Computational Electrostatics Predict Variations in SARS-CoV-2 Spike and Human ACE2 Interactions

Scott P. Morton*
Middle Tennessee State University
Murfreesboro, Tennessee
spm3c@mtmail.mtsu.edu

Joshua L. Phillips*†
Middle Tennessee State University
Murfreesboro, Tennessee
Joshua.Phillips@mtsu.edu

ABSTRACT

SARS-CoV-2 is a novel virus that is presumed to have emerged from bats to crossover into humans in late 2019. As the global pandemic ensues, scientist are working to evaluate the virus and develop a vaccine to counteract the deadly disease that has impacted lives across the entire globe. We perform computational electrostatic simulations on multiple variants of SARS-CoV-2 spike protein s1 in complex with human angiotensin-converting enzyme 2 (ACE2) variants to examine differences in electrostatic interactions across the various complexes. Calculations are performed across the physiological pH range to also examine the impact of pH on these interactions. Two of six spike protein s1 variations having greater electric forces at pH levels consistent with nasal secretions and significant variations in force across all five variants of ACE2. Five out of six spike protein s1 variations have relatively consistent forces at pH levels of the lung, and one spike protein s1 variant that has low potential across a wide range of pH. These predictions indicate that variants of SARS-CoV-2 spike proteins and human ACE2 in certain combinations could potentially play a role in increased binding efficacy of SARS-CoV-2 *in vivo*.

KEYWORDS

COVID-19; SARS-CoV-2; CoV-2; pandemic; electrostatics; binding; pH; mucosa

1 INTRODUCTION

SARS-CoV-2 (*CoV-2*) is the latest novel coronavirus to have crossed the species boundary into humans over the past twenty years [7, 22]. Currently available information strongly suggests an originating source of the CoV-2 virus to be the bat coronavirus SL-CoV-RaTG13 with 96% similarity to CoV-2 [21].

The World Health Organization (*WHO*) declared CoV-2 a public health emergency on January 30, 2020 and then pandemic status in the weeks following. WHO utilizes a tool to distinguish diseases based on the potential to bloom into epidemic/pandemic proportions and the availability of countermeasures to focus research and development (*R&D*) resources accordingly [19, 20].

We have previously performed extensive research using electrostatic analysis of HIV gp120 envelope glycoprotein interactions with human CD4 and broadly neutralizing antibody proteins [5, 6, 11–13, 15]. Electrostatics are a novel way to evaluate protein interactions at the structural level with a focus on how environmental conditions, like pH, effect binding efficacy. The processes can easily be adapted to any other protein-protein interactions which may be strengthened or weakened by modulating pH. The theoretical approaches have been further extended to predict electric force (Coulombs Law) modulation by pH using the same set of data generated by the electrostatic surface charge (*ESSC*) pipeline [10, 15].

2 RESULTS

For this research we utilized CoV-2 spike (s1) proteins: 6LZG_B, 6M17_E, 6VW1_2, 6W41_C, QIS30425.1, QIS60489.1 and angiotensin--converting enzyme 2 (*ACE2*) proteins: AAQ89076.1, BAG37592.1, BAJ21180.1, EAW98891.1, XP_011543854.1 obtained through sequence similarity searches using NCBI Protein BLAST and the Protein Data Bank [2]. Structurally, CoV-2 s1 shares similar traits with SARS-CoV s1 that emerged in 2003 with a major point of difference being an additional cleavage in the spike s1 protein sub-unit [17] making the substitution of similar SARS-CoV species difficult to justify in modeling and simulation work.

Computational methods of the pipeline performed as expected on s1 (see methods), but some results differed in content greatly when compared to those of HIV gp120. Past investigations predict stronger binding affinity between HIV gp120 and human CD4 proteins at lower physiological pH compared to higher physiological pH. Evidence of this was demonstrated by the pipeline EFP and BE data from those protein and complexes [5, 6, 11, 13]. However, the EFP and BE data for s1-ACE2 interactions predict no notable differences in binding efficacy across the physiologically relevant pH range. These data suggest that the s1-ACE2 interaction is therefore not pH-sensitive. For example, in Figure 1 we present a typical EFP for s1 and observe no notable shifts (away form zero) for these data in the physiologically relevant pH range. Color bars indicate the pH range of normal nasal secretions (red), pH 5.5 to 6.5; yellow indicates inflamed nasal secretions (rhinitis) from pH 7.2 to 8.3 [4]. Blue indicates the approximate lung pH listed as 7.38 to 7.43 [3], however, the graphs scale in 0.1 increments and displays the range as pH 7.3 to 7.5 as inclusive of the specific range. In Figure 2 we present the BEs for complexes showing the least and most variation, again observing no notable shifts for these data in the same relevant pH range.

One possible explanation for this lack of pH sensitivity is the minimal conformational change exhibited by s1-ACE2 interactions compared to gp120-CD4 interactions. The gp120-CD4 interaction requires a relatively large conformational change of the gp120 protein, leading to changes in the protein surface (residues) exposed to the external environment. In particular, an increase in negative surface change in the bound conformation (relative to the unbound conformation) of gp120, is thought to make the bound conformation

*Center for Computational Science
†Department of Computer Science

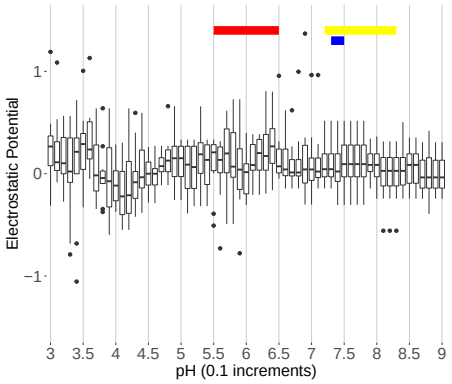

Figure 1: Typical electrostatic fingerprint of s1 subunit of CoV-2 envelope spike. Color bars near the top indicate the pH range of human nasal secretions (red), inflamed nasal secretions (yellow) and lung pH (blue). the fingerprint is basically flat from pH 5.0 and up indicating a stable structure. Color bars indicate pH range of normal nasal secretions (red), inflamed nasal secretions (yellow), and lung pH (blue).

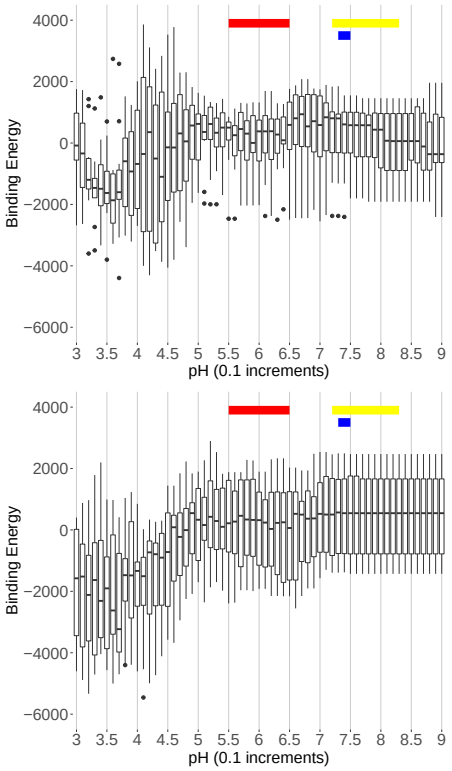

Figure 2: Predicted electrostatic contribution to binding energy across the physiological pH range shows some activity around nasal and lung pH levels in the most active plot (top), while a nearly flat result across the interesting range of pH 5.0 to 8.5 is observed in the most inactive plot (bottom). Color bars indicate pH range of normal nasal secretions (red), inflamed nasal secretions (yellow), and lung pH (blue).

energetically more favorable at low pH. However, the conformational change observed in s1-ACE2 complex binding appears to be either too small to alter, or simply results in too little alteration in the electrostatic surface charge of s1.

EFP and BE results predict that no or minimal differences exist in binding efficacy across the physiological pH range for s1-ACE2 interactions. These predictions leave unanswered questions, for instance, does the lack of differences in binding efficacy across the physiological pH range mean that s1 binds consistently regardless of pH? Another question, more pertinent to this research, what measurable forces are present between s1 and ACE2 that drive an interaction to taker place?

As a solution to the latter question, we expand the analysis with an alternative approach to electrostatic fingerprinting by taking the absolute values of the difference between s1 and ACE2 unbound conformation ESSC pipeline data to expose a predicted electric potential difference. We present results containing the highest and lowest potential differences from pH 4.5 to 9.0 in Figure 3. These data show an electric potential difference which varies across the examined pH range exists of varying quantity and across all complex combinations evaluated. Furthermore, the pipeline data predicts that s1 is positively charged throughout the pH range while ACE2 has a transition to negative between pH 4.5 and 5.0 as show in Figure 4. The presence of electric potential differences allow the results to be interpreted in terms of electric force.

The presence of electric potential difference introduces a force that can be determined by applying Coulombs Law to obtain the Electric Force (F_e) in Newtons (N). We can graph the changing forces, either repulsive (+) or attractive (-) between s1 and ACE2, across pH range of 3.0 to 9.0 in 0.1 increments at a distance of 10Å. Our model is *non-specific* in that specific interactions involve close-proximity contacts between proteins such as hydrogen bonds, unlike charges, and hydrophobic residue stacking. However, non-specific interactions are not close-proximity, and can be attractive or repulsive beyond the range of 3 angstroms or more. Electric force is a good example of the latter case. Details of the calculation are provided in the Methods section below.

Figure 5 displays the electric force from pH 3.0 to 9.0 in 0.1 increments for all complexes. Positive values indicate a repulsive force and negative values indicate an attractive force. For Figure 6 we discard the repulsive forces for the sake of figure clarity and on the basis that they exist outside the relevant pH range.

In Figure 6 we can see unique characteristics emerging where a group of complexes stand out from the rest with weakly attractive forces in a smooth curve across the examined range; another cluster of complexes emerges with a higher level of force in a slightly more erratic curve across the range and a group of complexes with a pronounced peak in force across the range of nasal secretions that peaks in the approximate middle.

While more details can be obtained from Figure 6, a clearer understanding is provided by Figure 7, which is grouped by s1. Clear patterns emerge that indicate higher forces presents for s1

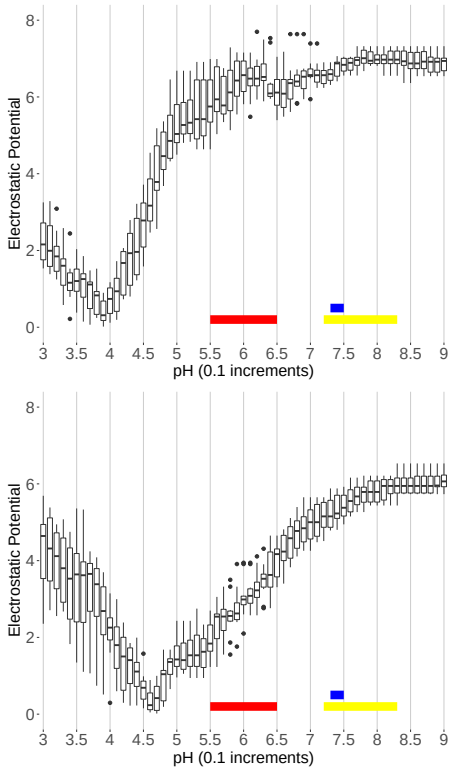

Figure 3: Comparison of potential differences between s1 and ACE2 at result extremes contrasting the highest potential differences (top) with the lowest (bottom). Color bars indicate pH range of normal nasal secretions (red), inflamed nasal secretions (yellow), and lung pH (blue).

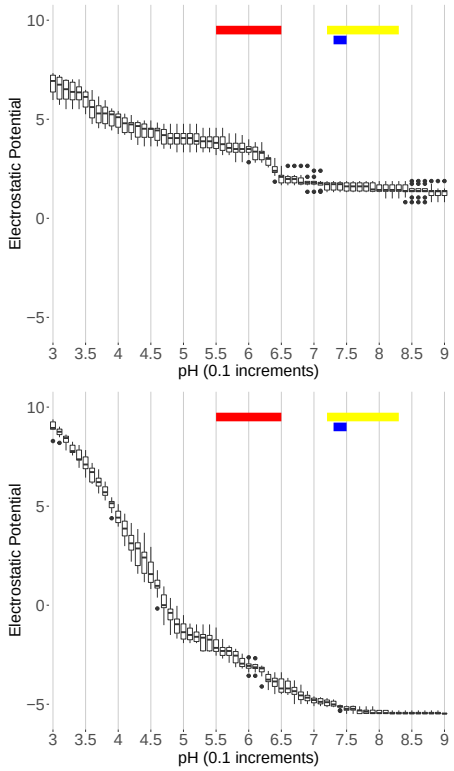

Figure 4: Typical ESSC pipeline results for s1 (top) and ACE2 (bottom). In all cases, s1 never crosses zero while ACE2 always transitions into negative values. Color bars indicate pH range of normal nasal secretions (red), inflamed nasal secretions (yellow), and lung pH (blue).

proteins 6LZG_B and 6W41_C at nasal secretion pH. Additionally, ACE2 genetic variations AAQ89076.1 and EAW98891.1 are most impacted by these two s1 variants. However, 6M17_E, QIS30425.1, and QIS60489.1 have nominal forces across the interesting pH range with marginal differences among ACE2 variants. 6VW1_2 has low potential across the pH range and little differences across ACE2 variants.

3 DISCUSSION

Our results clearly predict that sequence variations of s1 and ACE2 directly impact the attractive forces involved between s1 and ACE2. Furthermore, these results implicate human ACE2 variations that exacerbate those forces for specific s1 and ACE2 combinations. Additionally, we point out that three groupings emerge in this analysis showing two s1 variants (6LZG_B and 6W41_C) with large spikes in attractive force at pH 5.5 to 6.5 associated with nasal secretions. Three s1 variants (6M17_E, QIS30425.1, and QIS60489.1) have a relatively smooth curve at nominal attractive forces across ACE2 variants and one s1 variation (6VW1_2) has a smooth curve with low attractive force and minimal differences across ACE2 variations.

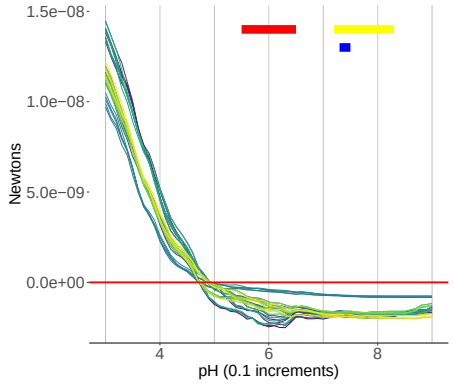

Figure 5: Aggregation of electric force for all complexes. Positive values are repulsive and negative values are attractive. Color bars indicate pH range of normal nasal secretions (red), inflamed nasal secretions (yellow), and lung pH (blue).

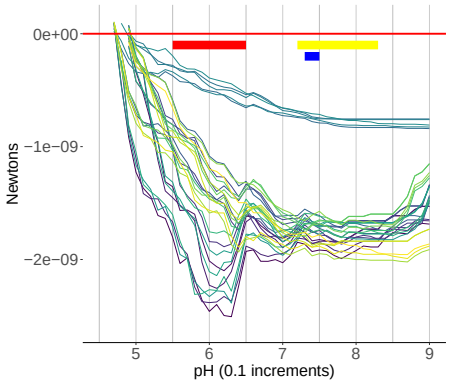

Figure 6: Aggregation of electric force for all complexes analyzed minus repulsive forces. This representation allows for a broad interpretation of the data where a single group of complexes stands out from the rest and another group has large spikes at nasal secretion pH levels. Color bars indicate pH range of normal nasal secretions (red), inflamed nasal secretions (yellow), and lung pH (blue).

Not all s1 and ACE2 protein sequences examined have solved crystal structures and were instead modeled based on the currently available s1 and ACE2 structures (see methods). Additional crystal structures for either or both s1 and ACE2 would allow for more precise predictions. Additionally, an analysis of electric force at the amino acid level would provide more detailed predictions at specific regions involved with s1 and ACE2 contact points and will be the subject of future studies.

We provide access to the full set of figures and results generated from this study as a tar package located at:
https://github.com/spmorton/SARS-CoV-2

4 METHODS

All computational methods related to the electrostatic pipeline are detailed in [5, 6, 11–13, 15]. Templates used to model s1 are: 6LZG, 6M0J, 6VW1, and 6VYB [9, 14, 17, 18] and for ACE2 the templates are: 1R4L, 1R42, 6M0J, 6VW1 [9, 14, 16]. To target Frodan for specific conformations of s1 we used: 6LGZ as bound and 6VYB as unbound target conformations. For ACE2 we used 6LZG for bound and unbound target conformations.

For the specific calculation of Electric Force, APBS returns a factor for V in:

$$V = \frac{k_b T}{e_c}$$

Where V is voltage of the system per unit, k_b is the Boltzmann's constant: 1.3806504 E-23 $JK^{-}1$, T is the temperature in Kelvin and e_c is the charge of an electron: 1.60217646 E-19 C.

For the calculation of Electric Force (F_e) we have:

$$F_e = k_c \left| \frac{q_1 q_2}{r^2} \right|$$

Where F_e is the Electric Force in Newtons, k_c is Coulombs constant: 8.987551E+09, q_1 is the charge in Coulombs of the first mass, q_2 is the charge in Coulombs of the second mass, and r is the distance in meters between the two masses.

To derive Coulombs of charge from APBS [1, 8] requires a simple transposition of the terms V and e_c and multiplication of the results with values returned by APBS. For this model, the variables required to complete the calculations are: $T = 310K$ and $r = 10$Å.

REFERENCES

[1] N. A. Baker, D. Sept, S. Joseph, M. J. Holst, and J. A. McCammon. 2001. Electrostatics of nanosystems: application to microtubules and the ribosome. *Proceedings of the National Academy of Sciences of the United States of America* 98, 18 (aug 2001), 10037–10041. https://doi.org/10.1073/pnas.181342398

[2] Helen Berman, Kim Henrick, and Haruki Nakamura. 2003. Announcing the worldwide Protein Data Bank. *Nature Structural & Molecular Biology* 10, 12 (dec 2003), 980–980. https://doi.org/10.1038/nsb1203-980

[3] Richard M. Effros and Francis P. Chinard. 1969. The in vivo pH of the extravascular space of the lung. *Journal of Clinical Investigation* 48, 11 (1969), 1983. https://doi.org/10.1172/JCI106164

[4] R. J. A. England, J. J. Homer, L. C. Knight, and S. R. Ell. 1999. Nasal pH measurement: a reliable and repeatable parameter. *Clinical Otolaryngology and Allied Sciences* 24, 1 (feb 1999), 67–68. https://doi.org/10.1046/j.1365-2273.1999.00223.x

[5] Jonathan Howton. 2017. *A Computational Electrostratic Modeling Pipeline for Comparing pH-dependent gp120-CD4 Interactions in Founder and Chronic HIV Strains.* Ph.D. Dissertation. Middle Tennessee State University, Murfreesboro, TN. http://jewlscholar.mtsu.edu/xmlui/handle/mtsu/5324

[6] Jonathan Howton and Joshua L Phillips. 2017. Computational Modeling of pH-dependent gp120-CD4 Interactions in Founder and Chronic HIV Strains. In *Proceedings of the 8th ACM International Conference on Bioinformatics, Computational Biology,and Health Informatics - ACM-BCB '17*. ACM Press, Boston, MA, USA, 644–649. https://doi.org/10.1145/3107411.3107506

[7] Chaolin Huang, Yeming Wang, Xingwang Li, Lili Ren, Jianping Zhao, Yi Hu, Li Zhang, Guohui Fan, Jiuyang Xu, Xiaoying Gu, Zhenshun Cheng, Ting Yu, Jiaan Xia, Yuan Wei, Wenjuan Wu, Xuelei Xie, Wen Yin, Hui Li, Min Liu, Yan Xiao, Hong Gao, Li Guo, Jungang Xie, Guangfa Wang, Rongmeng Jiang, Zhancheng Gao, Qi Jin, Jianwei Wang, and Bin Cao. 2020. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *The Lancet* 395, 10223 (feb 2020), 497–506. https://doi.org/10.1016/S0140-6736(20)30183-5

[8] Elizabeth Jurrus, Dave Engel, Keith Star, Kyle Monson, Juan Brandi, Lisa E. Felberg, David H. Brookes, Leighton Wilson, Jiahui Chen, Karina Liles, Minju Chun, Peter Li, David W. Gohara, Todd Dolinsky, Robert Konecny, David R. Koes, Jens Erik Nielsen, Teresa Head-Gordon, Weihua Geng, Robert Krasny, Guo-Wei Wei, Michael J. Holst, J. Andrew McCammon, and Nathan A. Baker. 2018. Improvements to the APBS biomolecular solvation software suite. *Protein Science* 27, 1 (jan 2018), 112–128. https://doi.org/10.1002/pro.3280

[9] Jun Lan, Jiwan Ge, Jinfang Yu, Sisi Shan, Huan Zhou, Shilong Fan, Qi Zhang, Xuanling Shi, Qisheng Wang, Linqi Zhang, and Xinquan Wang. 2020. Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *Nature* 581, 7807 (may 2020), 215–220. https://doi.org/10.1038/s41586-020-2180-5

[10] Scott P Morton, Salvador Barbosa, Ralph Butler, and Chrisila Pettey. 2016. A JSON-Based Markup Language for Deploying Virtual Clusters via Docker. In *Proceedings of the International Conference on Parallel and Distributed Processing Techniques and Applications PDPTA'16*. CSREA Press ©, 251–257. http://worldcomp-proceedings.com/proc/p2016/PDP3139.pdf

[11] Scott P. Morton, Jonathan Howton, and Joshua L. Phillips. 2018. Sub-Class Differences of PH-Dependent HIV GP120-CD4 Interactions. In *Proceedings of the 2018 ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics - BCB '18*. ACM Press, New York, New York, USA, 663–668. https://doi.org/10.1145/3233547.3233711

[12] Scott P Morton, Julie B Phillips, and Joshua L Phillips. 2017. High-Throughput Structural Modeling of the HIV Transmission Bottleneck. In *Proceedings of the 2017 IEEE International Conference on Bioinformatics and Biomedicine - BIBM-HPCB '17*, Vol. 2017-Janua. IEEE Press, Kansas City, MO, USA. https://doi.org/10.1109/BIBM.2017.8217952

[13] Scott P Morton, Julie B Phillips, and Joshua L Phillips. 2019. The Molecular Basis of pH-Modulated HIV gp120 Binding Revealed. *Evolutionary Bioinformatics* 15 (jan 2019), 117693431983130. https://doi.org/10.1177/1176934319831308

[14] Jian Shang, Gang Ye, Ke Shi, Yushun Wan, Chuming Luo, Hideki Aihara, Qibin Geng, Ashley Auerbach, and Fang Li. 2020. Structural basis of receptor recognition by SARS-CoV-2. *Nature* 581, 7807 (may 2020), 221–224. https://doi.org/10.1038/s41586-020-2179-y

[15] Daniel J Stieh, Joshua L Phillips, Paul M Rogers, Deborah F King, Gianguido C Cianci, Simon A Jeffs, Sandrasegaram Gnanakaran, and Robin J Shattock. 2013.

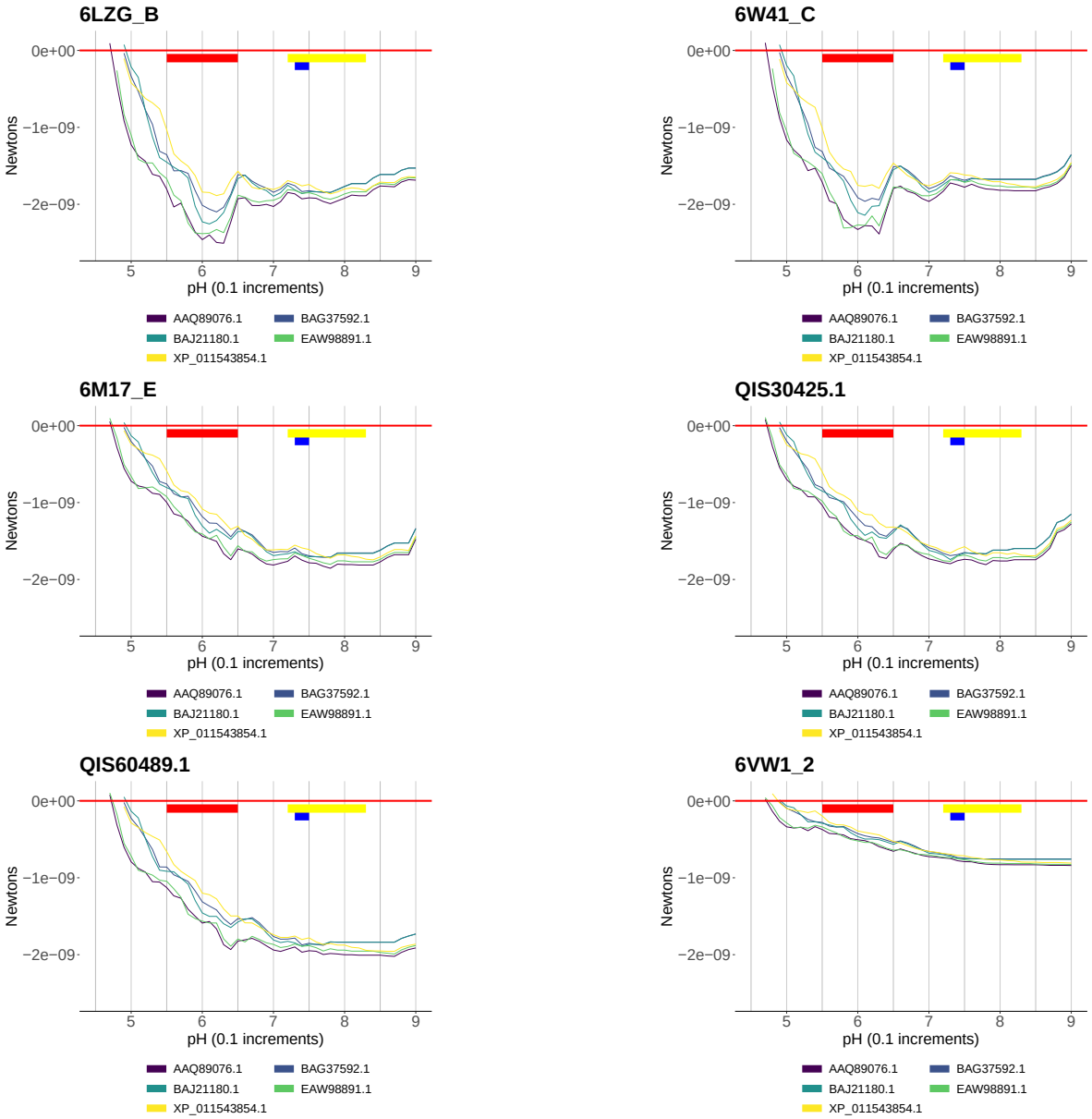

Figure 7: Grouped by s1, patterns emerge that indicate higher forces for s1 proteins 6LZG_B and 6W41_C at nasal secretion pH and distinct bias for ACE2 variants AAQ89076.1 and EAW98891.1, a lesser bias for ACE2 BAJ21180.1 and BAG37592.1. CoV-2 variants 6M17_E, QIS30425.1, and QIS60489.1 have nominal forces across the interesting pH range with marginal differences among ACE2 variants. 6VW1_2 has lower forces across the pH range and little differences across ACE2. Color bars indicate pH range of normal nasal secretions (red), inflamed nasal secretions (yellow), and lung pH (blue).

Dynamic electrophoretic fingerprinting of the HIV-1 envelope glycoprotein. *Retrovirology* 10, 1 (2013), 33. https://doi.org/10.1186/1742-4690-10-33

[16] Paul Towler, Bart Staker, Sridhar G Prasad, Saurabh Menon, Jin Tang, Thomas Parsons, Dominic Ryan, Martin Fisher, David Williams, Natalie A Dales, Michael A Patane, and Michael W Pantoliano. 2004. ACE2 X-ray structures reveal a large hinge-bending motion important for inhibitor binding and catalysis. *The Journal of biological chemistry* 279, 17 (apr 2004), 17996–8007. https://doi.org/10.1074/jbc.M311191200

[17] Alexandra C. Walls, Young-Jun Park, M. Alejandra Tortorici, Abigail Wall, Andrew T. McGuire, and David Veesler. 2020. Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. *Cell* 181, 2 (apr 2020), 281–292.e6. https://doi.org/10.1016/J.CELL.2020.02.058

[18] Qihui Wang, Yanfang Zhang, Lili Wu, Sheng Niu, Chunli Song, Zengyuan Zhang, Guangwen Lu, Chengpeng Qiao, Yu Hu, Kwok-Yung Yuen, Qisheng Wang, Huan Zhou, Jinghua Yan, and Jianxun Qi. 2020. Structural and Functional Basis of SARS-CoV-2 Entry by Using Human ACE2. *Cell* 181, 4 (may 2020), 894–904.e9. https://doi.org/10.1016/J.CELL.2020.03.045

[19] WHO. [n.d.]. Prioritizing diseases for research and development in emergency contexts. https://www.who.int/activities/prioritizing-diseases-for-research-and-development-in-emergency-contexts

[20] WHO. 2017. WHO | WHO publishes list of top emerging diseases likely to cause major epidemics. *WHO* (2017). https://www.who.int/medicines/ebola-treatment/WHO-list-of-top-emerging-diseases/en/

[21] Peng Zhou, Xing-Lou Yang, Xian-Guang Wang, Ben Hu, Lei Zhang, Wei Zhang, Hao-Rui Si, Yan Zhu, Bei Li, Chao-Lin Huang, Hui-Dong Chen, Jing Chen, Yun Luo, Hua Guo, Ren-Di Jiang, Mei-Qin Liu, Ying Chen, Xu-Rui Shen, Xi Wang, Xiao-Shuang Zheng, Kai Zhao, Quan-Jiao Chen, Fei Deng, Lin-Lin Liu, Bing Yan, Fa-Xian Zhan, Yan-Yi Wang, Geng-Fu Xiao, and Zheng-Li Shi. 2020. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* 579, 7798 (mar 2020), 270–273. https://doi.org/10.1038/s41586-020-2012-7

[22] Na Zhu, Dingyu Zhang, Wenling Wang, Xingwang Li, Bo Yang, Jingdong Song, Xiang Zhao, Baoying Huang, Weifeng Shi, Roujian Lu, Peihua Niu, Faxian Zhan, Xuejun Ma, Dayan Wang, Wenbo Xu, Guizhen Wu, George F. Gao, and Wenjie Tan. 2020. A Novel Coronavirus from Patients with Pneumonia in China, 2019. *New England Journal of Medicine* 382, 8 (feb 2020), 727–733. https://doi.org/10.1056/NEJMoa2001017