Accepted Article: 06/05/2020 **Uncorrected Proof**

Virtual screening assisted identification of small molecule against 2019 novel coronavirus protease enzyme.

Arkadeep Sarkar[1], Deepak Shilkar[2], Venkatesan Jayaprakash[2], Arindam Maity[3], Sarmi Sardar[1], Sudhan Debnath[4], Debanjan Sen[1c]

[1]BCDA College of Pharmacy and Technology, Hridaypur, Kolkata, 700127, India

[2]Department of Pharmaceutical Sciences & Technology, Birla Institute of Technology, Mesra, Ranchi 835215, Jharkhand, India

[3]School of Pharmaceutical Technology, Adamas University, Kolkata,

700126, India

[4]Department of Chemistry, Maharaja Bir Bikram College, Agartala,

799 004, Tripura, India

Abstract: The 2019 novel coronavirus infection or COVID-19 can be designated as a global threat. Till date, there is a lack of dedicated therapeutics available against this fatal infection. In the present work, we performed structure-based drug design studies in order to identify clinically used molecules exhibiting crucial binding with 2019-coronavirus main protease enzyme. Based on ligand binding energy and interaction with essential amino acids, two molecules were selected. The stability of the complexed molecules with main protease enzyme was further studied by performing molecular dynamics simulation.

Keyword: Drug Re-purposing, GROMACS, Molecular Dynamics, SARS-CoV-2

Introduction:

The alarming outbreak of Severe Acute Respiratory Syndrome, coronavirus 2 (SARS-CoV-2) earlier named 2019 novel coronavirus (2019-nCoV) or disease (COVID-19) in China at the

end of 2019 has rung alarm bells worldwide as a major public health situation. As we write these sentences, around 1,000,000 cases were confirmed all over the world. The maximum mortality cases were detected in Italy [1]. According to the WHO, a suspected patient will have symptoms such as acute respiratory infection, gastrointestinal tract infection, kidney damage, and multi-organ failure. In the case of COVID-19, the diagnosis was set in motion soon after isolation and characterization of the virus. The treatment regimen identified so far includes prophylactic antibiotics administration to prevent secondary infections and administration of broad spectrum antivirals to prevent viremia. To this date, no targeted therapy has been found [2]. The COVID-19 virus is the largest positive-stranded RNA virus consisting of approximately 30,000 nucleotides as a part of their genome, which is 76% similar to the SARS CoV betacoronavirus. There are three types of betacoronaviruses, which include SARS CoV (Severe Acute Respiratory Syndrome), MERS, and SARS CoV2. (SARS).[2] SARS bind to the ACE2 receptor of the lungs cell of the upper respiratory tract and MERS bind to the DPP4 receptor of the lung's cells of the lower respiratory tract. The genome of SARS coronavirus contains 29,727 nucleotide and Polyadenylated RNA. The replicase gene (ORF1a & ORF1b) of SARS CoV2 virus encodes two polyproteins, which undergo cotranslational proteolytic processing.[3] One sample from the Wuhan Institute of Virology (WIV), (WIV04), collected from bronchoalveolar lavage fluid (BALF), was analysed by metagenomics analysis. Of the 10,038,758 total reads—a total of 1,582 readings were retained after filtration of readings from the human genome—1,378 (87.1%) sequences matched that of SARSrCoV sequences.[4]

The complete analysis of the genome sequence revealed that there is a 76% similarity between 2019-nCoV and SARS (Severe Acute Respiratory Syndrome) Coronavirus (3). The endosomal uptake of the virus occurs after entering into the host cell by the ACE2 receptor through the clathrin-mediated pathway. The Lysosomal protein cathepsin B and L protein is produced at the acidic pH of the host cell endo-lysosome, which promotes the viral replicase genes ORF1a (Open ring frame) and ORF1b to encode 16 types of viral non-structural polyproteins (Nsp) including main protease (M_{pro} or $3CL_{pro}$), papain like protease (PLPpro) and RNA dependent RNA polymerase (RdRp) enzymes, and four structural polyproteins secreted from the endo-lysosome. M_{pro} and RdRp play a crucial role in viral RNA transcription in the host cell. In a recent study, it was revealed that, through the transmembrane serine 2 protease (TMPRSS2) receptor of the host cell, which is an independent endosomal pathway, the virus could use to enter the host cell (4).

As there is no antiviral drug is available for the treatment of COVID-19, the repurposing of existing drugs for treatment can be an effective approach for the treatment.

Previous research efforts for the development of an anti-viral drug for COVID-19 described that ACE2 receptor, RdRp, and M_{pro} can be an effective target for new drug discovery against COVID-19. Although there was observed a significant side effect due to the blocking of the ACE2 receptor, which allows the entry of SARS CoV2 into the host cell. Likewise, RdRp receptor inhibitors are not very specific and have lower potency, which also causes a significant side effect in patients (5). Recently, COVID-19 infected patients administered with potential protease inhibitors like lopinavir/ritonavir, have shown an improved therapeutic outcome (6), which demonstrates that M_{pro} or the main protease enzyme can be a promising target for drug discovery against COVID-19.

Recently published X-ray crystal structure of SARS CoV2 main protease enzyme (M_{pro}) by Liu et al. (2020), is considered as a significant breakthrough in novel coronavirus research (7). The structure explains that the main protease enzyme has a catalytic domain consisting HIS41 and CYS145 amino acid residues, which regulates the polyprotein to form single polypeptides that are required for replication and interpretation or transcription (8).

In this present study, through virtual screening based molecular docking analysis, a small molecule Argatroban and Piperacillin were identified as a potential SARS CoV2 main protease enzyme (M_{pro}) inhibitor. The stability of the protein-ligand complex was also analyzed through a 5 ns molecular dynamics study, and the RMSD and RMSF values of the protein-ligand complex and protein-ligand backbone suggest that the potential main protease inhibitors, Argatroban and Piperacillin are stable with the main protease enzyme protein complex of COVID-19.

Materials and Methods:

Preparation and purification of protein

The main protease enzyme protein of 2019 SARS CoV2 was collected from Protein Data Bank (PDB-ID 6LU7) (www.rcsbpdb.org). The protein was prepared by using AutoDock Tools (9). The water molecules were removed, and polar hydrogen were added by AutoDock Tools (9) and the protein was prepared in pdbqt format. Gasteiger and Kollman charges were added to the protein, and all other necessary preparative steps were performed. Along with co-crystalline ligand, all others non-amino acid residues were removed from the protein structure. By using the grid module of the software, the binding site information of the protein was collected, and a grid box text file was generated.

Preparation of ligand

For the recent experiment, the 3D structures of 10,6910 potential FDA approved drugs like ligands were collected from DrugBank (www.drugbank.ca). Then by following a script program written by *Samadani et.al 2018* (10), the lowest energy conformers of the collected ligands were generated in pdbqt format, which was further used for the virtual screening based molecular docking study. 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 co-crystalline ligand was also selected and prepared by following the above mention script program, to validate the pharmaco-informatics approach considered for this study. To generate the ligands, genetic algorithm conformations were used and 1000 steps minimization was carried away to prepare the ligands. The gaff force field was applied and 3D conformation of the input ligands were generated.

Virtual screening

For performing molecular docking-based analysis of prepared dataset with 2019 coronavirus main protease enzyme, a written script programmed by *Samadani et.al 2018* was used (11) by following the default operating process described in the software manual. After the molecular docking study, the binding energy was calculated, and interacting residue of the protein with the ligands were observed by using LigPlot software (12) and Pymol (13).

Molecular dynamics

Molecular dynamics study was conducted by GPU accelerated Gromacs 2019.5 software running over Linux Mint operating system supported by Intel Core i9-9900K processor system. SwissParam (http://www.swissparam.ch/), a server-based parameter generation tool was used to generate ligand parameters. Charm36 (www.charmm.org) force field was considered followed by using the TIP3P water model for this simulation.

SARS CoV2 main protease protein collected from PDB ID-6LU7

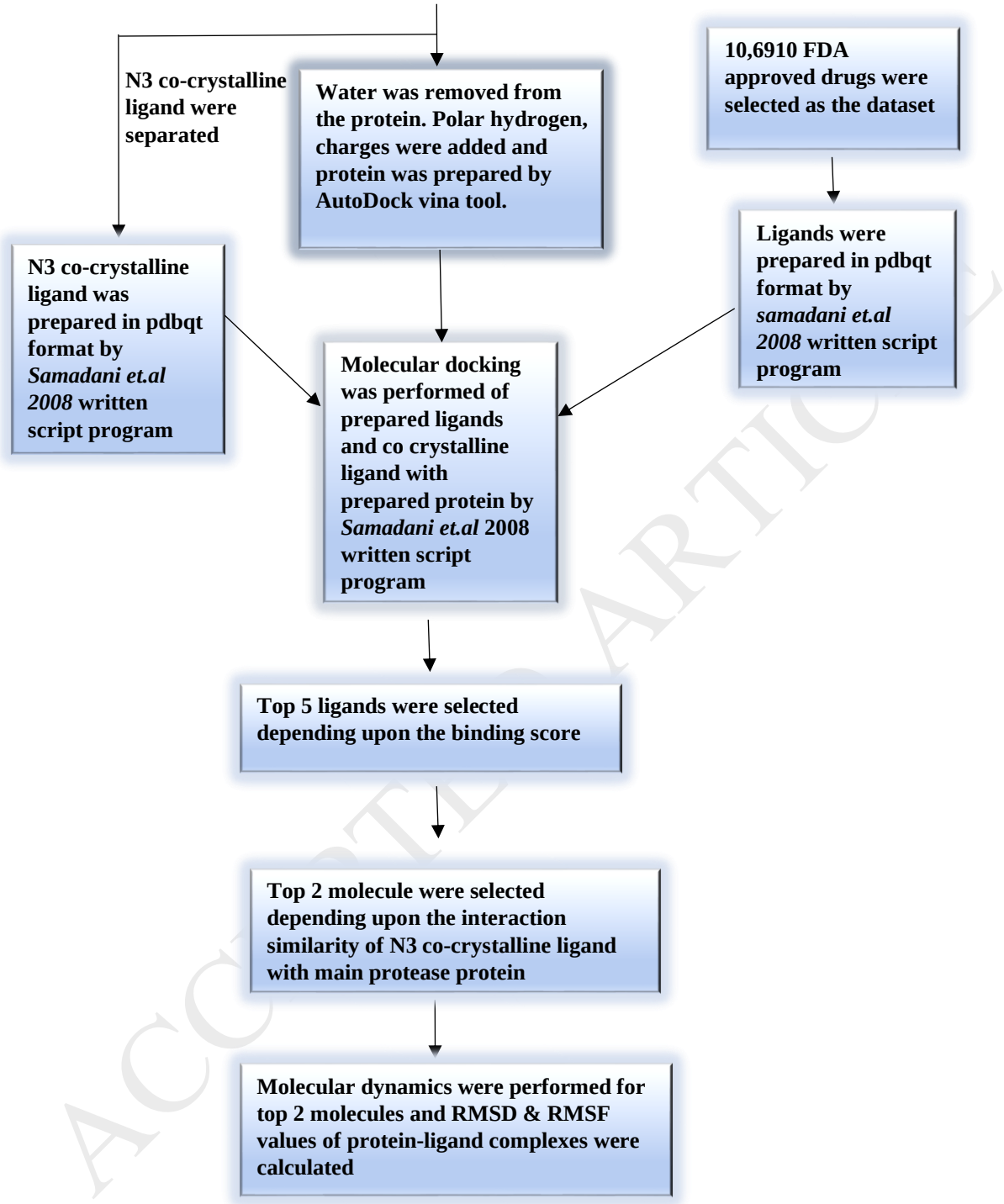


Figure1: Flow chart depicting the schematic strategy for identification of novel SARS CoV2 M_{pro} inhibitors

Results

Virtual screening

In this current study, virtual screening-based structure-based drug design approach was undertaken. A total of 10,691 compounds were obtained from DrugBank and screened in order to identify molecules exhibiting interaction with critical binding site residues of main protease enzyme protein (M_{pro}) of 2019 SARS CoV2. The co-crystalline ligand of main protease enzyme complex of SARS CoV2 was considered as the reference standard, and the interaction was analyzed by LigPlot software, which showed that the ligand N3 form hydrogen bond with HIS163, HIS164, PHE140, GLU166, THR190, GLN189, GLY143, CYS145, and shows hydrophobic interaction with LEU141, HIS172, THR25, THR26, THR24, ASN142, HIS41, SER144, MET165, ARG188, GLN192, ALA191, PRO168. Depending upon the equivalent interaction and binding energy, six compounds were considered for further analysis listed in Table 1. The pharmacological properties of the molecule, binding affinity, H-bond distance and interacting residue were further analyzed to select the best protease inhibitor and identified as Argatroban (14) and piperacillin (15). It was observed through LigPlot analysis that, Argatroban exhibit interaction by forming H-bond with HIS163, GLU166, HIS41, GLY143, LEU141. Hydrophobic interactions were observed with PHE140, HIS172, ASN142, THR26, CYS145, LEU27, HIS164, ARG188, ASP187, MET49, MET165 and GLU189 amino acid residues. Whereas, piperacillin interacted by forming the hydrogen bond with CYS145, HIS41, GLU166 residue and formed hydrophobic interaction with PHE140, LEU141, ASN142, MET165, ASP187, HIS164, TYE54, MET49, GLN189 and THR25 residues of main protease enzyme (M_{pro}) with a binding energy -8.7 kcal/mol and -7.9 Kcal/mol, respectively (**Figure 2**). Interaction in 2D molecular representation plots of the molecules are shown in **Figure 3(a-f)**.

Based on the above fact, that is interaction with critical amino acid residues present in M_{pro} binding pocket, Argatroban and Piperacillin were considered for further evaluation. Pharmacological properties of these molecules depict Argatroban is used as a selective thrombin inhibitor (Commercially available doses 1 mg/mL to 100 mg/mL) and exhibit 100% bioavailability after intravenous injection. It shows plasma half-life 39 - 51 min and 65% of drug eliminated through faeces and 22% of drug eliminated through urine. Mainly it is used for the management of Ischemic Stroke, Coronary Artery Disease, and Unstable Angina Pectoris. Piperacillin is a class of broad-spectrum penicillin antibiotics exhibiting biological half-life 36 - 72 minutes. From literature survey, the volume of distribution for this molecule was found to be 101 mL/kg, and in neonates, 50 mg/kg (5-minute infusion) was found after intravenous administration without any significant toxicity. The above pharmaceutical profile infers that these molecules can be re-purposed safely for developing new therapy against COVID-19.

Considering the above properties of these molecules, we undertake molecular dynamics strategies in order to evaluate the stability of these two protein-ligand complexes. A 5 ns simulation was conducted, and from the trajectory, various parameters like RMSD, RMSF, were calculated (**Figure 4**). Mean backbone RMSD of 0.1nm and 0.16 nm were found for Argatroban and Piperacillin, respectively. The mean RMSD of the protein-ligand complex was found to be 0.25 and 0.2 nm, respectively. RMSF plot depicts no significantly abnormal fluctuations in binding site residues. In addition to analyze the compactness of the system radius of gyration (Rg) plot was calculated (Figure 4d). The h-bond interaction observed during 5ns simulation was plotted against time in nano second and shown in Figure 4e.

Discussion:

From the analysis of the interacting residue of the co-crystalline ligand with M_{pro} through LigPlot study, it was observed that the co crystalline ligand formed a hydrogen bond with HIS163, HIS164, PHE140, GLU166, THR190, GLN189, GLY143, CYS145 and with the LEU141, HIS172, THR25, THR26, THR24, ASN142, HIS41, SER144, MET165, ARG188, GLN192, ALA191, PRO168 forming a hydrophobic interaction. Considering these interactions as the standard, the analysis of the top 6 molecules was performed through the LigPlot software. It was observed that the proposed molecules Argatroban interacted with main protease enzyme of SARS CoV2 by forming a hydrogen bond with HIS163, GLU166, HIS41, GLY143, LEU141, in which the bond length with HIS163, GLU166, HIS41, GLY143, LEU141 were 3.30Å , 3.13Å , 3.23Å , 3.34Å & 3.23 Å respectively and also the hydrophobic interaction was observed with PHE140, HIS172, ASN142, THR26, CYS145, LEU27, HIS164, ARG188, ASP187, MET49, MET165 and GLU189. Whereas, Piperacillin interacted by forming the hydrogen bond with CYS145, HIS41, GLU166 residue and formed hydrophobic interaction with PHE140, LEU141, ASN142, MET165, ASP187, HIS164, TYE54, MET49, GLN189 and THR25. The most similar hydrogen bond interaction with co-crystallised ligand was observed with HIS163, GLU166, and GLY143 amino acid for Argatroban with a binding affinity -8.7 kcal/mol and Piperacillin, the same as co-crystallied ligand hydrogen bond formation observed with CYS145, HIS41, and GLU166 amino acid residues at a bond length 2.99Å, 2.82Å, and 3.00Å respectively with a binding affinity of -7.9 kcal/mol other than four molecules. Further pharmacological analysis was performed, which proposed that the two molecules can be effective SARS CoV-2 protease inhibitor. Molecular dynamics analyse t protease enzyme protein of SARS CoV-2 complexed with ligands depict, up to ~2 ns both of the molecules show an identical pattern like crystalline ligand N3. After ~2.5 ns Argatroban follows an almost

identical deviation like N3, However piperacillin exhibits more stable RMSD in compare to the other two molecules. Besides RMSD pattern of the protein backbone, protein-ligand complex RMSD depict, both of the molecules follow an almost identical pattern like N3. However, after ~2 ns a sharp increment in RMSD was found with Argatroban which is stabilise after 4 ns. Piperacillin complex gradually exhibits increased RMSD after ~3 ns and stabilize after 4.5 ns. The binding site amino acid residues that is residue ID 140 to 160, Argatroban and Piperacillin complex exhibit comparatively less fluctuation in compared to co-crystallised ligand N3. Analysis of radius of gyration plot depict Argatroban shows identical compectness when compared with thermodynamically stable co-crystalline ligand N3, but remearkeble diviation was observed Pipperacillin – protein complex. H-bond interaction calculated from 5ns molecu;ardynamics trajectory indicates Argatroban forms average 4 numbers of H-bond with this protein throughout the simulation. The above facts infer both the drugs Argatroban and Piperacillin in complex with main protease enzyme show stability up to 5 ns molecular dynamics followed by acceptable binding interactions with this enzyme. This information can be used for futures to studies in order to discover new therapeutic options against COVID-19. However, a long simulation will be conducted by our group in order to explore even more dynamics behaviour of these two molecules.

Conclusion:

In this study, we have reported two potential SARS CoV2 main protease enzyme inhibitors Argatroban and Piperacillin with a binding affinity -8.7 kcal/mol and -7.9 kcal/mol which are showing interaction with the crucial amino acid residues present in the catalytic domain of COVID-19 main protease enzyme. A 5 ns molecular dynamics result concluded that the molecules can be further analysed to determine the molecular stability of the protein-ligand complex. However, in vitro and in vivo studies are the ultimate options to investigate these molecules further as potential SARS CoV-2 main protease enzyme inhibitor.

References:

1. Lai, Chih-Cheng, et al. "Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and corona virus disease-2019 (COVID-19): the epidemic and the challenges." *International journal of antimicrobial agents* (2020): 105924.

2. Peeri, Noah C., et al. "The SARS, MERS and novel coronavirus (COVID-19) epidemics, the newest and biggest global health threats: what lessons have we learned?." *International Journal of Epidemiology* (2020).

3. Brian, D. A., and R. S. Baric. "Coronavirus genome structure and replication." *Coronavirus replication and reverse genetics*. Springer, Berlin, Heidelberg, 2005. 1-30.

4. Arya R, Das A, Prashar V, Kumar M. Potential inhibitors against papain-like protease of novel coronavirus (SARS-CoV-2) from FDA approved drugs.

5. Ton AT, Gentile F, Hsing M, Ban F, Cherkasov A. Rapid Identification of Potential Inhibitors of SARS-CoV-2 Main Protease by Deep Docking of 1.3 Billion Compounds. Molecular Informatics. 2020 Feb 19

6. Chu CM, Cheng VC, Hung IF, Wong MM, Chan KH, Chan KS, Kao RY, Poon LL, Wong CL, Guan Y, Peiris JS. Role of lopinavir/ritonavir in the treatment of SARS: initial virological and clinical findings. Thorax. 2004 Mar 1;59(3):252-6.

7. Liu, X., Zhang, B., Jin, Z., Yang, H., Rao, Z The crystal structure of COVID-19 main protease in complex with an inhibitor N3,DOI**:** 10.2210/pdb6lu7/pdb

8. Bacha U, Barrila J, Velazquez-Campoy A, Leavitt SA, Freire E. Identification of novel inhibitors of the SARS coronavirus main protease 3CLpro. Biochemistry. 2004 May 4; 43(17):4906-12.

9. Ziebuhr J, Snijder EJ, Gorbalenya AE. Virus-encoded proteinases and proteolytic processing in the Nidovirales. Journal of General Virology. 2000 Apr 1;81(4):853-79.

10. Anand K, Palm GJ, Mesters JR, Siddell SG, Ziebuhr J, Hilgenfeld R. Structure of coronavirus main proteinase reveals combination of a chymotrypsin fold with an extra α-helical domain. The EMBO Journal. 2002 Jul 1;21(13):3213-24.

11. Samdani, A. and Vetrivel, U. (2018). POAP: A GNU parallel based multithreaded pipeline of open babel and AutoDock suite for boosted high throughput virtual screening. Computational Biology and Chemistry, 74, pp.39-48. DOI:10.1016/j.compbiolchem.2018.02.012

12. Wallace AC, Laskowski RA, Thornton JM. LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. Protein engineering, design and selection. 1995 Feb 1;8(2):127-34.

13.Schrödinger LL. The PyMOL Molecular Graphics System, Version 1.8. November 2015. URL https://pymol. org/2.

14. Cuker A, Arepally GM, Chong BH, et al. American Society of Hematology 2018 guidelines for management of venous thromboembolism: heparin-induced thrombocytopenia. *Blood Adv*. 2018;2(22):3360–3392.

15.Thabit AK, Fatani DF, Bamakhrama MS, Barnawi OA, Basudan LO, Alhejaili SF. Antibiotic penetration into bone and joints: An updated review. *Int J Infect Dis*. 2019;81:128–136.

16. Severin PN, Awad S, Shields B, Hoffman J, Bonney W, Cortez E, Ganesan R, Patel A, Barnes S, Barnes S, Al-Anani S. The pediatric cardiology pharmacopeia: 2013 update. Pediatric cardiology. 2013 Jan 1;34(1):1-29.
17. Occhipinti DJ, Pendland SL, Schoonover LL, Rypins EB, Danziger LH, Rodvold KA. Pharmacokinetics and pharmacodynamics of two multiple-dose piperacillin-tazobactam regimens. Antimicrobial agents and chemotherapy. 1997 Nov 1;41(11):2511-7.
18. Benaboud S, Urien S, Thervet E, Prié D, Legendre C, Souberbielle JC, Hirt D, Friedlander G, Treluyer JM, Courbebaisse M. Determination of optimal cholecalciferol treatment in renal transplant recipients using a population pharmacokinetic approach. European journal of clinical pharmacology. 2013 Mar 1;69(3):499-506.
19. Robin AL, Faulkner R, Curtis M, Patil S, McCarty G, Clifford W, Dahlin DC. Safety and Pharmacokinetics of Travoprost a Potent Prostaglandin F (FP) Receptor Agonist, in Patients With Renal and Hepatic Impairment. Investigative Ophthalmology & Visual Science. 2002 Dec 1;43(13):4106.
20. Leppik IE, Boucher BA, Wilder BJ, Murthy VS, Watridge C, Graves NM, Rangel RJ, Rask CA, Turlapaty P. Pharmacokinetics and safety of a phenytoin prodrug given IV or IM in patients. Neurology. 1990 Mar 1;40(3 Part 1):456.
21. Hsu A, Granneman GR, Cao G, Carothers L, Japour A, El-Shourbagy T, Dennis S, Berg J, Erdman K, Leonard JM, Sun E. Pharmacokinetic interaction between ritonavir and indinavir in healthy volunteers. Antimicrobial agents and chemotherapy. 1998 Nov 1;42(11):2784-91.

Table: Computational details of top ranking molecules

Sl No	Name of the drugs	Docking score	H-bond with residues	Hydrophobic Interaction	Pharmacological Use	Dose	Clearance	Adverse Effects
1	Argatroban	-8.7 kcal/mol	HIS163, GLU166, HIS41, GLY143, LEU141	PHE140, HIS172, ASN142, THR26, CYS145, LEU27, HIS164, ARG188, ASP187, MET49, MET165, GLU189	thrombin inhibitor	100 mg/1mL (16)	5.1 L/kg/hr [infusion doses up to 40 mcg/kg/min] (16)	Excessive bleeding (16)
2	Piperacillin	-7.9 kcal/mol	CYS145, HIS41, GLU166	PHE140, LEU141, ASN142, MET165, ASP187, HIS164, TYE54, MET49, GLN189	Beta lactum antibiotic	3-4 g/dose (17)	32 - 41 mL/min (17)	Constipation, Headache, Nausea (17)
3	Cholecalciferol	-7.5 kcal/mol	HIS163, GLY143, HIS41, GLN189, GLU166	CYS145, HIS164, ASP187,	Nutraceutical	400 intl units/mL (18)	2.5 L/day (18)	weakness, fatigue, somnolence, headache (18)
4	Travoprost	-8.0 kcal/mol	THR190, GLN192, GLU166, HIS164, THR26, HIS41, GLN189	CYS145, ASN142, ARG188	For the treatment of glaucoma	0.04 mg/1mL (19)	< 10 pg/mL/min (19)	A topical overdose of travoprost may be flushed from the eye(s) with lukewarm water
5	Phenytoin	-7.6 kcal/mol	GLU166, GLY143, SER144, LEU141, SER46	THR24, THR25, THR45, MET49, MET165,GLN189, PHE140, CYS145, THR28	anticonvulsant	50-300mg/ml (20)	less than 10 mg/L/min (20)	mental confusion, nervousness
6	Indinavir	-7.5 kcal/mol	HIS41	LEU141, GLY143, THR26, CYS145, HIS164, GLN189, ARG188, MET165, ASN142, MET49	HIV protease inhibitor	200 mg/ml (21)	300 to 400 ml/min (21)	myocardial infarction, angina pectoris

Figures:

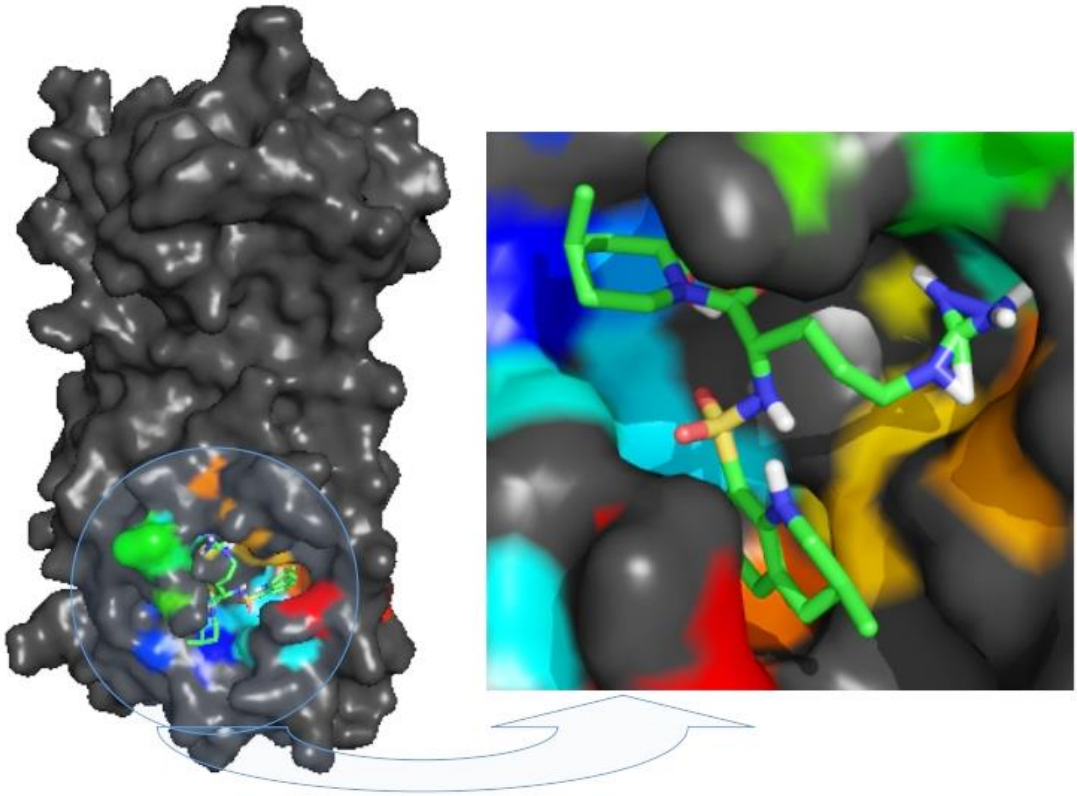
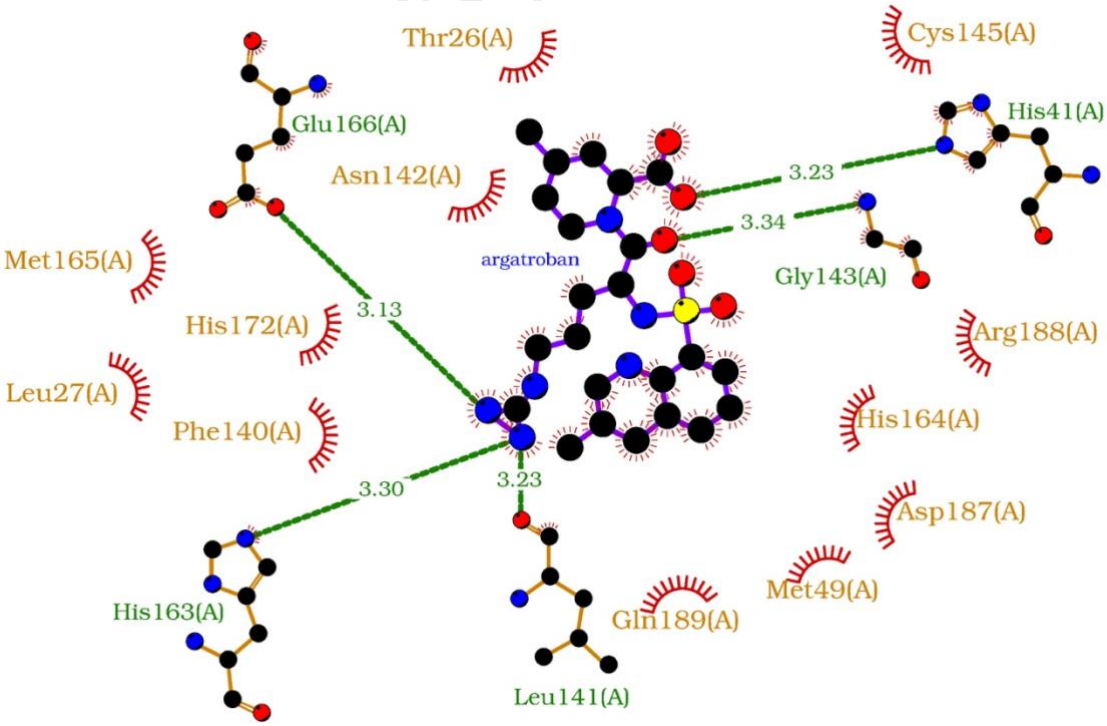
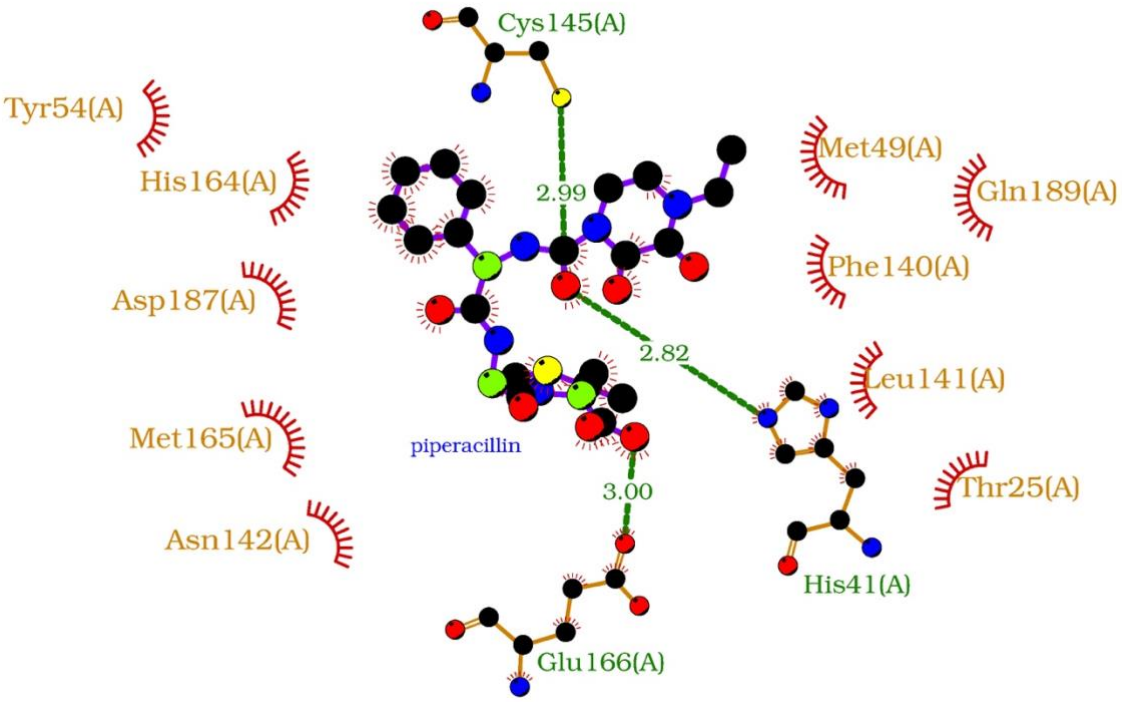


Figure 2: Binding interaction of argatroban with main protease enzyme (M_{pro}) (PDB-6LU7) of



novel SARS CoV2



complex

Figure 3b: Piperacillin-main protease enzyme of SARS CoV2 complex

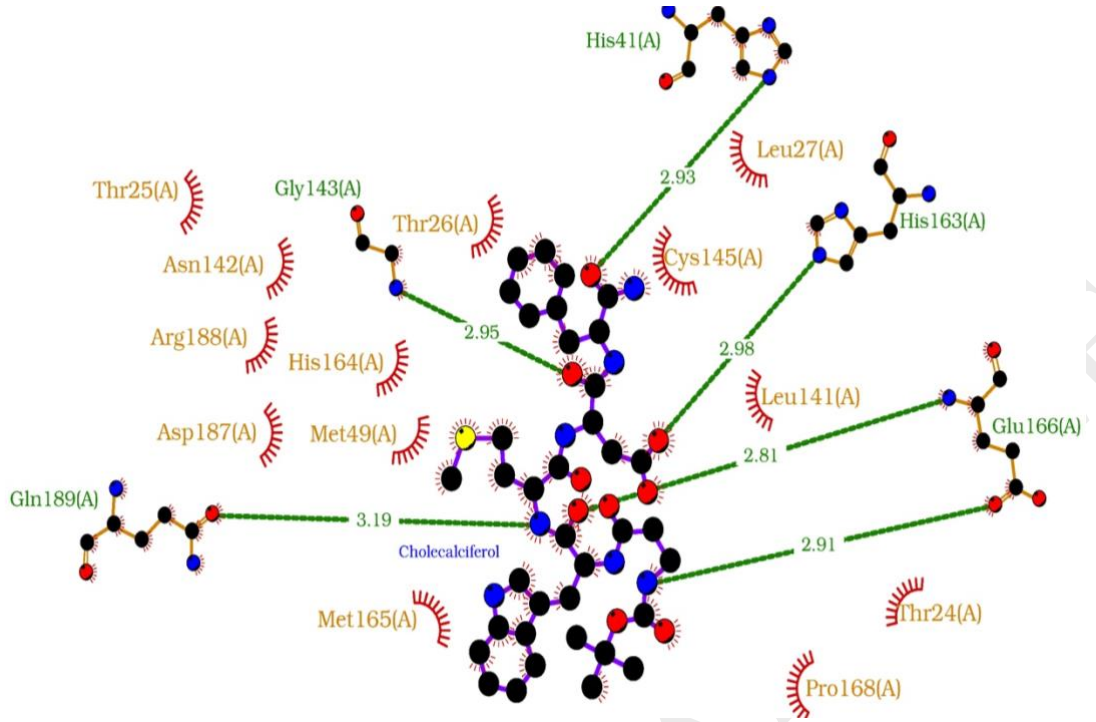
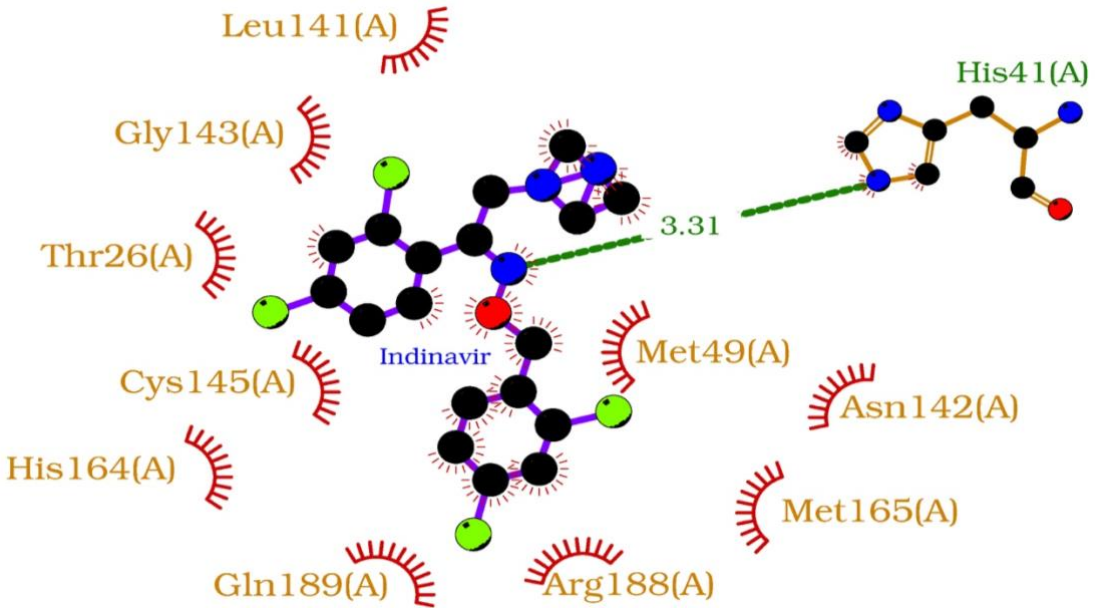


Figure 3c: Cholecalciferol-main protease enzyme of SARS CoV2 complex



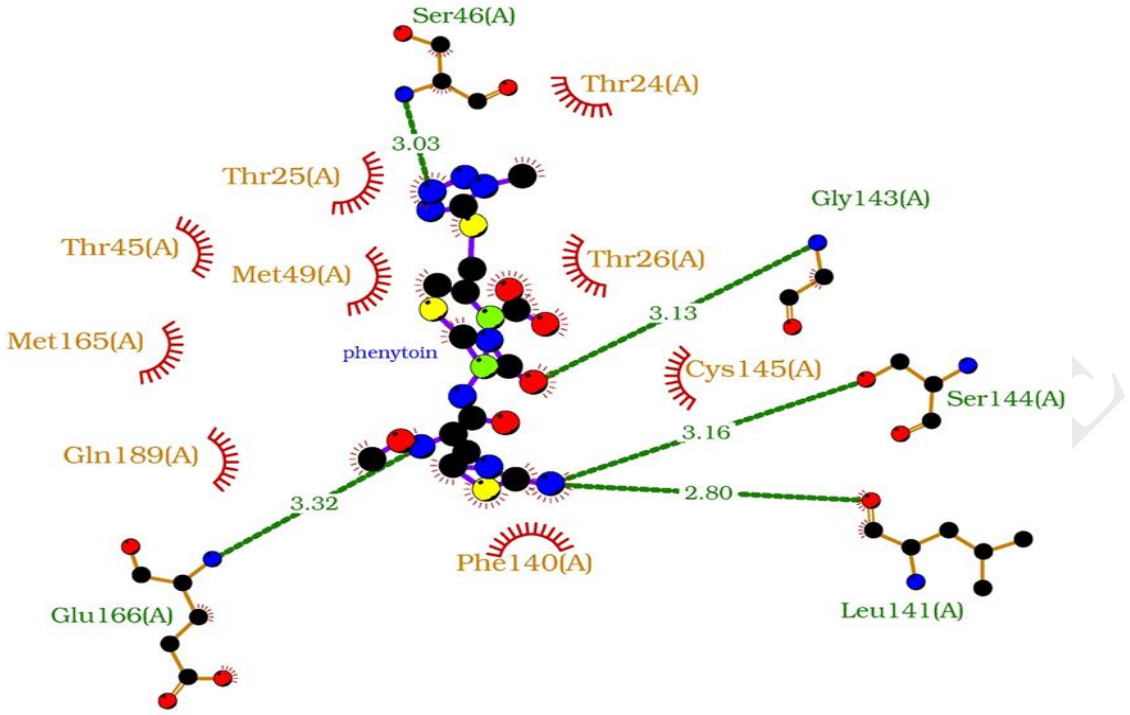


Figure 3e: Phenytoin-main protease enzyme of SARS CoV2 complex

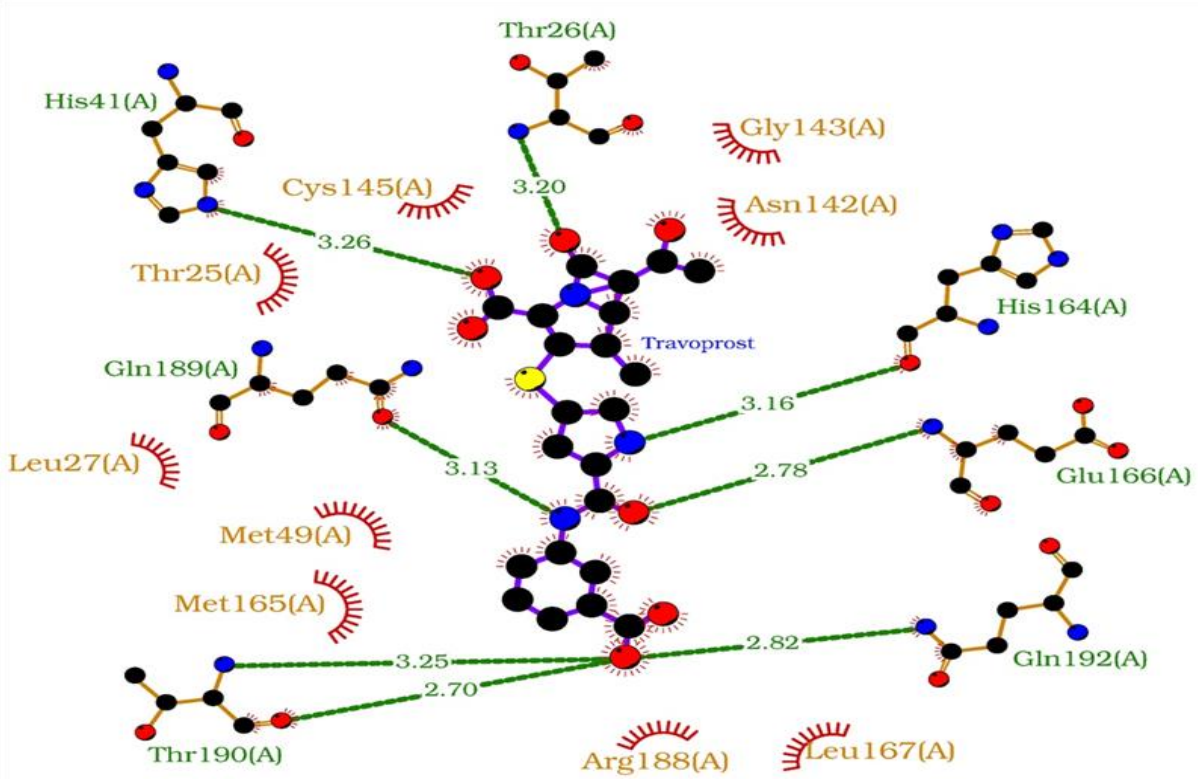


Figure3f: Travoprost-main protease enzyme of SARS CoV2 complex

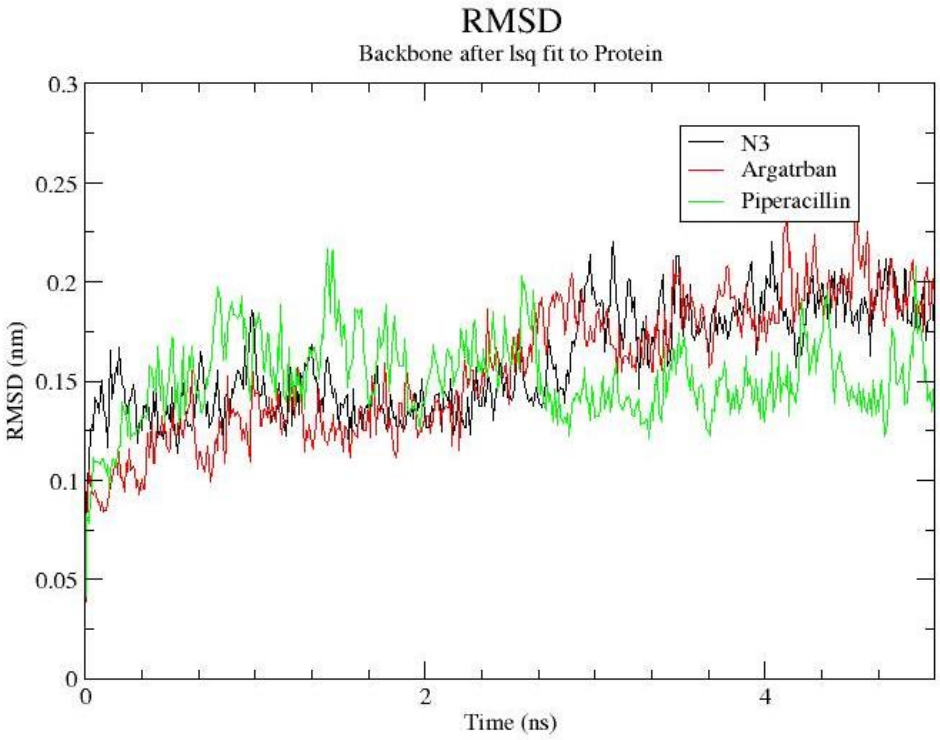


Figure 4a: RMSD backbone

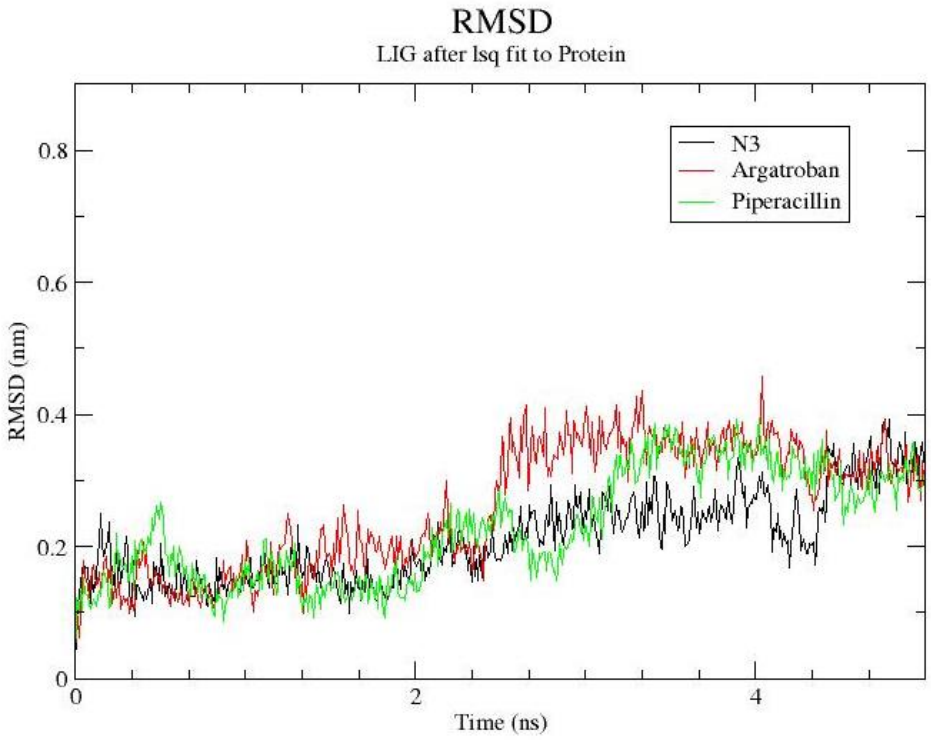


Figure4b: RMSD Protein-ligand

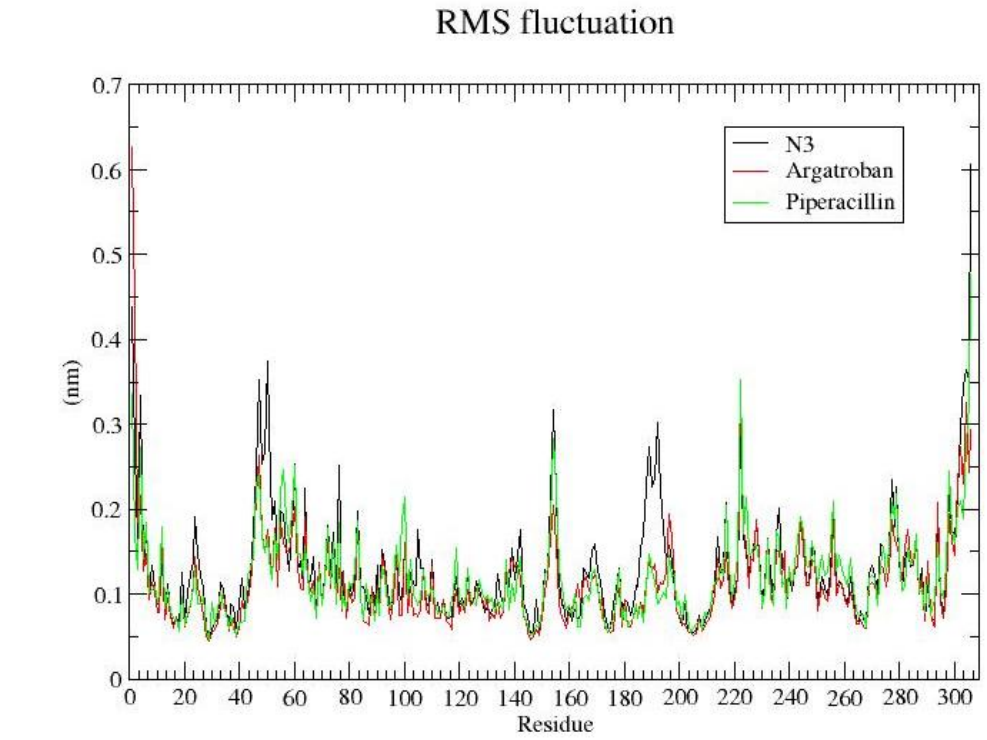


Figure4c: RMSF Protein-ligand complex

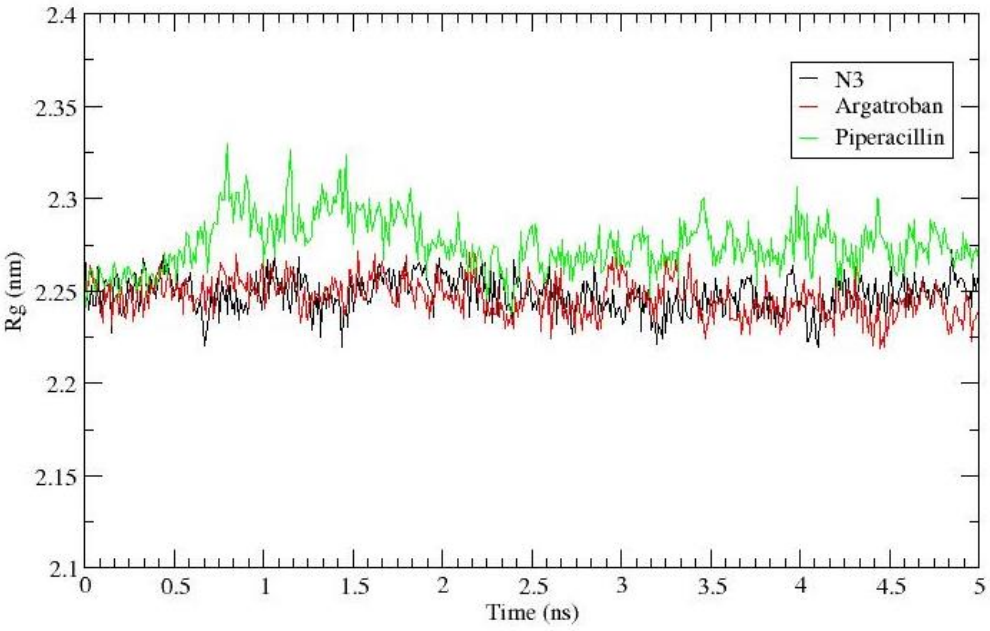


Figure 4d: Radious of gyration (total and around axis)

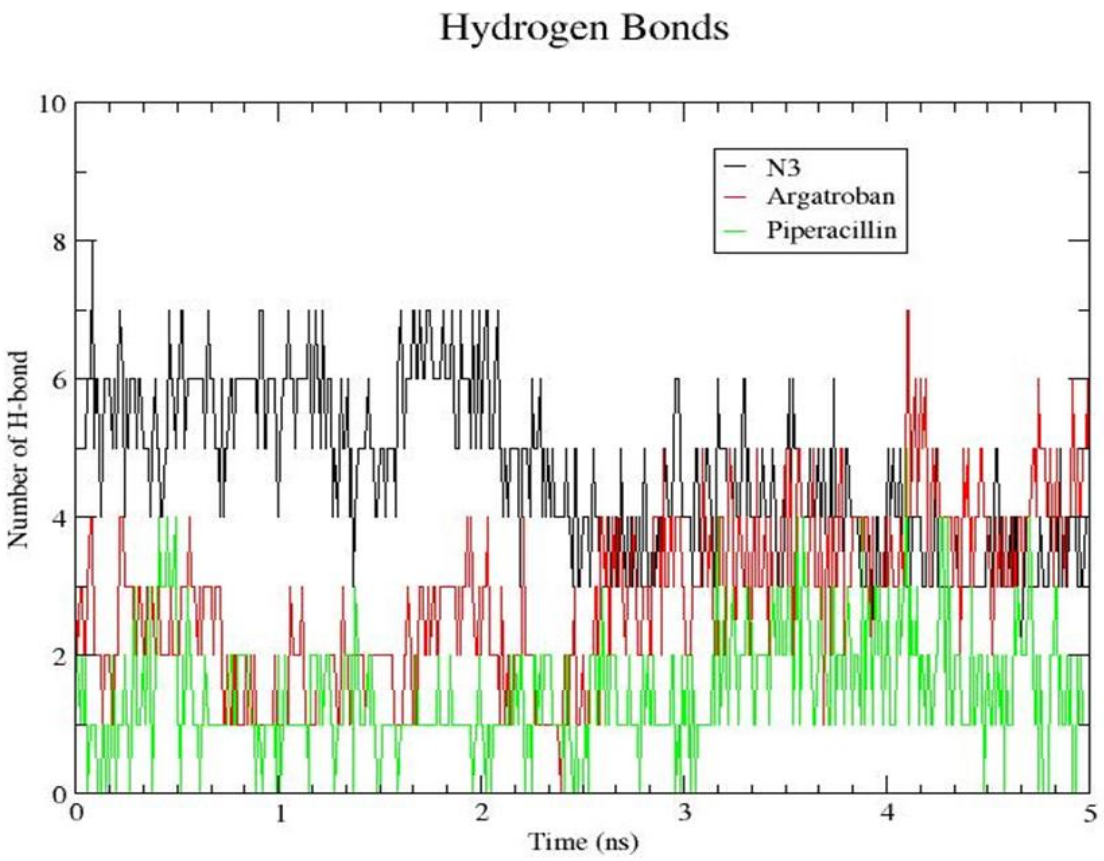


Figure 4e: Hydrogen bond formation of ligands with main protease protein complex