SARS-CoV-2 prion-like domains in spike proteins enable higher affinity to ACE2

George Tetz[1,2]* and Victor Tetz[1]

[1]Human Microbiology Institute, New York, NY 10016, USA. [2]Tetz Laboratories, New York, NY 10016, USA.

*Correspondence and requests for materials should be addressed to GT (email: g.tetz@hmi-us.com)

Abstract

Currently, the world is struggling with the coronavirus disease 2019 (COVID-19) pandemic, caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). Prion-like domains are critical for virulence and the development of therapeutic targets; however, the prion-like domains in the SARS-CoV-2 proteome have not been analyzed. In this in silico study, using the PLAAC algorithm, we identified the presence of prion-like domains in the SARS-CoV-2 spike protein. Compared with other viruses, a striking difference was observed in the distribution of prion-like domains in the spike protein, since SARS-CoV-2 was the only coronavirus with a prion-like domain found in the receptor-binding domain of the S1 region of the spike protein. The presence and unique distribution of prion-like domains in the SARS-CoV-2 receptor-binding domains of the spike protein is particularly interesting, since although the SARS-CoV-2 and SARS-CoV S proteins share the same host cell receptor, angiotensin-converting enzyme 2 (ACE2), SARS-CoV-2 demonstrates a 10- to 20-fold higher affinity for ACE2. Finally, we identified prion-like domains in the α1 helix of the ACE2 receptor that interact with the viral receptor-binding domain of SARS-CoV-2. Taken together, the present findings indicate that the identified PrDs in the SARS-CoV-2 receptor-binding domain (RBD) and ACE2 region that interact with RBD have important functional roles in viral adhesion and entry.

Introduction

The world is struggling with the pandemic caused by a novel coronavirus (now named severe acute respiratory syndrome-2 or SARS-CoV-2, causing the disease COVID-19) that has expanded from Wuhan throughout China (1). By March 30, 2020, the virus had caused over 775,000 confirmed cases worldwide and contributed to over 37,000 deaths (https://www.worldometers.info/coronavirus/).

SARS-CoV-2 is a new member of the Betacoronavirus (β-CoV) genus of large, enveloped single-stranded RNA viruses (2). This genus not only includes viruses that cause deadly human infections such as severe acute respiratory syndrome (SARS) and Middle East respiratory syndrome (MERS), but also encompasses viruses that cause non-life-threatening common colds, including human coronavirus OC43 (HCoV-OC43) and human coronavirus HKU1 (HCoV-HKU1) (3). Although these viruses predominantly infect lung epithelial cells, the clinical severity and pathogenesis of the infections they cause varies between different coronaviruses (4). While severe pneumonia and pulmonary fibrosis are fundamental to the pathogenesis of COVID-19, SARS, and MERS, these symptoms are not typical of infections caused by HCoV-OC43 and HCoV-HKU1 (5,6).

Like other β-CoVs, the genome of the novel SARS-CoV-2 virus encodes structural proteins required for the efficient formation of infectious virions; these include the spike (S), envelope (E), membrane (M), and nucleocapsid (N) proteins (7).

The key determinant of the host specificity of a β-CoVs is the surface-located S protein, which plays critical roles in infection by mediating viral attachment to host cell surface receptors and facilitating viral entry (8). The S protein consists of two large regions: N-terminal S1 and C-terminal S2 (9). S1 is responsible for recognizing host-cell receptors, including the receptor-binding domain (RBD), and has higher sequence variability than S2 (S1 shares around 70% identity with that of other human β-CoVs). Moreover, the membrane-embedded S2 region responsible for fusion is more highly conserved than that of S1 (8,9) In SARS-CoV-2, the RBD in S1 allows the virus to bind directly to the peptidase domain of the host angiotensin-converting enzyme 2 (ACE2) complex, mediating virus entry into sensitive cells (10). Notably, compared to SARS-CoV, SARS-CoV-2 has a higher binding affinity to ACE2 (which is the common receptor for both SARS-CoV-2 and SARS-CoV), with a broader interaction with ACE2 expressed not only in the lungs but also in kidney, testis, and heart (10,11).

Recently we have conducted an analysis and identified for the first time viral prion-like domains (PrDs), which we suggest are novel regulators of virion assembly with a role in virus-host cell interactions (12,13). These studies were in alignment with previous studies, showing that in addition to the pathological role of prions that they play in humans being implicated in Alzheimer's and Parkinson's diseases, diabetes, and many other human pathologies, protein misfolding plays important physiological roles in eukaryotes and prokaryotes (14-17).

Though the detailed molecular mechanisms underlying prion formation remain elusive, asparagine (Q)- and glutamine (N)-rich regions characterized by altered hydrophobicity and net sequence charge are known to drive prion formation. This is the basis for a number of algorithms for identifying candidate prionogenic domains (18,19). One such algorithm is prion-like amino acid composition (PLAAC) analysis, which allows the evaluation of prion-like domains based on the hidden Markov model (HMM) (20).

In this study, for the first time, we performed a detailed study of the prion-like domains in spike protein of SARS-CoV-2 and a comparison of SARS-CoV-2 to other human-pathogenic β-CoVs. Our findings can contribute to a better understanding of the pathogenicity of SARS-CoV-2 and will help to uncover new targets for the development of drugs and vaccines based on the prionogenic properties of particular viral protein regions.

Results

Using the prion-prediction PLAAC algorithm, we analyzed structural proteins derived from UniProtKB and NCBI databases and identified PrDs in the S proteins of all β-CoVs analyzed in this study (Supplementary figure S1). The LLR scores of PrDs of the S proteins were practically identical within the studied β-CoVs, ranging from 4.431 to 4.991 (Supplementary Figure S2). Notably, with more precise mapping of PrDs within these proteins, we found a striking difference in their localization with SARS-CoV-2 being the only virus with PrDs identified within the RBD of the S protein (Table 1).

Considering that although SARS-CoV-2 and SARS-CoV (which are the closest related human β-CoVs pathogens) share the same host-cell receptor ACE2, SARS-CoV-2 binds tighter to it; therefore, we hypothesized that the presence of PrDs in the RBD of the SARS-CoV-2 might explain this phenomenon (10). Consistent with this hypothesis, we found that SARS-CoV-2 along with other residue substitutions has five substituted amino acids in the RBD compared to SARS-CoV; the following are the substitutions in the RBD: S460 → Q474, T488 → N481, N480 → Q493, Y485 → Q498 and T488 → N501, therefore forming a hydrophobic Q/N rich region that enables the prionogenity of the SARS-CoV-2 RBD (Figure 1).

We next analyzed the presence of prion-like domains in ACE2 protein and found PrDs within the α1 helix of ACE2 (aa 40-65 and 93-106) (Supplementary figure S2). Based on previous analysis by Yan et al. we modulated the interface between the SARS-CoV-2 RBD and ACE2 identified in this study, by aligning their sequences (Figure 2A, B) (21). Interestingly, we identified a pattern in which five of seven amino acids that interact between the SARS-CoV-2 RBD and host cell ACE2 are localized within the PrDs of SARS-CoV-2 RBD, ACE2 or both of them (Figure 2A). Thus, Q498 and T500 from the PrD of the SARS-CoV2 RBD interact with Y41 and Q42 within the PrD of ACE2; while Q474, F486 and N501 from the PrD of the SARS-CoV2 RBD bind to Q24, M82 and K343 of a non-PrD of ACE2. Notably, only K417 and Y453 were the only residues of the SARS-CoV-2 RBD that were outside the viral PrD and bound to a non-PrD of ACE2 (Figure 2B).

Discussion

This study is the most complete evaluation of PrDs in the S protein of SARS-CoV-2. The results highlight some previously unknown, unique characteristics of SARS-CoV-2 that may play important roles in the pathogenesis and inform the development of new therapeutic strategies.

In this study, we used a high threshold of the PLAAC score for protein identification: only proteins with a high probability of prionogenic properties were included in the analysis. We found that all members of β-CoVs members contain PrDs in the S proteins. However, SARS-CoV-2 is the only member of β-CoVs that has a PrD in the RBD of the S protein that binds to the ACE2 receptor employed for host cell entry. Furthermore, we discovered specific amino acids (Q474, N481, Q493, Q498 and N501) that enable the prionogenity of the SARS-CoV-2 RBD that are not found in the RBD of SARS-CoV, of which Q474, Q498 and N501 directly contact within ACE2.

From these analyses, we conclude that the presence of these intrinsically disordered regions in the SARS-CoV-2 RBD, might be the reason for its optimized binding to the human ACE2 receptor in comparison to the RBD of SARS-CoV, since the distinguishing characteristic of PrDs is their ability to rapidly shift between multiple conformations due to residue hydrophobicity and net sequence charge (18, 22).

Notably, since five of seven amino-acid interactions that occur between the RBD of SARS-CoV-2 and ACE2 are within their PrDs, it is also interesting to consider whether the prion-prion interaction between the virus and human receptor participates in COVID-19 and does it add a special value for the higher affinity to their binding.   Since other β-CoVs were shown to lack the PrDs in the RBD, this means that the presence of PrDs is beneficial, but not necessary, for receptor-mediated virion attachment to the host cell. One of the critical goals of our previous studies was to show that PrDs identified in viruses may have important functional roles in virulence and are particularly associated with viral adhesion and entry.

This study provides a proof of this concept, showing that the presence of PrDs in the RBD of SARS-CoV-2 enhances viral binding to its host receptor compared to that of SARS-CoV, which lacks PrDs in its RBD structure. Further analyses of these PrD-containing proteins in SARS-CoV-2 may improve our understanding of COVID-19 infection and provide new insights into its pathophysiology novel targets for developing therapies.

Materials and Methods

Protein Sequences

To identify the PrDs present in viral proteomes, protein sequences were obtained from the UniProt Knowledge Base and National Center for Biotechnology Information (NCBI) database (23, http://www.ncbi.nlm.nih.gov/). Protein functions were manually curated using information from UniProt and NCBI databases. The structure of the RBD-ACE2 complex was established based on the data from PDB ID: 6VW1 and visualized using the YASARA software (http://www.yasara.com) (24, 25).

Identification of PrDs in viral proteomes

The presence of PrDs in β-CoV proteomes found using the PLAAC algorithm and the output probabilities for the PrDs were constructed based on amino-acid frequencies and similarities with PrDs in *Saccharomyces cerevisiae*. We used a cutoff of 3.0 log-likelihood ratio (LLR) and alpha = 1.0, representing *S. cerevisiae* background scanning, to identify the PrDs. Prion-like

domain amino acid positions were determined based on the PLAAC algorithm program analysis or manually.

Statistical analysis

All statistical analyses were conducted using the Statistical package for Windows (version 5.0) (StatSoft, Inc.). Data were compared between viruses using a χ^2 test or Fisher's exact test. To detect differences in multiple comparisons, one-way analysis of variance (ANOVA) was fitted with the standard confidence interval of 95%. P values < 0.05 were considered statistically significant.

Tables

Table 1. Comparison of the distribution of PrDs within the S protein among different β-CoV human pathogens

	S protein	
	Domain	**Prion-like domain AA position**
SARS-CoV-2	RBD	473-510
SARS-CoV	HR1	900-910
MERS-CoV	NA	Non-detectable
HCoV-OC43	NA	Non-detectable

Figures

Figure 1. Analysis and comparison of mutations in the RBD of SARS-CoV-2 and SARS-CoV.

The RBD of the SARS-CoV-2 spike protein was aligned against the closest related human βCoV, SARS-CoV. The PrDs of SARS-CoV-2 are red. Different residues are denoted by an "*" beneath the consensus position. The amino acids asparagine (Q) and glutamine (N) in the PrDs of the SARS-CoV-2 RBD that differ from the amino acids in the SARS-CoV RBD are denoted by red "**" beneath the consensus position. Amino acids of the SARS-CoV-2 RBD that bind to ACE2 are marked with red boxes. RBD - Receptor binding domain.

Figure 2. Interactions between amino acids of PrDs and non-prion-like regions of SARS-CoV-2 RBD and ACE2.
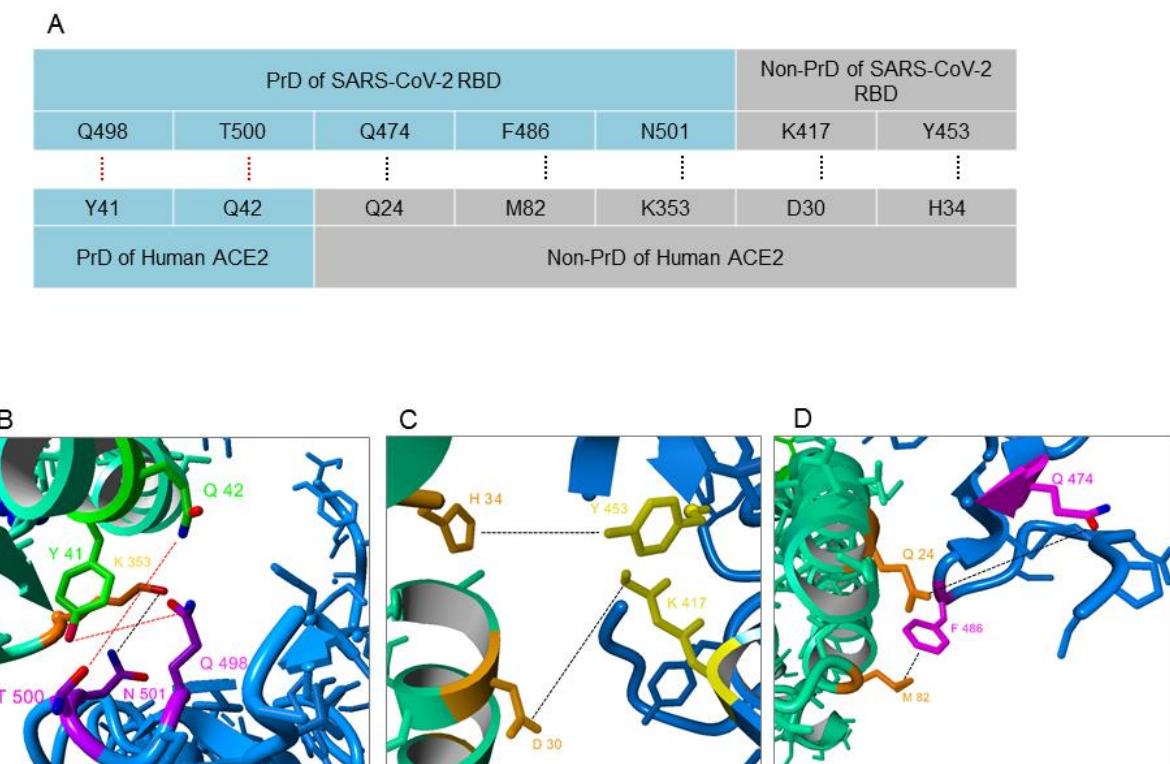
(A) The amino acids Q498 and T500 from the PrD of the SARS-CoV2 RBD interact with Y41 and Q42 within the PrD of ACE2, while Q474, F486 and N501 from the PrD of the SARS-CoV-2 RBD bind to Q24, M82 and K343 from the non-PrD of ACE2. K417 and Y453 were the only amino acids of the SARS-CoV-2 RBD that were outside the viral PrD and bound to ACE2.

(B to D) Detailed analysis of the interface between the SARS-CoV-2 RBD and ACE2. The structure of the RBD-ACE2 complex was established based on the data from PDB ID: 6VW1 and visualized using the YASARA software (http://www.yasara.com). The ACE2 and RBD molecules are stained green and blue respectively. Amino acids within the PrD of the RBD that interact with amino acid residues of ACE2 are stained purple, while those in the PrDs of ACE2 that interact with amino acid residues of the RBD are stained green; the interactions are indicated by red, dashed lines. Amino acids within the non-PrD of the RBD that interacts with amino acid residues of ACE2 are yellow, while those in the non-PrDs of ACE2 that interact with amino acid residues of the RBD are orange, and these interactions are indicated by a black dashed line.

Supplementary figure S1. Graphical representation of the LLR score in the PrDs of the S protein from different β-CoVs.
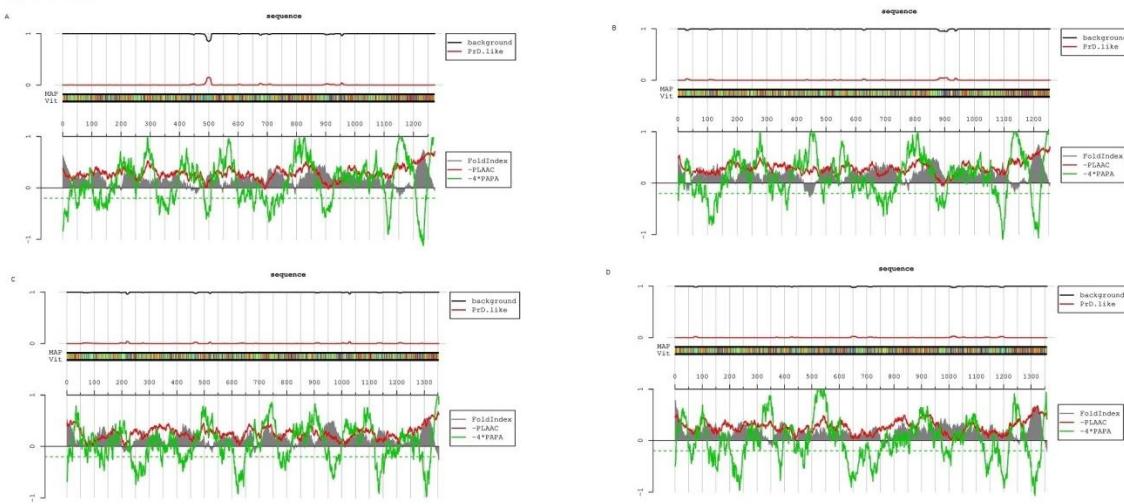

The LLR value of the S protein from (A) SARS-CoV-2, (B) SARS-CoV, (C) MERS-CoV and (D) HCoV-OC43.

Supplementary figure S2. LLR score showing the predicted putative PrDs in S proteins of β-CoVs.
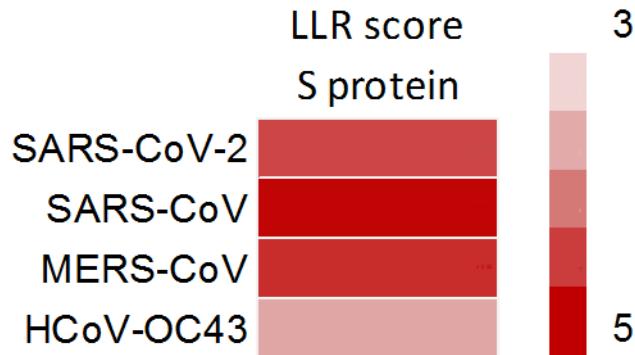
 Heatmap of PrD distribution in S proteins in β-CoVs. Cells indexed by rows and columns are marked using a color gradient, ranging from white (LLR < 3.0) to saturated red (LLR = 5). The results were analyzed using one-way ANOVA.

Supplementary figure S3. Graphical representation of the LLR score of PrDs in the ACE2 protein.
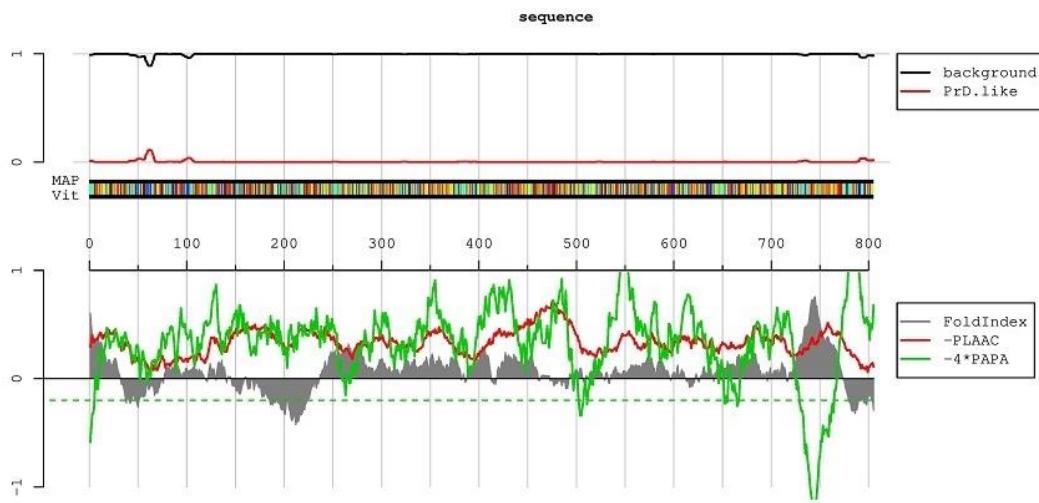
Author Contributions

GT and VT designed the experiments. GT performed the experiments and supervised data analysis. VT and GT analyzed the data and wrote the manuscript.

Competing interests

The authors declare no competing interests.

References

1. Lipsitch, M., Swerdlow, D.L. & Finelli, L. Defining the epidemiology of Covid-19—studies needed. *New England Journal of Medicine (*2020).

2. Xu, J., Zhao, S., Teng, T., Abdalla, A.E., Zhu, W., Xie, L., Wang, Y. & Guo, X. Systematic Comparison of Two Animal-to-Human Transmitted Human Coronaviruses: SARS-CoV-2 and SARS-CoV. *Viruses* **12***, .*244 (2020).

3. Chan, J.F.W., Kok, K.H., Zhu, Z., Chu, H., To, K.K.W., Yuan, S. & Yuen, K.Y. Genomic characterization of the 2019 novel human-pathogenic coronavirus isolated from a patient with atypical pneumonia after visiting Wuhan. *Emerging Microbes & Infections* **9**, 221-236 (2020).

4. Fehr, A.R., & Perlman, S. "Coronaviruses: an overview of their replication and pathogenesis." In *Coronaviruses* 1-23. Humana Press, New York, NY (2015).

5. Memish, Z.A., Perlman, S., Van Kerkhove, M.D & Zumla A. Middle East respiratory syndrome. *The Lancet* (2020).

6. Rothan, H.A. & Byrareddy, S.N. The epidemiology and pathogenesis of coronavirus disease (COVID-19) outbreak. *Journal of Autoimmunity* 102433, (2020).

7. Tai W , He L ,Zhang X ,Pu J, Voronin D , Jiang S, Zhou Y, Du L. Characterization of the receptor-binding domain (RBD) of 2019 novel coronavirus: implication for development of RBD protein as a viral attachment inhibitor and vaccine, Cellular & Molecular Immunology, 2020.

8. Lu, G., Wang, Q. and Gao, G.F., 2015. Bat-to-human: spike features determining 'host jump' of coronaviruses SARS-CoV, MERS-CoV, and beyond. *Trends in microbiology*, *23*(8), pp.468-478.

9. Shang W, Yang Y, Rao Y, Rao X. The outbreak of SARS-CoV-2 pneumonia calls for viral vaccines. npj Vaccines. 2020 Mar 6;5(1):1-3.

10. Wrapp, D., Wang, N., Corbett, K.S., Goldsmith, J.A., Hsieh, C.L., Abiona, O., Graham, B.S. and McLellan, J.S., 2020. Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science*, *367*(6483), pp.1260-1263.

11. Shen, Q., Xiao, X., Aierken, A., Liao, M. and Hua, J., 2020. The ACE2 Expression in Sertoli cells and Germ cells may cause male reproductive disorder after SARS-CoV-2 Infection.

12. Tetz G & Tetz V. Prion-Like Domains in Phagobiota. *Frontiers in Microbiology* 8: https://doi.org/10.3389/fmicb.2017.02239 (2017).

13. Tetz, G. and Tetz, V., 2018. Prion-like domains in eukaryotic viruses. *Scientific reports*, *8*(1), pp.1-10.

14. Soto C. Unfolding the role of protein misfolding in neurodegenerative diseases. Nature Reviews Neuroscience. 2003 Jan;4(1):49-60.

15. Tetz G, Brown SM, Hao Y, Tetz V. Type 1 diabetes: an association between autoimmunity, the dynamics of gut amyloid-producing E. coli and their phages. Scientific reports. 2019 Jul 4;9(1):1-1.

16. Tetz, G., Pinho, M., Pritzkow, S., Mendez, N., Soto, C. and Tetz, V., 2020. Bacterial DNA promotes Tau aggregation. *Scientific Reports*, *10*(1), pp.1-11.

17. Fowler, D., Koulov, A., Balch, W. & Kelly, J. Functional amyloid – from bacteria to humans. *Trends in Biochemical Sciences* **32**, 217–224, (2007).

18.   Michelitsch, M. & Weissman, J. A census of glutamine/asparagine-rich regions: Implications for their conserved function and the prediction of novel prions. *Proceedings of the National Academy of Sciences* **97**, 11910–11915, https://doi.org/10.1073/pnas.97.22.11910 (2000).

19.   Ross, E., MacLea, K., Anderson, C. & Ben-Hur, A. A Bioinformatics Method for Identifying Q/N-Rich Prion-Like Domains in Proteins. *Methods in Molecular Biology* 219–228, (2013).

20.   Lancaster, A., Nutter-Upham, A., Lindquist, S. & King, O. PLAAC: a web and command-line application to identify proteins with prion-like amino acid composition. *Bioinformatics* **30**, 2501–2502, (2014).

21.   Yan, R., Zhang, Y., Guo, Y., Xia, L. and Zhou, Q., 2020. Structural basis for the recognition of the 2019-nCoV by human ACE2. *bioRxiv*.

22.   Owen, I. and Shewmaker, F., 2019. The role of post-translational modifications in the phase transitions of intrinsically disordered proteins. *International journal of molecular sciences*, *20*(21), p.5501.

23.   UniProt Consortium. Reorganizing the protein space at the Universal Protein Resource (UniProt). *Nucleic acids research* **40.D1**, D71–D75, (2011).

24.   Shang, J., Ye, G., Shi, K., Wan, Y., Luo, C., Aihara, H., Geng, Q., Auerbach, A. and Li, F., Structural basis for receptor recognition by the novel coronavirus from Wuhan.

25.   Krieger, E. and Vriend, G., 2014. YASARA View—molecular graphics for all devices—from smartphones to workstations. *Bioinformatics*, *30*(20), pp.2981-2982.