

ON THE ORIGINS OF THE OMICRON VARIANT OF THE SARS-COV-2 VIRUS

ROBERT PENNER* AND MINUS VAN BAALEN

ABSTRACT. A possible explanation based on principles of speciation for the appearance of the Omicron variant of the SARS-CoV-2 virus is proposed involving coinfection with HIV. The gist of this is that the resultant HIV-induced immunocompromise allows SARS-CoV-2 greater latitude to explore its own mutational space. This latitude is not without restriction, and a specific constraint involving free energy and backbone hydrogen bonds is explored. Certain general considerations about viral mutation are considered.

INTRODUCTION

The Omicron variant of the SARS-CoV-2 virus has very recently arisen to global concern, owing to the unprecedented number of mutations in its spike glycoprotein and their locations in key regions governing both transmissibility and antibody binding. In fact, there are 36 such mutations according to the preliminary announcement of the proteome by Tulio de Oliveira, the director of the South African Center for Epidemic Response and Innovation. This is in contrast to the 2 to 8 spike mutations typical for earlier Variants of Interest or Concern or Under Monitoring, for example with 5 such mutations for the Delta variant [7], 2 of which are in common with Omicron. As is customary, these mutations are recorded relative to a specific initial variant called Wuhan-Hu-1 (UniProt Code P0DTC2).

One powerful aspect of mRNA or virus-vector vaccines is the ability to quickly develop multivalent vaccine cargos tailored to multiple viral variants using the same delivery system but with modified cargo, presumably with the same profile of side-effects. Moderna, BioNTech and Johnson and Johnson have all announced just that, with vaccines targeting Omicron expected by early 2022. But it is facile to be confident that this will surely work. Just compare the Pfizer/BioNTech and Moderna/NIAID vaccines, which deliver different mRNA instructions

Date: November 30, 2021.

It is a pleasure to thank Charles Swerdlow for critical comments.

*Corresponding author.

for exactly the same 2P stabilized spike, and yet present different efficacies with different side-effect profiles, highlighting the subtlety of these issues. However, one can at least be hopeful that such modifications will stay abreast of the mutating viral genome.

It is clearly important to understand the genesis of this latest mutational leap from Delta to Omicron, and this is the main topic of the current short note.

There are two key points we shall develop: one is a plausible scenario, which has its roots in the theory of speciation, involving coinfection by Covid and HIV/AIDS leading to Omicron; and another is the crucial energetic role that the Delta variant seems to have played in the evolution from Wuhan-Hu-1 to Delta to Omicron. The first point is speculative, for there can be no certainty of the specific biological circumstance that first led to Omicron. The second point, however, is borne out by explicit data involving hydrogen bonds and free energy, as we shall explain.

1. RESULTS

1.1. Holey adaptive landscapes. The highest national prevalence of about 20% of HIV infection among the general population occurs in South Africa [3], (compared to less than 1% globally [14]), with a corresponding SARS-CoV-2 infection rate of 4.99% (compared to 3.36% globally) [7] and presumed double-vaccination rate of 21.7% (compared to 42.7% globally). It follows that there have been and are vast numbers of South Africans coinfecting with HIV and SARS-CoV-2.

Typically in viral coinfection [2] whether by various strains or species, the usual antigenic drift of a subject genome such as SARS-CoV-2 is accelerated via antigenic shift by recombination with coinfecting viruses. We are aware of no evidence of this so far in the current case, but the presence of a coinfecting retrovirus such as HIV (with its machinery for address, transport, pore-forming, restriction and insertion of DNA that is blind as to its cargo) may enhance the scope for recombination. This is of course untested speculation but worthy of consideration.

But it is a different interaction between HIV and SARS-CoV-2 that we highlight here. Namely, a pernicious aspect of HIV/AIDS is the immediate burden placed on the immune system by the rapidly mutating HIV virions and the concomitant overwhelmed host immune system that typically advances the mortal blow to AIDS patients [8].

In any case, the HIV-infected host immune system is weakened and becomes ineffective. This allows another coinfecting pathogen greater

latitude to explore its own mutational landscape without being terminated by the enfeebled immune system. Surely there are still biophysical constraints on what mutations are possible, and we shall confront one in the next subsection. Such mutational latitude can allow for large jumps in genome, as we have just witnessed from the Delta variant (which accounted for essentially 100% of recent Covid infections in South Africa [7] until the advent of Omicron), to the Omicron variant (which accounts for upwards of 80% of current new infections there).

An analogy from speciation is useful, for imagine a species that can feed and procreate with no particular effort or skill. In the course of generations with this lack of selective pressure, the genome is free to explore countless mutations with no particular penalty or profit from any one over any other. All bets are off, and huge leaps are possible through the mutational landscape over time.

So it is for SARS-CoV-2 in a host with weakened immune system. Given sufficiently many generations of viral replication over sufficiently many coinfecting hosts, eventually there will be large salutary deviation from the original genome with potential for increased fitness.

This is a possible explanation for the monumental leap from the Delta to the Omicron variants. Such leaps have been discussed in the general context of speciation [5], from which we quote “Rapid speciation is most likely for populations that are subdivided into a large number of small subpopulations,” as is evidently the case in this viral analogy on several different scales.

Thus, a host population with strong immune systems are likely to effectively corral the virus population in phenotype space where structural (as discussed below) and other constraints limit antigenic drift. However with sufficiently many hosts with weakened immune systems, the virus population can explore more and further in this phenotypic landscape, eventually leading to sudden antigenic leaps.

This explanation seems more plausible than the classic antigenic shift in this particular case since the mutations that define Omicron are essentially point mutations and insertions or a couple of isolated deletions of 2-3 contiguous residues. In antigenic shift, one typically expects a series of backbone-contiguous recombinations, though the situation with a retrovirus may be more subtle as was already mentioned.

1.2. Wuhan-Hu-1 to Delta to Omicron. Despite the latitude just discussed of viral mutation in a subject coinfecting with HIV and SARS-CoV-2, there are also apparently certain fundamental constraints imposed by physics and chemistry. At a basic minimum, there must not

be so much free energy as to explode the molecule (typically for proteins on the order of 8-10 kcal/mole at laboratory conditions) nor so little as to compromise the metastability and hence function of the spike glycoprotein. More stringent constraints on the considered mutation are imposed by requiring increased fitness of the variant in the host population, but this two-fold energetic constraint on free energy is sacrosanct.

Here we discuss one particular piece of empirical evidence of this energetic constraint, which was first described in [9] for the Alpha-Gamma variants, then extended in [12] to include all Variants of Concern or Interest or under Monitoring before the advent of Omicron. Here we extend the discussion to include Omicron.

In order to state this finding, recall [4] that in any protein, hydrogen bonds form between backbone Nitrogen atoms N_i-H_i and Oxygen atoms $O_j=C_j$ in different peptide units, and these are called Backbone Hydrogen Bonds (or BHBs). (To be precise: A DSSP [6] hydrogen bond is accepted as a BHB provided that furthermore the distance between H_i and O_j is less than 2.7 Å and $\angle NHO$ and $\angle COH$ each exceed 90° .) A protein residue R_i itself is said to *participate* in a BHB if either the nearby Nitrogen N_i-H_i donates to or the nearby Oxygen $O_i=C_i$ accepts a BHB. (Again to be precise, if at least two monomers of the trimeric spike participate, then the residue itself participates.) In general in a protein, it is the BHBs that strongly contribute to the stability of secondary, tertiary and quaternary structures. On average for all proteins, roughly 70-80% of all residues participate in a BHB [4].

Our novel finding on the mutagenesis of viral glycoproteins is:

Basic Finding: In the receptor-binding subunit of the spike glycoprotein, it is the residues which do not participate in BHBs that more frequently mutate.

Our theory [9, 10, 11, 12, 13] of backbone free energy (BFE), which is based in its essence on the geometry of the protein backbone, assigns an absolute free energy in kcal/mole to each BHB, as specified by the 3d structure of the protein as determined by a PDB [1] file. The most trivial way to avoid upsetting molecular BFE metastability is to mutate only those residues which do not participate in BHBs, since they are completely neutral for BFE. Insofar as this explanation for our Basic Finding is completely general, it seems reasonable to expect analogous higher mutagenic pressure on residues which are free from BHBs more generally for other metastable viral glycoproteins.

We shall quantify the Basic Finding for the three mutational steps Wuhan-Hu-1 $\mapsto$ Delta, Wuhan-Hu-1 $\mapsto$ Omicron, and Delta $\mapsto$ Omicron. To this end, we introduce terminology from [12] as follows: A residue is *missing* if it is not modeled in the PDB file (which we argue can be interpreted as disorganized for high quality PDB files), it is *absent* if it occurs in the PDB file but does not participate in either nearby backbone hydrogen bond (along the backbone), and it is *unbonded* if it is neither missing nor absent in at least 5 of the PDB files in the database for that residue¹.

Let S_1 denote the sub-molecule of the spike S from N-terminus to the furin cleavage point at residue 681, and let S_2 denote the complementary C-terminal molecule, respectively the receptor-binding and fusion subunits of S. Several trends in any of the three mutational steps we study are expected for reasons of location and function, namely: S_1 is more disorganized than S_2 (i.e., # missing is larger); there are more loops in S_1 than S_2 (i.e., # absent is larger); and the average BFE of S_1 is larger than S_2 .

Table 1 quantifies our Basic Finding under various scenarios, with the $W \mapsto (*)$ and $\Delta \mapsto O$ mutations comparable. The $W \mapsto O$ transition is anomalous with its much smaller percentage of unbonded mutated residues. The explanation follows from the last column, showing the large percentages of residues that are bonded in W but not in Δ , and hence of higher mutagenic potential for their transition to O according to the Basic Finding.

2. DISCUSSION

We have proposed a plausible explanation for the genomic leap from the Delta to the Omicron variants of SARS-CoV-2 based upon coinfection with HIV. In effect, HIV weakens the host immune system, which allows SARS-CoV-2 to more freely explore its mutational landscape, as in certain models of speciation. With sufficiently many replications within the large coinfecting population in South Africa, a mutation that allows the virus to escape from the corraling effect of its host immune system would be bound to arise eventually.

However, this greater mutational latitude is still not without biophysical constraints, one such being our Basic Finding that residues in the receptor-binding subunit S_1 of the spike have higher mutagenic

¹Our findings are based upon the 15 high-quality PDB files 6VXX 6X29 6X7 6XLU 6XM0 6XM3 6XM4 6ZB5 6ZGE 7A4N 7AD1 7DDD 7DF3 7DWY 7JWY for Wuhan-Hu-1, and the 10 very good quality and as-yet unpublished PDB files 7V7O, 7V7P, ..., 7V7V for the Delta variant.

Percentage Unbonded Residues

<u>molecule</u>	<u>$W \mapsto (*)$</u>	<u>$\Delta \mapsto O$</u>	<u>$W \mapsto O$</u>	<u>$W \mapsto \Delta$</u>
S	35	40	35	30
$\bar{S}$	61	61	22	42
S_1	42	46	42	32
$\bar{S}_1$	74	67	27	43
S_2	25	32	25	28
$\bar{S}_2$	15	33	0	33

Table 1. All table entries are percentages of unbonded residues in each molecule. W , Δ , and O are respectively the Wuhan-Hu-1, Delta and Omicron variants, and $(*)$ denotes the collection of Variants of Concern or Interest or under Monitoring [7] before the advent of Omicron with combined mutated residues (5 9 12 18-20 26) 52 67 69-70 75-76 80 95 136 138 144-145 152 156-158 190 215 243-244 246-253 346 417 449 452 478 484 490 501 570 614 641 655 677 679 681 701 716 796 859 888 899 950 982 1027 1071 1092 1101 1118 (1176), where the residues in parentheses are outside PDB-coverage and are not reflected in this table. It is Tulio de Oliveira’s preliminary collection of mutated residues 67 69 70 142-145 211 212 214 339 371 373 375 417 440 446 477 478 484 493 496 498 501 505 547 614 655 679 681 764 796 856 954 969 981 that serve to define O for our purposes. We let $\bar{S} = S \cap M$, $\bar{S}_1 = S_1 \cap M$, $\bar{S}_2 = S_2 \cap M$ respectively denote the collection of mutagenic residues M in each of the molecules S, S_1, S_2 , with M determined by $(*)$ in the second column and by O in the third and fourth columns. The last column is simply the percentage of residues which are bonded in W and unbonded in Δ in each molecule.

pressure if they do not participate in backbone hydrogen bonds. This is in keeping with general principles of free energy and follows from metastability of the spike.

This finding does not hold for the fusion subunit S_2 , possibly because the underlying PDB files reflect the prefusion conformation of the spike at relatively higher pH than the fusion subunit activation and because these PDB files model the spike molecule pre-cleavage with the associated steric constraints imposed by S_1 sitting as a cap on top of S_2 .

Assuming the Basic Finding, the transition from the original Wuhan-Hu-1 variant directly to the Omicron variant seems most unlikely, with the Delta variant a seemingly necessary intermediary. The dynamics of this is that the transition from Wuhan-Hu-1 to Delta removes a multitude of backbone hydrogen bonds thereby allowing residues to further mutate to Omicron in keeping with the Basic Finding.

3. CONCLUSION

Our findings admit explanation by general principles, and so may hold more generally, to wit: weakened immune systems allow pathogens to capitalize on large mutational leaps; being free from backbone hydrogen bonds increases the mutagenic potential within the receptor-binding subunit of a viral glycoprotein; deleting backbone hydrogen

bonds, within the constraints of molecular functionality, can increase the mutagenic potential of a glycoprotein.

REFERENCES

- [1] Berman, H.M., Westbrook, J., Feng, Z., et al. 2000. The Protein Data Bank. *Nucleic Acids Research* **28**, 235–242. Available online: <http://www.rcsb.org/pdb/>.
- [2] Dimmock, N.J. et al. (2007). Introduction to Modern Virology, 6th edition. Blackwell, Oxford, UK.
- [3] Dwyer-Lindgren, L. et al. (2019). *Nature* **570**, 189–193.
- [4] Finkelstein, A.V., and Ptitsyn, O. 2016. Protein Physics, A Course of Lectures. 2nd edition. Academic Press, London, UK.
- [5] Gavrillets, S., 1999 A Dynamical Theory of Speciation on Holey Adaptive Landscapes, *American Naturalist* **154**, 1–22.
- [6] Kabsch, W. and Sander, C. 1983, Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features, *Biopolymers* **22**, 2577–637. Available online: <https://swift.cmbi.umcn.nl/gv/dssp/>.
- [7] Mullen, J.L., Center for Viral Systems Biology, et al. 2020, outbreak.info. Available online: <https://outbreak.info/>.
- [8] Nowak, M. A. (1992). Variability of HIV infections. *J. Theor. Biol.*, 155:1–20.
- [9] Penner, R. 2020 Backbone Free Energy Estimator Applied to Viral Glycoproteins, *Journal of Computational Biology*, **27**, 10, 1495–1508, Liebert.
- [10] Penner, R. 2020 Conserved High Free Energy Sites in Human Coronavirus Spike Glycoprotein Backbones, *Journal of Computational Biology*, **27**, 11, 1622–1630, Liebert.
- [11] Penner, R. 2021, Antiviral Resistance against Viral Mutation: Praxis and Policy for SARS-CoV-2, *Computational and Mathematical Biophysics* **9**(1), 81–89.
- [12] Penner, R. 2021, Mutagenic distinction between the receptor-binding and fusion subunits of the SARS-CoV-2 spike glycoprotein, bioRxiv 2021.11.15.468283; doi: <https://doi.org/10.1101/2021.11.15.468283>.
- [13] Penner, R., et al. 2014, Hydrogen bond rotations as a uniform structural tool for analyzing protein architecture, *Nature Communications*, **5**, 5803, Springer-Nature. Available online: <https://bion-server.au.dk/hbonds/>
- [14] UNAIDS. 2021. The Global HIV/AIDS Epidemic. <https://www.unaids.org/en/resources/fact-sheet>.

INSTITUT DES HAUTES ÉTUDES SCIENTIFIQUES, 35 ROUTE DES CHARTRES, LE BOIS MARIE, 91440 BURES-SUR-YVETTE, FRANCE, and MATHEMATICS DEPARTMENT, UCLA, LOS ANGELES, CA 90095, USA

E-mail address: rpenner@ihes.fr

CNRS & INSTITUT DE BIOLOGIE DE L'ÉCOLE NORMALE SUPÉRIEURE, PARIS SCIENCE LETTRES, 46 RUE D'ULM, 75005 PARIS, FRANCE,

E-mail address: minus.van.baalen@bio.ens.psl.eu