

---

## Structural and Mutational Analysis of SARS-CoV-2 Protein Interactome

Megha C Terdal<sup>1</sup>, Jayashree G<sup>2</sup>, Harshitha Priya R<sup>3</sup>, Himabindu B<sup>4</sup>, Sukruth K<sup>5</sup>,  
Manohar G M<sup>6</sup>, Kiran Kumar D J<sup>7</sup>

<sup>1-7</sup>*Department of PG studies in Biotechnology, Nrupatunga University, Bangalore, Karnataka, India.*

<sup>6</sup>*halliLabs, Ragihalli, Anekal Tq, Bangalore, Karnataka*

### ABSTRACT

SARS CoV-2, also known as the Severe Acute Respiratory Syndrome Coronavirus 2, is a highly contagious pathogenic virus that affects millions of people and causes respiratory diseases. Studying the intraviral protein-protein interactions of the virus will give us insight on the functioning mechanisms of the cell. And also, this is important to predict the severity of infection as and when mutations evolve. In order to establish which amino acids mutations, lead to stabilization and destabilization of protein-protein interactions in the virus, 15 intraviral proteins identified as interacting in literature were docked with their partners using ClusPro and the interface residues were mutated using FoldX. From the 15 interactions, the highest and lowest energy changes of each interaction yielded 30 models which showed 19 stabilizing mutations and 9 destabilizing mutations with an interaction complex showing neither. Stabilizing mutations may strengthen the interactions and therefore provide a more robust biomolecular interaction with the host while the opposite may be expected with destabilizing mutations. Our findings may be of support for deducing insights into the possible changes in the severity of the disease associated with the evolving mutations.

**Keywords:** SARS-CoV-2, protein-protein docking, intraviral protein interaction, interface residue, stabilizing energy, destabilizing energy, ClusPro, FoldX

### 1. Introduction

The coronavirus family includes SARS-CoV-2, which causes severe acute respiratory syndrome. The virus possesses a single-stranded, positive-sense RNA genome (+ ssRNA) of around 26 to 32 kilobases, and has been known to infect a variety of mammalian hosts including humans [6]. The SARS-CoV-2 virus expresses 29 proteins, which can be divided into three categories: structural proteins (S, E, M, and N proteins) involved in the structural characterization of the virus; non-structural proteins (NSP1-NSP16), which are proteins not typically involved with viral particles but are instead enzymes; transcription factors, which are essential for virus survival; and accessory proteins (ORF3a, ORF3b, ORF6, ORF7a, ORF7b, ORF8, ORF9, ORF10, ORF14) that plays an indirect role in viral functions [7]. The full set of physical interactions that occur between a cell's macromolecules is known as the interactome. Characterizing the roles of unidentified proteins requires the use of the interactome. Additionally, it aids in locating the molecules that play a key role in pathogenesis and therefore assume importance for therapeutic targeting.[8]. Any virus will evolve and accumulate mutations that will have a direct bearing on the intraviral interactome that influence the virulence, infectivity and host relationship. Here we dock all the interacting protein structures of Sars cov2 using Cluspro and analyse the stability of the interactions upon mutations using Foldx. From the above interactions, 15 interactions produced two models each with the highest and lowest total energies giving out 30 models, out of which 19 models displayed stabilizing mutations, 9 models displayed destabilizing mutations, and two models, did not exhibit any energy change. Stabilizing mutations may strengthen interactions, resulting in a stronger biomolecular interaction with the host, whereas destabilizing mutations may have the opposite effect. Our study may assist in improving the understanding of disease severity changes induced by evolving mutations.

## 2. Methods

29 proteins of SARS-CoV-2, and their intraviral interactions were obtained from UniProt (<https://covid-19.uniprot.org/>) [11]. 15 protein crystal [table S1] structures were obtained from Protein Data Bank (<https://www.rcsb.org/>) [10] except few proteins (NSP4, NSP5, NSP6, NSP11, M, Orf6, Orf7b, orf14) for which the structures were not available. The modeled structures of these proteins were obtained from the database from Zhang's lab (<https://zhanggroup.org/COVID-19/>).

Protein-Protein Docking of the intraviral binding partner proteins was done using ClusPro (<https://cluspro.bu.edu/login.php>) [1][9] between the interacting partners.

DockScore (<http://caps.ncbs.res.in/dockscore/>) [2] was used to rank the protein complexes from the many models obtained through ClusPro docking for each interacting pair.

PPCheck (<http://caps.ncbs.res.in/PPCheck/>) [3] was used to find which particular amino acids were interacting in the protein complexes. From the results hydrophobic, electrostatic, salt bridges, and short contacts were considered and the van der Waals was ignored.

To mutate the interacting proteins *in silico* conditions and analyze the protein complexes FoldX (<https://foldxsuite.crg.eu/>) [4] tool was used. Protein complexes were first “*repaired*” (FoldX --command=RepairPDB --pdb=RP.pdb). Next the “*PSSM command*” (FoldX --command=Pssm --analyseComplexChains=A, B --pdb=PM.pdb --positions=GA5a, GA14a) was used to mutate and analyze the complex. Lastly “*analyze complex*” command (FoldX --command=AnalyseComplex --pdb=AC.pdb --analyseComplexChains=A, B) was used to find the interaction energy for the individual mutated models which showed the highest and lowest total energy.

## 3. Result and Discussion

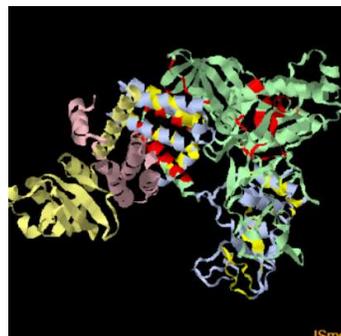
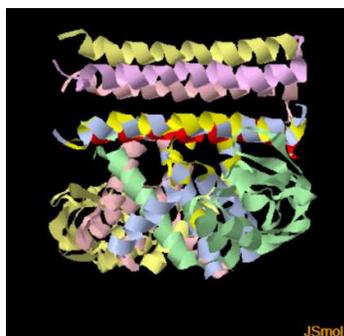
20 intraviral interaction of proteins positive for Co-immunoprecipitation from SARS-CoV2 was considered for the study (Table S2) [5]. Rigid docking software ClusPro [1] was used to dock protein-protein interactions of intraviral SARS-CoV-2 proteins as it enables the docking of proteins without any prior information on the complex structure and also due to its high accuracy among the other fully automated docking servers. Docked interacting protein complexes of input individual proteins were obtained from ClusPro. Among 20 interactions, 15 interactions were docked [Fig S(1), Fig S(2), Fig S(3), Fig S(4), S(5), Fig S(6), Fig S(7), Fig S(8), Fig S(9), Fig S(10), Fig S(11), Fig S(12), Fig S(13), Fig S(14), Fig S(15)] except the interactions E-ORF7b, ORF6-ORF7b, NSP2-NSP11, ORF14-ORF9b, NSP2-NSP6 because both the crystal and modeled structures of ORF7b, ORF14 and NSP 11 were unavailable. The interaction of NSP2-NSP6 was not docked due to an internal error.

To narrow down the models to native or near-native complexes so that the interactions between the proteins is as close to the ones occurring inside the cells, the DockScore web server was utilized. DockScore [2] was chosen because it considers various interface parameters such as surface area, evolutionary conservation, hydrophobicity, short contacts, and spatial clustering at the interface for scoring (Z score). The model with a high Z score was ranked as the top model. The Z scores for the top ranked models were between 4.7510 for NSP8 and E (Fig 1) to -5.8804 for NSP14 and ORF 9B (Fig 2). All the top models and their Z scores for each protein complex are given in Table S3.

Interface residues of protein complexes were identified using PPCheck [3] as it is an improvised version that considers only interface residue for normal energy calculation per residue as well as optimum distance cutoff implementation that can be used in short contacts, hydrophobic interactions, Van der Waals pairs, salt bridges as well as calculating electrostatic interactions. Interface residues for van der Waals were ignored as it was not possible for us to compute thousands of amino acid interactions. The number of amino acids of short contacts, hydrophobic interactions, Van der Waals pairs, salt bridges, and electrostatic interactions is listed in Table 1.

Amino acid mutations alter the 3D structure of proteins, affecting protein stability, function, and interaction with other biomolecules. Mutating the interacting interface amino acids with the rest of the 19 “natural” amino acids determines whether the protein-protein interaction is stabilized or

destabilized based on the changes in free energy. FoldX was used to perform these SARS-CoV-2 stabilizing and destabilizing mutations. Over the other traditional approaches, FoldX was considered as it uses high computational speed and protein engineering studies, carefully parameterized by actual experimental data for its energy terms. Foldx's PSSM command generates a large number of mutated models, analyzes the entire complex, and predicts protein stability in terms of total energy. For each interaction, the highest and lowest total energies were considered, which are compiled in (Table.S.5).



**Fig.1.** NSP8 and E, with a Z score of 4.7510 **Fig.2.** NSP14 and ORF9B, with a Z score of -5.8804

**Table.1.** Number of interface residues for interacting protein complexes obtained from

Protein Complex	Short Contacts	Hydrophobic Interactions	Van der Waals pairs	Salt Bridges	Electrostatic Interactions
NSP2 and ORF3a	0	12	4654	2	4
NSP1 and E	0	11	2815	4	1
NSP7 and NSP8	8	72	8006	24	0
NSP3 and ORF3a	0	0	1756	2	0
NSP2 and NSP16	2	8	4068	8	4
NSP2 and NSP8	1	2	2332	9	18
NSP2 and NSP3	2	9	3789	14	19
NSP5 and ORF 9B	33	55	6586	6	2
NSP8 and NSP12	5	48	9786	18	0
NSP8 and NSP13	3	69	6697	14	2
NSP8 and NSP14	2	60	1367	6	4
NSP8 and E	0	62	5893	7	4
NSP9 and NSP8	2	67	5655	14	0
NSP14 and ORF9B	70	21	18741	6	9
M and N	0	56	10477	2	0

**PPCheck**

Each individual complex's interaction energy was calculated for the models from Table 2. The energy released during the bond formation or interaction of ligand and protein is known as binding energy/interaction energy. The interaction energy of the entire complex was considered as  $\Delta G$  wildtype, and that of individual models was considered as  $\Delta G$  mutant, with the difference considered as  $\Delta\Delta G$ .  $\Delta\Delta G$  is a measure of the difference in energy between the folded and unfolded states ( $\Delta G$  folding) as well as the difference in  $\Delta G$  folding when a point mutation is present. This has been discovered to be an excellent predictor of whether a point mutation will be beneficial to protein

stability. With each interaction producing one highest and lowest model, a total of 15 interactions produced 30 interaction models, of which 19 had stabilized mutations, 9 had destabilized mutations, and the remaining two had zero mutations (Table 2)

**Table 2: List of amino acid mutations in SARS-CoV-2 along with its  $\Delta\Delta G$  value**

Protein complex	Mutating Amino Acids	$\Delta G$ wild type	$\Delta G$ Mutant	$\Delta\Delta G$ ( $\Delta G$ Wild type – $\Delta G$ Mutant)
NSP1 E Repair 28	AA22 - Ile	-4.86924	-7.15824	2.289
NSP1 E Repair 179	AB70 - Trp	-4.86924	1.3071	-6.1734
NSP2 NSP3 Repair 555	DB37 - Arg	-19.0276	22.6885	3.6609
NSP2 NSP3 Repair 99	AA596 - Trp	-19.0276	3.28935	-22.31695
NSP2 NSP8 Repair 272	RA246 - Met	-60.409	-62.0808	1.6718
NSP2 NSP8 Repair 559	DB161 - Trp	-60.409	-46.2581	-14.1509
NSP2 NSP16 Repair 191	EB60 - Leu	-36.5334	-37.4885	0.9551
NSP2 NSP16 Repair 19	DA391 - Trp	-36.5334	-16.3379	-20.1955
NSP2 Orf3a Repair 388	AB59 - Ile	-100.357	-102.212	1.855
NSP2 Orf3a Repair 119	VA459 - Trp	-100.357	-94.2445	-6.1125
NSP2 Orf3a Repair 29	KB66 - Lys	-5.73289	-5.73289	0.00
NSP2 Orf3a Repair 33	KB66 - Pro	-5.73289	-5.73289	0.00
NSP7 NSP8 Repair 341	RB190 - Ala	-42.7513	-40.5779	-2.1734
NSP7 NSP8 Repair 439	IB106 - Trp	-42.7513	-20.1724	-22.5789
NSP5 Orf9b Repair 341	RA217-ala	-50.9823	-50.1197	-0.8626
NSP5 Orf9b Repair 880	QB77-val	-50.9823	-48.6265	-2.3558
NSP9 NSP8 Repair 311	LA71-met	-45.1193	-45.2857	0.1664
NSP9 NSP8 Repair 439	RA10-tyr	-45.1193	-28.5926	-16.5267
NSP8 NSP12 Repair 471	KA2-met	-47.23	-47.9329	0.7029
NSP8 NSP12 Repair 715	YB135-arg	-47.23	-47.1856	-0.0444
NSP8 NSP13 Repair 559	LB91-tyr	-44.9471	-27.8553	-17.076
NSP8 NSP13 Repair 736	FB9-ser	-44.9471	-42.8711	-2.076
NSP8 NSP14 Repair 559	LB91-tyr	-44.5197	-35.2559	-9.2638
NSP8 NSP 14 Repair 721	AB110-ala	-44.5197	-42.4512	-2.0685
E NSP8 Repair 799	IB106-tyr	-53.1792	-14.3622	-38.817
E NSP8 Repair 911	VB115-met	-53.1792	-14.3622	-38.817
NSP 14 orf9B Repair 33	AA4-pro	-34.7523	-35.752	1.0057
NSP 14 orf9B Repair 459	LB7-tyr	-34.7523	-33.778	-0.9743
M N Repair 387	LA93-pro	-86.7171	-64.8202	-21.8969
M N Repair 463	RA107-asp	-86.7171	-88.4586	1.7415

Aspartic acid to arginine mutation at the 37th position in the complexes NSP2 and NSP3 (Fig.5) exhibited a maximum  $\Delta\Delta G$  value of 3.6609, interpreting the destabilizing mutation, while valine to methionine at the 115th position (Fig.4) and isoleucine to tyrosine mutations 106th position in the E and NSP 8 interacting complex showed lowest  $\Delta\Delta G$  value of -38.817, depicting the stabilizing mutations.  $\Delta\Delta G$  value of other complexes ranges between these values which is plotted in the graph (Fig.3) below.



Fig.3. Based on the  $\Delta\Delta G$  value, a graph shows stabilizing and destabilizing mutations of amino acid

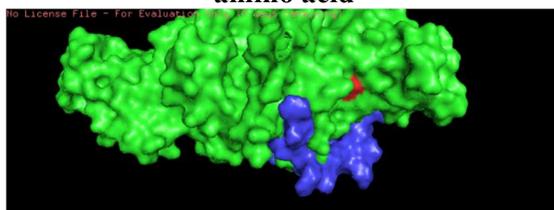


Fig.4: NSP8 (green) and E (blue) complex highlighted with valine (red) amino acid. Showing the stabilized mutation in the complex.

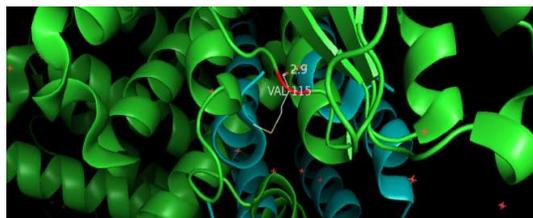


Fig 4(a): NSP 8 (green) and E (blue) complex showing valine at 115th position.

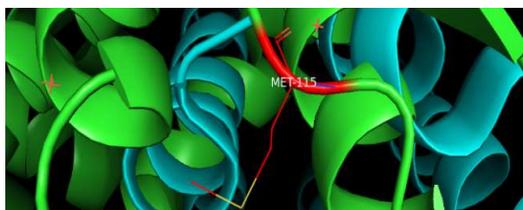


Fig 4(b): NSP 8(green) and E (blue) complex showing mutation of valine to methionine at 115th position with  $\Delta\Delta G$  value of -38.817 indicating stabilized mutation

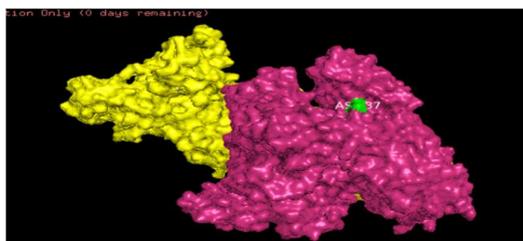
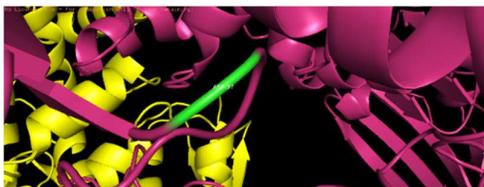
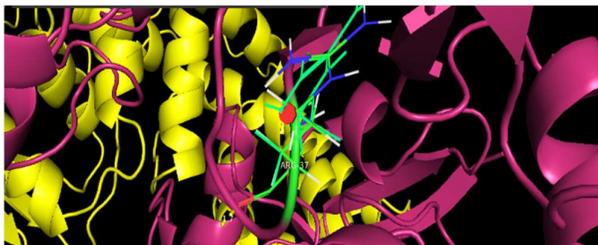


Fig.5: NSP2 (pink) and NSP3 (yellow) complex highlighted with the amino acid aspartic acid (green). Showing the destabilizing mutation in the complex.



**Fig 5(a): NSP2(Pink) and NSP3(yellow) complex showing aspartic acid at 37th position.**



**Fig 5(b): NSP2(Pink) and NSP3(yellow) complex showing mutation of aspartic acid to arginine at 37th position with  $\Delta\Delta G$  value of 3.6609 indicating destabilizing mutation.**

Comparing the modeled structures from the Zang lab with the crystal structures:

All of the information on the crystalline structures used in our work was taken from the PDB database, and Zhang Lab modelled structures that were created using homology modelling were referred to when crystalline structures were not accessible. The four nonstructural proteins—NSP4, NSP5, NSP6, and NSP11, structural proteins S and M, and accessory proteins include ORF3b, ORF6, ORF7a, ORF7b, ORF8, ORF9, ORF10, and ORF14 are the modeled structures used by Zhang Lab. In order to compare the similarities between the crystalline structure and the modelled structures, they were superimposed, and the RMSD value was obtained. Only the NSP2 value has less similarity, while NSP1, NSP3, NSP7, NSP8, NSP9, NSP10, NSP12, NSP13, NSP14, NSP15, NSP16, E, N, and ORF3a have high similarity, and these are listed in Table S4 below.

Superimposing SARS-CoV and SARS-CoV 2 protein structures:

PYMOL software was used to superimpose the protein structures of SARS-CoV and Sars CoV-2 to determine their similarity. The average distance between atoms and the degree of similarity of proteins are determined by the RMSD (Root Mean Square Deviation) value, which was obtained. The resemblance is greater the lower the RMSD. A good RMSD value is less than  $2\text{\AA}$  [9]. Most of the protein structures of SARS-CoV and Sars CoV-2 that were superimposed had RMSD values of less than  $2\text{\AA}$ , indicating greater similarity.

## CONCLUSION

The protein partners of the SARS-CoV-2 interactome were docked and the systematic mutational analyses was conducted. Several mutations either increased the interaction affinity or decreased it. These mutations may alter the viral protein's shape, binding affinity, and hot spots at the interface, which may have an effect on COVID-2 transmission, infection efficiency. These mutations can indicate the direction in which the current strains of the SARS-CoV2 may be evolving.

## Acknowledgement

We deeply appreciate the help and mentoring from Srikanth L, Research Scholar from University of Bristol, UK for regular inputs for this research.

**Reference**

1. Kozakov D, Hall DR, Xia B, Porter KA, Padhorny D, Yueh C, Beglov D, Vajda S. The ClusPro web server for protein-protein docking. *Nature Protocols*. 2017 Feb;12(2):255-278.
2. Malhotra S, Mathew O, Sowdhamini R.(2015) DOCKSCORE: a webserver for ranking protein-protein docked poses. *BMC Bioinformatics*. vol: 16:127. [DOI: [10.1186/s12859-015-0572-6](https://doi.org/10.1186/s12859-015-0572-6)] [PMID: [25902779](https://pubmed.ncbi.nlm.nih.gov/25902779/)]
3. Sukhwal, A.; Sowdhamini, R. Oligomerization status and evolutionary conservation of interfaces of protein structural domain superfamilies. *Mol. Biosyst*. 2013, 9, 1652-1661.
4. Schymkowitz J, Borg J, Stricher F, Nys R, Rousseau F, Serrano L. The FoldX web server: an online force field. *Nucleic Acids Res*. 2005 Jul 1;33(Web Server issue):W382-8. doi: 10.1093/nar/gki387. PMID: 15980494; PMCID: PMC1160148.
5. von Brunn A, Teepe C, Simpson JC, Pepperkok R, Friedel CC, Zimmer R, Roberts R, Baric R, Haas J. Analysis of intraviral protein-protein interactions of the SARS coronavirus ORFome. *PLoS One*. 2007 May 23;2(5):e459. doi: 10.1371/journal.pone.0000459. PMID: 17520018; PMCID: PMC1868897.
6. Yadav R, Chaudhary JK, Jain N, Chaudhary PK, Khanra S, Dhamija P, Sharma A, Kumar A, Handu S. Role of Structural and Non-Structural Proteins and Therapeutic Targets of SARS-CoV-2 for COVID-19. *Cells*. 2021 Apr 6;10(4):821. doi: 10.3390/cells10040821. PMID: 33917481; PMCID: PMC8067447.
7. Gordon DE, Jang GM, Bouhaddou M. *et al*. A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. *Nature*. 2020 Jul;583(7816):459-468. doi: 10.1038/s41586-020-2286-9. Epub 2020 Apr 30. PMID: 32353859; PMCID: PMC7431030.
8. P, Surat. (2018, August 24). Importance of Studying the Interactome. News-Medical. Retrieved on July 22, 2022 from <https://www.news-medical.net/life-sciences/Importance-of-Studying-the-Interactome.aspx>.
9. Vajda S, Yueh C, Beglov D, Bohnuud T, Mottarella SE, Xia B, Hall DR, Kozakov D. New additions to the ClusPro server motivated by CAPRI. *Proteins: Structure, Function, and Bioinformatics*. 2017 Mar; 85(3):435-444
10. Stephen K Burley et al,. RCSB Protein Data Bank: powerful new tools for exploring 3D structures of biological macromolecules for basic and applied research and education in fundamental biology, biomedicine, biotechnology, bioengineering and energy sciences, *Nucleic Acids Research*, Volume 49, Issue D1, 8 January 2021, Pages D437–D451, <https://doi.org/10.1093/nar/gkaa1038>
11. The UniProt Consortium, UniProt: the universal protein knowledgebase, *Nucleic Acids Research*, Volume 45, Issue D1, January 2017, Pages D158–D169, <https://doi.org/10.1093/nar/gkw1099>
12. Carugo O, Pongor S. A normalized root-mean-square distance for comparing protein three-dimensional structures. *Protein Sci*. 2001 Jul;10(7):1470-3. doi: 10.1110/ps.690101. PMID: 11420449; PMCID: PMC2374114.

**Supplementary Section:**

<https://docs.google.com/document/d/1TDwoHxF11VKoSV1Chtw2xXbhKPOWUEUQ--6XZw-X1AA/edit?usp=drivesdk>