Microsecond simulation unravel the structural dynamics of SARS-CoV-2 Spike-C-terminal cytoplasmic tail (residues 1242-1273)

Prateek Kumar[a], Taniya Bhardwaj[a], Neha Garg[b], Rajanish Giri[a*]

[a]School of Basic Sciences, Indian Institute of Technology Mandi, VPO Kamand, Himachal Pradesh, 175005, India.

[b]*Department of Medicinal Chemistry, Faculty of Ayurveda, Institute of Medical Sciences, Banaras Hindu University, Varanasi, Uttar Pradesh, 221005, India.*

***Correspondence Email:** rajanishgiri@iitmandi.ac.in. **Telephone number:** 01905-267134, Fax number: 01905-267138

Abstract:

Spike protein of human coronaviruses has been a vital drug and vaccine target. The multifunctionality of this protein including host receptor binding and apoptosis has been proved in several coronaviruses. It also interacts with other viral proteins such as membrane (M) protein through its C-terminal domain. The specific dibasic motif signal present in cytosolic region at C-terminal of spike protein helps it to localize within the endoplasmic reticulum (ER). However, the structural conformation of cytosolic region is not known in SARS-CoV-2 using which it interacts with other proteins and transporting vesicles. Therefore, we have demonstrated the conformation of cytosolic region and its dynamics through computer simulations up to microsecond timescale using OPLS and CHARMM forcefields. The simulations have revealed the unstructured conformation of cytosolic region (residues 1242-1273). Also, in temperature dependent replica-exchange molecular dynamics simulations it has shown to form secondary structures. We believe that our findings will surely help us understand the structure-function relationship of the spike protein's cytosolic region.

Keywords: Spike, Cytosolic domain, Conformational dynamics, SARS-CoV-2

Introduction

The importance of coronavirus spike protein is apparent from it surface-exposed location, rendering it a prime target after viral infection for cell-mediated and humoral immune responses as well as artificially designed vaccines and antiviral therapeutics. The SARS-CoV-2 homo-trimeric spike glycoprotein consists of an extracellular unit anchored by a transmembrane (TM) domain in viral membrane and a cytoplasmic domain (1). It is secreted as monomeric 1273 amino acid long protein from endoplasmic reticulum (ER) shortly after which it trimerizes to facilitate the transport to the Golgi complex (2, 3). Moreover, N-linked high mannose oligosaccharide side chains that are added to spike monomer in ER are further modified in Golgi compartments (2).

Spike protein comprises two independent subunits S1 and S2 that are associated through noncovalent interactions in a metastable prefusion state. During maturation in the trans-Golgi network, spike is cleaved by furin or furin-like host proteases at S1/S2 cleavage site yielding the two unequal subunits (4, 5). The distal S1 subunit (residues 14–685) contains a N-terminal domain, a C-terminal domain, and two subdomains (**Figure 1)**. The C-terminal domain is the receptor-binding domain or RBD, has a receptor-binding motif (RBM) which interacts with human angiotensin converting enzyme 2 (ACE2), chief target receptor of SARS-CoV-2 on human cells (6). RBM is present as an extended loop insertion which binds to bottom side of the small lobe of ACE2 receptor. RBD further consists of two subdomains: an external and a core subdomain which engages with the target receptor (2, 7). Several detailed crystallographic structures of SARS-CoV-2 RBD in complex with ACE2 are available on protein data bank revealing the network of key hydrophilic interactions (6, 8). The S2 subunit (residues 686-1273) has a hydrophobic fusion peptide, two heptad repeats, a transmembrane domain, and a cytoplasmic C-terminal tail (**Figure 1**). After virus attachment on host cellular membrane, S2 subunit enclosing a fusion machinery mediates the fusion of host cell and viral membranes (2, 9). It has been reported in SARS-CoV that the C-terminal domain or S2 subunit of spike induces apoptosis in Vero E6 cells in a dose and time dependent manner (10).

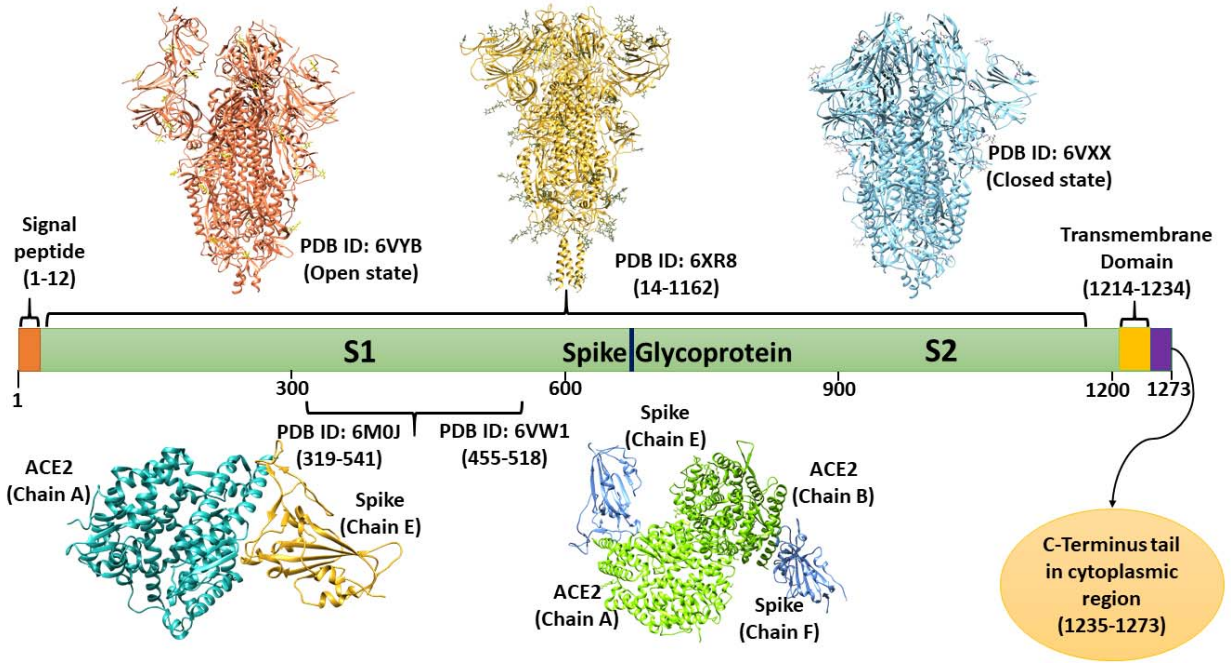


Figure 1: Domain architecture of spike Glycoprotein: depiction of available structures in open and closed states, transmembrane domain, and cytoplasmic C-terminal tail (based on UniProt database).

As of yet, cytoplasmic domain of spike protein is the least explored region despite of such extensive research in pandemic times. It is of particular importance as it contains a conserved ER retrieval signal (KKXX) (11). In SARS-CoV and SARS-CoV-2 spike proteins, a novel dibasic KLHYT (KXHXX) motif present at extreme ends of the C-terminus plays a crucial role in its subcellular localization (12–14). Also, deletions in cytoplasmic domain of coronavirus spike are implicated in viral infection in recent reports (15–18). SARS-CoV and SARS-CoV-2 spike having a deletion of last ~20 residues displayed increased infectivity of single-cycle vesicular stomatitis virus (VSV)-S pseudotypes (15, 16). Contrarily, short truncations of cytoplasmic domain of MHV spike protein (Δ12 and Δ25) had limited effect on viral infectivity while the long truncation of 35 residues interfered with both viral-host cell membrane fusion and assembly. Importantly, it is also shown to interact with the membrane protein inside host cells (17). In our previous report, the cytoplasmic tail is predicted to be a MoRF (Molecular Recognition Feature) region (residues 1265-1273) by a predictor MoRFchibi (14). The MoRF

regions in proteins are disorder-based binding regions that contribute to DNA, RNA, and other proteins. In the same report, it is also found to contain many DNA and RNA binding residues.

We aimed to understand the cytoplasmic domain (1242-1273 residues) of the SARS-CoV-2 spike protein to gain further insights. To this end, we computationally analyzed its behavioral dynamics using molecular dynamic (MD) simulations up to one microsecond (μs) using our in-house facility. This report's outcomes will help understand this domain's structure and function and provide an idea to explore the interaction of spike protein with other viral and host proteins.

Material and Methods

Transmembrane prediction: Before studying the cytoplasmic domain, we have applied multiple servers to predict the spike protein's transmembrane region. TMHMM (19), TMPred (20), SPLIT (21), PSIPRED (22, 23), and CCTOP (24) web predictors works using highly optimized and least biased algorithms which also takes into account the homology of sequences. Among the aforementioned predictors, CCTOP predicts the transmembrane regions based on the consensus of multiple predictors and experimental derived structural information from homologous proteins in the database. Therefore, it provides a better understanding and identification of transmembrane regions in the protein. Based on CCTOP prediction, we have chosen the C-terminal cytoplasmic tail region to elucidate its structural dynamics.

Disorder Prediction: The propensity of disorderedness in spike C-terminal cytoplasmic tail region is predicted using PONDR family (25–27), IUPred long (28), and PrDOS (29) servers. The detailed methodology is given in our previous reports (14, 30).

Structure Modelling: PEP-FOLD 3.5 webserver (31) is used to predict 3D structures of selected spike protein regions. By implementing *optimized potential for efficient structure prediction* (OPEP) coarse-grained forcefield-based simulations, an improved and minimized structure is obtained as described earlier (32–34). Then, the structure is prepared in Schrodinger suite where the missing hydrogens, improper bond orders, and protonation states are corrected. Further, the prepared structure is used for MD simulations.

Molecular Dynamic (MD) Simulations: To comprehend and comparable outcomes for intrinsically disordered regions, we used two different forcefields to analyze the structural dynamics of the cytosolic domain of spike protein.

Simulation with OPLS 2005

We used Desmond simulations package, where simulation setup is built by placing the protein structure in an orthorhombic box along with TIP3P water model, 0.15M NaCl salt concentration (35). After solvation, the system is charge neutralized with counterions using OPLS 2005 forcefield. To attain an energy minimized simulation system, the steepest descent method is used for 5000 iterations. Further, the equilibration of system is done to optimize solvent in the environment. Using NVT and NPT ensembles within periodic boundary conditions, the system is equilibrated for 100 ps each. The average temperature at 300K and pressure at 1 bar are maintained using Nose-Hoover and Martyna-Tobias-Klein (MTK) coupling methods during simulation (36, 37). All bond-related constraints are solved using SHAKE algorithm, and hydrogen bond constraints were solved using LINCS algorithm (38). The final production run is performed for 1 µs using our in-house facilities.

Simulation with CHARMM36m

Another forcefield we used, CHARMM36m in Gromacs, is an improved version of CHARMM36, which is effectively developed for analyzing IDP regions in the proteins in significant simulation timescale (39, 40). Using TIP3P water model and salt, the system is prepared for proper electrostatic distribution and then neutralized for charge using counterion (1 Na+ ion in this case). The energy minimization of simulation setup using steepest descent method is done for 50,000 steps. For temperature and pressure coupling, the V-rescale and Parrinello-Rahman algorithms are used where 300K and 1 bar is the average temperature and pressure respectively. After equilibration of 100 ps for NVT and NPT methods, the production run is then executed for 1 µs using our high performing cluster at IIT Mandi.

Replica-Exchange Molecular Dynamic (REMD) simulations

The enhanced conformation sampling using REMD simulations is widespread in protein folding. During REMD simulations, the swapping of conformations occurs and reduces the chances of entrapping simulations in local minimum energy states (41). Therefore, we have performed REMD using eight replicas (numbered from 0 to 7) at temperatures 298 K, 314 K, 330 K, 346 K, 362 K, 378 K, 394 K, and 410 K, calculated by linear mode of Desmond. The last frame of 1 µs of Desmond simulation trajectory is chosen as the initial conformation for REMD. The multigrator integrator of Langevin and Nose-Hoover as thermostats, whereas Langevin and

Martyna-Tobias-Klein (MTK) are used to equilibrate the systems. The accountability of conformation swaps to be accepted or rejected is done based on common Metropolis criterion using the following equation:

$$Q = (\beta 1U11 + \beta 2U22 - \beta 1U12 - \beta 2U21) + (\beta 1P1 - \beta 2P2)\,(V1 - V2)$$

Where U_{ij} = potential energy of replica i in the Hamiltonian of replica j,

P_i = the reference pressure of replica i,

V_i = instantaneous volume of replica i, and

β_i = the inverse reference temperature of replica i.

If $Q > 0$ = accept, or $Q < -20$ = reject the exchange,

else accept the exchange if $rand_N < \exp(Q)$, where $rand_N$ is a random variable on (0,1).

Results

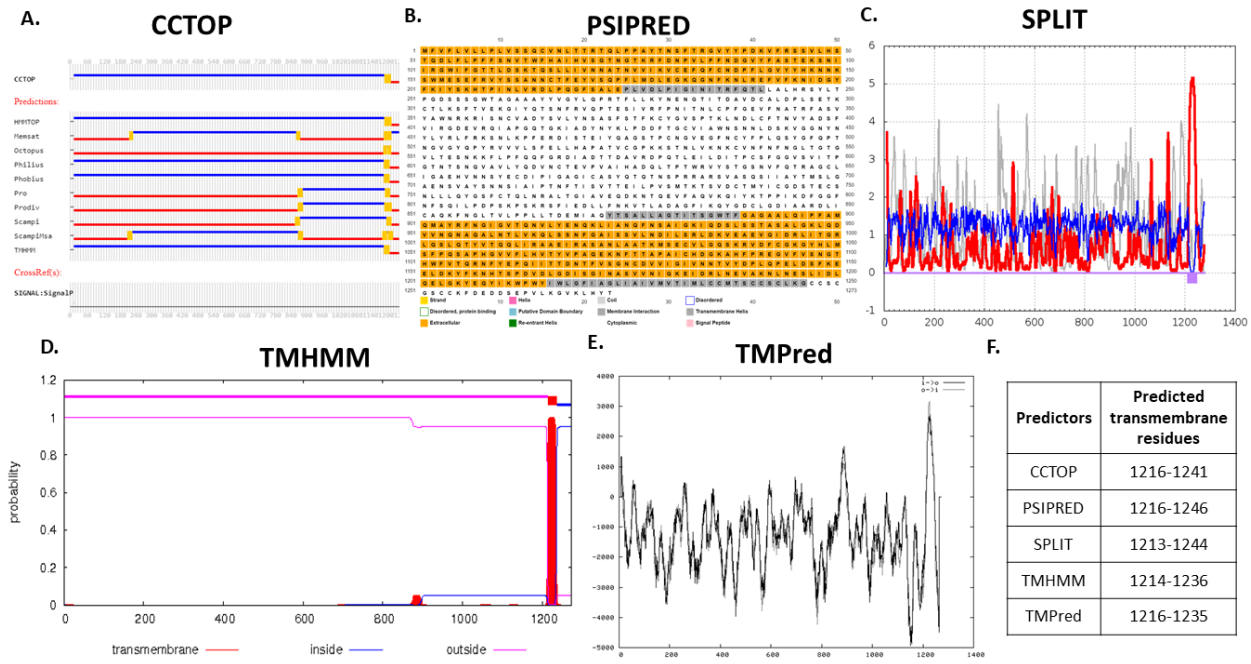
In recent times, computational approaches have been widely used to explore the secondary and tertiary structures of proteins and small peptides. It has immensely helped a lot in the ongoing COVID-19 pandemic to study structural conformations of protein and their interacting partners, i.e., protein, ligands, glycans, etc. Molecular dynamics simulation is an useful approach to answer such questions at the atomic level. Herein, we have studied the transmembrane region and performed rigorous simulations to unravel the least explored cytoplasmic domain of essential spike protein.

Transmembrane region analysis:

The sequence-based analysis of transmembrane region and disorder prone regions have also been analyzed. The subcellular localization of spike protein occurs in the extracellular, transmembrane, and cytoplasmic regions (42). However, based on SARS-CoV and SARS-CoV-2 proteins' sequence alignment, approximately 77% similarity is found among both viruses' spike proteins (14). The C-terminal has shown high similarity and conserved regions, while the N-terminal has vastly varying residues.

Based on multiple predictors used in this study, spike protein's transmembrane region lies within 1213-1246 residues (**Figure 2)**. A consensus-based server, CCTOP, has predicted the

transmembrane region from residues 1216-1241, which is more reliable as it compares and uses the previously available experimental information of related proteins. Therefore, the cytoplasmic region is selected from 1242 to 1273 amino acids (sequence: SCLKGCCSCGSCCKFDEDDSEPVLKGVKLHYT).



Predictors	Predicted transmembrane residues
CCTOP	1216-1241
PSIPRED	1216-1246
SPLIT	1213-1244
TMHMM	1214-1236
TMPred	1216-1235

Figure 2: Transmembrane region prediction from five web servers: A. CCTOP, **B.** PSIPRED, **C.** SPLIT, **D.** TMHMM, and **E.** TMPred. **F.** Table showing predicted transmembrane residues.

Disorder prediction

In our recent study, we have identified the disordered and disorder-based binding regions in SARS-CoV-2 where the cytoplasmic domain at C-terminal of spike protein is found to be disordered (14). Again, we analyzed the disorderedness in selected cytoplasmic region using multiple predictors, including PONDR family, IUPred long, and PrDOS predictors. Out of six predictors, three predictors from PONDR family have predicted it as highly disordered, PrDOS has predicted it as moderately disordered. In contrast, PONDR FIT and IUPred long predicted it as least disordered (**Figure 3**).

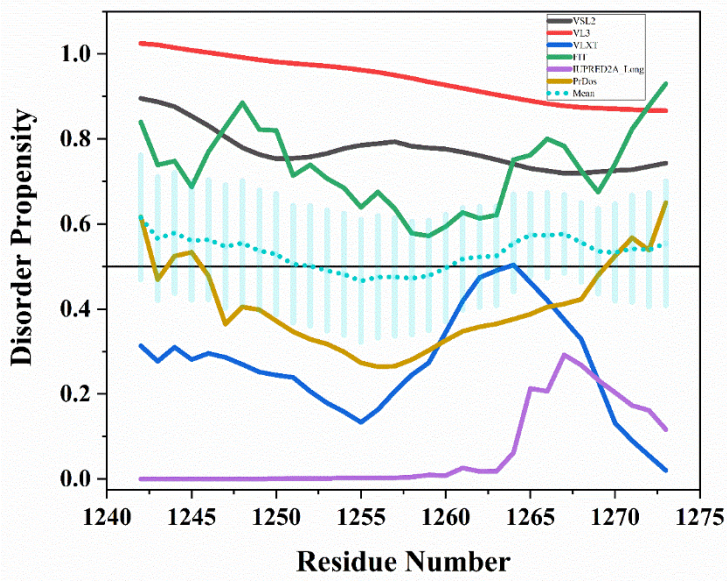


Figure 3: Intrinsic disorder analysis of spike C-terminal cytoplasmic tail (residues 1242-1273) region using six predictors including PONDR family (VSL2, VL3, VLXT, and FIT), IUPred long, and PrDOS servers. The mean line is denoted in short dots style, and the standard error bars on mean are also highlighted.

Structure modelling with PEPFOLD3

In absence of an experimentally determined 3D structure of protein, structure modeling provides an approximate structure model based on homology and properties of amino acids using a wide range of optimized algorithms. It generates the prototype fragments and assembles them by implementing *optimized potential for efficient structure prediction* (OPEP) coarse-grained forcefield-based simulations. The best-obtained structure of extreme C-terminus tail (1242-1273 amino acids) containing two small helical regions is further prepared for MD simulations in aqueous conditions. These helical regions are present at residues $_{1243}$CLKGC$_{1247}$ and $_{1265}$LKGV$_{1268}$ of spike glycoproteins C-terminus (**Figure 4A)**.

Simulation with OPLS 2005

In the last three decades, many advancements have been made in forcefields and hardware related to MD simulation to match the experimental events. Long MD simulations up to microseconds or milliseconds are incredibly insightful to study structural conformations occurring at the nanoscale level. We have recently explored various regions of different SARS-

CoV-2 proteins through computational simulations and experimental techniques that are very well correlated (32, 33). This study performed 1 µs MD simulations of C-terminal cytoplasmic domain of spike protein (1242-1273 residues) to understand its dynamic nature. As obtained from structure modeling through PEP-FOLD, the model contains two small helices at both terminals with residues $_{1243}$CLKGC$_{1247}$ and $_{1265}$LKGV$_{1268}$ (**Figure 4A & 4B**). According to *2struc* webserver (43), these helices contribute to 12.5% of total secondary structure while rest of the region is constituted by turns and extended coils. The secondary structure prediction of spike C-terminal tail region contains a -strand of five residues $_{1262}$EPVLK$_{1266}$, as predicted by PSIPRED webserver (22) (**Figure 4C**).

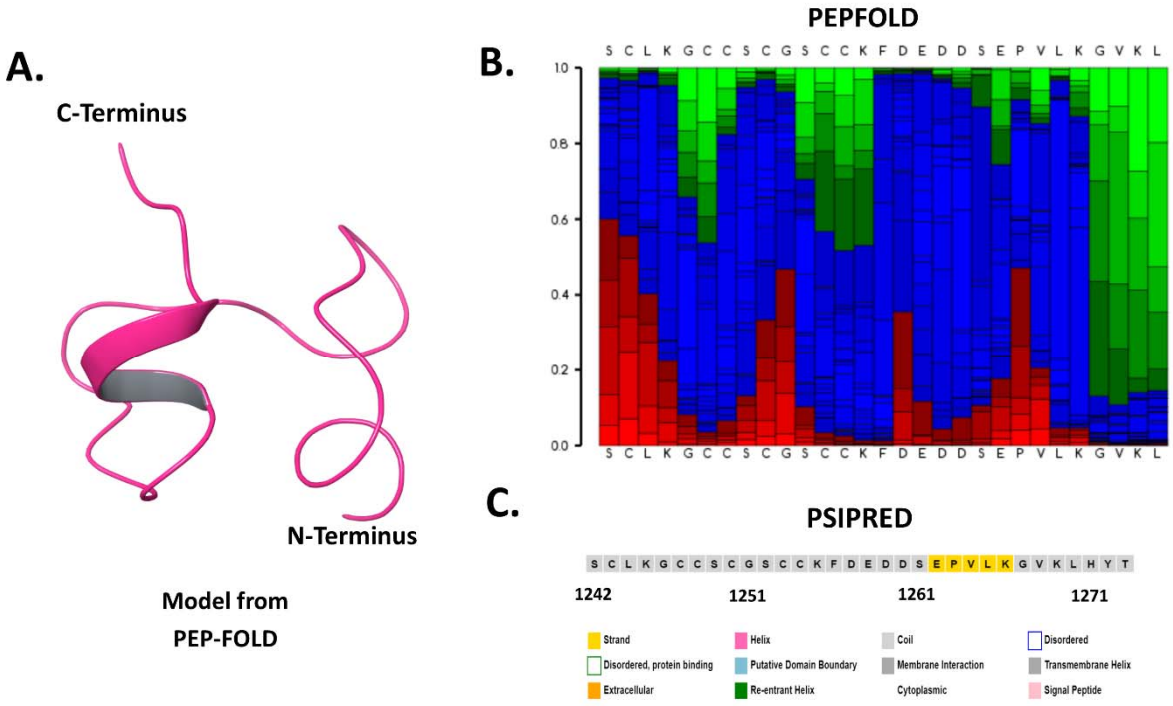


Figure 4: Sequence and structure-based analysis of spike C-terminal cytoplasmic domain (1242-1273 residues): A. Modeled structure through PEP-FOLD web server, **B.** Secondary structure analysis using PSIPRED web server, and **C.** PEP-FOLD structure analysis depicting helix (red), coil (blue), and extended (green).

After analyzing the disorder propensity and secondary structure composition, we performed a rigorous simulation of cytoplasmic region (residues 1242-1273) to understand its atomic movement and structural integrity. A total of 1 µs simulation was done after 50,000 steps of

steepest descent method-based energy minimization. It has been observed that the structure of spike C-terminal cytoplasmic region remains to be unstructured throughout the simulation. Based on mean distance analysis, the peptide simulation setup showed massive deviations up to 7.5 Å which does not attain any stable state (**Figure 5A**). As shown in **figure 5B**, mean fluctuation in residues is observed to be within the range of 1.6 – 6.4 Å. The secondary structure timeline (**Figure 5C & 5D**) also reveals the disordered state of spike C-terminal cytoplasmic region during the 1 µs simulation time (none of the frames captured α-helix or β-sheets) which is further depicted in the snapshot of 1 µs frame in **figure 5E**.

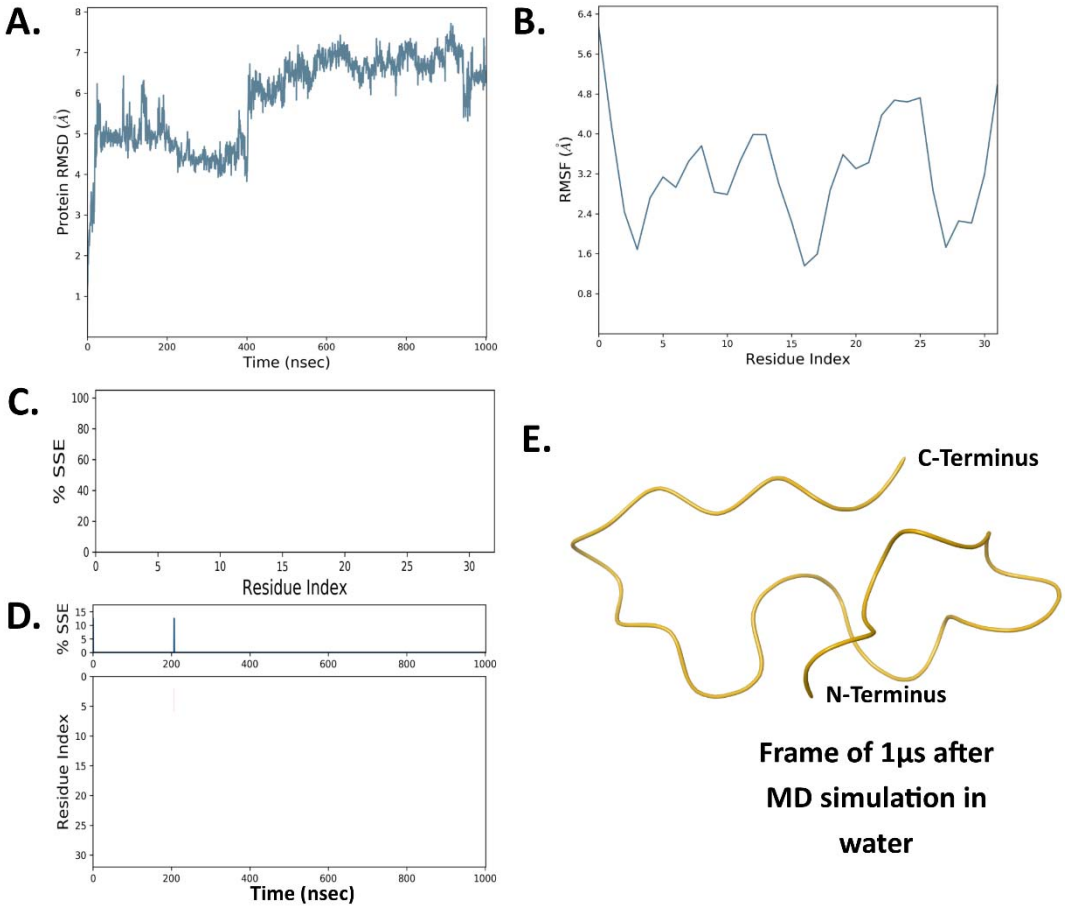


Figure 5: One microsecond MD Simulation analysis of spike C-terminal cytoplasmic domain (1242-1273): A. Root mean square deviation (RMSD), **B.** Root mean square fluctuation (RMSF), **C.** Secondary structure element (SSE) of residues, **D.** Timeline representation of secondary structure content during 1µs simulation time, and **E.** Last frame at 1µs.

Simulation with CHARMM36m

The characterization of intrinsically disordered proteins (IDPs) and regions (IDPRs) using MD simulations is very well persistent in literature. In last two decades, several forcefields for MD simulations are developed and improved at times. All forcefields have their advantages and limitations due to the proper evaluation of secondary structure composition. Here, we have used another forcefield, CHARMM36m, for determining the conformational dynamics of spike cytosolic region. As shown through trajectory snapshots at every 100 ns in **figure 6**, the cytosolic region has adopted a β-sheet conformation at its N-terminal. Two β-strands can be seen with varying amino acid length at every 0.1 μs frame. The only exceptions are 0.3 μs and 0.4 μs frames, which do not show any secondary structure. Further, after 0.5 μs frame, a gradual loss in two β-strands indicates a gain in disorder content in spike C-terminus. To this end, the 1 μs simulation frame comprises of only a short β-strand and rest unstructured residues.

These structural changes have been the reason for immensely varying atomic distances throughout the simulation. Likewise, RMSD values are found in a range of approx. 0.75 nm to 1.75 nm (**Figure 7A**); the mean residual fluctuations are in the range of 0.5 nm to 1.1 nm (**Figure 7C**), and the Rg values vary up to 2.4 nm (**Figure 7B**), which demonstrates the vastly changing structural compactness. The timeline representation of secondary structure composition at each frame also depicts the structural inconsistency throughout the simulation (**Figure 7D**). Convincingly, a major part of spike cytosolic region is disordered and might have the propensity to gain structure in physiological conditions.

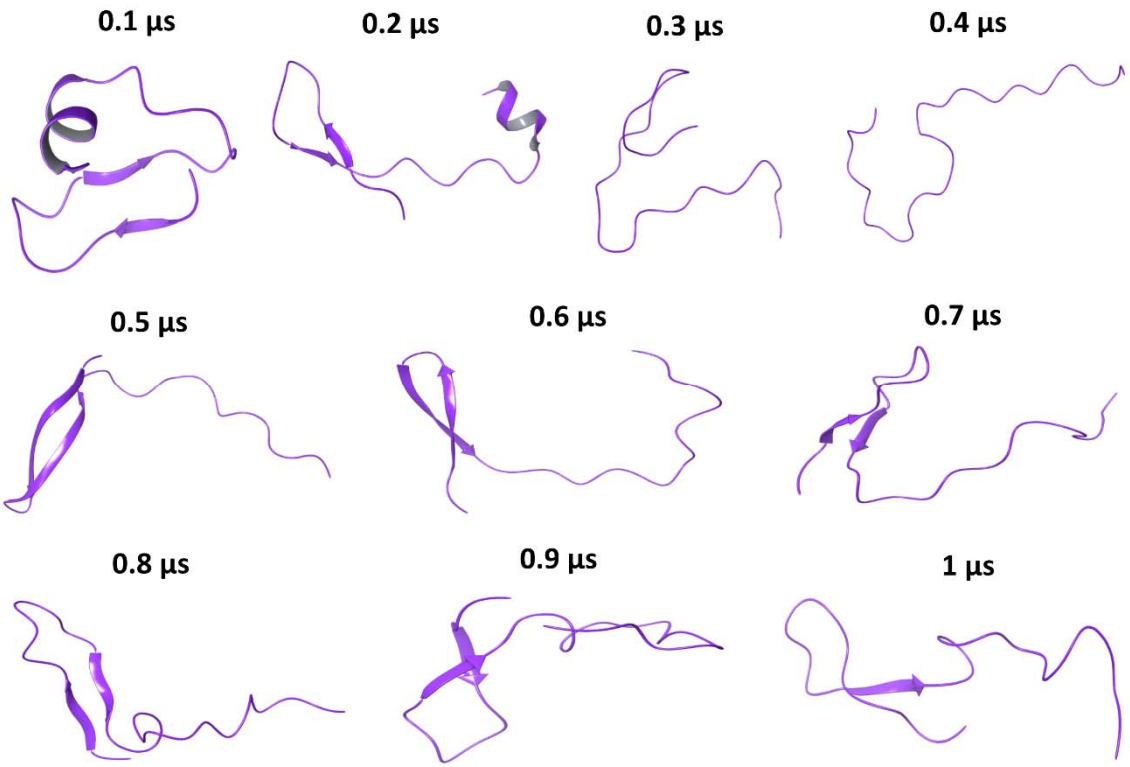


Figure 6: MD simulation with CHARMM36m forcefield: Snapshots at every 100 ns of 1 µs simulation trajectory.

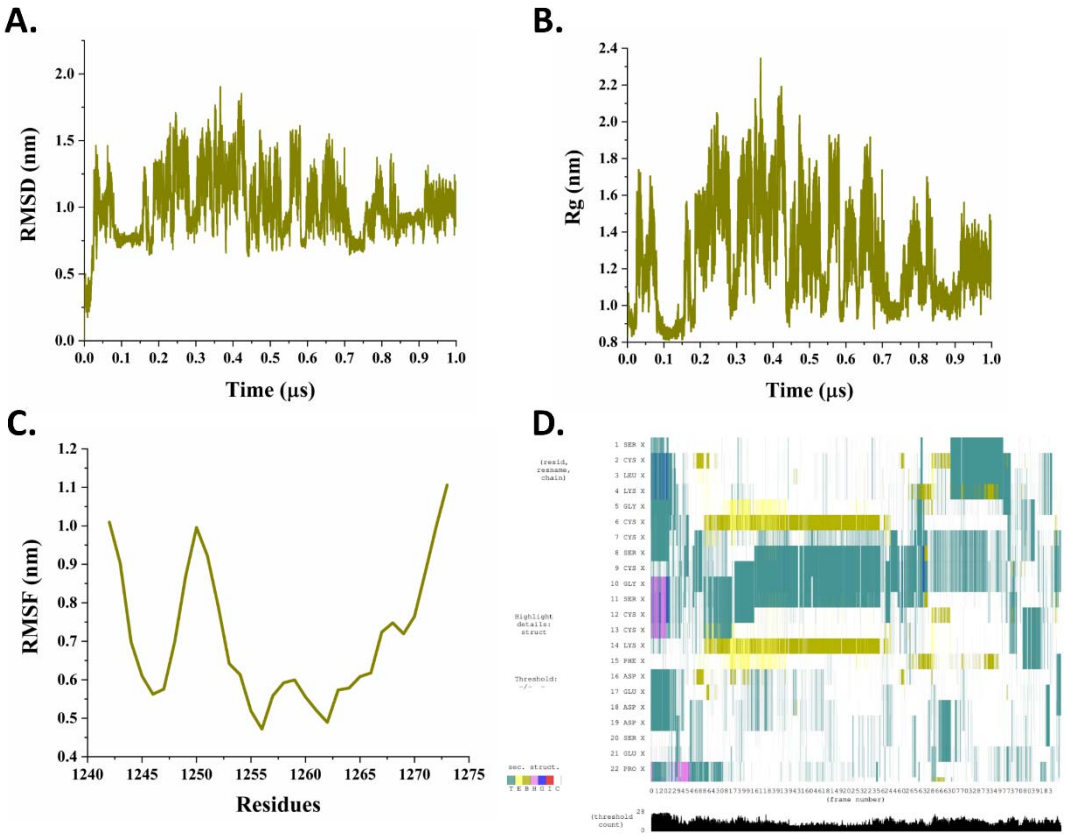


Figure 7: Trajectory analysis of spike cytoplasmic region with CHARMM36m forcefield: A. RMSD, **B.** Radius of gyration (Rg), **C.** RMSF, and **D.** Secondary structure element timeline during the course of 1 μs simulation time.

Conformation sampling using Replica-Exchange Molecular Dynamic (REMD) simulations

The disordered form (last frame) of spike cytoplasmic region from Desmond simulation trajectory is used for REMD at 8 temperatures *viz.* 298 K, 314 K, 330 K, 346 K, 362 K, 378 K, 394 K, and 410 K upto half a microsecond using 8 replicas (numbered as 0 to 7). As shown here in **figure 8**, the cytoplasmic region has adopted a β-sheet structure at increasing temperatures upto 394 K. In comparison, at 410 K in replica 7 these changes are found to be reversible. Although previous frames of replica 7 display the formation of multiple long β-strands throughout simulation time. According to snapshots illustrated in **figure 8**, the cytosolic region has gained three to four β-strands and appears to be in a well-folded manner as temperature increases.

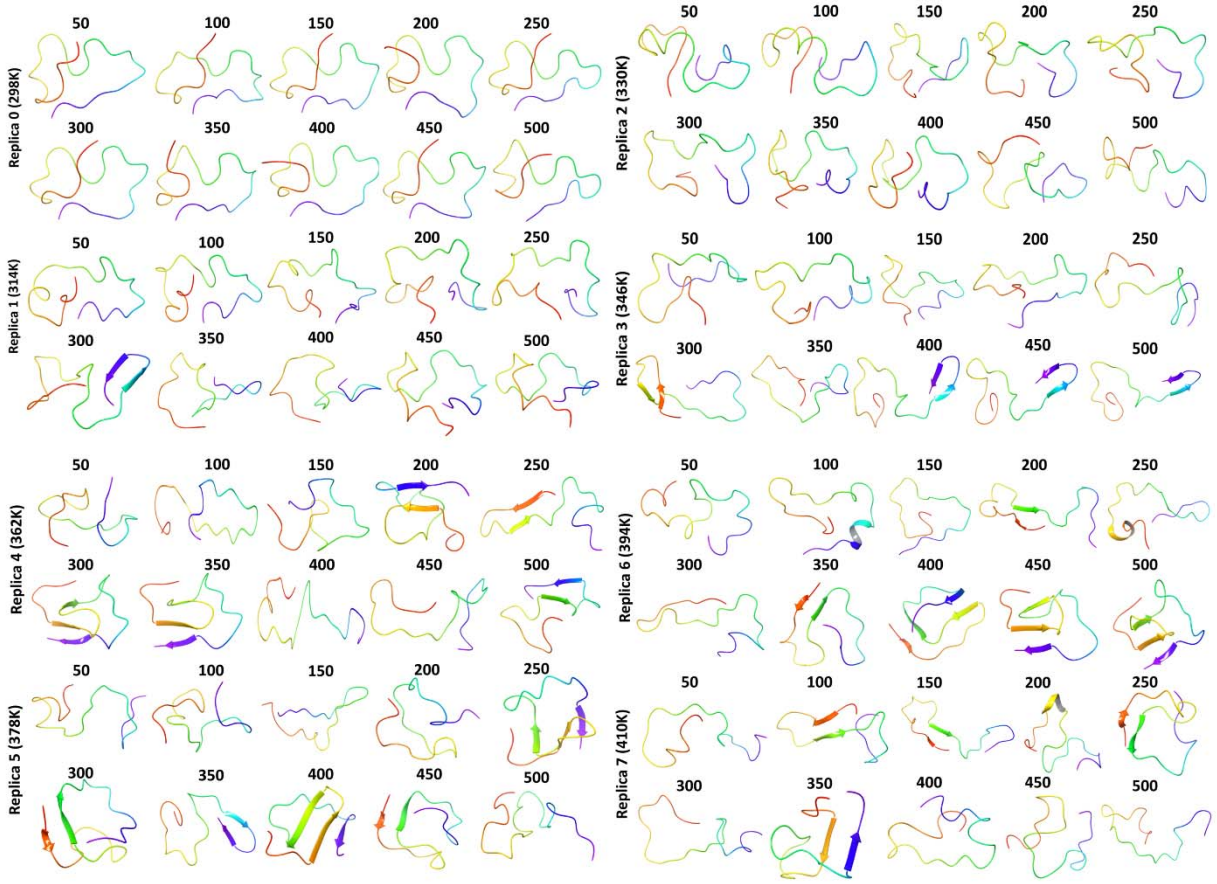
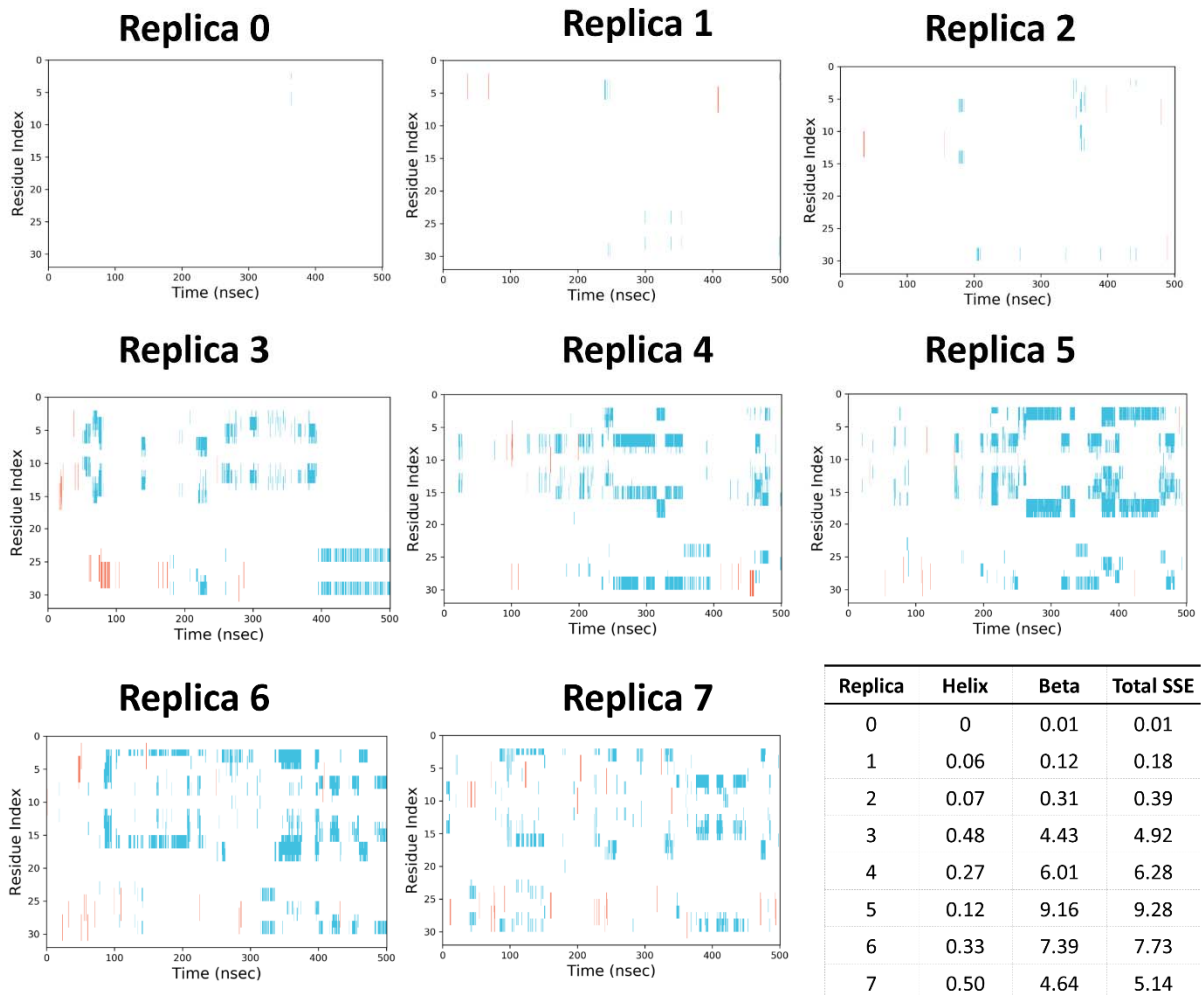


Figure 8: Snapshots from all replicas at every 50 ns during half a microsecond REMD simulation.



Replica	Helix	Beta	Total SSE
0	0	0.01	0.01
1	0.06	0.12	0.18
2	0.07	0.31	0.39
3	0.48	4.43	4.92
4	0.27	6.01	6.28
5	0.12	9.16	9.28
6	0.33	7.39	7.73
7	0.50	4.64	5.14

Figure 9: Timeline representation of simulation trajectories from all replicas upto half a microsecond simulation time. A table of percentage secondary structure elements in each simulation.

For a clear understanding of all frames in REMD, timelines of all 8 replicas are displayed in **figure 9** where the consistent formation of β-strands due to rising temperature is also validated. However, at extreme temperature (410K), some structural changes were reversed, and the total secondary structure element (SSE) gets reduced in comparison to previous replicas (**Figure 9 Table**). According to mean distances analyses, huge fluctuation is observed in replicas 3, 5, and 6 where structural changes occurred (**Figure 10A, 10B, 10D**). As elucidated through hydrogen bond analysis (**Figure 10C**), the highest numbers are 21, 20, 20 for replicas 2, 6, and 7, respectively. The superimposed last frame of each replica shows structural differences with

atomic distances (RMSD) in range of 3.7 Å to 8.5 Å with respect to the starting frame for REMD (**Figure 10E** and **F**).

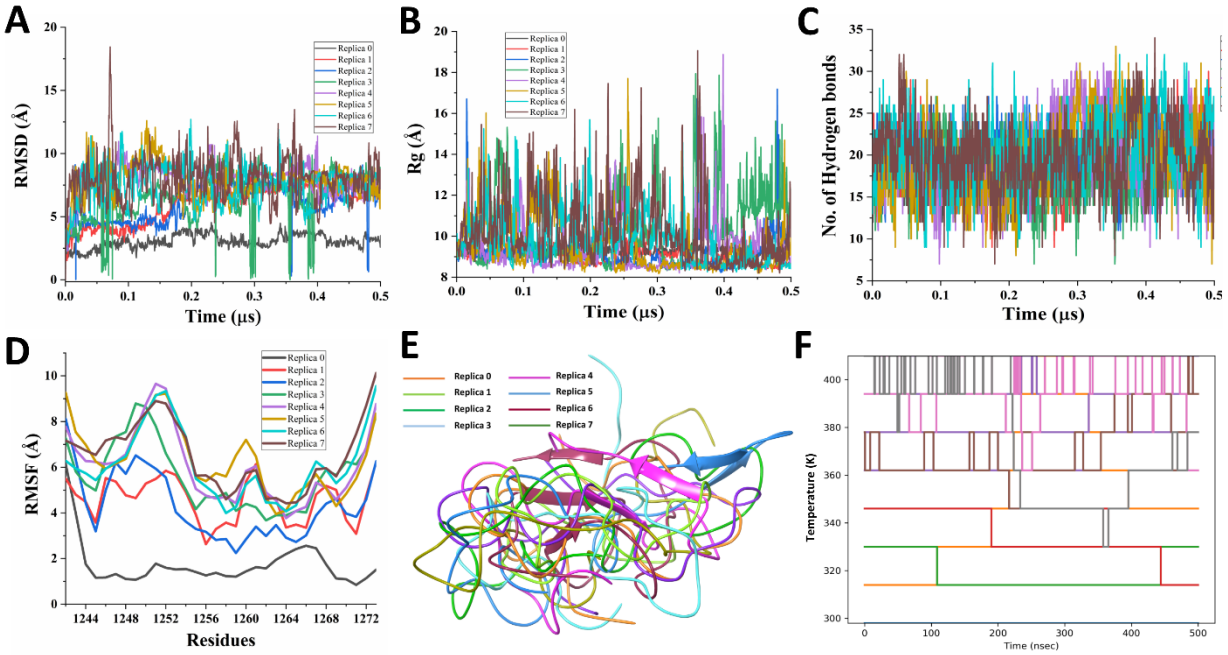


Figure 10: Trajectory analysis from all replicas simulated at different temperatures: A. RMSD, **B.** Rg, **C.** Number of hydrogen bonds, **D.** RMSF, **E.** Superimposed last frames, and **F.** Conformation exchange review during REMD.

Discussion

The functional importance of C-terminal domain (S2 subunit) of spike protein is very well established in SARS-CoV and SARS-CoV-2. Notably, the cytoplasmic domain is known to possess the localization signals for ER or endoplasmic reticulum-Golgi intermediate compartment in SARS-CoV as well as in other coronaviruses (11, 13, 44). The dibasic motif (KXHXX) present at the extreme C-terminal end of spike protein binds with COP1 coated transporting vesicles to transport it into ER from the Golgi complex result in virion assembly upon interacting with other viral proteins (13). In SARS-CoV, it has been reported that the cysteine-rich C-terminal domain of spike is also responsible for interaction with M protein as a mutation in this domain obstructed their interaction (17, 45). A report also suggests that during virus particle assembly, the localization signal remains active in unfolded conformation in ER while the well-folded structure interacts with M protein leading to the incorporation of spike into the virion (44). The additional spike molecules migrate to cell surface due to unrecognized

retention signals where it mediates cell fusion to spread infection (44). This is seen in some coronaviruses where the cytoplasmic domain with transmembrane domain embed spike protein in viral lipid envelope (17). Our recent report on protein-protein interactions of SARS-CoV-2 structural proteins have also demonstrated that the C-terminal domain or S2 subunit of spike protein interacts with M protein (46). In protein-protein docking and MD based study, it is revealed that the residues of spike protein Asp796, Lys921, Asn925, Tyr1209, Cys1241, and Cys1247 interacts with Trp31, Tyr39, Phe53, Trp55, and Trp58 residues of M protein through stable intermolecular bonds (46).

Our study on the structural dynamics of SARS-CoV-2 spike cytoplasmic domain demonstrates it to be a disordered region. Based on the outcomes of two forcefields, OPLS 2005 and CHARMM36m, spike cytosolic region remains majorly unstructured. Nevertheless, in REMD simulations, it adopted β-sheets at rising temperatures with time demonstrating its gain of structure-property. Generally, an IDPR gains any structure upon interacting with its interacting partner or in physiological conditions (47). In unstructured state, cytoplasmic domain may function as a MoRF to bind with COP1 coated transporting vesicles, which localizes spike protein into ER. As described earlier, the interaction of C-terminal domain of spike protein is reported with other structural proteins like M, which is highly likely to occur in its disordered form with extended radius.

Conclusion

The advancement in computational powers and excessive improvements in forcefields have empowered structural biology. Newly developed algorithms and their user-friendly approach allow correlating the outcomes with experimental observations. In this article, we have identified the transmembrane region in spike protein by employing distinguished web predictors. This cleared the composition of amino acids forming cytoplasmic domain. Further, the secondary structure and disorder predisposition analysis demonstrated it to be highly disordered. We have demonstrated the structural conformation of cytoplasmic domain (1242-1273 residues) of spike protein at a microsecond timescale using computational simulations. As revealed, this domain is purely unstructured or disordered after one microsecond and have not gained any structural conformation throughout the simulation period. Our results are in good correlation with previously published studies on coronaviruses. Based on our previous study (14), cytoplasmic

tail of spike glycoprotein has molecular recognition features therein. In this manuscript, the multiple conformations during the simulation process adds up to even more interesting speculations.

Acknowledgements: All the authors would like to thank IIT Mandi for the infrastructure. RG is thankful to IYBA award from DBT, Government of India (BT/11/IYBA/2018/06), MHRD-SPARC (SPARC/2018-2019/P37/SL), and Science and Engineering Research Board (SERB), India (Grant Number:CRG/2019/005603). TB is thankful to DST for her INSPIRE fellowship. NG acknowledges Seed grant IOE, from Banaras Hindu University.

Conflict of Interest: All authors affirm that there are no conflicts of interest.

Author contribution: RG, NG: Conception, design & study supervision. PK and TB: acquisition and interpretation of data. PK, TB, and RG: contributed to paper writing.

References

1. Huang, Y., Yang, C., Xu, X. feng, Xu, W., and Liu, S. wen (2020) Structural and functional properties of SARS-CoV-2 spike protein: potential antivirus drug development for COVID-19. *Acta Pharmacol. Sin.* **41**, 1141–1149

2. Duan, L., Zheng, Q., Zhang, H., Niu, Y., Lou, Y., and Wang, H. (2020) The SARS-CoV-2 Spike Glycoprotein Biosynthesis, Structure, Function, and Antigenicity: Implications for the Design of Spike-Based Vaccine Immunogens. *Front. Immunol.* **11**, 2593

3. Vennema, H., Rottier, P. J. M., Heijnen, L., Godeke, G. J., Horzinek, M. C., and Spaan, W. J. M. (1990) Biosynthesis and function of the coronavirus spike protein. in *Advances in Experimental Medicine and Biology*, pp. 9–19, Springer, Boston, MA, **276**, 9–19

4. Walls, A. C., Park, Y. J., Tortorici, M. A., Wall, A., McGuire, A. T., and Veesler, D. (2020) Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. *Cell*. **181**, 281-292.e6

5. Hoffmann, M., Kleine-Weber, H., and Pöhlmann, S. (2020) A Multibasic Cleavage Site in

the Spike Protein of SARS-CoV-2 Is Essential for Infection of Human Lung Cells. *Mol. Cell*. **78**, 779-784.e5

6. Lan, J., Ge, J., Yu, J., Shan, S., Zhou, H., Fan, S., Zhang, Q., Shi, X., Wang, Q., Zhang, L., and Wang, X. (2020) Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *Nature*. **581**, 215–220

7. Lan, J., Ge, J., Yu, J., Shan, S., Zhou, H., Fan, S., Zhang, Q., Shi, X., Wang, Q., Zhang, L., and Wang, X. (2020) Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *Nature*. **581**, 215–220

8. Yan, R., Zhang, Y., Li, Y., Xia, L., Guo, Y., and Zhou, Q. (2020) Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2. *Science (80-.).* **367**, 1444–1448

9. Wrapp Daniel, Wang Nianshuang, Corbett Kizzmekia S, Goldsmith Jory A, Hsieh Ching-Lin, Abiona Olubukola, Graham Barney S, and McLellan Jason S (2020) Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science (80-.).* **367**, 1260–1263

10. Chow, K. Y. C., Yeung, Y. S., Hon, C. C., Zeng, F., Law, K. M., and Leung, F. C. C. (2005) Adenovirus-mediated expression of the C-terminal domain of SARS-CoV spike protein is sufficient to induce apoptosis in Vero E6 cells. *FEBS Lett.* **579**, 6699–6704

11. Lontok, E., Corse, E., and Machamer, C. E. (2004) Intracellular Targeting Signals Contribute to Localization of Coronavirus Spike Proteins near the Virus Assembly Site. *J. Virol.* **78**, 5913–5922

12. Sadasivan, J., Singh, M., and Sarma, J. Das (2017) Cytoplasmic tail of coronavirus spike protein has intracellular targeting signals. *J. Biosci.* **42**, 231–244

13. McBride, C. E., Li, J., and Machamer, C. E. (2007) The Cytoplasmic Tail of the Severe Acute Respiratory Syndrome Coronavirus Spike Protein Contains a Novel Endoplasmic Reticulum Retrieval Signal That Binds COPI and Promotes Interaction with Membrane Protein. *J. Virol.* **81**, 2418–2428

14. Giri, R., Bhardwaj, T., Shegane, M., Gehi, B. R., Kumar, P., Gadhave, K., Oldfield, C. J., and Uversky, V. N. (2020) Understanding COVID-19 via comparative analysis of dark

proteomes of SARS-CoV-2, human SARS and bat SARS-like coronaviruses. *Cell. Mol. Life Sci.* 10.1007/s00018-020-03603-x

15. Dieterle, M. E., Haslwanter, D., Bortz, R. H., Wirchnianski, A. S., Lasso, G., Vergnolle, O., Abbasi, S. A., Fels, J. M., Laudermilch, E., Florez, C., Mengotto, A., Kimmel, D., Malonis, R. J., Georgiev, G., Quiroz, J., Barnhill, J., Pirofski, L. anne, Daily, J. P., Dye, J. M., Lai, J. R., Herbert, A. S., Chandran, K., and Jangra, R. K. (2020) A Replication-Competent Vesicular Stomatitis Virus for Studies of SARS-CoV-2 Spike-Mediated Cell Entry and Its Inhibition. *Cell Host Microbe*. **28**, 486-496.e6

16. Ou, X., Liu, Y., Lei, X., Li, P., Mi, D., Ren, L., Guo, L., Guo, R., Chen, T., Hu, J., Xiang, Z., Mu, Z., Chen, X., Chen, J., Hu, K., Jin, Q., Wang, J., and Qian, Z. (2020) Characterization of spike glycoprotein of SARS-CoV-2 on virus entry and its immune cross-reactivity with SARS-CoV. *Nat. Commun.* 10.1038/s41467-020-15562-9

17. Bosch, B. J., De Haan, C. A. M., Smits, S. L., and Rottier, P. J. M. (2005) Spike protein assembly into the coronavirion: Exploring the limits of its sequence requirements. *Virology*. **334**, 306–318

18. Ujike, M., Huang, C., Shirato, K., Makino, S., and Taguchi, F. (2016) The contribution of the cytoplasmic retrieval signal of severe acute respiratory syndrome coronavirus to intracellular accumulation of S proteins and incorporation of S protein into virus-like particles. *J. Gen. Virol.* **97**, 1853–1864

19. Krogh, A., È rn Larsson, B., von Heijne, G., and L Sonnhammer, E. L. (2001) Predicting Transmembrane Protein Topology with a Hidden Markov Model: Application to Complete Genomes. *J. Mol. Biol.* **305**, 567–580

20. Hofmann, K., and Stoffel, W. (1993) A Database of Membrane Spanning Protein Segments. *Biol. Chem.*

21. Juretić, D., Zoranić, L., and Zucić, D. (2002) Basic charge clusters and predictions of membrane protein topology. *J. Chem. Inf. Comput. Sci.* **42**, 620–632

22. Buchan, D. W. A., and Jones, D. T. (2019) The PSIPRED Protein Analysis Workbench: 20 years on. *Nucleic Acids Res.* **47**, W402–W407

23. Nugent, T., and Jones, D. T. (2009) Transmembrane protein topology prediction using support vector machines. *BMC Bioinformatics*. **10**, 159

24. Dobson, L., Reményi, I., and Tusnády, G. E. (2015) CCTOP: A Consensus Constrained TOPology prediction web server. *Nucleic Acids Res.* **43**, W408–W412

25. Obradovic, Z., Peng, K., Vucetic, S., Radivojac, P., Brown, C. J., and Dunker, A. K. (2003) Predicting Intrinsic Disorder From Amino Acid Sequence. in *Proteins: Structure, Function, and Genetics*, pp. 566–572, Proteins, **53**, 566–572

26. Xue, B., Dunbrack, R. L., Williams, R. W., Dunker, A. K., and Uversky, V. N. (2010) PONDR-FIT: A meta-predictor of intrinsically disordered amino acids. *Biochim. Biophys. Acta - Proteins Proteomics.* **1804**, 996–1010

27. P, R., Z, O., X, L., EC, G., CJ, B., AK, D., Romero, P., Obradovic, Z., Li, X., Garner, E. C., Brown, C. J., and Dunker, A. K. (2001) Sequence Complexity of Disordered Protein. *Proteins*. **42**, 38–48

28. Mészáros, B., Erdös, G., and Dosztányi, Z. (2018) IUPred2A: Context-dependent prediction of protein disorder as a function of redox state and protein binding. *Nucleic Acids Res.* **46**, W329–W337

29. Ishida, T., and Kinoshita, K. (2007) PrDOS: Prediction of disordered protein regions from amino acid sequence. *Nucleic Acids Res.* 10.1093/nar/gkm363

30. Kumar, D., Singh, A., Kumar, P., Uversky, V. N., Rao, C. D., and Giri, R. Understanding the penetrance of intrinsic protein disorder in rotavirus proteome. *Int. J. Biol. Macromol.* **144**, 892–908

31. Shen, Y., Maupetit, J., Derreumaux, P., and Tufféry, P. (2014) Improved PEP-FOLD approach for peptide and miniprotein structure prediction. *J. Chem. Theory Comput.* **10**, 4745–4758

32. Gadhave, K., Kumar, P., Kumar, A., Bhardwaj, T., Garg, N., and Giri, R. (2020) NSP 11 of SARS-CoV-2 is an Intrinsically Disordered Protein. *bioRxiv*. 10.1101/2020.10.07.330068

33. Kumar, A., Kumar, A., Kumar, P., Garg, N., and Giri, R. (2020) SARS-CoV-2 NSP1 C-terminal region (residues 130-180) is an intrinsically disordered region. *bioRxiv*. 10.1101/2020.09.10.290932

34. Kumar, A., Kumar, P., Kumari, S., Uversky, V. N., and Giri, R. (2020) Folding and structural polymorphism of p53 C-terminal domain: One peptide with many conformations. *Arch. Biochem. Biophys.* **684**, 108342

35. Shaw, D. E. (2005) A fast, scalable method for the parallel evaluation of distance-limited pairwise particle interactions. *J. Comput. Chem.* **26**, 1318–1328

36. Martyna, G. J., Klein, M. L., and Tuckerman, M. (1992) Nosé-Hoover chains: The canonical ensemble via continuous dynamics. *J. Chem. Phys.* **97**, 2635–2643

37. Martyna, G. J., Tobias, D. J., and Klein, M. L. (1994) Constant pressure molecular dynamics algorithms. *J. Chem. Phys.* **101**, 4177–4189

38. Hess, B., Bekker, H., Berendsen, H. J. C., and Fraaije, J. G. E. M. (1997) LINCS: A linear constraint solver for molecular simulations. *J. Comput. Chem.* **18**, 1463–1472

39. Huang, J., Rauscher, S., Nawrocki, G., Ran, T., Feig, M., De Groot, B. L., Grubmüller, H., and MacKerell, A. D. (2016) CHARMM36m: An improved force field for folded and intrinsically disordered proteins. *Nat. Methods.* **14**, 71–73

40. Berendsen, H. J. C., van der Spoel, D., and van Drunen, R. (1995) GROMACS: A message-passing parallel molecular dynamics implementation. *Comput. Phys. Commun.* **91**, 43–56

41. Sugita, Y., and Okamoto, Y. (1999) Replica-exchange molecular dynamics method for protein folding. *Chem. Phys. Lett.* **314**, 141–151

42. Cai, Y., Zhang, J., Xiao, T., Peng, H., Sterling, S. M., Walsh, R. M., Rawson, S., Rits-Volloch, S., and Chen, B. (2020) Distinct conformational states of SARS-CoV-2 spike protein. *Science (80-.).* **369**, 1586–1592

43. Klose, D. P., Wallace, B. A., and Janes, R. W. (2010) 2Struc: the secondary structure server. *Bioinformatics*. **26**, 2624–2625

44. Sadasivan, J., Singh, M., and Sarma, J. Das (2017) Cytoplasmic tail of coronavirus spike protein has intracellular targeting signals. *J. Biosci.* **42**, 231–244

45. de Haan, C. A. M., Smeets, M., Vernooij, F., Vennema, H., and Rottier, P. J. M. (1999) Mapping of the Coronavirus Membrane Protein Domains Involved in Interaction with the Spike Protein. *J. Virol.* **73**, 7441–7452

46. Kumar, A., Kumar, P., Garg, N., and Giri, R. (2020) An insight into SARS-CoV-2 Membrane protein interaction with Spike, Envelope, and Nucleocapsid proteins. *bioRxiv*. 10.1101/2020.10.30.363002

47. Wright, P. E., and Dyson, H. J. (1999) Intrinsically unstructured proteins: Re-assessing the protein structure-function paradigm. *J. Mol. Biol.* **293**, 321–331