Large-scale analysis of SARS-CoV2 spike protein variants and their impacts in Thailand

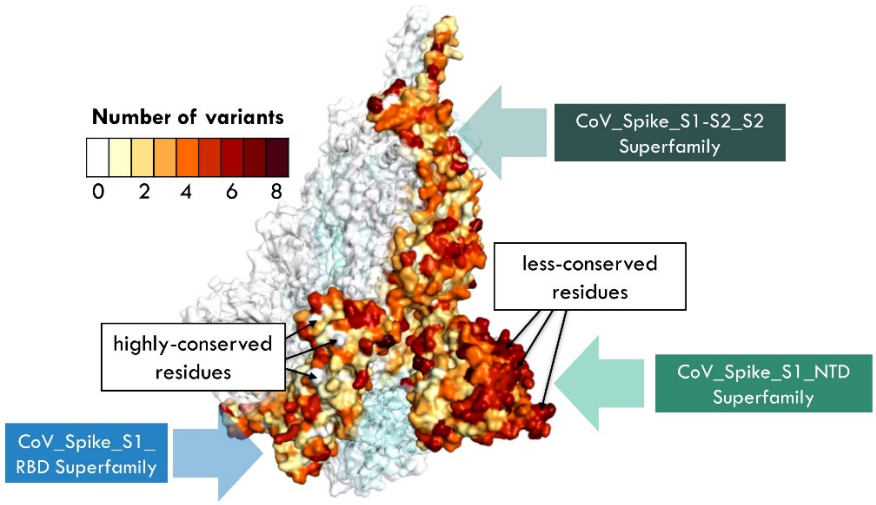
Shalip Yahangkiakan and Sirawit Ittisoponpisan *

Center for Genomics and Bioinformatics Research, Division of Biological Science, Faculty of Science, Prince of Songkla University, Songkhla, Thailand, 90112

* Correspondence: sirawit.i@psu.ac.th;

Abstract: With a large number of SARS-CoV2 sequences collected since the beginning of the COVID-19 outbreak, many genetic variations were yet to be analysed for their impacts. Moreover, little has been known about the effects of variants specific to Thai strains. In this study, we performed *in silico* analysis on 439,197 SARS-CoV2 spike protein sequences collected until 7 February 2021 from NCBI and GISAID databases to explore the distribution of variants on the SARS-CoV2 spike protein and their impacts. We identified mutation hotspots on the protein surface and several highly conserved residues located in the receptor-binding domain. These highly conserved residues could potentially serve as drug target candidates. Finally, we studied the effects of the variant of global concern Asp614Gly (D614G) and six variants exclusively found in Thailand (from 595 samples). Using 3D simulation, we showed that D614G disrupted inter-chain H bonds in the spike protein complex, while most Thai variants were structurally benign. The results from this study explain the mechanisms of variants observed in different strains of SARS-CoV2, aid variant prioritisation, and could provide implications for developing effective treatments or preventions for COVID-19.

Graphical abstract:



Keywords: coronavirus; SARS coronavirus 2; variant analysis; mutagenesis; spike protein; ACE2

1. Introduction

The outbreak of SARS-CoV2 has caused severe consequences to global health and economics since the beginning of 2020. As of February 2021, there had been reports of over a hundred million cases worldwide. Over 20,000 cases were reported in Thailand as a result of two major outbreaks. The fatality rates were ~3% globally and ~0.3% in Thailand [1, 2]. Despite having low fatality rates, SARS-CoV2 surpassed its coronavirus relatives such as SARS-CoV in 2002 and MERS-CoV in 2012 in terms of transmissibility [3, 4].

Many of the SARS-CoV2 sequences collected from patients worldwide had been made freely available through two databases NCBI [5] and GISAID [6], making it possible to perform *in silico* analysis on a large scale. The pre-screening for candidate sequences or variants is of particular importance as it significantly reduces the time and cost required for laboratory experiments. In light of the COVID-19 outbreak, most of the analyses were performed on variants found in Europe and America, where the infections were more severe, and the fatality rates were higher than the rest of the world. Spike protein is one of the most widely studied structural proteins of SARS-CoV2 as it binds directly to human ACE2 in order to invade the host cell. Unfortunately, studies on variants found locally in Thailand, where the first infection outside China was found, were at very limited availability [7–10]. Because of this, the mechanisms of many variants of the virus in Thailand remain to be elucidated.

In this study, we aim to provide more insights into the structures of SARS-CoV2 spike protein and its variants - with a particular focus on Thailand strains from the two major outbreaks collected until February 2021. We first explored the distribution of variants found on the SARS-CoV2 spike protein to discover mutation hotspots or any highly conserved residues which can be alternative drug-target candidates. Next, we analysed the effects of amino acid variants of global concern D614G and variants that were found exclusively in Thailand on spike protein complexes and the binding between spike protein and human ACE2. Insights into structural information of SARS-CoV2 could shed light on the viral mechanisms, explain the different transmissibility observed across different strains of SARS-CoV2, and guide molecular biologists and clinical researchers to develop effective treatments or preventions for COVID-19.

2. Materials and Methods

2.1 Dataset

2.1.1 SARS-CoV2 spike protein sequences

The sequences of SARS-CoV2 spike protein were retrieved on 7 February 2021 from NCBI (https://www.ncbi.nlm.nih.gov/sars-cov-2/) and GISAID (https://www.gisaid.org/). To facilitate the analysis, we extracted only the sequences of 1,273 amino acids, which is at the same length as the original Wuhan reference sequence (YP_009724390.1), from the NCBI database. However, in the GISAID database, the sequence length was set to 1,274 amino acids due to an extra asterisk (*) at the end of the sequence. In total, 53,056 sequences were obtained from NCBI (including the reference sequence YP_009724390.1) and 386,176 sequences from GISAID.

All sequences were then further screened by aligning with the reference sequence. Any sequence with less than 98% identity or more than 3 consecutive amino acid variants was removed. This is to minimise the possibility of the sequence containing insertion/deletion mutations. As a result, we removed six sequences from NCBI and 29 sequences from GISAID. The final dataset consisted of 53,050 sequences from NCBI and 386,147 from GISAID (439,197 total). 595 of these were collected in Thailand.

2.1.2 PDB structures of spike proteins

The 3D coordinate files used in this study were 6XR8 (resolution: 2.90 Å, coverage: 100%, released: 2020-07-22) for the SARS-CoV2 spike protein trimer at its inactive conformation and 6M0J (resolution: 2.45 Å, coverage: ~18%, released: 2020-03-18) for the spike receptor-binding domain (its active conformation) / human ACE2 complex.

2.2 Identification of buried, surface and interface residues

We used DSSP to calculated the solvent-accessible surface area (ASA) of each amino acid in a given PDB coordinate [11]. Next, we calculated the relative solvent accessibility (RSA) for each residue using the formula RSA = ASA/maxASA. The values for maximum solvent accessible surface area (maxASA) were determined by Rost and Sander [12]. Any residue is regarded as a "surface residue" when its RSA is ≥ 9 %, otherwise "buried".

Interface residues are defined as any surface residues that are within 4 Angstrom from any other residues of a different chain. In this study, the calculation of interface residues was performed in two structures: 1) the spike protein trimer (PDB: 6XR8) in order to determine the interface between identical chains (A-B and A-C), and 2) the spike-ACE2 complex (PDB: 6M0J) in order to determine the interface between the receptor-binding domain and the human ACE2. Full details of surface and interface residues are provided in Supplementary Table S1.

2.3 In silico analysis of variants on spike protein and ACE2-spike complex

Missense 3D [13], which was proven to be effective at detecting structural impact from amino acid variants on protein structures, was used to identify structural consequences of amino acid variants. Variants were simulated on PDB 6XR8. Further simulations were carried out on PDB 6M0J if variants were near the receptor-binding domain.

2.4 Statistical test

The two-tailed t-test for comparing means of two independent samples was used to determine any significant difference in the mean numbers of variants on surface residues vs buried residues and surface-interface residues vs surface non-interface residues. The test assumes equal population variances. A test result is regarded as significant when $P < 0.01$.

3. Results

3.1 Distribution of amino acid variants on SARS-COV2 spike complex.

In this part, we compared 439,196 sequences in our dataset with the Wuhan reference sequence (YP_009724390.1) and identified 4,040 amino acid variants, of which 3,412 (84.45%) could be mapped onto the spike complex structure (PDB: 6XR8). The rest were unmappable due to missing residues in the PDB coordinate (Supplementary Video S1).

There were 1,107 mappable residues in each spike protein chain. 326 residues were classified as "buried" and 781 residues as "surface" (of which 536 were further categorised into "surface & non-interface" and 245 into "surface & interface"). After mapping 3,412 amino acid variants, 781 of these variants were found on buried residues, while 2,631 were on surface residues. The average numbers of variants per buried residue vs surface residue were 2.40 and 3.37 ($P < 0.01$) (Figure 1a), suggesting that variants tend to be found more on surface residues than buried residues. Additionally, when considering "surface & non-interface" vs "surface & interface", we found that interface residues were less enriched in variants than non-interface residues. The average number of variants per residue were 2.40 and 3.37, respectively ($P < 0.01$) (Figure 1b).

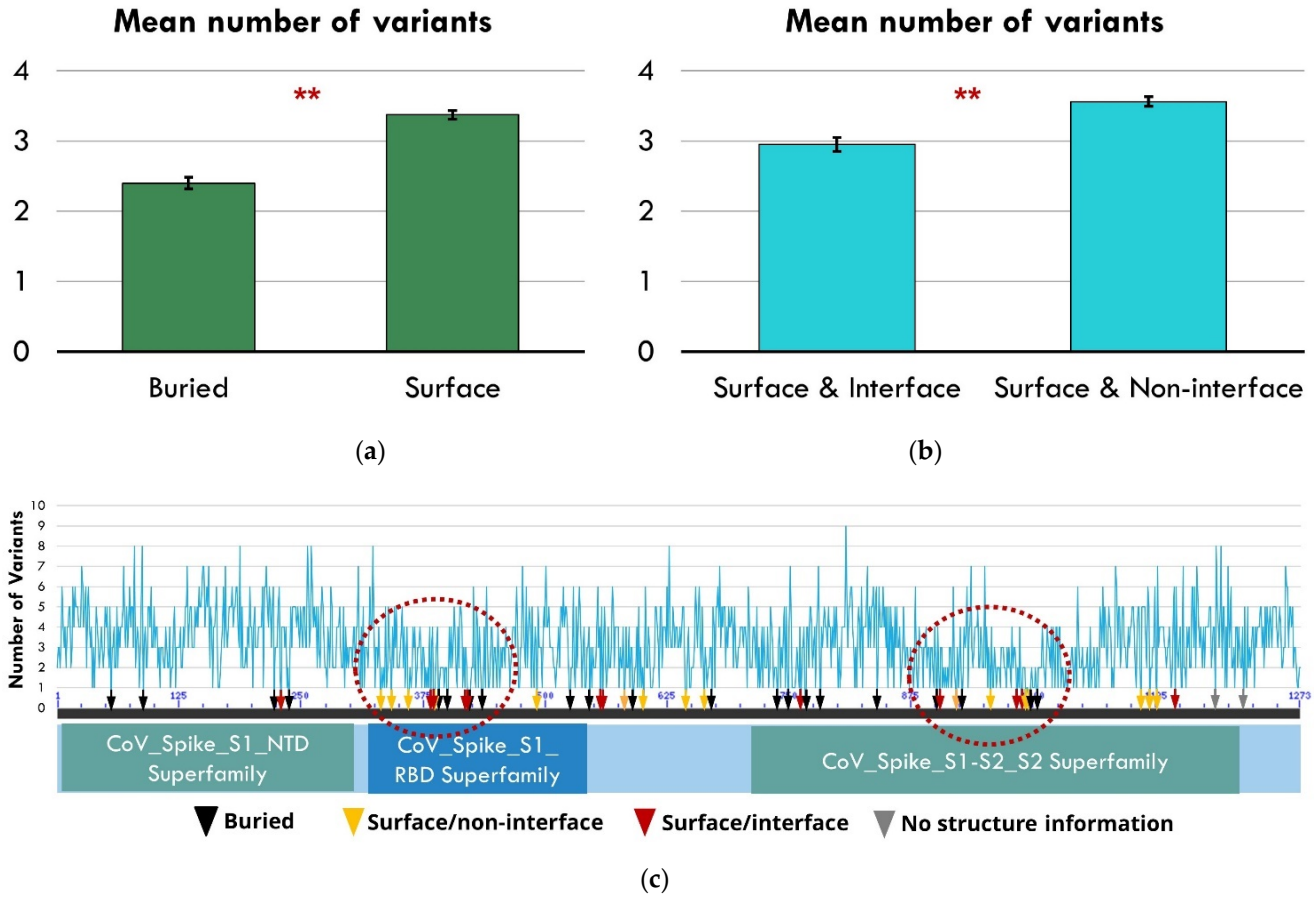


Figure 1. Comparisons of the mean number of variants per residue and the locations of highly conserved residues on the spike protein: (**a**) buried residues vs surface residues; (**b**) surface & interface residues vs surface & non-interface residues. Error bars represent the standard error of the mean (SEM). ** indicates a significant difference ($p < 0.01$) between the means using a two-tailed t-test; **(c)** distribution of highly conserved residues. Arrows represent locations where no variants were found. Colour boxes depict conserved domains identified by BLAST search (using the Wuhan reference sequence YP_009724390.1). Regions where conserved residues clustered are indicated by dotted circles.

When the conserved domain information from NCBI (based on Wuhan reference sequence) was considered, the NTD superfamily domain showed many mutation hotspots, suggesting that variants in this region were mostly tolerated, while the conserved residues were found more predominant in the RBD and the S2 superfamily domains (Figure 1c). No variants were found in 51 residues of the spike protein (Supplementary Table S2). Notably, many of these highly conserved residues were on the protein surface or served as interface residues.

3.2 Analysis of variants found in Thailand and the variant of global concern D614G

A total of 52 variants, including the variants of global concern Asp614Gly, were detected in 595 sequences collected in Thailand (Leu5Phe, Ser12Phe, Val47Ala, Thr51Asn^{T}, Gln52Lys, Leu54Phe, Phe55Leu, Pro57Thr^{T}, Ile68Arg^{T}, His69Tyr, Val70Leu, Ser71Pro, Gly75Val, Ser98Phe, Asp138His, Met153Ile, Met153Thr, Ala163Val, Gly181Val, Arg190Ser, Arg190Lys, Asp198Gly, Ser205Thr^{T}, Arg214Pro, Ala222Val, Lys278Thr^{T}, Ile358Thr, Ser359Asn, Val407Ala, Ser459Phe, Ser477Asn, Val483Phe, Glu484Asp, Asn501Thr, Lys537Arg, Asp614Gly*, Gln675His, Ala688Pro, Val785Leu, Thr791Ile, Ala829Thr, Gly832Cys^{T}, Phe855Leu, Asp936Tyr, Ser939Phe, Asp1118Tyr, Ser1147Leu, Pro1162Leu, Val1176Phe, Glu1202Gln, Gly1219Cys, and Asp1260Tyr). 6 of these (indicated as ^{T}) were not found elsewhere and were analysed in Table 1.

Table 1. Analysis of variants found exclusively in Thailand and the variant of global concern D614G (indicated by *)

Position	WT AA	MT AA	Location on Protein	Freq. (%)	Structural consequences
51	THR	ASN	surface	0.17	H-bond formed between Asn51 - His49 (no damage)
57	PRO	THR	buried	0.17	H-bond formed between Thr57- Gln271 (no damage)
68	ILE	ARG	surface	0.34	No structural damage
205	SER	THR	surface	0.17	H-bond formed between Thr205-Glu191 Cavity contraction of 213.624 $Å^3$, **Structural damage detected**
278	LYS	THR	surface	0.17	H-bond formed between Thr278-Thr286 (no damage)
614*	ASP	GLY	interface	48.74	H-bond disrupted between Asp614 (chain A) - Lys835 (chain B) and Asp614 (chain A) - Lys854 (chain B), Buried charge replaced, **Structural damage detected**
832	GLY	CYS	interface	0.34	No structural damage

None of the six Thailand-specific variants was on the receptor-binding domain; hence, no direct effect in the binding association between the spike protein and the human ACE2 can be anticipated. No structural damage was found in five variants (Table 1). Many Thailand-specific variants resulted in the formation of an H-bond between nearby amino acids, which might, in turn, slightly enhance the rigidity of the protein fold. Ser205Thr is the only variation that was predicted deleterious as it led to a surface cavity alteration. Previous studies have shown that cavity alterations tend to destabilise the protein fold [14–16]. Therefore, this variant is likely to be deleterious to the virus. Gly832Cys was found on the protein interface. Nevertheless, the variants were predicted to be structurally neutral. Unfortunately, with limited studies on Thailand-specific variants, not much is known about the clinical significance of these variants. Recently, Ile68Arg was discovered in a SARS-CoV2 strain that was reported to be associated with the ability to escape antibody [17]. However, the experiment was performed in mice, and this variant was due to mutation in the laboratory. It is still unclear whether Ile68Arg actually contributed to such an ability. The fact that all of the Thailand-specific variants were rare (found in < 1% of the samples collected) and not in the receptor-binding domain could imply that they are unlikely to affect the transmissibility and the human-ACE2 binding affinity.

In contrast, Asp614Gly (D614G) was found in 48.74% of the Thai samples. This variant was reported to be about 20% more transmissible [18]. As this mutation is not on the receptor-binding domain, it is unlikely to directly affect the binding affinity between the spike and the human-ACE2. This variant, when simulated on chain A, leads to the loss of two interchain H-bonds formed between Asp614 and the other two residues, Lys835 and Lys854, on chain B (Figure 2). This is likely to cause the chains to be less rigid when folded. This finding is in agreement with a study by Yurkovetskiy et al. using cryo-electron microscopy, which showed that the variant D614G disrupts an interprotomer contact, making the spike complex favour "open conformation", mimicking the conformation when bound to ACE2, more than the "closed conformation". Although it has been shown that only one single chain is required to be at the open conformation in order to bind to the human ACE2 and trigger the cell invasion mechanism [23], the open conformation can be found in up to three chains in the mutant structure with D614G [19]. This, in turn, increases the chance of host cell binding. In the wildtype structure, only one chain was found to be in the open conformation when binding to human-ACE2.

4. Discussion

Our study shows that SARS-CoV2 variants are more likely to be found on the surface residues than the core residues of the spike protein. Moreover, the spike protein interface residues are much less enriched in amino acid variations when compared to non-interface surface residues. Our finding is in agreement with many previous studies conducted in human proteins, suggesting that variants are less likely to be found on the interface residues [20–22].

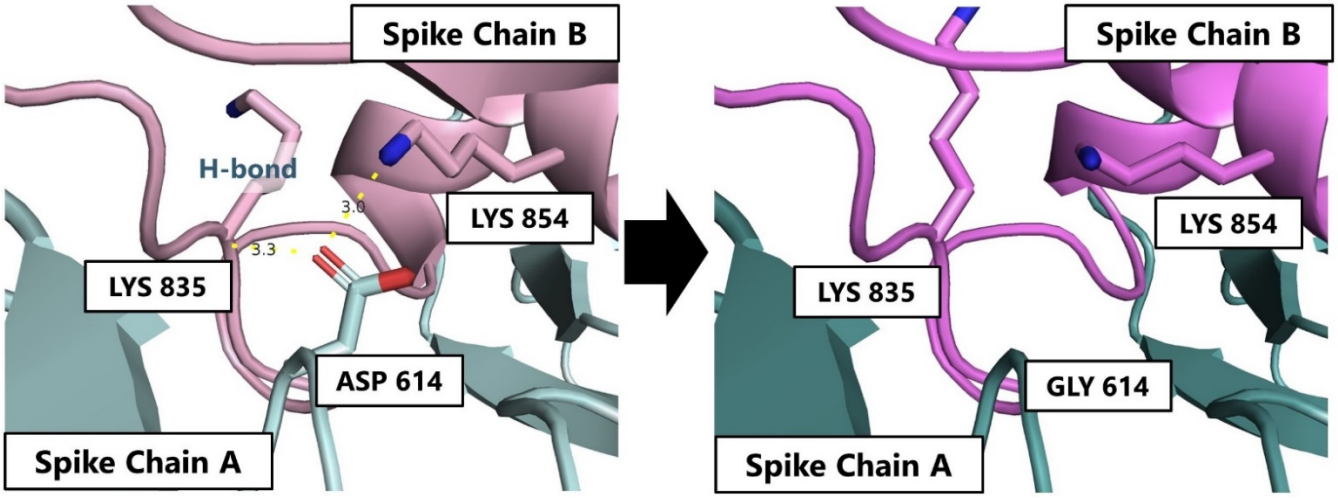


Figure 2. Structural consequence prediction of the variant of global concern Asp614Gly (D614G) using Missense3D. (left) pre-mutation; (right) post-mutation. The PDB used was 6XR8. Chain A is shown in grey and chain B in pink. H-bonds are shown as yellow dashed lines.

The extreme conservation of residues in some parts of the receptor-binding domain and the Coronavirus S2 domain could imply the biological importance of those particular residues for proper protein fold, protein-protein interactions, and functions. Furthermore, as genetic variations are generally maintained by natural selection, any alteration on these conserved residues may be deleterious to the survival of the virus. Therefore, further study and analyses on these residues need to be conducted as they could serve as promising drug target candidates or a biomarker for COVID-19 diagnosis.

One limitation in this study is that the publicly available Thai sequences were collected from two sources: the state quarantines (collected from individuals arriving in Thailand from overseas) and domestic hospitals/research institutions (collected from individuals infected by local transmission). Some strains collected in the state quarantine zones may have never been transmitted locally. Unfortunately, there is no way to distinguish the sources where the strains were collected from. Therefore, the number of Thai variants in this study could be an overestimation of the actual variants found in local transmission.

In Thailand, up until the second outbreak, the variant D614G was the only variant that considerably enhances transmission. Most of the variants detected in Thailand were predicted benign. Unfortunately, apart from strain reports, studies on the effect of Thai variants in terms of transmissibility and severity are very limited [7–10]. Therefore, some of the effects can only be implied from computational models. This can be helpful for prescreening or variants prioritisation. However, laboratory experiments or clinical studies are still required in order to confirm these *in silico* findings. It is also important to keep new strains under surveillance as the mutation rate is very high in single-stranded RNA viruses compared to DNA viruses [23]. Understanding molecular mechanisms of SARS-CoV2 variants could help scientists discover other alternative measures for COVID-19 diagnosis, prevention, and treatment.

5. Conclusions

In this study, we explored the amino acid variants on SARS-CoV2 spike protein through a large-scale analysis. We pinpointed mutation hotspots as well as highly conserved residues on the spike protein. Identification of conserved residues could suggest implications for alternative drug targets. We also identify variants found specifically in Thailand and show that some variants can affect the stability of the spike protein complexes while many were structurally neutral. The insights obtained from this study could help researchers further develop effective treatments for COVID-19.

Supplementary Materials: The following are available online at https://drive.google.com/drive/folders/1-PWgLbDQ0I_4BEskMiog-b26GhnjwYzR?usp=sharing , Table S1: surface-interface residues, Table S2: variant-list, Video S1: distribution of variants on the spike protein structure.

Author Contributions: Conceptualization and methodology, S.I.; data curation and investigation, S.Y.; formal analysis, S.I.; writing—original draft preparation, S.Y.; writing—review and editing, S.I.; visualization, S.I.; supervision, S.I. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Faculty of Science, Prince of Songkla University.

Acknowledgements: The authors would like to thank Prof. Michael Sternberg and Dr Alessia David from Structural Bioinformatics Group, Imperial College London, UK, for their supporting help.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Wang, C.; Wang, Z.; Wang, G.; Lau, J. Y. N.; Zhang, K.; Li, W. COVID-19 in Early 2021: Current Status and Looking Forward. *Signal transduction and targeted therapy*. 2021. https://doi.org/10.1038/s41392-021-00527-1.
2. Worldometers.info. COVID-19 Coronavirus Pandemix https://www.worldometers.info/coronavirus/ (accessed Apr 2, 2021).
3. Peiris, J. S. M.; Lai, S. T.; Poon, L. L. M.; Guan, Y.; Yam, L. Y. C.; Lim, W.; Nicholls, J.; Yee, W. K. S.; Yan, W. W.; Cheung, M. T.; et al. Coronavirus as a Possible Cause of Severe Acute Respiratory Syndrome. *Lancet*, **2003**, *361* (9366), 1319–1325. https://doi.org/10.1016/S0140-6736(03)13077-2.
4. de Groot, R. J.; Baker, S. C.; Baric, R. S.; Brown, C. S.; Drosten, C.; Enjuanes, L.; Fouchier, R. A. M.; Galiano, M.; Gorbalenya, A. E.; Memish, Z. A.; et al. Middle East Respiratory Syndrome Coronavirus (MERS-CoV): Announcement of the Coronavirus Study Group. *J. Virol.*, **2013**, *87* (14), 7790–7792. https://doi.org/10.1128/jvi.01244-13.
5. NCBI. SARS-CoV-2 Resources https://www.ncbi.nlm.nih.gov/sars-cov-2/ (accessed Apr 3, 2021).
6. Bogner, P.; Capua, I.; Cox, N. J.; Lipman, D. J. A Global Initiative on Sharing Avian Flu Data [1]. *Nature*. Nature Publishing Group August 31, 2006, p 981. https://doi.org/10.1038/442981a.
7. Okada, P.; Buathong, R.; Phuygun, S.; Thanadachakul, T.; Parnmen, S.; Wongboot, W.; Waicharoen, S.; Wacharapluesadee, S.; Uttayamakul, S.; Vachiraphan, A.; et al. Early Transmission Patterns of Coronavirus Disease 2019 (COVID-19) in Travellers from Wuhan to Thailand, January 2020. *Eurosurveillance*. European Centre for Disease Prevention and Control (ECDC) February 27, 2020. https://doi.org/10.2807/1560-7917.ES.2020.25.8.2000097.
8. Puenpa, J.; Suwannakarn, K.; Chansaenroj, J.; Nilyanimit, P.; Yorsaeng, R.; Auphimai, C.; Kitphati, R.; Mungaomklang, A.; Kongklieng, A.; Chirathaworn, C.; et al. Molecular Epidemiology of the First Wave of Severe Acute Respiratory Syndrome Coronavirus 2 Infection in Thailand in 2020. *Sci. Rep.*, **2020**, *10* (1). https://doi.org/10.1038/s41598-020-73554-7.
9. Buathong, R.; Chaifoo, W.; Iamsirithawon, S.; Wacharapluesadee, S.; Joyjinda, Y.; Rodpan, A.; Ampoot, W.; Putcharoen, O.; Paitoonpong, L.; Suwanpimolkul, G.; et al. Multiple Clades of SARS-CoV-2 Were Introduced to Thailand during the First Quarter of 2020. *Microbiol. Immunol.*, **2021**, 1348-0421.12883. https://doi.org/10.1111/1348-0421.12883.
10. Joonlasak, K.; Batty, E. M.; Kochakarn, T.; Panthan, B.; Kümpornsin, K.; Jiaranai, P.; Wangwiwatsin, A.; Huang, A.; Kotanan, N.; Jaru-Ampornpan, P.; et al. Genomic Surveillance of SARS-CoV-2 in Thailand Reveals Mixed Imported Populations, a Local Lineage Expansion and a Virus with Truncated ORF7a. *Virus Res.*, **2021**, *292*, 198233. https://doi.org/10.1016/j.virusres.2020.198233.
11. Kabsch, W.; Sander, C. Dictionary of Protein Secondary Structure: Pattern Recognition of Hydrogen-bonded and Geometrical Features. *Biopolymers*, **1983**, *22* (12). https://doi.org/10.1002/bip.360221211.
12. Rost, B.; Sander, C. Conservation and Prediction of Solvent Accessibility in Protein Families. *Proteins Struct. Funct. Bioinforma.*, **1994**, *20* (3), 216–226. https://doi.org/10.1002/prot.340200303.
13. Ittisoponpisan, S.; Islam, S. A.; Khanna, T.; Alhuzimi, E.; David, A.; Sternberg, M. J. E. Can Predicted Protein 3D Structures Provide Reliable Insights into Whether Missense Variants Are Disease Associated? *J. Mol. Biol.*, **2019**, *431* (11), 2197–2212. https://doi.org/10.1016/j.jmb.2019.04.009.
14. Lee, J.; Lee, K.; Shin, S. Theoretical Studies of the Response of a Protein Structure to Cavity- Creating Mutations. *Biophys. J.*, **2000**, *78* (4), 1665–1671. https://doi.org/10.1016/S0006-3495(00)76718-X.
15. Jenkins, K. A.; Fossat, M. J.; Zhang, S.; Rai, D. K.; Klein, S.; Gillilan, R.; White, Z.; Gerlich, G.; McCallum, S. A.; Winter, R.; et al. The Consequences of Cavity Creation on the Folding Landscape of a Repeat Protein Depend upon Context. *Proc. Natl. Acad. Sci. U. S. A.*, **2018**, *115* (35), E8153--E8161. https://doi.org/10.1073/pnas.1807379115.
16. Xue, M.; Wakamoto, T.; Kejlberg, C.; Yoshimura, Y.; Nielsen, T. A.; Risør, M. W.; Sanggaard, K. W.; Kitahara, R.; Mulder, F. A. A. How Internal Cavities Destabilize a Protein. *Proc. Natl. Acad. Sci. U. S. A.*, **2019**, *116* (42), 21031–21036. https://doi.org/10.1073/pnas.1911181116.

17. Peter, A. S.; Roth, E.; Schulz, S. R.; Fraedrich, K.; Steinmetz, T.; Damm, D.; Hauke, M.; Richel, E.; Mueller-Schmucker, S.; Habenicht, K.; et al. A Pair of Non-Competing Neutralizing Human Monoclonal Antibodies Protecting from Disease in a SARS-CoV-2 Infection Model. *bioRxiv*, **2021**, 2021.04.16.440101. https://doi.org/10.1101/2021.04.16.440101.
18. Volz, E.; Hill, V.; McCrone, J. T.; Price, A.; Jorgensen, D.; O'Toole, Á.; Southgate, J.; Johnson, R.; Jackson, B.; Nascimento, F. F.; et al. Evaluating the Effects of SARS-CoV-2 Spike Mutation D614G on Transmissibility and Pathogenicity. *Cell*, **2021**, *184* (1). https://doi.org/10.1016/j.cell.2020.11.020.
19. Yurkovetskiy, L.; Wang, X.; Pascal, K. E.; Tomkins-Tinch, C.; Nyalile, T. P.; Wang, Y.; Baum, A.; Diehl, W. E.; Dauphin, A.; Carbone, C.; et al. Structural and Functional Analysis of the D614G SARS-CoV-2 Spike Protein Variant. *Cell*, **2020**, *183* (3). https://doi.org/10.1016/j.cell.2020.09.032.
20. David, A.; Razali, R.; Wass, M. N.; Sternberg, M. J. E. Protein-Protein Interaction Sites Are Hot Spots for Disease-Associated Nonsynonymous SNPs. *Hum. Mutat.*, **2012**, *33* (2), 359–363. https://doi.org/10.1002/humu.21656.
21. David, A.; Sternberg, M. J. E. The Contribution of Missense Mutations in Core and Rim Residues of Protein-Protein Interfaces to Human Disease. *J. Mol. Biol.*, **2015**, *427* (17), 2886–2898. https://doi.org/10.1016/j.jmb.2015.07.004.
22. Jubb, H. C.; Pandurangan, A. P.; Turner, M. A.; Ochoa-Montaño, B.; Blundell, T. L.; Ascher, D. B. Mutations at Protein-Protein Interfaces: Small Changes over Big Surfaces Have Large Impacts on Human Health. *Progress in Biophysics and Molecular Biology*. Elsevier Ltd September 2017, pp 3–13. https://doi.org/10.1016/j.pbiomolbio.2016.10.002.
23. Sanjuán, R.; Nebot, M. R.; Chirico, N.; Mansky, L. M.; Belshaw, R. Viral Mutation Rates. *J. Virol.*, **2010**, *84* (19), 9733–9748. https://doi.org/10.1128/jvi.00694-10.