

Analysis of Seven Human Respiratory Coronavirus (CoV) S Proteins from a Bioinformatics Approach

LEONARD WHYE KIT LIM* & HUNG HUI CHUNG

Faculty of Resource Science and Technology, Universiti Malaysia Sarawak, 94300 Kota Samarahan, Sarawak, Malaysia

*Corresponding author: limwhyekitleonard@gmail.com

Received: 12 July 2023

Accepted: 11 August 2023

Published: 31 December 2023

ABSTRACT

The coronavirus disease 2019 (COVID-19) has caused a huge pandemic repercussion across the globe and it is mainly contributed by the human severe acute respiratory syndrome coronavirus (SARS-CoV-2). There are seven human respiratory coronaviruses identified to date, namely HCoV-229E, HCoV-NL63, HCoV-OC43, HCoV-HKU1, MERS-CoV, SARS-CoV and SARS-CoV-2. A recently published bioinformatic human CoV comparison only covered four human CoV. Therefore, in this study, a bioinformatics approach-based analyses route was taken to dissect the S proteins of all the available (seven) human respiratory coronaviruses publicly available in the GenBank database. The antigenic epitope amount is postulated to be the most accurate bioindicator among all in determining the severity of a particular human respiratory coronavirus. Other powerful bioinformatic indicators are global similarity index, maximum likelihood phylogenetic analysis as well as domain analysis. The data generated in this study can be channelled to the vaccine and antiviral drug development to combat the current and future spread of the human respiratory coronaviruses.

Keywords: Antigenic epitope, COVID-19, SARS-CoV, S protein

Copyright: This is an open access article distributed under the terms of the CC-BY-NC-SA (Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License) which permits unrestricted use, distribution, and reproduction in any medium, for non-commercial purposes, provided the original work of the author(s) is properly cited.

INTRODUCTION

The human severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is the major contributor to coronavirus disease 2019 (COVID-19), which had caused unprecedented challenges worldwide with its rapid spread and severe impact on mortalities, public health system and economies globally (Lim, 2023). This single-stranded RNA-enveloped virus is categorised under the order Nidovirales and family Coronaviridae, whereby the family encompasses alpha-, beta-, gamma- and delta-coronavirus genera (Cicaloni *et al.*, 2022). The alpha-coronavirus can infect both animals and human whereas the beta-coronavirus such as SARS-CoV, HCoV-HKU1, MERS-CoV and SARS-CoV-2 infect mainly human. The remaining two genera primarily infect animals. This virus contains four structural proteins, namely spike (S), membrane (M), nucleocapsid (N) and envelope (E). The non-structural proteins found within the virus include ORF3a, ORF8, ORF7a, Nsp3, Nsp6, Nsp4 and Nsp12 (Cicaloni *et al.*, 2022).

The surface of the viral envelope is where the S proteins are situated and it can be divided into S1 and S2 regions based on their functionalities. The receptor binding domain (RBD) that functions in the binding of viral-receptor is housed within the S1 region whereas the homology region structural domains 1 and 2 (HR1 and HR2) responsible for viral fusion is located within the S2 region (Xia *et al.*, 2020). Attachment of S proteins to the receptors on the host cells lead to fusion of host cell and virus membranes, subsequently replicase will be translated and the viral replicase complex will be assembled. This made S protein a potential target for vaccine development (Xia *et al.*, 2020). A recently published bioinformatic human CoV comparison only covered four human CoV (Niu *et al.*, 2023), namely SARS-CoV-2, HCoV-HKU1, MERS as well as SARS. Therefore, there is a research gap as not all seven human CoV were included in the analysis. Only the four aforementioned human CoV were studied using the methods employed in this study, this creates a gap in knowledge in several viral aspects such as the antigenic epitope characteristics, similarity, physical and chemical properties as

some other human CoV such as HCoV-229E, HCoV-OC43 and HCoV-NL63 were not included by the previous studies. In this study, a bioinformatics modus operandi was applied to dissect the protein profiles of all seven human respiratory coronavirus S proteins publicly available in the GenBank database. These protein functional analyses will aid in better comprehension of S proteins for vaccine development, drug design and protein modification.

MATERIALS AND METHODS

Virus Sequence Data Consolidation

The S protein sequences of all seven available human coronavirus were downloaded from the public GenBank database on 31 March 2023. Multiple sequence alignment was done across S proteins of all seven human coronavirus. A model test was done using MEGA X (Kumar *et al.*, 2018) to determine the best protein model for phylogenetic tree plotting. The selected model was Whelan and Goldman with Frequency and Gamma distribution (WAG+F+G) model. A maximum likelihood phylogenetic tree was constructed using MEGA X (Kumar *et al.*, 2018) with 1,000 bootstrap replications. Figtree was used to improve the visualisation of the final phylogenetic tree.

Protein Profiling

The bioinformatic analyses utilised in this study were emulated from Niu *et al.* (2023), Lim *et al.* (2022a & 2022b) and Lim (2022). ProtParam (Gasteiger *et al.*, 2005) was used to analyse the chemical and physical properties of all S proteins. The hydrophobicity and protein affinity were examined using ExPASy-ProtScale. The S protein transmembrane investigation and signal peptide prediction were performed using TMHMM Server v.2.0 and SignalP v.4.1 respectively. The O-type glycosylation sites, protein epitopes, N-type glycosylation sites, phosphorylation sites, functional motifs, domain analysis and subcellular localisation were conducted employing YinOYang v.1.2, Predicted Antigenic Peptides, NetNGlyc v.1.0, NetPhos v.3.1 server, PROSITE, SMART and Euk-mPLoc 2.0 (Chou & Shen, 2010), correspondingly. Phyre2 (Kelley *et al.*, 2015) was used to predict protein 3D structures before the plotting of Ramachandran graphs.

RESULTS AND DISCUSSION

HCoV-NL63, HCoV-OC43 and HCoV-HKU1 S Proteins Dominate the Highest Ranks in Most of the Protein Physical and Chemical Properties

The general information of all S proteins was consolidated in Table 1. The HCoV-229E has the smallest genome size (27,317 bp), shortest S protein (1,171 aa) and lowest theoretical isoelectric point (5.64) among all seven human respiratory coronaviruses examined. The HCoV-OC43 has the largest genome size (30,738 bp), the longest S protein (1,361 aa), as well as the greatest number of hydrogen atoms (10,370) and sulphur atoms (76) among all seven human respiratory coronaviruses. The HCoV-HKU1 S protein was discovered to have the greatest number of atoms (21,033), molecular weight (151,710.91 Da), carbon atoms (6,826), as well as oxygen atoms (2,038). The greatest theoretical isoelectric point was seen in S protein of HCoV-NL63 (6.83). The SARS-CoV S protein has the greatest number of negatively charged amino acid residues (Asp + Glu) (113). The SARS-CoV-2 S protein has the greatest number of positively charged amino acid residues (Arg + Lys) (103). Positively-charged amino acids have been found to enhance the human respiratory coronaviruses fusion with the negatively-charged host cell receptors (Pawlowski, 2021). This explains for the highest severity, morbidity and mortality of SARS-CoV-2 across the globe when compared to other human respiratory coronaviruses.

The highest instability index of 36.78 was observed in MERS-CoV S protein, despite that, all human respiratory coronavirus S proteins are stable as only proteins with more than 40 instability indexes are considered unstable (Guruprasad *et al.*, 1990). Most human respiratory coronaviruses have negative value for their grand average of hydropathicity (GRAVY), except for HCoV-229E and HCoV-NL63 S proteins. The highest GRAVY value (0.157) and aliphatic index (93.24) both belongs to that of the HCoV-229E S protein. In other words, the HCoV-NL63 S protein is the most thermally stable (based on aliphatic value) and most polar (based on GRAVY value). The amino acid with the greatest hydrophilicity is mostly asparagine (in HCoV-229E, HCoV-OC43 and SARS-CoV-2 S proteins) and arginine (in

HCoV-HKU1 and SARS-CoV S proteins). The serine and tyrosine act as the highest hydrophilicity amino acids in HCoV-NL63 and MERS-CoV S proteins, respectively. The most frequently observed amino acid with the highest hydrophobicity are leucine and valine, seen in all human respiratory coronaviruses, except for HCoV-OC43 (phenylalanine) and SARS-CoV (threonine) S proteins. All seven human respiratory coronavirus S proteins contain N-terminal signal peptide. The hydrophobicity of amino acids influences the protein folding capability significantly whereas the signal peptide analysis is useful to provide preliminary clues to the subcellular localisation, function and

structure of the S proteins. An interesting study by Matyášek *et al.* (2021) postulated that the elevated hydrophobicity of the human respiratory coronavirus viral proteins (specifically the SARS-CoV-2) was influenced by the mutational asymmetries in the viral genome via a cascade of protein-protein interactions. Therefore, the hydrophobicity of human respiratory coronavirus S protein might be used as a strong preliminary bioindicator as soon as a new human respiratory coronavirus emerges to compare on the severity of the new strain to that of the existing strains. This will aid greatly in the formulation of contingent plans to combat the new viral threats.

Table 1. The S protein characteristics of all seven human respiratory coronaviruses

S protein characteristics	HCoV-229E	HCoV-NL63	HCoV-OC43	HCoV-HKU1	SARS-CoV	SARS-CoV-2	MERS-CoV
GenBank accession number	NC_002645.1	AY567487.2	AY391777.1	NC_006577.2	AY463060.1	NC_045512.2	KF958702.1
Genome size (bp)	27317	27553	30738	29926	29013	29903	29851
S protein gene position in genome (bp)	20570-24091	20472-24542	23644-27729	22942-27012	20834-24601	21563-25384	21450-25511
Amino acid length	1173	1356	1361	1356	1255	1273	1353
Formula	C ₅₇₇₄ H ₈₃₃₉ N ₁₄₈₃ O ₁₇₄₂ S ₅₁	C ₆₇₄₀ H ₁₀₃₆₈ N ₁₇₅₂ O ₂₀₀₅ S ₅₇	C ₆₇₆₀ H ₁₀₃₇₀ N ₁₇₄₀ O ₂₀₂₉ S ₇₆	C ₆₈₂₆ H ₁₀₃₅₃ N ₁₇₄₇ O ₂₀₃₈ S ₅₉	C ₆₂₄₈ H ₉₅₈₈ N ₁₆₁₀ O ₁₈₆₆ S ₅₉	C ₆₃₃₆ H ₉₇₇₀ N ₁₆₃₆ O ₁₈₉₀ S ₅₄	C ₆₆₈₃ H ₁₀₂₄₄ N ₁₇₃₆ O ₂₀₂₈ S ₆₃
Total number of atoms	17989	20922	20984	21033	19371	19710	20754
Molecular weight (Da)	128639.44	149850.42	151025.80	151710.91	139006.07	141178.47	149377.05
Theoretical isoelectric point (pI)	5.64	6.83	5.72	6.20	5.67	6.24	5.73
Negatively charged residues (Asp + Glu)	88	86	111	106	113	110	112
Positively charged residues (Arg + Lys)	77	84	99	95	99	103	95
Instability index (II)	29.71	30.81	33.96	36.43	31.91	33.01	36.78
Aliphatic index	93.24	93.10	85.72	84.72	82.88	84.67	82.71
Grand average of hydropathicity (GRAVY)	0.157	0.116	-0.023	-0.025	-0.038	-0.079	-0.074
Amino acid with the highest hydrophilicity	842 Asparagine (2.011)	750 Serine (2.567)	761 Asparagine (3.033)	757 Arginine (2.789)	758 Arginine (2.822)	679 Asparagine (2.589)	540 Tyrosine (2.822)
Amino acid with the highest hydrophobicity	1126 Leucine (3.511)	1305 Valine & 1306 Valine (3.511)	1324 Phenylalanine (3.411)	1315 Leucine, 1316 Valine, & 1317 Leucine (3.667)	1213 Threonine (3.189)	7 Leucine (3.222)	1314 Valine (3.333)
Number of predicted serine phosphorylation sites	57	73	57	104	67	68	88
Number of predicted threonine phosphorylation sites	44	38	39	35	43	45	41
Number of predicted tyrosine phosphorylation sites	25	30	26	23	21	23	29
Potential N-type glycosylation sites	30	39	22	29	24	22	25

SARS-CoV, SARS-CoV-2 and MERS-CoV S Proteins Tend to Have More Antigenic Epitopes Than the Others

The highest number of predicted serine phosphorylation was unearthed in HCoV-HKU1 S protein (104). The SARS-CoV S protein has the greatest predicted threonine phosphorylation sites (45) whereas the HCoV-NL63 S protein encompasses the most amount of predicted

tyrosine phosphorylation sites (30). The greatest number of N-type glycosylation sites across all seven human respiratory coronaviruses is 39 in this study, and it was revealed in the HCoV-NL63 S protein. The HCoV-HKU1 S protein houses the most O-type glycosylation sites (243). Phosphorylation and glycosylation are two of the most essential post-translational modification mechanisms of viral proteins (Fung & Liu, 2018). Phosphorylation orchestrates

protein functionality, cellular signalling and protein viability whereas glycosylation regulates protein stability, immune evasion (by shielding viral antigenic epitopes from immune surveillance) as well as virus binding and entry to host cells (Reis *et al.*, 2021; Chatterjee & Thakur, 2022). The number of phosphorylation and glycosylation sites do not directly reflect the possible severity of a human respiratory coronavirus, as the most severe one (SARS-CoV-2) does not have the greatest number of sites among all seven human respiratory coronavirus S proteins. Thus, we look further into the number of antigenic epitopes.

Interestingly, the SARS-CoV, SARS-CoV-2 and MERS-CoV S proteins tend to have more antigenic epitopes (61, 63 and 60, correspondingly) as compared to the others (ranging from 47 to 60). The number of antigenic epitopes might be a better and more sensitive bioindicator to the measure of severity among these viruses as the top three most severe human respiratory coronaviruses (SARS-CoV, SARS-CoV-2 and MERS-CoV) S proteins were discovered to have more antigenic epitopes as compared to the others. These antigenic epitopes are powerful biomarkers for the measure of

potential immune escape mutations (PIEMs). For instance, the delta variant of SARS-CoV-2 was found to have the highest PIEMs, and this poses extreme predicaments in designing vaccines to effectively combat this variant (Jaiswal & Lee, 2022). The maximum likelihood phylogenetic tree encompassing all seven human respiratory coronaviruses revealed three distinctive clades (Figure 1). The first clade (Clade 1) houses the most members of the human respiratory coronaviruses (3), namely HCoV-OC43, HCoV-HKU1 and MERS-CoV S proteins. The SARS-CoV and SARS-CoV-2 S proteins shared the same clade (Clade 2) whereas the remaining two human respiratory coronavirus counterparts (HCoV-229E and HCoV-NL63 S proteins) reside the last clade (Clade 3). The phylogenetic tree does not provide direct information on the severity of the human respiratory coronaviruses but instead it provides clues on how closely related in terms of sequences between these viruses. For example, both SARS-CoV S proteins reside the same clade and if there is a new human respiratory coronavirus that are assigned into the same clade as them, it should be treated as high priority alert as both SARS-CoVs.

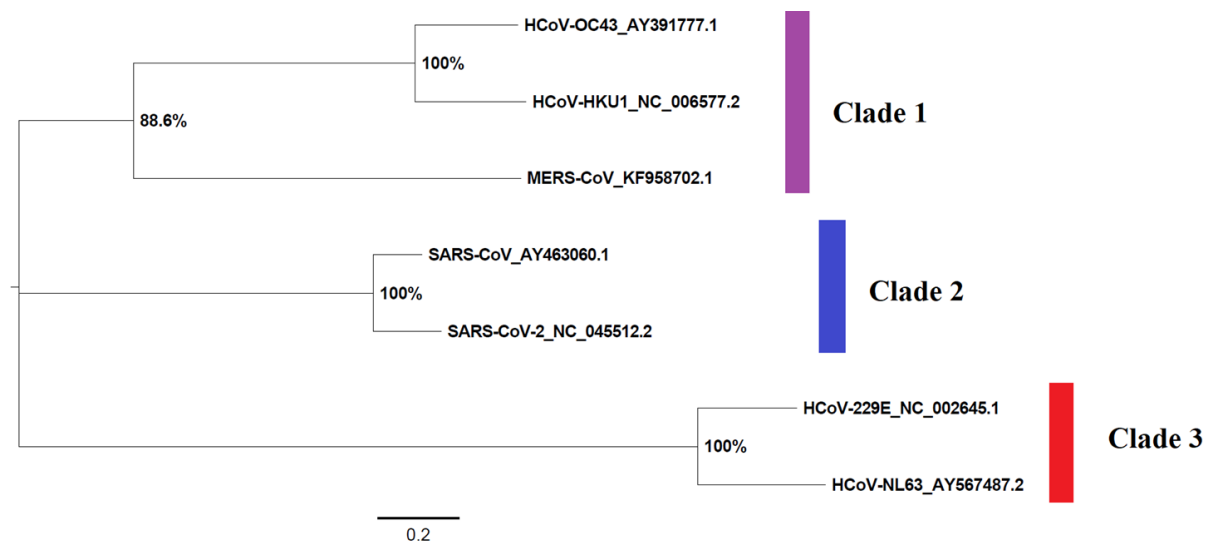


Figure 1. The maximum likelihood phylogenetic tree of all seven human respiratory coronaviruses

Corona_S2 Domain is Present in All Seven Human Respiratory Coronaviruses

The HCoV-229E and HCoV-NL63 S proteins do not have S1-amino terminal domain (N-terminal domain or NTD) and S1-carboxyl terminal domain (C-terminal domain or CTD) within their S protein sequences. The MERS-CoV S

protein have the lengthiest S1-NTD of 334 amino acid. The SARS-CoV and SARS-CoV-2 S proteins shared an almost similar length of S1-CTD (193 and 194 aa, respectively). The S2-homology region domain (HR) 1 was identical across HCoV-229E and HCoV-NL63 S proteins (120 aa), as well as the remaining five human respiratory coronaviruses (106 aa). The S2-HR2

domain has similar length in both HCoV-229E and HCoV-NL63 S proteins (97 aa) and this domain is highly conserved across the other five human respiratory coronaviruses (ranging between 82 to 84 aa). The corona_S1 domain is found only in HCoV-229E and HCoV-NL63 S proteins whereas the corona_S2 domain can be seen in all seven human respiratory coronaviruses with variable lengths. Interestingly, the corona_S2 domain length is identical in both SARS-CoV-2 and MERS-CoV S proteins. The spike-NTD domain was discovered in HCoV-OC43, HCoV-HKU1 and SARS-CoV S proteins with varying lengths. The spike-rec-bind functional domain was unravelled in all human respiratory coronaviruses with the exception to HCoV-229E and HCoV-NL63 S proteins, with distinctive lengths. The S1 and spike-NTD domains function in essential recognition and binding of the viral receptor-binding domain to the host receptor angiotensin-converting enzyme 2 (ACE2) whereas the S2 and spike-rec-bind domains facilitate cell entry and the cell-to-cell transmission that follows thereafter (Huang *et al.*, 2020; Xia, 2021). The lack of S1-NTD domain in S proteins of HCoV-229E and HCoV-NL63 may account for the less severity of these two human respiratory coronaviruses as compared to the others examined in this study.

Predicted Protein Structures of SARS-CoV, SARS-CoV-2 and MERS-CoV are Highly Similar

The subcellular localisation of S proteins of all seven human respiratory coronaviruses is identical, namely extracellular. The HCoV-NL63 S protein has additional predicted subcellular localisations such as cell membrane and endoplasmic reticulum. The HCoV-229E and HCoV-NL63 S proteins are different from one another and also from the others in the aspects of protein structure and number of alpha helices and beta sheets (Figure 2). The HCoV-OC43 and HCoV-HKU1 S proteins have highly identical protein structures as well as number of alpha helices and beta sheets (namely 10 – 12 and 73 – 75). The predicted protein structures of SARS-CoV, SARS-CoV-2 and MERS-CoV S proteins are highly similar, the number of alpha helices and beta sheets are also highly conserved across

these three human respiratory coronaviruses (11 – 12 and 65 – 67, correspondingly). The transmembrane helices prediction was also conducted across all seven human respiratory coronavirus S proteins. As a result, the SARS-CoV and SARS-CoV-2 S proteins shared the top rank in having the most transmembrane helices. The transmembrane regions are indispensable regions as primary drug targets in the drug development field of study as these regions are pivotal in determining the purification protocol and expression patterns of the viral proteins (Niu *et al.*, 2023).

The Ramachandran graphs of all seven human respiratory coronavirus S proteins displayed a sign of strong protein secondary structure in general. The least stable human respiratory coronavirus S protein belongs to that of the HCoV-HKU1 as the amino acid residues found within the generously allowed and disallowed regions are the highest among all others (2.7% and 1.3%, correspondingly). The second least stable S protein is from that of MERS-CoV with 1.9% and 0.5% of its amino acid residues situated within the generously allowed and disallowed regions. All other human respiratory coronaviruses showed high stability in structures of both alpha helices and beta sheets. The global similarity index (BLOSUM62) results depicted the similarity values across all seven human respiratory coronaviruses examined in this study. The SARS-CoV and SARS-CoV-2 S proteins shared the highest similarity value (0.77). The S proteins of both the HCoV-229E-HCoV-NL63 pair as well as HCoV-OC43-HCoV-HKU1 pair have the second highest similarity index of approximately 0.63. Placing focus onto the recently emerged SARS-CoV-2, its S protein is the most similar to that of SARS-CoV (0.77 similarity index), followed by the trio (HCoV-OC43, HCoV-HKU1 and MERS-CoV S proteins, namely 0.16 similarity index). The SARS-CoV-2 S protein is most distantly varied to that of HCoV-229E and HCoV-NL63 with similarity index of only 0.07. Similar to the phylogenetic analysis done in this study, the global similarity index is useful especially in important decision making in the selection of the best combat strategies to new human respiratory coronaviruses that emerge in the near future.

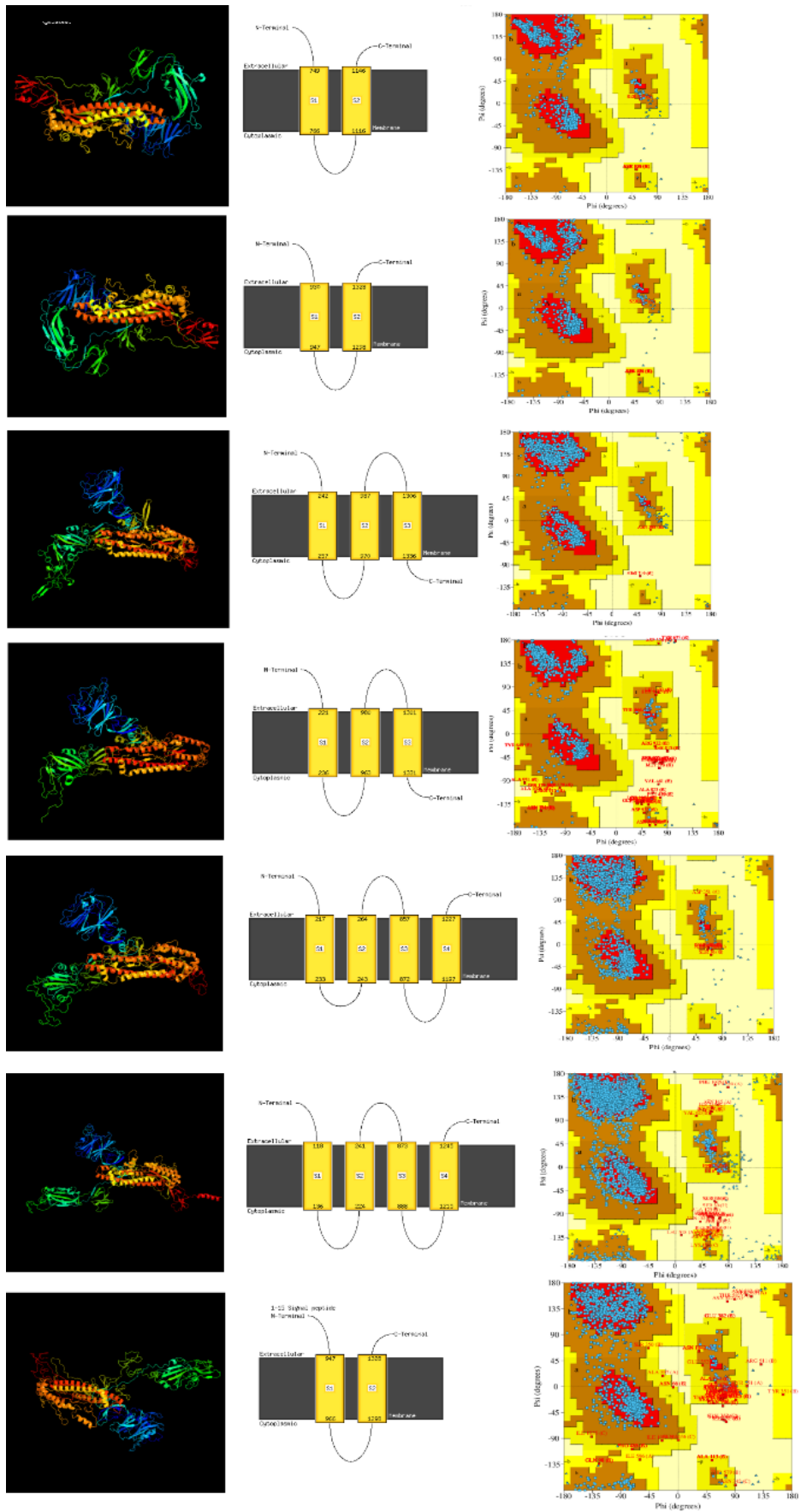


Figure 2. From top to bottom row: HCoV-229E, HCoV-NL63, HCoV-OC43, HCoV-HKU1, SARS-CoV, SARS-CoV-2 and MERS-CoV. From left to right column: predicted three-dimensional protein structure, transmembrane helix prediction, as well as Ramachandran plot

CONCLUSION

The current study presented a bioinformatic modus operandi analyses of both functions and structure of all seven human respiratory coronavirus S proteins. There are several key findings generated and deduced from this study. To name a few, the HCoV-NL63, HCoV-OC43 and HCoV-HKU1 S proteins dominate the highest ranks in most of the protein physical and chemical properties. SARS-CoV, SARS-CoV-2 and MERS-CoV S proteins tend to have more antigenic epitopes than the others. Interestingly, corona_S2 domain is present in all seven human respiratory coronaviruses. Predicted protein structures of SARS-CoV, SARS-CoV-2 and MERS-CoV are highly similar.

The data generated from this study are beneficial in enriching the human respiratory coronavirus database. In the long run, these data on S proteins will be channelled for the development and screening of coronavirus vaccines and antiviral drugs. However, this study has its limitations as it only covered the bioinformatic prediction tools, experimental validations are necessary to be conducted in the future. One of the major challenges is the lack of disclosure of the actual COVID-19 data and resources from certain countries and research institutions worldwide since the recent pandemic. It is envisioned that, with the advancement and reducing costs of next generation sequencing, individualised and personalised vaccines and drugs can be made catered closely to the patient's cell receptor proteins and S proteins of the coronavirus that infects the host, to reduce adverse side effects and achieve the greatest treatment efficacy.

ACKNOWLEDGEMENTS

This work was fully funded by The Ministry of Higher Education, Malaysia through the Fundamental Research Grant Scheme with grant number FRGS/1/2022/STG01/UNIMAS/02/2 awarded to H. H. Chung.

REFERENCES

- Chou, K.C. & Shen, H.B. (2010). A new method for predicting the subcellular localization of eukaryotic proteins with both single and multiple sites: Euk-mPLoc 2.0. *PLoS ONE*, 5: e9931. DOI: 10.1371/journal.pone.0009931
- Cicaloni, V., Costanti, F., Pasqui, A., Bianchini, M., Niccolai, N. & Bongini, P. (2022). A bioinformatics approach to investigate structural and non-structural proteins in human coronavirus. *Frontiers in Genetics*, 13: 891418. DOI: 10.3389/fgene.2022.891418
- Chatterjee, B. & Thakur, S.S. (2022). SARS-CoV-2 infection triggers phosphorylation: Potential target for anti-COVID-19 therapeutics. *Frontiers in Immunology*, 13: 829474. DOI: 10.3389/fimmu.2022.829474
- Fung, T.S., & Liu, D.X. (2018). Post-translational modifications of coronavirus proteins: Roles and function. *Future Virology*, 13(6): 405-430. DOI: 10.2217/fvl-2018-0008
- Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M.R., Appel, R.D. & Bairoch, A. (2005). Protein identification and analysis tools on the ExPASy server. In Walker, J.M. (Ed.), *The Proteomics Protocols Handbook*. Humana Press, New Jersey, pp. 571-607.
- Guruprasad, K., Reddy, B.V.B. & Pandit, M.W. (1990). Correlation between stability of a protein and its dipeptide composition: A novel approach for predicting *in vivo* stability of a protein from its primary sequence. *Protein Engineering, Design and Selection*, 4(2): 155-161. DOI: 10.1093/protein/4.2.155.
- Huang, Y., Yang, C., Xu, X.F., Xu, W. & Liu, S. (2020). Structural and functional properties of SARS-CoV-2 spike protein: Potential antiviral drug development for COVID-19. *Acta Pharmacologica Sinica*, 41: 1141-1149. DOI: 0.1038/s41401-020-0485-4
- Jaiswal, V.J. & Lee, H.J. (2022). Conservation and evolution of antigenic determinants of SARS-CoV-2: An insight for immune escape and vaccine design. *Frontiers in Immunology*, 13: 832106. DOI: 10.3389/fimmu.2022.832106.
- Kelley, L.A., Mezulis, S., Yates, C.M., Wass, M.N. & Sternberg, M.J.E. (2015). The Phyre2 web portal for protein modelling, prediction and analysis. *Nature Protocols*, 10(6): 845-858. DOI: 10.1038/nprot.2015.053
- Kumar, S., Stecher, G., Li, M., Knyaz, C. & Tamura, K. (2018). MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms. *Molecular Biology and Evolution*, 35(6): 1547-1549. DOI: 10.1093/molbev/msy096.

- Lim, L.W.K. (2023). Blended learning in animal biotechnology during pre-COVID-19, COVID-19 and post COVID-19 recovery phase periods across the globe: A step forward or backward. *International Journal of Zoology and Animal Biology*, 6(2): 1-5. DOI: 10.23880/izab-16000451
- Lim, L.W.K. (2022). Comparative genomic analysis reveals the origin and global distribution of melon necrotic virus isolates. *Gene Reports*, 29: 101685. DOI: 10.1016/j.genrep.2022.101685
- Lim, L.W.K., Hung, I.M. & Chung, H.H. (2022a). Cucumber mosaic virus: Global genome comparison and beyond. *Malaysian Journal of Microbiology*, 18(1): 79-92. DOI: 10.21161/mjm.211276
- Lim, L.W.K., Liew, J.X. & Chung, H.H. (2022b). Piper yellow mottle virus: A deep dive into the genome. *Gene Reports*, 29: 101680. DOI: 10.1016/j.genrep.2022.101680
- Matyášek, R., Řehůrková, K., Marošíová, K.B. & Kovařík, A. (2021). Mutational asymmetries in the SARS-CoV-2 genome may lead to increased hydrophobicity of virus proteins. *Genes*, 12(6): 826. DOI: 10.3390/genes12060826
- Niu, Z., Xu, S., Zhang, J., Zou, Z., Ren, L., Liu, X., Zhang, S., Zou, H., Hu, X., Wang, J., Zhang, L. & Zhou, Y. (2023). Bioinformatic analysis of the S protein of human respiratory coronavirus. *Molecular Phylogenetics and Evolution*, 181: 107704. DOI: 10.1016/j.ympev.2023.107704
- Pawlowski, P.H. (2021). Charged amino acids may promote coronavirus SARS-CoV-2 fusion with the host cell. *AIMS Biophysics*, 8(1): 111-120. DOI: 10.3934/biophy.2021008
- Reis, C.A., Tauber, R. & Blanchard, V. (2021). Glycosylation is a key in SARS-CoV-2 infection. *Journal of Molecular Medicine*, 99(8): 1023-1031. DOI: 10.1007/s00109-021-02092-0
- Xia, S., Zhu, Y., Liu, M., Lan, Q., Xu, W., Wu, Y., Ying, T., Liu, S., Shi, Z., Jiang, S. & Lu, L. (2020). Fusion mechanism of 2019-nCoV and fusion inhibitors targeting HR1 domain in spike protein. *Cellular & Molecular Immunology*, 17: 765-767. DOI: 10.1038/s41423-020-0374-2
- Xia, X. (2021). Domains and functions of spike protein in SARS-CoV-2 in the context of vaccine design. *Viruses*, 13(1): 109. DOI: 10.3390/v13010109.