Title: Studies on computational molecular interaction between SARS-CoV-2 main protease and natural products

Manish Manish[1]*

1. School of Computational and Integrative Sciences, Jawaharlal Nehru University, New Delhi, India-110067.

Email:

Manish Manish: manishjnu8@gmail.com, manish81_sbt@jnu.ac.in

*Corresponding author

Abstract

Our objective was to identify the molecule which can inhibit SARS-CoV-2 main protease and can be easily procured or available as a common household food ingredient. For this objective, natural products may provide such molecules and can supplement the current custom chemical synthesis based drug discovery. A combination of docking approaches, scoring functions, molecular dynamic simulation, and literature mining have been employed to screen readily available natural products (unique 27256 chemical entities, 598435 compounds), which can inhibit the SARS-CoV-2 main protease. Theaflavin digallate, present in black tea, has been observed to be as three top-scoring compounds after the virtual screening of 598435 compounds. The main protease-theaflavin digallate complex observed to be in the metastable stage and interact with critical active site residues of the main protease during molecular dynamics simulation for 200 ns. *Invitro* evidence on main protease inhibition of 2003 SARS-CoV by theaflavin digallate is available in the scientific literature. As evident by the dynamics of intermolecular interactions, theaflavin digallate, forms approximately three hydrogen bonds with Glu166 of main protease, mostly through hydroxyl groups in the benzene ring of benzo(7) annulen-6-one. Glu166 is the most critical amino acid for main protease dimerization, which in turn, is necessary for catalytic activity. We have employed epigallocatechin gallate (green tea component) as a control set because it has a similar structure, source, and biological activity as compared to TG. We have also included chloroquine as a control set as it has a different source, activity, and structure but known to inhibit SARS-CoV-2. Based on computational molecular interaction and data available in scientific literature, theaflavin digallate can inhibit the main protease of SARS-CoV-2.

Introduction

The current pandemic of coronavirus disease 2019 (COVID-19) is caused by *severe acute respiratory syndrome coronavirus 2* (SARS-CoV-2), which was previously tentatively named as 2019-nCoV. The 2012 epidemic of Middle East respiratory syndrome (MERS) was caused by *Middle East respiratory syndrome-related coronavirus* (MERS-CoV) and the 2003 epidemic of severe acute respiratory syndrome (SARS) was caused by *severe acute respiratory syndrome coronavirus* (SARS-CoV). SARS-CoV-2, SARS-CoV, and MERS-CoV belong to the same genus *Betacoronavirus (Coronaviridae Study Group of the International Committee on Taxonomy of Viruses 2020)*. The outbreak COVID-19 caused an unprecedented event to the world in the recent history of humanity. Most of the affected countries recommended to remain in home isolation, and cities have been shifted to lockdown mode as a part of a solution to this crisis.

SARS-CoV-2 is approximately 30 kilobases single-stranded positive sense enveloped RNA virus. Depending on the ribosomal frameshift, either replicase polyprotein pp1a (nsp1-11) or replicase polyprotein pp1b (nsp1-16) is translated from the genomic open reading frame 1a and 1b, respectively (Krichel et al. 2020). Main protease (also known as M^{pro}, $3CL^{pro}$) is necessary for the processing of these polyproteins from nsp 4-11 or nsp 4-16. Main proteases are chymotrypsin-like cysteine protease. Main protease has an essential role in proteolytic processing. Hence it has been extensively explored as a drug target against SARS-CoV or SARS-CoV-2 (Yang, Bartlam, and Rao 2006).

Natural products offer a wide range of opportunities in this scenario as it provides a diverse chemical scaffold which cannot be represented using conventional chemical libraries. Natural products can have known side effects/pharmacodynamics/pharmacokinetic properties as natural products may have been used traditionally for several indications. Natural products are chiral rich molecules and hence offer many stereoisomers whose biological exploration might be limited using conventional wet experimental techniques. Furthermore, natural products provide the opportunity to harness the chemical product, which can have resulted due to the long evolutionary interaction.

Hence, we have studied the computational molecular interaction of SARS-CoV-2 main protease with easily available natural products. We have employed an integrative strategy combining docking methods, scoring functions, and molecular dynamics. While analyzing interactions, we have observed that theaflavin digallate (TG) may inhibit SARS-CoV-2 main protease.

Methodology

Structural coordinates of easily available natural products in SDF file formats were downloaded from ZINC (Sterling and Irwin 2015). Chloroquine and epigallocatechin gallate (green tea component) were considered as control drugs. Ligand preparation was performed using the LigPrep module (Schrödinger Release 2018-4: LigPrep, Schrödinger, LLC, New York, NY, n.d.). The force field OPLS 2003e was used for the generation of three-dimensional conformation of natural products. Natural products were desalted and neutralized. Ionization states were prepared at neutral pH using Epik (Greenwood et al. 2010). PDB ID 6LU7 was used as SARS-CoV-2 main protease. Hydrogen atoms were added and side chains were optimized using prepwizard utility of Schrodinger suite. Ionizable groups at pH 7 were predicted using PROPKA, and hydrogen bonds were optimized using ProtAssign utility (Olsson et al. 2011). The structure was minimized with only restrains on heavy atoms using impref utility. The docking grid files were centered around Cys145 and His41.

The protein was docked using a virtual screen workflow combining glide module with high throughput virtual screening (Glide HTVS), single-precision (Glide SP), and extra precision (Glide XP). Glide XP has more extensive conformational sampling such as anchor-and-grow methodology and more stringent scoring function such as high penalty score for receptor-ligand noncomplementary than Glide SP. The compounds were ranked on the basis of docking score. The binding energy was estimated using Prime MMGBSA for the ligand-receptor complex derived after Glide XP docking. The VSGB2 was used as an implicit solvation model (Li et al. 2011). The complex was evaluated according to MMGBSA dG bind score where the dG bind = E_complex(minimized) - E_ligand(minimized) -E_receptor(minimized). The molecular dynamic simulation of the top-scoring complex after the Glide XP stage was performed using the desmond molecular dynamics package and OPLS 2005 force field. The solvation model used for molecular dynamics simulation was TIP3P. 10 Angstrom orthorhombic box has been used as boundary conditions. The charge was neutralized by adding Na+ ions, and 0.15 M NaCl was added as a salt. Simulations were run for 200 ns in NPT conditions and at 300 K at 1.01 bars.

Results and Discussion

We have used unique natural compounds that are reported to be easily available on the ZINC database. The possible isomeric/ionization/minimized conformation from this set resulted in a total of 598435 distinct ligands. We have used Virtual Screen Workflow default settings, i.e., a combination of HTVS, SP, XP, and prime-MMGBSA. However, we have additionally employed enhanced sampling of ligand conformation and calculation of XP descriptors. Data on the top 10 conformations is presented in table 1. Ligands were sorted according to respective docking scores. 2-D structures of the top 4 compounds observed as top 10 hits are represented in Figure 1. The top 3 hits are from ZINC ID ZINC000195838435, which corresponds to theaflavin digallate (TG) and having 100 percent similarity with ChEMBL IDs, CHEMBL1451483, CHEMBL402609, and CHEMBL434864. Theaflavin digallate has chromane and benzo(7)annulen-6-one as a chemical scaffold.

The exact chemical composition of tea is complex. However, it has mainly two classes of flavonoids/polyphenols viz. catechins and theaflavins. Catechins are the major component of green tea, whereas theaflavins are the major constituent of black tea. The major difference between green tea and black tea is the step of fermentation/oxidation. Green tea does not require fermentation, whereas black tea requires the complete step of fermentation. Theaflavins, as identified in this study, are the result of this oxidation process, and hence black tea contains more theaflavins as compared to green tea.

Using the similarity ensemble approach, TG appears to be in a protease inhibitor target class (Keiser et al. 2007). In the study of screening a natural products library against the 2003 SARS-CoV virus main protease, TG has a half-maximal inhibitory concentration of 7 µM against the main protease (Chen et al. 2005). Epigallocatechin gallate has a similar structure, the same natural source (tea), and known to have similar antioxidant properties as TG. However, epigallocatechin gallate did not inhibit the main protease enzymatic activity, which reflects the restricted stereo-selectivity of main protease. The other polyphenolic compounds of tea, such as epicatechin, epicatechin gallate, epigallocatechin, theophylline, and caffeine have also shown no inhibitory effect. Similarly, the crude extract of black tea has shown more enzymatic inhibition than the crude extract of green tea (Chen et al. 2005).

As molecular docking can have inherent inaccuracies, we have included two controls in our studies. The first control was epigallocatechin gallate (ZINC ID ZINC000003978478), the most abundant catechin in green tea, which has a similar structure, source, and activity as of TG. The second control was chloroquine, which is known to inhibit SARS-CoV-2. However, chloroquine has a different source,

activity, and structure (Vincent et al. 2005; Wang et al. 2020). Based on scoring functions, such as docking score, energy score, XP Gscore, glide gscore, TG observed to be a dominant inhibitor with docking score -15.263 as compared to epigallocatechin gallate for which docking scored varied from -9.401 to -7.541. The docking score of chloroquine was -4.477 to -4.300 and hence it can be inferred that chloroquine may not inhibit SARS-CoV-2 by targeting main protease. The molecular interactions of TG with main protease have been presented in figure 2A-C and for control set as figure 2K-2L which reflects TG has dominant interaction with main protease as compared to epigallocatechin gallate/chloroquine.

Whether the addition of milk influences the phenolic component of tea is not conclusively delineated (Rashidinejad et al. 2017). In an interventional clinical study on nine healthy male volunteers, infusion time has shown to increase the concentration of total phenolic concentration, and phenolic concentration was unaffected by the addition of milk (Kyle et al. 2007). In Swiss mice, antioxidative effects have been shown to be diminished by the addition of milk, sugar, and honey and the two-hour dose is recommended for health benefits (Korir et al. 2014)

The molecular interaction of TG with main protease has been evaluated by molecular dynamic simulations. The total number of amino acids in main protease (PDB ID 6LU7) was 308 and having a charge of -4. Four sodium ions were included to neutralize the system. The molecular dynamics has been performed in NPT at 300K. The total molecular system has 36432 atoms. The protein and ligand were simulated for 200 ns. During the simulation, protein fluctuated around 1.8 Å (Figure 3A left axis), whereas the ligand was fluctuated at around 4 Å (Figure 3A right axis) when superimposed to the docked protein at 0 ns. The protein complex was observed to be achieved a metastable stage throughout the simulation of 200 ns (Figure 3a). A movie of 1000 molecular frames during 200 ns simulation was also available as supplementary data and at YouTube https://www.youtube.com/watch?v=yYEwgQymsO4

TG approximately makes three hydrogen bonds, either directly or water-mediated with glutamate 166 throughout the simulation of 200 ns. Glu166 of each protomer interacts with "N-finger" of another protomer for dimerization, which is required for the catalytic activity of main protease (Zhang et al. 2020). There are at least 14 residues (Thr 26, His 41, Cys 44, Ser 46, Glu 47, Met 49, Asn 142, Gly143, Cys 145, His 164, Glu 166, Asp 187, Gln 189, Thr 190) in the main protease which interacts with TG for more than 100 ns either through hydrogen bonds, hydrophobic, water bridges or ionic interaction (Figure 3B). The catalytic residue Cys 145 predominantly interacts with water bridges, whereas His 41 interacts mainly through hydrophobic interaction (Figure 3B and 3C). The TG makes approximately 15 contacts during the whole simulation (Figure 3D). For the estimation of binding energy, we have employed the

prime MMGBSA using VSGB2 as an implicit solvent model on docked complexes and 1000 molecular dynamics trajectories snapshots. For the three top docked complexes that correspond with TG, the MMGBSA dG bind score varied from -69.06819456 kcal/mol to -70.85623647 kcal/mol. While during the 200 ns molecular dynamics simulation involving 1000 snapshots, the MMGBSA dG bind score varied from -51.79195471 kcal/mol to -122.7461939 kcal/mol. The MMGBSA dG bind scores for consecutive snapshots of molecular dynamics trajectories in 200 ns simulations were presented as Figure 3E. TG-main protease complexes have shown the average estimated binding energy of approximately -90 kcal/mol throughout the molecular dynamics simulation from 40 ns to 200 ns which may indicate the possible strong interaction between TG and main protease of SARS-CoV-2 (Figure 3E).

Tea is the second most-consumed beverage after the water. However, the implementation of tea as a public health measure is a long distant dream, and it requires extensive invitro, invivo, and clinical evidence. This study is based on mere computational calculations although robust and indicates that TG may inhibit SARS-CoV-2 main protease, maybe by inhibiting protease dimerization. Only invitro data is available in scientific literature to show that TG can inhibit SARS-CoV main protease. Natural products although easily available however proven to be difficult to be evaluated in robust clinical settings or as original natural sources such as with curcumin (Nelson et al. 2017). Similarly, an exact wet experiment or clinical studies providing the evidence on TG efficacy are difficult to conduct because of the huge number of variants available for black tea formulation. However, this study provide a scientific rationale to initiate an observational clinical study and in-vitro experiments with SARS-CoV-2 main protease for conclusive evidence.

Acknowledgment

References

Chen, Chia-Nan, Coney P C Lin, Kuo-Kuei Huang, Wei-Cheng Chen, Hsin-Pang Hsieh, Po-Huang Liang, and John T-A Hsu. 2005. "Inhibition of SARS-CoV 3C-like Protease Activity by Theaflavin-3,3'-Digallate (TF3)." *Evidence-Based Complementary and Alternative Medicine : ECAM* 2 (2): 209–15.

https://doi.org/10.1093/ecam/neh081.

Coronaviridae Study Group of the International Committee on Taxonomy of Viruses. 2020. "The Species
Severe Acute Respiratory Syndrome-Related Coronavirus: Classifying 2019-NCoV and Naming It
SARS-CoV-2." *Nature Microbiology* 5 (4): 536–44. https://doi.org/10.1038/s41564-020-0695-z.

Greenwood, Jeremy R., David Calkins, Arron P. Sullivan, and John C. Shelley. 2010. "Towards the
Comprehensive, Rapid, and Accurate Prediction of the Favorable Tautomeric States of Drug-like
Molecules in Aqueous Solution." *Journal of Computer-Aided Molecular Design* 24 (6–7): 591–604.
https://doi.org/10.1007/s10822-010-9349-1.

Keiser, Michael J., Bryan L. Roth, Blaine N. Armbruster, Paul Ernsberger, John J. Irwin, and Brian K.
Shoichet. 2007. "Relating Protein Pharmacology by Ligand Chemistry." *Nature Biotechnology* 25
(2): 197–206. https://doi.org/10.1038/nbt1284.

Korir, M W, F N Wachira, J K Wanyoko, R M Ngure, and R Khalid. 2014. "The Fortification of Tea with
Sweeteners and Milk and Its Effect on in Vitro Antioxidant Potential of Tea Product and Glutathione
Levels in an Animal Model." *Food Chemistry* 145 (February): 145–53.
https://doi.org/10.1016/j.foodchem.2013.08.016.

Krichel, Boris, Sven Falke, Rolf Hilgenfeld, Lars Redecke, and Charlotte Uetrecht. 2020. "Processing of
the SARS-CoV Pp1a/Ab Nsp7–10 Region." *Biochemical Journal* 477 (5): 1009–19.
https://doi.org/10.1042/bcj20200029.

Kyle, Janet A M, Philip C Morrice, Geraldine McNeill, and Garry G Duthie. 2007. "Effects of Infusion
Time and Addition of Milk on Content and Absorption of Polyphenols from Black Tea." *Journal of
Agricultural and Food Chemistry* 55 (12): 4889–94. https://doi.org/10.1021/jf070351y.

Li, Jianing, Robert Abel, Kai Zhu, Yixiang Cao, Suwen Zhao, and Richard A Friesner. 2011. "The VSGB
2.0 Model: A next Generation Energy Model for High Resolution Protein Structure Modeling."
Proteins 79 (10): 2794–2812. https://doi.org/10.1002/prot.23106.

Nelson, Kathryn M., Jayme L. Dahlin, Jonathan Bisson, James Graham, Guido F. Pauli, and Michael A.
Walters. 2017. "Curcumin May (Not) Defy Science." *ACS Medicinal Chemistry Letters*. American
Chemical Society. https://doi.org/10.1021/acsmedchemlett.7b00139.

Olsson, Mats H. M., Chresten R. Søndergaard, Michal Rostkowski, and Jan H. Jensen. 2011. "PROPKA3:
Consistent Treatment of Internal and Surface Residues in Empirical pK_a Predictions." *Journal of
Chemical Theory and Computation* 7 (2): 525–37. https://doi.org/10.1021/ct100578z.

Rashidinejad, Ali, E. John Birch, Dongxiao Sun-Waterhouse, and David W. Everett. 2017. "Addition of
Milk to Tea Infusions: Helpful or Harmful? Evidence from in Vitro and in Vivo Studies on
Antioxidant Properties." *Critical Reviews in Food Science and Nutrition* 57 (15): 3188–96.
https://doi.org/10.1080/10408398.2015.1099515.

Schrödinger Release 2018-1: LigPrep, Schrödinger, LLC, New York, NY, 2018. n.d. "No Title."

Sterling, Teague, and John J. Irwin. 2015. "ZINC 15 - Ligand Discovery for Everyone." *Journal of Chemical Information and Modeling* 55 (11): 2324–37. https://doi.org/10.1021/acs.jcim.5b00559.

Vincent, Martin J., Eric Bergeron, Suzanne Benjannet, Bobbie R. Erickson, Pierre E. Rollin, Thomas G. Ksiazek, Nabil G. Seidah, and Stuart T. Nichol. 2005. "Chloroquine Is a Potent Inhibitor of SARS Coronavirus Infection and Spread." *Virology Journal* 2 (August): 69. https://doi.org/10.1186/1743-422X-2-69.

Wang, Manli, Ruiyuan Cao, Leike Zhang, Xinglou Yang, Jia Liu, Mingyue Xu, Zhengli Shi, Zhihong Hu, Wu Zhong, and Gengfu Xiao. 2020. "Remdesivir and Chloroquine Effectively Inhibit the Recently Emerged Novel Coronavirus (2019-NCoV) in Vitro." *Cell Research*. Springer Nature. https://doi.org/10.1038/s41422-020-0282-0.

Yang, Haitao, Mark Bartlam, and Zihe Rao. 2006. "Drug Design Targeting the Main Protease, the Achilles Heel of Coronaviruses." *Current Pharmaceutical Design* 12 (35): 4573–90. https://doi.org/10.2174/138161206779010369.

ZINC_id	Energy	docking score	XP GScore	glide gscore	glide energy	glide emodel
ZINC00019583845 (Theaflavin Digallate)	66.45663	-15.263	-15.2929	-15.2929	-80.7874	-114.878
ZINC00019583845 (Theaflavin Digallate)	66.45663	-14.9138	-14.9437	-14.9437	-76.8608	-109.745
ZINC00019583845 (Theaflavin Digallate)	66.45663	-14.4314	-14.4613	-14.4613	-72.5158	-111.75
ZINC33861449	34.39056	-14.4268	-14.5059	-14.5059	-65.6439	-106.16
ZINC00008564535	33.10011	-14.24	-14.24	-14.24	-83.6571	-106.964
ZINC00008564535	30.89951	-14.2216	-14.2216	-14.2216	-72.4829	-105.12
ZINC00008564535	33.10011	-14.2128	-14.2128	-14.2128	-83.4784	-112.139
ZINC67903526	39.81764	-14.1658	-14.1944	-14.1944	-73.0363	-116.279
ZINC00008564535	33.10011	-14.1375	-14.1375	-14.1375	-82.697	-111.94
ZINC00008564535	33.10011	-14.0532	-14.0532	-14.0532	-81.513	-109.506

Table 1: The top 10 ligands after the virtual screening of 54386 unique compounds (909326 unique conformations). The ligands are sorted according to their docking score.