Comparison of random and site directed mutation effects on the efficacy between lead SARS-CoV2 anti-protease drugs Indinavir and Hydroxychloroquine

CURRENT STATUS: POSTED

Jithin S. Sunny
SRM-IST

jithin.sunny@srmuniv.edu.in

Corresponding Author

Saranya Balachandran
Institute of Human Genetics, Christian-Albrechts-University

Solaipriya S.
SRM-IST

Lilly M. Saleena
SRM-IST

saleenam@srmist.edu.in

Corresponding Author

DOI:
10.21203/rs.3.rs-22082/v1

SUBJECT AREAS
Bioinformatics

KEYWORDS
random substitution mutations, Indinavir, Hydroxychloroquine, pipeline, continuous distribution, mean, variance, t-test
Abstract
The diversification of virus can be attributed to random mutations leading to the development of drug resistance. The variations can be inherited from one generation to the other rendering the drug ineffective. However, a pharmacologically induced selection pressure can be countered by introducing drugs better adapted to work under rapid mutations. In this study we try to explore the effect of site directed and random substitution mutations simulated in the ligand binding region of SARS-CoV2 protease. Amongst six currently studied anti-protease drugs for COVID-19, Indinavir and Hydroxychloroquine were chosen for the study based on their high binding affinity scores, -6.81, and -4.81 respectively. The effect of mutations in protein-ligand binding was analysed in two steps. Initially, analysis of over 90 homologous protease and 100 SARS-CoV-2 orf1ab regions revealed un-conserved residues in the ligand binding sites. Gly170 and Thr190 were identified and interchanged with polar residues such as ARG, ASN and non-polar residues such as ALA, ILE. The resulting mutants were modelled, minimized and docked with Indinavir and Hydroxychloroquine. A higher binding affinity was observed for Indinavir; however, less variance in the binding affinity was observed for the latter. These results were consistent for random mutations as well. A Bio.seqIO based pipeline was build to simulate changes in the ligand binding site. Under the assumption that the ligand binding region has an equal probability of mutation over a given range for continuous distribution, 200 cycles of mutation was carried out in the nucleotide region corresponding to the ligand binding site. A paired t-test revealed a significant difference between the binding affinity of these mutant Indinavir and Hydroxychloroquine-protease complexes. Further, mean and variance was found to be higher for Indinavir-protease complex but Hydroxychloroquine displayed lesser variance pointing at a constant binding capability towards the mutant. Our study highlights the role of Hydroxychloroquine as a drug that can complement an evolving SARS-CoV2 main protease.

Introduction
The first coronavirus infection has been described as early as 1949 and has been tested in all species of animals including humans [1]. Of the many known coronaviruses, only 7 are known to cause disease in humans, out of which 3 cause severe respiratory infections [2]. SARS-CoV2 is the type of
coronavirus that has been identified as the cause of COVID-19 disease that began in Wuhan, China. The COVID-2019 epidemic has probably originated from bat which later underwent mutation in the spike glycoprotein leading to its transmission into humans [3]. Rapid investigation is required to study the impact of mutation and its influence on the potential drug targets.

Retroviral genomes evolve rapidly which helps it to evade selective pressure from the host immune system which otherwise can decrease their survival chance [4, 5]. Viral recombination rate, rate of replication, size of viral population, selective forces and viral mutation rate are the main players behind genetic variation [6]. The site of mutations is crucial in understanding the resulting viral viability. Viral replication is mediated by proteins encoded in open reading frame 1 (ORF1) which comprises nearly 2/3rds of the viral genome. ORF1a and ORF1ab are two polyproteins translated from this region. These are non-structural polyproteins processed through the main protease and papain-like protease. This crucial role of viral protease makes it a very promising target for drug discovery [7, 8]. In the present study, we focus on the SARS-CoV2 protease

Variation in the viral SARS-CoV2 protease can affect the drug binding capability rendering the host susceptible to infection in spite of medication [9]. Studies are currently being carried out to identify possible drugs capable of targeting it. Combinatorial drugs used as an anti-influenza and anti-HIV agent, Lopinavir and ritonavir have been suggested, with successful results [10]. Another study suggests protease inhibitor Indinavir which is currently used for HIV infections [11]. Non-specific Chloroquine analogs have been tested against a broad range of emerging virus. The use of Chloroquine has found no evidence of clinical benefit. Hydroxychloroquine along with azithromycin too does not show any significant change in antiviral effects [12, 13, 14]. In the present study, the effect of mutations on the drug binding capability of Indinavir and Hydroxychloroquine to the SARS-CoV2 protease has been studied. We wish to use both binding affinity and its variability to propose the flexibility of drug effect upon both site directed and random mutations.

Results

Protease Structural analysis

SARS-CoV2 protease consists of two protomers which associates to form a dimer. Each protomer is
composed of three domains. Domains I and II are six-stranded anti-parallel beta barrels and the substrate-binding site is located in a cleft between these two domains. The third domain is connected to domain I & II through a long loop. Domain III consists of 5 Alpha-helix (Figure 1). The sequence analysis showed 41.5% hydrophobic residues, 8.5% acidic, 9.48% basic and 40.52% of the residues to be neutral.

**Binding studies with retroviral protease drugs**

The minimized structure of SARS-CoV2 protease (PE = -1.55498e+06Kj/mol) predicted forty five residues to be actively involved in ligand binding (Figure 2). These residues were chosen for further analysis. The docking of six compounds Amprenavir, Atazanavir, Darunavir, Fosamprenavir, Indinavir and Hydroxychloroquine with the retroviral protease resulted in binding energies -4.74, -3.27, -4.17, -3.26, -6.81, and -4.81 respectively. Besides binding affinity, the comparative values for total energy, H.bond energy, and other parameters have been tabulated (Supplementary file 1). Indinavir and Hydroxychloroquine showed the maximum affinity towards the protease (Figure 2). Indinavir complex had the higher binding affinity with HIS163 involved in H-bonds and LEU 141, GLY 170, GLU 166 in hydrophobic interactions. THR190 and GLU166 showed H-bonds with Hydroxychloroquine and only one residue HIS41 were involved in hydrophobic interactions (Supplementary file 2).

**Site-directed mutation**

To check the effect of mutation in the un-conserved region, ConSurf server based on multiple sequence alignment of 98 coronavirus and its related polyprotein sequences predicted 20 un-conserved residues. However, the ligand binding site largely remained conserved. Amongst the 21 residues predicted with the ligand binding role, Gly170 and Thr190 were directly involved in interactions with Indinavir and Hydroxychloroquine respectively. The substitution of GLY and THR with polar and non-polar residues showed variation in the resulting modelled structure with respect to RMSD and the resulting binding affinity. The observed variation was in the range 0.375 Å to 0.435 Å and -5.98 to -8.04 for Indinavir complex, while for Hydroxychloroquine it was 0.376 Å to 0.404 Å and
-5.86 to -6.90 (Supplementary file 3). A very low correlation ($r = -0.04, 0.2$) stated the structural effects due to the mutations are negligible on the binding affinity. Paired t-test ($\alpha=0.05$), revealed a significant difference ($p=0.0016$) between the binding energies of Indinavir and Hydroxychloroquine.

**Random mutation effect**

General features of the SARS-CoV-2 complete genome (Gbk Id: MN908947.3) were retrieved and corresponding protein information was tabulated (Table 1). 12 translated gene products have been retrieved from the viral genome. 100 genomic regions belonging to the protease extracted from the orf1ab region were aligned and the changes were observed in the ligand binding regions. The ligand binding site was conserved throughout the 100 samples. Next, the SARS-CoV-2 protease gene ran with the uniform substitution algorithm, generated random mutations which produced 200 mutant proteases. RMSD deviations of the translated and modelled amino acids along with binding energy of the Indinavir and Hydroxychloroquine complex were tabulated (Supplementary file 4). Similar to site directed mutations, these results too showed a lower variance in the binding affinity of Hydroxychloroquine complex in spite of the amino acid changes. The t-test revealed a significant difference in the binding affinity of the two drugs. The difference in the means of binding affinity was much less compared to the ones observed in site directed mutation. On the other hand, the difference in variance increased. Hydroxychloroquine showed a variance of 0.45 compared to 1.53 for Indinavir-protease complex (Table 2).

**Discussion**

Viral proteases catalyses the cleavage of peptide bonds specific to viral polyprotein precursors [15]. The role of viral protease in mediating assembly and disassembly involves converting the polyproteins into capsids which disassembles upon its entry into a newly infected cell [16]. Without doubt, proteases are being considered widely as a potential target for the current novel Coronavirus infection. A general mechanism involving protease function is to recognize substrates through a conserved shape. A variant of protease thus might not bind to inhibitors efficiently even though it may still process the substrate, turning resistant [17]. To verify the property of inhibitor effect under
mutations, a set of six major protease inhibitors were selected. These six compounds Amprenavir, Atazanavir, Darunavir, Fosamprenavir, Indinavir and Hydroxychloroquine were docked with retroviral protease 6LU7 with binding energy -4.74, -3.27, -4.17, -3.26, -6.81, and -4.81 respectively. Compared to H-bonds, residues involved in hydrophobic interactions were relatively higher. The hydrophobic interactions are important contributors to ligand-receptor binding affinities. The tendency of nonpolar molecular surfaces to interact with other nonpolar molecular surfaces contributes significantly to the ligand-receptor binding affinities [18]. A higher binding affinity was assumed to be mainly responsible due to these hydrophobic interactions in Indinavir complex.

These hydrophobic residues and other adjoining ones were analysed further. The sequence alignment of homologous proteins from ORF1a polyprotein belonging to bat, human, Infectious bronchitis virus, Bottlenose dolphin coronavirus and others showed very few residues in the ligand binding sites to be variable amongst the other species. Only two mutations Gly170 and Thr190 were observed amongst the residues involved in the ligand binding. These results indicated a high degree of conservation amongst the viral protease. Nevertheless, the two residues found to be variable were further used to simulate changes in binding affinity upon substitution. Both polar and non-polar residues were substituted in their positions. The effect of these site directed mutation were quantified as structural (RMSD) and biophysical (Binding affinity). Since no direct correlation was established between the two, the binding affinity was further analysed. The site directed mutations pointed at a significant variation between the binding affinity of Indinavir and Hydroxychloroquine-protease complex. Even though Indinavir had a higher mean binding affinity, a comparatively less variance was observed with Hydroxychloroquine. This was further validated with 200 protease mutants. The difference between the variance observed with few mutations further increased. Interestingly the mean of binding affinity increased for Hydroxychloroquine.

Based on our observations, Hydroxychloroquine is observed to have less variance in the binding affinity as compared to Indinavir under similar set of mutations. The mean binding affinity of Indinavir is only 0.7 higher than Hydroxychloroquine. Considering a similar binding affinity profile and a comparatively less variance, Hydroxychloroquine can be assumed to be more stable under protease
mutation. Previous studies on using Hydroxychloroquine along with Azithromycin have provided no strong evidence in association with protease inhibition [19]. Few studies have even used Zn2+ for inhibiting coronavirus which alters RNA polymerase activity [20]. This potential inhibitor along with Hydroxychloroquine is being tried in few countries although a scientific reference is yet to be produced. Besides Hydroxychloroquine, Indinavir an anti-protease drug used for HIV infections have been tried too. This drug has shown high degree of specificity to HIV protease [21]. Through this study we have compared the binding affinity of two prominent drugs currently being tried worldwide. Indinavir and Hydroxychloroquine bind to SARS-CoV-2 protease with different binding affinities but under mutation effects, the observation has placed the later to be more stable. Based on the produced data, we propose Hydroxychloroquine to be a stable drug under rapid mutation conditions. Although, a combination of Indinavir and Hydroxychloroquine is worth being looked at

**Materials And Methods**

**Protease Structural analysis**

The X-ray crystal structure with PDB id: 6LU7 of SARS-CoV-2 protease was extracted from RCSB PDB database [22]. For structure analysis PyMOL [23] visualising tool was used.

**Docking studies**

Prior to further analysis the protein structure was energy minimised using GROMOS96 54a7 force field from the Gromacs 5.1 package [24]. To identify the ligand binding sites, residues showing binding activity in the ligand bounded PDB structure was identified. Along with this, online tool COACH [25-26] was also used to get the highest ranked ligand binding pockets. The top ranked site was then chosen for analysing the binding efficiency of six major ligands. These six compounds Amprenavir, Atazanavir, Darunavir, Fosamprenavir, Indinavir and Hydroxychloroquine were chosen based on literature survey of recent studies performed for finding suitable candidates as protease inhibitors (Supplementary figure 1). Molecular docking was performed using Autodock v.1.5.6 [27]. Kollman charges were used to process the protein. Maximum torsion of 6 was used for the ligands. The grid box was set based on ligand binding residues ranging from THR25 to GLN192. The docking was
initialised using Genetic Algorithm with 100 runs. Besides binding affinity, the comparative values for total energy, H.bond energy, and other were also computed. The residues taking part in hydrophobic interactions and H-bonds were identified using PoseView online tool [28].

**Conserved versus variable region in the ligand binding region**

To check the effect of mutation in the un-conserved region, the ConSurf [29] server was first used to find out variable regions based on multiple sequence alignment of 98 homologous coronavirus and other similar polyprotein sequences. Genes for protease belonging to orf1ab region was downloaded from 100 complete genomes of SARS-CoV-2 using NCBI [30]. These were aligned and analysed using MEGA [31].

**Site directed In-silico mutation**

Based on the homologous protein analysis, site directed mutagenesis were carried on the variable residues. The GLY170 and THR190 were changed with polar residues such as ARG, ASN and non-polar residues such as ALA, ILE, etc. In total 13 residues were used to observe the effect of mutation. The mutant proteins were modelled and its RMSD deviation was evaluated with respect to the wild type. Indinavir and Hydroxychloroquine were docked against the generated mutant proteases based on the above protocol.

**Random mutation effect**

Nucleotides that code for the ligand binding sites of the protease were subjected to random changes. A custom pipeline was written for this simulation. The start and end position of the mutational site along with the mutation rate was defined based on evolutionary analysis of similar sequences by aligning the protease and computing the variable region based on the above analysis (section). The mutational sites are designated based on uniform manner and the type of nucleotide in random manner. Bio.seqIO interface of Biopython was used to build the pipeline. The following inputs were used, fasta file of the nucleotide sequence of the organism, start (p1) and the end (pn) nucleotide
position of the active site and the mutational rate (μ). The selected mutational sites are defined as the (pn - p1)/μ+σ and (pn - p1)/μ-σ, where σ is the standard deviation of the uniform distribution. At each of these mutational sites, A/T/G/C are substituted to yield the mutated variant of the input protease. The simulated sequence was then translated to the protein sequence to study the changes in protein. 200 models were then generated then subjected to energy minimization using Steepest Descent. Further, Indinavir and Hydroxychloroquine were docked with these 200 structures. Pearson correlation was computed to analyse the dependence of RMSD changes on the binding affinity for Indinavir and Hydroxychloroquine complex. Mean, variance, and t-test was performed to evaluate significant difference between the binding affinity of the two ligand.

Acknowledgement
Authors thank SRM Supercomputer Centre, SRM Institute of Science and Technology for providing the computational facility.

Tables

| Gene               | Gene ID  | Length |
|--------------------|----------|--------|
| orf1ab polyprotein | 43740578 | 7096   |
| orf1a polyprotein  | 43740578 | 4405   |
| Surface glycoprotein | 43740568 | 1273   |
| ORF3a protein      | 43740569 | 275    |
| Envelope protein   | 43740570 | 75     |
| Membrane glycoprotein | 43740571 | 222    |
| ORF6 protein       | 43740572 | 61     |
| ORF7a protein      | 43740573 | 121    |
| ORF7b              | 43740574 | 43     |
| ORF8 protein       | 43740577 | 121    |
| Nucleocapsid phosphoprotein | 43740575 | 419    |
| ORF10 protein      | 43740576 | 38     |

Table 2. Statistical inference from random
| Pearson Correlation | RMSD | binding energy Indinavir |
|---------------------|------|--------------------------|
| RMSD                | 1    | -0.120781141             |
| binding energy Indinavir | 1    |                           |

| Pearson Correlation | RMSD | binding energy Hydroxychloroquine |
|---------------------|------|----------------------------------|
| RMSD                | 1    | -0.09                            |
| binding energy Indinavir | 1    |                                  |

|               | Indinavir-Protease complex | Hydroxychloroquine-protease complex |
|---------------|----------------------------|-----------------------------------|
| Mean          | -6.83085                   | -6.10905                          |
| variance      | 1.53418972                 | 0.448915173                       |
| ttest         | 3.45E-12                   |                                   |

References

1. Weiss, S.R. and Navas-Martin, S., 2005. Coronavirus pathogenesis and the emerging pathogen severe acute respiratory syndrome *Microbiol. Mol. Biol. Rev.*, 69(4), pp.635-664.

2. McIntosh, K., 1974. Coronaviruses: a comparative review. In *Current Topics in Microbiology and Immunology/Ergebnisse der Mikrobiologie und Immunitätsforschung* (pp. 85-129). Springer, Berlin,

3. Angeletti, S., Benvenuto, D., Bianchi, M., Giovanetti, M., Pascarella, S. and Ciccozzi, M., 2020. COVID-2019: The role of the nsp2 and nsp3 in its pathogenesis. *Journal of*
4. Woo, H.J. and Reifman, J., 2014. Quantitative modeling of virus evolutionary dynamics and adaptation in serial passages using empirically inferred fitness landscapes. *Journal of virology, 88*(2), 1039-1050.

5. Eigen, M., 1971. Selforganization of matter and the evolution of biological macromolecules. *Naturwissenschaften, 58*(10), 465-523.

6. Liang, P.H., 2006. Characterization and inhibition of SARS-coronavirus main protease. *Current topics in medicinal chemistry, 6*(4), 361-376.

7. Deming, D.J., Graham, R.L., Denison, M.R. and Baric, R.S., 2007. Processing of open reading frame 1a replicase proteins nsp7 to nsp10 in murine hepatitis virus strain A59 replication. *Journal of virology, 81*(19), 10280-10291.

8. Svarovskaia, E.S., Cheslock, S.R., Zhang, W.H., Hu, W.S. and Pathak, V.K., 2003. Retroviral mutation rates and reverse transcriptase fidelity. *Front Biosci, 8*(4), pp.d117-134.

9. Emamzadeh-Fard, S., Esmaeeli, S., Arefi, K., Moradbeigi, M., Heidari, B., E Fard, S., Paydary, K. and SeyedAlinaghi, S., 2013. Mechanisms of anti-retroviral drug resistance: implications for novel drug discovery and development. *Infectious Disorders-Drug Targets (Formerly Current Drug Targets-Infectious Disorders), 13*(5), 330-336.

10. Chang, Y.C., Tung, Y.A., Lee, K.H., Chen, T.F., Hsiao, Y.C., Chang, H.C., Hsieh, T.T., Su, C.H., Wang, S.S., Yu, J.Y. and Shih, S.S., 2020. Potential therapeutic agents for COVID-19 based on the analysis of protease and RNA polymerase

11. Contini, A., 2020. Virtual screening of an FDA approved drugs database on two COVID-19 coronavirus

12. Dayer, M.R., 2020. Old Drugs for Newly Emerging Viral Disease, COVID-19:
Bioinformatic Prospective. *arXiv preprint arXiv:2003.04524.*

13. Molina, J.M., Delaugerre, C., Goff, J.L., Mela-Lima, B., Ponscarme, D., Goldwirt, L. and de Castro, N., 2020. No Evidence of Rapid Antiviral Clearance or Clinical Benefit with the Combination of Hydroxychloroquine and Azithromycin in Patients with Severe COVID-19 Infection. *Médecine et Maladies Infectieuses.*

14. Molina, J.M., Delaugerre, C., Goff, J.L., Mela-Lima, B., Ponscarme, D., Goldwirt, L. and de Castro, N., 2020. No Evidence of Rapid Antiviral Clearance or Clinical Benefit with the Combination of Hydroxychloroquine and Azithromycin in Patients with Severe COVID-19 Infection. *Médecine et Maladies Infectieuses.*

15. Steinkühler (2008) Viral Proteases. In: Offermanns S., Rosenthal W. (eds) Encyclopedia of Molecular Pharmacology. Springer, Berlin, Heidelberg

16. Babé, L.M. and Craik, C.S., 1997. Viral proteases: evolution of diverse structural motifs to optimize function. *Cell,* 91(4), 427-430.

17. Yilmaz, N.K., Swanstrom, R. and Schiffer, C.A., 2016. Improving viral protease inhibitors to counter drug resistance. *Trends in microbiology,* 24(7), 547-557.

18. Patil, R., Das, S., Stanley, A., Yadav, L., Sudhakar, A. and Varma, A.K., 2010. Optimized hydrophobic interactions and hydrogen bonding at the target-ligand interface leads the pathways of drug-designing. *PloS one,* 5(8).

19. Molina, J.M., Delaugerre, C., Goff, J.L., Mela-Lima, B., Ponscarme, D., Goldwirt, L. and de Castro, N., 2020. No Evidence of Rapid Antiviral Clearance or Clinical Benefit with the Combination of Hydroxychloroquine and Azithromycin in Patients with Severe COVID-19 Infection. *Médecine et Maladies Infectieuses.*

20. Te Velthuis, A.J., van den Worm, S.H., Sims, A.C., Baric, R.S., Snijder, E.J. and van Hemert, M.J., 2010. Zn2+ inhibits coronavirus and arterivirus RNA polymerase activity in vitro and zinc ionophores block the replication of these viruses in cell
culture. *PLoS pathogens*, 6(11).

21. Chang, Y.C., Tung, Y.A., Lee, K.H., Chen, T.F., Hsiao, Y.C., Chang, H.C., Hsieh, T.T., Su, C.H., Wang, S.S., Yu, J.Y. and Shih, S.S., 2020. Potential therapeutic agents for COVID-19 based on the analysis of protease and RNA polymerase

22. The Protein Data Bank H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, Weissig, I.N. Shindyalov, P.E. Bourne (2000) *Nucleic Acids Research*, 28: 235- 242. doi:10.1093/nar/28.1.235

23. The PyMOL Molecular Graphics System, Version 2.0 Schrödinger,

24. Van Der Spoel, D., Lindahl, E., Hess, B., Groenhof, G., Mark, A.E. and Berendsen, H.J., 2005. GROMACS: fast, flexible, and free. *Journal of computational chemistry*, 26(16), 1701-1718.

25. Jianyi Yang, Ambrish Roy, and Yang Zhang. *Protein-ligand binding site recognition using complementary binding-specific substructure comparison and sequence profile alignment*, Bioinformatics, 29:2588-2595 (2013).

26. Jianyi Yang, Ambrish Roy, and Yang Zhang. *BioLiP: a semi-manually curated database for biologically relevant ligand-protein interactions*, Nucleic Acids Research, 41: D1096-D1103 (2013)

27. 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

28. Stierand, K. and Rarey, M., 2010. PoseView--molecular interaction patterns at a glance. *Journal of cheminformatics*, 2(1), 1-1.

29. Ashkenazy, H., Abadi, S., Martz, E., Chay, O., Mayrose, I., Pupko, T. and Ben-Tal, N., 2016. ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic acids research*, 44(W1),

13
pp.W344-W350.

30. National Center for Biotechnology Information (NCBI)[Internet]. Bethesda (MD): National Library of Medicine (US), National Center for Biotechnology Information; [1988] – [cited 2017 Apr 06]. Available from: https://ncbi.nlm.nih.gov/

31. Kumar, S., Nei, M., Dudley, J. and Tamura, K., 2008. MEGA: a biologist-centric software for evolutionary analysis of DNA and protein sequences. *Briefings in bioinformatics*, 9(4), 299-306.

Figures

![Monomer of SARS-CoV-2 protease protease and the predicted ligand binding sites.](image)
The active site depicted in blue, lies in the intersection of two domains.
Figure 3

Indinavir (blue) and Hydroxychloroquine (yellow) within the binding pocket of SARS-CoV-2 protease

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

Supplementary.xlsx
SFigure.PNG
