ANTIVIRAL RESISTANCE AGAINST VIRAL MUTATION: PRAXIS AND POLICY FOR SARS COV-2

ROBERT PENNER

Abstract. New tools developed by Moderna, BioNTech/Pfizer, and Oxford/Astrazeneca, among others, provide universal solutions to previously problematic aspects of drug or vaccine delivery, uptake and toxicity, portending new tools across the medical sciences. A novel method is presented based on estimating protein backbone free energy via geometry to predict effective antiviral targets, antigens and vaccine cargoes that are resistant to viral mutation. This method, partly described in earlier work of the author, is reviewed and reformulated here in light of the profusion of recent structural data on the SARS CoV-2 spike glycoprotein and its latest mutations. Particular attention to structures computed with Cryo Electron Microscopy allow the novel approach of probing the pH dependence of free energy in order to infer function. Scientific and regulatory challenges to nucleic acid therapeutic and vaccine development and deployment are also discussed.

Introduction

Breakthrough capabilities for cellular delivery of engineered nucleic acids have overcome major hurdles [1]. A critical question is which nucleic acid cargoes to deliver for what effect. Other aspects, such as cell-specific uptake or translation promoters, will surely be further refined. One can anticipate decades of immunological and protein-(re)placement therapeutic advances [2].

First applications have been the deployment of SARS CoV-2 vaccines, whose cargo is sensibly given by nucleic acid from the virus itself. Both the Moderna and BioNTech vaccines deliver mRNA for the full spike glycoprotein, albeit a prefusion stabilized mutation K986P/V987P, called 2P, patented in 2016 in the general context of β-coronaviruses, while the adenovirus-vectored vaccines developed by Oxford and others delivers DNA instructions for the wild-type spike.

The author has no competing interests.

Key words and phrases. mRNA technology | vectored-virus | virology | vaccinology | structural biology.

Thanks to Minus van Baalen, Pablo Guardado-Calvo and Guillaume Stewart-Jones for critical comments and to François Bachelier for computer support.

We have been fortunate so far in several regards: the 2P mutation was already known; effective reverse translation from protein to nucleic acid cargo was already solved; the derived antibodies are broadly neutralizing, including penetrating glycan shielding; and the inevitable viral mutations are only recently partly escaping vaccine efficacy [3].

Several Variants of Concern (VoCs) have arisen from the classical Wuhan strain, notably the D614G mutation, which achieved global prevalence, the U.K. strain B.1.1.7, the South African strain B.1.351 and the Brazilian strain P.1. The mutations characterizing these strains are annotated in [4] and together comprise the residues numbered in Fig. 1. See SI for variants A.23.1, B.1.429, B.1.525 and B.1.526.

Recall (cf. [5]) that the intracellular state is communicated to the extracellular adaptive immune system through several pathways including the major histocompatibility complexes (MHCs) and through intact-antigen presenting dendritic cells. These complex processes are too much to review here, culminating in the eventual production of antibodies and memory cells through B- and T-cell clonal expansion, both auspiciously provoked by nucleic acid vaccines.

Several points deserve amplification. Full-molecule cargo cannot be certain to target neutralizing epitopes, especially for a pathogen with high morbidity. Though other properties pertain, it is the protein geometry that is the driving force for immunological recognition and response, in intact-antigen display and in the proteolytically derived peptides for MHC presentation, so any vaccine cargo in isolation must determine a protein that has a similar three-dimensional folded structure to that in the full molecule. Viral evasion can present decoy non-neutralizing epitopes. Viral mutation selecting for favorable traits is especially brisk for RNA viruses such as SARS CoV-2, potentially thwarting antiviral strategies.

Some of these considerations can be addressed through protein backbone free energy (BFE), a predictive method for which is given in [6], which also establishes that high BFE predicts large conformational changes such as occur during fusion. The method is applied to coronavirus spike glycoproteins in [7], leading to five explicit so-called active sites of interest, depicted in red in Fig. 1 and described by the residue triples (131,117,134), (203,227,228), (392,524,525), (1029,1034,1035) and (1059,730,731), as detailed in SI, which are conserved across seven human coronavirus spikes aligned using both high BFE and amino acid sequence, a presumptive proxy for functional alignment. The idea is that high BFE is evolutionarily conserved only in case of functional dependence, and conservation across different coronaviruses implies critical function conserved also across variants of the same coronavirus.

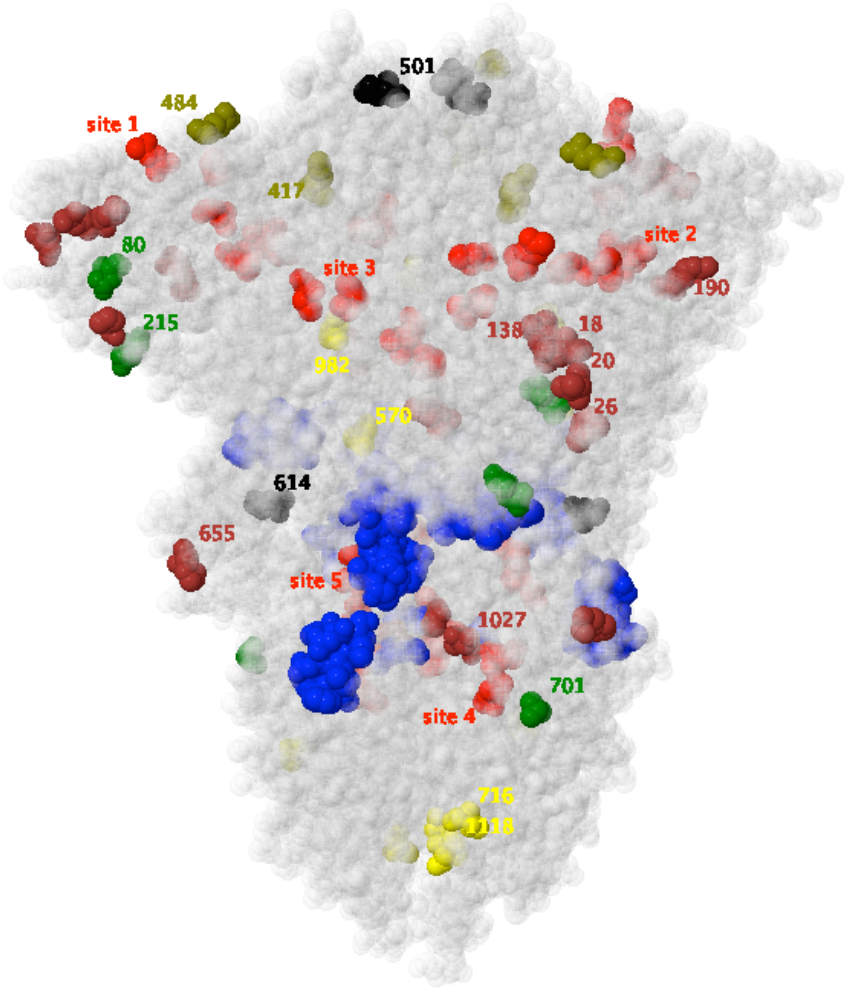

Figure 1. Mutated sites of the SARS CoV-2 spike glycoprotein for the variants of concern depicted in PDB file 7df3, which does not model residue 618. Yellow, brown and green residues are respective mutations for the U.K., Brazil and S. Africa variants, with olive residues 417 and 484 common to the latter two, and black residues 501 and 614 common to all three. Also depicted are red/blue residues for the active/passive sites of interest explained in the text.

These conserved and critical sites are presumably promising antiviral targets but are unsuitable as antigens since high BFE peptides cannot fold in isolation as they do in the full molecule, requiring low free energy scaffolding for meta-stability.

This leads to consideration of low BFE sites as targets and mutational loci and their nucleic acids as vaccine cargo. Low BFE peptides likely fold in isolation as they do in the full molecule and thus provide effective targets. Moreover by definition [6], low BFE geometry is stabilized by large numbers of peptides, so mutations should preserve the antigen structure; most mutations that occur in these regions are thus

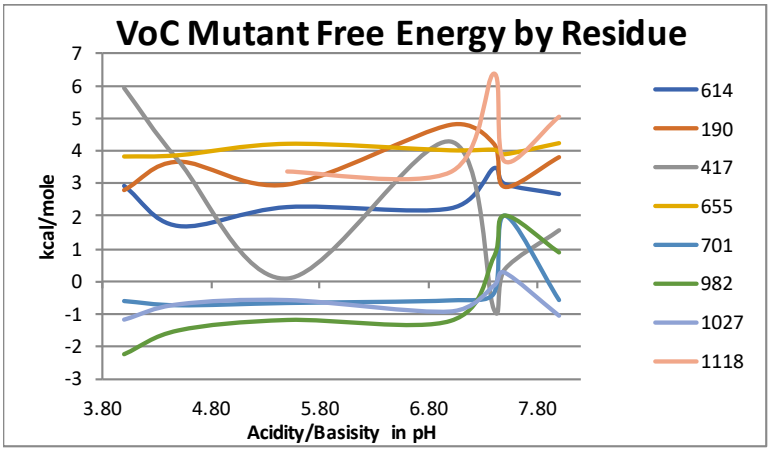

Figure 2. BFE as a function of pH for the VoC mutant residues which participate in a backbone hydrogen bond.

not critical and will not be selected by evolutionary pressures. The question thus becomes whether there are low BFE sites, so-called passive sites of interest, whose immobilization might interfere with nearby active sites.

These matters are treated in the next section in light of the proliferation of structural data and the recent appearance of the SARS CoV-2 VoCs. Scientific and regulatory strategies for effective vaccine development are finally discussed.

Results

The methods of [6] are applied to 29 structure files for uncleaved and ligand-free SARS CoV-2 spikes from [8]–[16]. First of all, this confirms the findings of [7] on active sites of interest as explicated in SI. Several among the VoC mutant residues, namely, 18, 20, 26, 484 and 681, are absent from most of these structures, cf. SI, and are subsequently ignored. One might conclude that these lie in disorganized regions of the protein which cannot be modeled, as also discussed in SI.

Furthermore, nearly half of the remaining VoC mutant residues, namely, 80, 138, 215, 501, 570 and 716, are unbonded in most structures in the sense that the C=O and N-H nearest to the C^{α} of the residue along the backbone do not participate in a backbone hydrogen bond. A plot of BFE as a function of pH for the remaining VoC mutant residues is rendered in Fig. 2. Residues 701, 928, 1027 have negative BFE, and 190, 1118, 614 have modest positive BFE in all cases.

Residue 417 has large positive BFE at low pH and low positive BFE at high pH, suggesting conformational activity in the endocytotic pathway, which is known to be acidifying, while residue 982 exhibits BFE that is very low at low pH and slightly positive at high pH, suggesting a

stabilizing influence in this pathway. Residue 1118 has no bond at pH below 5 and high BFE at higher extracellular pH, suggesting conformational activity during binding. For up/down conformations of the receptor binding domain (cf. [7]) at pH>7.0, residue 614 has respective average BFE 1.5/3.0 for Aspartic Acid and 2.6/3.2 for Glycine, consistent with the D614G mutation affecting up/down equilibrium [16] and increasing viral transmissibility [11].

Passive Sites of Interest

	site 1	site 2	site 3	site 4
Residues	295-304	816-825	949-962	1001-1028
Average	0.00	0.14	-1.16	-1.60
Percent	51.83	55.73	82.61	86.64

Table 1. Average BFE in kcal/mole and percentage of residues with negative BFE for each site.

Passive sites of interest common to the 29 structures are enumerated in Table 1 and illustrated in blue in Fig. 1. All sites occur on the molecular surface, and sites 2, 3 and 4 are nearby one another and combine into one supersite.

Discussion. It has already been argued that residues of high BFE are unlikely to mutate, since they would alter geometry and impede function, and of low BFE may randomly mutate but are unlikely to be selected by evolution, since geometry and hence function would be largely unchanged. The middle ground of no backbone hydrogen bonds and hence no BFE remains. Many of the VoC mutant residues present are of this type. Likewise, a multitude of residues are missing from all structures, suggesting that they lie in disorganized regions, cf. SI. This is a key finding here for designing vaccines targeting conserved epitopes: mutated residues (from 19 total) are typically either unbonded (6 occurrences) or disorganized (5 occurrences).

Among the remaining VoC mutant residues, three have negative free energy and three modest positive free energy, evidently with neither regime changing so substantially as to disrupt molecular meta-stability. Probable function can be inferred when the BFE depends upon pH.

These are general lessons that apply to any virus. For SARS CoV-2, it has already been argued that the existing full spike molecule vaccine target may not be optimal in light of mutations and potentially increased morbidity. Specific passive sites of interest are proposed as

an alternative, and mRNA or virus-vectored vaccine could be quickly developed targeting these sites.

As a proxy for an arbitrary virus, the current approach for SARS CoV-2 mutation is clear: keeping all other aspects fixed, substitute the spike of the VoCs for the earlier spike, whether 2P-stabilized or not. One might even sensibly deliver multi-variant cargo. In order that vaccine evolution keep pace with viral mutation, new and expedited regulatory pathways are required for the identical delivery of nucleic acid for the same protein from different variants. These remarks also apply to the targeted vaccine cargoes proposed here, though the precise regulatory specifications are more subtle.

Methods

As explained in SI, the density from [17] on rotations between planes of peptide groups participating in backbone hydrogen bonds is employed to estimate BFE as in [6, 7]. This approach is applied to the recent proliferation of structures for the SARS CoV-2 spike glycoprotein to study the VoC mutant residues.

References

[1] What the lightening-fast quest for covid vaccines means for other diseases, P. Ball, *Nature*, **589**, 7840, 16–18, 2021, Springer-Nature

[2] mRNA vaccines–a new era in vaccinology, N. Pardi et al., *Nature Reviews-Drug Discovery*, **17**, 7840, 261–279, 2020, Springer-Nature

[3] SARS-CoV-2 variants B.1.351 and B.1.1.248: Escape from therapeutic antibodies and antibodies induced by infection and vaccination, M. Hoffmann et al., *bioRxiv*, doi: https://doi.org/10.1101/2021.02.11.430787, 10.110/2021.01.11.430787, 2021, Cold Spring Harbor Laboratory

[4] A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology, A. Rambaut et al., *Nature Microbiology*, **5**, 1403–1407 2020, Springer-Nature

[5] Janeway's Immunobiology, K. Murphy and C. Weaver, 9th edition, 2017, Garland Science, Taylor & Francis Group, LLC

[6] Backbone Free Energy Estimator Applied to Viral Glycoproteins, R. Penner, *Journal of Computational Biology*, **27**, 10, 1495–1508, 2020, Liebert

[7] Conserved High Free Energy Sites in Human Coronavirus Spike Glycoprotein Backbones, R. Penner, *Journal of Computational Biology*, **27**, 11, 1622–1630, 2020, Liebert

[8] Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein, A.C. Walls et al., *Cell*, **181**, 2, 281–292, 2020, Cell Press

[9] Controlling the SARS-CoV-2 spike glycoprotein conformation, R. Henderson et al., *Nature Structural and Molecular Biology*, **27**, 10, 925-933, 2020, Springer-Nature

[10] Structure-guided covalent stabilization of coronavirus spike glycoprotein trimers in the closed conformation, M. McCallum et al., *Nature Structural and Molecular Biology*, **27**, 10, 942-949, 2020, Springer-Nature
[11] Cryo-EM Structures of SARS-CoV-2 Spike without and with ACE2 Reveal a pH-Dependent Switch to Mediate Endosomal Positioning of Receptor-Binding Domains, T. Zhou et al., *Cell Host Microbe*, **28**, 6, 867-979, 2020, Cell Press
[12] Free fatty acid binding pocket in the locked structure of SARS-CoV-2 spike protein, C. Toelzer et al., *Science*, **370**, 6517, 725–730, 2020, American Academy of Arts and Sciences
[13] SARS-CoV-2 and bat RaTG13 spike glycoprotein structures inform on virus evolution and furin-cleavage effects, A.G. Wrobel et al., *Nature Structural and Molecular Biology*, **27**, 8, 763-767, 2020, Springer-Nature
[14] Stabilizing the closed SARS-CoV-2 spike trimer, J. Juraszek et al., *Nature Communications*, **12**, 1, 244, 2021, Springer-Nature
[15] Conformational dynamics of SARS-CoV-2 trimeric spike glycoprotein in complex with receptor ACE2 revealed by cryo-EM, C. Xu et al., *Science Advances*, **7**, 1, 5575, 2021, American Academy of Arts and Sciences
[16] D614G Mutation Alters SARS-CoV-2 Spike Conformation and Enhances Protease Cleavage at the S1/S2 Junction, S.M. Gobeil et al., *Cell Reports*, **34**, 2, 108630, 2021, Cell Press
[17] Hydrogen bond rotations as a uniform structural tool for analyzing protein architecture, R. Penner et al., *Nature Communications*, **5**, 5803, 2014, Springer-Nature

Supplementary Information

Dataset Details. Since the publication of [7], there has been a proliferation of Cryo-Electron Microscopy PDB files for the SARS CoV-2 spike glycoprotein, which allow control of the pH of the sample through specified pH of the buffer at a variety of values, ranging from the acidic 4.5-5.5 of the endocytotic pathway to the 7.0-8.0 of the extracellular environment. 29 files of high quality that are ligand free and without S1/S2 cleavage were selected as the data from [8]–[16] on which to base the considerations of the paper. Specifically, the PDB accession codes of the structures studied here are: 6vxx, 6vyb, 6x2a, 6x2b, 6x2c, 6x29, 6x79, 6xlu, 6xm0, 6xm3, 6xm4, 6xm5, 6zb4, 6zb5, 6zge, 7a4n, 7ad1, 7df3, 7jwy, 7kdg, 7kdh, 7kdk, 7kdl, 7ke4, 7ke6, 7ke7, 7ke8, 7ke9, 7kea, 7keb, and 7kec. Several include various point mutations including 2P, as noted in their original literature, notably D614G in the last 10 files.

Results Details. Since the computations of [7] were performed before there was ample data on the SARS CoV-2 spike in the PDB, the first consideration here is confirmation that the active sites of interest remain so for the newly considered PDB files. This is indeed the case with the following stipulations: low pH <6.0 disrupts bifurcated hydrogen bond high BFE especially for sites 1,2,3; linoleic acid binding at pH 7.0 in 6bz4 and 6bz5 disrupts this bifurcated high BFE of all 5 sites; for site 1 even at high pH $\geqslant 7$, the mutation R685S disrupts bifurcated high BFE, and D614G shifts it to nearby residues; the structure file 6x29 at pH 7.4 inexplicably has all 5 sites disrupted.

A number of residues were ignored in part of the analysis since they were typically missing from the PDB files and hence presented insufficient data. To quantify this, here are these residue numbers followed in parentheses by the number of PDB files from which they were absent: 18(25), 20(27), 26(26), 484(20), 681(27). One might infer that these residues lie in disorganized regions of the protein, or perhaps that the experiments were less well controlled there. As an indication that the former pertains, one computes that in the specified collection of structures, the average clash score is 2.1, Ramachandran outliers amount to 0.21 percent and sidechain outliers amount to 0.79 percent, where the last datum is relatively insignificant here since the BFE depends only upon the protein backbone and its adjacent C=O.

A number of the residues are reported in the main text to be unbonded, meaning that the nearest C=O or N-H along the backbone does not participate in a BHB. To quantify this, here are these residue numbers followed in parentheses by the number of PDB files in which they are unbonded: 80(23), 138(21), 215(27), 501(28), 570(27), 716(27).

The BFE of the remaining residues is plotted as a function of pH in Fig. 2, and for these residues as well, some of the PDB files display no BHB. To quantify this, here are these residue numbers followed in parentheses by the number of PDB files in which they are unbonded: 614(0), 190(6), 417(5), 655(4), 701(1), 982(1), 1027(0), 1118(8).

The passive sites of interest all occur in alpha helices, sometimes with other secondary structure motifs at their ends. Alpha helices are not necessarily low BFE, and BFE occurs elsewhere beyond alpha helices. The percentages and averages in Table 1 of the text are computed without taking into consideration hydrogen bond donors in the first three or acceptors in the last three residues of the passive site of interest.

Further analysis (not presented), taking large/small B-factor to imply disorganization/order of structure, shows that VoC mutated residues which are present in structures are highly disorganized in S1 and ordered in S2. Moreover, the active sites of interest in S1 are not highly

disorganized and in S2 are highly ordered, while the passive sites of interest, all of which lie in S2, are also highly ordered. This is in keeping with the general trend in the prefusion spike that the membrane-distal molecular surface of S1 is largely disorganized and S2 is ordered as well as one key finding of this paper that VoC mutated residues are usually either unbonded or disorganized. It goes further to posit that the active sites of interest are not highly disorganized, and the passive sites of interest are furthermore highly ordered.

Recently reported variants B.1.525 (which shares the E484K mutation with B.1.351 and P.1), B.1.526, A.23.1, and B.1.429 lead to consideration of their 13 additional mutated residues, among which 13(21), 152(27),157(23), 253(29), 477(20), 677(29), 681(29) and 688(29) are usually missing in the same notation as before, with residue 13 missing exactly for pH $\geqslant 7$. Residue 367, which is mutated in A.23.1, is unbonded in 15 structures with no patten for BFE in pH, though D614G leads to bonding with negative BFE. These findings confirm those in the main text.

Most interestingly for its proximity to 614, residue 613, which is mutated in A.23.1 is bonded in all 29 structures and exhibits small positive BFE with no pattern in pH, just as for 614, but with up/down average BFE 2.0/1.7 for Aspartic Acid and with 2.8/2.2 for Glycine at 614. This opposite trend at 613 from 614 for the D614G mutation suggests similar functional changes for the two mutations with compensatory changes in BFE, as one might suspect in order to preserve meta-stability.

Likewise of interest, residue 452, which is mutated in B.1.429, is present in all 29 structures, is unbonded at pH below 7.4 (with 1 exception in file 7ad1 at pH 7.0) and at pH 7.5, and is always bonded at pH 8 with an average BFE of 5.87 kcal/mole that is maximal in most examples. For pH 7.4, this residue is either unbonded (in 6x2c and 6x29) or again bonded with a large average BFE of 4.37 kcal/mole (in 6x2a and 6x2b).

Thus, for B.1.429, all mutated residues are disorganized or unbonded *except residue 452*, which is always present and unbonded at neutral or low pH and bonded with extremely high BFE at pH 8. This high pH occurs in mammals only within mitichondria. Meanwhile, such high free energy is unlikely to be coincidental, suggesting that 452 may be involved in mitochondrial dysfunction, though its presence in the receptor binding domain suggests differently or additionally.

For variant B.1.526, the mutations E484K and S477N are approximately equally frequent and seemingly mutually exclusive. The BFE

of residue 477 is strikingly similar to 484, missing, so presumably disorganized, in 20 structures (19 in common with 484) and unbonded in 7 (5 in common with 484). Residue 95 is unremarkable: present in all structures, unbonded in 3, and with moderate average BFE of 2.64 kcal/mole when bonded. However, the adjacent residue 96 is quite special in that it is unremarkable at pH below 8, while at pH 8, and only at this high pH, it is bifurcated with maximum BFE in at least two chains, and often all three.

Methods

See [6, 7, 17] for further details on this overview of materials and methods. The atoms in a protein peptide group lie in a plane containing the unit displacement vector of the peptide bond. A backbone hydrogen bond (BHB) in a protein therefore determines a pair of such planes containing vectors, ordered from hydrogen bond donor to acceptor. There is a unique rotation of space carrying the first plane and the displacement vector within it to the second and preserving the natural orientations on these planes. It follows that a BHB determines a rotation of space.

A suitably unbiased and high quality subset of the Protein Data Bank (PDB) therefore determines the histogram of all such rotations for all the constituent BHBs as in [17]. A standard tool in protein theory is the so-called Pohl-Finkelstein quasi Boltzmann Ansatz, which allows the estimation of BHB free energy, called BFE in the main text, of a residue in any protein based upon this database, where in effect low density in the histogram corresponds to high free energy. Free energies in kcal/mole are estimated here in keeping with the discussion in [6]. By definition, the BFE of a residue is the maximum of the free energy of all BHBs between the C=O or N-H nearest it along the backbone. This method of estimating backbone free energy is a basic new tool from [6], of general utility across structural biology, that is employed in this paper.

Studying viral glycoproteins as a case in point where there is known to be large conformational activity during both receptor binding and fusion, the basic thesis that is proven in [6] is that if the BFE of a residue lies in the 90th percentile, i.e., exceeds 4.6 kcal/mole, then during molecular activation, at least one adjacent backbone conformational angle within one residue of it along the backbone changes by at least 180 degrees. In short: large BFE implies large backbone conformational change. The converse is not valid.

In [7], this method is applied to the spike glycoproteins of the seven human coronavirus diseases represented in the PDB. Residues of high BFE participating in bifurcated BHBs are discovered that are conserved across these different human coronaviruses in the sense that this structure persists in motifs that are nearby along the backbones aligned using sequence homology. The idea is that sequence alignment is not functional alignment, but nearly so, and this can be refined to functional alignment using this alignment of BFE. As recalled in the main text, one argues that this functional conservation should imply conservation across strains of a common coronavirus as well.

Five such conserved sites of high BFE are discovered in [7] for the SARS CoV-2 spike, given as triples (donor,acceptor,acceptor) of residues involved in high BFE bifurcated BHBs: (131,117,134), (203,227,228), (392,524,525), (1029,1034,1035) and (1059,730,731). These are called active sites of interest and colored red in Fig. 1 in the main text. The fundamental idea of [7] is that these active sites of interest provide promising antiviral targets since they are both of critical function due to high BFE and unlikely to be susceptible to mutation given also their presumed functional conservation across different coronaviruses.

The current paper tailors these tools to mRNA and vectored-virus vaccines, recognizing that active sites of interest represent unsuitable vaccine cargoes. The same tools from [6] can predict nearby low BFE peptides, the passive sites of interest colored blue in Fig. 1 and discussed in the main text, which should represent more suitable vaccine cargoes for the several reasons argued there.

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