



In silico characterization of mutations circulating in SARS-CoV-2 structural proteins

Neha Periwal^{a*}, Shravan B. Rathod^{b*} , Ranjan Pal^c, Priya Sharma^a, Lata Nebhnani^d, Ravi P. Barnwal^e , Pooja Arora^f, Kinshuk Raj Srivastava^c and Vikas Sood^a 

^aDepartment of Biochemistry, School of Chemical & Life Sciences, Jamia Hamdard, New Delhi, India; ^bDepartment of Chemistry, Smt. S. M. Panchal Science College, Talod, India; ^cBiocatalysis and Enzyme Engineering Lab, Regional Centre for Biotechnology, Faridabad, India; ^dDepartment of Chemistry, Gujarat University, Ahmedabad, India; ^eDepartment of Biophysics, Panjab University, Chandigarh, India; ^fDepartment of Zoology, Hansraj College, University of Delhi, New Delhi, India

Communicated by Ramaswamy H. Sarma

ABSTRACT

SARS-CoV-2 has recently emerged as a pandemic that has caused more than 2.4 million deaths worldwide. Since the onset of infections, several full-length sequences of viral genome have been made available which have been used to gain insights into viral dynamics. We utilised a meta-data driven comparative analysis tool for sequences (Meta-CATS) algorithm to identify mutations in 829 SARS-CoV-2 genomes from around the world. The algorithm predicted sixty-one mutations among SARS-CoV-2 genomes. We observed that most of the mutations were concentrated around three protein coding genes viz nsp3 (non-structural protein 3), RdRp (RNA-directed RNA polymerase) and Nucleocapsid (N) proteins of SARS-CoV-2. We used various computational tools including normal mode analysis (NMA), C- α discrete molecular dynamics (DMD) and all-atom molecular dynamic simulations (MD) to study the effect of mutations on functionality, stability and flexibility of SARS-CoV-2 structural proteins including envelope (E), N and spike (S) proteins. PredictSNP predictor suggested that four mutations (L37H in E, R203K and P344S in N and D614G in S) out of seven were predicted to be neutral whilst the remaining ones (P13L, S197L and G204R in N) were predicted to be deleterious in nature thereby impacting protein functionality. NMA, C- α DMD and all-atom MD suggested some mutations to have stabilizing roles (P13L, S197L and R203K in N protein) where remaining ones were predicted to destabilize mutant protein. In summary, we identified significant mutations in SARS-CoV-2 genomes as well as used computational approaches to further characterize the possible effect of highly significant mutations on SARS-CoV-2 structural proteins.

ARTICLE HISTORY

Received 9 June 2020
Accepted 19 March 2021

KEYWORDS

SARS-CoV-2; COVID-19; protein stability; mutations; nucleocapsid protein; spike protein; envelope protein; discrete molecular dynamics; molecular simulations; viral pathogenicity

Introduction

Coronaviruses are enveloped viruses having single-strand RNA as genetic material. First pandemic of coronaviruses was caused by SARS coronavirus in 2002 resulting in 919 deaths worldwide (Yang et al., 2020). Another wave of coronavirus pandemic was caused by MERS coronavirus in 2012 leading to 2499 infections and 858 deaths (Memish et al., 2020). Notably, MERS coronavirus was more pathogenic with around 34.3% mortality as compared to SARS coronavirus with mortality ratio of around 9.6%. Currently, we are witnessing another case of pandemic caused by a novel SARS coronavirus which originated in Wuhan, China (Zhou et al., 2020). The virus known as SARS-CoV-2 has infected millions and has caused more than 2.4 million deaths worldwide (<https://www.who.int/publications/m/item/weekly-epidemiological-update—23-february-2021>). Since the onset of infection, several full-length sequences of viral genome have

been made available with an aim to gain insights into the viral pathogenesis. These genomic sequences have been instrumental in gaining understanding into evolution and pathogenesis of SARS-CoV-2. Recent reports suggest that SARS-CoV-2 could have originated from bats and showed around 89.1-96% similarity with other bats originated coronaviruses (Wu et al., 2020; Zhou et al., 2020). Further analysis identified bat coronavirus RaTG13 as the closest relative of SARS-CoV-2. Scientists have obtained several insights from SARS-CoV-2 genomes which might help explain the rapid transmission of SARS-CoV-2. Apart from information obtained from viral genomes, the pattern of mutations that virus accumulate also sheds light into the possible evolution and pathogenesis of the virus. High rate of mutations among RNA viruses further hinders the development of potential therapeutics against them (Peck & Lauring, 2018).

As SARS-CoV-2 infections are progressing, there are chances that virus might accumulate several new mutations.

CONTACT Vikas Sood  vikas1101@gmail.com  Department of Biochemistry, School of Chemical & Life Sciences, Jamia Hamdard, New Delhi, India; Shravan B. Rathod  shravanathorizon93@gmail.com  Department of Chemistry, Smt. S. M. Panchal Science College, Talod, India

*Equal Contribution

 Supplemental data for this article can be accessed online at <https://doi.org/10.1080/07391102.2021.1908170>.

© 2021 Informa UK Limited, trading as Taylor & Francis Group

Nucleotide substitution plays a major role in virus evolution and several groups are studying viral genomic sequences to identify mutations driving SARS-CoV-2 evolution and pathogenesis. One of the earlier studies identified several nucleotide substitutions and deletions among SARS-CoV-2 genomic sequences from around the world (Phan, 2020). Data from phylogenetic analysis of SARS-CoV-2 identified three clusters within the virus circulating throughout the world. The study revealed that cluster A and C were more prevalent in Europe and America whereas cluster B was confined to East Asia (Forster et al., 2020). Another study identified 782 variant sites in 3067 genomes from 55 countries. Out of these 782 variants, around 65% of them were non-synonymous in nature suggesting that the majority of the mutations that occur in SARS-CoV-2 can result in a non-synonymous mutation (Laamarti et al., 2020). A recent report further aimed to characterize patient derived mutations on the pathogenesis of SARS-CoV-2 (Yao et al., 2020). The authors characterized eleven different mutations in patient derived viral isolates. These viral isolates differed significantly towards viral load as well as their infectivity of Vero cells thereby pointing towards the crucial role of mutations in viral pathogenesis. Another group utilized computational methods to characterize D614G mutation in SARS-CoV-2 spike protein and observed that the mutation led to generation of serine protease cleavage site near S1-S2 junction (Bhattacharyya et al., 2020). Further characterization of this mutation revealed that pseudovirus containing spike protein with D614G mutation entered Angiotensin-converting enzyme 2 (ACE2) expressing cells more efficiently as well as higher incorporation of S protein in virion particles (Zhang et al., 2020). Further reports suggested that D614G spike mutation led to increased infectivity of the SARS-CoV-2 (Korber et al., 2020). Since the onset of SARS-CoV-2 infections, it has accumulated several mutations and the recent emergence of new SARS-CoV-2 strain in the United Kingdom further attest to the fact. Further analysis revealed that some of the mutations in this new strain might lead to higher fitness of the virus (Singh et al., 2021).

As SARS-CoV-2 infections are increasing rapidly with each passing day, the virus is accumulating new mutations that might help the virus infectivity. Hence, we utilized meta-analysis based approaches to identify highly significant mutations among viral genomic sequences from around the world. In order to further understand the possible role of the mutations in viral structural proteins, we utilized a suite of algorithms including normal mode analysis, discrete molecular dynamics and all-atom molecular dynamic simulations to predict the possible effect of these mutations on parent proteins. Our extensive analysis of several mutations on SARS-CoV-2 structural proteins revealed that L37H mutation in E protein, G204R and P344S in N protein and D614G in S protein were destabilizing the parent protein whereas P13L, S197L, and R203K in N protein had stabilizing effect on the parent protein. This study provides valuable insights into possible changes of SARS-CoV-2 structural proteins due to novel mutations from around the world.

Material and methods

SARS-CoV-2 genomic sequences and meta-analysis

SARS-CoV-2 sequences used in this study were obtained from Virus Pathogen Resource (ViPR) (Pickett et al., 2012). A total of 829 sequences that were available to download till 18 April 2020 were used in this study. All the sequences were broadly grouped according to their geographic locations. A total of twelve groups were formed representing the six continents from where the virus was sequenced. We utilized a meta-data driven comparative analysis tool for sequences (Meta-CATS) algorithm to identify significant changes in various SARS-CoV-2 genomes (Pickett et al., 2013). All the analysis was performed on default settings and mutations having $p < 0.05$ were considered to be significant.

Protein structure retrieval and preparation

The crystal structures of the envelope protein (PDB: 7K3G) (Mandala et al., 2020), C-terminal domain (CTD) of nucleocapsid phosphoprotein (PDB: 6ZCO) (Zinzula et al., 2021), and spike glycoprotein (PDB: 6VYB) (Walls et al., 2020) were retrieved from the protein data bank (PDB) (<https://www.rcsb.org/>). We studied single mutation, L37H, and D614G in envelope and spike proteins respectively. However, in nucleocapsid protein, we analysed five mutations, P13L, S197L, R302K, G204R, and P344S that were spanning the whole protein. While searching for N protein crystal structures through the PDB database, we found that only Pro344 residue containing protein crystal structure (CTD, PDB: 6ZCO) has been reported. Since crystal structure of N protein spanning the remaining mutations were not available, hence we used modelled structure (Code: QHD43423) (<https://zhanglab.ccmb.med.umich.edu/COVID-19/>) to study the effect of remaining mutations on N protein. The structures used in the study i.e. envelope, nucleocapsid C-terminal domain, nucleocapsid modelled, and spike proteins have 31 (a.a.: 8-38), 135 (a.a.: 230-364), 419 (a.a.: 1-419) and 698 (a.a.: 1-698) amino acid residues respectively. Envelope, nucleocapsid CTD, and spike have 2.1 Å (Mandala et al., 2020), 1.36 Å (Zinzula et al., 2021), and 3.20 Å (Walls et al., 2020) resolution respectively. The missing residues in N protein CTD and spike protein were added using ModLoop (Fiser & Sali, 2003) and SWISS-MODEL web server (Schwede et al., 2003). To investigate the effects of mutation on the stability and flexibility of protein, mutations were inserted in all three wild-type proteins using the protein editing tool, PyMOL (Schrödinger LLC, 2010). Further, the modelled N protein was refined using Chiron (Ramachandran et al., 2011) web server without imposing any constraints. During the refinement, the van der Waals repulsion energy between the two residues was computed using the CHARMM19 force field (Brooks et al., 1983). The cut-off value of repulsion energy was set to 0.3 kcal mol⁻¹ (0.5 K_BT) hence, all of the steric clashes between the residues were removed that had greater than 0.3 kcal mol⁻¹ repulsion energy. The root mean square deviation (RMSD) between the unrefined and refined N modelled protein was found 1.28 Å. For further analysis, experimentally solved

structures, E, N protein CTD, and S proteins were directly used without any refinements.

Effects of mutation on protein function

The impact of mutation on protein functions were evaluated using PredictSNP (Bendl et al., 2014). This web server compiles six best prediction tools (MAPP, PhD-SNP, PolyPhen-1, PolyPhen-2, SIFT, and SNAP) and gives accurate prediction confidence scores for deleterious or neutral types of mutations. Embedded algorithms in PredictSNP use different approaches to predict the mutation effect. To evaluate accurate nature of the mutation, MAPP, PhD-SNP, PolyPhen-1, PolyPhen-2, SIFT, and SNAP utilize physicochemical properties and alignment score, support vector machine, expert set of empirical rules, naïve Bayes, alignment score, and neural network methods respectively (Bendl et al., 2014). PredictSNP uses the following mathematical expression to calculate the score,

$$\text{PredictSNPScore} = \frac{\sum_{i=1}^N (\delta_i \cdot S_i)}{\sum_{i=1}^N S_i}$$

Where N indicates the number of integrated tools, δ_i is an overall prediction (neutral prediction, -1 ; deleterious prediction, $+1$) and S_i represents the transformed confidence scores. PredictSNP Score is from -1 to $+1$ in which, -1 to 0 is denoted for neutral and 0 to $+1$ for deleterious nature of mutation.

Prediction of protein stability and flexibility upon mutation

To probe the effects of mutation on the stability and flexibility of proteins, free energy change ($\Delta\Delta G$) was calculated using the structure-based tool DynaMut (Rodrigues et al., 2018). DynaMut employs normal mode analysis (NMA) to calculate $\Delta\Delta G$ between the wild-type (WT) and mutant-type (MT) structures. Along with its own prediction, DynaMut also gives $\Delta\Delta G$ prediction of NMA based elastic network contact model (ENCoM) (Frappier & Najmanovich, 2014) and structure-based predictions of mCSM (Pires et al., 2014a), SDM (Worth et al., 2011), and DUET (Pires et al., 2014b). Furthermore, DynaMut predicts ENCoM based vibrational entropy difference ($\Delta\Delta S_{\text{vib}}$) of WT and MT which predicts whether the mutation will increase or decrease the flexibility of the protein. ENCoM utilizes coarse-grained normal mode analysis that calculates the $\Delta\Delta S_{\text{vib}}$ between WT and MT structures by screening the all pairwise atomic interactions. To calculate the number of total interactions in WT and MT proteins, the output protein structures from the DynaMut were uploaded to Arpeggio web server (Jubb et al., 2017).

Additionally, to validate the predictions of $\Delta\Delta G$ of WT and MT proteins calculated using DynaMut, sequence-based tools, single amino acid folding free energy changes-sequence (SAAFEC-SEQ) (Li et al., 2021) and I-Mutant2.0 (Capriotti et al., 2005) were used. SAAFEC-SEQ employs physicochemical properties at mutation position, sequence characteristics, and evolutionary data to predict $\Delta\Delta G$. However, I-Mutant2.0 employs the support vector machines (SVMs)

based predictors and predicts the sign and magnitude of $\Delta\Delta G$ at the different pH values. We used 7 pH to compute $\Delta\Delta G$ of WT and MT structures.

Discrete molecular dynamics

In order to understand the structural dynamics of WT and MT proteins, C- α discrete molecular dynamics implemented in MDWeb was used (Hospital et al., 2012). The DMD simulations of N and S proteins were performed for 500 ns at 300 K temperature. During the simulation, output frequency, sigma (square well amplitude for consecutive carbons), sigma_go (square well amplitude for non-consecutive carbons), and cut-off distance between pairs of atoms were kept fixed at 10 ps, 0.05, 0.1, and 8 Å respectively. Root mean square fluctuation (RMSF) and radius of gyration (Rg) were calculated and plotted for further investigation.

Molecular dynamics simulation

The WT and variants of novel SARS-CoV-2 protein were subjected to MD simulation using CHARMM36 (Huang et al., 2017) force field available with the GROMACS 2020 kit (Abraham et al., 2015; Berendsen et al., 1995). The protein was solvated with water molecules (spc216) and placed in the center of a dodecahedron box with a minimum of 1 nm from the box edge. Ions were added to neutralize the system by substituting water molecules. Energy minimization was performed using the steepest descent method with the Verlet cut-off scheme and particle mesh ewald (PME) long-range electrostatics. The system was then subjected to an equilibration (NVT and NPT) for 100 ps at 300 K. The final MD production runs were performed for 50 ns for each protein structure using the leapfrog integration algorithm, PME, and Verlet cut-off scheme at 300 K temperature and 1 bar pressure. Each trajectory was then subjected to further analysis of root mean square deviation (RMSD) of backbone, root mean square fluctuation (RMSF) for alpha-carbon of protein, radius of gyration (Rg) of whole protein structure, intramolecular hydrogen bonds of protein structure, and solvent accessible surface area (SASA) of entire protein to analyse the dynamic behaviour of residues and stability of proteins using inbuilt tools in GROMACS suit. Additionally, the Arpeggio web server was used to calculate the total intramolecular interactions (Jubb et al., 2017).

Results and discussion

Identification of mutations among SARS-CoV-2 circulating genomes

As the SARS-CoV-2 infections are increasing to new geographical locations, new mutations are arising among the viral genomes. In order to identify highly significant mutations among SARS-CoV-2 genomes, we utilized meta-analysis based approaches to identify novel mutations in 829 viral genomes from twelve countries (Supplementary Table S1). We observed that most of the mutations were confined to

three locations in the SARS-CoV-2 genome namely nsp3, RdRp and N proteins (Figure 1A and Supplementary Table S2). However, apart from these mutations, we identified several other mutations that were randomly distributed throughout the SARS-CoV-2 genome (Figure 1A). Next, we studied the effect of various mutations for possible amino acid substitutions. We identified nearly thirty-seven non-synonymous mutations in the genome of SARS-CoV-2 (Figure 1B-F, Supplementary Table S3). Out of these, twenty non-synonymous mutations were found in ORF1ab (open reading frame 1ab), five mutations were the part of S protein, one mutation was present in E protein and five mutations were present in N protein. In order to study the frequency of these mutations among circulating SARS-CoV-2 genomes, we calculated the percentage of genomes carrying the mutations. We observed that most of these mutations were present in >30% of genomes suggesting that these mutations are part of majority SARS-CoV-2 circulating genomes (Figure 1G).

Mutation analysis of structural proteins

Amino acid variation in proteins affects protein folding (Alfalah et al., 2009; Lorch et al., 1999, 2000), stability (Doss et al., 2012), and functioning (Yamada et al., 2006) that leads to various degenerative diseases (Blocquel et al., 2019). It has reported that around half of the total disease-related amino acid mutations arise from non-synonymous single nucleotide variants (SNVs) (Halushka et al., 1999). It is difficult to differentiate the nature of mutation using any experimental procedure due to the meteoric growth of SNVs (Capriotti et al., 2012; Tranchevent et al., 2011). To fill this gap, computational approaches are available to analyse the effects of various mutations on protein structure and function.

Hence, we studied the effect of mutations in circulating SARS-CoV-2 genomes and report the study of several mutations in envelope (E), nucleocapsid (N), and spike (S) structural proteins. The envelope protein of SARS-CoV-2 consists of 75 amino acids and three regions, N-terminal domain (NTD), transmembrane domain (TMD), and C-terminal domain (CTD). TMD is composed of 31 amino acids (a.a.: 8-38) (Mandala et al., 2020). The structures of E protein are illustrated in Figure 2A-C. N protein contains disordered regions and has several domains and flexible linkers (Peng et al., 2020) which is shown in Figure 2D. Recent report has shown that though the structure of the N protein of SARS-CoV-2 is similar to the previously reported SARS-CoV N protein, however, they differ in electrostatic properties (Kang et al., 2020). SARS-CoV-2 N protein contains two RNA binding domains, the C-terminal domain (CTD) and N-terminal domain (NTD) (Tomaszewski et al., 2020). These two domains are connected via a central linker. The terminal regions have been reported as intrinsically disordered regions (IDRs) (Wang et al., 2021). The crystal structure of N protein CTD (Figure 2E) has been solved and deposited in the PDB database (PDB: 6ZCO). The surface of the virus is decorated by the homotrimeric spike glycoproteins which looks like a crown. Each spike protomer consists of S1 and S2 domains which are separated by furin cleavage site. S1 and S2 domains have multiple regions like N-terminal domain (NTD), NTD to residue binding domain

(RBD) linker N2R, and subdomains SD1 and SD2 in S1 while, fusion peptide (FP), heptad repeat 1 (HR1), central helix (CH), connector domain (CD), heptad repeat 2 (HR2), transmembrane domain (TM), and cytoplasmic tail (CT) in S2 (Gobeil et al., 2021; Saputri et al., 2020). The crystal structure of spike protein trimer and S1 domain are shown in Figure 2F and Figure 2G respectively. Mutation sites and WT residues of three proteins (E, N and S) are shown in Figure 1I-VI. Interestingly it was observed that spike protein of SARS-CoV-2 harboured an RBD domain which was predicted to facilitate efficient binding of viral spike protein with human ACE-2 protein which acts as a receipt for this virus. SARS-CoV-2 spike protein analysis further revealed presence of a polybasic site at S1-S2 junction, a feature absent from SARS coronaviruses (Andersen et al., 2020).

Protein function and pathogenicity

We started with the L37H mutation that is located in the transmembrane domain (TMD) of the envelope protein. It is reported that this mutation has the potential to change relative solvent accessibility (Nguyen et al., 2021). Literature review suggested that the nature of this mutation has not been studied yet completely (Bianchi et al., 2020; Rahman et al., 2021; Sarkar & Saha, 2020). In this mutation, nonpolar leucine is substituted by a polar histidine residue. This mutation was predicted to be neutral at an 83% confidence score. In N protein, it has been reported that the mutations, P13L and R203K are among the top ten high-frequency SNPs (Wang et al., 2020). Moreover, P13L is present in the first disordered region, R203K and G204R are in the second disordered region, while P344S is in the CTD domain (Dasgupta, 2020). Three mutations in N protein (P13L, S197L, and G204R) were found to have a deleterious effect on protein functions whereas, the remaining two mutations, R203K and P344S were predicted to be neutral in nature. The single mutation, D614G in S protein was predicted neutral. Hence, mutations in all three proteins which are predicted to be deleterious, may alter the functions and reduce the fitness of coronavirus, while those which are found neutral do not have adverse effects on the protein function. PredictSNP results for the mutations are shown in Table 1 while the predicted results of all predictors with a confidence score of deleterious and neutral mutation for each protein are tabulated in Supplementary Table S4.

Normal mode analysis (DynaMut)

Free energy change ($\Delta\Delta G$)

To inspect the overall effects of missense mutation on protein stability, $\Delta\Delta G$ was calculated using structure and sequence based predictors tools including DynaMut, SAAFEC-SEQ, and I-Mutant2.0.

E protein

L37H mutation is part of the terminal region in a helix of the pentameric transmembrane protein. Among seven predictors

Table 1. Predicted results for the effects of mutation on functionality, stability, and flexibility of proteins.

Method Residue (WT-MT)	Mutation						
	E Protein	N Protein			S Protein		
	L37H Nonpolar- Polar	P13L Nonpolar- Nonpolar	S197L Polar- Nonpolar	R203K Polar- Polar	G204R Nonpolar- Polar	P344S Nonpolar- Polar	D614G Polar- Nonpolar
PredictSNP	83	72	64	63	61	63	63
Confidence Score (%) & Nature of Mutation	Neutral	Deleterious	Deleterious	Neutral	Deleterious	Neutral	Neutral
ENCoM $\Delta\Delta S_{vib}$ (kcal.mol ⁻¹ .K ⁻¹) & Flexibility	0.060 Increase	0.857 Increase	0.013 Increase	-4.166 Decrease	-1.340 Decrease	-0.036 Decrease	0.418 Increase
DynaMut $\Delta\Delta G$ (kcal.mol ⁻¹)	-0.029 (Destabilizing)	0.004 (Stabilizing)	0.528 (Stabilizing)	1.774 (Stabilizing)	-0.282 (Destabilizing)	0.110 (Stabilizing)	-0.316 (Destabilizing)
ENCoM $\Delta\Delta G$ (kcal.mol ⁻¹)	-0.048 (Destabilizing)	-0.686 (Destabilizing)	-0.010 (Destabilizing)	4.333 (Stabilizing)	1.072 (Stabilizing)	0.029 (Destabilizing)	-0.334 (Destabilizing)
mCSM $\Delta\Delta G$ (kcal.mol ⁻¹)	0.027 (Stabilizing)	-0.299 (Destabilizing)	-0.398 (Destabilizing)	-1.599 (Destabilizing)	-1.116 (Destabilizing)	-0.304 (Destabilizing)	-1.026 (Destabilizing)
SDM $\Delta\Delta G$ (kcal.mol ⁻¹)	-0.740 (Destabilizing)	0.840 (Stabilizing)	1.280 (Stabilizing)	-0.630 (Destabilizing)	-2.090 (Destabilizing)	-0.260 (Destabilizing)	-0.420 (Destabilizing)
DUET $\Delta\Delta G$ (kcal.mol ⁻¹)	-0.135 (Destabilizing)	0.221 (Stabilizing)	0.165 (Stabilizing)	-1.522 (Destabilizing)	-1.119 (Destabilizing)	-0.124 (Destabilizing)	-0.957 (Destabilizing)
SAAFEC-SEQ $\Delta\Delta G$ (kcal.mol ⁻¹)	-0.01 (Destabilizing)	-0.48 (Destabilizing)	-0.77 (Destabilizing)	-0.78 (Destabilizing)	-0.42 (Destabilizing)	-1.32 (Destabilizing)	-1.04 (Destabilizing)
I-Mutant2.0 $\Delta\Delta G$ (kcal.mol ⁻¹)	-1.81 (Destabilizing)	0.230 (Stabilizing)	0.290 (Stabilizing)	-2.160 (Destabilizing)	*0.000 (Destabilizing)	-0.45 (Destabilizing)	-1.94 (Destabilizing)

* $\Delta\Delta G$ value is zero but, sign of the $\Delta\Delta G$ is predicted negative hence, it is destabilizing.

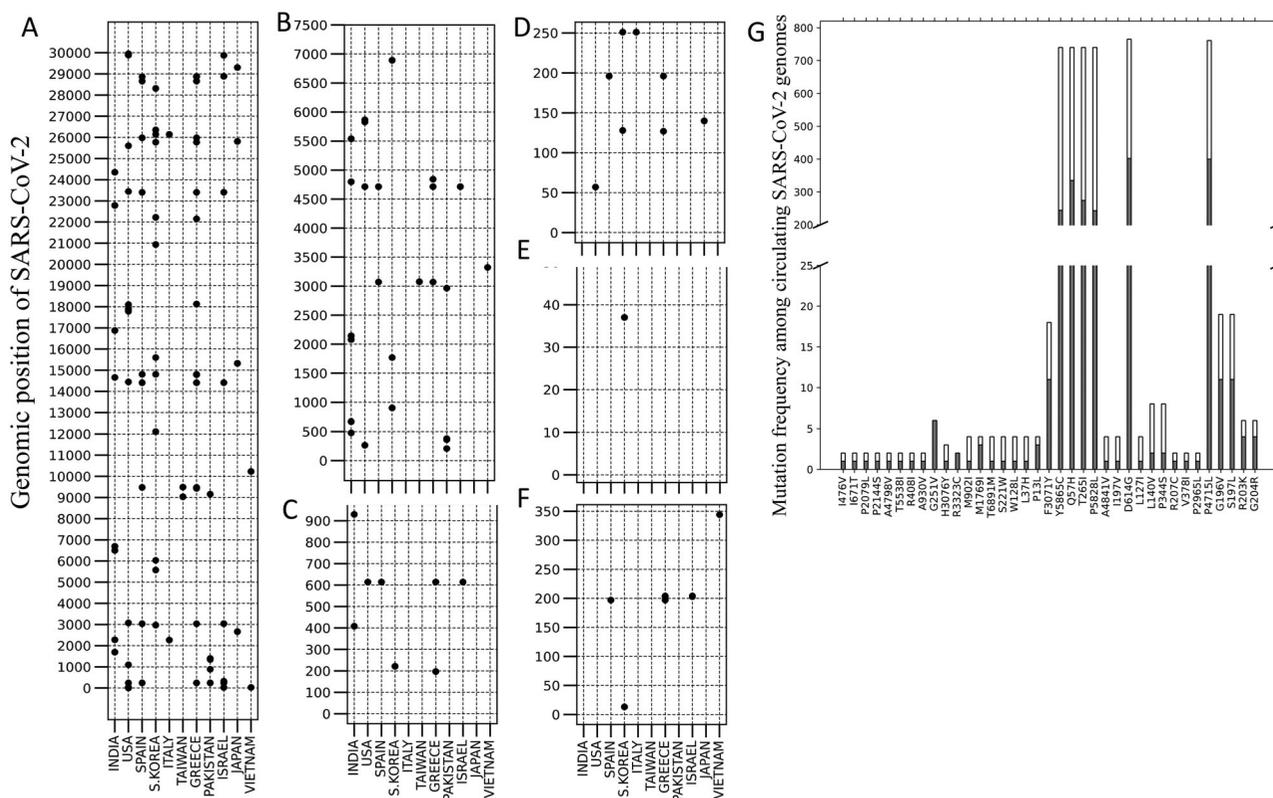


Figure 1. Identification of highly significant mutations among various SARS-CoV-2 genomes. (A) Figure showing nucleotide position of various mutations in SARS-CoV-2 genomes (B) Figure showing the non-synonymous mutations (positions in amino-acids) in ORF1ab (C) S protein (D) ORF3a protein (E) E protein and (F) N protein of SARS-CoV-2 genome (G) Figure showing the frequency of various non-synonymous mutations in SARS-CoV-2 genomes.

that we studied, six of them predicted that L37H mutation in E protein destabilizes the overall structure. I-Mutant2.0, SDM, and DUET also gave the significant magnitude of destabilization whereas the remaining predictors predicted minute destabilization for E protein. The majority of predictors predicted that the mutation results in destabilization of the protein. Looking at the interactions in WT and

MT, it can be observed that a weak hydrogen bond of Leu37 with Ile33 in WT is missing in MT. In WT, Leu37 has 33 proximal interactions at around 4-5 Å with surrounding residues, Ala32, Ile33, Leu34, Thr35, and Ala36. However, in MT these proximal interactions are reduced to 20 only. Hence, these two major factors play a significant role in destabilizing MT.

Table 2. Types and number of interactions of WT and MT residues with surrounding amino acids.

Protein	E protein			N protein (Modelled)					N protein CTD (crystal)			S protein		
	L37H_WT	L37H_MT	P13L_WT	P13L_MT	S197L_WT	S197L_MT	R203K_WT	R203K_MT	G204R_WT	G204R_MT	P344S_WT	P344S_MT	D614G_WT	D614G_MT
Type of Interactions														
VdW interactions	-	1	-	-	-	1	2	5	3	7	1	1	-	-
VdW clash interactions	2	1	-	-	1	13	12	7	7	72	-	-	1	-
Polar contacts	1	1	-	-	1	6	3	5	5	12	-	2	1	-
Weak polar contacts	1	1	-	-	1	4	4	3	3	10	-	-	1	-
Hydrogen bonds	1	1	-	-	1	3	-	5	5	11	-	-	1	-
Weak hydrogen bonds	1	-	-	-	2	2	-	-	-	2	-	-	-	1
Hydrophobic contacts	-	-	-	-	-	1	7	-	-	6	-	-	-	-
Carbonyl interactions	-	-	-	-	-	6	6	-	-	1	-	-	-	-
Proximal	33	20	12	7	35	41	172	127	80	189	14	13	35	18

N protein (modelled)

In N protein, the P13L mutation site is located at the centre of protein and at the loop region. $\Delta\Delta G$ values confirmed that P13L mutation gives stability to the MT. Although four predictors predict stabilization whereas three predict destabilization of the N protein, data from all-atom MD simulation (50 ns simulation) suggested that MT forms more number of intramolecular contacts as compared to WT thereby leading to stabilization of the N protein. The major contribution for the stability comes from proximal polar contacts. In WT, 12 proximal polar and hydrophobic contacts with Asn11, Asn48, Asn150, and Ile15 are observed but at longer distances (5-7.5 Å) as compared to seven contacts with Asn11, Asn48, and Ile15 at shorter distances (4-6 Å) in MT. So, the stability of MT protein is mainly attributed to the proximal polar and hydrophobic interactions. Our data is in agreement with another study which suggests that the P13L mutation in NTD stabilizes the N protein (Singh et al., 2020).

The second mutation, S197L is situated at a peripheral position of IDR in N protein. From the $\Delta\Delta G$ values as shown in Table 1, we can observe that the majority of the predictors predict stabilization of MT and that is mainly due to MT having 41 proximal polar contacts in comparison with WT that has 35. Additionally, in MT, Leu197 forms van der Waals bond with Ser193 and weak hydrogen bonds with Ser193 and Pro199. Hence, the overall stability is gained by polar contacts. Our In silico data is further supported by the experimental evidence which suggests that the S197L mutation has mild patient outcome ranging from 76% to 1% (Nagy et al., 2021). For the third mutation R203K, DynaMut and EnCOM based $\Delta\Delta G$ values were significantly higher, 1.774 kcal mol⁻¹ and 4.666 kcal mol⁻¹ respectively and predicted stabilization compared to the other predictors. Further analysis suggests reduced vibrational entropy in MT (-4.166 kcal mol⁻¹ K⁻¹) thereby indicating the rigidification of structure leading to reduced flexibility of MT as compared to WT. Additionally, the radius of gyration (Rg) from MD shows that the MT structure becomes compact at the end of the simulation. Another supporting evidence is the formation of new van der Waals bonds with Ala211 and Gly212 and hydrophobic contacts with Phe307, Leu339, and Asp216 in MT for the stabilization of overall structure due to R203K mutation. MD data shows the overall contacts in MT are higher compared to WT and that can be also considered supporting evidence of stabilizing effect of R203K mutation.

In the G204R mutation, all predictors predict the destabilization of mutant N protein except for ENCoM because negative free energy change is occurring after mutation. MD results show that MT has less number of intramolecular interactions compared to WT which is a good indication of the destabilizing effect of G204R mutation on N protein. Both the mutations, R203K and G204R have inferior outcomes according to the previously reported study (Nagy et al., 2021).

N protein C-terminal domain

In the crystal structure of the C-terminal domain (CTD) of N protein, P344S mutation was found to have a destabilizing

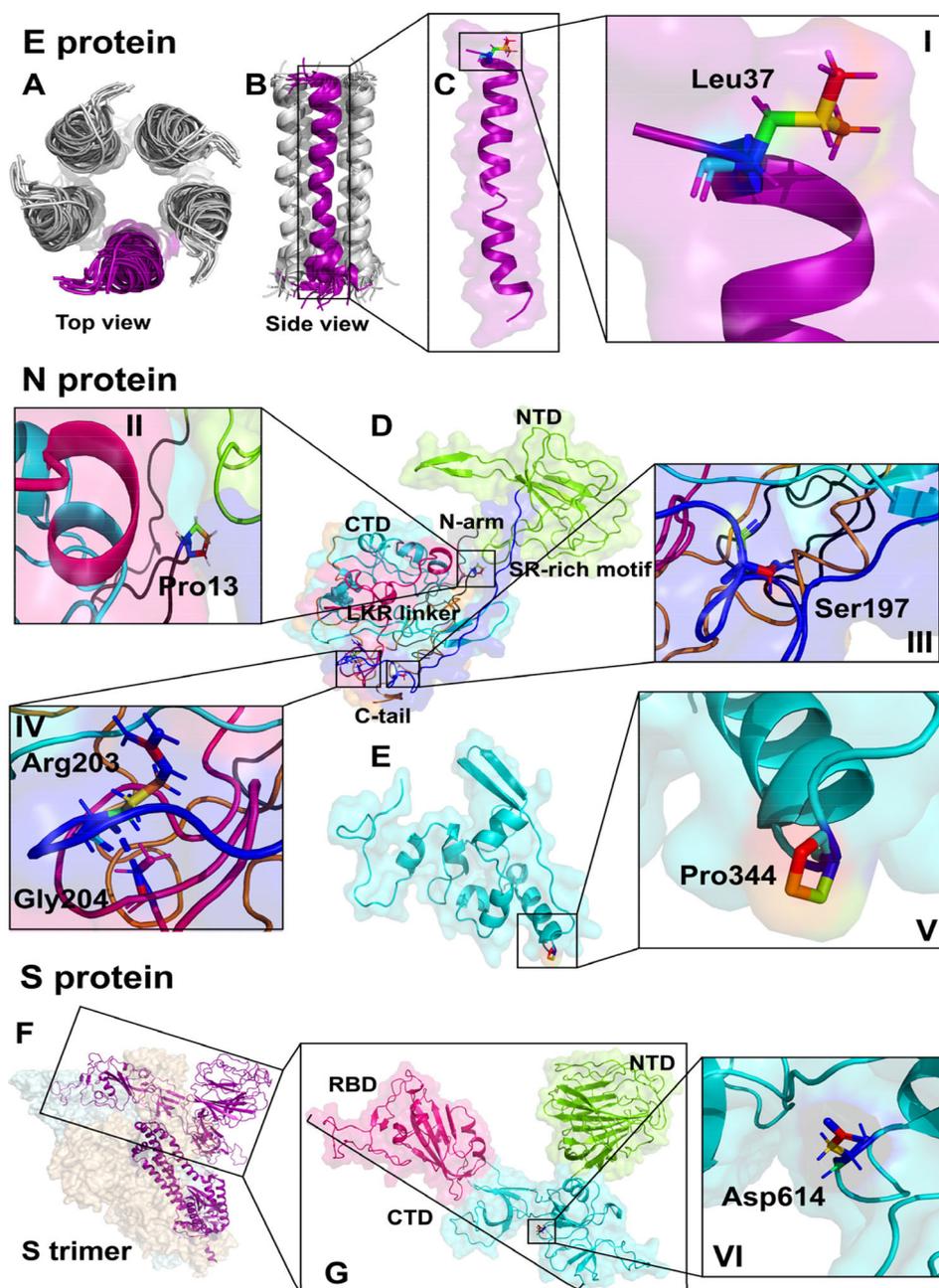


Figure 2. Protein structures and mutation sites. Alphabets (A-G) and Roman numerals (I-VI) represent the structures of protein and mutation sites respectively. (A-B) Represents the pentameric solution NMR structure of E protein (PDB: 7K3G) whereas (C) Monomeric E protein. (D-G) Illustrate the modelled N protein, experimental (crystal) C-terminal domain (PDB: 6ZCO) of N protein, trimeric S protein, and monomeric domain of S protein respectively. In N protein, D has the multiple fragments. Black (a.a.: 1-43), green (a.a.: 44-174), blue (a.a.: 175-203), pink (a.a.: 204-254), cyan (a.a.: 255-364), and orange (a.a.: 365-419) colours indicate N-arm, N-terminal domain (NTD), serine/arginine (SR) rich motif, linker region (LKR), C-terminal domain, and C-tail respectively. In S protein, G shows two different regions, green (a.a.: 27-291) for N-terminal domain (NTD) and pink (a.a.: 333-526) for residue binding domain (RBD) which is part of the C-terminal domain (a.a.: 294-700).

effect on mutant protein. Though the magnitudes of $\Delta\Delta G$ are slightly negative, however, this miniscule difference leads to destabilization of CTD. The destabilizing nature of this mutation was reported elsewhere as well (Rahman et al., 2021). Since the total number of contacts predicted by DynaMut in WT and MT are roughly similar, hence formation of contact could not be considered an evidence of destabilization. However, MD analysis revealed that the total number of interactions has been decreased in MT in comparison with WT. Furthermore, the structural features of amino acids can be considered one of the factors for the destabilization. Here, rigid and conformationally blocked proline is replaced

by flexible serine residue thereby resulting in a flexible CTD. Pro344 and Ser344 both are interacting with the same residues, Lys342, Phe346, and Lys347 in WT and MT respectively.

S protein

In spike protein, D614G mutation is one of the most widely studied mutations (Fernandez, 2020; Hou et al., 2020; Korber et al., 2020; Plante et al., 2021; Volz et al., 2021; Zhang et al., 2020). In this mutation, aspartic acid is replaced by glycine and the predicted values of free energy change show that

the mutation leads to less stable structure formation. $\Delta\Delta G$ values are negative in all the seven predictions thereby pointing towards the destabilizing effect of this mutation on parent protein. The higher stability of the WT structure is largely due to more numbers of polar contacts which are nearly double (contacts: 35) as compared to MT (contacts: 18). In WT protein, Asp614 forms approximately 40 polar contacts with Asp111, Gln115, Val120, Phe596, Tyr612, Asn616, and Cys649 residues whereas, only 16 proximal contacts are observed in MT with Gly614 with Lys97, Arg102, Trp104, Leu117, Leu118, Asn121, Tyr124, and Val130 residues. Also, a positive change in entropy indicates the highly dynamic nature of MT spike protein. Table 2 shows the number and types of interaction of mutant and wild-type residues with surrounding residues while visual representation is given in Figure 3.

Vibrational entropy change ($\Delta\Delta S_{vib}$)

To understand how the mutations affect the dynamic behaviour of various SARS-CoV-2 proteins, we further calculated vibrational entropy change between WT and MT structures of the proteins. It can be depicted from Figure 4 that mutations, L37H of E protein, P13L and P344S of N protein and D614G in S protein give flexibility to the MT structures. In E protein, L37H mutation is situated in the terminal part of the transmembrane domain and makes the whole transmembrane helix slightly flexible.

P13L mutation is located at the N-terminal IDR and substitution of proline with leucine gives the highest flexibility to the NTD as compared to the CTD with $\Delta\Delta S_{vib}$ 0.857 kcal mol⁻¹ K⁻¹. The second mutation, S197L is present at the central linker IDR which is the loop region of the protein. This mutation slightly increases the flexibility ($\Delta\Delta S_{vib}$: 0.013 kcal mol⁻¹ K⁻¹) of only the IDR region. Similarly, R203K and G204R mutations are also positioned at the central linker IDR and the loop of proteins. From the $\Delta\Delta S_{vib}$ values, it was confirmed that due to these two mutations, N protein gains the highest rigidity among all mutations. Like a P13L mutation, due to proline residue substitution in P344S mutation, MT attains flexibility but less extent at the CTD region. In S protein, the D614G mutation site is located at the loop region of S1 domain. In this mutation, charged aspartic acid is substituted by nonpolar glycine. Charged amino acids contribute to protein stability by forming electrostatic contacts or hydrogen bonds (Zhou & Pang, 2018). Hence, spike protein MT gains flexibility due to the lack of polar aspartic acid residue. Flexibility in MT is observed maximum at the CTD region and maximum at residue binding domain (RBD). The values of $\Delta\Delta S_{vib}$ of proteins are given in Table 1 whereas visual representations are shown in Figure 4.

Discrete molecular dynamics

The most problematic challenge in molecular dynamics (MD) and molecular mechanics (MM) is the demand for high computational power. To simplify and reduce the computational cost, Monte Carlo, Brownian dynamics, discrete molecular

dynamics (DMD) and their hybrids simulation techniques have been developed to study the dynamics of biomolecules (Proctor & Dokholyan, 2016). Dynamic behaviour of each amino acid in protein is closely related to the functionality of the protein. Mutation, which leads to decreased flexibility, may alter the activity of a protein (Boehr et al., 2013; Eisenmesser et al., 2005). The mutations are also known to increase flexibility and make protein less rigid (Verma et al., 2012; Wieczorek & Zielenkiewicz, 2008). Hence, to probe the impacts of mutation on flexibility, 500 ns C- α DMD was performed along with normal mode analysis (NMA). After running 500 ns DMD, we analysed trajectories and plotted RMSF and radius of gyration (Rg) for all mutations of N and a single mutation of S protein. Our results confirmed that C- α DMD results were in good agreement with the all-atom MD results.

N protein (modelled)

It was observed from the RMSF plots (Figure 5A), all the mutations in N protein give flexibility to the protein at loop regions of NTD. However, R203K and G204R mutations bring the highest flexibility in the protein whereas residues from 225 to 315 show the rigidity in the CTD region. The NTD loop (a.a.: 95-100) has the highest fluctuations (8-10 Å) among all the residues. S197L and P13L MT structures have the lowest fluctuations which shows the gaining of rigidity compared to WT protein. The radius of gyration plots (Figure 5D) show that the WT and MT structures initially open and close at the end of simulation.

N protein C-terminal domain

In CTD crystal structure, two regions, loop between two short helices (a.a.: 279-283) and beta-sheet (a.a.: 315-335) show the dynamic behaviour in WT and MT structures. In P344S MT, the values of RMSF (Figure 5B) are observed significantly greater (6-19 Å) than WT (6-13 Å) structure. It can be observed from the radius of gyration plots (Figure 5E) that the sequentially closing-opening-closing loop of MT whereas, WT closes initially and opens at the end of the simulation.

S protein

In spike protein, MT (D614G) has slightly lower values of RMSF (Figure 5C) compared to the WT structure. The radius of gyration plots (Figure 5F) indicate that the opening and closing of MT structure while WT shifts between opening and closing slightly in the middle of simulation and at the end of the simulation, it has little opening structure compared to WT structure. Residual analysis and Rg are illustrated in Figure 5.

Molecular dynamics simulation

To investigate the effects of point mutation on the stability and dynamics of S protein and N protein, the wild-type (WT) and mutant structure of S protein (WT, and D614G), and modelled N protein (WT, P13L, S197L, R203K, and G204R)

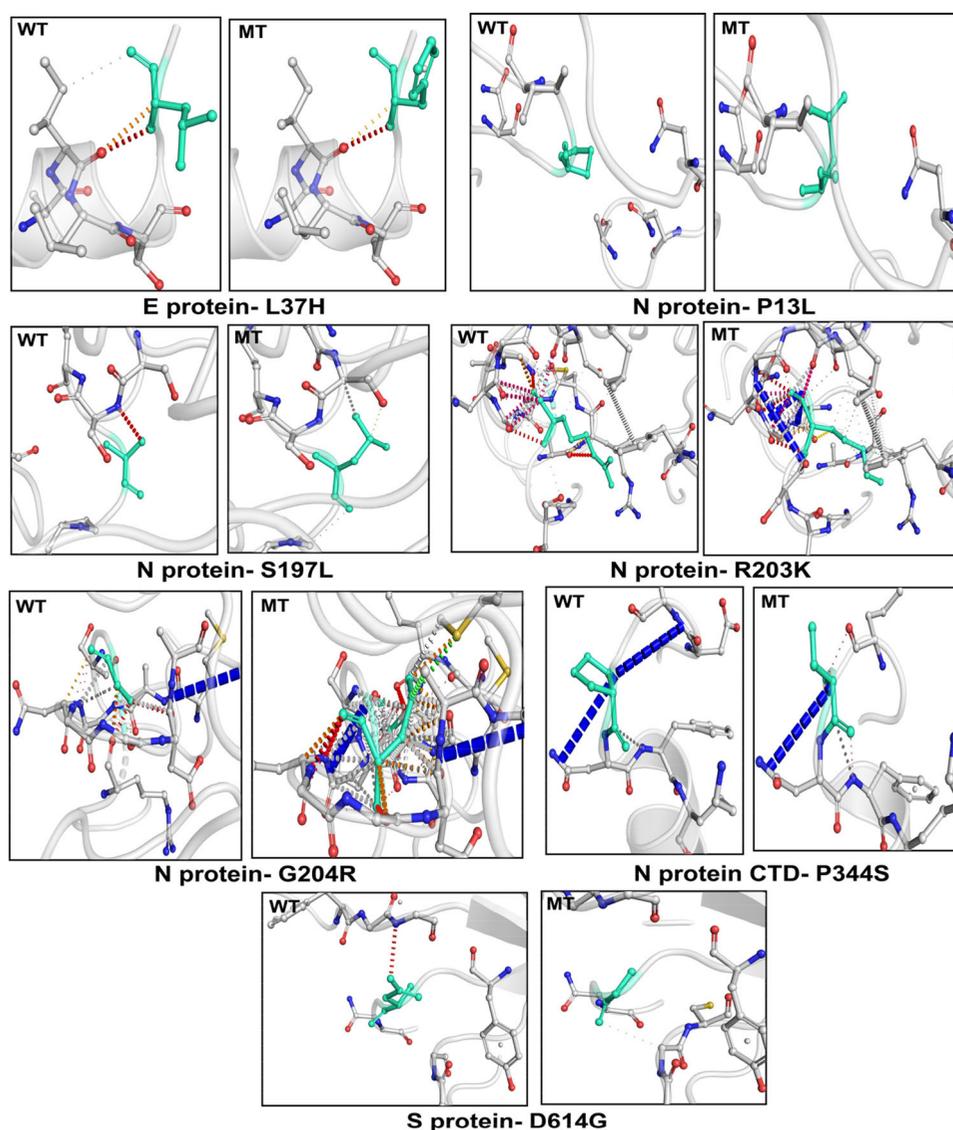


Figure 3. Intramolecular interactions of WT and MT residues with proximal amino acids. WT and MT residues are shown in cyan colour. Interactions are represented in different colours and for the further interpretation of interactions, readers are requested to visit the Web version of Arpeggio tool.

were subjected to 50 ns MD simulation. Additionally, we also simulated the available crystal structure of N protein CTD and its mutant P344S for 50 ns for the same. Furthermore, we calculated the total number of intramolecular interactions using the Arpeggio webserver and the results are summarized in [Supplementary Table S5](#) which will indicate the structural stability of protein variants.

N protein (modelled)

To examine the change in the protein dynamics and stability of WT and mutant N protein, we have performed 50 ns simulation of models of WT and mutant of N protein. Since the partial crystal structure of P344s mutant was available, we simulated that crystal structure and compared it with the corresponding WT model created based on crystal structure. The structure of 50 ns frames of WT and mutants along with the initial structure are presented in [Figure 6A](#), SASA and Rg plots of WT and mutant are presented in [Figure 6B](#), and 6C

respectively, RMSD, Hydrogen bonds, and RMSF are presented in [Supplementary Figure S1](#).

The Rg and SASA values of mutant P13L and S197L were observed to be lower as compared to WT and other mutants which indicates that the structure of these mutants are compact compared to WT and other mutants. Compaction of P13L and S197L mutant structures lead to significant increase in total number of contacts especially hydrogen bonds, polar and aromatics contacts as compared to WT and is indicated in [Supplementary Table S5](#). These results suggest that P13L and S197L mutation is possibly stabilizing the structure leading to compactness of protein N structure. Contrastingly, the Rg of R203K appears to have lower values compared to WT in the entire trajectory, while the SASA of mutant appears to have a lower profile than WT. The total number of intramolecular contacts is also higher in mutant R203K which suggests that at 50 ns, R203K adopts a stable structure compared to WT. However, Rg and SASA of G204R appear to have similar profiles like WT. Further the observed slight increase in hydrogen bonding and polar contacts compared to WT is compensated with

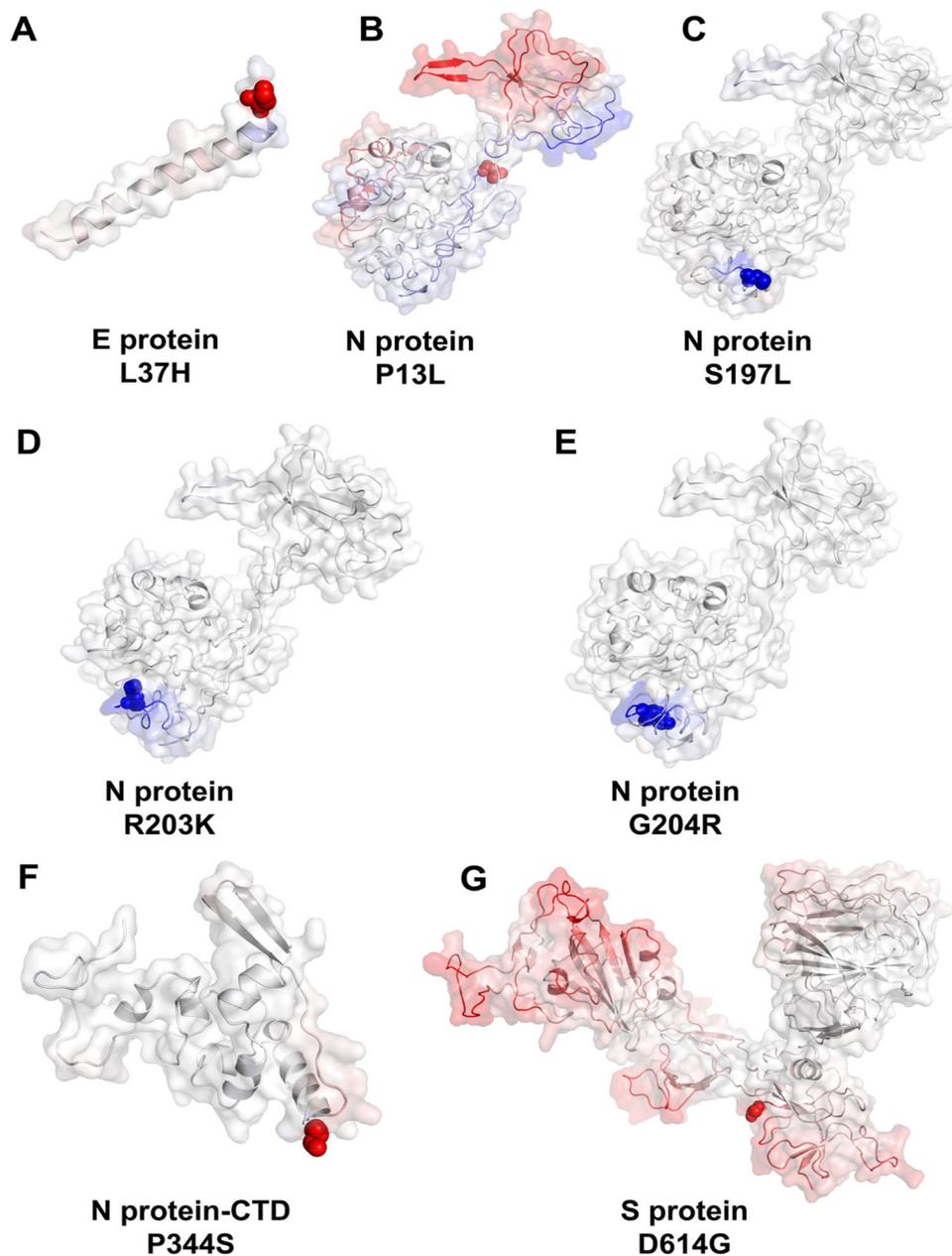


Figure 4. Visual representation of MT protein dynamics. Mutations are shown in sphere representation. Red and blue colours indicate the flexibility and rigidity respectively.

slight loss in total number of aromatic and non-polar contacts compared to WT structure populated at 50 ns. This suggests that WT and G204R mutants adopt similar compact structures during simulation. We observed the RMSD of WT and all the mutants are slightly different in the first half of simulation, but in the second half of simulation all the variants and WT display similar behaviour, except R203K which continuously have slightly higher values compared to all other variants. Furthermore, the RMSF of all the mutants are similar except for R203K and G204R mutants which shows higher continuous fluctuations compared to WT and other two mutants suggesting that R203K and G204R mutations had an effect on the flexibility of the protein structure throughout the simulations. Based on these data, it appears that R203K and G204R structures are more dynamic compared to WT and other two mutants.

In summary, mutations P13L and S197L appear to cause structural stabilization leading to an increase in compactness of structure as well as increase in polar and non-polar intramolecular interactions compared to WT and other two mutants. However, G204R appears to have similar compactness and less intramolecular contacts as of WT, but is more dynamic, therefore G204R mutations could be destabilizing the protein N structure. Based on simulation data only, it is difficult to predict whether the behaviour of R203K is stabilizing or destabilizing.

N protein C-terminal domain

A small segment (230-364 amino acids, C-terminal domain) of N protein has been experimentally crystallized (PDB:

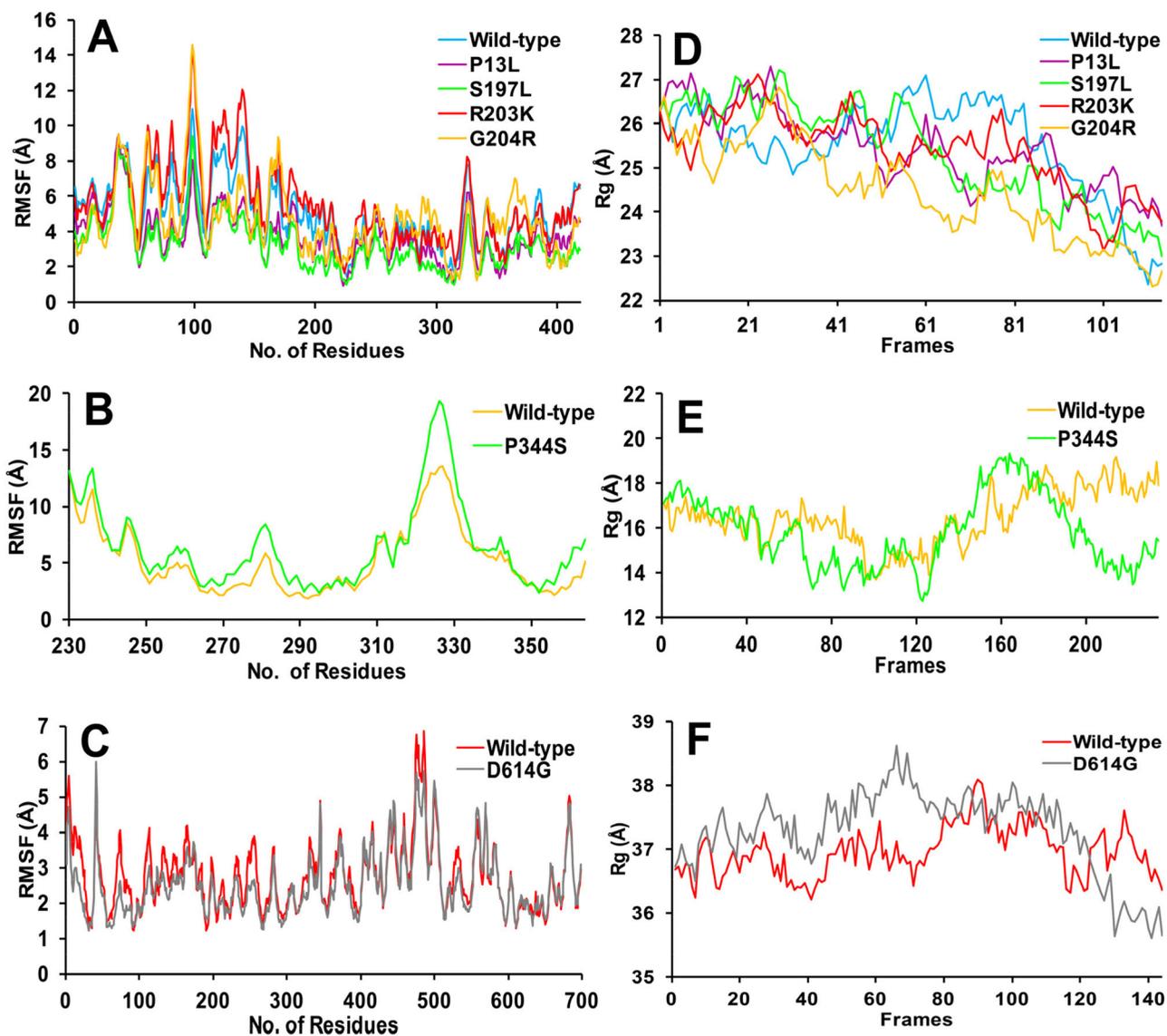


Figure 5. RMSF and Rg plot of WT and MT proteins. (A-C) RMSF of modelled N protein, CTD of N protein, and S protein respectively. (D-F) Rg plot of modelled N protein, CTD of N protein, and S protein respectively.

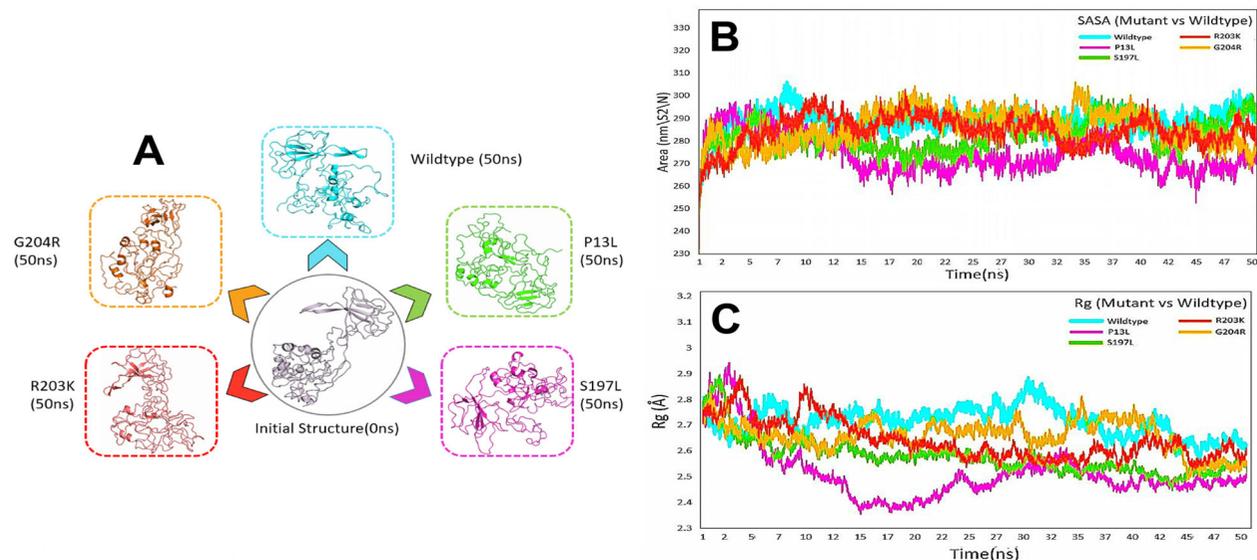


Figure 6. Comparative analysis for WT and various mutants of N protein. (A) Conformations of WT (cyan), P13L (green), S197L (pink), R203K (red), G204R (orange) populated at 50 ns timestep. (B) SASA plot of WT and their mutants against time. (C) Rg plot of WT and mutants showing compactness.

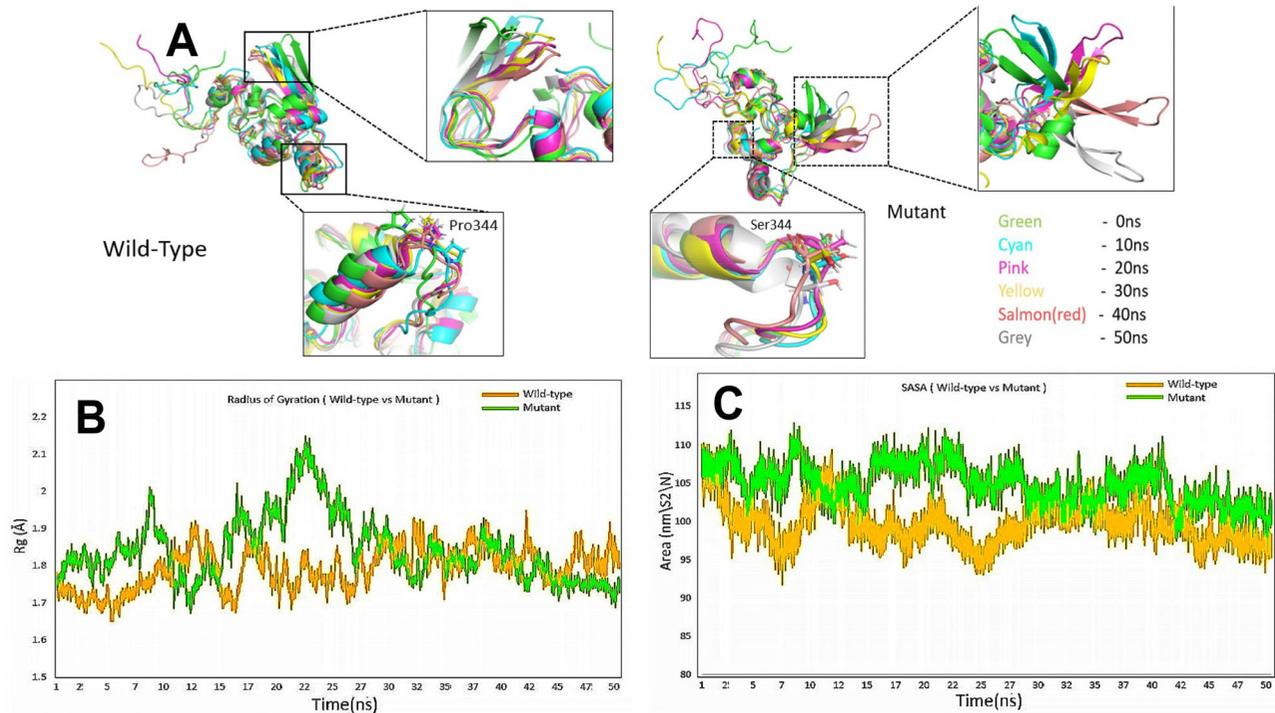


Figure 7. Stability analysis of crystallized N protein CTD and its mutant (P344S). (A) Showing dynamics in the structure of WT and P344S mutant at different intervals of time. (B) Rg plot of WT (green) and P344S mutant (yellow) displays their compactness. (C) SASA plot of variants of CTD.

6ZCO) recently, so we utilized that opportunity to investigate the effect of mutations P344S on the stability and dynamics of protein (Zinzula et al., 2021). The Rg and SASA values, described (Figure 7B and C), indicate that mutation leads to decrease in globularity of protein structure and statistically have less number of hydrogen bonds compared to WT. The mutants appear to have continuously higher fluctuations in RMSF plots (Supplementary Figure S2), indicating that the mutations had an effect on the flexibility of protein. To further investigate the structural basis of flexibility, we overlaid the structure populated at the interval of every 10 ns during the simulation of N protein WT and mutant and we observed that the mutation of proline at 344 lead to structural fluctuations at remote site and β -sheet region (a.a.: 318-334) changes its orientations dramatically during the simulation (Figure 7A). The results from the present work confirm that P344S mutations cause the structural destabilization of N protein leading to loss of protein compactness.

S protein

For MD simulation, we have taken the structural component (a.a.: 1-698) from the experimentally solved S protein (PDB: 6VYB). The initial structure (0 ns) and final structure (50 ns) of S protein variants are presented in Figure 8A. The calculated SASA, Rg, and hydrogen bonds were plotted against time of simulation and are presented in Figure 8B–D. It appears that Rg and SASA of WT is lower as compared to D614G mutant whereas the total number of intramolecular hydrogen bonds, calculated for the entire trajectory, are high in number in WT as compared to mutant. These results further provide supportive evidence that the structure of WT is compact

compared to D614G mutant. The observed high value of RMSD (Supplementary Figure S3A) for WT structure indicates the conformational dynamics associated with structural transition of open state (initial model) to compact state during the course of simulation. However, initial model structure for mutant protein is opening further only slightly, during the course of simulation thereby resulting in less deviation as compared to WT structure.

The dynamic behaviour of individual amino acid residues for WT and mutant was determined in terms of RMSF values, denoted by the peak elevation in Supplementary Figure S3B. RMSF plot indicated similar residue fluctuation profile for WT and mutant. We have highlighted the dynamic region in the model structure according to RMSF value (Supplementary Figure S3C). In conclusion, WT S protein was observed to adopt a compact fold in the latter half of the trajectory via a large conformational change in a process to adopt compact state during simulation from initial structure. However, the mutant was observed to adopt more open conformation compared to WT. Calculated value of total number of intramolecular interactions (Supplementary Table S5), further supports destabilizing effects of mutation on protein structure, because mutant protein structure at 50 ns comprises less number of intramolecular interactions compared to WT. The present results suggest that WT is stable and mutation D614G caused a destabilization leading to a loss of protein compactness. Previous study reported that the open conformation of spike protein increases the infectivity of virus (Wrapp et al., 2020). In our study, MD data shows that the mutant structure has open conformation at the end of simulation compared to wild-type structure and hence it was predicted destabilizing and shows the higher infectivity.

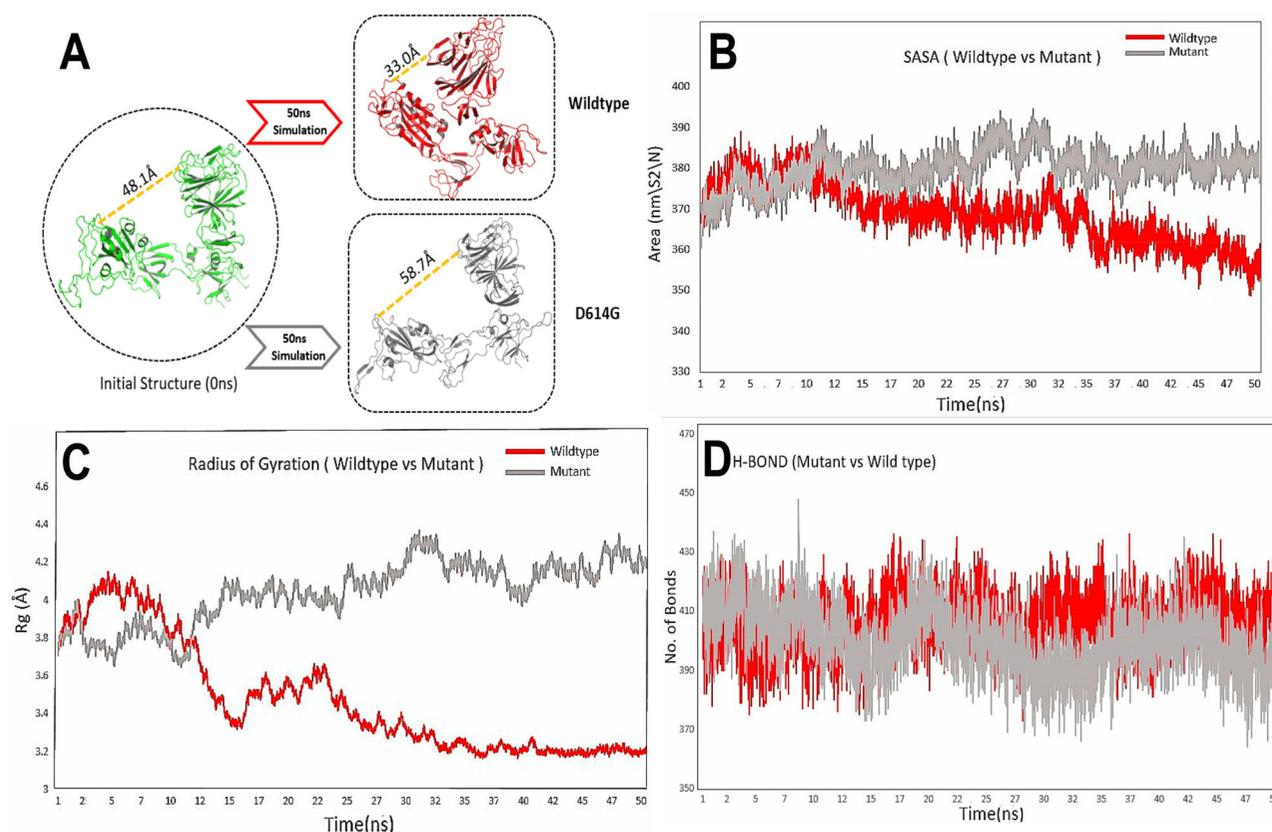


Figure 8. Comparative analysis of WT and D614S (mutant) of S protein. (A) Opening and closing states of WT and mutant during the course of 50 ns simulation. (B) SASA of WT (red) and D614S (grey). (C) Rg plot of WT and D614S. (D) Intramolecular Hydrogen bond calculation throughout the 50 ns simulation.

Conclusion

Since the first case reported on 1st December 2019 from China, SARS-CoV-2 has infected nearly all the countries throughout the globe. This alarming rate of spread of SARS-CoV-2 has prompted many countries to declare a complete shutdown, an event that was not witnessed till now. However, as the scientists from around the world are working hard to develop antivirals, the virus is spreading and accumulating mutations which might help it to combat host immune responses. Hence, it becomes imperative to gain understanding into viral evolution for the development of effective antivirals. In this study, we performed meta-analysis of SARS-CoV-2 genomes from around the world to identify highly significant mutations in the viral structural proteins. Our analysis identified sixty-one highly significant mutations spanning complete viral genomes. Further analysis revealed that most of these mutations were concentrated around few regions of viral genomes including NS3, RdRp and N protein. Since integrity of structural proteins is vital for immune responses, hence we sought to study the putative effect of highly significant mutations on viral structural proteins. Meta-CATS algorithm identified one mutation in spike protein, five mutations in N protein as well as one mutation in envelope protein. We then performed molecular simulations on these mutations to predict their effect on stability of parent protein. Among seven mutations in three proteins (E, N and S), four mutations (L13H in E, R203K and P344S in N and D614G in S) were predicted neutral while, remaining three

(P13L, S197L and G204R) found to have deleterious effects on protein functions. Discrete molecular dynamics and all-atom molecular dynamics results revealed that three mutations (P13L, S197L and R203K) in N protein give stability to the parent proteins. However, two mutations (G204R and P344S) in N protein and D614G in S protein destabilize the parent proteins. Though our analysis of D614G mutation in Spike protein predicted destabilization of the protein, several experimental studies have shown that the mutation contributes towards enhanced infectivity of the virus. D614G mutation opens up the spike protein which though might lead to destabilization of protein but might lead to enhanced viral infections. Hence, it seems that there is a trade-off between the stabilization of protein structure versus the infectivity of the virus. Therefore, our studies provide vital clues about the effect of highly significant mutations on the structure of viral structural proteins.

Acknowledgements

NP is thankful to UGC for PhD fellowship and PS is thankful to CSIR for PhD fellowship. VS received research grant from UGC, Govt. of India. SBR is thankful to his Chemistry Department for providing computational and infrastructure facilities.

Author contributions

NP, SBR, RP, PS, LN performed the experiments and analysed the data. RPB, PA, KRS and VS designed and supervised the study. VS and SBR

wrote the first draft of the manuscript. VS, RPB, SBR, PA and KRS reviewed the data and finalized the manuscript.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by grants from Department of Science and Technology (DST), India and University Grants Commission (UGC), India.

ORCID

Shravan B. Rathod  <http://orcid.org/0000-0002-1870-2508>

Ravi P. Barnwal  <http://orcid.org/0000-0003-3156-5357>

Vikas Sood  <http://orcid.org/0000-0001-6128-4279>

References

- Abraham, M. J., Murtola, T., Schulz, R., Páll, S., Smith, J. C., Hess, B., & Lindahl, E. (2015). Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*, 1–2, 19–25. <https://doi.org/10.1016/j.softx.2015.06.001>
- Alfalal, M., Keiser, M., Leeb, T., Zimmer, K. P., & Naim, H. Y. (2009). Compound heterozygous mutations affect protein folding and function in patients with congenital sucrase-isomaltase deficiency. *Gastroenterology*, 136(3), 883–892. <https://doi.org/10.1053/j.gastro.2008.11.038>
- Andersen, K. G., Rambaut, A., Lipkin, W. I., Holmes, E. C., & Garry, R. F. (2020). The proximal origin of SARS-CoV-2. *Nature Medicine*, 26(4), 450–452. <https://doi.org/10.1038/s41591-020-0820-9>
- Bendl, J., Stourac, J., Salanda, O., Pavelka, A., Wieben, E. D., Zundulka, J., Brezovsky, J., & Damborsky, J. (2014). PredictSNP: Robust and accurate consensus classifier for prediction of disease-related mutations. *PLoS Computational Biology*, 10(1), e1003440–11. <https://doi.org/10.1371/journal.pcbi.1003440>
- Berendsen, H. J. C., van der Spoel, D., & van Drunen, R. (1995). GROMACS: A message-passing parallel molecular dynamics implementation. *Computer Physics Communications*, 91(1–3), 43–56. [https://doi.org/10.1016/0010-4655\(95\)00042-E](https://doi.org/10.1016/0010-4655(95)00042-E)
- Bhattacharyya, C., Das, C., Ghosh, A., Singh, A. K., Nukherjee, S., Majumder, P. P., Basu, A., & Biswas, N. K. (2020). Global spread of SARS-CoV-2 subtype with spike protein mutation D614G is shaped by human genomic variations that regulate expression of TMPRSS2 and MX1 genes. *bioRxiv*. <https://doi.org/10.1101/2020.05.04.075911>
- Bianchi, M., Benvenuto, D., Giovanetti, M., Angeletti, S., Ciccozzi, M., & Pascarella, S. (2020). Sars-CoV-2 envelope and membrane proteins: structural differences linked to virus characteristics? *BioMed Research International*, 2020, 4389089. <https://doi.org/10.1155/2020/4389089>
- Blocquel, D., Sun, L., Matuszek, Z., Li, S., Weber, T., Kuhle, B., Kooi, G., Wei, N., Baets, J., Pan, T., Schimmel, P., & Yang, X. L. (2019). CMT disease severity correlates with mutation-induced open conformation of histidyl-tRNA synthetase, not aminoacylation loss, in patient cells. *Proceedings of the National Academy of Sciences of the United States of America*, 116(39), 19440–19448. <https://doi.org/10.1073/pnas.1908288116>
- Boehr, D. D., Schnell, J. R., McElheny, D., Bae, S. H., Duggan, B. M., Benkovic, S. J., Dyson, H. J., & Wright, P. E. (2013). A distal mutation perturbs dynamic amino acid networks in dihydrofolate reductase. *Biochemistry*, 52(27), 4605–4619. <https://doi.org/10.1021/bi400563c>
- Brooks, B. R., Brucoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S., & Karplus, M. (1983). CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *Journal of Computational Chemistry*, 4(2), 187–217. <https://doi.org/10.1002/jcc.540040211>
- Capriotti, E., Fariselli, P., & Casadio, R. (2005). I-Mutant2.0: Predicting stability changes upon mutation from the protein sequence or structure. *Nucleic Acids Research*, 33(suppl_2), W306–310. <https://doi.org/10.1093/nar/gki375>
- Capriotti, E., Nehrt, N. L., Kann, M. G., & Bromberg, Y. (2012). Bioinformatics for personal genome interpretation. *Brief Bioinform*, 13(4), 495–512. <https://doi.org/10.1093/bib/bbr070>
- Dasgupta, R. (2020). Mutations in structural proteins of SARS-CoV-2 and potential implications for the ongoing outbreak of infection in India. *AJIR Preprint*. <https://doi.org/10.21467/preprints.202>
- Doss, C. G. P., Rajith, B., Garwasis, N., Mathew, P. R., Raju, A. S., Apoorva, K., William, D., Sadhana, N. R., Himani, T., & Dike, I. P. (2012). Screening of mutations affecting protein stability and dynamics of FGFR1-A simulation analysis. *Applied & Translational Genomics*, 1, 37–43. <https://doi.org/10.1016/j.atg.2012.06.002>
- Eisenmesser, E. Z., Millet, O., Labeikovsky, W., Korzhnev, D. M., Wolf-Watz, M., Bosco, D. A., Skalicky, J. J., Kay, L. E., & Kern, D. (2005). Intrinsic dynamics of an enzyme underlies catalysis. *Nature*, 438(7064), 117–121. <https://doi.org/10.1038/nature04105>
- Fernandez, A. (2020). Structural impact of mutation D614G in SARS-CoV-2 spike protein: Enhanced infectivity and therapeutic opportunity. *ACS Medicinal Chemistry Letters*, 11(9), 1667–1670. <https://doi.org/10.1021/acsmchemlett.0c00410>
- Fiser, A., & Sali, A. (2003). ModLoop: Automated modeling of loops in protein structures. *Bioinformatics (Oxford, England)*, 19(18), 2500–2501. <https://doi.org/10.1093/bioinformatics/btg362>
- Forster, P., Forster, L., Renfrew, C., & Forster, M. (2020). Phylogenetic network analysis of SARS-CoV-2 genomes. *Proceedings of the National Academy of Sciences of the United States of America*, 117(17), 9241–9243. <https://doi.org/10.1073/pnas.2004999117>
- Frappier, V., & Najmanovich, R. J. (2014). A coarse-grained elastic network atom contact model and its use in the simulation of protein dynamics and the prediction of the effect of mutations. *PLoS Computational Biology*, 10(4), e1003569. <https://doi.org/10.1371/journal.pcbi.1003569>
- Gobeil, S. M. C., Janowska, K., McDowell, S., Mansouri, K., Parks, R., Manne, K., Stalls, V., Kopp, M. F., Henderson, R., Edwards, R. J., Haynes, B. F., & Acharya, P. (2021). D614G mutation alters SARS-CoV-2 spike conformation and enhances protease cleavage at the S1/S2 junction. *Cell Reports*, 34(2), 108630. <https://doi.org/10.1016/j.celrep.2020.108630>
- Halushka, M. K., Fan, J. B., Bentley, K., Hsie, L., Shen, N., Weder, A., Cooper, R., Lipshutz, R., & Chakravarti, A. (1999). Patterns of single-nucleotide polymorphisms in candidate genes for blood-pressure homeostasis. *Nature Genetics*, 22(3), 239–247. <https://doi.org/10.1038/10297>
- Hospital, A., Andrio, P., Fenollosa, C., Cicin-Sain, D., Orozco, M., & Gelpí, J. L. (2012). MDWeb and MDMoby: An integrated web-based platform for molecular dynamics simulations. *Bioinformatics (Oxford, England)*, 28(9), 1278–1279. <https://doi.org/10.1093/bioinformatics/bts139>
- Hou, Y. J., Chiba, S., Halfmann, P., Ehre, C., Kuroda, M., Dinnon, K. H., Leist, S. R., Schäfer, A., Nakajima, N., Takahashi, K., Lee, R. E., Mascenik, T. M., Graham, R., Edwards, C. E., Tse, L. V., Okuda, K., Markmann, A. J., Bartelt, L., de Silva, A., ... Baric, R. S. (2020). SARS-CoV-2 D614G variant exhibits efficient replication ex vivo and transmission in vivo. *Science (New York, N.Y.)*, 370(6523), 1464–1468. <https://doi.org/10.1126/science.abe8499>
- Huang, J., Rauscher, S., Nawrocki, G., Ran, T., Feig, M., De Groot, B. L., Grubmüller, H., & MacKerell, A. D. (2017). CHARMM36m: An improved force field for folded and intrinsically disordered proteins. *Nature Methods*, 14(1), 71–73. <https://doi.org/10.1038/nmeth.4067>
- Jubb, H. C., Higuero, A. P., Ochoa-Montano, B., Pitt, W. R., Ascher, D. B., & Blundell, T. L. (2017). Arpeggio: A web server for calculating and visualising interatomic interactions in protein structures. *Journal of Molecular Biology*, 429(3), 365–371. <https://doi.org/10.1016/j.jmb.2016.12.004>
- Kang, S., Yang, M., Hong, Z., Zhang, L., Huang, Z., Chen, X., He, S., Zhou, Z., Zhou, Z., Chen, Q., Yan, Y., Zhang, C., Shan, H., & Chen, S. (2020). Crystal structure of SARS-CoV-2 nucleocapsid protein RNA binding domain reveals potential unique drug targeting sites. *Acta Pharmaceutica Sinica B*, 10(7), 1228–1238. <https://doi.org/10.1016/j.apsb.2020.04.009>
- Korber, B., Fischer, W. M., Gnanakaran, S., Yoon, H., Theiler, J., Abfalterer, W., Hengartner, N., Giorgi, E. E., Bhattacharya, T., Foley, B., Hastie, K. M., Parker, M. D., Partridge, D. G., Evans, C. M., Freeman, T. M., de

- Silva, T. I., McDanal, C., Perez, L. G., Tang, H., ... Montefiori, D. C. (2020). Tracking changes in SARS-CoV-2 spike: Evidence that D614G increases infectivity of the COVID-19 virus. *Cell*, *182*(4), 812–827. <https://doi.org/10.1016/j.cell.2020.06.043>
- Laamarti, M., Alouane, T., Kartti, S., Chemao-Elfihri, M. W., Hakmi, M., Essabbar, A., Laamarti, M., Hlali, H., Bendani, H., Boumajdi, N., Benhrif, O., Allam, L., El Hafidi, N., El Jaoudi, R., Allali, L., Marchoudi, N., Fekkek, J., Benrahma, H., Nejari, C., ... Ibrahim, A. (2020). Large scale genomic analysis of 3067 SARS-CoV-2 genomes reveals a clonal geo-distribution and a rich genetic variations of hotspots mutations. *PLoS One*, *15*(11), e0240345. <https://doi.org/10.1371/journal.pone.0240345>
- Li, G., Panday, S. K., & Alexov, E. (2021). SAAFEC-SEQ: A sequence-based method for predicting the effect of single point mutations on protein thermodynamic stability. *International Journal of Molecular Sciences*, *22*(2), 606–616. <https://doi.org/10.3390/ijms22020606>
- Lorch, M., Mason, J. M., Clarke, A. R., & Parker, M. J. (1999). Effects of core mutations on the folding of a beta-sheet protein: Implications for backbone organization in the I-state. *Biochemistry*, *38*(4), 1377–1385. <https://doi.org/10.1021/bi9817820>
- Lorch, M., Mason, J. M., Sessions, R. B., & Clarke, A. R. (2000). Effects of mutations on the thermodynamics of a protein folding reaction: Implications for the mechanism of formation of the intermediate and transition states. *Biochemistry*, *39*(12), 3480–3485. <https://doi.org/10.1021/bi9923510>
- Mandala, V. S., McKay, M. J., Shcherbakov, A. A., Dregni, A. J., Kolocouris, A., & Hong, M. (2020). Structure and drug binding of the SARS-CoV-2 envelope protein transmembrane domain in lipid bilayers. *Nature Structural & Molecular Biology*, *27*(12), 1202–1208. <https://doi.org/10.1038/s41594-020-00536-8>
- Memish, Z. A., Perlman, S., Van Kerkhove, M. D., & Zumla, A. (2020). Middle East respiratory syndrome. *The Lancet*, *395*(10229), 1063–1077. [https://doi.org/10.1016/S0140-6736\(19\)33221-0](https://doi.org/10.1016/S0140-6736(19)33221-0)
- Nagy, Á., Pongor, S., & Gyórfy, B. (2021). Different mutations in SARS-CoV-2 associate with severe and mild outcome. *International Journal of Antimicrobial Agents*, *57*(2), 106272. <https://doi.org/10.1016/j.ijantimicag.2020.106272>
- Nguyen, T. T., Pathirana, P. N., Nguyen, T., Nguyen, Q. V. H., Bhatti, A., Nguyen, D. C., Nguyen, D. T., Nguyen, N. D., Creighton, D., & Abdelrazek, M. (2021). Genomic mutations and changes in protein secondary structure and solvent accessibility of SARS-CoV-2 (COVID-19 virus). *Scientific Reports*, *11*(1), 3487. <https://doi.org/10.1038/s41598-021-83105-3>
- Peck, K. M., & Luring, A. S. (2018). Complexities of viral mutation rates. *Journal of Virology*, *92*(14), 1–8. <https://doi.org/10.1128/JVI.01031-17>
- Peng, Y., Du, N., Lei, Y., Dorje, S., Qi, J., Luo, T., Gao, G. F., & Song, H. (2020). Structures of the SARS-CoV-2 nucleocapsid and their perspectives for drug design. *The EMBO Journal*, *39*(20), 1–12. <https://doi.org/10.15252/embj.2020105938>
- Phan, T. (2020). Genetic diversity and evolution of SARS-CoV-2. *Infection, Genetics and Evolution*, *81*, 104260. <https://doi.org/10.1016/j.meegid.2020.104260>
- Pickett, B. E., Liu, M., Sadat, E. L., Squires, R. B., Noronha, J. M., He, S., Jen, W., Zaremba, S., Gu, Z., Zhou, L., Larsen, C. N., Bosch, I., Gehrke, L., McGee, M., Klem, E. B., & Scheuermann, R. H. (2013). Metadata-driven comparative analysis tool for sequences (meta-CATS): An automated process for identifying significant sequence variations that correlate with virus attributes. *Virology*, *447*(1–2), 45–51. <https://doi.org/10.1016/j.virol.2013.08.021>
- Pickett, B. E., Sadat, E. L., Zhang, Y., Noronha, J. M., Squires, R. B., Hunt, V., Liu, M., Kumar, S., Zaremba, S., Gu, Z., Zhou, L., Larson, C. N., Dietrich, J., Klem, E. B., & Scheuermann, R. H. (2012). ViPR: An open bioinformatics database and analysis resource for virology research. *Nucleic Acids Research*, *40*(D1), D593–598. <https://doi.org/10.1093/nar/gkr859>
- Pires, D. E. V., Ascher, D. B., & Blundell, T. L. (2014a). DUET: A server for predicting effects of mutations on protein stability using an integrated computational approach. *Nucleic Acids Research*, *42*(W1), W314–319. <https://doi.org/10.1093/nar/gku411>
- Pires, D. E. V., Ascher, D. B., & Blundell, T. L. (2014b). MCSM: Predicting the effects of mutations in proteins using graph-based signatures. *Bioinformatics (Oxford, England)*, *30*(3), 335–342. <https://doi.org/10.1093/bioinformatics/btt691>
- Plante, J. A., Liu, Y., Liu, J., Xia, H., Johnson, B. A., Lokugamage, K. G., Zhang, X., Muruato, A. E., Zou, J., Fontes-Garfias, C. R., Mirchandani, D., Scharon, D., Bilello, J. P., Ku, Z., An, Z., Kalveram, B., Freiberg, A. N., Menachery, V. D., Xie, X., ... Shi, P. Y. (2021). Spike mutation D614G alters SARS-CoV-2 fitness. *Nature*, *592*, 116–121. <https://doi.org/10.1038/s41586-020-2895-3>
- Proctor, E. A., & Dokholyan, N. V. (2016). Applications of discrete molecular dynamics in biology and medicine. *Current Opinion in Structural Biology*, *37*, 9–13. <https://doi.org/10.1016/j.sbi.2015.11.001>
- Rahman, M. S., Hoque, M. N., Islam, M. R., Islam, I., Mishu, I. D., Rahaman, M. M., Sultana, M., & Hossain, M. A. (2021). Mutational insights into the envelope protein of SARS-CoV-2. *Gene Reports*, *22*, 100997. <https://doi.org/10.1016/j.genrep.2020.100997>
- Rahman, M. S., Islam, M. R., Alam, A. S. M. R. U., Islam, I., Hoque, M. N., Akter, S., Rahaman, M. M., Sultana, M., & Hossain, M. A. (2021). Evolutionary dynamics of SARS-CoV-2 nucleocapsid protein and its consequences. *Journal of Medical Virology*, *93*(4), 2177–2119. <https://doi.org/10.1002/jmv.26626>
- Ramachandran, S., Kota, P., Ding, F., & Dokholyan, N. V. (2011). Automated minimization of steric clashes in protein structures. *Proteins*, *79*(1), 261–270. <https://doi.org/10.1002/prot.22879>
- Rodrigues, C. H. M., Pires, D. E. V., & Ascher, D. B. (2018). DynaMut: Predicting the impact of mutations on protein conformation, flexibility and stability. *Nucleic Acids Research*, *46*(W1), W350–W355. <https://doi.org/10.1093/nar/gky300>
- Saputri, D. S., Li, S., van Eerden, F. J., Rozewicki, J., Xu, Z., Ismanto, H. S., Davila, A., Teraguchi, S., Katoh, K., & Standley, D. M. (2020). Flexible, functional, and familiar: Characteristics of SARS-CoV-2 spike protein evolution. *Frontiers in Microbiology*, *11*(September), 2112–2116. <https://doi.org/10.3389/fmicb.2020.02112>
- Sarkar, M., & Saha, S. (2020). Structural insight into the role of novel SARS-CoV-2 E protein: A potential target for vaccine development and other therapeutic strategies. *PLoS One*, *15*(8), e0237300–25. <https://doi.org/10.1371/journal.pone.0237300>
- Schrödinger LLC. (2010). The PyMOL molecular graphics system. Version 2.3.0.
- Schwede, T., Kopp, J., Guex, N., & Peitsch, M. C. (2003). SWISS-MODEL: An automated protein homology-modeling server. *Nucleic Acids Research*, *31*(13), 3381–3385. <https://doi.org/10.1093/nar/gkg520>
- Singh, J., Singh, H., Hasnain, S. E., & Rahman, S. A. (2020). Mutational signatures in countries affected by SARS-CoV-2: Implications in host-pathogen interactome. *bioRxiv*. <https://doi.org/10.1101/2020.09.17.301614>
- Singh, S., Ehtesham, N. Z., Rahman, S. A., & Hasnain, S. E. (2021). Structure function investigation of a new VUI-202012/01 SARS-CoV-2 variant. *bioRxiv*. <https://doi.org/10.1101/2021.01.01.425028>
- Tomaszewski, T., DeVries, R. S., Dong, M., Bhatia, G., Norsworthy, M. D., Zheng, X., & Caetano-Anollés, G. (2020). New pathways of mutational change in SARS-CoV-2 proteomes involve regions of intrinsic disorder important for virus replication and release. *Evolutionary Bioinformatics Online*, *16*, 1176934320965149–1176934320965118. <https://doi.org/10.1177/1176934320965149>
- Tranchevent, L. C., Capdevila, F. B., Nitsch, D., de Moor, B., de Causmaecker, P., & Moreau, Y. (2011). A guide to web tools to prioritize candidate genes. *Briefings in Bioinformatics*, *12*(1), 22–32. <https://doi.org/10.1093/bib/bbq007>
- Verma, D., Jacobs, D. J., & Livesay, D. R. (2012). Changes in lysozyme flexibility upon mutation are frequent, large and long-ranged. *PLoS Computational Biology*, *8*(3), e1002409. <https://doi.org/10.1371/journal.pcbi.1002409>
- Volz, E., Hill, V., McCrone, J. T., Price, A., Jorgensen, D., O'Toole, Á., Southgate, J., Johnson, R., Jackson, B., Nascimento, F. F., Rey, S. M., Nicholls, S. M., Colquhoun, R. M., da Silva Filipe, A., Shepherd, J., Pascall, D. J., Shah, R., Jesudason, N., Li, K., ... Connor, T. R. (2021). Evaluating the effects of SARS-CoV-2 spike mutation D614G on transmissibility and pathogenicity. *Cell*, *184*(1), 64–75. <https://doi.org/10.1016/j.cell.2020.11.020>
- Walls, A. C., Park, Y. J., Tortorici, M. A., Wall, A., McGuire, A. T., & Veersler, D. (2020). Structure, function, and antigenicity of the SARS-CoV-2

- spike glycoprotein. *Cell*, 181(2), 281–292.e6. <https://doi.org/10.1016/j.cell.2020.02.058>
- Wang, J., Shi, C., Xu, Q., & Yin, H. (2021). SARS-CoV-2 nucleocapsid protein undergoes liquid – liquid phase separation into stress granules through its N-terminal intrinsically disordered region. *Cell Discovery*, 7(1), 3–7. <https://doi.org/10.1038/s41421-020-00240-3>
- Wang, R., Hozumi, Y., Yin, C., & Wei, G. W. (2020). Decoding SARS-CoV-2 transmission, evolution and ramification on COVID-19 diagnosis, vaccine, and medicine. *Journal of Chemical Information and Modeling*, 60(12), 5853–5865. <https://doi.org/10.1021/acs.jcim.0c00501>
- Wieczorek, G., & Zielenkiewicz, P. (2008). DeltaF508 mutation increases conformational flexibility of CFTR protein. *Journal of Cystic Fibrosis*, 7(4), 295–300. <https://doi.org/10.1016/j.jcf.2007.11.008>
- Worth, C. L., Preissner, R., & Blundell, T. L. (2011). SDM - A server for predicting effects of mutations on protein stability and malfunction. *Nucleic Acids Research*, 39(suppl), W215–222. <https://doi.org/10.1093/nar/gkr363>
- 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*, 1263(March), 1260–1263. <https://doi.org/10.1101/2020.02.11.944462>
- 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. <https://doi.org/10.1038/s41586-020-2008-3>
- Yamada, Y., Banno, Y., Yoshida, H., Kikuchi, R., Akao, Y., Murate, T., & Nozawa, Y. (2006). Catalytic inactivation of human phospholipase D2 by a naturally occurring Gly901Asp mutation. *Archives of Medical Research*, 37(6), 696–699. <https://doi.org/10.1016/j.arcmed.2006.01.006>
- Yang, Y., Peng, F., Wang, R., Yange, M., Guan, K., Jiang, T., Xu, G., Sun, J., & Chang, C. (2020). The deadly coronaviruses: The 2003 SARS pandemic and the 2020 novel coronavirus epidemic in China. *Journal of Autoimmunity*, 109, 102434. <https://doi.org/10.1016/j.jaut.2020.102434>
- Yao, H., Lu, X., Chen, Q., Xu, K., Chen, Y., Cheng, M., Chen, K., Cheng, L., Weng, T., Shi, D., Liu, F., Wu, Z., Xie, M., Wu, H., Jin, C., Zheng, M., Wu, N., Jiang, C., & Li, L. (2020). Patient-derived SARS-CoV-2 mutations impact viral replication dynamics and infectivity in vitro and with clinical implications in vivo. *Cell Discovery*, 6(1), 76. <https://doi.org/10.1038/s41421-020-00226-1>
- Zhang, L., Jackson, C. B., Mou, H., Ojha, A., Peng, H., Quinlan, B. D., Rangarajan, E. S., Pan, A., Vanderheiden, A., Suthar, M. S., Li, W., Izard, T., Rader, C., Farzan, M., & Choe, H. (2020). SARS-CoV-2 spike-protein D614G mutation increases virion spike density and infectivity. *Nature Communications*, 11(1), 1–9. <https://doi.org/10.1038/s41467-020-19808-4>
- Zhou, H. X., & Pang, X. (2018). Electrostatic interactions in protein structure, folding, binding, and condensation. *Chemical Reviews*, 118(4), 1691–1741. <https://doi.org/10.1021/acs.chemrev.7b00305>
- Zhou, P., Yang, X.-L., Wang, X.-G., Hu, B., Zhang, L., Zhang, W., Si, H.-R., Zhu, Y., Li, B., Huang, C.-L., Chen, H.-D., Chen, J., Luo, Y., Guo, H., Jiang, R.-D., Liu, M.-Q., Chen, Y., Shen, X.-R., Wang, X., ... Shi, Z.-L. (2020). A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature*, 579(7798), 270–273. <https://doi.org/10.1038/s41586-020-2012-7>
- Zinzula, L., Basquin, J., Bohn, S., Beck, F., Klumpe, S., Pfeifer, G., Nagy, I., Bracher, A., Hartl, F. U., & Baumeister, W. (2021). High-resolution structure and biophysical characterization of the nucleocapsid phosphoprotein dimerization domain from the Covid-19 severe acute respiratory syndrome coronavirus 2. *Biochemical and Biophysical Research Communications*, 538, 54–62. <https://doi.org/10.1016/j.bbrc.2020.09.131>