Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
Phase transitions may explain why SARS-CoV-2 spreads so fast and why new variants are spreading faster

J.C. Phillips a, Marcelo A. Moret b, Gilney F. Zebende c, Carson C. Chow d,∗

a Department of Physics and Astronomy, Rutgers University, Piscataway, NJ 08854, United States of America
b SENAI CIMATEC Salvador, BA, Brazil
c Department of Physics, State University of Feira de Santana, BA, Brazil
d Mathematical Biology Section, NIDDK, NIH, Bethesda, Md 20892, United States of America

A R T I C L E   I N F O

Article history:
Received 19 February 2021
Received in revised form 26 October 2021
Available online 12 April 2022

Keywords:
Proteins
Spike
Virus
Evolution
Vaccines

A B S T R A C T

The novel coronavirus SARS-CoV-2 responsible for the COVID-19 pandemic and SARS CoV-1 responsible for the SARS epidemic of 2002–2003 share an ancestor yet evolved to have much different transmissibility and global impact. A previously developed thermodynamic model of protein conformations hypothesized that SARS-CoV-2 is very close to a new thermodynamic critical point, which makes it highly infectious but also easily displaced by a spike-based vaccine because there is a tradeoff between transmissibility and robustness. The model identified a small cluster of four key mutations of SARS-CoV-2 that predicts much stronger viral attachment and viral spreading compared to SARS-CoV-1. Here we apply the model to the SARS-CoV-2 variants Alpha (B.1.1.7), Beta (B.1.351), Gamma (P.1) and Delta (B.1.617.2) and predict, using no free parameters, how the new mutations will not diminish the effectiveness of current spike based vaccines and may even further enhance infectiousness by augmenting the binding ability of the virus.

Published by Elsevier B.V.

1. Introduction

The novel coronavirus SARS-CoV-2, responsible for the COVID-19 pandemic, enters cells by binding to the Angiotensin-converting enzyme 2 (ACE2) receptor [1,2]. The binding is done via a structure called the spike (S), which is a glycosylated fusion protein that operates as a trimer. The S glycoprotein rests in a metastable prefusion state that must undergo conformational changes before the virus can fuse to the cell membrane. Given the importance of dynamics for S function it must be understood as a thermodynamic object immersed in water, which is extremely difficult to compute ab initio. One approach to circumvent this intractability is to apply a hydrophobic score to the residues of the protein with the idea that hydrophobic residues are more likely to be in the interior while hydrophilic residues are more likely to reside on the exterior. This approach has been attempted many times with varying success [3]. We adopt an approach used previously that predicted that vaccines based on S would be very effective and finding four key mutations that enhance viral attachment and infectiousness [4]. This is nontrivial as vaccine efficacy is not always guaranteed. For example, flu vaccine efficacy is usually around 50% [5].

It has long been hypothesized that some biological systems including proteins may extract important functional benefits from operating at the edge of instability, halfway between order and disorder, i.e., in the vicinity of the critical
Fig. 1. (a) Moret and Zebende [3] began with a set of 5526 protein segments of varying length M up to 45. A few short examples are shown here, with their amino acid side chains. (b) These segments were grouped into 20 subsets. Each subset has the same amino acid at its center, so each subset contains about 250 segments. (c) The amino acids side chains are each surrounded by a sphere with radius set by van der Waals interactions. Where the spheres overlap, they are cutoff by planes equidistant from their chain contacts (dotted lines). The surface area of the central amino acid that is accessible to water molecules is then calculated. These surface areas are then plotted for each amino acid as functions of chain length M. Increasing M decreases surface areas as the chains fold back upon themselves. These decreases were fitted well by power-laws; such power-laws are known to be characteristic of fractal structures and second-order phase transitions very close to critical points [6,7].

point of a phase transition [6–8]. Proteins act as folded self-organized networks that must obey steric and topological constraints. In particular, inside and outside shape extremals as measured by the interactions of amino acid side groups with water are paramount [7,9]. Each S chain consists of over 1200 amino acid residues, with approximately 300 of these having mutated from SARS-CoV-1 (CoV-1, GenBank: AY278741, UniProtKB: P59594) to SARS-CoV-2 (CoV-2, GenBank: MN908947.3, UniProtKB: P0DTC2) [10]. Of the 300 mutations between CoV-1 and CoV-2, the previous application of the model predicted that four spike mutations affected its infectiousness [4]. The variants contain twelve or fewer spike mutations. Here, we show how the additional S amino acid mutations in the variants will not diminish the efficacy of existing S-based vaccines while increasing the binding ability of the Delta variant [11]. Our work continues a long tradition of applying concepts from thermodynamics and statistical mechanics to virus and virus–host interactions [12–21].

2. Materials and methods

The method utilizes a thermodynamic amino acid scale [3] that considers the loss of solvent-accessible surface area (ASA) surrounding a central residue. Moret and Zebende [3] found in over 5000 different protein segments that the ASA for a specified amino acid residue at the center of a protein fragment scales as a power-law of the length of the fragment with a well-ordered negative exponent for all 20 amino acids (See Fig. 1 for a schematic of the analysis and Table 1 for the values). Power-law scaling is the hallmark for the critical point of a second order phase transition and the existence of universal amino acid–water interaction parameters is evidence for a thermodynamic phase transition model for novel coronavirus function [6–8]. The average accuracy of each of the 20 exponents is better than 1% [3]. In the absence of a phase transition critical point, the likelihood that 20 such accurate power-law exponents would coexist independently is astronomically small (less than \(10^{-40}\)). To our knowledge, proteins are the only large-scale networks that exhibit both first-order unfolding phase transitions and second-order conformational phase transitions described by fractals.

The 20 exponents give a measure of the average hydrophathy of each residue at the center of an arbitrary background neighborhood. The smaller the magnitude of the exponent, the more hydrophilic the amino acid will be on average. This differs from the many attempts to assign hydrophathy based on chemical properties of an amino acid in isolation [3]. Hydrophilic residues are more likely to reside near the outside of the protein and vice versa for the more hydrophobic. However, given that the residues are situated sequentially on a chain, the effect of neighboring residues must be considered. This requires deducing an effective domain length that dominates the protein’s conformation, whose details at the molecular level are unknown. This length may differ from protein to protein although there may be preferred lengths. It captures the scale at which dynamics of the protein may take place at a scale larger than secondary structures such as alpha helices. We conjecture that natural selection acting on a given class of proteins, such as the novel coronavirus spike, favors sequences such that the protein operates close to a thermodynamic critical point at an optimized length scale. Evidence for positive selection has been demonstrated in CoV-2 compared to CoV-1 [4]. Nearer the critical point,
Table 1
The shifted and rescaled hydropathic values $\psi$ for the Moret Zebende (MZ) and Kyte-Doolittle (KD) scales. The overall correlation is 86%, and in practice the differences are large enough to be reflected in protein function. For different proteins, one or the other of the two scales is better. Here and previously only the MZ scale gives synchronous edges that optimize conformational changes for faster spreading for CoV-2 and its mutants.

|     | MZ  | KD  | $(\text{KD} - \text{MZ}) / (\text{KD} + \text{MZ})$ |
|-----|-----|-----|-----------------------------------------------------|
| A   | 157 | 200.5 | 0.12                                               |
| C   | 246 | 214.2 | −0.07                                              |
| D   | 87  | 96.0  | 0.05                                               |
| E   | 94  | 96.0  | 0.01                                               |
| F   | 218 | 220.2 | 0.00                                               |
| G   | 156 | 157.1 | 0.00                                               |
| H   | 152 | 102.0 | −0.20                                              |
| I   | 222 | 253.7 | 0.07                                               |
| K   | 69  | 88.2  | 0.12                                               |
| L   | 197 | 239.9 | 0.10                                               |
| M   | 221 | 202.4 | −0.04                                              |
| N   | 113 | 96.1  | −0.08                                              |
| P   | 121 | 133.5 | 0.05                                               |
| Q   | 105 | 96.0  | −0.04                                              |
| R   | 78  | 76.4  | −0.01                                              |
| S   | 100 | 149.2 | 0.20                                               |
| T   | 135 | 151.2 | 0.06                                               |
| V   | 238 | 247.7 | 0.02                                               |
| W   | 174 | 147.3 | −0.08                                              |
| Y   | 222 | 139.4 | −0.23                                              |

it can function more reversibly and at lower temperatures [6]. This conjecture has been applied successfully to other proteins [9, 22].

We thus posit a mechanism for the increase in infectiousness of CoV-2 based on synchronized attachment [4]. The analogy is to a network of coupled oscillators that have a transition between synchrony and asynchrony. The transition is facilitated when the intrinsic frequencies of the oscillators are more symmetric allowing the network to sit nearer to the critical point [23]. In the case of the oscillators, the bifurcation parameter is the coupling strength between the oscillators. As the oscillators become more symmetric the critical transition strength becomes smaller and the network synchronizes much more readily. We note that we use infectiousness here only to imply more effective binding of the virus to the cell to the ACE2. Transmissibility could also increase due to an improvement in the replication ability of the virus within the host; our model does not directly address this very important possibility.

The predictions are made by computing the coarse-grained (shifted and rescaled, see Table S1) magnitude of the exponent along $S$ over a domain length $W$ to create a matrix of hydropathy scores, $\Psi(R, W)$, where a larger value means that the residue at site $R$ acts more hydrophobically. $W$ is the width of a sliding window centered on each $R$ to average nearby interactions. $\Psi(R, W)$ fluctuates between hydrophilic to hydrophobic values with decreasing amplitude and increasing length scale as $W$ increases. Hence, $W$ sets a length scale for which protein dynamics act at the tertiary and quaternary structural level. This scale is an emergent property of the protein which we deduce by finding the $W$ that minimizes the coefficient of variation (standard deviation divided by mean) between six isolated local hydropathic minima (hydrophilic maxima). The minima were the nearest local minima corresponding to those in CoV-2 at a window length of $W = 35$, which was used previously [4]. These minima are often connected to observed static structural features in other proteins [7]. The conclusion that the spike attachment dynamics are synchronized by hydrophilic extrema is preserved [4].

3. Results

Fig. 2 shows the hydropathy score as a function of $W$, $\Psi(R, W)$, for CoV-1, CoV-2, Alpha, and Delta [24]. The score trends downwards and reaches a minimum at a $W$ near 40. However, there is a clear difference in the curves for CoV-1 versus CoV-2 and its variants. CoV-2 declines faster to a more defined minimum. In fact, CoV-1 appears to have two minima.

Fig. 3 shows the profile of $\Psi(R, W)$ as a function of $R$ at the optimal $W$ for the same strains as Fig. 2. (The profiles of the other variants are similar to CoV-2). The notable features include a large hydrophobic maximum near residue 1231 and 6 deep minima (hydrophilic maxima) between 400 and 1200. The large hydrophobic maximum is located in the virus transmembrane domain and likely acts as an anchor. The minima represent the boundaries or edges of hydrophobic regions. Minimum 1 (456 positions are for CoV-2) is located in the Receptor Binding Domain (RBD), 2 (572) in C-Terminal Domain 1 (CTD1), 3 in the Linker between S1 and S2 (692), 4 in the Fusion peptide in S2 (792), 5 (937) in heptad repeat 1 (HR1), and 6 (1163) in the Linker to the stem of S2 [25]. Although these minima need not necessarily be associated with any observable structural features, they seem to be located in salient regions for dynamics. The minima are much more symmetric in depth for CoV-2 and the variants compared with CoV-1 (see Table 1). This symmetry is a signature of a critical point and lends the hydrophilic domains more conformational mobility, particularly for acting in synchrony. The
Fig. 2. Hydropathy as a function of window length, W. The optimal W became smaller for CoV-2 and variants compared to CoV-1. Note that W is always odd by convention.

Fig. 3. Hydropathic score $\Psi(R,W)$ for CoV-1, CoV-2, Alpha, and Delta, at the optimal W. The six local hydropathic minima (hydrophilic maxima) are much more symmetric in CoV-2 and variants compared to CoV-1. Minimum 1 is located within the RBD (residues 331-524), which also contains other local minima and maxima.

Symmetry is also fragile and can be easily broken by mutations as confirmed by mutation simulations (data not shown). While only one minimum is located within the receptor binding domain (RBD) of the spike (residues 331-524), there are smaller minima and maxima within the RBD (see Fig. 3). We also note that minima 1 and 2 are near regions where several predicted antibody epitopes may reside [26].

Table 2 gives the hydropathic values for the six minima as well as the optimal W and the coefficient of variation (CV) of the minima (all six or the four deepest). There is a significant decrease in CV between CoV-1 and CoV-2 and its variants and also a decrease in W from 41 to 39 (note that W is always an odd number by definition). The CV for Delta decreased even more but that of Alpha, Beta, and Gamma increased slightly compared to CoV-2. Not much is noticeable in the hydropathy profiles in Fig. 2 between CoV-2 and the variants except that minimum 1 has shifted lower to align better with minimum 4. Such long-range or allosteric interactions occur in motor proteins, where they were quantified by hydropathic scaling [27]. They are known to occur in principle but when small they are only detectable using the hydropathic scale with 20 exponents. Equalization of hydrophobic minima was recognized as the cause of domain attachment synchronization in the evolution of CoV-1 to CoV-2 for second, but not first, order phase transitions [28,29]. While the minima have not become more leveled in Alpha, Beta, and Gamma, the hydrophobic maximum between minima 1 and 2 is higher in all
Table 2
Scores for main hydrophobic minima (based on CoV-2 sites) at the optimal window, W. CV is the coefficient of variation for the six minima and CV* is the coefficient of variation for the four most hydrophilic (deepest) minima.

|     | 1   | 2   | 3   | 4   | 5   | 6   | W  | CV | CV* |
|-----|-----|-----|-----|-----|-----|-----|----|----|-----|
| CoV-1| 137.4| 135.9| 147.2| 141.6| 137.0| 138.1| 41 | 0.03| 0.007|
| CoV-2| 141.2| 139.7| 140.2| 138.5| 139.6| 139.4| 39 | 0.006| 0.004|
| Alpha| 141.2| 137.9| 140.9| 138.5| 139.6| 139.4| 39 | 0.009| 0.006|
| Beta | 141.2| 139.7| 142.2| 138.5| 139.6| 139.4| 39 | 0.001| 0.004|
| Gamma| 141.2| 139.7| 140.2| 138.5| 139.6| 138.9| 39 | 0.007| 0.004|
| Delta| 138.1| 139.7| 139.1| 138.5| 140.3| 139.4| 39 | 0.005| 0.004|

three compared to CoV-2 and Delta. This increased hydrophobicity could possibly stabilize the virus in aerosols, which then results in higher transmissibility [29]. The near leveling of the hydrophobic peak near 227 with the receptor binding peak near 380 suggests that the two peaks could bind together to aerosol surfaces. It had been noted previously that CoV-2 mutations favored hydrophobic residues [30]. Thus, we hypothesize that Delta may have become more transmissible by improving its binding ability to ACE2 while the others have preserved their ability to bind but possibly improved function elsewhere, such as being more stable outside of the host or replicating at a faster rate within the host. We do note that Delta, which seems much more transmissible than the other variants and CoV-2, likely also has improved function elsewhere. The important point is that for all variants the changes to the S protein are slight and thus leave the spike vulnerable to existing vaccines.

The increased binding ability of CoV-2 and variants compared to CoV-1 may also partially explain why CoV-1 had a much higher case fatality rate than CoV-2 but with lower transmissibility [31]. It is well known that viral attachment in the upper respiratory tract, URT, (e.g. throat) leads to a less severe illness compared to attachment to the lower respiratory tract, LRT, (e.g. lung) infection [17]. Both CoV-1 and CoV-2 are equally severe when they attach to the LRT. The virus must pass through the URT to get to the LRT [32]. The optimized spike of CoV-2 and its variants may lead to a higher probability of binding to the URT and thus to a much higher transmissibility with lower total case fatality rate.

Could more key mutations bring the edges into better agreement, increasing transmission further? This has low probability as it would require coordinated hydrophilic mutations within the narrow range of W residues surrounding an edge. Additionally, the great success of vaccines based on S was predicted because even a small disturbance in S will tend to drive it away from the critical point [33]. For the virus to attach to a cell, S must act in a coordinated fashion and this is impaired away from the critical point. Thus, the S-based vaccines that are already available are expected to be equally successful not only for the current variants, but for any future mutation as well that increases transmissibility by moving the virus still closer to its critical point.

All S mutations are easily evaluated using our measure, which can also be used to design animal experiments to test S mutations for their nearness to criticality and thus transmissibility. Our score evaluates transmissibility due to efficacy of viral attachment to cells but does not address the effect of mutations that may alter mechanisms after the viral material has entered the cell, such as replication rate. Finally the predictions (with no new parameters) were made possible through experience gained from the many previous studies that utilized global 21st century protein databases [34].

Julia code used for all computations is available at https://github.com/ccc1685/SARS-CoV-2-Spike.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We thank Mario Borgnia for helpful discussions. CC was supported by the Intramural Program of the NIH/NIDDK, Brazil. MAM received financial support from the National Council of Technological and Scientific Development - CNP, Brazil (grant number 305291/2018-1).

References

[1] D. Wrapp, et al., Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation, Science 367 (2020) 1260–1263.
[2] A.C. Walls, et al., Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein, Cell 181 (2020) 281–292.e6.
[3] M.A. Moret, G.F. Zebende, Amino acid hydrophobicity and accessible surface area, Phys. Rev. E 75 (2007) 011920.
[4] J.C. Phillips, Synchronized attachment and the Darwinian evolution of coronaviruses CoV-1 and CoV-2, Phys. A 581 (2021) 126202.
[5] N.L. Mazur, et al., The respiratory syncytial virus vaccine landscape: lessons from the graveyard and promising candidates, Lancet Infect. Dis. 18 (2018) e295–311.
[6] M.A. Muñoz, Colloquium: Criticality and dynamical scaling in living systems, Rev. Modern Phys. 90 (2018) 031001.
[7] J.C. Phillips, Fractals and self-organized criticality in proteins, Phys. Stat. Mech. Appl. 415 (2014) 440–448.
[8] J.C. Phillips, Thermodynamic scaling of interfering Hemoglobin strain field waves, J. Phys. Chem. B 122 (2018) 9324–9330.
[9] J.C. Phillips, Scaling and self-organized criticality in proteins: Lysozyme c, Phys. Rev. E 80 (2009) 051916.
[10] Severe acute respiratory syndrome coronavirus 2 (2019-nCoV) (SARS-CoV-2). https://www.uniprot.org/taxonomy/2697049.
[11] N.G. Davies, et al., Estimated transmissibility and severity of novel SARS-CoV-2 Variant of Concern 202012/01 in England. 35.
[12] S. Katen, A. Zlotnick, The thermodynamics of virus capsid assembly, Methods Enzymol. 455 (2009) 395–417.
[13] P. Ceres, A. Zlotnick, Weak protein-protein interactions are sufficient to drive assembly of hepatitis B virus capsids, Biochemistry 41 (2002) 11525–11531.
[14] J.M. Casasnovas, T.A. Springer, Kinetics and thermodynamics of virus binding to receptor. Studies with rhinovirus, intercellular adhesion molecule-1 (ICAM-1), and surface plasmon resonance, J. Biol. Chem. 270 (1995) 13216–13224.
[15] P. Gale, Thermodynamic equilibrium dose–response models for MERS-CoV infection reveal a potential protective role of human lung mucus but not for SARS-CoV-2, Microb. Risk Anal. 16 (2020) 100140.
[16] P. Gale, Towards a thermodynamic mechanistic model for the effect of temperature on arthropod vector competence for transmission of arboviruses, Microb. Risk Anal. 12 (2019) 27–43.
[17] P. Gale, How virus size and attachment parameters affect the temperature sensitivity of virus binding to host cells: Predictions of a thermodynamic model for arboviruses and HIV, Microb. Risk Anal. 15 (2020) 100104.
[18] G. Mahmoudabadi, R. Milo, R. Phillips, Energetic cost of building a virus, Proc. Natl. Acad. Sci. 114 (2017) E4324–E4333.
[19] S. Tzili, M. Deserno, W.M. Gelbart, A. Ben-Shaul, A statistical-thermodynamic model of viral budding, Biophys. J. 86 (2004) 2037–2048.
[20] M. Popovic, M. Minceva, Thermodynamic insight into viral infections 2: empirical formulas, molecular compositions and thermodynamic properties of SARS, MERS and SARS-CoV-2 (COVID-19) viruses, Heliyon 6 (2020) e04943.
[21] M. Popovic, M. Minceva, A thermodynamic insight into viral infections: do viruses in a lytic cycle hijack cell metabolism due to their low gibbs energy? Heliyon 6 (2020) e03933.
[22] J.C. Phillips, Darwinian evolution of taste, 2020, ArXiv200812984 Q-Bio.
[23] C. Chow, Phase-locking in weakly heterogeneous neuronal networks, Phys. Nonlinear Phenom. (1998).
[24] Preliminary genomic characterisation of an emergent SARS-CoV-2 lineage in the UK defined by a novel set of spike mutations, in: Virological, 2020, https://virological.org/t/preliminary-genomic-characterisation-of-an-emergent-sars-cov-2-lineage-in-the-uk-defined-by-a-novel-set-of-spike-mutations/563.
[25] Y. Cai, et al., Distinct conformational states of SARS-CoV-2 spike protein, Science 369 (2020) 1586–1592.
[26] M. Sikora, et al., Computational epitope map of SARS-CoV-2 spike protein, PLOS Comput. Biol. 17 (2021) e1008790.
[27] J.C. Phillips, Self-organized networks: Darwinian evolution of dynein rings, stalks, and stalk heads, Proc. Natl. Acad. Sci. 117 (2020) 7799–7802.
[28] J.S. Marvin, et al., The rational design of allosteric interactions in a monomeric protein and its applications to the construction of biosensors, Proc. Natl. Acad. Sci. 94 (1997) 4366–4371.
[29] M. Jayaweera, H. Perera, B. Gunawardana, J. Manatunge, Transmission of COVID-19 virus by droplets and aerosols: A critical review on the unresolved dichotomy, Environ. Res. 188 (2020) 109819.
[30] R. Matyášek, K. Rehůřková, K. Berta Marošiová, A. Kovařík, Mutational asymmetries in the SARS-CoV-2 genome may lead to increased hydrophobicity of virus proteins, Genes 12 (2021) 826.
[31] C.A. Donnelly, et al., Epidemiological determinants of spread of causal agent of severe acute respiratory syndrome in Hong Kong, Lancet 361 (2003) 1761–1766.
[32] R. Lu, et al., Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding, Lancet 395 (2020) 565–574.
[33] J.C. Phillips, Synchronized attachment and the darwinian evolution of coronaviruses CoV-1 and CoV-2, 2020, ArXiv200812168 Q-Bio.
[34] A.E. Guttmacher, F.S. Collins, Welcome to the genomic era, N. Engl. J. Med. 349 (2003) 996–998.