Computational and comparative investigation of hydrophobic profile of spike protein of SARS-CoV-2 and SARS-CoV

Uma Shekhawat1,2 · Anindita Roy Chowdhury (Chakravarty)1

© The Author(s), under exclusive licence to Springer Nature B.V. 2022

Abstract
The hydrophobic force is one of the most dominant factors in protein folding. A protein becomes functional only when it achieves its three-dimensional structure and stability upon folding. For a better understanding of the hydrophobic effects and their function in protein folding, quantitative measurement of the hydrophobicity of amino acid side chains is crucial. Spike protein is the primary structural protein in SARS-CoV-2 and SARS-CoV. This study explores how protein sequences in SARS-CoV-2 and SARS-CoV spike proteins encode hydrophobic interactions. Computational tools/techniques have been utilized to investigate the protein sequences of the spike proteins of SARS-CoV-2 and SARS-CoV. Investigations provided an estimate of hydrophobic distribution and its relative strength, indicating a hydrophobic pattern. Analysis of the spike protein’s hydrophobic profile may help identify and treat the virus-caused disease; additionally, it can give an insight into the transmissibility and pathogenicity of the virus.

Keywords Aliphatic · Aromatic · Hydrophobicity · SARS-CoV · SARS-CoV-2 · Spike protein

1 Introduction

The hydrophobic force is one of the primary forces which drives for protein folding and protein three-dimensional structure, which has been used to predict where different protein folds would be located. As a physicochemical property, hydrophobicity is the dominant factor responsible for the protein’s three-dimensional structure, which in turn would make it functional [1, 2]. Hydrophobic interaction is essential in maintaining a stable and biologically active protein because that facilitates the protein to reduce surface area and unfavorable interactions in water [3] and the presence of water near hydrophobic surfaces has a
significant impact on protein folding [4, 5]. The rate of protein folding is an important aspect for developing an insight into the hydrophobic interaction at molecular length scales [6].

The cause of the epidemic of new coronavirus disease 2019 (COVID-19) is severe acute respiratory syndrome corona-virus 2 (SARS-CoV-2) [7, 8]. Before December 2019, only six human coronaviruses (HCoV) species that cause human illness were identified as HCoV-229E, HCoV-OC43, HCoV-NL63, HCoV-HKU1, SARS-CoV, and MERS-CoV. The SARS-CoV and SARS-CoV-2 are, respectively, the fifth and seventh coronavirus that caused life-threatening infections. In coronavirus, the structural proteins are responsible for the infection. Among the structural proteins of SARS-CoV-2, spike protein is a major target for obtaining antibodies as it is common among all coronaviruses [9–11]. Structural and molecular details of spike protein and differences in spike protein length, which is longer in SARS-CoV-2, are expected to have a role in the pathogenesis and treatment of the disease caused by the virus [10–13]. To mediate host-cell entry, SARS-CoV-2 uses a highly glycosylated spike (S) protein that projects from the viral surface and attaches to angiotensin-converting enzyme 2 (ACE2) [8, 13, 14]. In disease pathogenesis, knowledge of the glycosylation of recombinant viral spikes plays a variety of roles, including regulating protein folding and stability, determining viral tropism, and enabling immune recognition [7, 13, 15]. Casalino et al. estimated the surface area covered by glycan shield over the spike protein and how it changes based on the protein’s conformational state [14].

In a protein molecule, mapping of hydrophobic amino acids also indicates secondary structural configuration, tertiary properties, folding patterns, etc. all of which contribute to the molecule’s functional attainment [16]. Investigation regarding hydrophobic residue distribution would give an insight into the maintenance of structural integrity and stability of the spike protein [3]. In this paper, the authors aim to investigate the integral and relative contribution of varied strengths of hydrophobic aromatic and aliphatic amino acids. Further, this study examines the hydrophobic profile of the protein sequences and the sequence alignment of protein sequences of the spike proteins of SARS-CoV-2 and SARS-CoV in comparison to allied sequences. The investigation has been carried out based on designed programs as well as computational and statistical tools.

2 Methodology

2.1 Selection of sequences

Protein sequences of the spike proteins of SARS-CoV-2 (UniProt ID: P0DTC2) which originated from the Wuhan seafood market [17–19] and SARS-CoV (UniProt ID: P59594) in fasta format were obtained from the Universal Protein Resource [20, 21]. These two sequences were then blasted by using the sequence alignment tool BLASTP [22, 23] with respect to the Protein Data Bank (PDB) [24] as a database.

2.2 Hydrophobic scale

To determine the hydrophobic strength of protein sequences, a hydrophobic scale was required. Various types of experimental, statistical, or combinational hydrophobic scales are available [25–27]. As the statistical distribution of hydrophobic residues was involved in this investigation, an experimental method-based scale known as Fauchere-Pliska hydrophobicity scale was
preferred [25, 28]. In many biophysical, bioinformatics, and pharmacological investigations, this scale is effective [27, 29–33].

2.3 Calculation of hydrophobic contribution

C++ program was executed on the downloaded fasta sequences of the spike proteins of SARS-CoV-2 and SARS-CoV as well as their aligned sequences to determine the hydrophobic strength [29–31]. The output files of the program showed the specific hydrophobic details for each protein sequence of the spike proteins of SARS-CoV-2 and SARS-CoV. The total aromatic and aliphatic hydrophobic content due to specific amino acid (XH) and hydrophobic contribution of specific amino acid per unit length (XHL) was calculated (where X indicated the specific amino acid).

Hydrophobic contribution was calculated on the basis of the frequency of occurrence of respective amino acids in a protein sequence and their hydrophobic scale value [29–31]. The total hydrophobic contribution of all aromatic residues was considered and mentioned as TAroH (i.e., WH, FH, YH, and HH); TAliH (IH, LH, CH, VH, PH, AH, and MH) stated the same for all aliphatic residues. The contribution of respective aromatic amino acids was considered for TAroH, like tryptophan’s contribution to aromatic hydrophobic content (WH%) and for all other hydrophobic amino acids considered here; TAliH replaces TAroH for aliphatic category.

2.4 Sequence alignment

Using BLASTP, sequences similar to protein sequences of the spike proteins of SARS-CoV-2 (UniProt ID: P0DTC2) and SARS (UniProt ID: P59594) from the PDB database were retrieved. Since the domain of this study involved the protein sequences, BLASTP program was thus chosen to compare protein sequences with a protein sequence database [34, 35]. To endorse sequence alignment of all the retrieved sequences that were similar to the protein sequences of spike proteins of SARS-CoV-2 (UniProt ID: P0DTC2) and SARS-CoV (UniProt ID: P59594), a multiple sequence alignment tool, ClustalW, was used as well [36, 37].

3 Results and discussion

Implementing the previously stated methodology which involved execution of a designed computer program and a number of computational tools, hydrophobic profile of spike proteins from two coronavirus strains, SARS-CoV-2 and SARS-CoV, and their aligned sequences were investigated and compared.

3.1 Hydrophobic profile of spike proteins of SARS-CoV-2 and SARS-CoV

On executing the C++ program on the protein sequences of the spike proteins of SARS-CoV-2 (UniProt ID: P0DTC2) and SARS-CoV (UniProt ID: P59594), every amino acid’s contribution to their relevant category was determined. A more in-depth understanding of the contribution of distinct hydrophobic amino acids to total aromatic hydrophobic content (Table 1a, Fig. 1a) and per unit length of the sequence (Fig. 1b) showed that the highest
Table 1 Contribution of different (a) aromatic residues and (b) aliphatic residues to respective hydrophobic content in spike proteins of SARS-CoV-2 (UniProt ID: P0DTC2) and SARS-CoV (UniProt ID: P59594)

(a) Aromatic residues

| S. No. | UniProt ID | Seq Len | Tryptophan No. (W) | Contr. (WH%) | Phenylalanine No. (F) | Contr. (FH%) | Tyrosine No. (Y) | Contr. (YH%) | Histidine No. (H) | Contr. (HH%) |
|--------|------------|---------|-------------------|--------------|----------------------|--------------|-----------------|--------------|-----------------|-------------|
| 1      | P0DTC2     | 1273    | 12                | 12.33        | 77                   | 62.97        | 54              | 23.68        | 17              | 1.00        |
| 2      | P59594     | 1255    | 11                | 10.89        | 83                   | 65.41        | 54              | 22.82        | 15              | 0.85        |

(b) Aliphatic residues

| S. No. | UniProt ID | Seq Len | Isoleucine No. (I) | Contr. (IH%) | Leucine No. (L) | Contr. (LH%) | Cysteine No. (C) | Contr. (CH%) | Valine No. (V) | Contr. (VH%) | Proline No. (P) | Contr. (PH%) | Alanine No. (A) | Contr. (AH%) | Methionine No. (M) | Contr. (MQ%) |
|--------|------------|---------|-------------------|--------------|-----------------|--------------|-----------------|--------------|----------------|--------------|----------------|--------------|----------------|-------------|----------------|-------------|
| 1      | P0DTC2     | 1273    | 76                | 23.43        | 108             | 31.44        | 10.55           | 97           | 20.27          | 58           | 7.15           | 79           | 4.19           | 14          | 2.94          |
| 2      | P59594     | 1255    | 78                | 24.56        | 99              | 29.45        | 10.50           | 91           | 19.42          | 57           | 7.18           | 84           | 4.55           | 20          | 4.30          |
The precise contribution of each aliphatic hydrophobic content to total aliphatic hydrophobic content (Table 1b, Fig. 2a) and per unit length of the sequence (Fig. 2b) showed that leucine with the second-highest scale value in aliphatic category was the maximum aliphatic hydrophobic contributor, both in terms of total aliphatic hydrophobic content and per unit length. In the aliphatic category, methionine was the least contributor. According to the hydrophobic scale chosen, methionine does not have the least scale value but because of its least frequency of occurrence, it emerged as the lowest aliphatic hydrophobic contributor. From Table 1a, it can be observed that the hydrophobic contribution of
aromatic amino acids, tryptophan, phenylalanine, and tyrosine are numerically quite close to each other, within range of 3%, and that of histidine within 1%. Similarly, Table 1b showed that hydrophobic contribution of aliphatic amino acids isoleucine, leucine, and methionine were very close to each other, within the 3% range, whereas cysteine, proline, valine, and alanine’s contribution for both spike proteins were within 1% range of variation. A minimal variation in these amino acid contributions may cause a significant variation in structure and function [38]. Qiu et al. have revealed how single residue diversity can cause substantial changes in intermolecular interactions hinting at the significance of hydrophobic interactions. Intricate details regarding hydrophobic profile as obtained in this research endeavor can play a role in peptide therapeutics. Results here indicate that computational approach for hydrophobic characterization helps in understanding protein folding and related aspects [39, 40].

3.2 BLASTP results of spike proteins of SARS-CoV-2 and SARS-CoV

Using the sequence alignment tool, BLASTP [22], protein sequences similar to the spike proteins of SARS-CoV-2 (UniProt ID: P0DTC2) as well as SARS-CoV (UniProt ID: P59594) were obtained and have been described based on query coverage and average E-Value. For spike proteins of SARS-CoV-2, 42 protein sequences and, for SARS-CoV, 48 protein sequences emerged, respectively, during this simulation.

Based on the alignment description, it has been observed that the lower the E-value, or closer it is to zero, the better the match with query sequence [34, 41]. The E-value of the sequence represents a better statistical indicator of the significance of the sequence match, and query coverage refers to the percentage of query length that is covered in the aligned segments [35, 37]. Usually, an E-value $< 10^{-5}$ is considered reliable [37, 42]; accordingly, protein sequences of the spike proteins of SARS-CoV-2 and SARS-CoV along with their respective parameters have been selected (SI Table 1).

Based on maximum query coverage, for UniProt ID: P0DTC2, among the 42 PDB sequences, the intervals of higher query coverage ranging from 61 to 75% have been taken, and similarly for UniProt ID: P59594, among the 48 PDB sequences, the ranges of 60% to 95% have been considered to maintain parity in sequence number with the other PDB ID. Table 2 highlights the parameters significant for this analysis of the spike protein sequences of SARS-CoV-2 and SARS-CoV. Table 2 revealed the best seven aligned protein sequences from PDB [24] in BLASTP resource [22] with different parameters, which indicated that the spike proteins of SARS-CoV-2 and SARS-CoV sequences were the best matches with these seven protein database sequences, taking into consideration both the total alignment length and the number of matching residues. In addition, all these selected PDB sequences have the highest alignment score that is calculated from the sum of amino acid matching rewards and penalties for mismatches and gaps according to the BLASTP tool.

The spike protein of SARS-CoV-2 sequence was best aligned with PDB IDs 6VSJ_A, 5I08_A, 3JCL_A, 6NZK_A, 5X59_A, 5W9H_A, and 6NB3_A. The spike protein of SARS-CoV protein sequence was best aligned with PDB IDs 6VSJ_A, 7KIP_A, 3JCL_A, 6NZK_A, 5X59_A, 5W9H_A, and 6NB3_A. The PDB IDs 5X59_A, 5W9H_A, and 6NB3_A were related to MERS-CoV (Middle East respiratory syndrome coronavirus). PDB IDs 5I08_A, 6NZK_A, and 7KIP_A were related to human coronavirus (HCoV-HKU1, HCoV-OC43, and HCoV-NL63). PDB IDs 6VSJ_A and 3JCL_A belong to murine hepatitis virus strain (MHV-A59). Work done by Zhou and their team supported this close
relationship through their pairwise protein sequence analysis [43]; the statistical parameters in Table 2 highlight the same.

### 3.3 Multiple sequence alignment results

The sequence identity and similarity among the best seven aligned sequences (Table 2) for spike proteins of SARS-CoV-2 (UniProt ID: P0DTC2) and SARS-CoV (UniProt ID: P59594), respectively, have been determined by a bioinformatic program ClustalW [36] (SI Tables 2 and 3). Some of the aligned sequences have very high sequence similarity, while others have very low similarity over certain stretches. Multiple sequence alignment scores of the spike protein of SARS-CoV-2 (SI Table 2) indicated the protein sequence of 6VSJ_A has been best aligned with the protein sequence of 3JCL_A on two different stretches with 100% and 99% alignment scores, respectively. Interestingly, both of them were found to be related to murine hepatitis virus strain. Similarly, the protein sequence of 5X59_A, 5W9H_A, and 6NB3_A have a 99% alignment score with each other in certain stretches, which means that all these three IDs share nearly identical sequence stretches and are related to MERS-CoV (SI Table 2).

In the case of the multiple sequence alignment score of the spike protein of SARS-CoV-2 (SI Table 3), a similar result is reflected. The hydrophobic profile analysis has been done for the best-aligned protein sequences of SARS-CoV-2 and SARS-CoV. Zheng and Song’s work also showed the similarity between these protein sequences through phylogenetic analysis [44]. Results, as revealed by this multiple sequence alignment, also get supported by the work of Lu et al. [19] which, too, revealed the phylogenetic relationship of
these viruses. ClustalW results have been seen to support earlier investigations carried out for coronaviruses also [45].

3.4 Hydrophobic profile analysis of best aligned protein sequences of SARS-CoV-2 (UniProt ID: P0DTC2) and SARS-CoV (UniProt ID: P59594)

Hydrophobic profile analysis of best-aligned sequences of SARS-CoV-2 (UniProt ID: P0DTC2) and SARS-CoV (UniProt ID: P59594) revealed the detailed aromatic and aliphatic hydrophobic constitution of each of the amino acid protein sequences. Every amino acid’s contribution to its respective category has been determined using the designed C++ program [29, 31] on the best-matched sequences. All the best-aligned protein sequences have a specific PDB ID that belongs to the protein database and has a particular aligned stretch length that is identical/similar to the query sequence. Clusters of aligned stretches observed in Fig. 3 indicate identical amino acids on specific stretches; hydrophobic contribution seems to be more dominant in SARS-CoV-2 than SARS-CoV coronavirus plots. Figure 3a (SI Tables 4 and 5) established phenylalanine as the highest hydrophobic aromatic contributor and histidine as the lowest one and, thus, synchronizes with the findings of Table 1a and Fig. 1 for the full-length sequences. The plot of hydrophobic content per unit length shown in Fig. 3b supported the same relative variation as in Fig. 3a.

Analogous hydrophobic profile analysis about aliphatic category has been shown in Fig. 4 (SI Tables 6 and 7). Similar to the full-length sequences (Table 1b and Fig. 2), even for the aligned stretches, leucine contributes the most and methionine the least; frequency of occurrence is the reason behind it. Comparing the plots of aligned stretches of SARS-CoV-2 and SARS-CoV, it can be said that the length of the aligned stretches varies in case of SARS-CoV, whereas that of SARS-CoV-2 are nearly identical to each other. The variations in aliphatic hydrophobic content per unit length for aligned protein sequences of SARS-CoV-2 and SARS-CoV (Fig. 4b) showed that the overall contribution range for proline and alanine is nearly the same and this feature is more significant for SARS-CoV-2 coronavirus than SARS-CoV. All tables and graph plots reporting the investigation carried out on the segments of different sequences of spike proteins show that contributors are the same whether it is the lowest or highest contributor for both aromatic and aliphatic categories indicating the conservation pattern of hydrophobicity [10]. Conservation of certain stretches in protein sequences of different strains of coronaviruses has been shown in the work of Robson as well [45].

The amino acid contributions of the spike proteins from SARS-CoV-2 and SARS-CoV were analyzed to evaluate the compositional variations. The difference in the contributions of aromatic amino acids was noted (Fig. 3, SI Tables 4 and 5) for the SARS-CoV-2 aligned protein sequences; the overall variation of contribution range is in the range of 1% to 14%, and the range for SARS-CoV varies from 1 to 11%. However, a considerable difference in the contributions of aliphatic amino acids was also observed (Fig. 4, SI Tables 6 and 7). The overall variation of contribution range for all aliphatic amino acids in the case of SARS-CoV-2 lies from 3 to 15%, whereas in SARS-CoV, it lies from 2 to 4% only. This indicates though the aromatic hydrophobic contribution of both is within a similar range but that for aliphatic varies over a much wider range in case of SARS-CoV-2. The result in this paper indicated the significance of specific hydrophobic aromatic and aliphatic contributors, like, phenylalanine and leucine, respectively. Findings of Seyren et al. supported this result with their identification of the dominant presence of phenylalanine and tyrosine in the aromatic and leucine, proline, and cysteine in the aliphatic categories [7].
The highest contributor phenylalanine of aromatic and leucine of aliphatic category of the spike protein of SARS-CoV-2 and SARS-CoV is shown in Figs. 5 and 6 respectively. PDB protein sequence (PDB ID: 5I08_A) was uniquely identical to the UniProt protein sequence of SARS-CoV-2 (UniProt ID: P0DTC2) and the other PDB protein sequence (PDB ID: 7KIP_A) that was uniquely identical to UniProt protein sequence of SARS-CoV (UniProt ID: P59594) to a great extent has been considered (Table 2) to show that the highest hydrophobic contributors as the 3-dimensional structure was available in PDB of both IDs 5I08_A and 7KIP_A. Moreover, protein sequences of the PDB IDs 5I08_A and 7KIP_A belong to the human coronavirus category [24]. So here on ChimeraX [46–48], the three-dimensional structure of these particular IDs has been investigated to show the structural positions of the highest contributor of hydrophobic aromatic (phenylalanine) and aliphatic (leucine) categories which have been highlighted in green. The front view of the
full sequence (PDB ID: 5I08_A and 7KIP_A) and its respective core view are visualized in Figs. 5 and 6.

Leucine and phenylalanine, which are found inside the protein’s core, are crucial for structural stability, and these occur most frequently in the whole spike protein when comparing SARS-CoV-2 and SARS-CoV [49]. These variations to the spike protein’s amino acids may have an impact on how well the spike protein binds to other proteins during the viral fusion process [39]. Also, Srinivasulu and colleagues as well as Zheng and Song reported
that the SARS-CoV-2 and SARS-CoV spike proteins share nearly 73.5% amino acid sequence identity, implying that the remaining 24.5% non-conserved amino acid sequences may be accountable for the antigenic variations between these two proteins which supports the findings for the hydrophobic profile of spike protein of SARS-CoV-2 and SARS-CoV [14, 19, 38, 44], discussed here.

Results here indicate that though conservation of hydrophobic pattern can be observed, but intricate analysis of the hydrophobic profile of SARS-CoV-2 and SARS-CoV revealed certain individual variations as well which possibly is indicative of the fact of their respective nature and maybe even virulence [44]. Through this investigation, hydrophobic profile of the spike protein of SARS-CoV-2 and its comparison with SARS-CoV coronavirus could be studied, and thus, the significance of hydrophobic characterization could be recognized.

Examination of hydrophobic profile analysis through computational approaches facilitates the in silico investigation for drug designing [50]. It has been reported that the glycosylation
of the SARS-CoV-2 spike protein can prevent immune response; glycan shielding has been extensively explored in a variety of viral glycoproteins, including HIV-1 envelope protein (Env), Lassa virus glycoprotein complex (LASV GPC), and influenza hemagglutinin (HA) [13–15, 51]. Structural and functional significance of hydrophobicity has also been highlighted by Leung et al., and Seyran et al.’s increasing hydrophobicity of residues improves peptide binding in an anti-HIV-1 Env peptide which in turn improves effectiveness [7, 51, 52].

In carrying out hydrophobic profile investigation of spike protein of SARS-CoV-2, significant insights can be gained which might help in improving development of vaccine targets and also anti-viral cures. In carrying out hydrophobic profile investigation of spike protein of SARS-CoV-2, significant insights can be gained which might help in improving development of vaccine targets and also anti-viral cures [38]. Drug design and drug delivery are highly impacted by hydrophobicity. Estimation about hydrophobicity and examination of intricate details about hydrophobic profile is capable of giving a breakthrough for related aspects of drug designing and computational life sciences [1]. Computational aspects of hydrophobic profile of spike proteins of SARS-CoV-2 and SARS-CoV will aid the understanding at the molecular level which might give a further opportunity to investigate the treatment of diseases caused by them [49].
4 Conclusion

In this paper, the detailed hydrophobic profile for spike proteins of SARS-CoV-2 and SARS-CoV has been determined, and also the maximum and the minimum, aromatic, and aliphatic hydrophobic contributors have been identified. The sequences having close similarity with protein sequences of the spike proteins of SARS-CoV-2 and SARS-CoV have been analyzed. It can be concluded that a considerable amount of hydrophobic contribution is required for a stable and functional protein structure which in this study has been obtained majorly from phenylalanine for the aromatic category and leucine for the aliphatic category; histidine and methionine are minimum aromatic and aliphatic contributors, respectively. A focused investigation showed the contribution of particular hydrophobic amino acids in the different protein sequences obtained for the spike protein of SARS-CoV-2 and its comparison with the spike protein of SARS-CoV. Conservation of hydrophobic pattern has been indicated and some specific individual variations, too, were visible. The computational method will be a rapid and cost-effective approach for the early prediction of newly emerging viral variant impact at the molecular level which would probably benefit the scientific community.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10867-022-09615-x.

Author contribution Uma Shekhawat wrote the first draft of the manuscript. Anindita Roy Chowdhury supervised the work and reviewed the manuscript.

Declarations

Ethical approval This is a theoretical study. No ethical approval is required.

Informed consent Informed consent was not desirable as no living subject was involved in the study; moreover, open access databases were used and duly acknowledged.

Conflict of interest The authors declare no competing interests.

References

1. Sarkar, A., Kellogg, G.: Hydrophobicity - shake flasks, protein folding and drug discovery. CTMC 10, 67–83 (2010). https://doi.org/10.2174/15680261079023233
2. Simm, S., Einloft, J., Mirus, O., Schleiff, E.: 50 years of amino acid hydrophobicity scales: revisiting the capacity for peptide classification. Biol. Res. 49, 1–19 (2016). https://doi.org/10.1186/s40659-016-0092-5
3. Aydin, H., Al-Khooly, D., Lee, J.E.: Influence of hydrophobic and electrostatic residues on SARS-coronavirus S2 protein stability: insights into mechanisms of general viral fusion and inhibitor design: biophysical characterization of the SARS S2 helical core. Protein. Sci. 23, 603–617 (2014). https://doi.org/10.1002/pro.2442
4. Michael, M. Cox, D.L.N.: Lehninger Principles of Biochemistry. Sara Tenney (2008)
5. Wang, H.-J., Xi, X.-K., Kleinhammes, A., Wu, Y.: Temperature-induced hydrophobic-hydrophilic transition observed by water adsorption. Science 322, 80–83 (2008). https://doi.org/10.1126/science.1162412
6. Meyer, E.E., Rosenberg, K.J., Israelechvili, J.: Recent progress in understanding hydrophobic interactions. Proc. Natl. Acad. Sci. U.S.A. 103, 15739–15746 (2006)
7. Seyran, M., Takayama, K., Uversky, V.N., Lundstrom, K., Palù, G., Sherchan, S.P., Attrish, D., Rezaei, N., Aljabali, A.A.A., Ghosh, S., Pizzol, D., Chauhan, G., Adadi, P., Mohamed Abd El-Aziz, T., Soares, A.G., Kandimalla, R., Tambuwala, M., Hassan, S.S., Azad, G.K., Pal Choudhury, P., Baetas-da-Cruz, W., Serrano-Aroca, Á., Brufsky, A.M., Uhal, B.D.: The structural basis of
accelerated host cell entry by SARS-CoV-2†. FEBS J. 288, 5010–5020 (2021). https://doi.org/10.1111/febs.15651
8. Schultz, J.L., Aschenbrenner, A.C.: COVID-19 and the human innate immune system. Cell 184, 1671–1692 (2021). https://doi.org/10.1016/j.cell.2021.02.029
9. Huang, Y., Yang, C., Xu, X., Xu, W., Liu, S.: Structural and functional properties of SARS-CoV-2 spike protein: potential antiviral drug development for COVID-19. Acta. Pharmacol. Sin. 41, 1141–1149 (2020). https://doi.org/10.1038/s41401-020-0485-4
10. Mittal, A., Manjunath, K., Ranjan, R.K., Kaushik, S., Kumar, S., Verma, V.: COVID-19 pandemic: insights into structure, function, and hACE2 receptor recognition by SARS-CoV-2. PLoS Pathog. 16, 1–19 (2020). https://doi.org/10.1371/journal.ppat.1008762
11. Mousavizadedeh, L., Ghasemi, S.: Genotype and phenotype of COVID-19: their roles in pathogenesis. J. Microbiol. Immunol. Infect. 54, 159–163 (2021). https://doi.org/10.1016/j.jmii.2020.03.022
12. McIntosh, K.: Proving etiologic relationships to disease. Pediatr. Infect. Dis. J. 41, e102–e103 (2022). https://doi.org/10.1097/INF.0000000000003443
13. Watanabe, Y., Allen, J.D., Wrapp, D., McLeLLan, J.S., Crispin, M.: Site-specific glycans analysis of the SARS-CoV-2 spike. Science 369, 330–333 (2020). https://doi.org/10.1126/science.abb9983
14. Casalino, L., Gaieb, Z., Goldsmith, J.A., Hjorth, C.K., Dommer, A.C., Harbison, A.M., Fogarty, C.A., Barros, E.P., Taylor, B.C., McLeLLan, J.S., Fadda, E., Amaro, R.E.: Beyond shielding: the roles of glycans in the SARS-CoV-2 spike protein. ACS. Cent. Sci. 6, 1722–1734 (2020). https://doi.org/10.1021/acscentsci.0c01056
15. Watanabe, Y., Berndsen, Z.T., Raghwani, J., Seabright, G.E., Allen, J.D., Pybus, O.G., McLeLLan, J.S., Wilson, I.A., Bowden, T.A., Ward, A.B., Crispin, M.: Vulnerabilities in coronavirus glycan shields despite extensive glycosylation. Nat. Commun. 11, 2688 (2020). https://doi.org/10.1038/s41467-020-16567-0
16. Dill, K.A.: Dominant forces in protein folding. Biochemistry 29, 7133–7155 (1990). https://doi.org/10.1021/bi00483a001
17. Kumar, A., Singh, R., Kaur, J., Pandey, S., Sharma, V., Thakur, L., Sati, S., Mani, S., Asthana, S., Chandra, T.K., Chaudhuri, S., Bhattacharyya, S., Kumar, N.: Wuhan to world: the COVID-19 pandemic. Front. Cell. Infect. Microbiol. 11, 65010–65020 (2021). https://doi.org/10.3389/fcimb.2021.65010
18. Dill, K.A., Doolittle, R.F.: A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. 215, 403–410 (1990). https://doi.org/10.1016/S0022-2836(05)80360-2
19. J.B.: Basic local alignment search tool. Basic. Local. Alignment. Search. Tool (BLAST). J. Mol. Biol. 215, 403–410 (1990). https://doi.org/10.1016/S0022-2836(05)80360-2
20. McLeLLan, J.S., Wilson, I.A., Bowden, T.A., Ward, A.B., Crispin, M.: Vulnerabilities in coronavirus glycan shields despite extensive glycosylation. Nat. Commun. 11, 2688 (2020). https://doi.org/10.1038/s41467-020-16567-0
21. The UniProt Consortium: UniProt: a worldwide hub of protein knowledge. Nucleic. Acids. Res. 1111/2020. https://doi.org/10.1038/s41401-020-03526-w
22. The UniProt Consortium: UniProt: a worldwide hub of protein knowledge. Nucleic. Acids. Res. 47, D506–D515 (2019). https://doi.org/10.1038/s41401-020-03526-w
23. The UniProt Consortium: UniProt: a worldwide hub of protein knowledge. Nucleic. Acids. Res. 47, D506–D515 (2019). https://doi.org/10.1038/s41401-020-03526-w
30. Chowdhury, A.R., Kothari, A.: Hydrophobic mapping of chlorobium tepidum, the energy generating bacteria. J. Harmonized Res. Appl. Sci. 7, 98 (2019). https://doi.org/10.30876/johr.7.3.2019.98-106
31. Chowdhury, A.R., Seal, A., Nagendra, H.G.: Computational analysis of hydrophobicity across six enzyme classes revealing relative contribution of aliphatic and aromatic residues. Biotechnol. Bioinf. Bioeng. 1, 83–91 (2011)
32. Homaean, L., Kurgan, L.A., Ruan, J., Cios, K.J., Chen, K.: Prediction of protein secondary structure content for the twilight zone sequences. Proteins: Struct. Funct. Bioinformatics 69, 486–498 (2007). https://doi.org/10.1002/prot.21527
33. Margalit, H., Spouge, J.L., Cornette, J.L., Cease, K.B., Delisi, C., Berzofsky, J.A.: Prediction of immunodominant helper T cell antigenic sites from the primary sequence. J. Immunol. 138, 2213–2229 (1987)
34. Syngai, G.G., Barman, P., Bharali, R., Dey, S.: BLAST: An introductory tool for students to Bioinformatics Applications. Keanean J. Sci. 2, 67–76 (2013)
35. Eric, S.D., Nicholas, T.K.D.D., Theophilus, K.A.: Bioinformatics with basic local alignment search tool (BLAST) and fast alignment (FASTA). J. Bioinform. Seq. Anal. 6, 1–6 (2014). https://doi.org/10.5897/JBC2013.0086
36. Multiple Sequence Alignment: CLUSTALW. https://www.genome.jp/tools-bin/clustalw. Accessed 31 March 2021
37. Hung, J.-H., Weng, Z.: Sequence Alignment and Homology Search with BLAST and ClustalW. Cold Spring Harb. Protoc. 1–6 (2016). https://doi.org/10.1101/pdb.prot093088
38. Yerukala Sathipati, S., Shukla, S.K., Ho, S.-Y.: Tracking the amino acid changes of spike proteins across diverse host species of severe acute respiratory syndrome coronavirus 2. iScience 25, 103560 (2022). https://doi.org/10.1016/j.isci.2021.103560
39. Qiu, C., Whittaker, G.R., Gelman, S.H., Daniel, S., Abbott, N.L.: Characterization of hydrophobic interactions of SARS-CoV-2 and MERS-CoV spike protein fusion peptides using single molecule force measurements. (2022). https://www.bioxiv.org/content/10.1101/2022.03.05.483104v1. Accessed 16 June 2022
40. Zhu, C., Gao, Y., Li, H., Meng, S., Li, L., Francisco, J.S., Zeng, X.C.: Characterizing hydrophobicity of amino acid side chains in a protein environment via measuring contact angle of a water nanodroplet on planar peptide network. Proc. Natl. Acad. Sci. U.S.A. 113, 12946–12951 (2016). https://doi.org/10.1073/pnas.1616381113
41. Newell, P.D., Fricker, A.D., Roco, C.A., Chandransu, P., Merkel, S.M.: A small-group activity introducing the use and interpretation of BLAST. J. Microbiol. Biol. Educ. 14, 238–243 (2013). https://doi.org/10.1128/jmbe.v14i2.637
42. Kerfeld, C.A., Scott, K.M.: Using BLAST to teach “E-value-tionary” Concepts. PLoS Biol. 9, 1–4 (2011). https://doi.org/10.1371/journal.pbio.1001014
43. Zhou, P., Yang, X.-L., Wang, X.-G., Hu, B., Zhang, L., Zhang, W., Si, H.-R., Zhu, Y., Li, B., Huang, C.-L., Chen, H.-D., Chen, J., Luo, Y., Guo, H., Jiang, R.-D., Liu, M.-Q., Chen, Y., Shen, X.-R., Wang, X., Zheng, X.-S., Zhao, K., Chen, Q.-J., Deng, F., Liu, L.-L., Yan, B., Zhan, F.-X., Wang, Y.-Y., Xiao, G.-F., Shi, Z.-L.: A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature 579, 270–273 (2020). https://doi.org/10.1038/s41586-020-2012-7
44. Zheng, M., Song, L.: Novel antibody epitopes dominate the antigenicity of spike glycoprotein in SARS-CoV-2 compared to SARS-CoV. Cell. Mol. Immunol. 17, 536–538 (2020). https://doi.org/10.1038/s41423-020-0385-z
45. Robson, B.: Computers and viral diseases. Preliminary bioinformatics studies on the design of a synthetic vaccine and a preventative peptidomimetic antagonist against the SARS-CoV-2 (2019-nCoV, COVID-19) coronavirus. Comput. Biol. Med. 119, 103670 (2020). https://doi.org/10.1016/j.compbiomed.2020.103670
46. UCSF ChimeraX Home Page, https://www.rbvi.ucsf.edu/chimerax/. Accessed 15 September 2022
47. Goddard, T.D., Huang, C.C., Beng, E.C., Pettersen, E.F., Couch, G.S., Morris, J.H., Ferrin, T.E.: UCSF ChimeraX: Meeting modern challenges in visualization and analysis. Protein Sci. 27, 14–25 (2018). https://doi.org/10.1002/pro.3235
48. Pettersen, E.F., Goddard, T.D., Huang, C.C., Beng, E.C., Couch, G.S., Croll, T.I., Morris, J.H., Ferrin, T.E.: UCSF ChimeraX: Structure visualization for researchers, educators, and developers. Protein Sci. 30, 70–82 (2021). https://doi.org/10.1002/pro.3943
49. Kumar, S., Thambiraja, T.S., Karuppanan, K., Subramaniam, G.: Omicron and delta variant of SARS-CoV-2: a comparative computational study of spike protein. J. Med. Virol. 94, 1641–1649 (2022). https://doi.org/10.1002/jmv.27526
50. Brogi, S., Ramalho, T.C., Kuca, K., Medina-Franco, J.L., Valko, M.: Editorial: In silico methods for drug design and discovery. Front. Chem. 8, (2020)
51. Leung, M.Y.K., Cohen, F.S.: Increasing hydrophobicity of residues in an anti-HIV-1Env peptide synergistically improves potency. Biophys. J. 100, 1960–1968 (2011). https://doi.org/10.1016/j.bpj.2011.02.053

52. Zhou, X., Han, X., Zhao, K., Du, J., Evans, S.L., Wang, H., Li, P., Zheng, W., Rui, Y., Kang, J., Yu, X.-F.: Dispersed and conserved hydrophobic residues of HIV-1 Vif are essential for CBFβ recruitment and A3G suppression. J. Virol. 88, 2555–2563 (2014). https://doi.org/10.1128/JVI.03604-13

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.