Electrostatic Characteristics of SARS-CoV-2 Spike and Human ACE2 Protein Variations Predict Mutable Binding Efficacy

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 pneumonia virus that 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 present an electrostatic analysis at the structural level that predicts different variations of SARS-CoV-2 in complex with human angiotensin-converting enzyme 2 (ACE2) variants impact forces involved with interactions between the proteins. Two of six SARS-CoV-2 variations having greater electric forces at pH levels consistent with nasal secretions and significant variations in force across all five variants of ACE2. While five out of six SARS-CoV-2 variations have relatively consistent forces at pH levels associated with normal lung function, and one SARS-CoV-2 variant that has low potential across a wide range of pH. These predictions indicate that variants of SARS-CoV-2 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; Wuhan; electrostatics; binding; pH; mucosa

ACM Reference Format:
Scott P. Morton and Joshua L. Phillips. 2020. Electrostatic Characteristics of SARS-CoV-2 Spike and Human ACE2 Protein Variations Predict Mutable Binding Efficacy. In *Proceedings of ACM Conference (Conference'17).* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

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, 17]. The World Health Organization (*WHO*) declared CoV-2 a public health emergency on January 30, 2020 and then pandemic status in the weeks following.

*Center for Computational Science
†Department of Computer Science

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
Conference'17, July 2017, Washington, DC, USA
© 2020 Association for Computing Machinery.
ACM ISBN 978-x-xxxx-xxxx-x/YY/MM…$15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

The current available information strongly suggests an originating source of the CoV-2 virus to be the bat caronavirus SL-CoV-RaTG13 with 96% similarity to CoV-2 [16]. WHO objectives include the measure of impact an emerging disease has to the public based on the potential to bloom into epidemic/pandemic proportions and the availability of countermeasures [14]. WHO ranked CoV-2 as a top priority for research and development (*R&D*) activities in their tools designed to provide focused R&D [15].

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, 9–12]. 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.

Although the focus has been HIV, the processes can easily be adapted to any domain for which molecular interactions may be altered 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 [8, 12].

2 RESULTS

For this research we utilized CoV-2 '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 such as NCBI Protein BLAST and the Protein Data Bank [2]. Each structure is unique in sequence arrangement out of all available crystal structures. Structurally, CoV-2 shares similar traits with SARS-CoV that emerged in 2003 with a major point of difference being an additional cleave in the spike sub-unit 's1' [13] making the substitution of similar species difficult to justify in model and simulation work.

Methods of charge calculation functioned as desired through the pipeline on CoV-2, but some analysis methods returned unexpected results. Where electrostatic fingerprinting (*EFP*) relies on structural displacement of the binding interface, the 's1' subunit of CoV-2's envelope spike assembly is structurally stable as indicated by the EFP of Figure 1; this graph shows minor fluctuations across zero with most activity associated with lower pH values likely caused by denaturing of the protein. Some evidence of fluctuation can be observed around the pH range of nasal secretions indicated by the red bar near the top and ranging from pH 5.5 to 6.5[4]. The yellow bar indicates inflamed nasal secretions (rhinitis) pH from 7.2 to 8.3 [4] and blue indicates the approximate normal lung pH listed as

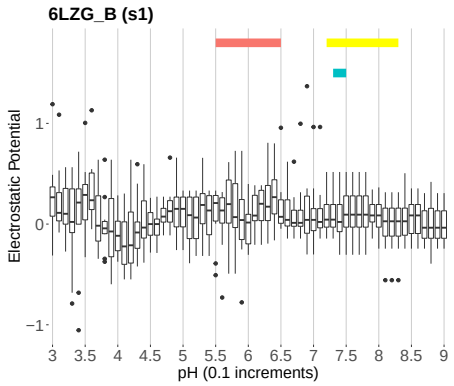

Figure 1: Typical electrostatic fingerprint of 's1' subunit of CoV-2 envelope spike. Color bars near the top indicate the pH range of human nasal secretions (red), inflamed nasal secretions (yellow) and normal lung pH (blue). the fingerprint is basically flat from pH 5.0 and up indicating a stable structure

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.

Application of the Adaptive Poisson-Boltzmann Solver (APBS) [1] to derive the binding energies provided marginally better results in some cases where other returns were also unexpected. We present the best and worst case returns across the all combinations in Figure 2. Again, most of the information derived is basically flat with the exception of lower pH as previously described.

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 APBS data. With all combinations available, the extremes are presented in Figure 3. There is a clear contrast between the two extremes across pH 4.5 to 9.0.

An examination of APBS results for any 's1' and ACE2 structure establishes the fact that 's1' of CoV-2 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. This knowledge allows the application of fundamental physics to express the results in terms of force.

By applying Coulombs Law to obtain the Electric Force (F_e) in Newtons (N), that either repel or attract two masses, we can graph the changing forces across the pH range of 3.0 to 9.0 in 0.1 increments to clearly represent the potential forces involved with two molecules in close proximity (10Å). Our model excludes all interference involved with solvent contaminates that may produce competing forces, orientation, or any other form that can be contemplated and is a purely *non-specific interaction*. Details of the calculation are provided in the Methods section below.

Figure 5 is an unlabeled graph of all complexes that displays a large repulsive force from pH 3.0 to approximately pH 4.7 where values cross the zero boundary into an attractive measure of force. For the purposes of this research we will discard the repulsive forces for all other representations.

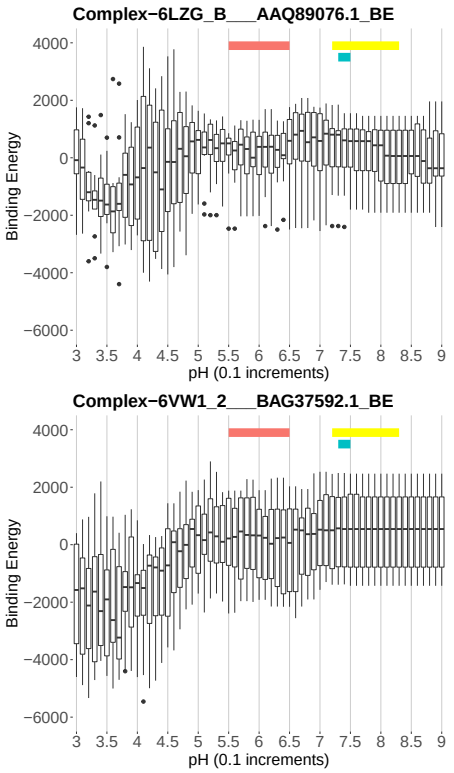

Figure 2: Predicted electrostatic contribution to binding energy across the physiological pH range shows some activity around nasal and lung pH levels in the most active plot (top), while a nearly flat result across the interesting range of pH 5.0 to 8.5 is observed in the most inactive plot (bottom).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

Figure 6 is the same information provided by Figure 5 with the repelling forces clipped. From this representation we can see unique characteristics emerging where a group of complexes stand out from the rest with low potential energy 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 clear understanding is provided by Figure 7 grouped by 's1.' Clear patterns emerge that indicate higher forces presents for 's1' 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. 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.

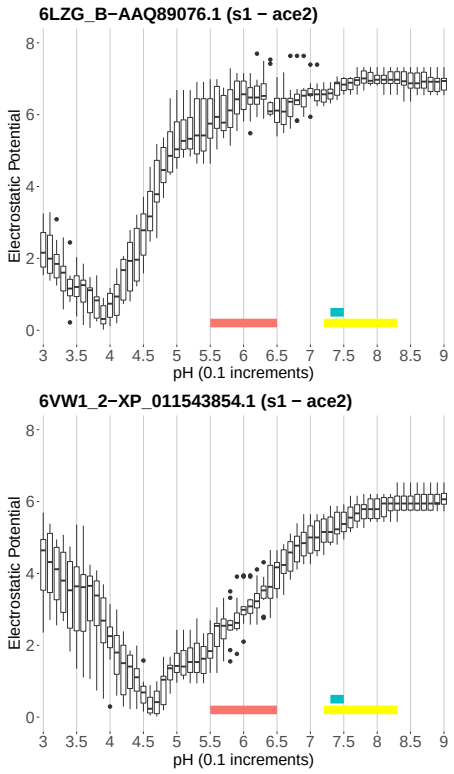

Figure 3: Comparison of potential differences between 's1' and ACE2 at result extremes contrasting the highest potential differences (top) with the lowest (bottom). Color bars indicate normal nasal secretion pH (red), inflamed nasal secretions (yellow) and normal lung pH (blue)

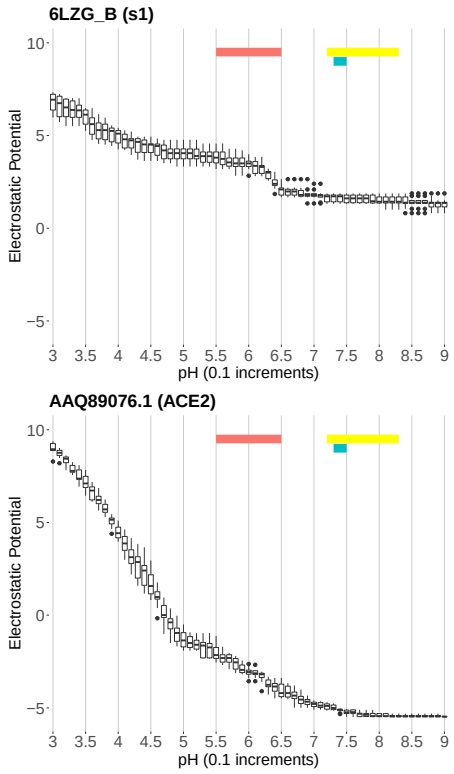

Figure 4: Typical APBS results for 's1' (top) and ACE2 (bottom). In all cases, 's1' never crosses zero while ACE2 always transitions into negative values. Color bars indicate normal nasal secretion pH (red), inflamed nasal secretions (yellow) and normal lung pH (blue)

3 DISCUSSION

Our results clearly predict that variations of CoV-2 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 with large spikes in force at pH 5.5 to 6.5 associated with nasal secretions. Three 's1' variants have a relatively smooth curve at nominal forces and one variation has a smooth curve with low force and minimal differences across ACE2 variations.

While we are limited in the available crystal structures required to analyze through our pipeline, the current level of diversity allowed for detailed analysis adequate to express differences that expose potential alignments of increased force for certain combinations of CoV-2 and ACE2 variants. Additional crystal structures of larger diversity for both proteins would allow a greater understanding.

We acknowledge the fact that our model is non-specific in that a point-source charge is used to extract the forces involved, however, this is a common scenario in such calculations and is a logical use of point source data returned by APBS. An analysis of the same nature at the amino acid level would provide more specific information and is the next logical step in the discovery process.

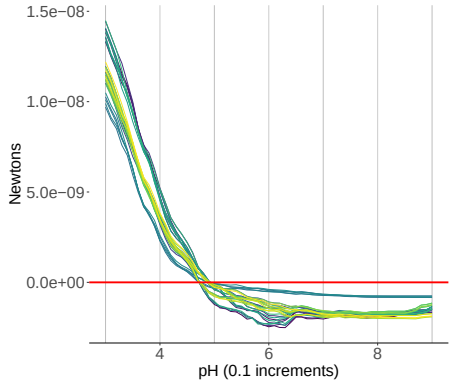

Figure 5: Aggregation of electric force for all complexes. Values above 0 (red line) are repulsive and values below are attractive.

We provide access to the full set of graphs generated from this data as a tar package located at: SARS-CoV-2

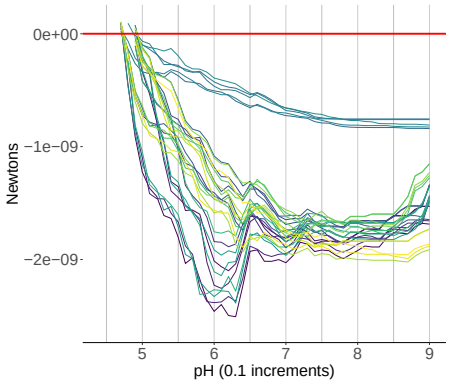

Figure 6: Aggregation of electric force for all complexes analyzed minus repulsive forces. This representation allows for a broad interpretation of the data where a single group of complexes stands out from the rest and another group has large spikes at nasal secretion pH levels.

4 METHODS

All computational methods related to the electrostatic pipeline are detailed in [5, 6, 9–12]. For the specific calculation of Electric Force, APBS returns a factor for result of the following formula:

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

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^2 is the distance in Meters between the two masses.

To derive Coulombs of charge from APBS requires a simple transposition of the terms V and e_c and multiply results with the 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] 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

[9] 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

[10] 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

[11] 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

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

[13] 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* 0, 0 (mar 2020). https://doi.org/10.1016/j.cell.2020.02.058

[14] 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

[15] 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/

[16] 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

[17] 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

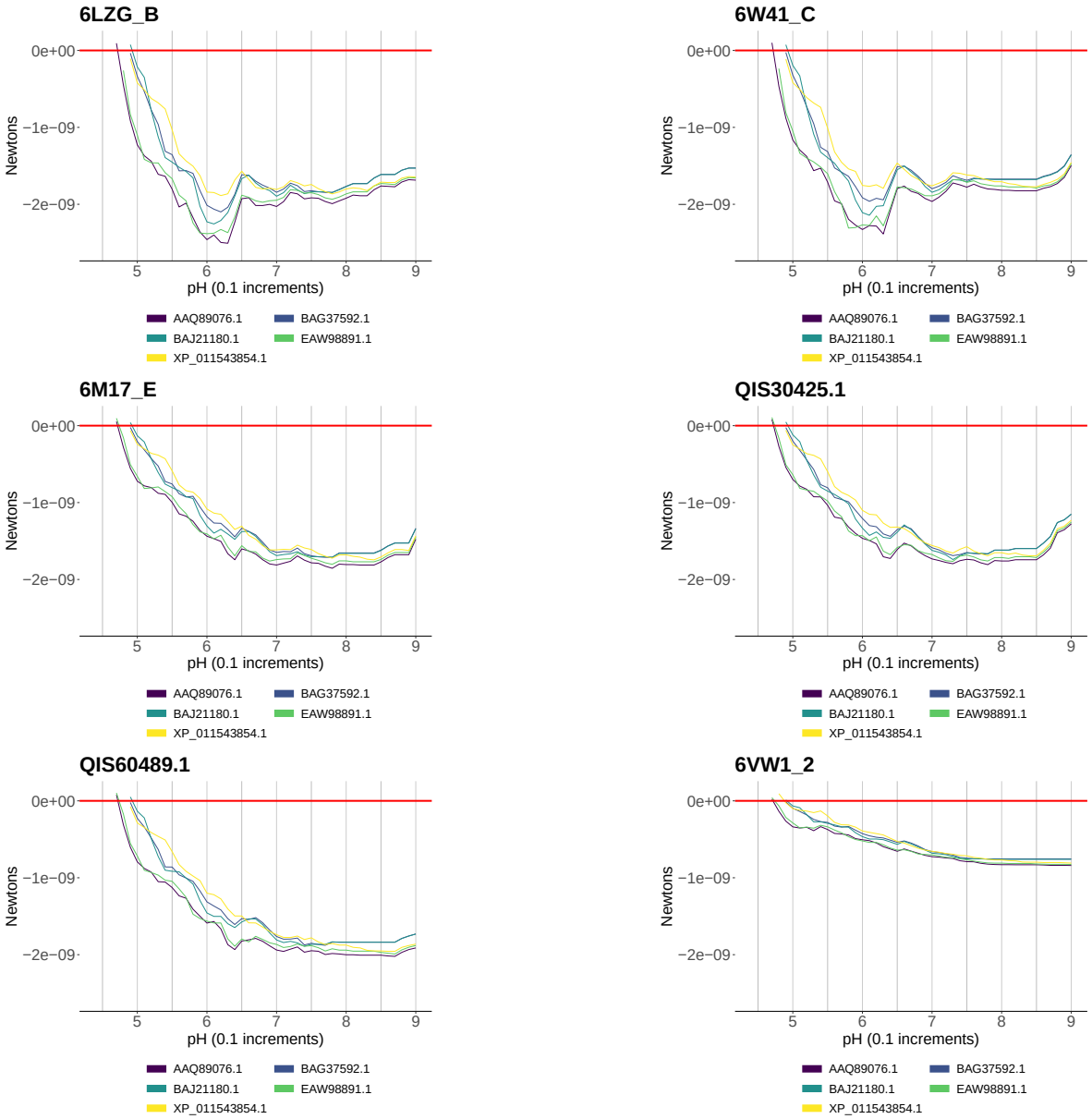

Figure 7: Grouped by 's1', patterns emerge that indicate higher forces for 's1' proteins 6LZG_B and 6W41_C at nasal secretion pH and distinct bias for ACE2 variants AAQ89076.1 and EAW98891.1, a lesser bias for ACE2 BAJ21180.1 and BAG37592.1. CoV-2 variants 6M17_E, QIS30425.1, and QIS60489.1 have nominal 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.