Computational Docking of Potential Drug Inhibitors of SARS-CoV-2 Main Protease

John V Molthen

B.S. Candidate, Department of Chemistry, California State University Stanislaus, 1 University Circle, Turlock, CA 95382

Received 14 May 2021; accepted 20 July 2021

Abstract

Computational biology and chemistry have become more relevant with the new technology of the 21st century. Many programs are used as replicas for experimental results, and visualizations of the chemical and microbiological world. This project is focused on docking drugs to the Main Protease of SARS-CoV-2 using AutoDock Vina, a computational docking program. With the quantum mechanical equations done by the program, it should give results within 5% of the actual experimental results. If it does do this, it can be considered a good model for the actual experiment.

Keywords: AutoDock, Main Protease, Protease Inhibitors

Introduction

This project is focused on the main protease of SARS-CoV-2 (severe acute respiratory syndrome coronavirus 2), more commonly known as COVID-19. Before diving into what this protease is and what it does, a little background of SARS-CoV-2 should be given. SARS-CoV-2 is a coronavirus. Coronaviruses are members of the subfamily Coronavirinae in the family Coronaviridae[1]. This subfamily can be divided into four different genera based on their genomic structures and physiological activities[1]. The four genera are: Alphacoronavirus, Betacoronavirus, Gammacoronavirus, and Deltacoronavirus[1,2]. Out of these four, only Alphacoronavius and Betacoronavirus have been found to cause human diseases (Marty). Both of these genera usually will cause respiratory illness in humans[1]. Alphacoronavirus are usually not deadly to humans and cause symptoms of the common cold, while Betacoronavirus is known to be the deadliest to humans[2]. Within the Betacoronavirus genus falls Severe Acute Respiratory Syndrome (SARS-CoV), Middle East Respiratory Syndrome (MERS-CoV), and the latest addition to the list of Severe Acute Respiratory Syndrome 2 (SARS-CoV-2). This new coronavirus, however, has turned into a world-wide pandemic. The first reports of SARS-CoV-2 came from Wuhan, China[3,4]. This coronavirus is theorized to have come from a Rhinplophus affinis bat, as RaTG13 (the coronavirus in bats) is about 96% identical to it, but this has not been proven[4,5]. What is known about this coronavirus is that it can be deadly[6]. This implies that as a community, scientists have to work together to stop this virus and end the pandemic.

One approach suggested to stop this virus is inhibiting its main protease. This protease is a 3-chymotrypsin-like protease[6]. This main protease is about 306 amino acids long and its responsible for allowing the polypeptide to be processed into functional proteins[7,8]. The activities of this protease are activated by a substate that bind to the proteins active site. However, there are some substrates that could block this action from occurring called inhibitors. If an inhibitor could disable the main protease from carrying out its conventional function, then that could be a huge step towards ending this pandemic.

Possible protease inhibitors have been suggested to be HIV/AIDS drug inhibitors[9]. This experiment will be putting that suggestion to the test, through computational work. All of the drugs listed in *Table 1* are FDA approved protease inhibitors and were chosen to dock to the main protease[10]. Computational is being used more and more often in these times. Not only does it drastically decrease the cost and time put into drug development, but it is also an efficient way to be safe way to do research away from the lab during this pandemic[11].

Table 1: FDA Approved Protease Inhibitors[10]

Drug
Darunavir
Fosamprenavir
Indinavir
Nelfinavir

Tipranavir

AutoDock Vina is a molecular docking program that was used to determine the affinities of the drugs to the protease. This program is based on a series of different quantum mechanical calculations[12,13]. It will use the input protein-ligand complex and will calculate the lowest probable configuration in terms of ΔG values[12,13]. The ΔG values are calculated using the equation in *Figure 1*. This equation was created by taking into consideration of five key parameters. It takes into consideration van der Waals interactions, hydrogen bonding, electrostatics, desolvation, and the tortional component of these protein-ligand interactions[14]. The protease-drug complex with the lowest ΔG value will be determined to be the optimal complex out of the complexes tested. The results from all of the complexes obtained should be compared to experimental data to determine if they are within 5% error. The results from this work could possibly help strengthen the development of halting SARS-CoV-2 activity.

$$\begin{aligned}\Delta G = {} & \Delta G_{\text{vdW}} \sum_{i,j} \left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}} \right) \\ & + \Delta G_{\text{hbond}} \sum_{i,j} E(t) \left(\frac{C_{ij}}{r_{ij}^{12}} - \frac{D_{ij}}{r_{ij}^{10}} + E_{\text{hbond}} \right) \\ & + \Delta G_{\text{elec}} \sum_{i,j} \frac{q_i q_j}{\varepsilon(r_{ij}) r_{ij}} \\ & + \Delta G_{\text{tor}} N_{\text{tor}} \\ & + \Delta G_{\text{sol}} \sum_{i_C,j} S_i V_j e^{(-r_{ij}^2/2\sigma^2)}\end{aligned}$$

Figure 1: AutoDock Vina ΔG Calculation

Methods

The procedure followed in this lab was based off of Prasanth et. al.'s article[15]. This procedure included the use of multiple different computer programs. These programs include PyMol, AutoDock Tools, and AutoDock Vina[16,17,18]. PyMol was used to visualize the protein-drug complex's, AutoDock Tools was used to prepare the complexes to be tested by AutoDock Vina, and AutoDock Vina was then used to test the affinities of the protein drug complexes.

Before using any programs though, multiple things had to be determined. First, the correct structure of the main protease, its active site coordinates, and its dimensions for the gridbox for the AutoDock programs had to be determined. These were found using Prasanth et. al.'s article[15]. The PDB ID for the main protease was 6LU7. The active site and the grid box dimensions for this structure was determined to be $x = -17.59,\ y = 15.81,\ z = 63.53$ and 30 x 30 x 30 Å^3 respectfully. These values were inserted into AutoDock Tools and the polar hydrogen atoms and Kollman Charges were added to the protein. All of the water molecules were also deleted from the structure using AutoDock Tools. The next thing that had to be done was to find possible inhibitors.

The possible inhibitors found were when looking into the FDA's website for protease inhibitors. The inhibitors found and used in this experiment were Darunavir, Fosamprenavir, Indinavir, Nelfinavir, and Tipranavir. To use these drugs, their 3D structures were downloaded off of PubChem[19]. Their PubChem ID's were: 213039 for Darunavir; 131536 for Fosamprenavir; 5362440 for Indinavir; 64143 for Nelfinavir; and 54682461 for Tipranavir. Once everything was downloaded AutoDock Vina was used to dock the ligands to the protein and each affinity was tested. PyMol was then used to visualize all of these interactions.

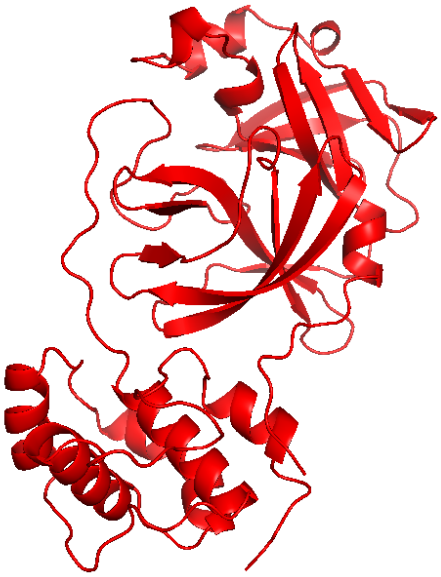


Figure 2: A visualization from PyMol of the Main Protease

Table 2: Visualizations from PyMol of all of the drugs before they are docked to the Main Protease.

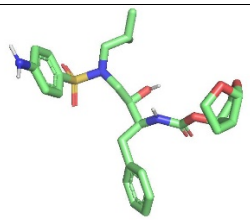
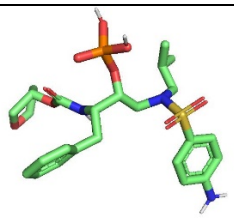
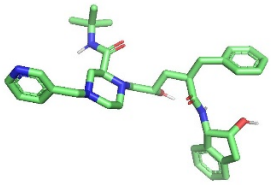
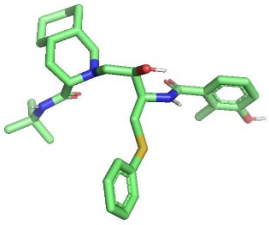
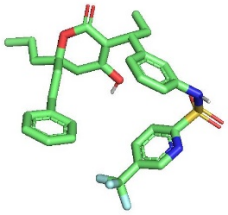
	Darunavir
	Fosamprenavir
	Indinavir
	Nelfinavir
	Tipranavir

Table 2 Legend

Atom	Color
Carbon	Green
Hydrogen	White
Oxygen	Red
Nitrogen	Blue
Sulfur	Yellow
Fluoride	Teal

Results

In the experiment the main protease of SARS-CoV-2, shown in *Figure 2*, was the protein that I used to dock the five protease inhibitors to. These drugs are shown in *Table 2*. Using the calculations shown in *Figure 1*, AutoDock Vina was able to calculate the ΔG values for all of the protein-drug complexes. The energies of these complexes were calculated and further tabulated into *Tables 3* and *4*. The lowest energy complex was determined to be the nelfinavir-main protease complex. Its binding energy was calculated to be -8.2 kcal/mol.

In *Tables 3* and *4*, the standard deviations of all of the trials for the complexes were calculated. All of the standard deviations calculated were 0.25 or less. These are very low standard deviations, which means that the precision of this program is very high. However, relative to each other the standard deviations are very different. This can be explained using a couple of different concepts. First, looking at the calculations that AutoDock Vina performs, in Figure 1, to obtain ΔG values, we can see that there is a torsional component. This tortional component is important because in molecules the bonds between atoms have different vibrational and rotational modes. With these different rotations and vibrations, the drug can orient itself in different positions, causing the functional groups of the drugs to interact with different residues of the protein. It is also possible that there are variations of the standard deviations, because the complexes have multiple minimum energy positions. If AutoDock Vina is calculating the minimum energy and thinks that it has found the global minimum, because the orientations of the molecule that it tries after it finds the minimum, then it is possible for it to be wrong. AutoDock Vina may mistake a local minimum as a global minimum. This would ultimately change the energy values that the program calculates.

Table 3: Energies of Main Protease-Drug Complexes

Drug	Darunavir	Fosamprenavir	Indinavir
Trial 1 /kcal*mol^{-1}	-7.6	-7.6	-8
Trial 2 /kcal*mol^{-1}	-7.7	-7.7	-8.1
Trial 3 /kcal*mol^{-1}	-7.7	-7.4	-8
Trial 4 /kcal*mol^{-1}	-7.8	-7.4	-8
Average Energy /kcal*mol^{-1}	-7.7	-7.5	-8
Standard Deviation	0.082	0.15	0.05

Table 4: Energies of Main Protease-Drug Complexes

Drug	Nelfinavir	Tipranavir
Trial 1 (kcal/mol)	-8.2	-7.4
Trial 2 (kcal/mol)	-7.9	-7.6
Trial 3 (kcal/mol)	-8.3	-7
Trial 4 (kcal/mol)	-8.5	-7.4
Average Energy (kcal/mol)	-8.2	-7.4
Standard Deviation	0.25	0.25

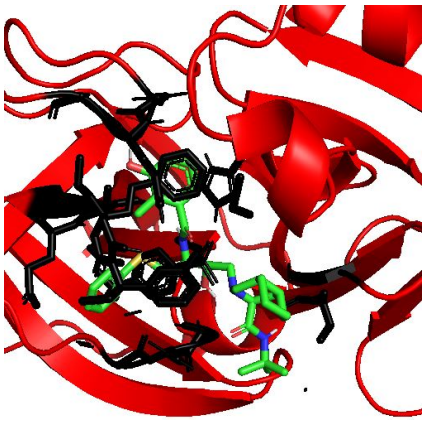


Figure 3: Nelfinavir-Main Protease Complex with Active Site Residues Shown in Black, Generated by PyMol

Figure 4 shows the nelfinavir-main protease complex. Since the nelfinavir-main protease complex has the lowest binding energy, nelfinavir can be considered the best inhibitor for the main protease. However, this doesn't mean that it will be the best inhibitor if they were all tested in a lab, but it is predicted to be the best.

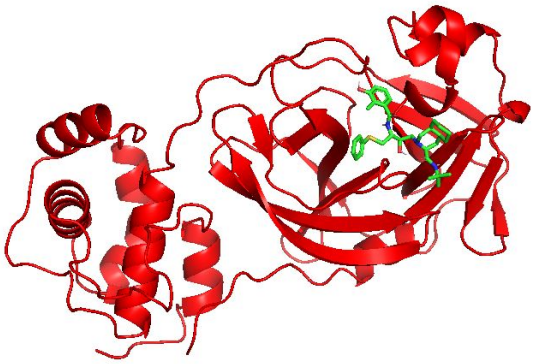


Figure 4: Nelfinavir-Main Protease Complex Generated by PyMol

Once all of the energies were calculated they could be compared to the literature values of the complexes. However, only one value was recovered from the literature was for the indinavir-main protease complex. The literature value found was -7.6 kcal/mol[20]. When this value was compared to the value calculated by AutoDock Vina, it was determined to have a 5.30 percent error. This was slightly greater than the predicted percent error. Without any more data from the literature, these comparisons cannot be made.

Discussion

Before testing the procedure stated in this article another procedure was tested. This procedure was based off of Monthay's and Ramesh's article[7]. In this article a protein model of the main protease was made using the Swiss Model. This model was created by inputting the DNA sequence for the main protease (region 1541-1858), downloaded from GenBank (accession number: P0C6X7.1), into the Swiss Model program[7]. This article was published prior to there being an accurate structure on the PDB website, so the first thing that had to be done was to check the accuracy of the Swiss Model to the PDB structure. Both of these models were uploaded to PyMol and aligned to each other. In *Table 4* it can be seen that the RMSD value that PyMol calculated to be 14.650 Å. This number was very high for an RMSD value. Identical proteins should give RMSD values ranging from 0-1.2 Å[21]. Since the RMSD value obtained was so high, this model protein was determined to me inaccurate. This Swiss Model protein structure was not used to dock the possible inhibitors to.

Table 4: RMSD of Swiss Model and Main Protease (6LU7)

RMSD /Å	14.650

Another problem that arrived when doing this project was that there was some trouble finding binding energy values to compare to the computational results. However, there was some data found in the literature for kinetics data for these complexes. I believe that in the future, I can calculate the binding energy values from the kinetics data and compare the rest of the AutoDock results to the literature. I would be able to furthermore prove whether my thesis is correct or not.

I plan to continue working on this research topic; however, I would like to take it in a new direction. I would like to study different variants of the main protease with these same protease inhibitors. It would be interesting to see how the binding energy changes with these variants. I would also like to work on drug design with the wild type enzyme as well as with the variants that I test. Working on drug design for these proteases could possibly help end the current pandemic.

Conclusion

This project was successful and unsuccessful in a few different ways. I was unsuccessful determining whether AutoDock Vina was an accurate enough program to use for most of these complexes. However, I did find that the complex I was able to compare to the literature, the indinavir-main protease complex, had a percent error of 5.30%, which was just over the predicted value of 5% error. I was also successful in determining the best predicted inhibitor for the main protease. The nelfinavir-main protease complex had a ΔG value of -8.2 kcal/mol. This was the lowest energy value I calculated, making nelfinavir the best inhibitor of the ones tested.

Acknowledgements

This project would not have been possible without the funding and guidance of the McNair Scholars Program. I would also like to recognize Dr. Koni Stone from California State University, Stanislaus for guidance on this project as well.

References

[1] Cui, J.; Li, F.; Shi, Z.-L. Origin and Evolution of Pathogenic Coronaviruses. *Nat. Rev. Microbiol.* **2019**, *17* (3), 181–192.

[2] Marty, A. M.; Jones, M. K. The Novel Coronavirus (SARS-CoV-2) Is a One Health Issue. *One Health* **2020**, *9* (100123), 100123.

[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* (7798), 265–269.

[4] 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* (7798), 270–273.

[5] Andersen, K. G.; Rambaut, A.; Lipkin, W. I.; Holmes, E. C.; Garry, R. F. The Proximal Origin of SARS-CoV-2. *Nat. Med.* **2020**, *26* (4), 450–452.

[6] Mori, M.; Capasso, C.; Carta, F.; Donald, W. A.; Supuran, C. T. A Deadly Spillover: SARS-CoV-2 Outbreak. *Expert Opin. Ther. Pat.* **2020**, *30* (7), 481–485.

[7] Mothay, D.; Ramesh, K. V. Binding Site Analysis of Potential Protease Inhibitors of COVID-19 Using AutoDock. *Virusdisease* **2020**, *31* (2), 1–6.

[8] Wang, F.; Chen, C.; Tan, W.; Yang, K.; Yang, H. Structure of Main Protease from Human Coronavirus NL63: Insights for Wide Spectrum Anti-Coronavirus Drug Design. *Sci. Rep.* **2016**, *6* (1), 22677.

[9] Ngo, S. T.; Quynh Anh Pham, N.; Thi Le, L.; Pham, D.-H.; Vu, V. V. Computational Determination of Potential Inhibitors of SARS-CoV-2 Main Protease. *J. Chem. Inf. Model.* **2020**, No. acs.jcim.0c00491. https://doi.org/10.1021/acs.jcim.0c00491.

[10] HIV and AIDS: Medicines to Help You. *Food and Drug Administration*. https://www.fda.gov/consumers/free-

publications-women/hiv-and-aids-medicines-help-you (accessed Sep 6, 2020).

[11] Marshall, G. R. Computer-Aided Drug Design. *Annu. Rev. Pharmacol. Toxicol.* **1987**, *27* (1), 193–213.

[12] Ryde, U.; Söderhjelm, P. Ligand-Binding Affinity Estimates Supported by Quantum-Mechanical Methods. *Chem. Rev.* **2016**, *116* (9), 5520–5566.

[13] Kuhn, B.; Kollman, P. A. Binding of a Diverse Set of Ligands to Avidin and Streptavidin: An Accurate Quantitative Prediction of Their Relative Affinities by a Combination of Molecular Mechanics and Continuum Solvent Models. *J. Med. Chem.* **2000**, *43* (20), 3786–3791.

[14] Morris, G. M. Equations — AutoDock http://autodock.scripps.edu/resources/science/equations (accessed Apr 30, 2021).

[15] Prasanth, D. S. N. B. K.; Murahari, M.; Chandramohan, V.; Panda, S. P.; Atmakuri, L. R.; Guntupalli, C. In Silico Identification of Potential Inhibitors from Cinnamon against Main Protease and Spike Glycoprotein of SARS CoV-2. *J. Biomol. Struct. Dyn.* **2020**, 1–15.

[16] The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC.

[17] Goodsell, D. S.; Olson, A. J. Automated Docking of Substrates to Proteins by Simulated Annealing. *Proteins* **1990**, *8* (3), 195–202.

[18] O. Trott, A. J. Olson, AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization and multithreading, *Journal of Computational Chemistry 31 (2010) 455-461* PubChem. PubChem https://pubchem.ncbi.nlm.nih.gov/ (accessed Oct 18, 2020)

[19] Li, Z.; Li, X.; Huang, Y.-Y.; Wu, Y.; Liu, R.; Zhou, L.; Lin, Y.; Wu, D.; Zhang, L.; Liu, H.; Xu, X.; Yu, K.; Zhang, Y.; Cui, J.; Zhan, C.-G.; Wang, X.; Luo, H.-B. Identify Potent SARS-CoV-2 Main Protease Inhibitors via Accelerated Free Energy Perturbation-Based Virtual Screening of Existing Drugs. *bioRxiv*, 2020. https://doi.org/10.1101/2020.03.23.004580.

[20] Kufareva, I.; Abagyan, R. Methods of Protein Structure Comparison. *Methods Mol. Biol.* **2012**, *857*, 231–257.