

**A Comprehensive investigation of protein structural parameters
& dynamic features of therapeutic targets from SARS-CoV-2 &
Mycobacterium tuberculosis: Implications towards drug-design**

Thesis submitted for the degree of

Doctor of Philosophy

By

Surabhi Lata

(18LBPH05)



**Department of Biochemistry
School of Life Sciences
University of Hyderabad
(P.O.) Central University, Gachibowli,
Hyderabad-500 046, Telangana, India**

November, 2022



University Of Hyderabad
Hyderabad- 500046, India

DECLARATION

I Surabhi Lata hereby declare that this thesis entitled “A Comprehensive investigation of protein structural parameters & dynamic features of therapeutic targets from *SARS-CoV-2* & *Mycobacterium tuberculosis*: Implications towards drug design”. Submitted by me under the guidance and supervision of Dr.Mohd Akif, is a bonafide research work. I also declare that it has not been submitted previously in part or in full to this University or any other University or Institution for the award of any degree or diploma.

Date: 14/11/2022

Name: Surabhi Lata

Surabhi Lata
Signature of the Student:

Regd. No. 18LBPH05

Mohd Akif
Signature of Supervisor
MOHD. AKIF Ph.D.

Assistant Professor
DEPARTMENT OF BIOCHEMISTRY
SCHOOL OF LIFE SCIENCES
UNIVERSITY OF HYDERABAD
HYDERABAD-500 046. (INDIA)



University of Hyderabad
Hyderabad -500046, India

Certificate

This is to certify that this thesis entitled "A Comprehensive investigation of protein structural parameters & dynamic features of therapeutic targets from SARS-CoV-2 & *Mycobacterium tuberculosis*: Implications towards drug design." submitted by Surabhi Lata bearing registration number 18LBPH05 in partial fulfilment of the requirement for the award of Doctor of Philosophy in the Department of Biochemistry, School of Life Sciences, is a bonafide work carried out by her under my supervision and guidance. This thesis is free from plagiarism and has not been submitted previously in part or in full to this University or any other University or Institution for the award of any degree or diploma.

Published in the following journal

1. Lata, S., & Akif, M. (2021). Structure-based identification of natural compound inhibitor against *M. tuberculosis* thioredoxin reductase: insight from molecular docking and dynamics simulation. *Journal of biomolecular structure & dynamics*, 39(12), 4480–4489.
2. Lata, S., & Akif, M. (2021). Comparative protein structure network analysis on 3CL^{pro} from SARS-CoV-1 and SARS-CoV-2. *Proteins*, 89(9), 1216–1225.
3. Lata, S., & Akif, M. (2022). Probing structural basis for enhanced binding of SARS-CoV-2 P.1 variant spike protein with the human ACE2 receptor. *Journal of cellular biochemistry*, 123(7), 1207–1221.
4. Kumar, P., Lata, S., Shankar, U. N., & Akif, M. (2021). Immunoinformatics-Based Designing of a Multi-Epitope Chimeric Vaccine from Multi-Domain Outer Surface Antigens of *Leptospira*. *Frontiers in immunology*, 12, 35373. (#equal contribution).
5. Shiraz, M., Lata, S., Kumar, P., Shankar, U. N., & Akif, M. (2021). Immunoinformatics analysis of antigenic epitopes and designing of a multi-epitope peptide vaccine from putative nitro-reductases of *Mycobacterium tuberculosis* DosR. *Infection, genetics and evolution*, 94, 105017.
6. Priya B.V., Rao D. H. S., Gilani R., Lata S., Rai N*, Akif M., & Padhi S. K (2022). Enzyme engineering improves catalytic efficiency and enantioselectivity of hydroxynitrile lyase for promiscuous retro-nitroaldolase activity. *Bioorganic Chemistry*, 120. (*equal contribution).
7. Kolligundla L. P., Gupta S., Lata S., Mulukala S.K.N., Killaka P., Akif M., & Pasupulati A. K. (2022). Identification of Novel GTP Analogs as Potent and Specific Reversible Inhibitors for Transglutaminase 2, *Molecular Simulations*, 1-11.

Manuscript Under preparation

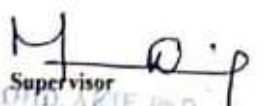
1. Lata S. & Akif M. Computationally designed analogs as the potential disrupters of the RBD-ACE2 complex of SARS-CoV-2 P.1 variant.

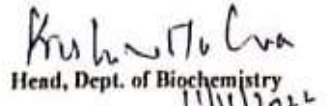
Part of this thesis has been presented at the following conferences

1. Poster presentation in EMBO workshop on "Advances and Challenges in Biomolecular Simulations" 18th - 21st Oct, 2021.
2. Poster presentation in "International Conference on Drug Discovery" 29th Feb -2nd March, 2020, BITS-Hyderabad.
3. Presentation in 44th Indian Biophysical Society Meeting on "Conceptual Advances in Biophysics and its Applications" 30th March-1st April, 2022, TMC ACTREC, Mumbai

Further, the student has passed the following courses towards fulfilment of the coursework requirement for a Ph.D.

	Course Code	Name	Credits	Pass/Fail
1.	S801	Research Methodology/Analytical Techniques	4	Pass
2.	S802	Research Ethics, Biosafety, Data Analysis	4	Pass
3.	S803	Research Proposal & Scientific Writing	4	Pass


Supervisor
M. AKIF Ph.D.
Assistant Professor
Department of Biochemistry


Head, Dept. of Biochemistry
11/11/2022
HEAD
Department of Biochemistry


Dean, School of Life Sciences
DEAN
School of Life Sciences
University of Hyderabad

ACKNOWLEDGEMENTS

An indeed thanks to my supervisor **Dr. Mohd Akif** for his supervision, interests in my work and providing me the facilities to do the work. Working with you was very exciting sir! Your interest in my work and our scientific tuning helped a lot to achieve all the success in my PhD journey. Thanks for giving me chance to work on TrxR project which gave me my first career break and successfully we published the first paper of our lab and after that we didn't stopped, even in the bad time of lockdown also. I don't know it was by chance or destined all chapters of my thesis work got published in May month of each consecutive year. So yes we have made a hatrick since 2020-2022 and cheers and thanks for that sir! Overall you are very nice person sir. In your lab I have learnt many practical life things which will be helpful for my future endeavors. I am thankful to you for all the efforts you have put in my PhD marathon race!

I am thankful to my Doctoral Committee members **Prof. H.A Nagarajaram and Dr. P.Anil Kumar** for their valuable suggestions. I especially thanks to Nagarajram sir, for his trust in me and his appreciations always encouraged me a lot. Very patiently he listens my presentations and never said no to give his time for my academic stuffs.

I want to thanks to previous and present Deans of SLS **Prof K.V.A Ramaiah, Prof. Dayananda, Prof. N Siva Kumar.**

I would like to thanks to previous and present HOD of Biochemistry Department **Prof. Mrinal Kanti Bhattacharya, Prof. Krishnaveni Mishra.** Thank you for maintaining all the scientific gatherings in the department which have always cheered us a lot.

I would like to extend my thanks to all the professors of **School of Life Sciences.** I am really grateful that I have trained under such a caring and friendly environment.

I am really thankful and appreciate the **HPC facility** provided by **CMSD of UOH.**

I would like to thanks my present and previous lab members **Faisal, Nachiket, Shiraz, Yash, Ruchira, Hitakshi, Haneesh, Shagufa, Pradeep, Sripriya.** Thank you **Faisal** for all the scientific discussions and making the environment more cheering in the lab. Thank you **Hitakshi** for always being available to me in the midnight also for all the medical emergencies that I have faced in my last days of PhD thesis writing time.

A ton of thanks to my friends **Rachana, Shubhangi, Neha, Yashasawi, Smita, Aishwarya, Anusha, Pritikana, Faisal, Gufrana, Vishnu Priya, Shaumashish, Srinivas, Kiran,** and all my PhD course work friends & Many thanks to all the smiling faces of SLS. You guys are like detox element for me. With you all I have learnt how to maintain balance between both work and

personal life. You guys always scolded me to take care of health. Without the support and love from you guys' journey wouldn't have been so easy for me.

I want to thanks to my PhD seniors **Pratap anna, Sarat Anna, Santhosh Anna, Kannan.**

I owe my thanks to all **office and non-teaching staffs** of SLS. Thank you **Chary and Prabhu sir** for always helping me in admin work.

Thank you to all the cleaning **Akkas and Annas & security staffs** of SLS. I always admired by the sincerity you show for your job!

I am thankful to my **B.Sc & M.Sc professors & friends**. After leaving the place also you all have always cheered me up for my success and achievements. In my M.Sc I wasn't having the idea of any techinal things but you all have cooperated a lot to make me understand the things.

A special thanks to **Prof. Somdatta Sinha** your published papers helped a lot in furnishing the theme of my PhD work. Extending the thanks to my previous working place members **Dr. Monika Sharma, Ashutosh for providing his thesis, Nitika di, all friends of IISER Mohali.** Thank you **Somdatta Ma'am** for your scoldings & concerns. Your scoldings taught me how to be an independent reseracher! I am thankful that I got the opportunity to work under you! I always remembered your advice which you had given me while I was leaving your lab "Surabhi I will keep asking about your progress, don't get induldge in any other activity and complete your PhD on time! Whenever I felt sad and disappointed you always reminded me with the goal I had come here. I can never forget your efforts you have put me to learn that how to do research. I can never forget any of your advice like how to deal with different types of people and always encouraged me to accept both good and bad that comes along my righteousness decisions. I really feel myself lucky that I got the opportunities to work under you. Your life journey and passion for research has always inspired me that how to be consistent and focused inspite of all difficulties. I awlays thanks to my destiny for giving me a guru like you **Somdatta Ma'am!!**

I am grateful to my both late maternal (nana) and paternal (baba) side grandfathers. I am sure today you both would be happiest person after seeing me getting this highest degree of education. Thank you to my Nani & late Aaji. And a final heartfelt thanks to my **Parents (Mummy & Papa)** for believing in me and seeding the career importance in me. !! I am really blessed with luck to have parents like you. I remember the decision when you have taken to sent me out for my studies after my 10th Standadrd especially then when we were financially not stable. You always encouraged me to prioritize my career more than any thing. Thanks to my brother Saurabh for his background support. And thanks to all my loving and charming family members who has always pampered me a lot.

I am really thankful to my destiny for always favouring me for the things what was right for me.

I dedicate me thesis to all the taxpayers of India. Because of their contribution only I was able to work in such a safe environment of university and providing me the University fellowship.

*A thesis dedicated to all the tax-payers
of India!!*

Table of Contents

List of Figures	1-3
List of Tables	4
Preface	5-14

CHAPTER 1:

Introduction	16
1.1 Protein molecules as therapeutic target: perspective of structure-function	16
1.2 Therapeutic Targets	18
1.2.1 Severe Acute Respiratory Syndrome- CoV-2 (SARS CoV-2)	19
1.2.1.1 Genomic organization and potential virulence factors	20
1.2.1.2 Molecular Basis of pathogenesis	21
1.2.1.3 Spike Protein as potential therapeutic target	23
1.2.1.3.1 Role of glycosylation sites in the spike protein	25
1.2.1.3.2 Spike protein in vaccine design & therapeutic development	26
1.2.2 No-Structural Protein	29
1.2.2.1 Main Proteases (M ^{pro})	30
1.3 Variants of SARS CoV-2	32
1.4 <i>Mycobacterium tuberculosis</i> and its pathogenesis	34
14.1 Events in disease process	35
1.4.2 Treatments so far and the challenges	37
1.4.3 Drug Resistance in Mtb	40
1.4.4 <i>Antioxidant system</i>	41
1.4.5 <i>Thioredoxin system</i>	42
1.5 Aim of the thesis	45

CHAPTER 2:

Investigation of protein structural parameters and dynamic features in Spike protein

2.1 Introduction	50
2.2 Methods	53
2.2.1 Construction of Protein Structure Graph (PSG)	53
2.2.2 Network Parameters	54
2.2.3 Protein stability and flexibility analysis	54
2.2.4 Perturbation Residue Scanning	55
2.2.5 Molecular Dynamics Simulation	55
2.2.6 Principal component analysis (PCA)	56

2.2.7 Binding free energy calculation	56
2.2.8 Normal Mode Analysis	56
2.3 Results	57
2.3.1 Protein structure network analysis of wild type and P.1 variant protein structures	58
2.3.2 Perturbation in PSG due to mutation outside the RBD domain.	63
2.3.3 Effect of mutations on the thermodynamic of protein.	67
2.3.4 Probing of conformational dynamics of RBD-ACE2 complex	70
2.3.5 Probing the Conformational Variations in the truncated S1 domain	80
2.4 Conclusions	85

CHAPTER 3:

Structure-based drug design of potential inhibitor targeting the SARS-CoV-2 Spike Protein: ACE2 interface

3.1 Introduction	88
3.2 Methods	89
3.2.1 Preparation of the target protein coordinates	89
3.2.2 Fragment substructure selection and analogue generation	89
3.2.3 Molecular Docking and Ligand Scoring	90
3.2.4 Molecular Dynamics Simulation Studies	90
3.2.5 Binding Free energy Calculations	91
3.3 Results	91
3.3.1 Binding mode analysis of SSAA09E2 on RBD-ACE2 interface for WT and P.1 variant.	91
3.3.2 Computationally designed SSAA09E2 analogues and its interactions mechanism.	93
3.3.3 Structural stability of compounds in the RBD-ACE2 complex	97
3.3.4 Effect of designed analogues as potential disruptors against the RBD-ACE2 binding	102
3.3.5 Effect of analogs as potential disruptors on interfacial residues	104
3.3.6 Effect of analogs as potential disruptors on RBD-ACE2 SASA interface surface	107
3.3.7 Effect of the compounds on the ED of protein complex	107
3.4 Conclusions	110

CHAPTER 4:

Investigation of protein structural parameters in SARS-CoV-2 M pros

4.1. Introduction	112
4.2 Methods	115
4.2.1 Protein Structure Network Construction	115
4.2.2 Network parameters	115
4.2.3 Structural Communication Analysis	116

4.3 Results	117
4.3.1 Protein Structure Network analysis of the SARS-CoV-1 and SARS-CoV-2 M ^{pro}	117
4.3.2 Community Structure Analysis of the PSNs	122
4.3.3 PSN analysis of inhibitor (ketoamide) bound complexes of M ^{pro} s	123
4.3.3.1 Inhibitor binding perturbs the PSG and the communication pathway in M ^{pro}	123
4.3.3.2 Community analysis of apo and inhibitor bound complex	127
4.3.4 PSN analysis on quaternary structure of SARS-CoV M ^{pro}	129
4.4 Conclusion	133

CHAPTER 5:

Structure-based inhibitor design against *M. tuberculosis* Thioredoxin Reductase

5.1 Introduction	135
5.2 Methods	138
5.2.1 Preparation of the target Coordinate	138
5.2.2 Virtual screening and molecular docking	138
5.2.3 Molecular dynamic simulation	139
5.2.4 Binding-free energy calculations of compounds	140
5.2.5 Principal component analysis	141
5.2.6 Protein contact network analysis	142
5.3 Results	142
5.3.1 Structure-based virtual screening and docking	143
5.3.2 Binding interaction analysis of hit compounds	144
5.3.3 Analysis of Molecular Dynamics Simulations	146
5.3.4 Evaluation of binding affinity with the protein molecule	149
5.3.5 Analyses of protein conformational variation	150
5.4 Conclusion	156

CHAPTER 6:

6.1 Summary	159
-------------------	-----

Bibliography	167
---------------------	-----

Appendix	195
-----------------	-----

Plagiarism Report	204
--------------------------	-----

List of Figures

- Figure1.1 SARS-CoV-2 genome organization
- Figure1.2 Pathogenesis mechanism of SARS-CoV-2
- Figure1.3 Domain organization of Spike protein
- Figure1.4 Spike protein mechanism from prefusion to postfusion
- Figure1.5 3D structural arrangement of SARS-CoV-2 Mpro in its monomeric form
- Figure1.6 3D structural arrangement of SARS-CoV-2 Mpro in its dimeric form
- Figure1.7 WHO TB report for estimated cases in 2020 reported for countries with a minimum of 100,000 cases
- Figure1.8 Events in Tuberculosis disease process
- Figure1.9 Progression of LTBI or active TB
- Figure1.10 Open and closed state conformation of TrxR
- Figure1.11 Thioredoxin system in *Mtb*
- Figure2.1 Structural alignment of the RBD-ACE2 complex from SARS-CoV-2 wild type and P.1 variant
- Figure2.2 Representation of hubs arrangement on RBD-ACE2 complexes (A) wild type and (B) P.1 variant
- Figure2.3 Comparative community structures in RBD-ACE2 complexes. (A) wild type and (B) P.1 variant
- Figure2.4 Shortest path communication channel on the RBD-ACE2 complex. A) wild type and B) P.1 variant
- Figure2.5 Comparative hubs arrangement in the S1 domain-ACE2 complexes from wild type and P.1 variant
- Figure2.6 Mapping of the communities on the truncated S1 domain-ACE2 complexes of A) wild type and B) P.1 variant
- Figure2.7 Shortest path communication channel in the truncated S1 domain-ACE2 complexes of A) wild type and B) P.1 variant
- Figure2.8 Visual representation of the interface mutations on dynamicity and plasticity the RBD. (A) RBD structure with the K417T mutation RBD structure with the E484K mutation (C) RBD structure with the N501Y
- Figure2.9 Visual representation of effects of the mutations present in the truncated S1 domain on protein dynamicity
- Figure2.10 Residue-wise Perturbation Residue Scanning profile of three interface mutations present on the RBD
- Figure2.11 Graphical representation of backbone RMSD trajectory of RBD-ACE2 complex from wild type and P.1 variant for three independent replicates of 100ns simulations
- Figure2.12 RMSF trajectories generated for the wild type and P.1 variant
- Figure2.13 Graphical representation of the conformational dynamics
- Figure2.14 Plots representing strength and compactness in the RBD-ACE2 complexes
- Figure2.15 Covariance matrix plot for RBD-ACE2 complexes
- Figure2.16 Motion Mode Analysis of the RBD for A) wild type and B) P.1 variant
- Figure2.17 Total number of interactions
- Figure2.18 Representation of interface interacting residues involved in hydrogen bond interactions in the representative structure of wild type and P.1 variant

- complex obtained after clustering analysis
- Figure2.19 Representation of the conformational variation
- Figure2.20 Representation of the conformational variation
- Figure2.21 Secondary structure content analysis during the course of 100ns MDS for A) wild type and B) P.1 variant
- Figure2.22 Electrostatic charge distribution at the interacting interface of S1-ACE2 complex of wild type and P.1 variant at different time frames of 100ns MDS trajectories
- Figure3.1 Representation of docking of the SSAA09E2
- Figure3.2 2D representation A) 2D structure of parent compound SSAA09E2. B) two analogues designed from FragRep search
- Figure3.3 A) The best docking pose of Compound 1 in the RBD-ACE2 interface binding cavity. B) 3D interaction plot between compound and RBD ACE2 complex residues. C) 2D interaction plot between compound and RBD-ACE2 complex residues
- Figure3.4 A) The best docking pose of Compound 2 in the RBD-ACE2 interface binding cavity. B) 3D interaction plot between compound 2 and RBD-ACE2 complex residues. C) 2D interaction plot between compound 2 and RBD-ACE2 complex residues.
- Figure3.5 RMSD plot at 200ns (A) protein backbone and compound1. (B) protein backbone and compound2.
- Figure3.6 Protein compound interaction profile on the representative structure after 200ns of MDS
- Figure3.7 Hydrogen plot during 200ns MDS. (A) Compound1 bound complex (B) Compound2 bound complex
- Figure3.8 Residue wise energy decomposition plot. A) Compound 1. B) Compound2
- Figure3.9 Rmsf plot (A) ACE2 (B) RBD)
- Figure3.10 Residue-wise RMSD for apo and compound bound complex
- Figure3.11 B factor analysis on representative structure obtained after 200ns MDS
- Figure3.12 Percentage of variance for eigenvectors
- Figure3.13 PCA plot along its two major eigenvectors
- Figure3.14 Porcupine plot of apo and compound bound complexes
- Figure3.15 Radius of gyration plot during 200ns of MDS
- Figure3.16 Hydrogen bond distance bar plot for key interface residues between RBD and ACE2
- Figure3.17 Centre of mass measurement between $\alpha 3$ - $\alpha 4$ helix of ace2 and RBM of spike protein. A) apo, B) compound1 complex C) compound2 complex
- Figure3.18 Superimposed structure of interacting ace2 and RBM complex for A) complex, B) compound2 complex
- Figure4.1 Contact maps in M^{pro} (A) New contacts in SARS-CoV-2 near the active site (B) Residues showing the highest increase in BC along with the new contacts mapped on the SARS-CoV-2 structure
- Figure4.2 Residue wise changes in betweenness centrality of (A) SARS-CoV-1 M^{pro} and (B) SARS-CoV-2 M^{pro} ; Residue wise Clustering coefficient of (C) SARS CoV-1 M^{pro} and (D) SARS-CoV-2 M^{pro}
- Figure4.3 Community structure in the M^{pro} . All communities are mapped on their

tertiary structure of M^{pro}s and depicted with different color modules (A) SARS-CoV-2 and (B) SARS-CoV-1

Figure4.4 3-D PSG representations of the inhibitor bound complexes of M^{pro} (A) SARS-CoV-1 (B) SARS-CoV-2

Figure4.5 Residues involved in global meta-path mapped on 3-D structure of M^{pro}. (A) SARS-CoV-1 and B) SARS-CoV-2

Figure4.6 Community structure in the inhibitor bound states of M^{pro}

Figure4.7 PSN analysis on the quaternary structure of SARS-CoV M^{pro}s

Figure5.1 (A) Predicted binding cavity on the *Mtb* TrxR structure. (B) The chemical structure of two screened compounds: compound1, officinaruminane_B and compound2, dichotomine_D

Figure5.2 LIGPLOT and a 3-D representation of protein–ligand interactions of the screened compounds in the selected cavity of *Mtb* TrxR. A) compound1 B) Compound2

Figure5.3 A) Backbone RMSD vs time. B) RMSF vs residue number. Secondary structure arrangement of TrxR according to residues shown on the top of the RMSF plot. C) SASA vs time of apo-TrxR complex with compound1 and Compound2

Figure5.4 Covariance matrix plot of apo and complex form of *Mtb* TrxR during 20 ns MD simulation. A) apo–TrxR. B) complex with compound1. C) complex with compound2

Figure5.5 PCA plot constructed by eigenvector 1 and eigenvector 2

Figure5.6 Protein contact network of the apo and compound bound *Mtb* TrxR. Contact network of among the residues of the apo and the compound-bound forms of TrxR

Figure5.7 Center of mass distance calculation between the NADPH and the FAD Interface domain residues. A) Apo form B) Compound1 bound form C) Compound2 bound form

Figure5.8 Superimposition of structure of the apo–TrxR and compound1 complex form of TrxR

List of Tables

Table1.1	Functional role of Nsps1-16 of SARS-CoV-2
Table1.2	List of Mpro identified inhibitors
Table1.3	Early variants of SARS CoV-2 and its list of mutations
Table1.4	List of first line drugs against TB
Table1.5	List of second line drugs
Table2.1	List of global average network parameters
Table2.2	Hubs residues participating in the RBD-ACE2 complex of the wild type and P.1 variant
Table2.3	Network components and parameters involved in average shortest path for RBD-ACE2 complexes from the wild type and P.1 variant
Table2.4	Predicted free energy values of the mutants
Table2.5	Predicted change in entropy values associated with the interface mutations on the RBD
Table2.6	Average MMPBSA calculations for the wild type and P.1 variant complexes
Table3.1	Docking score for SSAA09E2 with the RBD-ACE2 complex of the WT and P.1 Variant
Table3.2	Binding free energy analysis of compound1 and compound 2 with the protein Complex
Table4.1	Average network parameters of M ^{pro} monomer and dimer units, generated through PSNs.
Table4.2	Residues showing largest change in degree
Table4.3	List of hubs in SARS-CoV-1 and SARS-CoV-2 M ^{pro}
Table4.4	Residues with highest BC value
Table4.5	Network components and its parameters for both SARS-CoV-1 and SARS-CoV-2 M ^{pro} monomer in apo and holo states and M ^{pro} dimer in apo state
Table4.6	Residues with highest BC value in dimeric form of M ^{pro}
Table5.1	Docking score of top two screened compounds from the TCM database
Table5.2	Average MM/PBSA free energies of <i>Mtb</i> TrxR complexes with compounds, calculated from the MD simulation performed at 20 ns
Table5.3	Dominant motion of atoms in NADPH and FAD domains and the hinge region of the protein

PREFACE

Protein characterizes the property of every living organism. It regulates the various cellular processes inside the cell and serves as an important macromolecule in the cell. Many proteins are being utilized as therapeutic targets to combat disease conditions. Hence, understanding the structure-function relationship of protein molecules is very important [Redfern et al. 2008]. The function of a protein is determined by its 3-D structure, which adopts several conformational changes while performing a different function. Moreover, the conformation of a protein is modulated by several factors, such as mutations of amino acids, interaction with other protein molecules, or the crowding effect of other large molecules inside the cell [Schmid et al. 2020]. So, the structure-function paradigm states that the function of a protein is determined by its 3-D structure, and a change in protein structure results in a change in function. Hence, the structure-function paradigm is crucial for studying the functional mechanism of any therapeutic proteins. However, there have been many cases where significant alteration of protein function has been observed due to subtle changes in the protein structure [Srivastava et al., 2014; Srivastava et al., 2016]. For example, a mutant form of HIV-reverse transcriptase retains 3-D structural integrity as the wild-type. Still, it displays a functional change regarding the binding of proposed inhibitors [Srivastava et al. 2020]. This suggests that there must be some missing link in understanding the structure-function relationships. There have been reports that in the absence of noticeable structural change in the protein, the changes at the protein contact level play an important role in establishing the structure-functional relationship. However, it is challenging to characterize these small changes in the protein structure and to understand the role of these changes in the physiological change in the function. Protein molecules are known to be flexible in nature, only a 3-D structure sometimes is not enough to probe the molecular mechanism of the protein.

Moreover, the 3-D structure does not provide complete details of functional change in the protein due to modulations and changing physiological conditions. There must be a missing link between structure and function. This lies in the dynamics of proteins, which determine the molecular behavior of protein molecules with respect to time [Rader et al. 2005]. So, the theme of this thesis is to use integrative computational approaches such as graph theory and Molecular Dynamics Simulation to study subtle but functionally significant variations in protein structures. For the study, three different therapeutic targets were chosen. (1) **Severe acute respiratory syndrome-Coronavirus-2 (SARS-CoV-2) spike protein** where the effect of mutants on host-pathogen interaction without much change in the 3-D structure was studied. (2) **SARS-CoV-2 main proteases (M^{Pro})** where changes in amino acid residues result in the differential binding and sensitivity of the inhibitors without significant change in the 3-D structure are reported. (3) ***Mycobacterium tuberculosis* (Mtb) thioredoxin reductase (TrxR)**, where modulations on protein structure to change the function are studied. Structures of all three topical interest targets are shown in **Figure 1**.

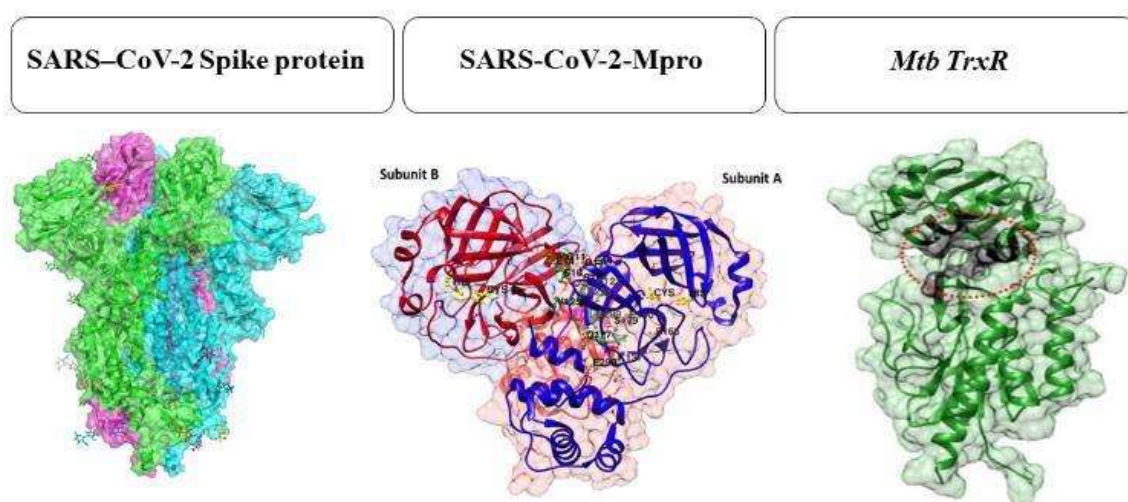


Figure 1. Three-dimensional structures of three tropical interest therapeutic targets were chosen for the study.

Chapter 1 reviews all the literature related to pathogens and therapeutic targets. The novel Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) belongs to the beta coronavirus genus, which also includes SARS-CoV and Middle East respiratory syndrome coronavirus (MERS-CoV) [Cui et al. 2019]. Coronaviruses enter the host cell with the help of spike glycoprotein on the outer envelope through a host cellular receptor, Angiotensin Converting Enzyme 2 (ACE2) [Gui et al. 2017]. Spike protein has been one of the important therapeutic targets for drug design and vaccine development. It is a homo-trimeric protein consisting of S1 and S2 subunits, the S1 subunit aids in the binding to the ACE2 receptor, and the S2 subunit is involved in the fusion of the virus to the host cellular membrane [Wang et al. 2020a]. The structural organization of the S1 domain includes the N-terminal domain (NTD), C-terminal domain (CTD), sub-domain 1 (SD1), and sub-domain 2 (SD2). The CTD consists of the receptor binding domain (RBD, the main component that interacts with the ACE2 receptor. Several studies have demonstrated that virus replication, infectivity, and transmission ability depend on the binding affinities of the RBD to the ACE2 receptor [Wang et al. 2020a].

The COVID-19 pandemic has been more challenging due to SARS-CoV-2 emerging variants. Many variants have been detected with many mutations, mainly in the spike protein. The world health organization (WHO) has classified these variants either as Variants of Concern (VOC) or Variants of Interest (VOI), where three variants have gained more attention: B.1.1.7 (VOC 202012), B.1.351 (501Y.V2) and P.1 (B.1.1.28.1). The B.1.1.7 was first detected in the UK in September 2020 and has been reported with seven mutations in the Spike protein [Davies et al. 2021]. Within a month, a South African variant was detected with nine mutations [Tegally et al. 2021]. Later on 2 January 2021, Japan/Brazil variant named P.1 (B.1.1.28.1) was in circulation with 12 mutations in the Spike protein [Sabino et al. 2021]. The two mutations N501Y and D614G

in spike protein were common in all VOC. Very recently, Omicron has also been identified as a VOC and reported to have almost 30 mutations in the spike protein [Kannan et al. 2022]. The functional role of many mutations has been demonstrated using deep mutational scanning and other biochemical methods. Most of the mutations associated with the variants have been reported to have enhanced affinity towards the ACE2 receptor and are crucial for virus transmission. These mutations are also ineffective in neutralizing antibodies and evading neutralization [Dejnirattisai et al., 2021]. The underlying cause of the enhanced binding affinity in the variants has not been associated with a significant change in the three-dimensional structures of the spike as well as the ACE2 receptor. However, local and global rearrangement of the interaction site and allosteric sites cannot be overlooked. Mutations away from the RBD facilitate spike protein in the open state to bind efficiently with the ACE2 receptor. Most of the previous studies focus on the effect of individual mutation. However, the cumulative effect of the mutations present on the RBD and away from the RBD such as NTD, SD1, and SD2 of the S1 domain was not investigated in detail. We asked the questions for this study. What are the cumulative effects of the mutations on rearrangement in the protein contact networks in the spike protein? What is the structural basis of the increased binding affinity of mutant spike protein with the ACE2 receptor?

Chapter 2 addresses these questions by using integrative network and dynamics approaches. This reports subtle conformational changes in the RBD and the truncated S1 domain of the P.1 variant spike protein responsible for the increased affinity with the ACE2 receptor. Our study on the protein contact network identifies changes in the protein structures at the residue contact level which are otherwise not easily detectable. The network parameters suggested that mutations have perturbed the node-wise network rearrangement in Protein Structure Graph, reflecting the subtle conformational change in the mutant structure. Additionally, to probe the effect of mutation

outside the RBD domain we have constructed the PSG on truncated S1 domain in complex with ACE2, which suggested significant rearrangement of network parameters in the NTD of the spike protein. Then further, we applied a dynamics approach to probe the conformational dynamics of the RBD-ACE2 complex and the S1-ACE2 complex. Our 100ns dynamics study suggests that there is global flexibility and conformational plasticity in the RBD and S1 domain, which can increase the conformational heterogeneity of the complete S1 domain. And additionally, we have identified unique interactions in the P.1 variant, which may provide energetically favorable binding with the ACE2 receptor. Overall, the study provides additional information that may be utilized for designing better therapeutics against the circulating P.1 variants and other future variants.

Chapter 3 explains the results obtained from screening and designing a small molecule that can destabilize the RBD-ACE2 complex of the P.1 variant. The study applied a fragment replacement approach against the parent compound SSAA09E2, which has the potential to block the SARS-CoV-1 infection [Adedji et al. 2013]. But our docking study of the same parent compound with WT and P.1 variant ACE2-RBD complex results in a decrease in affinity and changes in the binding geometry inside the cavity. So we hypothesized to design an analog of SSAA09E2 (**Figure 2**).

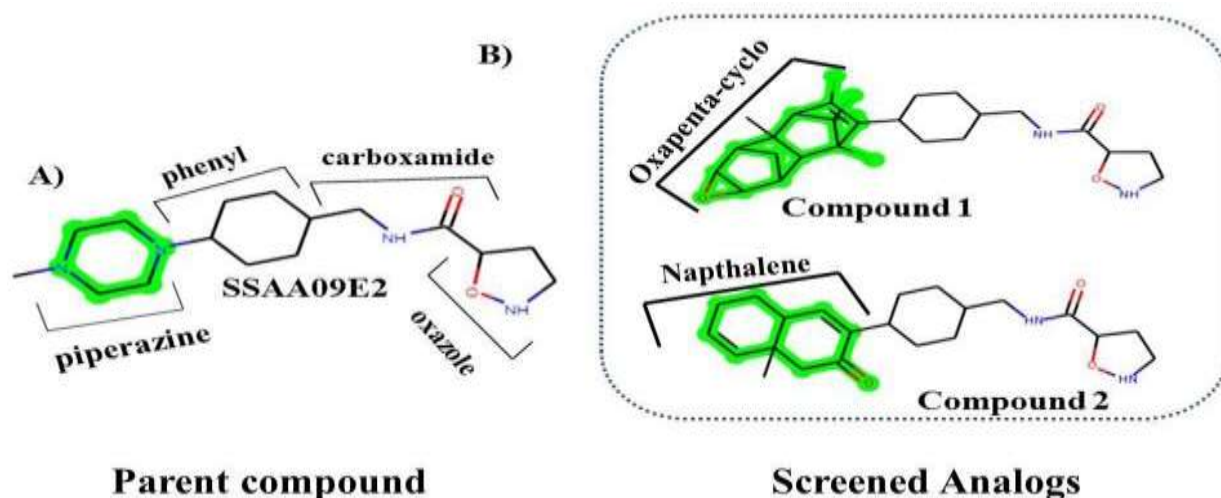


Figure 2. Parent compound and screened analogs are named compound 1 and compound 2. Group colored green in the parent compound is replaced with other groups in compounds 1 and 2.

The FragRep database was screened against the piperazine group of the parent molecule. Out of 1900 screened compounds, we have selected the top two compounds according to binding score and geometry. Both screened compounds had better binding affinity compared to the parent molecule. Also, they made important hydrogen bond interaction with interface residues of the RBD and van-der-Waal interaction with the ACE2 interface residues. Additionally, both compounds interacted with those residues identified as crucial in our previous chapter. It suggests that both compounds can potentially interfere with the interface residues. Further, to check the stability of screened analogs against the ACE2-RBD interface cavity, we have performed 200ns long simulations. The RMSD trajectory, Hydrogen bond trajectory, and MMPBSA analysis suggest that both compounds are stable throughout the simulations with high negative binding free energy where interface residues contributed more negative free energy. In addition, we evaluated the effect of the designed analogs as potential disruptors against the RBD-ACE2 binding. The RMSF analysis of ACE2 and RBD residues indicates that the flexibility of RBD interface residues was

high in both compound-bound complexes, leading to higher entropy penalty and negatively influencing the binding energy. Further, we have calculated the distance between those interface residues participating in the hydrogen bond. Expectedly, we have observed apparent disruption in the distance of interfacial interacting residues of the complex. Interface SASA trajectory was also increased for both compound-bound complexes. It indicates that both compounds have a primary disrupting effect on the interface of ACE2 and RBD. Further, we have also analyzed the secondary effect of compounds on the complex. For this, we have performed PCA on apo and compound-bound complexes where the trace value of compound1 was high compared to apo, and also, the compound-bound complexes have altered the Essential dynamics of the RBD-ACE2 complex. Overall, the designed analogs can induce structural changes locally and globally.

Chapter 4 describes the therapeutic investigation of SARS-CoV-2 main protease (M^{pro}) or 3-chymotrypsin-like protease ($3CL^{pro}$). It is a non-structural protein involved in the processing of polyproteins cleaved into 16 NSPs [Anand et al. 2002]. The M^{pro} has been designated as an important drug target because of its essential role in the processing of polyproteins translated from the viral RNA. Recent crystal structure of SARS-CoV-2 M^{pro} [Zhang et al. 2020] reveals its structural similarity with M^{pro} of SARS-CoV-1 and has a high degree of sequence identity (96.1%) among the two [Savarino et al. 2005]. Despite high identity and lack of structural change, the identified inhibitors against SARS-CoV-1 M^{pro} are reported to have a differential binding affinity against SARS-CoV-2 M^{pro} . The reasons for the differential affinity have not been probed in detail. Moreover, as reported earlier, a small rearrangement of protein at the structural level by substituting a few amino acids at the substrate binding pockets or allosteric sites results in changes in internal interactions, which may lead to differing patterns of inhibitors sensitivity [Mahdi et al. 2020]. Since no significant structural changes are noticed at the active site, a subtle change of

interactions at the allosteric sites of the proteins may have an effect on the sensitivity of the inhibitors. Hence, a protein structure network (PSN) based approach can investigate the negligible conformational changes associated with the protein structure.

To address this, we have performed a comparative PSN analysis to study the subtle change in interactions due to a change in the amino acid at the allosteric site of protein which may affect the sensitivity of inhibitors. We have observed a noticeable difference in each network parameter. The change in contact pattern at the active site was observed, suggesting a conformational change in SARS-CoV-2 M^{pro}. Unique hub residues were observed at the active site of SARS-CoV-2 M^{pro}, which offers inter-domain communication. Residues with high Betweenness Centrality values were observed to make new contacts in SARS-Cov-2 M^{pro}. Differential perturbation in network parameters was also observed in inhibitor-bound complexes, SARS-CoV-1 and SARS-CoV-2 M^{pro}, which was not evident at the structural level. Additionally, the PSN analysis of the quaternary complex suggests network changes at the interface and long-range interactions and has identified crucial residues involved in the complex formation.

Chapter 5 demonstrates structure-based drug design against a therapeutic target from bacterial systems, i.e., *Mtb* TrxR, which provides a reducing environment to *Mtb* inside the macrophages [Fahey et al., 2001]. The thioredoxin system in *Mtb* reduces peroxides and dinitrobenzenes and detoxifies dinitrobenzene hydroperoxides [Lu et al. 2014]. It has been validated as an essential gene in the *Mtb* genome [Zhang et al. 2012]. It shares only 23% similarity with *HsTrxR*, which makes it a unique therapeutic target. Its 3D structure consists of the NADP and FAD domains, which are accompanied by two antiparallel hinge regions [Akif et al. 2005]. Electron transfer occurs from the nicotinamide ring of NADP to the isoalloxazine ring of FAD. Then electrons are transferred to two active site cysteine residues, of which a reduced form is available for thioredoxin

substrate. But the distance between two rings is more than 16 Angstrom; at this distance, the electron transfer mechanism is impossible. It was proposed in *E. coli* TrxR that 66-degree rotation in the NADP domain brings it closer to the FAD domain, and two buried cysteine residues come on the surface, readily available for the Trx substrate [Lennon et al. 2000]. In our study, we aimed to restrict this conformational change by designing a small molecule that can target the interface residues of NADP and FAD along with the hinge region residues. An allosteric cavity was probed, including the interface's residues and the hinge region. We have used this cavity for the virtual screening of natural compounds from the Traditional Chinese Medicine Database. The top two compounds were selected against the cavity according to their docking score. MDS was applied to apo and both compound bound-complexes, where the RMSD trajectory of the apo complex was high compared to compound-bound complexes. Further, through RMSF calculation, we have observed an increase in flexibility of the interface loop region from the FAD domain, which was stable in the compound-bound complex. This decrease in flexibility indicates that compounds are making favorable interactions in the cavity. Both the compounds have high negative binding free energy values, which suggests that compounds form stable molecular interactions with amino acid residues of the TrxR. Our PCA analysis indicates that both compounds have reduced conformational sampling along their two eigenvectors compared to the apo complex. There is a shift in the 3-alpha helix of the NADP interface, which is engaged in hydrogen bond interactions with the loop region of the FAD. These interactions were absent in the apo, which indicates that both compounds are locking the conformational change of the NADP domain. Further, we have performed PCN on apo and compound bound complex, which suggests the rearrangement contacts between the interface domains contribute to the conformational changes in the complexes along with the reduction in the Center of the mass distance between the NADP and FAD domain. Overall,

this study suggests that both compounds can potentially interfere with *Mtb* TrxR, whose binding leads to significant conformational change at the interface of both domains.

Chapter 6 briefs the summary of the work. The study on three different therapeutic targets from two other systems highlights that the subtle change in protein conformation due to modulation caused by mutations, inhibitor binding, or interactions with other molecules alters the protein function. The PCN and MDS are helpful approaches to identifying the subtle changes that are impossible to locate at the structural level. Proteins' dynamics help to determine the conformational changes obtained from static protein structures. Moreover, our PCN and MDS studies provide insight into identifying the critical residues contributing to change in function while maintaining structural integrity.

CHAPTER-1

Introduction & Review of Literature

1. INTRODUCTION

Proteins are the essential entity of a living organism. Proteins catalyze various biochemical reactions and are known as the workhorse of the cell. It does diverse essential functions like (i) catalysis of chemical reactions: enzymes are proteins that act as catalysts in any chemical reaction by increasing the rate of chemical reaction, (ii) Regulation: It regulates various biosynthetic pathways, (iii) DNA synthesis and repairing: many proteins are involved in DNA synthesis, repairing, and genome stability, (iv) cellular signaling and transport: membrane proteins act as receptors that detect the external stimuli and trigger specific cellular responses. Proteins as channels transport substrates/ions across the cellular membrane, (v) Metabolism and maintenance: proteins involved in the synthesis and degradation of cellular components. Proteins provide structural support to the cells. Microtubules and actin proteins maintain the structural integrity of the cell, (vi) cell division: proteins such as spindle fibers separate the chromosomes during cell division, (vi) Oxygen transport: two well-known oxygen binding proteins, hemoglobin, and myoglobin, transport oxygen across the tissues. Proteins are essential macromolecules and are responsible for regulating almost all cellular life processes.

1.1 Protein molecules as therapeutic targets: a perspective of the structure-function relationship

The current paradigm for treating diseases is associated with targeting protein molecules. Targeting proteins involved in virulence and pathogenesis is an important strategy for controlling infectious diseases [Alberts et al. 2002]. This focuses on developing drugs that may inhibit the pathogen's virulence rather than combating its viability [Clatworthy et al., 2007]. Moreover, targeting virulence factors reduces the chances of developing resistance due to selective pressure.

So, protein molecules responsible for virulence factors are the best therapeutic targets in terms of specificity, absence of toxicity, and side effects. For a protein to be a therapeutic target, it is required to fulfill mentioned four criteria: (1) it needs to be essential for pathogen survival, (2) the absence of homologous proteins in humans, (3) it should be druggable, and 4) biological role of the target in the diseased state.

Identifying protein molecules as therapeutic targets involved in the virulence mechanism is the primary step in infectious disease control. To control infectious diseases, a detailed structural and functional understanding of a particular therapeutic target is required. The three-dimensional structure of a protein provides an essential insight into the structure and function relationship [Anfinsen et al. 1961; Anfinsen 1973]. Protein's conformational flexibility and dynamics contribute toward understanding the molecular mechanism. This knowledge generally aids in designing therapeutics. The 3-D structures of many therapeutic targets from the PDB have been used for drug discovery and design [Berman et al. 2000]. These structures usually do not represent the functional states of the protein. Instead, the functional forms of proteins are associated with the conformational change required to perform an activity. The conformational changes usually result from interactions with other entities such as substrates, inhibitors, or inter and intra-domain interactions. These conformational changes in the protein 3-D structure modulate the functional site, which may have an effect on protein function. This general process comes under the "structure-function-paradigm," which means that the function of any biomolecule is determined by its three-dimensional structure [Redfern et al. 2008]. The structure-function paradigm is also essential for studying therapeutic proteins' functional mechanisms. Conformational changes in a protein and residues involved in the conformational variations are usually considered a druggable site in the drug design process. Modulations in the protein structures usually result from mutations

at the functional or ligand binding or protein-protein interaction sites. There are few cases where structural modulation exerts unnoticeable changes in the protein 3-dimensional structure. It has been observed that mutations of a few important residues in protein lead to changes in the function with negligible changes in its three-dimensional structures [Srivastava et al. 2015; Srivastava et al. 2020]. Hence it is important to understand how these modulations maintain the structural integrity and regulate the protein's function. The structure-function paradigm states that proteins acquire specific shapes for performing a particular function efficiently. Proteins are usually flexible and change their shape like a tree in the wind. So, the only static structural knowledge obtained through crystallography is insufficient to probe the molecular mechanism of protein's biological functions. Instead, the dynamics of proteins provide molecular insights into their action. Moreover, dynamics play a crucial role in bridging the missing link between the structure and function paradigm [Rader et al. 2006]. This study is an attempt to understand the effect of modulations on the protein structure, which is not detectable at the structural level. Over two decades, many computational efforts were employed to understand subtle structural changes at the contact level, which may have significant functional alterations in the protein.

1.2 Therapeutic targets

Three different therapeutic targets belonging to two phylogenetically diverse pathogens (viral and bacterial) were considered for this study. These are (1) **Spike protein from Severe Acute Respiratory Syndrome-2 (SARS-CoV-2)**: an important therapeutic target present on the outer envelope of coronaviruses. (2) **Main protease (M^{Pro}) from SARS-CoV-2**: is an essential drug target that plays a role in polyprotein processing, and (3) **Thioredoxin Reductase from *Mycobacterium tuberculosis* (Mtb)**: a member of the thioredoxin system, a well-known antioxidant system of *Mtb*.

1.2.1 SARS-CoV-2

SARS-CoV-2 belongs to the coronaviridae family of viruses, usually known as enveloped and positive sense RNA viruses. The viruses from this family are classified into four genera named Alpha coronavirus (α -CoV), Beta coronavirus (β -CoV), Gamma coronavirus (γ -CoV), and Delta coronavirus (δ -CoV). Alpha coronavirus and Beta coronavirus mainly infect mammals. Gamma coronavirus comes under the subfamily of orthocoronavirinae. Gamma and delta coronavirus infect mammals but also infect birds. Alpha and Beta genera originated from the bat gene pools, while gamma and delta are originated from avian and pig viral gene pools. β -coronavirus generally includes three highly pathogenic viruses Severe Acute Respiratory Syndrome Coronavirus 1 (SARS-CoV-1), Middle East Respiratory Syndrome Coronavirus (MERS-CoV), and SARS-CoV-2 [Cui et al. 2019]. SARS-CoV-1 originated in China and caused a global pandemic in 2003 with about a 10% fatality rate [Zhong et al., 2003; Ksiazek et al., 2003]. MERS-CoV was first reported in Saudi Arabia in 2012 and has infected the human population with limited human-to-human transmission. SARS-CoV-2 a new coronavirus reported for the first time from Wuhan, China, in December 2019 causes severe human respiratory disease [Wu et al. 2020]. The new coronavirus is very contagious and has a rapid transmission capability among human-to-humans. The virus has caused an outbreak of severe pulmonary diseases in almost 216 countries, resulting in approximately 814 438 confirmed deaths globally (WHO report, 2020). The world health organization (WHO) has coined SARS-CoV-2 causing disease as the COVID-19 pandemic, which has now become a global health emergency. Currently, COVID-19 has led to severe ramifications for the worldwide economy.

1.2.1.1 Genomic organization and potential virulence factors of SARS-CoV-2

The genome of SARS-CoV-2 is approximately 30kb in length (Astuti et al. 2020). It consists mainly of two ORFs, ORF1a and ORF1 cover almost two-thirds of its entire genome and code the non-structural proteins. The rest of the genome encodes for structural proteins such as Spike protein (S), an Envelope protein (E), a Membrane protein (M), and a Nucleocapsid protein (N) with some accessory proteins. Ribosomal frameshifting between these two ORFs yields two polypeptides, pp1a and pp1ab, which further process into 16 NSPs with the help of either 3CLpro (3 cysteine-like proteases)/M^{pro} (main protease) or papain-like protease (PLpro). **(Figure 1.1)** [Astuti et al. 2020]. The genome of the SARS-CoV-2 share sequence identity of about 89% with other reported CoVs. It is closely related to MERS-CoV [Guo et al. 2020]. Comparative genome analysis indicates its similarity to the BAT-CoV, suggesting bats as a natural reservoir of SARS-CoV-2 [Shereen et al. 2020].

The structural proteins coded from its genome serve potential virulent factors. Outer membrane spike protein protrudes outside the virus's surface and facilitates the virus entry after attachment with the host ACE2 receptor. The E-protein is a transmembrane protein consisting of an N-terminal transmembrane domain and a C-terminal ectodomain. It performs an essential role in structural integrity and virus assembly. Membrane protein (M-protein) is present in high amounts compared to all other proteins of coronaviruses. It consists of three major domains: the N-terminal domain, three transmembrane helices, and the C-terminal domain. The M-protein is a potential immunogen therapeutic and interacts with other structural proteins. The N-protein consists of two domains, N-terminal and C-terminal domains. Each domain has the potential to bind with the RNA genome and provides structural integrity to the viral genome.

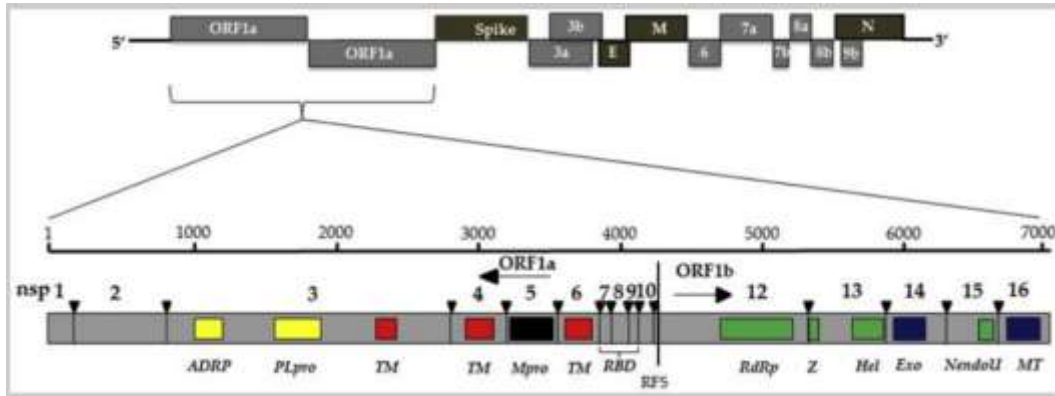


Figure 1.1. Genome architecture of SARS-CoV-2 [Adopted from Astuti et al. 2021].

1.2.1.2 Molecular Basis of pathogenesis

SARS-CoV-2 enters the host cell through its spike protein binding to the ACE2- a host cellular receptor [Lamers et al. 2022]. The binding brings about the conformational change in the spike protein, dissociating the S1 domain of the protein. The conformational change also exposes the S2 domain's cleavage site to be cleaved by the human TMPRSS2. After the cleavage, the viral envelope fuses to the host cell membrane and releases viral ssRNA into the host cells (**Figures 1.2a & b**). The primary site of SARS-CoV-2 infection is the nasopharynx multi-ciliated cells in the nasal mucosa. After the entry of the SARS-CoV-2 genome triggers the formation of viral proteins from the replication factories and gets detected by the cytoplasmic patterns recognition receptors (PRRs). The viral replicase genes (ORF1a and ORF1b) of the viral RNA genome undergo translation and lead to the production of polyproteins (pp1a and pp1b). These polyproteins are, further, processed by viral proteases to produce other Nsps. The Nsps have vital role in the discontinuous transcription of other RNA genes of the viral genome with various mRNA fragments. Detecting translated viral proteins inside host cells by cytoplasmic pattern recognition

receptors initiates the type I and III interferon cascade. This leads to the production of cytokines and triggers T cells and B cells responses (**Figure 1.2c**). Further, JAK/STAT1/2 signaling cascades activate and induce interferon-stimulated gene expression (ISGs) [Lamers et al 2022].

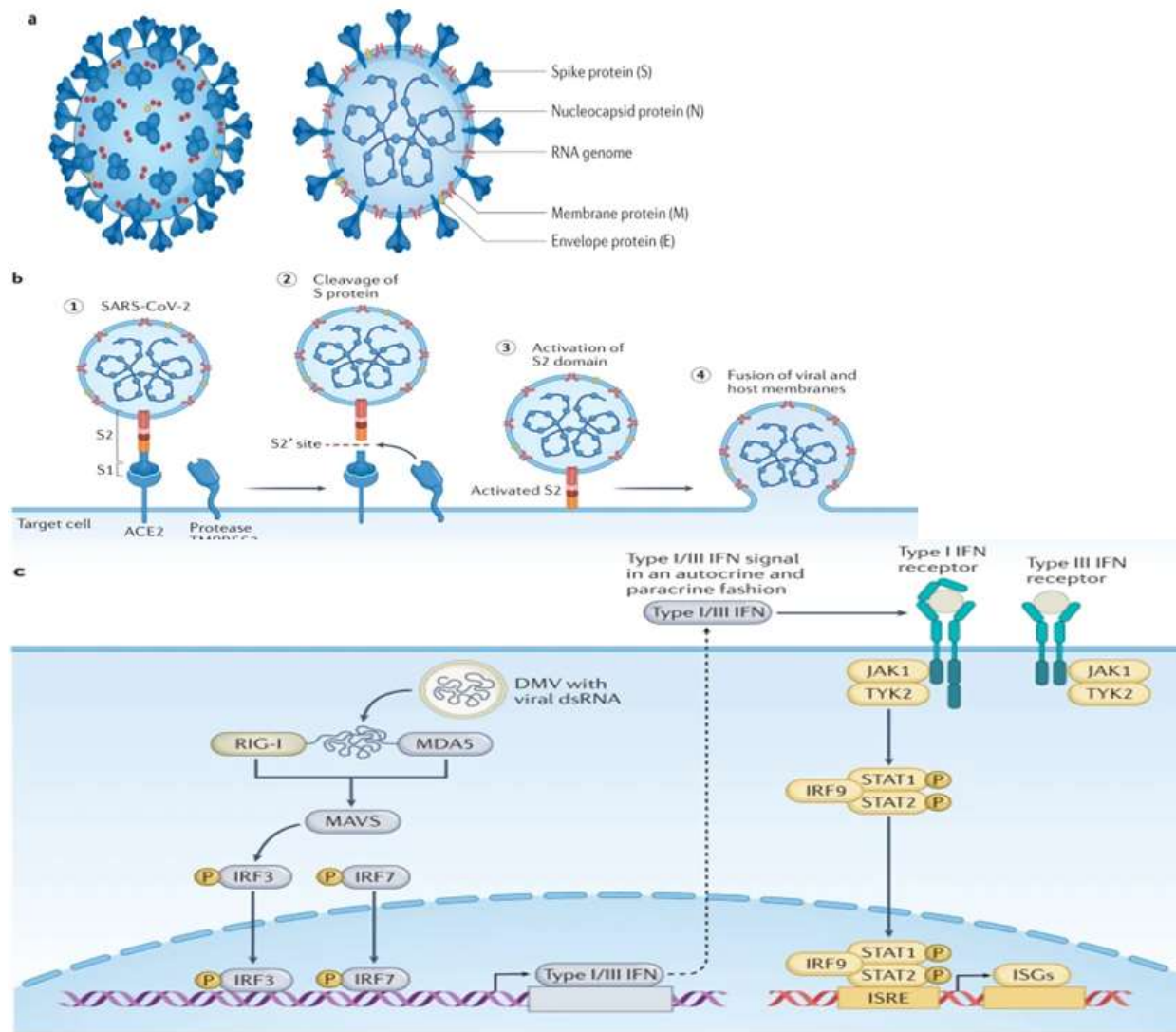


Figure 1.2. Pathogenesis mechanism of SARS-CoV-2. (a) General genome organization. (b) Entry mechanism of SARS-CoV-2. (c) Molecular mechanism after the infection. (Lamers et al. 2022).

1.2.1.3 Spike Protein as a potential therapeutic target

Spike glycoprotein is synthesized as a 1273 amino acids polypeptide precursor on the rough endoplasmic reticulum. The unprocessed precursor anchors the signal sequence present on the endoplasmic reticulum, sending Spike protein to the membrane of the rough endoplasmic reticulum. During the synthesis of N-linked, high-mannose oligosaccharide side chains are added. After the synthesis, the trimeric unit of S-protein is formed and transported from the endoplasmic reticulum to Golgi bodies [Duan et al. 2020]. Each protomer contains S1 and S2 subunits. The trimeric topology resembles mushroom-like homotrimers. The S1 subunit consists of four distinct domains, N-terminal domain (NTD), C-terminal domain (CTD), which is also called Receptor Binding Domain (RBD), sub-domain 1 (SD1), and sub-domain 2 (SD2).

The S2 subunit consists of a core of three helices, a transmembrane domain of a single helix, and a short intracellular domain. This subunit possesses S2' cleavage site, which is important for virus entry after cleaving at this site (**Figure 1.3A & B**) [Heald et al. 2012; Millet et al. 2015]. Although the SARS-CoV-2 spike protein shows 76% sequence identity with the spike protein of SARS-CoV-1, the S1/S2 protease cleavage site is different. The furin cleavage site “RRRR” in the spike protein of SARS-CoV-1 is replaced with the “RRAR” sequence in the spike protein of SARS-CoV-2 [Coutard et al. 2020]. Various structural studies show that RBD of spike protein exists in two different conformations, “up” and “down” conformation [Wrapp et al. 2020; Song et al. 2018; Walls et al. 2020; Yuan et al. 2017]. One RBD of one of the protomers stays up, and the other two are maintained in the down conformation. A dynamic equilibrium exists between up and down states [Gui et al. 2017]. Usually, in the prefusion state, all three RBD is in the down state. The fusion of the virus with the host membrane through the S protein is explained in **Figure 1.4**. Only one RBD goes in the up conformation and binds with the ACE2 receptor. The binding promotes

the dissociation from the S1 subunit of the trimeric spike and disturbs the equilibrium of the other two RBD, resulting in shedding off from the complex. Dissociation of the S1 subunit from the complex exposes the S2' cleavage site. This results in an extension of the S2 subunit with fusion peptides into the host cell. Once the fusion peptides get inserted into the host cell, the other regions of the S2 subunit, like HR1(Heptad Repeat 1) and HR2 (Heptad Repeat 2) fold back to interact with each other, which brings the viral and host membrane close to each other to let the membrane fusion happen [Wang et al. 2020].

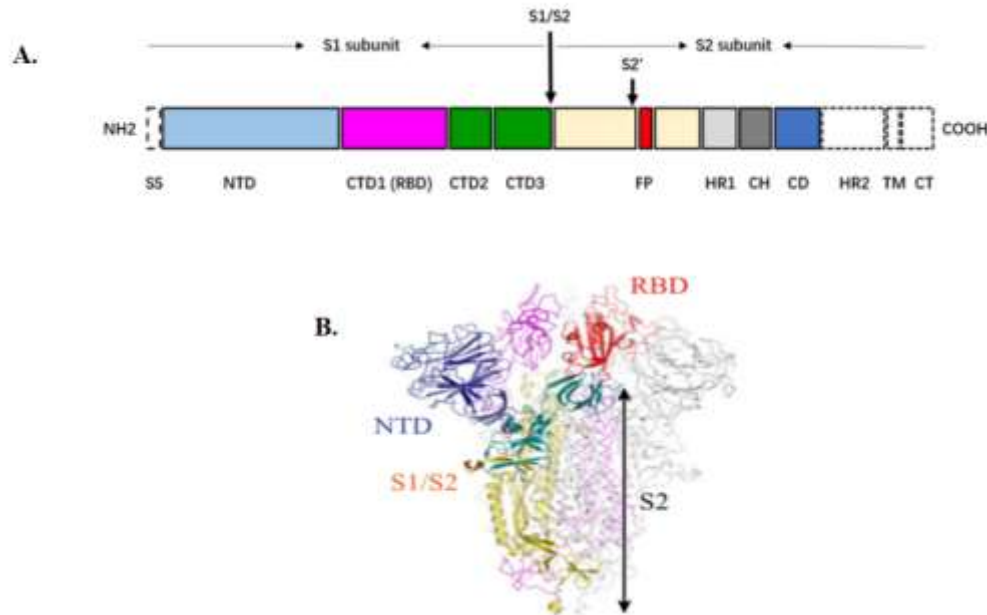


Figure 1.3. (A) Domain organization of Spike protein. (B) 3D structure of S protein [Wang et al. 2020].

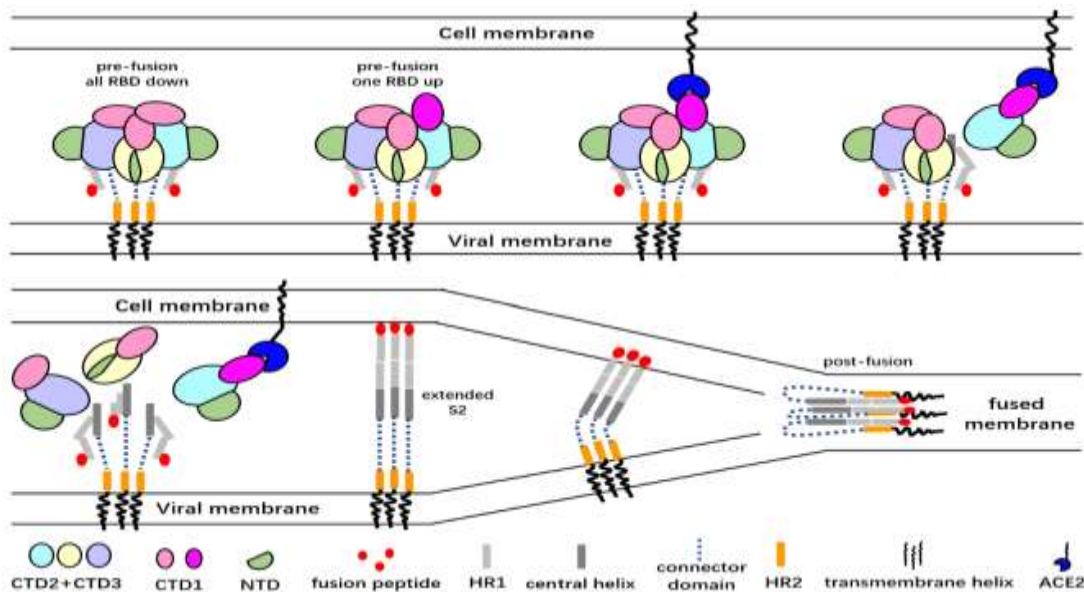


Figure 1.4. Spike protein mechanism from prefusion to postfusion [Adopted from Wang et al. 2020].

1.2.1.3.1 Role of glycosylation sites in the Spike protein

Glycosylation of spike protein has a very essential role in the pathogenesis process. Each protomer consists of 22-N-linked glucans attached to asparagine residue by the host cellular glycosylation apparatus. These N-glycosylations are distributed across the spike protein's two functional subunits, S1 and S2 [Zhao et al. 2021]. The glycans on the spike protein are important moieties to modulate the interaction with the host ACE2 receptor [Zhao et al. 2021]. Moreover, these glycans also play the role of masking the antigenic epitopes from the host-neutralizing antibodies. Glycosylation at the N165 and N234 sites of the spike protein provides an essential structural role in the modulation of RBD's conformation [Casalino et al. 2020]. Mutation of these two residues causes the glycans deletion at their respective sites and reduces the binding of the spike protein

with the ACE2 receptor [Casalino et al. 2020]. A study of these glycans may provide a better perspective on vaccine designs. In addition to N-glycosylation, O-glycosylation is also present in the spike protein. The O-glycosylation is mainly the attachment of *N*-acetylgalactosamine (GalNAc) to the hydroxyl group of serine (Ser, S) or threonine (Thr, T). This modification also plays a role in various biological functions [Bagdonaite et al. 2018].

1.2.1.3.2 Spike protein in vaccine design and therapeutic development

Spike protein facilitates SARS-CoV-2 entry inside the host cell. It is well known virulent factor and has been a primary target for vaccine design efforts. Spike protein has high antigenicity potency and induces robust immune responses. Since the last two years, a large number of vaccine candidates have been in clinical trials [Table 1.1]. Most of the vaccines made against SARS-CoV-2 are primarily based on spike protein. The most popular vaccine made by Pfizer/BioNTech and Moderna is based on the mRNA of spike protein [Mascellino et al. 2021]. The vaccines developed by AstraZeneca-Oxford and Janssen (Johnsons & Johnsons) work on the same mechanism. These vaccines are based on manipulated adenovirus vectors that carry the genetic material for the spike protein in viral DNA which generates specific mRNA. These vaccines are usually associated with the full-length spike protein. In addition, other domains of spike proteins such as RBD, NTD, and S1 domains have been used as vaccines and applied in clinical trials [Arashkia et al. 2020].

Vaccine name	Antigen	Mechanism of antigen
Pfizer-BioNTech	Full-length S protein	mRNA based vaccine
Moderna	Full-length S protein	mRNA
Astra-Zeneca Oxford	Spike protein	Adenovirus viral vector
Janssen(Johnson & Johnson)	Spike protein	Adenovirus viral vector
Sputnik V	Spike glycoprotein	Two adenoviruses viral vector
Novavax	Recombinant full-length Spike protein	Protein subunit
CUREVAC	Prefusion stabilized full-length Spike protein	Modified mRNA

Table: 1.1: List of Principal COVID-19 vaccine and their features

Besides the vaccine, there are also several different types of therapies that target the Spike protein of SARS-CoV-2. 1) Natural product-based therapies: natural therapies like Linoleic acid, Bisoxatin, and Emodin have the potential to inhibit the interaction of S protein and ACE2 [Unni et al. 2020]. 2) Antiviral Based therapies: in this strategy, potential peptides are used to interact with more functional groups of binding sites. The peptides mimic the protein sub-structure and block the native protein-protein interactions. EK1C4 is a fusion inhibitor of coronavirus that targets the Heptad Repeats 1 domain of the Spike protein [Ke et al. 2020]. 3) Recombinant-based therapy: administration of soluble ACE2 in excess amounts that can bind with the Spike protein, distributing the virus binding with the membrane ACE2 receptor. Various human recombinant soluble ACE2 has been developed (rhACE2, APN01), which have an affinity to bind with human ACE2 receptors [Zhang et al. 2020]. 4) Monoclonal neutralizing antibodies: Synthetic neutralizing

monoclonal antibodies such as Casirvimab and Imdevimab inhibit the virus attachment with the ACE2 receptor. These monoclonal antibodies bind with the non-overlapping epitopes of the RBD region of Spike protein [Taylor et al. 2021]. Monoclonal antibody therapy is effective in neutralizing variants of SARS-CoV-2 (Variant of concern), such as B.1.1.7 and B.1.351 variants [Taylor et al. 2021]. 5) TMPRSS2 target-based therapies: TMPRSS is a well-known protease that cleaves the junction of the S1/S2 domain of spike protein and induces the conformational change which favors its binding to the ACE2 receptor. Many inhibitors have been developed for inhibiting the TMPRSS2 that prevent SARS-CoV-2 entry. 6) Small molecule inhibitors: Maraviroc is a reported inhibitor against HIV-1 and has been used to inhibit the S protein fusion in cell culture experiments [Hu et al. 2021]. Various computational and experimental pipeline screening of novel and repurposed drugs have been demonstrated against the binding interface of the RBD-ACE2, furin cleavage site, and fusion. An inhibitor, SSAA09E2, has been reported to inhibit the early interaction of the RBD to hACE2. The SSAA09E1 blocks the CathepsinA, a host protease involved in processing Spike protein. The detailed mechanism of inhibition of SSAA09E2 was studied for SARS-CoV-2 infection using computational-based techniques [Adedji et al. 2021]. Various structure-based drug design processes and supervised MD are being applied to find a suitable drug-like molecule having the potential to bind with the RBD interface and inhibit the interaction process. But still, FDA has not approved any drug against COVID infection. Various repurposed drugs which have shown potency to block spike protein are under research like Camostat: which is used in chronic pancreatitis, it can block ACE2 and TMPRSS2 receptors, Nafamostat: which is a blood thinner used against acute pancreatitis, it can block spike protein from binding to TMPRSS2 and ACE2, Folic acid: it is a well-known source for vitamin B and

reported to bind with furin, an enzyme which converts inactive spike protein to its active form [Santiano et al. 2021].

1.2.2 Non-structural proteins

Non-structural proteins (Nsps) are encoded by genes present within the 5'-region of the viral genome. The 5' -end, consists of replicase gene called as Open Reading Frame 1a and ab (ORF1ab), which encode the nonstructural proteins (nsps) referred to as pp1a and pp1ab polyproteins, respectively. The pp1a non-structural protein corresponds to NSP1 to NSP11 and pp1ab non-structural protein comprises of NSP12 to NSP16. The corresponding functions and features of the Nsps are tabulated in [Table 1.2].

Name	Protein Full name	Function	References
Nsp1	N-terminal product of viral replicase	Inhibiting IFN signaling	Huang et al. 2011
Nsp2	N-terminal product	Binds to prohibitin protein	Cornillez et al. 2009
Nsp3	Papain like proteinase	Promote cytokine expression	Lei et al. 2018
Nsp4	Membrane-spanning protein containing transmembrane domain 2	Viral replication- transcription complex	Sakai et al. 2017
Nsp5	Main proteinase	Cleavage of viral protein	Ullrich et al. 2020
Nsp6	Putative transmembrane domain	Induces the formation of ER-derived autophagosomes	Cottam et al. 2014
Nsp7	RNA-dependent RNA polymerase	Provides clamps for RNA polymerase	Te et al. 2012
Nsp8	Multimeric RNA polymerase	Makes heterodimer with nsp8 & 12	Shi et al. 2020
Nsp9	RNA binding viral protein	Binds to helicase	
Nsp10	Growth factor-like protein with Zinc motifs	Not known	
Nsp11	Consists of 13 amino acids anti-identical to the first segment of nsp12	Not known	
Nsp12	RNA-dependent RNA polymerase	Replication and methylation	Subissi et al. 2014
Nsp13	RNA polymerase proofreading	Replication and transcription	Jang et al. 2020
Nsp14	Exoribonuclease domain	Exoribonuclease activity	Yadav et al.2021
Nsp15	NSP15-A1	Endoribonuclease activity	Yadav et al. 2021
Nsp16	2'O ribose methyl transferase	Inhibit innate immunity regulation	Decroly et al. 2011

Table1.2. Functional role of Nsps1-16 of SARS-CoV-2.

1.2.2.1 Main Protease (M^{pro})

Main protease is a non-structural protein of 33.8kDa size. It is also referred to as 3CLpro, or Nsp5. It is a critical enzyme in the viral replication cycle of SARS-CoV-2. The main protease M^{pro} exclusively cleaves polypeptide, pp1a, and pp1ab, sequences after a glutamine residue and generates 16 non-structural proteins [Ullrich et al. 2020]. It also autoclaves itself between nsp4 and nsp6 before the polyprotein process. There are no closely related homologs in humans, and it is considered a most attractive therapeutic target. SARS-CoV-2 M^{pro} exists as a biologically active dimer. However, it is closely related to the Main proteases of other coronaviruses. It shares 99% sequence identity with BatCoV Mpro and 96% identity with SARS-CoV-1, while it shares only 50% identity with MERS-CoV [Ullrich et al. 2020].

Structurally, the main protease of SARS-CoV-2 is almost identical and similar to the previous Main proteases of another coronavirus [Cui et al., 2020]. The 3-D structure of a protomer contains three domains (**Figure 1.5**). Domain I and II consist of antiparallel β -barrel structural arrangement, and the substrate binding site is located in a cleft between the two domains. Domain III consists of five α -helices arranged in an antiparallel globular cluster. The dimeric interface is formed due to the interaction between domain II of one protomer and the N-terminal residues of the other (Reference). The active catalytic site is situated between domains I and II (**Figure 1.5**). The M^{pro} is a cysteine protease with a catalytic dyad consisting of conserved cysteine and histidine residues in its active site. In contrast to other cysteine and serine proteases, the third residue at the catalytic site of M^{pro} occupies a buried water molecule. The M^{pro} follows a multi-step mechanism for catalysis.

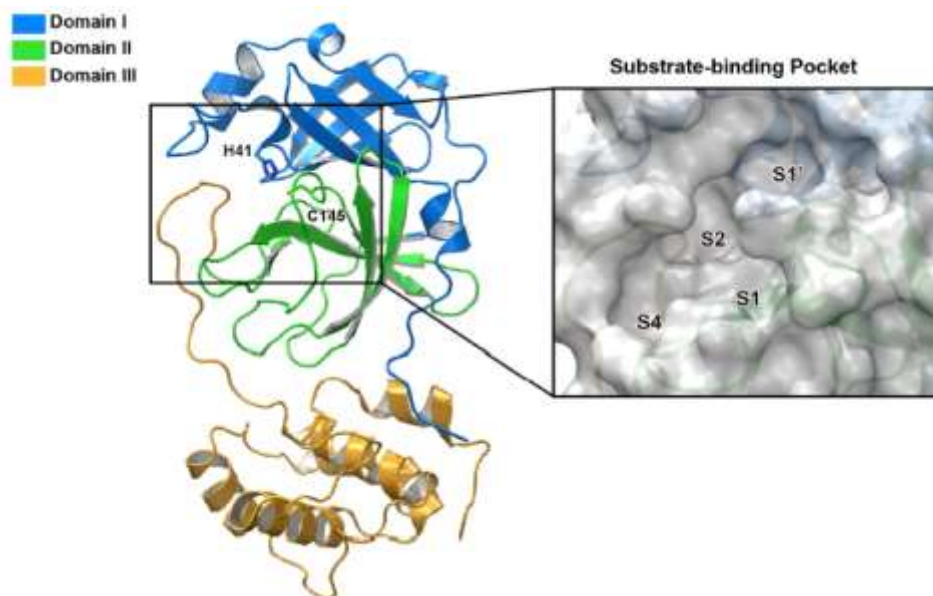


Figure 1.5. The 3-D structural arrangement of SARS-CoV-2 M^{pro} in its monomeric form [Astuti et al. 2020].

Main proteases identify the substrate residues that efficiently bind to the S4 to S1 cavity. The S1, S2, and S1' sites determine the specificity of substrates. Glutamine residue at the S1 has the highest degree of conservation in the cleavage site of MERS, SARS-CoV-1, and SARS-CoV-2 [Schechter et al. 1967]. Hydrophobic residues are preferred at the S2 site, while the S1' site tolerates serine or alanine. Several compounds have been identified that can inhibit the main protease's activity. Various covalent and non-covalent inhibitors (like N3, α -ketoamide, etc.) have been identified (**Figure 1.6**). Many crystal structures in complex with the bound inhibitors have been solved, and the inhibition mechanism has been deduced [Cui et al. 2020].

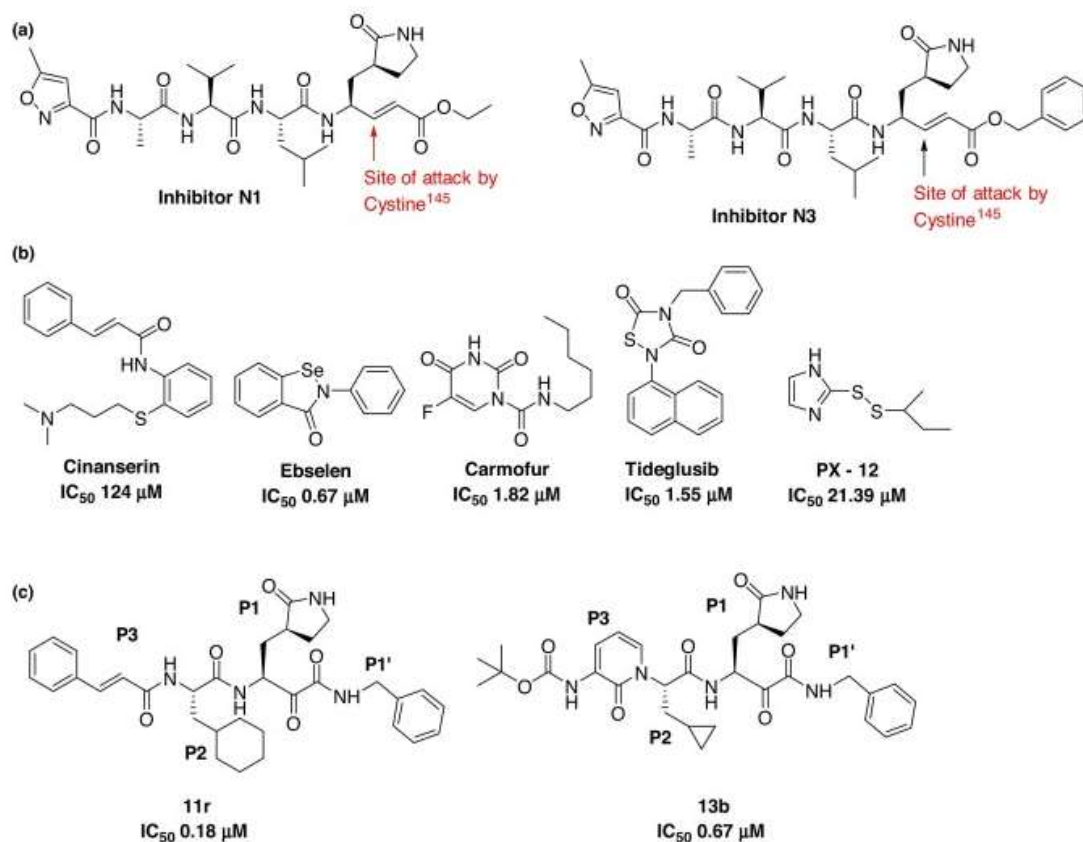


Figure 1.6. List of M^{pro} identified inhibitors [Adopted from Banerjee et al, 2021].

1.3. Variants of SARS-CoV-2

Since December 2020, many mutations have been detected in the SARS-CoV-2 genome, mainly in the spike protein, which accumulates the highest number of mutations. WHO has put these variants in two groups 1) Variants of Concern (VOC) and Variants of Interest (VOI) based on their potential risk to human health [Gomez et al. 2021]. At the start, three VOCs have gained much interest: B.1.1.7 (UK variant), also called alpha variant (first reported in the UK), B.1.351, also called Beta variant (South Africa variant), and P.1 variant, also called Gamma variant (Japan/Brazil). The list of mutations in all VOC has been listed in **(Table 1.3)**. These variants were reported to have higher transmission rates with increased disease severity and reinfection rates and

become dominant within the population. Most importantly, they have the potential to neutralize the antibody generated from the vaccinations and the convalescence sera. In all, two VOC mutations were in common and reported to be critical mutations (N501Y & D614G). The mutations in the RBD region of spike protein are of great interest as they may directly affect the binding of ACE2 protein. The P.1 variant, also known as the Gamma variant, was first detected in Japan and later identified in Brazil in January 2021. The P.1 variant exhibits 12 mutations in the S protein, which are implicated in increased transmission, severity, immune evasiveness, and reinfection. Since the P.1 variant accumulates many mutations in spike protein, it is more resistant to antibody protection than the B.1.351 variant. E484K mutation in B.1.351 and P.1 can escape the immune invasion. Studies have shown that this mutation is critical to monitor because of decreased vaccine-induced antibody neutralization titers against variants carrying the E484K substitution [Gomez et al. 2021].

Variants	B.1.1.7	B.1.351	P.1
1st detection	September 2020	8 October 2020	2 January 2021
Detection site	United Kingdom	South Africa	Japan/Brazil
Mutations in S protein	7 mutations: N501Y, A570D, D614G, P681H, T716I, S982A, D1118H 2 deletions: H69-V70del, Y144del	9 mutations: L18F, D80A, D215G, R246I, K417N, E484K, N501Y, D614G, A701V 1 deletion: LAL 242-244 del	12 mutations: L18F, T20N, P26S, D138Y, R190S, K417T, E484K, N501Y, D614G, H655Y, T1027I, V1176F
Countries reported cases	82	40	19
Countries with sequences	64	35	14

Table 1.3. Early variants of SARS CoV-2 and its list of mutations [Adopted from Gomez et al. 2021].

1.4 *Mycobacterium tuberculosis* and its pathogenesis

Mycobacterium tuberculosis (*Mtb*) is the causative agent of tuberculosis, which a major threat to human health. Before COVID infection, it was the leading cause of death, ranking above HIV. It spreads when infected people expel the bacteria by coughing or sneezing. The primary site of infection is the lung but can also affect the other targets. Its primary symptoms are a heavy cough, chest pain, and bloody sputum in the cough, while other symptoms are also observed, like weakness, weight loss, loss of appetite, chill, fever, or sweating at night. TB can be cured with six to nine months of regimen treatment. Every year the World Health Organization (WHO) publishes global TB reports. According to the 2021 TB report, cases of TB have declined, possibly because the infected targets are off-track from the population (WHO report 2021). The ratio has fallen from 7.1 million in 2019 to 5.8million in 2020. But inversely, the estimates for the TB death rate for 2020 have increased due to poor diagnosis and treatment. Globally in 2020, 1.3 million deaths were reported in HIV-infected patients compared to 1.2 million in 2019. Global TB status shows countries with a minimum of 100 000 cases from South-East Asia & Africa (**Figure 1.7**). The 30 high TB-burden countries accounted for 86% of all estimated incident cases worldwide. Eight of these countries responsible for two-thirds of the global total: India (26%), China (8.5%), Indonesia (8.4%), Philippines (6.0%), Pakistan (5.8%), Nigeria (4.6%), Bangladesh (3.6%) and South Africa (3.3%) (WHO report 2021).

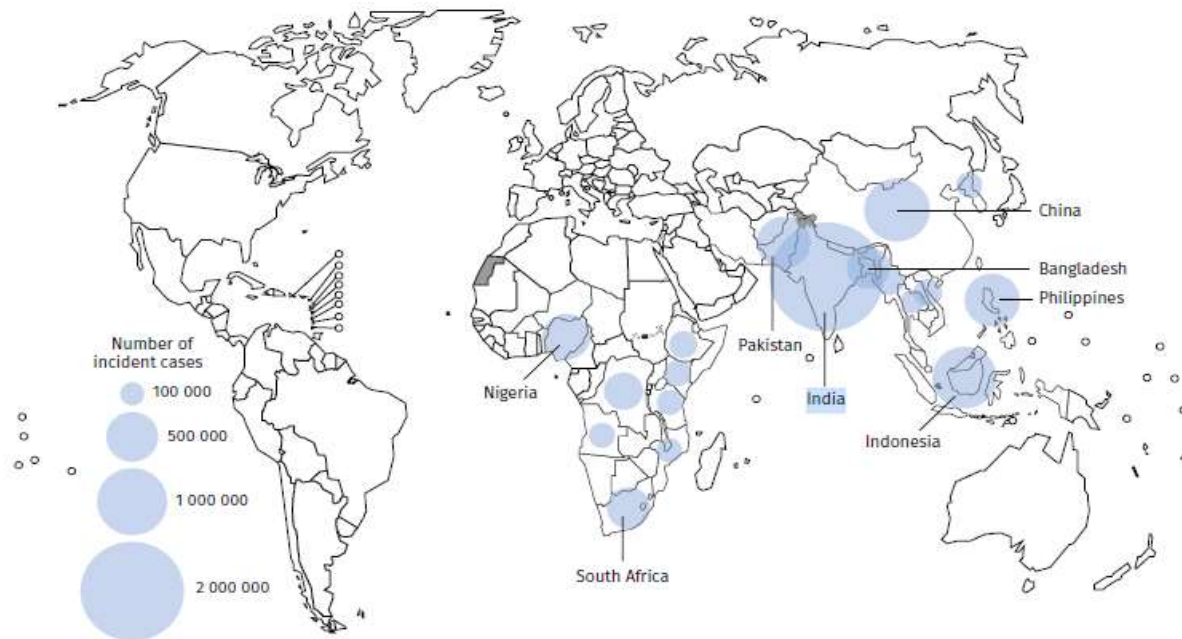


Figure 1.7. WHO TB report for estimated cases in 2020 reported for countries with a minimum of 100,000 cases [Global tuberculosis report, 2021].

1.4.1 Events in the disease process

TB infection occurs when a person inhales a droplet containing tubercle bacilli that reach the vicinity of the lungs' alveoli [Natarajan et al. 2020]. Further, host alveolar macrophages ingest these tubercle bacilli and destroy them. But, in a few cases, bacilli survive and multiply and cause an active infection. The active infection also spreads to other tissue through lymphatic channels. There are two states of *Mtb* infection- active TB infection or pulmonary TB infection and Latent TB infection (**Figure 1.8**). Less than 10% of the infected individuals actually develop the active disease; and in >90% of infected individuals, *Mtb* persists within infected macrophages for a long period of time in a metabolically inactive and non-transmissible form but reversible state known as “dormancy” or latent TB infection (LTBI) (WHO, 2009). However, reactivation of latent *Mtb*

either due to malnutrition, immunosuppression, steroid use; TNF therapy, or HIV infection account for the majority of adult pulmonary TB cases.

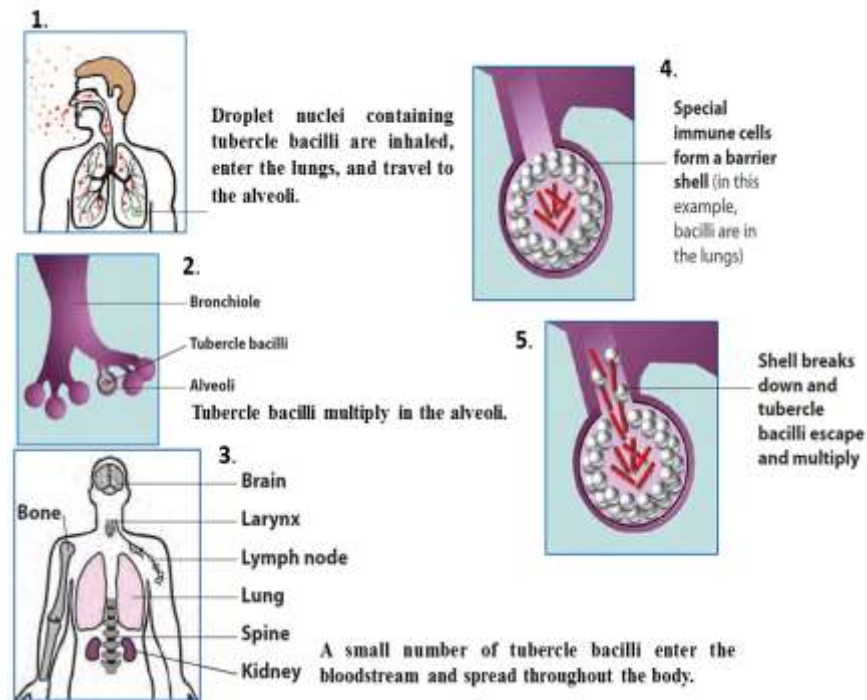


Figure 1.8. Events in the Tuberculosis disease process (Adopted from transmission and pathogenesis of tuberculosis, Book chapter).

In LTBI states, the infected persons don't spread the infection. The LTBI initiates when the macrophages ingest extracellular bacilli, which triggers the immune response, leading to granuloma formation. The latently infected individuals indeed constitute a major impediment to TB control efforts. The major problems about latent TB are i) lack of proper diagnosis, ii) unavailability of efficacious vaccines, iii) nexus with HIV, and the emergence of MDR and XDR strains, further aggravating the problem. In a compromised immune system, the tubercle bacilli in some people relapse from the latent states and multiply inside the host cell, resulting in progression

from LTBI to active TB state (**Figure 1.9**). Such persons are capable of spreading the infection to other people.

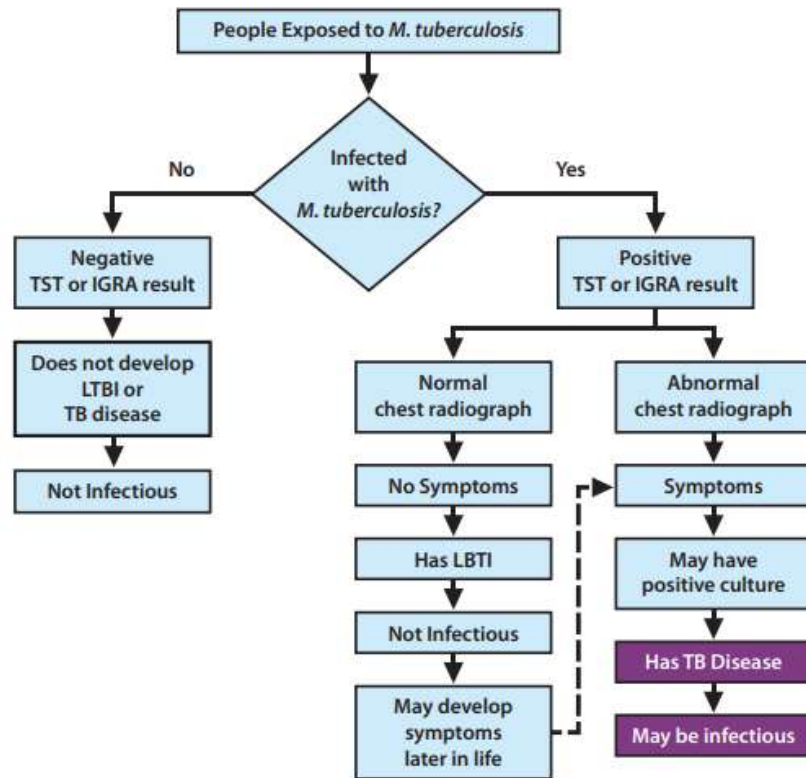


Figure 1.9. Progression of LTBI or active TB.

1.4.2 Treatments so far and the challenges

Mtb is curable after taking a combination of first-line drugs for 6-9 months [Miggiano et al. 2020]. It includes isoniazid, rifampicin, streptomycin, etc. But *Mtb* may become resistant to these first lines of drugs, and such strains are called MDR (Multi-drug Resistant) strains. In MDR, TB patients are given a combination of second-line drugs (like amikacin, kanamycin, etc.). The treatment takes a much longer time, and it is costly also [Mogashoa et al. 2019]. Challenges are imposed due to the development of extensively drug-resistant (EDR) TB strains. These strains are even resistant to second-line drugs, and the therapy becomes more complicated [Maslov et al.

2020]. Besides drug treatment, DOTS (Directly Observed Treatments, Short) therapy is an effective strategy for controlling *Mtb* infection [Karumbi et al. 2015]. It requires political commitments like systematic reporting and recording systems to assess the treatment results of patients. Moreover, the drugs offer standard care for active TB, but their effectiveness against latent TB is doubtful. First and second-line drugs being used for TB treatment and their respective targets are listed in (Table 1.4 and Table 1.5). The current protective TB vaccine – a live-attenuated *Mycobacterium bovis* Bacillus-Calmette-Guerin (BCG) vaccine - is known to protect against severe forms of TB in young children and leprosy. However, it is not efficiently and consistently protective against pulmonary TB in adults; and from reactivation of the latent TB infection. Currently, most of the vaccines under clinical trial are aimed to replace the BCG, like rBCG30 which is a genetically modified BCG strain. In addition, many subunit vaccines like MVA85A, Crucell AD35, and other protein-in-adjuvant vaccines under clinical trials, are being used as boosters of BCG vaccines.

Drugs	Chemical class	Targets
Isoniazid (INH)	Isonicotinic acid	Enoyl-ACP reductase, mycolic acid elongation
Rifampicin (RIF)	Rifamycin	DNA-primed RNA polymerase
Pyrazinamide (PZA)	Pyrazine	Fatty acid biosynthesis
Ethambutol (EMB)	Ethylenediamine	Cell wall arabinan deposition
Streptomycin	Aminoglycosides	Protein synthesis inhibition

Table 1.4. List of first-line drugs against TB.

Second Line Drugs	Class	Targets
Amikacin (AMK)	Aminoglycosides	Misreading of mRNA and inhibition of synthesis of proteins
Kanamycin (KM)	Aminoglycosides	Mistranslation during protein synthesis.
para-Aminosalicylic acid (PAS)	Aminophenol	Inhibition of folic acid biosynthesis.
Cycloserine (CYS)	D-alanine	Inhibition of bacterial cell wall synthesis.
Ethionamide (ETA)	Thioamide	Inhibition of mycolic acid synthesis by binding to ACP Reductase
Capreomycin (CPR)	Polypeptides	Inhibition of protein synthesis.
Ciprofloxacin	Fluoroquinolones	Interactions with DNA gyrase, DNA replication and Transcription

Table 1.5. List of second-line drugs.

Besides first and second-line drugs, other drugs like Moxifloxacin, Levofloxacin, and Linezolid have proven to be effective in the early phase of MDR-*Mtb* treatment. Bedaquiline and Delamanid have been approved as the best drug that enhances the treatment options for MDR-TB [Kwon et al. 2014]. Other new chemicals like AZD5847, SQ109, and PA824 also have also been shown potency against the *Mtb* [Kwon et al. 2014] and are in different stages of clinical trials. Many of the drugs so far designed to control crucial processes of *Mtb*. DNA replication, RNA synthesis, Cell wall synthesis, protein synthesis, energy metabolism, and folate metabolism are potential drug targets that have been exploited in TB drug discovery. In spite of this development in drug discovery, *Mtb* possesses a great threat because of the emergence of MDR and XDR strains. In order to avoid antibiotic onslaughts, *Mtb* rewires its metabolic gene expression and overexpresses many virulent factors that are involved in the pathogenesis and help the bacteria to survive.

1.4.3 Drug resistance in *Mtb*

Drug resistance in TB makes the bacteria resistant to the most commonly used anti-TB drugs. Resistance against Rifampicin arises due to mutations in the *rpoB* gene. The mutation brings a conformational change in RNA polymerase that decreases its affinity for the drugs (Telenti et al. 1993). Isoniazid is a well-known prodrug converted to an active drug by the action of the peroxidase enzyme, KatG [Zhang et al. 1992]. Mutations in genes like *katG*, *inhA*, *ahpC*, *kasA*, and NADH dehydrogenase cannot convert a prodrug to an active drug [Rawat et al. 2003]. Similarly, ethambutol serves as a bacteriostatic by interfering with the biosynthesis of arabinogalactan in the cell wall [Takayama et al. 1989]. Mutations in the gene *embCAB* that encodes arabinogalactan biosynthesis enzymes provide resistance to ethambutol. The nicotinamidase enzyme coded by the *pncA* gene converts Pyrazinamide prodrug to pyrazinoic acid, inhibiting the bacilli membrane transport [Konno et al. 1967]. Mutations in the *pncA* gene may make bacilli resistant to this drug. Streptomycin inhibits the initiation of protein translation of protein synthesis. Mutations in *rpsL* and *rrs* genes cause a high level of resistance to streptomycin [Crofton et al. 1948]. Chromosomal mutations in *gyrA* or *gyrB* make the bacteria quinolone-resistant [Fabrega et al. 2009]. Many second-line anti-TB-drugs inhibit protein synthesis, and mutations at positions 1400 and 1401 of the *rrs* gene confer resistance to kanamycin and amikacin [Palomino et al. 2014]. The involvement of other mechanisms, like the efflux system, which extrudes drugs to the cell's exterior, also provides resistance towards anti-tubercle drug therapy [Von et al. 2009]. Modifications in the known targets due to mutations provide resistance and generate resistant strains of *Mtb*. Hence, exploring new targets and designing inhibitors or drugs against the new target is always warranted.

1.4.4 Antioxidant system

An antioxidant system works to protect the cellular components from the pro-oxidants. Microorganisms are always exposed to reactive oxygen species such as oxidizing radicals and other free radicals. To cope with the damaging effect of the radicals, microorganisms developed both enzymatic and non-enzymatic anti-oxidant systems to protect themselves from the oxidizing environment. *Mtb* resides in the host macrophages. Macrophages kill bacilli by producing reactive oxidizing species. *Mtb* survives in the macrophage's hostile environment by producing its antioxidant systems. Generally in other microorganisms and *E. coli*, glutathione, NADPH, and ascorbic acid are non-enzymatic intracellular antioxidant molecules. The glutathione system constitutes a well-known enzymatic antioxidant system, which includes glutathione, glutathione reductase, and glutaredoxins (Holmgren et al. 1995). Glutaredoxins are low molecular weight enzymes which is characterized by the thiol active site (Cys-Pro-Tyr-Cys). It belongs to the oxidoreductase family and protects the proteins from oxidative stress (Holmgren et al. 2000). Other enzymatic antioxidants include superoxide dismutase (SOD), catalase, glutathione peroxidase, and peroxiredoxins. Catalase found in peroxisomes protects the cell in an aerobic environment. It catalyzes hydrogen peroxide into water and oxygen (Kirkman et al. 1999). *Mtb* lacks the glutathione system rather it contains mycothiol system, thioredoxin system, disulfide oxidoreductase, catalase-peroxidase, superoxide dismutase, and alkyl hydroperoxide reductase as antioxidant molecules. Mycothiol (MSH) is a low molecular weight thiol that has an important role in maintaining the reducing environment inside the bacterial cell. It protects the cell from the cellular oxidant and also detoxifies the thiol-reactive compounds (Newton et al. 1995). It is usually used as a cofactor or a substrate of MSH-dependent enzymes. Mycothiol disulfide reductase (Mtr) is a key enzyme that maintains reduced MSH levels (Patel et al. 1999). Mycothiol disulfide

reductase is validated as an essential gene for *Mtb* growth. In addition to mycothiol, *Mtb* also possesses a thioredoxin system that plays a key role in maintaining a reducing environment and protecting the cellular proteins from oxidative damage.

1.4.5 Thioredoxin system

The Thioredoxin system in *Mtb* consists of two thioredoxins (Trx), thioredoxin reductase (TrxR), and NADPH. Thioredoxin is a superior thiol reductant that reduces the disulfide bonds [Trivedi et al. 2011]. Trx is a small protein of 12kda in molecular weight. It maintains the reducing environment inside the host cell. Thioredoxin has been isolated from both prokaryotic and eukaryotic cells. It was first purified from *E.coli* [Laurent et al. 1964]. It helps in catalyzing the reduction of ribonucleotides to deoxyribonucleotides. Despite high diversity, the active site of thioredoxin is conserved and consists of Cys-Gly-Pro-Cys. The 3-D structure possesses five β -strands surrounded by four α -helices. The reduced form of Trx performs diverse functions such as disulfide reduction, sulfur assimilation, detoxification of ROS, protein repair, redox regulation, transcription factor, and regulatory role [Laurent et al. 1964].

The other member of the *Mtb* thioredoxin system is Thioredoxin reductase (TrxR). Only a single copy of TrxR is reported in *Mtb*. *Mtb* TrxR reduces peroxides and nitrobenzene (Jaeger et al., 2004; Zhang et al., 1999). TrxR utilizes electron potential from cellular NADPH and reduces to Trx. As reported earlier for *E. coli* and other species, reduced Trxs are essential for nucleotide biosynthesis as they donate the reducing equivalents to ribonucleotide reductase. The transfer mechanism of reducing equivalents from NADPH to the target protein is well characterized in *E. coli*. The NADPH reduces TrxR, and TrxR reduces Trxs, which are available for reducing target proteins in the cell (**Figure 1.10**).

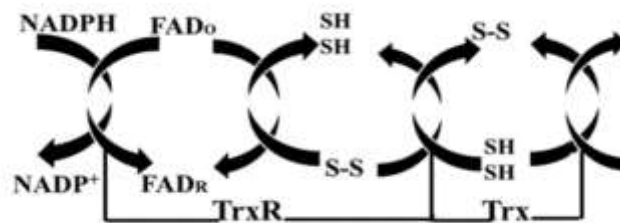


Figure 1.10. Mechanism of electron transfer through the thioredoxin system

Transposon-mediated analysis has validated *Mtb* TrxR as an essential gene [Zhang et al., 2012]. Moreover, it has been reported that deletion of the *trxR* gene results in a hyper-susceptible strain. TrxR is also essential for the growth of other organisms, such as *S. aureus* and *B. subtilis* [Uziel et al. 2004; Scharf et al. 1998]. Thus, considering its crucial role, the bacterial TrxR has been demonstrated as a promising drug target [Lu et al., 2013]. TrxR is of two types (1) low molecular weight (low MW), 35kDa TrxR, occurring in prokaryotes, fungi, and plants, and (2) High molecular weight (high MW), 55kDa TrxR, identified in mammals, *Caenorhabditis elegans*, *Drosophila melanogaster*, and *Plasmodium falciparum*. In contrast to low MW TrxR, High MW TrxR contains extra C-terminal peripheral redox-active catalytic sites (Becker et al. 2000). Mammalian TrxR possesses a Cys-Sec (Cysteine-Selenocysteine) sequence at the C-terminal redox pair (Becker et al. 2000), which has broad substrate specificity that also reduces non-disulfide substrate such as hydroperoxides, vitamin C or selenite. *Mtb* TrxR exists as a dimer in solution, and each protomer is 36kDa. The crystal structure also shows the presence of two subunits. Each subunit has two domains, NADPH and FAD, connected by a flexible hinge (**Figure 1.11**) (Akif et al. 2008). The CXXC motif constitutes the active site present in the NADPH domain. The NADPH binding site is 15 Å away from the FAD binding site and active site CXXC motif.

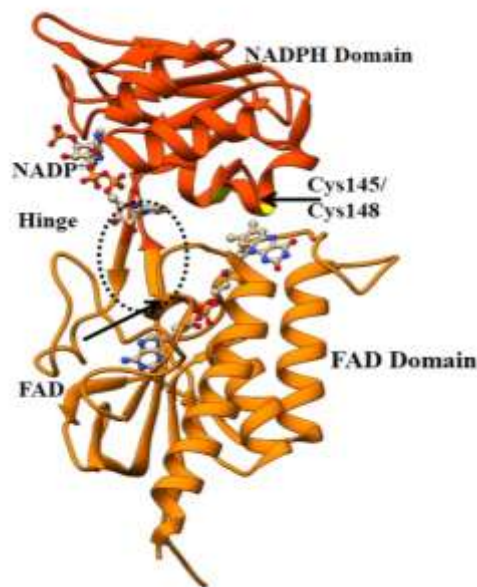


Figure 1.11. Structure of *Mtb* TrxR

To transfer an electron from NADPH to FAD, the NADPH domain takes 66° rotation with respect to the FAD domain across the flexible hinge region that brings the NADPH ring close to the FAD domain. This conformational change is necessary to bring the buried active site to the surface of the enzyme and make it available for Trx reduction. With the availability of high-resolution crystal structures of *E. coli* TrxR as the stand-alone structure (FO conformation) and in complex with its substrate Trx and cofactor analog (FR conformation), a detailed understanding of the structural basis of the reaction mechanism has already been demonstrated. The conformational flexibility of TrxR and its cofactors are the key to its activity. The Trx system has multiple roles in the cell. Thus, inhibiting TrxR will likely impact various functions necessary for *Mtb* survival and proliferation.

1.5 AIM OF THE THESIS

Protein molecules are the best therapeutic agent as they have been involved in various virulence factors. But it requires detailed molecular knowledge of protein structure which is responsible for the specific function of protein molecules. The primary step in characterizing proteins is the availability of their three-dimensional structure either by crystallization process or homology modeling. Proteins undergo conformational changes to perform a particular function. These changes can be initiated due to substrate binding, drug binding, or domain-domain interaction. The Structure-Function Paradigm plays a central role in structural biology. An understanding of the structure-function relationship provides the molecular basis of conformational changes. It is helpful to design a drug molecule that can have inhibitory effects on the protein target. There are cases where this structure-function paradigm breaks down, meaning there is a significant change in protein function due to mutations or inhibitor binding without any substantial changes at the structural level. Hence, it is necessary to probe minute changes at the contact level, which are not easily recognizable at the structural level. This will help identify the hotspot residues that can be targeted as druggable targets.

This thesis aims to identify the negligible changes responsible for a change in function without having any significant change at the structural level. This particular case of the structure-function paradigm has been applied to three different therapeutic targets from two phylogenetically diverse pathogenic systems. One is from the viral system, i.e., spike protein and main protease from SARS-CoV-2; another target is from the bacterial system, i.e., Thioredoxin Reductase from *Mtb*.

The SARS-CoV-2 pandemic is more challenging because of its emerging variants. Every new variant imposes a great challenge to therapeutic development and vaccine design. The variants usually evolved to survive by modifying their genome. Most of the mutations were reported to be

present in the Spike protein, the primary target of infection. This makes spike protein a primary and topical interest target in this pandemic. In the early pandemic, three Variants of Concerns have gained more interest (alpha, beta, and gamma). The P.1 variant had the highest number of spike protein mutations in all three variants. Experimental studies suggest that the binding affinity of the P.1 variant towards ACE2 is more as compared to the WT, and other VOCs. Previous studies suggest that mutations in RBD and the sites outside the RBD (like the S1 domain) of spike protein are responsible for enhanced binding affinity with the ACE2 receptor. Surprisingly many of the mutations do not induce any significant variations at the structural level compared to its wild-type protein. However, it is known that mutations at critical functional sites bring changes at the structural level in the protein compared to its wild-type structure. Previously studies have investigated the effect of mutations on spike protein. But the impact of mutations on a significant change in function in terms of its binding without losing its structural integrity was not probed at the molecular level for the P.1 variant. Additionally, the dynamic analysis of residue fluctuations involved during the folding and unfolding of native protein structure was not elucidated for the P.1 variant. Analysis of subtle but functionally significant variations in protein structures can address the molecular mechanism of cumulative effects of the mutations on rearrangement in the protein contact networks and the dynamic variability in the spike protein. So it would be a matter of interest and concern to uncover the structural basis of spike protein variants which is responsible for increasing affinity towards its host ACE2 receptor and can pave the way for designing better therapeutics against the SARS CoV-2 variants.

Though many inhibitors have been designed to disturb the RBD-ACE2 complex. However, a limited study is reported for the P.1 variant of SARS-CoV-2. After identifying the structural basis of the enhanced binding of the P.1 spike protein with the ACE2, the subsequent study aims to

design potential small molecule inhibitors that can disturb the RBD-ACE2 complex of the P.1 variant. SSAA09E2 a known inhibitor against SARS-CoV-1 which blocks the early interaction of RBD of spike protein with ACE2. This molecule showed a lesser binding with the interface of the RBD-ACE2 complex of the P.1 variant as compared to the wild type complex. This chapter also aims to modify the SSAA09E2 using different computational approaches to make it a potent inhibitor with a better binding affinity towards the RBD-ACE2 complex from P.1 variant.

In the subsequent chapter, the breakdown of the structure-function-paradigm was investigated for the Main protease, M^{pro}, which is an important therapeutic target from SARS-CoV-2. It performs the polyprotein processing encoded from viral RNA. It has a sequence identity of 96% with M^{pro} from SARS-CoV-1 and is structurally conserved also. With this conservation, the inhibitors effective for SARS-CoV-1 must also show the same inhibitory effect for SARS-CoV-2 M^{pro}. However, the inhibitors were having differential inhibitory effects on SARS-CoV-2 M^{pro}. In spite of differences in a few amino acids among the two M^{pros}, the overall three-dimensional crystal structure was very similar. Then the question arises why the inhibitors of SAR-CoV-1 M^{pro} have different inhibitory efficiency towards SARS-CoV-2 M^{pro}? Do they have differences in the structural parameters? Hence, the matter of interest in this study is to identify the minute conformational changes at the contacts level which are not detectable at the structural level between SARS-CoV-1 and SARS-CoV-2 M^{pro} in its apo and inhibitor bound states that may guide towards specific drug development against SARS-CoV-2 M^{pro}.

In chapter 5, the case study of Thioredoxin reductase from *Mycobacterium tuberculosis* was explored, which helps *Mtb* to survive in the harsh environment of host macrophage cells. A structural study of this target demonstrates that TrxR exists in two conformational states, open (oxidized) and closed (reduced). Both conformations consist of the NADPH domain and the FAD

domain. TrxR reduces Trx by utilizing the reducing potential from the NADPH via the FAD. But in the oxidized state, the distance between the NADPH and FAD is 15Å, which is too far for an electron transfer mechanism. Hence, the NADPH domain rotates 66° with respect to the FAD domain to bring the NADPH to FAD. This rotation is necessary to bring the buried active site to the enzyme's surface. The reduced active site then be available to reduce Trx. The conformational flexibility of TrxR is provided due to the hinge region, which connects the NADPH and FAD domains. Since the conformational rotation of the NADPH domain is a critical step in the catalytic activity of TrxR. So, the hypothesis of this study is to restrict this conformational change of the NADPH domain by designing a novel compound. To study the effect of screened compounds on the protein's structural dynamics. To analyze the inhibitory mechanism of the compound on the oxidized form of TrxR.

CHAPTER-2

Investigation of protein structural parameters and dynamics features of Spike protein

THIS CHAPTER IS PUBLISHED IN:

Lata, S., & Akif, M. (2022). Probing structural basis for enhanced binding of SARS-CoV-2 P.1 variant spike protein with the human ACE2 receptor. *Journal of cellular biochemistry*, 123(7), 1207–1221.

2.1 INTRODUCTION

Coronaviruses get entry inside the host cell with the help of spike glycoprotein present on the outer envelope through a host cellular receptor, Angiotensin Converting Enzyme 2 (ACE2) [Gui et al. 2017; Song et al. 2018]. Spike protein has been one of the important therapeutic targets for drug design and vaccine development. It is a homo-trimeric protein consisting of S1 and S2 subunits, the S1 subunit aids in the binding to the ACE2 receptor and the S2 subunit is involved in the fusion of virus to the host cellular membrane [Wang et al., 2020a]. The structural organization of the S1 domain includes N-terminal domain (NTD), C-terminal domain (CTD), sub-domain 1 (SD1) and sub-domain 2 (SD2). The CTD consists of Receptor Binding Domain (RBD) which is a main component that interacts with the ACE2 receptor. Previously, several studies have demonstrated that virus replication, infectivity, and transmission ability depend on the binding affinities of the RBD to the ACE2 receptor [Lan et al., 2020; Wang et al., 2020a].

COVID-19 pandemic has been more challenging due to SARS-CoV-2 emerging variants. Since December 2020, many variants have been detected with a high number of mutations mainly in the spike protein [Gómez et al., 2021]. The world health organization (WHO) has classified these variants either as Variants of Concern (VOC) or Variants of Interest (VOI), where three variants have gained more attention: B.1.1.7 (VOC 202012), B.1.351 (501Y.V2) and P.1 (B.1.1.28.1) [Buss et al., 2021; Fiorentini et al., 2021; Tegally et al., 2021]. The B.1.1.7 was first detected in the UK in September 2020 and has been reported with 7 mutations in the Spike protein [Davies et al., 2021; Gómez et al., 2021]. Within a month a South African variant was detected with 9 mutations [Gómez et al., 2021; Tegally et al., 2021]. Later on 2 January 2021 Japan/Brazil variant named P.1 (B.1.1.28.1) was in circulation with 12 mutations in the Spike protein [Gómez et al.,

2021; Sabino et al., 2021]. The two mutations N501Y and D614G in spike protein were common in all the three VOC. Very recently, Omicron has also been identified as a VOC and reported to have almost 30 mutations in the spike protein [Kannan et al., 2022]. The functional role of many mutations have been demonstrated using deep mutational scanning and other biochemical methods. Most of the mutations associated with the variants have been reported to have enhanced affinity towards the ACE2 receptor as well as are also crucial for the virus transmission [Chan et al., 2020; Starr et al., 2020]. These mutations are also demonstrated to be ineffective towards neutralizing antibodies and evading from neutralization [Dejnirattisai et al., 2021; Wibmer et al., 2021; Xie et al., 2021]. The underlying cause of the enhanced binding affinity in the variants has not been associated with significant change in three dimensional structures of spike as well as ACE2 receptor. However, local and global rearrangement of the interaction site and allosteric site cannot be overlooked. Mutations away from the RBD facilitate spike protein in the open state to bind efficiently with the ACE2 receptor [Ray et al., 2021]. Most of the previous studies focus on the effect of individual mutation. However, the cumulative effect of the mutations present on the RBD and away from the RBD such as NTD, SD1 and SD2 of the S1 domain was not investigated in detail. It is assumed that the cumulative effect of the mutations may provide a significant rearrangement in the protein contact networks that may have an effect on increased binding affinity with the ACE2 receptor.

Protein contact network (PCN) and molecular dynamic simulation (MDS) have been very useful approaches for analyzing nominal conformational changes crucial for large conformational effects on the function and inhibitor binding of proteins [Srivastava et al., 2020]. These methods have already been used in various studies such as understanding the dynamics of mutant proteins

[Kandhari and Sinha, 2017; Srivastava and Sinha, 2014], structural flexibility [Jacobs et al., 2001] structure-function relationship [Srivastava and Sinha, 2017] identifying crucial functional residues in proteins [Amitai et al., 2004]. Recently, the PCN approach has also been applied to study a structural comparison of SARS-CoV-1 and SARS-CoV-2 M^{pro} in apo and inhibitor bound states [Lata and Akif, 2021]. These approaches have been important for analyzing evolution as well conformational dynamics of SARS-CoV-2 spike protein [Verkhivker, 2020a] and prediction of its binding affinity with ACE2 receptor [Laurini et al., 2020; Wang et al., 2020b]. Recently, the PCN along with perturbation scanning based elastic network methods provided evidence of the allosteric sites and their involvement in the regulation of functional dynamics of the spike protein [Di Paola et al., 2020; Ray et al., 2021; Verkhivker et al., 2021; Verkhivker and Di Paola, 2021]. There have been many integrative computational efforts towards better understanding of SARS-CoV-2 and its variants interaction with the ACE2 [Khan et al., 2021; Periwai et al., 2021; Teruel et al., 2021; Verkhivker, 2020b]. Moreover, efforts are also being put into designing inhibitors focusing on the RBD domain of spike protein [Khan et al., 2021; Periwai et al., 2021; Razizadeh et al., 2021]. However, the mutations associated in the spike protein render the current therapeutics a challenge to combat SARS-CoV-2 and possibly be ineffective to other similar future pandemics. It is therefore really warranted to have an insight into the atomic details and effect of these mutations on the structural communication and dynamics of the RBD and truncated S1 domain of the spike protein in context to its interaction with the ACE2 receptor.

In order to probe the effect of all mutations present on the RBD as well as mutations from other part of the truncated S1 domain, we report, here, a comparative analysis on the structural communication and dynamics of the RBD and truncated S1 domain of spike proteins from the SARS-CoV-2 wild type and P.1 variant. However, we emphasize a limitation of our study, we

emphasize a limitation of our study, we have not considered the full length of spike protein. Since our aim was to understand the cumulative effect of mutations on the RBD as well as truncated S1 domain, assuming that this will not affect our comparison with the available experimental data for the P.1 variants, as the experimental studies were also performed mostly with the same domains [Dejnirattisai et al., 2021]. We applied an integrative network and dynamics approach to investigate the subtle conformational changes in the RBD as well as the truncated S1 domain of the P.1 variant spike protein responsible for the increased affinity with the ACE2 receptor. Our study on the protein contact network identifies changes in the protein structures at the residue contact level which are otherwise not easily detectable. Moreover, our study also highlights the identification of commonalities and differences in the dynamics of the interactions between spike proteins of SARS-CoV-2 and its P.1 variant with the ACE2 receptor. Together with our structural communication analysis and dynamics study provide an understanding towards interaction with the ACE2 receptor. Overall, the study provides additional information that may be utilized for designing better therapeutics against the circulating P.1 variants as well as other future variants.

2.2 METHODS

2.2.1 Construction of Protein Structure Graph (PSG)

The PCN identifies the change in protein structures at contact level which are otherwise not easily detectable [Kandhari and Sinha, 2017; Lata and Akif, 2021]. The 3-D coordinates of the RBD-ACE2 complexes spike protein from SARS-CoV-2 wild type [6MOJ; [Lan et al., 2020]] and its P.1 variants [7NXC; [Dejnirattisai et al., 2021]] were extracted from the PDB. The PSG was constructed on the complex structures using the PSN and ENM-NMA approaches implemented in the WebPSN [Felline et al., 2020]. Each amino acid is depicted as a node and each node is

connected to other nodes in a protein structure through an edge. The interaction strength between two connecting nodes is defined as

$$I_{ij} = \frac{n_{ij}}{\sqrt{N_i N_j}} 100$$

Where (I_{ij}) interaction percentage of nodes i and j . It follows that the number of side chain atoms pairs within (4.5 Å) cutoff, N_i and N_j are normalization factors [Brinda and Vishveshwara, 2005; Kannan and Vishveshwara, 1999]. It constructs a PSG on the basis of the atomic cross correlation motions using the ENM-NMA.

2.2.2 Network Parameters

Three important network parameters such as Hubs, Modularity and Structural communication pathways were calculated for the RBD-ACE2 and S1-ACE2 complexes from the wild type as well as P.1 variant. PYMOL was used to visualize these parameters. Briefly, **hubs** are nodes with the highest degree. **Modularity** is represented as communities with more interconnected nodes and the nodes of the same community are highly connected to each other than the poorly connected nodes of outside the community. **Shortest Path** is the smallest number of links required to travel from one node to another. It is calculated on the basis of Dijkstra's algorithm.

2.2.3 Protein stability and flexibility analysis

To analyze the effect of all mutations present on the RBD domain and outside the RBD (truncated S1 domain) of the spike protein, free energy changes ($\Delta\Delta G$) and vibrational entropy difference ($\Delta\Delta S$) were calculated using the DYNAMUT tool [Rodrigues et al., 2018]. Based on these calculations, protein stability and flexibility were analyzed on the complexes by utilizing the NMA based elastic network contact model (ENCoM).

2.2.4 Perturbation Residue Scanning

Perturbation Residue Scanning was performed using the pPerturb server [Gopi et al., 2020]. It allows the mutation of one or more residues to alanine and generates a perturbation profile (ΔQ Vs Calpha-Calpha) distance from the perturb site. The perturbation effect can be analyzed as a distance connecting the perturb residue to its nearby residues or on the interaction network strength.

2.2.5 Molecular Dynamics Simulation

To study the dynamic behavior of the complexes (RBD-ACE2 and truncated S1-ACE2) from the wild type and P.1 variant with respect to time, MDS was performed for the time scale of 100ns. To ensure the convergence of simulations, we performed three independent 100 ns production runs for the wild type and P.1 variant RBD-ACE2 complex. The coordinates of the complexes were obtained from the PDB and used as starting models for the MDS using the GROMACS package [Pronk et al., 2013]. The crystal waters were removed from the co-ordinates. For setting the periodic boundary conditions the system was soaked in a cubic box with a dimension of 1.5 nm, further the box was filled with SPC water molecules and OPLSAA force field was applied [Berendsen et al., 1987]. For neutralization the appropriate water molecules were replaced by counter ions. To remove the steric clash the systems were subjected to energy minimization with the steepest descent method for 50000 steps until the largest force was smaller than 1000 kj/mol/nm. Then the minimized system was equilibrated in the NVT and NPT ensemble for 100 ps each. Lastly the production run was performed for 100ns at 300K where the leapfrog integrator was used for time evolution trajectories. A constant temperature on the system was maintained at 300K using modified Berendson Thermostat [Bussi et al., 2007] and Parrinello-Rahman barostat [Parrinello and Rahman, 1981] pressure was maintained at 1bar during the simulations. The

analysis of the trajectory files were done with gmx rmsd, gmx rmsf, gmx gyrate, gmx sasa, gmx hbond, gmx covar, gmx anaeig, gmx do_dssp for the root mean square deviation (RMSD), root mean square fluctuation (RMSF), Radius of gyration (Rg), Solvent accessible surface area (SASA), hydrogen bond (H-bond), principal component and secondary structures, respectively. Graphs were generated using pymol, XMGRACE and GNUPLOT.

2.2.6 Principal component analysis (PCA)

In order to analyze dominant and collective motions in the RBD-ACE2 and S1-ACE complexes from the wild type and P.1 variant, PCA was performed. Mass weighted covariance matrix on the coordinates of the MD trajectories was calculated. Diagonalization of the covariance matrix was used to get a set of eigenvectors and eigenvalues from the G_covar function implemented in the GROMACS. The g_anaeig function was utilized to obtain trajectories on the eigenvectors [J., 2014]. Backbone atoms were considered for the PCA analysis. The coordinates of the residues were projected on the first two Principal Components that represent the highest eigenvalues and captured the overall motion of the protein.

2.2.7 Binding free energy calculations

Molecular mechanism-Poisson-Boltzmann surface analysis (MMPBSA) method was used to estimate the binding free energy between the RBD-ACE2 complexes from the wild type and P.1 variant. Last 10ns time frames of MD simulations were integrated to calculate the free energy using the g_mmpbsa tool of GROMACS [Kumari et al., 2014]. The molecular mechanics energy includes electrostatic, Van-der-Waals interactions and polar and non-polar solvation energy.

2.2.8 Normal Mode Analysis

Normal mode analysis (NMA) was performed on the stable simulated trajectories coordinates by using the I-MODS tool [López-Blanco et al., 2014]. The NMA calculated the lowest frequency

modes and identified atoms moving together as a rigid body. The internal modes were estimated on the basis of the Lagrangian equation of motions.

2.3 RESULTS

Due to emerging variants of the COVID-19, the current pandemic seems to be quite alarming with continuously rising waves. The emerging variants are associated with many mutations mostly accumulated in the spike protein of SARS-CoV-2. These mutations have been functionally mapped for the increased affinity toward the ACE2 receptor. The increased affinity attributes variants to be more infectious and have a high transmission. This renders the current therapeutics a challenge to combat SARS-CoV-2. Therefore, it is warranted to have a better understanding of the structural basis of interaction in variant spike protein with the ACE receptor. In our study, we chose the P.1 variant of SARS-CoV-2 that has almost ten mutations in the spike protein. In spite of having three mutations on the binding interface of the RBD, the calculated cross structure RMSD between the complex of the RBD-ACE2 from the wild type and P.1 variant was 0.29 Å which is almost negligible (**Figure 2.1**), suggesting no overall structural changes among the two. Still it is a matter of concern. Hence, in order to investigate further, we considered the PCN and dynamics of the complexes, assuming that the changes at that level may cause the significant variation noted for the increased affinity with the ACE2 receptor [Verkhivker, 2020a; Verkhivker and Di Paola, 2021]. Since, the overall structural changes are negligible hence we performed a protein structure network based approach to examine the small variations spanned throughout the protein.

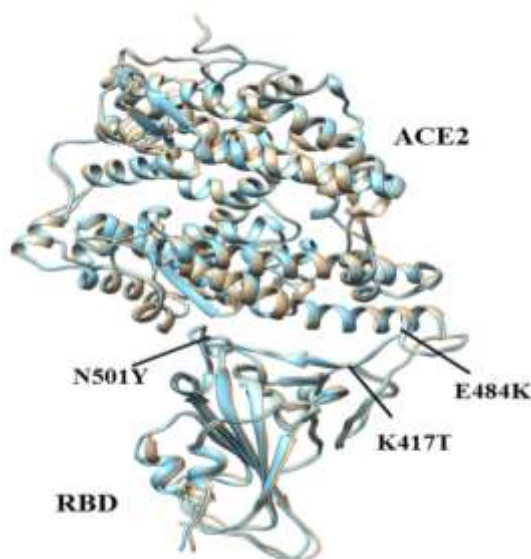


Figure 2.1. Structural alignment of the RBD-ACE2 complex from SARS-CoV-2 wild type (sandy brown color) and P.1 variant (cyan). Positions of three interface mutations are indicated with black lines.

2.3.1 Protein structure network analysis of wild type and P.1 variant protein structures

The PSN construction calculated three important network parameters such as hubs, modularity and shortest communication pathway and yielded global average network parameter values for the wild type and P.1 variant (**Table 2.1**). These values were quite comparable suggesting diminutive changes among the two structures. This also supports the observation of negligible cross structure RMSD deviation between the wild type and P.1 variant complexes. Further, for the better understanding of the effect of mutations on the subtle conformational variations, residue specific local variations among the network parameters were calculated for the wild type and P.1 variant.

	SARS-CoV-2	
	Wild type	P.1 variant
No. of edges	317	317
Degree	8.02	8.03
Clustering coefficient	0.51	0.51
Shortest path	9.66	9.80

Table 2.1. List of global average network parameters.

Hubs. The calculated hubs were mapped on the RBD-ACE2 complex from the wild type and P.1 variant. A total number of 172 and 86 hubs were observed in the wild type and P.1 variant complex, respectively. This indicates a significant change in the hub formation among the two. Moreover, a reduction of hub residues in the P.1 variant complex is observed to be localized in the RBD region. A significant loss of hub residues was also observed at the interface region of RBD-ACE2 complex from the P.1 variant and these regions are highlighted in the circle as shown in the (Figure 2.2) The interacting residues from this region, Y449, Y453, L455, Q493, Y495, F497, Q498, N501, Y505, were found to be involved in hubs formation in the wild type, but were absent in the P.1 variant. Additionally, a difference in the hub formation was also observed on the centrally located antiparallel β -sheets (Figure 2.2). This difference suggests that three interface mutations (K417T, E484K and N501Y) associated in this region of the P.1 variant may have significantly perturbed these hub zones. All the residues involved in the hub formation among the RBD from wild type and P.1 variant are listed in (Table 2.2). Few unique hubs (Y41, K353, and R357) were also observed at the interface of the ACE2 from the P.1 variant complex (Table 2.2). The observation of the variation among the hub residues from wild type and P.1 variant suggests differences in the structural connectivity. This may be the consequence of loss of interactions among the domain which may account for the flexibility and better conformational adaptability in

case of the P.1 variant. Structural dynamic/variability in the spike protein in context to open and close state has been investigated by analyzing unique hubs among the two states [Halder et al., 2020].

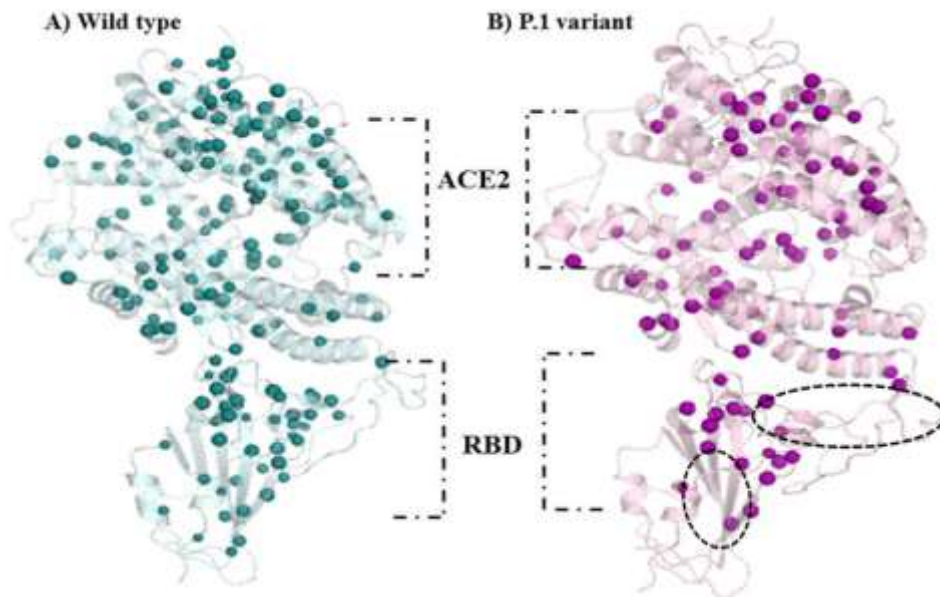


Figure 2.2. Representation of hubs arrangement on RBD-ACE2 complexes (A) wild type and (B) P.1 variant. Hubs are represented as cyan and purple spheres in wild type and P.1 variant, respectively. Loss of hubs from the interface RBM and betasheet core of the RBD of P.1 variant are encircled. ACE2, angiotensin converting enzyme 2; RBD, receptor binding domain.

Wild type	P.1 variant
ACE2: Y41, K353, R357	ACE2: Q24, F32, Y41, Y83, K353, R357
RBD: C336, V341, F342, Y351, W353, R355, S359, Y365, F374, N394, Y396, F400, I402, R403, E406, 420-423, P426, F429, N439, D442, Y449, 453-457, 464, N487, Q493, Y495, Q498, P 499, N501, 505-509, F515	RBD: F342, V350, Y351, W353, R355, Y396, F400, R403, E406, I418, Y423, N439, D442, N487, Y495, Q498, P507, R509

Table 2.2. Hubs residues participating in the RBD-ACE2 complex of the wild type and P.1 variant.

Unique ACE2 hub residues in the P.1 variant are highlighted red.

Modularity. The community structure analysis results showed a total 22 communities in the wild type and 20 communities in the P.1 variant. The residue-wise communities in the wild type and P.1 variant are shown in Figure 3. The largest community (C1) was observed in the ACE2 from the both complexes. Interestingly, the total number of nodes and links involved in this community formation were observed to be reduced in the ACE2 of the P.1 variant complex. This community consisted of 114 nodes, 192 links and 74 hubs in the wild type complex, while the same was reduced to 54 nodes, 82 links and 30 hubs in the P.1 variant complex. The second largest community (C2) was observed in the RBD domain and the same was extended throughout the interface of the ACE2 in the wild type complex. In contrast, this region in the P.1 variant was divided into two different communities (**Figure 2.3**). Overall this change in community arrangement reflects subtle conformational changes due to mutations which were not evident at the structural level. Moreover, the rearrangements in modules were spanned throughout the structure which reflects the perturbation at global level.

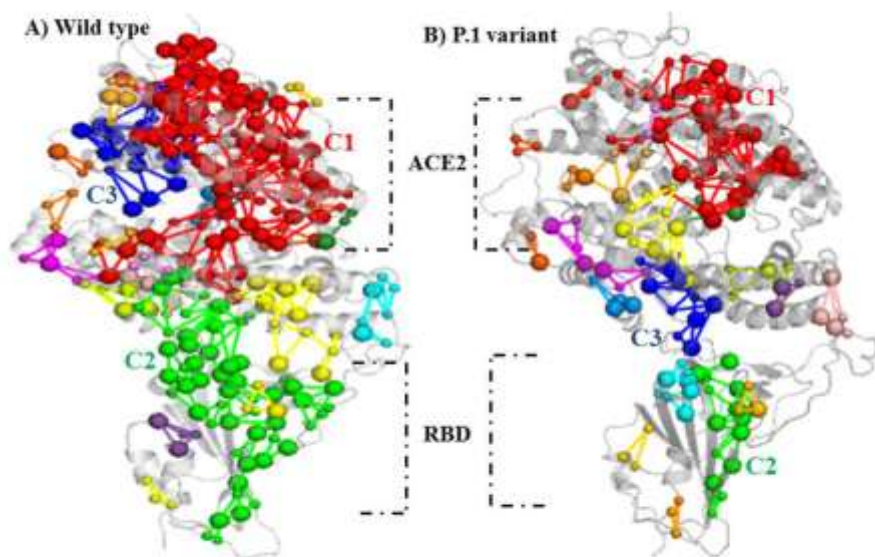


Figure 2.3. Comparative community structures in RBD-ACE2 complexes. (A) Wild type (B) P.1 variant. Communities are represented in different colors where the first three largest communities (C1, C2, C3 are in red, green, blue colors), respectively. Nodes and links involved in these communities were reduced in the P.1 variant. ACE2, angiotensin converting enzyme 2; RBD receptor binding domain

Shortest Communication Pathways. The total length of the shortest path in wild type was 1210620 while a total of 523853 paths were observed in the communication channel of the P.1 variant. This observation indicated a decrease in the pathways of the RBD-ACE2 complex in the P.1 variant. Moreover, the average path hub% was also significantly decreased in the P.1 variant (**Table 2.3**). Major difference was observed in the interface residues nearby mutation N501Y and the N-terminal region of the RBD. The interface residues such as E497, Q498, P499, and T500 from the wild type were observed as nodes in the shortest path. In contrast, these residues

were not observed to be involved in the shortest pathway in the P.1 variant (**Figure 2.4**). Possibly the N501Y mutation from the P.1 variant may have perturbed this channel. Another prominent change was observed in the N-terminal region (Res. C336, F338, F374) of the wild type RBD were involved on paths but were absent in the P.1 variant (**Figure 2.4**). These changes clearly suggest a significant perturbation in inter and intra communication within the complex due to the mutations.

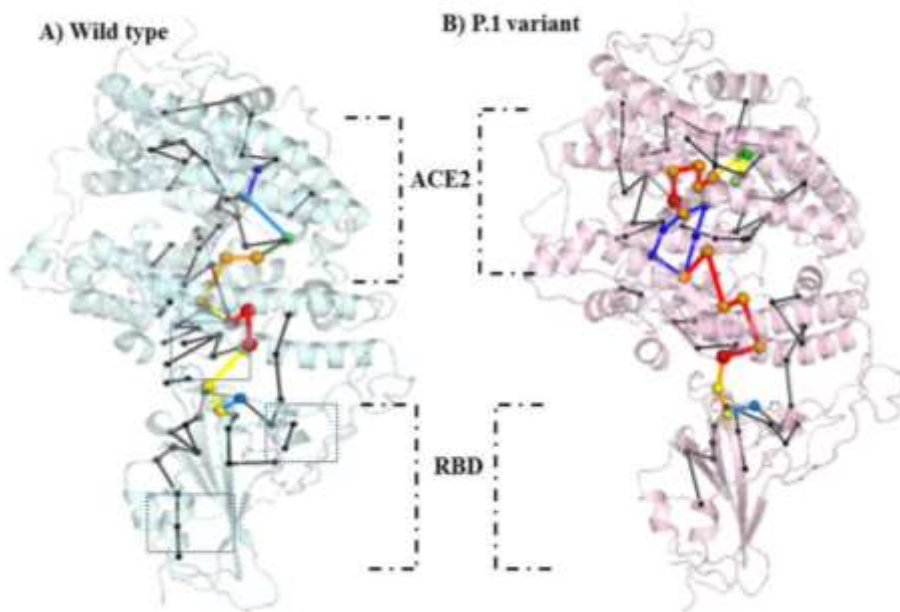


Figure 2.4. Shortest path communication channel on the RBD-ACE2 complex. A) wild type B) P.1 variant.

2.3.2 Perturbation in PSG due to mutation outside the RBD domain

Mutations present outside the RBD of spike protein may also confer an effect on the binding affinity with the ACE2 receptor. Hence, in order to analyze the effect of these mutations on the structural flexibility, the PSG for the truncated S1 domain-ACE2 complex from the wild type and P.1 variant was generated. Except for a few differences, the hub arrangement was almost similar

in both the complexes (**Figure 2.5**). Five out of seven mutations in the NTD of the S1 domain were found to have altered hub arrangement compared to wild type. Interestingly, few interface hubs were observed in the same complex of the P.1 variant, which were absent in the PCN analysis of the RBD-ACE2 complex. Moreover, the N501Y mutation in P.1 variant was also observed to engage in hub formation. The altered hub in the P.1 variant may have some role in providing a better efficiency towards the ACE2 receptor. In fact, the functional role of N501Y towards the binding affinity with the ACE2 receptor has already been demonstrated [Dejnirattisai et al., 2021].

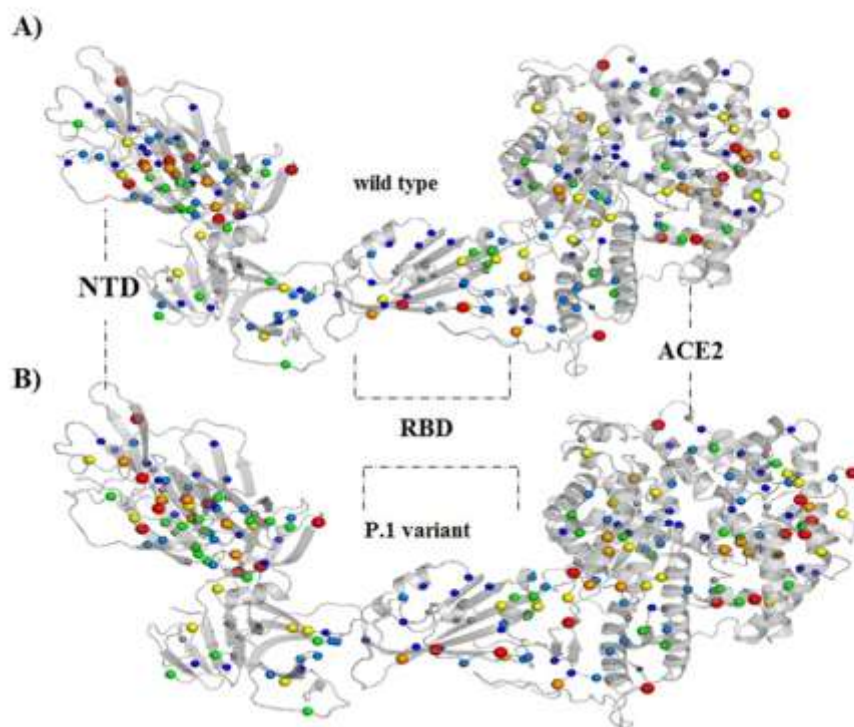


Figure 2.5. Comparative hubs arrangement in the S1 domain-ACE2 complexes from the wild type and P.1 variant. The hubs are represented as spheres and the colour of the spheres indicates the interaction strength.

A total of 39 communities were observed in the truncated S1 domain of wild type while the corresponding domain in the P.1 variant showed up only 41 communities. In both the complexes, the ACE2 region was observed to form the largest community and RBD region contained a second largest community (**Figure 2.6**). A significant perturbation was observed in the community arrangement of the NTD domain of P.1 variant complex suggesting an extensive reorganization in node's membership as compared to the wild type. Alteration of nodes and links was observed in the NTD of the P.1 variant compared to the wild type complex.

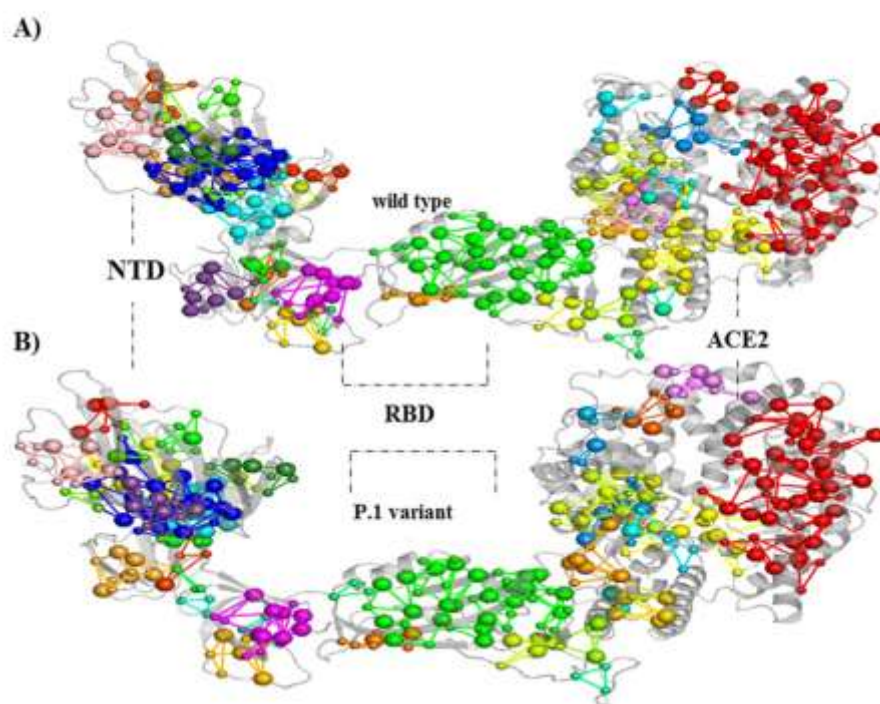


Figure 2.6. Mapping of the communities on the truncated S1 domain-ACE2 complexes of A) wild type and B) P.1 variant.

Moreover, residues participating in the shortest path of the NTD from both the cases were also different. The pathways from this region were observed to be slightly shifted in the P.1 variant (**Figure 2.7**). A total number of nodes and links in the communication pathway from the P.1

variant were more compared to the wild type. The total numbers of the shortest pathways along with the average path length were also slightly high in the P.1 variant (**Table 2.3**). These observations from this region clearly indicate that the mutations may have perturbed the core of the NTD communication in the P.1 variant.

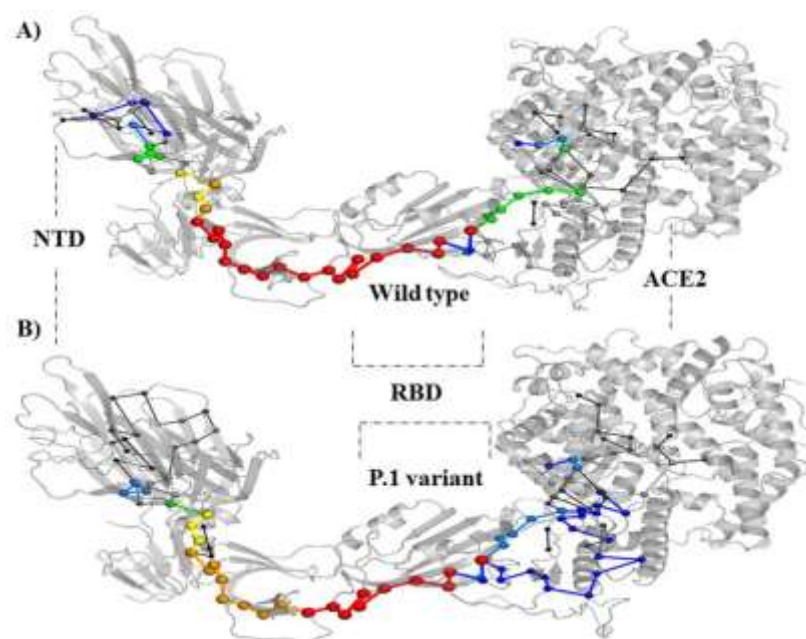


Figure 2.7. Shortest path communication channel in the truncated S1 domain-ACE2 complexes of A) wild type and B) P.1 variant.

Path Summary	Wild type	P.1 variant
No. of nodes in path	98	109
No. of links in path	97	108
No. of Shortest Path	4609389	5347490
Avg. Path Length	44.24	45.62
Avg. Path Hub %	43.65	42.13

Table 2.3. Network components and parameters involved in average shortest path for RBD-ACE2 complexes from the wild type and P.1 variant.

2.3.3 Effect of mutations on the thermodynamic of protein

To analyze the effect of mutations on protein structures, change in vibrational entropy ($\Delta\Delta S$) and change in free energy ($\Delta\Delta G$) were calculated. All three interface mutations (E484K, K417T and N501Y) of the RBD provided a destabilizing effect on the protein and were found to have negative free energy values (Table 2.4).

Mutants	$\Delta\Delta G$ (kcal/mol)
L18F	0.8
T20N	-0.27
P26S	-0.18
D138Y	1.86
R190S	-0.9
K417T	-0.058
E484K	-0.496
N501Y	-0.299
D614G	-0.3
H655Y	1.08

Table 2.4. Predicted free energy values of the mutants

Moreover, change of vibrational entropy value for the K417T mutant was observed as 0.514 kcal mol⁻¹ K⁻¹ which was comparatively higher than the E484K and N501Y mutations (Table 2.5).

Mutations	$\Delta\Delta S$ (kcal/mol/K)
K417T	0.514
E484K	0.179
N501Y	0.145

Table 2.5. Predicted change in entropy values associated with the interface mutations on the RBD.

The K417T mutation attributed towards more flexibility as compared to the other two mutations and its effect was extended to the other part of the protein as well. Though, the K417T is present

near 3/10 α -helix, but observed to impart flexibility to the loop region of RBM as well as N-terminal region residues (**Figure 2.8A**). In contrast to the K417T mutant, the effects of the E484K and N501Y mutations were localized at the nearby residues only (**Figure 2.8 B & C**). Further, the vibrational entropy change was also observed for the mutations localized outside the RBD domain (**Figure 2.9**). The effect of the R190S mutation from the NTD was significant and observed to make the complete domain flexible. Slightly less flexibility was observed for the T20N mutation. Interestingly, a major effect was noticed for the D614G mutation that provided a flexibility to the RBD as well as the NTD region of S1 domain, suggesting that the D614G mutation is associated with a global flexibility in the protein. These changes may provide a conformational plasticity to the RBD domain and other parts of the protein enabling efficient binding with the ACE2 receptor. Similar observation was reported in many previous studies signifying that D614G mutation has increased the binding affinity for ACE2 receptors [Gobeil et al., 2021; Ozono et al., 2021].

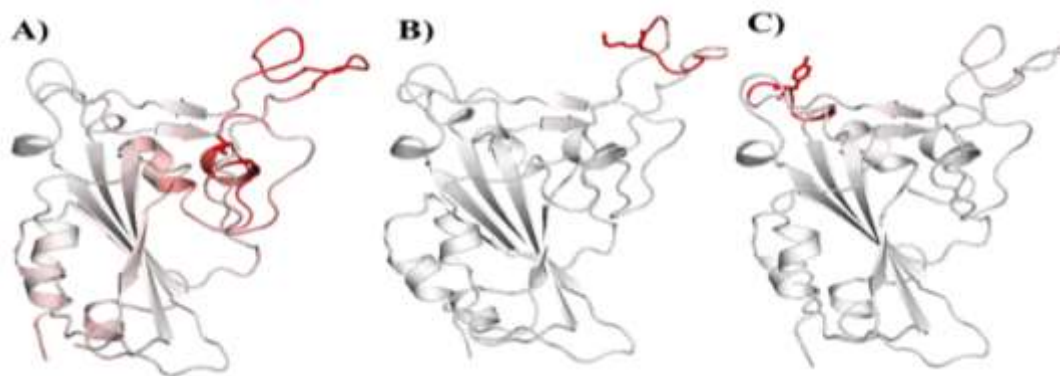


Figure 2.8. Visual representation of the interface mutations on dynamicity and plasticity of the RBD. (A) RBD structure with the K417T mutation (B) RBD structure with the E484K mutation (C) RBD structure with the N501Y mutation. The RBD is represented as a cartoon structure and mutations are shown as a stick model. Red color in the cartoon structure indicates the flexibility in the protein. RBD, receptor binding domain.

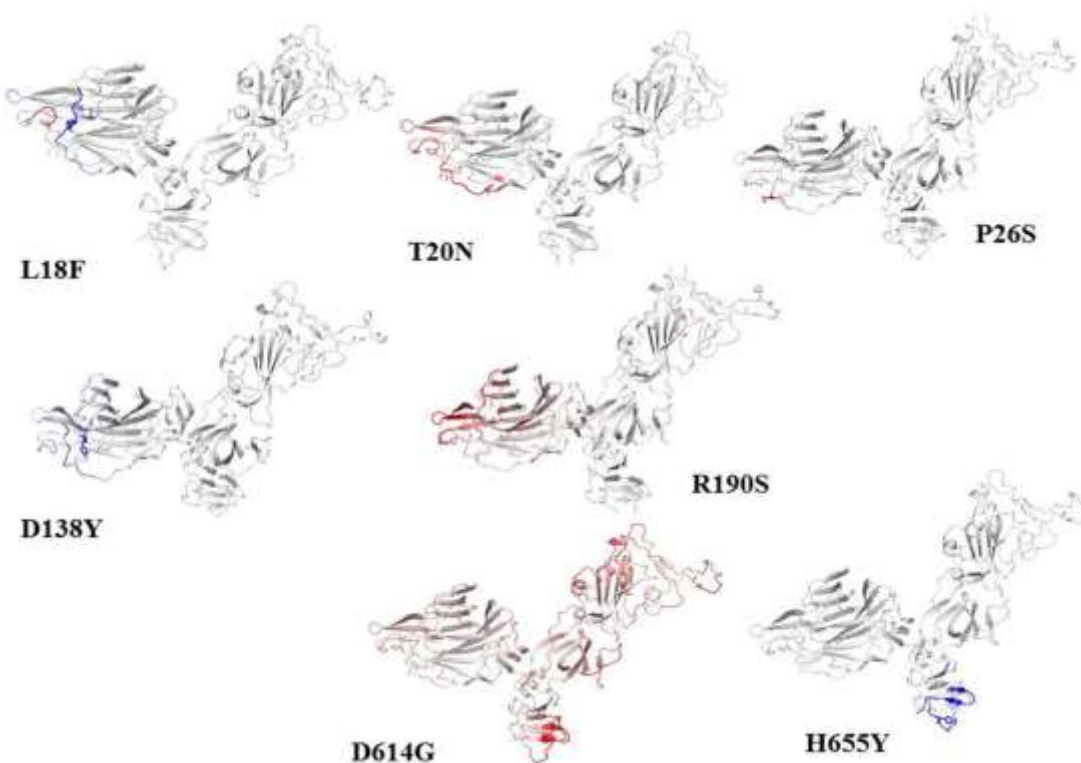


Figure 2.9. Visual representation of effects of the mutations present in the truncated S1 domain on protein dynamics. Flexibility is represented in red color.

Additionally, in order to probe the effect of the interface mutations on the allosteric residue, perturbation residue scanning (PRS) was analyzed. Similar to the entropy change analysis, the PRS analysis of the K417T mutation showed maximum effect on the allosteric residues (**Figure 2.10**). This observation suggests that mutation at one part of the protein may provide an effect on the other part to make it more adaptable. The flexibility associated with the mutations in the ACE2 complexes was further investigated by MDS at 100ns time-frames. The cumulative local as well as allosteric effects of the mutations on the conformational dynamics behavior of the interfacial binding residues of RBD-ACE2 complexes were mapped.

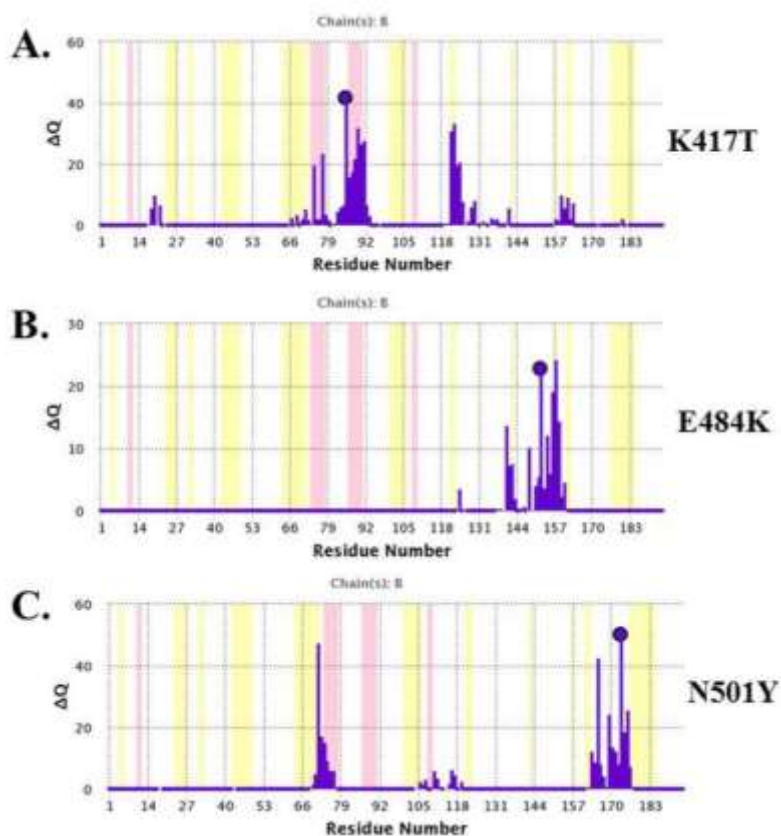


Figure 2.10. Residue-wise Perturbation Residue Scanning profile of three interface mutations present on the RBD.

2.3.4 Probing of conformational dynamics of RBD-ACE2 complex

The calculated average backbone RMSD values of the RBD-ACE2 complexes from the wild type and P.1 variant were observed as 0.3 and 0.35, respectively. The RMSD values clearly indicate that there is no significant change in the trajectory of both the complexes (**Figure 2.11 A**). Moreover, the probability distribution function of the RMSD values of the wild type and P.1 variant also appeared to fall within the same range of 0.1-0.55 (**Figure 2.11 B**). These observations suggest that the backbone stability for both the complexes was maintained throughout the simulations with an RMSD value of less than 4.5Å.

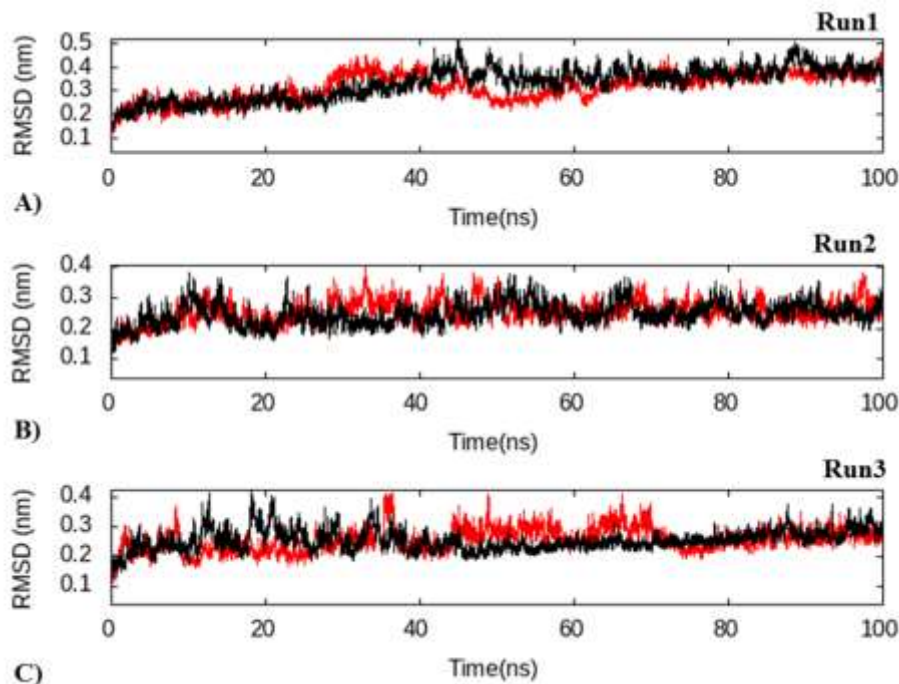


Figure 2.11. Graphical representation of backbone RMSD trajectory of RBD-ACE2 complex from wild type and P.1 variant for three independent replicates of 100ns simulations. Red and black colored trajectories represent wild type and P.1 variant complex, respectively

Similarly, the RMSF profile of the RBD-ACE2 complexes from the wild type and P.1 variant did not show any significant change throughout the simulation time (**Figure 2.12A**). Except for a few differences, the trajectory of both the complexes was observed quite overlapping to each other with a maximum fluctuation of less than 0.45nm. The differences in the RMSF were observed when only RBD (apo-RBD) or only ACE2 (apo-ACE2) was considered for the analysis. For apo-RBD, the RMSF values of the regions - Res380-388 from the P.1 variant and Res470-480 from the wild type, were slightly high (**Figure 2.12B**). Additionally, the fluctuations of the mutated residues at the interface of the apo-RBD from the P.1 variant were observed to be reduced. The

region, Res19-200, of the ACE2 from the wild type complex displayed comparatively higher fluctuation (**Figure 2.12C**). Expectedly the interface interacting residues from apo-ACE2 were less fluctuating in the P.1 variant complex suggesting a rigidity in the residues upon binding with the RBD of the P.1 variant and results in better binding.

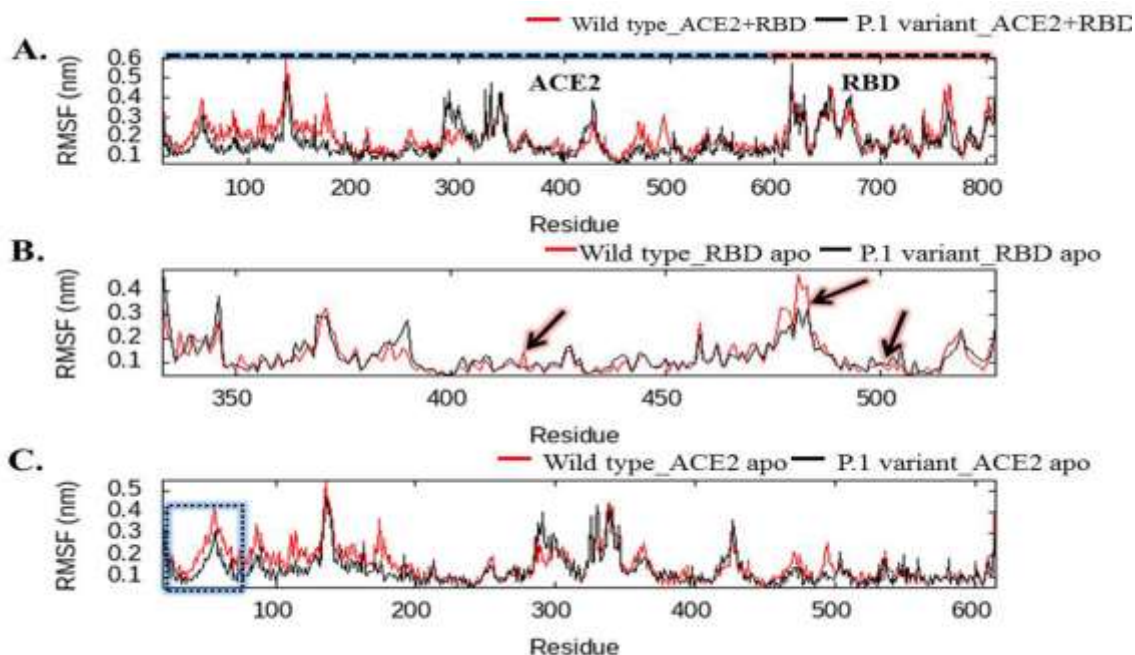


Figure 2.12. RMSF trajectories generated for the wild type and P.1 variant. (A) RBD-ACE2 complex. Region of ACE2 and RBD from the complex is highlighted and labeled at the top of the graph. (B) Apo RBD. Arrows indicate three mutations of the RBD interface (C) Apo ACE2. Interface interacting residues of the ACE2 are depicted in the dotted rectangle. ACE2, angiotensin converting enzyme 2; RBD, receptor binding domain; RMSF, root mean square fluctuation.

The SASA analysis of both the complexes indicated a slight difference in the overall values (Figure 2.13A). In fact, the SASA value for the RBD-ACE2 complex of the P.1 variant was observed to be higher than the same complex from the wild type. The probability distribution function of the SASA also showed a slightly different distribution pattern for the two (Figure 2.13B). The residues which are buried in the hydrophobic core of the protein are the driving force in the protein folding and the slight change of SASA suggests a conformational change in the protein during the course of 100ns of simulation.

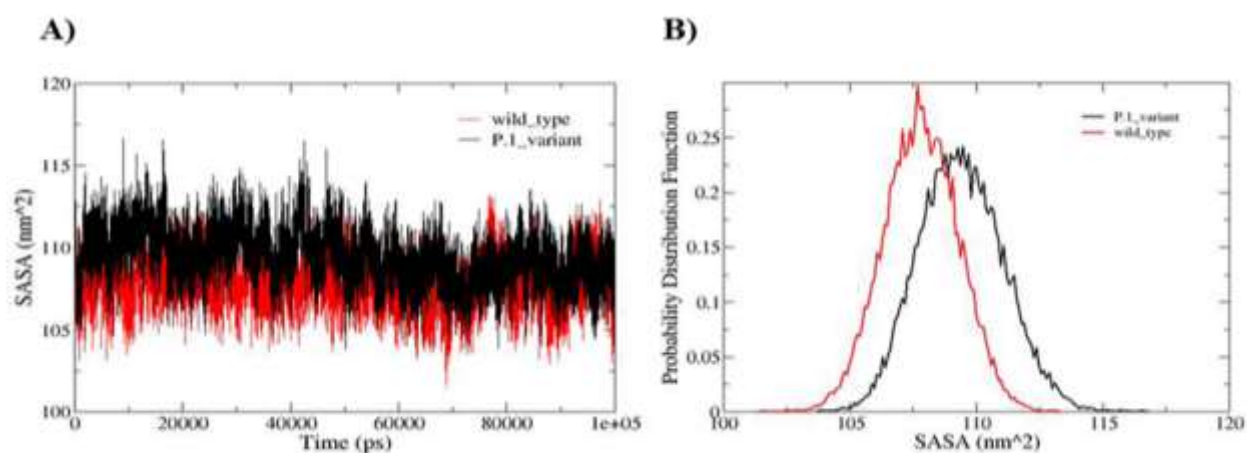


Figure 2.13. Graphical representation of the conformational dynamics. (A) SASA values against time for wild type and P.1 variant RBD-ACE2 complex. (B) Probability distribution function of SASA values. ACE2, angiotensin converting enzyme 2; RBD, receptor binding domain; SASA, solvent accessible surface area.

Inter/intramolecular hydrogen bond interactions play an important role for the stability and integrity of protein. Compared to the RBD-ACE2 complex from the P.1 variant, the same complex from the wild type possessed a significantly higher number of hydrogen bonds (H-bonds) (Figure

2.14A&B). The average number of inter H-bonds for the wild type and P.1 variant were 13 and 9, respectively. Similarly, the average number of intra H-bonds in the RBD-ACE2 complex of the P.1 variant was 113 while the same for the wild type was 118.4. The decrease in total H-bonds in the RBD-ACE2 complex of the P.1 variant indicates comparatively lesser stability which suggests that this complex may have slightly higher overall structural flexibility that may be required for the better adaptability. Further, a change in the Rg was observed among the two indicating a variation in the compactness also. The compactness during 0-20ns for the wild type and P.1 variant were almost same and the average value of complexes was observed approx. 3.15 Å (**Figure 2.14C**). Later, the structural compactness of the RBD-ACE2 complex from the P.1 was lost. Its average value was increased to 3.25Å and maintained till the end of simulations. In contrast, the RBD-ACE2 complex from the wild type was observed to have higher average value during the initial frame of the simulation (0-40ns) and later it decreased. This was maintained till the end of the 100ns simulation. This suggests that the wild type complex is comparatively stable and the loss of compactness of the P.1 variant complex may be associated with more structural flexibility and plasticity.

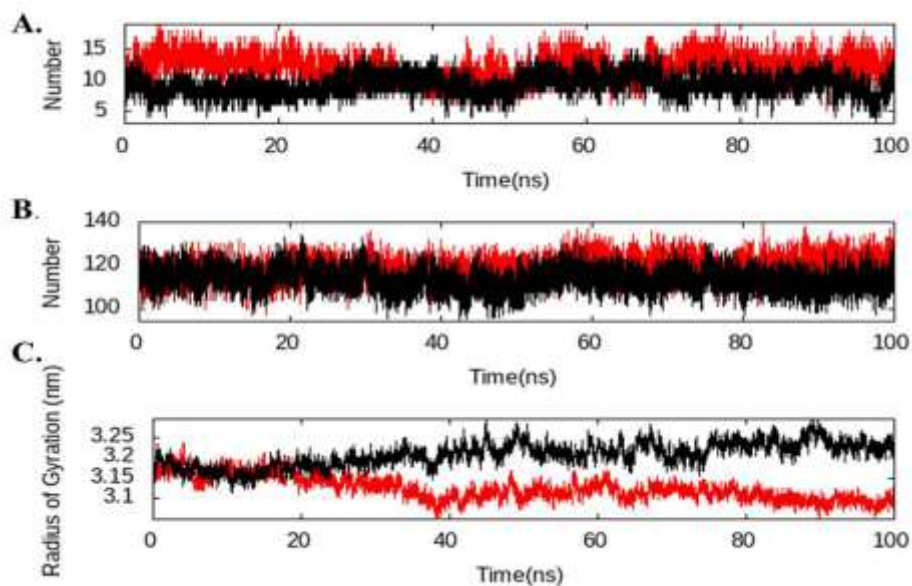


Figure 2.14. Plots representing strength and compactness in the RBD-ACE2 complexes. (A) Number of Inter hydrogen bonds vs time between RBD and ACE2. (B) Number of Intra hydrogen bonds vs time within the RBD. (C). Radius of gyration of RBD-ACE2 complex. Wild type (red plot) and P.1 variant (black plot). ACE2, angiotensin converting enzyme 2; RBD, receptor binding domain.

In order to probe dominant motions and conformational changes among the RBD-ACE2 complexes from wild type and P.1 variant, the principal components were extracted from the MD simulation trajectories. Covariance analysis of the atomic correlation plot indicates that the RBD-ACE2 complex of the wild type mostly possessed dominant anti-correlation motion. Whereas the same complex from the P.1 variant was associated with uncorrelated motion (**Figure 2.15A & B**). This difference may be accounted for due to the loss of few contacts and slight conformational variations associated in the P.1 variant. This observation is in correlation with our PCN and MDS analysis.

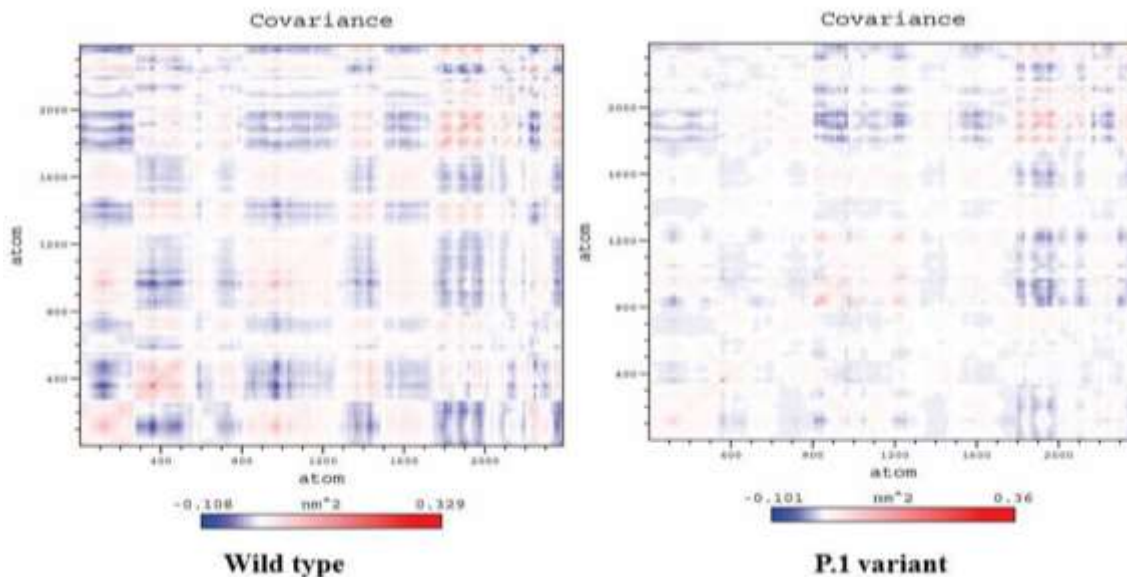


Figure 2.15. Covariance matrix plot for RBD-ACE2 complexes (A) wild type (B) P. 1 variant

Further, nature of motions of the RBD from the wild type and P.1 variant was analyzed through the NMA. The constructed porcupine plots extracted from the stable trajectories indicate an opposite motion for the RBD from the complex from wild type and P.1 variant. Interestingly, the RBM domain and N-terminal region of the RBD from the wild type appeared to move in the same direction but the same regions in P.1 variant was observed to have motion opposite to each other (Figure 2.16A & B).

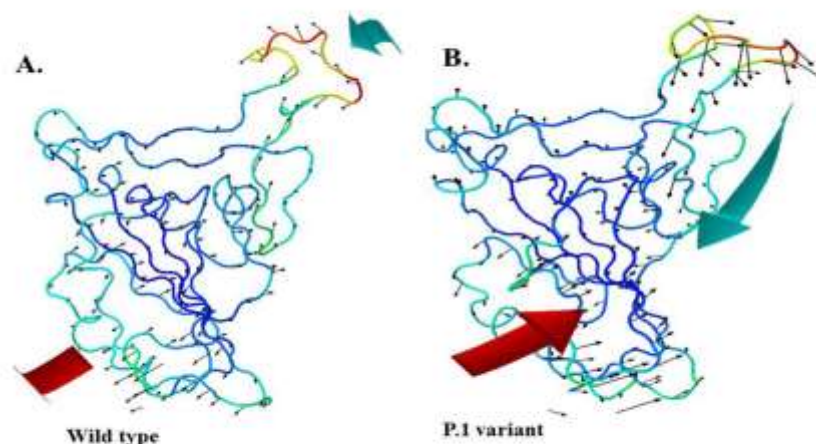


Figure 2.16. Motion Mode Analysis of the RBD for A) wild type and B) P.1 variant.

To estimate the binding affinities between RBD and ACE2 complex from the wild type and P.1 variant, free energy of association for the complex was estimated using the MMPBSA method. The MMPBSA calculation clearly indicates that the complex from P.1 variant possessed a higher negative free energy value of -1804 kJ/mol compared to the complex from wild type which was estimated as -1325 kJ/mol (**Table 2.6**). Comparatively better binding energy for the P.1 variant suggests a better affinity towards ACE2. This observation is in accordance with the experimental data reported about the P.1 variant. The better binding is mostly contributed by the Van-der-Waals, the SASA and the electrostatic interactions. All these interactions provide a negative binding energy towards the total free energy. However, the polar solvation energy was positively contributing toward the total binding free energy (**Table 2.6**). Additionally, the number and type of interactions from the interface of the RBD-ACE2 complexes of the wild type and P.1 variant during 70-100ns time frames were plotted. The plot clearly indicates that the hydrophobic

interactions were significantly high in P.1 variant as compared to wild type. In contrast, the interface H-bonds in the complex from the P.1 variant were less (**Figure 2.17**).

	Vander Waal Energy	Electrostatic Energy	Polar Solvation Energy	SASA Energy	Binding Energy (kJ/mol)
Wild type	-448.272	-1416.22	591.499	-52.442	-1325.44
P.1 variant	-351.868	-2058.80	652.073	-45.786	-1804.38

Table 2.6. Average MMPBSA calculations for the wild type and P.1 variant complexes.

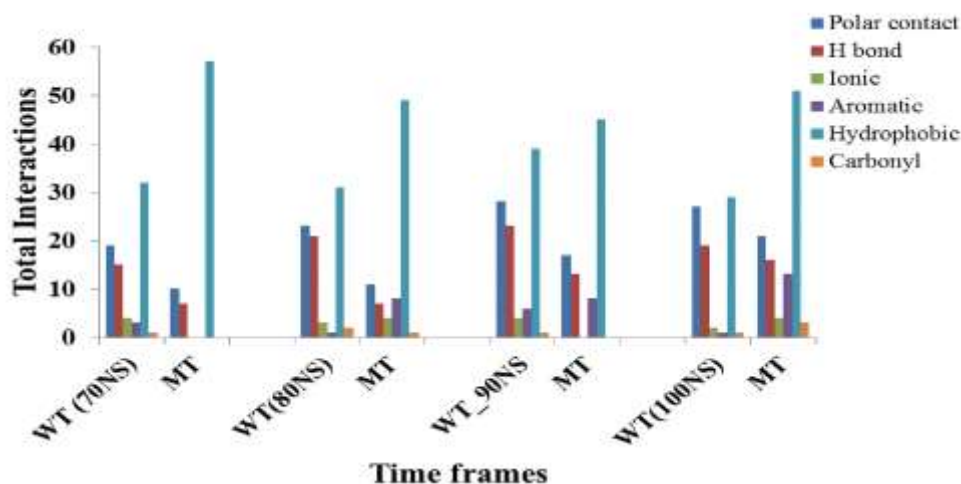


Figure 2.17. Total number of interactions. The RBD-ACE2 complexes from wild type and P.1 variant at different time frames. Types of interactions are depicted with different color as shown. WT, MT and NS are abbreviated here for the wild type, P.1 variant and Nano seconds, respectively.

Comparison of the RBD-ACE2 interface residues from the wild type and P.1 variant after 100ns simulations provided information about the several hydrophobic and hydrogen bond interacting residues. The rearrangement of the hydrogen bonds were observed in the interface of the complexes. Apart from several common hydrogen bonds, few unique hydrogen bonds appeared in the P.1 variant. The Thr500 from the RBD of the P.1 variant was found to make hydrogen bonds

with Asp355 and Tyr41 of the ACE2 (**Figure 2.18**). Similarly, hydrophobic interactions such as Ala475, Lys489, Tyr505 and Gln498 from the RBD interface and Met82, Gly352 and Gly354 from the ACE2 interface were found to be unique in the P.1 variant (**Figure 2.18**). Involvement of these unique interactions in the P.1 variant may provide an energetically favorable binding with the ACE2. Hence, these interactions may be utilized to target specific therapeutics to disturb the RBD-ACE2 binding in the variant.

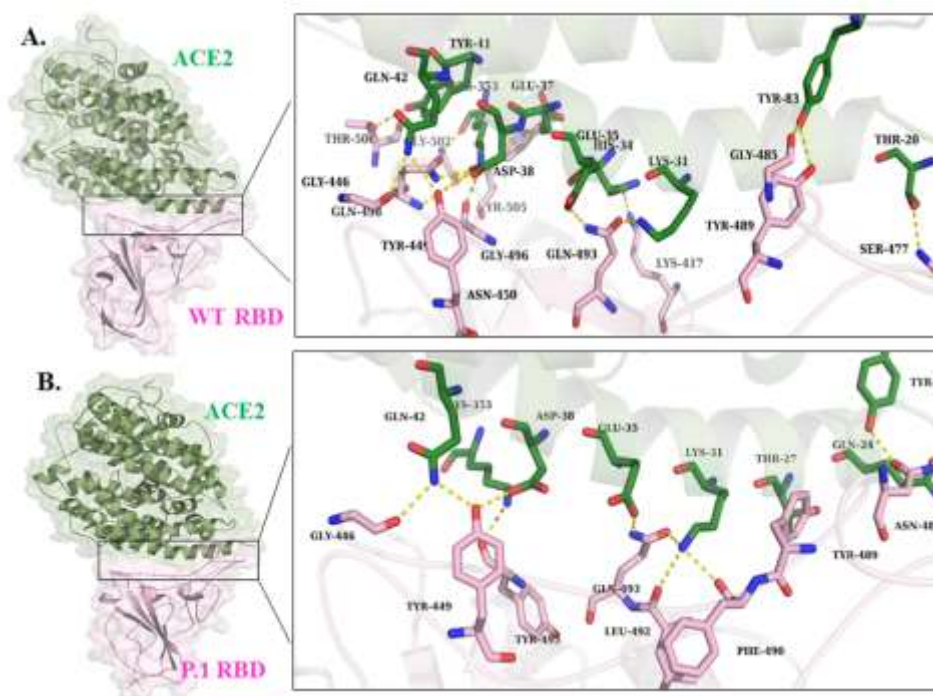


Figure 2.18. Representation of interface interacting residues involved in hydrogen bond interaction in the representative structure of wild type and P.1 variant complex obtained after clustering analysis. RBD and ACE2 are represented as cartoon structure and surface diagram with color hot-pink and smudge green, respectively. The interacting region is denoted in a rectangular box. Interacting residues are represented in the stick model with the respective color. Hydrogen bonds are shown in yellow dashed lines. ACE2, angiotensin converting enzyme 2; RBD, receptor binding domain.

2.3.5 Probing the Conformational Variations in the truncated S1 domain

To study the conformational variations attributed with the mutations present outside the RBD, 100ns simulation was performed for the truncated S1 domain-ACE2 complexes from the wild type and P.1 variant. Residue-wise RMSD calculation quantified the flexibility due to mutations and its cumulative effect was analyzed. For both wild type and P.1 variant, the NTD of truncated S1 domain represented a high RMSD but the extent of deviation was relatively slightly more in the P.1 variant (**Figure 2.19A**). Interestingly we observed that the anchoring RBM loop (Res 437-508) region showed comparatively high flexibility in the P.1 variant, possibly because of the interface mutations (K417T and E484K, N501Y) present on the loop. In contrast, this loop was observed to be slightly rigid in the wild type. Moreover, D614G mutation on the SD2 region of the P.1 variant also incorporates a significant high RMSD (**Figure 2.19A**). These observations suggest that mutations outside the RBD have a role in the conformational heterogeneity and may provide a conformational plasticity for the efficient binding with the ACE2 receptor. Similar studies have been investigated to analyze conformational variation in RBD as well as truncated S1 domain in complex with ACE2. Ray et. al. stated the effect of D614G on the conformational variation facilitating the RBD opening process from down to up state [Ray et al., 2021]. Conformational flexibility has also been analyzed within three important loops of the RBD (Res 474-485, 488-490, 494-505) and speculated to have a conformational variation for better binding with the ACE2 [Khan et al., 2021]. MDS of similar complexes has been investigated by another group [Verkhivker et al., 2021]. The observed flexibility in the RBM of the truncated S1 domain complex was similar to our observation. Overall, the effect of the mutations on the conformation

flexibility in the spike protein supports the earlier studies. This feature in the variants especially P.1 variant may provide an improved adaptability for binding with the ACE2.

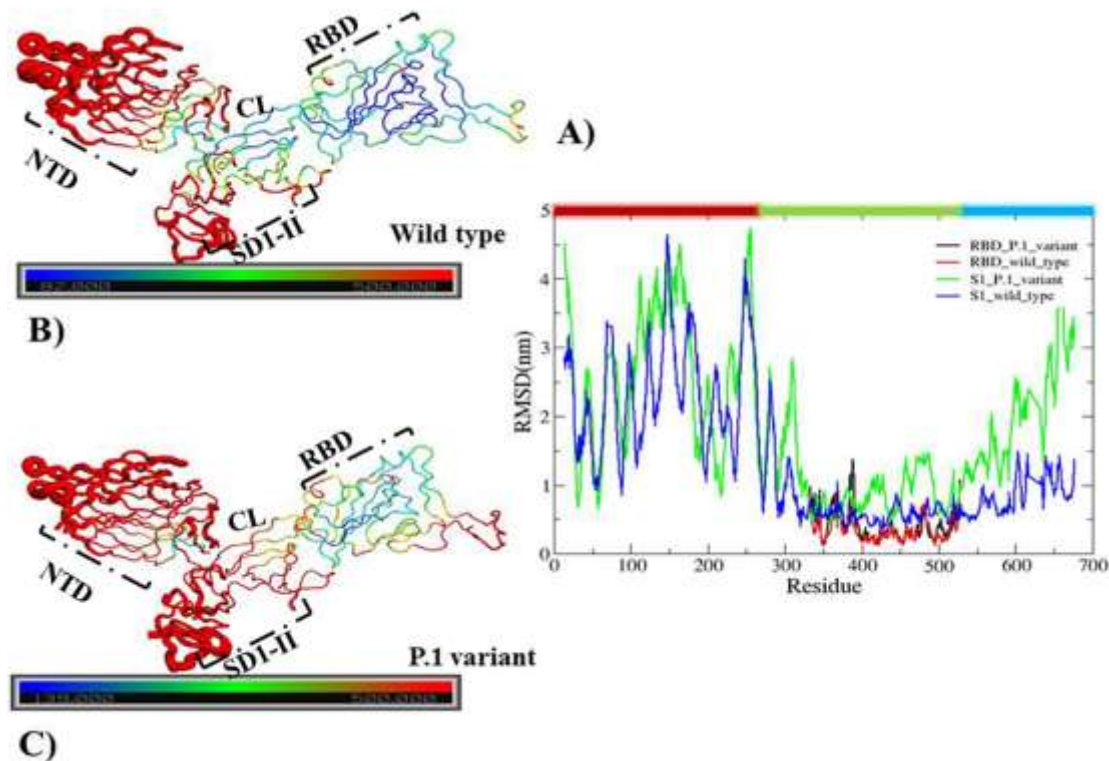


Figure 2.19. Representation of the conformational variation. (A) Residue-wise RMSD for the RBD as well as the truncated S1 domain. Shaded part at the top of the plot shows different regions of the S1 domain. Maroon, green, and cyan represent NTD, RBD, and CTD, respectively. (B) Residue-wise B-factor for the wild type (C) Residue-wise B-factor for the P.1 variant of the S1 domain. The B factor is represented according to the color thickness scheme where the blue region represents the rigid residues and red corresponds to the most flexible residues. Green and blue line is for the complete S1 domain of Spike protein. Black and red is only for RBD of Spike protein. CL, connecting loop; NTD, N-terminal domain; RBD, receptor binding domain; SD1-II, Sub domain 1 and 2.

Further, structural flexibility of the truncated S1 domain was also probed in terms of B-factor or thermal displacement. This has been an important parameter to analyze displacement of amino acid residues due to the thermal vibrations. The B-factor of each amino acid of the wild type and P.1 variant clearly indicates few variations among the two (**Figure 2.19 B&C**). The NTD of the both wild type and P.1 variant showed comparable B-factor values. However, significant thermal fluctuations were noticed in the connecting loop of the NTD and RBD. Similar fluctuation was also observed in the RBM region as well as the SD1-2 region of the P.1 variant (**Figure 2.19C**). These regions showed reduced B-Factor in the wild type (**Figure 2.19B**). In brief, our B-factor analysis for the truncated S1 domain structures also supports the similar flexibility changes with respect to our residue-wise RMSD deviation (**Figure 2.19**).

The PCA of the truncated S1 domain complex with the ACE2 from wild type and P.1 variant indicates that the P.1 variant covered comparatively larger space along the principal component 1 (PC1) and the PC2 (**Figure 2.20**). The global flexibility of both the complexes were examined by the trace of diagonalised covariance matrix of backbone atomic fluctuations. The diagonalised covariance matrix trace was higher for the truncated S1 domain complex from the P.1 variant compared to the wild type. This observation along with the PCA plots suggest that mutations outside the RBD are also involved in providing conformational variations or flexibility to the P.1 variant complex.

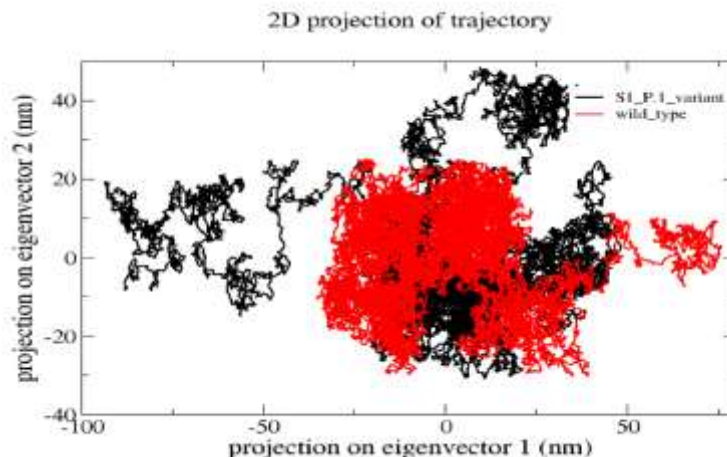


Figure 2.20. Representation of the conformational variation. Projection of dominant motions of backbone atoms in essential subspace along the first two principal eigenvectors of wild type and P.1 variant.

The change of secondary structure content analysis in the truncated S1 domain complexes from the both did not provide any noticeable difference in 100 ns time scale simulation. This suggests that the mutations do not incorporate any noticeable deviation in the secondary structure content. The mutations has sustained the stability at secondary structure content level throughout the 100 ns of MDS (**Figure 2.21**). However, calculated electrostatic charge distribution in S1-ACE2 complex in wild type and P.1 variant display a change in electrostatic charge distribution. Complex from P.1 variant observed to display more positive charge patches at the interface, compared to wild type (**Figure 2.22**). Positive charge in the P.1 variant may contribute to the better binding affinity. Earlier reports also suggested that positive charge patch on RBD of SARS-CoV-2 provide better binding with ACE-2 compared to SARS-CoV [Lan et al. 2020] Other study have also demonstrated that mutations at RBM affect the charge distribution at the interface [Dehury et al. 2021].

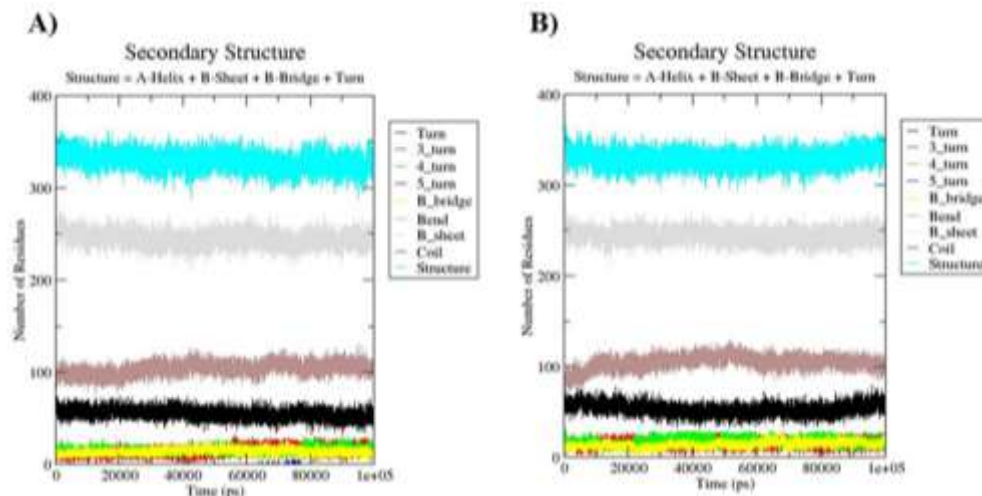


Figure 2.21. Secondary structure content analysis during the course of 100ns MDS for A) wild type and B) P.1 variant.

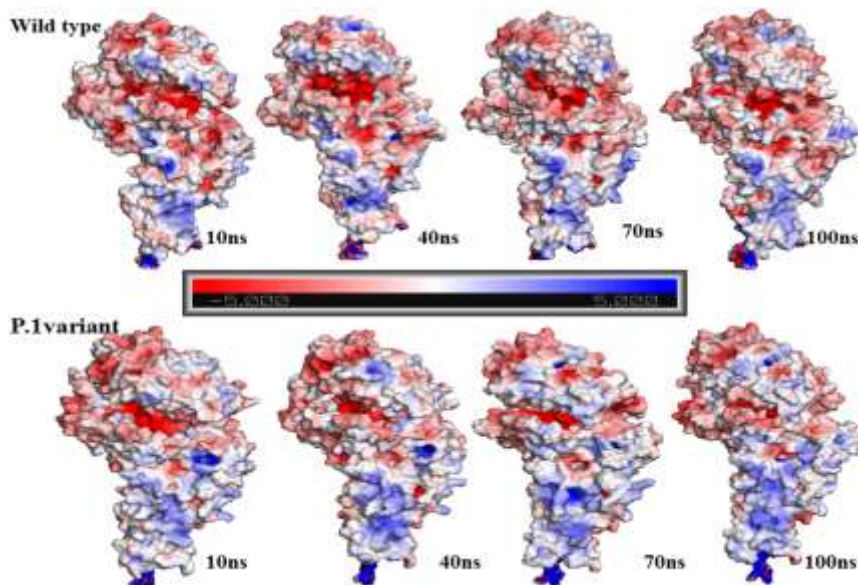


Figure 2.22. Electrostatic charge distribution at the interacting interface of S1-ACE2 complex of wild type and P.1 variant at different time frames of 100ns MDS trajectories (Color Scale: blue, red, and white colors represent positively charged, negatively charged, and neutral surfaces, respectively).

2.4 CONCLUSIONS

Emerging SARS-CoV-2 variants in each successive wave of infection have made this pandemic more threatening; where most of the mutations accumulated mainly in the Spike (S) glycoprotein. Since the Spike protein is the primary target of infection, concentrating on mutations of S protein leads to increase in transmission efficiency and escape from neutralizing antibodies. P.1 variant contains ten mutations on the S1 domain of the spike protein. Mutations on the RBD were reported to have an impact on increased affinity for receptor ACE2 receptor. So in our study we have used SARS-CoV-2 S protein wild type and its P.1 variant RBD complex with ACE2 to uncover the structural basis responsible for the increased affinity towards the ACE2 receptor.

We have applied integrative networking and dynamics approaches to study the small conformational variations. Our networking study highlighted a node wise rearrangement in the network parameters. Significant reduction in nodes and links was observed related to each parameter of the PCN. The reduction of nodes and links in P.1 variant signifies a loss of interactions. Additionally the mutations outside the RBD were observed to perturb the network properties of the NTD of the truncated S1 domain of P.1 variant. Further the effect of mutations were also predicted on the protein stability and flexibility through the vibrational entropy change ($\Delta\Delta S$) and free energy change ($\Delta\Delta G$) calculations. Both the analysis concluded that the interface RBD mutations are having a destabilizing effect. Comparative conformational variations among the RBD-ACE2 and truncated S1 domain-ACE2 were highlighted by MDS studies. Our MMPBSA study suggested the energetically favorable binding of the RBD with the ACE2 in the P.1 variant. Moreover, the conformational variations have also rearranged the type of interaction and interface interacting residues. The PCA of the S1 domain-ACE2 complex indicates an increased motion in the P.1 variant, which may be associated with the conformational heterogeneity in the

P.1 variant complex. Overall, our study provides the structural basis of better interaction of the P.1 variant and identifies unique RBD residues crucial for the interaction with the ACE2 receptor. This study may be useful for designing better therapeutics against the circulating P.1 variant and other future variants as well.

CHAPTER -3

Structure-based drug design of potential inhibitor targeting the SARS-CoV-2 Spike Protein: ACE2 interface

3.1 INTRODUCTION

Binding of the RBD of spike protein to the human ACE2 receptor is the primary step in the SARS-coronavirus infection process [Wang et al. 2020]. Several efforts have been put forward to inhibit this initial interaction by designing small molecules as a therapeutics [Razizadeh et al. 2021; Deganutti et al. 2020; Sisk et al. 2018]. But the emerging variants and their mutations, especially in the RBD region of spike protein, are making the efforts more strategic. The functional significance of the mutations has already been discussed in the previous chapter. Our previous chapter has highlighted the changes between the Wild type SARS-CoV-2 (WT) and its P.1 variant at their structural and contact level that are responsible for the increased affinity towards human ACE2 receptor [Lata et al. 2021]. This intrigues us to target the RBD-ACE2 complex of P.1 variant. The P.1 variant consists of three RBD interface mutations (K417T, E484K, N501Y) and have gained much attention since early of the variant storms [Dejnirattisai et al. 2021]. This chapter, briefs about the computationally designing of a potent inhibitor against the RBD-ACE2 complex of the P.1 variant.

It has been reported that SSAA09E2 compound {N-[[4-(4-methyl-piperazine-1-yl)phenyl]methyl]-1,2-oxazole-5-carboxamide} blocks the early interactions of SARS-CoV Spike protein with the human ACE2 receptor [Adedeji et al. 2013]. In another study, the mechanism of interference of the compound and its potency has been investigated [Roth et al. 2021]. The compound SSAA09E2 is reported to be effective in destabilizing the ACE2-RBD complex.

In our observation, the compound SSAA09E2 showed less affinity with the RBD-ACE2 complex from the P.1 variant suggesting its inefficiency towards the P.1 variant. In this study, we have

applied a fragment replacement approach to probe the fragment's space to be replaced with other active compounds [Shan et al. 2020]. We have screened various fragments and identified suitable fragments that meet the geometric requirements to be fitted within the specified local environments of the protein. Among all, two designed analogs showed a better affinity compared to the parent compound. Further we have performed 200ns Molecular Dynamics Simulation to analyze the structural changes of the RBD-ACE2 complex due to the binding of designed analogs. Our computational analysis highlights a stable interaction of analogs with the interface residues of the RBD-ACE2 complex. The binding of analogs in the complex led the interfacial disruption and triggered large-scale conformational changes in the complex.

3.2 METHODS

3.2.1 Preparation of the target protein coordinates

The coordinates of the RBD-ACE2 complex of WT SARS-CoV-2 (6MOJ) [Lan et al. 2020] and P.1 variant complex (7NXC) were downloaded from the protein database [Dejnirattisai et al. 2021]. Further, the structure was subjected to the what-if server to add missing residue side chain atoms. The crystal waters and hetero molecules were removed from the PDB file.

3.2.2 Fragment substructure selection and analog generation

A new analog was generated by using the fragment replacement approach implemented in the FragRep tool [Shan et al. 2020]. The piperazine ring of the SSAA09E2 inhibitor was substituted with the analogs present in the FragRep database. The FragRep extracts over half a million of fragments with three-dimensional structures present in the ChEMBL-25 database [Bento et al. 2014]. Suitable fragments that replace the substructure and meet the geometric requirements to remain a part of the compound and also fit in the local protein environment.

3.2.3 Molecular Docking and Ligand scoring

Molecular docking of the RBD-ACE2 complex with the selected two analogs as inhibitors was performed using MGL AutoDock4.2 tools [Morris et al. 2009]. The initial coordinates and topology parameters were generated for further dynamics analysis. Molecular docking was done between the RBD-ACE2 complex and the SSAA09E2 inhibitor. The polar hydrogen atoms and Kollman charges were added to the protein complex and the gasteiger charges were assigned to both the analogs. The grid box with dimension 68×68×54 was centered on the interface region of the RBD-ACE2 complex. The Lamarckian genetic algorithm (LGA) was used and generated 150 docked conformations for each compound. Then finally the best representative binding pose for both the analogs in complex with protein was selected based on the lowest binding energy.

3.2.4 Molecular Dynamics Simulation

The selected docked complex of analog and RBD-ACE2 protein were used to carry out classical MDS using GROMACS 5.1.4. The topology files of the protein and compound-complex were parameterized using GROMOS96 45a3 force field and PRODRG2 server, respectively. The protein compound complexes were implemented for solvation using the SPCE water model to add water molecules contained in a cubic box with water molecules of 1.5nm to the box wall from the protein's surface. Further, to neutralize the solvated systems, Na atoms as counter ions were added to the solvated systems. To minimize the wrong short-range contacts after adding counter ions, a quick steepest descent energy minimization algorithm for 50,000 steps was applied until the largest acting force on the system becomes smaller than 1000 kJ/mol/nm. The minimized complex was set for NVT and NPT equilibration by restraining a constant number of particles, volume, and temperature ensemble for 500ps and a constant number of particles, pressure, and temperature

ensemble for 500ps pressure equilibration. To ensure the proper equilibration of the complex, the system's thermodynamic properties (like density, pressure, temperature, and potential energy) were monitored. The PME (Particle Mesh Ewald) method was used to calculate the long-range electrostatics. Modified Berendsen thermostat and Parrinello-Rahman barostat were used for the temperature and pressure coupling, respectively. Finally, the equilibrated complex system was set to MD production runs for 100ns at 310K and 1bar atmospheric pressure. The graphs were generated using the XMGRACE, and GNUPLOT.

Principal component Analysis (PCA) was performed to calculate the mass-weighted covariance matrix on the MD trajectories by calculating and diagonalising the covariance matrix to get a set of eigenvectors and eigenvalues which reflects the initial motion of the atoms. G_covar package is in-built in the GROMACS and is used to extract the covariance matrix of the backbone atoms. The principal components were projected through the g_anaeig package.

3.2.5 Binding Free energy Calculations

The MMPBSA method implemented in the g_mmpbsa package was used to calculate the binding free energy between protein-compound complexes. The last 100ns MDS frames were integrated for binding free energy calculation. The molecular mechanic's energy includes electrostatics, Van-der-Waals interactions, polar, non-polar solvation energy.

3.3 RESULTS

3.3.1 Binding of SSAA09E2 on RBD-ACE2 interface for WT and P.1 variant

The molecular docking results show higher binding affinity of SSAA09E2 towards the RBD-ACE2 complex of WT as compared to the P.1 variant (**Table 3.1**). The orientation of SSAA09E2 was changed in the P.1 variant complex compared to the WT complex. The oxazole group of

SSAA09E2 was observed to interact with the ACE2 interface residues in the WT complex. In addition, the piperazine group of SSAA09E2 was facing towards the RBD interface region in the WT complex (**Figure 3.1A&C**). However the similar geometry was not observed in the P.1 complex. The geometry of the SSAA09E2 was flipped and the core group of the SSAA09E2 shifted from the binding cavity (**Figure 3.1B&D**) in the P.1 variant complex. The differential binding affinity and orientation of SSAA09E2 may have different inhibitory effects on the RBD-ACE2 complex of the WT and P.1 variant. Hence, we hypothesized to design an analog of SSAA09E2 to improve the interactions with the RBD-ACE2 interface residues of the P.1 variant.

Complex	Binding score (Kcal/mol)
SSAA09E2_WT	-4.82
SSAA09E2_P.1	-4.20

Table 3.1. Docking score for SSAA09E2 with the RBD-ACE2 complex of the WT and P.1 variant.

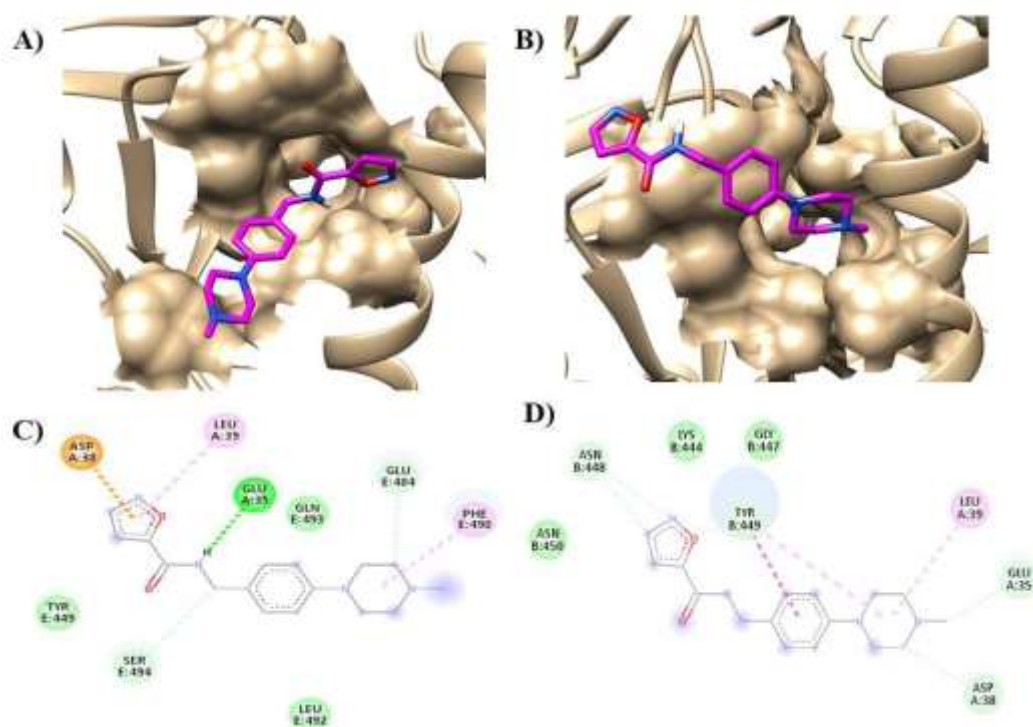


Figure 3.1. Representation of docking of the SSAA09E2 A) Best docking pose for SSAA09E2 in the binding cavity of RBD-ACE2 interface for WT. B) Best docking pose for SSAA09E2 in the binding cavity of RBD-ACE2 interface for P.1 variant. C) 2-D interactions plot for SSAA09E2 with RBD-ACE2 complex for WT, chain A and chain E indicates ACE2 and RBD respectively. D) 2-D interactions plot for SSAA09E2 with RBD-ACE2 complex for P.1 variant, chain A and chain B indicates ACE2 and RBD respectively.

3.3.2 Computationally designed SSAA09E2 analogs and Binding conformation

The FragRep tool probed the suitable fragments to replace the piperazine group of the SSAA09E2 compound and generated a library of 1,924 analogs. The top two analogs named compound 1 and

compound 2, were selected on the basis of binding energy and geometry in the target site (**Figure 3.2**). Compound 1 consists of a tetrachloro oxapentacyclo group, a substituted piperazine group of the SSAA09E2. This group is linked to the rest of the fragment of the parent compound which includes a cyclohexyl, carboxamide and oxazole group in the tail region. Compound 2 is generated by attaching a naphthalene group in place of the piperazine group of the SSAA09E2. These two compounds fall within Lipinski's rule of five criteria. Compound 1 docked inside the interface cavity of RBD-ACE2 and displayed a binding score of -5.83 Kcal/mol (**Figure 3.3A**). The oxazole group compound 1 was observed to interact by making three potential hydrogen bonds with the RBD interface residues such as GLN 493, SER 494 and TYR453. These interactions were not observed in the parent molecule-protein complex (**Figure 3.3B&C**). Moreover, the N-atom of carboxamide of compound 1 is involved in the H-bond interaction with TYR 449 of the RBD. In addition, the epoxy group from the head region is also observed to form a H-bond with the LYS 484 of the RBD interface. The LYS484 residue is an important mutation in the variants. The 2D and 3D interaction plot also displayed one pi-sigma and pi-alkyl bond with the PHE490 residue of the RBD. The ACE2 interface residues are mostly involved in the van-der-Waals interactions except for HIS34 residue, which forms a carbon-hydrogen bond with the carboxamide group (**Figure 3.3B &C**).

The best-docked pose of Compound 2 is displayed in the interface cavity (**Figure 3.4A**) with a binding score of -6.24 Kcal/mol (**Figure 3.4A**). The oxygen atom of the naphthalene group is making H-bond with the ARG403 residue of the RBD. The Oxygen atom of the carboxamide group also forms a H-bond with the ARG403 of the RBD (**Figure 3.4 B & C**). The head region of Compound 2 was flipped towards the ACE2 interface and HIS34 of the ACE2 is participating in the pi-cation and pi-alkyl interactions with the substituent group. Additionally, two H-bonds are

formed with the TYR495 and TYR505 with the RBD interface. Overall the molecular docking observation suggests that the analogs can potentially interfere with the important RBD-ACE2 interface interaction. The interface interacting residues have already been characterized as important drugable targets as mentioned in the previously reported study.

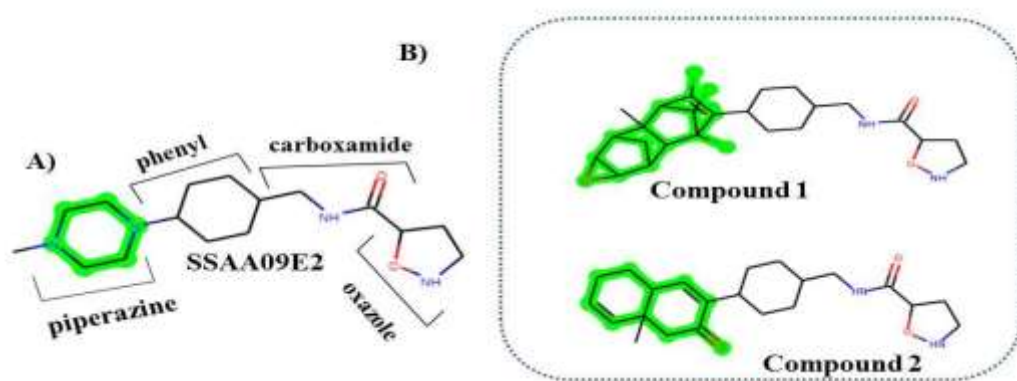


Figure 3.2. 2D representation A) 2D structure of parent compound SSAA09E2. B) Two analogues designed from FragRep search (namely Compound 1 and Compound 2). Substituted region has been highlighted in green color.

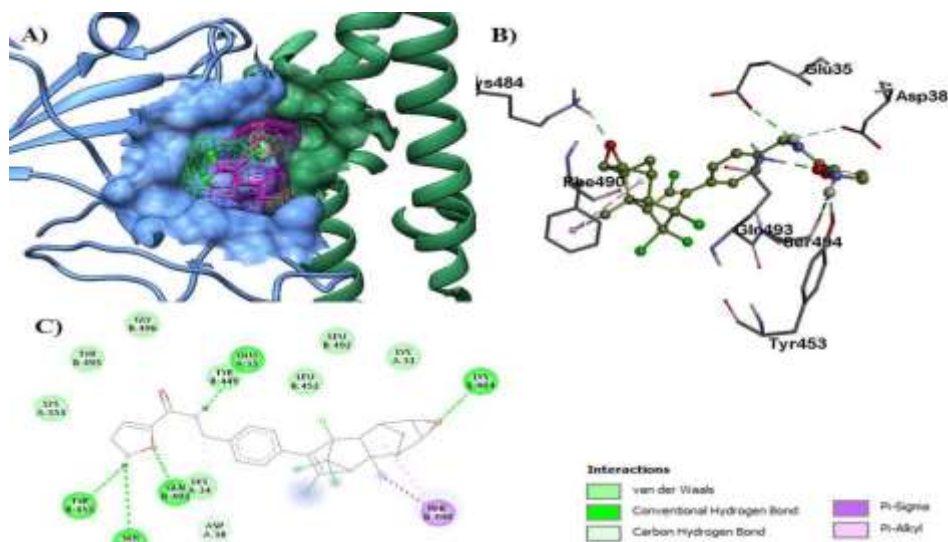


Figure 3.3. A) The best docking pose of Compound 1 in the RBD-ACE2 interface binding cavity. B) 3D interaction plot between compound and RBD-ACE2 complex residues. C) 2D interaction plot between compound and RBD-ACE2 complex residues.

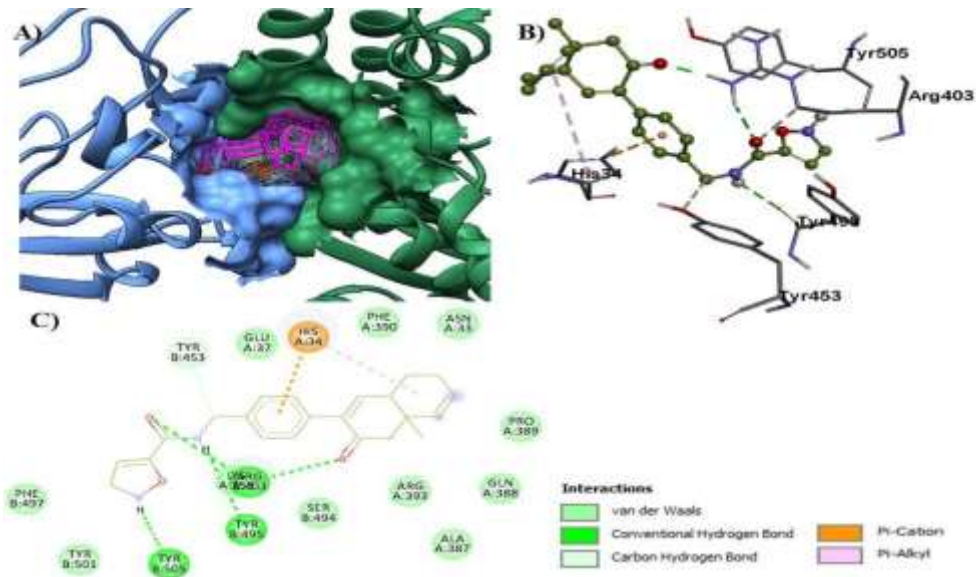


Figure 3.4. A) The best docking pose of Compound 2 in the RBD-ACE2 interface binding cavity. B) 3D interaction plot between compound 2 and RBD-ACE2 complex residues. C) 2D interaction plot between compound 2 and RBD-ACE2 complex residues.

3.3.3 Structural stability of compounds in the RBD-ACE complex

The stability of protein-compound complexes was evaluated and backbone RMSD for the protein and compounds were calculated during the time scale of 200ns MDS. The protein backbone RMSD for both the complexes was stable after 50ns of MDS. In the case of the compound 1 bound complex, the compound was re-equilibrated at 100ns time frames, and its RMSD was increased to 0.2nm. Later, the RMSD dropped to 0.2Å after 110ns. The RMSD trajectory of compound 2 inside the interface cavity of the RBD-CE2 was relatively stable throughout 200ns of MDS (Figure 3.5A & B). The protein ligand interaction profile was generated on the representative structure obtained after clustering analysis from MD simulation (Figure 3.6). The interaction profile indicates that compound 1 is stable inside the cavity and the N-atom of carboxamide and oxazole group of compound 1 maintains hydrogen bonds interaction with Tyr449 & yr451) of the RBD. The oxygen atom of the carboxamide group of compound 2 was forming three H-bonds- one with His34 residue of the ACE2 and two with Arg403 and Tyr453 residues of the RBD. These interactions provide stability to the compounds in the interface and suggest that the compounds have the potential to interfere with the interface residues of the RBD-ACE2 complex.

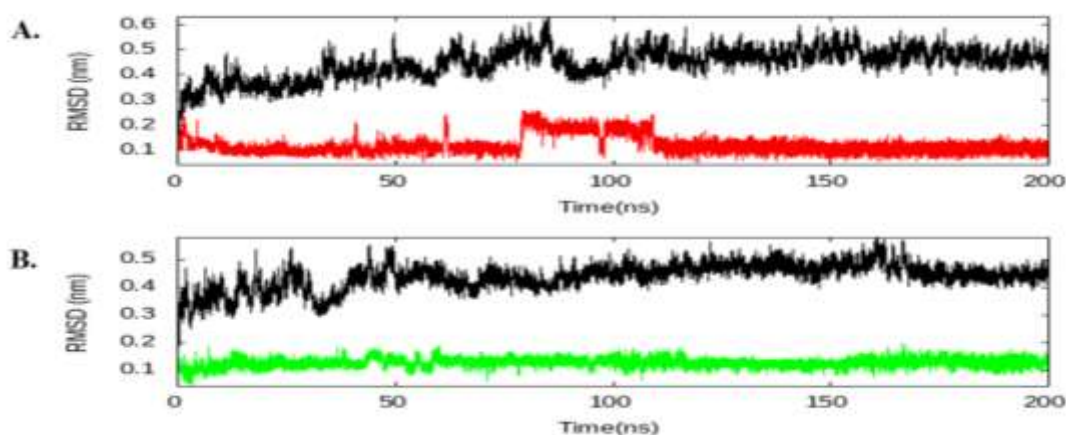


Figure 3.5. RMSD plot at 200ns (A) protein backbone (black) and compound1 (red). (B) Protein backbone (black) and compound2 (green).

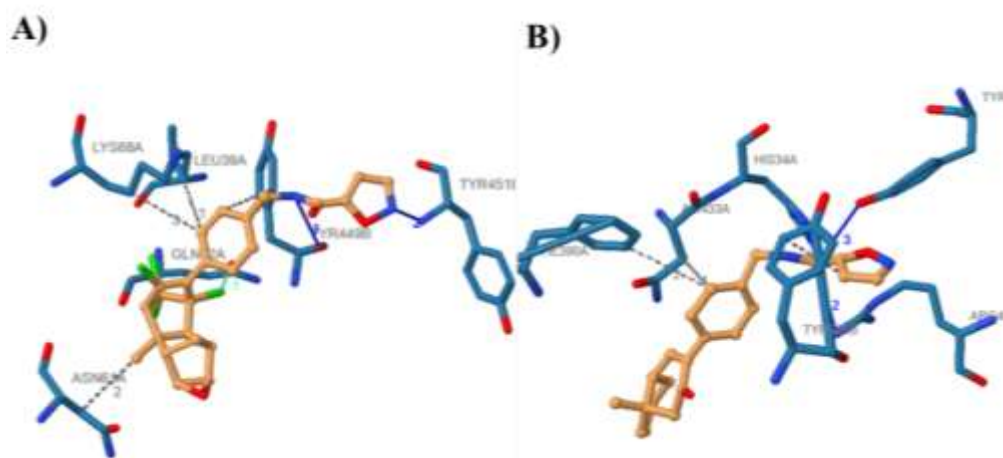


Figure 3.6. Protein compound interaction profile on the representative structure after 200ns of MDs. Hydrogen bonds are in blue, hydrophobic interactions are in dashed grey, halogen interactions are in green.

Additionally, the average number of hydrogen bonds formed between the compounds and the RBD-ACE2 complex was analyzed (Figure 3.7). On an average, compound 2 has 0.13 hydrogen bonds during the 200ns with the ACE2. The average hydrogen bonds between compound 1 and the ACE2 were relatively lower and displayed approximately 0.06 bonds. . The average number of hydrogen bonds with the RBD was 0.87 and 1.2 for compounds 1 and 2, respectively. Both compounds form at least one hydrogen bond with the RBD and the ACE2 during the 200ns of simulations trajectory. This observation further confirms the stability of both the compounds inside the cavity.

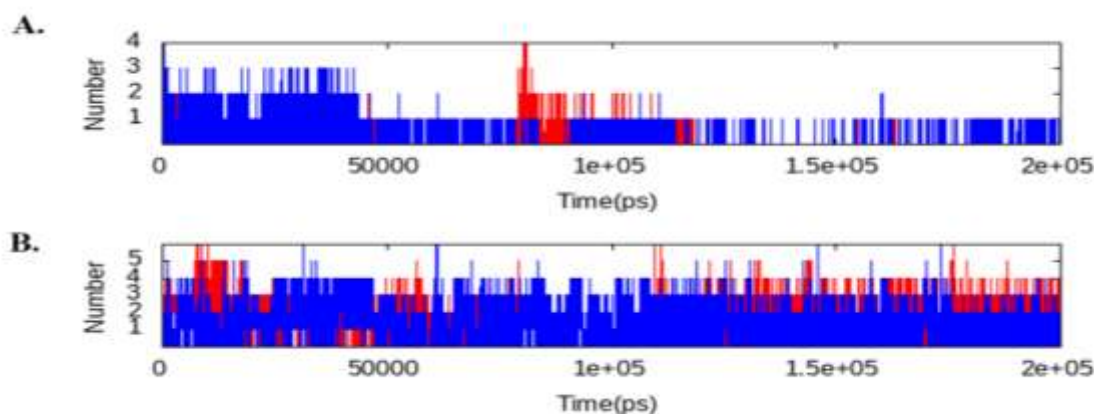


Figure 3.7. Hydrogen plot during 200ns MD. (A) Compound 1 bound complex (B) Compound 2 bound complex. Compound 1 is in red and compound 2 is in blue.

Further to analyze the compactness of proteins, the Radius of Gyration (Rg) was analyzed (Figure 3.8). The overall Rg value of compound 2 complex was observed relatively decreased compared to the apo and compound 1 bound complex. The Apo and compound 1 displayed a similar

trajectory to each other. This observation suggests that the binding of compounds doesn't influence the shape and size of the complex during the 200ns MDS.

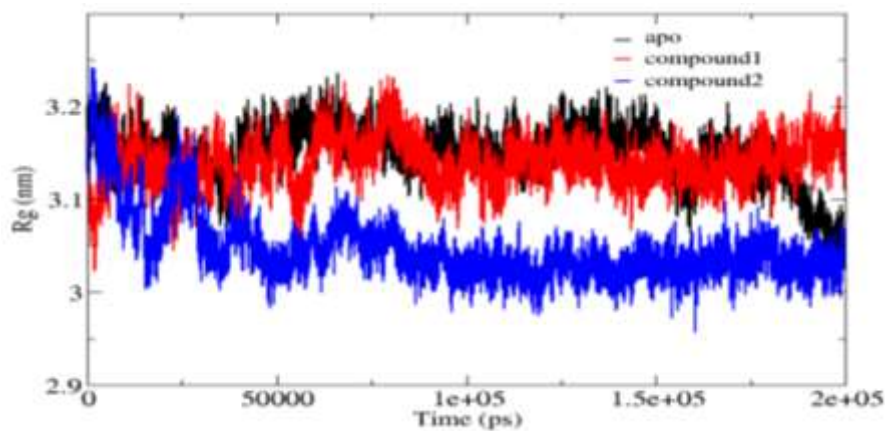


Figure 3.8. Radius of gyration plot during 200ns of MDS. Apo is in black, compound1 bound complex is in red and compound2 complex is in blue.

The stability of the compounds was further analyzed by binding free energy, calculated using MMPBSA on the last 30ns of the RBD-ACE2 complex. The thermodynamics parameters of the complex were also calculated and mentioned in (Table 3.2). Both the compounds displayed negative free energy inside the binding pocket. Compound 1 showed a relatively more negative free energy value. The electrostatic and Van-der-Waals interactions were mainly contributing to this negative energy. In contrast, compound 2 possessed slightly higher binding free energy. These observations demonstrate that the residue-wise energy contribution from the ACE2-RBD complex is different in both cases. In order to extract residue-wise energy contribution for both the ACE2-RBD complex, the total binding free energy was decomposed into per amino acid residue energy contribution as indicated in (Figure 3.9). Compound 1 bound complex, residues 13, 38, and 39 of the ACE2 were observed to contribute more to the negative free energy. Similarly, residues

431, 436, and 466 of the RBD were found to contribute to free energy. In the compound 2 bound complex residue numbers 12, 16, 335, 364, 368, and 370-372 of the ACE2 were observed to contribute to the free energy. These observations suggest that these interface residues play a significant role in the complex formation.

System	Van-der-waal energy	Electrostatic energy	Polar energy	SASA energy	Total binding energy (kJ/mol)
Compound 1	-179.6 +/- 10.6	-206 +/- 31.6	201.9 +/- 31.9	-16.9 +/- 1.05	-201.2 +/- 18.7
Compound 2	-207.3 +/- 12.3	-80.17 +/- 20.3	157.4 +/- 24.4	-16.8 +/- 0.9	-146.9 +/- 17.5

Table 3.2. Binding free energy analysis of compound1 and compound 2 with the protein complex.

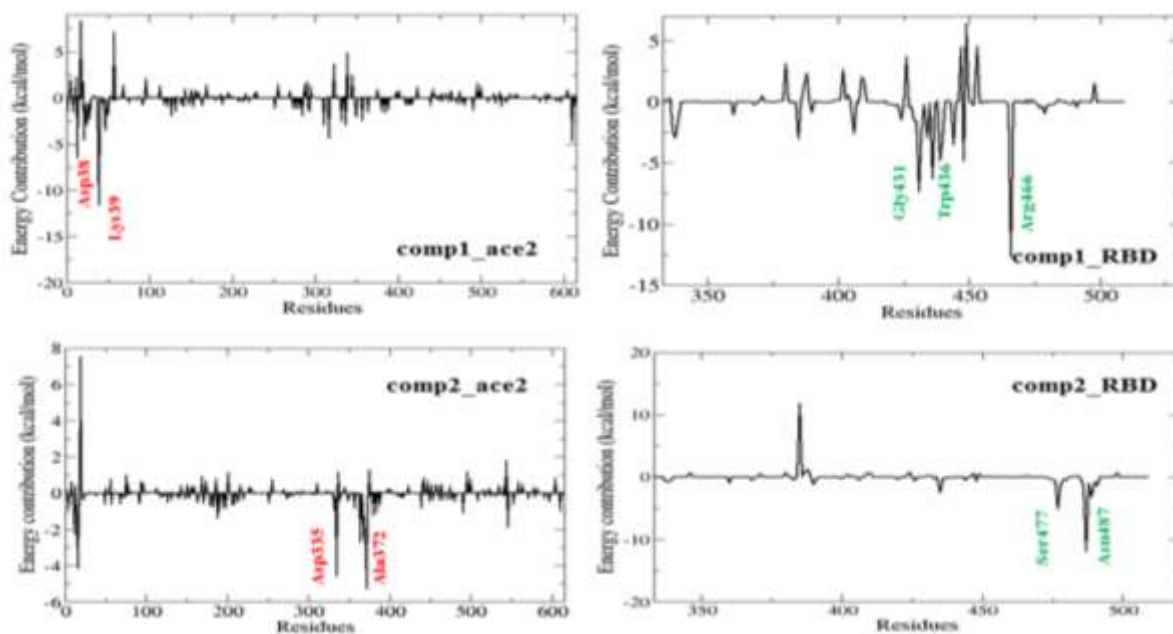


Figure 3.9. Residue wise energy decomposition plot. A) Compound 1. B) Compound 2

3.3.4. Effect of designed analogs on the RBD-ACE2 binding

The post simulations results indicated that both the compounds have direct interaction with the interface residues of the RBD-ACE2 complex. It suggests that both analogs may have the potential to disrupt important interfacial interactions. Considering this, changes in RMSF due to the binding of both compounds on the RBD-ACE2 complex were analyzed (Figure 3.10). The addition of both compounds doesn't cause any significant change in the RMSF of the ACE2 compared to the apo complex. However, the RMSF of residues 380-500 of the RBD region was increased significantly for both compound 1 and compound 2 bound complexes, suggesting high flexibility in that region. The mentioned region of the RBD mainly includes the interface residues. The increased flexibility can lead to higher entropy penalty which negatively influences the binding energy. Moreover, changes in flexibility were also observed in the region far from the binding interface. A noticeable increase in flexibility was in the few loops and extended β -sheet core of the RBD, suggesting that both the compounds may have some secondary effect on the functional conformation of the RBD of the spike protein. . Our residue-wise RMSD plot was also correlating with the RMSF analysis (Figure 3.11).

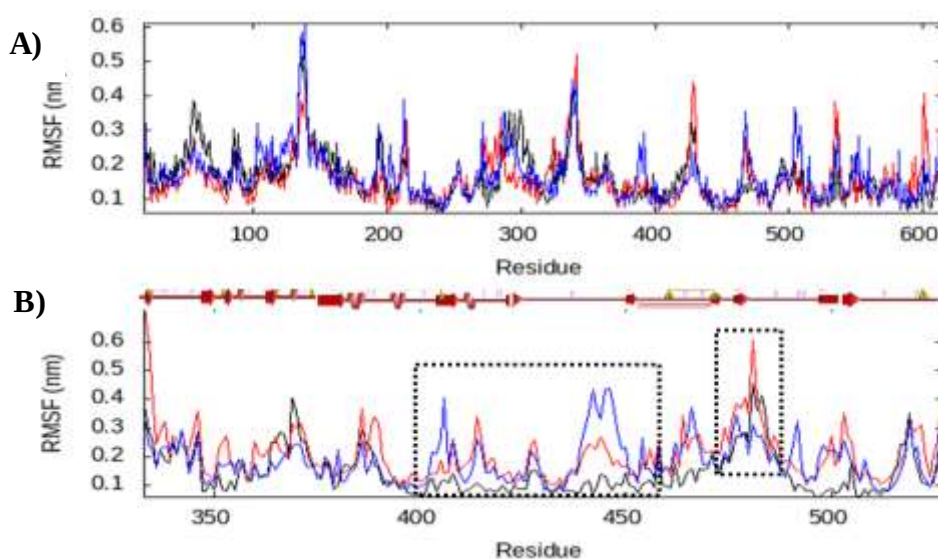


Figure 3.10. RMSF plot (A) ACE2 (B) RBD) apo is in black, compound 1 bound complex is in red and compound 2 bound complex is in blue.

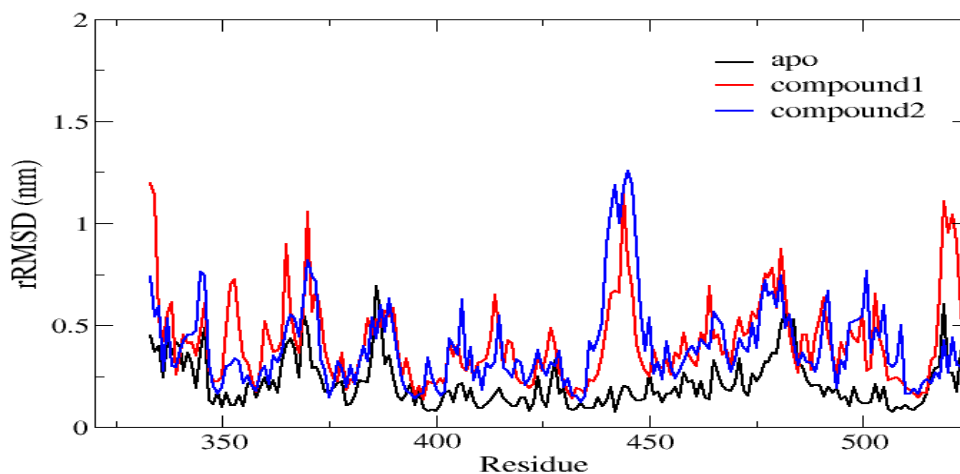


Figure 3.11. Residue-wise RMSD for apo and compound bound complex.

A consequence of compound binding was also investigated in terms of structural flexibility, by probing the changes in the B-factor. This is an important parameter to analyze the displacement of amino acid residues due to thermal vibrations. The thermal fluctuation was observed to increase

in the NTD of the RBD in compound 1 bound complex (Figure 3.12B). This domain was relatively less flexible in the compound 2 bound complex (Figure 3.12C). A significant increase in the B-factor of the 3 α -helix of the RBM region was observed in compound 2 bound complexes compared to the Apo form (Figure 3.12A). The B-factor analysis suggests that there is an alteration in the movement of particular segment of the RBD due to the binding of compounds.

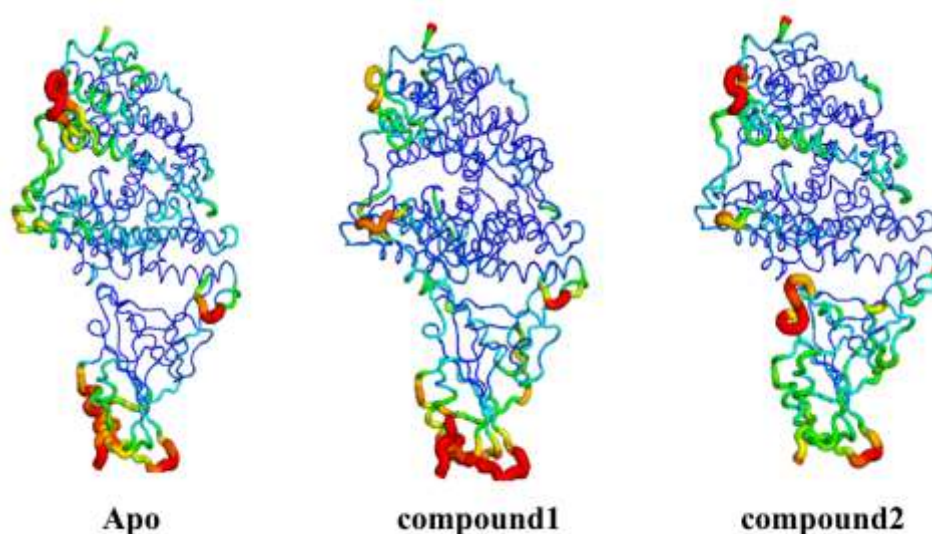


Figure 3.12. B factor analysis on representative structure obtained after 200ns MDS.

3.3.5. Effect of analogs as potential disruptors on interfacial residues

To measure the effect of compounds on interfacial residues disruption, distance between the key interface interacting residues involved in Hydrogen bond interaction was investigated. The H-bond distance among the key interacting residues was increased significantly for both the compound bound complexes except for the RBD487-ACE283 (Figure 3.13). The increase in the distance may bring about a disruption in the interfacial residues interaction. This disruption may reduce the binding of RBD to ACE2. Additionally, a change in the center of mass between RBM of S-

protein and $\alpha 3$ - $\alpha 4$ helix of the ACE2 in the compound-bound state was analyzed. It has been known that the ACE2 consists of two domains sub-domain I and sub-domain II; $\alpha 1$ and $\alpha 2$ helices interact with the RBM lies in subdomain I while subdomain II consists of $\alpha 3$ - $\alpha 4$ helices. These subdomains undergo a hinge binding movement and exhibit a closed and open conformation to accommodate the substrate/inhibitors. Binding of compounds bound in the RBD-ACE2 complexes brings about change in the distance of center of mass of the RBM and $\alpha 3$ - $\alpha 4$ helices of the ACE2 (Figure 3.14). The significant increase of distance due to binding of compounds may have some impact on the RBD-ACE2 interaction and may destabilize the interaction. The superimposed structure of the RBM region of S-protein interacting with $\alpha 1$ & $\alpha 2$ helix of ACE2 shows a significant displacement of $\alpha 1$ and $\alpha 2$ helix along with the RBM region (Figure 3.15).

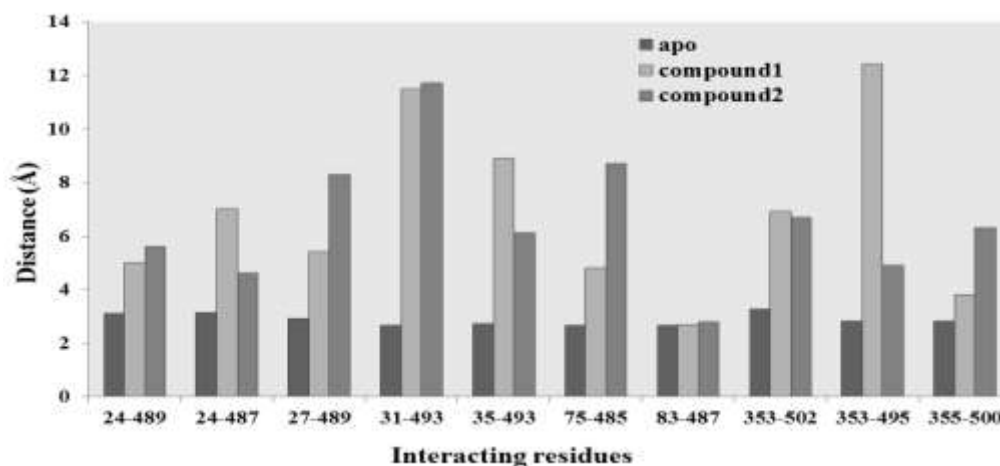


Figure 3.13. Hydrogen bond distance among key RBD: ACE2 interface residues

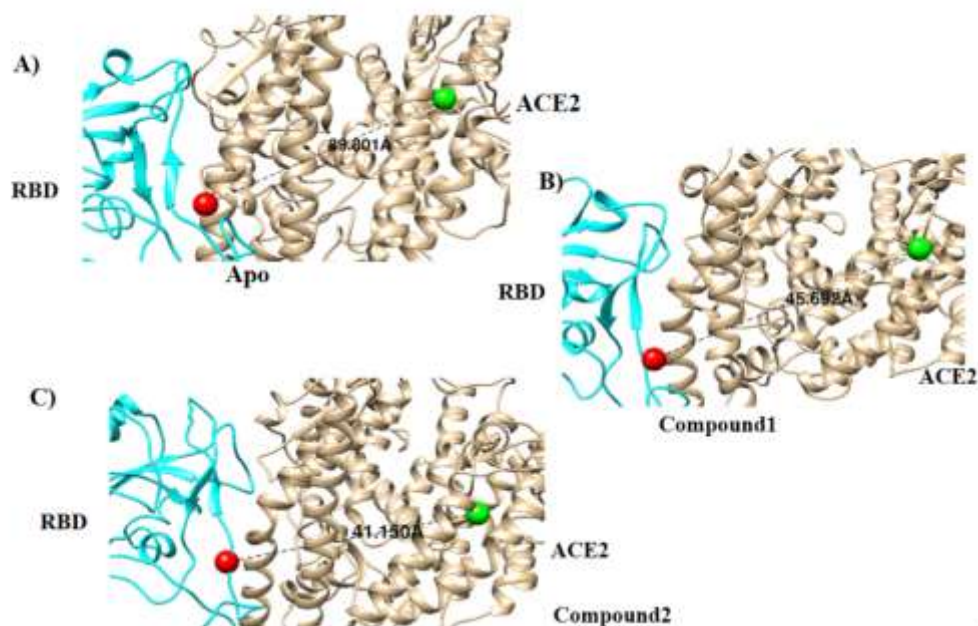


Figure 3.14. Centre of mass measurement between a3-a4 helix of ace2 and RBM of spike protein.

A) apo, B) compound 1 complex C) compound 2 complex

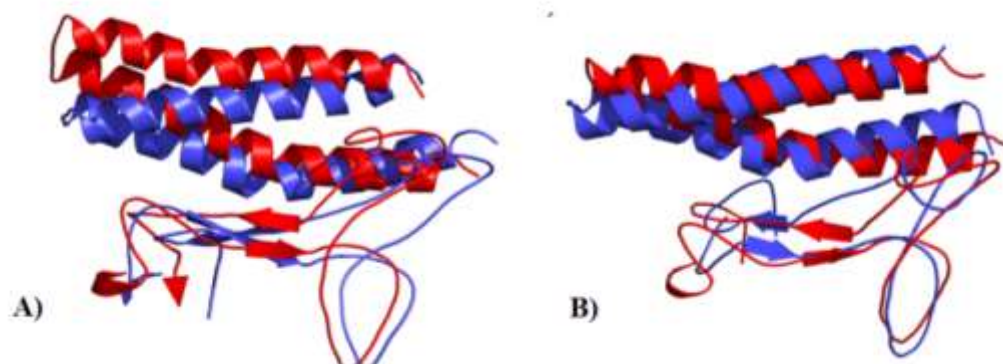


Figure 3.15. Superimposed structure of interacting ace2 and RBM complex for A) compound1 complex, B) compound2 complex. Apo is in red and compound bound complexes are in blue.

3.3.6. Effect of analogs as potential disruptors on the RBD-ACE2 SASA interface surface

We have measured the interface SASA area for apo and compound bound complexes. For WT the average SASA value is 118 nm² while for both the complexes it has increased to 116 nm² (Figure 3.16). It clearly indicates that the binding of compounds have increase the SASA value for the RBD-ACE2 interface segment as compare to the WT. It means binding of both the compounds interfere with the interface residues and weakens the potential interacting strength of RBD and ACE2 which leads to the opening of interfacial area and getting more exposed to solvent molecules.

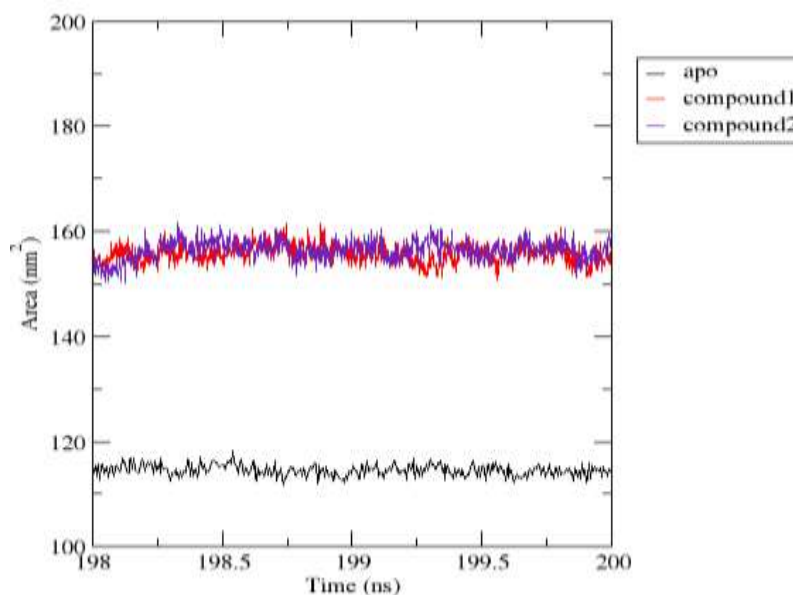


Figure 3.16. Interface SASA profile for apo and Compound bound complexes.

3.3.7 Effect of the compounds on the Essential Dynamics of protein complex

The effect of compounds in the complex was further analyzed on the internal motion of the RBD-ACE2 complex. The dominant motion modes of the RBD and ACE2 were extracted and employed for the Essential dynamics or PCA. The covariance matrix of fluctuations was calculated for the

backbone atoms and diagonalized to the first eigenvectors and eigenvalues. The sampling area of the apo and compound bound system was probed by projecting the trajectory into the essential space corresponding to their eigenvectors. The majority of the global fluctuations come from the first few numbers of low-frequency eigenvectors. The eigenvalues and their percentage of variance are shown in Figure 3.17. The trace of covariance matrix was 84, 92.7, and 84.9 for the apo, compound 1, and compound 2 bound complexes, respectively. This indicates that the first three eigenvectors account for 59%, 51%, and 40% of overall fluctuations in the apo, compound 1, and compound 2 bound complexes, respectively. The first eigenvector has a more significant percentage of variance for each case, reflecting the importance of the first eigenvector in the system dynamics.

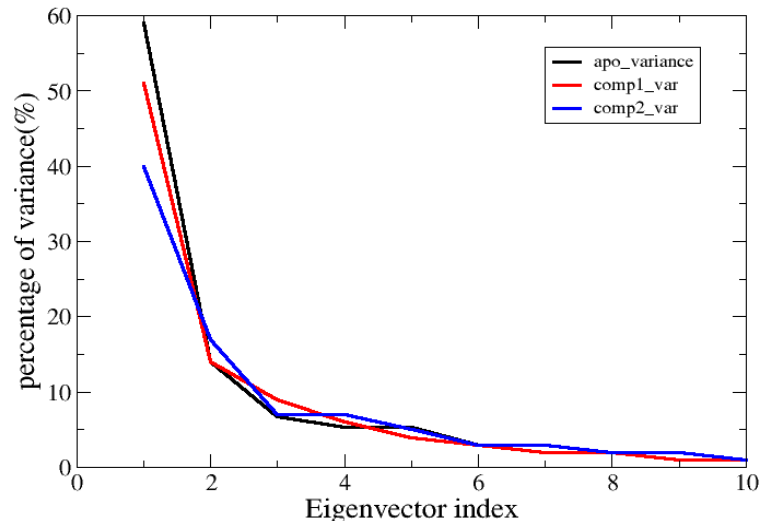


Figure 3.17. Percentage of variance for eigenvectors.

The Apo and compound bound RBD-ACE2 complexes were subjected to conformational space sampling. The conformational space of the compound 1 bound complex was observed to occupy

a greater area along its first and second principal components (Figure 3.18). At the same time, the apo and compound 2 bound complex showed almost similar conformational flexibility along the PC1 and PC2 (Figure 3.18). The porcupine plots were used for the first eigenvector to characterize dominant dynamics. The mode vectors for backbone atoms along the direction of its eigenvector are represented in (Figure 3.19). The length and direction of each arrow represent the magnitude and direction, respectively. For the compound 1 case, a significant motion was observed in the NTD and RBM of the RBD. The direction of motions in the RBD and interacting α -helix for the apo and compound bound complexes were observed differently. In addition, a change in the direction of the ACE2 movement was also noticed. The porcupine plot indicates the effect of bound compounds on the essential dynamics of the RBD-ACE2 complex. These observations suggest that the binding of compounds induce completely different large-scale motions in the RBD-ACE2 complex.

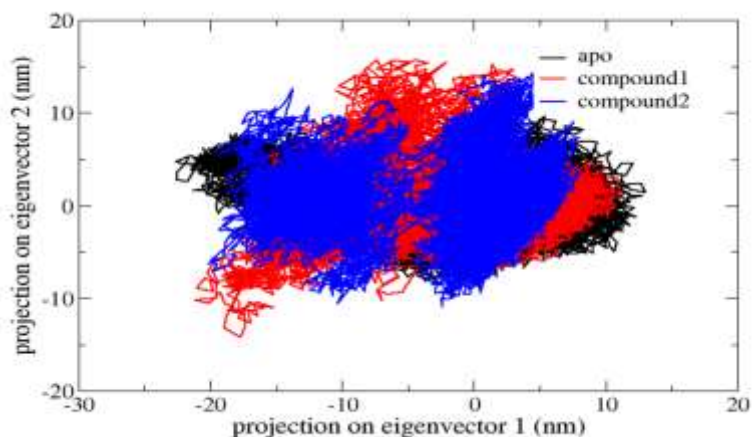


Figure 3.18. PCA plot along its two major eigenvectors.

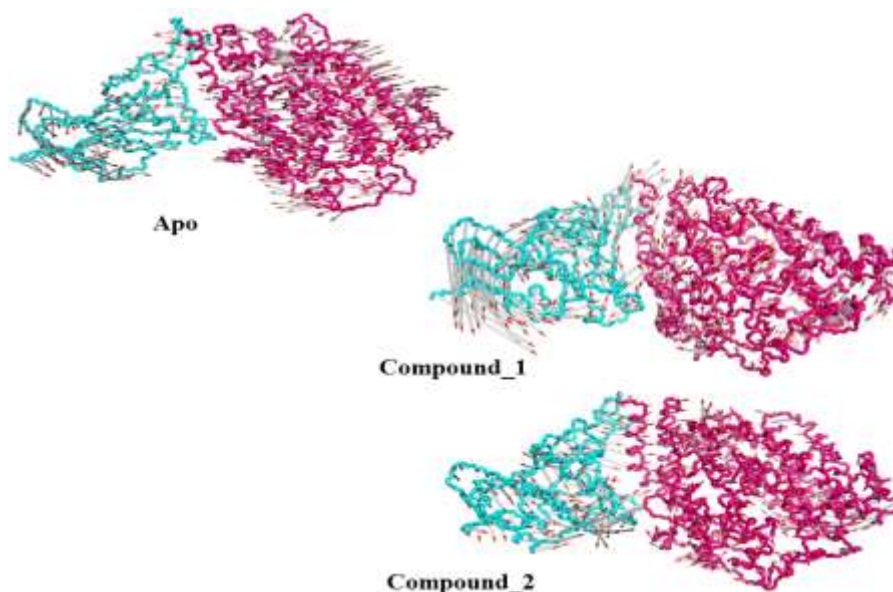


Figure 3.19. Porcupine plot of apo and compound bound complexes.

3.4 CONCLUSIONS

Our computationally designed analogs of parent compound SSAA09E2 which was a validated drug against SARS-CoV1, were employed in Molecular docking and Molecular dynamics study to analyze the disruption mechanism of RBD-ACE2 interaction. Our results suggest that both the compounds are having stable interaction with the interface residues which can intervene in the binding process of the protein complex. Overall our work suggests that both the compounds have potential to induce the structural changes both locally and globally which needs to be tested further for a better therapeutic against the emerging variants.

CHAPTER 4

Investigation of protein structural parameters in SARS-CoV-2 M^{pros}

THIS CHAPTER IS PUBLISHED IN:

Lata, S., & Akif, M. (2021). Comparative protein structure network analysis on 3CL^{pro} from SARS-CoV-1 and SARS-CoV-2. *Proteins*, 89(9), 1216–1225

4.1 INTRODUCTION

Coronaviridae family of virus usually possesses enveloped, positive sense RNA virus that generally includes three highly pathogenic viruses such as Severe Acute Respiratory Syndrome Coronavirus 1 (SARS-CoV-1), Middle East Respiratory Syndrome Coronavirus (MERS-CoV) and SARS-CoV-2 [Cui et al. 2019]. SARS-CoV-1 originated in China and caused a global pandemic in 2003 with about a 10% fatality rate [Zhong et al. 2020; Ksiazek et al. 2003]. MERS-CoV was first reported from Saudi Arabia in 2012 and has infected the human population with limited human-to-human transmission [Zaki et al. 2012]. SARS-CoV-2, a new coronavirus reported for the first time from Wuhan, China in December 2019, causes severe human respiratory disease [Wu et al. 2020]. It has also been characterized as a very contagious pathogenic virus with rapid transmission capability among human-to-human that has caused an outbreak of the severe pulmonary diseases in almost 216 countries, resulting in approximately 814,438 confirmed deaths globally till date (WHO report, 2020). WHO has coined SARS-CoV-2 causing disease as COVID-19 pandemic and that has now become a global health emergency and expected to have severe ramification on the global economy.

Currently, there is no specific treatment available to control COVID-19 pandemic. Efforts are being made towards design of vaccines as well as drugs against COVID-19. In the past, therapies have been developed against SARS-CoV-1 targets such as proteases, helicases and polymerases. Moreover, Immune modulators such as interferons and corticosteroids have also been used as therapeutics [Zumla et al. 2016]. Among the viral targets, main protease (M^{pro}, 3CL^{pro}) has been designated as an important drug target because of its essential role in the processing of polyprotein translated from the viral RNA [Anand et al. 2002; Yang et al. 2003]. The M^{pro} is a homodimer cysteine protease where each protomer consists of three domains, domain I (residue

8-101), domain II (residues 102-184) and domain III (residues 201-303) and catalytic residues and the substrate binding sites are situated between domains I and II of M^{pro} [Anand et al. 2002]. Recent crystal structure of SARS-CoV-2 M^{pro} ⁹ reveals its structural similarity with M^{pro} of SARS-CoV-1 and has a high degree of sequence identity (96.1%) among the two [Zhang et al. 2020]. Previous studies showed that HIV-1 protease inhibitors block SARS-CoV-1 M^{pro} [Savarino et al. 2005]. Hence, having a structural similarity to the M^{pro} from SARS-CoV-1 and SARS-CoV-2, the known inhibitors could also impart a similar effect on the M^{pro} of SARS-CoV-2. However, the HIV protease inhibitors show different binding effect on M^{pro} of SARS-CoV-2 [Ortega et al. 2020]. One of the HIV protease inhibitors, Lopinavir, was shown to inhibit M^{pro} of SARS-CoV-1, *in-vitro* [Wu et al. 2004]. While, none of the HIV inhibitors was able to significantly inhibit M^{pro} of SARS-CoV-2, *in-vitro* [Mahdi et al. 2020]. Other known potent inhibitors such as α -ketoamide and N3 are also reported to have differential inhibition on the activity of M^{pro}s from SARS-CoV-1- and SARS-CoV-2 [Zhang et al. 2020; Jin et al. 2020; Xue et al. 2007].

A small rearrangement of protein at the structural level by the substitution of a few amino acids at the substrate binding pockets or allosteric sites, results in changes in internal interactions which may lead to differing patterns of inhibitor sensitivity. Similar might be the case with the M^{pro} from SARS-CoV-2, where few changes in the amino acid sequence in comparison to M^{pro} of SARS-CoV-1, may contribute towards the differential effect on M^{pro}s. Since no significant structural changes are noticed at the active site, a subtle change of interactions at the allosteric sites of the proteins may have an effect of sensitivity of the inhibitors. Hence, a protein structure network (PSN) based approach can investigate the negligible conformational changes associated in the protein structure. A PSN is mainly depicted on a protein structure as a system of networks that comprises nodes and links. Nodes are represented by amino acid residues and links are represented

as long and short range interactions among the nodes. Interestingly, this method identifies small changes in structures of protein which are otherwise not easily detectable [Srivastava et al. 2014; Srivastava et al. 2017; Kandhari et al. 2017]. Moreover, similar methods have already been implicated in investigating various features of a protein such structural flexibility [Jacobs et al. 2001], protein domain folding [Dokholyan et al. 2002], key residue in folding [Wangikar et al. 2003], structural pattern [Vendruscolo et al. 2002], cluster of residue flexibility [Atilgan et al. 2004] and identification of functional residues [Amitai et al. 2004]. Similarly, a PSN-ENM based method has also been used to construct a PSN on a protein 3-D structure by integrating the information from systems dynamic supplied from the Elastic Network mode analysis (ENM-NMA) [Seeber et al. 2015]. A global (average) network parameters generated from these methods reveal diminutive structural changes among proteins.

In order to probe subtle conformational changes occurring due to differences of few amino acids in the M^{pro} sequences, alteration of local contacts as well as residue specific network parameters were investigated on the structures of M^{pro} (apo and inhibitor bound states) from both SARS-CoV-1 and SARS-CoV-2 by using the PSN and PSN-EMA methods. Recently, topological interactions properties of M^{pro} were analyzed [Estrada et al. 2020]. In another similar study, an analysis of changes in residue interactions of M^{pro} when bound to an N3 inhibitor was also investigated [Griffin et al. 2020]. However, in both the cases, the monomeric unit of M^{pro} was considered for the analysis. While, it is well known that a biological active M^{pro} molecule exists as a dimer. In fact, as reported earlier that the subunits interfacial region of the M^{pro} can be a possible target for a rational drug design against the SARS-CoV [Fan et al. 2004; Shi et al. 2004]. Hence, it is essential to understand the network connectivity in the biologically active state. Here, we analyzed a comparative PSN in both the proteases (in a biological active state). Further, we

elucidated the negligible changes throughout the protein structure by quantifying their residues connectivity pattern and mapped the network parameters on 3D structures of protein. We also applied a graph theory centrality concept such as betweenness, closeness, hubs and modularity to highlight critical residues for complex formation. This study will provide an understanding about the sensitivity and effectiveness of the existing inhibitors and this would further be helpful to design specific inhibitors.

4.2 METHODS

4.2.1 Protein Structure Network Construction

The 3-D coordinates of 3CL^{pro} were downloaded from Protein Data Bank (1Z1I [Hsu et al. 2005] & 6M03 [Zhang et al. 2020] for apo M^{pro} of SARS-CoV 1 & 2, respectively; 5N19 & 6Y2G [Zhang et al. 2020] for the inhibitor bound complex of M^{pro} from SARS-CoV-1 & 2, respectively). Here, C_{alpha} atom of amino acid residues is considered as a node and it forms an edge with another C_{alpha} atom if the distance cutoff is 7 Å. Edge weighted C_{alpha} network that is based on Euclidean distance was constructed using NAPS [Chakrabarty et. al. 2016] and protein network global parameters such as degree, Betweenness Centrality, and Clustering Coefficient were analyzed.

4.2.2 Network parameters

Degree is the total number of direct links between two connected nodes where C_{alpha} atoms were considered as contact type. The average degree (D) of a network with N nodes can be computed as $D = 1/N \sum_{i=1}^N D_i$

Hubs are nodes with higher degrees. **Shortest Path Length (SPL)** is the minimum number of links required to span through one node to another in a protein topological network. **Clustering Coefficient (CC)** computes the cliquishness for each node in the protein network graph.

Cliquishness is defined with respect to total possible edges between them. CC varies between 0 (for no clustering) and 1 (for maximum clustering). **Betweenness Centrality** (BC) is a centrality measurement in a network graph which is based on the total number of shortest path passing between connected nodes in such a way that edges passing for weighted graphs is minimized. **Closeness Centrality** (CCen) represents the closeness of a node to other nodes. It is a centrality measurement which calculates the sum of the shortest path. **Community or Modularity** is the region in a network where nodes are more connected to each other.

4.2.3 Structural Communication Analysis

Protein Structure Network and Elastic network model- Normal Mode Analysis (ENM-NMA) approaches were used for long-space communication and effect of allostery on network connectivity [Boccaletti Boccaletti et al. 2006; Tang et al. 2020; Artymiuk et al. 1990]. Previously, the ENM-NMA approach for PSN was applied to characterize the topological and allosteric communication pathways in proteins [Raimondi et al. 2013]. Other network parameters such as hubs, community, and structural communication analysis were analyzed using a mixed PSN ENM-NMA approach implemented in WebPSN [Seeber et al. 2015]. It constructs Protein Structure Graph based on interaction strength of two connected nodes

$$I_{ij} = \frac{n_{ij}}{\sqrt{N_i N_j}} 100$$

where interaction percentage (I_{ij}) of nodes i and j represents the number of side chain atoms pairs with given cut off (4.5Å), N_i and N_j are normalization factors [Kannan et al. 1999; Brinda et al. 2005; Feline et al. 2020]. The interaction strength (represented as percent) between residues i and j (I_{ij}) is calculated for all node pairs. If I_{ij} is more than the minimum interaction strength cutoff (I_{min}) among the residue pairs, then is considered to be interacting and hence represented as a

connection in the PSG. It builds PSG on atomic cross-correlation motions using ENM-NMA [Felline et al. 2020]. All network parameters were visualized using PYMOL.

4.3 RESULTS

4.3.1 Protein Structure Network analysis of the SARS-CoV-1 and SARS-CoV-2 M^{pro}

A protein structure network depicts a network of nodes and links. These nodes are represented by amino acids and links are represented as long and short range interactions among the nodes and that provide useful information at the contact level in protein structure. M^{pro} has been designated as an attractive drug target. In spite of having 96% sequence identity and negligible variation in 3-D structure compared to SARS-CoV-1, the drugs/inhibitors developed so far against M^{pro} of SARS-CoV-1 showed different inhibitory effects on the M^{pro} of SARS-CoV-2. Since, structural changes among the two M^{pro}s are negligible, hence a network based approach has been utilized to map subtle conformational alteration arising in the protein structure. Network parameters such as Degree, BC, C Cen, CC, SP and Modularity were analyzed for both free and inhibitor bound forms of SARS-CoV M^{pro} structures. A little difference was observed in the average network parameters of the M^{pro} structures (Table 4.1) suggests a diminutive change in the overall structures.

	Monomer		Dimer	
	SARS-CoV-1	SARS-CoV-2	SARS-CoV-1	SARS-CoV-2
No. of edges	1152	1185	2372	2436
Degree	7.65	7.75	7.87	7.96
Clustering coefficient	0.49	0.49	0.49	0.48
Shortest path	6.40	6.36	7.95	7.48

Table 4.1 Average network parameters of M^{pro} monomer and dimer units, generated through PSNs.

Calculated degrees are compared among the two structures. It was found that the near active site residues (T26, I43, Q189 and Q192) of SARS-CoV-2 M^{pro} showed an increase in degree by 2, compared to SARS-CoV-1 M^{pro}, and while degree of D187 was observed decreased by 2 in SARS-CoV-2. It was also observed that the N and C-terminal residues (G2 and E290), crucial for dimerization, were associated with higher degree compared to the same residues of M^{pro} in SARS-CoV-1. Additionally, few other residues of domain II & III of SARS-CoV-2 M^{pro} also found to have changes in the calculated degree. List of residues showing the largest change in degrees among the two structures are listed in (Table 4.2).

Residue No.	Degrees in M ^{pro}	
	SARS-CoV-1	SARS-CoV-2
2	2	4
26	8	10
43	6	8
61	4	6
119	7	9
128	10	12
153	4	6
187	6	4
189	7	5
192	4	6
232	6	8
264	11	9
290	6	8

Table 4.2: Residues showing largest change in degree.

Interestingly, replacement of A46 in SARS-CoV-1 M^{pro} with S46 in SARS-CoV-2 M^{pro} resulted in the rearrangement of contacts and observed to form new contacts with L27 and H41 from domain I of SARS-CoV-2 M^{pro}. A recent report states that SARS-CoV-2 M^{pro} possesses an active site with a solvent surface accessible area of 356 Å² and the solvent accessible surface area in case of SARS-CoV-1 M^{pro} was observed to be only 256 Å² [Griffin et al. 2020]. These changes may be attributed due to the variation of amino acid, S46 which resulted in the rearrangement of contacts

in SARS-CoV-2 M^{pro} . Additionally, few residues of domain II of SARS-CoV-2 M^{pro} are observed to form five new contacts (Figure 4.1A). Changes in the contact patterns surrounding the active site of SARS-CoV-2 M^{pro} is due to change in amino acid, suggesting a subtle conformational change in the SARS-CoV-2 M^{pro} , which may contribute towards the efficiency of inhibitors on M^{pro} s.

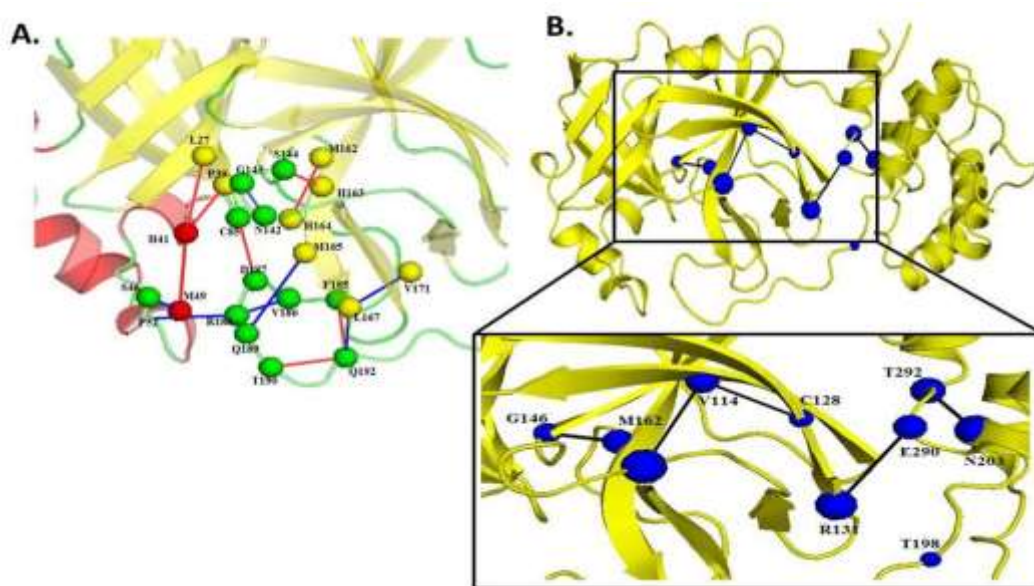


Figure 4.1. Contact maps in M^{pro} (A) New contacts in SARS-CoV-2 near the active site (new contacts made are in red line while contacts lost are in blue line). (B) Residues showing the highest increase in BC along with the new contacts mapped on the SARS-CoV-2 structure.

Hubs were also analyzed for the SARS-CoV-1 and SARS-CoV-2 M^{pro} s. Interestingly, a significant difference in the total number of hubs was observed among the two main proteases. The M^{pro} from SARS-CoV-1 possesses 42 hubs, whereas the same from SARS-CoV-2 consists of 47 hubs, in total (Table 4.3). Many hubs were found to be similar among the two structures. Few hubs were distinctive to each structure, suggesting their important role in interactions and stability. Hubs near

the active site region such as H41, H163, D187 and Q192 from the SARS-CoV-2 are assumed to be crucial for the catalysis. The unique hubs are distributed in all the three domains of SARS-CoV-2 M^{pro} , and may suggest a subtle change in inter domain communication within the protease.

Hubs in M^{pro}	
SARS-CoV-1	SARS-CoV-2
3, 8, 19, 31, 32 , 38 , 39, 54, 88 , 95, 103, 106, 112, 113, 126, 131, 132 , 140, 159, 161, 164, 167 , 171 , 181, 182, 188 , 200, 207, 209, 213, 218 , 219, 229 , 239, 256, 268 , 273, 281, 282, 288 , 292, 293	3, 8, 19, 39, 40 , 41 , 54, 61 , 66 , 95, 101 , 103, 106, 112, 113, 118, 122 , 126, 131, 140, 150 , 156 , 157 , 159, 161, 163 , 164, 181, 182, 185 , 187 , 192 , 200, 207, 209, 213, 219, 228 , 239, 242 , 256, 269 , 273, 281, 282, 292

Table 4.3. List of hubs in SARS-CoV-1 and SARS-CoV-2 M^{pro} , residues in bold are unique to corresponding PSN.

Betweenness Centrality (BC) have been reported to play an important role in the structural complexes. In our study, the residues from the both M^{pro} structures and their corresponding BC scores are plotted in (Figure 4.2 A&B). The trend of the plots are quite comparable in both the structures except few residues show significant change in the BC scores. Residues with significantly high BC scores (z scores ≥ 0.4) from each M^{pro} structure are listed in (Table 4.4). Significantly high BC value of a residue signifies its involvement in the communication among different modules of the PCN. The residue V114 with high BC value is observed to make a new contact with F140 in case of SARS-CoV-2. In addition, other residues such as C128, G146, and T292 found to have high BC values that are also involved in the formation of a new set of contacts in the SARS-CoV-2 (Figure 1B). The new contacts formed in the SARS-CoV-2 M^{pro} , suggest their role in providing connectivity among residues of the network.

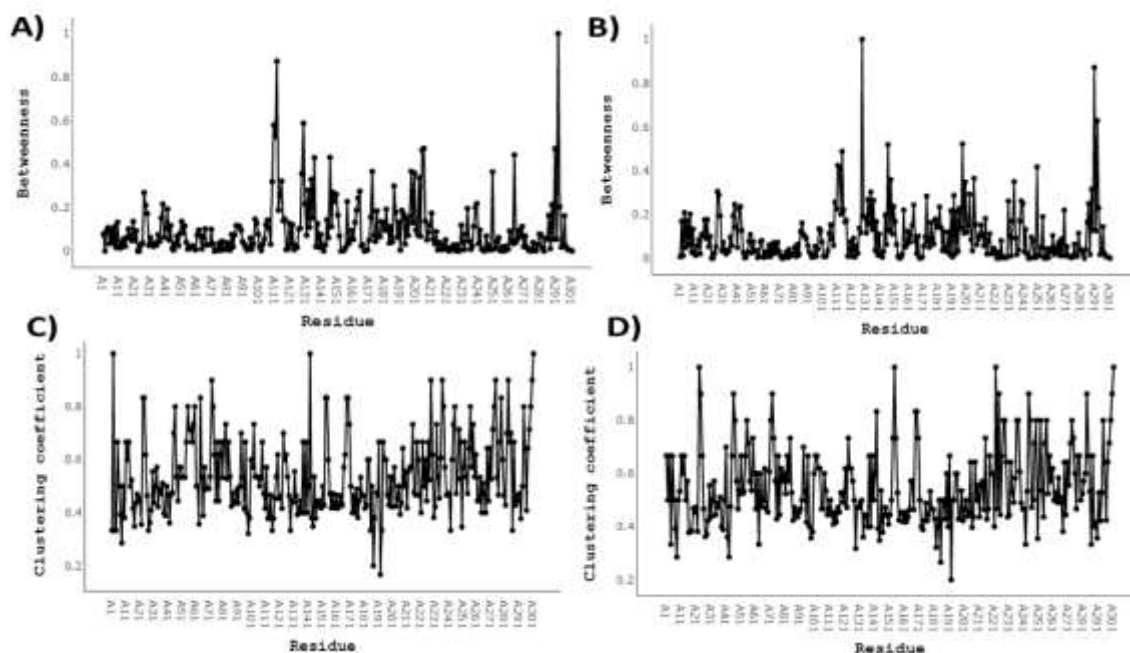


Figure 4.2. Residue wise changes in betweenness centrality of (A) SARS-CoV-1 M^{pro} and (B) SARS-CoV-2 M^{pro}; Residue wise Clustering coefficient of (C) SARS-CoV-1 M^{pro} and (D) SARS-CoV-2 M^{pro}.

M ^{pro}			
SARS-CoV-1		SARS-CoV-2	
Residue No.	BC score	Residue No.	BC score
110	0.5	111	0.5
111	0.5	114	0.5
112	0.8	128	1
129	0.5	146	0.5
205	0.5	198	0.5
206	0.5	290	0.8
264	0.5	292	0.6
290	0.5		
292	1		

Table 4.4. Residues with highest BC value. Residues with high BC values in both the PCNs are colored as Yellow.

Residues wise CC were analyzed for both protease structures and values are depicted in (Figure 4.2 C&D). The residue at 46 positions in SARS-CoV-2 compared to SARS-CoV-1, resulted in an increase in CC of nearby N-finger active site residues such as G23, T24, and S46. Interestingly, T24 is also observed to form a direct contact with the active site residues in SARS-CoV-2. This suggests that changes in the interconnectedness among the residues at and near the active site region may play a role towards selectivity of the inhibitors. Though the average parameters calculated from the PSN of M^{pro} s from both SARS-CoV-1 & 2, did not show significant changes. However, residue wise comparison of degree and BC values among the two exhibited noticeable change (Table 4.2 & 4.4). Moreover, these observations on the change in the network parameters suggested their effect on the local conformations of M^{pro} s, which is assumed to provide an insight into the sensitivity and selectivity of inhibitors.

4.3.2 Community Structure Analysis of the PSNs

M^{pro} structures from SARS-CoV-1 and SARS-CoV-2 do not show much difference, however, an analysis of the contact points either generated or lost due to change of few amino acids, may provide an insight into restructuring of modules within the M^{pro} . Hence, the community structure for both SARS-CoV-1 & SARS-CoV-2 M^{pro} s was analyzed. The analysis resulted in twelve communities in case of SARS-CoV-1 M^{pro} and eleven communities in SARS-CoV-2 M^{pro} and residues of each community are shown in (Figure 4.3). The residues at the active site region of SARS-CoV-2 M^{pro} are observed to constitute the largest community shown as C1 red module in Figure 4.3A, that consists of 12 nodes, 18 links and seven hubs. Unlike SARS-CoV-2, the largest community (formed with 8 nodes, 11 links, 4 hubs) in case of SARS-CoV-1 M^{pro} is located at the interfacial residues of domain I and II, instead of active site region (Figure 4.3B). Moreover, the community formed at the active site region of SARS-CoV-1 M^{pro} is found to be smaller than that

of SARS-CoV-2 M^{pro} . Rearrangement of modules was also observed throughout the structure which indicates the perturbation at global level in the 3-D structure of two proteins.

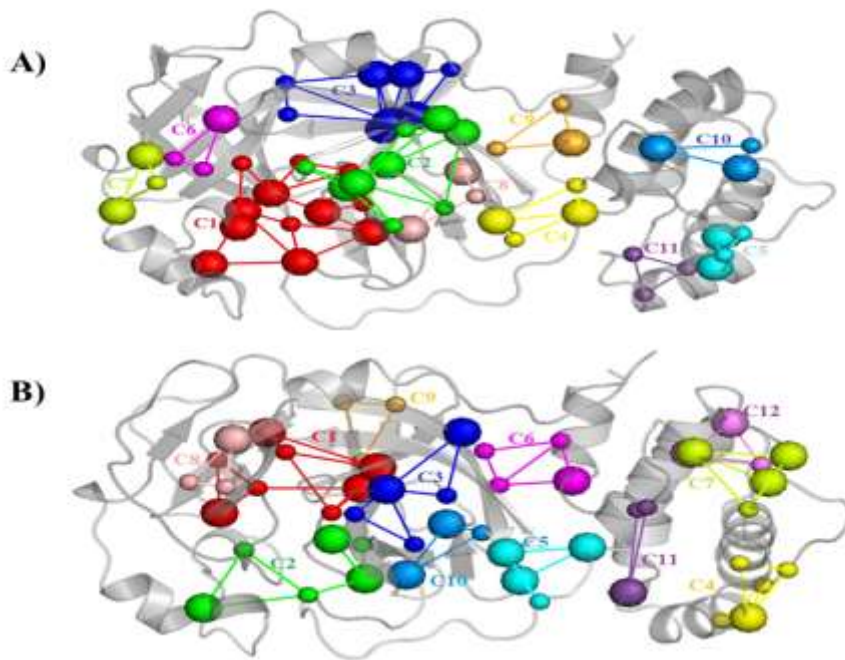


Figure 4.3. Community structure in the M^{pro} . All communities are mapped on their tertiary structure of M^{pro} s and depicted with different color modules (A) SARS-CoV-2 and (B) SARS-CoV-1.

4.3.3 PSN analysis of inhibitor (ketoamide) bound complexes of M^{pros}

4.3.3.1 Inhibitor binding perturbs the PSG and the communication pathway in M^{pro}

The Protein Structure Graph (PSG) of SARS-CoV-1 M^{pro} inhibitor bound complex is richer in nodes, links as well as hubs compared to its unbound state (Table 4.5). Binding of inhibitors to the M^{pro} generated many hubs at the inhibitor binding site and these hubs are associated with residues such as H41, Y54, F140, S144, H163, H172, and Q192. Additionally, inhibitor complex specific hubs are also formed between the interface of domain I & II (residues C16, Y101, F150, and L115),

while invariant hubs spanned throughout the structure. Interestingly, similar trends for nodes and links were not observed for SARS-CoV-2 M^{pro} inhibitor bound complexes when compared with apo form of the same (Table 4.5). However, the total number of hub residues in the apo form of SARS-CoV-2 M^{pro} was found to be 36, whereas the inhibitor complex of the same possesses 35 hub residues. Few hubs are found to be unique in each structure, suggesting their role towards the specificity of inhibitors. Unlike SARS-CoV-1, the residue H41 and Q192 from SARS-CoV-2 M^{pro} inhibitor complex form do not participate in the active site hubs formation.

	M ^{pro} monomer				M ^{pro} Dimer	
	SARS-CoV-1		SARS-CoV-2		SARS-CoV-1	SARS-CoV-2
	Apo	Holo	Apo	Holo	Apo	Apo
No. of linked nodes	282	290	288	291	577	578
No. of Links	311	328	324	327	644	637
No. of Hubs	30	37	36	35	58	53
No. of Link mediated hubs	118	149	140	143	229	202

Table 4.5. Network components and its parameters for both SARS-CoV-1 and SARS-CoV-2 M^{pro} monomer in apo and holo states and M^{pro} dimer in apo state.

We mapped the perturbations on the 3-D structure which considers nodes and links unique to each structure (Figure 4.4 A&B). In the case of the apo and inhibitor bound states of SARS-CoV-1 M^{pro}, the bound inhibitor was observed to induce perturbations which are essentially consistent with a gain of intermolecular links and nodes. The perturbations associated with a gain of links are mostly located in the region of small helix near P2 group consists of residues S46- L50, β -hairpin loop

near P3-P4 (res E166-G170) and P5 loop residues (T190- A194). Additional gains of links are also observed in the interfacial region of domain I and II, along with N-finger residues making new links with C-terminal of domain III. However, in case of SARS-CoV-2 M^{pro} , the comparison of the apo and inhibitor bound states was observed to have not very significant perturbations and the specific contacts show changes to a lesser extent than SARS-CoV-1 M^{pro} .

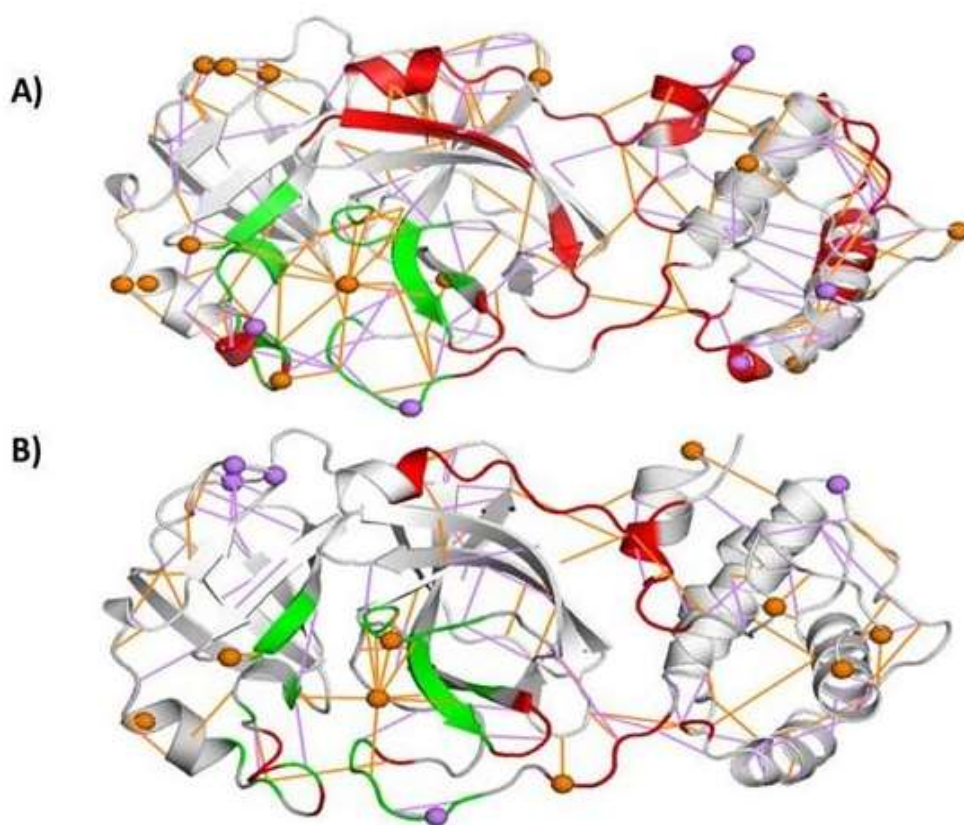


Figure 4.4. 3-D PSG representations of the inhibitor bound complexes of M^{pro} (A) SARS-CoV-1 (B) SARS-CoV-2. Nodes and links peculiar to the unbound and bound states are in violet and orange, respectively. Major changes near the active site and other allosteric sites are in green and red, respectively.

Perturbations in inter and intra subunit communication due to binding of inhibitors were also analyzed (Table 4.5). To investigate more into the communication pattern within the structure, we analyzed meta-path and mapped residues participating in each path. The length of the shortest communication paths in SARS-CoV-1 M^{pro} apo form was 62345 and a total of 77391 paths were observed in the inhibitor bound form of M^{pro}. This indicates an increase in the pathways upon the inhibitor binding to SARS-CoV-1 M^{pro}. In contrast, a decrease in possible pathways was observed in case of SARS-CoV-2 M^{pro} inhibitor bound complex. However, the average path length increased in both inhibitor bound states of SARS-CoV-1 and SARS-CoV-2 M^{pro}. Changes in the most frequent nodes and links in the structural communication upon inhibitor binding were also observed (Figure 4.5). The significant inhibitor induced perturbations in the form of loss or frequency reduction of nodes were observed within the active site region of the complexes in either cases (for SARS-CoV-1: C44, P52, Y54, F140, S144, L167, R187, Q192 and for SARS-CoV-2: F140, S144, H163 and R187). Moreover, a redistribution of nodes was observed around the N-finger of M^{pro} inhibitor complex from SARS-CoV-1, suggesting a role of intercommunication exchange between domain II & III, which may be crucial for the dimerization of M^{pro}.

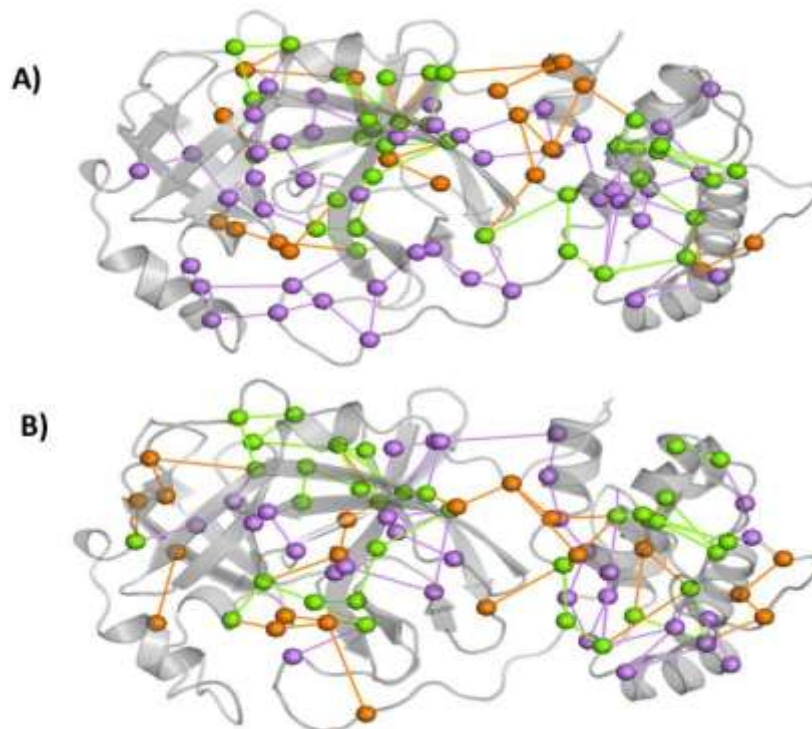


Figure 4.5. Residues involved in global meta-path mapped on 3-D structure of M^{pro} . A) SARS-CoV-1 and B) SARS-CoV-2. Residues involved in the meta path for apo state are in purple nodes and links while for inhibitor bound states are in orange, shared nodes and links are in green.

4.3.3.2 Community analysis of apo and inhibitor bound complex

The community structure of PSNs of the M^{pro} from SARS-CoV-1 and SARS-CoV-2, in apo and inhibitor bound states were analyzed. The SARS-CoV-1 M^{pro} bound complex is found to have 11 communities, whereas the SARS-CoV-2 M^{pro} inhibitor complex possesses six communities and these communities are mapped on PSNs. In both the complexes, inhibitor binding sites were part of a large community, C1 (Figure 4.6). The long loop connecting domain II & III was involved in the second largest community for SARS-CoV-1 complex which includes 7 nodes, 11 links, and 3 hubs. This connecting loop is a part of community C3 in the case of SARS-CoV-2 complex which includes 5 nodes, 7 links and 3 hubs. The rearrangement of communities in the inhibitor complex

forms in comparison to the apo forms, suggests that the inhibitors induce perturbations in the network connectivity. Our observations show a correlation with the previous report of RIN on M^{pro} of SARS-CoV-1 and SARS-CoV-2, with and without inhibitor N3 [Griffin et al. 2020]. Similarly, a recent study on topologies of M^{pro} by PCN methods highlights sensitive structural perturbations.

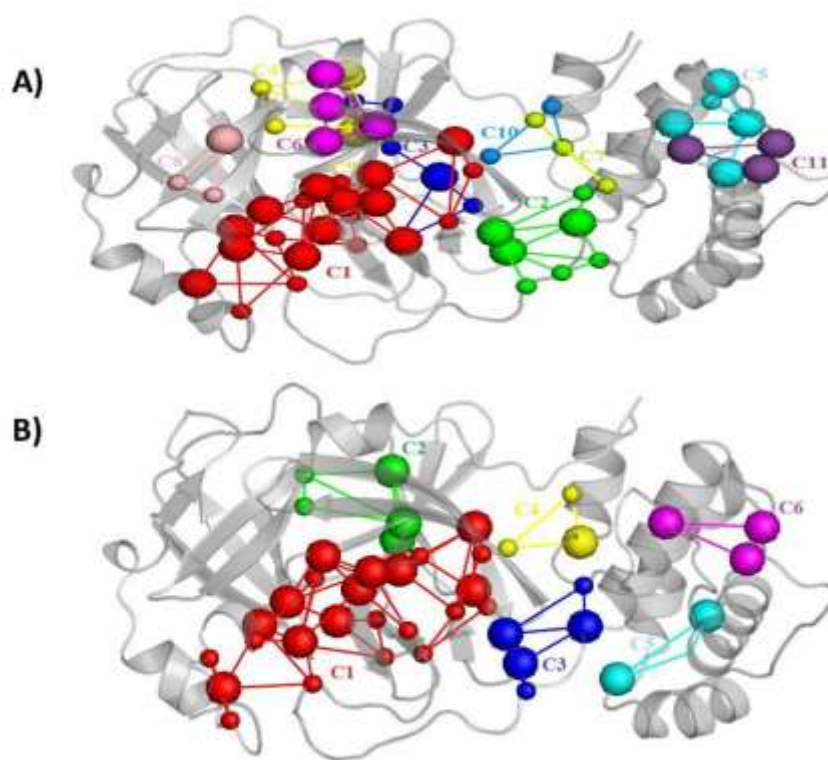


Figure 4.6. Community structure in the inhibitor bound states of M^{pro} . All communities are mapped on their tertiary structure in different color module A) SARS-CoV-1 and B) SARS-CoV-2.

4.3.4 PSN analysis on quaternary structure of SARS-CoV-2 *M_{pro}*

It has been known that biologically active SARS-CoV-2 *M_{pro}* exists as a dimer. Previously, mutagenesis of E290A in SARS-CoV-2 *M_{pro}* reported loss of catalytic activity, indicating importance of domain III in dimerization [Chou et al. 2004]. Hence, In order to highlight structural differences and commonalities in the quaternary structures (homodimers) of *M_{pro}*^S, PSN parameters were computed and identified the crucial residues from the subunit interfacial region. Average network parameters for the two subunits of *M_{pro}*^S don't show any significant changes, hence no correlation was drawn (Table 4.1). So, we further evaluated the previously mentioned network components like links, hubs, and link mediated hubs. Interestingly, these parameters were observed to be slightly higher for SARS-CoV-1 dimeric form (Table 4.5). Moreover significant changes were observed in the hubs and links mediated hubs. A total number of 58 hubs residues were observed in the quaternary SARS-CoV-1 *M_{pro}*, whereas 53 hubs were noted in the SARS-CoV-2 *M_{pro}* homodimer (Figure 4.7A). Interestingly, subtle rearrangements of hub residues in domain III in both homodimers were also observed. In case of SARS-CoV-1 *M_{pro}*, the residues such as 206: A, 259: A, 289: A, 218: B, 230: B (the alphabets A & B represent sub-units) were observed to form a specific hub. Similarly, the residues such as 288: A, 273: B, 288: B, 290: B from SARS-CoV-2 *M_{pro}* were involved in a hub formation. In addition, specific hub residues were also observed in the domain II of SARS-CoV-2 (161: B, 181: B) and unfortunately these were not seen to form hubs in SARS-CoV-1 *M_{pro}* homodimer. Other residues such as 39: A, 141: B, 163: B, 172: B were also involved in hub formation in SARS-CoV-1 *M_{pro}*. It was observed that residues such as 185: A, 192: A, 192: B from SARS-CoV-2 *M_{pro}* were engaged in forming active site hubs. These rearrangement in hub residues suggest some disruption in the inter-domain communication between both proteases in their quaternary structure.

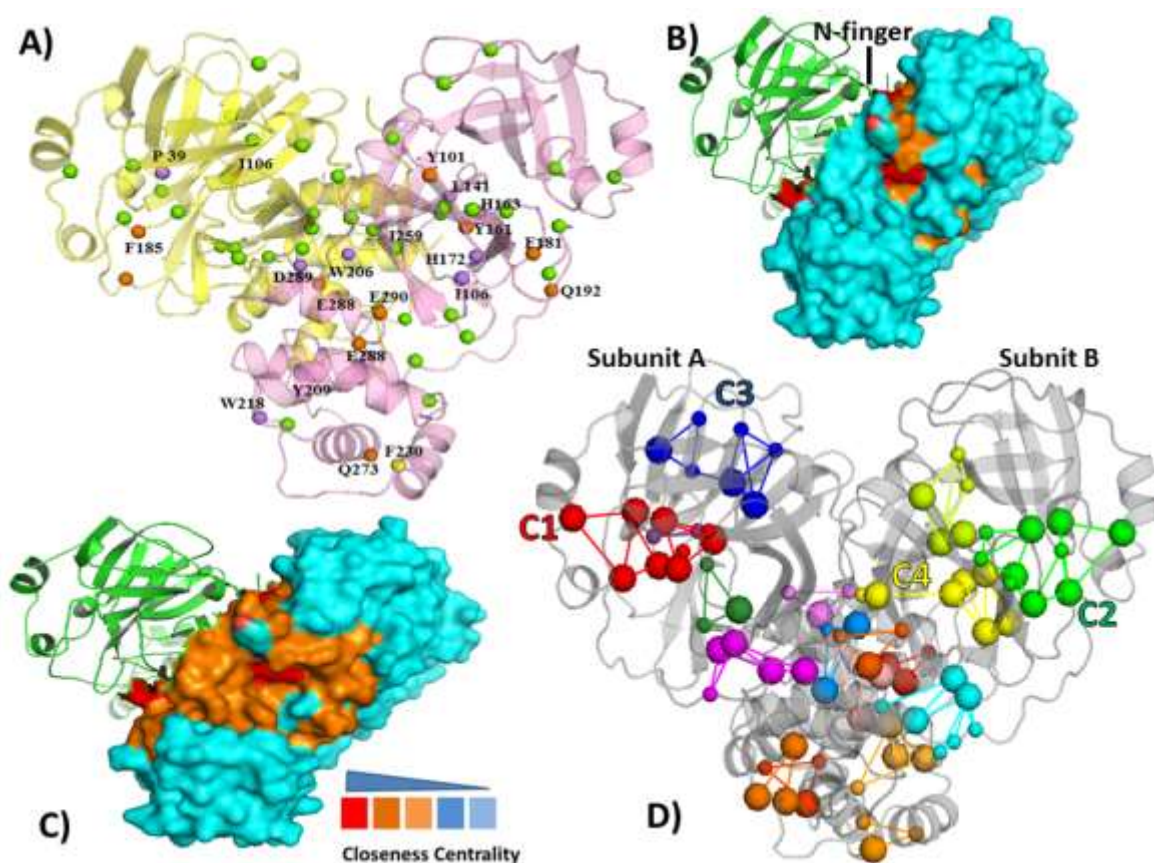


Figure 4.7. PSN analysis on the quaternary structure of SARS-CoV M^{pro} . Chain A is in yellow, chain B is in light pink A). Residues involved in hub formation SARS-CoV-1 dimeric state are in purple nodes and links while for SARS-CoV-2 it is in orange color. Nodes and links shared in common are in green. B) & C) The residues are colored by closeness values with red, orange and cyan corresponding to the high (top 20%), intermediate (20~60%) and low (below than 60%) closeness values, SARS-CoV-1 & SARS-CoV-2., respectively. D) Consensus community structure analysis between both Proteases. Major communities in the complex are numbered with respect to its specific module color.

The BC for each residue of M^{pro} from SARS-CoV-1 as well as SARS-CoV-2 dimeric complexes was analyzed and residues with high BC are listed in (Table 4.6). N-finger residues (Res. No.3 &

8) and C-terminal residues of domain III (Res. No. 282 & 290) that are at the interface of both monomers were showing high betweenness values in both the structures, indicating the importance of these residues in the dimer formation [Felline et al. 2020]. Additionally, these interfacial residues showing high BC are also involved in hub formation which suggests their role in catalysis. The residues at positions 28 and 144 in SARS-CoV-2 M^{pro} showed high BC value and their essentiality for enzyme activity and dimerization has been confirmed with experimental mutagenesis studies in homodimer formation [Barrila et al. 2010].

M ^{pro} dimer			
SARS-CoV-1		SARS-CoV-2	
Residue No.	BC score	Residue No.	BC score
3:A	1	2:A	0.6
8:A	0.5	8:A	0.7
128:A	0.6	14:A	0.5
138:A	0.5	17:A	0.5
282:A	0.9	28:A	0.7
290:A	0.5	113:A	0.5
3:B	0.8	118:A	0.5
8:B	0.5	128:A	0.6
128:B	0.6	144:A	0.5
138:B	0.6	214:A	1
211:B	0.6	285:A	0.8
282:B	0.8	290:A	0.5
290:B	0.5	292:A	0.5
		293:A	0.5
		2:B	0.6
		8:B	0.6
		14:B	0.5
		17:B	0.5
		28:B	0.6
		123:B	0.7
		128:B	0.6
		144:B	0.6
		214:B	0.9
		285:B	0.8
		290:B	0.5
		292:B	0.5

Table 4.6. Residues with highest BC value in dimeric form of M^{pro}

The closeness values of all residues were computed and classified them into three categories: i . High closeness value, ii . Intermediate closeness values, and iii. Small closeness values. Our results suggested that residues from N-finger (3-11), B2 (112-117), B3 (122-130) and B4 (149-151) of domain II might be considered as the most likely recognition sites (Figure 4.7 B&C). Previously, it has been reported that residue C117 makes direct interaction with N28 and plays a major role in the dimer stability and enzymatic activity of SARS-CoV-1 M^{pro} [Barrila et al. 2010]. Experimentally, it has been identified that N28A mutant plays a critical role in active site structural integrity and positions the important residues involved in dimer interface binding and catalysis of substrate [Hu et al. 2009]. This suggests that residues showing high closeness might be responsible for long range interactions that are crucial for dimerization.

The search for shortest communication pathway led to a total of 290236 and 468290 paths for the M^{pro}s of SARS-CoV-1 and SARS-CoV-2, respectively which indicates a significant increase in paths for SARS-CoV-2 M^{pro} dimeric form (Table 4.5). The total number of nodes and links along with the specific nodes and links in global Metapath were observed to be 60% and 21.79% for SARS-CoV-1 and SARS-CoV-2, respectively. Additionally, we observed some of the interface residues are specific to SARS-CoV-1 and those are frequent nodes in communication pathways (Chain A: Res. 3, 6, 123, 126, 140, 290; Chain B: Res. 4, 6, 116, 141, 122, 126, 299). Few substrate binding residues (Chain A: Res. 41, 49, 144, 163, 165; Chain B: Res.163, 167) were also involved in the communication pathway of SARS-CoV-1 M^{pro}, while these corresponding frequent nodes were absent in SARS-CoV-2 M^{pro}. In addition, average hub percent involved in the communication pathway was also observed to be decreased for SARS-CoV-2 (Table 4.5). These observations among the two suggest a change at structural communication level in the dimeric form of M^{pro}.

Common modules are shared among the two homo-dimeric M^{proS} and depicted in Figure 4.7D. Two large communities CI and CII consisting of active site residues from both monomers possess 10 nodes, 15 links and 7 hubs. Third large community CIII is distributed on the strands of domain II in both monomers. Additionally, a fourth community was observed at the interface residues of both monomers; N-term residue M6 of chain A and β -strands of domain II from chain B. One of the residues F140 from this community has been previously reported to present on the dimer interface of SARS-CoV M^{pro} and mutation of this residue resulted in the conformational change of M^{pro} [Chou et al. 2004].

4.4 CONCLUSIONS

Our study on comparative protein structure network analysis of M^{proS} from SARS-CoV-1 and SARS-CoV-2, investigated the noticeable difference in the network parameters among the two proteases. Moreover, the study also highlights differential perturbation among the community structures in inhibitor bound form of proteins. Interestingly, the investigations helped us to probe subtle conformation changes associated throughout the structure of the two proteases, which otherwise are not evident from the crystals structures. Our observations gauge an insight into the diminutive structural changes which may provide an understanding towards selectivity of inhibitors towards M^{proS} of SARS-CoV-2. In addition, the investigation of PSN on the quaternary structure of M^{proS} suggests structural and network changes at the interface as well as long range interactions and highlights critical residue pairs for the complex formation using three centrality measurement parameters. This study is a thorough comparative investigation of subtle structural changes that may provide an insight into designing a specific inhibitor/drug.

CHAPTER 5

Structure-based inhibitor design against *M. tuberculosis* Thioredoxin Reductase

THIS CHAPTER IS PUBLISHED IN:

Lata, S., & Akif, M. (2021). Structure-based identification of natural compound inhibitor against *M. tuberculosis* thioredoxin reductase: insight from molecular docking and dynamics simulation. *Journal of biomolecular structure & dynamics*, 39(12), 4480–4489

5.1 INTRODUCTION

Mycobacterium tuberculosis (*Mtb*), the causative agent of tuberculosis that accounts for 1.3 million deaths worldwide every year (World Health Organization [WHO], 2018). Although many novel therapeutics have been developed, still the evolving resistance strains such as MDR (multi-drug-resistant) and XDR (extensively-drug-resistant) pose a great challenge to counter TB globally (Libardo et al. 2018). Moreover, the current first line of drugs offers a long treatment regimen. Hence, a search for new drugs or drug-like molecules that can augment the effectiveness of the current treatment and reduce the emerging resistance is warranted.

Oxidative stress and accumulation of reactive oxygen species (ROS) provide a hostile environment in the cell, which is an inevitable challenge to invading pathogens (Imlay et al. 2013). Pathogens such as *Mtb* resides in the host phagocytes and copes with the hostile environment by expressing a number of antioxidant systems to ensure their survival inside its host (Trivedi et al. 2012). The thioredoxin and glutathione systems are two well-known antioxidant systems that provide reducing environments and regulate many important cellular processes such as antioxidant pathways, DNA and protein repair enzymes, and the activation of redox dependent transcription factors (Fahey et al. 2001; Lu et al. 2014). Hence, antioxidant systems have been considered as a potential drug target.

Mtb thioredoxin system consists of two typical thioredoxins (TrxB and TrxC) and a single copy of thioredoxin reductase (TrxR) (Akif et al. 2008). TrxR reduces the Trxs by typically utilizing the reducing potential from the cellular NADPH. The reduced Trxs function in reducing the peroxides and dinitrobenzenes, and also play an important role in

detoxifying hydroperoxides. This signifies the importance of the thioredoxin system for the survival of a pathogen in the hostile environment of macrophages (Jaeger et al., 2004; Zhang et al. 1999). Interestingly, as reported earlier for *E. coli* and other species, reduced Trxs are essential for nucleotide biosynthesis as they donate the reducing equivalents to ribonucleotide reductase (Lu et al. 2014). Transposon-mediated analysis has validated *Mtb* TrxR as an essential gene. Moreover, it has been reported that deletion of the *trxR* gene results in a hyper-susceptible strain (Li et al., 2016; Zhang et al., 2012). TrxR is also shown to be essential for the growth of other organisms such as *S. aureus* (Uziel, Borovok, Schreiber, Cohen & Aharonowitz, 2004) and *B. subtilis* (Scharf et al., 1998). Thus, considering its crucial role, the bacterial TrxR has recently been demonstrated as a promising drug target (Lu et al., 2013).

The success of a structure-based drug design depends on a selective inhibition of the target compared to its counterpart in the human host. Fortunately, *Mtb* TrxR differs with the human TrxRs in terms of structure, sequence, and the mode of transfer of reducing equivalent. Human and mammal TrxR is a high molecular weight protein which, apart from NADPH- and FAD-binding domains, consists of an extra flexible C-terminal extension of the cysteine-selenocysteine-glycine (CUG) motif as a redox center that transfers the reduction potential to Trxs (Lu et al. 2014). In contrast, *Mtb* and other prokaryotes TrxR are a low molecular weight protein with distinct NADPH and FAD domains connected through a two-stranded β -sheet known as a hinge region (Akif et al. 2005; Waksman et al. 1994). The buried redox cysteine residues (CXXC) come out on the surface by a 67° rotation of the NADPH domain with respect to the FAD domain during catalysis to provide reducing potential to Trxs. This unique feature of TrxR has been demonstrated with complex crystal structure of the *E. coli* TrxR–Trx and *Mtb* TrxR structure (Akif et al., 2005; Lennon et al. 2000).

In earlier reports, TrxR–Trx interaction site of *Mtb* was selected as a druggable target (Koch et al., 2013). But the main obstacle was a low hit rate of protein–protein interaction site and was considered not to be easily druggable. Thus, targeting an allosteric pocket near the NADPH domain, consisting of the interface of NADPH and FAD domains, active site, and hinge region, may bring about structural changes in the protein such that active site residues cannot be available on the surface to interact with Trxs. In recent years, TrxR from various bacterial species has been targeted, which has yielded some compounds demonstrated to inhibit TrxR, includes ebselene (Gustafsson et al., 2016; Lu et al., 2013), auranofin (Harbut et al., 2015; Owings et al., 2016) and gold (I)-alkynyl chromones (Hikisz et al., 2015). Still, an investigation of inhibitors with a strong specificity towards *Mtb* TrxR is needed.

Natural compounds have been the single most productive source of leads for the development of drugs. Traditional Chinese Medicine (TCM) (Chen et al. 2011) is a database of natural compounds that follow Lipinski's rule of five and almost all compounds have a therapeutic effect. The current study aimed for the identification of natural compounds suitable for the potential inhibitors of *Mtb* TrxR. All natural compounds present in the TCM database were used to carry out virtual screening against *Mtb* TrxR. This led to the identification of several compounds on the basis of docking scores and the top two compounds were selected for further study. A detailed study of docking and molecular dynamics simulation (MDS) along with MM/PBSA of the selected compounds showed that they form a stable protein–ligand complex. Principal component analysis (PCA) and alteration in solvent accessibility studies confirmed the change in conformation of protein. Protein contact network (PCN) graphs of simulated structures of both the apo- and compound-bound complexes showed the changes in the degree of connectivity within the hinge region residues among the two structures, which signifies that binding of compounds affect the protein structure

and function. This is the first report where natural compounds have been screened as an inhibitor scaffold against *Mtb* TrxR.

5.2 METHODS

5.2.1 Preparation of the target coordinate

The target crystal structure of *Mtb* TrxR (PDB ID 2A87) (Akif et al., 2005) was downloaded from the protein database and subjected for preparation by adding/fixing the missing side chains using what-if server (Vriend et al. 1990). Water and heteroatom molecules were removed from the coordinates. Energy minimization of 25,000 cycles of the steepest descent and 20,000 cycles of the conjugate gradient were performed through the Swiss PDB viewer.

5.2.2. Virtual screening and molecular docking

Crystal structure of *Mtb* TrxR was used to probe surface cavity predictions by CastP analysis (Tian et al. 2018). One of the well-defined allosteric cavities (area 1301.594 Å² and volume 855.756 Å³) was selected for the virtual screening of natural compounds using iScreen web server (Tsai et al. 2011). iScreen is a robust screening and docking server with 1) TCM integrated into CADD (computer aided drug design) services, 2) PLANTS module to evaluate the docking score, and 3) E-LEAD3D for *de-novo* docking. More than 20,000 compounds present in the TCM database that follow the Lipinski's rule of five were screened to bind on the cavity of TrxR through PLANTS docking module. The docking algorithm of PLANTS is based on ant colony optimization and provides various conditions of docking. Out of the 200 docked compounds, 2 best compounds were sorted out based on the docking scores of binding in the cavity. These two compounds with the highest scores were further re-docked in the binding cavity

by a *de-novo* drug design method implemented in LEAD3D software module of the iScreen package.

5.2.3. Molecular dynamics simulation

To study the structural stability of *Mtb* TrxR-compound complexes, MD simulations were carried out using GROMACS 5.1.4 (Hess et al. 2008; Pronk et.al. 2013). The topology files of the protein and compounds were generated using GROMOS96 45a3 force field (Oostenbrink et al. 2004) and the PRODRG2 server (Schüttelkopf et al. 2004), respectively. The TIP3P water model was implemented for water molecules. The complex system was solvated in a cubic box with water molecules of 1.5nm to the box wall from the surface of the protein and the system was further neutralized by adding Na⁺ counter ions. To minimize the short-range bad contacts, energy minimization was carried out using the steepest descent method for 50,000 steps until the largest force acting in the system was smaller than 10,000 kJ/mol/nm. After energy minimization, the temperature was equilibrated first in an NVT ensemble at 300 K for 50 ps using a modified V-rescale Berendsen thermostat with a time constant of 0.1 ps, followed by NPT ensemble to 1atm using Parrinello–Rahman coupling method with a time constant of 2 ps for 50 ps. After the systems were equilibrated, the production run was performed for 20 ns at 300 K. The equations of motion were integrated with time steps of 2 fs and the coordinates were saved for every 2500 time steps (5 ps), which resulted in a total 4000 frames for a 20 ns simulation. The long-range electrostatics was controlled using Particle Mesh Ewald (PME) method with a space cut-off of 10 Å. The hydrogen bonds were constrained by implementing the P-LINCS algorithm (Hess et al. 2008). Whole analysis was done using the frames from the production run (4000 for 300 K). RMSD, RMSF, and SASA were respectively calculated using `g_rms`, `g_rmsf`, and `g_sasa` functions of

GROMACS. The protein structures were visualized and the figures were generated using PyMOL, XMGRACE, GNUPLOT, and VMD.

5.2.4. Binding-free energy calculations of compounds

Molecular mechanics Poisson Boltzmann surface area (MMPBSA) remains the most widely used method for binding-free energy calculations from the snapshots of the MD trajectory (Kollman et al. 2000). MMPBSA calculates free energy interaction of protein–ligand complexes in three steps: 1) calculates potential energy in the vacuum, 2) calculates polar solvation energy, and 3) calculates non-polar solvation energy. The binding-free energies of the complexes between compounds and TrxR were analyzed during equilibrium phase by capturing snapshots from the last frame 15 to 20 ns MD simulations, using *g_mmpbsa* tool of GROMACS (Kumari et al. 2014). Particularly, the binding-free energy of protein–ligand complex in the solvent was expressed as:

$$\Delta G_{\text{binding}} = G_{\text{complex}} - (G_{\text{protein}} + G_{\text{ligand}})$$

where G_{complex} is the total free energy of the protein-ligand complex, G_{protein} and G_{ligand} respectively are total energy of separated protein and ligand in the solvent. The free energy for each individual G_{complex} , G_{protein} and G_{ligand} were estimated by:

$$G_X = E_{\text{mm}} + G_{\text{solvation}}$$

where x is the protein, ligand, or complex, E_{mm} is the average molecular mechanics potential energy in vacuum, and $G_{\text{solvation}}$ is the free energy of the solvation. The molecular mechanics potential energy was calculated in vacuum as:

$$E_{\text{mm}} = E_{\text{bonded}} + E_{\text{non-bonded}} = E_{\text{bonded}} + (E_{\text{vdw}} + E_{\text{elec}})$$

where E_{bonded} is bonded interaction including bond, angle, dihedral, and improper interactions and

$E_{\text{non-bonded}}$ is non-bonded interactions consisting of van der Waals (E_{vdw}) and electrostatic (E_{elec}) interactions. ΔE_{bonded} is always taken as zero.

The solvation free energy ($G_{\text{solvation}}$) was estimated as the sum of electrostatic solvation free energy (G_{polar}) and apolar solvation free energy ($G_{\text{non-polar}}$):

$$G_{\text{solvation}} = G_{\text{polar}} + G_{\text{non-polar}}$$

where G_{polar} was computed using the Poisson-Boltzmann (PB) equation and $G_{\text{non-polar}}$ was estimated from the solvent-accessible surface area (SASA) as:

$$G_{\text{nonpolar}} = \gamma \text{SASA} + b$$

where γ is a coefficient related to the surface tension of the solvent and b is a fitting parameter.

The values of the constant are as follows:

$$\gamma = 0.02267 \text{ kJ/mol/\AA}^2 \text{ or } 0.0054 \text{ Kcal/mol/\AA}^2$$

$$b = 3.849 \text{ kJ/mol or } 0.916 \text{ Kcal/mol}$$

5.2.5. Principal component analysis

Principal component analysis (PCA) was performed to obtain a mass-weighted covariance matrix of the protein atom displacement, which is indicative of the dominant and collective modes of the protein. This covariance matrix is diagonalized to extract a set of eigenvectors and eigenvalues that accurately reflect the concerted motion of the molecule. The GROMACS inbuilt tool `g_covar` was used to yield distinct eigenvalues and eigenvectors by calculating and diagonalizing the covariance matrix, whereas the `g_anaeig` tool was used to analyze and plot the eigenvectors (David et al. 2014). The first two eigenvectors represent the highest eigenvalues and are adequate to capture the overall motion of the protein.

5.2.6 Protein contact networks analysis

Protein contact network (PCN) was performed to carefully analyze changes between the interacting residues by the comparing contact maps of the apo and holo complex structures of the TrxR using CMView (Vehlow et al. 2011). The network is a graph where each residue corresponds to a node, and two nodes are connected by an edge if and only if the two residues are in contact. Two residues are considered to be in contact if they are spatially closed in the three-dimensional structure and are specified by two key parameters: contact type and distance cutoff. The contact type defines a subset of atoms of the residue. The most commonly used cutoff values for C α -based PCNs are 0.7 nm to 0.8 nm. RING 2.0 web server was used for the identification of both covalent and non-covalent bonds in protein structures (Piovesan et al. 2016). The RING output was visualized directly using Pymol and the python script of RING-Viz script.

5.3 RESULTS AND DISCUSSION

Discovery of anti-tuberculosis drugs has been a challenge due to the alarming emergence of MDR and XDR strains of *Mtb*, making tuberculosis a global health threat. Moreover, the existing anti-tuberculosis drugs have been associated with various side effects. Therefore, extended efforts are needed for the identification of a novel target and the discovery of a specific anti-tuberculosis drug. Many computational tools are available for the structure-based drug design and optimization of lead compounds. *Mtb* TrxR has been shown to be a druggable target, which helps mycobacteria to survive against oxidative killing in the host cell. Earlier inhibitors of *Mtb* TrxR were designed based on a protein–protein interaction site. It has been known that protein–protein interaction site has a low hit rate in inhibitor designing efforts. Moreover, it is not easily druggable. Allosteric sites on protein are considered to be a druggable target for designing a selective inhibitor. Hence,

an allosteric site is chosen for the specific targeting of *Mtb* TrxR. *Mtb* TrxR has a unique feature of catalysis compared to the higher eukaryotic TrxR. It requires rotating the NADPH domain through the hinge region to bring about the catalysis. Hence, the cavity around the hinge region is targeted in our study using virtual screening of natural compounds from the TCM database.

5.3.1 Structure-based virtual screening and docking

The virtual screening function in iScreen identifies potential TCM compounds by a docking algorithm based on the protein structure and the binding site information. The predicted cavity with an area of $1301.554\text{\AA}^2$ and volume $855.756\text{\AA}^3$ located between the NADPH and FAD domains, as shown in **(Figure 5.1 A)**, was utilized for the binding analysis of TCM compounds and which in turn generated a docking score of each compound **(Table 5.1)**. Among all compounds, two promising compounds (called here compound1 and compound2) were selected based on their docking score value and interaction orientation in the complex. Pharmacokinetic study of inhibitor compounds plays a critical role in the development of a drug. Since these compounds are screened from the TCM database, they have acceptable absorption, distribution, metabolism, and excretion criteria. The two-dimensional chemical structures of the screened compounds are shown in **(Figure 5.1 B)**.

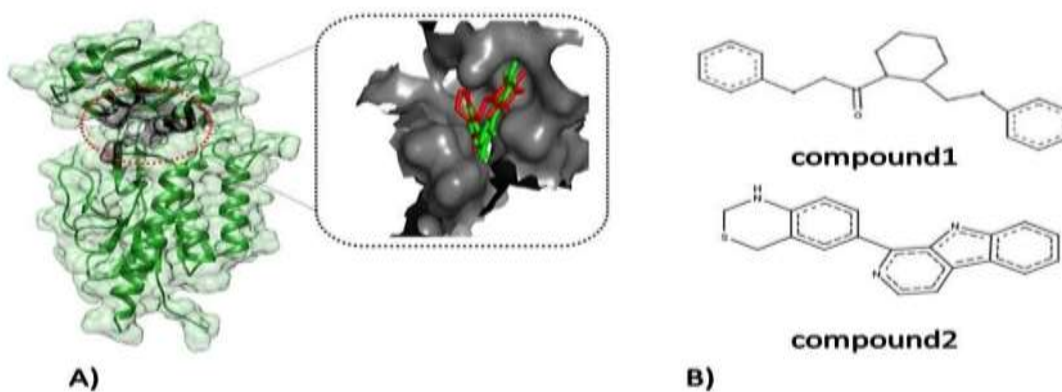


Figure 5.1. Predicted binding cavity on the *Mtb* TrxR structure shown in surface representation (circled in red). The screened compounds docked in the cavity are shown in the blowup. B) The chemical structure of two screened compounds: compound1, officinaruminane_B and compound2, dichotomine_D.

TCM compound	iScreen Dock Score	Autodock Score (kcal/mol)
Compound 1	-96.9407	-5.54
Compound 2	-85.5386	-5.66

Table 5.1. Docking score of top two screened compounds from the TCM database.

5.3.2 Binding interaction analysis of hit compounds

Binding interactions of hit compounds with the binding cavity in TrxR were analyzed with LIGPLOT (Wallace et al. 1996) as well as manually (**Figures 5.2A & 5.2B**). Different types of molecular binding interactions were observed to stabilize the compounds in the binding cavity and the most prominent is the hydrophobic bonding interaction with important residues near the active site and hinge region as well as the interface residues of NADPH and FAD domains. The binding pattern within the cavity is almost similar in both compounds except that three additional hydrogen bonds were observed between compound2 and the side chains of Arg295, Glu170, and His250. Interestingly, it was also observed that cation pi and pi-pi interactions in both cases are crucial for an enhanced binding stability in the cavity. It is not surprising that hydrophobic interactions also contribute to the stable binding of the inhibitor to the protein target, as it has been reported earlier

for the inhibition of MDM2 by polyphenols (Verma, Grover, Tyagi, Goyal, Jamal, & Singh et al., 2016).

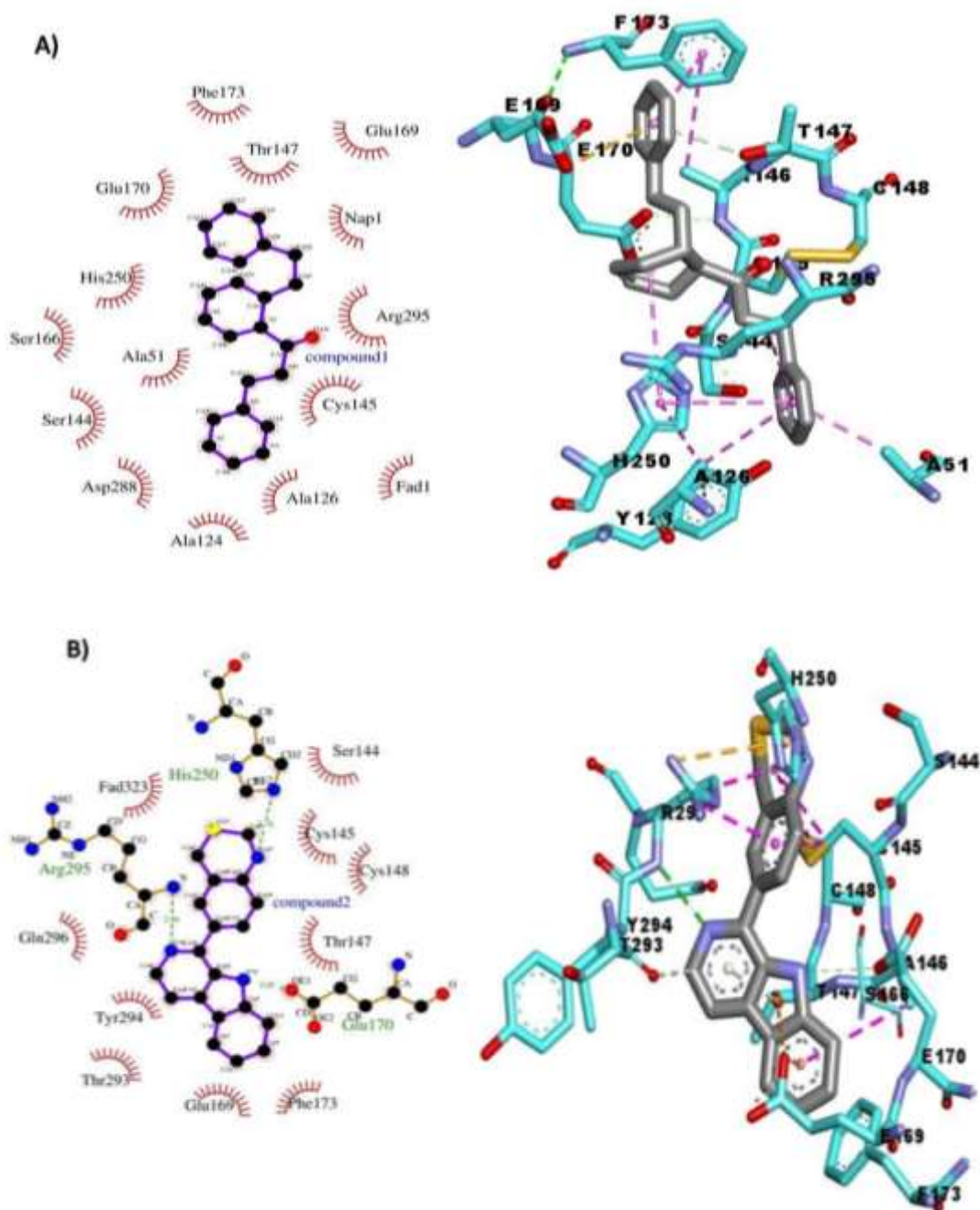


Figure 5.2. LIGPLOT and a 3-D representation of protein–ligand interactions of the screened compounds in the selected cavity of *Mtb* TrxR. A) Compound1 B) Compound 2. Hydrophobic interactions are shown in brick red arch and hydrogen bonds are shown with green dashed lines.

5.3.3 Analysis of Molecular Dynamics Simulations

In order to study the stability and dynamic behavior of the two hit compounds in the complex with *Mtb* TrxR, a molecular dynamics simulation study was performed for 20 ns. The conformational stability and convergence of TrxR and TrxR bound with hit compounds (TrxR^{comp1}, TrxR^{comp2}) were evaluated by calculating the root mean square deviation (RMSD) of the backbone atoms. RMSD of the protein backbone against the time of the simulation is represented as trajectories (**Figure 5.3A**). These trajectories show that the system obtained adequate stability after 15 ns of the molecular dynamics simulations, explained by the low variation of the TrxR protein backbone after forming a complex with the individual compound. Interestingly, while both compound complex forms showed similar RMSD trends, the apo form of TrxR tended to have a higher RMSD after 15 ns of the simulation. The higher RMSD is due to enhanced motion of one of the loops (region 47–72) from the FAD binding domain. This loop is observed to be stabilized by forming new hydrogen bond interactions with one of the α -helices of the NADPH domain in the compound1 complex form of TrxR.

Individual amino acid residues of a protein play a significant role in providing stability to selective inhibitor binding to the protein. Hence, the position and the relative flexibility of each amino acid residue of TrxR were analyzed in the molecular dynamics simulations. The RMSF trajectories give information about each amino acid fluctuation in TrxR as shown in (**Figure 5.3 B**). A movement (>6 Å) can be observed in the loop region (residues 47–72) of the FAD domain of the apo-TrxR

protein, which overlaps with the high RMSD. Binding of the compounds altered the RMSF of this region, suggesting that fluctuation of the loop region facilitates the compounds to bind. The binding induces conformational changes in both NADPH and FAD domains. The compounds binding residues in TrxR, such as A51, A124, A126, S144, C145, T147, C148, S166, E169, E170, and F173, are relatively stable and fluctuate little. In fact, no significant fluctuations of amino acid residues were observed during the entire simulation. As reported earlier (Akif et al. 2005; Waksman et al. 1994), the motion of NADPH domain with respect to the FAD domain is functionally important for the catalysis. The hinge region is less flexible but provides a pivotal point for functionally relevant motion of the NADPH domain. However, to test the effect of the compounds in the hinge region, we performed position restrained MDS, in which hinge region residues were position restrained. Simulation of the restrained apo-TrxR for 20 ns provided an average RMSD of 0.1 nm whereas the average RMSD of the unrestrained TrxR was observed to be more than 0.3 nm. Further, similarity in the RMSD profile of the restrained apo-TrxR and the compound bound complexes suggests the rigidification of the hinge region upon binding of the compounds (Dantu et al. 2016). Binding of compounds with TrxR, mostly through hydrophobic interactions, caused conformational changes and formed a new set of interactions between the NADPH and the FAD domains. This suggests that the new interactions between the two domains would resist functionally relevant motion for the catalysis. In addition, compound1 was also observed to interact with the active site residues of TrxR.

Solvent accessible surface area (SASA) is a parameter computed using the `gmx_sasa` module of GROMACS, which measures the proportion of protein surface exposed to the water solvent. The buried amino acid residues in the hydrophobic core of the protein are the driving force for protein

folding. The relative value of SASA can predict the extent of conformational changes in the protein that have occurred during the course of binding (Marsh et al. 2011). The SASA value for an unbound protein was calculated to be 180 nm² whereas the compound-bound protein exhibited a slight decline in the solvent residual exposure. The average SASA value of the bound compounds 1 and 2 was calculated as 160 and 163 nm² respectively (**Figure 5.3 C**). This indicates that bound forms were least exposed to the water solvent during the 20 ns of the MDS, which could suggest that the binding of compounds changed the conformation of the protein and that it could provide an irreversible inhibitory effect by binding to the interactive surface residues near the protein hydrophobic core.

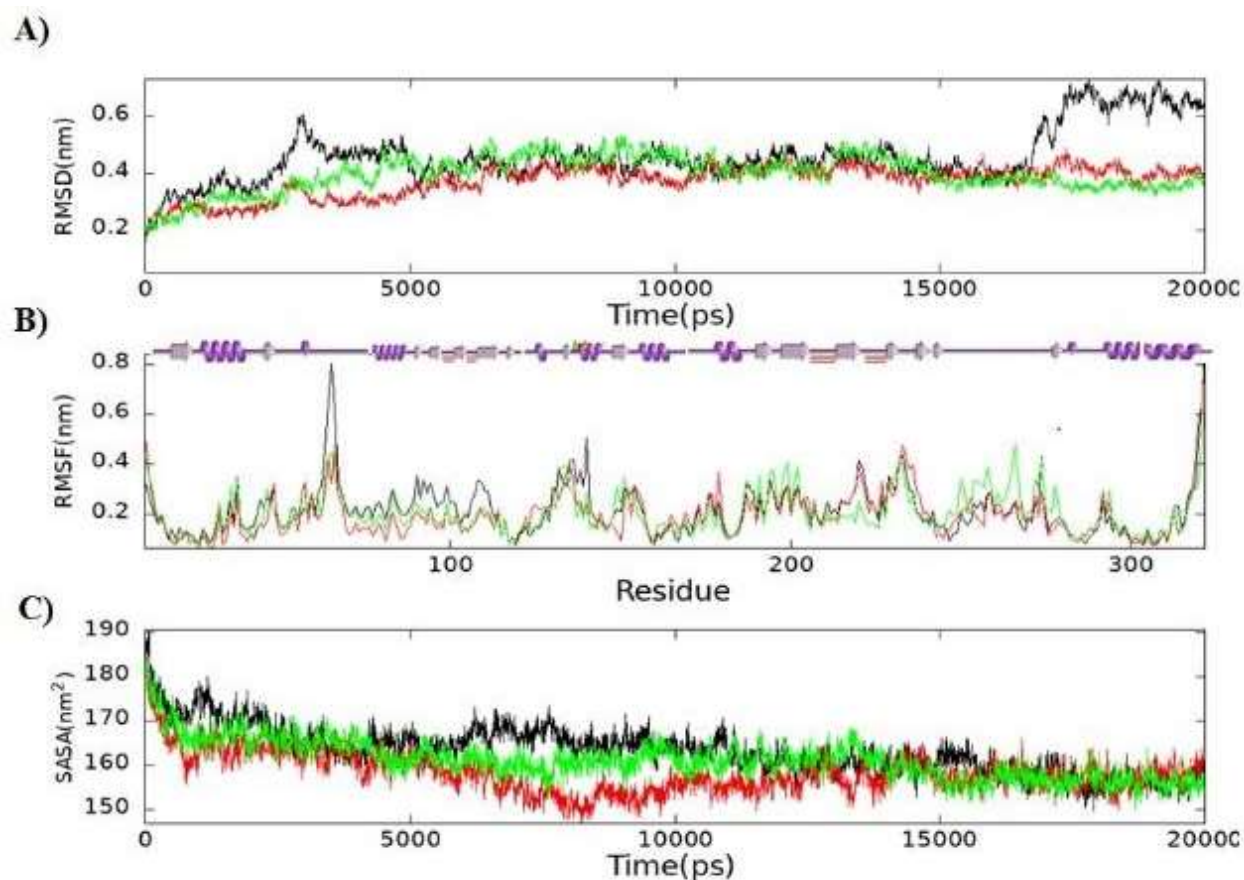


Figure 5.3. A) Backbone RMSD vs time. B) RMSF vs residue number. Secondary structure arrangement of TrxR according to residues shown on the top of the RMSF plot. C) SASA vs time of apo-TrxR complex with compound1 and compound2 and the trajectories are represented in black, red, and green respectively.

5.3.4 Evaluation of binding affinity with the protein molecule

The free energy calculations using MM/PBSA method provides an important parameter for estimating the binding affinities of ligands with a protein molecule (Wang et al. 2013). The

MM/PBSA calculations yielded the binding-free energies of compounds 1 and 2 with TrxR, listed in Table 5.2. The results indicate that compound1 possessed a higher negative binding free energy value of -151.385 kJ/mol compared to compound2, which had the value of -121.613 kJ/mol. These binding energies suggest a significant potential for the formation of stable molecular interactions with amino acid residues of the TrxR. Interestingly, it was noted that Van der Waals, electrostatic interactions and non-polar solvation energy negatively contributed to the total interaction energy while only polar solvation energy positively contributed to the total free-binding energy. Overall, the high negative value of Van der Waals energy suggests massive hydrophobic interactions are dominant in the formation of a stable protein–ligand complex.

System	Energy (kJmol ⁻¹)				
	Van der Waals	Electrostatic	Polar Solvation	SASA	Total
TrxR_compound 1	-208.35 ±10.43	-5.30 ±4.79	80.86 ±15.74	-18.58± 1.15	-151.38 ±14.43
TrxR_compound 2	-166.76 ±6.71	-15.00 ±8.69	74.79 ±9.55	-14.63 ±1.00	-121.61 ±9.57

Table 5.2. Average MM/PBSA free energies of *Mtb* TrxR complexes with compounds, calculated from the MD simulation performed at 20 ns.

5.3.5 Analyses of protein conformational variation

The MD trajectories of the system were inspected with the principal components to get a better

understanding of the conformational changes of *Mtb* TrxR in the complex with two compounds. Correlated motion plot shows how atoms move relative to each other. Motions can be correlated (in the same direction), anti-correlated (in the opposite direction), or uncorrelated (Kasahara et al. 2014). Anti-correlated motions were observed to be significantly dominant in the complex forms (Figure 5.4). A summary of significant motions between two domains/regions such as NADPH–FAD, NADPH–Hinge and FAD–Hinge is presented in Table 5.3.

System	Regions		
	NADPH-FAD	NADPH-Hinge	FAD-Hinge
Apo TrxR	Weakly correlated	Weakly correlated	Weakly correlated
Compound 1	Mixed	Mixed	Strongly-anti correlated
Compound 2	Weakly-anti-correlated	Weakly-anti-correlated	Weakly-anti-correlated

Table 5.3. Dominant motion of atoms in NADPH and FAD domains and the hinge region of the protein

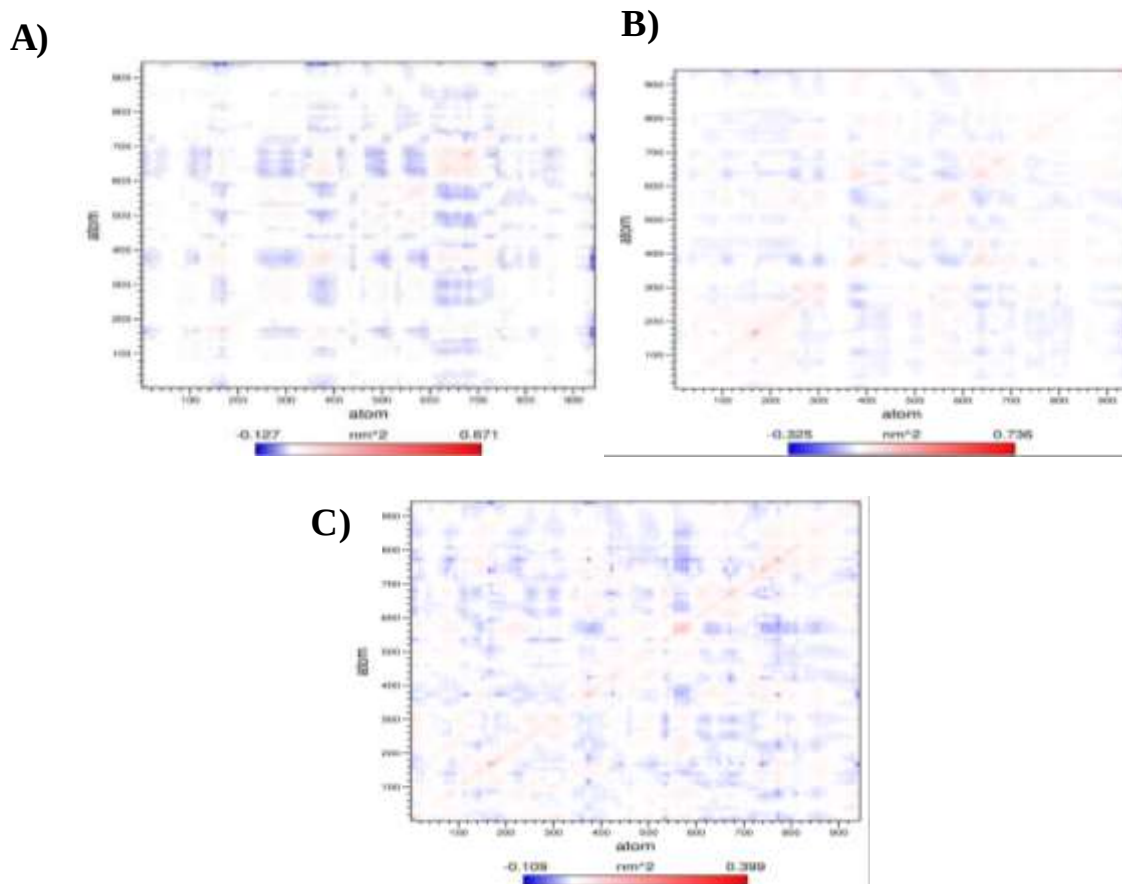


Figure 5.4. Covariance matrix plot of apo and complex form of *Mtb* TrxR during 20 ns MD simulation. A) apo-TrxR. B) complex with compound1. C) Complex with compound2. The positive and negative limits are shown.

The magnitude (eigenvalue) and direction (eigenvector) of overall atomic motions in the apo- and complex bound forms of TrxR were evaluated using principal components (PC) (Peng et al. 2014). The first few components can be interpreted as directions, which represent the maximum variance in the backbone atoms. The first two eigenvectors of the apo-TrxR and complex-bound form represent a significant number of conformational dynamics (Figure 5.5). The projections of the

motion on the first two eigenvectors imply that the TrxR complex covers a small space and shows a reduced number of conformational motions along PC1 and PC2 projections as compared to the apo-TrxR. In complex-bound forms of TrxR, each point in phase space describes the specific conformation and reduced displacement in phase space highlights lower conformational sampling upon binding with the compounds. Additionally, the conformational sampling of both systems was inspected by tracing the covariance matrix for backbone atom positions. For apo-TrxR, the covariance trace value was observed to be 40.44 nm² and lower covariance trace values of 31.82 and 32.98 nm² were observed for both compound1 and compound2 complex bound forms respectively (Figure 5.5). Overall, PCA suggests that binding of both compounds with TrxR results in a significant change in the overall motion of TrxR with a compressed conformational space.

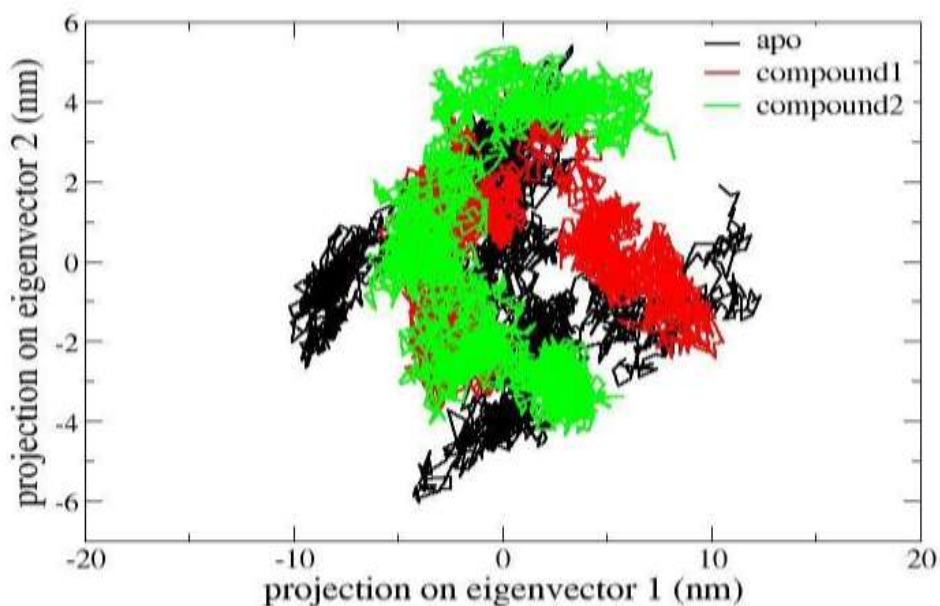


Figure 5.5. PCA plot constructed by eigenvector 1 and eigenvector 2. The conformational sampling of apo state, compound1 and compound2 bound to the TrxR is represented in black, red, and green respectively.

The simulated structures of both complexes causing the conformational variations in the protein domains were also analyzed using the PCN (Protein Contact Network) method that reveals differences in contacts among the nodes (Piovesan et al. 2016). It indicates that the total number of contacts were slightly increased in the active site region as well as in the interface of NADPH and FAD domains in the complex structure as compared to the apo-TrxR. While the number of contacts in the hinge region remains unchanged in both cases as compared to the apo-TrxR (Figure 5.6). It suggests that the rearrangement of contacts between the interface domains contributes to the conformational changes in the complexes. Moreover, it was also observed that the binding of compounds impacted the local interaction network between the NADPH and FAD interface domains and the number of linked nodes, links, and links mediated by hubs were slightly changed. Furthermore, the center of mass distance between NADPH and FAD interface regions of the compound-bound forms of TrxR was found to change significantly as compared to the apo-TrxR form. The center of the mass moved closer in the compound-bound forms (Figure 5.7). This clearly suggests that the binding of the compounds to TrxR restricts the conformational flexibility.

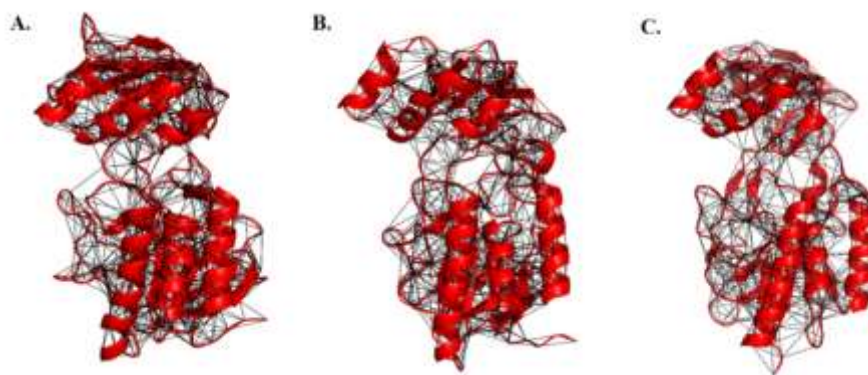


Figure 5.6. Protein contact network of the apo and compound bound *Mtb* TrxR. Contact network of among the residues of the apo and the compound-bound forms of TrxR.

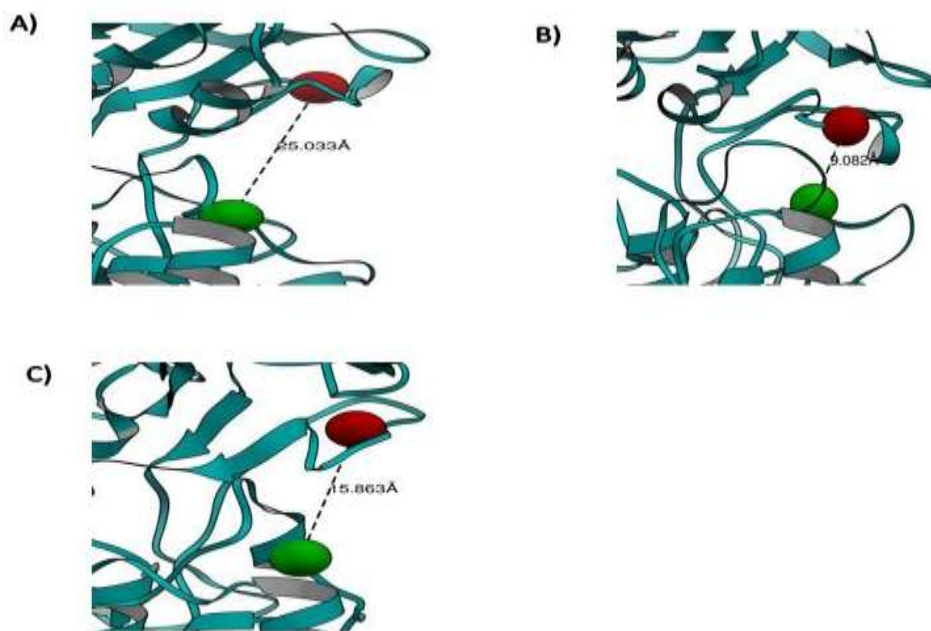


Figure 5.7. Center of mass distance calculation between the NADPH and the FAD interface domain residues. A) Apo form B) Compound1 bound form C) Compound2 bound form.

Furthermore, the comparison of the structures of apo-TrxR with the compound-bound forms of TrxR revealed a significant shift of one of the α -helices from the NADPH domain of TrxR–compound1 towards FAD domain and was observed to stabilize by many new hydrogen bonds. The carbonyl oxygen of Gly139 of the NADPH domain forms a H-bond with the side chain of Thr54 of the FAD domain. In addition, Glu134, Glu135 of the NADPH domain formed H-bonds with Ser47, Gly49, 50 of the FAD domain, respectively (Figure 5.8). The formation of additional contacts in the complexes may restrict the conformational flexibility of the NADPH binding domain, which in turn may inhibit the ability of *Mtb* TrxR for transferring the electrons to the substrate thioredoxin.

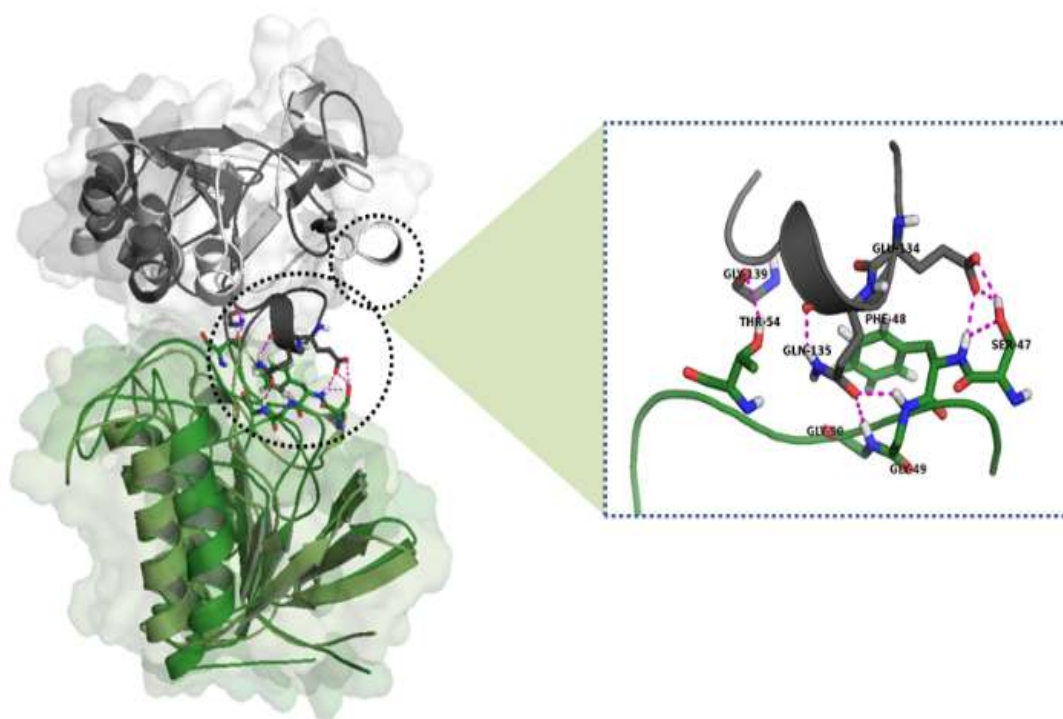


Figure 5.8. Superimposition of structure of the apo-TrxR and compound1 complex form of TrxR and their respective domains are shown in different colors. The NADPH domain of complex and apo-TrxR is represented in dark gray and light gray respectively. The FAD domain of complex and FAD domain is shown in dark green and light green respectively. Dotted circles show the orientation of the α -helix in both. The H-bonds formed between the residues from NADPH and FAD domains in the complex are shown in the blowup.

5.4 CONCLUSIONS

Mtb TrxR has been validated as an essential gene of *Mtb* and also shown to be a druggable candidate. This study reports screening of natural compounds from the TCM database against *Mtb* TrxR. This led to the identification of 200 compounds and the top two compounds, compound1

and compound2, were selected. Both compounds bound to the hydrophobic groove of the *Mtb* TrxR and the binding was stable throughout the MDS. Finally, the data obtained in the study indicate that both these natural compounds may have the potential to bind with *Mtb* TrxR and can be used as lead molecules for the inhibition of *Mtb* TrxR.

CHAPTER-6

Summary

6. 1 SUMMARY

In the current paradigm of treating infectious diseases, protein molecules are the best therapeutic targets. They are involved as essential virulence factors of many microorganisms. Studying the structure-function relationship of any therapeutic target is important for understanding its role in the diseased state. The function of a protein is determined by its three-dimensional structure. This information is required for designing therapeutics. The static structure obtained from X-ray crystallography or NMR is not enough to define the functional states of the protein. It has been observed that specific modulations in the protein structure, like mutations and interaction with other entities, may influence the protein's three-dimensional shape, which can result in changes in function. However, some cases have been observed where modulations bring change in function with little change in its 3D structure. So, this suggests there is a missing link between the structure and function paradigm. It is challenging to extract the subtle protein structure differences due to modulations that maintain the structural integrity. Protein's conformational flexibility and dynamics contribute towards understanding the molecular mechanism, which suggests that the dynamics of protein structure bridge the gap between structure and function. The thesis work attempts to investigate the effect of modulations on the protein structure, which is not detectable at the structural level but influences the role in binding. For the investigation, three well-known therapeutic targets were taken into consideration. Integrative Protein Contact Network (PCN) and Molecular Dynamics Simulation (MDS) approaches were utilized to study the dynamics of small conformational changes that lead to the change in the protein's function.

Amino acids interaction networks in a protein provide an insight into the molecular mechanism of protein folding and unfolding, probing hotspot residues, and identifying allosteric effects. Our

study constructed the Protein Structure Graph on a single static structure and applied the ENM-NMA approach for building a network based on the cross-correlation of motion between the atoms. The study also used Molecular Dynamics Simulations to investigate protein dynamics at the atomistic level. It is an important approach to extracting the molecular mechanism of biomolecules in different physiological conditions. The protein undergoes different modulations, which result in conformational changes due to the gain or loss of contacts. Using these combinatorial approaches, one can uncover various changes that are not even identified at the structural level.

In our study, three different therapeutic targets from two different systems were chosen. The first system is a topical interest target, spike protein, from SARS-CoV-2. The spike protein is an important drugable and vaccine candidate. It consists of two domains: S1 and S2. The S1 domain includes NTD (N-terminal Domain), CTD (C-terminal Domain), SDI, and SDII (sub-domain I&II) domain. RBD is present in CTD, which binds with the ACE2 receptor. In this current pandemic storm, several mutations have been identified in Spike protein, making the available vaccines ineffective. In the early days of the pandemic, three Variants of Concern (VOCs) of SARS-CoV-2 gained much attention; alpha, beta, and gamma strains. The gamma variant, also known as the P.1 variant of SARS-CoV-2, possesses 12 mutations on the spike protein. Previous In-vitro kinetics data shows that the P.1 variant displays greater affinity towards the ACE2 receptor than the prototype strain. However, the cross-structure RMSD between spike proteins of wild-type (WT) and P.1 variant was almost negligible to 0.2 Å. Hence, it intrigued us to identify the factors responsible for the change in the affinity of the P.1 variant. How are mutations making the available drugs/vaccines ineffective without any change at the structural level? To address this, Protein Contact Network and dynamics approaches were applied to identify the subtle changes responsible for the increase in affinity with the ACE2 receptor. Protein contact network analysis

has suggested that mutations on the RBD have perturbed the network parameters like hubs, communities, and shortest communication pathways, which results in the loss of nodes and links participating in the network parameters in the P.1 variant. Further, the effect of mutations outside the RBD domain was also investigated. The observation highlighted the rearrangement of node-wise contacts, but the alteration was more confined to the core of the NTD in the P.1 variant. A comparison of the thermodynamics of spike protein of wild-type and P.1 variant also suggests that mutations provide a destabilizing effect. Our comparative dynamics study highlighted the conformational variations among the RBD-ACE2 and truncated S1 domain-ACE2. The PCA results suggested more conformational heterogeneity in the P.1 variant than in the wild type. Overall, the results indicated a loss of interactions among the domains, which account for greater flexibility and better conformational adaptability in the case of the P.1 variant.

In the subsequent section, a small molecule therapeutic was designed to target the interface region of the RBD-ACE2 complex from the P.1 variant. The chosen parent small molecule (SSAA09E2) was already a validated potent inhibitor against the interface RBD-ACE2 complex of the SARS-CoV-2. But the same inhibitor showed less affinity for the same complex of the P.1 variant. This observation emphasized our theme that modulations in the protein have maintained structural integrity, but the binding potency of the inhibitor differs. Using the parent molecule SSAA09E2, various analogs were designed through a fragment replacement approach. The approach used to probe the fragment's space to be replaced with other active compounds. Among all the screened compounds, only two analogs were chosen that showed better affinity than the parent compound. Binding conformational analysis suggested that both compounds are sitting on the interface of RBD-ACE2. They were also involved in potential interactions like hydrogen or other covalent interaction with the interface residues. Molecular Dynamics Studies for 200ns

demonstrated the stability of designed analogs inside the cavity. The RMSD profile of both compounds suggested that the protein backbone trajectory was stable throughout the simulations and also maintained some crucial interactions involved with the interface residues. Binding free energy calculation also indicated that both compounds have high negative free energy values with the protein complex. The residue-wise energy contribution suggested that the interface residues had the most significant contribution in the binding of designed compounds at the interface of the RBD-ACE2 complex. Additionally, the results highlighted the molecular mechanism of designed compounds on the RBD-ACE2 binding. The RMSF of the interface region of RBD has increased significantly for the compound-bound complexes. The increased flexibility can lead to higher entropy penalty, negatively influencing the binding energy. Additionally, both compounds also showed some secondary effect on the functional conformation of the RBD of the spike protein for some region which was away from the interface. Our residue-wise RMSD analysis also corroborates with RMSF results. The effect of the compounds in the complex was further analyzed on the internal motion of the RBD-ACE2 complex. The conformational space sampling confirms that the compound 1 bound complex was observed to occupy a greater area along its first and second principal components. The porcupine plots indicated a change in the internal motion for both compound-bound complexes. Hydrogen distances between key interface residues increased after binding both the compound bounds. This may bring about a disruption in the interfacial residues interaction. Overall our work suggests that both compounds have the potential to induce structural changes both locally and globally, which can interfere with the critical interface residues.

The second system chosen for our study was M^{pro} from SARS-CoV-2. It has been designated an important drug target because of its essential role in processing polyproteins translated from the

viral RNA. M^{pro} from SARS-CoV-1 and SARS-Cov-2 shares a high degree sequence identity of 96.1%. So early in the pandemic, it was hypothesized that designed inhibitors against SARS-CoV-1 M^{pro} may also have the same effect. Unfortunately, the reported inhibitors had shown differential inhibition on the activity of M^{pro}s from SARS-CoV-1- and SARS-CoV-2. The M^{pro} is a homodimer cysteine protease where each protomer consists of three domains, domain I (residue 8-101), domain II (residues 102-184) and domain III (residues 201-303) and catalytic residues and the substrate binding sites are situated between domains I and II. Though, M^{pro}s from SARS-CoV-2 and SARS-CoV-2 show slight variation at the amino acid sequence level. But no significant structural changes were noticed. Hence it was hypothesized that the changes in amino acid could contribute to the differential effect of inhibitors on M^{pro}s. In this study, with the application of a protein structure network (PSN), minor conformational changes associated with the protein structure were investigated. The PSN analyzed important network parameters like degree, hubs, betweenness centralities, communities, and shortest paths. First, the analysis was focused on the apo form of M^{pro} from SARS-CoV-1 and SARS-CoV-2. A change in the degree of residues at the active site regions was observed. In addition, a change in the degree of residues was also noticed at the N and C-terminal residues, which were crucial for the dimerization of M^{pro}. Interestingly replacement of A46 in SARS-CoV-1 M^{pro} with S46 in SARS-CoV-2 M^{pro} resulted in the rearrangement of contacts and was observed to form new connections. The total number of hubs was also changed in both cases. Few hubs were distinctive to each structure, suggesting their important role in interactions and stability. We have analyzed the residues involved in the high betweenness centrality. Residues with high BC were involved in the formation of new contacts. Further, we have analyzed the community structure analysis on both structures. The rearrangement of modules was observed throughout the structure, indicating the perturbation at

the global level in the 3-D structure of two proteins. Then, we constructed PCN on inhibitor-bound complexes to analyze the network parameters of inhibitor-bound complexes. The binding of inhibitors to the M^{pro} generated many hubs at the inhibitor binding site. Few hubs are unique in each structure, suggesting their role in the specificity of inhibitors. Further, we have analyzed perturbations in inter and intra-subunit communication due to the binding of inhibitors. We have analyzed the meta-path and mapped the residues participating in each path. There was an increase in the length of communication paths for the SARS-CoV-1 inhibitor-bound complex. We have also observed the changes in the most frequent nodes and links in structural communication. Community rearrangements suggest that the inhibitors induce perturbations in the network connectivity compared to the apo form. As it has been reported that biologically active SARS-CoV M^{pro} exists as a dimer, we have constructed PCN on its dimeric state and computed the PCN parameters. The result highlighted the identification of the crucial residues from the subunit interfacial region.

The third target, Thioredoxin Reductase (TrxR), was chosen for our study from the bacterial system, *Mycobacterium tuberculosis* (*Mtb*). TrxR is a well-known therapeutic target and antioxidant system involved in redox potential and provides a reducing environment to the *Mtb* inside macrophages. *Mtb* TrxR consists of two domains NADPH and FAD binding domain. The distance between NADPH and FAD binding region is 15 Å apart, which is very large for any electron transfer mechanism. Hence, a rotation of 66° in the NADPH domain with respect to the FAD domain is required for the catalysis. This rotation is possible due to a hinge region in the protein. It has already been demonstrated that this conformational rotation is essential for the activity of TrxR. Our study hypothesized to restrict this conformational change by designing a compound-like molecule targeting the interface residues from NADPH and FAD along with the

hinge region residues. First, we probed a druggable pocket and a cavity of area 1301.554Å² and volume 855.756Å³ located between the NADPH and FAD domains subjected to the screening process. We screened 20,000 natural compounds from the Traditional Chinese Medicine database and took the first two compounds for our study, which showed high docking scores. Then we applied 20ns MD on the apo and compound-bound complexes to analyze the stability and molecular mechanism behind the screened compounds. The RMSD trajectory was very stable for all three systems. But for the apo form, it was quite high after 15ns. The increased flexibility region stabilized in the compound-bound complexes. Protein residues interacting with compounds had reduced flexibility compared to apo form. The average SASA value for the apo state was also high in the apo case, indicating the change in conformation where compound-bound complexes were less exposed to water. The binding free energy calculation showed compound1 had a higher affinity than compound2. Further, the PCA analysis has suggested that the anti-correlation between the atoms was more dominant between the compound-bound complexes. The magnitude and direction of motion were also changed, where the compound-bound complexes covered less space. Additionally, we have analyzed the change in contact pattern by constructing PCN on all three complexes. It indicated that the total number of contacts was slightly increased in the active site region and the interface of NADPH and FAD domains in the complex structure compared to the apo-TrxR. So there was a rearrangement of contacts in both domains, indicating that both compounds were causing the conformational change in the protein. So it suggested that both compounds have the potency to bind with *Mtb* TrxR and can be used as a lead molecule to inhibit *Mtb* TrxR.

Overall, the study of the three different therapeutic targets highlights the following conclusions

- A subtle change in conformation induces a change in function due to modulation caused by several factors like mutations, inhibitor binding, or other types of interactions.
- PCN and MDS are useful approaches to identifying the subtle changes which are not identified at the structural level.
- An understanding of the dynamics of proteins can help in identifying the conformational changes obtained from static protein structures.
- Integrating PCN and MDS studies can help in identifying the important residues which are contributing to change in function but maintain structural integrity.

BIBLIOGRAPHY

- Adedeji, A. O., Severson, W., Jonsson, C., Singh, K., Weiss, S. R., & Sarafianos, S. G. (2013). Novel inhibitors of severe acute respiratory syndrome coronavirus entry that act by three distinct mechanisms. *Journal of virology*, 87(14), 8017–8028.
- Akif, M., Khare, G., Tyagi, A. K., Mande, S. C., & Sardesai, A. A. (2008). Functional studies of multiple thioredoxins from *Mycobacterium tuberculosis*. *Journal of bacteriology*, 190(21), 7087–7095.
- Akif, M., Suhre, K., Verma, C., & Mande, S. C. (2005). Conformational flexibility of *Mycobacterium tuberculosis* thioredoxin reductase: crystal structure and normal-mode analysis. *Acta crystallographica. Section D, Biological crystallography*, 61(Pt 12), 1603–1611.
- Alberts B, Johnson A, Lewis J, et al. *Molecular Biology of the Cell*. 4th edition. New York: Garland Science; 2002. Introduction.
- Amitai, G., Shemesh, A., Sitbon, E., Shklar, M., Netanel, D., Venger, I., & Pietrokovski, S. (2004). Network analysis of protein structures identifies functional residues. *Journal of molecular biology*, 344(4), 1135–1146.
- Anand, K., Palm, G. J., Mesters, J. R., Siddell, S. G., Ziebuhr, J., & Hilgenfeld, R. (2002). Structure of coronavirus main proteinase reveals combination of a chymotrypsin fold with an extra alpha-helical domain. *The EMBO journal*, 21(13), 3213–3224.
- Anfinsen, C. B., Haber, E., Sela, M., & White, F. H., Jr (1961). The kinetics of formation of native ribonuclease during oxidation of the reduced polypeptide chain. *Proceedings of the National Academy of Sciences of the United States of America*, 47(9), 1309–1314.
- Arashkia, A., Jalilvand, S., Mohajel, N., Afchangi, A., Azadmanesh, K., Salehi-Vaziri, M., Fazlalipour, M., Pouriaeyali, M. H., Jalali, T., Mousavi Nasab, S. D., Roohvand, F., Shoja, Z., & SARS CoV-2 Rapid Response Team of Pasteur Institute of Iran (PII) (2021). Severe acute

respiratory syndrome-coronavirus-2 spike (S) protein based vaccine candidates: State of the art and future prospects. *Reviews in medical virology*, 31(3), e2183.

Artymiuk, P. J., Rice, D. W., Mitchell, E. M., & Willett, P. (1990). Structural resemblance between the families of bacterial signal-transduction proteins and of G proteins revealed by graph theoretical techniques. *Protein engineering*, 4(1), 39–43.

Astuti, I., & Ysrafil (2020). Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2): An overview of viral structure and host response. *Diabetes & metabolic syndrome*, 14(4), 407–412.

Atilgan, A. R., Akan, P., & Baysal, C. (2004). Small-world communication of residues and significance for protein dynamics. *Biophysical journal*, 86(1 Pt 1), 85–91.

B., Li, L., Hurley, T. D., & Meroueh, S. O. (2013). Molecular recognition in a diverse set of protein-ligand interactions studied with molecular dynamics simulations and end-point free energy calculations. *J Chem Inf Model*, 53, 2659-2670.

Bagdonaite, I., & Wandall, H. H. (2018). Global aspects of viral glycosylation. *Glycobiology*, 28(7), 443–467.

Banerjee, R., Perera, L., & Tillekeratne, L. (2021). Potential SARS-CoV-2 main protease inhibitors. *Drug discovery today*, 26(3), 804–816.

Barrila J, Gabelli SB, Bacha U, Amzel LM, Freire E.(2010). Mutation of Asn28 disrupts the dimerization and enzymatic activity of SARS 3CLpro. *Biochemistry*, 49(20):4308-4317.

Becker, K., Gromer S., Schirmer R. H. and Muller S. (2000). Thioredoxin reductase as a pathophysiological factor and drug target. *Eur J Biochem*, 267: 6118-6125.

Bento, A. P., Gaulton, A., Hersey, A., Bellis, L. J., Chambers, J., Davies, M., Krüger, F. A., Light, Y., Mak, L., McGlinchey, S., Nowotka, M., Papadatos, G., Santos, R., & Overington, J. P. (2014).

The ChEMBL bioactivity database: an update. *Nucleic acids research*, 42(Database issue),

Berendsen HJC, Grigera JR, Straatsma TP. 1987. The missing term in effective pair potentials. *J Phys Chem* 91(24):6269–6271.

Berman, H.M. et al., 2000. The Protein Data Bank. *Nucleic Acid Research*, 28(1), pp.235–242.

Boccaletti S, Latora V, Moreno Y, Chavez M, Hwang DU. (2006). Complex networks: Structure and dynamics. *Phys Rep*, 424(4-5):175-308.

Brinda K V., Vishveshwara S. (2005). A network representation of protein structures: Implications for protein stability. *Biophys J*, 89(6):4159–4170.

Buss LF, Prete CA, Abraham CMM, Mendrone A, Salomon T, Almeida-Neto C De, França RFO, Belotti MC, Carvalho MPSS, Costa AG, Crispim MAE, Ferreira SC, Fraiji NA, Gurzenda S, Whittaker C, Kamaura LT, Takecian PL, Silva Peixoto P Da, Oikawa MK, Nishiya AS, Rocha V, Salles NA, Souza Santos AA De, Silva MA Da, Custer B, Parag K V., Barral-Netto M, Kraemer MUG, Pereira RHM, Pybus OG, Busch MP, Castro MC, Dye C, Nascimento VH, Faria NR, Sabino EC. (2021). Three-quarters attack rate of SARS-CoV-2 in the Brazilian Amazon during a largely unmitigated epidemic. *Science* (80-) 371(6526):288–292.

Bussi G, Donadio D, Parrinello M. (2007). Canonical sampling through velocity rescaling. *J Chem Phys* 126(1).

Casalino, L., Gaieb, Z., Goldsmith, J. A., Hjorth, C. K., Dommer, A. C., Harbison, A. M., Fogarty, C. A., Barros, E. P., Taylor, B. C., McLellan, J. S., Fadda, E., & Amaro, R. E. (2020). Beyond Shielding: The Roles of Glycans in the SARS-CoV-2 Spike Protein. *ACS central science*, 6(10), 1722–1734.

Chakrabarty B, Parekh. (2016). N. NAPS: Network analysis of protein structures. *Nucleic Acids Res*, 44(W1):W375-W382.

- Chan KK, Dorosky D, Sharma P, Abbasi SA, Dye JM, Kranz DM, Herbert AS, Procko E. (2020). Engineering human ACE2 to optimize binding to the spike protein of SARS coronavirus 2. *Science* (80-) 369(6508):1261–1265.
- Chen, C. Y. (2011). TCM Database @Taiwan: the world's largest traditional Chinese medicine database for drug screening in silico. *PLoS One*, 6:e15939.
- Chou CY, Chang HC, Hsu WC, Lin TZ, Lin CH, Chang GG. (2004). Quaternary structure of the severe acute respiratory syndrome (SARS) coronavirus main protease. *Biochemistry*, 43(47):14958-14970.
- Clatworthy AE, Pierson E, Hung DT. (2007). Targeting virulence: a new paradigm for antimicrobial therapy. *Nat Chem Biol*, 3(9):541-8. doi:10.1038/nchembio.2007.24.
- Cornillez-Ty, C. T., Liao, L., Yates, J. R., 3rd, Kuhn, P., & Buchmeier, M. J. (2009). Severe acute respiratory syndrome coronavirus nonstructural protein 2 interacts with a host protein complex involved in mitochondrial biogenesis and intracellular signaling. *Journal of virology*, 83(19), 10314–10318.
- Cottam, E. M., Whelband, M. C., & Wileman, T. (2014). Coronavirus NSP6 restricts autophagosome expansion. *Autophagy*, 10(8), 1426–1441.
- Coutard, B., Valle, C., de Lamballerie, X., Canard, B., Seidah, N. G., & Decroly, E. (2020). The spike glycoprotein of the new coronavirus 2019-nCoV contains a furin-like cleavage site absent in CoV of the same clade. *Antiviral research*, 176, 104742.
- CROFTON, J., & MITCHISON, D. A. (1948). Streptomycin resistance in pulmonary tuberculosis. *British medical journal*, 2(4588), 1009–1015.
- Cui J, Li F, Shi ZL. (2019). Origin and evolution of pathogenic coronaviruses. *Nat Rev Microbiol* 17(3):181–192.

- Cui W, Yang K, Yang H. (2020). Recent Progress in the Drug Development Targeting SARS-CoV-2 Main Protease as Treatment for COVID-19. *Front Mol Biosci*, 4;7:616341.
- Cui, J., Li, F., & Shi, Z. L. (2019). Origin and evolution of pathogenic coronaviruses. *Nature reviews. Microbiology*, 17(3), 181–192. D1083–D1090.
- Dantu, S. C., Khavnekar, S. & Kale, A. (2016). Conformational dynamics of Peb4 exhibit “mother’s arms” chain model: a molecular dynamics study. *J Biomol Struct Dyn*, 35, 2186–2196.
- David, C. C. & Jacobs, D. J. (2014). Principal component analysis: a method for determining the essential dynamics of proteins. *Methods Mol Biol*, 1084, 193-226.
- Davies NG, Abbott S, Barnard RC, Jarvis CI, Kucharski AJ, Munday JD, Pearson CAB, Russell TW, Tully DC, Washburne AD, Wenseleers T, Gimma A, Waite W, Wong KLM, Zandvoort K van, Silverman JD, Diaz-Ordaz K, Keogh R, Eggo RM, Funk S, Jit M, Atkins KE, Edmunds WJ. (2021). Estimated transmissibility and impact of SARS-CoV-2 lineage B.1.1.7 in England. *Science* (80-) 372(6538).
- Davies, N. G., Abbott, S., Barnard, R. C., Jarvis, C. I., Kucharski, A. J., Munday, J. D., Pearson, C., Russell, T. W., Tully, D. C., Washburne, A. D., Wenseleers, T., Gimma, A., Waite, W., Wong, K., van Zandvoort, K., Silverman, J. D., CMMID COVID-19 Working Group, COVID-19 Genomics UK (COG-UK) Consortium, Diaz-Ordaz, K., Keogh, R., ... Edmunds, W. J. (2021). Estimated transmissibility and impact of SARS-CoV-2 lineage B.1.1.7 in England. *Science(NewYork,N.Y.)*, 372(6538), eabg3055.
- Decroly, E., Debarnot, C., Ferron, F., Bouvet, M., Coutard, B., Imbert, I., Gluais, L., Papageorgiou, N., Sharff, A., Bricogne, G., Ortiz-Lombardia, M., Lescar, J., & Canard, B. (2011). Crystal structure and functional analysis of the SARS-coronavirus RNA cap 2'-O-methyltransferase nsp10/nsp16 complex. *PLoS pathogens*, 7(5), e1002059.

Deganutti, G., Moro, S., & Reynolds, C. A. (2020). A Supervised Molecular Dynamics Approach to Unbiased Ligand-Protein Unbinding. *Journal of chemical information and modeling*, 60(3), 1804–1817.

Dehury, B., Raina, V., Misra, N., & Suar, M. (2021). Effect of mutation on structure, function and dynamics of receptor binding domain of human SARS-CoV-2 with host cell receptor ACE2: a molecular dynamics simulations study. *Journal of biomolecular structure & dynamics*, 39(18), 7231–7245.

Dejnirattisai W, Zhou D, Supasa P, Liu C, Mentzer AJ, Ginn HM, Zhao Y, Duyvesteyn HME, Tuekprakhon A, Nutalai R, Wang B, López-Camacho C, Slon-Campos J, Walter TS, Skelly D, Costa Clemens SA, Naveca FG, Nascimento V, Nascimento F, Fernandes da Costa C, Resende PC, Pauvolid-Correa A, Siqueira MM, Dold C, Levin R, Dong T, Pollard AJ, Knight JC, Crook D, Lambe T, Clutterbuck E, Bibi S, Flaxman A, Bittaye M, Belij-Rammerstorfer S, Gilbert SC, Carroll MW, Klennerman P, Barnes E, Dunachie SJ, Paterson NG, Williams MA, Hall DR, Hulswit RJG, Bowden TA, Fry EE, Mongkolsapaya J, Ren J, Stuart DI, Screaton GR. (2021). Antibody evasion by the P.1 strain of SARS-CoV-2. *Cell* 184(11):2939-2954.e9.

Dokholyan N V., Li L, Ding F, Shakhnovich EI. (2002). Topological determinants of protein folding. *Proc Natl Acad Sci U S A*, 99(13):8637-8641.

Duan, H., Chen, X., Boyington, J. C., Cheng, C., Zhang, Y., Jafari, A. J., et al. (2018). Glycan masking focuses immune responses to the HIV-1 CD4-binding site and enhances elicitation of VRC01-class precursor antibodies. *Immunity* 49 (2),301–311.e5.

E. J. (2012). Global assessment of genomic regions required for growth in *Mycobacterium tuberculosis*. *PLoS pathogens*, 8(9), e1002946.

Estrada E. (2020). Topological analysis of SARS CoV-2 main protease. *Chaos*, 30(6).

Fàbrega, A., Madurga, S., Giralt, E., & Vila, J. (2009). Mechanism of action of and resistance to quinolones. *Microbial biotechnology*, 2(1), 40–61.

Fahey, R. C. (2001). Novel thiols of prokaryotes. *Annu. Rev. Microbiol*, 55, 333–356.

Fahey, R. C. (2001). Novel thiols of prokaryotes. *Annual Review of Microbiology*, 55, 333–356.

Fan K, Wei P, Feng Q, et al. (2004). Biosynthesis, Purification, and Substrate Specificity of Severe Acute Respiratory Syndrome Coronavirus 3C-like Proteinase. *J Biol Chem*, 279(3):1637-1642.

Felline A, Seeber M, Fanelli F. (2020). webPSN v2.0: A webserver to infer fingerprints of structural communication in biomacromolecules. *Nucleic Acids Res* 48(W1):W94–W103.

Felline A, Seeber M, Fanelli F. (2020). webPSN v2.0: a webserver to infer fingerprints of structural communication in biomacromolecules. *Nucleic Acids Res*, 48(W1):W94-W103.

Fiorentini S, Messali S, Zani A, Caccuri F, Giovanetti M, Ciccozzi M, Caruso A. (2021). First detection of SARS-CoV-2 spike protein N501 mutation in Italy in August, 2020. *Lancet Infect Dis* 21(6):e147.

Global tuberculosis report 2018; *World Health Organization*: Geneva, (2018), pp 27-64

Gobeil SMC, Janowska K, McDowell S, Mansouri K, Parks R, Manne K, Stalls V, Kopp MF, Henderson R, Edwards RJ, Haynes BF, Acharya P. (2021). D614G Mutation Alters SARS-CoV-2 Spike Conformation and Enhances Protease Cleavage at the S1/S2 Junction. *Cell Rep* 34(2):108630.

Gómez CE, Perdiguer B, Esteban M. (2021). Emerging sars-cov-2 variants and impact in global vaccination programs against sars-cov-2/covid-19. *Vaccines* 9(3):1–13.

Gopi S, Devanshu D, Rajasekaran N, Anantakrishnan S, Naganathan AN. (2020). PPerturb: A

Server for Predicting Long-Distance Energetic Couplings and Mutation-Induced Stability Changes in Proteins via Perturbations. *ACS Omega* 5(2):1142–1146.

Griffin JWD. (2020). SARS-CoV and SARS-CoV-2 main protease residue interaction networks change when bound to inhibitor N3. *J Struct Biol*, 211(3):107575.

Gui M, Song W, Zhou H, Xu J, Chen S, Xiang Y, Wang X. (2017). Cryo-electron microscopy structures of the SARS-CoV spike glycoprotein reveal a prerequisite conformational state for receptor binding. *Cell Res* 27(1):119–129.

Guo, Y. R., Cao, Q. D., Hong, Z. S., Tan, Y. Y., Chen, S. D., Jin, H. J., Tan, K. S., Wang, D. Y., & Yan, Y. (2020). The origin, transmission and clinical therapies on coronavirus disease 2019 (COVID-19) outbreak - an update on the status. *Military Medical Research*, 7(1), 11.

Gustafsson, T. N., Osman, H., Werngren, J., Hoffner, S., Engman, L. & Holmgren, A. (2016). Ebselen and analogs as inhibitors of *Bacillus anthracis* thioredoxin reductase and bactericidal antibacterials targeting *Bacillus* species, *Staphylococcus aureus* and *Mycobacterium tuberculosis*. *Biochem. Biophys. Acta*, 1860, 1265–1271.

Halder A, Anto A, Subramanyan V, Bhattacharyya M, Vishveshwara S, Vishveshwara S. (2020). Surveying the Side-Chain Network Approach to Protein Structure and Dynamics: The SARS-CoV-2 Spike Protein as an Illustrative Case. *Front Mol Biosci* 7(December):1–15.

Harbut, M. B., Vilcheze, C., Luo, X., Hensler, M. E., Guo, H., Yang, B., Chatterjee, A. K., Nizet, V., Jacobs, W. R. Jr., Schultz, P. G. & Wang, F. (2015). Auranofin exerts broad-spectrum bactericidal activities by targeting thiol-redox homeostasis. *Proc. Natl. Acad. Sci. USA*, 112, 4453–4458.

Heald-Sargent, T., & Gallagher, T. (2012). Ready, set, fuse! The coronavirus spike protein and acquisition of fusion competence. *Viruses*, 4(4), 557–580.

Hess, B. (2008). P-LINCS: A Parallel Linear Constraint Solver for Molecular Simulation. *J. Chem. Theory Comput*, 4, 116-122.

Hess, B., Kutzner, C., van der Spoel, D. & Lindahl, E. (2008). GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theory Comput*, 4, 435-447.

Hikisz, P., Szczupak, L., Koceva-Chyla, A., Gu Spiel, A., Oehninger, L., Ott, I., Therrien, B., Solecka, J. & Kowalski, K. (2015). Anticancer and Antibacterial Activity Studies of Gold (I)-Alkynyl Chromones. *Molecules*, 20, 19699–19718.

Holmgren A. (2000). Antioxidant function of thioredoxin and glutaredoxin systems. *Antioxidants & redox signaling*, 2(4), 811–820.

Holmgren, A. (1995). Thioredoxin structure and mechanism: conformational changes on oxidation of the active-site sulfhydryls to a disulfide. *Structure*. 3: 239-243.

Hsu MF, Kuo CJ, Chang KT, et al. (2005). Mechanism of the maturation process of SARS-CoV 3CL protease. *J Biol Chem*, 280(35):31257-31266.

<https://drjessesantiano.com/5-spike-protein-blocking-drugs-under-research-for-covid-19/>

Hu T, Zhang Y, Li L, et al. (2009). Two adjacent mutations on the dimer interface of SARS coronavirus 3C-like protease cause different conformational changes in crystal structure. *Virology*. 388(2):324-334.

Hu, X., Chen, C. Z., Xu, M., Hu, Z., Guo, H., Itkin, Z., Shinn, P., Ivin, P., Leek, M., Liang, T. J., Shen, M., Zheng, W., & Hall, M. D. (2021). Discovery of Small Molecule Entry Inhibitors Targeting the Fusion Peptide of SARS-CoV-2 Spike Protein. *ACS medicinal chemistry letters*, 12(8), 1267–1274.

- Huang, C., Lokugamage, K. G., Rozovics, J. M., Narayanan, K., Semler, B. L., & Makino, S. (2011). SARS coronavirus nsp1 protein induces template-dependent endonucleolytic cleavage of mRNAs: viral mRNAs are resistant to nsp1-induced RNA cleavage. *PLoS pathogens*, 7(12), e1002433.
- Imlay, J. A. (2013). The molecular mechanism and physiological consequences of oxidative stress: lessons from model bacterium. *Nat. Rev. Micro*, 11, 443-454
- J. CCD and D. (2014). Principal Component Analysis: A Method for Determining the Essential Dynamics of Proteins. 193–226 p.
- Jacobs DJ, Rader AJ, Kuhn LA, Thorpe MF. (2001). Protein flexibility predictions using graph theory. *Proteins Struct Funct Genet* 44(2):150–165.
- Jacobs DJ, Rader AJ, Kuhn LA, Thorpe MF. (2001). Protein flexibility predictions using graph theory. *Proteins Struct Funct Genet*, 44(2):150-165.
- Jaeger T. (2007). Peroxiredoxin systems in mycobacteria. *Subcell Biochem*, 44:207-17.
- Jaeger, T., Budde, H., Flohe, L., Menge, U., Singh, M, Trujillo, M. & Radi, R. (2004). Multiple thioredoxin-mediated routes to detoxify hydroperoxides in *Mycobacterium tuberculosis*. *Arch. Biochem. Biophys*, 423,182–191.
- Jang, K. J., Jeong, S., Kang, D. Y., Sp, N., Yang, Y. M., & Kim, D. E. (2020). A high ATP concentration enhances the cooperative translocation of the SARS coronavirus helicase nsP13 in the unwinding of duplex RNA. *Scientific reports*, 10(1), 4481.
- Jin Z, Du X, Xu Y, et al. (2020). Structure of Mpro from SARS-CoV-2 and discovery of its inhibitors. *Nature*, 582(7811):289-293.
- Kandhari N, Sinha S. (2017). Complex network analysis of thermostable mutants of *Bacillus subtilis* Lipase A. *Appl Netw Sci* 2(1).

- Kannan N, Vishveshwara S. (1999). Identification of side-chain clusters in protein structures by a graph spectral method. *J Mol Biol* 292(2):441–464.
- Kannan, S. R., Spratt, A. N., Sharma, K., Chand, H. S., Byrareddy, S. N., & Singh, K. (2022). Omicron SARS-CoV-2 variant: Unique features and their impact on pre-existing antibodies. *Journal of autoimmunity*, 126, 102779. <https://doi.org/10.1016/j.jaut.2021.102779>
- Karumbi J, Garner P. (2015). Directly observed therapy for treating tuberculosis. *Cochrane Database Syst Rev*. 2015 May 29, (5):CD003343.
- Kasahara, K., Fukuda, I. & Nakamura, H. (2014). A Novel Approach of Dynamic Cross Correlation Analysis on Molecular Dynamics Simulations and Its Application to Ets1 Dimer–DNA Complex. *PLoS ONE*, 9(11): e112419.
- Ke, Z., Oton, J., Qu, K., Cortese, M., Zila, V., McKeane, L., Nakane, T., Zivanov, J., Neufeldt, C. J., Cerikan, B., Lu, J. M., Peukes, J., Xiong, X., Kräusslich, H. G., Scheres, S., Bartenschlager, R., & Briggs, J. (2020). Structures and distributions of SARS-CoV-2 spike proteins on intact virions. *Nature*, 588(7838), 498–502.
- Khan A, Gui J, Ahmad W, Haq I, Shahid M, Khan AA, Shah A, Khan A, Ali L, Anwar Z, Safdar M, Abubaker J, Uddin NN, Cao L, Wei DQ, Mohammad A. (2021). The SARS-CoV-2 B.1.618 variant slightly alters the spike RBD-ACE2 binding affinity and is an antibody escaping variant: a computational structural perspective. *RSC Adv* 11(48):30132–30147.
- Kirkman, H. N., Rolfo, M., Ferraris, A. M., & Gaetani, G. F. (1999). Mechanisms of protection of catalase by NADPH. Kinetics and stoichiometry. *The Journal of biological chemistry*, 274(20), 13908–13914.

- Koch, O., Jäger, T., Heller, K., Khandavalli, P. C., Pretzel, J., Becker, K., Flohé, L. & Selzer, P. M. (2013). Identification of *M. tuberculosis* thioredoxin reductase inhibitors based on high-throughput docking using constraints. *J Med Chem.*, 56, 4849-4859.
- Kollman, P. A., Massova, I., Reyes, C., Kuhn, B., Huo, S., Chong, L., Lee, M., Lee, T., Duan, Y., Wang, W., Donini, O., Cieplak, P., Srinivasan, J., Case, D. A., & Cheatham, T. E. 3rd. (2000). Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. *Acc. Chem Res.*, 33, 889-897.
- Konno, K., Feldmann, F. M., & McDermott, W. (1967). Pyrazinamide susceptibility and amidase activity of tubercle bacilli. *The American review of respiratory disease*, 95(3), 461–469.
- Ksiazek, T. G., Erdman, D., Goldsmith, C. S., Zaki, S. R., Peret, T., Emery, S., Tong, S., Urbani, C., Comer, J. A., Lim, W., Rollin, P. E., Dowell, S. F., Ling, A. E., Humphrey, C. D., Shieh, W. J., Guarner, J., Paddock, C. D., Rota, P., Fields, B., DeRisi, J., ... SARS Working Group (2003). A novel coronavirus associated with severe acute respiratory syndrome. *The New England journal of medicine*, 348(20), 1953–1966.
- Kumari, R., Kumar, R., Open Source Drug Discovery Consortium & Lynn, A. (2014). g_mmpbsa -a GROMACS tool for high-throughput MM-PBSA calculations. *J. Chem. Inf. Model*, 54, 1951-1962.
- Kwon, Y. S., Jeong, B. H., & Koh, W. J. (2014). Tuberculosis: clinical trials and new drug regimens. *Current opinion in pulmonary medicine*, 20(3), 280–286.
- Lamers, M.M., Haagmans, B.L. (2022). SARS-CoV-2 pathogenesis. *Nat Rev Microbiol* **20**, 270–284.

- Lan J, Ge J, Yu J, Shan S, Zhou H, Fan S, Zhang Q, Shi X, Wang Q, Zhang L, Wang X. (2020). Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *Nature*. May;581(7807):215-220.
- Lata S, Akif M. (2021). Comparative protein structure network analysis on 3CLpro from SARS-CoV-1 and SARS-CoV-2. *Proteins Struct Funct Bioinforma* 89(9):1216–1225.
- Lata, S., & Akif, M. (2022). Probing structural basis for enhanced binding of SARS-CoV-2 P.1 variant spike protein with the human ACE2 receptor. *Journal of cellular biochemistry*, 123(7), 1207–1221.
- Laurent, T. C., Moore, E. C. and Reichard, P. (1964). Enzymatic synthesis of deoxyribonucleotides. IV. Isolation and characterization of thioredoxin, the hydrogen donor from *Escherichia coli* B. *J Biol Chem*. 239: 3436-3444.
- Laurini E, Marson D, Aulic S, Fermeglia M, Pricl S. (2020). Computational Alanine Scanning and Structural Analysis of the SARS-CoV-2 Spike Protein/Angiotensin-Converting Enzyme 2 Complex. *ACS Nano* 14(9):11821–11830.
- Lei, J., Kusov, Y., & Hilgenfeld, R. (2018). Nsp3 of coronaviruses: Structures and functions of a large multi-domain protein. *Antiviral research*, 149, 58–74.
- Lennon, B. W., Williams, C. H., Jr, & Ludwig, M. L. (2000). Twists in catalysis: alternating conformations of *Escherichia coli* thioredoxin reductase. *Science (New York, N.Y.)*, 289(5482), 1190– 1194.
- Libardo, M.D.J., Boshoff, H.I.M., & Barry, C.E. (2018). The present state of the tuberculosis drug development pipeline. *Curr. Opin. Pharmacol*, 42, 81-94

- Lin, K., O'Brien, K. M., Trujillo, C., Wang, R., Wallach, J. B., Schnappinger, D. & Ehrt, S. (2016). *Mycobacterium tuberculosis* Thioredoxin Reductase Is Essential for Thiol Redox Homeostasis but Plays a Minor Role in Antioxidant Defense. *PLoS Pathog*, 12(6):e1005675
- López-Blanco JR, Aliaga JI, Quintana-Ortí ES, Chacón P. (2014). IMODS: Internal coordinates normal mode analysis server. *Nucleic Acids Res* 42(W1):271–276.
- Lu, J. & Holmgren, A. (2014). The thioredoxin antioxidant system. *Free Radic Biol Med.*, 66, 75–87
- Lu, J., & Holmgren, A. (2014). The thioredoxin antioxidant system. *Free radical biology & medicine*, 66, 75–87.
- Lu, J., Vlamis-Gardikas, A., Kandasamy, K., Zhao, R., Gustafsson, T.N., Engstrand, L., Hoffner, S., Engman, L., & Holmgren, A. (2013). Inhibition of bacterial thioredoxin reductase: an antibiotic mechanism targeting bacteria lacking glutathione. *The FASEB Journal*, 27, 1394 - 1403.
- Mahdi, M., Mótyán, J. A., Szojka, Z. I., Golda, M., Miczi, M., & Tózsér, J. (2020). Analysis of the efficacy of HIV protease inhibitors against SARS-CoV-2's main protease. *Virology journal*, 17(1), 190.
- Marsh, J. A. & Teichmann, S. A. (2011). Relative solvent accessible surface area predicts protein conformational changes upon binding. *Structure*, 19, 859-867.
- Mascellino, M. T., Di Timoteo, F., De Angelis, M., & Oliva, A. (2021). Overview of the Main Anti-SARS-CoV-2 Vaccines: Mechanism of Action, Efficacy and Safety. *Infection and drug resistance*, 14, 3459–3476.
- Maslov, D. A., Shur, K. V., Vatlin, A. A., & Danilenko, V. N. (2020). MmpS5-MmpL5 Transporters Provide *Mycobacterium smegmatis* Resistance to imidazo[1,2-b][1,2,4,5]tetrazines. *Pathogens*(Basel,Switzerland), 9(3),166.

- Miggiano R, Rizzi M, Ferraris DM. (2020). Mycobacterium tuberculosis Pathogenesis, Infection Prevention and Treatment. *Pathogens*, 18;9(5):385.
- Millet, J. K., & Whittaker, G. R. (2015). Host cell proteases: Critical determinants of coronavirus tropism and pathogenesis. *Virus research*, 202, 120–134.
- Mogashoa T, Melamu P, Derendinger B, Ley SD, Streicher EM, Iketleng T, Mupfumi L, Mokomane M, Kgwaadira B, Rankgoane-Pono G, Tsholofelo TT, Kasvosve I, Moyo S, Warren RM, Gaseitsiwe S. (2019). Detection of Second Line Drug Resistance among Drug Resistant Mycobacterium Tuberculosis Isolates in Botswana. *Pathogens*, 28;8(4):208.
- Morris, G. M., Huey, R., Lindstrom, W., Sanner, M. F., Belew, R. K., Goodsell, D. S., & Olson, A. J. (2009). AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *Journal of computational chemistry*, 30(16), 2785–2791
- Natarajan A, Beena PM, Devnikar AV, Mali S. (2020). A systemic review on tuberculosis. *Indian J Tuberc*, 67(3):295-311.
- Newton, G. L., Bewley C. A., Dwyer T. J., Horn R., Aharonowitz Y., Cohen G., Davies J., Faulkner D. J. and Fahey R. C. (1995). The structure of U17 isolated from *Streptomyces clavuligerus* and its properties as an antioxidant thiol. *Eur J Biochem*. 230: 821-825.
- Oostenbrink, C., Villa, A., Mark, A. E., & van Gunsteren, W. F. (2004). A biomolecular force field based on the free enthalpy of hydration and solvation: the GROMOS force-field parameter sets 53A5 and 53A6. *J Comput Chem.*, 25, 1656-1676.
- Ortega JT, Serrano ML, Pujol FH, Rangel HR. (2020). Unrevealing sequence and structural features of novel coronavirus using in silico approaches: The main protease as molecular target. *EXCLI J*, 19(Savarino 2005):400-409.

Owings, J. P., McNair, N. N., Mui, Y. F., Gustafsson, T. N., Holmgren, A., Contel, M., Goldberg, J. B. & Mead, J. R. (2016). Auranofin and N-heterocyclic carbene gold-analogs are potent inhibitors of the bacteria *Helicobacter pylori*. *FEMS Microbiol. Lett.*, 363(14).

Ozono S, Zhang Y, Ode H, Sano K, Tan TS, Imai K, Miyoshi K, Kishigami S, Ueno T, Iwatani Y, Suzuki T, Tokunaga K. (2021). SARS-CoV-2 D614G spike mutation increases entry efficiency with enhanced ACE2-binding affinity. *Nat Commun* 12(1).

Palomino, J. C., & Martin, A. (2013). Tuberculosis clinical trial update and the current anti-tuberculosis drug portfolio. *Current medicinal chemistry*, 20(30), 3785–3796.

Paola L Di, Hadi-Alijanvand H, Song X, Hu G, Giuliani A. (2020). The discovery of a putative allosteric site in the SARS-CoV-2 spike protein using an integrated structural/dynamic approach. *J Proteome Res* 19(11):4576–4586.

Parrinello M, Rahman A. (1981). Polymorphic transitions in single crystals: A new molecular dynamics method. *J Appl Phys* 52(12):7182–7190.

Patel, M. P. and Blanchard J. S. (1999). Expression, purification, and characterization of *Mycobacterium tuberculosis* mycothione reductase. *Biochemistry*. 38: 11827-11833

Peng, J. & Zhang, Z. (2014). Simulating Large-Scale Conformational Changes of Proteins by Accelerating Collective Motions Obtained from Principal Component Analysis. *J. Chem. Theory Comput.* 10, 3449-3458.

Periwal N, Rathod SB, Pal R, Sharma P, Nebhnani L, Barnwal RP, Arora P, Srivastava KR, Sood V. (2021). In silico characterization of mutations circulating in SARS-CoV-2 structural proteins. *J Biomol Struct Dyn* 0(0):1–16.

Piovesan, D., Minervin, G. & Tosatto, S.C.E. (2016). The RING 2.0 web server for high quality residue interaction networks. *Nucleic acids research*, 44, W367-W374.

Pronk S, Páll S, Schulz R, Larsson P, Bjelkmar P, Apostolov R, Shirts MR, Smith JC, Kasson PM, Spoel D Van Der, Hess B, Lindahl E. 2013. GROMACS 4.5: A high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* 29(7):845–854.

R. (2020). Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved α -ketoamideinhibitors. *Science (New York,N.Y.)*, 368(6489), 409–412.

Rader, A.J., Chennubhotla, C., Yang, L., & Bahar, I. (2005). The Gaussian Network Model: Theory and Applications.

Raimondi F, Felling A, Seeber M, Mariani S, Fanelli F. (2013). A mixed protein structure network and elastic network model approach to predict the structural communication in biomolecular systems: The PDZ2 domain from tyrosine phosphatase 1E as a case study. *J Chem Theory Comput*, 9(5):2504-2518.

Raugi DN, Smith RA, Gottlieb GS. (2016). Four Amino Acid Changes in HIV-2 Protease Confer Class-Wide Sensitivity to Protease Inhibitors. *J Virol*, 90(2):1062-1069.

Rawat, R., Whitty, A., & Tonge, P. J. (2003). The isoniazid-NAD adduct is a slow, tight-binding inhibitor of InhA, the Mycobacterium tuberculosis enoyl reductase: adduct affinity and drug resistance. *Proceedings of the National Academy of Sciences of the United States of America*, 100(24), 13881–13886.

Ray D, Le L, Andricioaei I. (2021). Distant residues modulate conformational opening in SARS-CoV-2 spike protein. *Proc Natl Acad Sci U S A* 118(43).

Razizadeh, M., Nikfar, M., & Liu, Y. (2021). Small molecule therapeutics to destabilize the ACE2-RBD complex: A molecular dynamics study. *Biophysical journal*, 120(14), 2793–2804.

Redfern, O. C., Dessailly, B., & Orengo, C. A. (2008). Exploring the structure and function

paradigm. *Current opinion in structural biology*, 18(3), 394–402.

Rodrigues CHM, Pires DEV, Ascher DB. (2018). DynaMut: Predicting the impact of mutations on protein conformation, flexibility and stability. *Nucleic Acids Res* 46(W1):W350–W355.

Roth, S., & Danielli, A. (2021). Rapid and Sensitive Inhibitor Screening Using Magnetically Modulated Biosensors. *Sensors (Basel, Switzerland)*, 21(14), 4814.

Sabino EC, Buss LF, Carvalho MPS, Prete CA, Crispim MAE, Fraiji NA, Pereira RHM, Parag K V., Silva Peixoto P da, Kraemer MUG, Oikawa MK, Salomon T, Cucunuba ZM, Castro MC, Souza Santos AA de, Nascimento VH, Pereira HS, Ferguson NM, Pybus OG, Kucharski A, Busch MP, Dye C, Faria NR. (2021). Resurgence of COVID-19 in Manaus, Brazil, despite high seroprevalence. *Lancet* 397(10273):452–455.

Sabino, E. C., Buss, L. F., Carvalho, M., Prete, C. A., Jr, Crispim, M., Fraiji, N. A., Pereira, R., Parag, K. V., da Silva Peixoto, P., Kraemer, M., Oikawa, M. K., Salomon, T., Cucunuba, Z. M., Castro, M. C., de Souza Santos, A. A., Nascimento, V. H., Pereira, H. S., Ferguson, N. M., Pybus, O. G., Kucharski, A., ... Faria, N. R. (2021). Resurgence of COVID-19 in Manaus, Brazil, despite high seroprevalence. *Lancet (London, England)*, 397(10273), 452–455.

Sakai, Y., Kawachi, K., Terada, Y., Omori, H., Matsuura, Y., & Kamitani, W. (2017). Two-amino acids change in the nsp4 of SARS coronavirus abolishes viral replication. *Virology*, 510, 165–174.

Savarino A. (2005). Expanding the frontiers of existing antiviral drugs: possible effects of HIV-1 protease inhibitors against SARS and avian influenza. *Journal of clinical virology: the official publication of the Pan American Society for Clinical Virology*, 34(3), 170–178.

Scharf, C., Riethdorf, S., Ernst, H., Engelmann, S., Völker, U., & Hecker, M. (1998). Thioredoxin is an essential protein induced by multiple stresses in *Bacillus subtilis*. *Journal of bacteriology*,

180(7), 1869–1877.

Schechter, I., & Berger, A. (1967). On the size of the active site in proteases. I. Papain. *Biochemical and biophysical research communications*, 27(2), 157–162.

Schmid, S., & Hugel, T. (2020). Controlling protein function by fine-tuning conformational flexibility. *eLife*, 9, e57180.

Schüttelkopf, A. W. & van Aalten, D. M. F. (2004). PRODRG: a tool for high-throughput crystallography of protein-ligand complexes, *Acta Crystallogr, D60*, 1355–1363.

Seeber M, Feline A, Raimondi F, Mariani S, Fanelli F. WebPSN: (2015). A web server for high-throughput investigation of structural communication in biomacromolecules. *Bioinformatics*, 31(5):779-781.

Shan, J., Pan, X., Wang, X., Xiao, X., & Ji, C. (2020). FragRep: A Web Server for Structure-Based Drug Design by Fragment Replacement. *Journal of chemical information and modeling*, 60(12), 5900–5906.

Shereen, M. A., Khan, S., Kazmi, A., Bashir, N., & Siddique, R. (2020). COVID-19 infection: Origin, transmission, and characteristics of human coronaviruses. *Journal of advanced research*, 24, 91–98.

Shi J, Wei Z, Song J. (2004). Dissection study on the severe acute respiratory syndrome 3C-like protease reveals the critical role of the extra domain in dimerization of the enzyme. Defining the extra domain as a new target for design of highly specific protease inhibitors. *J Biol Chem*, 279(23):24765-24773.

Shi, Z., Gao, H., Bai, X. C., & Yu, H. (2020). Cryo-EM structure of the human cohesin-NIPBL-DNA complex. *Science (New York, N.Y.)*, 368(6498), 1454–1459.

Sisk, J. M., Frieman, M. B., & Machamer, C. E. (2018). Coronavirus S protein-induced fusion is

blocked prior to hemifusion by Abl kinase inhibitors. *The Journal of general virology*, 99(5), 619–630.

Song, W., Gui, M., Wang, X., & Xiang, Y. (2018). Cryo-EM structure of the SARS coronavirus spike glycoprotein in complex with its host cell receptor ACE2. *PLoS pathogens*, 14(8), e1007236.

Srivastava A, Birari V, Sinha S. (2020). Small Conformational Changes Underlie Evolution of Resistance to NNRTI in HIV Reverse Transcriptase. *Biophys J* 118(10):2489–2501.

Srivastava A, Sinha S. (2014). Thermostability of in vitro evolved *Bacillus subtilis* Lipase a: A network and dynamics perspective. *PLoS One* 9(8).

Srivastava A, Sinha S. (2017). Uncoupling of an ammonia channel as a mechanism of allosteric inhibition in anthranilate synthase of *Serratia marcescens*: dynamic and graph theoretical analysis. *Mol Biosyst* 13(1):142–155.

Starr TN, Greaney AJ, Hilton SK, Ellis D, Crawford KHD, Dingens AS, Navarro MJ, Bowen JE, Tortorici MA, Walls AC, King NP, Veasler D, Bloom JD. (2020). Deep Mutational Scanning of SARS-CoV-2 Receptor Binding Domain Reveals Constraints on Folding and ACE2 Binding. *Cell* 182(5):1295-1310.e20.

Subissi, L., Posthuma, C. C., Collet, A., Zevenhoven-Dobbe, J. C., Gorbalenya, A. E., Decroly, E., Snijder, E. J., Canard, B., & Imbert, I. (2014). One severe acute respiratory syndrome coronavirus protein complex integrates processive RNA polymerase and exonuclease activities. *Proceedings of the National Academy of Sciences of the United States of America*, 111(37), E3900–E3909.

Supasa, P., Zhou, D., Dejnirattisai, W., Liu, C., Mentzer, A. J., Ginn, H. M., Zhao, Y., Duyvesteyn, H., Nutalai, R., Tuekprakhon, A., Wang, B., Paesen, G. C., Slon-Campos, J., López-

Camacho, C., Hallis, B., Coombes, N., Bewley, K. R., Charlton, S., Walter, T. S., Barnes, E., ... Screaton, G. R. (2021). Reduced neutralization of SARS-CoV-2 B.1.1.7 variant by convalescent and vaccine sera. *Cell*, 184(8), 2201–2211.e7.

Takayama, K., & Kilburn, J. O. (1989). Inhibition of synthesis of arabinogalactan by ethambutol in *Mycobacterium smegmatis*. *Antimicrobial agents and chemotherapy*, 33(9), 1493–1499.

Tang QY, Kaneko K. Long-range correlation in protein dynamics: Confirmation by structural data and normal mode analysis. *PLoS Comput Biol*. 2020;16(2):1-17.

Taylor, P. C., Adams, A. C., Hufford, M. M., de la Torre, I., Winthrop, K., & Gottlieb, R. L. (2021). Neutralizing monoclonal antibodies for treatment of COVID-19. *Nature reviews. Immunology*, 21(6), 382–393.

te Velhuis, A. J., van den Worm, S. H., & Snijder, E. J. (2012). The SARS-coronavirus nsp7+nsp8 complex is a unique multimeric RNA polymerase capable of both de novo initiation and primer extension. *Nucleic acids research*, 40(4), 1737–1747.

Tegally, H., Wilkinson, E., Giovanetti, M., Iranzadeh, A., Fonseca, V., Giandhari, J., Doolabh, D., Pillay, S., San, E. J., Msomi, N., Mlisana, K., von Gottberg, A., Walaza, S., Allam, M., Ismail, A., Mohale, T., Glass, A. J., Engelbrecht, S., Van Zyl, G., Preiser, W., ... de Oliveira, T. (2021). Detection of a SARS-CoV-2 variant of concern in South Africa. *Nature*, 592(7854), 438–443.

Telenti, A., Imboden, P., Marchesi, F., Lowrie, D., Cole, S., Colston, M. J., Matter, L., Schopfer, K., & Bodmer, T. (1993). Detection of rifampicin-resistance mutations in *Mycobacterium tuberculosis*. *Lancet (London, England)*, 341(8846), 647–650.

Teruel N, Mailhot O, Najmanovich RJ. (2021). Modelling conformational state dynamics and its role on infection for SARS-CoV-2 Spike protein variants. *PLoS Comput Biol* 17(8):1–25.

Tian, W., Chen, C., Lei, X., Zhao, J. & Liang, J. (2018). CASTp 3.0: computed atlas of surface topography of proteins. *Nucleic Acids Res*, 46, W363-W367.

Trivedi, A., Singh, N., Bhat, S. A., Gupta, P., & Kumar, A. (2012). Redox biology of tuberculosis pathogenesis. *Advances in microbial physiology*, 60, 263–324.

Tsai, T. Y., Chang, K. W. & Chen, C. Y. C. (2011). iScreen: World's first cloud computing web server for virtual screening and de novo drug design based on TCM database@Taiwan. *Journal of Computer-Aided Molecular Design*, 25, 525-531.

Ullrich, S., & Nitsche, C. (2020). The SARS-CoV-2 main protease as drug target. *Bioorganic & medicinal chemistry letters*, 30(17), 127377.

Unni, S., Aouti, S., Thiagarajan, S., & Padmanabhan, B. (2020). Identification of a repurposed drug as an inhibitor of Spike protein of human coronavirus SARS-CoV-2 by computational methods. *Journal of biosciences*, 45(1), 130.

Uziel O, Borovok I, Schreiber R, Cohen G, Aharonowitz Y. (2004). Transcriptional regulation of the *Staphylococcus aureus* thioredoxin and thioredoxin reductase genes in response to oxygen and disulfide stress. *J Bacteriol*, 186(2):326-34.

Vehlow, C., Stehr, H., Winkelmann, M., Duarte, J.M., Petzold, L., Dinse, J., & Lappe, M. (2011). CMView: Interactive contact map visualization and analysis. *Bioinformatics*, 27, 1573–1574

Vendruscolo M, Dokholyan N V., Paci E, Karplus M. (2002). Small-world view of the amino acids that play a key role in protein folding. *Phys Rev E Stat Nonlin Soft Matter Phys*, 65(6):4.

Verkhivker G. (2020a). Coevolution, Dynamics and Allostery Conspire in Shaping Cooperative Binding and Signal Transmission of the SARS-CoV-2 Spike Protein with Human Angiotensin-Converting Enzyme 2. 1–31 p.

Verkhivker GM, Agajanian S, Oztas DY, Gupta G. (2021). Comparative Perturbation-Based

Modeling of the SARS-CoV-2 Spike Protein Binding with Host Receptor and Neutralizing Antibodies: Structurally Adaptable Allosteric Communication Hotspots Define Spike Sites Targeted by Global Circulating Mutations. *Biochemistry* 60(19):1459–1484.

Verkhivker GM, Paola L Di. (2021). Dynamic Network Modeling of Allosteric Interactions and Communication Pathways in the SARS-CoV-2 Spike Trimer Mutants: Differential Modulation of Conformational Landscapes and Signal Transmission via Cascades of Regulatory Switches. *J Phys Chem B* 125(3):850–873.

Verkhivker GM. (2020b). Molecular Simulations and Network Modeling Reveal an Allosteric Signaling in the SARS-CoV-2 Spike Proteins. *J Proteome Res* 19(11):4587–4608.

Verma, S., Grover, S., Tyagi, C., Goyal, S., Jamal, S, Singh, A. & Grover, A. (2016). Hydrophobic interactions are key to MDM2 inhibition by polyphenols as revealed by molecular dynamics simulations and MM/PBSA free energy calculations. *PLoS One*, 11 (2): e0149014.

Von Groll, A., Martin, A., Jureen, P., Hoffner, S., Vandamme, P., Portaels, F., Palomino, J. C., & da Silva, P. A. (2009). Fluoroquinolone resistance in *Mycobacterium tuberculosis* and mutations in *gyrA* and *gyrB*. *Antimicrobial agents and chemotherapy*, 53(10), 4498–4500.

Vriend, G. (1990). WHAT IF: a molecular modeling and drug design program. *J Mol Graph*, 8, 52-56.

Waksman, G., Krishna, T. S., Williams, C. H. Jr. & Kuriyan, J. (1994). Crystal structure of *Escherichia coli* thioredoxin reductase refined at 2Å resolution- Implications for a large conformational change during catalysis. *J Mol Biol*, 236, 800-816.

Wallace, A. C., Laskowski, R. A. & Thornton, J. M. (1996). LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. *Protein Eng.*, 8, 127-134.

Walls, A. C., Park, Y. J., Tortorici, M. A., Wall, A., McGuire, A. T., & Veerler, D. (2020). Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. *Cell*, 183(6), 1735.

Wang Q, Zhang Y, Wu L, Niu S, Song C, Zhang Z, Lu G, Qiao C, Hu Y, Yuen KY, Wang Q, Zhou H, Yan J, Qi J. (2020a). Structural and Functional Basis of SARS-CoV-2 Entry by Using Human ACE2. *Cell* 181(4):894-904.e9.

Wang Y, Liu M, Gao J. (2020b). Enhanced receptor binding of SARS-CoV-2 through networks of hydrogen-bonding and hydrophobic interactions. *Proc Natl Acad Sci U S A* 117(25):13967–13974.

Wang, Q., Zhang, Y., Wu, L., Niu, S., Song, C., Zhang, Z., Lu, G., Qiao, C., Hu, Y., Yuen, K. Y., Wang, Q., Zhou, H., Yan, J., & Qi, J. (2020). Structural and Functional Basis of SARS-CoV-2 Entry by Using Human ACE2. *Cell*, 181(4), 894–904.e9.

Wangkar PP, Tendulkar A V., Ramya S, Mali DN, Sarawagi S. (2003). Functional sites in protein families uncovered via an objective and automated graph theoretic approach. *J Mol Biol*, 326(3):955-978.

Wibmer CK, Ayres F, Hermanus T, Madzivhandila M, Kgagudi P, Oosthuysen B, Lambson BE, Oliveira T de, Vermeulen M, Berg K van der, Rossouw T, Boswell M, Ueckermann V, Meiring S, Gottberg A von, Cohen C, Morris L, Bhiman JN, Moore PL. (2021). SARS-CoV-2 501Y.V2 escapes neutralization by South African COVID-19 donor plasma. *Nat Med* 27(4):622–625.

Wrapp, D., Wang, N., Corbett, K. S., Goldsmith, J. A., Hsieh, C. L., Abiona, O., Graham, B. S., & McLellan, J. S. (2020). Cryo-EM Structure of the 2019-nCoV Spike in the Prefusion Conformation. *bioRxiv : the preprint server for biology*, 2020.02.11.944462.

Wu CY, Jan JT, Ma SH, et al. (2004). Small molecules targeting severe acute respiratory

syndrome human coronavirus. *Proc Natl Acad Sci U S A*. 101(27):10012-10017.

Wu F, Zhao S, Yu B, Chen YM, Wang W, Song ZG, Hu Y, Tao ZW, Tian JH, Pei YY, Yuan ML, Zhang YL, Dai FH, Liu Y, Wang QM, Zheng JJ, Xu L, Holmes EC, Zhang YZ. (2020). A new coronavirus associated with human respiratory disease in China. *Nature* 579(7798):265–269.

Wu F, Zhao S, Yu B, et al. (2020). A new coronavirus associated with human respiratory disease in China. *Nature*, 579(7798):265-269.

Wu, F., Zhao, S., Yu, B., Chen, Y. M., Wang, W., Song, Z. G., Hu, Y., Tao, Z. W., Tian, J. H., Pei, Y. Y., Yuan, M. L., Zhang, Y. L., Dai, F. H., Liu, Y., Wang, Q. M., Zheng, J. J., Xu, L., Holmes, E. C., & Zhang, Y. Z. (2020). A new coronavirus associated with human respiratory disease in China. *Nature*, 579(7798), 265–269.

Xie X, Liu Y, Liu J, Zhang X, Zou J, Fontes-Garfias CR, Xia H, Swanson KA, Cutler M, Cooper D, Menachery VD, Weaver SC, Dormitzer PR, Shi PY. (2021). Neutralization of SARS-CoV-2 spike 69/70 deletion, E484K and N501Y variants by BNT162b2 vaccine-elicited sera. *Nat Med* 27(4):620–621.

Xue X, Yang H, Shen W, et al. (2007). Production of Authentic SARS-CoV Mpro with Enhanced Activity: Application as a Novel Tag-cleavage Endopeptidase for Protein Overproduction. *J Mol Biol*. 366(3):965-975.

Yadav, R., Chaudhary, J. K., Jain, N., Chaudhary, P. K., Khanra, S., Dhamija, P., Sharma, A., Kumar, A., & Handu, S. (2021). Role of Structural and Non-Structural Proteins and Therapeutic Targets of SARS-CoV-2 for COVID-19. *Cells*, 10(4), 821.

Yang H, Yang M, Ding Y, et al. (2003). The crystal structures of severe acute respiratory syndrome virus main protease and its complex with an inhibitor. *Proc Natl Acad Sci U S A*. 100(23):13190-13195.

Yuan, Y., Cao, D., Zhang, Y., Ma, J., Qi, J., Wang, Q., Lu, G., Wu, Y., Yan, J., Shi, Y., Zhang, X., & Gao, G. F. (2017). Cryo-EM structures of MERS-CoV and SARS-CoV spike glycoproteins reveal the dynamic receptor binding domains. *Nature communications*, 8, 15092.

Yuan Y, Cao D, Zhang Y, Ma J, Qi J, Wang Q, Lu G, Wu Y, Yan J, Shi Y, Zhang X, Gao GF. (2017). Cryo- EM structures of MERS-CoV and SARS-CoV spike glycoproteins reveal the dynamic receptor binding domains. *Nat Commun*, 10;8:15092.

Zaki AM, Van Boheemen S, Bestebroer TM, Osterhaus ADME, Fouchier RAM. (2012). Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia. *N Engl J Med*. 367(19):1814-1820.

Zhang L, Lin D, Sun X, et al. (2020). Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved α -ketoamide inhibitors. *Science* (80-), 368(6489):409-412.

Zhang, H., Penninger, J. M., Li, Y., Zhong, N., & Slutsky, A. S. (2020). Angiotensin-converting enzyme 2 (ACE2) as a SARS-CoV-2 receptor: molecular mechanisms and potential therapeutic target. *Intensive care medicine*, 46(4), 586–590.

Zhang, L., Lin, D., Sun, X., Curth, U., Drosten, C., Sauerhering, L., Becker, S., Rox, K., & Hilgenfeld, Zhang, Y. J., Ioerger, T. R., Huttenhower, C., Long, J. E., Sassetti, C. M., Sacchettini, J. C. & Rubin, E. J. (2012). Global assessment of genomic regions required for growth in *Mycobacterium tuberculosis*. *PLoS Pathog*, 8(9):

Zhang, Y., Heym, B., Allen, B., Young, D., & Cole, S. (1992). The catalase-peroxidase gene and isoniazid resistance of *Mycobacterium tuberculosis*. *Nature*, 358(6387), 591–593.

Zhang, Z.H., Hillas, P.J., & Ortiz de Montellano, P.R. (1999). Reduction of peroxides and dinitrobenzenes by *Mycobacterium tuberculosis* thioredoxin and thioredoxin reductase. *Archives of biochemistry and biophysics*, 363 1, 19-26 .

Zhao, X., Chen, H., & Wang, H. (2021). Glycans of SARS-CoV-2 Spike Protein in Virus Infection and Antibody Production. *Frontiers in molecular biosciences*, 8, 629873.

Zhong, N. S., Zheng, B. J., Li, Y. M., Poon, Xie, Z. H., Chan, K. H., Li, P. H., Tan, S. Y., Chang, Q., Xie, J. P., Liu, X. Q., Xu, J., Li, D. X., Yuen, K. Y., Peiris, & Guan, Y. (2003). Epidemiology and cause of severe acute respiratory syndrome (SARS) in Guangdong, People's Republic of China, in February, 2003. *Lancet (London, England)*, 362(9393), 1353–1358.

Zhou P, Yang X Lou, Wang XG, Hu B, Zhang L, Zhang W, Si HR, Zhu Y, Li B, Huang CL, Chen HD, Chen J, Luo Y, Guo H, Jiang R Di, Liu MQ, Chen Y, Shen XR, Wang X, Zheng XS, Zhao K, Chen QJ, Deng F, Liu LL, Yan B, Zhan FX, Wang YY, Xiao GF, Shi ZL. 2020. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* 579(7798):270–273.

Zumla A, Chan JFW, Azhar EI, Hui DSC, Yuen KY. (2016). Coronaviruses-drug discovery and therapeutic options. *Nat Rev Drug Discov*.15(5):327-347.

PUBLICATIONS

Probing structural basis for enhanced binding of SARS-CoV-2 P.1 variant spike protein with the human ACE2 receptor

Surabhi Lata | Mohd. Akif

Laboratory of Structural Biology,
Department of Biochemistry, School of
Life Sciences, University of Hyderabad,
Hyderabad, Telangana, India

Correspondence

Mohd. Akif, Laboratory of Structural
Biology, Department of Biochemistry,
School of Life Sciences, University of
Hyderabad, Hyderabad, Telangana, India.
Email: akif@uohyd.ac.in

Funding information

IoE, University of Hyderabad, MHRD,
India, Grant/Award Number: UoH/IoE/
RC1/RC1-2-017

Abstract

The initial step of infection by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) involves the binding of receptor binding domain (RBD) of the spike protein to the angiotensin converting enzyme 2 (ACE2) receptor. Each successive wave of SARS-CoV-2 reports emergence of many new variants, which is associated with mutations in the RBD as well as other parts of the spike protein. These mutations are reported to have enhanced affinity towards the ACE2 receptor as well as are also crucial for the virus transmission. Many computational and experimental studies have demonstrated the effect of individual mutation on the RBD-ACE2 binding. However, the cumulative effect of mutations on the RBD and away from the RBD was not investigated in detail. We report here a comparative analysis on the structural communication and dynamics of the RBD and truncated S1 domain of spike protein in complex with the ACE2 receptor from SARS-CoV-2 wild type and its P.1 variant. Our integrative network and dynamics approaches highlighted a subtle conformational changes in the RBD as well as truncated S1 domain of spike protein at the protein contact level, responsible for the increased affinity with the ACE2 receptor. Moreover, our study also identified the commonalities and differences in the dynamics of the interactions between spike protein of SARS-CoV-2 wild type and its P.1 variant with the ACE2 receptor. Further, our investigation yielded an understanding towards identification of the unique RBD residues crucial for the interaction with the ACE2 host receptor. Overall, the study provides an insight for designing better therapeutics against the circulating P.1 variants as well as other future variants.

KEYWORDS

ACE2 receptor, COVID-19, molecular dynamic simulation, protein network analysis, SARS-CoV-2, spike protein, variant

RESEARCH ARTICLE

PROTEINS WILEY

Comparative protein structure network analysis on 3CL^{pro} from SARS-CoV-1 and SARS-CoV-2

Surabhi Lata | Mohd. Akif

Department of Biochemistry, School of Life Sciences, University of Hyderabad, Hyderabad, India

Correspondence

Mohd. Akif, Department of Biochemistry, School of Life Sciences, University of Hyderabad, Prof. CR Rao Road, Gachibowli, Hyderabad, 500046 India.
Email: akif@uohyd.ac.in

Funding information

SERB; University of Hyderabad

Abstract

The main protease M^{pro}, 3CL^{pro} is an important target from coronaviruses. In spite of having 96% sequence identity among M^{pro}s from SARS-CoV-1 and SARS-CoV-2; the inhibitors used to block the activity of SARS-CoV-1 M^{pro} so far, were found to have differential inhibitory effect on M^{pro} of SARS-CoV-2. The possible reason could be due to the difference of few amino acids among the peptidases. Since, overall 3-D crystallographic structure of M^{pro} from SARS-CoV-1 and SARS-CoV-2 is quite similar and mapping a subtle structural variation is seemingly impossible. Hence, we have attempted to study a structural comparison of SARS-CoV-1 and SARS-CoV-2 M^{pro} in apo and inhibitor bound states using protein structure network (PSN) based approach at contacts level. The comparative PSNs analysis of apo M^{pro}s from SARS-CoV-1 and SARS-CoV-2 uncovers small but significant local changes occurring near the active site region and distributed throughout the structure. Additionally, we have shown how inhibitor binding perturbs the PSG and the communication pathways in M^{pro}s. Moreover, we have also investigated the network connectivity on the quaternary structure of M^{pro} and identified critical residue pairs for complex formation using three centrality measurement parameters along with the modularity analysis. Taken together, these results on the comparative PSN provide an insight into conformational changes that may be used as an additional guidance towards specific drug development.

KEYWORDS

Covid19, main protease (M^{pro}, 3CL^{pro}), protein structure graph, protein structure network, SARS coronavirus

1 | INTRODUCTION

Coronaviridae family of virus usually possesses enveloped, positive sense RNA virus that generally includes three highly pathogenic viruses such as Severe Acute Respiratory Syndrome Coronavirus 1 (SARS-CoV-1), Middle East Respiratory Syndrome Coronavirus (MERS-CoV) and SARS-CoV-2.¹ SARS-CoV-1 originated in China and caused a global pandemic in 2003 with about a 10% fatality rate.^{2,3} MERS-CoV was first reported from Saudi Arabia in 2012 and has infected the human population with limited human-to-human transmission.⁴ SARS-CoV-2, a new coronavirus reported for the first time

from Wuhan, China in December 2019, causes severe human respiratory disease.⁵ It has also been characterized as a very contagious pathogenic virus with rapid transmission capability among human-to-human that has caused an outbreak of the severe pulmonary diseases in almost 216 countries, resulting approximately 814 438 confirmed deaths globally till date (WHO report, 2020). WHO has coined SARS-CoV-2 causing disease as COVID-19 pandemic and that has now become a global health emergency and expected to have severe ramification on the global economy.

Currently, there is no specific treatment available to control COVID-19 pandemic. Efforts are being made towards design of



Structure-based identification of natural compound inhibitor against *M. tuberculosis* thioredoxin reductase: insight from molecular docking and dynamics simulation

Surabhi Lata and Mohd. Akif 

Laboratory of Structural Biology, Department of Biochemistry, School of Life Sciences, University of Hyderabad, Gachibowli, Hyderabad, Telangana, India

Communicated by Ramaswamy H. Sarma

ABSTRACT

Antioxidant systems of *M. tuberculosis* (*Mtb*) play an important role in providing resistance in the hostile environment of mononuclear phagocytes. Thioredoxin system is a known antioxidant system that consists of three copies of thioredoxins (Trxs) and a single copy of thioredoxin reductase (TrxR). TrxR has been validated as an essential gene known to be involved in the reduction of peroxides, dinitrobenzenes and hydroperoxides, and is crucial in maintaining the survival of *Mtb* in macrophages. Recently, it has been demonstrated to be a druggable target. In this study, molecular docking was applied to screen more than 20,000 natural compounds from the Traditional Chinese Medicine database. Theoretical calculation of $\Delta G_{\text{binding}}$ by the Molecular Mechanics Poisson-Boltzmann Surface Area (MM-PBSA) methods indicated two top-hit compounds that bind with a high affinity to the allosteric site, consisting of a hinge region, of TrxR. Further, stability and binding analysis of both compounds were carried out with molecular dynamics simulation. An analysis of conformational variation by principal component analysis (PCA) and protein contact network (PCN) uncovered the conformational changes in the compound-bound forms of protein. The NADPH domain formed many new interactions with the FAD domain in the compound-bound form, signifying that the binding may render an effect on the protein structure and function. Our results suggest that these two compounds could potentially be used for structure-based lead inhibitors against TrxR. The inhibitor selected as lead compound will be used further as a scaffold to optimize as novel anti-tuberculosis therapeutic.

Abbreviations: *Mtb*: *Mycobacterium tuberculosis*; MDR: Multiple Drug Resistance; XDR: Extensively Drug Resistance; Trxs: Thioredoxins; TrxR: Thioredoxin reductase; MD: Molecular dynamics; MM-PBSA: Molecular mechanics Poisson-Boltzmann Surface Area; PCN: Protein Contact Network; TCM: Traditional Chinese Medicine; ROS: Reactive oxygen species; CADD: Computer Aided Drug Discovery; PCA: Principal Component Analysis

ARTICLE HISTORY

Received 5 February 2020
Accepted 30 May 2020

KEYWORDS

Antioxidant; *Mtb* TrxR;
Druggable target; Virtual
Screening; Traditional
Chinese Medicine Database;
Molecular
Dynamics Simulation

1. Introduction

Mycobacterium tuberculosis (*Mtb*), the causative agent of tuberculosis that accounts for 1.3 million deaths worldwide every year (World Health Organization [WHO], 2018). Although many novel therapeutics have been developed, still the evolving resistance strains such as MDR (multi-drug-resistant) and XDR (extensively-drug-resistant) pose a great challenge to counter TB globally (Libardo et al., 2018). Moreover, the current first line of drugs offers a long treatment regimen. Hence, a search for new drugs or drug-like molecules that can augment the effectiveness of the current treatment and reduce the emerging resistance is warranted.

Oxidative stress and accumulation of reactive oxygen species (ROS) provide a hostile environment in the cell, which is an inevitable challenge to invading pathogens (Imlay, 2013).

Pathogens such as *Mtb* resides in the host phagocytes and copes with the hostile environment by expressing a number of antioxidant systems to ensure their survival inside its host (Trivedi et al., 2012). The thioredoxin and glutathione systems are two well-known antioxidant systems that provide reducing environments and regulate many important cellular processes such as antioxidant pathways, DNA and protein repair enzymes, and the activation of redox dependent transcription factors (Fahey, 2001; Lu & Holmgren, 2014). Hence, antioxidant systems have been considered as a potential drug target.

Mtb thioredoxin system consists of two typical thioredoxins (TrxB and TrxC) and a single copy of thioredoxin reductase (TrxR) (Akif, Khare, Tayagi, Mande, & Sardesai, 2008). TrxR reduces the Trxs by typically utilizing the reducing potential from the cellular NADPH. The reduced Trxs function in reducing the peroxides and dinitrobenzenes, and also play

CONTACT Mohd Akif  akimohd@gmail.com  Laboratory of Structural Biology, Department of Biochemistry, School of Life Sciences, University of Hyderabad, Hyderabad, Telangana 500 046, India.

 Supplemental data for this article can be accessed online at <https://doi.org/10.1080/07391102.2020.1778530>.

© 2020 Informa UK Limited, trading as Taylor & Francis Group



Research paper

Immunoinformatics analysis of antigenic epitopes and designing of a multi-epitope peptide vaccine from putative nitro-reductases of *Mycobacterium tuberculosis* DosR

Mohd. Shiraz, Surabhi Lata, Pankaj Kumar, Umate Nachiket Shankar, Mohd. Akif^{*}

Department of Biochemistry, School of Life Sciences, University of Hyderabad, Hyderabad 500046, India

ARTICLE INFO

Keywords:
Mycobacterium tuberculosis
 Nitro-reductases
 Antigenic T-cell epitopes
 Peptide vaccine
 Dormancy
 DosR

ABSTRACT

Mycobacterium tuberculosis (*Mtb*) resides in alveolar macrophages as a non-dividing and dormant state causing latent tuberculosis. Currently, no vaccine is available against the latent tuberculosis. Latent *Mtb* expresses ~48 genes under the control of DosR regulon. Among these, putative nitroreductases have significantly high expression levels, help *Mtb* to cope up with nitrogen stresses and possess antigenic properties. In the current study, immunoinformatics methodologies are applied to predict promiscuous antigenic T-cell epitopes from putative nitro-reductases of the DosR regulon. The promiscuous antigenic T-cell epitopes prediction was performed on the basis of their potential to induce an immune response and forming a stable interaction with the HLA alleles. The highest antigenic promiscuous epitopes were assembled for designing an *in silico* vaccine construct. A TLR-2 agonist Phenol-soluble modulin alpha 4 was exploited as an adjuvant. Molecular docking and Molecular Dynamics Simulations were used to predict the stability of vaccine construct with the immune receptor. The predicted promiscuous epitopes may be helpful in the construction of a subunit vaccine against latent tuberculosis, which can also be administered along with the BCG to increase its efficacy. Experimental validation is a prerequisite for the *in silico* designed vaccine construct against TB infection.

1. Introduction

Mycobacterium tuberculosis (*Mtb*) is a multifaceted pathogen causing Tuberculosis (TB), which remains one of the prominent reasons for mortality worldwide. According to the WHO global TB 2020 report, millions of people are currently infected with *Mtb*. Among that less than 10% of the total infected individuals are affected with active TB; and while more than 90% of individuals are afflicted with latent tuberculosis where *Mtb* resides inside infected macrophages for a longer period of time in an inactive metabolic state and non-transmissible form but reversible state, known as "dormancy". The population of latent tuberculosis infection (LTBI) individuals indeed responsible for a major obstruction to TB control strategies. Although antibiotic treatment offers standard care for active TB, its effectiveness against latent TB is doubtful. The current protective Bacillus-Calmette-Guerin (BCG) TB vaccine – a live-attenuated *Mycobacterium bovis* vaccine (Eddine and Kaufmann, 2005) – is known to protect against dreadful forms of TB in young children (Corbett et al., 2003). However, it does not show efficient and consistent protection in adults against pulmonary TB and also

it doesn't protect from reactivation of the latent TB infection. Hence, significant efforts for the development of new vaccines/therapies that can prevent latent *Mtb* infection are desperately needed. Dormancy in *Mtb* is characterized as a state of low pH, nutrient deprivation, hypoxia, and nitric oxide, which triggers an upregulation of a set of ~46 genes (~1.2% of the *Mtb* genome), known as Dormancy Survival Regulon or DosR (Voskuil et al., 2003). Functions of most DosR regulon gene products are unknown, but many of them are found to be immunodominant that elicit a strong T-cell response. It has been reported that DosR regulon triggers a T-cell response and IFN- γ inducing capability. Rv1733c, Rv2029c, Rv2627c, and Rv2628 were found to elicit a strong immune response as compared to CFP-10, a well-recognized antigen for *Mtb* infection (Leyten et al., 2006). Rv1813c, Rv2628, Rv2029c, and Rv2659c were also reported as latency antigens with a good humoral immune response and produce a higher number of CD4⁺ cells (Liang et al., 2019). Therapeutic effects of these antigens were studied in the endogenous resurgence mouse TB model. In fact, in the same study, Rv2626c and Rv2032 latency antigens were reported to induce a significant effect on CD4⁺ and CD8⁺ cells (Liang et al., 2019). These two

^{*} Corresponding author.

E-mail address: akif@unhyr.ac.in (Mohd. Akif).

<https://doi.org/10.1016/j.mgeid.2021.105017>

Received 10 April 2021; Received in revised form 13 June 2021; Accepted 22 July 2021

Available online 28 July 2021

1567-1348/© 2021 Elsevier B.V. All rights reserved.



Immunoinformatics-Based Designing of a Multi-Epitope Chimeric Vaccine From Multi-Domain Outer Surface Antigens of *Leptospira*

Pankaj Kumar[†], Surabhi Lata[†], Umate Nachiket Shankar and Mohd. Akif^{*}

Laboratory of Structural Biology, Department of Biochemistry, School of Life Sciences, University of Hyderabad, Hyderabad, India

OPEN ACCESS

Edited by:

Jochen Matzner,
University of Erlangen—Nuremberg,
Germany

Reviewed by:

Abdelrahman Hamza Abdelmoneim,
Al-Nasrah University, Sudan
Alok Jain,
Birla Institute of Technology—Mesra,
India

*Correspondence:

Mohd. Akif
akif@uohyd.ac.in

[†]These authors have contributed
equally to this work

Specialty section:

This article was submitted to
Vaccines and Molecular Therapeutics,
a section of the journal
Frontiers in Immunology

Received: 02 July 2021

Accepted: 08 November 2021

Published: 30 November 2021

Citation:

Kumar P, Lata S, Shankar UN and
Akif M (2021) Immunoinformatics-
Based Designing of a Multi-Epitope
Chimeric Vaccine From Multi-Domain
Outer Surface Antigens of *Leptospira*.
Front. Immunol. 12:735373.
doi: 10.3389/fimmu.2021.735373

Accurate information on antigenic epitopes within a multi-domain antigen would provide insights into vaccine design and immunotherapy. The multi-domain outer surface *Leptospira* immunoglobulin-like (Lig) proteins LigA and LigB, consisting of 12–13 homologous bacterial Ig (Big)-like domains, are potential antigens of *Leptospira interrogans*. Currently, no effective vaccine is available against pathogenic *Leptospira*. Both the humoral immunity and cell-mediated immunity of the host play critical roles in defending against *Leptospira* infection. Here, we used immunoinformatics approaches to evaluate antigenic B-cell lymphocyte (BCL) and cytotoxic T-lymphocyte (CTL) epitopes from Lig proteins. Based on certain crucial parameters, potential epitopes that can stimulate both types of adaptive immune responses were selected to design a chimeric vaccine construct. Additionally, an adjuvant, the mycobacterial heparin-binding hemagglutinin adhesin (HBHA), was incorporated into the final multi-epitope vaccine construct with a suitable linker. The final construct was further scored for its antigenicity, allergenicity, and physicochemical parameters. A three-dimensional (3D) modeled construct of the vaccine was implied to interact with Toll-like receptor 4 (TLR4) using molecular docking. The stability of the vaccine construct with TLR4 was predicted with molecular dynamics simulation. Our results demonstrate the application of immunoinformatics and structure biology strategies to develop an epitope-specific chimeric vaccine from multi-domain proteins. The current findings will be useful for future experimental validation to ratify the immunogenicity of the chimera.

Keywords: *Leptospira interrogans*, antigenic epitope, outer surface antigen, vaccine, *Leptospira* immunoglobulin-like protein, subunit vaccine, immunoinformatics

INTRODUCTION

Leptospirosis is categorized as an emerging and neglected tropical zoonotic disease worldwide. It is considered a public health problem globally, with an estimated 1 million leptospirosis cases reported each year, causing deaths of around 60,000 (1–3). The infection usually shows symptoms such as headache, chills, illness, and muscle aches, and a more severe form of disease is associated with



Enzyme engineering improves catalytic efficiency and enantioselectivity of hydroxynitrile lyase for promiscuous retro-nitroaldolase activity

Badipatla Vishnu Priya^a, D.H. Sreenivasa Rao^{a,1}, Rubina Gilani^{a,1}, Surabhi Lata^{b,1}, Nivedita Rai^a, Mohd. Akif^b, Santosh Kumar Padhi^{a,*}

^a Biocatalysis and Enzyme Engineering Laboratory, Department of Biochemistry, School of Life Sciences, University of Hyderabad, Hyderabad 500 046, India

^b Laboratory of Structural Biology, Department of Biochemistry, School of Life Sciences, University of Hyderabad, Hyderabad 500 046, India

ARTICLE INFO

Keywords:

Protein engineering
Catalytic promiscuity
Catalytic efficiency
Chiral β -nitro alcohol
Hydroxynitrile lyase
Asymmetric synthesis
Retro-Henry reaction

ABSTRACT

Protein engineering to improve promiscuous catalytic activity is important for biocatalytic application of enzymes in green synthesis. We uncovered the significance of binding site residues in *Arabidopsis thaliana* hydroxynitrile lyase (AtHNL) for promiscuous retro-nitroaldolase activity. Engineering of AtHNL has improved enantioselective retro-nitroaldolase activity, a synthetically important biotransformation, for the production of enantiopure β -nitroalcohols having absolute configuration opposite to that of the stereopreference of the HNL. The variant F179A has shown ~ 12 fold increased selectivity towards the retro-nitroaldol reaction over cyanogenesis, the natural activity of the parent enzyme. Screening of the two saturation libraries of Phe179 and Tyr14 revealed several variants with higher k_{cat} , while F179N showed ~ 2.4-fold k_{cat}/K_M than the native enzyme towards retro-nitroaldol reaction. Variants F179N, F179M, F179W, F179V, F179I, Y14L, and Y14M have shown > 99% ee in the preparation of (S)-2-nitro-1-phenylethanol (NPE) from the racemic substrate, while F179N has shown the E value of 138 vs. 81 by the wild type. Our molecular docking and dynamics simulations (MDS) studies results provided insights into the molecular basis of higher enantioselectivity by the F179N toward the retro-nitroaldolase activity than the other mutants. Binding energy calculations also showed the higher negative binding free energy in the case of F179N-(R)-NPE compared to other complexes that support our experimental low K_M by the F179N for NPE. A plausible retro-nitroaldol reaction mechanism was proposed based on the MDS study of enzyme-substrate interaction.

1. Introduction

In recent years enzymes have been increasingly exploited in the green synthesis of non-natural molecules and for different abiobiotic reactions, by virtue of their substrate and catalytic promiscuity [1]. However, the catalytic efficiency of natural enzymes for promiscuous reactions are often found to be low. This has triggered the research for laboratory evolution of enzymes with improved catalytic properties. While design and engineering of enzymes for improved promiscuous catalytic activity can be facilitated with mechanistic understanding, in several cases, the mechanism of promiscuous catalytic activity is not well understood. Here we have tried to address a similar case of improvement of promiscuous retro-nitroaldolase activity of a hydroxynitrile lyase (HNL) by protein engineering and predicted its plausible catalytic mechanism.

HNLs in nature catalyze cyanogenesis from cyanohydrins [2,3]. In biocatalysis, they carry out the reverse transformation of nucleophilic addition of cyanide to the carbonyl center in the synthesis of enantiopure cyanohydrins [4,5]. A few of them show promiscuity in addition of nucleophiles other than cyanides, e.g., nitromethane leading to the stereoselective synthesis of nitroaldol reaction products, i.e., β -nitroalcohols [6–10]. Both the enantiopure cyanohydrins and β -nitroalcohols are important chiral synthons used in the preparation of pharmaceuticals, agrochemicals, and biologically active molecules. The reverse transformation of the above two synthesis reactions are cyanogenesis and retro-nitroaldol or retro-Henry (Scheme 1). The latter is a synthetically important biotransformation, because a HNL catalyzed retro-nitroaldol reaction produces enantioenriched β -nitroalcohols having absolute configuration opposite to that of the stereopreference of the HNL [11–13]. This approach enables the enzyme to produce opposite

* Corresponding author.

E-mail address: skpadhi@uhyd.ac.in (S. Kumar Padhi).

¹ Authors equally contributed.

<https://doi.org/10.1016/j.bioorg.2021.105594>

Received 8 October 2021; Received in revised form 8 December 2021; Accepted 30 December 2021

Available online 4 January 2022

0045-2068/© 2022 Elsevier Inc. All rights reserved.



Identification of Novel GTP Analogs as Potent and Specific Reversible Inhibitors for Transglutaminase 2

Lakshmi P. Kolligundla, Samridhi Gupta, Surabhi Lata, Sandeep K. N. Mulukala, Praneeth Killaka, Mohd Akif  and Anil K. Pasupulati 

Department of Biochemistry, School of Life Sciences, University of Hyderabad, Hyderabad, India

ABSTRACT

Transglutaminase 2 (TG2) is a calcium-dependent enzyme that catalyzes the N- ϵ -(γ -glutamyl) lysine bonds between side chains of glutamine and lysine residues resulting in proteolytically resistant cross-links. Increased TG2 activity and levels are involved in the pathophysiology of various diseases, including liver injury, cystic fibrosis, celiac sprue, metastatic cancers, and several neurodegenerative conditions. Inhibiting TG2 activity is considered a potential strategy to combat these diseases. Although guanine nucleotide (GTP) could inhibit TG2, its inhibitory activity decreased with increased calcium concentration. Search for GTP analogs that could strongly bind and inhibit TG2 activity is intense. This study screened the PubChem database for about two thousand GTP-like compounds for TG2. Using docking and molecular dynamics simulations we identified three compounds (C4959, C4215, and C9560) that could selectively interact with TG2. These three compounds have less affinity for several other intracellular and extracellular GTP-binding proteins suggesting selectivity for TG2. Interestingly, C9560 showed stronger interactions and better binding energy with TG2 than C4959 and C4215, suggesting that C9560 can form a more stable complex with TG2. Our study indicates that C9560, a GTP analog, could be exploited as a promising candidate to inhibit TG2-mediated fibrotic conditions.

ARTICLE HISTORY

Received 23 September 2021
Accepted 16 August 2022

KEYWORDS

Transglutaminase 2; fibrosis;
GTP; Docking and molecular
dynamics simulations;
toxicology analysis

1. Introduction

Transglutaminase 2 (TG2), or tissue transglutaminase, is an extracellular calcium-dependent transferase enzyme (EC: 2.3.2.13). TG2 catalyzes the formation of N- ϵ -(γ -glutamyl) lysine crosslinks in peptide side chains [1]. TG2 is a ubiquitous member of the TG family and calcium is essential for its catalytic activity. Several extracellular matrix (ECM) proteins, including collagen, fibronectin, fibrinogen, laminin, and osteopontin, are client proteins of TG2. The intra- and inter-cross-linking of the client proteins by TG2 yields proteolytic-resistant products that accumulate at the cell surface and surrounding matrix. TG2 is essential for the stabilisation of the ECM, and several tissues (skin and hair) and processes (blood clotting and wound healing) involve TG2 cross-linked products [2]. Although TG2 is primarily considered a cytosolic protein, its presence in the nucleus, plasma membrane, cell surface, and extracellular space is reported [3].

The human TG2 is 686 amino acids, 76 kDa protein consisting of GTPase pocket capable of hydrolysing ATP and GTP. Crystal structure of TG2 in a complex with an inhibitor that mimics inflammatory gluten peptide substrates (PDB ID: 2Q3Z) revealed that TG2 has four distinct domains, namely N-terminal β -sandwich (1–139), catalytic core (140–454), and two C-terminal β -barrels (479–585 and 586–687) connected by the loop (455–478) [4]. Cys277, His335, and Asp358 (the catalytic triad) and a conserved Trp241 make up the catalytic core of TG2. TG2 alternates between open (active) and closed

(inactive) conformation based on the ligand binding [5,6]. TG2 is allosterically activated by calcium and deactivated by GTP [7]. In the TG2-GTP bound closed state, interactions between the catalytic domain and the two C-terminal β -barrels severely restrict access to the catalytic site [6]. Upon binding, calcium structurally alters TG2 to open conformation by moving the β -barrels apart and exposing the catalytic site [8]. Cys277 in the active site is crucial for TG2's enzymatic role. TG2 catalyzes the formation N- ϵ -(γ -glutamyl) lysine crosslinks in peptide side chains making them into isopeptidic products. This reaction occurs in two steps. In the first part, the thiol group of Cys277 attacks the substrate's acyl-donor group (e.g. glutamine), forming a thioester intermediate and releasing ammonia. In the second step, the acyl group from the Cys277 is transferred to acyl-acceptor amine substrate (e.g. lysine), forming an isopeptidic product and recycling the TG2 enzyme [8]. This catalytic activity requires millimolar concentrations of calcium [8].

Regulated isopeptide bond formation among TG2 client proteins is crucial for ECM turnover. However, in TG2 related pathologies, abnormal crosslinking of ECM components leads to diseases like celiac sprue, neurodegenerative disorders, diabetes, liver cirrhosis, pulmonary fibrosis, and renal scarring [1,9]. For example, known substrates of TG2 such as polyglutamine repeats and α -synuclein cause Huntington's disease [10,11]. Further, elevated TG2 expression causes chemotherapeutic resistance and metastatic potential in melanomas,

CONTACT Anil K Pasupulati  pasupulati.anilkumar@gmail.com

 Supplemental data for this article can be accessed online at <https://doi.org/10.1080/08927022.2022.2123917>.

© 2022 Informa UK Limited, trading as Taylor & Francis Group

PLAGIARISM REPORT

A Comprehensive investigation of protein structural parameters & dynamic features of therapeutic targets from SARS-CoV-2 & Mycobacterium tuberculosis: Implications towards drug-design

ORIGINALITY REPORT

25%

SIMILARITY INDEX

20%

INTERNET SOURCES

24%

PUBLICATIONS

16%

STUDENT PAPERS

PRIMARY SOURCES

1

www.ncbi.nlm.nih.gov

Internet Source

8%

2

Submitted to University of Hyderabad,
Hyderabad

Student Paper

7%

3

www.frontiersin.org

Internet Source

1%

4

Surabhi Lata, Mohd. Akif. " Structure-based
identification of natural compound inhibitor
against thioredoxin reductase: insight from
molecular docking and dynamics simulation ",
Journal of Biomolecular Structure and
Dynamics, 2020

Publication

1%

5

www.researchgate.net

Internet Source

1%



University of Hyderabad
Hyderabad- 500 046, India

PLAGIARISM FREE CERTIFICATE

This is to certify that the similarity index of this thesis as checked by the Library of the University of Hyderabad is 25%. Out of this, 15% similarity has been found from the candidate's (**Surabhi Lata**) own publications which form the adequate part of the thesis. The details the of student's publication are as follows:

Paper	% Similarity
1. Lata, S., & Akif, M. (2022). Probing structural basis for enhanced binding of SARS-CoV-2 P.1 variant spike protein with the human ACE2 receptor. <i>Journal of cellular biochemistry</i> , 123(7), 1207–1221. https://doi.org/10.1002/jcb.30276	7% (Source No. 1)
2. Lata, S., & Akif, M. (2021). Comparative protein structure network analysis on 3CL ^{pro} from SARS-CoV-1 and SARS-CoV-2. <i>Proteins</i> , 89(9), 1216–1225. https://doi.org/10.1002/prot.26143	7% (Source No. 2)
3. Lata, S., & Akif, M. (2021). Structure-based identification of natural compound inhibitor against <i>M. tuberculosis</i> thioredoxin reductase: insight from molecular docking and dynamics simulation. <i>Journal of biomolecular structure & dynamics</i> , 39(12), 4480–4489. https://doi.org/10.1080/07391102.2020.1778530	1% (Source No. 4)

About 10% similarity was identified from external sources in the present thesis which is in accordance with the regulations of the University. All the publications related to this thesis have been appended at the end of the thesis. Hence the present thesis may be considered to be plagiarism free.


Dr. Mohd. Akif

MOHD. AKIF
Assistant Professor
DEPARTMENT OF BIOCHEMISTRY
SCHOOL OF LIFE SCIENCES
UNIVERSITY OF HYDERABAD
HYDERABAD-500 046, (INDIA)

Date of Registration: 2/8/2018
Date of Expiry: 2/8/2023

UNIVERSITY OF HYDERABAD
SCHOOL OF LIFE SCIENCES
DEPARTMENT OF BIOCHEMISTRY

PROFORMA FOR SUBMISSION OF THESIS/DISSERTATION

1. Name of the Candidate : Anabhi Lato
2. Roll No. : 18LRPH05
3. Year of Registration : 2018
4. Topic : A comprehensive investigation of protein structural parameters & dynamic features of therapeutic target from SARS-CoV-2 & its implications towards drug design.
5. Supervisor(s) and affiliations : Dr. Mohd A Kif
6. Whether extension/re-registration granted: No
(Specify below)
7. Date of submission : 15/11/2022
8. Last class attended on (M.Phil/Pre-Ph.D): 7th Sept, 2022
9. Whether tuition fee for the current period Paid : Receipt No. & Dt :
10. Whether thesis submission fee paid : Receipt No & Dt : 221109135817246, 9th Nov, 2022
Rs 1200/-
11. Whether hostel dues cleared : Receipt No & Dt :
12. Whether a summary of the dissertation enclosed:

Supervisor(s)
Signature:

Signature of the
Dean, of the School

Forwarded on

Signature of the Head of the
Department.

HEAD
Dept. of Biochemistry
SCHOOL OF LIFE SCIENCES
UNIVERSITY OF HYDERABAD
HYDERABAD-500 046

School of Life Sciences
University of Hyderabad
Hyderabad-500 046.

12:59

65%

← VD-ICICIB



Dear Payer, you have made a payment of Rs. 5020.00 to FINANCE OFFICER UNIVERSITY OF HYDERABAD vide eazypay reference ID 221108135780399.

8/11/2022

Transaction • [Change category](#)



Dear Payer, you have made a payment of Rs. 8216.00 to WARDEN LADIES HOSTEL IX SOUTH UOH vide eazypay reference ID 221108135780568.

8/11/2022

12:55

Transaction • [Change category](#)



Dear Payer, you have made a payment of Rs. 1200.00 to FINANCE OFFICER UNIVERSITY OF HYDERABAD vide eazypay reference ID 221109135817246.

9/11/2022