CHAPTER 4

Study on the salient structural features of SARS-CoV-2 Main protease (Mpro) and Spike protein (S protein) along with the conformational accessibility of S-protein

Study on the salient structural features of SARS-CoV-2 Main protease (Mpro) and Spike protein (S protein) along with the conformational accessibility S-protein

4.1. Abstract:

The COVID-19 pandemic has driven extensive research efforts aimed at developing preventive and therapeutic strategies. While the rapid development of multiple vaccines has significantly contributed to controlling the spread of the virus, the emergence of new SARS-CoV-2 variants continues to challenge global health efforts. Therapeutic development has particularly focused on two key target proteins: the main protease (Mpro) and the spike glycoprotein (S protein).

Mpro, also known as 3CLpro, is a crucial cysteine hydrolase involved in viral replication by cleaving polyproteins into functional units. Given its essential role in the viral life cycle, it has been widely recognized as a promising target for broad-spectrum antiviral agents. The second key protein, S protein, facilitates viral entry into host cells through its interaction with the angiotensin-converting enzyme-2 (ACE2) receptor, making it a critical focus for antiviral drug development and vaccine design.

This study presents a preliminary bioinformatics analysis of Mpro and S protein using various computational tools, including Expasy ProtParam, GOR IV, Clustal Omega, PONDR, PDBsum, and the PRODIGY server. Structural assessments revealed the thermostability of both proteins, along with insights into their secondary structures, conserved regions, and intrinsically disordered regions. Conformational accessibility analysis of the S protein indicated that its open state exhibits the highest affinity for the ACE2 receptor, while the closed state shows minimal accessibility. These findings contribute to a deeper understanding of SARS-CoV-2 structural dynamics, offering valuable insights for future therapeutic and vaccine development.

4.2. Introduction:

The main protease (Mpro), also known as the 3-chymotrypsin-like protease (3CLpro), is a highly conserved cysteine hydrolase found in β-coronaviruses. Studies have shown that 3CLpro plays a critical role in viral replication, making it a key target for developing treatments against coronavirus-related infectious diseases, including COVID-19 [1]. Given its significance, Mpro has been extensively studied as a potential therapeutic target. In SARS-CoV-2, Mpro is essential for processing viral polyproteins, ensuring they are properly cleaved to produce functional viral RNA. Specifically, it facilitates the cleavage of the 1ab polyprotein, a crucial step in viral replication [2].

Once Mpro completes this process, the resulting nonstructural proteins (nsps) contribute to assembling the viral replication-transcription complex, which is responsible for synthesizing new viral RNA [3,4]. Additionally, Mpro plays a key role in encoding two polyproteins, pp1b and pp1ab, and is responsible for cleaving 11 other sites, ultimately leading to the production of 16 essential nonstructural proteins [5,6]. Due to its vital function in the viral life cycle, Mpro remains a prime target for antiviral drug development, as inhibiting its activity could significantly disrupt the virus's ability to replicate and spread.

In addition to the Mpro, another key target for SARS-CoV-2 is the spike protein (S protein). This glycosylated protein densely covers the surface of the virus and plays a crucial role in facilitating viral entry into host cells. It specifically binds to the ACE2 receptor on the host cell surface, initiating the infection process [7]. Once the S protein attaches to the receptor, the host enzyme TM protease serine 2 (TMPRSS2), a type II transmembrane serine protease, activates the S protein and enables the virus to enter the cell. Following viral entry, the viral RNA is released into the host cell, where it serves as a template for polyprotein translation. These polyproteins undergo cleavage and assembly, forming the replicase-transcriptase complex, which drives the replication and transcription of the viral genome. As the viral RNA replicates, structural proteins are synthesized, assembled, and packaged within the host cell before newly formed viral particles are released [8]. Among all SARS-CoV-2 structural proteins, the S protein is particularly significant as it serves as the primary antigen responsible for triggering host immune responses and stimulating the production of neutralizing antibodies. Due to its critical role in viral infection, the S protein has been a major focus for vaccine development and antiviral drug research [9].

Given their essential functions in the viral life cycle, both Mpro and the S protein are considered prime targets for therapeutic intervention against SARS-CoV-2.

One of the key characteristics of the S protein is its trimeric structure, which can adopt different conformations depending on the positioning of the receptor-binding domain (RBD). These conformations determine whether the S protein is in an "open" state—allowing interaction with the ACE2 receptor—or a "closed" state, where the receptor remains inaccessible. However, the precise conformational dynamics and the transition pathway between these states are still not fully understood [10].

The first part of this study primarily focuses on the structural characteristics of the Mpro. The second part of the study consists of the salient structural feature analysis of the S protein. Additionally, the second part also includes investigation on the conformational accessibility of the S protein by analysing its interaction with the ACE2 receptor in three distinct states—open, closed, and

intermediate. By evaluating the extent to which the ACE2 receptor can access the S protein in these different states, we aim to gain insights into its role in viral entry.

A. Salient structural features of SARS-CoV-2 Main protease (Mpro)

In the first part of the study, we will be discussing about the salient structural features of the SARS-CoV-2 Mpro.

4.3. Materials and methods:

4.3.1. Data retrieval

The 3-D structure as well as the amino acid sequence of the SARS-CoV-2 Mpro was downloaded from the RCSB-PDB as shown in **Figure 4.1.**

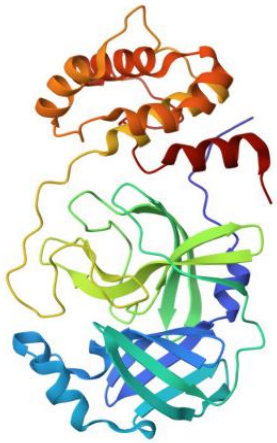

Figure 4.1. *SARS-CoV-2 main protease (2019-nCoV, coronavirus disease 2019, COVID-19)*

4.3.2. Salient structural features analysis

4.3.2.1. Analysis of physicochemical parameter

The ExPASy ProtParam online tool [11] was utilized to analyze the physicochemical properties of the SARS-CoV-2 Mpro. This tool calculates a range of properties based on the given protein sequence, including molecular weight, theoretical isoelectric point (pI), amino acid and atomic composition, extinction coefficient, and estimated half-life. Additionally, it determines the instability index, aliphatic index, and grand average of hydropathicity (GRAVY), providing valuable insights into the protein's structural and functional characteristics.

4.3.2.2. Prediction of secondary structural changes

The secondary structure of SARS-CoV-2 Mpro was analyzed using the GOR IV tool [12]. The Garnier-Osguthorpe-Robson (GOR) method utilizes information theory and Bayesian statistics to predict protein secondary structures. By incorporating multiple sequence alignments, GOR IV enhances the accuracy of structural predictions, helping to distinguish different secondary structure elements more effectively.

4.3.2.3. Identification of conserved residues

Clustal Omega [13] a bioinformatics program, was used to identify the conserved residues by aligning the SARS-CoV-2 Mpro amino acid sequence with the amino acid sequence of SARS-CoV Mpro.

4.3.2.4. Intrinsically disordered protein prediction

Intrinsic disorder regions are segments of a protein that exist as a dynamic ensemble of conformations without forming a stable three-dimensional structure under physiological conditions. To identify these disordered regions in SARS-CoV-2 Mpro, the Predictor of Naturally Disordered Regions (PONDR® VLXT) [14] was used. This tool provides insights into protein flexibility and structural adaptability, which are essential for understanding protein function and interactions.

4.4. Results and Discussions:

4.4.1. Analysis of physicochemical parameter

Table 4.1. *ExPASY ProtParam data of Mpro.*

Formula	No. of amino acids	MW* (Da)	Theoretical pI	Extinction coefficients (M^{-1} cm^{-1})	Estimated half-life (hours)	Instability index	Aliphatic index	GRAVY*
C1499H2318N402O445S22	306	33796.64	5.95	33640	>1.9 hours	27.65	82.12	-0.019

Note: *MW*: Molecular Weight; GRAVY*: Grand average of hydropathicity*

Table 4.1 presents key physicochemical properties of the SARS-CoV-2 Mpro. The extinction coefficient measures the amount of light a protein absorbs at a specific wavelength, providing insights into its structural characteristics.

The estimated in vivo half-life indicates how long it takes for half of the synthesized protein within a cell to degrade. This value follows the "N-end rule," which associates a protein's stability with its N-terminal residue. The stability of the N-terminal residue plays a crucial role in determining the

overall stability of the protein in vivo—proteins with more stable N-terminal residues tend to have a longer half-life.

The instability index is another important parameter used to predict protein stability. A value below 40 suggests that the protein is stable, while a value above 40 indicates potential instability. For Mpro, the instability index is calculated as 27.65, suggesting that the protein is relatively stable.

The aliphatic index, which represents the relative volume occupied by aliphatic side chains (such as alanine, valine, isoleucine, and leucine), is an indicator of protein thermostability. A higher aliphatic index generally correlates with increased thermal stability of globular proteins. The aliphatic index of Mpro, calculated using ExPASy, is 82.12, indicating good thermostability.

Additionally, the Grand Average of Hydropathy (GRAVY) value provides insight into the protein's overall hydrophobicity. It is determined by summing the hydropathy values of all amino acids in the sequence and dividing by the total number of residues. For Mpro, a negative GRAVY value suggests a hydrophilic nature, which contributes to its overall stability in aqueous environments.

4.4.2. Prediction of secondary structural changes

The secondary structure prediction of Mpro, analyzed using GOR IV, reveals that the protein predominantly consists of three types of secondary structures: alpha helices (**h**), extended strands (**e**), and random coils (**c**) with 14.05%, 37.58% and 48.37% respectively as shown in **Table 4.2**.

Table 4.2. *Secondary structure predicted using GOR IV for Mpro.*

Secondary structure	Percentage (%)
Alpha helix (h)	**14.05%**
3_{10} helixes (g)	**0.00%**
Pi helix (i)	**0.00%**
Beta bridge (b)	**0.00%**
Extended strand (e)	**37.58%**
Beta turn (t)	**0.00%**
Bend region (s)	**0.00%**
Random coil (c)	**48.37%**
Ambiguous stress (?)	**0.00%**
Other states	**0.00%**

Additionally, a visual representation of the secondary structure of Mpro, generated using GOR IV, is shown in **Figure 4.2.** In this illustration, different colors are used to distinguish various

types of secondary structures, providing a clearer understanding of the protein's structural composition.

Figure 4.2. *Pictorial representation of the secondary structure predicted using GOR IV for Mpro.*

4.4.3. Identification of conserved residues

The conserved regions of Mpro in both SARS-CoV and SARS-CoV-2 were analyzed using Clustal Omega. The results indicate a sequence identity of 96.08% and a sequence similarity of 98.69% between SARS-CoV Mpro and SARS-CoV-2 Mpro. This high degree of conservation suggests that Mpro is highly conserved in the SARS-CoV and SARS-CoV-2 as shown in **Figure 4.3.**

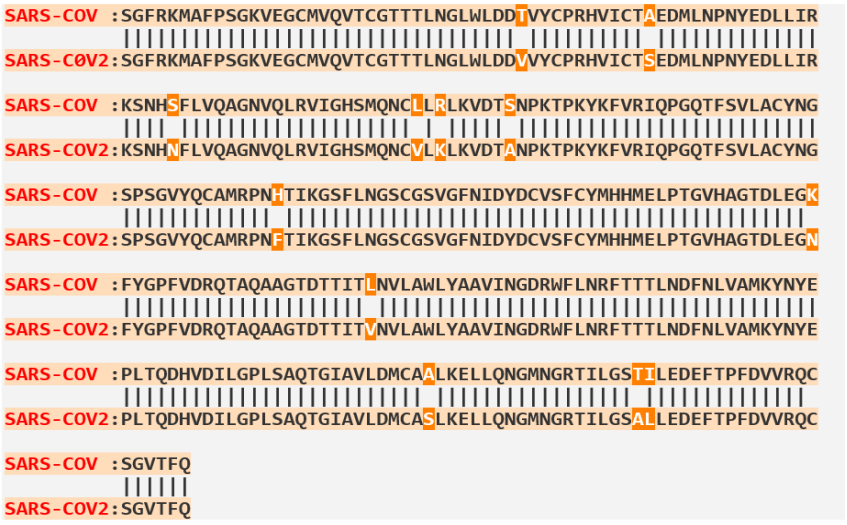

Figure 4.3. *Showing the alignments of the SARS-CoV and SARS-CoV-2 Mpro sequences aligned using Clustal Omega*

4.4.4. Intrinsically disordered protein prediction

The disordered regions of Mpro were predicted using PONDR® VLXT, revealing that the protein is predominantly structured, with only 2.94% of its regions classified as disordered. This indicates a high degree of structural order, as illustrated in **Figure 4.4.**

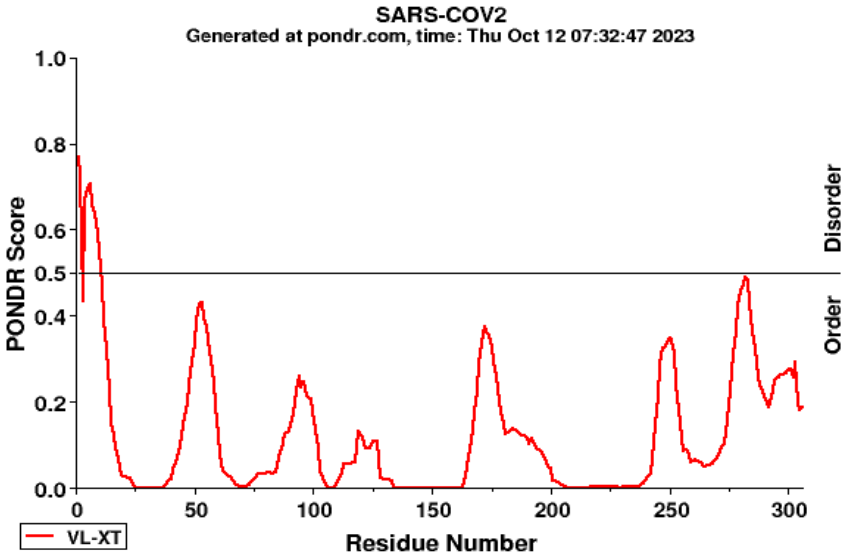

Figure 4.4. *Intrinsically disordered prediction for Mpro using PONDR® VLXT*

4.5. Conclusion:

A preliminary analysis of the main protease (Mpro) was conducted to examine its key structural characteristics. This analysis included the evaluation of physicochemical properties, secondary structure prediction, conserved region identification, and the assessment of intrinsically disordered regions.

The physicochemical properties of Mpro were analysed using the ExPASy ProtParam tool, which provided insights into various parameters such as molecular weight, theoretical isoelectric point (pI), amino acid and atomic composition, extinction coefficient, estimated half-life, instability index, aliphatic index, and the grand average of hydropathicity (GRAVY). These parameters help assess the overall stability of the protein. Secondary structure prediction, performed using GOR IV, revealed that Mpro primarily consists of three types of secondary structures: alpha helices, extended strands, and random coils. To identify conserved regions, the amino acid sequences of SARS-CoV Mpro and SARS-CoV-2 Mpro were aligned. The results showed a sequence identity of 96.08% and a sequence similarity of 98.69%, indicating that Mpro

is highly conserved between the two viruses. Lastly, the prediction of disordered regions was carried out using PONDR® VLXT, which demonstrated that Mpro is predominantly structured, with only 2.94% of its regions classified as disordered.

B. Salient structural features and conformational accessibility of SARS-CoV-2 Spike protein (S-protein).

The second part of the study involves the study of the salient structural features of the SARS-CoV-2 S protein. Additionally, it explores the conformational accessibility of the S protein, which varies depending on its state—either receptor-accessible or inaccessible. To gain deeper insights, we analyzed the interaction between the SARS-CoV-2 S protein and the ACE2 receptor by evaluating three distinct conformational states of the Spike protein: open, closed, and intermediate. This analysis allowed us to assess the extent to which the ACE2 receptor is accessible or inaccessible in each state.

4.6. Materials and methods

4.6.1. Data retrieval

The 3D structure and amino acid sequence of the SARS-CoV-2 Spike (S) protein were obtained from the RCSB Protein Data Bank (RCSB-PDB), as illustrated in **Figure 4.5**. For the conformational accessibility analysis, the 3D structures of the S protein in its open, closed, and intermediate (transition) states were also retrieved from the RCSB-PDB, as depicted in **Figure 4.6.** Additionally, the ACE2 receptor was extracted from the SARS-CoV-2 Spike receptor-binding domain bound to the ACE2 receptor complex (PDB ID: 6LZG) using the UCSF Chimera package alpha v.1.12 [15].

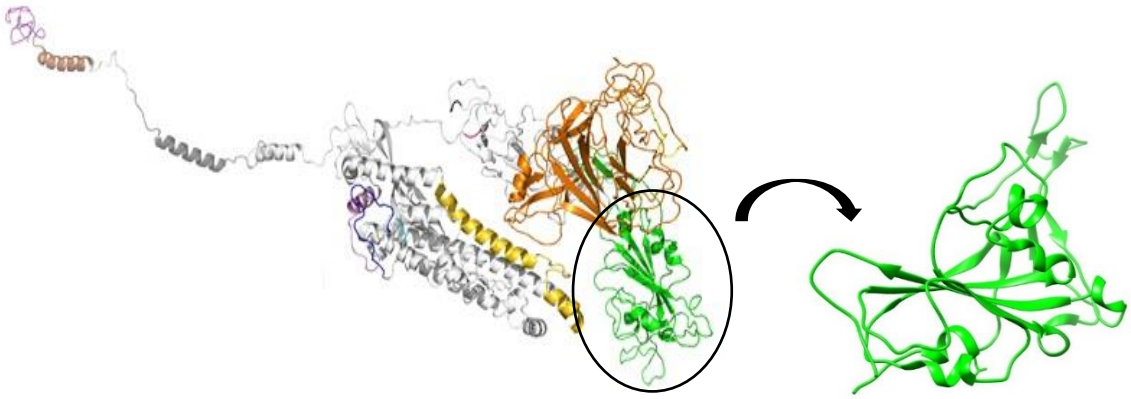

Figure 4.5. *3D structure of the entire SARS-CoV-2 Spike protein along with the Receptor binding domain (RBD) highlighted in green.*

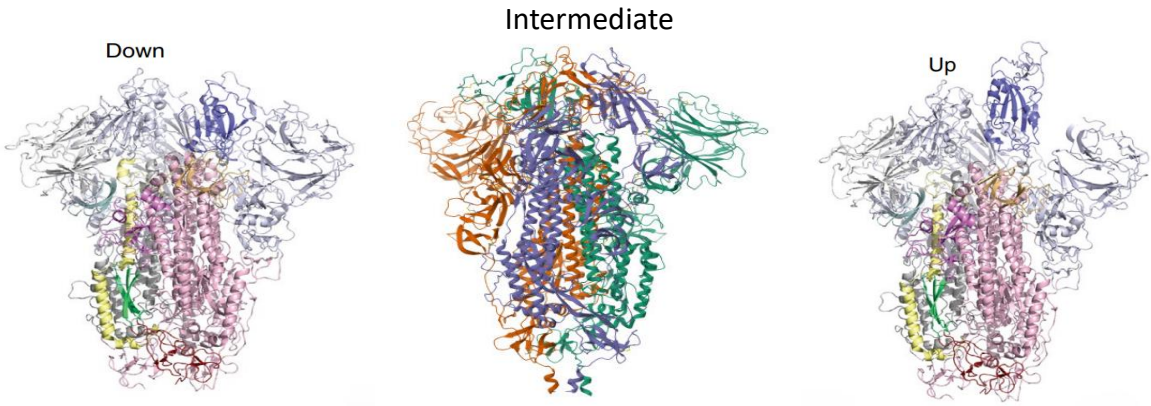

Figure 4.6. *Cartoon representations of the down (left), Intermediate (middle) and up (right) state SARS-CoV-2 structures (PDB 6VXX, 7W94 and 6VYB, respectively)*

4.6.2. Salient structural features analysis

4.6.2.1. Analysis of physicochemical parameter

The physico-chemical properties of the SARS-CoV-2 S protein and its Receptor Binding Domain (RBD) were analyzed using the ExPASy ProtParam online tool [11]. This tool calculates various physico-chemical characteristics derived from a given protein sequence. The parameters assessed include molecular weight, theoretical isoelectric point (pI), amino acid composition, atomic composition, extinction coefficient, estimated half-life, instability index, aliphatic index, and the grand average of hydropathicity (GRAVY).

4.6.2.2. Prediction of secondary structural changes

The secondary structure of the SARS-CoV-2 S protein and its Receptor Binding Domain (RBD) was analyzed using the GOR IV tool [12]. The Garnier-Osguthorpe-Robson (GOR) method utilizes information theory and Bayesian statistics to predict protein secondary structures. This tool integrates multiple sequence alignments to enhance the accuracy of secondary structure differentiation and prediction.

4.6.2.3. Identification of conserved residues

The identification of conserved residues was performed using Clustal Omega [13], a bioinformatics tool designed for multiple sequence alignment. To analyze conserved regions, the amino acid sequence of the RBD of SARS-CoV-2 S protein was aligned with that of the SARS-CoV S protein. Additionally, the wild-type (WT) RBD sequence of the S protein was aligned with

the RBD sequences of various mutants that had a significant impact on human health during the pandemic.

4.6.2.4. Intrinsically disordered protein prediction

Intrinsic disorder regions are segments of a protein that exist as a dynamic ensemble of conformations without forming a stable three-dimensional structure under physiological conditions. The Predictor of Naturally Disordered Regions (PONDR® VLXT) [14] was utilized to identify these disordered regions in both the SARS-CoV-2 S protein as well as the RBD exclusively.

4.6.2.5. Docking

To analyze conformational accessibility, the ACE2 receptor was individually docked to the three distinct states of the SARS-CoV-2 S protein using the ClusPro docking server [16]. Subsequently, three docked complexes were obtained, followed by an interaction analysis and binding energy calculation.

4.6.2.6. Interaction study

The Spike-ACE2 interaction analysis was conducted using the PDBsum server [17], a web-based platform that provides structural insights into Protein Data Bank (PDB) entries. PDBsum primarily offers image-based analyses, including protein secondary structure visualization, protein-ligand and protein-DNA interaction studies, PROCHECK assessments of structural quality, and various other structural evaluations.

4.6.2.7. Binding free energy calculation

The binding affinity between the S protein in its open, intermediate, and closed states and the ACE2 receptor was assessed using the PRODIGY web server [18]. PRODIGY (PROtein binDIng enerGY prediction) is a suite of web-based tools designed for predicting binding affinity in biological complexes and distinguishing biological interfaces from crystallographic ones.

4.7. Results and Discussions:

4.7.1. Analysis of physicochemical parameter

We have examined the physio-chemical properties of SARS-CoV-2 S protein and exclusively for the RBD of the Spike protein using the ExPASy ProtParam tool as shown in **Table 4.3**.

Table 4.3. *ExPASY ProtParam data of entire spike and RBD.*

	No. of amino acids	MW* (Da)	Theoretical pI	Extinction coefficients (M^{-1} cm^{-1})	Estimated half-life (hours)	Instability index	Aliphatic index	GRAVY*
Spike protein	1273	141178.4	6.24	148960	30 hours	33.01	84.67	-0.079
RBD	195	23493.4	8.39	33850	1 hours	24.34	71.77	-0.22

Note: *MW*: Molecular Weight; GRAVY*: Grand average of hydropathicity*

4.7.2. Prediction of secondary structural changes in spike protein

The secondary structure prediction analyzed using GOR IV for the Spike as well as exclusively for the RBD shows that the secondary structure is contributed maximum by the Random coil (c) followed by the Extended strand (e) and Alpha helix (h) in both the cases as shown in **Table 4.4**.

Table 4.4. *Secondary structure predicted using GOR IV for entire spike and RBD.*

Secondary structure	Spike protein (%)	RBD (%)
Alpha helix (h)	21.52%	7.18%
3_{10} helixes (g)	0.00%	0.00%
Pi helix (i)	0.00%	0.00%
Beta bridge (b)	0.00%	0.00%
Extended strand (e)	22.07%	25.36%
Beta turn (t)	0.00%	0.00%
Bend region (s)	0.00%	0.00%
Random coil (c)	56.40%	67.46%
Ambiguous stress	0.00%	0.00%
Other states	0.00%	0.00%

Moreover, a pictorial representation of the secondary structure exhibited by the S protein as well as RBD is obtained from the GOR IV as shown in **Figure 4.7 & 4.8 respectively**. Here different colors depict different form of secondary structures.

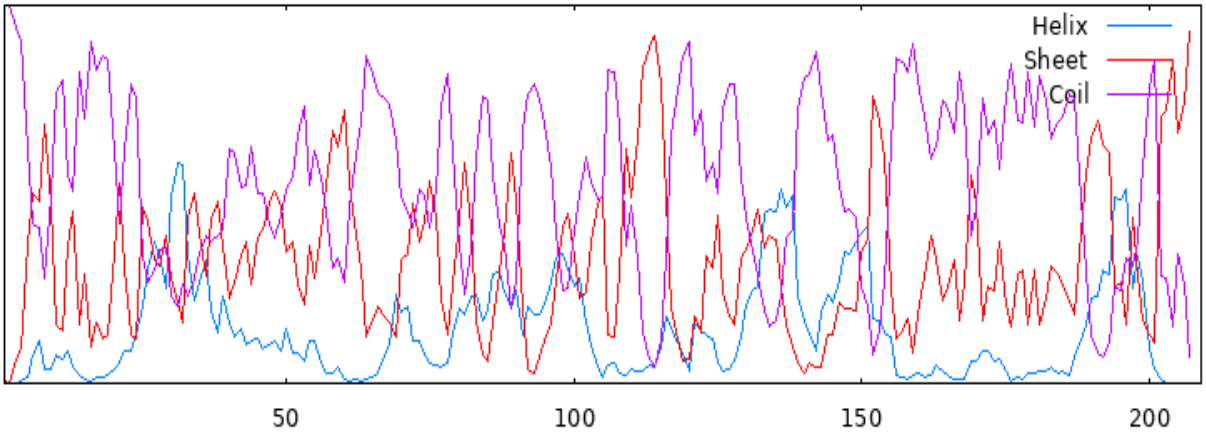

Figure 4.7. *Pictorial representation of the secondary structure predicted using GOR IV for the RBD of the spike*

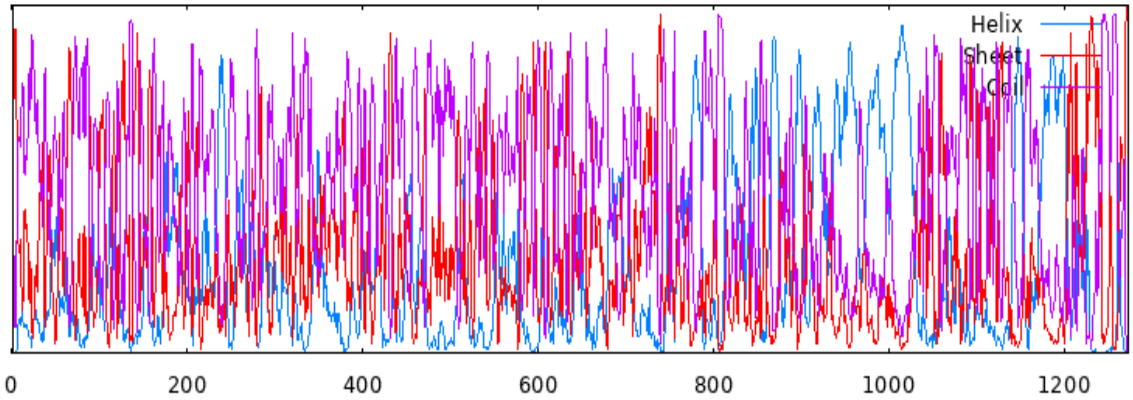

Figure 4.8. *Pictorial representation of the secondary structure predicted using GOR IV for the entire spike.*

4.7.3. Identification of conserved residues

The conserved regions of the Receptor Binding Domain (RBD) of the Spike (S) protein in both SARS-CoV and SARS-CoV-2 were analyzed using Clustal Omega [13]. The results indicate a sequence identity of 72.09% and a sequence similarity of 75.22% between the RBDs of the S proteins from SARS-CoV and SARS-CoV-2. These findings suggest that the RBD of the S protein is relatively poorly conserved between the two viruses, as illustrated in **Figure 4.9.**

```
sp|RBD|SARS-COV        TNLCPFGEVFNATKFPSVYAWERKKISNCVADYSVLYNSTFFSTFKCYGVSATKLNDLCF
sp|RBD|SARS-COV-2      TNLCPFGEVFNATRFASVYAWNRKRISNCVADYSVLYNSASFSTFKCYGVSPTKLNDLCF
                       **************:* *****:**:***************: ********** ********

sp|RBD|SARS-COV        SNVYADSFVVKGDDVRQIAPGQTGVIADYNYKLPDDFMGCVLAWNTRNIDATSTGNYNYK
sp|RBD|SARS-COV-2      TNVYADSFVIRGDEVRQIAPGQTGKIADYNYKLPDDFTGCVIAWNSNNLDSKVGGNYNYL
                       :********::**:********** ************ ***:***:.*:*:.  *****

sp|RBD|SARS-COV        YRYLRHGKLRPFERDISNVPFSPDGKPCTP-PALNCYWPLNDYGFYTTTGIGYQPYRVVV
sp|RBD|SARS-COV-2      YRLFRKSNLKPFERDISTEIYQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVV
                       ** :*:.:*:*******.  :. ...**.   .:***:**:.***  *.*:**********

sp|RBD|SARS-COV        LSFELLNAPATVCGP
sp|RBD|SARS-COV-2      LSFELLHAPATVCGP
                       ******:********
```


Figure 4.9. *Showing the alignments of the SARS-CoV and SARS-CoV-2 RBD sequences aligned using Clustal Omega*

Additionally, we aligned the wild-type (WT) RBD sequence of the Spike protein with the RBD sequences of various mutants that had a significant impact on public health during the pandemic. The analysis included the Double Mutant (DM), Delta, Delta-Plus (DP), and several Omicron variants (BA.1, BA.2, BA.4, BA.2.12.1, BA.2.75, BA.2.75.2), which are further examined in detail in the subsequent chapters. The results revealed a progressive decline in sequence identity as new variants emerged, with a notable reduction observed in the Omicron variants compared to the WT, as illustrated in **Figure 4.10.**

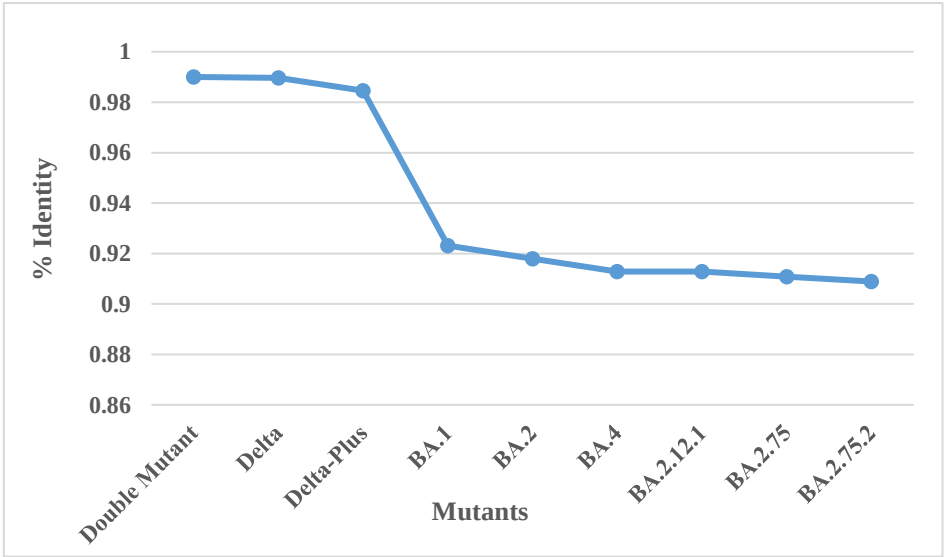

Figure 4.10. *Showing the % identity of several variants of SARS-CoV-2 compared to the wild type.*

4.7.4. Intrinsically disordered protein prediction

The disordered regions of the S protein and its Receptor Binding Domain (RBD) were predicted using PONDR® VLXT. The results indicate that both the RBD and the S protein are highly ordered, with an overall percentage of disordered regions calculated as 6.70% and 7.70%, respectively, as illustrated in **Figures 4.11 and 4.12.**

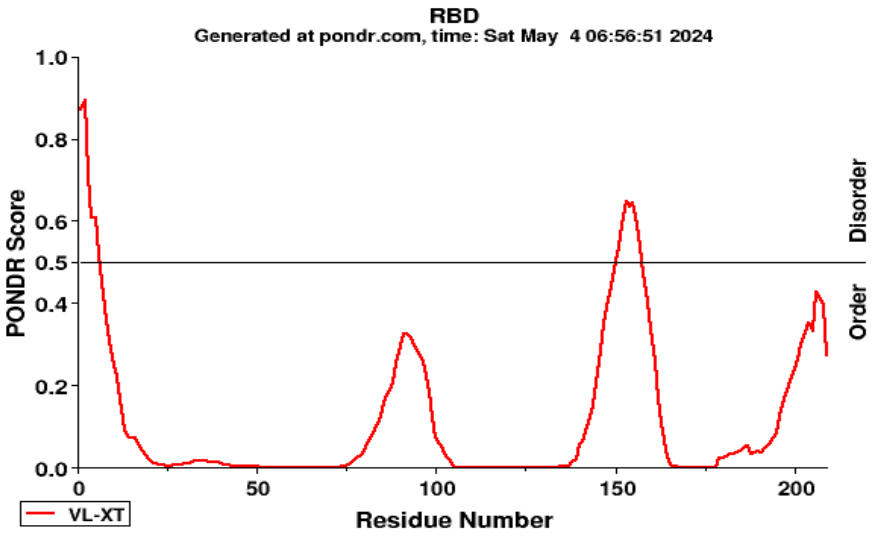

Figure 4.11. *Intrinsically disordered prediction for the RBD of spike using PONDR® VLXT*

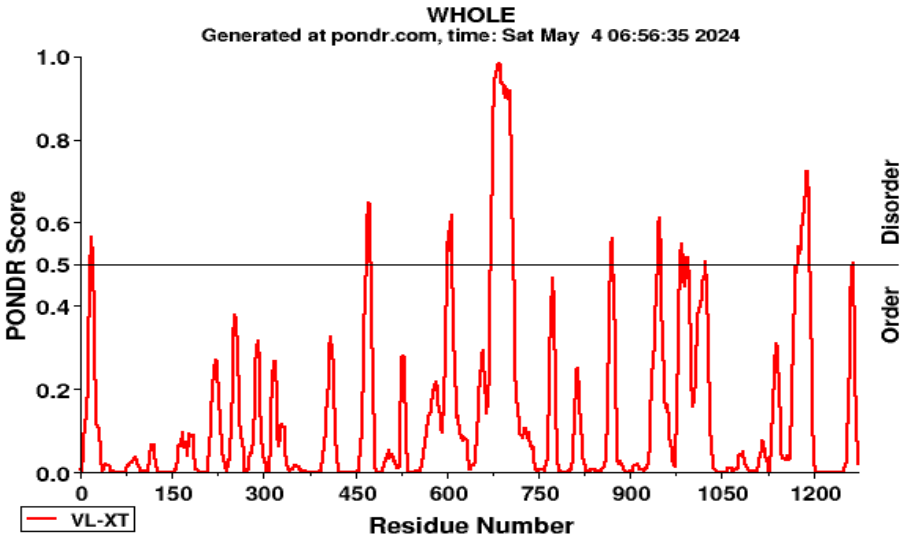

Figure 4.12. *Intrinsically disordered prediction for the entire spike protein using PONDR® VLXT*

Conformational accessibility of SARS-CoV-2 S-protein:

4.7.5. Determination of the interface interactions of the S protein-ACE2 complexes

A region where two sets of proteins come into contact is commonly referred to as an interface area. Surface residues with significant surface areas that are exposed to the solvent are frequently characterized. To analyze the interactions between the ACE2 receptor and the different conformational states of the S protein, interface statistics for all three complexes were obtained by submitting the docked structures to the PDBsum server. The interface statistics for these complexes are presented in **Table 4.5.**

Table 4.5. *Interface statistics for the S protein (Intermediate)-ACE2, S protein (Closed)-ACE2 and S protein (Open)-ACE2 complexes*

Complex system	**No. of salt bridges**	**No. of Disulphide bonds**	**No. of hydrogen bonds**	**No. of non-bonded contacts**
Intermediate spike-ACE2	2	9	12	122
Closed Spike-ACE2	-	-	1	39
Open Spike-ACE2	1	16	15	167

Table 4.5 shows molecular interactions like salt bridges, disulphide bonding, hydrogen bonding, and non-bonded contacts stabilized the wild-type and mutant type complexes. From the interface interaction statistics we found that the number of interface residues in the S protein(intermediate)-ACE2 complex was contributed by two, nine, twelve and one hundred and twenty-two salt bridges, disulphide bonds, hydrogen bonding, and non-bonded contacts respectively. In the S protein(closed)-ACE2 complex we can see the presence of one hydrogen bond and thirty-nine non-bonded contacts at the interface.

There are no salt bridges and disulphide contacts in the closed complex indicating no strong interactions between the closed S protein and ACE2 indicating that the closed state of S protein is inaccessible to the ACE2. However, at the interface of S protein and ACE2 in the S protein(open)-ACE2 complex, we observed one hundred and sixty-seven non-bonded contacts, one salt bridge, thirteen disulphide contacts and fifteen hydrogen bonds. An increase in the number of interface interaction between the S protein and ACE2 in the open state suggest that the S protein is the most

accessible to the ACE2 in the open state followed by the intermediate (transition) state and the least accessible in the closed state.

The pictorial representation of the summarized intermolecular interactions between S protein and ACE2 of all the complexed structure at the three states are shown in **Figure 4.13.**

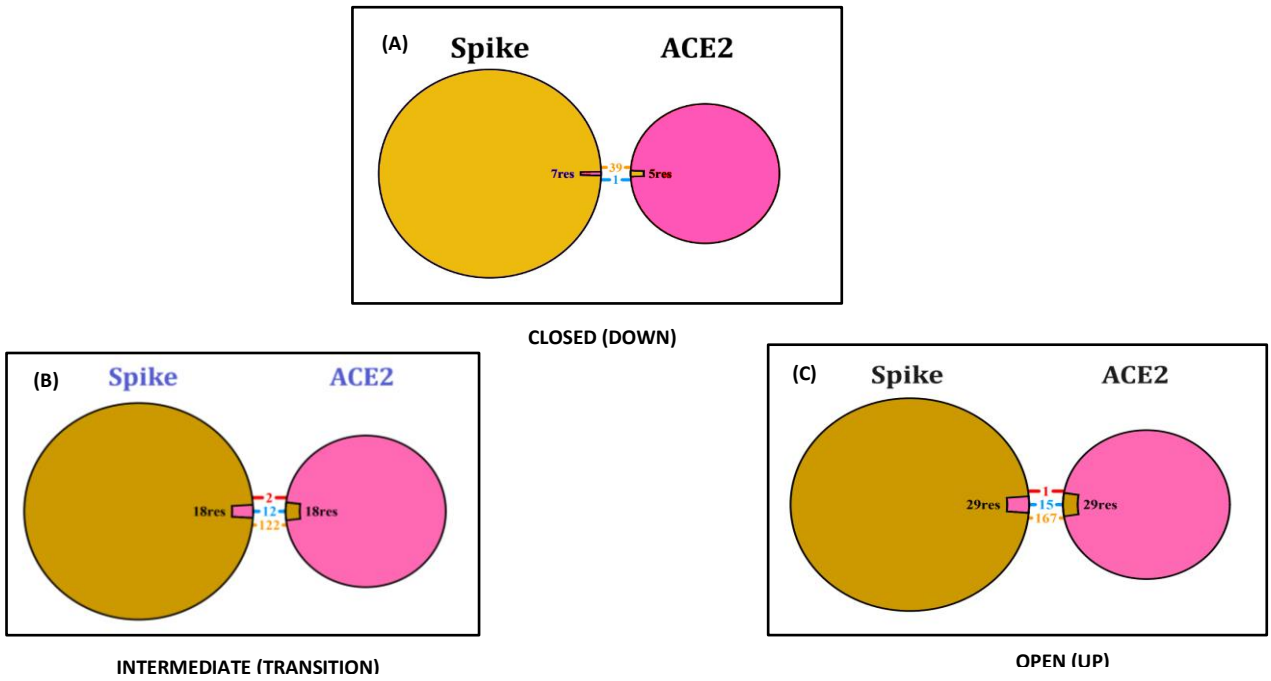

Figure 4.13. *The pictorial representation of the summarized intermolecular interactions between S protein and ACE2 in* ***(A)*** *S protein (closed)-ACE2 complex* ***(B)*** *S protein (Intermediate)-ACE2 complex and* ***(C)*** *S protein (Open)-ACE2 complex*

4.7.6. Binding free energy (BFE) calculation

We analyzed the binding affinity between the Spike (open, intermediate and closed) and ACE2 receptor using the PRODIGY web server. **Table 4.6** shows the ΔG values for the three complexes.

Table 4.6. *Binding free energy (ΔG) values between the S protein and ACE2 in the closed, intermediate and open state.*

Complex	BFE (kcal/mol)
S protein (closed)-ACE2 complex	-9.2
S protein (Intermediate)-ACE2 complex	-13.2
S protein (open)-ACE2 complex	-18.5

The binding affinity analysis revealed that the S protein in its open state exhibits the highest binding affinity with the ACE2 receptor compared to its intermediate and closed states. This indicates a strong interaction between the S protein and ACE2 in the open conformation. Based

on these findings, it can be inferred that the S protein is most accessible to the ACE2 receptor in its open form, where the RBD is positioned in the "up" conformation, facilitating binding. Conversely, the S protein in its closed state is the least accessible to the ACE2 receptor.

4.8. Conclusion:

Salient structural features followed by the conformational accessibility analysis of S protein was analyzed. First of all, we analyzed the physio-chemical properties, secondary structure prediction, conserved region prediction followed by the prediction of intrinsically disordered regions in S protein and the RBD exclusively. Physio-chemical parameters like molecular weight, theoretical pI (isoelectric point), amino acid composition, atomic composition, extinction coefficient, estimated half-life, instability index, aliphatic index and grand average of hydropathicity (GRAVY) was analyzed using ExPASy ProtParam tool. The parameters obtained from the ExPASy ProtParam provides us an idea about the overall stability of the protein. The secondary structure prediction analyzed using GOR IV provide us information on the secondary structure exhibited by the S protein and exclusively for the RBD and results showed that the secondary structure is contributed maximum by the Random coil followed by the Extended strand and Alpha helix in both the cases.

The conserved regions of the RBD of S protein for both the SARS-CoV and SARS CoV-2 was analyzed using Clustal omega and results depicts that there is a sequence identity of 72.09% and sequence similarity of 75.22 %. We also aligned the wild type (WT) sequence of RBD of spike with the RBD sequence of various mutants namely Double mutant, Delta, Delta-plus and few omicron variants that had a devastating impact on the human health during the pandemic and results showed that the percentage identity kept on decreasing as the new variants were emerging and a significant decline in the percentage identity can be observed in the omicron variants compared to the WT.

Lastly, we analyzed the conformational accessibility of SARS-CoV-2 S-protein with the ACE2 receptor using three different states (open, closed and Intermediate) of the Spike and on performing the interaction study and binding free energy calculation it depicted that the S protein is the most accessible to the ACE2 receptor in its open form with highest number of interface interaction and binding affinity between the followed by the transition(intermediate) state and the S protein in its closed state is the least accessible to the ACE2 receptor.

4.9. Bibliography:

[1]. Hu, Q., Xiong, Y., Zhu, G. H., Zhang, Y. N., Zhang, Y. W., Huang, P., and Ge, G. B. The SARS-CoV-2 main protease (Mpro): Structure, function, and emerging therapies for COVID-19. *Medical Communications (2020)*, 3(3): e151, 2022. https://doi.org/10.1002/mco2.151

[2]. Zhang, L., Lin, D., Sun, X., Curth, U., Drosten, C., Sauerhering, L., Becker, S., Rox, K., and Hilgenfeld, R. Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved α-ketoamide inhibitors. *Science*, 368(6489): 409–412, 2020. https://doi.org/10.1126/science.abb3405

[3]. Jin, Z., Du, X., Xu, Y., Deng, Y., Liu, M., Zhao, Y., Zhang, B., Li, X., Zhang, L., Peng, C., Duan, Y., Yu, J., Wang, L., Yang, K., Liu, F., Jiang, R., Yang, X., You, T., Liu, X., Yang, X., Bai, F., Liu, H., Liu, X., Guddat, L. W., Xu, W., Xiao, G., Qin, C., Shi, Z., Jiang, H., Rao, Z., and Yang, H. Structure of Mpro from SARS-CoV-2 and discovery of its inhibitors. *Nature*, 582(7811): 289–293, 2020. https://doi.org/10.1038/s41586-020-2223-y

[4]. Grottesi, A., Beškǒr, N., Emerson, A., Manelfi, C., Beccari, A. R., Frigerio, F., Lindahl, E., Cerchia, C., and Talarico, C. Computational studies of SARS-CoV-2 3CLpro: Insights from MD simulations. *International Journal of Molecular Sciences*, 21(15): 5346, 2020. https://doi.org/10.3390/ijms21155346

[5]. Samrat, S. K., Xu, J., Xie, X., Gianti, E., Chen, H., Zou, J., Pattis, J. G., Elokely, K., Lee, H., Li, Z., Klein, M. L., Shi, P. Y., Zhou, J., and Li, H. Allosteric inhibitors of the main protease of SARS-CoV-2. *Antiviral Research*, 205: 105381, 2022. https://doi.org/10.1016/j.antiviral.2022.105381

[6]. Rungruangmaitree, R., Phoochaijaroen, S., Chimprasit, A., Saparpakorn, P., Pootanakit, K., and Tanramluk, D. Structural analysis of the coronavirus main protease for the design of pan-variant inhibitors. *Scientific Reports*, 13(1): 7055, 2023. https://doi.org/10.1038/s41598-023-34305-6

[7]. Letko, M., Marzi, A., and Munster, V. Functional assessment of cell entry and receptor usage for SARS-CoV-2 and other lineage B betacoronaviruses. *Nature Microbiology*, 5(4): 562–9, 2020. https://doi.org/10.1038/s41564-020-0688-y

[8]. Fehr, A. R., and Perlman, S. Coronaviruses: an overview of their replication and pathogenesis. *Methods in Molecular Biology*, 1282: 1–23, 2015. https://doi.org/10.1007/978-1-4939-2438-7_1

[9]. Du, L., He, Y., Zhou, Y., Liu, S., Zheng, B. J., and Jiang, S. The spike protein of SARS-CoV--a target for vaccine and therapeutic development. *Nature Reviews Microbiology*, 7(3): 226–236, 2009. https://doi.org/10.1038/nrmicro2090

[10]. Gur, M., Taka, E., Yilmaz, S. Z., Kilinc, C., Aktas, U., and Golcuk, M. Conformational transition of SARS-CoV-2 spike glycoprotein between its closed and open states. *The Journal of Chemical Physics*, 153(7): 075101, 2020. https://doi.org/10.1063/5.0011141

[11]. Gasteiger, E., Gattiker, A., Hoogland, C., Ivanyi, I., Appel, R. D., and Bairoch, A. ExPASy: the proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Research*, 31(13): 3784-3788, 2003. https://doi.org/10.1093/nar/gkg563

[12]. Sen, T. Z., Jernigan, R. L., Garnier, J., and Kloczkowski, A. Gor IV server for protein secondary structure prediction. *Bioinformatics*, 21(11): 2787–2788, 2005. https://doi.org/10.1093/bioinformatics/bti408

[13]. Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., Thompson, J. D., and Higgins, D. G. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology, 7*(1): 539, 2011. https://doi.org/10.1038/msb.2011.75

[14]. Xue, B., Dunbrack, R. L., Williams, R. W., Dunker, A. K., and Uversky, V. N. Pondr-fit: A meta-predictor of intrinsically disordered amino acids, *Proteins and Proteomics*, 1804(4): 996–1010, 2010. https://doi.org/10.1016/j.bbapap.2010.01.011

[15]. Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., and Ferrin, T. E. UCSF Chimera—A visualization system for exploratory research and analysis. *Journal of Computational Chemistry*, 25(13): 1605-1612, 2004. https://doi.org/10.1002/jcc.20084

[16]. Kozakov, D., Hall, D. R., Xia, B., Porter, K. A., Padhorny, D., Yueh, C., Beglov, D., and Vajda, S. The ClusPro web server for protein-protein docking. *Nature Protocols*, 12(2): 255-278, 2017. https://doi.org/10.1038/nprot.2016.169

[17]. Laskowski, R. A., Hutchinson, E. G., Michie, A. D., Wallace, A. C., Jones, M. L., and Thornton, J. M. PDBsum: A Web-based database of summaries and analyses of all PDB structures. *Trends in Biochemical Sciences*, 22(12): 488–490, 1997. https://doi.org/10.1016/s0968-0004(97)01140-7

[18]. Xue, L. C., Rodrigues, J. P., Kastritis, P. L., Bonvin, A. M., and Vangone, A. Prodigy: A web server for predicting the binding affinity of protein–protein complexes. *Bioinformatics*, 32(23): 3676–3678, 2016. https://doi.org/10.1093/bioinformatics/btw514