Understanding the Structure and Function of Viral Glycosylation by Molecular Simulations: State-of-the-Art and Recent Case Studies

Elisa Fadda, Department of Chemistry and Hamilton Institute, Maynooth University, Maynooth, Kildare, Ireland

1 Introduction

The COVID-19 pandemic has renewed the scientific community's interest into viral infection worldwide, triggering extraordinarily rapid advancements in the discovery of new therapeutics and prevention and diagnostic strategies.[1] As a direct measure of the scientific engagement worldwide, since January 2020 to September 2020, 1707 articles focusing on SARS-CoV-2 have been posted on bioRxiv (www.biorxiv.org) and 6368 on medRxiv (www.medrxiv.org) online public preprint servers for biology and health sciences, respectively.[2] Of these preprints, 244 are focused specifically on various aspects concerning SARS-CoV-2 protein glycosylation, such as the glycosylation of the spike (S) fusion glycoprotein.[3–7] (Data from @glycopreprint (https://twitter.com/glycopreprint).) The surface of the SARS-CoV-2 S is covered with a dense N-glycan coat, known also as a "glycan shield,"[3,8] which is a hallmark of viral fusion glycoproteins.[3] This shield contributes to the intrinsic properties of the biomolecule, such as its folding, structural stability and trafficking,[9] and confers to the whole virus the ability to hide from the immune system. It can contribute to enhancing binding specificity and affinity to glycan epitopes, antibodies and other protein receptors[3] and can also, as recently discovered specifically for the SARS-CoV-2 S, actively contribute to the mechanistic function of the protein.[10] Structural biology approaches have been instrumental in these discoveries, revealing at the atomistic level details of the fusion proteins in different stages of viral entry.[4,5,7,11–15] However, from the structural, dynamics and energetic point of view of our understanding of how protein glycosylation can contribute to these events, experimental structural biology approaches are limited. The chemical nature of complex carbohydrates confers on them a high degree of conformational flexibility, or intrinsic disorder, which translates into highly dynamic behavior at room temperature that is retained in cryogenic environments. These aspects, together with the characteristic micro (or macro) heterogeneity at each glycosylation site,[16,17] complicate the determination of accurate glycan structures. Molecular simulations have now reached the technological maturity to provide a powerful alternative method, complementary to experimental structural biology, to advance our knowledge at the atomistic level of detail of viral glycosylation and in glycoscience at large. In this article I discuss the methodological foundations of complex carbohydrate modelling and illustrate a few contemporary examples in viral glycosylation research, namely studies of the SARS-CoV-2 S, the influenza A hemagglutinin (HA) and the HIV-1 envelope (Env) fusion trimer.

2 Molecular simulations

The physico-chemical properties of all molecular systems depend on their electronic and 3D structure. The latter is determined by the spatial arrangement of the atomic nuclei with connectivity, or topology, determined by the electronic structure within their molecular orbitals framework. According to the principles of quantum mechanics (QM), the electronic wave function of the system, or its electron density within a density functional theory (DFT) representation,[18–20] completely and uniquely defines all molecular properties.[21] Unfortunately, because of limitations imposed by the mathematical formulation of these theories, this knowledge cannot be extracted in an exact fashion and approximations are necessary to find solutions. The specific type of approximation defines different levels of theory that correspond to progressively more sophisticated ab-initio QM and DFT-based computational approaches. These in turn should provide increasingly higher accuracy in the calculation of molecular properties. Therefore, to completely understand the structure, spectroscopic properties and reactivity of any molecule we should ideally determine its electronic structure through a QM (or DFT) description. In practice this is impossible because of the
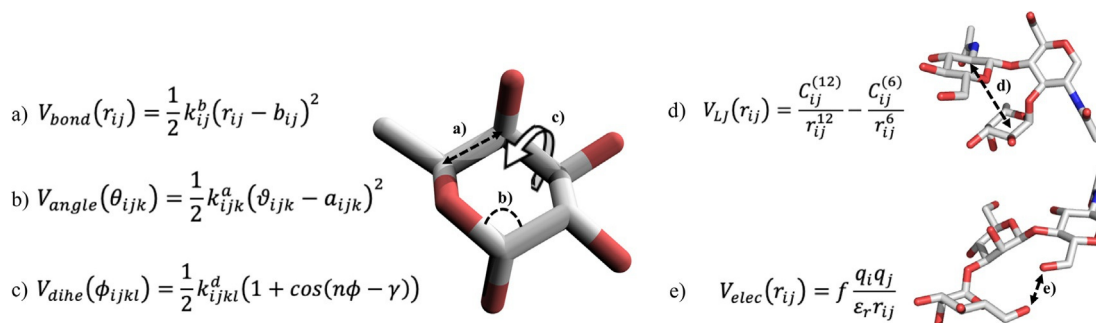
computational costs of these methods, which does not scale linearly with the number of atoms (N), but follows N^3, N^4, or up to N^7 dependencies for the highest levels of theory such as the "gold standard" QM approach Coupled-Cluster Singles and Doubles with perturbative Triples (CCSD(T)) method. The better scaling ab initio methods (e.g. most DFT implementations, which scale with N^3) can now handle systems with up to a few hundred atoms at most.[20] This excludes the large majority of the biomolecules of interest in this specific context, which have at least tens of thousands of atoms. Also, the properties of biomolecular systems are inextricably linked to their activity at physiological temperatures and in the presence of heterogeneous solvent environments, neither of which can be represented through QM calculations. Fortunately, if we are not interested in the reactivity and/or spectroscopic properties of a molecular system, its electronic structure can be disregarded. Indeed, according the Born-Oppenheimer (BO) approximation, nuclear and electronic motions can be treated separately, because of the large mass difference between them.

"Biomolecular simulations" is an umbrella term covering different calculation approaches that allow description of the structure, dynamics and function of large systems, with sizes ranging from a few thousand to millions of atoms by classical mechanics or by hybrid methods. What all these methods have in common is that the description of the majority of the system does not consider explicitly the electron density and that the nuclear motions are approximated through Newtonian, or classical, mechanics. In fact, the same laws of physics that govern the motion of all macroscopic objects, up to the size of planets and galaxies, work remarkably well in describing the motion of objects as small as atoms, at the very frontier where classical mechanics and QM coexist. In biomolecular simulations, classical mechanics is used to describe the atomic motions of a system represented by a potential energy function known as the force field (see the subsection below for an overview of the most popular carbohydrate-specific force fields).

According to Newton's 2nd law of motion the time rate of change of the momentum (*mv*) of a body is equal to the force (*F*) acting on it, both in terms of value and direction. More simply put, the structure and dynamics of the biomolecular system obtained from a simulation is determined by the forces acting on each atom at any given time step, in case of a molecular dynamics (MD) simulation,[22] and affects the stochastic probability of visiting different conformations in a Monte Carlo (MC) approach.[23] The choice of a force field appropriate to describe the system at hand is paramount in this context, as the force acting on the atoms is obtained as the derivative of the potential energy relative to the atoms positions. In the next two subsections, I briefly outline some of the fundamental elements in a successful simulation experiment, namely the choice of force field and conformational space sampling. For an in-depth discussion of these topics, the reader is referred to an excellent recent review.[22]

2.1 Carbohydrate force fields

A force field is a collection of parameters that within a specific mathematical framework produce the potential energy of a molecular system as a function of the position of all its atoms, *V(r)*. The most commonly used force fields in classical simulations are empirical force fields, where the term empirical refers to the nature of the parameters set. Within a typical empirical force field formalism, the interactions between atoms are divided into covalent and non-covalent, where covalent interactions comprise bond stretching, angles bending and dihedral torsion potentials, while non-covalent interactions include an electrostatic term and a repulsion and dispersion potential, the latter two usually combined in one term. The functional forms of these terms are shown with a corresponding representation of the specific interactions within and between monosaccharides in Fig. 1. These terms will be briefly described together with the specific approach used to derive them, for two of the most widely used all-atom carbohydrates empirical force fields nowadays, namely GLYCAM06[24] and CHARMM.[25–28] For a more comprehensive review of carbohydrate force fields and an historical background the reader is referred to ref. [29].

a) $V_{bond}(r_{ij}) = \frac{1}{2}k_{ij}^b(r_{ij} - b_{ij})^2$

b) $V_{angle}(\theta_{ijk}) = \frac{1}{2}k_{ijk}^a(\vartheta_{ijk} - a_{ijk})^2$

c) $V_{dihe}(\phi_{ijkl}) = \frac{1}{2}k_{ijkl}^d(1 + cos(n\phi - \gamma))$

d) $V_{LJ}(r_{ij}) = \frac{C_{ij}^{(12)}}{r_{ij}^{12}} - \frac{C_{ij}^{(6)}}{r_{ij}^6}$

e) $V_{elec}(r_{ij}) = f\frac{q_i q_j}{\varepsilon_r r_{ij}}$

Fig. 1 Potential energy terms included in an empirical force field. Covalent (bonded) interactions, listed on the left-hand side, include Hook-type harmonic potentials, used to represent (a) bond stretching and (b) angle bending. A periodic harmonic potential can be used to represent (c) dihedral torsions. The function of these potentials is graphically represented on the structure of α-L-fucose. Non-covalent (non-bonded) interactions, shown on the right-hand side, include: (d) a Lennard-Jones (LJ) potential that includes a short-range repulsion and long-range attraction (dispersion) terms and (e) a Coulomb potential to represent electrostatic interactions. The interactions are represented through arrows through stacking (top right) and hydrogen bonded (bottom right) monosaccharides.

2.1.1 GLYCAM06

GLYCAM06 is the first complete, stand-alone all-atom empirical force field for carbohydrates and glycoproteins.[24] Historically associated with the AMBER family of force fields,[30] the GLYCAM06 parameters set was independently developed in a way that remains consistent with AMBER parameters, while also being a stand-alone set to be used in combination with any other protein force field.[24] As a note of caution, in case of glycoproteins the linking residues with parameters developed to match the GLYCAM06 set can be found in all AMBER protein force fields.[24] As one of its distinctive features, the GLYCAM06 set does not have different α and β-anomeric atom types. This allows changes in ring conformations (or ring pucker) that are energetically accessible to different monosaccharides[31] and can be required in transitions between oligosaccharide conformations.[32-34] As shown in Fig. 1, the harmonic nature of the covalent terms requires the development of bonds, angles and dihedral constants and equilibrium values. In GLYCAM06 these are all derived from gas-phase QM calculations on appropriate molecular fragments. The QM and MM torsional energy barriers were fitted to minimize the error throughout the rotational potential energy curve for the force field to correctly reproduce the high degree of flexibility of oligo- and polysaccharides. As shown in Fig. 1, the electrostatic term in an empirical force field is represented by a Coulomb potential, where q are fixed point charges, centered on i and j atom pairs located at a distance r_{ij}, in an environment with dielectric ε_r. In GLYCAM06 the electrostatic charges are obtained by fitting the QM restrained electrostatic potential (RESP) and by averaging over all anomer-specific charges, which results in accurate evaluations of relative anomeric stability compared to QM values.[24] The electrostatic potential contribution is generally not calculated for bonded pairs (1–2,3) and in the AMBER force field is scaled for 1–4 pairs.[30] As a note of caution, to reproduce correctly the unique structural features of the glycosidic (1–6) linkage in disaccharides, the GLYCAM06 (1–4) electrostatic interactions are not scaled. Finally, the parameters for the Lennard-Jones (LJ) potential, representing repulsion and dispersion interactions, see Fig. 1, are taken directly from the AMBER PARM94 set, used in the current distributions.[30]

2.1.2 CHARMM

CHARMM parameter set for carbohydrates and glycoconjugates[25-28] was developed to be compatible with the CHARMM all-atom biomolecular force fields. As for GLYCAM06, parameters were derived from fragments and then transferred to monosaccharides, where missing parameters were developed and the set adjusted to fit gas phase QM and condensed phase experimental data. A final refinement was based on a comparison to MD simulation of infinite crystals and of differently diluted solutions aimed at reproducing crystal geometries of monosaccharides[25,27] and disaccharides.[28] In terms of fundamental analogies with the GLYCAM06 force field, in the CHARMM parameters set the same atom types are used for all diastereoisomers, such as α and β anomers, while consistently with the CHARMM all-atom biomolecular force fields, but not with the AMBER all-atom force fields, there is no scaling of electrostatic, including (1–4), interactions or of dispersion Lennard-Jones interactions. In an important more recent development, the CHARMM carbohydrate force field family includes now the first carbohydrate polarizable empirical force field[35] based on the Drude oscillator model.[36] Within this framework the point charge on the atoms responds to the electrostatic induction from the environment, adding a charge polarization effect to the empirical force field.[36] The parameters available to date cover 8 hexopyranose monosaccharides in α and β anomeric conformations,[35] 16 forms in total, 5 aldopentofuranoses and methyl derivatives,[37] and respective glycosidic linkages for the simulation of polysaccharides.[38]

2.2 Conformational sampling

The appropriate choice of empirical, additive or polarizable, force field is only one of the elements controlling the validity and reproducibility of a molecular simulation. Indeed, just as for wet lab-based experiments, simulations need to be reproducible, at least within the limitation of each specific force field model, to be meaningful and therefore useful. Exhaustive (or sufficient) conformational sampling is at the very core of this issue and can be translated into a common question often asked of MD simulations: "How long is long enough?". Unfortunately, there is not a standard answer to this question, as the problem is highly system dependent. In the specific case of complex carbohydrates and especially in cases where the glycan is not bound to a receptor, thereby retaining a high degree of flexibility, complete sampling may require considerable computational resources. An example of the problem is shown in Fig. 2, where the heat maps correspond to the free energy associated with different conformations of two glycosidic linkages, namely a β(1–4) and a α(1–6). MD simulations within a conventional implementation[22] allow exploration of the force field-defined conformational potential energy surface through a continuous walk in time. Let us assume we are using conventional MD to determine the structure(s) corresponding to the energy maps in Fig. 2. In the case of the β(1–4) linkage on the left-hand side, if the starting structure of our MD trajectory is the one corresponding to the red dot, a relatively short MD will be sufficient to characterize the most stable structure and to estimate its degree of flexibility in terms of standard deviations.

In the case of the α(1–6) linkage on the right-hand side of Fig. 2, there are several minima, as is expected from a flexible linkage. Some of these minima are connected by low energy paths, but others are not. The dashed line in Fig. 2 shows a potential MD trajectory that allows us to identify some, but not all, of the stable structures (minima). Indeed, the higher the energy barrier separating local minima, the lower the probability that a conventional MD trajectory will cross it at 300 K, and thus the longer the MD simulation should be. A better approach towards exploring the conformational space more effectively through conventional MD is to start uncorrelated MD runs from different structures, for example from the ones corresponding to the red dots in Fig. 2 panel B. Within this framework, each MD run can be potentially shorter and sampling completeness can be estimated from the
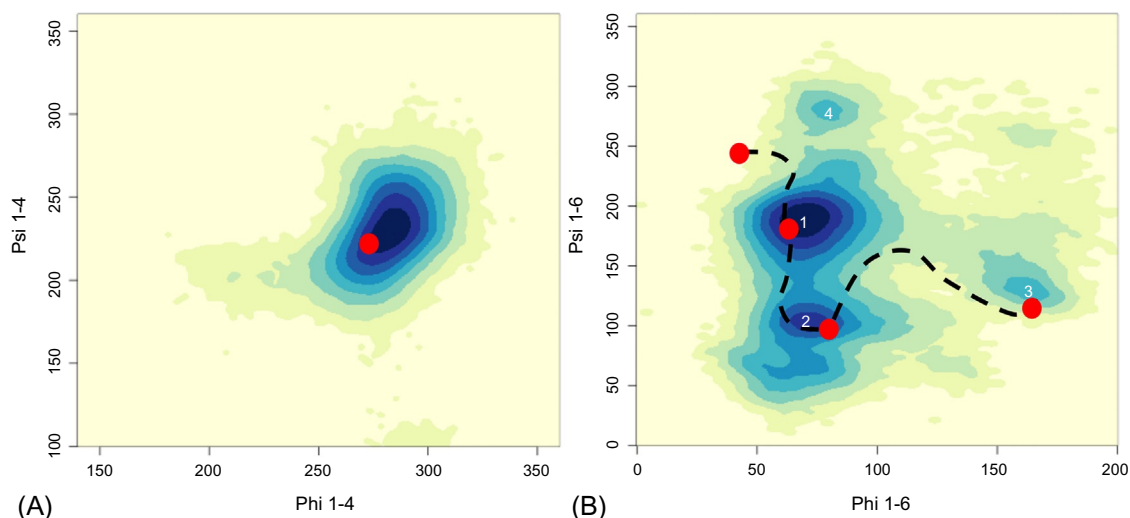
Fig. 2 Conformational heat maps corresponding to (A) a β(1−4) linkage and (B) an α(1−6) linkage obtained for a N-glycan through complete conformational mapping performed with a series of conventional MD simulations run in parallel. The red dots and black dashed lines represent a hypothetical walk in time space obtained by MD simulation. The numbers correspond to the minima identified during complete sampling,[39] with the darkest shade of blue corresponding to the most stable structure. Maps rendered with *RStudio* (https://rstudio.com/index2/).

degree of structural interconversion between the runs. This approach is particularly useful for the characterization of free carbohydrates,[33,39] where conformational interconversion does not generally involve crossing high energy barriers and therefore occurs without the need for enhanced sampling techniques within reasonable trajectory lengths, except in particularly difficult cases.[32,34]

In cases when the degrees of freedom of the glycan are restricted, because of interactions with the protein they are linked or bound to, or with each other in glycan-glycan complex networks that can occur in highly dense viral glycan shields, the accessible conformational space and its energetics can change dramatically. The number of minima can be reduced, with some conformations becoming inaccessible and separated by very high energy barriers. These issues are especially important to consider when, as almost always is the case in computational studies, complex carbohydrates are rebuilt de novo on the glycoprotein in conformations that are not necessarily relevant or occupied. As we will see in specific case studies described in the next sections, when conformational degrees of freedom are constrained, enhanced sampling techniques[40−45] can help to overcome high energy barriers and to achieve potentially complete sampling.

## 3   Glycosylation of viral envelope glycoproteins

Enveloped viruses are obligate pathogens encased in a lipid membrane, or envelope (Env).[3,46] The fusion of the viral and host-cell membranes results in infection,[47] with the viral fusion proteins facilitating entry.[6] Fusion is triggered through different, virus-specific mechanisms of recognition of cell surface protein or glycan receptors by envelope (Env) glycoproteins on the viral surface.[3,6,46−48] Viral Env proteins are all heavily glycosylated, with varying degrees of protein surface coverage, or shielding, and with specific N- and O-glycoforms.[3] These glycans can play very different functions, ranging from stabilization of the protein structure and assembly and trafficking through the ER/Golgi, to evasion of the immune system and enhancement of immune cell infection.[3] Specific glycans on the SARS-CoV-2 S glycoprotein play an intrinsic functional role, supporting the S protein active conformation in the prefusion stage.[10] Because viruses use the host-cell glycosylation machinery, the nature of the Env protein glycans is dictated by the host-cell origin and its characterization is important for understanding the immune response, infectivity and viral tropism.[3] Glycoprotein glycosylation analysis involves sequential enzymatic digestion of the recombinant viral proteins, followed by liquid chromatography (LC) coupled to mass spectrometry (MS). Recent progress in bioinformatics and sequencing technologies have greatly advanced glycoanalytics.[49−51] Nevertheless, the very nature and complexity of these steps leads to heavy dependencies on the specific details of laboratory protocols.[52] Molecular modelling has now reached the technological maturity and methodological sophistication to be able to play a crucial role in understanding the different roles of glycans in biology at the atomistic level. Complex glycan structures can be rebuilt on proteins and their structure, dynamics, interactions and recognition can be fully characterized through extensive conventional MD simulations and enhanced sampling techniques to provide a direct comparison with experimental data. In the following sections I will describe contributions that molecular simulations provided to our knowledge of the glycan shield in the SARS-CoV-2 S, the influenza HA and in the HIV-1 Env trimer.

3.1   The SARS-CoV-2 S glycoprotein

The COVID-19 pandemic is caused by a type of severe acute respiratory syndrome (SARS) coronavirus named CoV-2,[53] an enveloped positive-sense RNA virus similar to previously identified species, such as Middle-Eastern respiratory syndrome (MERS) and SARS-CoV.[3] These are the only highly pathogenic coronaviruses out of the 7 known to infect humans, while the remaining 4 have been found to cause much less severe "common cold"-type symptoms.[54] The primary mechanism of SARS-CoV-2 infection relies on the spike, or S, glycoprotein, protruding from the viral envelope membrane, see Fig. 3, and responsible for the virus' initial attachment to and fusion with the host cell. The S protein is a class I fusion protein homotrimer,[5–7,55,56] where each protomer is synthesized as a single polypeptide chain. The initial step of infection hinges on the interaction of the S protein receptor binding domain (RBD) with the angiotensin-converting enzyme 2 (ACE2), a host-cell bound receptor.[5,57–59] This interaction requires the activation of the S protein through a complex conformational transition, from a "closed" to an "open" conformation,[5,7] where the RBD on one protomer, or possibly two[4] or even all three in specific conditions, become exposed and accessible for recognition and binding. Important information on the RBD opening mechanism and pathway was obtained through unbiased enhanced sampling simulations in a model of the non-glycosylated S protein[60]; an effort that required enormous and unprecedented computational resources available through the Folding@Home distributed computing platform (www.foldingathome.org), corresponding to the first-to-date exascale computer.[60]

The SARS-CoV-2 S protein is heavily glycosylated, with each protomer bearing 22 potential N-glycosylation sites, 15–22 of which have been seen occupied in recombinant models,[5,7,8,61–63] and at least 3 mucin-type O-glycosylation sites, also with different degrees of occupancy observed.[61–63] All analytical studies concur in observing the prevalence of complex N-glycans in a typically dense shield,[3,8,61–63] a unusual trait where most evasion-strong viruses show a large majority of oligomannose type glycans.[3,64–67] Information on the structure and dynamics of the SARS-CoV-2 S glycan shield is unavailable through cryo-EM studies. Nevertheless, the presence of one N-linked GlcNAc or of a chitobiose at some sites[5,7] can be instrumental in indicating occupancy and demonstrating the spatial orientation of the N-glycan root, due to the rigidity of the GlcNAc-β(1–4)-GlcNAc linkage.[33,39] The complete reconstruction of the glycan shield was the result of several computational studies that used cryo-EM data, often coupled to homology modelling, to reconstruct the glycans with carbohydrate builder tools,[68,69] with glycoforms chosen based on available glycoanalytic data,[8,61,62,70] followed by conformational sampling through MD simulations.[10,62,71,72] Although the SARS-CoV-2 S glycan shield appears to be less dense than envelope glycoproteins from evasion-strong viruses, such as HIV-1[11,73–76] and specific strains of influenza HAs,[77–83] it provides a very effective coverage against large molecules, with over 90% of the accessible area shielded against 15 Å radius probes, corresponding to the approximate size of antibodies.[10] This percentage decreases down to an average of 20% for smaller molecule probes, such as water with a 1.4 Å radius.[10] The same study has revealed important vulnerabilities in the SARS-CoV-2 S glycan shield that can be exploited for the development of therapeutic, diagnostic and prevention strategies. The SARS-CoV-2 S RBD is sparsely glycosylated, bearing only two sequons at N331 and N343 and no O-glycosylation sites,[8] and is particularly exposed when in the open conformation.[10] Indeed, when the RBD is "up", an average
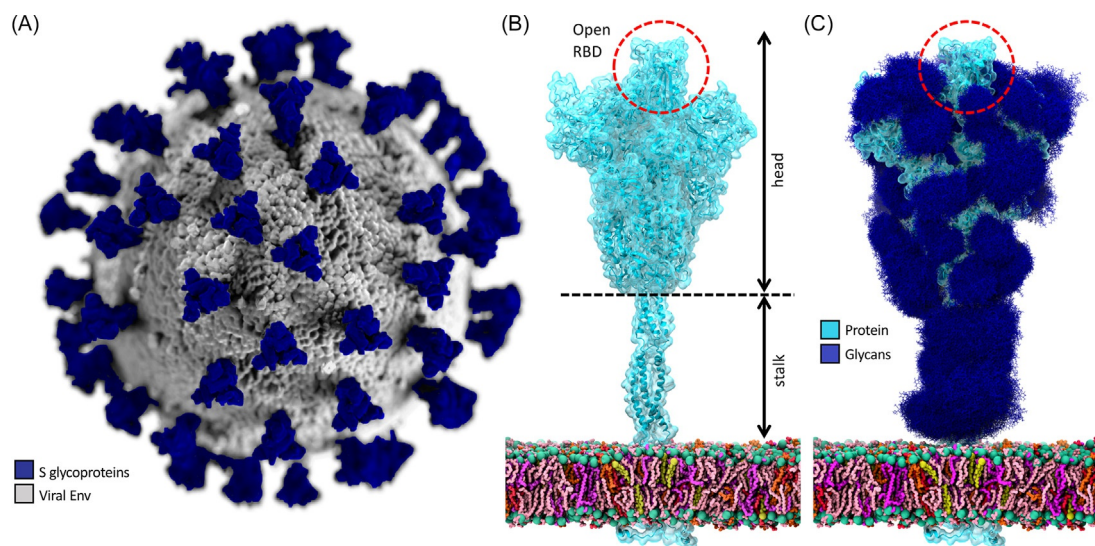

Fig. 3   (A) Graphical representation of the SARS-CoV 2 viral particle with the S glycoproteins highlighted in blue and the viral membrane and Env proteins embedded in grey. Original image from Wikipedia Commons, colorized with the GNU Image Manipulation Program GIMP (www.gimp.org). (B) Protein structure of the SARS-CoV 2 S trimer (cyan) embedded in the viral membrane (multicolored). The open conformation of the RBD is highlighted in the red circle. (C) Structure of the SARS-CoV 2 S trimer with the glycan shield highlighted in blue from MD simulations,[10] the open conformation of the RBD is highlighted in the red circle. Images (B) and (C) courtesy of L. Casalino and RE. Amaro. (UCSD).

of only 9% of the receptor binding motif (RBM, i.e., the RBD region that binds ACE2) surface area across all probes, is concealed by glycans, compared to 35% in the closed system, when the RBD is "down".[10] Furthermore, because the type of glycosylation depends directly on the nature of the host cell and tissue origin within the same species, it is important to assess if and how this shielding pattern can vary. Within this framework, reconstruction of the glycan shield with different glycoforms indicates similar degrees of average shielding.[71]

One important and unique feature of the SARS-CoV-2 S glycan shield that was discovered through extensive MD simulations is its integral role in stabilizing the S active "open" conformation, thus modulating its ACE2 binding activity.[10] Despite the majority of the N-glycans on the S surface being of the complex type, all available glycan analysis data on recombinant CoV-2 S from human cell lines[8,61−63] show the presence of a large oligomannose-type glycan (Man7 to Man9) at position N234 in each protomer's N-terminal domain (NTD). Unbiased MD simulations run in parallel on uncorrelated starting structures and with different force fields[10] have shown that a Man9 at N234 is able to rapidly crawl to fill the large volume left vacant by the opening of the RBD, see Fig. 4. The oligomannose glycoforms are perfectly sized and spatially shaped to fill the large volume and to reach the core of the trimer.[10] The same study also indicates that the glycans at positions N165 and N343 play a role in stabilizing the RBD open conformation and that single mutants N234A and double mutants N234A/N165A are characterized by a highly unstable RBD.[10] These computational predictions have been recently confirmed by biolayer interferometry experiments,[10] and supported by cryo-EM,[84] indicating that the absence of a glycan at N234 promotes the RBD closed conformation, which is inaccessible to ACE2.

The functional role of the glycans in the SARS-CoV-2 S-ACE2 interaction not only involves stabilization of the S glycoprotein's active conformation, but also directly contributes to the ACE2-RBD binding affinity.[62,85] Molecular simulation studies[62,85] of the fully glycosylated complexes have independently highlighted important contacts of the glycans at N90 and N322 on the ACE2 receptor with the S RBD in the complex and also extensive glycan-glycan contacts between the ACE2 glycan at N546 and the S glycans at N74 and N165.[62]

Molecular simulations are highly informative also to understand the roles of glycosaminoglycans (GAGs) in SARS-CoV-2 S activity and infection. To date there is extensive evidence that cellular heparan sulfate (HS) promotes ACE2 binding to SARS-CoV-2 S, potentially by forming a ternary complex that involves HS binding at the S RBD,[86] and that heparin also binds the S glycoprotein, possibly in different locations,[87] and inhibits SARS-CoV-2 entry.[88−90] Because structural data is still scarce, simulation studies need to be run in parallel with complementary mutagenesis and binding assays. Furthermore, the variable nature of GAG sequences and sulfation patterns and the characterization difficulties inherent in this context complicate the number of different simulations that need to be run to address the many questions at hand.
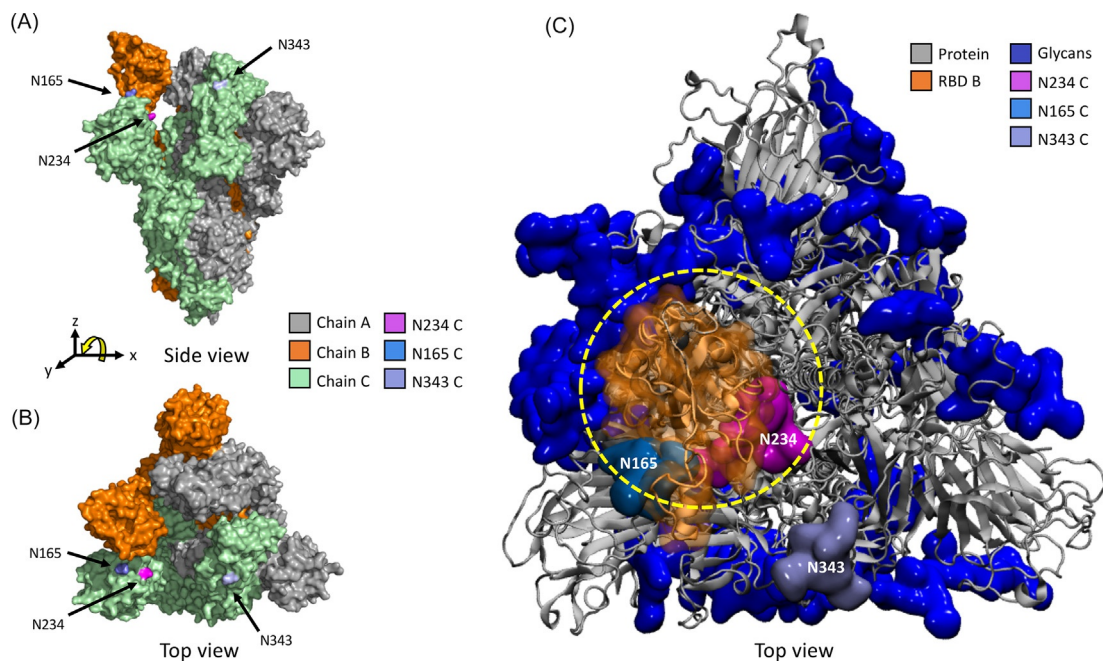
Fig. 4 (A) Side view of the SARS-CoV-2 S protein from PDB 6VYB, rendered by solvent accessible surface (pyMol) and color-coded by chain (protomer). In this structure the RBD in chain B (orange) is open, while the RBDs in chain A (grey) and C (light green) are closed. The positions of the key N-glycans in chain C are color-coded according to the legend. Structure rendered with pyMol. (B) Top view of the SARS-CoV-2 S protein from PDB 6VYB. (C) Reconstruction of the glycans (blue) followed by MD simulations show that the Man9 at N234 (magenta) inserts itself in the cleft left vacant by opening of the RBD, ultimately stabilizing the complex.[10] Structure rendered with VMD.

3.2   The influenza A hemagglutinin (HA)

An exceptional degree of adaptability and antigenic evolution has enabled the spread of the influenza A virus (IAV) across many different species[3,91–93] with occasionally devastating and regularly threatening consequences to human life.[94,95] IAV infection is initiated by the recognition of sialic acid-terminating sugars on the host cell surface by the envelope glycoprotein hemagglutinin (HA),[96,97] a class I fusion glycoprotein.[3,55] The distribution and dynamics of HAs on the IAV membrane surface was studied computationally through coarse-grained MD[98] and was illustrated recently in a landmark mesoscale simulation of the whole virus,[99] where HA competes for sialic acid with another envelope glycoprotein, neuraminidase (NA), which is responsible for the release of newly developed virions from the host cell. A Brownian Dynamics[100] (BD) study of different IAV models provided important information on the HA and NA binding kinetics on the viral surface[101] elucidating how substrate competition between these two proteins is balanced and how it can be exploited in therapeutic strategies.

HA is a trimer of heterodimers (Fig. 5) synthesized as a single polypeptide HA0 and then cleaved by host-cell proteases[3] into two subunits, HA1 and HA2, to produce the fully functional form of the protein.[102] HA1 contains the sialyllactosamine binding site, while HA2 is responsible for the fusion of the viral and cellular membranes, through a pH-dependent mechanism based on a complex conformational change.[103] 18 subtypes of HA (H1-18) have been characterized, with aquatic birds acting as primary reservoirs.[104] HA binds sialic acid-terminating epitopes (Fig. 5A) found on the surface of epithelial cells in the upper respiratory tract[97] with affinity and specificity changing between $\alpha(2–3)$ and $\alpha(2–6)$-linked sialic acids, depending on the IAV subtype.[105] Specificity for one linkage versus the other is important for interspecies transmission. In fact, the $\alpha(2–3)$-linked sialic acid is more common in the respiratory tract of avian species, while the $\alpha(2–6)$-linked is more common in mammalian species.[14,106–108]

Sequence hypervariability[109] especially on the HA1 subunit, the main target of antibody recognition,[3,110] has posed numerous challenges to the development of specific HA-targeted therapeutics and antiadhesives.[82] The HA1 mutations can occur within protein epitopes, and can also result in a marked increase of N-glycosylation sites, which result in radical antigenic changes.[111,112] As an example of the variability in the HA glycan shield, while the H1N1 from 1918[113] HA has 3 N-glycosites per protomer, the current H3N2[3,110] HA has at least 33 glycosylation sites per trimer. As opposed to the SARS-CoV-2 S, the majority of HA N-glycans are of the oligomannose type,[67] reflecting the potential restriction of the accessible space due to molecular crowding. As shown by an MD simulation study of H5N1,[80] the same molecular crowding can have a non-productive downside, reducing viral replication fitness[111] by decreasing IAV efficacy in binding sialylated receptors. MD studies of the H3N2 (3C.2a clade)[114] and of H3N2 (A/Aichi/68)[79] indicate that rearrangement of the N-glycan shield by loss and acquisition of different sequons in the HA1 domain
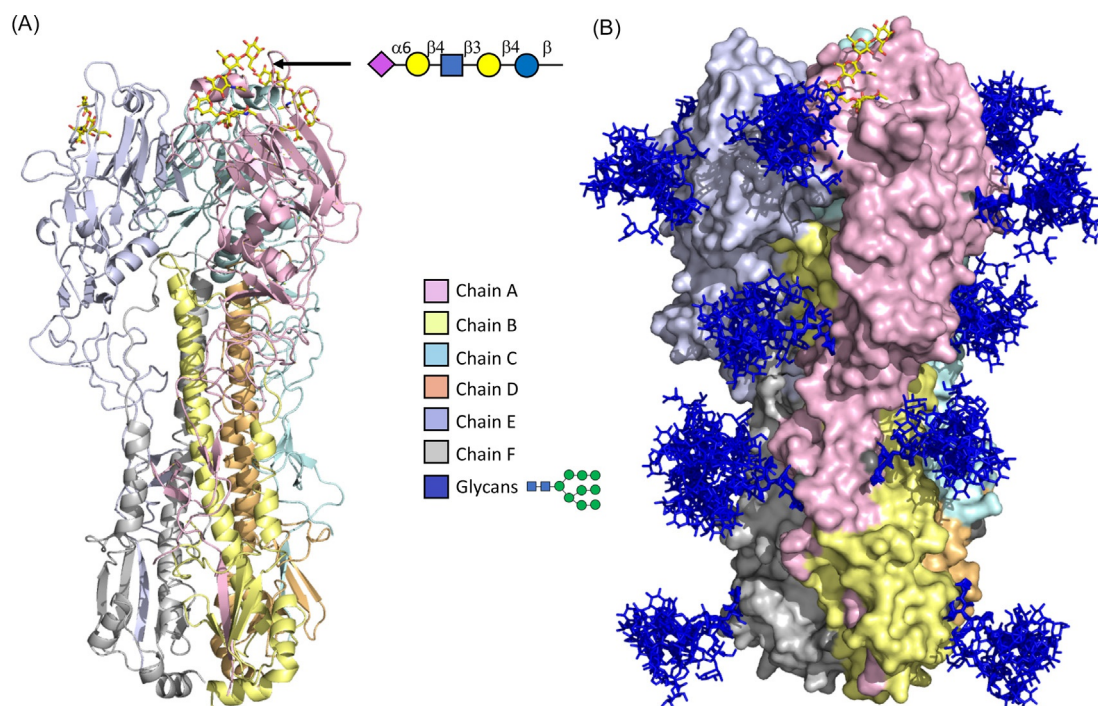
Fig. 5   (A) Structure of HA from the 1968 human H3N2 virus in complex with human receptor analogue LSTc (PDB 2YPG). The protein structure is rendered with cartoons with different coloring corresponding to each chain as shown in the legend. The human glycan epitope is rendered in sticks, with yellow C atoms, with sequence highlighted according to the SNFG nomenclature. (B) Protein structure rendered with solvent accessible surface with 16 N-glycosylation sites per trimer reconstructed for graphical purposes by structural alignment of the glycan fragments resolved in PDB 4FNK with Man 9 (blue) from snapshots selected from a 180 ns MD trajectory of the isolated glycans. All structures rendered with pyMol.

can change the IAV specificity and affinity by re-morphing the accessible space in the receptor binding site through the rearrangement of the glycan shield[79,114] and by directly contributing to binding through glycan-glycan interactions.[79]

3.3   The HIV-1 Env fusion trimer

Human immunodeficiency virus 1 (HIV-1) is an enveloped retrovirus responsible for the large majority of HIV/AIDS cases worldwide. HIV-1 primarily targets T-cells with high levels of CD4 glycoproteins and CCR5 (R5) co-receptors,[115,116] and then evolves to adhere to different cell types when immunodeficiency sets in.[115,117] This is due to the main HIV-1 Env fusion protein trimer switching preferential binding from the chemokine CCR5 (R5) co-receptor to CXCR4 (X4), which allows it to infect different CD4+ T-cells.[115–117] The HIV-1 fusion protein is a trimer of heterodimers, synthesized as a single polypeptide, namely gp160, and then cleaved in two parts, namely gp120 and gp41, in the active form of the protein (Fig. 6). The HIV-1 Env trimer is heavily glycosylated, with ~90 potentially occupied N-glycosites,[11,118] predominantly of the oligomannose type.[3,11,75,76,119,120]

In the prefusion conformation the HIV-1 Env trimer binds the CD4 cell-bound receptor.[13] This leads to a conformational change that promotes binding to the secondary co-receptors CCR5 or CXCR4, leading to fusion with the host cell.[116] Understanding the underlying mechanism governing these processes from the structural point of view is made particularly difficult not only by the extensive glycosylation of the HIV-1 Env trimer, but also by the presence of highly flexible loops directly involved in recognition and binding. Protein crystallization often requires removal of these carbohydrates and sequence modifications and truncations,[12,13,15,118,121,122] which may preclude the acquisition of important information. Indeed, the hypervariable V3 loop is a key determinant of HIV-1 cellular tropism.[123] Also, V1/2 and V3 loops are key targets of neutralizing antibodies.[124,125] Within this framework molecular simulations have contributed a great deal to elucidating the structure and dynamics of the HIV-1 Env trimer glycan shield[11,73,126–128] with sampling limitations in some cases determined by the system's size in relation to the HPC infrastructure available at the time the work was performed.

Crystallographic analysis of the complete and fully glycosylated trimeric HIV-1 Env fusion proteins from clades A, B and G[11] shows that the glycan shield density changes from highly crowded regions around the solvent exposed gp120, to a low crowding region around the gp41 stem. In high density regions, the oligomannose N-glycans outstretch from the gp120 surface through their rigid chitobiose core,[33] while the α(1–3) and α(1–6) arms extend perpendicularly, interacting with neighboring N-glycans to form a dense and ordered glycan-glycan network that effectively shields the protein surface for immune evasion.[11] This landmark structural work[11] was complemented by canonical MD simulations, more specifically by three 500 ns trajectories with either Man5, Man7 or Man9 glycans at every sequon. These simulations confirm the stability of the glycan-glycan interactions and support their contribution to the shield in high density regions and also inform on the volume of the binding regions occupied by N-glycans required for binding CD4 and broadly neutralizing antibodies.[11] To provide further insight, these MD simulations were extended to 2 μs in later work from the same group,[126] which delivered a more complete picture of the collective dynamics of the shield and of protein surface accessibility.

Although informative in terms of immediate relaxation and stability of the glycan shield, the results obtained from canonical MD simulations of highly crowded, interacting glycans patches, such as the ones found on the gp120s of the HIV-1 Env fusion
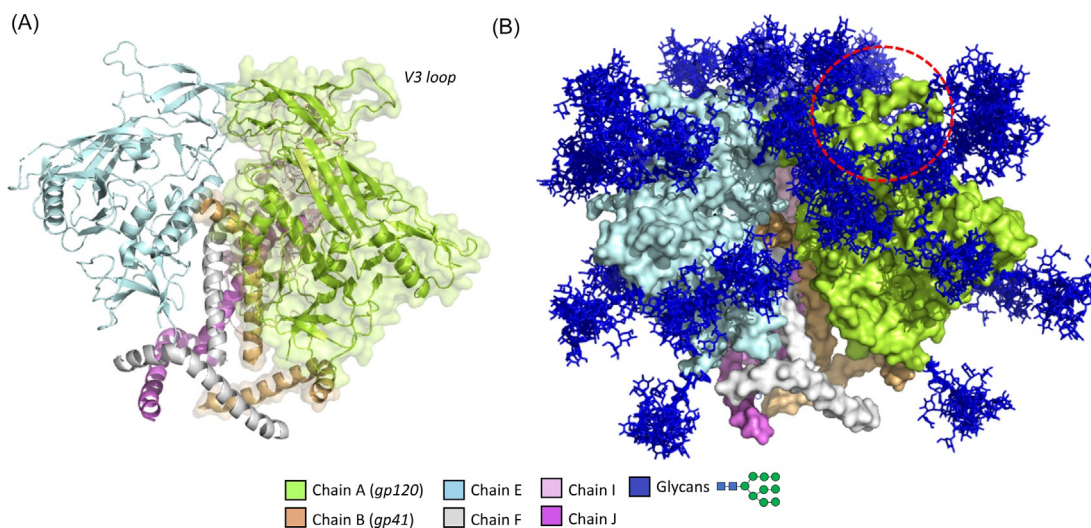
Fig. 6   (A) Structure of the HIV-1 Env trimer glycoprotein from PDB 4NCO rendered in cartoons with coloring corresponding to the legend at the bottom. The location of the gp120 (chain A)/gp41 (chain B) protomer is highlighted with a transparent solvent accessible surface area. The position of the V3 loop in chain A is also highlighted. (B) Structure of the HIV-1 Env trimer glycoprotein from PDB 4NCO with glycosylation (blue) reconstructed for graphical purposes in 41 sites by structural alignment of the resolved glycan fragments with Man 9 snapshots selected from a 180 ns MD trajectory of the isolated glycans. The position of the V3 loop is highlighted within the red dotted circle. All structures rendered with pyMol.

trimer, may not always provide a significant advance relative to analysis of the structural data, when available. Notable in this particular context is the excellent work by Yang et al.[73] that shows how enhanced sampling MD achieved through Hamiltonian Replica Exchange (HREX) is a much more informative approach, when the size of the system and computational resources limit sampling through conventional MD. Indeed, the HREX simulations reveal a complete image of the N-glycan conformational space and heterogeneity, providing important insight on the pre-structuring of the N-glycans for recognition and binding by broadly neutralizing antibodies and of their accessibility, as well as the accessibility of the protein surface to CD4 receptors and to CCR5 and CXCR4 co-receptors.[73]

4 Conclusions and perspectives

Extensive glycosylation is a typical feature of viral envelope fusion proteins that has both intrinsic functions, regulating folding and trafficking, and extrinsic functions, modulating immune evasion, substrate and receptor recognition and binding affinity[3] and actively participating in the protein activation mechanism.[10] Advances in HPC technology, simulation software and in force fields sophistication now allow molecular simulations to provide not only a platform technology complementary to standard structural biology experimental techniques, but also to be an independent, primary research tool. Molecular modelling studies of the SARS-CoV-2 S, HA and HIV-1 Env fusion trimer have supported a rapid advance in our knowledge of viral glycoscience. Similar points apply to all cases where the structural and dynamic role of glycans are difficult to characterize by any other research method.

In the case studies discussed, the highly dynamic nature of the glycans generally allows for their conformational relaxation in much more rapid timescales relative to the structured proteins to which they are linked. This is fundamentally true when the glycans interact very little with each other or with the protein(s). Nevertheless, it is important to be aware that highly dense glycan networks and/or in cases with multiple interactions with the protein, ligand or receptor, the conformational energy landscape can dramatically change and enhanced sampling techniques can provide a much more complete picture. The wider availability of very large computational resources now allows all-atom simulations of large proteins and even mesoscale simulations of the whole viral capsid to provide a detailed atomistic view of glycobiology and an alternative way to uncover all (or most of) its very many secrets. Moreover, the enormous molecular weight of transmembrane viral fusion proteins that has limited the investigative power of molecular modelling for many years, no longer represents a barrier and the future of computational glycoscience has never looked brighter.

Acknowledgments

References

1. Graham, B. S.; Gilman, M. S. A.; McLellan, J. S. *Annu. Rev. Med.* **2019**, *70*, 91–104.
2. bioRxiv COVID-19 SARS-CoV-2 preprints from medRxiv and bioRxiv. https://connect.biorxiv.org/relate/content/181.
3. Watanabe, Y.; Bowden, T. A.; Wilson, I. A.; Crispin, M. *Biochim. Biophys. Acta Gen. Subj.* **2019**, *1863* (10), 1480–1497.
4. Ke, Z.; Oton, J.; Qu, K.; Cortese, M.; Zila, V.; McKeane, L.; Nakane, T.; Zivanov, J.; Neufeldt, C. J.; Cerikan, B.; Lu, J. M.; Peukes, J.; Xiong, X.; Kräusslich, H. G.; Scheres, S. H. W.; Bartenschlager, R.; Briggs, J. A. G. *Nature* **2020**. https://doi.org/10.1038/s41586-020-2665-2.
5. Walls, A. C.; Park, Y. J.; Tortorici, M. A.; Wall, A.; McGuire, A. T.; Veesler, D. *Cell* **2020**, *181* (2), 281–292.e6.
6. Tortorici, M. A.; Veesler, D. *Adv. Virus Res.* **2019**, *105*, 93–116.
7. Wrapp, D.; Wang, N.; Corbett, K. S.; Goldsmith, J. A.; Hsieh, C. L.; Abiona, O.; Graham, B. S.; McLellan, J. S. *Science* **2020**, *367* (6483), 1260–1263.
8. Watanabe, Y.; Allen, J. D.; Wrapp, D.; McLellan, J. S.; Crispin, M. *Science* **2020**, *369* (6501), 330–333.
9. Varki, A.; Lowe, J. In *Essentials of Glycobiology*, 2nd Ed.; Varki, A., Cummings, R. D., Esko, J. D., et al. Cold Spring Harbor Laboratory Press: New York, 2009.
10. Casalino, L.; Gaieb, Z.; Goldsmith, J.; Hjort, C.; Dommer, A.; Harbison, A.; Fogarty, C.; Barros, E.; Taylor, B.; McLellan, J.; Fadda, E.; Amaro, R. *ACS Cent. Sci.* **2020**, *6* (10), 1722–1734.
11. Stewart-Jones, G.; Soto, C.; Lemmin, T.; Chuang, G.; Druz, A.; Kong, R.; Thomas, P.; Wagh, K.; Zhou, T.; Behrens, A.; Bylund, T.; Choi, C.; Davison, J.; Georgiev, I.; Joyce, M.; Do Kwon, Y.; Pancera, M.; Taft, J.; Yang, Y.; Zhang, B.; Shivatare, S.; Shivatare, V.; Lee, C.; Wu, C.; Bewley, C.; Burton, D.; Koff, W.; Connors, M.; Crispin, M.; Baxa, U.; Korber, B.; Wong, C.; Mascola, J.; Kwong, P. *Cell* **2016**, *165* (4), 813–826.
12. Kwon, Y. D.; Pancera, M.; Acharya, P.; Georgiev, I. S.; Crooks, E. T.; Gorman, J.; Joyce, M. G.; Guttman, M.; Ma, X.; Narpala, S.; Soto, C.; Terry, D. S.; Yang, Y.; Zhou, T.; Ahlsen, G.; Bailer, R. T.; Chambers, M.; Chuang, G. Y.; Doria-Rose, N. A.; Druz, A.; Hallen, M. A.; Harned, A.; Kirys, T.; Louder, M. K.; O'Dell, S.; Ofek, G.; Osawa, K.; Prabhakaran, M.; Sastry, M.; Stewart-Jones, G. B.; Stuckey, J.; Thomas, P. V.; Tittley, T.; Williams, C.; Zhang, B.; Zhao, H.; Zhou, Z.; Donald, B. R.; Lee, L. K.; Zolla-Pazner, S.; Baxa, U.; Schön, A.; Freire, E.; Lee, K. K.; Arthos, J.; Munro, J. B.; Blanchard, S. C.; Mothes, W.; Binley, J. M.; McDermott, A. B.; Mascola, J. R.; Kwong, P. D. *Nat. Struct. Mol. Biol.* **2015**, *22* (7), 522–531.
13. Kwong, P. D.; Wyatt, R.; Robinson, J.; Sweet, R. W.; Sodroski, J.; Hendrickson, W. A. *Nature* **1998**, *393* (6686), 648–659.
14. Nicholls, J. M.; Chan, M. C.; Chan, W. Y.; Wong, H. K.; Cheung, C. Y.; Kwong, D. L.; Wong, M. P.; Chui, W. H.; Poon, L. L.; Tsao, S. W.; Guan, Y.; Peiris, J. S. *Nat. Med.* **2007**, *13* (2), 147–149.

15. Pancera, M.; Zhou, T.; Druz, A.; Georgiev, I. S.; Soto, C.; Gorman, J.; Huang, J.; Acharya, P.; Chuang, G. Y.; Ofek, G.; Stewart-Jones, G. B.; Stuckey, J.; Bailer, R. T.; Joyce, M. G.; Louder, M. K.; Tumba, N.; Yang, Y.; Zhang, B.; Cohen, M. S.; Haynes, B. F.; Mascola, J. R.; Morris, L.; Munro, J. B.; Blanchard, S. C.; Mothes, W.; Connors, M.; Kwong, P. D. *Nature* **2014**, *514* (7523), 455–461.
16. Thaysen-Andersen, M.; Packer, N. H. *Glycobiology* **2012**, *22* (11), 1440–1452.
17. Zacchi, L. F.; Schulz, B. L. *Glycoconj. J.* **2016**, *33* (3), 359–376.
18. van Mourik, T.; Bühl, M.; Gaigeot, M. P. *Phil. Trans. A Math. Phys. Eng. Sci.* **2014**, *372* (2011), 20120488.
19. Koch, W.; Holthausen, M. *A Chemist's Guide to Density Functional Theory*; Wiley-VCH Verlag GmbH, 2001.
20. Burke, K. *J. Chem. Phys.* **2012**, *136* (15), 150901.
21. Friesner, R. A. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102* (19), 6648–6653.
22. Braun, E.; Gilmer, J.; Mayes, H. B.; Mobley, D. L.; Monroe, J. I.; Prasad, S.; Zuckerman, D. M. *LiveCoMS* **2018**, *1* (1), 5957. https://doi.org/10.33011/livecoms.1.1.5957.
23. Hansmann, U. H.; Okamoto, Y. *Curr. Opin. Struct. Biol.* **1999**, *9* (2), 177–183.
24. Kirschner, K. N.; Yongye, A. B.; Tschampel, S. M.; González-Outeiriño, J.; Daniels, C. R.; Foley, B. L.; Woods, R. J. *J. Comput. Chem.* **2008**, *29* (4), 622–655.
25. Guvench, O.; Mallajosyula, S.; Raman, E.; Hatcher, E.; Vanommeslaeghe, K.; Foster, T.; Jamison, F.; MacKerell, A. *J. Chem. Theory Comput.* **2011**, *7* (10), 3162–3180.
26. Mallajosyula, S. S.; Jo, S.; Im, W.; MacKerell, A. D. *Methods Mol. Biol.* **2015**, *1273*, 407–429.
27. Guvench, O.; Greene, S. N.; Kamath, G.; Brady, J. W.; Venable, R. M.; Pastor, R. W.; Mackerell, A. D. *J. Comput. Chem.* **2008**, *29* (15), 2543–2564.
28. Guvench, O.; Hatcher, E. R.; Venable, R. M.; Pastor, R. W.; Mackerell, A. D. *J. Chem. Theory Comput.* **2009**, *5* (9), 2353–2370.
29. Fadda, E.; Woods, R. J. *Drug Discov. Today* **2010**, *15* (15–16), 596–609.
30. Case, D.; Ben-Shalom, I.; Brozell, S.; Cerutti, D.; Cheatham, T., III; Cruzeiro, V.; Darden, T.; Duke, R.; Ghoreishi, D.; Gilson, M.; Gohlke, H.; Goetz, A.; Greene, D.; Harris, R.; Homeyer, N.; Izadi, S.; Kovalenko, A.; Kurtzman, T.; Lee, T.; LeGrand, S.; Li, P.; Lin, C.; Liu, J.; Luchko, T.; Luo, R.; Mermelstein, D.; Merz, K.; Miao, Y.; Monard, G.; Nguyen, C.; Nguyen, H.; Omelyan, I.; Onufriev, A.; Pan, F.; Qi, R.; Roe, D.; Roitberg, A.; Sagui, C.; Schott-Verdugo, S.; Shen, J.; Simmerling, C.; Smith, J.; Salomon-Ferrer, R.; Swails, J.; Walker, R.; Wang, J.; Wei, H.; Wolf, R.; Wu, X.; Xiao, L.; York, D.; Kollman, P. *AMBER*; **2018**. University of California: San Francisco, 2018.
31. Mayes, H. B.; Broadbelt, L. J.; Beckham, G. T. *J. Am. Chem. Soc.* **2014**, *136* (3), 1008–1022.
32. Alibay, I.; Burusco, K.; Bruce, N.; Bryce, R. *J. Phys. Chem. B* **2018**, *122* (9), 2462–2474.
33. Harbison, A. M.; Brosnan, L. P.; Fenlon, K.; Fadda, E. *Glycobiology* **2019**, *29* (1), 94–103.
34. Topin, J.; Lelimousin, M.; Arnaud, J.; Audfray, A.; Perez, S.; Varrot, A.; Imberty, A. *ACS Chem. Biol.* **2016**, *11* (7), 2011–2020.
35. Patel, D. S.; He, X.; MacKerell, A. D. *J. Phys. Chem. B* **2015**, *119* (3), 637–652.
36. Lemkul, J. A.; Huang, J.; Roux, B.; MacKerell, A. D. *Chem. Rev.* **2016**, *116* (9), 4983–5013.
37. Jana, M.; MacKerell, A. D. J. *J. Phys. Chem. B* **2015**, *119* (25), 7846–7859.
38. Aytenfisu, A.; Yang, M.; MacKerell, A. *J. Chem. Theory Comput.* **2018**, *14* (6), 3132–3143.
39. Fogarty, C.; Harbison, A.; Dugdale, A.; Fadda, E. *Beilstein J. Org. Chem.* **2020**, *16*, 2046–2056.
40. Tiwary, P.; van der Walle, A.; Weinberger, C. R.; Tucker, G. J., Eds.; In *Multiscale Materials Modeling for Nanomechanics*; Springer International: Switzerland, 2016; pp 195–221.
41. Bernardi, R. C.; Melo, M. C. R.; Schulten, K. *Biochim. Biophys. Acta* **2015**, *1850* (5), 872–877.
42. Okamoto, Y. *J. Mol. Graph. Model.* **2004**, *22* (5), 425–439.
43. Abrams, C.; Bussi, G. *Entropy* **2014**, *16* (1), 163–199.
44. Spiwok, V.; Sucur, Z.; Hosek, P. *Biotechnol. Adv.* **2015**, *33* (6 Pt 2), 1130–1140.
45. Allison, J. R. *Biochem. Soc. Trans.* **2020**.
46. Rey, F. A.; Lok, S. M. *Cell* **2018**, *172* (6), 1319–1334.
47. Plemper, R. K. *Curr. Opin. Virol.* **2011**, *1* (2), 92–100.
48. Morizono, K.; Chen, I. S. *Curr. Opin. Virol.* **2011**, *1* (1), 13–18.
49. Alocci, D.; Mariethoz, J.; Gastaldello, A.; Gasteiger, E.; Karlsson, N. G.; Kolarich, D.; Packer, N. H.; Lisacek, F. *J. Proteome Res.* **2019**, *18* (2), 664–677.
50. Rojas-Macias, M. A.; Mariethoz, J.; Andersson, P.; Jin, C.; Venkatakrishnan, V.; Aoki, N. P.; Shinmachi, D.; Ashwood, C.; Madunic, K.; Zhang, T.; Miller, R. L.; Horlacher, O.; Struwe, W. B.; Watanabe, Y.; Okuda, S.; Levander, F.; Kolarich, D.; Rudd, P. M.; Wuhrer, M.; Kettner, C.; Packer, N. H.; Aoki-Kinoshita, K. F.; Lisacek, F.; Karlsson, N. G. *Nat. Commun.* **2019**, *10* (1), 3275.
51. Zhang, P.; Woen, S.; Wang, T.; Liau, B.; Zhao, S.; Chen, C.; Yang, Y.; Song, Z.; Wormald, M. R.; Yu, C.; Rudd, P. M. *Drug Discov. Today* **2016**, *21* (5), 740–765.
52. De Leoz, M. L. A.; Duewer, D. L.; Fung, A.; Liu, L.; Yau, H. K.; Potter, O.; Staples, G. O.; Furuki, K.; Frenkel, R.; Hu, Y.; Sosic, Z.; Zhang, P.; Altmann, F.; Gru Nwald-Grube, C.; Shao, C.; Zaia, J.; Evers, W.; Pengelley, S.; Suckau, D.; Wiechmann, A.; Resemann, A.; Jabs, W.; Beck, A.; Froehlich, J. W.; Huang, C.; Li, Y.; Liu, Y.; Sun, S.; Wang, Y.; Seo, Y.; An, H. J.; Reichardt, N. C.; Ruiz, J. E.; Archer-Hartmann, S.; Azadi, P.; Bell, L.; Lakos, Z.; An, Y.; Cipollo, J. F.; Pucic-Bakovic, M.; Štambuk, J.; Lauc, G.; Li, X.; Wang, P. G.; Bock, A.; Hennig, R.; Rapp, E.; Creskey, M.; Cyr, T. D.; Nakano, M.; Sugiyama, T.; Leung, P. A.; Link-Lenczowski, P.; Jaworek, J.; Yang, S.; Zhang, H.; Kelly, T.; Klapoetke, S.; Cao, R.; Kim, J. Y.; Lee, H. K.; Lee, J. Y.; Yoo, J. S.; Kim, S. R.; Suh, S. K.; de Haan, N.; Falck, D.; Lageveen-Kammeijer, G. S. M.; Wuhrer, M.; Emery, R. J.; Kozak, R. P.; Liew, L. P.; Royle, L.; Urbanowicz, P. A.; Packer, N. H.; Song, X.; Everest-Dass, A.; Lattová, E.; Cajic, S.; Alagesan, K.; Kolarich, D.; Kasali, T.; Lindo, V.; Chen, Y.; Goswami, K.; Gau, B.; Amunugama, R.; Jones, R.; Stroop, C. J. M.; Kato, K.; Yagi, H.; Kondo, S.; Yuen, C. T.; Harazono, A.; Shi, X.; Magnelli, P. E.; Kasper, B. T.; Mahal, L.; Harvey, D. J.; O'Flaherty, R.; Rudd, P. M.; Saldova, R.; Hecht, E. S.; Muddiman, D. C.; Kang, J.; Bhoskar, P.; Menard, D.; Saati, A.; Merle, C.; Mast, S.; Tep, S.; Truong, J.; Nishikaze, T.; Sekiya, S.; Shafer, A.; Funaoka, S.; Toyoda, M.; de Vreugd, P.; Caron, C.; Pradhan, P.; Tan, N. C.; Mechref, Y.; Patil, S.; Rohrer, J. S.; Chakrabarti, R.; Dadke, D.; Lahori, M.; Zou, C.; Cairo, C.; Reiz, B.; Whittal, R. M.; Lebrilla, C. B.; Wu, L.; Guttman, A.; Szigeti, M.; Kremkow, B. G.; Lee, K. H.; Sihlbom, C.; Adamczyk, B.; Jin, C.; Karlsson, N. G.; Örnros, J.; Larson, G.; Nilsson, J.; Meyer, B.; Wiegandt, A.; Komatsu, E.; Perreault, H.; Bodnar, E. D.; Said, N.; Francois, Y. N.; Leize-Wagner, E.; Maier, S.; Zeck, A.; Heck, A. J. R.; Yang, Y.; Haselberg, R.; Yu, Y. Q.; Alley, W.; Leone, J. W.; Yuan, H.; Stein, S. E. *Mol. Cell. Proteomics* **2020**, *19* (1), 11–30.
53. Chan, J. F.; Yuan, S.; Kok, K. H.; To, K. K.; Chu, H.; Yang, J.; Xing, F.; Liu, J.; Yip, C. C.; Poon, R. W.; Tsoi, H. W.; Lo, S. K.; Chan, K. H.; Poon, V. K.; Chan, W. M.; Ip, J. D.; Cai, J. P.; Cheng, V. C.; Chen, H.; Hui, C. K.; Yuen, K. Y. *Lancet* **2020**, *395* (10223), 514–523.
54. Li, F. *J. Virol.* **2015**, *89* (4), 1954–1964.
55. White, J. M.; Delos, S. E.; Brecher, M.; Schornberg, K. *Crit. Rev. Biochem. Mol. Biol.* **2008**, *43* (3), 189–219.
56. Bosch, B. J.; van der Zee, R.; de Haan, C. A.; Rottier, P. J. *J. Virol.* **2003**, *77* (16), 8801–8811.
57. Lan, J.; Ge, J.; Yu, J.; Shan, S.; Zhou, H.; Fan, S.; Zhang, Q.; Shi, X.; Wang, Q.; Zhang, L.; Wang, X. *Nature* **2020**, *581* (7807), 215–220.
58. Wan, Y.; Shang, J.; Graham, R.; Baric, R. S.; Li, F. *J. Virol.* **2020**, *94* (7).
59. Shang, J.; Ye, G.; Shi, K.; Wan, Y.; Luo, C.; Aihara, H.; Geng, Q.; Auerbach, A.; Li, F. *Nature* **2020**, *581* (7807), 221–224.
60. Zimmerman, M. I.; Porter, J. R.; Ward, M. D.; Singh, S.; Vithani, N.; Meller, A.; Mallimadugula, U. L.; Kuhn, C. E.; Borowsky, J. H.; Wiewiora, R. P.; Hurley, M. F. D.; Harbison, A. M.; Fogarty, C. A.; Coffland, J. E.; Fadda, E.; Voelz, V. A.; Chodera, J. D.; Bowman, G. R. *bioRxiv* **2020**.
61. Sanda, M.; Morrison, L.; Goldman, R. *bioRxiv* **2020**.
62. Zhao, P.; Praissman, J. L.; Grant, O. C.; Cai, Y.; Xiao, T.; Rosenbalm, K. E.; Aoki, K.; Kellman, B. P.; Bridger, R.; Barouch, D. H.; Brindley, M. A.; Lewis, N. E.; Tiemeyer, M.; Chen, B.; Woods, R. J.; Wells, L. *bioRxiv* **2020**.
63. Shajahan, A.; Supekar, N. T.; Gleinich, A. S.; Azadi, P. *Glycobiology* **2020**.
64. Doores, K. J.; Bonomelli, C.; Harvey, D. J.; Vasiljevic, S.; Dwek, R. A.; Burton, D. R.; Crispin, M.; Scanlan, C. N. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107* (31), 13800–13805.
65. Struwe, W. B.; Chertova, E.; Allen, J. D.; Seabright, G. E.; Watanabe, Y.; Harvey, D. J.; Medina-Ramirez, M.; Roser, J. D.; Smith, R.; Westcott, D.; Keele, B. F.; Bess, J. W.; Sanders, R. W.; Lifson, J. D.; Moore, J. P.; Crispin, M. *Cell Rep.* **2018**, *24* (8), 1958–1966.e5.

66. Watanabe, Y.; Raghwani, J.; Allen, J. D.; Seabright, G. E.; Li, S.; Moser, F.; Huiskonen, J. T.; Strecker, T.; Bowden, T. A.; Crispin, M. *Proc. Natl. Acad. Sci. U. S. A.* **2018**, *115* (28), 7320–7325.
67. An, Y.; McCullers, J. A.; Alymova, I.; Parsons, L. M.; Cipollo, J. F. *J. Proteome Res.* **2015**, *14* (9), 3957–3969.
68. Park, S. J.; Lee, J.; Qi, Y.; Kern, N. R.; Lee, H. S.; Jo, S.; Joung, I.; Joo, K.; Im, W. *Glycobiology* **2019**, *29* (4), 320–331.
69. Woods R. GLYCAM Web (http://glycam.org); Complex Carbohydrate Research Center, University of Georgia: Athens, GA, 2013.
70. Shajahan, A.; Supekar, N.; Gleinich, A.; Azadi, P. *Glycobiology* **2020**. https://doi.org/10.1093/glycob/cwaa042.
71. Grant, O. C.; Montgomery, D.; Ito, K.; Woods, R. J. *bioRxiv* **2020**.
72. Turoňová, B.; Sikora, M.; Schürmann, C.; Hagen, W.; Welsch, S.; Blanc, F.; von Bülow, S.; Gecht, M.; Bagola, K.; Hörner, C.; van Zandbergen, G.; Landry, J.; Trevisan Doimo de Azevedo, N.; Mosalaganti, S.; Schwarz, A.; Covino, R.; Mühlebach, M.; Hummer, G.; Krijnse Locker, J.; Beck, M. *Science* **2020**eabd5223.
73. Yang, M.; Huang, J.; Simon, R.; Wang, L. X.; MacKerell, A. D. *Sci. Rep.* **2017**, *7* (1), 4435.
74. Behrens, A. J.; Crispin, M. *Curr. Opin. Struct. Biol.* **2017**, *44*, 125–133.
75. Bonomelli, C.; Doores, K. J.; Dunlop, D. C.; Thaney, V.; Dwek, R. A.; Burton, D. R.; Crispin, M.; Scanlan, C. N. *PLoS One* **2011**, *6* (8)e23521.
76. Crispin, M.; Ward, A.; Wilson, I.; Dill, K. *Annu. Rev. Biophys.* **2018**, *47*, 499–523.
77. Wu, C. Y.; Lin, C. W.; Tsai, T. I.; Lee, C. D.; Chuang, H. Y.; Chen, J. B.; Tsai, M. H.; Chen, B. R.; Lo, P. W.; Liu, C. P.; Shivatare, V. S.; Wong, C. H. *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *114* (2), 280–285.
78. Amaro, R. E.; Li, W. W. *Methods Mol. Biol.* **2012**, *819*, 575–594.
79. Kasson, P. M.; Pande, V. S. *Biophys. J.* **2008**, *95* (7), L48–L50.
80. Chen, W.; Sun, S.; Li, Z. *PLoS One* **2012**, *7* (6)e38794.
81. Wang, C. C.; Chen, J. R.; Tseng, Y. C.; Hsu, C. H.; Hung, Y. F.; Chen, S. W.; Chen, C. M.; Khoo, K. H.; Cheng, T. J.; Cheng, Y. S.; Jan, J. T.; Wu, C. Y.; Ma, C.; Wong, C. H. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106* (43), 18137–18142.
82. Ji, Y.; White, Y. J.; Hadden, J. A.; Grant, O. C.; Woods, R. J. *Curr. Opin. Struct. Biol.* **2017**, *44*, 219–231.
83. Tate, M. D.; Job, E. R.; Deng, Y. M.; Gunalan, V.; Maurer-Stroh, S.; Reading, P. C. *Viruses* **2014**, *6* (3), 1294–1316.
84. Henderson, R.; Edwards, R. J.; Mansouri, K.; Janowska, K.; Stalls, V.; Kopp, M.; Haynes, B. F.; Acharya, P. *bioRxiv* **2020**.
85. Mehdipour, A.; Hummer, G. *bioRxiv* **2020**.
86. Clausen, T. M.; Sandoval, D. R.; Spliid, C. B.; Pihl, J.; Painter, C. D.; Thacker, B. E.; Glass, C. A.; Narayanan, A.; Majowicz, S. A.; Zhang, Y.; Torres, J. L.; Golden, G. J.; Porell, R.; Garretson, A. F.; Laubach, L.; Feldman, J.; Yin, X.; Pu, Y.; Hauser, B.; Caradonna, T. M.; Kellman, B. P.; Martino, C.; Gordts, P. L. S. M.; Leibel, S. L.; Chanda, S. K.; Schmidt, A. G.; Godula, K.; Jose, J.; Corbett, K. D.; Ward, A. B.; Carlin, A. F.; Esko, J. D. *bioRxiv* **2020**.
87. Kim, S. Y.; Jin, W.; Sood, A.; Montgomery, D. W.; Grant, O. C.; Fuster, M. M.; Fu, L.; Dordick, J. S.; Woods, R. J.; Zhang, F.; Linhardt, R. J. *Antiviral Res.* **2020**, 104873.
88. Kwon, P. S.; Oh, H.; Kwon, S. J.; Jin, W.; Zhang, F.; Fraser, K.; Hong, J. J.; Linhardt, R. J.; Dordick, J. S. *Cell Discov.* **2020**, *6*, 50.
89. Tandon, R.; Sharp, J. S.; Zhang, F.; Pomin, V. H.; Ashpole, N. M.; Mitra, D.; Jin, W.; Liu, H.; Sharma, P.; Linhardt, R. J. *J. Vir.* **2020**.
90. Mycroft-West, C.; Su, D.; Pagani, I.; Rudd, T.; Elli, S.; Guimond, S.; Miller, G.; Meneghetti, M.; Nader, H.; Li, Y.; Nunes, Q.; Procter, P.; Mancini, N.; Clementi, M.; Bisio, A.; Forsyth, N.; Turnbull, J.; Guerrini, M.; Fernig, D.; Vicenzi, E.; Yates, E.; Lima, M.; Skidmore, M. *bioRxiv* **2020**.
91. Boni, M. F. *Vaccine* **2008**, *26* (Suppl. 3), C8–14.
92. Carrat, F.; Flahault, A. *Vaccine* **2007**, *25* (39–40), 6852–6862.
93. Nelson, M. I.; Holmes, E. C. *Nat. Rev. Genet.* **2007**, *8* (3), 196–205.
94. Johnson, N. P.; Mueller, J. *Bull. Hist. Med.* **2002**, *76* (1), 105–115.
95. Ferguson, N. M.; Fraser, C.; Donnelly, C. A.; Ghani, A. C.; Anderson, R. M. *Science* **2004**, *304* (5673), 968–969.
96. Air, G. M. *Curr. Opin. Virol.* **2014**, *7*, 128–133.
97. Gamblin, S. J.; Skehel, J. J. *J. Biol. Chem.* **2010**, *285* (37), 28403–28409.
98. Parton, D. L.; Tek, A.; Baaden, M.; Sansom, M. S. *PLoS Comput. Biol.* **2013**, *9* (4)e1003034.
99. Durrant, J. D.; Kochanek, S. E.; Casalino, L.; Ieong, P. U.; Dommer, A. C.; Amaro, R. E. *ACS Cent. Sci.* **2020**, *6* (2), 189–196.
100. Huber, G. A.; McCammon, J. A. *Trends Chem.* **2019**, *1* (8), 727–738.
101. Amaro, R. E.; Ieong, P. U.; Huber, G.; Dommer, A.; Steven, A. C.; Bush, R. M.; Durrant, J. D.; Votapka, L. W. *ACS Cent. Sci.* **2018**, *4* (11), 1570–1577.
102. Taubenberger, J. K. *Proc. Natl. Acad. Sci. U. S. A.* **1998**, *95* (17), 9713–9715.
103. Das, D. K.; Govindan, R.; Nikić-Spiegel, I.; Krammer, F.; Lemke, E. A.; Munro, J. B. *Cell* **2018**, *174* (4), 926–937.e12.
104. Lee, P. S.; Ohshima, N.; Stanfield, R. L.; Yu, W.; Iba, Y.; Okuno, Y.; Kurosawa, Y.; Wilson, I. A. *Nat. Commun.* **2014**, *5*, 3614.
105. Stencel-Baerenwald, J. E.; Reiss, K.; Reiter, D. M.; Stehle, T.; Dermody, T. S. *Nat. Rev. Microbiol.* **2014**, *12* (11), 739–749.
106. Shinya, K.; Ebina, M.; Yamada, S.; Ono, M.; Kasai, N.; Kawaoka, Y. *Nature* **2006**, *440* (7083), 435–436.
107. Matrosovich, M. N.; Matrosovich, T. Y.; Gray, T.; Roberts, N. A.; Klenk, H. D. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101* (13), 4620–4624.
108. van Riel, D.; Munster, V. J.; de Wit, E.; Rimmelzwaan, G. F.; Fouchier, R. A.; Osterhaus, A. D.; Kuiken, T. *Science* **2006**, *312* (5772), 399.
109. Bhatt, S.; Holmes, E. C.; Pybus, O. G. *Mol. Biol. Evol.* **2011**, *28* (9), 2443–2451.
110. Wu, N. C.; Wilson, I. A. *J. Mol. Biol.* **2017**, *429* (17), 2694–2709.
111. Altman, M. O.; Angel, M.; Košík, I.; Trovão, N. S.; Zost, S. J.; Gibbs, J. S.; Casalino, L.; Amaro, R. E.; Hensley, S. E.; Nelson, M. I.; Yewdell, J. W. *MBio* **2019**, *10* (2).
112. Pentiah, K.; Lees, W. D.; Moss, D. S.; Shepherd, A. J. *Glycobiology* **2015**, *25* (1), 124–132.
113. Zhang, M.; Gaschen, B.; Blay, W.; Foley, B.; Haigwood, N.; Kuiken, C.; Korber, B. *Glycobiology* **2004**, *14* (12), 1229–1246.
114. Yokoyama, M.; Fujisaki, S.; Shirakura, M.; Watanabe, S.; Odagiri, T.; Ito, K.; Sato, H. *Front. Microbiol.* **2017**, *8*, 584.
115. Arrildt, K. T.; Joseph, S. B.; Swanstrom, R. *Curr. HIV/AIDS Rep.* **2012**, *9* (1), 52–63.
116. Deng, H.; Liu, R.; Ellmeier, W.; Choe, S.; Unutmaz, D.; Burkhart, M.; Di Marzio, P.; Marmon, S.; Sutton, R. E.; Hill, C. M.; Davis, C. B.; Peiper, S. C.; Schall, T. J.; Littman, D. R.; Landau, N. R. *Nature* **1996**, *381* (6584), 661–666.
117. Swanstrom, R.; Coffin, J. *Cold Spring Harb. Perspect. Med.* **2012**, *2* (12), a007443.
118. Julien, J. P.; Cupo, A.; Sok, D.; Stanfield, R. L.; Lyumkis, D.; Deller, M. C.; Klasse, P. J.; Burton, D. R.; Sanders, R. W.; Moore, J. P.; Ward, A. B.; Wilson, I. A. *Science* **2013**, *342* (6165), 1477–1483.
119. Behrens, A. J.; Harvey, D. J.; Milne, E.; Cupo, A.; Kumar, A.; Zitzmann, N.; Struwe, W. B.; Moore, J. P.; Crispin, M. *J. Virol.* **2017**, *91* (2).
120. Mizuochi, T.; Matthews, T. J.; Kato, M.; Hamako, J.; Titani, K.; Solomon, J.; Feizi, T. *J. Biol. Chem.* **1990**, *265* (15), 8519–8524.
121. Garces, F.; Lee, J. H.; de Val, N.; de la Pena, A. T.; Kong, L.; Puchades, C.; Hua, Y.; Stanfield, R. L.; Burton, D. R.; Moore, J. P.; Sanders, R. W.; Ward, A. B.; Wilson, I. A. *Immunity* **2015**, *43* (6), 1053–1063.
122. Lee, J. H.; de Val, N.; Lyumkis, D.; Ward, A. B. *Structure* **2015**, *23* (10), 1943–1951.
123. Hung, C. S.; Vander Heyden, N.; Ratner, L. *J. Virol.* **1999**, *73* (10), 8216–8226.
124. Parren, P. W.; Moore, J. P.; Burton, D. R.; Sattentau, Q. J. *AIDS* **1999**, *13* (Suppl. A), S137–S162.
125. Wei, X.; Decker, J.; Wang, S.; Hui, H.; Kappes, J.; Wu, X.; Salazar-Gonzalez, J.; Salazar, M.; Kilby, J.; Saag, M.; Komarova, N.; Nowak, M.; Hahn, B.; Kwong, P.; Shaw, G. *Nature* **2003**, *422* (6929), 307–312.
126. Lemmin, T.; Soto, C.; Stuckey, J.; Kwong, P. *Structure* **2017**, *25* (10), 1631.
127. Ferreira, R. C.; Grant, O. C.; Moyo, T.; Dorfman, J. R.; Woods, R. J.; Travers, S. A.; Wood, N. T. *Sci. Rep.* **2018**, *8* (1), 15031.
128. Wood, N. T.; Fadda, E.; Davis, R.; Grant, O. C.; Martin, J. C.; Woods, R. J.; Travers, S. A. *PLoS One* **2013**, *8* (11)e80301.