Contents lists available at ScienceDirect

Virology

journal homepage: www.elsevier.com/locate/virology

Microsecond simulations and CD spectroscopy reveals the intrinsically disordered nature of SARS-CoV-2 spike-C-terminal cytoplasmic tail (residues 1242–1273) in isolation

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

ARTICLE INFO

Keywords:
Spike
Cytosolic domain
Conformational dynamics
Secondary structure
SARS-CoV-2

ABSTRACT

All available SARS-CoV-2 spike protein crystal and cryo-EM structures have shown missing electron densities for cytosolic C-terminal regions (CTR). Generally, the missing electron densities point towards the intrinsically disordered nature of the protein region (IDPR). This curiosity has led us to investigate the cytosolic CTR of the spike glycoprotein of SARS-CoV-2 in isolation. The spike CTR is supposed to be from 1235 to 1273 residues or 1242–1273 residues based on our used prediction. Therefore, we have demonstrated the structural 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. Further, we have validated our computational observations with circular dichroism (CD) spectroscopy-based experiments and found its signature spectra at 198 nm. We believe that our findings will surely help in understanding the structure-function relationship of the spike protein's cytosolic region.

1. Introduction

The importance of coronavirus spike protein is apparent from it surface-exposed location, suggesting it is 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 (Walls et al., 2020). 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 (Duan et al., 2020; Walls et al., 2020). Moreover, N-linked high mannose oligosaccharide side chains that are added to spike monomer in ER are further modified in Golgi compartments (Duan et al., 2020).

Spike is one of the most extensively studied protein among all of SARS-CoV-2 proteome. So far, based on Uniprot database, approximately two hundred structures have been reported using X-ray crystallography and cryo-electron microscopy techniques. However, these structures consist of S1 subunit of spike but lacks the transmembrane and cytoplasmic C-terminal regions present in S2 subunit or with missing electron densities in cytoplasmic region. The distal S1 subunit (residues 14–685) contains a N-terminal domain, a C-terminal domain, and two subdomains (Fig. 1). The C-terminal domain of S1 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 (Lan et al., 2020). RBM is present as an extended loop insertion which binds to bottom side of the small lobe of ACE2 receptor. The S2 subunit (residues 686–1273) has a hydrophobic fusion peptide, two heptad repeats, a transmembrane domain, and a cytoplasmic C-terminal tail (Fig. 1).

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) (Lontok et al., 2004). 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 (Giri et al., 2020; McBride et al., 2007; Sadasivan et al., 2017). Also, deletions in cytoplasmic domain of coronavirus spike are implicated in viral

; CD, Circular Dichroism; cryo-EM, Cryo-electron microscopy; CTR, C-terminal Region; DTT, Dithiothreitol; IDPR, Intrinsically disordered protein region; MD, Molecular Dynamics; PEG, Polyethylene glycol; REMD, Replica-Exchange Molecular Dynamics; SDS, Sodium Dodecyl Sulfate; TFE, 2,2,2-Trifluoroethanol.

* Corresponding author.
E-mail address: rajanishgiri@iitmandi.ac.in (R. Giri).

https://doi.org/10.1016/j.virol.2021.11.005
Received 12 August 2021; Received in revised form 1 November 2021; Accepted 17 November 2021
Available online 27 November 2021

infection in recent reports (Bosch et al., 2005; Dieterle et al., 2020; Ou et al., 2020; Ujike et al., 2016). 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 (Dieterle et al., 2020; Ou et al., 2020). Contrarily, short truncations of cytoplasmic domain of Mouse Hepatitis Virus (MHV) spike protein (△12 and △25) 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 (Bosch et al., 2005). In our previous report, the cytoplasmic tail is predicted to be a MoRF (Molecular Recognition Feature) region (residues 1265–1273) by a predictor MoRFchibi (Giri et al., 2020). The MoRF regions in proteins are disorder-based binding regions that contribute the binding to DNA, RNA, and other proteins. In the same report, it is also found to contain many DNA and RNA binding residues (Giri et al., 2020).

Despite of availability of several structures of spike protein using advanced techniques like cryo-EM, the structure of cytoplasmic domain is not yet clear due to its 'missing electron density'. Generally, intrinsically disordered proteins show such characteristic of missing electron density and lacks a well-defined three-dimensional structure (Uversky, 2020). Additionally, the consensus-based disorder prediction by MobiDB has shown this region to be disordered (Piovesan et al., 2021). Considering these arguments, we aimed to understand the cytoplasmic domain of the SARS-CoV-2 spike protein to gain further insights. To this end, we computationally analyzed its behavioural dynamics using molecular dynamic (MD) simulations up to 1 microsecond (μs) and validated it with CD spectroscopy based experiments. This report's outcomes will help to understand this domain's structure and function and provide knowledge to explore the interaction of spike protein with other viral and host proteins.

2. Material and methods

2.1. Materials

For CD based experiments, the synthesized peptide of spike cytoplasmic region (residues 1242–1273) with purity of 85.3% was procured in lyophilized form from GeneScript, USA. Chemicals and solvents i.e., Dithiothreitol (DTT), 2,2,2-Trifluoroethanol (TFE), Sodium Dodecyl Sulfate (SDS), and Sucrose with more than 99% of purity were purchased from Sigma-Aldrich, USA. In addition, macromolecular crowding agents, polyethylene glycol (PEG 8000), and Dextran-70 are used to examine the structural behavior of spike-CTR in cell-like crowding conditions. Further, the lyophilized peptide of spike cytoplasmic region was dissolved in nuclease free water with a concentration of 1 mg/ml and prepared a stock concentration of 289.35 μM for all experimental measurements.

2.2. Transmembrane prediction

Before studying the cytoplasmic domain, we have applied multiple servers to predict the spike protein's transmembrane region. TMHMM (Krogh et al., 2001), TMPred (Hofmann and Stoffel, 1993), SPLIT (Juretić et al., 2002), PSIPRED (Buchan and Jones, 2019; Nugent and Jones, 2009), and CCTOP (Dobson et al., 2015) web predictors works using highly optimized and least biased algorithms which also takes into account the homology of sequences. Among the 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.

2.3. Disorder prediction

The propensity of disorderedness in spike C-terminal cytoplasmic tail region is predicted using PONDR family (Obradovic et al., 2003; P et al., 2001; Xue et al., 2010), IUPred2A (Mészáros et al., 2018), and PrDOS (Takashi Ishida et al., 2007) servers. The detailed methodology is given in our previous reports (Giri et al., 2020; Kumar et al., 2020).

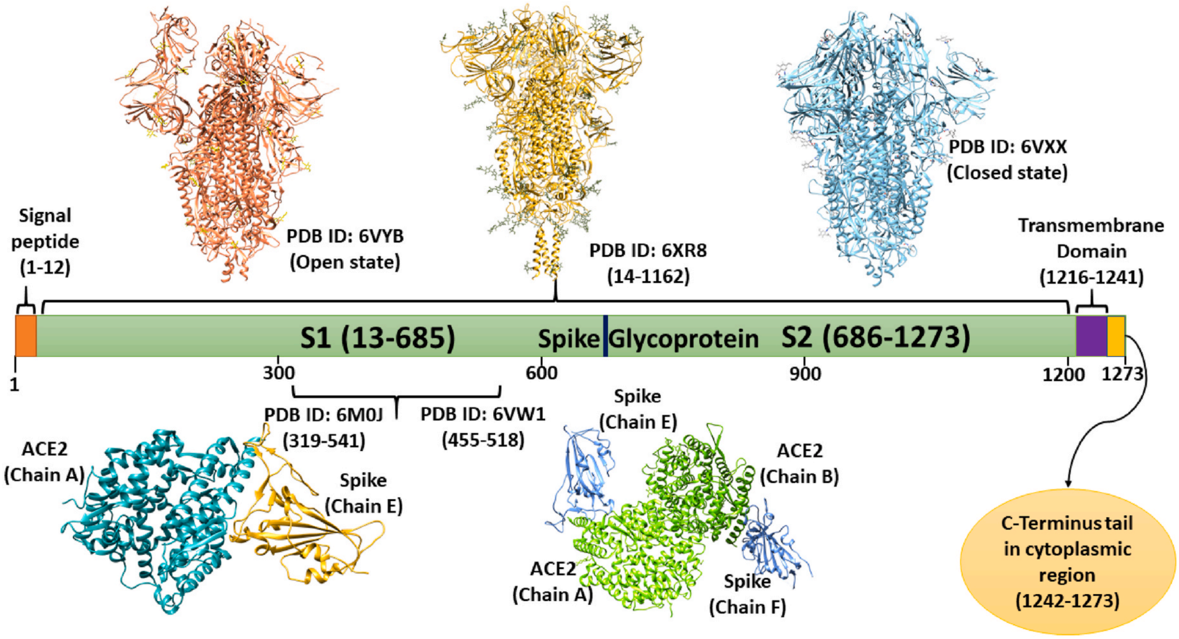


Fig. 1. Domain architecture of spike Glycoprotein: depiction of available structures in open and closed states, transmembrane domain, and cytoplasmic C-terminal tail. Based on prediction of transmembrane region in the spike protein by CCTOP, a consensus-based predictor, the boundaries of all domains have been defined. As per CCTOP prediction, the transmembrane region of spike lies within the residues 1216–1241, and so, the cytoplasmic region of spike has been used in this study with the residues 1242–1273.

2.4. *Structure modelling*

For full-length protein structure modelling, *AlphaFold2* tool is employed (Jumper et al., 2021). AlphaFold2 is a new, highly-advanced, and reliable tool which utilizes artificial intelligence-based algorithms to model the 3D structure of a protein. It used multiple available structures of spike protein as templates and generated the five high-quality models.

PEP-FOLD 3.5 webserver (Shen et al., 2014) is used to predict 3D structures of selected spike protein's cytoplasmic region. By implementing *optimized potential for efficient structure prediction* (OPEP) coarse-grained forcefield-based simulations, an improved and minimized structure is obtained as described earlier (Gadhave et al., 2020b; A. Kumar et al., 2021). 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.

2.5. *Molecular dynamic (MD) simulations*

To comprehend and comparable outcomes for intrinsically disordered regions, we have used two different forcefields to analyze the structural dynamics of the cytosolic domain of spike protein. The improved and optimized forcefields, OPLS 2005 and Charmm36 m are developed and now generally used for investigation of conformational dynamics of IDPs. Both forcefields produce a balanced evaluation for secondary structure evaluation (P. Kumar et al., 2021).

2.5.1. *Simulation with OPLS 2005*

We have used Desmond simulations package, where simulation setup is built by placing the protein structure in an orthorhombic box along with TIP3P water model, 0.15 M NaCl salt concentration (Shaw, 2005). 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 300 K and pressure at 1 bar are maintained using Nose-Hoover thermostat and Martyna-Tobias-Klein (MTK) coupling methods during simulation (Martyna et al., 1992, 1994). All bond-related constraints are solved using SHAKE algorithm, and hydrogen bond constraints are solved using LINCS algorithm (Hess et al., 1997). The final production run is performed for 1 μs using our in-house facilities.

2.5.2. *Simulation with CHARMM36 m*

Another forcefield we used, CHARMM36 m in Gromacs, is an improved version of CHARMM36, which is effectively developed for analyzing IDP regions in the proteins in significant simulation timescale (Berendsen et al., 1995; Huang et al., 2016). Using TIP3P water model, 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 300 K 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.

2.5.3. *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 (Sugita and Okamoto, 1999). 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 (Martyna et al., 1992, 1994). The accountability of conformation swaps to be accepted or rejected is done based on common Metropolis criterion using the following equation:

$$Q = (\beta1U11 + \beta2U22 - \beta1U12 - \beta2U21) + (\beta1P1 - \beta2P2)\ (V1 - V2)$$

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).

2.6. *CD spectroscopy*

To obtain the secondary structure based spectral information, we employed JASCO CD instrument (Jasco J-1500 CD spectrometer, USA). 25 μM of spike cytoplasmic region was prepared in 20 mM sodium phosphate buffer (physiological pH 7.4) to record the spectra in isolation. Further, to observe the structural changes in peptide, CD spectra (with same concentration) was recorded in increasing concentration of organic solvent TFE (0–90%). Similarly, conformational changes in peptide were monitored in presence of an anionic detergent, Sodium Dodecyl Sulfate (SDS) (below and above critical micellar concentration (CMC)), and Sucrose. Additionally, the intracellular crowding environment mimicking agents PEG 8000 and Dextran-70 were utilized (at concentrations from 50 to 300 g/L) to examine the structural changes in the peptide. All spectra were recorded in far-UV region from 190 nm to 240 nm in 1 mm quartz cuvette. The scan speed was kept at 50 nm/min with a response time of 1s and 1 nm bandwidth. With similar parameters, baseline spectra of buffers were recorded and subtracted from all spectral measurements. In case of macromolecular crowders experiments, the baseline spectra consisted buffers and crowding agents. The resultant data were plotted as wavelengths vs ellipticity (with HT value, < 600 V) at x and y axis respectively using Origin software.

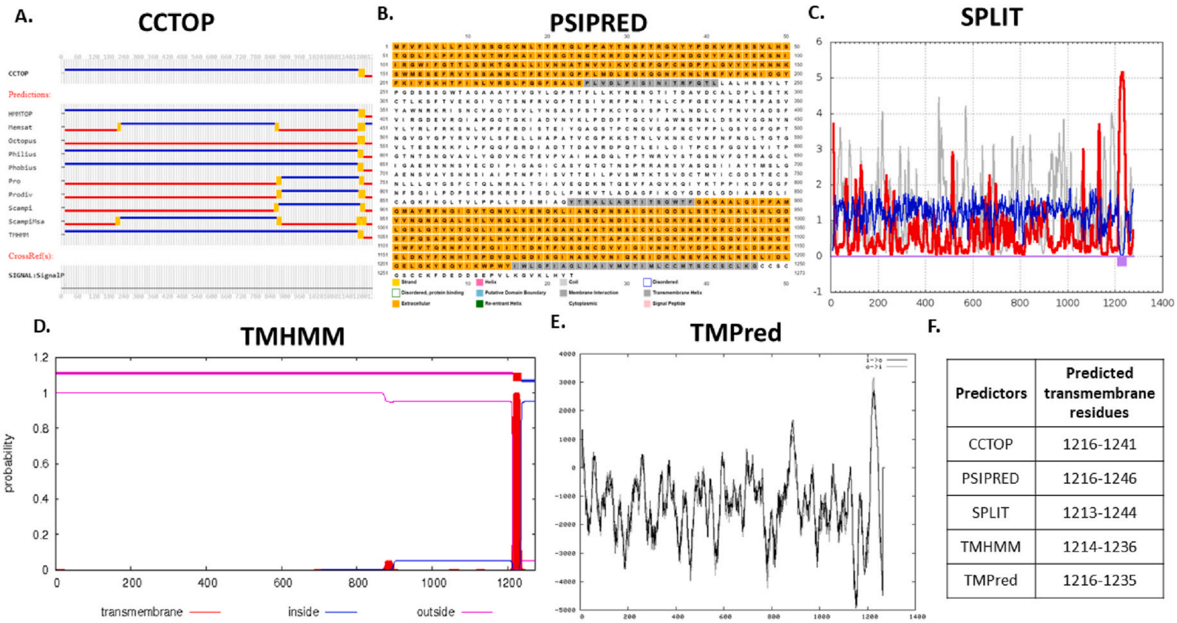
3. Results

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 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 a useful approach to answer such questions at the atomic level. Herein, we have studied the cytoplasmic domain and performed rigorous simulations to unravel the structural dynamics of least explored cytoplasmic domain of essential spike protein.

3.1. *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 (Cai et al., 2020). However, based on SARS-CoV and SARS-CoV-2 proteins sequence alignment, approximately 77% similarity is found among both viruses spike proteins (Giri et al., 2020). 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 (Fig. 2). A



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

Fig. 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. All predictors work as standalone except CCTOP which works based on consensus from multiple predictors. Considering this, the cytoplasmic region of spike is chosen from residues 1242–1273 as CCTOP has predicted residues 1216–1241 in transmembrane region.

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: NH2-SCLKGCCSCGSCCKFDEDDSEPVLKGVKLHYT-COOH).

3.1.1. *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 (Giri et al., 2020). Again, we analyzed the disorderedness in selected cytoplasmic region using multiple predictors, including PONDR family, IUPred2A (redox state), and PrDOS predictors. Out of six predictors, three predictors from PONDR family have predicted it as highly disordered, PrDOS has

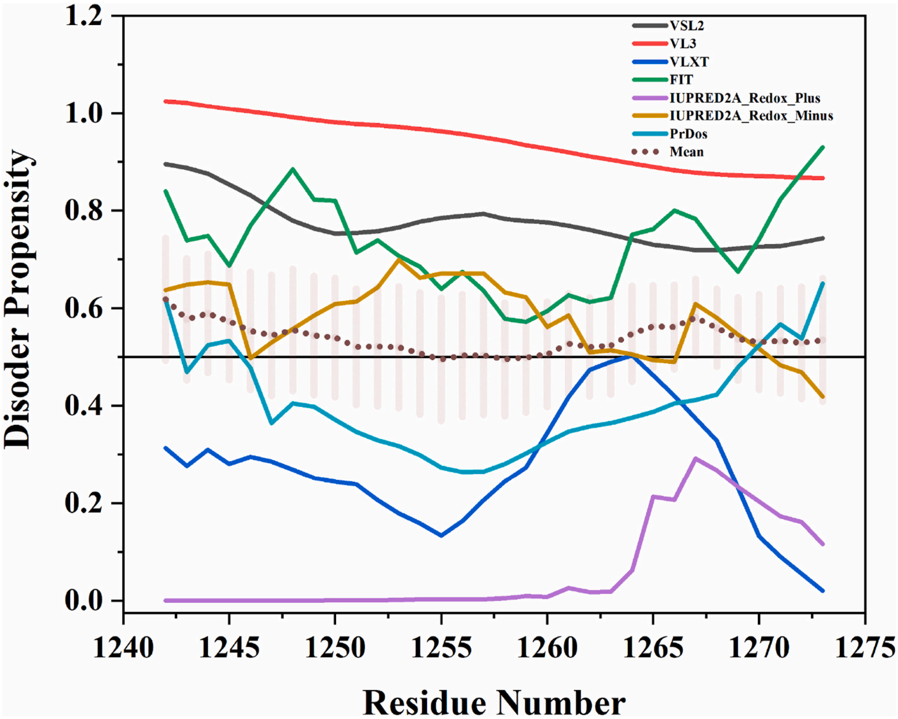


Fig. 3. Intrinsic disorder analysis of spike C-terminal cytoplasmic tail (residues 1242–1273) region using six predictors including PONDR family (VSL2, VL3, VLXT, and FIT), IUPred2A (Redox Plus and Minus), and PrDOS servers. The mean line is denoted in short dots style, and the standard error bars on mean are also highlighted.

predicted it as moderately disordered, and PONDR FIT has predicted it as least disordered. Additionally, IUPred2A has been used with its redox-state calculation function due to high number of cysteine residues present in the peptide (Fig. 3). As per calculations, the redox minus (where all cysteines are replaced by serine) state has shown high disorder propensity while redox plus has shown least propensity.

3.1.2. *Structure modelling with AlphaFold2 and PEPFOLD3*

In absence of an experimentally determined 3D structure of protein, structure modelling provides an approximate structure model based on homology and properties of amino acids using a wide range of optimized algorithms. So far, multiple 3D structures are made available for spike glycoprotein but they do not compose transmembrane and subsequent cytoplasmic regions. Therefore, we have modelled a full-length (residues 1–1273) structure using AlphaFold2, where a single membrane-pass like helical region is observed with unstructured cytoplasmic region (Fig. 4A).

Further, for cytoplasmic region in isolation, we have used PEP-FOLD3, which 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 one small helical region is further prepared for MD simulations in aqueous conditions. The helical region is present at residues $_{\mathbf{1265}}$LKGV$_{\mathbf{1268}}$ of spike glycoproteins C-terminus (Fig. 4B).

3.1.3. *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 (Gadhave et al., 2020a, 2020b, 2020a). 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 modelling through PEP-FOLD, the model contains one small helix at C-terminal with residues $_{\mathbf{1265}}$LKGV$_{\mathbf{1268}}$ (Fig. 4B). According to *2struc* webserver (Klose et al., 2010), the total helix propensity contribute approximately 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 β-strand of five residues $_{\mathbf{1262}}$EPVLK$_{\mathbf{1266}}$, as predicted by PSIPRED webserver (Buchan and Jones, 2019) (Fig. 4C).

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 (Fig. 5A). As shown in Fig. 5B, mean fluctuation in residues is observed to be within the range of 1.6–6.4 Å. The intramolecular hydrogen bond analysis demonstrates the highly fluctuating trend portraying no stable helical or beta sheet conformation adoption by the residues (Fig. 5E). The secondary structure timeline (Fig. 5C & D) 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 Fig. 5F and the trajectory movie up to 1 μs (supplementary movie 1**).**

We have also modelled the cytosolic part of Spike protein from 1235 to 1273 residues as defined in Uniprot database and two predictors (TMPred: 1216–1235, and TMHMM: 1214–1236) used in this study. In modelled structure, the helical propensity in cytosolic region was shown by 1237–1245 residues. Using above described OPLS 2005 forcefield parameters, the all-atoms explicit solvent MD simulation was carried out for 1 μs. The trajectory analysis has been shown in Supplementary Fig. 1, the cytosolic region has revealed majorly unstructured region along with a small β-strand of two residues $_{1258}$FD$_{1259}$ after 1 μs. The upward trend of RMSD values illustrates the highly deviating atomic positions and fluctuating RMSF shows the change in structural property of residues (Supplementary Figs. 1A and 1B). Also, the decreasing number of hydrogen bonds demonstrates the breaking of helices in the structure (Supplementary Fig. 1C). The time-dependent secondary structure element analysis illustrates that a total of 15% secondary structure was formed that includes mainly alpha helix and small percentage of beta strands (Supplementary Figs. 1D and 1E; red: alpha helix and blue: beta strands). After huge structural transitions, the structural composition of

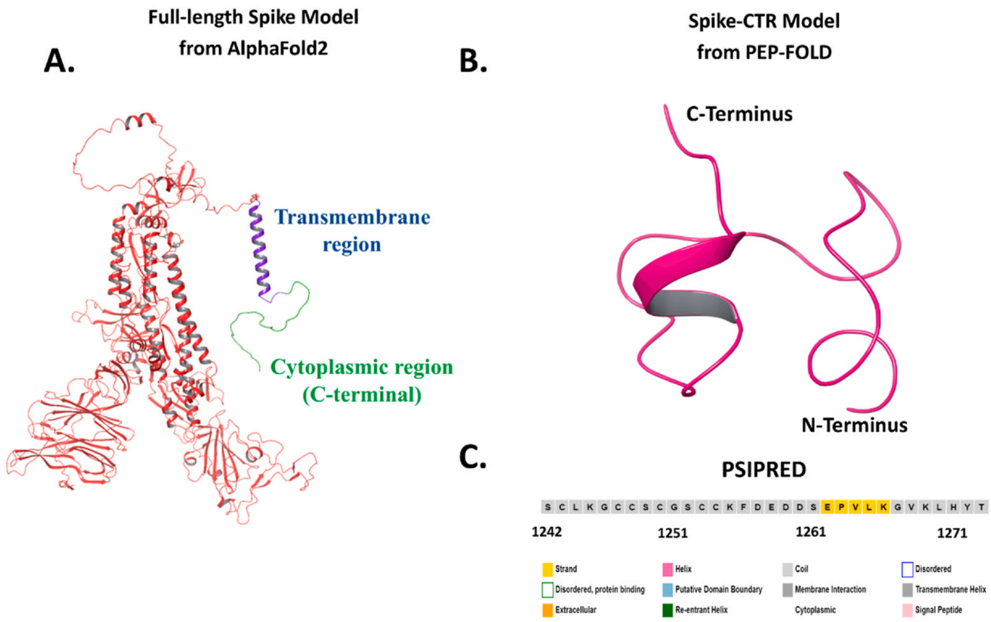


Fig. 4. Structure models of spike full-length and C-terminal cytoplasmic domain (1242–1273 residues): A. Full-length spike protein model using AlphaFold2, **B.** Modelled structure through PEP-FOLD web server, visualized in Maestro, and **C.** Secondary structure analysis using PSIPRED web server.

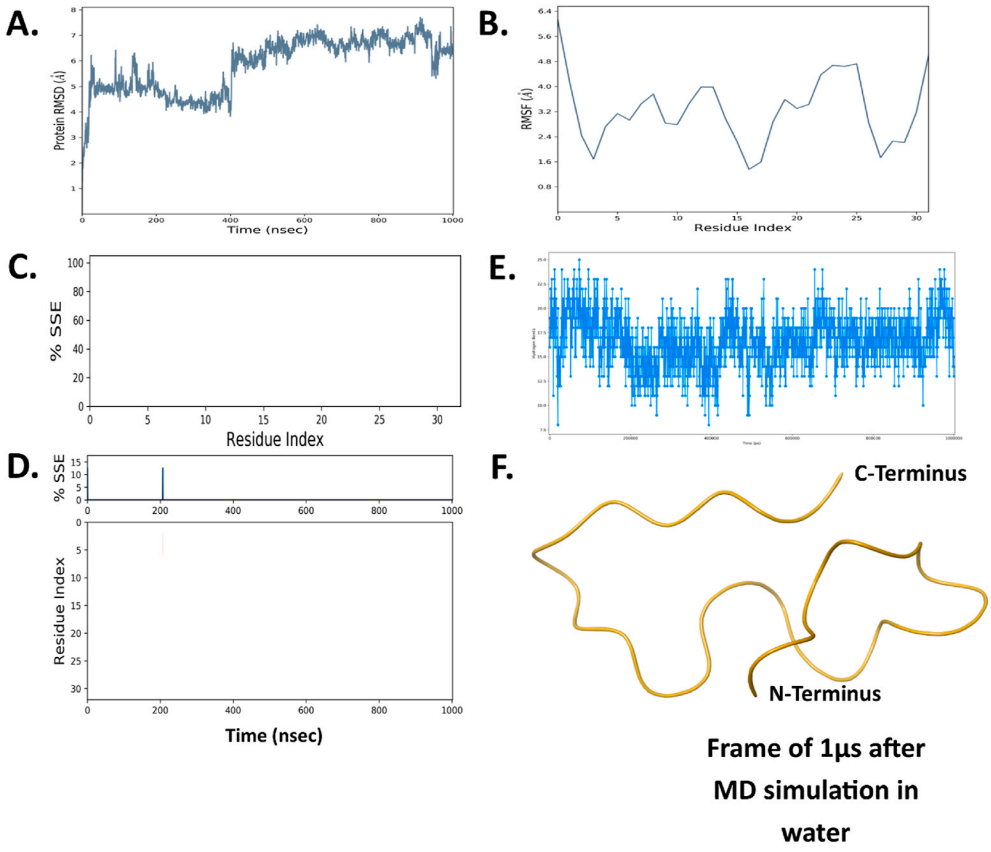


Fig. 5. One microsecond MD Simulation analysis of spike C-terminal cytoplasmic domain (1242–1273) using OPLS 2005 forcefield: A. Root mean square deviation (RMSD), B. Root mean square fluctuation (RMSF), C. Secondary structure element (SSE) of residues, D. Timeline representation of secondary structure content during 1 μs simulation time, E. Hydrogen bonds of protein and F. Last frame at 1 μs.

last frame of simulation is shown with a small beta strand of two residues and other regions to be disordered (Supplementary Fig. 1F). The snapshots at every 100 ns till 1 μs show the structural transitions in Spike cytoplasmic region (Supplementary Fig. 2).

3.1.4. *Desmond trajectory clustering*

In a microsecond simulation trajectory, a large amount of data can be sampled and simply understood using techniques like clustering (Shao et al., 2007). Desmond applies affinity propagation algorithm to produce the representatives of each identified cluster (Frey and Dueck, 2007). It calculates a similarity matrix for sampling the frames of trajectory. Here, we have done trajectory clustering based on RMSD values of backbone atoms of Spike cytoplasmic tail region with a timer threshold of 0.2. A total 15 clusters with size 39 to 4 structures were identified, out of which top 10 are shown here in Fig. 6. The RMSD values of these identified structures in reference to the first frame of trajectory is in the range of 4.53–6.85 Å. All cluster representative structures are fully unstructured and show no propensity for helical or beta sheet conformations in aqueous condition.

3.1.5. *Simulation with CHARMM36 m*

The characterization of intrinsically disordered proteins (IDPs) and regions (IDPRs) using MD simulations is now 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, CHARMM36 m, for determining the conformational dynamics of spike cytosolic region. As shown through trajectory snapshots at every 100 ns in Fig. 7, 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 at residues ${}_{1257}$KFD${}_{1259}$ 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–1.75 nm (Fig. 8A); the mean residual fluctuations are in the range of 0.5 nm–1.1 nm (Fig. 8C), and the Rg values vary up to 2.4 nm (Fig. 8B), 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 (Fig. 8E). As calculated in VMD (Humphrey et al., 1996), the total number of salt bridges are found to be 11 in the trajectory between the residues Asp1257-Lys1269, Glu1258--Lys1266, Glu1258-Lys1245, Asp1259-Lys1269, Glu1262-His1271, Glu1262-Lys1266, Glu1262-Lys1269, Asp1260-Lys1266, Asp1259-His1271, Asp1259-Lys1266, and Asp1260-Lys1269. The free energy landscape of complexation of Spike cytosolic region with variables RMSD and Radius of gyration is calculated using a Python script *generateFES.py* (http://www.strodel.info/index_files/lecture/html/media/generateFES.py) which reveals that the most of the conformations are lying with least energy from 0 to 4 kcal/mol. Convincingly, it is evident that a major part of spike cytosolic region is disordered.

3.1.6. *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 up to half a microsecond using 8 replicas (numbered as 0 to 7). As shown here in

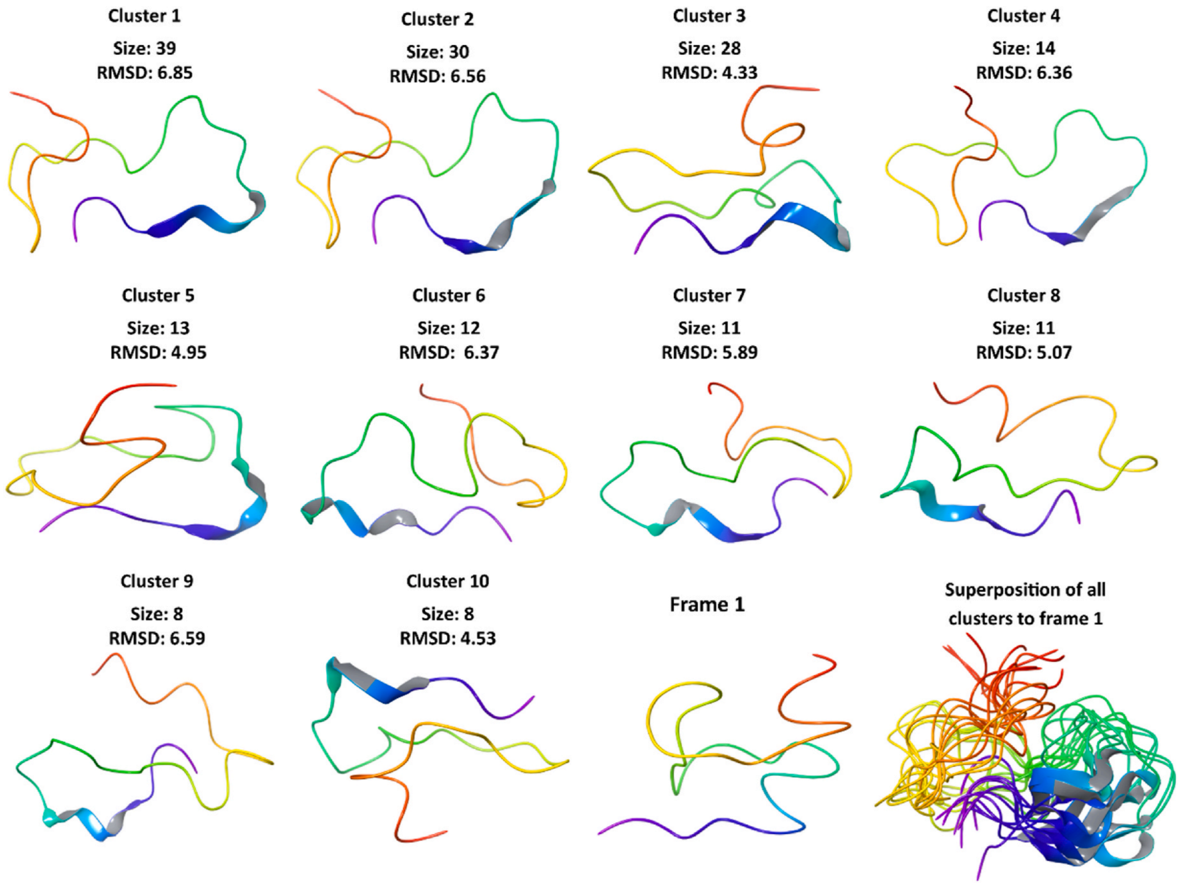


Fig. 6. Depiction of representative from top 10 clusters from 1 μs simulation trajectory. Size of each cluster is shown which represent the total number of frames in the trajectory based on RMSD calculated with reference structure (frame 1). The protein backbone of all frames is shown as superimposition. The N- to C-terminal protein structures are colored from red to blue, respectively.

Fig. 9, the cytoplasmic region has adopted a β-sheet structure at increasing temperatures up to 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 Fig. 9, the cytosolic region has gained three to four β-strands and appears to be in a well-folded manner as temperature increases.

For a clear understanding of all frames in REMD, timelines of all 8 replicas are displayed in Fig. 10 where the consistent formation of β-strands due to rising temperature is also validated. However, at extreme temperature (410 K), some structural changes are reversed, and the total secondary structure element (SSE) gets reduced in comparison to previous replicas (Fig. 10 Table). According to mean distances analyses, huge fluctuation is observed in replicas 3, 5, and 6 where structural changes occurred (Fig. 11A, B, 11D). As elucidated through hydrogen bond analysis (Fig. 11C), 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 (Fig. 11E).

3.2. CD spectroscopy analysis

3.2.1. Spike cytoplasmic region is unstructured in isolation

As observed through MD simulations using two different forcefields, the spike cytoplasmic region (residues 1242–1273) is unstructured or disordered. To validate our findings experimentally, we have performed CD spectroscopy-based experiments of the synthesized peptide of same region at 25 μM concentration in sodium phosphate buffer at physiological pH 7.4. In a good correlation with computer simulations, the spike cytoplasmic region has been confirmed to be disordered with a signature negative ellipticity peak at 198 nm (Fig. 12). Moreover, we recorded the CD spectra both in absence and presence of reducing agent. As a reducing agent, 1 mM 'Dithiothreitol (DTT)' was used with Spike-CTR. Interestingly, in the presence of DTT, the CD spectra of Spike-CTR peptide (25 μM) has shown a disordered nature with the negative peak at 197 nm with ellipticity −7.5 deg cm^2 dmol^{-1} in DTT, respectively (Fig. 12).

Further, we have also investigated the gain-of-structure property of spike cytoplasmic region. For this purpose, we have employed a well-known secondary structure inducer, 2,2,2-trifluoroethanol (TFE). TFE weakens H-bonds between NH and CO of the protein backbone with surrounding H_2O and if observed to stabilize the intra-chain H-bonds during secondary structure formation (Luo and Baldwin, 1997). In presence of the organic solvent TFE, spike cytoplasmic region (at 25 μM) is observed to attain an α-helical structure. The peptide slowly started bending towards the hallmark helical negative peaks at 208 nm and 222 nm before 30% of TFE concentration (Fig. 13A and B). After increment to 40% TFE, the helical peaks having negative ellipticity at 208 nm and 222 nm are significantly clear representing the structural transformation of spike terminal tail from disordered to α-helical indicating its gain-of-structure property.

Next, we used sodium dodecyl sulfate (SDS), an artificial membrane-mimicking micelle forming ionic detergent to evaluate the fundamental gain-of-structure property of spike C-terminal tail. It is known to bind positively charged and hydrophobic residues of proteins using its alkyl chains and sulfate groups (Hansen et al., 2009). Contrarily to TFE, on addition of SDS, no significant change in disordered nature is observed.

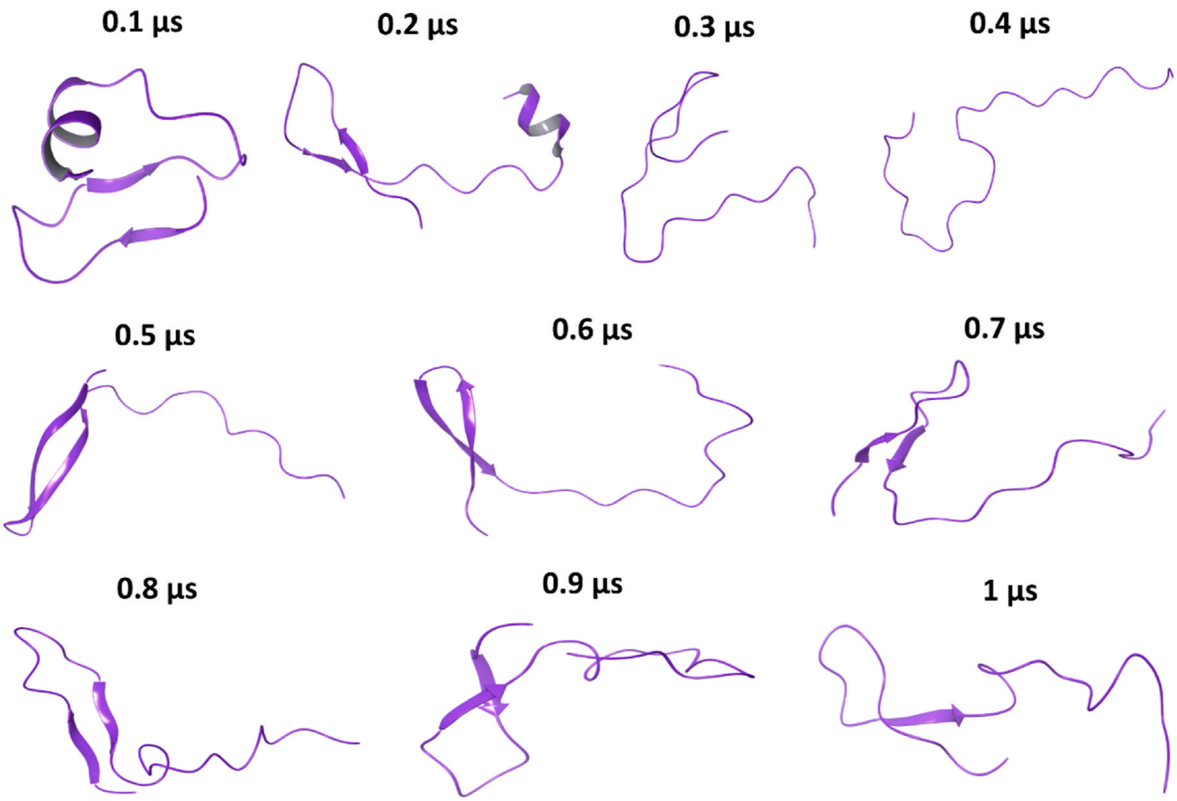


Fig. 7. MD simulation with CHARMM36m forcefield: Snapshots at every 100 ns of 1 µs simulation trajectory.

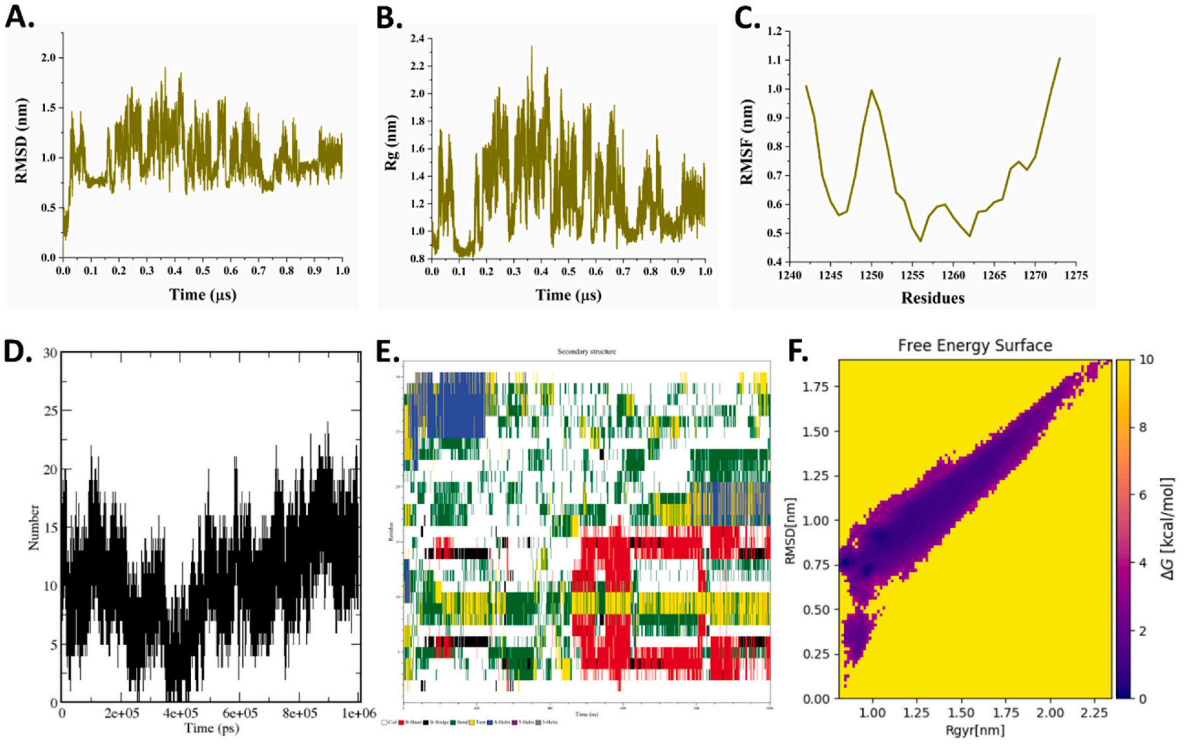


Fig. 8. Trajectory analysis of spike cytoplasmic region with CHARMM36m forcefield: A. RMSD, B. Radius of gyration (Rg), C. RMSF, D. Hydrogen bonds and E. Secondary structure element timeline during the course of 1 µs simulation time, and F. Free energy landscape of simulation trajectory using RMSD and Radius of gyration as variables.

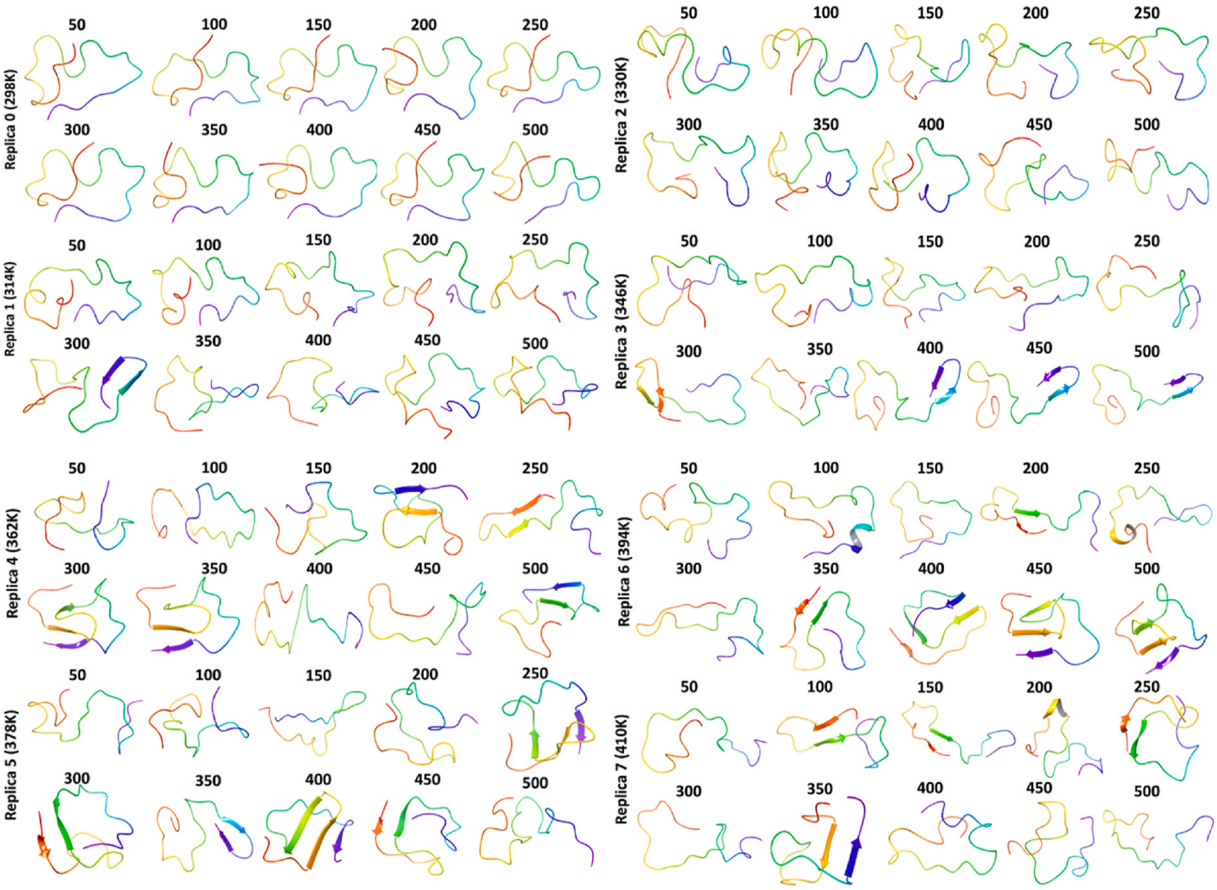


Fig. 9. Snapshots from all replicas at every 50 ns during half a microsecond REMD simulation. The N- to C-terminal protein structures are colored from red to blue, respectively.

As evident from Fig. 13C and D, not much gain in negative ellipticity at 208 nm and 222 nm is detected using CD spectroscopy.

In presence of sucrose, an increment in negative ellipticity from −2 deg cm^2 dmol^{-1} to −10 deg cm^2 dmol^{-1} is observed in the far-UV CD spectra (with wavelength 200–240 nm) with increasing concentrations up to 300 g/L. These results also demonstrate the unstructured nature of the peptide in the presence of sucrose.

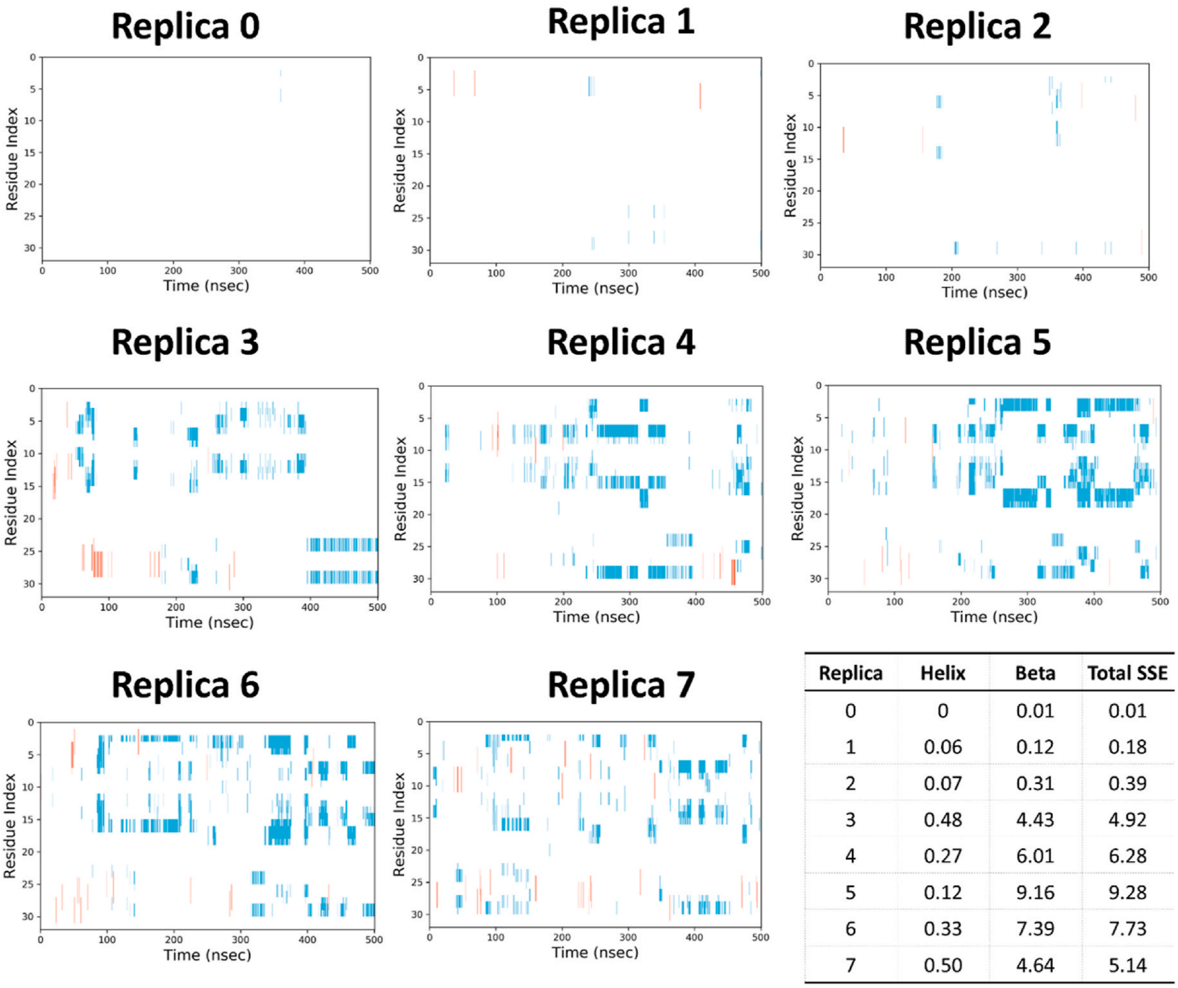
3.3. *Spike cytoplasmic region shows no change in presence of macromolecular crowders*

It is a known fact that intracellular region is highly crowded and occupies a volume of 5–40% in the cell (Kuznetsova et al., 2014). Crowding conditions can substantially affect several thermodynamic processes such as protein folding, change of conformation, and protein aggregation, etc. (Kuznetsova et al., 2014). Therefore, in order to extrapolate the cellular conditions and behavior of spike cytoplasmic tail in crowded environment, we have used two widely employed macromolecular crowders PEG 8000 and Dextran-70 to investigate the change in conformation of spike endo-domain. Here, according to our observations, in presence of either of the crowders at high concentrations up to 300 g/L (generally 80–400 g/L in cells (Biswas et al., 2018)), no change in secondary structure is observed. At all concentrations of PEG 8000, the negative ellipticity at 198 nm is seen to be similar to disordered spectra with only a minimal shift in wavelength (Fig. 14A). Whereas, in presence of another crowder, Dextran-70, the disordered like nature is appeared to remain unchanged as there is no shift in the spectra at any concentration up to 300 g/L of Dextran-70. Due to high absorption by Dextran-70 at lower wavelengths, the spectra are truncated after 200 nm. (Fig. 14B). Overall, no noticeable structural transitions are observed in spike cytoplasmic region in presence of crowders. These results are indicative of absence of weaker hydrophobic forces and electrostatic interactions among residues. These observations clearly explain that the spike endodomain or cytoplasmic domain is intrinsically disordered even under macromolecular crowding conditions.

4. Discussion

The existence of IDPs/IDPRs was established based on non-existing electron density of some regions in proteins during crystallography (Uversky, 2019, 2020). This unmapped electron density of characterized protein was later described as intrinsic disorderness of proteins. With increasingly published articles and revelation of presence of IDPs in frequently studied living systems, IDPs have been found prevalent in all three domains of life (Uversky, 2019). For a long period, the structure elucidation of transmembrane regions was not feasible due to their complex process of purification and crystallization (Tusnády et al., 2015). It was also reported in a study on survey of disordered regions in eukaryotic proteins that the water-soluble cytoplasmic tails of membrane spanning proteins constitute intrinsic disorder (Minezaki et al., 2007). These disordered regions are generally rich in certain residues such as proline, glycine, lysine and serine, and contains hydrophobic residues to form a core while folding (Hansen et al., 2006). The presence of disordered regions in proteins makes crystallization process an unsuccessful task and compel to produce a truncated structure.

In past decade, several reports of IDPs and IDRs in virus proteomes have exposed the “disordered” side of viral systems. It has answered some intriguing questions related to numerous protein-protein interactions led by viral proteins with host proteins required for hijacking



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

Fig. 10. Timeline representation of simulation trajectories from all replicas up to half a microsecond simulation time. Red and blue lines in the timeline represent alpha-helix and beta-strands, respectively. A table of percentage secondary structure elements in each simulation.

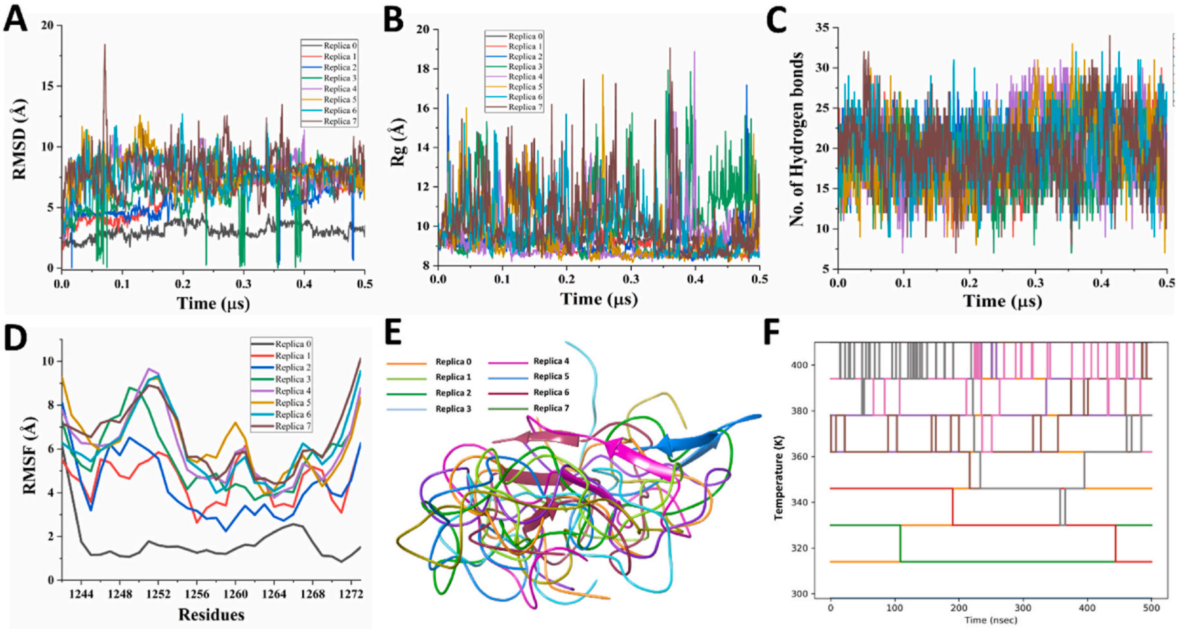


Fig. 11. Trajectory analysis from all replicas simulated at different temperatures: A. RMSD, B. Rg, C. Number of hydrogen bonds, D. RMSF, E. Superimposed last frames, and F. Conformation exchange review during REMD.

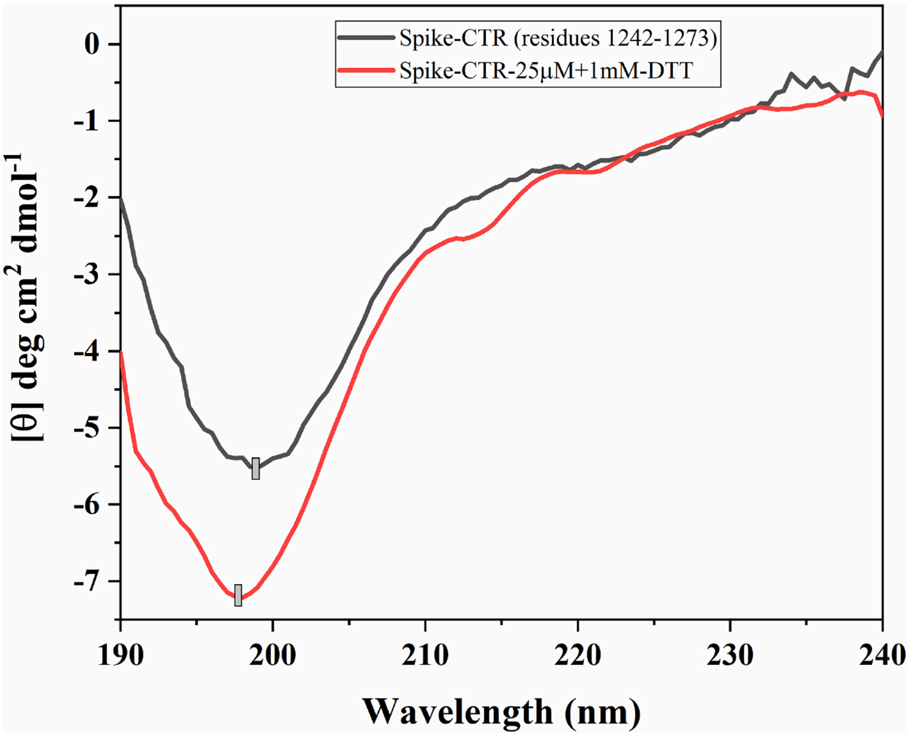


Fig. 12. Disordered nature of spike cytoplasmic region (residues 1242–1273): Far UV-CD spectra of 25 μM spike cytoplasmic region in presence of 20 mM sodium phosphate buffer (pH 7.4) is recorded from 190 nm to 240 nm (in black). Additionally, CD spectra are also recorded in presence of reducing agent, Dithiothreitol (DTT; in red). The significant negative ellipticity at 198 nm is characteristic of a disordered conformation.

the entire host cell (Uversky, 2015). Our detailed investigation on disordered regions in SARS-CoV and SARS-CoV-2 proteomes showed the presence of large flexible region in various proteins like nucleocapsid, ORF6, and Nsp8 proteins. The report also revealed the presence of intrinsic disorder in N- and C-terminal tails of most of the proteins (Giri et al., 2020). It also exposed that the C-terminal region of spike protein of SARS-CoV-2 contain disordered regions. Recently, we have shown that the C-terminal regions of SARS-CoV-2 NSP1 and Envelope proteins are disordered in isolation (Gadhave et al., 2020a; A. Kumar et al., 2021).

Due to their dynamic nature, IDPs cannot be crystallized or frozen to be detected using such advanced structural biology techniques. X-ray crystallography and cryo-electron microscopy-based structures of SARS-CoV-2 spike proteins on PDB show missing electron density in transmembrane and cytosolic regions of C-terminal. Some of the structures of spike protein mentioned here with PDB IDs – 6VXX, 6ZGH, 6ZGG, 6XM3 and 6XM4, have reported the C-terminal tail as either unmodeled or as an artefact. Moreover, the consensus prediction of disorderness by MobiDB has also supported this observation by showing the missing residues consensus of region 1147–1273 which include transmembrane as well as cytosolic regions. 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 (Lontok et al., 2004; McBride et al., 2007; Sadasivan et al., 2017). Notably, spike C-terminal tail contains multiple cysteine residues which may have implications in protein-protein interactions. In SARS-CoV, this cysteine-rich C-terminal domain of spike is responsible for interaction with M protein as a mutation in this domain obstructed their interaction (Bosch et al., 2005; de Haan et al., 1999).

Our study on the structural dynamics of SARS-CoV-2 spike cytoplasmic domain (residues 1242–1273) demonstrates it to be a disordered region. Based on the outcomes of two forcefields, OPLS 2005 and CHARMM36 m, spike cytosolic region remains majorly unstructured. Additionally, in another simulation run of 1 μs, the cytoplasmic region with residues 1235–1273 have also shown a large part to be disordered and a small beta strand in few frames. As observed, the residues $_{1257}$KFD$_{1259}$ have shown propensity to form beta strands in simulations. Nevertheless, in REMD simulations, it adopted β-sheets at rising temperatures with time demonstrating its gain-of-structure property. However, we also tried to get the synthesized peptide of residues 1235–1273 of spike but due to multiple cysteine residues it was not feasible.

Further, it was of utmost importance to validate MD simulation outcomes using experimental techniques. The water-soluble peptide of spike residues 1242–1273 at 25 μM concentration exhibits a prominent negative peak at approximately 198 nm in far-UV CD spectra which defines the unstructured nature of a protein. Infact, we have also checked the secondary structure state in presence of a reducing agent, DTT, then also, the peptide is observed to be disordered with significant negative ellipticity. Further, in presence of helix inducer solvent, TFE, the peptide adopts helical structure. However, SDS micelles in surroundings of peptide generates little changes in the peptide structure which may signify its inability to gain structure. Also, in presence of sucrose, the CD spectra of peptide corresponds to the disordered conformation. Under the influence of crowding agents like Dextran-70 and PEG (8000), conservation of disordered structure indicates that no -intra chain forces are acting in between the residues. Based on this combination of facts, we have interpreted that spike C-terminal cytosolic tail (residues 1242–1273) as an intrinsically disordered region. Generally, an IDPR gains any structure upon interacting with its interacting partner or in physiological conditions (Wright and Dyson, 1999). In its unstructured state, it may function as a MoRF to bind with COP1 coated transporting vesicles which localizes the 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.

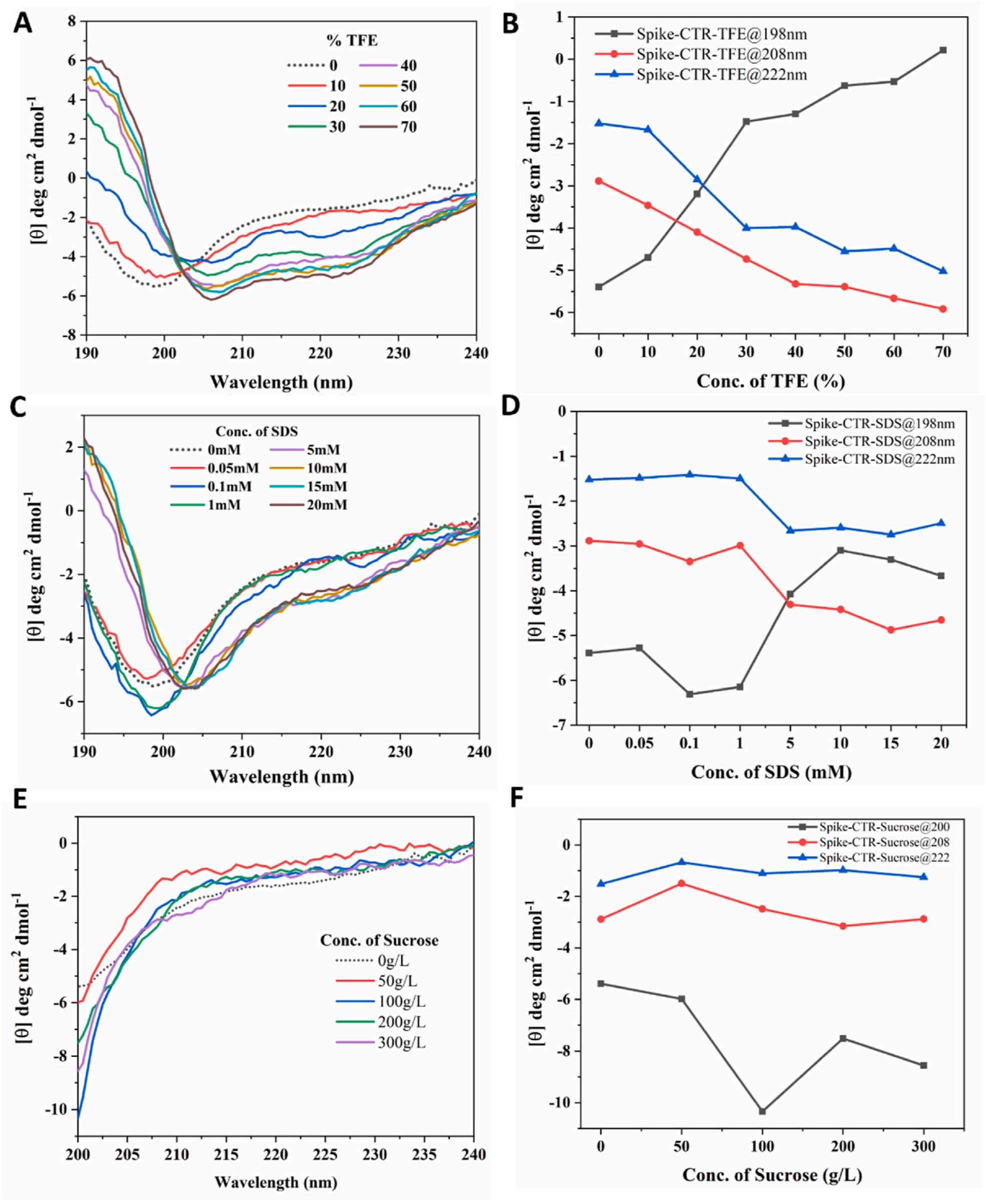


Fig. 13. Secondary structure analysis of spike cytoplasmic region (residues 1242–1273) using CD spectroscopy: A. Gain in secondary structure of spike cytoplasmic region in presence of different concentrations of a secondary structure inducer TFE (2,2,2-trifluroethanol). The spike cytoplasmic region started gaining helical structure (negative ellipticity peaks near to 208 nm and 222 nm) around 30% of TFE concentration. B. Plot showing the change in ellipticity at 198 nm, 208 nm and 222 nm in CD spectra of spike cytoplasmic region in presence of TFE C. In presence of SDS (Sodium Dodecyl Sulfate), no prominent change in unstructured nature is noticed. D. Plot represents the change in ellipticity at 198 nm, 208 nm and 222 nm in CD spectra of spike cytoplasmic region in presence of SDS. E. In presence of sucrose, no significant change is observed. F. Plot depicts the relationship of ellipticity at 198 nm, 208 nm and 222 nm in CD spectra with increasing concentration of sucrose.

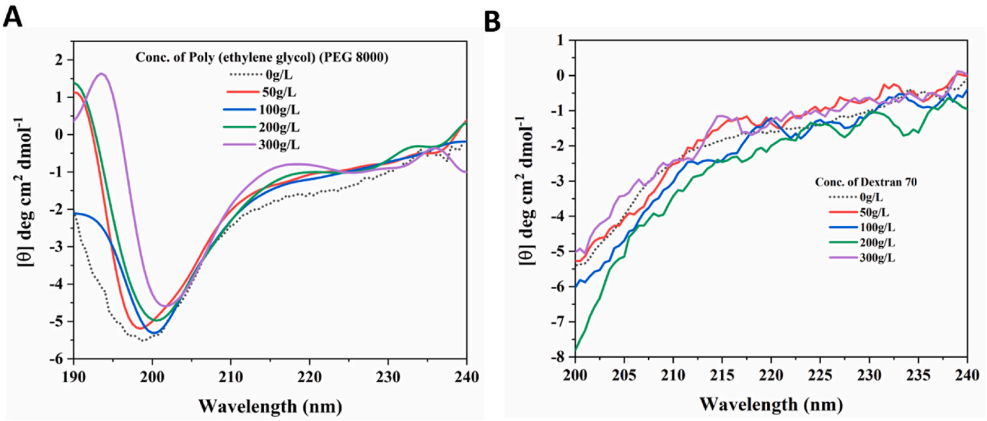


Fig. 14. Effect of macromolecular crowding on spike cytoplasmic region (residues 1242–1273): A. and B. represent the far UV CD spectra spike cytoplasmic region in presence of poly ethylene glycol (8000) and Dextran-70, respectively.

5. Conclusion

The cytoplasmic region of spike glycoprotein of SARS-CoV-2 has not been studied yet. Given its extreme importance in functioning of spike protein, the structure and its dynamics has been investigated here. 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 1 μs and have not gained any structural conformation throughout the simulation period. Experimental outcomes also confirm the intrinsic disordered state of cytoplasmic domain of spike. The intrinsic disordered nature of peptide is shown in presence of macromolecular crowders. Based on our previous study (Giri et al., 2020), cytoplasmic tail of spike glycoprotein has molecular recognition features therein which needs to be explored further. The disordered nature of cytosolic region may possibly have implications to interact with other viral proteins during virion assembly as well as host proteins and transporting vesicles during localization in ERs. In this study, the multiple conformations during the simulation process adds up to even more interesting speculations.

Data and software availability

All simulations and analysis were performed using Desmond (Driver v2.3) (https://www.deshawresearch.com/resources_desmond.html) and Gromacs (v2018.8) (https://www.gromacs.org) simulation packages under academic licence. Charmm36 (charmm36-mar2019.ff.tgz) forcefield was downloaded from http://mackerell.umaryland.edu/charmm_ff.shtml#gromacs. No new algorithm or forcefield was developed. Disorder and 3D model predictions were done with freely accessible web servers, mentioned in method section.

CRediT authorship contribution statement

Prateek Kumar: acquisition and interpretation of data, contributed to paper writing. **Taniya Bhardwaj:** acquisition and interpretation of data, contributed to paper writing. **Neha Garg:** Conception, design, Supervision. **Rajanish Giri:** Conception, design, Supervision, contributed to paper writing.

Declaration of competing interest

All authors affirm that there are no conflicts of interest.

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.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.virol.2021.11.005.

References

Berendsen, H.J.C., van der Spoel, D., van Drunen, R., 1995. GROMACS: a message-passing parallel molecular dynamics implementation. Comput. Phys. Commun. 91, 43–56. https://doi.org/10.1016/0010-4655(95)00042-E.

Biswas, S., Kundu, J., Mukherjee, S.K., Chowdhury, P.K., 2018. Mixed macromolecular crowding: a protein and solvent perspective. ACS Omega 3, 4316–4330. https://doi.org/10.1021/acsomega.7b01864.

Bosch, B.J., De Haan, C.A.M., Smits, S.L., Rottier, P.J.M., 2005. Spike protein assembly into the coronavirion: exploring the limits of its sequence requirements. Virology 334, 306–318. https://doi.org/10.1016/j.virol.2005.02.001.

Buchan, D.W.A., Jones, D.T., 2019. The PSIPRED protein analysis workbench: 20 years on. Nucleic Acids Res. 47, W402–W407. https://doi.org/10.1093/nar/gkz297.

Cai, Y., Zhang, J., Xiao, T., Peng, H., Sterling, S.M., Walsh, R.M., Rawson, S., Rits-Volloch, S., Chen, B., 2020. Distinct conformational states of SARS-CoV-2 spike protein. Science 84 369, 1586–1592. https://doi.org/10.1126/science.abd4251.

de Haan, C.A.M., Smeets, M., Vernooij, F., Vennema, H., Rottier, P.J.M., 1999. Mapping of the coronavirus membrane protein domains involved in interaction with the spike protein. J. Virol. 73, 7441–7452. https://doi.org/10.1128/jvi.73.9.7441-7452.1999.

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., 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. https://doi.org/10.1016/j.chom.2020.06.020 e6.

Dobson, L., Reményi, I., Tusnády, G.E., 2015. CCTOP: a Consensus Constrained TOPology prediction web server. Nucleic Acids Res. 43, W408–W412. https://doi.org/10.1093/nar/gkv451.

Duan, L., Zheng, Q., Zhang, H., Niu, Y., Lou, Y., 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. https://doi.org/10.3389/fimmu.2020.576622.

Frey, B.J., Dueck, D., 2007. Clustering by passing messages between data points. Science 315, 972–976. https://doi.org/10.1126/science.1136800.

Gadhave, K., Kumar, A., Kumar, P., Kapuganti, S.K., Garg, N., Vendruscolo, M., Giri, R., 2020a. Environmental dependence of the structure of the c-terminal domain of the SARS-CoV-2 envelope protein, bioRxiv. bioRxiv. https://doi.org/10.1101/2020.12.29.424646.

Gadhave, K., Kumar, P., Kumar, A., Bhardwaj, T., Garg, N., Giri, R., 2020b. NSP 11 of SARS-CoV-2 is an intrinsically disordered protein, 2020 bioRxiv. https://doi.org/10.1101/2020.10.07.330068, 10.07.330068.

Giri, R., Bhardwaj, T., Shegane, M., Gehi, B.R., Kumar, P., Gadhave, K., Oldfield, C.J., 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. https://doi.org/10.1007/s00018-020-03603-x.

Hansen, J.C., Lu, X., Ross, E.D., Woody, R.W., 2006. Intrinsic protein disorder, amino acid composition, and histone terminal domains. J. Biol. Chem. https://doi.org/10.1074/jbc.R500022200.

Hansen, J.H., Petersen, S.V., Andersen, K.K., Enghild, J.J., Damhus, T., Otzen, D., 2009. Stable intermediates determine proteins' primary unfolding sites in the presence of surfactants. Biopolymers 91, 221–231. https://doi.org/10.1002/bip.21125.

Hess, B., Bekker, H., Berendsen, H.J.C., Fraaije, J.G.E.M., 1997. LINCS: a linear constraint solver for molecular simulations. J. Comput. Chem. 18, 1463–1472. https://doi.org/10.1002/(SICI)1096-987X (199709)18:12<1463::AID-JCC4>3.0.CO;2-H.

Hofmann, K., Stoffel, W., 1993. A database of membrane spanning protein segments. Biol. Chem. 374.

Huang, J., Rauscher, S., Nawrocki, G., Ran, T., Feig, M., De Groot, B.L., Grubmüller, H., MacKerell, A.D., 2016. CHARMM36m: an improved force field for folded and intrinsically disordered proteins. Nat. Methods 14, 71–73. https://doi.org/10.1038/nmeth.4067.

Humphrey, W., Dalke, A., Schulten, K., 1996. VMD: visual molecular dynamics. J. Mol. Graph. 14, 33–38. https://doi.org/10.1016/0263-7855(96)00018-5.

Ishida, Takashi, Kinoshita, Kengo, Ishida, T., Kinoshita, K., 2007. PrDOS: prediction of disordered protein regions from amino acid sequence. Nucleic Acids Res. 35, W460–W464. https://doi.org/10.1093/nar/gkm363.

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S.A.A., Ballard, A.J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., Back, T., Petersen, S., Reiman, D., Clancy, E., Zielinski, M., Steinegger, M., Pacholska, M., Berghammer, T., Bodenstein, S., Silver, D., Vinyals, O., Senior, A.W., Kavukcuoglu, K., Kohli, P., Hassabis, D., 2021. Highly accurate protein structure prediction with AlphaFold. Nature 596, 583–589. https://doi.org/10.1038/s41586-021-03819-2.

Juretić, D., Zoranić, L., Zucić, D., 2002. Basic charge clusters and predictions of membrane protein topology. J. Chem. Inf. Comput. Sci. 42, 620–632. https://doi.org/10.1021/ci010263s.

Klose, D.P., Wallace, B.A., Janes, R.W., 2010. 2Struc: the secondary structure server. Bioinformatics 26, 2624–2625. https://doi.org/10.1093/bioinformatics/btq480.

Krogh, A., È rn Larsson, B., von Heijne, G., L Sonnhammer, E.L., 2001. Predicting transmembrane protein topology with a hidden markov model: application to complete genomes. J. Mol. Biol. 305, 567–580. https://doi.org/10.1006/jmbi.2000.4315.

Kumar, D., Singh, A., Kumar, P., Uversky, V.N.V.N., Rao, C.D.D., Giri, R., 2020. Understanding the penetrance of intrinsic protein disorder in rotavirus proteome.

Int. J. Biol. Macromol. 144, 892–908. https://doi.org/10.1016/j.ijbiomac.2019.09.166.

Kumar, A., Kumar, Ankur, Kumar, P., Garg, N., Giri, R., 2021a. SARS-CoV-2 NSP1 C-terminal (residues 131–180) is an intrinsically disordered region in isolation. Curr. Res. Virol. Sci. 2, 100007. https://doi.org/10.1016/j.crviro.2021.100007.

Kumar, P., Sharma, N., Kumar, A., Giri, R., 2021b. Molecular dynamic simulation of intrinsically disordered proteins and relevant forcefields. Innov. Implementations Comput. Aided Drug Discov. Strateg. Ration. Drug Des. 317–333. https://doi.org/10.1007/978-981-15-8936-2_13.

Kuznetsova, I.M., Turoverov, K.K., Uversky, V.N., 2014. What macromolecular crowding can do to a protein. Int. J. Mol. Sci. https://doi.org/10.3390/ijms151223090.

Lan, J., Ge, J., Yu, J., Shan, S., Zhou, H., Fan, S., Zhang, Q., Shi, X., Wang, Q., Zhang, L., Wang, X., 2020. Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. Nature 581, 215–220. https://doi.org/10.1038/s41586-020-2180-5.

Lontok, E., Corse, E., Machamer, C.E., 2004. Intracellular targeting signals contribute to localization of coronavirus spike proteins near the virus assembly site. J. Virol. 78, 5913–5922. https://doi.org/10.1128/jvi.78.11.5913-5922.2004.

Luo, P., Baldwin, R.L., 1997. Mechanism of helix induction by trifluoroethanol: a framework for extrapolating the helix-forming properties of peptides from trifluoroethanol/water mixtures back to water. Biochemistry 36, 8413–8421. https://doi.org/10.1021/bi9707133.

Martyna, G.J., Klein, M.L., Tuckerman, M., 1992. Nosé-Hoover chains: the canonical ensemble via continuous dynamics. J. Chem. Phys. 97, 2635–2643. https://doi.org/10.1063/1.463940.

Martyna, G.J., Tobias, D.J., Klein, M.L., 1994. Constant pressure molecular dynamics algorithms. J. Chem. Phys. 101, 4177–4189. https://doi.org/10.1063/1.467468.

McBride, C.E., Li, J., 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. https://doi.org/10.1128/jvi.02146-06.

Mészáros, B., Erdös, G., 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. https://doi.org/10.1093/nar/gky384.

Minezaki, Y., Homma, K., Nishikawa, K., 2007. Intrinsically disordered regions of human plasma membrane proteins preferentially occur in the cytoplasmic segment. J. Mol. Biol. 368, 902–913. https://doi.org/10.1016/j.jmb.2007.02.033.

Nugent, T., Jones, D.T., 2009. Transmembrane protein topology prediction using support vector machines. BMC Bioinf. 10, 159. https://doi.org/10.1186/1471-2105-10-159.

Obradovic, Z., Peng, K., Vucetic, S., Radivojac, P., Brown, C.J., Dunker, A.K., 2003. Predicting intrinsic disorder from amino acid sequence. In: Proteins: Structure, Function, and Genetics. Proteins, pp. 566–572. https://doi.org/10.1002/prot.10532.

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., Qian, Z., 2020. Characterization of spike glycoprotein of SARS-CoV-2 on virus entry and its immune cross-reactivity with SARS-CoV. Nat. Commun. 11 https://doi.org/10.1038/s41467-020-15562-9.

P, R., Z, O., X, L., Ec, G., Cj, B., Ak, D., Romero, P., Obradovic, Z., Li, X., Garner, E.C., Brown, C.J., Dunker, A.K., 2001. Sequence complexity of disordered protein. Proteins 42, 38–48. https://doi.org/10.1002/1097-0134(20010101)42:1<38::AID-PROT50>3.0.CO, 2-3.

Piovesan, D., Necci, M., Escobedo, N., Monzon, A.M., Hatos, A., Mičetić, I., Quaglia, F., Paladin, L., Ramasamy, P., Dosztányi, Z., Vranken, W.F., Davey, N.E., Parisi, G., Fuxreiter, M., Tosatto, S.C.E., 2021. MobiDB: intrinsically disordered proteins in 2021. Nucleic Acids Res. 49, D361–D367. https://doi.org/10.1093/nar/gkaa1058.

Sadasivan, J., Singh, M., Sarma, J. Das, 2017. Cytoplasmic tail of coronavirus spike protein has intracellular targeting signals. J. Biosci. 42, 231–244. https://doi.org/10.1007/s12038-017-9676-7.

Shao, J., Tanner, S.W., Thompson, N., Cheatham, T.E., 2007. Clustering molecular dynamics trajectories: 1. Characterizing the performance of different clustering algorithms. J. Chem. Theor. Comput. 3, 2312–2334. https://doi.org/10.1021/ct700119m.

Shaw, D.E., 2005. A fast, scalable method for the parallel evaluation of distance-limited pairwise particle interactions. J. Comput. Chem. 26, 1318–1328. https://doi.org/10.1002/jcc.20267.

Shen, Y., Maupetit, J., Derreumaux, P., Tufféry, P., 2014. Improved PEP-FOLD approach for peptide and miniprotein structure prediction. J. Chem. Theor. Comput. 10, 4745–4758. https://doi.org/10.1021/ct500592m.

Sugita, Y., Okamoto, Y., 1999. Replica-exchange molecular dynamics method for protein folding. Chem. Phys. Lett. 314, 141–151. https://doi.org/10.1016/S0009-2614(99)01123-9.

Tusnády, G.E., Dobson, L., Tompa, P., 2015. Disordered regions in transmembrane proteins. Biochim. Biophys. Acta Biomembr. 2839–2848. https://doi.org/10.1016/j.bbamem.2015.08.002, 1848.

Ujike, M., Huang, C., Shirato, K., Makino, S., 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. https://doi.org/10.1099/jgv.0.000494.

Uversky, V.N., 2015. Paradoxes and wonders of intrinsic disorder: prevalence of exceptionality. Intrinsically Disord. Proteins. https://doi.org/10.1080/21690707.2015.1065029.

Uversky, V.N., 2019. Intrinsically disordered proteins and their “Mysterious” (meta) physics. Front. Physiol. 7, 10. https://doi.org/10.3389/fphy.2019.00010.

Uversky, V.N., 2020. Intrinsically disordered proteins. In: Structural Biology in Drug Discovery. Wiley, pp. 587–612. https://doi.org/10.1002/9781118681121.ch25.

Walls, A.C., Park, Y.J., Tortorici, M.A., Wall, A., McGuire, A.T., Veesler, D., 2020. Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein. Cell 181, 281–292. https://doi.org/10.1016/j.cell.2020.02.058 e6.

Wright, P.E., Dyson, H.J., 1999. Intrinsically unstructured proteins: Re-assessing the protein structure-function paradigm. J. Mol. Biol. 293, 321–331. https://doi.org/10.1006/jmbi.1999.3110.

Xue, B., Dunbrack, R.L., Williams, R.W., Dunker, A.K., Uversky, V.N., 2010. PONDR-FIT: a meta-predictor of intrinsically disordered amino acids. Biochim. Biophys. Acta Protein Proteonomics 996–1010. https://doi.org/10.1016/j.bbapap.2010.01.011, 1804.