European Review for Medical and Pharmacological Sciences 2021; 25: 1708-1723

Evaluation of molecular interaction, physicochemical parameters and conserved pattern of SARS-CoV-2 Spike RBD and hACE2: in silico and molecular dynamics approach

C. CHAKRABORTY[1], A.R. SHARMA[2], B. MALLICK[3], M. BHATTACHARYA[4], G. SHARMA[5], S.-S. LEE[3]

[1]Department of Biotechnology, School of Life Science and Biotechnology, Adamas University, Kolkata, West Bengal, India
[2]Institute for Skeletal Aging and Orthopedic Surgery, Hallym University-Chuncheon Sacred Heart Hospital, Chuncheon-si, Gangwon-do, Republic of Korea
[3]Departments of Physics, Galgotias College of Engineering and Technology, Greater Noida, India
[4]Department of Zoology, Fakir Mohan University, Vyasa Vihar, Balasore, Odisha, India
[5]Neuropsychopharmacology and Toxicology Program, College of Pharmacy, Kangwon National University, Republic of Korea

Chiranjib Chakraborty and Ashish Ranjan Sharma contributed equally

Abstract. – OBJECTIVE: Recent pandemic virus SARS-CoV-2 is a global warning for the healthcare system. The spike protein of virus SARS-CoV-2 is significant because of two reasons. Firstly, the spike protein of this virus binds with the human ACE2 (hACE2) receptor. Secondly, it has several antigenic regions that might be targeted for vaccine development. However, the structural analytical data for the spike protein of this virus is not available.

MATERIALS AND METHODS: Here, we performed an analysis to understand the structural two subunits of S glycoprotein (S gp) of SARS-CoV-2. Further, an analysis of secondary structure components and the tertiary structure analysis of RBD was carried out. We also performed molecular interaction analysis between S gp of this virus and hACE2 as well as between SARS-CoV S gp and hACE2 to compare the binding properties of these two viruses.

RESULTS: We noted that the molecular interaction of SARS-CoV-2 S gp and hACE2 form eleven hydrogen bonds, while the molecular interaction of SARS-CoV S gp and hACE2 receptor form seven hydrogen bonds, indicating that the molecular interaction of SARS-CoV-2 S gp and hACE2 receptor is more stable than SARS-CoV S gp and hACE2 receptor. The pairwise sequence alignment of S gp SARS-CoV and SARS-CoV-2 shows several conserved residues of these two proteins. Besides, conserved pattern analysis of SARS-CoV-2 S gp and hACE2 revealed the presence of several highly conserved regions for these two proteins. The molecular dynamics simulation shows a stable interplay between SARS-CoV-2 S gp with the hACE2 receptor.

CONCLUSIONS: The present study might help determine the SARS-CoV-2 virus entrance mechanism into the human cell. Moreover, the understanding of the conserved regions may help in the process of therapeutic development from the infection of the deadly virus.

Key Words:
S glycoprotein, SARS-CoV-2, hACE2 receptor, Molecular interaction, Structural analysis.

Introduction

The present coronavirus causes Coronavirus disease (COVID-19), which originated in China[1,2]. On January 9, 2020, the SARS-CoV-2 was affirmed as the causative agent for COVID-19. WHO announced the COVID-19 as an epidemic, requiring a health emergency, on January 30, 2020. On 12th March 2020, WHO confirmed COVID-19 as a pandemic as it mushrooms worldwide[3]. Presently, it has continued to spread over 200 countries, and this ongoing outbreak has accounted for more than 1,536,482 human lives till now.

Corresponding Authors: Chiranjib Chakraborty, Ph.D; e-mail: drchiranjib@yahoo.com
Sang-Soo Lee, MD, Ph.D; e-mail: 123sslee@gmail.com

SARS-CoV-2 is distantly linked with the MERS-CoV and SARS-CoV[4,5]. Earlier, the SARS-CoV caused respiratory outbreaks in China in 2002 and 2003[6,7]. Later MERS-CoV caused a deadly outbreak in the Middle East in 2012[8]. The SARS-CoV-2 belongings to the genus betacoronavirus, order Nidovirales[9].

SARS-CoV-2 S glycoprotein (S gp) is a structural protein. It is a transmembrane glycoprotein that forms overhanging homotrimers on the viral surface[10]. The spike protein of coronavirus is approximately 150 kDa in size and forms a typical spike-like appearance on the virus's outer portion. It is a trimeric glycoprotein that helps in spike protein-mediated amalgamation to the host receptor. The S gp has two subunits, S1 and S2 subunit. In the S1 subunit, the loop structure may change structurally in size between CoV species of the betacoronavirus genus[11,12]. There are two functional domains in the S1 subunit: NTD (N-terminal domain) and C-domain. C-domain acts as RBD (receptor-binding domain)[13]. Both of these domains are accountable for the interaction between the virion and the receptor part of the host[14]. This part contains several epitopes that might serve as a potent target for designing the vaccines and developing the antibodies[15]. Conversely, there are three types of functional domains in the S2 subunit, such as FP (fusion peptide), HR1 (heptad repeat 1), and HR2 (heptad repeat 2). The binding phenomenon of the virion to the host cell's receptor occurs in two steps. First, the S1 subunit of the RBD domain interacts with the receptor. Second, immediately after binding, the S2 subunit modifies its configuration by introducing the FP domain into the host cell membrane. Subsequently, it can lead to combining two domains, such as HR1 and HR2, in generating 6-HB (six-helical bundle). Finally, virion fuses with the cellular membranes, resulting in the entry of the viral genetic elements inside the host cell along the fusion pore. The cell of the host acts as machinery for viral replication[16,17].

Due to several epitopic sites, spike protein has a distinct role in the human body. Spike protein can generate neutralizing antibodies in the host and can develop protective immunity[18]. Thus, spike protein can be a target protein for vaccine development[16]. Utilizing spike protein antigenic epitopes, Bhattacharya et al[19] suggested an epitope-based vaccine model against this virus. Therefore, the structural study of all the spike protein domains is necessary for further vaccine developments.

ACE2, a cellular receptor, is expressed in diverse epithelial cells, such as the nasal, lung, intestine, heart, kidney, immune cells, and blood vessels[20,21]. To enter the human body, both SARS-CoV-2 and SARS-CoV interacts with the ACE2 with the help of spike protein[22-24]. However, it is necessary to understand more about the interplay between the S gp of SARS-CoV-2 and the humanized form of ACE2 (hACE2).

Here, we performed structural analysis and depiction of the SARS-CoV-2 S gp and hACE2 protein interaction *via in silico* tools. Further, we also examined the binding part of S gp of both SARS-CoV and SARS-CoV-2 with hACE2 receptor and compared the receptor-binding properties of these two viruses. Subsequently, we also studied various physicochemical properties [i.e., glycosylation site, hydrophobicity prediction, grand average of hydropathicity (GRAVY)] of SARS-CoV-2 S gp and hACE2 interaction to understand the key factors that are involved during the molecular interactions of both proteins. Finally, we performed the molecular dynamics simulation of SARS-CoV-2 S gp with hACE2 complex to understand the interaction properties.

Materials and Methods

Collection of Target Protein Sequences

The sequences of the S gp of SARS-CoV-2 and hACE2 were collected from the NCBI[25]. The FASTA format of the sequence was collected for additional analysis. The accession number of the S gp of SARS-CoV-2 protein sequence was 6VYB_A, and the accession number of the ACE2 protein sequence was NP_068576.1.

Collection of the PDB Files and Assortment of Its Information for 3D Structure Analysis

The PDB files of the S gp of SARS-CoV-2 with hACE2 complex (PDB ID: 6M0J), S gp of SARS-CoV with hACE2 protein (6ACG), only spike protein SARS-CoV-2 (PDB ID: 6VYB) and ACE2 (PDB ID:1R42) were extracted from the Protein Data Bank for advance analysis[26].

Prediction of Domain Structure, Distribution of Amino Acid Residues in Different Domains in S1 and S2 Subunit of SARS-CoV-2 S gp, Complete SARS-CoV-2 S gp and hACE2 Receptor

For analysis of the domain structure and amino acid distribution in different domains of spike protein as well as complete SARS-CoV-2 S gp and

hACE2 protein, the ExPASy server[27] and Statistical Analysis of Protein Sequences (SAPS) server were used[28]. These two servers are considered as the central servers for analyzing the details about protein sequence characteristics. We have used their data for further analysis.

Secondary Structure Architecture and Motif Architecture Interface of S-Protein and hACE2 Receptor

For the prediction of the secondary structure of S-protein and ACE2, we retrieved the PDBsum database. It is a significant database of graphical summaries of the secondary structure architecture and motif architecture of proteins linked with the Protein Data Bank[29,30]. Finally, secondary structure architecture such as α helices, β sheets, and motif architecture such as β hairpins, β turns, and γ turns from the graphical figures were calculated.

Analysis of Tertiary Structure Conformation of SARS-CoV-2 S gp and Its Architecture

For analysis of tertiary structure conformation of S gp of SARS-CoV-2 and its architecture, the UCSF Chimera software package was used. It also helps to understand the molecular structure[31,32] and the molecular interaction of tertiary structure conformation.

Analysis of the Tertiary Structure Conformation of RBD

For this study, again**,** the UCSF Chimera software package was used[31,32].

Analysis of Tertiary Structure Interface of S gp of SARS-CoV-2 with hACE2 Receptor Interaction and S gp of SARS-CoV with hACE2 Receptor Interaction

To understand the tertiary structure interface of S gp of SARS-CoV-2 with hACE2 receptor and S gp of SARS-CoV with hACE2, we used UCSF Chimera software package[31,32]. It shows molecular structures of the interactions and an interactive platform for visualization of tertiary structures.

Sequence Alignment (SA) of S gp of Two Viruses (SARS-CoV-2 and SARS-CoV)

The SA of S gp proteins of SARS-CoV-2 and SARS-CoV was performed using Clustal Omega software[33]. This server uses the mBed10 algorithm, which can calculate the distance matrix.

Analysis of Conservation Pattern with Highly Conserved Amino Acids of the Tertiary Structure of SARS-CoV-2 S gp and hACE2 Receptor

The conservation prototype of the shape of S gp of SARS-CoV-2 was developed using the ConSurf server[34,35]. The conservation scores were calculated at each location of the aa residue of S gp of SARS-CoV-2 and hACE2 receptor using this server. For calculation of conservation prototype, the server used an empirical Bayesian inference from protein sequence to protein structure. The results were used for additional analysis.

Prediction of Physicochemical Parameters Such as Glycosylation Site, Hydrophobicity Prediction, Grand Average of Hydropathicity (GRAVY) Analysis of SARS-CoV-2 S gp and hACE2 Protein

Glycosylation of protein is a post-translational modification necessary for its folding, transport, and interactions with several receptor proteins. We have analyzed O-glycosylation as well as N-glycosylation sites of the S-protein and hACE2 receptor using the NetOglyc 4.0 and NetNGlyc 1.0 server[36,37].

It has been observed that the hydrophobicity of protein has a vital role in biomolecular interactions[38]. The hydrophobicity of S-protein (each domain of S-protein, complete S-protein) and hACE2 receptor was analyzed according to the Kyte-Doolittle algorithm used by ProtScale server[39] to produce a diagram of percent accessible amino acid residues for every S gp and hACE2 receptor in linear weight variation model. GRAVY analysis of complete S-protein, each S-protein domain, and hACE2 receptor residues was performed using ExPASy server[39,40].

Molecular Dynamics Simulation (MDS) Analysis

In order to validate the conformational stability of the docking complex of two proteins, i.e., S gp of SARS-CoV-2 with hACE2 complex (PDB ID:6M0J), we performed MDS analysis using GROMACS v. 2020.2 software, selecting GROMOS54A7 force-field as a parameter[41]. After the topology building, the complex solvate module was applied to make the system solvated with a simple point charge water model (SPC216). Later, we added Na^+ counter ions and stepped to energy minimization of the system using the steepest descent minimization algorithm with 50000n steps. The machine was then calibrated in two steps;

NVT ensemble with fixed particle number, temperature, volume, and NPT ensemble with fixed particle number, temperature, and pressure, set for 100 picoseconds. Finally, MD simulation was performed for 1 nanosecond run. The result was further used to measure Root Mean Square Deviation (RMSD) for the Cα backbone atoms of the protein and Root Mean Square Fluctuation (RMSF) of the ligand. The radius of gyration along with the number of hydrogen bonds of the interacting molecules, later the solvent-accessible surface was analyzed, and finally, we used Xmgrace to plot and visualize all the analysis reports.

Results

Prediction of Domain Structure, Distribution of Amino Acid Residues in Different Domains in S1 and S2 Subunits of SARS-CoV-2 S gp Complete SARS-CoV-2 S gp and hACE2

Thematic map of the spike protein with different regions regarding S1 and S2 subunit is recorded in Figure 1A. The three-dimensional arrangement of the S gp and its subunits, such as S1 and S2, are illustrated in Figure 1B. The sequence length of the different functional domains of spike proteins observed was of the signal peptide (SP, 1-22 aa), N-terminal domain (22-314 aa), receptor-binding domain (327-600 aa), fusion peptide (797-815 aa), heptad repeat 1(919-1040 aa), heptad repeat 2 (1172-1222 aa), transmembrane domain (1222-1246 aa) and a cytoplasmic domain (1246-1281aa) (Figure 1, **Supplementary Table I**). From the amino acid distribution analysis of different functional domains of the S1 subunit, we found that the amino acid distribution is more scattered in RBD (Figure 2A[A1-A3]). Conversely, the amino acid distribution is less in SP. On the other hand, from the amino acid distribution of HR1 as well as HR2 functional domains of the S2 subunit, we found more scattered amino acid distribution (Figure 2B[B1-B5]). Similarly, the distribution of amino acid residues of complete SARS-CoV-2 S gp and hACE2 is shown in **Supplementary Figure 1 (A and B)**.

Secondary Structure Architecture and Motif Architecture Interface of SARS-CoV-2 S gp and hACE2

Secondary structure architecture of SARS-CoV-2 spike structural components such as α helices, β sheets, and motif architecture such as β hairpins, β turns, and γ turns was analyzed. Figure 3 (A1, A2) shows the number of different components such as α helices (21 numbers), β sheets (51 numbers) of secondary structure architecture and the number of different components of β hairpins (17 numbers), β turns (71 numbers) and γ turns (12 numbers) of motif architecture of SARS-CoV-2 S gp. The several components of secondary structure architecture (α helices, β sheets) and different components of motif architecture (β hairpins, β turns, and γ turns) are shown in Figure 3B. Similarly, secondary structure architecture of hACE2 receptor such as α helices, β sheets, and motif architecture such as β hairpins, β turns, and γ turns were analyzed (Figure S2). Figure S2 (A1, A2) shows the number of different components, such as α helices (31 numbers), β sheets (4 numbers) of

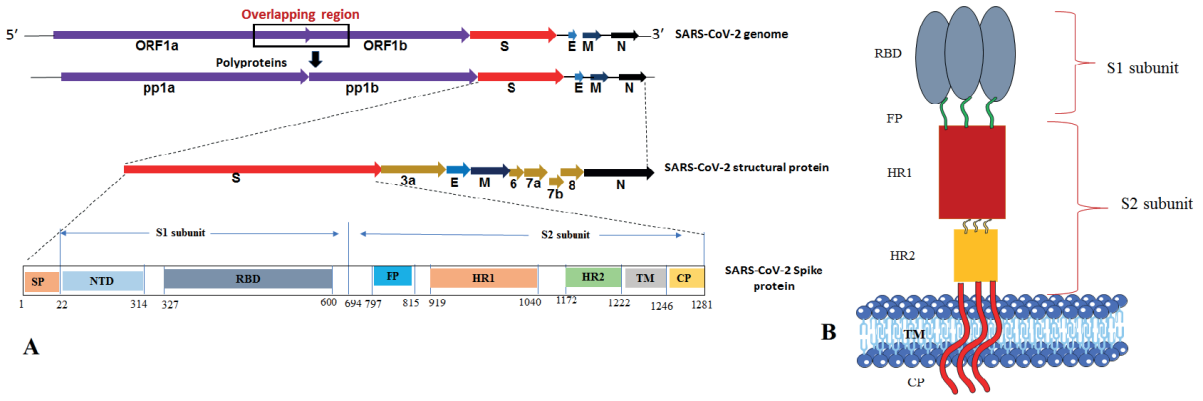

Figure 1. Schematic diagram of spike protein with S1 and S2 subunit. **A,** A schematic diagram shows the total genome of the structural protein of SARS-CoV-2. It also shows spike protein with different regions with respect to S1 and S2 subunit. **B,** Spike protein and its three-dimensional arrangement in ball-shaped diagram which shows S1 and S2 subunits. It also shows the transmembrane (TM), anchor.

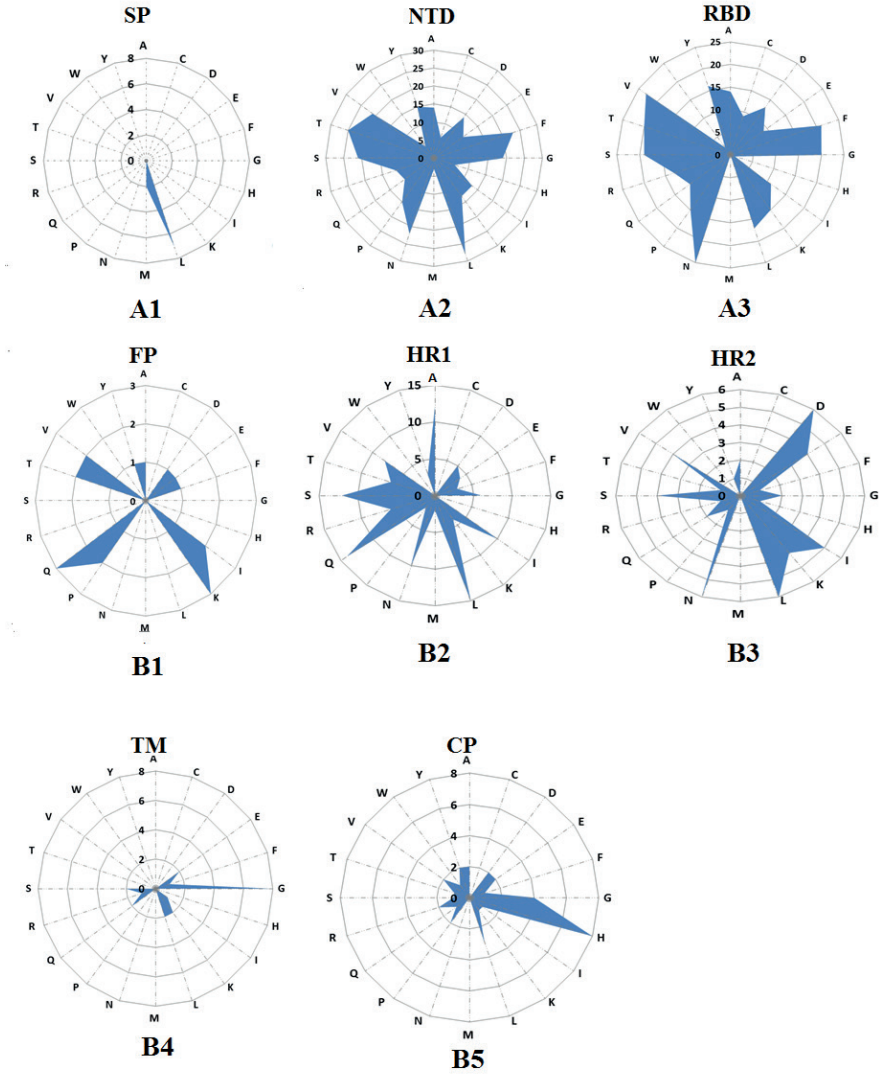

Figure 2. Amino acid residue distribution of the spike protein different functional domains (**A**) S1 subunit (**B**) S1 subunit.

secondary structure architecture and the number of different components of β hairpins (2 numbers), β turns (36 numbers) and γ turns (6 numbers) of motif architecture of hACE2 receptor. The different components of secondary structure architecture (α helices, β sheets) and different components of motif architecture (β hairpins, β turns, and γ turns) are shown in Figure S2B.

Analysis of Tertiary Structure Conformation of S gp of SARS-CoV-2 and Its Architecture

The tertiary configuration of S gp shows the trimeric structure conformation of SARS-CoV-2 (Figure 4). The structure shows the partly open conformation of RBD of S gp of this virus. Figure 4 also shows the conformation of spike protein after 90° rotation, where we have noted the presence of RBD in different chains. These two trimeric structure models of the spike protein show two different orthogonal views. The first one is a side view, and the second one is a top view of the spike protein.

Analysis of the Tertiary Structure Conformation of RBD from S gp of SARS-CoV-2

We documented the position of RBD in the tertiary structure conformation in a single chain of SARS-CoV-2 S gp, and it is recorded in Figure 5A.

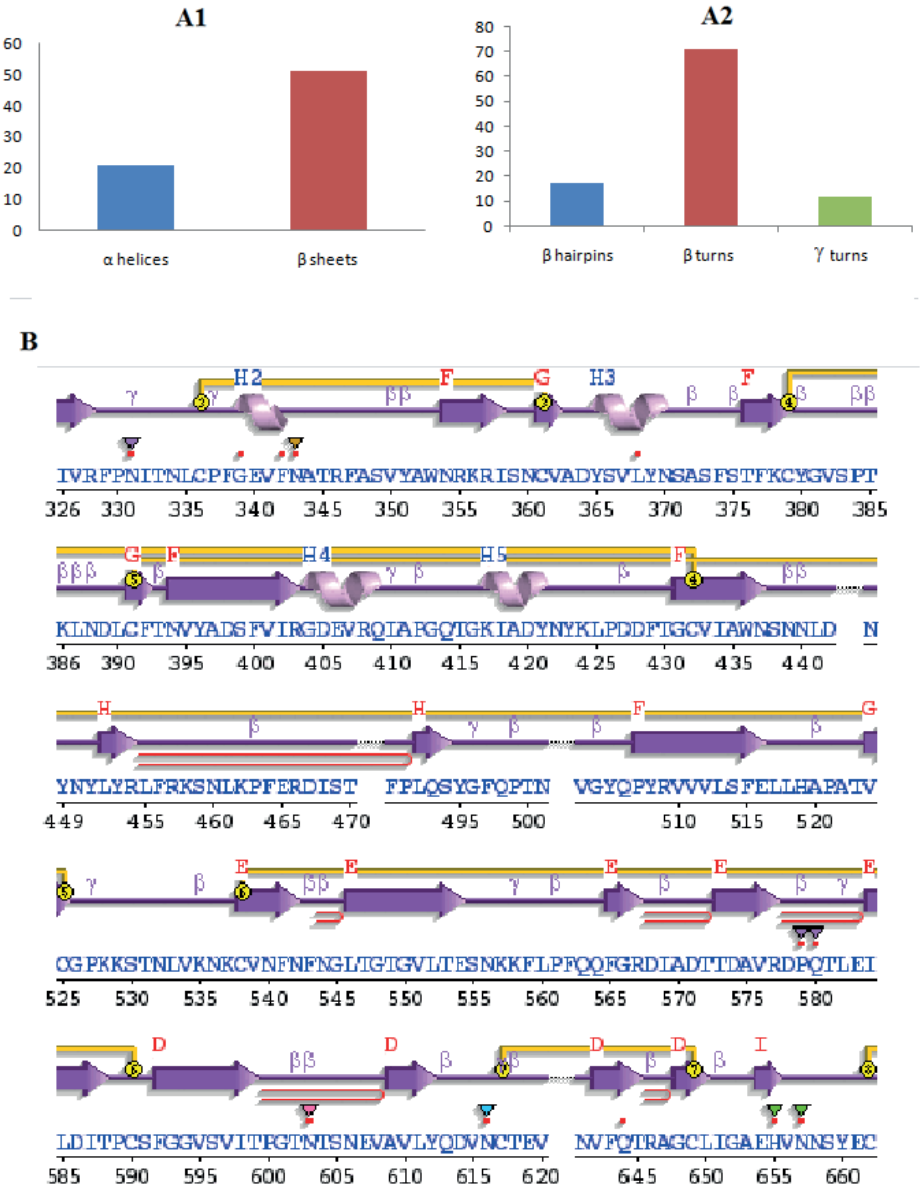

Figure 3. Secondary structure architecture and motif architecture interface of SARS-CoV-2 spike protein (**A**) Number of different components of secondary structure architecture (**A1**) and the number of different components of motif architecture (**A2**). **B,** Different components of secondary structure architecture and different components of the motif architecture of RBD.

Here, we have noted that two single-chain structural models of the spike protein show two different orthogonal views. One is the side view, and another is the top view of the spike protein. Here we have also located RBD on the single chain of the spike protein, and the depiction is recorded. For this analysis, we have separated the particular structure of the RBD, and it was used for further analysis. Using this portion, we have developed a tertiary structure conformation of the RBD part, illustrated in Figure 5B. Here, all the secondary structures were also observed in the RBD.

Tertiary Structure Interface SARS-CoV-2 S gp with hACE2 Protein Interaction

The molecular interaction of SARS-CoV-2 S gp with hACE2 in the tertiary structure interface is shown in Figure 6. We observed eleven hydrogen bonds in the molecular interaction in this virus S gp and hACE2 receptor. The primary interaction residues of spike protein with ACE2 and its hydrogen bond length are noted in **Supplementary Table II**.

The spike protein interaction residues include LYS417, GLY446, TYR449, ASN487, GLN493, GLY496, THR500, and GLY502. Similarly, inter-

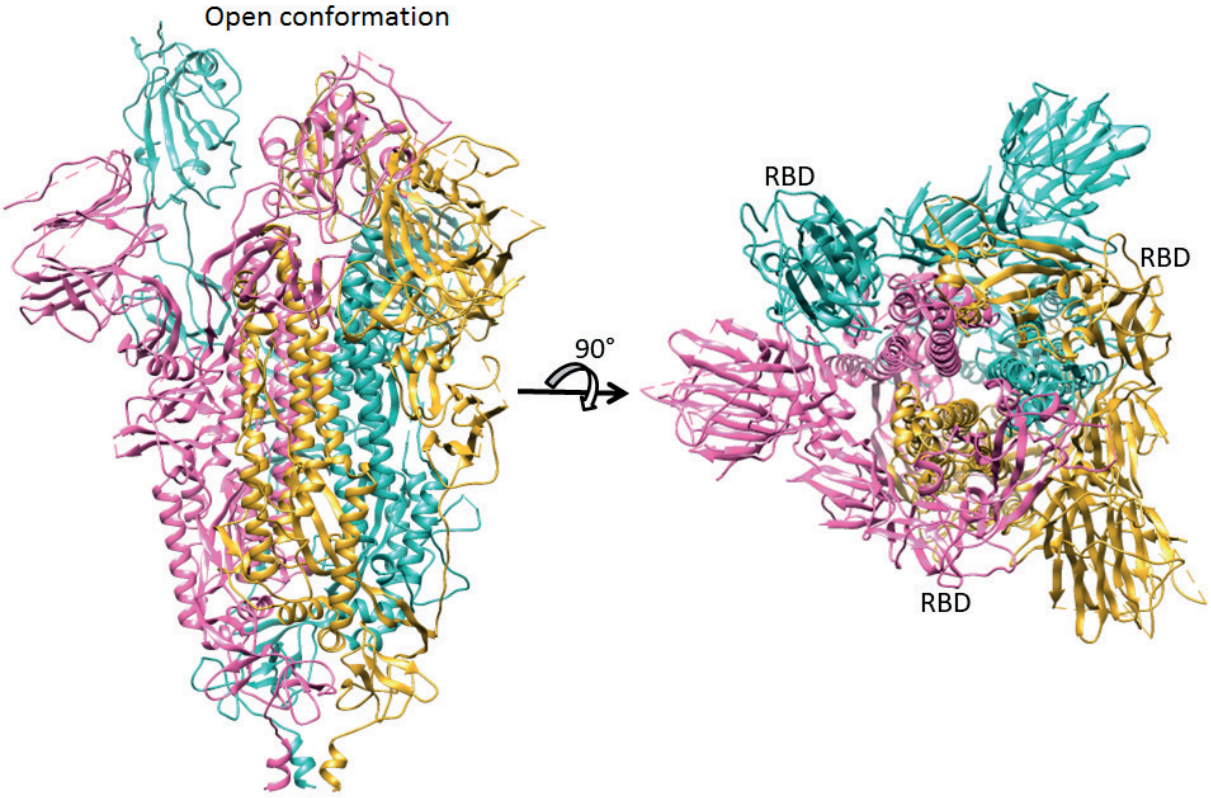

Figure 4. The tertiary structure of spike proteins of SARS-CoV-2 in trimeric conformation. Another part of the figure also shows the conformation of spike proteins after 90° rotation.

action residues of ACE2 include ASP30, GLN42, ASP38, GLN24, TYR83, GLU35, LYS353, TYR41, and LYS353.

Tertiary Structure Interface SARS-CoV S gp with hACE2 Protein Interaction

Further, we aimed to observe the difference between the molecular interactions of SARS-CoV-2 S gp and SARS-CoV S gp with hACE2 protein. We have found a tertiary structure interface of SARS-CoV S gp with hACE2 protein interaction (Figure 7A). We have noted that the spike protein is a glycoprotein trimer, having three chains (Chain-A, Chain-B, and Chain-C). Conversely, the molecular interaction between spike protein and the ACE2 receptor has been recorded in Figure 7B. There are several residues which have taken part in this interaction. The core interaction residues of spike protein with ACE2 and its hydrogen bond length are noted in **Supplementary Table III**. The spike protein interaction residues include TYR436, ASN473, GLY482, TYR484, THR486, and GLN492. Similarly, interaction residues of ACE2 include ASP38, GLN42, TYR83, LYS353, ASN330, TYR41, and GLN325. Seven hydrogen bonds were noted in this molecular interaction.

Sequence Alignment (SA) of S gp (SARS-CoV RBD and SARS-CoV-2 RBD)

We have performed SA between the RBD of S gp of SARS-CoV and SARS-CoV-2, which is recorded in Figure 7C. In this figure, identical and similar positions of amino acids are highlighted within RBD. In this study, we found several conserved regions between these two proteins.

Analysis of Conservation Prototype with Highly Conserved Amino Acids of S gp of SARS-CoV-2 and hACE2 Receptor

We found highly conserved amino acids of S gp in trimeric form, as shown in Figure 8 (A1-A3). We documented three types of conserved protein structures. The first one is a cartoon structure of the conserved regions (Figure 8A1). The second one is the surface conservation of the protein ranging from low to high conserved regions (Figure 8A2). Third one is the highly conserved residues on the surface of the protein (Figure

8A3). Again, we noted conserved amino acids of SARS-CoV-2 in one chain (B chain) (Figure 8 [B1, B2]). In this case, we noted the conserved surface regions of one chain, i.e., chain B (Figure 8B1). We also documented highly conserved amino acids of one chain, i.e., chain B (Figure 8B2).

Some highly conserved residues are: PRO139, CYS136, PHE133, PHE106, GLY107, PHE86, ASN188, LEU270, PHE55, ARG44, CYS291, GLU298, GLU988, GLY971, CYS538, THR1006, CYS617, SER596, ILE312, CYS671, ILE312, LYS947, PRO1057, ALA944, THR723, VAL1065, THR1105, PHE1095, VAL1128, CYS361, ASP398, TYR423 etc. We also noted highly conserved amino acids of hACE2 receptor in the trimeric form as shown in Figure 9 (A and B). The first one is a cartoon structure of the conserved regions of the hACE2 receptor (Figure 9A). The second one is the surface conservation of the hACE2 receptor (Figure 9B). The amino acid conservation of the hACE2 receptor was noted in Figure 9C. Several highly conserved residues can be predicted from the structure.

Prediction of Physicochemical Parameters Such as Glycosylation Site, Hydrophobicity Prediction, Determination Of Grand Average Of Hydropathicity (GRAVY) of SARS-CoV-2 S gp and hACE2

Our result has shown 10 numbers of N-glycosylation sites in SARS-CoV-2 S gp and 4 numbers of N-glycosylation sites in the hACE2 receptor **(Figure S3 A, B, C)**. Similarly, 4 numbers of O-glycosylation were detected in this virus S gp, and no O-glycosylation sites were detected in the hACE2 receptor **(Supplementary Figure 4)**.

The hydrophobicity pattern of different domains in both subunits (S1 and S2 subunit) of this

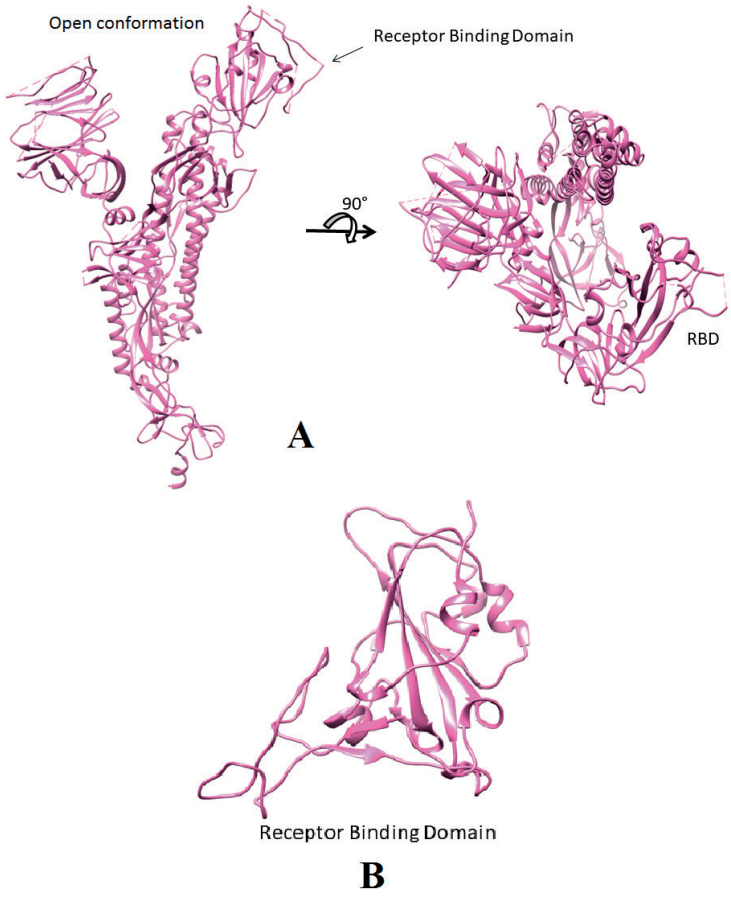

Figure 5. The tertiary structure conformation of RBD in single-chain from spike protein of SARS-CoV-2 (**A**) The position of RBD in the tertiary structure conformation in a single chain of the spike proteins of SARS-CoV-2. The single chain is again is rotated at 90^0 rotation and shows the location of RBD. **B,** The figure shows the tertiary structure of the separated particular structure of RBD.

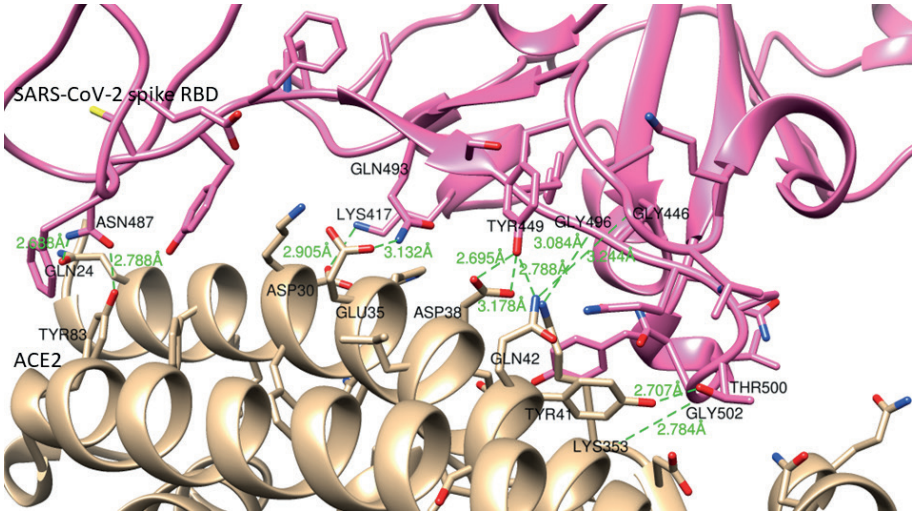

Figure 6. Molecular interaction of SARS-CoV-2 spike protein with hACE2 receptor in tertiary architecture interface. It also shows the hydrogen bond between interaction residue between SARS-CoV-2 spike protein with hACE2.

virus S gp is also noted in **Supplementary Figure 5**. We found a higher peak in NTD and RBD, which shows a positive hydrophobicity score **(Supplementary Figure 5)**. Similarly, hydrophobicity of complete hACE2 and S gp of this virus is recorded in **Supplementary Figure 6 (A, B)**.

The GRAVY pattern of different domains in both subunits (S1 and S2) of S gp of this virus is depicted in **Supplementary Figure 7.** We found GRAVY positive values in the SP domain, and other domains showed negative GRAVY score **(Supplementary Figure 7)**. Similarly, GRAVY of complete SARS-CoV-2 S gp and hACE2 are recorded in **Figure S8**. Both have shown negative GRAVY values.

MDS Analysis

The different parameters (RMSD, RMSF, hydrogen bonds of the interacting molecules, solvent accessible surface) of MDS analysis of S gp of this virus with hACE2 complex are noted in Figure 10 (A-D.)

The result in Figure 10A is visualizing an average RMSD value ranging from 0.00 to 0.25 nm with more or less stability. The RMSF of the ligand is also depicted in the plot, and it fluctuates within a range of 0.05 to 0.35 nm (Figure 10B). While Figure 10C and Figure 10D indicate the compactness of the structure and the number of hydrogen bonds and solvent accessible surface, making the protein complex stable.

Discussion

The recent pandemic of SARS-CoV-2 virus is now a major concern in public health throughout the world. It has been noted that S gp of SARS-CoV-2 is a crucial element produced as the homotrimers at the viral surface. There are two structural and functional subunits (S1 and S2 subunits). S1 subunit is necessary for interaction with the protein of the host cell. It has been noted that there is a cleavage site at the borderline between the S1 and S2 subunit. That means that spike protein is split into the S1 and S2 subunits from this cleavage site. Recently, Walls et al[42] also found one furin cleavage site within S gp of SARS-CoV-2, which is located within the border among the S1 subunit and S2 subunit. In this work, we have characterized the secondary structure architecture of S1 and S2 subunits of S gp. Here, we also performed an analysis of the amino acid distribution of S1 and S2 subunits. Protein architecture plays a vital role in adaptive evolution. Therefore, we need to understand the protein architecture to understand adaptive evolution[43]. Conversely, Patino-Galindo et al[44] have shown that the RNA virus's secondary structure plays an essential role in understanding the RNA virus phylogenies.

The structure of RBD has a distinct role in influencing the binding of hACE2[42]. Therefore, we analyzed the secondary structure architecture and amino acid distribution of S gp RBD of SARS-

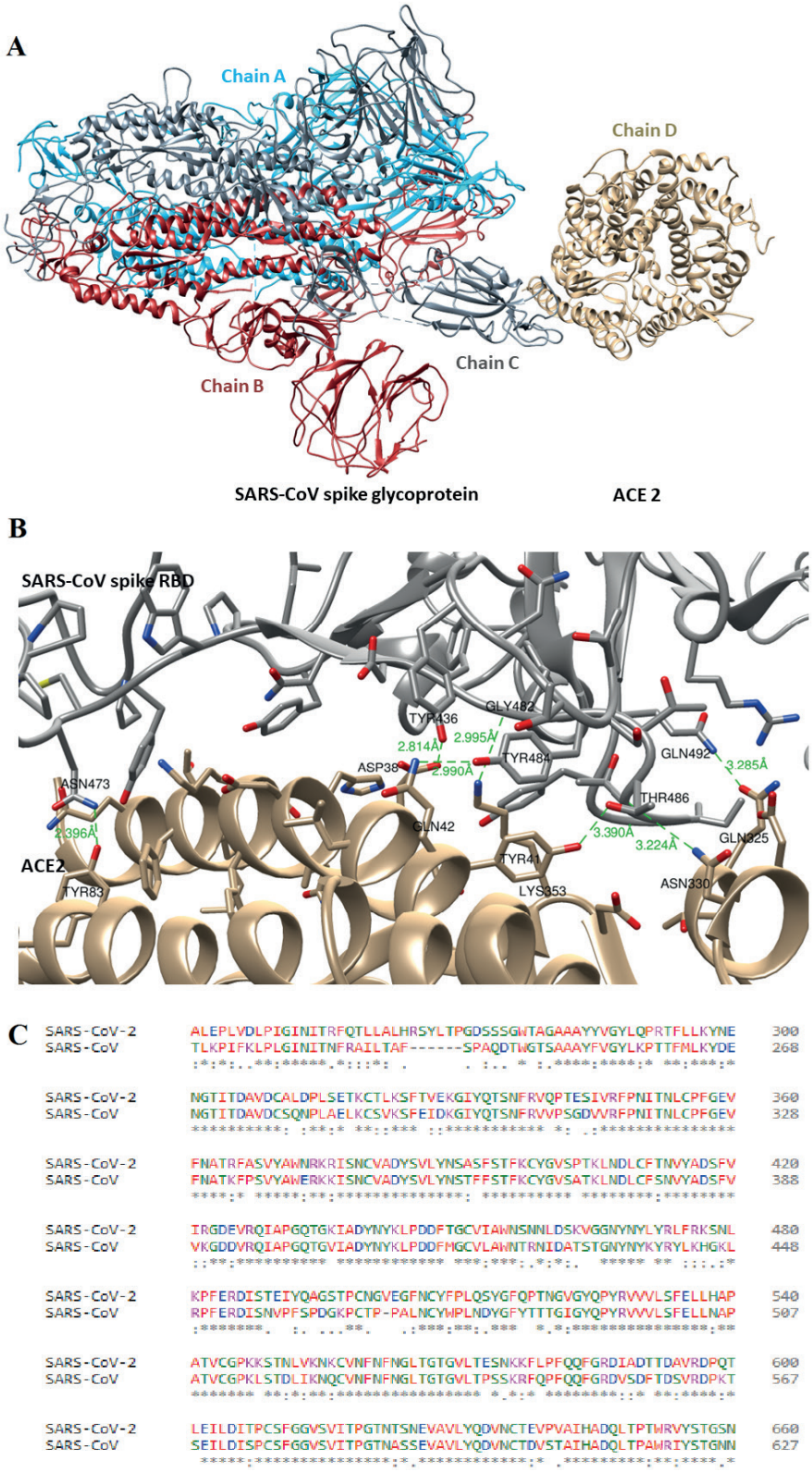

Figure 7. Tertiary structure interface SARS-CoV spike protein with hACE2 receptor interaction and pairwise alignment between the spike proteins of SARS-CoV-2 and SARS-CoV. **A,** Tertiary structure interface of SARS-CoV spike protein with hACE2 receptor interaction. **B,** Molecular interaction of SARS-CoV spike protein with hACE2 receptor in tertiary architecture interface. It also shows the hydrogen bond between interaction residue between SARS-CoV spike protein with hACE2. **C,** Sequence alignment between SARS-CoV-2 and SARS-CoV spike proteins.

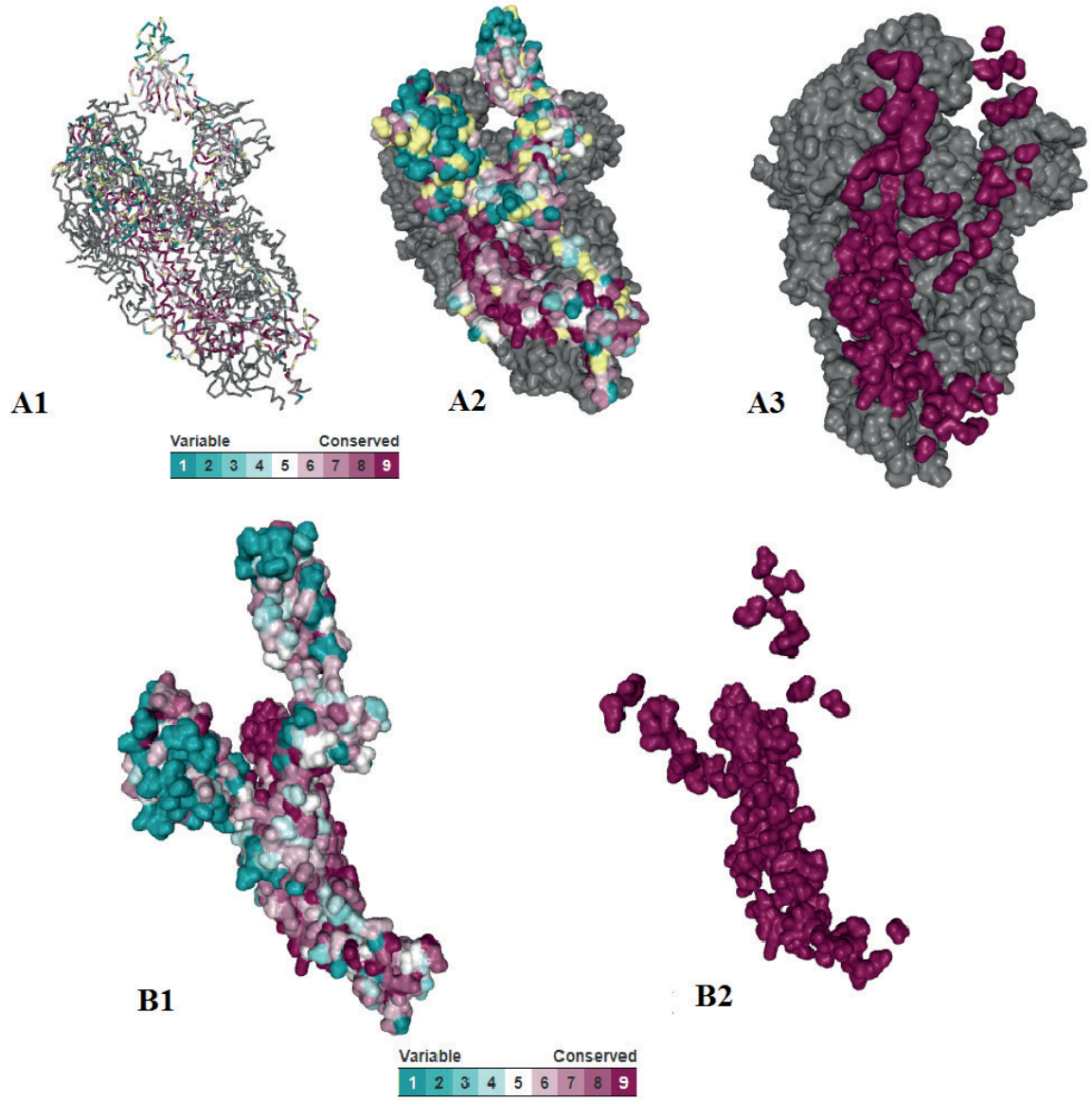

Figure 8. Conserved regions and highly conserved amino acids of spike protein SARS-CoV-2 in trimeric form and only one chain (B chain). **(A), (A1)** cartoon structure of the conserved regions (**A2**) the surface conservation of the protein shows conserved regions (**A3**) highly conserved residues in the surface of the protein in a trimeric form **(B)** (**B1**) Low to high surface conservation in one chain (**B2**) highly conserved residues in the surface of the protein in one protein chain.

CoV-2. Walls et al[42] described that among S gp of this virus, residues 473-486 affect the binding of RBD and hACE2.

An understanding of the tertiary structure of the trimeric form of SARS-CoV-2 spike protein is essential. There are three shapes of the tertiary structure of the trimeric structure of S gp of this virus: open form, partially open form, and closed form. These forms can help us understand the mechanism of interaction with hACE2 and the entry of SARS-CoV-2. Recently, Shang et al[45] developed a structure of mouse coronavirus S gp complex form with ACE2 and proposed a method for mouse coronavirus entry into the body. In the same way, this analysis might help future researchers to develop different assays and understand the process of this viral entry into the human body.

Actually, there is a very specific question: are the interaction of SARS-CoV-2 S gp with hACE2 and the interaction of SARS-CoV S gp with hACE2 alike or not. To answer it, we have per-

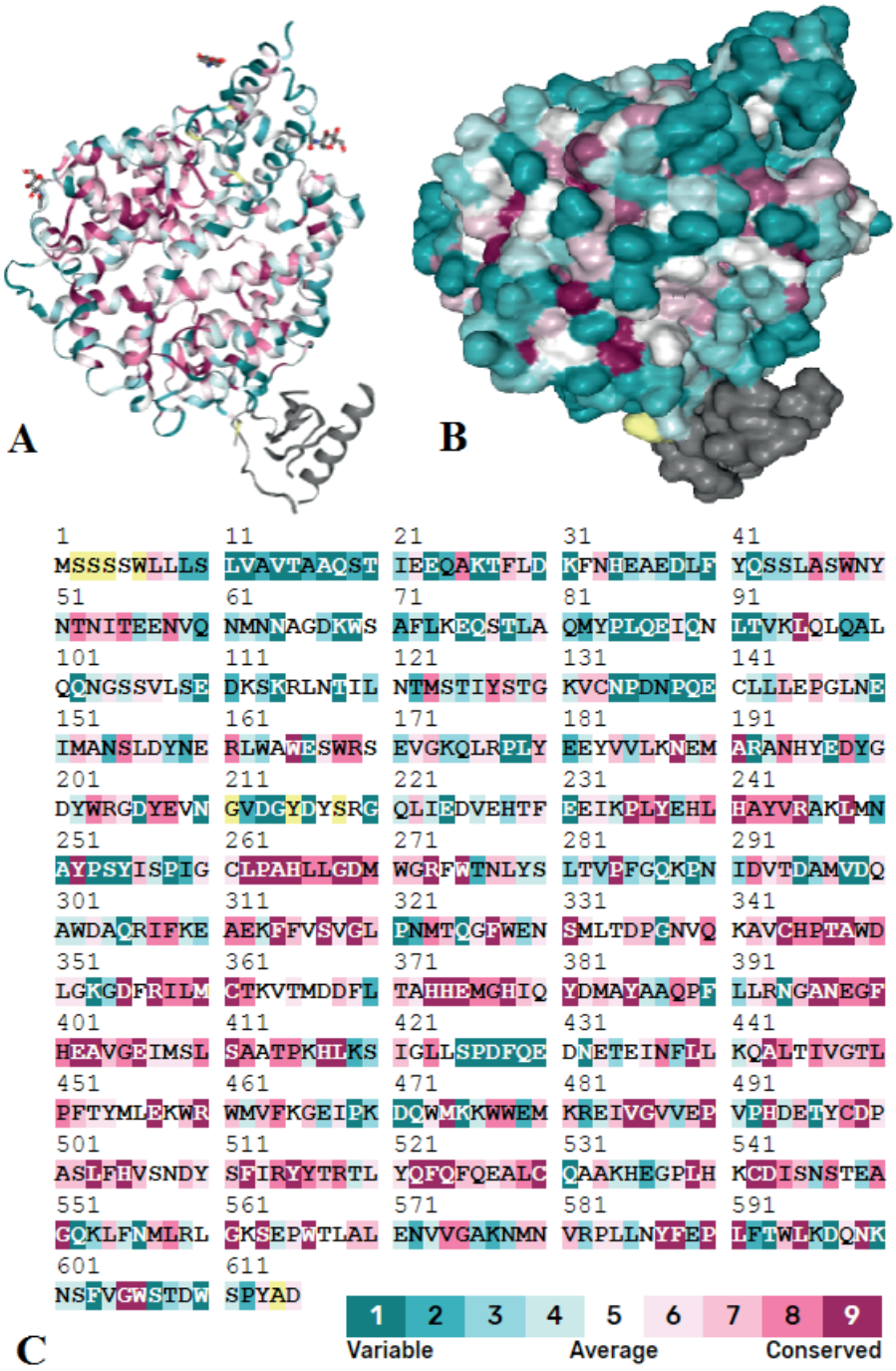

Figure 9. Conserved regions and highly conserved amino acids of hACE2 receptor in trimeric form. **A,** Cartoon structure of the conserved regions. **B,** The surface conservation of the protein shows conserved regions (**C**) Conserved in residues of hACE2 receptor.

formed the interaction analysis. Here, we found that more hydrogen bonds are formed between SARS-CoV-2 spike protein with hACE2 receptor. It also shows that this interaction is more robust than that of SARS-CoV S gp with hACE2.

Sequence alignment between two CoVs showed several identical and similar positions of amino acids. This analysis indicates that there are several highly conserved regions between these two CoVs. Our result corroborates Walls et al[42] about

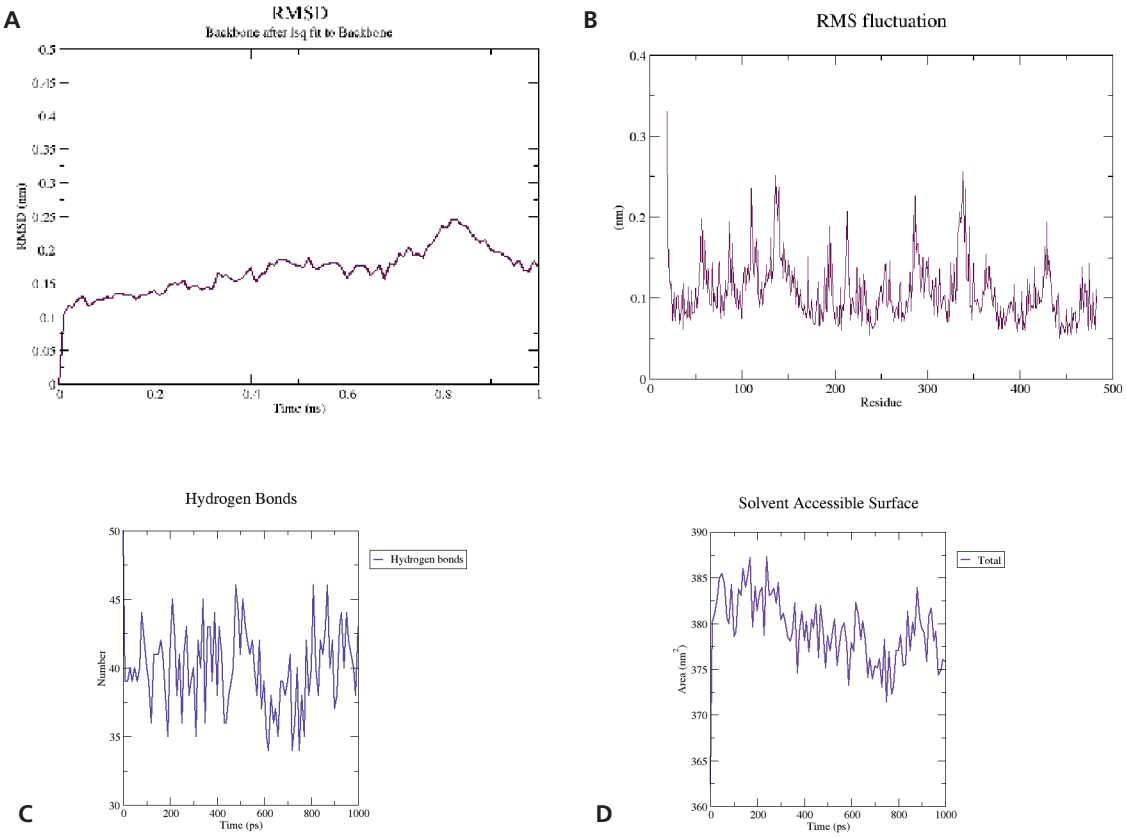

Figure 10. Molecular dynamics simulation (MDS) analyses spike protein of SARS-CoV-2 with hACE2 receptor complex (**A**) RMSD of the spike protein of SARS-CoV-2 with hACE2 receptor complex (**B**) RMSF of S -protein of SARS-CoV-2 with hACE2 receptor complex (**C**) hydrogen bonds of the interacting molecules of S-protein of SARS-CoV-2 with hACE2 receptor complex (**D**) solvent accessible surface of S-protein of SARS-CoV-2 with hACE2 receptor complex.

the same kind of sequence similarity and RBD identity among these two viruses.

Finally, we found low to high conservation residues of the spike protein in the trimeric protein and a single chain format. Usually, the conservation scores represent the evolutionary rate of a particular position of a protein. It has been noted that some parts of a particular protein can evolve at a rapid rate. These protein regions are called the "variable region"[46]. On the other hand, some protein parts evolve slowly and are called "conserved region". We found several highly conserved residues in the surface of the trimeric protein and a single chain of S gp of SARS-CoV-2. Our result corroborates Walls et al[42] about the same type of highly conserved amino acids exists in the surface of S gp of SARS-CoV-2. We also found highly conserved residues in the hACE2 receptor in surface conservation and cartoon structure. These conserved regions in both proteins (SARS-CoV-2 S gp and hACE2) play a vital role in the interaction between this virus S gp and hACE2.

We next analyzed physicochemical parameters such as glycosylation site, hydrophobicity prediction, GRAVY of S gp SARS-CoV-2, and hACE2. These physicochemical parameters play an important role in binding between SARS-CoV-2 S gp and hACE2. We observed high hydrophobicity at its higher peak in RBD, and it has a vital role in hACE2 receptor binding.

Our MDS results showed the stability of S gp of SARS-CoV-2 with hACE2 receptor complex with respect to the time scale. Moreover, there is a possibility to carry out calculations on a prolonged time scale, on a higher molecular system, or with advanced physical perfection. Current simulations showed a time scale of 1 ns. The entire atom molecular system was an RBD of S gp of this virus bound with hACE2

protein: analysis of same times scale and same molecular size in this field. The dynamic motion of H bonds in the plot (Figure 10C) represents the correlated motions, a block-of-time interval of 1000 ps. The RSMD map (nm) is represented in the block-of-time-interval of 0-1 ns, which shows the results within a limit of 0.25 ns. The arrangement between the two different time intervals suggests that a joined picture of the correlated motion is with respect to 1ns. Our result shows that the RMS fluctuation is within a significant range (long to the middle) with a 500 residues limit. The studied RMSF highest fluctuations value was noted as 0.35 nm (Figure 10B). This work has an extensive converged in light of MD simulation that reveal the RBD-hACE2 protein complex's dynamic characteristics.

Conclusions

Our research shows some interesting findings for the interaction of SARS-CoV-2 with hACE2. Firstly, increased number hydrogen bonds were formed between S gp of SARS-CoV-2 with hACE2 compared to SARS-CoV S gp with hACE2. Here, we can conclude that an interaction of S gp of SARS-CoV-2 with hACE2 is steadier than SARS-CoV S gp with hACE2 receptor. That is why single particle of the SARS-CoV-2 virus can attach tightly to the surface of the cells for entry. Secondly, our secondary and tertiary structure analysis of spike protein will provide a platform for future researchers to answer the virus's critical entry mechanism into the host cell and ultimately into the human body. Thirdly, our conservation prototype analysis will help understand the conserved and accessible epitopes that can help in the vaccine development process with the more stable antigenic part of the S gp. Fourthly, MD simulation of SARS-CoV-2 S gp with hACE2 protein shows a stable interaction between the two proteins. Overall, this study will provide important structural information to help develop the therapeutic molecules against this killer virus.

Acknowledgments

This study was supported by Hallym University Research Fund and by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2020R1C1C1008694 & NRF-2020R1I1A3074575).

Authors Contributions

C.C. collected, analyzed, and interpreted the data. A.R.S., M.B., and B.M. evaluated data, developed figures. C.C., A.R.S., and G.S. wrote, edited, and revised the manuscript. C.C. and S.S.L. supervised the whole work. All authors have read and approved the final manuscript.

Orcid Number

Chiranjib Chakraborty ORCID ID = 0000-0002-3958-239X.

Conflict of Interest

The Authors declare that they have no conflict of interests.

Data Availability Statement

All the data, score and model are generated or used during the study appear in the submitted article.

References

1) Chan JF, Yuan S, Kok KH, To KK, Chu H, Yang J, Xing F, Liu J, Yip CC, Poon RW, Tsoi HW. A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster. Lancet 2020; 395: 514-523.
2) Chakraborty C, Sharma AR, Sharma G, Bhattacharya M, Lee SS. SARS-CoV-2 causing pneumonia-associated respiratory disorder (COVID-19): diagnostic and proposed therapeutic options. Eur Rev Med Pharmacol Sci 2020: 24: 4016-4026.
3) Chakraborty C, Sharma AR, Bhattacharya M, Sharma G, Lee SS. The 2019 novel coronavirus disease (COVID-19) pandemic: a zoonotic prospective. Asian Pac J Trop Med 2020; 6: 242-246.
4) Liu SL, Saif L. Emerging viruses without borders: the Wuhan Coronavirus. Viruses 2020; 12: 130-131.
5) Gralinski LE, Menachery VD. Return of the Coronavirus: 2019-nCoV. Viruses 2020; 12: 135-143.
6) Ksiazek TG, Erdman D, Goldsmith CS, Zaki SR, Peret T, Emery S, Tong S, Urbani C, Comer JA, Lim W, Rollin PE. A novel coronavirus associated with severe acute respiratory syndrome. N Engl J Med 2003; 348: 1953-66.
7) Drosten C, Günther S, Preiser W, Van Der Werf S, Brodt HR, Becker S, Rabenau H, Panning M, Kolesnikova L, Fouchier RA, Berger A. Identification of a novel coronavirus in patients with severe acute respiratory syndrome. N Engl J Med 2003; 348: 1967-1976.
8) Zaki AM, Van Boheemen S, Bestebroer TM, Osterhaus AD, Fouchier RA. Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia. N Engl J Med 2012; 367: 1814-1820.
9) Gorbalenya AE, Baker SC, Baric R, Groot RJ, Drosten C, Gulyaeva AA, Haagmans BL, Lau-

ber C, Leontovich AM, Neuman BW, Penzar D. Coronaviridae Study Group of the International Committee on Taxonomy of Viruses. The species severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. Nat Microbiol 2020; 5: 536-544.
10) Tortorici MA, Veesler D. Structural insights into coronavirus entry. Adv Virus Res 2019; 105: 93-116.
11) Beniac DR, Andonov A, Grudeski E, Booth TF. Architecture of the SARS coronavirus prefusion spike. Nat Struct Mol Biol 2006; 13: 751-752.
12) Hulswit RJ, De Haan CA, Bosch BJ. Coronavirus spike protein and tropism changes. Adv Virus Res 2016; 96: 29-57.
13) Li F. Evidence for a common evolutionary origin of coronavirus spike protein receptor-binding subunits. J Virol 2012; 86: 2856-2858.
14) Jiang S, Du L, Shi Z. An emerging coronavirus causing pneumonia outbreak in Wuhan, China: calling for developing therapeutic and prophylactic strategies. Emerg Microbes Infect 2020; 9: 275-277.
15) Ou X, Liu Y, Lei X, Li P, Mi D, Ren L, Guo L, Guo R, Chen T, Hu J, Xiang Z. Characterization of spike glycoprotein of SARS-CoV-2 on virus entry and its immune cross-reactivity with SARS-CoV. Nat Commun 2020; 11: 1-12.
16) Du L, He Y, Zhou Y, Liu S, Zheng BJ, Jiang S. The spike protein of SARS-CoV--a target for vaccine and therapeutic development. Nat Rev Microbiol 2009; 7: 226-236.
17) Liu S, Xiao G, Chen Y, He Y, Niu J, Escalante CR, Xiong H, Farmar J, Debnath AK, Tien P, Jiang S. Interaction between heptad repeat 1 and 2 regions in spike protein of SARS-associated coronavirus: implications for virus fusogenic mechanism and identification of fusion inhibitors. Lancet 2004; 363: 938-947.
18) Du L, Zhao G, Chan CC, Sun S, Chen M, Liu Z, Guo H, He Y, Zhou Y, Zheng BJ, Jiang S. Recombinant receptor-binding domain of SARS-CoV spike protein expressed in mammalian, insect and E. coli cells elicits potent neutralizing antibody and protective immunity. Virology 2009; 393: 144-150.
19) Bhattacharya M, Sharma AR, Patra P, Ghosh P, Sharma G, Patra BC, Lee SS, Chakraborty C. Development of epitope-based peptide vaccine against novel coronavirus 2019 (SARS-COV-2): Immunoinformatics approach. J Med Virol 2020; 92: 618-631.
20) 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: 1-9.
21) Song Z, Xu Y, Bao L, Zhang L, Yu P, Qu Y, Zhu H, Zhao W, Han Y, Qin C. From SARS to MERS, Thrusting Coronaviruses into the Spotlight. Viruses 2019; 11: 1-28.
22) Wang Q, Zhang Y, Wu L, Niu S, Song C, Zhang Z, Lu G, Qiao C, Hu Y, Yuen KY, Wang Q. Structural and functional basis of SARS-CoV-2 entry by using human ACE2. Cell 2020; 181: 894-904.
23) Yan R, Zhang Y, Li Y, Xia L, Guo Y, Zhou Q. Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2. Science 2020; 367: 1444-1448.
24) Lan J, Ge J, Yu J, Shan S, Zhou H, Fan S, Zhang Q, Shi X, Wang Q, Zhang L, Wang X. Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. Nature 2020; 581: 215-220.
25) Coordinators NR. Database resources of the national center for biotechnology information. Nucleic Acids Res 2018; (Database issue) 46: D8-D13.
26) Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. The protein data bank. Nucleic Acids Res 2000; 28: 235-242.
27) Artimo P, Jonnalagedda M, Arnold K, Baratin D, Csardi G, De Castro E, Duvaud S, Flegel V, Fortier A, Gasteiger E, Grosdidier A. ExPASy: SIB bioinformatics resource portal. Nucleic Acids Res 2012; 40: W597-603.
28) Brendel V, Bucher P, Nourbakhsh IR, Blaisdell BE, Karlin S. Methods and algorithms for statistical analysis of protein sequences. Proc Natl Acad Sci U S A 1992; 89: 2002-2006.
29) Laskowski RA, Jabłońska J, Pravda L, Vařeková RS, Thornton JM. PDBsum: structural summaries of PDB entries. Protein Sci 2018; 27: 129-134.
30) de Beer TA, Berka K, Thornton JM, Laskowski RA. PDBsum additions. Nucleic Acids Res 2014; 42: 292-296.
31) Rodríguez-Guerra Pedregal J, Maréchal JD. Marechal, PyChimera: use UCSF Chimera modules in any Python 2.7 project. Bioinformatics 2018; 34: 1784-1785.
32) Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. UCSF Chimera--a visualization system for exploratory research and analysis. J Comput Chem 2004; 25: 1605-1612.
33) Sievers F, Higgins DG. Clustal Omega for making accurate alignments of many protein sequences. Protein Sci 2018; 27: 135-145.
34) Ashkenazy H, Erez E, Martz E, Pupko T, Ben-Tal N. ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. Nucleic Acids Res 2010; 38: 529-533.
35) Ashkenazy H, Abadi S, Martz E, Chay O, Mayrose I, Pupko T, Ben-Tal N. ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. Nucleic Acids Res 2016; 44: 344-350.
36) Julenius K, Mølgaard A, Gupta R, Brunak S. Prediction, conservation analysis, and structural characterization of mammalian mucin-type O-glycosylation sites. Glycobiology 2005; 15: 153-164.
37) Hamby SE, Hirst JD. Prediction of glycosylation sites using random forests. BMC Bioinformatics 2008; 9: 1-13.
38) Xi E, Venkateshwaran V, Li L, Rego N, Patel AJ, Garde S. Hydrophobicity of proteins and nanostructured solutes is governed by topographical and chemical context. Proc Natl Acad Sci U S A 2017; 114: 13345-13350.
39) Kyte J, Doolittle RF. A simple method for displaying the hydropathic character of a protein. J Mol Biol 1982; 157: 105-132.
40) Gasteiger E, Gattiker A, Hoogland C, Ivanyi I, Appel RD, Bairoch A. ExPASy: the proteomics

server for in-depth protein knowledge and analysis. Nucleic Acids Res 2003; 31: 3784-3788.
41) Van Der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE, Berendsen HJ. GROMACS: fast, flexible, and free. J Comput Chem 2005; 26: 1701-1718.
42) 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; 181: 281-292e6.
43) Moutinho AF, Trancoso FF, Dutheil JY. The impact of protein architecture on adaptive evolution. Mol Biol Evol 2019; 36: 2013-2028.
44) Patino-Galindo JÁ, González-Candelas F, Pybus OG. The Effect of RNA Substitution Models on Viroid and RNA Virus Phylogenies. Genome Biol Evol 2018; 10: 657-666.
45) Shang J, Wan Y, Liu C, Yount B, Gully K, Yang Y, Auerbach A, Peng G, Baric R, Li F. Structure of mouse coronavirus spike protein complexed with receptor reveals mechanism for viral entry. PLoS Pathog 2020; 16: e1008392.
46) Chakraborty C, Agoramoorthy G, Hsu MJ. Exploring the evolutionary relationship of insulin receptor substrate family using computational biology. PLoS One 2011; 6: e16580.