial D-Amino Acid Substitution: Improved Enzymatic Stability and Preserved Ab Recognition of a MUC2 Epitope Peptide." Proceedings of the National Academy of Sciences, vol. 102, no. 2, 2005, pp. 413–418., doi:10.1073/pnas.0407677102.

(17) Hanwell, Marcus D, et al. "Avogadro: an Advanced Semantic Chemical Editor, Visualization, and Analysis Platform." Journal of Cheminformatics, vol. 4, no. 1, 2012, doi:10.1186/1758-2946-4-17.

(18) Neese, Frank. "The ORCA Program System." WIREs Computational Molecular Science, vol. 2, no. 1, 2011, pp. 73–78., doi:10.1002/wcms.81.

(19) Grosdidier, A., et al. "SwissDock, a Protein-Small Molecule Docking Web Service Based on EADock DSS." Nucleic Acids Research, vol. 39, no. suppl, 2011, doi:10.1093/nar/gkr366.

(20) Owen, C.d., et al. "SARS-CoV-2 Main Protease with Unliganded Active Site (2019-NCoV, Coronavirus Disease 2019, COVID-19)." 2020, doi:10.2210/pdb6y84/pdb.

(21) Pettersen, Eric F., et al. "UCSF Chimera-A Visualization System for Exploratory Research and Analysis." Journal of Computational Chemistry, vol. 25, no. 13, 2004, pp. 1605–1612., doi:10.1002/jcc.20084.