Structure-Based Virtual Screening of a Natural Product Database to Identify Several Possible SARS-CoV-2 Main Protease Inhibitors

Ettayapuram Ramaprasad Azhagiya Singam[1], Michele A. La Merrill[2], Kathleen A. Durkin[1], and Martyn T. Smith[3]*

[1]Molecular Graphics and Computation Facility, College of Chemistry, University of California, Berkeley, CA, USA

[2] Department of Environmental Toxicology, University of California, Davis, CA, USA

[3] Division of Environmental Health Sciences, School of Public Health, University of California, Berkeley, CA, USA

Abstract

A novel coronavirus (SARS-CoV-2) has been the cause of a recent pandemic of respiratory illness known as COVID-19. The lack of anti-viral drugs or vaccines to control the infection has resulted in an enormous number of seriously ill patients requiring hospitalization. In the absence of an effective vaccine, there is an urgent need for therapies which can fight COVID-19 infection. Readily available compounds in foods and plants may be one source of anti-viral compounds. Here, natural product chemicals from the Nuclei of Bioassays, Ecophysiology and Biosynthesis of Natural Products Database ($NuBBE_{DB}$) were screened against the main protease (Mpro) of SARS-CoV-2. This protease was chosen as a target due to its importance in the replication of SARS-CoV-2. Molecular docking was used to screen the natural products against Mpro to identify potential candidates. The identified candidates were further filtered using molecular dynamics simulation investigation. Nine natural compounds were identified for experimental validation, with carlinoside and quercetin 3-o-sophoroside being the top candidates.

Introduction

The coronavirus disease 2019 (COVID-19) presents as respiratory illness with symptoms such as a cough, fever, and in more severe cases, difficulty breathing and death.[1, 2] On March 11, 2020, the World Health Organization (WHO) declared the COVID-19 outbreak a pandemic, and it has infected over 2 million people worldwide as of April 15, 2020.[2] New antiviral drugs or vaccines against the novel coronavirus (SARS-CoV-2) which causes COVID-19 will not be available in the short term to treat or prevent COVID-19 as they require clinical trials and regulatory agency approval. As a stopgap measure, we chose to search for effective antivirals against SARS-CoV-2 that are natural products found in foods and plants, such that foods rich in antivirals could be added to the diet or taken as nutritional supplements, if available and reasonably safe to use.

The SARS-CoV-2 genome is composed of a long RNA strand that act as a messenger RNA when it infects a host cell and directs the synthesis of polyproteins required for multiplication of new viruses.[3] These proteins include a replication/transcription complex that synthesizes more RNA, several structural proteins that construct new virions, and two proteases. The SARS-CoV-2 main protease (Mpro) and papain-like protease (PLpro) are responsible for processing the viral proteins at a specific site into functional units for virus replication. Viral replication can be blocked by inhibiting the Mpro enzyme and thus Mpro is a target for drug design.[4] Recently, X-ray structures of the SARS-CoV-2 Mpro have been reported, showing that it is a dimer with two identical subunits that together form two active sites. These structures have small molecule or peptide-based inhibitors bound in the active site, showing the Mpro to be a suitable target for small molecule inhibitors.[5, 6] Recently, Yang and coworkers utilized structure based virtual screening combined with high throughput screening to identify new drug candidates such as ebselen, carmofur, etc, that target this Mpro of SARS-CoV-2.[6] Computational modeling has been used

to screen chemical libraries against SARS-CoV-2 protein targets with clearly defined 3D structures including Mpro.[4, 7-12] In this investigation, a natural product library from Nuclei of Bioassays, Ecophysiology and Biosynthesis of Natural Products Database ($NuBBE_{DB}$)[13] was screened against the active site of Mpro using molecular docking, molecular dynamics simulations, and MM-GBSA binding free energy calculations. From this data, several potential natural antiviral compounds were identified.

2. MATERIALS AND METHODS

2.1. Computational Methods

2.1.1. Virtual Screening Using Molecular Docking

SMILES notation of natural product chemicals was obtained from $NuBBE_{DB}$ (https://nubbe.iq.unesp.br/portal/nubbedb.html)[13]. The dataset contains 2,526 natural products which were then prepared for docking simulation using the LigPrep module[14] (LigPrep) in the Maestro software suite (Schrödinger; New York, NY, 2018-4). Prepared chemicals were then docked to Mpro (PDB ID: 6Y2G)[5] using the Glide docking program. Glide considers the protein to be a rigid entity whereas ligands can move flexibly relative to the binding site of the receptor. The structure of the Mpro ((PDB ID: 6Y2G)) was energy minimized using the protein preparation wizard, applying the OPLS3E force field[15] with default parameters. The docking grid was centered on the active site of the protease using default parameters for receptor grid generation. Docking was performed using Glide's XP (Extra Precision)[16, 17]. The $NuBBE_{DB}$ compounds were docked in the active site of Mpro and visually inspected for interaction between the catalytic residues and the ligands using the Pose Viewer module of Maestro. All compounds were ranked based on their docking score values and those with a score $\leq$ -10 and ligand having contact with catalytic residues were taken forward for investigation by molecular dynamics simulation.

Molecular Dynamics (MD) Simulation

MD simulations were performed using Desmond 3.2[18, 19] for the Mpro-ligand complexes shortlisted from the docking simulation, incorporating the OPLS3e force field for 100 nanosecond (nsec) simulation time. Each complex was set up for simulation using a TIP3P[20] water model as solvent in an orthorhombic box (sized for a 10 Å × 10 Å × 10 Å buffer distance around the complex) with periodic boundary conditions. Total charge of the system was neutralized by adding ions at a salt concentration of 0.15 M NaCl.

After building the solvated system, we performed minimization and relaxation of the protease-ligand complex using a default protocol of Desmond. This includes a total of 8 stages as follows: Stage 1 – Minimization; Stage 2 - Simulate, Brownian Dynamics NVT, T = 10 K, small timesteps, and restraints on solute heavy atoms, 100 picosecond (psec); Stage 3 - Simulate, NVT, T = 10 K, small timesteps, and restraints on solute heavy atoms, 12psec; Stage 4 - Simulate, NPT, T = 10 K, and restraints on solute heavy atoms, 12psec; Stage 5 – Solvate pocket; Stage 6 - Simulate, NPT and restraints on solute heavy atoms, 12psec; Stage 7 - Simulate, NPT and no restraints, 24psec; and, Stage 8 – Production run for 100nsec. The production run was performed for 100nsec using a 2 femtosecond (fsec) time step for integration of the equation of motion in the NPT ensemble at 300 K and at 1 atmospheric pressure, which were controlled by Nose-Hoover thermostat algorithm[21] and Martyna-Tobias-Klein Barostat algorithm[22]. The trajectories were saved every 50ps for a total of 2000 frames for each simulation.

Binding Free Energy Calculation

The Molecular Mechanics Generalized Born Surface Area (MM-GBSA)[23] method in the Prime module was used to calculate the binding free energy (ΔG_{bind}) using equation 1

$$\Delta G_{bind} = E_{Complex} - E_{Ligand} - E_{Receptor} \quad (1)$$

where $E_{Complex}$, E_{Ligand}, $E_{Receptor}$ are the energies calculated from the complex, free ligand and free receptor, respectively. These energies were calculated using the OPLS3E force field and VSGB solvation model.[23] 50 frames were extracted at from the last 25nsec MD trajectories to calculate average ΔG_{bind}. The ligands with average ΔG_{bind} values of ≤ -45 kcal/mol were further analyzed for critical interactions.

Results and Discussion

Virtual Screening

The main objective of the present study was to screen for natural product chemicals which potentially binds to Mpro. Virtual screening of 2,526 compounds from $NuBBE_{DB}$ was performed using molecular docking and MD simulation. In 2020, two X-ray crystal structures of Mpro-ligand complexes were reported.[5, 6] The RCSB PDB codes were 6LU7 and 6Y2G. 6LU7 and 6Y2G was reported with N-[(5-methylisoxazol-3-yl)carbonyl]alanyl-l-valyl-n~1~-((1R,2Z)-4-(benzyloxy)-4-oxo-1-{[(3R)-2-oxopyrrolidin-3-yl]methyl}but-2-enyl)-l- leucinamide and ~{tert}-butyl ~{N}-[1-[(2~{S})-3-cyclopropyl-1-oxidanylidene-1-[[(2~{S},3~{R})-3-oxidanyl-4-oxidanylidene-1-[(3~{S})-2-oxidanylidenepyrrolidin-3-yl]-4-[(phenylmethyl)amino]butan-2-yl]amino]propan-2-yl]-2-oxidanylidene-pyridin-3-yl]carbamate ligands.[5, 6] The interaction of both ligands with the Mpro catalytic residues, including His4, Cys145 and Gln166, was confirmed in molecular docking. We then carried forward only those $NuBBE_{DB}$ ligands which had interactions with these key residues and had a docking score ≤ -9.5. From this first stage of the calculations, 18 compounds were identified as initial hits based on the docking score and type of interactions with active site residues. **Figure 1** summarizes the shortlisted ligand structures, their docking scores, and the Mpro active site amino acid residues with which these chemicals interact.

Stability of complexes – MD simulation Analysis

To explore the stability of shortlisted ligands in the catalytic site of the protease, the 100nsec MD trajectories was analyzed using the root mean standard deviation (RMSD). RMSD was calculated for all backbone atoms of the protein and for all heavy atoms of the ligand. The protein-ligand complex was first aligned to the protein backbone of the reference (first MD frame) and then both the backbone RMSD and the RMSD of the ligand heavy atoms were measured as a function of time. If the ligand RMSD values fluctuated significantly, then the ligand diffused away from the binding pocket. The calculated RMSD for all Mpro-ligand complexes is given in supporting information **Figure S1**. Visual inspection of the trajectory and the RMSD plots reveal that NuBBE 286, NuBBE 282, NUBBE 271 fluctuated significantly from their initial docked conformation, which suggested that these ligands diffuse away from the Mpro binding pocket. Stable binding was observed for all other ligands. Hence, NuBBE 286, NuBBE 282, NuBBE 271 were ignored for further analysis. We then calculated the mean binding free energy for the remaining 15 ligands using MM-GBSA (**Table 1**). We identified 11 chemicals with average ΔG_{bind} values of ≤ -45 kcal/mol. NuBBE 278, NuBBE 273, NuBBE 420, and NuBBE 1178 with ΔG_{bind} values of ≥ -45 consistent with their instability inside the binding pocket of Mpro. Hence, we omitted these four chemicals for further analysis.

Protein-Ligand Interaction – A Contact Analysis

The interactions between ligand and residues of Mpro were calculated as a function of time using MD stimulations (**Figure 2**). For example, analysis of contacts revealed that all ligands remained bound to Mpro by predominantly interacting with Glu166, Thr190, Gln192 and/or Gln189 (**Figure 2**). NuBBE 204 lacked an important contact with the catalytic residues His41 and Cys145 whereas

remining ten ligands had constant contact with these residues. The nature of the predicted interaction between ligand and active site residue was analyzed according the prevalence of four types of ligand protein interactions: hydrogen-bonding, hydrophobic, ionic bonds and water bridges (**Figure 3**). Hydrogen bonding interactions are important because of their strong influence on drug specificity, metabolization and adsorption. Hence, we further ignored the ligand, NuBBE 1182 (**Figure 3(j)**), which lacked stable hydrogen bonds with catalytic residues during the MD simulations. Indeed, this approach revealed that nine other ligands form a hydrogen bond with His41, Cys145, and Gln166 (**Figure 3(a-i)**). In addition to these key interactions, ligands also establish contacts with the protease via water mediated hydrogen bonds, hydrophobic and other stabilizing interactions. After evaluating these interactions, nine chemicals (NuBBE 1245, NuBBE 142, NuBBE 360, NuBBE 46, NuBBE 1170, NuBBE 602 and NuBBE 283, NuBBE 125, and NuBBE 361) were shown to be promising ligands for Mpro and were identified for experimental validation, with carlinoside (NuBBE 1245) and quercetin 3-O-sophoroside (NUBBE 142) being the best potential ligands of Mpro on the basis of docking, MM-GBSA scores, and binding stability. Both carlinoside and quercetin 3-O-sophoroside are readily absorbed[24, 25] and are good antiviral candidates[26, 27] with several available natural sources including unfermented Rooibos tea[28]. The results of this investigation should not be taken as medical advice or encouragement for people to take supplements of these compounds as excessive intake of flavonoids may cause adverse effects on human health, including the inhibition of thyroid hormone synthesis[29]. However, elevating dietary intake of flavonoids by eating plant materials rich in these substances, such as bell peppers, brassica plants like broccoli, pigeon pea leaves and drinking rooibos tea is unlikely to be harmful and may be helpful. More research is needed to confirm the anti-viral effects of the candidate compounds in biological systems and to determine the safe doses of these flavonoids.

Conclusion

There is an urgent need to discover methods for treating the early stages of COVID-19 caused by SARS-CoV-2. Mpro is one of the potential targets for antiviral treatment against SARS-CoV-2. Therefore, we used molecular docking and molecular dynamics simulation techniques to screen chemicals from a large natural product database to identify novel inhibitors. Molecular docking and MD simulation showed 9 natural compounds having strong predicted binding affinities for the catalytically important residues of Mpro, with carlinoside and quercetin 3-O-sophoroside being the strongest.

Acknowledgements

We thank the Office of Environmental Health Hazard Assessment of the California EPA and the College of Chemistry at UC Berkeley for their support of our capabilities in computational chemistry. The Molecular Graphics and Computational Facility is funded by NIH S10OD023532. MTS was supported in part by an endowment from Kenneth and Marjorie Kaiser. MAL was supported in part by USDA National Institute of Food and Agriculture, Hatch project: 1002182. The authors have no financial conflicts affecting this work.

Table 1. Calculated Binding free energy ΔG_{bind} (kcal/mol)

Common Name	$NuBBE_{DB}$ ID	ΔG_{bind} (kcal/mol)
Carlinoside	NuBBE 1245	-74.32±4.77
Quercetin 3-O-sophoroside	NuBBE 142	-72.03±7.51
Kaempferol 3-O-α-L-rhamnopyranosyl (1→6)-O-[β -D-glucopyranosyl (1→3)-O-α-L-rhamnopyranosyl-(1→2)]-O-β -D-glucopyranosyl	NuBBE 360	-69.28±10.02
Rutin	NuBBE 46	-59.36±7.08
Isocarlinoside	NuBBE 1170	-58.35±7.48
6-Hydroxy-rutin	NuBBE 283	-55.12±8.71
Nitensoside B; Pedalitin 6-O-α-rhamnopyranosyl(1'''→6'')-β-glucopyranoside	NuBBE 125	-54.28±8.07
N/A	NuBBE 602	-52.25±5.95
Quercetin 3-O-α-L-rhamnopyranosyl (1→6)-O-[β -D-glucopyranosyl (1→3)-O-α-L-rhamnopyranosyl-(1→2)]-O-β -D-glucopyranosyl	NuBBE 361	-49.98±6.36
3,4,5-Trimethoxyphenyl-1-O-β-D-(5-O-syringoyl)-apiofuranosyl-(1→6)-β-D-glucopyranoside	NuBBE 204	-49.41±5.46
Ellagic acid 4-O-α-L-4''-O-acetylrhamnopyranoside	NuBBE 1182	-45.42±6.46
1-O-(E)-caffeoyl-β-D-glucopyranoside	NuBBE 273	-42.32±6.26
Alboside IV	NuBBE 420	-38.95±8.65
Casuarinin	NuBBE 1178	-37.59±16.65
Chimarrhoside	NuBBE 278	-29.10±23.98

N/A Not Available

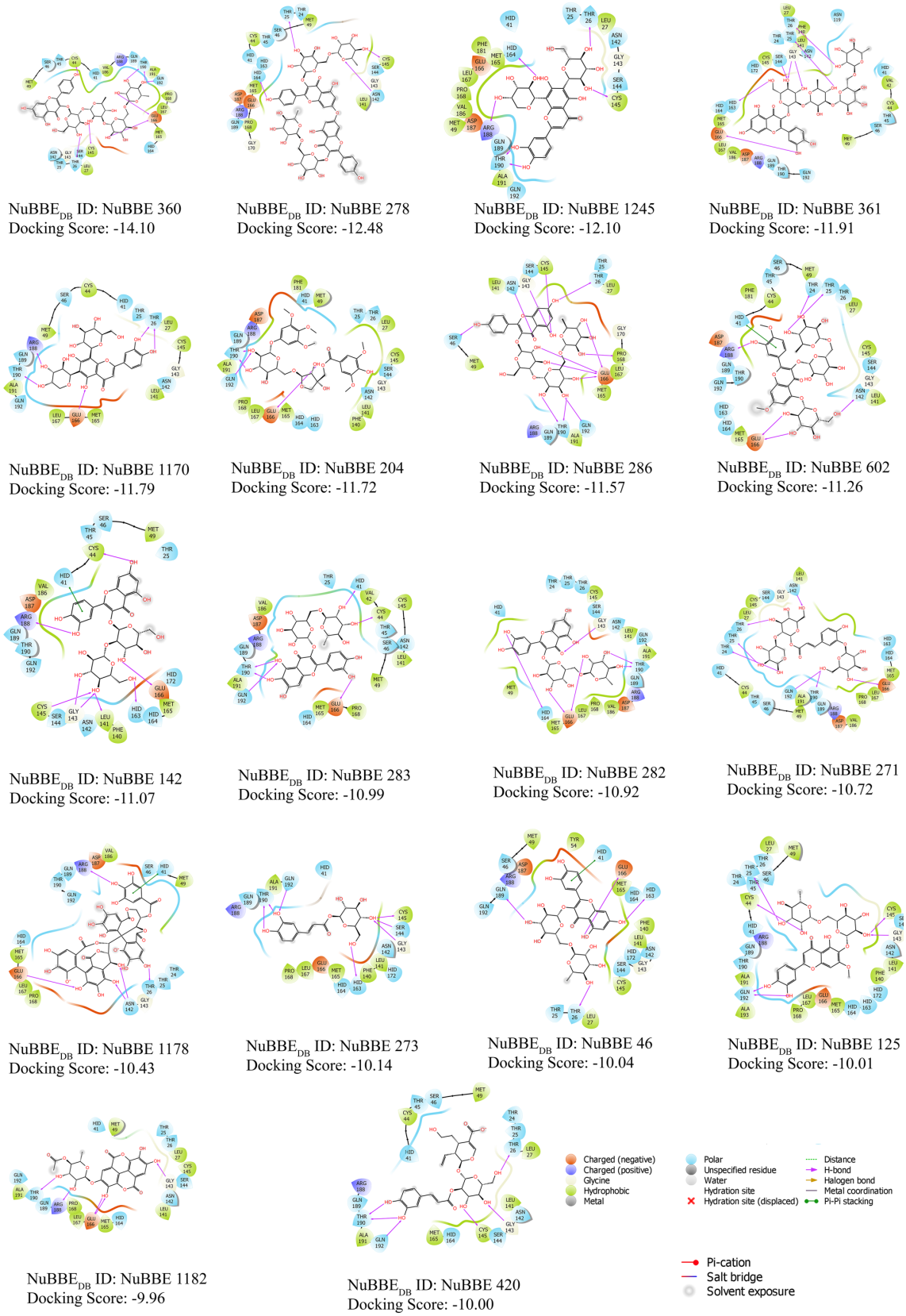

Figure 1: Ligand interaction diagram of initial shortlisted chemicals from NuBBE$_{DB}$ using molecular docking calculations against COVID-19 Mpro

Figure 2. Contacts between residues of protein and ligand a) NuBBE1245, b) NuBBE1182, c) NuBBE1170, d) NuBBE602 e) NuBBE361, f) NuBBE360, g) NuBBE283, h) NuBBE204 i) NuBBE142 j) NuBBE125 and k) NuBBE46 as a function of time. Time in nsec and Mpro residue number in x-axis and y-axis respectively

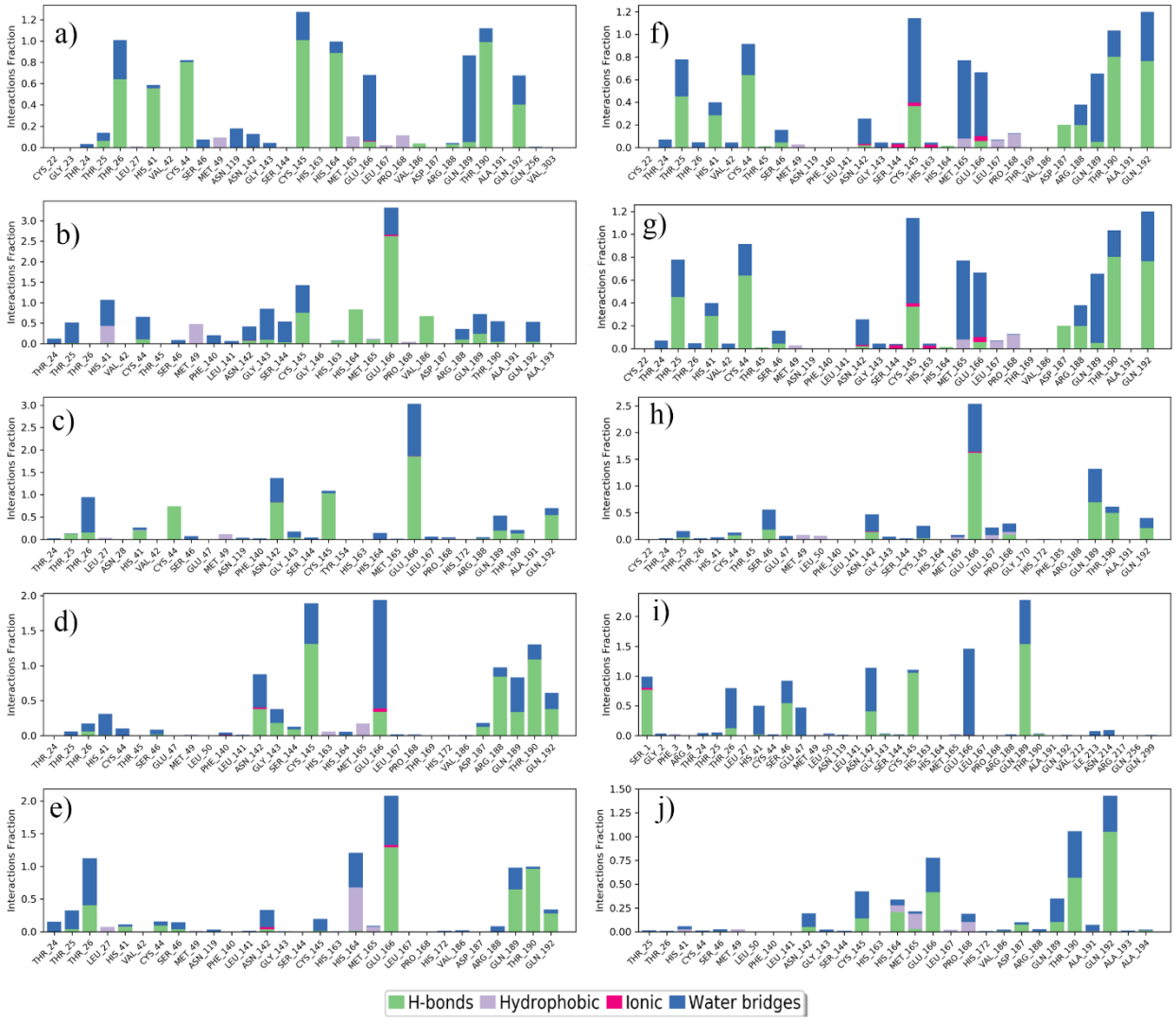

Figure 3. Interaction analysis between the active site amino acids and the different ligands (a) NuBBE 1245, (b) NuBBE 142, (c) NuBBE 360 (d) NuBBE 46 (e) NuBBE 1170 (f) NuBBE 283 (g) NuBBE 125 (h) NuBBE 602 (i) NuBBE 361 (j) NuBBE 1182.The average values of the occupancy interactions were calculated along the 100 ns MD simulations. The stacked bar charts are normalized over time, if a value 0.5 means 50% of the simulation time a particular interaction is maintained in the MD trajectory. Values over 1.0 are possible as same amino acid residue making multiple contacts with the ligand.

References

1. Li, Q., et al., *Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus-Infected Pneumonia.* N Engl J Med, 2020. **382**(13): p. 1199-1207.
2. Sohrabi, C., et al., *World Health Organization declares global emergency: A review of the 2019 novel coronavirus (COVID-19).* International Journal of Surgery, 2020.
3. Malik, Y.S., et al., *Emerging novel coronavirus (2019-nCoV)-current scenario, evolutionary perspective based on genome analysis and recent developments.* Vet Q, 2020. **40**(1): p. 68-76.
4. Ahmed, S.F., A.A. Quadeer, and M.R. McKay, *Preliminary Identification of Potential Vaccine Targets for the COVID-19 Coronavirus (SARS-CoV-2) Based on SARS-CoV Immunological Studies.* Viruses, 2020. **12**(3).
5. Zhang, L., et al., *Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved α-ketoamide inhibitors.* Science, 2020.
6. Jin, Z., et al., *Structure of M(pro) from COVID-19 virus and discovery of its inhibitors.* Nature, 2020.
7. Ortega, J.T., et al., *Unrevealing sequence and structural features of novel coronavirus using in silico approaches: The main protease as molecular target.* EXCLI J, 2020. **19**: p. 400-409.
8. ul Qamar, M.T., et al., *Structural basis of SARS-CoV-2 3CLpro and anti-COVID-19 drug discovery from medicinal plants.* Journal of Pharmaceutical Analysis, 2020.
9. Liu, C., et al., *Research and Development on Therapeutic Agents and Vaccines for COVID-19 and Related Human Coronavirus Diseases.* ACS Cent Sci, 2020. **6**(3): p. 315-331.
10. Sun, N., W.-L. Wong, and J. Guo, *Prediction of Potential 3CLpro-Targeting Anti-SARS-CoV-2 Compounds from Chinese Medicine.* 2020.
11. Dai, W., et al., *Structure-Based Design, Synthesis and Biological Evaluation of Peptidomimetic Aldehydes as a Novel Series of Antiviral Drug Candidates Targeting the SARS-CoV-2 Main Protease.* bioRxiv, 2020, https://doi.org/10.1101/2020.03.25.996348.
12. Huang, X., R. Pearce, and Y. Zhang, *Computational Design of Peptides to Block Binding of the SARS-CoV-2 Spike Protein to Human ACE2.* bioRxiv, 2020, https://doi.org/10.1101/2020.03.28.013607.
13. Pilon, A.C., et al., *NuBBEDB: an updated database to uncover chemical and biological information from Brazilian biodiversity.* Sci Rep, 2017. **7**(1): p. 7215.
14. LigPrep, *Schrödinger release 2019–2; Schrödinger, LLC: New York, NY, 2019.*
15. Harder, E., et al., *OPLS3: A Force Field Providing Broad Coverage of Drug-like Small Molecules and Proteins.* J Chem Theory Comput, 2016. **12**(1): p. 281-96.
16. Friesner, R.A., et al., *Extra precision glide: Docking and scoring incorporating a model of hydrophobic enclosure for protein-ligand complexes.* Journal of Medicinal Chemistry, 2006. **49**(21): p. 6177-6196.
17. Friesner, R.A., et al., *Glide: A new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy.* Journal of Medicinal Chemistry, 2004. **47**(7): p. 1739-1749.
18. 2020-1:, S.R., *Desmond Molecular Dynamics System, D. E. Shaw Research, New York, NY, 2020. Maestro-Desmond Interoperability Tools, Schrödinger, New York, NY.* 2020.
19. Bowers, K.J., et al. *Scalable algorithms for molecular dynamics simulations on commodity clusters*. in *SC'06: Proceedings of the 2006 ACM/IEEE Conference on Supercomputing*. 2006. IEEE.
20. Jorgensen, W.L., et al., *Comparison of simple potential functions for simulating liquid water.* The Journal of chemical physics, 1983. **79**(2): p. 926-935.
21. Evans, D.J. and B.L. Holian, *The nose–hoover thermostat.* The Journal of chemical physics, 1985. **83**(8): p. 4069-4074.
22. Martyna, G.J., D.J. Tobias, and M.L. Klein, *Constant pressure molecular dynamics algorithms.* The Journal of chemical physics, 1994. **101**(5): p. 4177-4189.

23. Li, J., et al., *The VSGB 2.0 model: a next generation energy model for high resolution protein structure modeling.* Proteins, 2011. **79**(10): p. 2794-812.
24. Duke, J. and M.J. Bogenschutz, *Dr. Duke's phytochemical and ethnobotanical databases*. 1994: USDA, Agricultural Research Service.
25. Shinbo, Y., et al., *KNApSAcK: a comprehensive species-metabolite relationship database*, in *Plant metabolomics*. 2006, Springer. p. 165-181.
26. Das, S., et al., *Impact of edaphic factors and nutrient management on the hepatoprotective efficiency of Carlinoside purified from pigeon pea leaves: An evaluation of UGT1A1 activity in hepatitis induced organelles.* Environ Res, 2018. **161**: p. 512-523.
27. Yan, X., et al., *Traditional Chinese Medicines: Molecular Structures, Natural Sources and Applications: Molecular Structures, Natural Sources and Applications*. 2018: Routledge.
28. Iswaldi, I., et al., *Identification of phenolic compounds in aqueous and ethanolic rooibos extracts (Aspalathus linearis) by HPLC-ESI-MS (TOF/IT).* Anal Bioanal Chem, 2011. **400**(10): p. 3643-54.
29. Skibola, C.F. and M.T. Smith, *Potential health impacts of excessive flavonoid intake.* Free Radic Biol Med, 2000. **29**(3-4): p. 375-83.