A perspective on the PDB's impact on the field of glycobiology

James H. Prestegard*

From the Complex Carbohydrate Research Center, University of Georgia, Athens, Georgia, USA

Edited by Wolfgang Peti

Structures deposited in the Protein Data Bank (PDB) facilitate our understanding of many biological processes including those that fall under the general category of glycobiology. However, structure-based studies of how glycans affect protein structure, how they are synthesized, and how they regulate other biological processes remain challenging. Despite the abundant presence of glycans on proteins and the dense layers of glycans that surround most of our cells, structures containing glycans are underrepresented in the PDB. There are sound reasons for this, including difficulties in producing proteins with well-defined glycosylation and the tendency of mobile and heterogeneous glycans to inhibit crystallization. Nevertheless, the structures we do find in the PDB, even some of the earliest deposited structures, have had an impact on our understanding of function. I highlight a few examples in this review and point to some promises for the future. Promises include new structures from methodologies, such as cryo-EM, that are less affected by the presence of glycans and experiment-aided computational methods that build on existing structures to provide insight into the many ways glycans affect biological function.

Glycobiology

Glycobiology is the study of how glycans, also called carbohydrates or oligosaccharides, result from, or have an impact on, a wide array of biological processes (1). Their impact is broad, with glycans occurring on the surfaces of most cells in addition to being excreted into the environment and incorporated into structural materials that constitute most of the biomass on earth (2). Many glycans are covalently attached to proteins, making them glycoproteins. It is estimated that, among all proteins in the Swiss-Prot database, between 20% and 50%, are glycosylated (3, 4); among eukaryotes (about one-third of the proteins in the database), the percentage is likely to be much higher (5, 6). In addition to glycoproteins, numerous other proteins bind glycans in the process of transforming them for a source of energy or for substrates needed to build other biological macromolecules, including the structural carbohydrates of plants and microbes. Some proteins also bind to glycans on cell surfaces to

initiate various signaling and adhesion events, and pathogens, both bacterial and viral, have proteins that bind to cell surface glycans as a step in the infection process. This adds to the biomedical importance of glycans and has raised the interest of the pharmaceutical industry in both glycoproteins and glycomimetic glycans (7).

Glycans are structurally very diverse, more so than polypeptides and nucleic acids of comparable size. Although the number of residue types (sugars) making up glycans is similar to the number of amino acid types making up proteins, residues can be linked in multiple ways and they can exist as branched and linear oligomers. Differences between some glycan residues seem subtle. For example, glucose (Glc), mannose (Man), and galactose (Gal) are all composed of six-membered pyranose rings with exactly the same composition, $C_6H_{12}O_6$, from which the name carbohydrate is derived. However, structural consequences associated with differences in chirality are significant. Also, unlike proteins and nucleic acids, glycan synthesis is not template driven but occurs *via* the combined action of hundreds of enzymes that add or remove individual residues. As a result, glycoproteins seldom carry a single type of glycan, making them heterogeneous even at the single site level. These properties have certainly impeded characterization of glycans and perhaps the interest of the broader scientific community in studying of how glycans influence biological processes.

Recently, some steps have been taken to make glycobiology more accessible to a broad audience and to attract a new generation of scientists who will tackle challenging glycobiology problems. One step is a text, now in its third edition, *Essentials of Glycobiology*, that begins with a useful historical review of the field (1). The authors and editors receive no financial benefit from sales of this book, and they have arranged to have electronic versions freely available through the NCBI Bookshelf (https://www.ncbi.nlm.nih.gov/books/NBK310274/). Another step is the adoption of a set of symbols for residues that make up glycans. Chemical structures are, of course, important to understanding glycan interactions with proteins and other glycoconjugates, but in many cases replacing atomic depictions of sugar residues with simple symbols is sufficient to show differences in glycan structures and place glycans in the larger biological contest where they function. Symbol Nomenclature for Glycans (SNFG) provides different symbols for each type of sugar (8). Software

* For correspondence: James H. Prestegard, jpresteg@ccrc.uga.edu.

James H. Prestegard, an eminent scholar of NMR Spectroscopy and professor of Chemistry and Biochemistry, Emeritus, at the Complex Carbohydrate Research Center, University of Georgia, is a leading researcher in structural biology who has contributed many innovations to nuclear magnetic resonance methods. He has focused much of his later career on the structural biology of carbohydrates and glycoproteins.

facilitating the depiction of these symbols in three dimensions has also been devised (9).

Figure 1A illustrates the use of the 3D symbols (http://glycam.org/3d-snfg) to depict a possible spatial arrangement of glycans in a particularly heavily glycosylated protein, the 11-kDa *N*-terminal domain of human carcinoembryonic antigen–related cell adhesion molecule 1. For comparison, the same structure is depicted in Figure 1B, showing the atomic detail of glycans. The domain has three *N*-glycosylation sites in which glycans are linked to the side-chain nitrogen of an asparagine in a consensus amino acid sequence, NXS/T. Three of the glycans frequently found on this domain are depicted, one high Man glycan composed of two GlcNAc residues (blue squares) and five Man residues (green circles) and two complex glycans that have additional Gal residues (yellow circles), fucose residues (red triangles), and a particular type of sialic acid residue, *N*-acetylneuraminic acid (Neu5Ac, purple diamonds). These glycans have been modeled onto the crystal structure of a nonglycosylated version of the protein (Protein Data Bank [PDB] ID: 4QXW) using computational modeling (10). As with other figures in this article, this figure has been made with readily available visualization tools (11, 12). The spatial extension of the glycans is well represented in both depictions, but residue types are more easily recognized using SNFG symbols. It is striking how much space glycans can occupy compared with the underlying protein in this example. In many cases, SNFG symbols now appear in the 3D views supplied when first accessing a glycoprotein deposition in the PDB.

Structural glycobiology and the PDB

Most of the structural biology community may well view glycobiology as a recent entry into the world of structure-based investigation. Nothing could be further from the truth. Emil Fischer, a 1902 recipient of the Nobel prize in Chemistry, developed the Fisher projection as means of describing the chirality of successive carbons in the simple sugars that make up glycans. He also introduced the lock and key model of enzyme action while contemplating how enzymes ferment glucosides, in which the simple sugar, Glc, is linked to other sugars or aglycons (13). Interactions with the PDB also started surprisingly early; 10 of the first 51 structures deposited in the PDB (starting in November of 1977) came from fiber diffraction studies of polysaccharides. Since that time, with

appropriate emphasis on macromolecular structure, proteins and nucleic acids have dominated depositions. However, glycan structures persist as ligands in the binding sites of proteins and as the covalent adducts that turn proteins into glycoproteins. The percentage of protein depositions in the PDB annotated with the structure description "glycoprotein" has risen steadily from 0.2% in the 1990 to 2000 period to 0.5% in the 2000 to 2010 period and to 0.9% in the 2010 to 2020 period. The percentage annotated with "saccharide" as a chemical component type (mostly bound ligands) has remained fairly constant between 1.2 and 1.5%.

Some of the difference between deposition statistics and potential impact may reside in the difficulty in retrospectively identifying structures that contain glycans. For years, naming of glycan residues and the use of keywords to identify the glycan content were left entirely up to authors of the depositions. Much of this has now been corrected through the use of improved deposition and remediation tools for carbohydrates (https://www.wwPDB.org/documentation/carbohydrate-remediation). Also, there are now software search tools that use either atom connectivity data (14) or actual structural footprints https://dev.glycam.org/portal/gf_home/ (15) to find relevant structures in the PDB. A recent application of one of these tools suggests that the percentage of depositions containing a carbohydrate moiety is actually 7.7% (14).

There are some real difficulties in producing structures of glycoproteins. For X-ray crystallography, producing suitable crystals can be problematic. Native glycosylation is very heterogeneous; attached glycans are very mobile, and they can actually dominate a glycoprotein structure as seen in Figure 1. The heterogeneity, mobility, and bulk of glycans can all
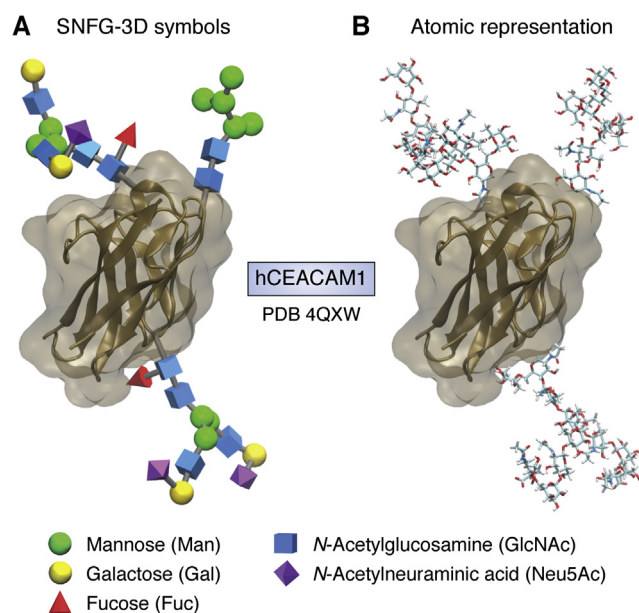

Figure 1. N-terminal domain of hCEACAM1 (4QXW). N-glycans are shown in SNFG-3D symbols (*A*) and full atomic representations (*B*). Spatial occupancies of protein and glycans are equally well represented in both depictions. SNFG, Symbol Nomenclature for Glycans.

contribute to a failure to crystallize (16). Fortunately, methods for engineering glycans have advanced, and producing glycoproteins with less extended and more homogeneous glycosylation now exist (17, 18). For NMR, most structural studies depend on uniform isotope labeling with ^{15}N and ^{13}C. This is economical for proteins that can be expressed in bacterial cultures which synthesize all amino acids from simple substrates (*e.g.*, $^{15}NH_4Cl$ and ^{13}C-Glc), but for glycoproteins expressed in mammalian cells, which produce near-native glycosylation, this can be very expensive, and perdeuteration needed to work on larger proteins is usually not possible. Fortunately, there are advances that make expression in mammalian cells less costly (19), and there are resonance assignment strategies that do not depend on uniform isotope labeling (20). NMR also excels when bound ligands are of interest. Glycan ligands tend to have relatively low affinities, something that is often compensated by multivalency in biological contexts. For crystallography, this leaves sites in crystals unoccupied, but weak binding ligands are not a problem for NMR-based investigations.

Importantly, cryo-EM structures, which have come on the scene more recently, do not require crystallization and provide a possible route to increased numbers of structures of glycoproteins, particularly large ones (21). For the last 6 months, nearly 15% of the deposited electron microscopy structures have a structure description of glycoprotein. With this new technology, and despite the challenges to producing structures with X-ray and NMR methods, there are now many structures deposited in the PDB that have had an impact on our understanding of how glycans influence biological function. In this review, I will touch on a few of these examples.

Erythropoietin—a glycosylated recombinant pharmaceutical

Erythropoietin (EPO) is a heavily glycosylated human protein hormone responsible for inducing differentiation of bone marrow erythroid progenitor cells to form red blood cells. The native protein has three *N*-glycosylation sites, easily identified by their conserved consensus sequence (NXS/T), and one O-glycosylation site. These are all at least partially occupied with a heterogeneous array of glycans. EPO was the first biopharmaceutical produced as a recombinant protein in mammalian cells, being approved by the FDA for the treatment of anemia in 1989. It was also the first glycosylated recombinant protein to surpass annual sales of $1 billion (22). The importance of glycosylation was determined early in its development (23), and efforts immediately began to alter glycosylation to improve activity and extend lifetime in the blood stream. The presence of complex glycans, particularly those terminated with sialic acid, was found to do both, extending lifetime and activity by almost an order of magnitude.

Structure certainly played a role in this development. There are only two structures of EPO in the PDB, both deposited relatively early in the history of the PDB. One was determined by NMR methods (1BUY, deposited in 1998) (24) and one by X-ray crystallography, in which EPO is complexed with its dimeric receptor (1EER, also deposited in 1998) (25). Interestingly, neither of these contains coordinates for glycans. All three of the potentially glycosylated asparagines in EPO were replaced with lysine to produce a soluble analog that could be expressed in a nonglycosylating *Escherichia coli* culture. Although the receptor was expressed in a potentially glycosylating yeast culture, the asparagine in its single *N*-glycosylation site was mutated to glutamine to minimize interference with crystal formation. The structures provided insight into how binding may stimulate receptor signaling. Both receptor and EPO appear to alter structure on complex formation.

Despite the lack of glycans in these structures, they provided a basis for understanding the potential role of the glycans and how additional glycans might improve efficacy of EPO. Figure 2 shows how Elliott *et al.* (23) modeled in native glycans to identify a region in the protein sequence where an additional *N*-glycosylation sequon could be added without inhibiting the receptor interaction but potentially increasing the protein's stability and resistance to clearance from the blood stream. Since that time, these same structures have been used repeatedly to facilitate the design of new glycosylated species and rationalize their effects on efficacy of the drug. Much effort has gone into producing constructs with homogeneous glycosylation using purely chemical methods (26, 27) and enzyme facilitated methods (28). Glycans clearly stabilize the structure, improve solubility, and particularly with glycans terminated in sialic acid, prolong residence in the blood stream. The latter presumably occurs because of reduced clearance through the asialoglycoprotein receptor. A certain level of intrinsic disorder in the binding elements of EPO also appears to be important to its ability to interact with other receptors and elicit other physiological responses (22). Glycosylation may well play a role in regulating structures of these regions as well.

Modeling glycans into proteins in which glycosylation sites have been removed by mutation, or into proteins where there is simply insufficient electron density to place glycan residues experimentally, continues today. Recent examples of addition of glycans by modeling include the spike proteins of the coronavirus SARS-CoV-2, where heavy glycosylation obscures sites that might have been used for antibody development (29, 30).

Glycans and antibody function

Although glycans on the surface of glycoproteins, being flexible and heterogenous, often work against crystal formation, there are exceptions. The dimer formed by the fragment crystallizable (*Fc*) region of human immunoglobulin G (IgG) antibodies is one. The *Fc* region found at the C-terminus of the heavy chain of IgG antibodies has a single *N*-linked glycan, which proves to be reasonably homogeneous when isolated from pooled human serum. The dominant glycans are core-fucosylated biantennary structures with zero, one, or two Gal residues terminating the branches (31). In 1976, Huber *et al.* (32) reported a crystal structure of *Fc*; it was refined and deposited in the PDB in 1981 (1FC1) (33). This is possibly the
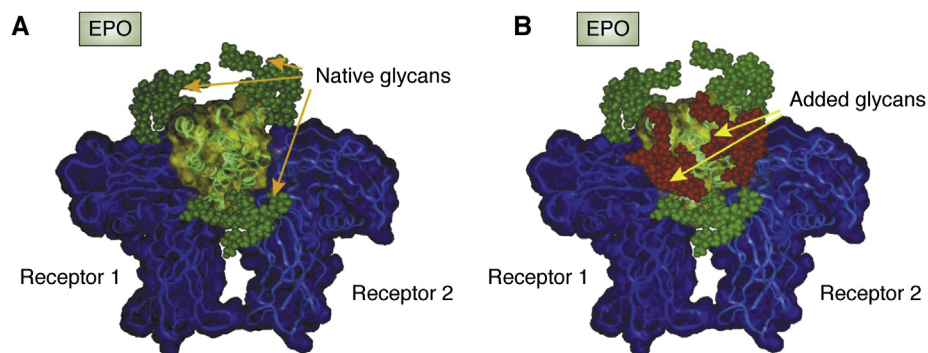
Figure 2. Comparison of EPO with native glycans and added glycans. The crystal structure is of the nonglycosylated extracellular domain of the EPO receptor bound to an EPO analog (1EER). A structure of a tetra-antennary glycan terminated with sialic acids was produced by molecular modeling and attached to the original sites (*A*) or new N-linked glycosylation sites (*B*) on the EPO structure. Structures are EPO (*yellow/green*), EPO receptors (*blue*), original three N-linked glycans (*green*), and new glycans (*red*). Reproduced with permission, Nature Biotechnology, 21:417, 2003. EPO, erythropoietin.

first glycoprotein deposited in the PDB. Many structures of this fragment have been deposited since. Enzymatically engineering glycans has improved resolution substantially, but the general placement of glycans has remained the same. Depicted in Figure 3 is a more recent structure (4KU1 (34)). The glycans are inside a cavity formed between the two *F*c chains of the dimer and are closely packed against the protein surface. Immobilization of the glycans in this way clearly contributes to the quality of the structure.

The glycans are linked to asparagine 297 near the top of the structure as depicted in SNFG-3D symbols. There is a mobile disulfide linkage forming a hinge between monomers near the top (not resolved in this crystal structure). The rest of the heavy chains extend out from this region to interact with
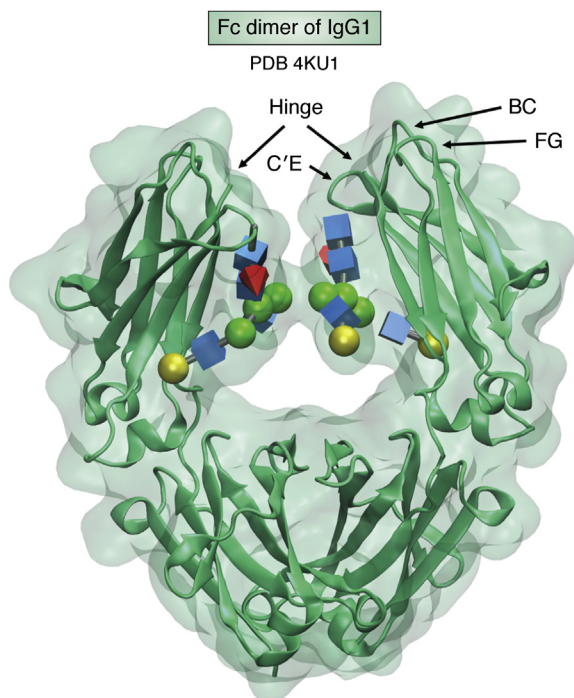

Figure 3. Fc dimer of IgG1 depicted using PDB deposition 4KU1. Glycans are shown in 3D-SNFG symbols. The Gal terminating the 6 branch of one monomer is shown as a *yellow sphere* at the center left, and the N-acetylglucosamine (GlcNAc) terminating the 3 branch of the other monomer is shown as a *blue cube* near the center of the cavity. PDB, Protein Data Bank; SNFG, Symbol Nomenclature for Glycans.

another pair of chains to form the Fab domains responsible for antigen binding. The hinge region, along with BC, C'E, and FG loops, is also the area involved in binding the Fc gamma receptors that turn antigen binding into a physiological response.

Given the proximity of glycans to the hinge region, it is not surprising that early work found that producing IgG antibodies without glycosylation resulted in loss of an ability to activate complement and induce cellular toxicity (35) but did not affect antigen binding or binding of protein A (a bacterial protein that recognizes structural elements of *F*c near the outer midpoint of the depicted structure). In subsequent work, it became clear that even small changes in glycan structure had substantial physiological effects. Absence of terminal Gal residues is associated with inflammation, and the addition of a sialic acid to the terminal Gal residues produces an anti-inflammatory molecule that appears to be the active component in intravenous IgG treatments for severe cases of rheumatoid arthritis (36). The exclusive use of a particular sialic acid to terminate IgG Gal residues, Neu5Ac, is characteristically human. Other mammals, including other primates, use both Neu5Ac and *N*-glycolylneuraminic acid. Specific recognition of Neu5Ac is important for glycoproteins implicated in diseases outside those associated with immune response (37), and an understanding of the role of sialic acids in immune response may contribute to understanding those diseases as well.

The mechanism by which changes in physiological function of IgGs occurs is still an important target of investigation (38). Chimeric constructs having an *F*c component, as well as the antibodies themselves, now dominate sales of the pharmaceutical industry, and having a rational approach to altering glycan composition could have an enormous impact. Structures of *F*c having engineered complements of glycans, or in complex with various Fc gamma receptors, are providing snapshots of interactions and structural changes in the hinge region that corelate with glycan substitutions. There are now more than 100 such depositions in the PDB. There is clearly considerable variation in the conformation of the receptor binding region. However, the picture may be more complex than static snapshots can provide. Solution-based methods such as NMR and SAXS suggest that the region may dynamically sample these conformations (39). Also, the

glycans themselves do not stay in the position depicted. Instead, they transiently sample extended conformers where enzymatic modification of glycans can occur (40). Ultimately, shifts in distributions in response to changes in glycan composition may be a better description of how glycans affect IgG function.

Glycans as ligands in protein structures

Far more abundant among structures deposited in the PDB, than glycans covalently attached to glycoproteins, are glycans as ligands of proteins that bind glycans, modify glycans, or add glycans to other proteins. The first example of an enzyme with a bound substrate was hen egg white lysozyme, deposited in 1979 as one of the first 65 structures in the PDB. Lysozyme is a hydrolytic enzyme with homologues produced by many animals as an antibacterial agent. The 1979 structure (9LYZ) was a complex with a trisaccharide, containing *N*-acetylmuramic acid (MurNAc) and GlcNAc, MurNAcβ1-4GlcNAcβ1-4MurNAc. Mechanistically, there had been much speculation about capturing a transition state with a distorted MurNAc pyranose ring. The crystal structure failed to show this distortion (41). Many structures of lysozymes followed (now nearly 1000 structures from various species). However, direct experimental evidence for a distorted transition state remains elusive (42). Resolution of detailed mechanistic questions may fall to computational work that builds on these many high-quality crystal structures (43).

In general, enzymes that build or degrade glycans represent a major class of protein structures deposited in the PDB. There is, in fact, a separate database, the Carbohydrate-Active enZYmes (CAZy) database (http://www.cazy.org), that assembles domains of these proteins into families. This was originally based on sequence and evolutionary relationships; however, it has evolved to incorporate biochemical data and structural data from the PDB. Early on, it provided a resource for finding enzymes, systematically naming enzymes, and displaying bound glycan ligands. It is now apparent that families are as much structurally related as they are sequence related (44). There are now nearly a million modules classified into families (a single modular protein can be classified into many families), and there are nearly 10,000 PDB IDs associated with these modules. While the CAZy database does classify noncatalytic, glycan-binding modules, there are also databases devoted specifically to proteins containing only these modules, namely lectins. The UniLectin3D database (https://www.unilectin.eu/unilectin3D) classifies more than 2000 PDB structures into 35 families that share protein folds (45). These databases have become important tools for understanding how structure relates to function. Both are highly dependent on the PDB, with CAZy linking its entries to the PDB and UniLectin3D actually classifying entries based on structural folds.

It seems appropriate to highlight one recent example of structures providing insight into the function of a carbohydrate active enzyme (actually a pair of enzymes). The pair is the *O*-GlcNAc transferase (OGT) and the *O*-GlcNAc hydrolase (OGA) responsible for respectively adding and removing a single GlcNAc residue at serine and threonine sites that are often phosphorylated in higher organisms. This unusual addition of a single sugar with no further extension was not discovered until 1984 because of the size, lack of charge, and lability of the addition (46). However, it proves very ubiquitous and functionally important to a host of processes including nutrient sensing, response to stress, cell division, and transcription (47). Moreover, it dispelled the common assumption that glycosylation was a process relegated to extracellular proteins and proteins involved in, or destined for, the secretory pathway. The modification occurs in both the cytosol and the nucleus. Its competition with phosphorylation has suggested that it is a master regulator of many signaling processes, but in contrast to the hundreds of kinases and phosphatases involved in phosphorylation, there is just a single pair of glycosyltransferase and glycosyl hydrolase enzymes. This raises many questions about how these enzymes find their many targets and to what extent modifications are selective. The structure of these two enzymes provides a starting point for answers to these questions.

The structure of OGT is depicted in Figure 4A (48). The catalytic domain (in blue and cornflower blue) may look similar to a number of other glycosyltransferases. It belongs to the family of GT-B glycosyltransferases and consists of a pair of Rossman folds forming a binding site for the sugar donor, UDP–GlcNAc. It is, however, unusual in that it has a second activity, a proteolytic cleavage of HCF1 to produce a mature cell cycle coregulator (49). Also, there is a highly negatively charged domain between the two Rossman folds (green) that may suggest interactions with nucleic acids, and the *N*-terminal domain (cyan) is a series of tetratricopeptide repeats, domains that are well-known for their protein–protein interactions.

Recently, two groups have produced structures of human OGA (50, 51). The one depicted in Figure 4B (5UN8) crystallized as a dimer with an O-GlcNAc-containing peptide from p53. The catalytic domain of one monomer is shown in cyan along with its stalk domain in forest green. The stalk domain of the other monomer is shown in green. Interestingly, both the catalytic domain of one monomer and the stalk domain of the other monomer make contact with the substrate. The extended surface contacts suggest a mechanism whereby activities toward certain glycosylated peptides may be enhanced. These depictions of surface contacts are proving useful in the design of a number of OGA inhibitors (52). Although there are structures of both OGT and OGA with peptides from potential substrates bound, a full understanding of target selection may await structures of complexes containing the actual protein substrates or complementary data from techniques that can identify regions involving more transient protein–protein contacts.

Large protein assemblies from cryo-EM

Although the structural detail offered by high-resolution X-ray structures of individual proteins remains essential for mechanistic studies and inhibitor design, other challenges
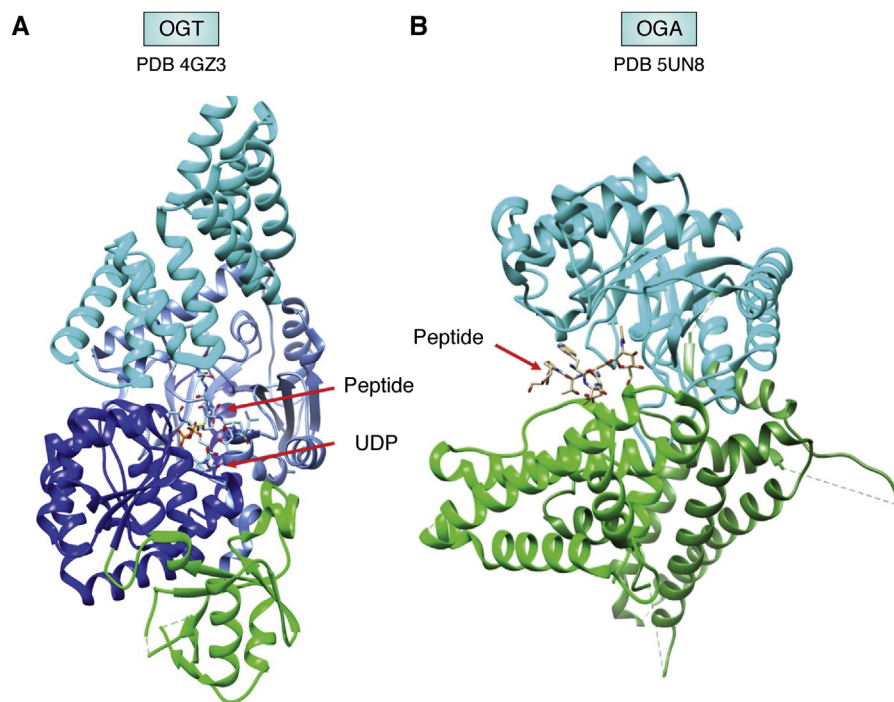
Figure 4. Structures of O-GlcNAc enzymes. *A,* OGT (4GZ3) catalytic units are *blue* and *cornflower blue* with the donor product (UDP) and acceptor peptide shown in *stick figures.* The positively charged domain inserted between two parts of the catalytic domain is in *green* and the tetra-tricorepeat domain is in *cyan. B,* an OGA structure (5UN8) is shown with the O-GlcNAc modified peptide from p53 inserted into the catalytic domain of one monomer of the dimeric structure (*cyan*). The peptide makes additional contacts with the stalk region of the second monomer (*green,* as opposed to *forest green,* for the first monomer). OGT, *O*-GlcNAc transferase.

remain that require new technology. Advances in cryo-EM technology allowing production of large structures with sub 3Å resolution is revolutionizing many areas of structural biology, including structural glycobiology. Without the requirement for crystallization, the mobility and heterogeneity of glycosylation imposes few impediments. In many cases, sufficient electron density is not seen for glycans beyond the first glycan attached, but protein structures can be produced and glycans added computationally. Also, because the size is more of an advantage than disadvantage, many of these structures are of multiprotein complexes, including the now numerous structures of whole virus particles, all of which are heavily glycosylated.

The impact of cryo-EM on glycoscience is, however, best illustrated with a nonviral example of multiprotein assemblies, namely the oligosaccharide transferases (OSTs) which are embedded in the endoplasmic reticulum (ER). The *N*-glycans, that play important roles in protein folding and extracellular recognition events, are not added stepwise starting with the first GlcNAc attached to the asparagine side chain of the NXS/T sequons. Instead, they are first transferred by the action of OST enzymes as a large oligosaccharide (Glc$_3$Man$_9$GlcNAc$_2$) from a dolichol pyrophosphate donor anchored to the luminal side of the ER. From there, they are trimmed and modified before transfer to the cisternae of the Golgi apparatus for further modification and eventual secretion. The study of complexes embedded in membranes, whether membrane isolates (for EM tomography) or membrane mimetics (for single-particle EM), is an area where cryo-EM approaches excel (53, 54). One of the earliest medium-resolution cryo-EM articles on an OST focused on interactions with the ribosome. It documented

the interaction of an OST with the translocon, another multiproton complex, that includes Sec61 which provides a tunnel through the ER (55).

N-glycan sequons are not glycosylated to an equal extent, and the extent clearly varies with the context in which the sequons are found. Part of the challenge in understanding the origin of variations in glycosyl addition arises from mammals having two OSTs, one that operates cotranslationally (OST-A) and one that operates post-translationally (OST-B). Both OSTs are multiprotein complexes, having catalytic units (STT3A and STT3B) that differ slightly in sequence. These units are complexed with 5 other proteins that they share, but OST-A has an additional protein that appears essential for interaction with the ribosome, DC2, and OST-B has one of two oxidoreductases that facilitate post-translation glycosylation, MAGT1 or TUSC3. The entire assembly of each is embedded in the ER membrane with a cluster of transmembrane alpha helices. Recently, high-resolution structures (3.5 Å) of both OST-A and OST-B were obtained. They provide interesting details explaining why more specific associations with the translocon occur for OST-A than for OST-B (56).

The two human OST assemblies, OST-A and OST-B, are depicted in Figure 5, *A* and *B*, respectively. The respective catalytic units are depicted in green. In both cases, a dolichol phosphate molecule is present (red spheres) identifying the active site. One apparent difference is that the OST-A structure shows a four-helix bundle depicted in cyan at the lower left that extends into the cytosol. This is the C-terminus of ribophorin-I, a protein that is known to interact with ribosomes. This segment is present but disordered and unobservable in OST-B, presumably because of different interactions with STT3A and STT3B. Colored blue in OST-A
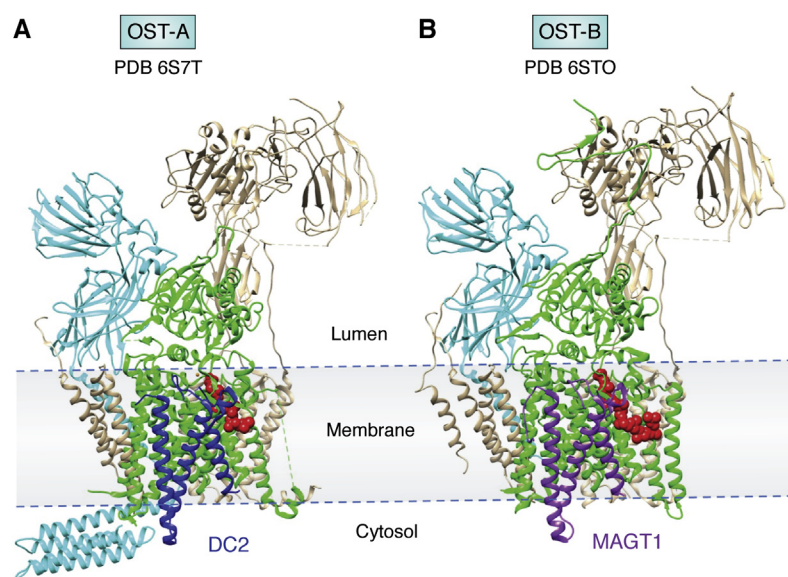
Figure 5. Cryo-EM structures for oligosaccharyltransferase enzymes. *A*, OST-A (6S7T) is the cotranslational enzyme containing DC2 (*blue*). *B*, OST-B (6S7O) is the post-translational enzyme containing MAGT1 (*purple*). Both enzymes have catalytic units (STT3) shown in *green* and ribophorin-I shown in *cyan*. Both have a dolichol phosphate reaction product shown in *red spheres*. OST, oligosaccharide transferase.

and purple in OST-B are the transmembrane helices of DC2 and MAGT1 proteins. Despite substantial sequence variations, these helices sit in homologous sites in the OSTs. In OST-A, DC2 would facilitate interaction with the ribosome and the cotranslational addition of an *N*-glycan. In OST-B, the segment of MAGT1 shown has a substantially different sequence and likely inhibits interaction with the ribosome. Instead, the MAGT1 protein has the capacity to disrupt and reform disulfides, something that may be needed in post-translational *N*-glycan additions. Hence, there are clear structural differences that explain why the two OSTs function cotranslationally *versus* post-translationally.

However, there is also much work to be done. There are, for example, several domains in the OST-B and OST-A structures that lack sufficient electron density for structural determination. One is the large *N*-terminal catalytic domain of MAGT1. This would sit on the luminal side of the ER near the dolichol phosphate. There is a crystal structure of a closely homologous domain from the TUSC3 protein (4M91, 72% identity) (57). Positioning it in the structure may clarify how post-translational glycan addition proceeds and whether extensive unfolding is required. Positioning this may require cooperation of scientists using a host of other structural technologies.

PDB structures leveraged by active site modeling

Despite the tremendous advances in structure determination methodology, there are aspects of glycosylated and glycan processing systems that cannot be captured in experimental structures. The dynamic motions of these systems, the transient binding of ligands and the complete structures of systems that cannot be produced in the amounts or levels of purity required for experimental determination, are a few examples. In these cases, PDB structures provide starting points for building additional structures by homology modeling, for displaying motion in molecular dynamics simulations, and for

ligand docking using purely computational methods or computational methods reinforced by limited amounts of experimental data.

One common source of complementary data is NMR. Of course, NMR can provide complete *de novo* structures. In fact, more than 13,000 structures produced by NMR methods have been added to the PDB, many as the result of the Protein Structure Initiative (58). Some of the NMR structures are actually glycoproteins that contain glycans and require these glycans for stability; the first deposited glycoprotein structure by NMR was one of the human CD2 adhesion protein in 1993 (59). More recently, solid-state NMR has come on the scene, providing access to insoluble aggregates of proteins and other amorphous material, which often contain glycans (60). However, because glycoproteins usually require expression in organisms other than *E. coli*, the uniform labeling with ^{13}C,^{15}N, and ^{2}H commonly used for complete structure determination by NMR is often considered impractical, and applications to large glycosylated systems have been slow in coming.

Applications of NMR to locate binding sites for glycan ligands on proteins are more numerous. These applications use very basic experiments where the positions of cross-peaks in two dimensional plots are perturbed by ligand addition (chemical shift perturbation), providing binding-site locations. Resonances from ligands can also suffer intensity changes on irradiation of protein protons with Rf identifying protein-binding epitopes (saturation transfer difference experiments), and cross-peaks connecting one ligand resonance to another in the bound state can be selectively detected, providing bound ligand conformations (transfer nuclear Overhauser effect experiments). More recently, paramagnetic perturbations are being used to locate ligand-binding sites and determine glycan conformations. Although there are some notable cases of structure deposition using the above methods (61), and even some exploiting related methods of solid-state NMR (62), there are many more structures that are better described in
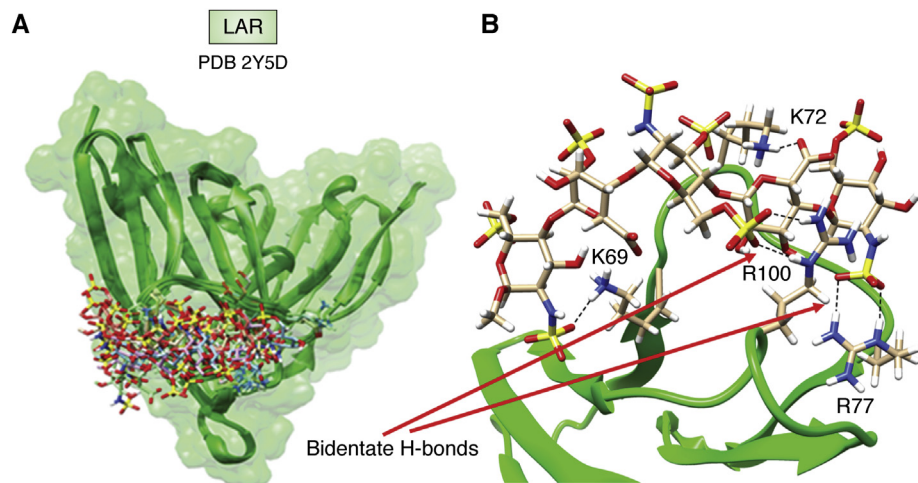
Figure 6. Docked structures of LAR using PDB deposition 2Y5D. *A*, the cluster of docked HS structures. *B*, the snapshot from molecular dynamics refinement showing the bidentate hydrogen bond between R77 and N-sulfate oxygens of the terminal GlcNS residue, as well as that between R100 and O6 sulfate oxygens of the central GlcNS6S residue (*dotted lines*, indicated with *arrows*). LAR, leukocyte common antigen-related protein; PDB, Protein Data Bank.

words, pictures, and molecular dynamics movies. The latter cases are no less important and rely no less on deposited structures of the underlying proteins.

One example drawn from my laboratory's own work involves a short piece of heparan sulfate (HS), actually the commercial anticoagulant, fondaparinux (GlcNS6S-GlcA-GlcNS3,6S-IdoA2S-GlcNS6S-OMe), bound to the terminal two domains of the leukocyte common antigen-related protein (LAR) (63). This system is of considerable interest because of its involvement in regulation of axon outgrowth and nerve regeneration after injury (64). Among the factors involved, is LAR's competitive interaction with chondroitin sulfate and HS components of the glycocalyx that surrounds most human cells (65). There are more than a dozen crystal

structures containing the two *N*-terminal domains of LAR, including one with sucrose octasulfate as an HS mimic (2YD8). While sucrose octasulfate and HS both have extensively sulfated sugar residues, the degree to which sucrose octasulfate mimics a native HS ligand is clearly in question. To produce a better model, we undertook an NMR study that included chemical shift perturbation, saturation transfer difference, and transfer nuclear Overhauser effect data. These data were used in a docking study that began with the unliganded crystal structure, 2YD5. The docking software, HADDOCK (66), was used to produce the cluster of structures shown in Figure 6A. HADDOCK can use a variety of experimental data as constraints, including those mentioned above, and it has become one of several platforms for
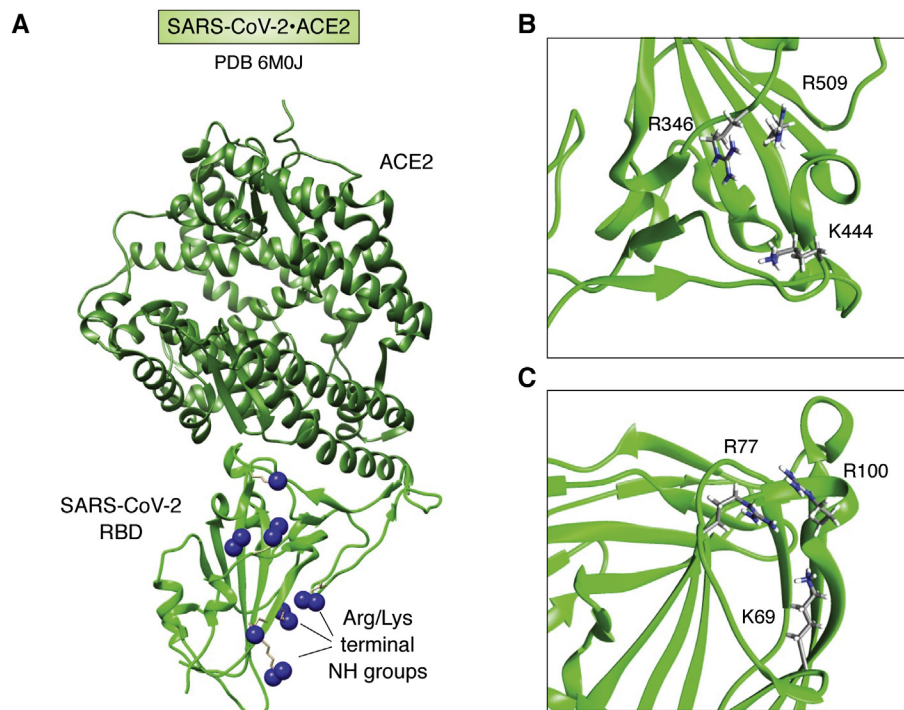

Figure 7. HS-binding clusters in SARS-CoV-2 RBD and LAR. *A*, the complex of SARS-CoV2 RBD (*green*) with ACE2 (*forest green*) (6M0J). *Blue spheres* are terminal NH groups of arginines and lysines. *B*, expansion of SARS-Cov2-RBD showing charged clusters: R509, R346, and K444. *C*, expansion of LAR (2YD5) showing charged clusters R100, R77, and K69. HS, heparan sulfate; LAR, leukocyte common antigen-related protein; RBD, receptor-binding domain.

integrating various types of data in the structural determination of large complexes (67).

Subsequent to publication, a low-energy structure that satisfied all experimental data was used as a starting point for a long (1 μs) molecular dynamics simulation. This allowed examination of the many structures sampled near this low-energy structure. A snapshot showing a frequent interaction between the negatively charged *N*-sulfate on the GlcNS residue at the nonreducing terminus of fondaparinux and the positively charged side chain of Arg 77 is shown in Figure 6*B*. This has a bidentate hydrogen bond involving 2 sulfate oxygens and 2 N-H groups of the arginine. A similar interaction occurs transiently between Arg 100 and the O-6 sulfate of the interior GlcNS. Additional stabilization comes from lysine side-chain interactions with other negatively charged groups of the ligand. Of course, there are many documented cases of lysine- and arginine-rich pockets binding negatively charged glycosaminoglycans such as chondroitin sulfate and HS. However, the structural details are intriguing, suggesting a search for structurally similar interactions that might occur in other systems.

Another example of how deposited structures, combined with computational docking of ligands, can provide important biological insight comes from recent work on the SARS-CoV-2 virus. The angiotensin-converting enzyme-2 is currently regarded as the cell-surface receptor for the virus. In fact, there is a crystal structure of the receptor-binding domain of the viral spike in complex with angiotensin-converting enzyme-2 (6M0J). However, Clausen *et al.* (68) noticed a cluster of positively charged residues near the interface of these two proteins that suggested a possible HS binding site and proceeded with docking studies to verify the interaction. These were followed by an extensive set of cell-based studies to convincingly document cell-surface HS as a required coreceptor for the virus. Thus, docking may suggest new molecular targets for disruption of virus–host cell interactions.

The region identified on the receptor-binding domain and the cluster of positively charged residues are depicted in Figure 7*A*. Figure 7, *B* and *C* present an interesting comparison of a subset of these residues to those implicated in the study of LAR binding to HS. Despite the very different sequence distribution of residues (K69, R77, R100 for LAR and K444, R346, and R509 for SARS-CoV-2), the triad of a lysine residue and two arginine residues have inter-residue separations of β-carbons that differ by less than 1.5 Å. Observations like this suggest that mining the PDB for structural details of binding sites could impact our ability to anticipate ligand-mediated cell–cell or cell–pathogen interactions in other systems. The future will tell if such suggestions have merit.

Conclusion

Glycobiology has clearly benefitted from decades of structural deposition in the PDB. However, obstacles associated with preparation of samples having homogeneous glycans, as well as recognizing glycans when they are present in a deposited structure, may have diminished impact both within

and outside the glycobiology community. Many of these obstacles have now been overcome. Annotation of deposited structures is much improved, there are new glycan-specific tools for searching the PDB, and symbolic representation of glycans is facilitating depiction of protein–glycan interactions in complex systems. New structural technologies, such as cryo-EM and solid-state NMR, which do not require crystallization, promise to provide many new glycan-containing structures. However, the primary goal of most glycobiologists, and structural biologists in general, remains a functional understanding. Structures of glycoproteins and glycan-binding proteins can provide a basis for this understanding, but integration of data relating to dynamics, energetics, and transient interactions will be required. Continued efforts to facilitate access to deposited structures by the broad range of scientists who integrate these data will be key to achieving this goal.

Author contributions—J. H. P. is responsible for all aspects of this article.

Conflict of interest—The author declares that he has no conflicts of interest with the contents of this article.

Abbreviations—The abbreviations used are: CAZy, Carbohydrate-Active enZYmes; EPO, erythropoietin; ER, endoplasmic reticulum; Gal, galactose; Glc, glucose; HS, heparan sulfate; IgG, immunoglobulin G; LAR, leukocyte common antigen-related protein; Man, mannose; Neu5Ac, *N*-acetylneuraminic acid; OGA, O-GlcNAc hydrolase; OGT, O-GlcNAc transferase; OST, oligosaccharide transferase; PDB, Protein Data Bank; SNFG, Symbol Nomenclature for Glycans.

References

1. Varki, A., and Kornfeld, S. (2017) Historical background and overview. In: Varki, A., Cummings, R., Esko, J., Stanley, P., Hart, G., Aebi, M., Darvill, A., Kinoshita, T., Packer, N., Prestegard, J., Schnaar, R., Seeberger, P., eds. *Essentials of Glycobiology*, 3rd Ed, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY: 387–400
2. Williams, C. L., Westover, T. L., Emerson, R. M., Tumuluru, J. S., and Li, C. L. (2016) Sources of biomass feedstock variability and the potential impact on biofuels production. *Bioenerg. Res.* **9**, 1–14
3. Apweiler, R., Hermjakob, H., and Sharon, N. (1999) On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database. *Biochim. Biophys. Acta* **1473**, 4–8
4. Spiro, R. G. (2002) Protein glycosylation: Nature, distribution, enzymatic formation, and disease implications of glycopeptide bonds. *Glycobiology* **12**, 43R–56R
5. Cummings, R. D., and Pierce, J. M. (2014) The challenge and promise of glycomics. *Chem. Biol.* **21**, 1–15
6. Steentoft, C., Vakhrushev, S. Y., Joshi, H. J., Kong, Y., Vester-Christensen, M. B., Schjoldager, K., Lavrsen, K., Dabelsteen, S., Pedersen, N. B.,

Marcos-Silva, L., Gupta, R., Bennett, E. P., Mandel, U., Brunak, S., Wandall, H. H., *et al.* (2013) Precision mapping of the human O-GalNAc glycoproteome through SimpleCell technology. *EMBO J.* **32**, 1478–1488

7. Valverde, P., Arda, A., Reichardt, N. C., Jimenez-Barbero, J., and Gimeno, A. (2019) Glycans in drug discovery. *MedChemComm* **10**, 1678–1691

8. Neelamegham, S., Aoki-Kinoshita, K., Bolton, E., Frank, M., Lisacek, F., Lutteke, T., O'Boyle, N., Packer, N. H., Stanley, P., Toukach, P., Varki, A., Woods, R. J., Darvill, A., Dell, A., Henrissat, B., *et al.* (2019) Updates to the symbol nomenclature for glycans guidelines. *Glycobiology* **29**, 620–624

9. Thieker, D. F., Hadden, J. A., Schulten, K., and Woods, R. J. (2016) 3D implementation of the symbol nomenclature for graphical representation of glycans. *Glycobiology* **26**, 786–787

10. Woods, R. J. (2018) Predicting the structures of glycans, glycoproteins, and their complexes. *Chem. Rev.* **118**, 8005–8024

11. Humphrey, W., Dalke, A., and Schulten, K. (1996) VMD: Visual molecular dynamics. *J. Mol. Graph. Model.* **14**, 33–38

12. Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., and Ferrin, T. E. (2004) UCSF chimera - a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612

13. Lemieux, R. U., and Spohr, U. (1994) How Emil Fischer was led to the lock and key concept for enzyme specificity. *Adv. Carbohydr. Chem. Biochem.* **50**, 1–20

14. Dashti, H., Westler, W. M., Wedell, J. R., Demler, O. V., Eghbalnia, H. R., Markley, J. L., and Mora, S. (2020) Probabilistic identification of saccharide moieties in biomolecules and their protein complexes. *Sci. Data* **7**, 210

15. Woods, R. J. (2020) *Woods Group. (2005-2020) GLYCAM Web*, Complex Carbohydrate Research Center, University of Georgia, Athens, GA

16. Kwong, P. D., Wyatt, R., Desjardins, E., Robinson, J., Culp, J. S., Hellmig, B. D., Sweet, R. W., Sodroski, J., and Hendrickson, W. A. (1999) Probability analysis of variational crystallization and its application to gp120, the exterior envelope glycoprotein of type 1 human immunodeficiency virus (HIV-1). *J. Biol. Chem.* **274**, 4115–4123

17. Chang, V. T., Crispin, M., Aricescu, A. R., Harvey, D. J., Nettleship, J. E., Fennelly, J. A., Yu, C., Boles, K. S., Evans, E. J., Stuart, D. I., Dwek, R. A., Jones, E. Y., Owens, R. J., and Davis, S. J. (2007) Glycoprotein structural genomics: Solving the glycosylation problem. *Structure* **15**, 267–273

18. Moremen, K. W., Ramiah, A., Stuart, M., Steel, J., Meng, L., Forouhar, F., Moniz, H. A., Gahlay, G., Gao, Z. W., Chapla, D., Wang, S., Yang, J. Y., Prabhakar, P. K., Johnson, R., dela Rosa, M., *et al.* (2018) Expression system for structural and functional studies of human glycosylation enzymes. *Nat. Chem. Biol.* **14**, 156–162

19. Yanaka, S., Yagi, H., Yogo, R., Yagi-Utsumi, M., and Kato, K. (2018) Stable isotope labeling approaches for NMR characterization of glycoproteins using eukaryotic expression systems. *J. Biomol. NMR* **71**, 193–202

20. Chalmers, G. R., Eletsky, A., Morris, L. C., Yang, J. Y., Tian, F., Woods, R. J., Moremen, K. W., and Prestegard, J. H. (2019) NMR resonance assignment methodology: Characterizing large sparsely labeled glycoproteins. *J. Mol. Biol.* **431**, 2369–2382

21. Murin, C. D., Wilson, I. A., and Ward, A. B. (2019) Antibody responses to viral infections: A structural perspective across three different enveloped viruses. *Nat. Microbiol.* **4**, 734–747

22. Uversky, V. N., and Redwan, E. M. (2017) Erythropoietin and co.: Intrinsic structure and functional disorder. *Mol. Biosyst.* **13**, 56–72

23. Elliott, S., Lorenzini, T., Asher, S., Aoki, K., Brankow, D., Buck, L., Busse, L., Chang, D., Fuller, J., Grant, J., Hernday, N., Hokum, M., Hu, S., Knudten, A., Levin, N., *et al.* (2003) Enhancement of therapeutic protein *in vivo* activities through glycoengineering. *Nat. Biotechnol.* **21**, 414–421

24. Cheetham, J. C., Smith, D. M., Aoki, K. H., Stevenson, J. L., Hoeffel, T. J., Syed, R. S., Egrie, J., and Harvey, T. S. (1998) NMR structure of human erythropoietin and a comparison with its receptor bound conformation. *Nat. Struct. Biol.* **5**, 861–866

25. Syed, R. S., Reid, S. W., Li, C. W., Cheetham, J. C., Aoki, K. H., Liu, B. S., Zhan, H. J., Osslund, T. D., Chirino, A. J., Zhang, J. D., Finer-Moore, J., Elliott, S., Sitney, K., Katz, B. A., Matthews, D. J., *et al.* (1998) Efficiency of

26. Wang, Y. X., Yang, S. H., Brimble, M. A., and Harris, P. W. R. (2020) Recent progress in the synthesis of homogeneous erythropoietin (EPO) glycoforms. *Chembiochem* **21**, 3301–3312

27. Wang, P., Dong, S. W., Shieh, J. H., Peguero, E., Hendrickson, R., Moore, M. A. S., and Danishefsky, S. J. (2013) Erythropoietin derived by chemical synthesis. *Science* **342**, 1357–1360

28. Yang, Q., An, Y. M., Zhu, S. L., Zhang, R. S., Loke, C. M., Cipollo, J. F., and Wang, L. X. (2017) Glycan remodeling of human erythropoietin (EPO) through combined mammalian cell engineering and chemoenzymatic transglycosylation. *ACS Chem. Biol.* **12**, 1665–1673

29. Grant, O. C., Montgomery, D., Ito, K., and Woods, R. J. (2020) Analysis of the SARS-CoV-2 spike protein glycan shield reveals implications for immune recognition. *Sci. Rep.* **10**, 14991

30. Casalino, L., Gaieb, Z., Goldsmith, J. A., Hjorth, C. K., Dommer, A. C., Harbison, A. M., Fogarty, C. A., Barros, E. P., Taylor, B. C., McLellan, J. S., Fadda, E., and Amaro, R. E. (2020) Beyond shielding: The roles of glycans in the SARS-CoV-2 spike protein. *ACS Cent. Sci.* **6**, 1722–1734

31. Cobb, B. A. (2020) The history of IgG glycosylation and where we are now. *Glycobiology* **30**, 202–213

32. Huber, R., Deisenhofer, J., Colman, P. M., Matsushima, M., and Palm, W. (1976) Crystallographic structure studies of an IgG molecule fc fragment. *Nature* **264**, 415–420

33. Deisenhofer, J. (1981) Crystallographic refinement and atomic models of a human fc fragment and its complex with fragment-B of protein-A from staphylococcus-aureus at 2.9-A and 2.8-A resolution. *Biochemistry* **20**, 2361–2370

34. Frank, M., Walker, R. C., Lanzilotta, W. N., Prestegard, J. H., and Barb, A. W. (2014) Immunoglobulin G1 Fc domain motions: Implications for Fc engineering. *J. Mol. Biol.* **426**, 1799–1811

35. Nose, M., and Wigzell, H. (1983) Biological significance of carbohydrate chains on monoclonal-antibodies. *Proc. Natl. Acad. Sci. U. S. A.* **80**, 6632–6636

36. Kaneko, Y., Nimmerjahn, F., and Ravetch, E. V. (2006) Anti-inflammatory activity of immunoglobulin G resulting from Fc sialylation. *Science* **313**, 670–673

37. Okerblom, J., and Varki, A. (2017) Biochemical, cellular, physiological, and pathological consequences of human loss of *N*-glycolylneuraminic acid. *Chembiochem* **18**, 1155–1171

38. Irvine, E. B., and Alter, G. (2020) Understanding the role of antibody glycosylation through the lens of severe viral and bacterial diseases. *Glycobiology* **30**, 241–253

39. Yamaguchi, Y., and Barb, A. W. (2020) A synopsis of recent developments defining how N-glycosylation impacts immunoglobulin G structure and function. *Glycobiology* **30**, 214–225

40. Barb, A. W., and Prestegard, J. H. (2011) NMR analysis demonstrates immunoglobulin G *N*-glycans are accessible and dynamic. *Nat. Chem. Biol.* **7**, 147–153

41. Kelly, J. A., Sielecki, A. R., Sykes, B. D., James, M. N. G., and Phillips, D. C. (1979) X-ray crystallography of the bacterial cell wall trisaccharide NAM-NAG-NAM to lysozyme. *Nature* **282**, 875–878

42. Davies, G. J., Withers, S. G., and Vocadlo, D. J. (2009) The chitopentaose complex of a mutant hen egg-white lysozyme displays no distortion of the-1 sugar away from a C-4(1) chair conformation. *Aust. J. Chem.* **62**, 528–532

43. Limb, M. A. L., Suardiaz, R., Grant, I. M., and Mulholland, A. J. (2019) Quantum mechanics/molecular mechanics simulations show saccharide distortion is required for reaction in hen egg-white lysozyme. *Chemistry* **25**, 764–768

44. Lombard, V., Ramulu, H. G., Drula, E., Coutinho, P. M., and Henrissat, B. (2014) The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.* **42**, D490–D495

45. Bonnardel, F., Mariethoz, J., Salentin, S., Robin, X., Schroeder, M., Perez, S., Lisacek, F., and Imberty, A. (2019) UniLectin3D, a database of carbohydrate binding proteins with curated information on 3D structures and interacting ligands. *Nucleic Acids Res.* **47**, D1236–D1244

46. Torres, C. R., and Hart, G. W. (1984) Topography and polypeptide distribution of terminal *N*-acetylglucosamine residues on the surfaces of intact lymphocyte - evidence for O-linked GlcNAc. *J. Biol. Chem.* **259**, 3308–3317

47. Hart, G. W. (2019) Nutrient regulation of signaling and transcription. *J. Biol. Chem.* **294**, 2211–2231

48. Lazarus, M. B., Jiang, J. Y., Gloster, T. M., Zandberg, W. F., Whitworth, G. E., Vocadlo, D. J., and Walker, S. (2012) Structural snapshots of the reaction coordinate for O-GlcNAc transferase. *Nat. Chem. Biol.* **8**, 966–968

49. Levine, Z. G., and Walker, S. (2016) The biochemistry of O-GlcNAc transferase: Which functions make it essential in mammalian cells? *Annu. Rev. Biochem.* **85**, 631–657

50. Roth, C., Chan, S., Offen, W. A., Hemsworth, G. R., Willems, L. I., King, D. T., Varghese, V., Britton, R., Vocadlo, D. J., and Davies, G. J. (2017) Structural and functional insight into human O-GlcNAcase. *Nat. Chem. Biol.* **13**, 610–612

51. Li, B. B., Li, H., Hu, C. W., and Jiang, J. Y. (2017) Structural insights into the substrate binding adaptability and specificity of human O-GlcNAcase. *Nat. Commun.* **8**, 666

52. Alonso, J., Schimpl, M., and van Aalten, D. M. F. (2014) O-GlcNAcase: Promiscuous hexosaminidase or key regulator of O-GlcNAc signaling? *J. Biol. Chem.* **289**, 34433–34439

53. Danev, R., Yanagisawa, H., and Kikkawa, M. (2019) Cryo-electron microscopy methodology: Current aspects and future directions. *Trends Biochem. Sci.* **44**, 837–848

54. Ognjenovic, J., Grisshammer, R., and Subramaniam, S. (2019) Frontiers in cryo electron microscopy of complex macromolecular assemblies. *Ann. Rev. Biomed. Eng.* **21**, 395–415

55. Braunger, K., Pfeffer, S., Shrimal, S., Gilmore, R., Berninghausen, O., Mandon, E. C., Becker, T., Forster, F., and Beckmann, R. (2018) Structural basis for coupling protein transport and *N*-glycosylation at the mammalian endoplasmic reticulum. *Science* **360**, 215–218

56. Ramirez, A. S., Kowal, J., and Locher, K. P. (2019) Cryo-electron microscopy structures of human oligosaccharyltransferase complexes OST-A and OST-B. *Science* **366**, 1372–1375

57. Mohorko, E., Owen, R. L., Malojcic, G., Brozzo, M. S., Aebi, M., and Glockshuber, R. (2014) Structural basis of substrate specificity of human oligosaccharyl transferase subunit N33/Tusc3 and its role in regulating protein *N*-glycosylation. *Structure* **22**, 590–601

58. Everett, J. K., Tejero, R., Murthy, S. B. K., Acton, T. B., Aramini, J. M., Baran, M. C., Benach, J., Cort, J. R., Eletsky, A., Forouhar, F., Guan, R. J., Kuzin, A. P., Lee, H. W., Liu, G. H., Mani, R., *et al.* (2016) A community resource of experimental data for NMR/X-ray crystal structure pairs. *Protein Sci.* **25**, 30–45

59. Withka, J. M., Wyss, D. F., Wagner, G., Arulanandam, A. R. N., Reinherz, E. L., and Recny, M. A. (1993) Structure of the glycosylated adhesion domain of human T-lymphocyte glycoprotein CD2. *Structure* **1**, 69–81

60. Chakraborty, A., Deligey, F., Quach, J., Mentink-Vigier, F., Wang, P., and Wang, T. (2020) Biomolecular complex viewed by dynamic nuclear polarization solid-state NMR spectroscopy. *Biochem. Soc. Trans.* **48**, 1089–1099

61. Fulcher, Y. G., Prior, S. H., Masuko, S., Li, L. Y., Pu, D., Zhang, F. M., Linhardt, R. J., and Van Doren, S. R. (2017) Glycan activation of a sheddase: Electrostatic recognition between heparin and proMMP-7. *Structure* **25**, 1100–1110

62. Shukla, R., Medeiros-Silva, J., Parmar, A., Vermeulen, B. J. A., Das, S., Paioni, A. L., Jekhmane, S., Lorent, J., Bonvin, A., Baldus, M., Lelli, M., Veldhuizen, E. J. A., Breukink, E., Singh, I., and Weingarth, M. (2020) Mode of action of teixobactins in cellular membranes. *Nat. Commun.* **11**, 2848

63. Gao, Q., Yang, J. Y., Moremen, K. W., Flanagan, J. G., and Prestegard, J. H. (2018) Structural characterization of a heparan sulfate pentamer interacting with LAR-Ig1-2. *Biochemistry* **57**, 2189–2199

64. Ohtake, Y., and Li, S. X. (2015) Molecular mechanisms of scar-sourced axon growth inhibitors. *Brain Res.* **1619**, 22–35

65. Coles, C. H., Shen, Y. J., Tenney, A. P., Siebold, C., Sutton, G. C., Lu, W. X., Gallagher, J. T., Jones, E. Y., Flanagan, J. G., and Aricescu, A. R. (2011) Proteoglycan-specific molecular switch for RPTP sigma clustering and neuronal extension. *Science* **332**, 484–488

66. van Zundert, G. C. P., Rodrigues, J., Trellet, M., Schmitz, C., Kastritis, P. L., Karaca, E., Melquiond, A. S. J., van Dijk, M., de Vries, S. J., and Bonvin, A. (2016) The HADDOCK2.2 web server: User-friendly integrative modeling of biomolecular complexes. *J. Mol. Biol.* **428**, 720–725

67. Koukos, P. I., and Bonvin, A. (2020) Integrative modelling of biomolecular complexes. *J. Mol. Biol.* **432**, 2861–2881

68. Clausen, T. M., Sandoval, D. R., Spliid, C. B., Pihl, J., Perrett, H. R., Painter, C. D., Narayanan, A., Majowicz, S. A., Kwong, E. M., McVicar, R. N., Thacker, B. E., Glass, C. A., Yang, Z., Torres, J. L., Golden, G. J., *et al.* (2020) SARS-CoV-2 infection depends on cellular heparan sulfate and ACE2. *Cell* **183**, 1043–1057