Received: 23 November 2020 | Revised: 18 February 2021 | Accepted: 9 April 2021

DOI: 10.1002/prot.26088

RESEARCH ARTICLE

PROTEINS WILEY

Tackling Covid-19 using disordered-to-order transition of residues in the spike protein upon angiotensin-converting enzyme 2 binding

Dhanusha Yesudhas[1] | Ambuj Srivastava[1] | Masakazu Sekijima[2] | M. Michael Gromiha[1,2]

[1]Department of Biotechnology, Bhupat and Jyoti Mehta School of Biosciences, Indian Institute of Technology Madras, Chennai, India

[2]School of Computing, Tokyo Tech World Research Hub Initiative (WRHI), Institute of Innovative Research, Tokyo Institute of Technology, Yokohama, Kanagawa, Japan

Correspondence
M. Michael Gromiha, Department of Biotechnology, Bhupat and Jyoti Mehta School of Biosciences, Indian Institute of Technology Madras, Chennai 600036, India.
Email: gromiha@iitm.ac.in

Funding information
Department of Science and Technology, Government of India, Grant/Award Number: MSC/2020/000319; Japan Agency for Medical Research and Development, Grant/Award Number: JP20am0101112

Abstract

The 2019-novel coronavirus also known as severe acute respiratory syndrome-coronavirus 2 (SARS-CoV-2) is a common threat to animals and humans, and is responsible for the human SARS pandemic in 2019 to 2021. The infection of SARS-CoV-2 in humans involves a viral surface glycoprotein named as spike proteins, which bind to the human angiotensin-converting enzyme 2 (ACE2) proteins. Particularly, the receptor binding domains (RBDs) mediate the interaction and contain several disordered regions, which help in the binding. Investigations on the influence of disordered residues/regions in stability and binding of spike protein with ACE2 help to understand the disease pathogenesis, which has not yet been studied. In this study, we have used molecular-dynamics simulations to characterize the structural changes in disordered regions of the spike protein that result from ACE2 binding. We observed that the disordered regions undergo disorder-to-order transition (DOT) upon binding with ACE2, and the DOT residues are located at functionally important regions of RBD. Although the RBD is having rigid structure, DOT residues make conformational rearrangements for the spike protein to attach with ACE2. The binding is strengthened via hydrophilic and aromatic amino acids mainly present in the DOTs. The positively correlated motions of the DOT residues with its nearby residues also explain the binding profile of RBD with ACE2, and the residues are observed to be contributing more favorable binding energies for the spike-ACE2 complex formation. This study emphasizes that intrinsically disordered residues in the RBD of spike protein may provide insights into its etiology and be useful for drug and vaccine discovery.

KEYWORDS
coronavirus, COVID-19, disorder-to-order transition, MD simulation, spike protein

1 | INTRODUCTION

Human respiratory disease with symptoms of pneumonia has been diagnosed at the end of 2019, in Wuhan city, China, which is identified as COVID-19 caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). Coronaviruses have been known to cause

Abbreviations: ACE2, angiotensin-converting enzyme 2; C-HR, C-terminal heptad repeat; DOT, disorder-to-order transition; GAFF, general amber force field; HKU1, human coronavirus HKU1; IDPs, intrinsically disordered proteins; IDRs, intrinsically disordered regions/residues; MMPBSA, the molecular mechanics-Poisson Boltzmann surface area; N-HR, N-terminal heptad repeat; RBD, receptor binding domain; SARS-CoV, severe acute respiratory syndrome coronavirus; SARS-CoV-2, severe acute respiratory syndrome coronavirus 2.

Dhanusha Yesudhas and Ambuj Srivastava contributed equally on this work.

respiratory diseases and gastroenteritis in animals and humans.[1] The SARS-CoV and SARS-CoV-2 belong to the betacoronavirus family, and share 80% whole genome sequence similarity.[2]

At the time of viral invasion, spike protein (~1300 amino acid residues) is cleaved by host cell proteases into S1 (the peripheral fragment) and S2 (membrane-spanning fragment) subunits. The S1 subunit contains the receptor binding domain (RBD), which facilitates the interactions with host cell receptors whereas, the S2 subunit fuses with host cell membrane and releases the viral genome into the cell[2,3] (Figure S1). The heptads repeat regions N-HR, C-HR and the fusion peptides are involved in membrane fusion. The viral entry to the host cells takes place by an endocytic and nonendosomal pathways.[4] The spike protein priming by furin at the S1/S2 site is important and unique in SARS-CoV-2 membrane fusion, which is absent in SARS-CoV.[5] Spike proteins have a good binding affinity for sugar receptors of human cells, which uses them as a mechanism of cell entry.[6] Notably, an SARS-CoV-2 has a stronger affinity to human ACE2 than an SARS-CoV. SARS-CoV-2 also has a higher pathogenicity and mortality rate than SARS-CoV. The clinical spectrum of SARS-CoV-2 is broader than SARS-CoV including asymptomatic infections, mild upper respiratory tract illness, severe viral pneumonia with respiratory failure and death. In addition, SARS-CoV-2 infected patients suffer prolonged illness and have a low neutralizing antibody capacity, suggesting a greater need for vaccine interventions.[7]

The intrinsic disorder regions (IDRs) are the regions, which have a dynamic ensemble of conformations that do not adopt stable three-dimensional structure in physiological conditions.[8] IDRs are known to be involved in many biological functions, including signaling, regulation and binding molecules including proteins, DNA, RNA and small molecules. IDPs are known to be associated with various diseases including neurological disorders and cancer.[9,10] IDPs also tune their binding affinity for specific and non-specific binding with other molecules by changing its conformation.[9,11] They also act as hub proteins in various signaling networks, bind multiple partners and are involved in many pathways.[12,13]

IDRs are reported in SARS and other coronaviruses, for example, in human coronavirus 229E (hCoV-229E) an intrinsically disordered domain present near the C-terminal tail peptide, plays an important role in dimer-dimer association. Similarly, coronavirus HKU1 has a large structurally conserved linker loop (amino acids 428-587), which is partially disordered.[14] The S2 subunit of HKU1 also shows the presence of disordered residues at its protease cleavage site.[15] The structure reported for the recent SARS-CoV-2 spike protein, shows that S1/S2 junction is in a disordered, solvent-exposed loop.[1] Based on the literature reports, more number of databases and tools have been developed for predicting the disordered regions in proteins.[16-18] For example, the DisProt database contains the disorder content of spike and other viral protein structures.[17]

Wrapp et al determined the structure of the pre-fusion spike trimeric protein using cryogenic electron microscopy at a resolution of 3.46 Å (PDB ID: 6VSB).[1] In this trimeric protein, many residues are missing their structural coordinate information in the RBD (residues 329-334, 444-448, and 455-490), especially in "A" chain, which has an elevated structure and is primarily responsible for binding. The 329 to 334 residues are present at the N-terminal of RBD and disordered in free protein as well as in the complex and are termed as DR (disordered residues). However, the intrinsic disordered residues 444 to 448 and 455 to 490 have stable structure upon ACE2 binding (PDB ID: 6M0J)[19] and are termed as disorder-to-order transition (DOT) residues. Despite the lack of a stable three-dimensional structure of these regions in free protein, the DOTs facilitate the spike-ACE2 recognition and help for both the proteins interactions.

On the other hand, coronavirus spike protein is a glycoprotein, and the glycosylation site has high flexibility and variable secondary structures.[20] SARS-CoV-2 spike protein exists in both closed and open states, referred as inactive and active states of spike protein. Casalino et al suggested that the glycans help in conformational change from closed-to-open state, allowing the RBD to bind with ACE2 receptors.[21]

The infection and pathogenicity are often associated with the disordered regions in the viral proteins.[22-25] However, the role of disordered residues/regions in stability and binding of spike protein with ACE2, which could help in understanding the disease pathogenesis, has not yet been explored. In this work, we investigate the importance of disordered residues in spike-ACE2 recognition using molecular dynamics (MD) simulations. We observed that the DOTs in the free form of the spike protein experience very high fluctuations more than the complex structure. The DOTs undergo a structural transition upon binding with ACE2, which facilitates the stable compact complex structure. In addition, DOT regions are also contributing significant binding energy upon interactions with ACE2 receptors. Our study helps in understanding the functionality of IDRs in spike protein, and its interactions with host cell proteins.

2 | METHODS

2.1 | Disordered residues in spike protein

A region is considered to be disordered if at least five continuous non-terminal residues are missing in the PDB structure.[26,27] The DOT residues in RBD domain are obtained by comparing the missing residues from free protein (PDB ID: 6VSB) and spike-ACE2 complex (PDB ID: 6M0J). We observed that the A chain of 6VSB is an elevated structure and it contains missing residues from 444 to 448 and 455 to 490, which obtained an ordered structure upon binding to the ACE2 receptor (PDB ID: 6M0J). Thus, in this study, the residues 444 to 448 are considered as DOT1 and 455 to 490 as DOT2.

2.2 | Modeling

To obtain the structures for missing residues of the RBD present in free spike protein we have used modeler version 9.24[28] software and generated 10 models and selected the model, which has the lowest discrete optimized protein energy score. The missing residues (DOT residues) of the prefusion structure were modeled using 6VSB as a template.

2.3 | MD simulations

MD simulations are performed for free spike protein and the complex with ACE2 for 100 ns using AMBER18 tools with FF14SB[29] and GAFF force fields.[30] We have repeated the simulations three times to evaluate the consistency of results. The repetitions have been carried out with a seed generated from a random number generator and hence the initial velocities for each system are different. Molecule is solvated using TIP3PBOX water model with edge of the box at least 10 Å away from the solute molecules.[31] The Na + ions are added to neutralize the system in order to maintain the physiological condition. Energy minimization of each molecule is done for 50 000 steps using the steepest descent minimization with the step size of 2 fs at constant temperature (300 K), and pressure (1 atm). The description of the simulation systems has been listed in Table S1.

2.4 | Trajectory analysis

The trajectories obtained from simulations are visualized using VMD,[32] Pymol,[33] and Chimera.[34] Properties such as root mean square fluctuation (RMSF) and root mean square deviation (RMSD) are calculated using "rms" and "atomicfluct" commands available in CPPTRAJ program, respectively. The average minimum distance is calculated by plotting the distance of each residue in spike RBD with ACE2 during the simulation using "nativecontacts" command and generating average value using a python program. The interaction energy between spike and ACE2 is calculated using the the molecular mechanics-Poisson Boltzmann surface area (MMPBSA)[35] program for the last 2 ns. The dynamic residue movement is monitored via the dynamic cross-correction map (DCCM) using the Bio3D package.[36]

2.5 | Binding free energy calculation

MM-PBSA method is applied to calculate the binding free energy between RBD of spike protein and the ACE2 receptor. The energy contribution of each residue toward binding is measured.

$$G = E_{bond} + E_{el} + E_{vdW} + G_{pol} + G_{np} - TS \quad (1)$$

where E_{bond}, E_{el}, and E_{vdW} are the molecular mechanics (MM) energy terms from bonded (bond length, angle, and dihedral), electrostatic, and van der Waals interactions, respectively. G_{pol} and G_{np} are the polar and nonpolar contributions, respectively, to the solvation free energies. TS is the absolute temperature, T, multiplied by the entropy, S.

$$\Delta G_{bind} = \Delta G_{complex} - \Delta G_{protein1} - \Delta G_{protein2} \quad (2)$$

where $\Delta G_{complex}$ represents the free energy of a protein-protein complex, and $\Delta G_{protein1}$ and $G_{protein2}$ are the free energies of RBD and ACE2 proteins, respectively.

3 | RESULTS AND DISCUSSION

Recognition and interaction of spike with ACE2 is regulated by the RBD (319-541), in which residues from 436 to 506 are termed as receptor binding motif (RBM) because these residues are actively participating in the interactions. The DOT1 (residues 444-448) and DOT2 (residues 455-490) are grouped in the RBM motif which is responsible for ACE2 recognition. As shown in Figures 1 and S1, the DOT1 and DOT2 are located in the interface of the spike protein and ACE2. The crucial interacting DOT residues as well the residues, which are important to maintain the binding conformation, are heighted in Figure 1. The DOT regions help in providing the shape complementarity, strength and stability. In addition, the results from the previous studies state that the disordered regions of viral proteins are associated with their virulence.[24-27] Hence, atomic-level MD simulations studies of disordered regions in spike protein can provide better insights into the host tropism and transmission capacity of SARS-CoV-2.

3.1 | Flexibility of disordered residues in free and complex structure of spike protein

The flexibility for the overall structure of RBD spike protein in the free and complex form is observed from the backbone RMSD during the simulations. For the free protein, a major global conformational change occurs around 50 ns, increasing the RMSD up to 8 Å (Figure S2). This conformational change is principally caused by the C-terminal loop structures of the DOT regions. We confirm this observation by plotting the RMSF values and superimposed structures of RBD domain at different time frames (30, 50 and 80 ns). The RMSF calculations showed that the C-terminal of DOT2 region is highly flexible from 48 to 52 ns of simulations, and the superimposed structures show that DOT regions are highly deviated while the other regions are marginally deviating in different time frames (Figure S3).

Overall, both the simulations reached an equilibrium state, and the RMSD value of the complex structure shows lesser fluctuations (3 Å) than the free form (8 Å) (Figure S2). The unconstrained DOT regions have an impact on the overall structural stability when the protein binds with its partner. Thus, the RMSD of the DOT residues alone was calculated throughout the trajectories in order to observe its dynamic nature during simulation. Figure 2A,B displays the RMSDs of DOT1 and DOT2, respectively, of the complex and free protein structures. The RMSD of the DOTs in complex structure showed less dynamics, and maintains the stable trajectories throughout the simulation. However, the DOTs of the free protein show fluctuating movement, especially the DOT2 residues experience deviations up to 8 Å distance (Figure 2B). In addition, the RMSD and RMSF results from each simulation repetition have been provided for better insights (Figure S4). Moreover, the compactness of the structures is computed from the radius of gyration analysis and the results showed that RBD in both free and complex structures have similar compactness

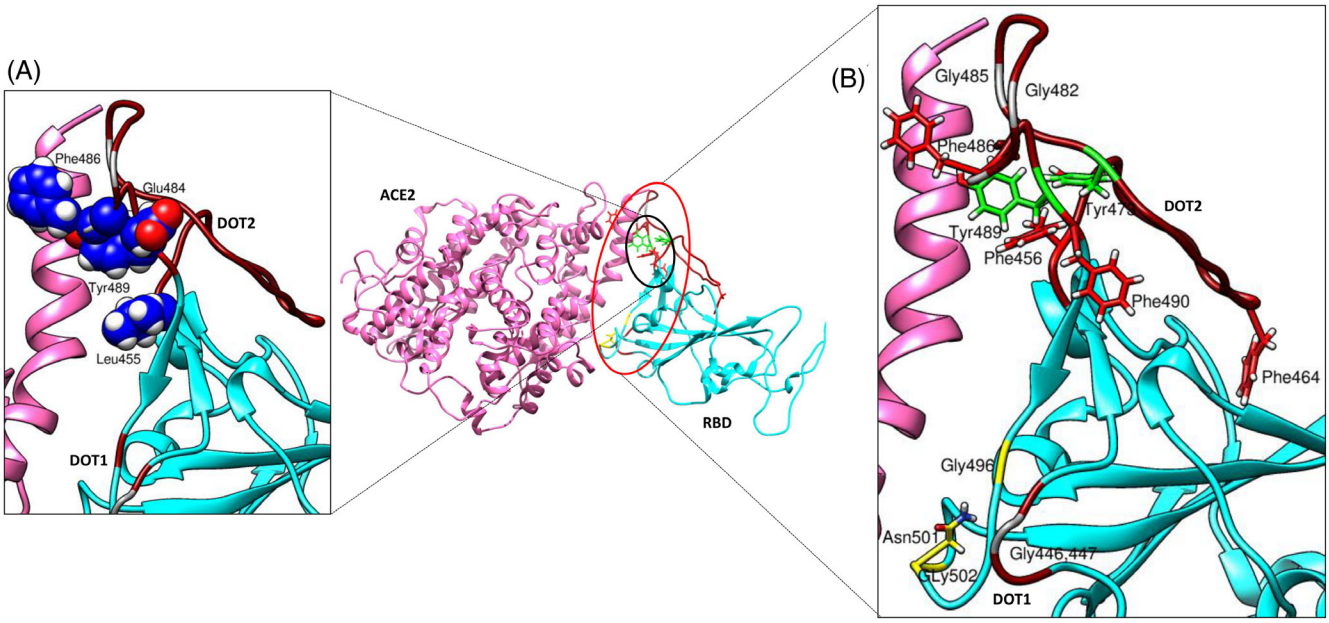

FIGURE 1 Structure of spike receptor binding domain (RBD)-ACE2 complex. A, Important interacting residues (Leu455, Glu484, Phe486, and Tyr489) at the interface, disorder-to-order transition (DOT) regions (DOT1 and DOT2) are highlighted. B, The interface residues between spike RBD and angiotensin-converting enzyme (ACE). Gly, Phe, and Tyr residues in DOT1 and DOT2 are represented in sticks and colored with silver, red, and green, respectively; the conserved residues are marked in yellow

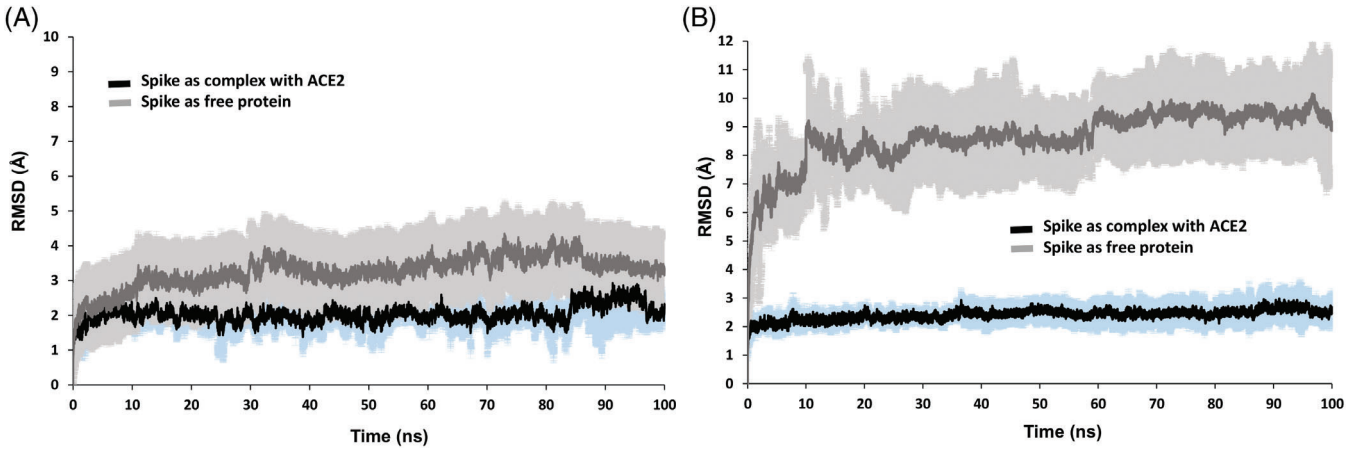

FIGURE 2 Overall backbone fluctuation of disorder-to-order transition (DOT) residues throughout the simulation. A, Root mean square deviation (RMSD) of DOT1 in free and complex structure. B, RMSD of DOT2 in free and complex structure. The blue and gray shades represent the standard deviations

throughout the simulations. The results are consistent in all three repetitions of simulations (Figure S5).

Similarly, the RMSF analysis explains the fluctuations of each residue during the simulation. The fluctuations of residues show that the DOTs in the free protein structure are experiencing high fluctuations, especially in the DOT2, some of the residues (470-490) are fluctuating up to 9 Å distance (Figure 3). However, the DOTs in the complex structure show low fluctuations, particularly the DOT2 experiences a maximum of 2.5 Å fluctuation for few residues (476-483) (Figure 3). Based on the RMSD and RMSF calculations, it clearly explains that the DOT regions greatly alter the dynamic stability of structure. The interactions of DOT residues with ACE2 make the complex structure compact and stable, thereby experiencing less fluctuation.

3.2 | Contributions of ordered residues vs disordered residues

SARS-CoV and SARS-CoV-2 do not fall in the same clade and the whole genome sequence similarity is less than 80%. The major sequence differences arise from their RBD sequences.[37] Regardless of this sequence diversity, the SARS-CoV and SARS-CoV-2 are having

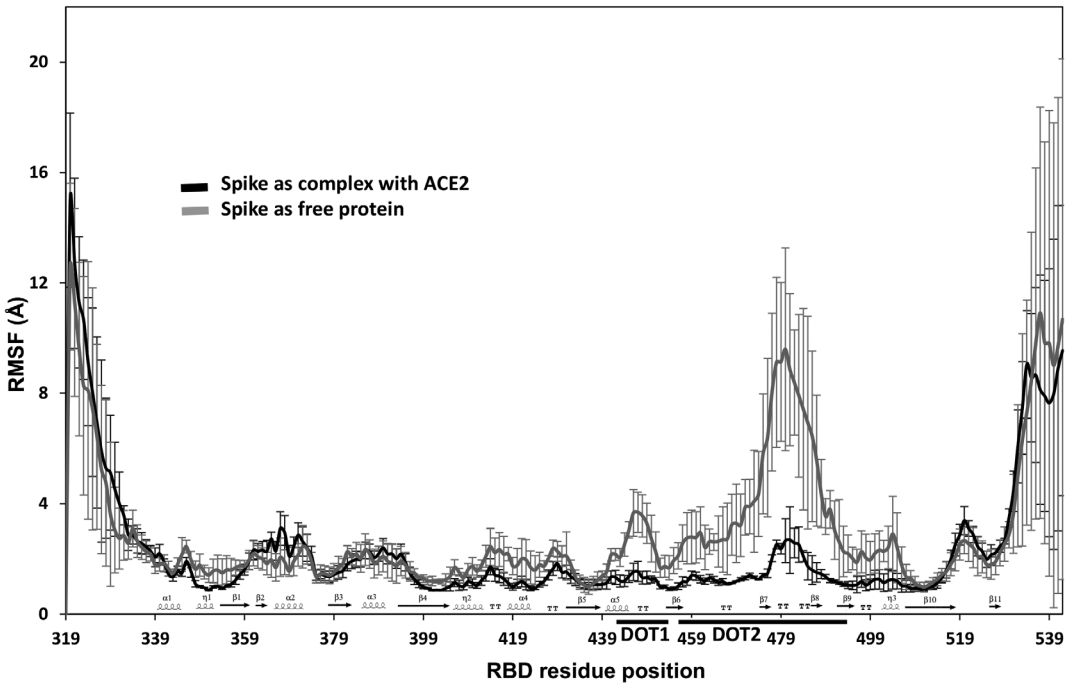

FIGURE 3 The residue fluctuation of spike protein receptor binding domain (RBD). The DOT1 and DOT2 residues are marked by black line below the X-axis. The, secondary structure information is displayed above the X-axis. The symbols α, β, η, and *T* represent α-helix, β-strand, 3_{10}-helix, and turn, respectively

100% structure identities in the RBD domain, and they share few conserved residues. Based on the interactions that exist between the RBD domain and ACE2, residues such as Tyr449, Gly502, and Tyr505 are observed to be conserved among SARS-CoV, NL-63, and SARS-CoV-2.[38,39] All these conserved residues are located at the flanking regions of DOTs. Studies have been reported that often the DOTs are located near the actual binding sites.[10]

The average minimum distance calculation of each residue of a spike protein with ACE2 receptor is measured (Figure 4). Most of the DOT residues are maintaining less than 5 Å distance from the corresponding ACE2 residues throughout the simulation. Apart from the DOTs, the ordered residues in the interface (436-443; 449-454, and 491-506), which are located adjacent to the DOTs, are showing lesser distance with ACE2 (Table 1). Particularly, the hot spot residues (Lue455, Glu484, Phe486, Asn487, Tyr489, Gln493, Gly496, Asn501, and Gly502), which are mentioned in the literature[38-41] maintain less than 3 Å distance from ACE2 (Table 1 and Figure 1). The average minimum distance between the spike RBD and ACE2 has been computed for the three repetitions of simulations, and the results are observed to be consistent (Figure S6). The location and the flexibility of these DOT residues provide better shape complementary for the spike-ACE2 recognition, and maintain its binding conformation for a longer time[42] demonstrated that RBM is an intrinsically flexible region, which is recognized by the ACE2 receptor, and the binding region is either in a closed (active) or open (inactive) state. The DOTs, which are responsible for the compact and stable structure, and their interactions with ACE2 support the open conformation.

Likewise, the binding free energy contribution from each residue of the RBD domain is calculated using AMBER-MMPBSA. As expected, the residues, which are maintaining lesser distance with ACE2 are showing more favorable interaction energies. The ordered and the disordered residues and their interaction energies are given in Figure 4B and are also listed in Table 1. Particularly, the residues located at the flanking regions of DOTs are shown to be contributing more favorable interaction energies. Among the ordered residues in the interface, seven residues are maintaining less than 3 Å distance with ACE2 and contributing more for the binding energy whereas among DOT residues, six of them are contributing favorable binding free energies. The alanine scanning mutagenesis of DOT residues has been performed to examine the change in binding free energy of the complex. The mutation of DOT residues Leu455, Phe456, Phe486, and Tyr489 decreased the binding affinity between spike RBD and ACE2, which confirms the importance of these residues in maintaining the binding affinity of the complex (Figure S7).

3.3 | Interactions of DOTs with ACE2 receptor

The DOT2, which makes the longest interacting loop region, is rich in Phe (11%) residues. This aromatic amino acid makes bonded and non-bonded interactions with the other aromatic residues of the partner ACE2 protein (Figure 1B). The residues involved in hydrogen bond interactions between the DOTs and ACE2 receptors are reported in Table 2, and the other hydrophobic interactions are listed in Table S2. Mainly, the residues present in RBD of spike protein Ala475, Gly476,

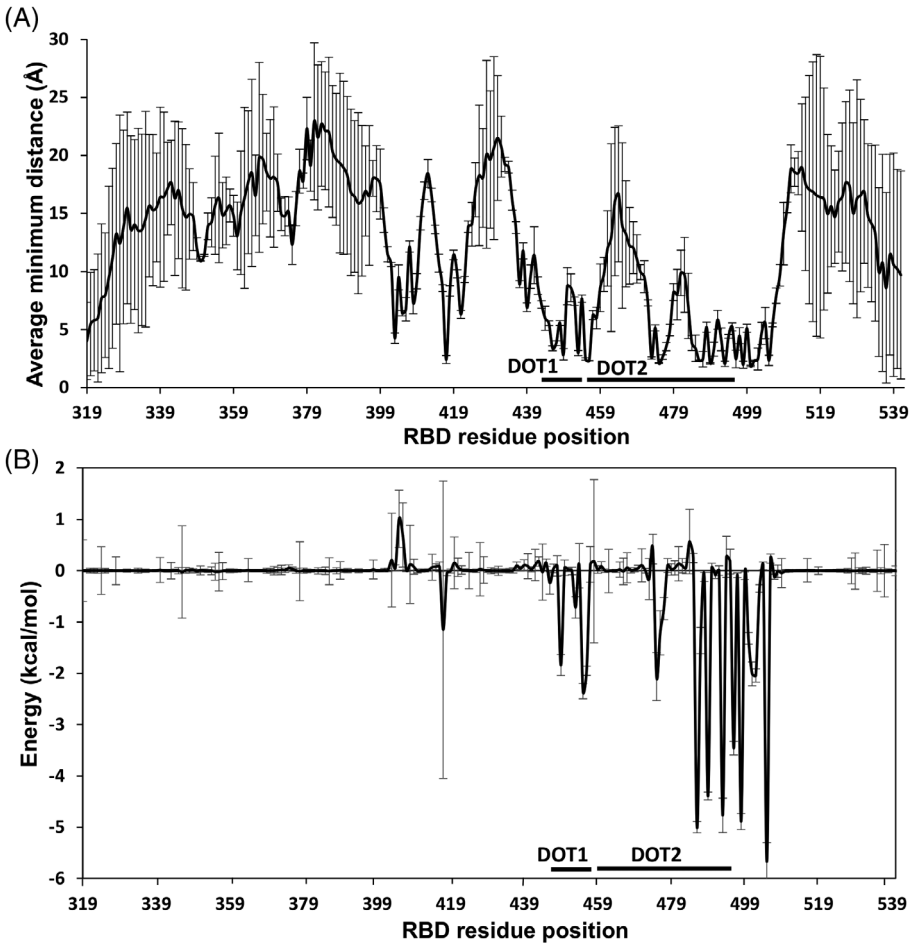

FIGURE 4 A, The average minimum distance (deviations are indicated with error bars) and, B, the decomposition energy of each residue between the receptor binding domain (RBD) residues against angiotensin-converting enzyme 2 (ACE2) throughout the simulations. Secondary structures of residues are the same as in Figure 3

TABLE 1 Residues, which are less than 3 Å distance from ACE2 and their energetic contribution upon binding

No.	Residues	Energy contribution (kcal/mol)
DOT residues		
1	Leu455[a]	−2.3
2	Phe456	−1.9
3	Arg457	−2.06
4	Gly476	−1.21
5	Phe486[a]	−5
6	Tyr489[a]	−4.4
Ordered residues		
1	Gln493[a]	−4.77
2	Gly496[a]	−3.46
3	Gln498	−4.89
4	Thr500	−1.32
5	Asn501[a]	−2.01
6	**Gly502**	−2.04
7	**Tyr505**[a]	−5.67

Abbreviations: ACE2, angiotensin-converting enzyme 2; DOT, disorder-to-order transition; SARS-CoV-2, severe acute respiratory syndrome-coronavirus 2.

[a]The key residues reported in the literature. Conserved residues among SARS-CoV-2, NL-63, and SARS-CoV are in bold.

Asn487, and Tyr489 are making interactions with Ser19, Gln24 and Tyr83 of ACE2 (Table 2). Part of the experimentally determined interacting residues which are reported in previous studies are present in the DOT regions (Glu484, Asn487, and Tyr489), and in the flanking region of the DOTs (Figure 1).

Both SARS-CoV and SARS-CoV-2 RBM (436-506) are enriched in Gly (11%), Tyr (8%), and Phe (7%) residues. Based on previous studies, these three residues are playing an important role in interactions with ACE2.[43] In the current study, these residues form hydrogen bonds and other non-covalent interactions (Tables 2 and S2). Notably, the decomposition energy calculations also stated that glycine, tyrosine, and phenylalanine residues are contributing favorable affinities for the complex structure. However, a recent study revealed that in SARS-CoV RBD, the tyrosine residues, which are present at the interface (Tyr442, Tyr475, and Tyr491), are not actively taking part in the interaction with ACE2.[43] Whereas in SARS-CoV-2, even though the Tyr and Phe residues of the DOT2 are present in the loop, they are still stable via pi-pi interactions of their phenyl ring structures, for example, Phe486 of spike protein interacts with Tyr83 of ACE2. These residues are actively taking part in intraprotein and interprotein interactions. Figure 1B represents the binding interface, in which all the Gly, Phe, and Tyr residues in DOT2 are marked in silver, red, and green, respectively, and the conserved residues are also marked in yellow. The side chains of Tyr and Phe of spike protein make favorable

TABLE 2 Hydrogen bond and salt bridge interacting residues between DOTs and ACE2

	Hydrogen bond interacting residues		Salt bridge interactions
No	DOTs of spike	ACE2 receptor	DOTs of spike-ACE2
1	Leu455 (N—H)[a]	Lys31 (O)	Lys458 (NZ)-Glu23 (OE2)
2	Lys458 (N—H)	Glu23 (O)	Glu484 (OE1)[a]-Lys31 (NZ)
		Ser19 (O)	
3	Tyr473 (O—H)	Thr27 (O)	
		Glu23 (O)	
		Ser19 (N)	
4	Ala475 (O—H)	Ser19 (N)	
		Gln24 (O)	
5	Gly476 (N—H)	Ser19 (O)	
		Gln24 (N)	
6	Ser477 (N—H)	Ser19 (O)	
		Gln24 (N)	
7	Thr478 (O—H)	Gln24 (O)	
		Ser19 (O)	
8	Asn487 (N—H)[a]	Gln24 (O)	
		Tyr83 (O)	
9	Tyr489 (O—H)[a]	Tyr83 (O)	
		Gln24 (O)	
		Phe28 (N)	
10	Phe486 (N—H)[a]	Leu79 (O)	
		Met82 (S)	

Abbreviations: ACE2, angiotensin-converting enzyme 2; DOT, disorder-to-order transition; SARS-CoV-2, severe acute respiratory syndrome-coronavirus 2.
[a]The experimentally reported residues. The donor and the acceptor atoms are also included.

electrostatic interactions with α1 helix of ACE2, especially with Tyr and Gln amino acids (Table 2). Followed by Tyr and Phe, Gly is also abundant in the RBD domain of both SARS-CoV and SARS-CoV-2 spike proteins and is located at the binding site. The prevalence of aromatic-aromatic interactions and presence of DOT regions in SARS-CoV-2 makes it different from SARS coronavirus.

Moreover, recently, Ju et al identified antibodies that potentially neutralize SARS-CoV-2, and the residues Lys444, Gly446, Gly447, Asn448, Tyr449, Asn450, Lys452, Val483, Glu484, Gly485, Phe490, and Ser494 are epitope residues.[44] Interestingly, all these residues are in the DOT regions of the RBD receptor-binding motif. This implies the importance of targeting DOT regions/residues in designing the drugs and vaccines.

3.4 | Cross-correlation map of DOTs of spike protein

In addition, the DCCM of RBD residues in free and complex structures revealed that most of the residues in the complex structure experience positive correlation motion with nearby residues which are represented in blue boxes. The pink and white color regions show negative and no correlation motion, respectively. The DOTs, especially DOT2 in the complex structure, experience positive correlation with the nearby residues, (highlighted with a red circle) which shows its stability and compactness upon ACE2 binding (Figure 5). Other than the DOT residues, the N-terminal region of the spike RBD domain in complex structure shows positive correlation with other residues. Since the N-terminal also took part in interactions with the ACE2 receptor, the ordered residues in that N-terminal are also

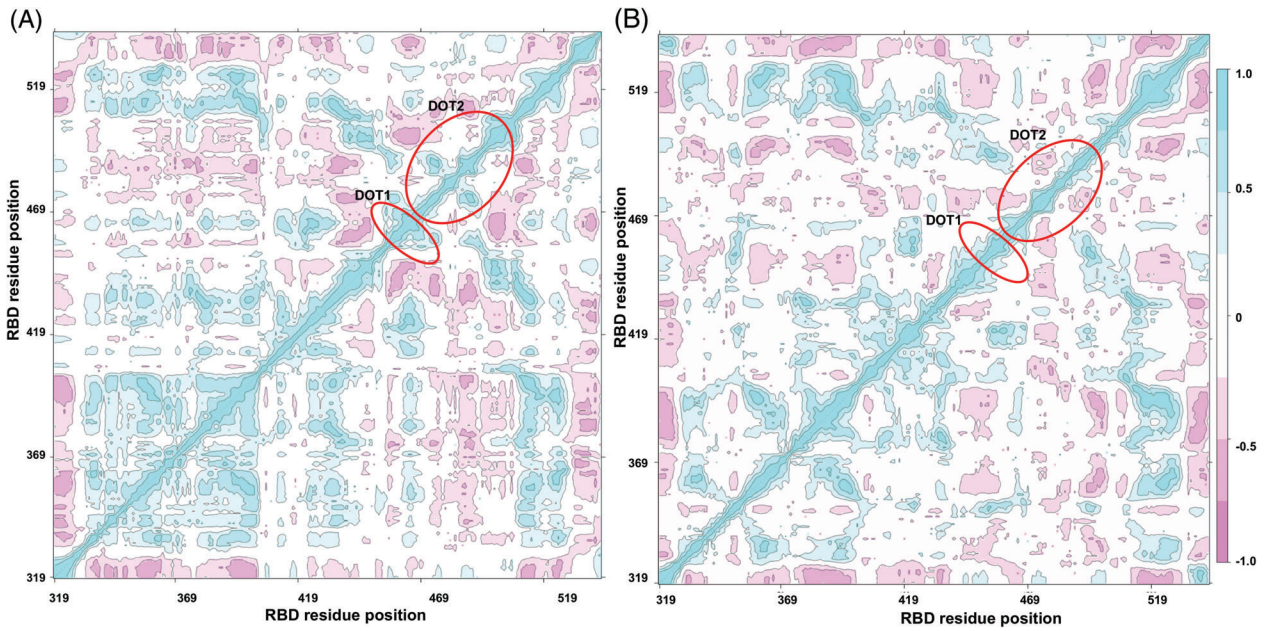

FIGURE 5 A, Dynamic cross-correction map (DCCM) of receptor binding domain (RBD) residues in the complex structure and it represents that the disorder-to-order transition (DOT) residues exist in positive correlative motions with nearby residues, which suggest the stable interactions and compact structure of the complex. B, DCCM of RBD residues in the free form structure

showing positive correlated motion. However, in Figure 5B, the DOT residues of the free protein (marked in red circle) are not having positively correlated motion with the nearby residues, which explains its higher structural flexibility (Figure 2). Similarly, the residue contact plot also explained that the crucial DOT residues are maintaining the interactions with ACE2 throughout the simulations (Figure S8).

The host cell invasion of SARS-CoV-2 is happening by glycosylated spike protein. This fusion protein exists in metastable prefusion form, which undergoes more structural rearrangements to fuse with the host cell membrane. Receptor binding destabilizes the prefusion trimer, and the S1 subunit undergoes open and close conformational changes which represent the receptor accessible (less stable conformation) and inaccessible states, respectively.[1,42-47] These conformational rearrangements could be possibly controlled by the DOT residues which are present in the interface between the RBD domain and the ACE2. The positive correlation motion of residues in the complex structure, especially the DOTs and N-terminal region of RBD provides the precise geometry for the binding and explains the stable and compact structure of spike-ACE2 complex.

Hence, focusing on the behavior of disordered residues of spike protein in free and complex forms may shed some light to identify the reason for transmission and higher binding affinities. The flexibility and the conformational changes adapted by these DOTs upon binding with ACE2 have to be focused while identifying drugs.

4 | SUMMARY

The intrinsically flexible RBD recognizes the ACE2 receptor, and the binding region is either in an open or a closed state. The ACE2 recognition by the RBD is mainly based on the conformational state adapted by the RBD, and the conformational switch is governed by the highly flexible residues. Due to the highly flexible nature of DOTs, the DOT residues can govern the precise geometry for ACE2 binding. The DOT of the residues and their higher number of aromatic and ionic interactions make the complex structure more stable and experience higher binding energies. The positive correlative motion of DOTs with the nearby residues also explains the compact and stable structure of the RBD with ACE2. The uniquely reported DOTs in SARS-CoV-2 spike protein, which are located in the binding interface, require attention for the development of small molecule inhibitors. In addition, these spike protein DOT residues act as potential epitope residues for designing new antibodies. Thus, an atomic level characterization of the spike protein and its unique DOTs structure can guide for mechanism study, which is associated with its transmission. The inhibition of viral entry needs to focus on the structural aspects of DOTs in spike RBD domain, which may help in therapeutic applications for COVID-19.

ACKNOWLEDGMENTS

The authors thank Indian Institute of Technology Madras, Department of Biotechnology for the computational facilities. This work is partially supported by the Platform Project for Supporting Drug Discovery and Life Science Research (Basis for Supporting Innovative Drug Discovery and Life Science Research [BINDS]) from AMED under Grant Number JP20am0101112 to M. S. and the Department of Science and Technology to M. M. G. (No. MSC/2020/000319).

CONFLICT OF INTEREST

The authors declare no conflicts of interest.

AUTHOR CONTRIBUTIONS

Dhanusha Yesudhas and **Ambuj Srivastava**: Designed and carried out the work. **Dhanusha Yesudhas**, **Ambuj Srivastava**, **Masakazu Sekijima**, and **M. Michael Gromiha**: Contributed in discussions. **Dhanusha Yesudhas** and **Ambuj Srivastava**: Wrote the manuscript. **Masakazu Sekijima** and **M. Michael Gromiha**: Edited the manuscript. All authors read and approved the manuscript.

PEER REVIEW

The peer review history for this article is available at https://publons.com/publon/10.1002/prot.26088.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID

M. Michael Gromiha https://orcid.org/0000-0002-1776-4096

REFERENCES

1. Wrapp D, Wang N, Corbett KS, et al. Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science*. 2020;367(6483):1260-1263.
2. Zhou P, Yang XL, Wang XG, et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature*. 2020;579:270-273.
3. Wang Q, Zhang Y, Wu L, et al. Structural and functional basis of SARS-CoV-2 entry by using human ACE2. *Cell*. 2020;181(4):894-904.
4. Zumla A, Chan JF, Azhar EI, Hui DS, Yuen KY. Coronaviruses—drug discovery and therapeutic options. *Nat Rev Drug Discov*. 2016;15:327-347.
5. Yesudhas D, Srivastava A, Gromiha MM. COVID-19 outbreak: history, mechanism, transmission, structural studies and therapeutics. *Infection*. 2021;49:199-213.
6. Li F. Structure, function, and evolution of coronavirus spike proteins. *Annu Rev Virol*. 2016;3:237-261.
7. Rossi GA, Sacco O, Mancino E, Cristiani L, Midulla F. Differences and similarities between SARS-CoV and SARS-CoV-2: spike receptor-binding domain recognition and host cell infection with support of cellular serine proteases. *Infection*. 2020;48:665-669.
8. Wright PE, Dyson HJ. Intrinsically disordered proteins in cellular signaling and regulation. *Nat Rev Mol Cell Biol*. 2015;16:18-29.
9. Babu MM, van der Lee R, de Groot NS, Gsponer J. Intrinsically disordered proteins: regulation and disease. *Curr Opin Struct Biol*. 2011;21:432-440.
10. Srivastava A, Ahmad S, Gromiha MM. Deciphering RNA-recognition patterns of intrinsically disordered proteins. *Int J Mol Sci*. 2018;19(6):1595.
11. Arai M. 2018. Unified understanding of folding and binding mechanisms of globular and intrinsically disordered proteins. *Biophysics*. 2018;10(2):163-181.

12. Krieger JM, Fusco G, Lewitzky M, et al. Conformational recognition of an intrinsically disordered protein. *Biophys J*. 2014;106(8):1771-1779.
13. Uversky VN. The multifaceted roles of intrinsic disorder in protein complexes. *FEBS Lett*. 2015;589(19):2498-2506.
14. Hamming I, Timens W, Bulthuis ML, Lely AT, Navis G, van Goor H. Tissue distribution of ACE2 protein, the functional receptor for SARS coronavirus. A first step in understanding SARS pathogenesis. *J Pathol*. 2004;203:631-637.
15. Kirchdoerfer RN, Cottrell CA, Wang N, et al. Pre-fusion structure of a human coronavirus spike protein. *Nature*. 2016;531(7592):118-121.
16. Oates ME, Romero P, Ishida T, et al. D^2P^2: database of disordered protein predictions. *Nucleic Acids Res*. 2013;41:D508-D516.
17. Vucetic S, Obradovic Z, Vacic V, et al. DisProt: a database of protein disorder. *Bioinformatics*. 2005;21(1):137-140.
18. Potenza E, Domenico TD, Walsh I, Tosatto SC. MobiDB 2.0: an improved database of intrinsically disordered and mobile proteins. *Nucleic Acids Res*. 2015;43(D1):D315-D320.
19. Lan J, Ge J, Yu J, et al. Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *Nature*. 2020;581(7807):215-220.
20. Petrescu AJ, Milac AL, Petrescu SM, Dwek RA, Wormald MR. Statistical analysis of the protein environment of N-glycosylation sites: implications for occupancy, structure, and folding. *Glycobiology*. 2004;14(2):103-114.
21. Casalino L, Gaieb Z, Goldsmith JA, et al. Beyond shielding: the roles of Glycans in the SARS-CoV-2 spike protein. *ACS Cent Sci*. 2020;6(10):1722-1734.
22. Giri R, Kumar D, Sharma N, Uversky VN. Intrinsically disordered side of the Zika virus proteome. *Front Cell Infect Microbiol*. 2016;6:144.
23. Xue B, Williams RW, Oldfield CJ, Kian-Meng Goh G, Keith Dunker A, Uversky VN. Viral disorder or disordered viruses: do viral proteins possess unique features? *Protein Pept Lett*. 2010;17(8):932-951.
24. Singh A, Kumar A, Yadav R, Uversky VN, Giri R. Deciphering the dark proteome of Chikungunya virus. *Sci Rep*. 2018;8:5822.
25. Giri R, Bhardwaj T, Shegane M, et al. Understanding covid-19 via comparative analysis of dark proteomes of sars-cov-2, human sars and bat sars-like coronaviruses. *Cell Mol Life Sci*. 2021;78(4):1655-1688.
26. Ishida T, Kinoshita K. PrDOS: prediction of disordered protein regions from amino acid sequence. *Nucleic Acids Res*. 2007;35:W460-W464.
27. Uversky VN, Dunker AK. Understanding protein non-folding. *Biochim Biophys Acta*. 2010;1804(6):1231-1264.
28. Webb B, Sali A. Comparative protein structure modeling using modeller. *Curr Protoc Bioinformatics*. 2016;54(1):5-6.
29. Maier JA, Martinez C, Kasavajhala K, Wickstrom L, Hauser KE, Simmerling C. ff14SB: improving the accuracy of protein side chain and backbone parameters from ff99SB. *J Chem Theory Comput*. 2015;11(8):3696-3713.
30. Wang J, Wolf RM, Caldwell JW, Kollman PA, Case DA. Development and testing of a general amber force field. *J Comput Chem*. 2004;25(9):1157-1174.
31. Hornak V, Abel R, Okur A, Strockbine B, Roitberg A, Simmerling C. Comparison of multiple amber force fields and development of improved protein backbone parameters. *Proteins*. 2006;65(3):712-725.
32. Humphrey W, Dalke A, Schulten K. VMD: visual molecular dynamics. *J Mol Graph Model*. 1996;14(1):33-38.
33. Schrödinger L, DeLano W. PyMOL. http://www.pymol.org/pymol
34. Goddard TD, Huang CC, Meng EC, et al. UCSF ChimeraX: meeting modern challenges in visualization and analysis. *Protein Sci*. 2018;27(1):14-25.
35. Miller BR III, McGee TD Jr, Swails JM, Homeyer N, Gohlke H, Roitberg AE. MMPBSA. Py: an efficient program for end-state free energy calculations. *J Chem Theory Comput*. 2012;8(9):3314-3321.
36. Grant BJ, Rodrigues AP, ElSawy KM, McCammon JA, Caves LS. Bio3d: an R package for the comparative analysis of protein structures. *Bioinformatics*. 2006;22(21):2695-2696.
37. Lu R, Zhao X, Li J, et al. Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *The Lancet*. 2020;395(10224):565-574.
38. Rawat P, Jemimah S, Ponnuswamy PK, Gromiha MM. Why are ACE2 binding coronavirus strains SARS-CoV/SARS-CoV-2 wild and NL63 mild? *Proteins*. 2021;89(4):389-398.
39. Walls AC, Park YJ, Tortorici MA, Wall A, McGuire AT, Veesler D. Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein. *Cell*. 2020;183(6):281-292.
40. Yi C, Sun X, Ye J, et al. Key residues of the receptor binding motif in the spike protein of SARS-CoV-2 that interact with ACE2 and neutralizing antibodies. *Cell Mol Immunol*. 2020;17(6):621-630.
41. Wan Y, Shang J, Graham R, Baric RS, Li F. Receptor recognition by the novel coronavirus from Wuhan: an analysis based on decade-long structural studies of SARS coronavirus. *J Virol*. 2020;94(7):e00127-e00120.
42. Yuan Y, Cao D, Zhang Y, et al. Cryo-EM structures of MERS-CoV and SARS-CoV spike glycoproteins reveal the dynamic receptor binding domains. *Nat Commun*. 2017;8:15092.
43. Chowdhury R, Maranas CD. Biophysical characterization of the SARS-CoV2 spike protein binding with the ACE2 receptor explains increased COVID-19 pathogenesis. *Comput Struct Biotechol J*. 2020;18:2573-2582.
44. Ju B, Zhang Q, Ge J, et al. Human neutralizing antibodies elicited by SARS-CoV-2 infection. *Nature*. 2020;584(7819):115-119.
45. Walls AC, Xiong X, Park YJ, et al. Unexpected receptor functional mimicry elucidates activation of coronavirus fusion. *Cell*. 2019;176(5):1026-1039.
46. Gui M, Song W, Zhou H, et al. Cryo-electron microscopy structures of the SARS-CoV spike glycoprotein reveal a prerequisite conformational state for receptor binding. *Cell Res*. 2017;27:119-129.
47. Pallesen J, Wang N, Corbett KS, et al. Immunogenicity and structures of a rationally designed prefusion MERS-CoV spike antigen. *Proc Natl Acad Sci U S A*. 2017;114(35):E7348-E7357.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Yesudhas D, Srivastava A, Sekijima M, Gromiha MM. Tackling Covid-19 using disordered-to-order transition of residues in the spike protein upon angiotensin-converting enzyme 2 binding. *Proteins*. 2021;89:1158–1166. https://doi.org/10.1002/prot.26088