

New tool for structural biology: backbone free energy estimator applied to viral glycoproteins

Robert C. Penner^{1,2}

¹Institut des Hautes Études Scientifiques,

Le Bois-Marie, 35 route de Chartres, 91440 Bures-sur-Yvette, France

²Mathematics Department, University of California at Los Angeles,
Los Angeles, CA 90095, USA

Earlier work analyzed the geometry of protein hydrogen bonds between peptide groups and derived from the Protein Data Bank (PDB) the associated distribution of rotations from the plane of the hydrogen bond donor peptide group to the plane of the acceptor peptide group. Here the Pohl-Finkelstein quasi Boltzmann formalism is applied to estimate the free energies of protein elements with these hydrogen bonds, and this is brought to bear in particular for viral glycoproteins as well as capsids. The 90th-plus percentiles of free energies determine residues that correlate well with viral fusion peptides, receptor binding and other functional domains in analyzed known cases and provide a novel method of predicting them from a PDB file in general. Moreover, protein primary structure correlates strongly across the free energy spectrum. The method, implemented at <https://bion-server.au.dk/hbonds/> from an uploaded PDB file, should apply still more generally to predict other protein domains with high propensity for conformational change.

Introduction

The lifecycle of a virus [1, 2] involves several activities: adsorption, entry, uncoating, transcription/mRNA production, synthesis of viral components, virion assembly and release. Here are studied the first two stages which might be more explicitly characterized as recognition/binding [3]-[5] with the host cell and the subsequent fusion/penetration [6]-[11] of cell or endosomal membrane. Recognition and binding can be by chemical cue presumably enhanced by geometric features of virion or host. This is typically followed by dramatic reformation in order to fashion characteristic fusion/penetration motifs. This reformation is triggered in manners ranging from the binding itself to acidification along the endocytic pathway sometimes with the support of onboard or native chaperones.

Viruses can be enveloped in a lipid bilayer, whose presumptive role is to evade recognition by the immune system, non-enveloped and simply enclosed in a protein capsid, or may be quasi-enveloped, that is, enveloped for only part of their lifecycle. Enveloped viruses are the best understood, and their envelopes support specific glycoproteins [12, 13] which orchestrate both recognition/binding and fusion/penetration. With this case in mind, one might rightly think of the glycoprotein as a mechanical device primed for reformation with appropriate stimuli.

In the more general context of the host organism, for instance in the bloodstream or lungs, recognition/binding and subsequent reformation must occur quickly and with specificity. Since protein hydrogen bonds in water (accounting for the conformational entropy paid to form them) lie just at the limits of stability, it is natural to imagine that their ephemeral nature could be exploited by the virus to these ends. Following this intuition leads to the consideration of

It is a pleasure to acknowledge Minus van Baalen, Thibault Damour, Misha Gromov, Pablo Guardado-Calvo, Nadya Morozova, Sergiy Garbuzynskiy and Mike Waterman for valuable comments and inspiration. But the deepest thanks are to Alexeii Finkelstein for his guidance, patience, critical readings of earlier versions of this manuscript and general mentoring through this project. Thanks are furthermore due to Karim Ben Abdallah, Jakob Trousdal Nielsen and especially Ebbe Sloth Andersen and Greg McShane for computer assistance.

hydrogen bonds in viral glycoproteins undertaken here.

The free energy of a protein feature provides a measure of its stability [14]. While most features of a protein must have low free energy in order to stabilize the structure, there are also energy defects as reflected by *exotic* features, as they shall be called here, with high free energy. Such exotic features occur only rather rarely and may arise for functional reasons.

Such defect may be tolerated, preserved by evolution and compensated by other low free energy regions, because it is required for protein function, especially in cases when the function consists of some conformational change: an unstable feature will more likely change conformation in a biologically reasonable time while a stable structure without defects would likely take too long to suitably reorganize.

Receptor binding and fusion peptides are just such a case as their function is connected with conformational change. In particular, the fusion peptide has several different conformations: first immature before virion release and usually shielded by other proteins, next in the pre-binding state, where this region is often hidden from the host immune system or may not yet be fully comprised, another in the fusion-ready state primed for activity, and a final receptor-bound state with this region extended to fuse with or penetrate the host or endosomal membrane.

These considerations lead to the scrutiny of exotic features of viral glycoproteins undertaken here. This regime is probed by applying the quasi Boltzmann formalism, observed by Pohl [15] and explained by Finkelstein et al. [16, 17], to the distribution of hydrogen bond geometry compiled from an unbiased subset of the Protein Data Bank (PDB) [18]. Hydrogen bonds of a subject protein might be analyzed and free energy differences of corresponding features computed via relative densities in the distribution with residues determined where conformational change is likely, which it has been argued should comprise significant functional domains.

In multiple cases where these regions have been determined, the method discussed here succeeds in accurately identifying them. This therefore offers the prospect of prediction in cases

where they have not been so determined. After first reviewing the background material, namely the distribution derived from PDB and its application using the quasi Boltzmann formalism, several viral glycoproteins are studied in detail to establish credibility of the method. The bulk of this manuscript is then a table of exotic residues, namely those involved in features with high free energy hydrogen bonds between peptide groups, for a multitude of viral glycoproteins as well as for several non-enveloped viral capsids.

This table of residues offers prediction of recognition/binding, fusion/penetration and other functional sites for viruses as argued above. At the very least, it provides potential experimental targets for mutational knockdown of functional domains. From another point of view, these residues may also provide appealing targets for drugs or vaccines, not only because their obstruction should interrupt function, but also because exotic peptides by their very nature occur rarely in the host organism and therefore might minimize the likelihood of drug or vaccine side-effects; however, the PDB may not truly be representative of the human proteome, and indeed it likely is not, thus mitigating the latter value of this drug or vaccine design strategy at this time.

1 Background

As introduced and developed in [19] and illustrated in Figure 1, two peptide groups sharing a backbone hydrogen bond (BHB) ordered from donor to acceptor provide a unique rotation of 3d space as is determined by an axis of rotation and the amount of rotation about it. The “collection of all 3d rotations” is abbreviated simply by the group $\text{SO}(3)$ following mathematical traditions. Mathematics furthermore endows $\text{SO}(3)$ itself with intrinsic notions of distance, angle and volume.

Upon choosing an unbiased representative subset of PDB called HQ60 for high-quality 3d structures with 60% and below homology identity which is culled from PDB using PISCES [20], one might study the histogram of all BHBs that occur, some 1166165 in number. A

Dictionary of Secondary Structure for Proteins (DSSP) [21] prospective BHB N–H :: O = C is accepted provided that furthermore $|H-O| < 2.7\text{\AA}$, $|N-O| < 3.5\text{\AA}$ and $\angle NHO, \angle COH > 90^\circ$. The results reveal that the rotations that occur for these BHBs (or as is abbreviated simply the BHBs themselves) in HQ60 occupy only about 32.5% of the volume of SO(3); this distribution in SO(3) is depicted in Figure 2. Using lower quality 3d structures and both higher and lower homology identity establish robustness of the basic properties of this distribution on SO(3) over the data employed to compute it. Moreover, one must confirm that these constraints are not simply steric in nature, and indeed in excess of 95% of SO(3) is achievable by pairs of peptide groups at the distance scale of hydrogen bonds. On the other hand, the constraints are partly quantum physical insofar as a Density Functional Theory solution of the Schrödinger equation [19] for pairs of peptide groups essentially recovers the empirically discovered region. In fact, within this subspace containing all BHBs in HQ60, there is evident grouping into 30 distinct regions, various attributes of which are given [19, Table 1]. However, this clustering is entirely immaterial to the considerations of the current manuscript. Indeed, a recent further analysis within clusters (which is not presented here) reveals that they are highly anisotropic and fail to remotely resemble a normal distribution therein, thus the attention here only on the PDB-derived distribution depicted in Figure 2. Given two peptide groups, there is not only the rotation between them, but also the displacement between their N-terminal alpha carbons, and one might wonder about including these translations as a further aspect of peptide group comparison; it was determined already in [19] that this adds nothing since the translation is essentially determined by the rotation.

2 Exotic BHBs

As explained in [14, Lecture 16], specific features of proteins obey a so-called *quasi Boltzmann* law in the sense that feature occurrence is proportional to $\exp(-F/kT_C)$, where F is the free

energy of the feature, k is the Boltzmann constant and T_C is an effective temperature, the *conformational temperature* of approximately 350 degrees Kelvin, roughly the melting temperature of protein, with kT_C about 0.7 kcal/mole at room temperature, compared to $kT = 0.6$ kcal/mole with T the temperature of about 300 in degrees Kelvin. These are not Boltzmann statistics in the usual sense of a particle visiting energy states with a probability proportional to the energy divided by $-kT$, but rather reflect the statistics of words in the alphabet of amino acids which stabilize proteins with the particular feature, cf. [14, 16, 17].

More explicitly, consider again the distribution on $\text{SO}(3)$ illustrated in Figure 2. $\text{SO}(3)$ is dissected into roughly a quarter million *boxes* of small equal Euclidean volume, and the *density* $d(p)$ at any BHB rotation p in $\text{SO}(3)$ is the number of points of the distribution in the box containing p divided by the $\text{SO}(3)$ volume of the box. Note that roughly 65% of the boxes are empty owing to the clustering of the distribution. Thus is the density determined as a function on $\text{SO}(3)$ that takes a constant value on each box. There is a point m in $\text{SO}(3)$ of highest density $d(m) = 19000$ at the rotation unsurprisingly corresponding to the ideal α helix. To fix an overall scale, the quantity

$$\Pi(p) = \ln(d(m)/d(p))$$

is taken as a descriptor of the point p in $\text{SO}(3)$. By the quasi Boltzmann ansatz, differences $\Pi(p_1) - \Pi(p_2)$ agree with free energy differences in kT_C units between protein features corresponding to p_1, p_2 .

The histogram of $\Pi(p)$ over HQ60 in kT_C units (henceforth the units kT_C in Π -values are usually suppressed) is given in Figure 3A. Taking a normalizing shift to the left of -2 kcal/mole= $-2.9kT_C$ for the nominal free energy [14] of an α helix as that of the ideal α helix, which has $\Pi = \ln 1 = 0$, the free energy for the protein feature stabilized by p is given by $[\Pi(p) - 2.9]kT_C$, and the respective Π -values 7.5, 8.5, 9.5 and 9.85 corresponding to 90th, 95th, 99th and 100th percentiles give 4.6, 5.6, 6.6 and 6.95 in kT_C units. This scheme assigns absolute free energies

to protein features and justifies computing these quantities separately for sub-units of an entire protein. This is also consistent with the most exotic protein features a few years kT_C below the limits of protein stability [14]. The Π -values themselves will be employed in the sequel. A BHB with $p \in SO(3)$ is *exotic* if $\Pi(p) \geq 7.5$, and a residue $N_i - C_i^\alpha - C_i$ is *exotic* if either N_i or C_i participates in an exotic BHB.

The distribution of DSSP flanking secondary structure types across HQ60 is shown in Figure 3B. Note the predominance of α helices for small and the mixture of all types for large free energy. The distribution of flanking primary structures for exotic BHBs across HQ60 is presented in Figure 4. Notice that certain subsets of residues vary together in certain regions of the free energy spectrum. This strongly suggests characteristic primary structure motifs of large free energy which may be uncovered using machine learning, an ongoing research initiative.

3 Findings

Several well-studied examples of fusion glycoproteins are considered in detail, namely those in the survey paper [9]. Specifically, the exotic BHBs for both pre- and postfusion conformations of fusion glycoproteins of two Class I fusion glycoproteins (Influenza HA and Paramyxovirus F), one Class II (Tick-borne Encephalitis E) and one Class III (Vesicular Stomatitis G) are compared with the fusion mechanisms explained in [9]. These data are presented in Table 1 and provide test cases for proof-of-concept of the method as explained in detail in the Supplementary Material (SM).

Table 2 provides a summary of the predictive success of the method to establish the implication: if the Π -value of the BHB $N_i-H_i::O_j-C_j$ exceeds 7.5, then at least one of the standard backbone conformational angles ϕ_i and ψ_j varies by at least one hundred degrees in the change of conformation from pre- to postfusion, that is, exotic free energy implies big nearby conformational change. Details on the multinomial p -values are given in SM Table 3.

Also presented in SM is a narrative account relative to the function described in [9], and the predictive power of the method for influenza hemagglutinin is striking: all of the predicted exotic residues prefusion are explained by the function in [9], and only these arise from the method. Narratives for the other three viruses are also substantive.

SM Table 1 provides the exotic residues for a host of viral glycoproteins in the same notation as in Table 1, where the residues lying in the generally agreed upon fusion loops are indicated in boldface. There is evidently fine agreement between the tables and the known fusion loops with a few exceptions: MERS and yellow fever postfusion, for which there are no explanations, and possibly also murine alpha corona virus and metapneumovirus. Several table entries are *not* fusion peptides but are included in the table to illuminate for instance receptor binding domains.

SM Table 2 supplies analogous data for a selection of non-enveloped viral capsids, about whose recognition/binding and penetration mechanisms much less is known [10, 11]. For the best studied polio virus, the exotic regions of VP1 adjacent to VP4 interior to the capsid are consistent with what is in the literature [22, 5], where VP1 and VP4 are implicated in penetration, although VP4 itself contains only one exotic residue. Moreover, appropriate residues in the canyon walls presumed to be associated with receptor binding [23, 5] are found to be exotic for both polio and rhinovirus. By analogy for the other entries in SM Table 2 under the assumption that penetration peptides must be shielded from the immune system, the exotic residues interior to the capsids provide natural predictions for penetration peptides as do the exotic exterior residues for receptor binding domains.

In all cases scrutinized, the tables compare favorably with the literature for example in detecting receptor binding domains. There is furthermore sometimes controversy regarding which residues are involved in specific functions, which the findings here might elucidate.

4 Discussion

Detailed examples provide credibility to the method, further substantiated since in all but the few cases mentioned already of SM Table 1, the fusion loop contains exotic residues when it is present and not disordered.

There is the general pattern that fusion peptides and receptor binding domains are exotic, the latter typically less so than the former and the fusion loop hidden prefusion as for influenza and flaviviruses or only partially composed and exposed as for vesicular stomatitis or hidden as for tick-borne encephalitis. Based on this scant circumstantial evidence, one might ask whether the host immune system can detect exotic protein features.

It is worth emphasizing that this idea of using exotic protein features to estimate free energies in order to locate conformationally active sites should surely be more widely applicable in structural biology, for example in tyrosine kinase receptors, for which there are promising preliminary results. Other seemingly natural candidates for the method include certain prion, transmembrane, signal transduction and cell motility proteins.

References

- [1] N. J. Dimmock et al., *Introduction to Modern Virology, 6th edition*, Blackwell Publishing (2007).
- [2] A. J. Levine, *Viruses*, Scientific American Library (1992).
- [3] S. Shanker et al., Structural features of glycan recognition among viral pathogens, *Current Opinions in Structural Biology* **44**, 211-218 (2017).
- [4] S. Boulant et al., Dynamics of Virus-Receptor Interactions in Virus Binding, Signaling, and Endocytosis, *Viruses* **7**(6), 2794-2815 (2015).
- [5] M. G. Rossman, Viral cell recognition and entry, *Protein Science* **3**(10), 1712-1725 (1994).
- [6] L. V. Chernomordik and M. M. Kozlov, Mechanics of membrane fusion, *Nature Structural and Molecular Biology* **15**(7), 675-683 (2009).
- [7] S. C. Harrison, Viral membrane fusion, *Nature Structural Molecular Biology* **15**(7), 690-698 (2008).
- [8] J. A. Thorley et al., Mechanisms of viral entry: sneaking in the front door, *Protoplasma* **244**, 15-24 (2010).
- [9] J. M. White, et al., Structures and mechanisms of viral membrane fusion proteins: multiple variations on a common theme, *Critical reviews in Biochemistry and Molecular Biology* **43**, 189-219 (2008).
- [10] B. Tsai, Penetration of nonenveloped viruses into the cytoplasm, *Annual Reviews of Cellular and Developmental Biology* **23**, 23-43 (2007).

- [11] C. L. Moyer and G. R. Nemerow, Viral weapons of membrane destruction: variable modes of membrane penetration by non-enveloped viruses, *Current Opinions in Virology* **1**(1), 44-49 (2011).
- [12] P. W. Choppin and A. Scheid, The Role of Viral Glycoproteins in Adsorption, Penetration, and Pathogenicity of Viruses, *Reviews of Infectious Diseases* **2**(1), 40-61 (1980).
- [13] *Viral Glycoprotein Structure*, *Viruses-Special Issue*, editor A. Ward (2015).
- [14] A. V. Finkelstein and O. Ptitsyn, *Protein Physics, A Course of Lectures, Second Edition*, Academic Press (2016).
- [15] F. M. Pohl, Empirical protein energy maps, *Nature New Biology* **234**, 277-279 (1971).
- [16] A. V. Finkelstein et al., Boltzmann-like statistics of protein architectures. Origins and consequences. In: Biswas, B.B., Roy, S. (Eds.), *Subcellular Biochemistry. Proteins: Structure, Function and Protein Engineering*, vol. 24. Plenum Press, New York, pp. 1-26. (1995).
- [17] Finkelstein et al., Why do protein architectures have Boltzmann-like statistics? *Proteins* **23**, 142-150 (1995)
- [18] J. Westbrook et al., The Protein Data Bank, *Nucleic Acids Research* **28**, 235-242 (2000).
- [19] R. C. Penner et al., Hydrogen bond rotations as a uniform structural tool for analyzing protein architecture, *Nature Communications* **5**, 5803 (2014).
- [20] G. Wang and R. L. Dunbrack, Jr., PISCES: a protein sequence culling server, *Bioinformatics* **19**, 1589-1591 (2003).

- [21] W. Kabsch and C. Sander, DSSP: definition of secondary structure of proteins given a set of 3D coordinates, *Biopolymers* **22**, 2577-2637 (1983).
- [22] T. J. Tuthill et al., Characterization of Early Steps in the Poliovirus Infection Process: Receptor-Decorated Liposomes Induce Conversion of the Virus to Membrane-Anchored Entry-Intermediate Particles, *Journal of Virology* **80**(1), 172-180 (2006).
- [23] Y. He et al., Interaction of the poliovirus receptor with poliovirus, *Proceedings of the National Academy of Sciences* **97**(1), 79-84 (2000).

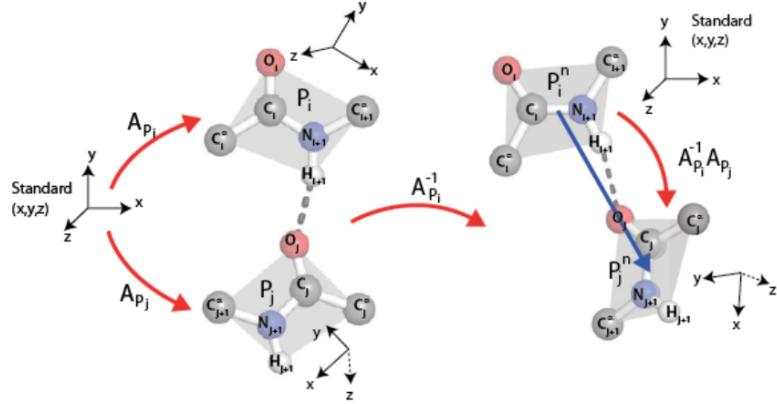


FIGURE 1. On the left are depicted two peptide groups P_i, P_j participating in a hydrogen bond with P_i the donor and P_j the acceptor. The planes of these peptide groups are illustrated in grey. Two data are not indicated here, namely, the distance $|i - j|$ of residues along the backbone and the length $|O_j - H_{i+1}|$ of the BHB. The backbone conformational angles ψ_i and ϕ_{i+1} , respectively indicating rotation about the C_i^α - C_i and N_{i+1} - C_{i+1}^α axes, are likewise not illustrated. The cross product (in this order) of displacement vectors $\overrightarrow{C_i^\alpha C_i}$ and $\overrightarrow{C_i O_i}$ determines a unit vector perpendicular to the plane of peptide group P_i , and this plane contains the unit vector parallel to the displacement vector $\overrightarrow{C_i N_{i+1}}$ of the peptide bond. The cross product of these (in this order) determines a third vector. There is a unique 3d rotation A_{P_i} mapping unit vectors parallel to the z-, x-, and y-axes, respectively, to these three vectors (in these orders), and likewise A_{P_j} for the peptide group P_j . In order to obtain a result which is independent of the position of the pair P_i, P_j in space, one applies to the entire configuration the rotation $A_{P_i}^{-1}$ as illustrated on the right and achieves the result $A_{P_i}^{-1}A_{P_j}$ as the rotation in $SO(3)$ associated to the pair P_i, P_j .

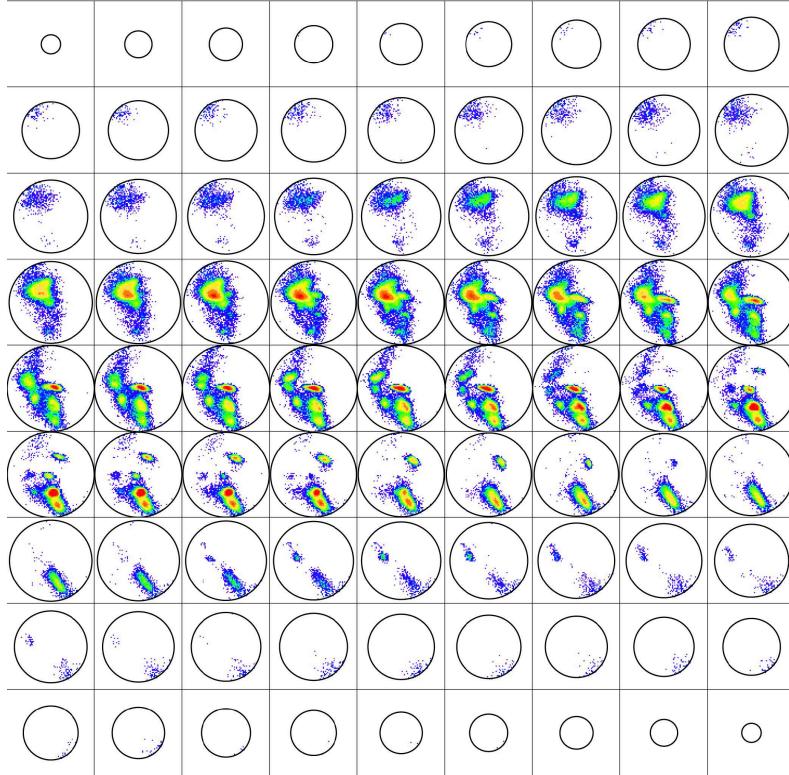
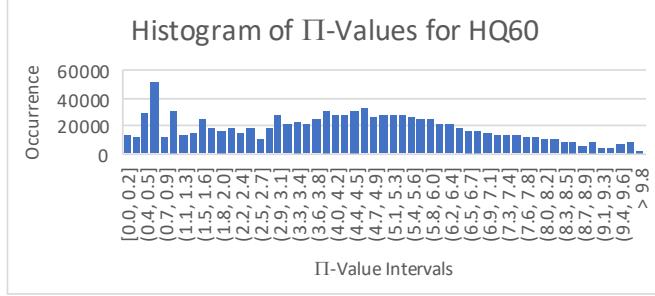
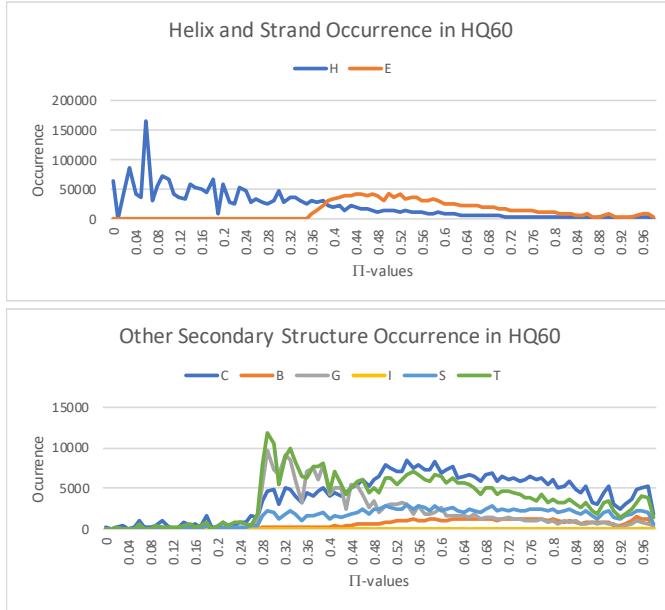


FIGURE 2. A 3d rotation is determined by an axis L of rotation and an angle $-\pi \leq \theta \leq \pi$ of rotation about it. If the unit vector \vec{u} is parallel to L , then the interval of all multiples $\theta\vec{u}$ therefore corresponds to all rotations with axis L including the trivial rotation corresponding to $\theta = 0$, where $\pi\vec{u}$ and $-\pi\vec{u}$ evidently describe the same 3d rotation, namely by π or by $-\pi$ about L . The collection $\text{SO}(3)$ of all 3d rotations can therefore be visualized as a 3d ball of radius π with each pair $\pm\pi\vec{u}$ of points in its boundary 2d sphere identified to a separate single point. Presented here are eighty-one horizontal slices of this ball from north to south pole colored by population density from [19], where the R-Y-G-B color is linear in the density ranging from 19000 to 1. This illustrates the fundamental distribution on $\text{SO}(3)$ of the BHGs in HQ60 that provides the basic dataset upon which the free energy estimates of this paper are based. One must keep in mind that pairs of antipodal points in the boundary of the depicted ball are to be identified to a single point, and this particular representation of the distribution in $\text{SO}(3)$ was chosen to minimize the density proximal to the boundary. The *ideal* (right-handed) alpha helix has its conformational angles $\phi = -65^\circ$ and $\psi = -40^\circ$ and here has its 3d rotation described by $\theta = 1.086$, $\vec{u} = (-0.315, 0.935, -0.164)$. This element of $\text{SO}(3)$ occurs at the point of highest density in HQ60, as is evident in the middle of the fourth row from the top. Other local maxima for density which are clear in the figure are studied in the cluster analysis of [19], which is not pertinent here.



(A) Histogram of $\Pi(p) = \ln(d(m)/d(p))$ for all BHBs across HQ60. The x-axis corresponds to the indicated intervals of Π -values achieved for the BHBs in HQ60, and the y-axis indicates the number of occurrences in HQ60 within each interval of size 0.18.



(B) Population of flanking DSSP secondary structure types H (α helix), E (β strand), C (coil), B (β bridge), G (3_{10} helix), I (π helix), S (bend), and T (turn) for the four residues $i, i + 1, j - 1, j$ flanking a BHB $N_i - H_i :: O_j = C_j$ across the range of Π -values along the x-axis.

FIGURE 3. Histogram of Π -values and of flanking DSSP secondary structure types across HQ60.

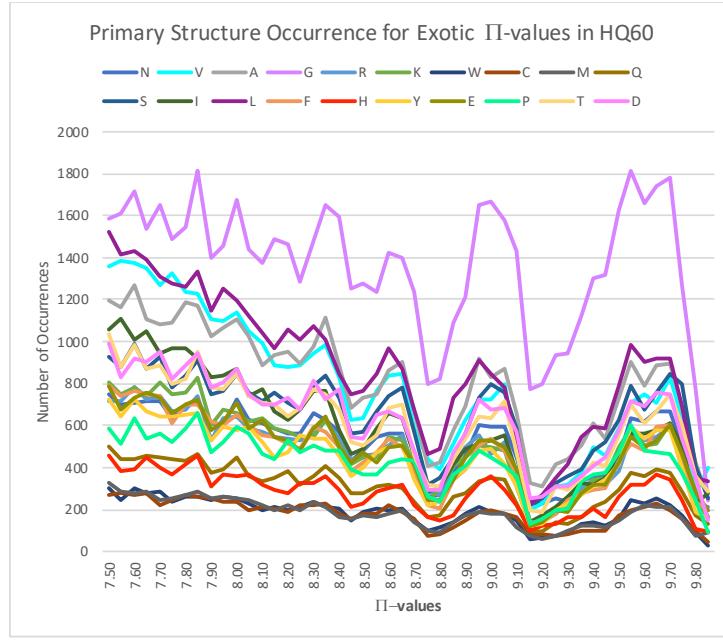


FIGURE 4. Histogram of Π -values and of flanking primary structure for all exotic BHBs across HQ60. Curves are colored by residue as indicated. Notice the comparatively frequent occurrence of glycine reflecting the presumably increasingly contorted exotic features that the primary structure must support. SM Figure 3 gives the analogous plots across the entire free energy spectrum illustrating similar phenomena.

TABLE 1. Exotic BHB Donor/Acceptor Residues of Viral Glycoproteins

Influenza Virus Type 2 Glycoprotein HA Prefusion (2HMG)
Chain E
90-94% 116/111 241/170 258/121 16/F136 256/150 86/57 158/160 221/227
95-98% 284/286 288/50 150/72 253/181 65/61 308/293 114/109 304/F62 135/153 157/194 137/146 142/144
99% F15/17 161/157 F24/16 74/68
100% 19/F21 20/F14 29/31 95/63 106/102 124/255 147/136 198/195 207/209 254/152 F63/303
Chain F
90-94% E16/136 36/24 101/96 9/5
95-98% 130/126
99% 15/E17 24/E16
100% E19/21 E20/14 10/4 11/5 63/E303 134/137 175/172
†Influenza Virus Type 2 Glycoprotein HA2 Postfusion (1HTM)
Chain F
90-94% 128/123 61/56 92/87
95-98% 63/59 107/103 134/137 50/45
99% 132/139
100% 44/40 57/52 108/102 160/157
Paramyxovirus Glycoprotein F Prefusion (2B9B)
Chain A
90-94% 259/272 170/166 241/237 408/424 269/263 334/39 295/301 294/367 98/95 423/411 362/300 352/348
95-98% 38/329 313/315 422/B106 354/351 328/330 297/299
99% 258/219 373/375 300/296 262/270 491/486 24/21 441/437
100% 26/22 92/87 94/91 95/90 96/90 160/156 188/184 189/184 264/268 353/347 374/B114 376/372 416/418 419/415
Chain B
90-94% 132/127 167/150 68/65 483/478 300/296 296/401 357/353 269/263135/130 449/445 258/219 129/124 496/491 390/412 181/60
70/66 313/315 408/424 377/405
95-98% 334/39 31/25 374/C114 113/109 376/372387/414 82/77 492/487 A422/106 145/141 315/312 297/299 319/339
99% 388/392 93/88 90/85 262/270 188/184
100% 27/23 46/275 92/87 95/90235/231 236/231 264/268 328/330 393/387 416/418 A374/114
Chain C
90-94% 484/479 271/261 29/25 384/379 416/418 220/257 377/405 373/37526/23 38/329 313/315 170/166 411/407 300/296 258/219 296/401
319/339 353/347
95-98% 297/299 387/414 95/89 157/159 31/25 269/263 94/91
99% 269/263 94/91 328/330 473/469
100% 96/90 102/96 113/109 188/184 189/184 264/268 278/43 388/392 393/387 408/424 485/480
†Paramyxovirus Glycoprotein F Postfusion (1ZTM)
Chain A
90-94% C59/443 323/319 29/25 261/257 210/205 264/281 326/346 408/404 158/153 262/256 361/358 366/363 271/275 C53/438 193/189
215/210 244/240 243/239
95-98% B229/219 254/249 84/79 423/425 395/399 289/38 363/359 229/C219 303/408
99% 360/354 167/162 185/179 307/303448/445 353/349 469/464 460/456 63/59
100% 30/26 31/25 38/336 155/150 166/161 199/194 216/211 235/231 242/238 320/322 354/349 380/382 421/427 470/465
Chain B
90-94% 53/C438 88/83 191/186 426/422 174/169 276/270 303/408 264/28195/90 462/457 C229/219 395/399 184/178271/275 234/230
95-98% 229/A219 215/210 233/228 380/382 320/322 187/182 87/83 421/427
99% 423/425 335/337 38/336 366/363 172/167 262/256361/358
100% 31/25 83/7984/79 84/80 185/179 196/192 197/192 244/240 305/405 327/315 383/379 484/479
Chain C
90-94% B53/438 59/A44329/25 301/374 155/150 42/340 235/231 31/25 83/79 463/458 233/228 271/275 151/146 229/B219 53/A438 476/472
283/45 383/380 408/404326/346
95-98% 328/344 30/26 22/203/198 387/375 185/179 184/178
99% 380/382 474/469 149/144 472/467
100% 38/336 95/90 148/143 244/240 264/281 303/408 320/322 332/328 360/354 361/358 363/359 394/421 421/427 470/465

Donor/Acceptor residues of BHBs in order of non-decreasing Π -values with 7.5, 8.5, 9.5 and 9.85 respectively corresponding to percentiles 90, 95, 99 and 100. Marked in boldface are the residues lying in generally agreed upon fusion loops. † indicates that the fusion loop is missing from the structure and therefore a fortiori contains no exotic BHBs.

TABLE 1. (continued) Exotic BHB Donor/Acceptor Residues of Viral Glycoproteins

Tick-Borne Encephalitis Virus Glycoprotein E Prefusion (1SVB)
Chain A
90-94% 218/196 188/289 386/388 330/316 389/385 65/120 177/180 308/339 322/325 339/364 355/344
95-98% 167/169 278/280 184/293380/394 372/148 388/309
99% 366/368 360/373
Tick-Borne Encephalitis Virus Glycoprotein E Postfusion (1URZ)
Chain A
90-94% 258/241 28/286 29/45 380/394 218/196 278/280 181/177 388/309 317/329 355/344 78/74 106/100 360/373
95-98% 249/251 389/385147/40 330/316
99% 322/325 180/176 243/238 366/368 371/363 15/18 9/302
100% 16/B13 167/169 177/180 192/285 251/248 252/248 C16/13
Vesicular Stomatitis Virus Glycoprotein G Prefusion (5I2S)
Chain A
90-94% 224/226 313/262 332/6 320/322 185/43 314/328 37/33 324/403 9/329 55/134 183/45 136/144 372/316 150/159 16/325 345/342 298/400
95-98% 104/98 333/208 38/190 51/47
99% 323/319 254/220 370/373 355/345 367/363 142/137 33/29 312/330 208/210
100% 138/142 146/151152/148 261/234 348/352 351/347 359/10 364/366 404/321
Vesicular Stomatitis Virus Glycoprotein G Postfusion (5I2M)
Chain A
90-94% 72/75 332/6 119/115 254/220 225/138 377/379 258/254 261/234 216/203 219/215 38/190
95-98% 312/330 374/370 71/118 34/29 153/149 404/261
99% 348/352 15/12 138/142 146/151 152/148 208/210
100% 105/98 364/366 370/373

Donor/Acceptor residues of BHBs in order of non-decreasing II-values with 7.5, 8.5, 9.5 and 9.85 respectively corresponding to percentiles 90, 95, 99 and 100. Marked in boldface are the residues lying in generally agreed upon fusion loops.

TABLE 2. Predicting Conformational Activity from Exotic Residues in Test Cases

Viral Glycoprotein	Active Hits	1-Off Active Hits	2-Off Active Hits	≥ 3 -Off Active Hits	First p-value	Second p-value	Third p-value
influenza glycoprotein HA2 chain F	2/1	2/1	0/0	0/1	6.2×10^{-3}	2.8×10^{-2}	5.2×10^{-3}
paramyxovirus glycoprotein F chain A	27/15	6/8	1/1	3/1	2.3×10^{-2}	7.2×10^{-2}	3.7×10^{-3}
tick-borne encephalitis glycoprotein E chain A	7/7	2/9	0/5	2/2	2.3×10^{-4}	1.2×10^{-1}	2.0×10^{-3}
vesicular stomatitis glycoprotein G chain A	17/4	12/15	2/10	3/9	4.8×10^{-1}	4.8×10^{-3}	2.8×10^{-25}

For each of the four viral glycoprotein single chains considered in [9], it is tested whether an exotic BHB $N_i - H_i :: O_j - C_j$ is active, namely, effects a change of at least 100 degrees in at least one of the nearby conformational angles ϕ_i and ψ_j . Active Hits indicate the number of exotic residues which are active, and k -Off Active Hits indicates the number of exotic residues whose nearest active residue is $k \geq 1$ residues away. Results are listed as fractions Dissipative/Conservative, where dissipative indicates that the free energy is no longer exotic postfusion, and conservative indicates that the free energy remains exotic postfusion, each determination within one residue of the prefusion exotic residue. For each row, the number of predictions is the sum over second through fourth columns of the number of dissipative plus the number of conservative trials. Probabilities and p -values are computed with the trinomial distribution discussed in SM. The first p -value tests the significance of the implication: if a residue is exotic prefusion, then it is at most one residue away from an active residue; for the second p -value, all conservative results are discarded; and for the third p -value, only conservative negative results are discarded. It is arguably only the dissipative case that provides possible false positives. The method is quite effective with acceptable first p -values except for vesicular stomatitis glycoprotein G, which is remarkable since it can oscillate between its pre- and postfusion conformations. This is reflected in its conservative high free energy residues for which there is no large nearby conformational change, hence the corrections in this case with second and third p -values.

Supplementary Material

1. METHODS

The server <https://bion-server.au.dk/hbonds/> for a given PDB file returns a list of its BHBs together with the adjusted density $\Pi(p)$ relative to HQ60 of each BHB as previously discussed. The server also supplies a PyMol file of residues colored by free energies from white to red for small to large Π -values in the given protein. The BHBs are then rank-ordered by Π -values, and only those exceeding the percentile cutoffs are considered.

Extensive tables of viral glycoproteins in SM Table 1 and capsids in SM Table 2 are presented in the same notation as Table 1 in the main text. Indicated in boldface are the residues lying in generally agreed upon fusion loops, which are taken ± 2 residues to reflect uncertainty in precise peptide boundaries.

A few words are in order about the method in general and these tables in particular: Comparison with the residue B-values reported in the PDB files should be taken into account with large B-values (which measure the disorder of the protein, cf. [O. Carugo, How large B-factors can be in protein crystal structures, *BMC Bioinformatics* **19**(61) (2018)]) at a residue presumably casting potential doubt upon the verity of the reported high free energy. At the same time, the exotic residues determined here for PDB files with large reported resolutions might likewise be questioned though an extensive study of influenza virus type A hemagglutinin (not reported here) found the resulting exotic residues basically insensitive to reasonable resolutions, say, less than about 3.5-4.0 Angstrom. Note that the server very rarely, perhaps two percent of the time, returns data for only some of the constituent chains, explaining the few omissions of certain PDB files here.

2. COMPUTATION OF p -VALUES

Turning next to the computation of p -values, SM Table 3 provides the data upon which are based the p -values of Table 2 in the main text. For each of the four examples let R denote the number of residues and a, b and c denote the respective number of residues further than one away from an active residue, the number of active residues and the number of inactive residues adjacent to an active one, so the trinomial probability density of n_a, n_b, n_c respective results is given by

$$P(n_a, n_b, n_c) = \binom{n_a + n_b + n_c}{n_a \ n_b \ n_c} \left(\frac{a}{R}\right)^{n_a} \left(\frac{b}{R}\right)^{n_b} \left(\frac{c}{R}\right)^{n_c}.$$

These data R, a, b, c are reported in the first four columns of SM Table 3 and then used in this way to compute probabilities for the trials given in Table 2. The computation of p -values requires a linear ordering for tails, which is naturally given by $(n_a, n_b, n_c) \leq (n'_a, n'_b, n'_c)$ provided $n_a \leq n'_a$ with equality only if $n_b \geq n'_b$ with equality only if $n_c \geq n'_c$, or in other words, lexicographic ordering on triples (n_a, n_b, n_c) derived from \leq on the first and \geq on the remaining two entries.

3. NARRATIVE DISCUSSION OF TEST CASES

For the test cases, essentially all of the prefusion exotic residues in Table 1 of the main text accord perfectly with expectations from [J. M. White, et al., Structures and mechanisms of viral membrane fusion proteins: multiple variations on a common theme, *Critical reviews in Biochemistry and Molecular Biology* **43**, 189-219 (2008)] as next detailed.

Figure 2 depicts the various glycoproteins aligned to Figure 6 of [ibid] to which it should be compared. The color scheme here is that blue indicates non-exotic, yellow above the 7.5 cutoff, orange above the 8.5 cutoff and red above the 9.5 cutoff for II-values, which respectively correspond to the 90th, 95th, and 99th percentiles. As a notational convenience for this discussion, if the residue N is involved in an exotic BHB, then one writes Nb, Ny, No, Nr to indicate this discretization of II-values into colors also letting Nx indicate that N is involved in an exotic BHB with the 100th percentile II-value of 9.85.

Influenza (PDB files 2HMG, 1HTM): See Figure 1A-D and 2A-B. As per [ibid] for chain F prefusion: residues 4x,5x form the fusion peptide with 9y,10x,11x and 14x,15r the nearby loop; 62o,63x account for helix extension; 96y,101y account for C-terminus inversion; 172x,175x form the C-terminus linker; 21x,24r,36y account for movement of the fusion peptide; 126o,130o,134x,136x,137x are of function unmentioned in [ibid] prefusion and reorganize postfusion to form the C-helix. For chain E prefusion, the residues 135o,136x,137o and 221y,227y pinpoint the sialic acid binding sites with the others of unknown function. Illustrated in Figure 2A-B is the full hemagglutinin glycoprotein including the sialic acid receptor binding domain, which is also comprised of exotic BHBs.

Paramyxovirus (PDB files 2B9B, 1ZTM): See Figures 1E-H and 2C-D. There is only approximate consensus among the three chains A,B,C. For the prefusion chain A,B,C consensus exotic residues as per [ibid] and using the color scheme of Chain C depicted in Figure 3: 90x,91x,92,93,94r,95o,96x lie in the fusion peptide; 264x,269r lie at the C-terminus of the helix extension domain; 297o,299o lie adjacent to the C-terminal inversion domain; 484y lies in the C-helix; 330r lies in a loop in Domain II; 414o,416y lie in a loop in Domain I. Concentrating just on Chain C and considering only colors R,X: 43x lies at the beginning of a β strand in DIII prefusion and in the middle of a β sheet postfusion; 90x,91r,94r,96x,102x,109x,113x lie in the fusion peptide, 184x,188x,189x lie in the C-terminus extension domain; 263r,264r,268x,269x lie in a loop and β turn region prefusion and comprise a β sheet postfusion; 278x lies in a loop between DI and DII prefusion and comprise a β sheet postfusion; 328r,330r lie in a β turn region prefusion and in a short β sheet postfusion; 387x,388x,392x,393x,408x,424x lie in a short β sheet prefusion and comprise a loop postfusion; 469r,473r,480x,485x lie in the C-terminus inversion domain. All R,X bonds of Chain C prefusion exhibit postfusion DSSP secondary structure reformation consistent with expectations.

Tick-borne Encephalitis (PDB files 1SVB, 1URZ): See Figure 1I-K. Concentrating here primarily on O,R prefusion as per [ibid]: 307y,309o lie in the inversion loop; the fusion peptide is not exotic prefusion although residues 74,78,100,101,105,106 all have II-values above 7.0, significant but below the cutoff, however with colors 74y,78y,100x,106x postfusion; the ij loop is unremarkable prefusion, but postfusion

contains 248x,249x,251x,252x; 148o prefusion lies in a loop in DI that is not exotic postfusion; in contrast, 167o,169o prefusion also lie in a loop in DI which however becomes red postfusion; 184o lies in the middle of a β strand both pre- and postfusion; 278o,280o lie in a loop prefusion and in a β turn postfusion; 360r,366r,368r lie in a loop in DII prefusion that remains red postfusion; 372o,373r lie in the middle of a β strand in DI both pre- and postfusion but colored 372b,373y postfusion; 388o,394o lie at the beginning and end, respectively, of a β strand prefusion and likewise postfusion but with an orange residue now between them. It appears that the fusion peptide is not composed of exotic residues until after the pre- to postfusion transition and that the ij loop is unremarkable pre- and exotic postfusion. Moreover, all of 148o,372o,373r appear to lose free energy in the pre- to postfusion transition. In contrast 167o,169o and 360r,366r,368r become or remain red postfusion suggesting either possible false positives or some further activity involving them to follow the postfusion conformation. Meanwhile 278o,280o undergo transition from loop to β turn consistent with losing free energy for reconnection; in contrast 388o,394o retain their β strand conformation but decrease free energy and produce an orange residue between them postfusion.

Vesicular Stomatitis (PDB files 5I2S, 5I2M): See Figure 1L-O. As per [ibid] prefusion: extension domain 1 has no BHBs at all except for the nearby exotic 183y,185y, while extension domain 2 contains the exotic 29r,33r,38o; the fusion peptide is not exotic prefusion and contains 115y,118o,119y and 71o,72y,75y postfusion. Considering only O,R,X there are two general rules from pre- to postfusion transition: DSSP secondary structure conformation is preserved; and the free energy is non-increasing. The notable exceptions are: the loop 261x changes conformation to the end of a short β 261o; the loop 404x at the C-terminus becomes the short β 404o; the end of the β strand plus loop 370r,373r becomes the more exotic 370x,373x; the loop 10x becomes 12r,15r. Except for these few cases and the fusion peptide, the free energy of exotic peptides is again diminished or preserved from pre- to post-fusion, and all residues which are exotic postfusion are already exotic in the prefusion conformation. This finding is consistent with the fact that this glycoprotein is capable of oscillating between its pre- and postfusion conformations.

4. PRIMARY STRUCTURE MOTIFS

Turning finally to a companion plot for Figure 4 of the main text, SM Figure 3 illustrates the distribution of primary structure across the entire free energy spectrum. Owing to their different magnitudes, the collection of residues is partitioned into 3 comparable sets as indicated for ready comparisons. As in the exotic tail in Figure 4 of the main text, again here there appear to be characteristic peptides not as there of high but here of low and other free energies which might be analyzed using machine learning.

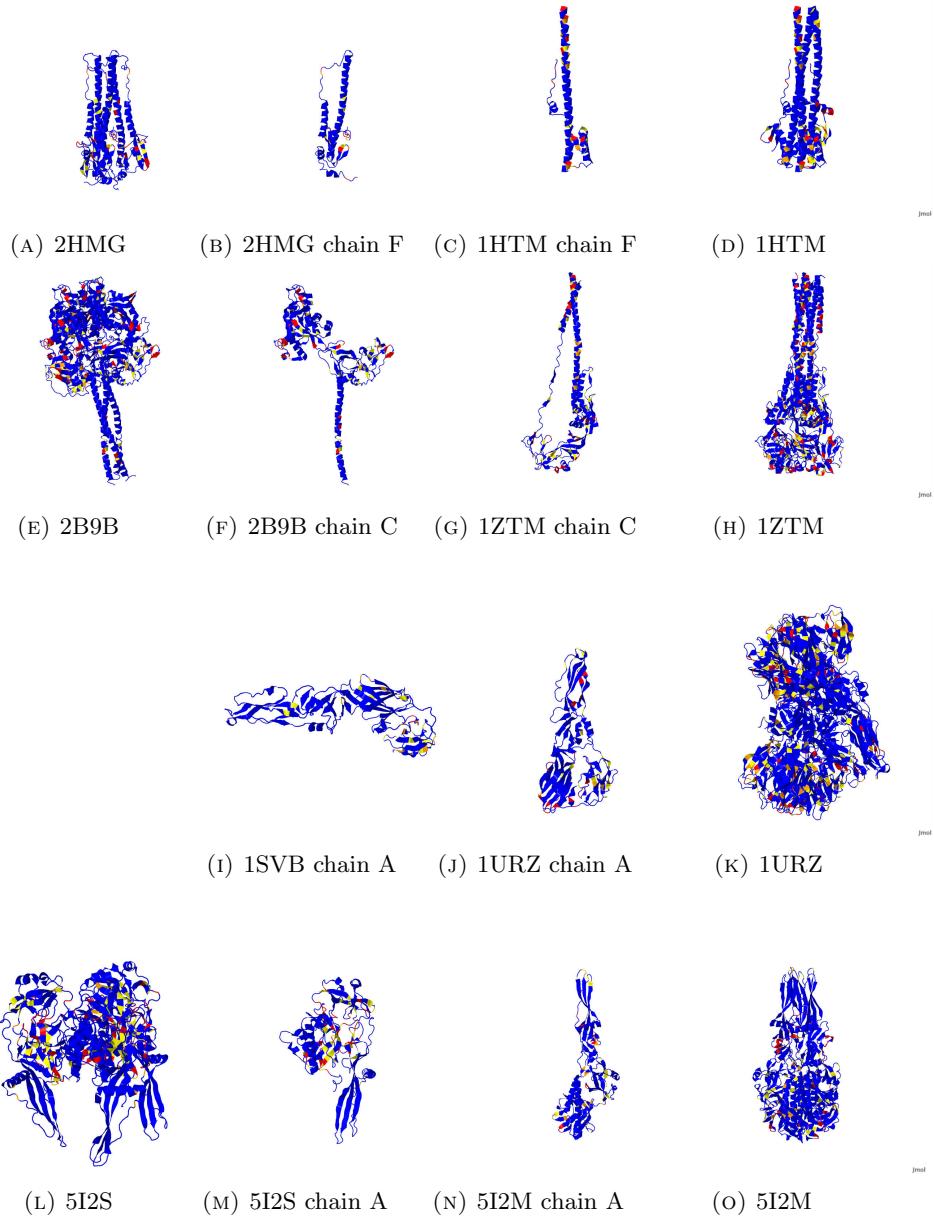
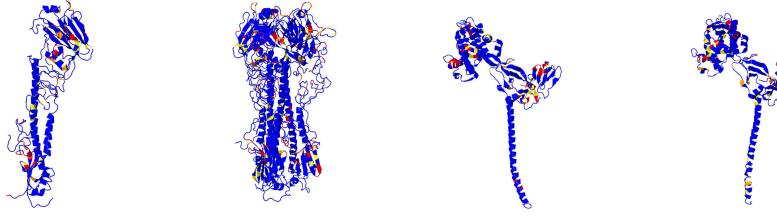


FIGURE 1. Compare with Figure 6 in [ibid], to which these images are aligned. Blue indicates non-exotic, and yellow, orange and red respectively correspond to Π -value at least 7.5, 8.5 and 9.5. Parts A,B are influenza hemagglutinin pre- and parts C,D postfusion. Parts E,F are paramyxovirus glycoprotein F pre- and parts G,H postfusion. Part I is tick-borne encephalitis glycoprotein E pre- and parts J,K postfusion. Parts L,M are vesicular stomatitis glycoprotein G pre- and parts N,O postfusion.



(A) 2HMG chain E,F (B) 2HMG (C) 2B9B chain A (D) 2B9B chain B

FIGURE 2. Blue indicates non-exotic, and yellow, orange and red respectively correspond to Π -value at least 7.5, 8.5 and 9.5. Part A illustrates the exotic regions for influenza hemagglutinin HA1 and HA2 chains E and F, and B illustrates the entire glycoprotein. Parts C and D respectively depict 2B9B chains A and B for comparison with 2B9B chain C, which is illustrated in part F of Figure 1.

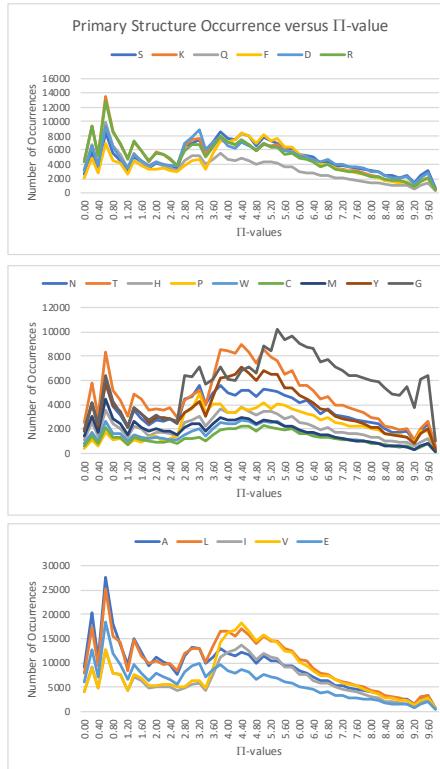


FIGURE 3. Histogram of Π -values and of flanking primary structure for all BHBs across HQ60. Compare to Figure 4 of main text.

TABLE 1. Exotic BHB Donor/Acceptor Residues of Viral Glycoproteins

Bourbon Virus Glycoprotein Env Postfusion (5ZKX)**Chain A**

90-94% 189/185 117/93 168/171 140/97 144/121 109/105 155/254 239/241 98/139 145/48 445/127
 95-98% 253/271 358/341 **68/57** 101/137
 99% 348/351 414/240 9/5
 100% 44/148 **103/135** 106/101 116/93 238/207 349/351

Chikugunkya Virus Glycoprotein E1-E2-E3 Prefusion (3N41)**Chain A**

90-94% 18/6
 95-98%
 99%
 100% 13/10

Chain B

90-94% 58/61 245/F58 44/154 158/260 334/330 237/170 96/102 150/268 109/131 196/192
 189/217 236/11 130/109 116/123
 95-98% 31/17 27/23 16/51 55/98 75/71
 99% 240/237 99/54 280/282 98/100
 100% 23/26 72/74 74/71 88/89 92/85 118/120 166/163 185/187 200/209 283/279 300/304 F258/301

Chain F

90-94% 160/156 157/159 252/184 109/74 370/372 125/175 373/369 381/361 248/243 203/199 **85/100** 102/61 77/73 **89/91** 60/103
 95-98% 323/348 263/267 382/306 273/257 321/350 138/140 129/39
 99% 315/356 88/227
 100% **101/98** 151/163 161/281 258/B301 268/262

Dengue Virus Glycoprotein E-M Prefusion (3J2P)**Chain A**

90-94% 405/400 390/378 382/336 242/249 263/258 187/284 **74/112** 86/82 430/427 366/363 468/463 349/339 119/66 61/219 181/291 22/287
 95-98% 412/407 443/438 335/300 411/406 250/241 41/38 189/186 252/239 410/405 180/175 286/23 433/428 190/186 330/327 233/220 447/441
 163/139 78/75 341/377 50/136
 99% 488/482 120/89 176/179 484/479 280/277 463/458 481/476 394/374 492/487 136/132 216/212 346/343 433/429 490/485
 100% 9/6 21/16 29/26 62/123 79/76 87/83 91/87 **104/100** **106/101** 126/59 156/153 179/175 182/173 183/288 199/128 202/204 215/211
 262/257 265/261 277/272 284/25 287/184 290/19 308/325 321/369 326/307 328/305 345/342 395/373 396/373 398/395 400/396 401/396 404/399
 408/403 413/408 415/412 416/411 420/416 421/417 434/429 434/428 435/430 436/431 439/434 447/443 448/444 451/9 461/456 469/464 475/472
 477/473 479/475 483/479 489/484 493/489 493/488 494/489

Chain B

90-94% 49/45 48/43 54/49 52/47
 95-98%
 99% 51/46 31/26 46/41
 100% 35/30 53/48 59/56 60/56 68/63 71/66

Dengue Virus Glycoprotein E Postfusion (1OK8)**Chain A**

90-94% 284/25 161/141 298/6 392/376 382/336 334/358 245/247 363/326 303/334 119/66 202/204 354/368 **106/100**
 95-98% 324/310 179/175 342/345
 99% 166/168 316/319 274/276 37/293
 100% 15/18 176/179 180/175 239/234 268/264 375/393 386/383 387/383

Dhori Virus Glycoprotein G_P Postfusion (5XEB)**Chain A**

90-94% 188/191 160/117 **121/157** 94/71 175/281 118/159
 95-98% 32/29 64/168
 99% 164/141
 100% 137/113

Donor/Acceptor residues of BHBs in order of non-decreasing Π -values with 7.5, 8.5, 9.5 and 9.85 respectively corresponding to percentiles 90, 95, 99 and 100. Marked in boldface are the residues lying in generally agreed upon fusion loops.

TABLE 1. (continued) Exotic BHB Donor/Acceptor Residues of Viral Glycoproteins

Eastern Equine Encephalitis Virus Glycoprotein E1-E2 and Capsid C-terminal Prefusion (6MX4)	
Chain A	
90-94%	280/5 182/184 55/52 120/48 310/382 330/345 171/168 78/75 123/178 218/203 319/303 60/103 256/251 6/1 118/113 255/251 306/316
95-98%	65/99 328/347 253/185 162/282 282/3 424/419 302/320 315/312 271/261 B337/389 322/351
99%	259/B298 84/98 257/252 201/197
100%	4/281 4/2 14/11 18/15 77/73 88/228 89/91 105/79 141/137 142/137 146/133 153/150 184/181 185/181 230/226 244/239 246/242 316/357 356/317 371/373 389/386 412/408 416/411 422/417
Chain B	
90-94%	373/368 28/25 243/A58 249/246 230/227 411/408 96/98 33/29 77/65 381/376 201/222 319/278 42/151 180/186 279/276 285/312 225/198 94/100 404/400 358/353 182/184
95-98%	234/11 25/22 73/69 104/43 282/274 315/282 14/233 180/177 337/A389
99%	195/228 90/83 280/276 160/157 241/237
100%	16/50 18/29 23/25 24/21 57/60 65/54 65/61 70/72 75/67 82/78 86/87 151/148 198/206 235/167 255/162 277/279 293/306 297/301 370/365 403/400 251/396
Chain C	
90-94%	246/249 229/240 185/182 107/102 242/253 118/142 138/160 124/116 122/118 189/191 225/216
95-98%	131/174 234/236 187/193 119/121
99%	
100%	133/130 143/117 157/161 161/156 162/156 215/211 230/240 235/236 240/200 251/B396
Ebola Virus Glycoprotein GP Prefusion (5JQ3)	
Chain A	
90-94%	186/36 100/164 115/143 71/178 160/179 266/260 43/39 477/276 173/123
95-98%	40/42 226/144 144/112 133/175 238/240 475/274 145/222
99%	237/240 192/189 101/65
100%	111/139 174/111 B511/73
Chain B	
90-94%	531/527
95-98%	599/594 622/617 555/551
99%	628/623
100%	511/A73
*Epstein-Barr Virus Glycoprotein gp350 Prefusion (2H6O)	
Chain A	
90-94%	197/218 193/190 83/62 405/351 227/245 420/403 54/124 148/16 249/261 142/10 55/57 47/66 101/313 128/51 362/357 135/138 235/19 121/117 272/177 80/64 422/401 15/143 279/303
95-98%	385/388 397/394 208/204 17/145 334/389 432/428 210/202 33/30 260/257 86/112
99%	176/178 46/133 389/385 223/220
100%	44/41 58/54 73/69 74/69 98/101 194/191 195/192 209/23 234/210 243/267 277/273 314/311 339/336 355/38
Hanta Virus Glycoprotein G_c Postfusion with scFv A5 (5LJY)	
Chain A	
90-94%	216/224 336/411 222/239 162/164 313/310 361/398 45/19 22/44 318/315 388/368 364/366 319/314 290/287 253/250
95-98%	293/285 63/33 379/349 95/91 301/298 179/181 227/223 396/408
99%	46/51 50/47
100%	74/229 288/290 335/340 337/388 412/333
Hanta Virus Glycoprotein G_c Postfusion (5LJZ)	
Chain A	
90-94%	336/411 162/164 361/398 288/290 400/402 379/349 378/353 174/185 388/368 79/137 274/270 222/239 127/113 316/17
95-98%	240/219 216/224 63/33 175/171 364/366 315/179
99%	229/236 412/333 227/223
100%	74/229 179/181 205/202 253/118 335/340 396/408

Donor/Acceptor residues of BHBs in order of non-decreasing Π -values with 7.5, 8.5, 9.5 and 9.85 respectively corresponding to percentiles 90, 95, 99 and 100. Marked in boldface are the residues lying in generally agreed upon fusion loops. * indicates a glycoprotein other than the fusion peptide which does not and is not expected to contain the fusion loop.

TABLE 1. (continued) Exotic BHB Donor/Acceptor Residues of Viral Glycoproteins

*Hendra Virus Glycoprotein G Receptor Bound (2VSK)	
Chain A	
90-94%	453/445 490/529 232/251 226/228 409/370 481/466 190/598 535/556 463/497 572/568 410/399 291/287 465/485 293/285 375/371
359/361 480/477 269/252 371/408 528/534 496/529	
95-98%	385/382 589/215 558/581 352/365 432/411 413/366 521/515 316/295 542/544 194/546 484/542 257/264
99%	471/476 515/520 265/256 288/290 520/516 391/387
100%	245/237 255/266 345/342 347/369 362/358 368/349 376/372 440/409 448/450 477/471 507/122 516/520 562/508 569/571
Chain B	
90-94%	103/81 102/81 46/42 69/107
95-98%	51/46
99%	130/118 80/104 44/41133/167
100%	47/50 79/143 99/95
Hendra Virus Glycoprotein F Prefusion (5EJB)	
Chain A	
90-94%	320/322 321/290 293/318 224/221 439/435 396/419 307/303 197/192 361/358 262/256 42/295 184/178 415/430 285/50 235/231
283/53 46/336 271/275 70/66 123/120 323/319 384/412	
95-98%	304/306 276/270 335/337 326/346 450/365 431/381 301/374 30/356 303/408
99%	82/77
100%	166/163 238/232 264/281 360/354 373/370 383/379 395/399 417/413 418/413 425/422 477/472
HIV Glycoprotein gp41 Prefusion (6MTJ)	
Chain B	
90-94%	583/578 627/622
95-98%	532/623 543/538 624/619
99%	
100%	531/626
Influenza Virus D Glycoprotein HEF Prefusion (5E64)	
Chain A	
90-94%	155/103 382/399 44/367 278/127 356/359 34/414 393/406 130/277 407/390 82/53 B75/402 305/155 288/210 41/384 378/385 17/27
221/307 248/260 295/246 106/152 54/79 159/302 22/B101 403/B74	
95-98%	182/93 425/24 215/217 193/199 200/192 38/391 B25/10
99%	B18/11 69/65 72/58
100%	12/B15 70/65 122/327 127/174 173/178 178/172 179/175 205/293 261/247 270/231 312/148 342/B77 354/357 355/359 358/C28 374/370
382/397 404/377 358/C28	
Chain B	
90-94%	145/142 75/A402 A22/102 A403/74
95-98%	48/45 25/10
99%	18/A11
100%	16/A9 40/29 135/138 A12/15 A342/77
Chain C	
90-94%	82/53 183/93 250/258 288/210 9/D26 295/246 72/58 305/155 155/103 205/293 41/384 130/277 215/217 179/175 200/192 22/D102
106/152	
95-98%	248/260 159/302 69/65 34/414 178/172 193/199 261/247 38/391 221/307 54/79 127/174
99%	278/127 70/65
100%	12/D15 109/50 122/327 173/178 182/93 274/270 312/148 12/D15
Chain D	
90-94%	C9/26
95-98%	
99%	
100%	C12/15
*Lassa Fever Virus Glycoprotein GP1 Receptor Bound (4ZJF)	
Chain B	
90-94%	202/214 87/231 103/225 169/218 197/191
95-98%	216/170 156/117
99%	162/158 211/207 99/101
100%	199/79 97/228

Donor/Acceptor residues of BHBs in order of non-decreasing Π -values with 7.5, 8.5, 9.5 and 9.85 respectively corresponding to percentiles 90, 95, 99 and 100. Marked in boldface are the residues lying in generally agreed upon fusion loops. * indicates a glycoprotein other than the fusion peptide which does not and is not expected to contain the fusion loop.

TABLE 1. (continued) Exotic BHB Donor/Acceptor Residues of Viral Glycoproteins

Lymphatic Choriomeningitis Virus Glycoprotein GPC Prefusion (5INE)**Chain A**

90-94% 311/306 **298/286** 104/106 74/379 248/244 233/166 **293/78** 368/362 166/162 69/72 391/376 382/71 161/122
 95-98% 381/385
 99% 201/196 401/393 168/161
 100% 73/68 96/98 105/85 162/167 208/364 210/365 249/244 **300/283** 348/343 396/398

Measles Virus Glycoprotein F Prefusion (5YXW)**Chain A**

90-94% 74/B343 69/65
 95-98% B280/56 53/B283 27/B359
 99%
 100% 102/B115 B286/49
Chain B
 90-94% A47/343 230/267 307/309 444/409 241/235 232/265 274/278 326/322 188/182 398/402 **126/123**
 95-98% 418/433 383/385 338/335 280/A56 187/182 363/357 235/230 53/283 27/359 338/340 310/306
 99% 323/325 279/273 400/422
 100% 286/B49 295/291 329/349 373/370 434/384 **A102/115**

Measles Virus Glycoprotein H Receptor Bound (3ALX)*Chain A**

90-94% 595/592 601/579 98/106 359/291 388/383 40/138 262/269 486/496 574/576 387/382 444/355 552/542 37/334 455/468 213/208 392/388
 411/430 315/312 273/203
 95-98% 106/45 593/590 111/41 77/66 232/220 427/414 338/319 548/544 115/111 44/107
 99% 274/257 307/349 365/357 383/378 472/477 225/227 125/127 86/83 592/589 460/463 75/72 441/438
 100% 54/51 104/99 117/135 204/270 214/210 215/209 229/258 287/301 293/295 313/310 357/289 463/459 464/459 478/471 482/467 493/489
 494/491 499/483 518/523 588/591

MERS Virus Spike Glycoprotein Prefusion Conformation 1 (5X5C)**Chain A**

90-94% 981/976 440/576 59/278 169/185 91/300 506/514 934/803 1152/778 1036/1031 493/391 302/208 329/332 256/266 148/171
 400/396 273/268 280/267 325/337 297/146 824/819 1032/1027 246/178 991/987 215/210 1143/786 343/696 305/93 1037/1031 198/190 454/450
 539/558 384/408 399/395 292/150
 95-98% 913/908 94/96 1125/1139 522/470 907/676 301/88 320/75 54/50 127/123 108/104 185/239 76/319 113/291 133/130 674/660 508/512
 939/936 1181/1178 669/664 1137/793 375/372 372/604 584/581 1035/1030 117/256 392/490 164/154
 99% 1109/1103 1131/1133 857/854 370/688 1192/1189 855/850 1042/1037 970/779 809/805
 100% 131/136 175/179 179/174 180/174 187/235 195/199 214/209 337/50 354/351 375/605 376/597 389/385 408/583 470/464 519/466 551/546
 552/546 599/596 609/611 651/618 670/664 741/761 759/721 765/762 773/859 798/795 799/796 808/810 870/866 986/983 987/984 1020/1016
 1107/1102 1118/1114 1146/1143 1157/1203 1158/1159 1167/1192 1173/1177

MERS Virus Spike Glycoprotein Prefusion Conformation 2 (5X5F)**Chain A**

90-94% 981/976 351/346 59/278 169/185 320/75 91/300 506/514 934/803 1152/778 1036/1031 493/391 302/208 329/332 256/266 478/426
 148/171 400/396 273/268 280/267 297/146 824/819 1032/1027 246/178 522/470 991/987 215/210 1143/786 305/93 1037/1031 198/190 113/291
 407/583 454/450 539/558 384/408 399/395 292/150
 95-98% 347/342 913/908 94/96 1125/1139 907/C676 80/48 301/88 54/50 127/123 108/104 185/239 133/130 674/660 508/512 939/936 1181/1178
 669/664 1137/793 375/372 604 1035/1030 117/256 392/490 164/154
 99% 1109/1103 857/854 370/688 1192/1189 855/850 1042/1037 585/438 970/C779 809/805
 100% 131/136 175/179 179/174 180/174 187/235 195/199 214/209
 337/50 353/348 375/605 376/597 389/385 470/464 519/466 551/546 552/546 584/439 599/596 609/611 651/618 670/664 741/761 759/721
 765/762 773/B859 798/795 799/796 808/810 870/866 986/983 987/984 1020/1016 1107/1102 1118/1114 1131/1133 1146/1143 1157/1203
 1158/1159 1167/1192 1173/1177

Donor/Acceptor residues of BHBs in order of non-decreasing II-values with 7.5, 8.5, 9.5 and 9.85 respectively corresponding to percentiles 90, 95, 99 and 100. * indicates a glycoprotein other than the fusion peptide which does not and is not expected to contain the fusion loop. For 5X5C and 5X5F, the expected fusion loop is part of the structure at residues 880-900 but contains no exotic residues.

TABLE 1. (continued) Exotic BHB Donor/Acceptor Residues of Viral Glycoproteins

MERS Virus Spike Glycoprotein Receptor Bound (4L72)*Chain A**

90-94% 466/462 575/541 492/489 592/ 587 621/540 668/663 687/683 572/567 649/624 303/314 580/525 509/41 104/117 585/552
 485/489 214/197 545/625 654/628 633/629 76/72
95-98% 523/519 109/113 541/618 739/735 137/141 318/322 405/382 463/467 412/414 477/453 308/310 629/651 430/455 153/130 456/474
 355/320 520/522
 99% 260/662 210/204 159/163 307/310 332/336 455/450 322/319 195/190 262/258
 100% 64/67 65/67 81/85 124/127 191/194 197/170 209/203 218/222 344/340 366/370 378/381 426/422 436/441 458/429 464/64 489/484 552/582
 558/427 560/473 574/567 583/ 590 597/670 624/620 680/673 681/675

Chain B

90-94% 498/561 508/512 440/576 401/445 503/557
 95-98% 522/470 468/463 470/464
 99%
 100% 408/583 416/411 417/411 519/466 552/546 584/581

Metapneumovirus Glycoprotein F Prefusion (5WB0)**Chain F**

90-94% 264/266 313/282 380/376 393/401 62/180 309/286 174/168 29/24 257/272 418/393 413/378 281/37 262/268 338/40 314/316
 214/210 284/35 407/384 39/333 30/24 434/19 175/169
95-98% 43/275 333/329 310/320 330/332 373/383 389/385 417/393
 99% 377/379
 100% 25/29 167/51 443/26

†Metapneumovirus Glycoprotein F Postfusion (5L1X)**Chain A**

90-94% 48/D433 B334/36 79/74 88/B260 B284/35
 95-98% B281/37 B279/39 30/24 66/63
 99% 25/29
 100% 36/B283 39/B333

Chain B

90-94% 334/A36 421/423 258/214 220/F208 413/378 174/168 394/417 454/451 284/A35 407/384 309/286 218/252 373/383
 95-98% 302/366 314/316 354/349 281/A37 303/291 155/150 175/169 279/A39 213/207 251/247
 99% 252/247 372/F315 158/153
 100% 264/266 333/329 377/379 417/393 A36/283 A39/333 D372/315

Mumps Virus Glycoprotein HN Prefusion (5B2C)*Chain A**

90-94% 323/378 385/316 259/242 179/198 483/541 409/307 315/317 518/529 197/180 211/235
 95-98% 561/573 304/325 375/326 231/214 308/321 549/553 138/217 301/297 395/295 572/562 137/189 B168/171
 99% 569/174 391/386
 100% 168/B171 221/224 250/252 283/286 310/408 406/375 411/310 417/413 455/452 462/509 468/509 542/480 548/143

Murine Corona Virus Spike Glycoprotein Prefusion (5I08)**Chain A**

90-94% 427/367 82/240 1132/1146 650/B55 272/275 1146/799 720/692 807/1139 215/199 434/431 716/718 805/802 402/398 660/646
 671/632 678/306 462/459 1114/1109 413/408 581/455 99/233 283/280 1131/1146 606/354 768/741 137/151 263/67 463/574 262/101 1115/1109
 814/942 78/256 583/585 200/214 72/40 221/48 613/318 341/437
95-98% 1000/995 234/127 1001/996 626/319 1052/B832 1062/1058 117/114 695/717 690/680 745/736 826/822 1007/1002 661/665 158/18 320/623
 88/26 205/100 823/820 742/B949 345/341 328/353 112/118
 99% 828/823 396/579 1138/1140 291/286 79/256 941/936
 100% 17/156 47/218 55/51 59/271 95/91 96/91 109/196 143/145 146/142 230/207 280/43 375/B1064 414/408 415/409 416/409 449/444 586/582
 600/626 605/381 621/631 651/654 728/764 739/742 767/729 817/813 827/823 927/923 937/932 973/968 1061/1057 1116/1110

Donor/Acceptor residues of BHBs in order of non-decreasing Π -values with 7.5, 8.5, 9.5 and 9.85 respectively corresponding to percentiles 90, 95, 99 and 100. Marked in boldface are the residues lying in generally agreed upon fusion loops. * indicates a glycoprotein other than the fusion peptide which does not and is not expected to contain the fusion loop. Part of the fusion loop is missing from 5L1X, and the residues 100-125 presumed to lie in the fusion loop a fortiori do not appear in exotic BHBs; in 5WB0 which contains the fusion loop, there is a bond with sub-exotic Π -value 6.6, roughly the 82nd percentile, for a BHB with residues 112/109 in the fusion loop. † indicates that the fusion loop, though part of the peptide, is either missing from the structure or disordered and in either case a fortiori can provide no reliable exotic BHBs. For murine coronavirus, the fusion loop at residues 900-920 is disordered in 6B3O and contains no exotic BHBs in 5I08.

TABLE 1. (continued) Exotic BHB Donor/Acceptor Residues of Viral Glycoproteins

†Murine Corona Virus Spike Glycoprotein Postfusion (6B3O)	
Chain A	
<i>90-94%</i>	<i>1235/1231 750/1125 805/800 1153/1149 776/773 800/796 1176/B755 C1176/755 1143/1170 1175/1138</i>
<i>95-98%</i>	<i>1134/1137 1165/1155 1122/754 767/1108 1085/1079 1236/1231 1121/754</i>
<i>99%</i>	<i>1208/1203 755/1121 1110/1106</i>
<i>100%</i>	<i>782/776 1092/1096 1107/1109 1150/1152 1159/1143 1181/757 1200/780 C1181/757 C1200/780</i>
*Newcastle Virus Glycoprotein HN Prefusion Complexed With Thiosialoside (1USX)	
Chain A	
<i>90-94%</i>	<i>473/527 276/247 223/210 225/208 554/551 253/236 378/384 205/229 547/559 539/535 405/304 298/319 178/188 560/546 133/129</i>
	<i>285/272 436/432 292/287 535/539 162/159 249/241 136/132</i>
<i>95-98%</i>	<i>182/184 452/495 428/413 528/470 130/211 129/183 B217/230</i>
<i>99%</i>	<i>456/452 555/169 540/534 368/364 408/410 215/218 550/523 505/489 294/290 275/282</i>
<i>100%</i>	<i>456/452 555/169 540/534 368/364 408/410 215/218 550/523 505/489 294/290 275/282</i>
*Newcastle Virus Glycoprotein HN Prefusion (3T1E)	
Chain A	
<i>90-94%</i>	<i>301/314 408/137 136/132 311/307 319/296 472/526 276/279 210/221 451/494 457/494 377/383 274/281 561/543 236/251 142/475</i>
	<i>479/482 517/500 491/501 493/499 133/129</i>
<i>95-98%</i>	<i>549/522 284/271 474/487 190/174 305/239 185/208 524/497 402/300 504/488 399/368 404/303 535/131 158/561</i>
<i>99%</i>	<i>534/538 533/135 201/197 482/478</i>
<i>100%</i>	<i>180/185 182/183 186/179 197/192 243/245 303/401 364/360 396/391 455/451 507/509 539/533 550/553 554/549 555/B551 B555/551</i>
Chain B	
<i>90-94%</i>	<i>402/300 545/527 527/469 377/383 136/132 444/423 305/239 185/208 372/315 319/296 426/441 214/217 533/135 485/476 301/314</i>
	<i>474/487 201/197 243/245 197/193</i>
<i>95-98%</i>	<i>479/482 534/538 195/170 158/561 467/417 408/137 357/352 554/549 404/303</i>
<i>99%</i>	<i>432/434 523/548 435/431 451/494 504/488 134/130 327/323</i>
<i>100%</i>	<i>180/185 182/183 186/179 197/192 210/221 224/207 252/236 292/287 295/291 358/353 395/391 399/368 424/443 455/451 472/526</i>
	<i>507/509 536/131 539/533 550/553 555/A551 A555/551</i>
Chain E	
<i>90-94%</i>	<i>86/81 102/97</i>
<i>95-98%</i>	
<i>99%</i>	
<i>100%</i>	<i>89/85 98/93</i>
Chain F, D	
<i>90-94%</i>	
<i>95-98%</i>	<i>93/87</i>
<i>99%</i>	
<i>100%</i>	<i>86/82 100/95 104/99</i>
†Newcastle Virus Glycoprotein F Postfusion (3MAW)	
Chain A	
<i>90-94%</i>	<i>54/347 316/308 234/269 437/425 432/429 402/406 245/239 247/244 368/365 394/382 94/89 284/63</i>
<i>95-98%</i>	<i>83/78 333/353 288/59 327/329 329/326</i>
<i>99%</i>	<i>401/428 165/160 50/343 47/302 474/469 342/344 41/37</i>
<i>100%</i>	<i>187/182 189/184 278/282 367/361 387/389</i>

Donor/Acceptor residues of BHBs in order of non-decreasing Π -values with 7.5, 8.5, 9.5 and 9.85 respectively corresponding to percentiles 90, 95, 99 and 100. Marked in boldface are the residues lying in generally agreed upon fusion loops. * indicates a glycoprotein other than the fusion peptide which does not and is not expected to contain the fusion loop. † indicates that the fusion loop, though part of the peptide, is missing from the structure or disordered and in either case a fortiori can provide no reliable BHBs. For murine coronavirus, the fusion loop at residues 900-920 is disordered in 6B3O and contains no exotic BHBs in 5I08.

TABLE 1. (continued) Exotic BHB Donor/Acceptor Residues of Viral Glycoproteins

Nipah Virus Glycoprotein G Prefusion (2VWD)*Chain A**

90-94% 297/280 359/361 226/228 307/302 235/218 212/209 311/307 222/232 357/363 213/209 385/382 484/542 521/515 265/256 433/475
 288/290 569/571 506/458
 95-98% 194/546 542/544 481/466 190/598 413/366 218/588 562/508 463/497
 99% 410/399 269/252 471/476 515/520 328/324 270/252
 100% 259/262 291/287 331/327 440/409 445/354 448/450 477/471 511/561 516/520

Nipah Virus Glycoprotein G Receptor Bound (2VSM)*Chain A**

90-94% 297/280 307/302 359/361 545/541 385/382 235/218 445/354 357/363 481/466 451/447 269/252 288/290 490/529 245/237 291/287
 589/215 226/228 328/324 562/508 507/122
 95-98% 484/542 542/544 347/369 463/497 432/411 569/571 413/366 440/409 194/546 511/561 215/586
 99% 259/262 410/399 471/476 332/327 515/520 190/598 448/450 483/480 516/520 472/476 477/471
 100%

Chain B

90-94% 51/46 133/167 59/117 103/81 69/107 162/155 102/81 79/143 80/104 170/166
 95-98% 130/118 156/141 47/50 157/139
 99%
 100%

Parainfluenza Type 5 Virus Glycoprotein HN Sialyllactose Soaked (1Z4X)*Chain A**

90-94% 298/300 414/440 171/173 553/160 520/516 445/492 185/159 436/415 443/411 214/197 301/297 525/463
 95-98% 362/305 466/524 552/157 309/286 461/407 358/309 351/356 505/507 269/265 533/122 424/426
 99% 480/472 204/207 356/350 392/290 531/126 266/269 537/531
 100% 120/172 233/235 293/391 357/353 389/358 394/293 398/128 439/505 449/445 501/512 548/552 559/541

Pseudorabies Virus Glycoprotein B Postfusion (6ESC)**Chain A**

90-94% 239/280 681/665 198/185 223/243 248/218 249/261 258/254 379/423 214/209 219/214 461/410 216/210 451/469 567/562
 343/347 539/534 408/402 685/677 219/213 628/657 675/671 680/682 595/597 644/631
 95-98% 190/**192** 303/306 435/430 312/324 460/463 647/650 417/456
 99% 688/675 215/209 474/446 298/294 241/237 619/624 161/385 403/407 **275/188**
 100% 122/124 217/211 339/353 445/439 447/443 464/459 672/674

Puumala Virus Glycoprotein G_n (5FXU)*Chain A**

90-94% 138/115 224/252 190/192 369/335 335/139 74/133 134/71 45/263
 95-98% 321/310 182/34 243/238 143/145
 99% 57/144 116/76 118/137
 100% 46/154 198/377 211/378 291/371 323/363 352/141

Puumala Virus Glycoprotein G_c Postfusion at pH 6.0 (5J81)**Chain A**

90-94% 1057/1059 737/795 993/1068 **772/785** 837/839 887/893 831/811 897/877 823/819
 95-98% 945/947 1036/1006 885/881 759/755 880/896 931/927 **785/771**
 99% 1021/1023 874/882 721/691 708/703
 100% 820/822 **910/776** 919/898 992/997 994/1045 1035/1010 1053/1065 1069/990

Puumala Virus Glycoprotein G_c Postfusion at pH 8.0 (5J9H)**Chain A**

90-94% 973/675 837/839 975/673 737/795 822/819 820/822 889/891 993/1068
 95-98% 931/927 887/893 1057/1059 1069/990 945/947 1036/1006 **788/744** 1053/1065 1060/1056
 99% 721/691 874/882
 100% 704/709 885/881 992/997 994/1045 1021/1023

Donor/Acceptor residues of BHBs in order of non-decreasing II-values with 7.5, 8.5, 9.5 and 9.85 respectively corresponding to percentiles 90, 95, 99 and 100. Marked in boldface are the residues lying in generally agreed upon fusion loops. * indicates a glycoprotein other than the fusion peptide which does not and is not expected to contain the fusion loop.

TABLE 1. (continued) Exotic BHB Donor/Acceptor Residues of Viral Glycoproteins

Respiratory Sphinctial Virus Glycoprotein F Prefusion at pH 5.5 (4MMS)	
Chain A	
90-94% B320/37 B475/35 B311/46 86/81 34/38 61/B193	
95-98% 51/B305 103/B241 B197/60	
99% 62/B295	
100% 48/B365 68/65	
Chain B	
90-94% 265/260 267/264 324/332 320/A37 439/416 287/302 248/282 341/316 484/479 475/A35 203/198 311/A46 479/475 A61/193 398/487	
373/369 405/415	
95-98% 365/361 D408/144 A51/305 205/199 342/352 409/411 F145/405 243/237 A103/241 408/F144 197/A60	
99% 346/348 A62/295	
100% 145/D405 294/296 421/417 445/410 A48/365	
Respiratory Sphinctial Virus Glycoprotein F Prefusion at pH 9.5 (4MMR)	
Chain A	
90-94% 39/33 31/B465 93/89 B370/49 51/B305 38/33 B193/58	
95-98% B314/44	
99% B466/28 81/76 34/38 84/81 98/95	
100% 48/B365 52/B305 62/B295 78/74 88/83 91/86 B306/51 B311/46	
Chain B	
90-94% 338/394 A31/465 497/492 150/146 244/240 512/507 453/455 267/264 272/267 433/424 398/487 370/A49 182/185 341/316	
A51/305 364/361 193/A58 220/217 445/410 301/288 405/415 203/199 202/198	
95-98% 399/333 291/298 449/425 314/A44 402/399 475/A35 345/312 362/364 188/179	
99% 466/A28 308/305 346/348	
100% 173/169 297/293 306/A51 311/A46 373/369 409/411 431/426 506/501 508/503 A48/365 A52/305 A62/295	
†Respiratory Sphinctial Virus Glycoprotein F Postfusion (3RRR)	
Chain A	
90-94% 62/B295 90/86 34/38 51/B305 D465/54	
95-98% B308/48	
99%	
100% 39/33	
Chain B	
90-94% 362/364 303/286 A62/295 170/165 186/181 441/414 459/E50 436/422 465/E54 250/F238 A51/305 450/425 299/291	
95-98% 348/345 439/416 449/425 308/A48 205/199 433/424 499/494 341/316 457/F307	
99% 342/352 248/282 213/208 427/430 409/411 294/296 241/236 227/222 172/167 223/218 405/415 202/197 373/369 380/376 358/354	
100% 219/214 226/221 460/448	
Rift Valley Fever Virus Glycoprotein G_c Postfusion (6EGT)	
Chain A	
90-94% 893/1010 1116/1038 1113/1093 832/817 1026/704 747/864 1059/1075 1008/724 813/793 1017/885 802/798 909/912 1115/1091	
1026/C704 972/1122 1041/1115 854/757 857/755 1076/1058 866/869 932/928 986/988 785/782 764/934 810/943 784/780	
95-98% 693/729 1019/1015 960/962 997/993 726/723	
99% 946/807 826/820 884/886 855/757 737/741	
100% 718/1 018 780/783 788/784 821/827 899/C897 904/B900 962/959 963/959 994/998 1040/1045 1079/1054 1125/969 C904/900	
Rubella Virus Glycoprotein E1 Postfusion (4ADI)	
Chain A	
90-94% 67/62 49/142 89/139 238/39 107/103 169/25 327/3 384/362 79/147 123/86 140/88 149/42 32/161 157/207 286/131	
95-98% 35/351 91/97 289/279 312/307 349/411 357/389	
99% 351/413 324/195 273/294	
100% 16/21 22/15 110/101 172/174 180/182 188/190 248/297 282/129 C248/297	

Donor/Acceptor residues of BHBs in order of non-decreasing Π -values with 7.5, 8.5, 9.5 and 9.85 respectively corresponding to percentiles 90, 95, 99 and 100. Marked in boldface are the residues lying in generally agreed upon fusion loops. † indicates that the fusion loop, though part of the peptide, is missing from the structure or disordered and in either case a fortiori can provide no reliable BHBs.

TABLE 1. (continued) Exotic BHB Donor/Acceptor Residues of Viral Glycoproteins

SARS Virus Spike Glycoprotein Prefusion Conformation 1 (5X58)**Chain A**

90-94% 286/281 937/932 562/552 1015/1010 184/91 554/47 1039/1036 259/59 89/255 1033/1047 94/181 420/499 981/976 985/980
 1047/705 68/254 658/677 239/246 261/57 702/1050 256/88 423/363 597/584 599/582 725/720 363/422 452/411 133/130 370/367 615/609
 443/477 680/655 496/423 335/332 1079/1085 656/846 879/865 162/127 476/460 740/736
 95-98% 235/100 932/927 350/321 636/599 461/475 43/216 565/569 116/126 240/140 1042/1038 726/721 299/585 621/280 230/104 431/425
 60/258 82/231 561/552 869/866
 99% 115/103 510/379 120/122 655/845 676/659 64/35 129/113 480/440 330/325 579/576 924/919 73/75 293/287 146/141
 100% 51/267 223/194 226/192 379/530 392/490 398/393 454/225 507/504 526/536 530/532 635/631 674/661 754/749 765/760 **780/782**
783/779 868/862 931/927 1016/1010 1017/1011 1039/1041 1056/696 1083/1080

SARS Virus Spike Glycoprotein Prefusion Conformation 2 (5X5B)**Chain A**

90-94% 120/122 859/854 657/653 696/B877 B48/551 872/868 643/676 **783/779** 420/499 561/552 562/552 82/231 261/57 528/525
 289/284 423/363 1083/1080 363/422 268/271 452/411 1033/1047 952/B737 939/935 370/367 949/945 553/550 **778/775** 1060/892 443/477
 804/799 496/423 100/234 335/332 526/536 615/612 932/927 67/33 569/564 867/862 324/349 476/460
 95-98% B50/553 937/932 621/280 461/475 223/194 315/376 **780/782** 733/728 293/287 674/661 1072/1103 431/425 632/634
 99% 736/731 879/865 975/970 480/440 609/604 553/561 924/919 162/127 330/325 599/582 735/731 146/141
 100% 39/35 43/216 48/C551 51/267 54/264 115/103 129/113 185/199 200/184 239/246 240/140 258/86 264/276 314/376 392/490 398/393
 530/532 555/556 579/576 582/303 597/584 635/631 741/738 754/749 765/760 805/800 868/862 869/866 873/868 892/887 931/927 972/966
 1017/1011 1042/1038 1056/696

†SARS Virus Spike Glycoprotein Receptor Bound (2AJF)**Chain A**

90-94% 207/203 107/103 42/37 97/92 461/456 55/341 536/530 393/389 263/486 216/212 156/151 352/355 348/345 366/363 564/560 344/51
 95-98% 488/481 44/39 331/326 125/120 218/209 153/148 284/279 361/332 507/503 280/276 174/168
 99% 341/338 563/392 293/290 560/555 40/35 210/217 603/599
 100% 25/20 71/66 135/140 149/146 149/143 150/146 150/144 170/165 197/191 276/444 333/358 363/333 425/418 447/442 449/443 467/462
 479/474 487/481 511/505 559/554

Chain E

90-94% 430/425 389/335 356/352 421/365 394/390
 95-98% 452/411 411/405 357/352 447/444 431/426 449/407
 99% 407/403 472/468
 100% 393/490 408/403

Semliki Forest Virus Glycoprotein E1 Postfusion (1RER)**Chain A**

90-94% 321/350 53/109 31/136 160/156 157/159 181/183 21/286 10/274 261/269 327/346 366/376 128/39 370/372 **104/59** 47/121 20/27
 95-98% 271/36 263/267 77/73 **100/97** 274/271 167/147 217/202 **85/100 92/89**
 99% 28/19 315/356 308/313 33/134 24/21 314/310 255/251 316/305
 100% 45/309 **102/62** 138/140 141/137 190/187 203/199 254/250 256/251 259/340 260/256 268/262 392/389

Sindbis Virus Glycoprotein E1-E2 Prefusion (3MUU)**Chain A**

90-94% 332/294 605/600 368/636 525/637 53/63 299/327 **444/460** 583/592 286/317 597/577 155/262 329/336 64/60 510/523 424/421 13/48
 612/544 563/559 621/629 314/289
 95-98% 131/108 61/52 327/338 287/275 402/485 484/403 608/603 105/92 45/35 623/627 374/370 276/286
 99% 588/585 95/102 489/399 385/381 **452/449** 437/433 40/153
 100% 20/21 54/63 98/100 100/97 101/97 263/154 280/282 290/313 300/304 322/281 324/321 337/334 380/387 432/429 472/410 517/514

Thogoto Virus Glycoprotein G_P Prefusion (5XEA)**Chain A**

90-94% 46/221 258/260 177/173 32/412 **150/73 83/72** 116/153 184/187 120/116 132/108 202/204 156/112 160/137 180/191
 95-98% **71/84** 176/173
 99% 174/176 **74/150** 87/68 243/222
 100% 171/258 261/257 292/186

Donor/Acceptor residues of BHBs in order of non-decreasing Π-values with 7.5, 8.5, 9.5 and 9.85 respectively corresponding to percentiles 90, 95, 99 and 100. Marked in boldface are the residues lying in generally agreed upon fusion loops. † indicates that the fusion loop, though part of the peptide, is missing from the structure or disordered and in either case a fortiori can provide no reliable BHBs.

TABLE 1. (continued) Exotic BHB Donor/Acceptor Residues of Viral Glycoproteins

West Nile Virus Glycoprotein E Prefusion (2HG0)**Chain A**

90-94% 127/124 129/203 297/183 338/303 392/388 280/50 358/374 353/342 306/337 383/397 139/47 262/258 266/262 181/184 186/294 52/134
 146/143 163/159 205/211
95-98% 177/188 319/322 276/278 144/163 154/150 185/181 **106/100**
 99% 114/95 369/329 267/263 279/275 30/26
 100% 48/138 57/127 72/114 74/112 **86/82** 171/173 202/198 209/206 210/206 220/215 239/235 244/241 264/259 270/266 354/342 373/147
 381/399 393/388 400/380

Yellow Fever Virus Glycoprotein E Prefusion (6IW4)**Chain A**

90-94% **98/111** 63/123 251/61 248/231 349/368 202/198 373/391 292/175 119/66 **112/97** 180/286
 95-98% 352/366 347/336 265/202 302/299 148/145 364/322
 99% **106/103** 192/187 361/324 7/3
 100% 173/176 178/289 314/317

Yellow Fever Virus Glycoprotein E Postfusion (6IW1)**Chain A**

90-94% 144/155 B19/11 119/66 123/62 361/324 7/32 120/89 351/366 183/166 127/197 373/391 177/172 352/366
 95-98% 19/C11 331/356 202/198 301/331 199/201 239/241 63/123
 99% 173/176 296/6 205/262
 100% 14/B16 15/18 16/C13 19/C8 163/165 165/133 221/56 314/317 347/336 B16/13 B19/8 C14/16

Zika Virus Glycoprotein E Prefusion (6CO8)**Chain A**

90-94% 340/337 179/186 205/211 500/496 216/200 323/320 290/25 31/8 17/20 295/187 389/344 221/217 51/282 **114/95** 466/461 124/62
 92/117 207/209
 95-98% 384/402 443/438 333/372 309/340 335/370 322/325 185/180 361/377 238/225 138/169
 99% 376/147 452/447 28/287 471/466
 100% 50/136 **112/97** 184/180 188/294 249/252 270/265 450/445 458/453 459/453 460/457 B13/271
Chain B
 90-94% 56/53 37/32
 95-98%
 99%
 100% 13/A271

Donor/Acceptor residues of BHBs in order of non-decreasing Π -values with 7.5, 8.5, 9.5 and 9.85 respectively corresponding to percentiles 90, 95, 99 and 100. Marked in boldface are the residues lying in generally agreed upon fusion loops. For yellow fever postfusion, the fusion loop is not missing from the structure and yet contains no exotic BHBs.

TABLE 2. Exotic BHB Free Energy Donor/Acceptor Residues of Viral Capsids

Human Adenovirus 2 (1X9P)**Chain A**

90-94% 464/439 227/285 183/178 263/424 237/212 474/516
 95-98% 114/533 249/273 276/246 131/553 546/549 538/109 235/232 446/441 72/559 217/214 565/120 560/71 234/231 452/448 174/134
 99% 240/211 386/389 198/192 280/277 206/202
 100% 53/49 61/58 123/562 152/199 182/178 194/189 195/190 200/197 201/197 221/225 230/226 390/385 395/400 460/456 471/468 485/481
 549/545

Foot and Mouth Virus (1FOD)**Chain 1**

90-94% 57/53 89/104 43/176 3:167/122 182/75 70/66
 95-98% 71/186 172/83 174/46 21/15 190/68 2:148/6 3:168/122
 99% 115/3:10 39/36 20/16 151/147
 100% 134/2:173 139/136 142/2:172

Chain 2

90-94% 170/101 217/3:138 105/206 77/183 189/118 154/149 20/22
 95-98% 23/19 121/185 148/1:6
 99% 11/8 194/189

100% 13/6 30/13 139/135 1:134/173 1:142/172

Chain 3

90-94% 57/85 2:217/138 54/203 80/182 100/213 167/1:122 50/45 97/92 81/78 126/143
 95-98% 150/119 58/61 207/104 168/1:122 119/192 121/190
 99% 1:115/10 197/66
 100% 100/212 206/48 12/8

Chain 4

90-94%
 95-98%
 99%
 100% 70/65 70/66

Hepatitis A Virus (5WTE)**Chain A**

90-94% 267/C96 184/179 126/270 75/72 245/148 250/93 77/26 96/247 71/43 151/242 230/109 137/258 107/173
 95-98% 143/250 136/259 136/258 52/48 B138/263 89/84 210/B166 C30/201
 99% 129/124 224/219
 100% C27/199

Chain B

90-94% 171/104 218/100 124/188
 95-98% 138/A263 C61/139 58/53 188/123
 99%
 100% 30/13 76/186 103/214 151/154 210/107

Chain C

90-94% A267/96 54/50 202/136 139/199 88/61 75/71 245/86 208/130 229/55
 95-98% 184/196 61/B139 91/95 30/A201
 99% 83/80 74/206 69/218
 100% 27/A199 47/43

Hepatitis E Virus (2ZTN)**Chain A Acceptors**

90-94% 144/298 172/320 156/278 286/203 313/181 341/332 424/418 432/331 532/534 437/327 324/440 500/474 600/471 581/570 392/219
 252/248 261/270 567/564 373/439 328/436 603/540 397/371 354/446
 95-98% 141/300 178/173 393/220 374/394 408/405 132/308 195/242 378/149 487/484 451/447 333/340 287/291 273/268 351/346 507/502 515/510
 99% 283/151 258/254 437/434 503/507 590/587 469/601 357/353
 100% 142/166 199/297 203/199 315/312 323/169 330/434 338/334 361/358 364/360 393/388 395/390 396/390 449/351 485/482 522/524 573/578

Donor/Acceptor residues of BHBs in order of non-decreasing Π -values with 7.5, 8.5, 9.5 and 9.85 respectively corresponding to percentiles 90, 95, 99 and 100.

TABLE 2. (continued) Exotic BHB Donor/Acceptor Residues of Viral Capsids

Polio Virus (5O5B)**Chain 1**

90-94% 204/126 194/3:22 92/101 163/159 269/125

95-98% 276/2:183 717/2:267 125/268 130/263

99% 134/259

100% 3:25/193

Chain 2

90-94% 37/33

107/209 235/121 175/133

95-98% 3:125/116 17/23 1:276/183 123/197 1:217/267 178/157 144/171 226/82 103/218 55/259

99% 261/101

100% 86/81 101/260 182/176 184/179 203/198

Chain 3

90-94% 196/129 63/58 97/92 215/51 1:194/22

95-98% 178/110 56/67 125/2:116 197/129 22/4:38

99%

100% 25/1:193 53/2:187 57/67 109/223

Chain 4

90-94% 95-98% 3:22/38 99% 100%

Rhinovirus (4AYM)**Chain 1**

90-94% 85/230 76/238 68/3:41 120/241 181/3:22 189/2:207 233/82 236/125 280/3:57 3:25/180 24/53 114/110

95-98% 254/2:176 115/246 3:85/282

99% 3:163/35 2:75/15

100% 75/3:15 241/71

Chain 2

90-94% 256/98 175/169 141/137 220/82 86/81 106/247 189/207 103/211 122/227 219/82 134/174 102/252 196/191 38/34 168/133 176/172

95-98% 144/164 60/56

99% 94/89 228/121 190/1:38

100% 29/12 177/172 178/172 3:53/180

Chain 3

90-94% 1:68/41 221/110 1:181/22 1:280/57 171/168 11/7 25/1:180

95-98% 108/222 216/47 65/61 85/1:282 152/133

99% 163/1:35 96/91 95/91 196/128

100% 53/2:180 57/67 1:75/15

Chain 4

90-94% 29/3 4/28

95-98%

99%

100%

Donor/Acceptor residues of BHBs in order of non-decreasing II-values with 7.5, 8.5, 9.5 and 9.85 respectively corresponding to percentiles 90, 95, 99 and 100.

TABLE 3. Conformationally Active and Exotic Residues in Test Cases

Viral Glycoprotein	#Residues	#Further Than One Away From Active	#Active	#One Away from Active	#Exotic
influenza glycoprotein HA2 chain F	122	70	33	19	7
paramyxovirus glycoprotein F chain A	422	81	251	90	62
tick-borne encephalitis glycoprotein E chain A	376	120	148	108	34
vesicular stomatitis glycoprotein G chain A	409	140	138	131	72

Here are described the data upon which the *p*-values in Table 2 of the main text are based. For each virus, the pre- and postfusion PDB files are aligned in order to compare the change of conformational angles during reconnection. For paramyxovirus, there are different strains of the virus pre- and postfusion, and the Smith-Waterman algorithm [T. Smith and M. S. Waterman, Identification of Common Molecular Subsequences, *Journal of Molecular Biology* **147**(1), 195-197 (1981)] is used to align the two primary structures. The # Residues *R* is the number of residues common to the aligned pre- and postfusion conformations, the # Exotic is the number of exotic prefusion residues, namely the number of predictions to be made.