Omicron variant, sub-variants and Delta variant of SARS-CoV-2:

A comparative study of N-terminal domain (NTD) and Receptor-Binding Domain (RBD) of Spike protein.

ABSTRACT

The severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) Omicron variant has now spread over the world. In this study, computational techniques were employed to evaluate the spike infectivity, transmission, and pathogenicity of the Omicron BA.1.1 variant. It was found that the Omicron BA.1.1 variant has 40 mutations when compared with the original Wuhan-Hu-1 virus. The mutations that had occurs to the sequence was identified which consequently causes changes in the properties of the virus. We discovered that the Omicron BA.1.1 variant exhibited a larger positive electrostatic surface potential due to the significant effect of mutations in the receptor-binding domain (RBD). This could boost RBD's interaction with the electronegative human angiotensin-converting enzyme 2 (hACE2). When compared to the wild type, the Omicron BA.1.1 variant demonstrated a stronger affinity for hACE2 and the potential for increased transmission. The stability of the Omicron BA.1.1 was also found to be more stable than the wildtype as it contains lesser intrinsically disordered segments and regions. Molecular docking was conducted between the wildtype and Omicron variant that identified the difference in the binding energy of the both viruses, ultimately showing that the BA.1.1 variant has increased binding affinity towards the human ACE2 protein.

Keywords: Omicron, COVID-19, SARS-CoV-2, BA.1.1, Wildtype

1 INTRODUCTION

The world has been experiencing a pandemic caused by the novel coronavirus disease 2019 since late 2019. (COVID-19). COVID-19 is caused by the severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2) infection, which has had a devastating impact on the mortality and morbidity of its victims as well as the entire human race. By early 2020, the virus outbreak, which began in late 2019 in Wuhan, China, had spread fast around the world, prompting the designation of a worldwide pandemic as a public health emergency of international significance [1]. The virus's sickness was dreaded by everyone because of its unknown genetic composition and properties. The lack of comprehensive understanding of the virus and its interaction with the hosts had caused many lives to be lost at the beginning of the pandemic as this information is crucial and fundamental information that is required to fight against the SARS-CoV-2 virus [1]. As of 13th May 2022, it has been confirmed by the World Health Organization (WHO) that there are a total of 517,648,631 COVID-19 cases worldwide and a total of 6,261,708 causalities of this virus.

The SARS-CoV-2 virus in transmitted in the form of fomites and droplets in the events of close contacts between infected and uninfected people. The virus is able to infect asymptomatic individuals who are within 6 feet of an infected individual. Individuals with COVID-19 have a wide variety of clinical symptoms, ranging from mild to severe, with rapid progression and acute disease. COVID-19 has no defined symptoms and can express itself in a variety of ways, ranging from no symptoms (asymptomatic) to severe pneumonia and death. The manner of development of the SARS-CoV-2 virus is dependent on the viral spike protein binding to the receptors, the angiotensin-converting enzyme-2 (ACE2) [2].

SARS-CoV-2 is a single-stranded, positive-sense RNA (+ssRNA) virus in the Coronaviridae family that belongs to lineage B of the genus Beta-coronavirus. Over a year ago, the genome size of the SARS-CoV-2 virus was sequences. Based on this data it was deduced and found that ORF1a and ORF1b are two major open reading frames (ORFs), which covers two-thirds of the genome and are translated into pp1a and pp1b proteins, are included in the SARS-CoV-2 genomic RNA. The remaining one-third of the genome is made up of overlapping ORFs that codes for four primary structural proteins: S (spike glycoprotein), N (nucleocapsid protein), M (membrane protein), and E (envelope protein), as well as other auxiliary proteins [1].

Overall the duration of 3 years, the SARS-CoV-2 virus has undergone multiple mutations and evolution in its spike protein. The genetic lineages of the virus are continuously monitored to identify presence of mutations in the virus. Beginning from it first variant, the Alpha variant to the current variant, the Omicron variant. As of now, the Omicron variant (B.1.1.529) is considered as the variant of concern (VOC) for the SARS-CoV-2 virus. This implies that the Omicron variant has mutations in the spike protein receptor-binding domain (RBD) that significantly increase its binding affinity while also inducing rapid spread in human populations [2]. Compared to the other variants, the Omicron variant has more mutations, in which most of them occurred in the spike protein's receptor-binding domain.

The Omicron variant and its sub-variants can be categorized into several sub-lineages – BA.1, B.A.1.1, BA.2, BA.2.1 BA.3, BA.4 and BA.5 [2]. The lineage shows the collection of closely related variations that share a common ancestor and its branches in Omicron. The properties of each variant lineages differ in terms of transmission and virulence. The major contributing factor for the difference in properties is the mutations found in the spike protein of the virus which is responsible for host cell penetration and viral replication.

In order to understand the comprehensive function of the SARS-CoV-2 virus, the original virus, the Delta variant, the Omicron variant (BA.1) and its sub lineages are BA.1.1, BA.2 and etc, are to be studied. Comparison of the spike proteins structures of the virus and its variants would bring about answers to the properties of the virus and its ability to penetrate host cells and replicate. The BA.1 variant has 39 mutations, BA.1.1 has 40 mutations, BA.2 has 31 mutations, BA.3 has 34 mutations, BA.4 has 30 mutations and BA.5 has 32 mutations that were determined by computational tools that studied the spike infectivity, transmission and pathogenicity of the virus.

The purpose of this study is to examine and contrast the N-terminal domain (NTD) and receptor-binding domain (RBD) of spike protein between the original SARS-Cov-2 virus and its variants – Delta, Omicron BA.1, BA.1.1, BA.2, BA.2.12, BA.2.12.1, BA.3, BA.4 and BA.5. Computational methodologies and tools are used to investigate the sequencing and structural properties of the NTD and RBD, which are critical for viral transmission.

2 METHODOLOGY

2.1 SARS-CoV-2 Virus, Delta, Omicron and Sub-Variants Protein Sequence Retrieval

The protein sequences for all the viruses were obtained from Expasy, ViralZone (https://viralzone.expasy.org/9556). The FASTA sequence for the spike protein for all the virus and variants were obtained from NCBI. The sequences were downloaded from the NCBI database.

2.2 Multiple alignment of Delta and Omicron Variants with Wild Type

The Clustal OMEGA alignment tool with the default settings was utilised to conduct the alignments between the wild type and its variants. The protein sequence of Wuhan-Hu-1 (Wild type) was aligned with the protein sequences of delta, omicron variant and sub-lineages BA.1, BA.1.1, BA.2, BA.2.12, BA.2.12.1, BA.3, BA.4 and BA.5. Based on the multiple alignment, mutations were identified.

2.3 Physiochemical Characterization

Using the Expasy protparam tool, the protein sequences of BA.1, BA.1.1, BA.2, BA.2.12, BA.2.12.1, BA.3, BA.4 and BA.5's whole spike protein, RBD and NTD were compared to Wuhan-Hu-1 (Wild type). The number of amino acids, molecular weight, theoretical pI, amino acid composition, charged residues, instability index, aliphatic index, and Grand average of hydrophathicity (GRAVY) were all analysed.

2.4 Secondary structure predictions

GOR IV secondary structure prediction software was used to determine the secondary structure of the Wuhan-Hu-1, Delta variant, Omicron variant (BA.1) and sub-variants (BA.1.1, BA.2, BA.2.12, BA.2.12.1, BA.3, BA.4 and BA.5). By utilizing information theory and Bayesian statistics, the Garnier–Osguthorpe–Robson (GOR) programme analyses secondary protein structure. The goal of combining several sequence alignments with GOR is to gather information for improved secondary structure identification.

2.5 Natural disordered regions predictions

The intrinsically disordered regions of the spike protein, RBD and the NTD protein sequences between the Wuhan-Hu-1 (wild type) and its variants (Delta, Omicron BA.1, BA.1.1, BA.2, BA.2.12, BA.2.12.1, BA.3, BA.4 and BA.5) was predicted. The PONDR®VLXT tool was utilised to conduct the prediction.

2.6 Structure prediction using AlphaFold

Instead of the classical homology modelling technique, the AlphaFold2 with the MMseqs2 suite was utilised for computing the sequence alignments of the SARS-CoV-19 study. The structures of the RBD and the NTD sequences of the Covid-19 variants (wild type, Delta, Omicron BA.1, BA.1.1, BA.2, BA.2.12, BA.2.12.1, BA.3, BA.4 and BA.5) were predicted using the AlphaFold software via ColabFold, the first formal release of the tool. It is able to predict the structures of monomeric proteins as well as their homo- and hetero-complexes, refine them using molecular dynamics simulations, and score them using model quality criteria.

2.7 Structure analysis via PyMOL

The Protein Data Bank was used to obtain the crystal structure of the SARS-CoV-2 spike receptor-binding domain coupled to ACE2 of wuhan-HU-1. The electrostatic potential of the Receptor-binding domain and the N-Terminal Domain of Omicron and sub-variants was

calculated using the Adaptive Poisson–Boltzmann Solver (APBS) tool implemented in PyMOL.

2.8 Protein-Protein Docking and stability analysis of RBD-hACE2

Using the Hawkdock docking tool, protein-protein docking between hACE2 and Omicron variants RBD was compared to wild type. To aid in the analysis of binding structures, MM/GBSA is used to forecast the binding free energy and decomforse the free energy contributions to the binding free energy of a protein–protein complex in per-residue.

3 RESULTS AND DISCUSSION

3.1 Multiple Sequence Alignment

Sequence Alignment of the spike protein of the SARS-CoV-2 (Wuhan-Hu) virus and the Omicron variant (BA.1.1) was conducted in order to identify the mutations between the two sequences. Identifying the mutations between the original Wuhan-Hu virus spike protein and the BA.1.1 Omicron variant aids in determining the difference in the protein properties and their functions. Based on the sequence alignment software, it was determined that the identity between the two sequences are 96.9% and the similarity is 97.4%. the high percentages show us that the Omicron variant (BA.1.1) and the wildtype have a resemblance. With the use of the BLOSUM62 substitution matrix, the gaps in the sequences were found as 0.7%. Table S1 shows the mutations that were found between the wildtype spike protein sequence and the BA.1.1 variant. There are a total of 40 mutations that were present between the BA.1.1 Omicron variant and the Wuha-Hu-1 (wildtype). There were also presence of insertion and deletion mutation at the respective positions, 69-70 (deletion) and 214 EPE (insertion). The unique deletion mutation of 69-70del that is found in the BA.1.1 is the H69/V70 deletion that compensates for immune escape mutations that reduce infectivity, according to previous studies; consequently, it is vital to keep an eye out for deletions that have functional effects [3]. Another significant mutation is the N501Y mutation which enhances the receptor binding strength. It was found that the stabilising effect of mutations on receptor binding and their population occurrence have a generally positive relationship [2].

3.2 Physio-chemical characterization

The various physical and chemical parameters of the Wuhan-HU (wildtype) and the Omicron BA.1.1 variant was computed using the ProtParam computational tool. Based on the data obtained in Table S4.1, it was observed that the Wuhan-Hu-1 whole spike protein is inclusive of 1273 amino acids whereas the variant of study, the Omicron BA.1.1 variant lacks a few amino acids at 1270 amino acids. The BA.1.1 variant is three amino acids deficient to the wildtype. This is observed in the multiple sequence alignment that was conducted about where the mutated amino acids were identified. In terms of the molecular weight between the Wuhan-Hu-1 and the BA.1.1 variant, the variant (141300.09) is seen to have a higher molecular weight that the Wuhan-Hu-1 (1411178.47), despite the deficiency in amino acids. The theoretical pI of the Wuhan-Hu-1 is also lower compared to the BA.1.1 variant. The BA.1.1 has a pI of 7.14 and this indicates that the protein is alkaline. The Wuhan-Hu-1 on the other hand which has a pI value of 6.24 is acidic. Similar to how the pH value functions, a pI value more than 7 is alkaline and a value lower than 7 is acidic. According to prior research, the SARS-CoV-2 S protein is found to be rather more positively charged than the SARS-CoV protein, which leads to a higher propensity for attaching to negatively charged regions of other molecules via nonspecific and selective interactions [4]. Based on the data obtained the expected charges of the residues (Arg + Lys) was found in a similar trend where the BA.1.1 is higher than the wildtype. The wildtype's value is 103 whereas the BA.1.1 is 111. This goes to show that the BA.1.1 variant contains signifincalty more positively charged amino acids than the wildtype. This aids in improving the propensity to bind to the negatively charged areas of other molecules such as the ACE2 [2]. The number of negatively charged residue on the other hand did not have much difference in value as the wildtype is 110 and the variant is 111. The instability index of the wildtype is 33.01 whereas he variant is 34.69. This shows that the BA.1.1 has a slight improvement in stability compared to the wildtype as a stability score of lesser than 40 is considered to have a stable structure [2]. The aliphatic index, a positive factor linked to enhanced protein thermostability, is 84.67 for wild type and 84.95 for BA.1.1, indicating that it can withstand a wide range of temperatures. According to a recent study, increased Omicron

thermal stability may lead to greater Omicron persistence in exposed environments, posing a higher risk of transmission among household contacts than the Delta form [5]. Increased viral stability may aid viral attachment to host cells by improving receptor recognition efficiency, but it may hinder viral membrane fusion [2]. Finally, the grand average of hydropathicity index (GRAVY) was calculated for the spike protein of both wildtype and BA.1.1 variant. This data was determined via the Kyte and Doolittle's hydropathy values which ranged from -2 to +2 for most proteins and the positively scored proteins have a higher hydrophobicity [2]. It goes to define that proteins with a hydrophobicity score below 0 are more likely to be hydrophilic, while proteins with a score above 0 are more likely to be membranous hydrophobic proteins. The GRAVY was estimated to be -0.079 for wild type and -0.080 for BA.1.1, indicating that they're both hydrophilic in nature. The BA.1.1 variant was determined to be more hydrophilic and less hydrophobic, which indicates a decrease in the pathogenicity of the viral protein [6].

The physiochemical properties between the receptor-binding domain of the wildtype and the BA.1.1 variant was also identified and recorded in Table S4.2. Due to the mutation that had occurred, the number of amino acids between both are different. The wildtype has 229 amino acids whereas the BA.1.1 variant has 214 amino acids. The variant is 15 amino acids deficient from the wildtype. This could contribute to the difference in the molecular weight of both RBD sequence. Unlike the spike protein above, the wildtype's molecular weight, 25745.11, is higher than the BA.1.1 variant, 24441.05. As for the theoretical pI value, despite the difference both are alkaline. However, the BA.1.1 variant seems to be more alkaline than the wildtype where wildtype with 8.91 and BA.1.1 with 9.64 pI value. The difference between the positively charged residues (Arg + Lys) and negatively charged residue (Asp +Glu) between the both is quite significant. Both are more positively charge but the BA.1.1 is higher with 29 residues unlike the wildtype which has 23 residues. It can be speculated that the increase in positively charged residues in the Omicron variant would most likely increased its affinity towards the human ACE2 protein. Furthermore, there is also a large difference between the instability index of the wildtype, 21.04, and BA.1.1, 32.60. It can be said that even though the variant is stable, the wildtype RBD is far more stable. The aliphatic index of 71.44 for wildtype and 73.27 shows that both rather have a similar thermostability and can withstand a range of temperature. Finally, the GRAVY was estimated to be -0.259 for wild type and -0.289 for BA.1.1, indicating that they're both hydrophilic in nature.

The N-Terminal Domain (NTD) functions as a "wedge" in the conformational changes of the S protein, regulating them [7]. Therefore, it remains as one of the important parts of the SARS-CoV-2 sequence. The physiochemical data obtained for the NTD for both wildtype and BA.1.1 shows that there is three amino acids deficiency between them and can be seen in Table S4.3. The wildtype has 293 amino acids and the variant has 290 amino acids. Their molecular weight shows a difference as well, where the wildtype is higher with 33224.64 and BA.1.1 with 32902.27. The NTD data has a different theoretical pI than the RBD data. The wildtype has an alkaline pI of 8.13, while BA.1.1 has an acidic pI of 6.51. There remains to be no difference in the positively charged residues (Arg + Lys) for both NTD sequences, 26 residues. However, the BA.1.1 NTD is more negatively charged as it has 27 negatively charged residues (Asp + Glu). On the other hand, wildtype has more positively charged residues (26), than negatively charged residues (24). The instability index, aliphatic index, and grand average of hydropathicity are not significantly different between the wildtype and BA.1.1. This indicates that both NTD sequences are stable, have a similar thermostability and are hydrophilic in nature.

From the amino acid composition comparison of the wildtype against the BA.1.1 variant (Table S2.1, S2.2 and S2.3) for the spike protein, RBD and NTD shows the increase and decrease of

important charged residues in the side chains of the sequence. Overall, from Table S1.2, it can be seen that the spike protein comparison doesn't have much significant difference. There weren't many increases or decreases that were above 5%, except for Lys (K) which has an increase in 6% between the wildtype (4.8%) and BA.1.1 (5.4%). It is also similar for the NTD amino acid composition comparison from Table S2.3. there weren't any significant differences between the wildtype and BA.1.1 except for Serine (S) which has a huge increase of 7.6%. The amino acid composition for the RBD sequences showed several notable difference in percentage. The first would be the increase in the charged residue of the Arginine (R) and Lysine (K) side chains. It was found that these side chains are important factors that contributes the formation of salt bridges in the BA.1.1 variant [8]. In the receptor-binding domain (RBD), there is also an increase in Asparagine (N) residues, which leads to the formation of hydrogen bonds in BA. 1.1 [2]. There is an increase in hydrophobic residues such as phenylalanine (F), Alanine (A), Leucine (L), Proline (P), and leucine (L) in the receptor binding domains of B.A.1.1 compared to the wild type, which are generally buried inside the protein core. When compared to all sub-lineages, there is an increase in polar amino acids such Tryptophan (W), which is commonly found near the protein's surface, Serine (S), which forms hydrogen bonds, and Valine (V), which is hydrophobic [2].

3.3 Secondary structure and Intrinsically disordered prediction

The secondary structure prediction (Table S3) was conducted for the spike protein, RBD and NTD sequences between the Wuhan-HU-1 and the Omicron BA.1.1 variant. In the whole spike protein comparison, it can be observed that the extended strand and the random coil values has decreased slightly unlike the alpha-helix which exhibited an increase. The uniqueness of the mutation that was present in the BA.1.1 lineage caused this result in the whole spike protein. The increased alpha-helix indicates that alpha-helices are more susceptible to be influenced by mutation than beta strands [2]. In the RBD on the other hand, the opposite is observed. The extended stand and the random coil has increased whereas the alpha-helix has decreased compared against the wildtype. Finally, the NTD sequence shows a decrease in the extended strand and random coil, as well as an increase in the alpha-helix of the BA.1.1 variant, similar to the entire spike protein data.

In an overview, it can be observed that the Omicron variant BA.1.1 whole spike protein is less disordered than the wildtype based on the Intrinsically disordered prediction in Table S5. There are 98 residues disordered in the wildtype whereas only 85 in the BA.1.1 variant. An important disorder in the receptor binding motif, 468-473, is seen in the wildtype and not in the Omicron variant BA.1.1 [9]. This disordered region is important as it is required for the human ACE2 binding. In terms of the effect of disordered residues/regions on spike protein stability and ACE2 binding, this transition is critical [10]. This would mean that the wildtype is more unstable than the BA.1.1 variant. The series of mutations that had occurred have made the BA.1.1 variant to be more stable than the wildtype.

3.4 Protein structure and Electrostatic surface potential analysis.

The tertiary structures of the BA.1.1 RBD and NTD sequences were predicted using the AlphaFold, a computational tool that utilised the deep learning algorithm [11]. For both the NTD and RBD, 5 models were predicted and ranked based on their accuracy. In all the models that were produced the disordered regions were identified by a pLDDT < 50 and are graphically presented as long filaments [11]. Figures 2.1 and 2.2 shows the Rank 1 models for both RBD and NTD respectively. These high ranked models were used to identify the electrostatic surface potential of the sequences [2]. In order to conduct the comparison, the electrostatic surface potential of the human ACE2 and the wildtype sequences were also identified. This is to

identify the properties of the BA.1.1 against the wildtype on the human ACE2 protein. It aids in identifying the RBD interaction of the wildtype and BA.1.1 with the human ACE2 protein. The PyMOL tool was utilised with the implementation of the Adaptive Poisson-Boltzmann Solver (APBS) to calculate the electrostatic potential for the hACE2, wildtype RBD, BA.1.1 RBD and NTD (Figure 3.0). It is known that the result from PyMOL that the colour red is associated with a negative electrostatic potential, whereas blue is associated with a positive electrostatic potential. The hACE2 is seen with higher concentration of red colour than blue, which shows that it is a highly negative protein. The wildtype on the other hand has both red and blue colours visible. The blue colour seems to be more abundant but it doesn't have a thick concentration of red colour as well. Therefore, it can be said that the wildtype RBD is to be slightly positive protein. The BA.1.1 RBD is found to have a higher concentration of blue colour overall. Thus it being an electropositive protein. Comparing it with the wildtype, the BA.1. I more electropositive. The NTD of the BA.1.1 appears to have more red colour signifying that it is more electronegative. The identification of the electrostatic potential for the proteins are important to identify the viral pathogenicity, infectivity, and transmission of the protein. Since positive and negative attracts each other, the negatively charged hACE2 would attract the positively charged wildtype RBD and BA.1.1 RBD. However, since the BA.1.1 RBD is more electropositive, the potential for it to be attracted and transmit is higher than the wildtype. Since the wildtype RBD has a rather large proportion of red colour, it has weaker energy. This makes the Omicron BA.1.1 variant to be more transmissible. It has a higher affinity towards the hACE2 protein. The mutation that causes the birth of the BA.1.1 variant has changed its electrostatic potential, affinity towards the hACE2, infectivity and the transmission of the virus. The SARS-CoV-2 virus is able to increase its binding affinity towards the human ACE2 and avoid antibody detection due to the evolutionary advantages at the RBD sequence [12]. Previous studies show that given the virus's increased infectivity in human cells, a single mutation is unlikely to result in a considerable increase in viral infectivity [2]. As observed from the Omicron BA.1.1 variant, several mutations at the RBD has increase infectivity and also appear to represent a suitable infection pathway. As for the NTD domain of the BA.1.1, since it is more electronegative, it has lower affinity to the other protein connection. The NTD domain has lower affinity to other attaching to the human cells and interacting proteins unlike the RBD. The NTD domain is known to be related with the pathogenicity of the virus. Since the BA.1.1 NTD is more electronegative, it would not be attracted to the negatively charged human cells as the would repel each other. This reduces the severity of the BA.1.1 variant, as it the probability of it binding to the proteins in the lungs are lower.

3.5 Molecular Docking

The HawkDock protein-protein docking tool was used to identify binding pockets and interaction residues in RBD-ACE2, wildtype (WT) and the BA.1.1 variant. In order to identify the binding free energy calculations, the tool combines the ATTRACT docking algorithm, the HawkRank scoring function, and the MM/GBSA free energy decomposition analysis [2]. Figure 4.0 shows the binding free energy complex of the wildtype and the BA.1.1 RBD which was run against the hACE2/ the binding energy for the wildtype is -8.28 (kcal/mol) whereas the binding energy for the BA.1.1 is -66.29 (kcal/mol). The binding energy shows the interaction between the ligand and the BA.1.1 variant and the wildtype. Comparing both, it can be clearly seen that the BA.1.1 has the higher binding free energy compared to the wildtype. The higher energy in the BA.1.1 variant shows that it has increased binding affinity towards the human ACE2 protein.

4 CONCLUSION

To summarize the following information, it can be determined that the Omicron sub-variant BA.1.1 was analysed against the Wuhan-Hu-1 (wildtype) with the aid of different computational bioinformatics tools. Based on the results, it can be observed that the BA.1.1 variant has got different properties compared to the wildtype which was caused due to the mutations that it has been through. The mutations that the BA.1.1 variant has undergone is the factor which is responsible for the changes caused in terms of the transmission rate, pathogenicity and infectivity of the virus. Not only the spike protein, but also the N-Terminal Domain and the Receptor Binding Domains were studied as they pose a significant function in the viruses' ability to infect and transmit. The results from above have proven that the BA.1.1 variant has increase in transmission rate which allows it to spread faster between the infected and uninfected. It is highly hoped that this Omicron sub-variant would be studied experimentally in efforts to produce improved antiviral therapies, vaccines and drugs that would aid in curbing the virus from spreading further and taking people's lives.

5 LIMITATIONS OF THE STUDY

The comparison of the BA.1.1 Omicron variation to the Wuhan-Hu-1 was primarily based on computational sequence and structural predictions, which should be studied and confirmed in future experiments. The BA.1.1 Omicron protein structure was disclosed in this study, establishing the foundation for future research on the SARS-CoV-2 Omicron and its variations.

6 REFERENCES

1. Basu, S., Mukhopadhyay, S., Das, R., Mukhopadhyay, S., Singh, P. K., & Ganguli, S. (2020). Impact of clade specific mutations on structural fidelity of SARS-CoV-2 proteins. *Impact of Clade Specific Mutations on Structural Fidelity of SARS-CoV-2 Proteins*. https://doi.org/10.1101/2020.10.20.347021
2. Bendl, J., Stourac, J., Salanda, O., Pavelka, A., Wieben, E. D., Zendulka, J., Brezovsky, J., & Damborsky, J. (2014). PredictSNP: Robust and Accurate Consensus Classifier for Prediction of Disease-Related Mutations. *PLoS Computational Biology*, *10*(1), e1003440. https://doi.org/10.1371/journal.pcbi.1003440
3. Cui, Z., Liu, P., Wang, N., Wang, L., Fan, K., Zhu, Q., Wang, K., Chen, R., Feng, R., Jia, Z., Yang, M., Xu, G., Zhu, B., Fu, W., Chu, T., Feng, L., Wang, Y., Pei, X., Yang, P., . . . Wang, X. (2022). Structural and functional characterizations of infectivity and immune evasion of SARS-CoV-2 Omicron. *Cell*, *185*(5), 860–871.e13. https://doi.org/10.1016/j.cell.2022.01.019
4. David, A., Islam, S., Tankhilevich, E., & Sternberg, M. J. (2022). The AlphaFold Database of Protein Structures: A Biologist's Guide. *Journal of Molecular Biology*, *434*(2), 167336. https://doi.org/10.1016/j.jmb.2021.167336
5. Kemp, S., Meng, B., Ferriera, I., Datir, R., Harvey, W., Collier, D. A., Lytras, S., Papa, G., Carabelli, A., Kenyon, J., Lever, A., James, L. C., Robertson, D., & Gupta, R. K. (2021). Recurrent Emergence and Transmission of a SARS-CoV-2 Spike Deletion H69/V70. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3780277
6. Kumar, S., Karuppanan, K., & Subramaniam, G. (2022). Omicron (BA.1) and Sub-Variants (BA.1, BA.2 and BA.3) of SARS-CoV-2 Spike Infectivity and Pathogenicity: A Comparative Sequence and Structural-based Computational Assessment. *Omicron (BA.1) and Sub-Variants (BA.1, BA.2 and BA.3) of SARS-CoV-2 Spike Infectivity and Pathogenicity: A Comparative Sequence and Structural-Based Computational Assessment*. https://doi.org/10.1101/2022.02.11.480029

7. Li, Y., Wang, T., Zhang, J., Shao, B., Gong, H., Wang, Y., He, X., Liu, S., & Liu, T. (2021). Exploring the Regulatory Function of the *N* -terminal Domain of SARS-CoV-2 Spike Protein through Molecular Dynamics Simulation. *Advanced Theory and Simulations*, *4*(10), 2100152. https://doi.org/10.1002/adts.202100152
8. Mendoza-Espinosa, P., García-González, V., Moreno, A., Castillo, R., & Mas-Oliva, J. (2009). Disorder-to-order conformational transitions in protein structure and its relationship to disease. *Molecular and Cellular Biochemistry*, *330*(1–2), 105–120. https://doi.org/10.1007/s11010-009-0105-6
9. Shang, J., Ye, G., Shi, K., Wan, Y., Luo, C., Aihara, H., Geng, Q., Auerbach, A., & Li, F. (2020). Structural basis of receptor recognition by SARS-CoV-2. *Nature*, *581*(7807), 221–224. https://doi.org/10.1038/s41586-020-2179-y
10. Socher, E., Conrad, M., Heger, L., Paulsen, F., Sticht, H., Zunke, F., & Arnold, P. (2021). Mutations in the B.1.1.7 SARS-CoV-2 Spike Protein Reduce Receptor-Binding Affinity and Induce a Flexible Link to the Fusion Peptide. *Biomedicines*, *9*(5), 525. https://doi.org/10.3390/biomedicines9050525
11. Wang, M. Y., Zhao, R., Gao, L. J., Gao, X. F., Wang, D. P., & Cao, J. M. (2020). SARS-CoV-2: Structure, Biology, and Structure-Based Therapeutics Development. *Frontiers in Cellular and Infection Microbiology*, *10*. https://doi.org/10.3389/fcimb.2020.587269
12. Xu, C., Wang, Y., Liu, C., Zhang, C., Han, W., Hong, X., Wang, Y., Hong, Q., Wang, S., Zhao, Q., Wang, Y., Yang, Y., Chen, K., Zheng, W., Kong, L., Wang, F., Zuo, Q., Huang, Z., & Cong, Y. (2021). Conformational dynamics of SARS-CoV-2 trimeric spike glycoprotein in complex with receptor ACE2 revealed by cryo-EM. *Science Advances*, *7*(1). https://doi.org/10.1126/sciadv.abe5575