

Modeling glycans with AlphaFold 3: capabilities, caveats, and limitations

Chin Huang^{1,2,*} , Natarajan Kannan^{1,3} , Kelley W. Moremen^{1,2,*} 

¹Department of Biochemistry and Molecular Biology, University of Georgia, 120 Green Street, Athens, GA 30602, United States, ²Complex Carbohydrate Research Center, University of Georgia, 315 Riverbend Road, Athens, GA 30602, United States, ³Institute of Bioinformatics, University of Georgia, 120 Green Street, Athens, GA 30602, United States

*Corresponding authors: Kelley W. Moremen, moremen@uga.edu; Chin Huang, chin@uga.edu

Abstract

Glycans are complex carbohydrates that exhibit extraordinary structural complexity and stereochemical diversity while playing essential roles in many biological processes, including immune regulation, pathogen recognition, and cell communication. In humans, more than half of all proteins are glycosylated, particularly those in secretory and membrane-associated pathways, highlighting the importance of glycans in health and disease. The recent release of the AlphaFold 3 source code enables customizable modeling not only of proteins but also glycan-containing biomolecular complexes. We assessed the capacity of AlphaFold 3 to model glycans using several input formats and identified a hybrid syntax employing Chemical Component Dictionary (CCD)-based molecular building blocks linked by “bondedAtomPairs” (BAP) as most effective in generating stereochemically valid glycan models. This workflow was used to create a library of AlphaFold 3 input templates and corresponding structural models for various glycan classes. We further explored capabilities, limitations, and remediation strategies for modeling problematic structures. Glycan interactions were also modeled with glycosylation enzymes and lectins with benchmarking and validation against known crystal structures. This protocol-driven approach is valuable for generating stereochemically valid, static models of glycan-protein interactions to support hypothesis development and subsequent structural and functional validation. However, caution should be observed in overinterpretation of the static models since glycans are known to exhibit considerable conformational dynamics that can be further captured by equilibrium sampling using molecular dynamics-based approaches. By sharing benchmarked examples using the BAP syntax we aim to support broader evaluation of AlphaFold 3 in studying glycan-related mechanisms in biosynthesis, signaling, infection, and disease.

Keywords: AlphaFold 3; glycan modeling; glycoprotein modeling; glycan stereochemistry; glycan-protein interaction.

Introduction

Glycans are ubiquitous in all domains of life (Varki 2017) and play essential roles in processes such as immune modulation (Pinho et al. 2023), cancer progression (Smith and Bertozzi 2021), and neural development (Schnaar et al. 2014). Glycan structures are assemblies of monosaccharide building blocks linked through the C1 or C2 anomeric carbon in glycosidic linkages to hydroxyl groups of other monosaccharides in linear or branched configurations or to functionalities in other molecule types (Moremen and Haltiwanger 2019). These structures can be found free in solution (e.g. human milk oligosaccharides (HMOs) and hyaluronic acid (HA)), covalently linked to proteins (N- and O-linked glycans), lipids (glycosphingolipids (GSLs), dolichol phosphate (Dol-P)-linked oligosaccharides), in bridging linkages between lipids and proteins (GPI anchors) (Moremen et al. 2012), to other macromolecules (xenobiotics) (Allain et al. 2020), and numerous other classes of glycoconjugates. Glycan structures are assembled by glycosyltransferases (Moremen and Haltiwanger 2019) and deconstructed by glycoside hydrolases (Moremen et al. 1994) and lyases (Garron and Cygler 2010), and play numerous roles in modulation of function including recognition by glycan-binding proteins (lectins) (Stowell et al. 2010).

The structural complexities of glycans arise from the diversity of glycan linkages between monosaccharides,

numerous possibilities for glycan branching, two possible anomeric linkages, multiple sugar ring configurations and puckering, and diverse additional modifications (Woods 2018). Rotational flexibility of glycans at the anomeric linkage also leads to conformational dynamics that present challenges in structure determination. This glycan structural diversity, complexity, and dynamics remain a formidable challenge in glycobiology research. Computational approaches, including molecular docking (Grant and Woods 2014), molecular dynamics (MD) (Case et al. 2005; Kirschner et al. 2008; Fadda 2022), and quantum mechanics/molecular mechanics (QM/MM) simulations (Mendoza and Masgrau 2021; Perez and Makshakova 2022), can generate predictive models that complement empirical structural studies. Recent developments in AI approaches, including RoseTTAFold All-Atom (Krishna et al. 2024), AlphaFold 3 (Abramson et al. 2024), DeepGlycanSite (He et al. 2024), Chai-1 (Chai Discovery et al. 2024), CLIMBS (Luo and Parmeggiani 2025), and Boltz-2 (Passaro et al. 2025) now present an additional and complementary set of in silico tools for modeling glycan-protein interaction.

In 2018, AlphaFold initially demonstrated highly accurate predictive modeling capabilities for proteins, particularly when homologous protein templates were available in the PDB (Senior et al. 2020). AlphaFold 2 expanded its

Received: May 09, 2025. Revised: July 28, 2025. Accepted: August 17, 2025

© The Author(s) 2025. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

capabilities to novel proteins and dimers (Jumper et al. 2021; Evans et al. 2022). AlphaFold 3 (AF3), with an updated diffusion-based generative framework, enabled the modeling of diverse biological molecules, including a growing number of bound ligands. Initially released as a web server format (<https://alphafoldserver.com/>), AF3 provides support for modeling multimers, a limited library of ligand structures (e.g. nucleic acids, selected metal ions, and cofactors), and post-translational modifications (PTMs) (Abramson et al. 2024). While protein glycosylation was included among PTMs, the ambiguity in specifying linkages and higher-order stereochemistry presented limitations in evaluating the application of AF3 in modeling protein-glycan interactions. AF3 was also released as a command-line interface, standalone version in late 2024 (<https://github.com/google-deepmind/alphafold3>) that presented opportunities for universal ligand input, but the syntax for specifying glycan stereochemistry remained a challenge. Concurrently, the Protein Data Bank (PDB) underwent an effort in carbohydrate remediation (Shao et al. 2021) (7.7% of entries contain glycan structures (Dashti et al. 2020, Prestegard 2021)). Monosaccharides have been standardized, accompanied by the removal of obsolete polysaccharide entries, implementation of consistent linkage annotations, and the integration of cross-references to external glycoinformatics resources. Although not originally intended for deep learning applications, this curated dataset now serves as a valuable resource for model training and evaluation of AF3 predictions.

Here, we investigate the input syntax for specifying glycan structures using standalone AF3 and its impact in modeling key glycan conformational features, including anomeric configurations, epimeric orientations (axial versus equatorial), and ring puckering. We compare modeling outcomes across multiple input formats and identify the *bondedAtomPairs* (BAP) syntax that yields the most consistent and stereochemically plausible glycan models. Fidelity of modeling is assessed through comparison with empirical structures. The BAP syntax framework supports the generation of starting glycan structures for downstream analysis and expands the applicability of AF3 in modeling complex macromolecules beyond glycoconjugates.

Results

Enhanced structural fidelity through *bondedAtomPairs* (BAP)-defined glycosidic linkages

Monosaccharides possess distinct structural features, including absolute configurations (D/L nomenclature), ring forms (linear, pyranose, or furanose), and anomeric centers (α/β), which pose substantial challenges in cheminformatics (Seiberger 2022). Linkage of monosaccharide building blocks into larger glycan structures presents further complications, since multiple hydroxyl positions are available for glycosidic linkages, and combined modification of multiple hydroxyls on a single monosaccharide can lead to branched structures (Lebrilla et al. 2022). AF3 generally allows ligand definition through three input formats: Simplified Molecular Input Line Entry System (SMILES) (Weininger 1988), Chemical Component Dictionary (CCD) codes (Feng et al. 2004), and user-defined CCDs (*userCCD*). Although these formats are

intended to enable universal ligand input, their performance in handling stereochemically complex glycans has yet to be systematically evaluated.

We first assessed the modeling quality in AF3 using a simple linear HMO, lacto-*N*-neotetraose (LNnT) (Fig. 1a). Given the simplicity of the SMILES format, we initially tested its feasibility. AF3 correctly represented the absolute configurations, ring forms, linkage order, and anomeric centers of LNnT. However, a Gal residue (residue 2) was incorrectly modeled as Glc, due to misassignment of the C4 hydroxyl from axial to equatorial (Fig. 1b). Furthermore, the SMILES input format does not support atom indexing, thereby limiting its applicability for specifying covalent linkages in multicomponent assemblies.

To circumvent this, the AF3 development team introduced *rdkit_utils* (https://github.com/google-deepmind/alphafold3/blob/main/src/alphafold3/data/tools/rdkit_utils.py), which converts SMILES into a CCD-like format compatible with *userCCD* input. This tool generates idealized coordinates designed to enhance model performance. However, models generated with this approach still exhibited stereochemical problems. Specifically, the LNnT Gal residue 2 retained the erroneous C4 hydroxyl orientation, and the glycosidic bond between Glc and Gal (residues 1 and 2) was incorrectly rendered as an α - rather than β -linkage (Fig. 1c). Additional attempts to convert an energy-minimized PDB structure of LNnT (computed by GLYCAM (Grant et al. 2025); Fig. 1e) into *userCCD* format using *rdkit_utils* introduced further errors (data not shown). Alternative programs for CCD generation (Agirre 2017; Steiner and Tucker 2017) such as AceDRG (Long et al. 2017) yielded similarly incorrect configurations in AF3 (data not shown).

An alternative strategy for glycan structure input in AF3 is to specify individual monosaccharide building blocks in a glycan structure using their unique CCD identifiers. By leveraging the curated monosaccharide CCD library, larger glycan assemblies can be generated by explicitly defining glycosidic linkages between the building blocks using the *bondedAtomPairs* (BAP) syntax in the input JavaScript Object Notation (JSON) file (described in detail below). Using this approach, we were able to model the correct stereochemical structure of LNnT, including all anomeric configurations and axial/equatorial orientations (Fig. 1d).

To evaluate whether this approach generalizes to more complicated branched structures, we modeled a complex biantennary *N*-glycan G2 (Fig. 1f). SMILES input resulted in multiple structural errors: GlcNAc (residue 1) C4 hydroxyl was incorrectly modeled as axial; C2 hydroxyls of Man residues (residue 3–5) and C4 hydroxyls of Gal residues (residue 8 and 9) exhibited erroneous equatorial orientations; residues 3, 4, 8 and 9 were depicted in opposite anomeric configurations (α instead of β , or vice versa) (Fig. 1g). Models generated from *userCCD* converted through *rdkit_utils* were improved, though not flawless (Fig. 1h); of note, the C2 hydroxyls on Man residues 3–5 were still incorrect. Due to limitations in *rdkit_utils*, default parameters failed to generate correct conformers for G2; setting *useRandomCoords = True* allowed generation of idealized coordinates. Ultimately, using CCD input linked with *bondedAtomPairs* (BAP) enabled effective modeling of G2, recapitulating all stereochemical features correctly (Fig. 1i). These predictions were in strong agreement with energy-minimized structures simulated by GLYCAM (Grant et al. 2025) (Fig. 1j).

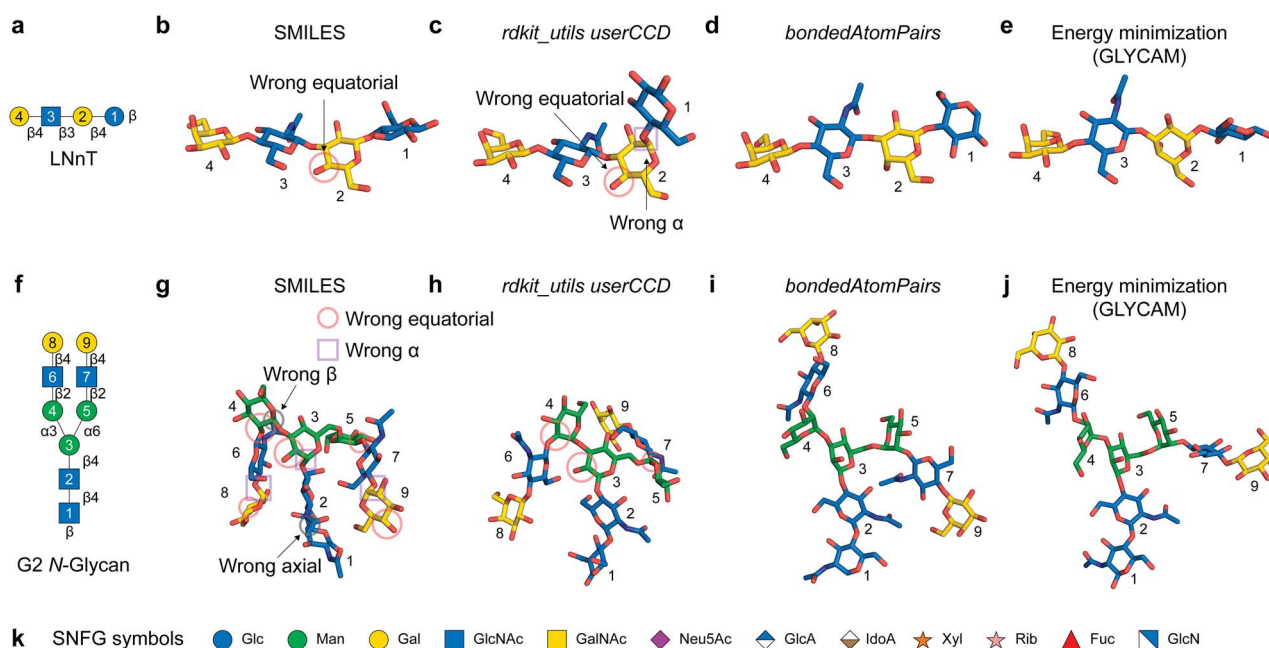


Fig. 1. Comparison of ligand input formats for glycan modeling in AlphaFold 3. **a)** Symbol nomenclature for glycans (SNFG) cartoon of lacto-*N*-neotetraose (LNnT) with annotated residue numbers. **b)** AF3-modeled LNnT using simplified molecular input line entry system (SMILES) input, **c)** user-defined chemical component dictionary (*userCCD*) converted by *rdkit_utils*, **d)** CCD with *bondedAtomPairs*, **e)** energy minimized LNnT simulated by GLYCAM. **f)** SNFG cartoon of biantennary G2 *N*-linked glycan. **g)** AF3-modeled G2 using SMILES, **h)** *userCCD*, **i)** CCD with *bondedAtomPairs*, **j)** GLYCAM-simulated energy-minimized G2. **k)** SNFG legend used in this article. Gray and pink circles, or purple square indicate wrong stereochemistry as labeled.

Formalization of *bondedAtomPairs* (BAP) syntax for glycan topology specification

AF3 accepts protein sequence input in the form of a JSON file. This file also allows specification of additional inputs, including modeling random seeds, DNA, RNA, modified amino acid residues, and small-molecule ligands using SMILES, CCD, or *userCCD* formats (<https://github.com/google-deepmind/alphafold3/blob/main/docs/input.md>). In particular, the *bondedAtomPairs* (BAP) syntax can be used to define covalent linkages between distinct molecular entities; either between proteins and ligands or between individual ligand components. This syntax offers greater detail and flexibility compared to the simplified format used in the AF3 Server interface. Surprisingly, AF3 can also process input consisting solely of ligands without a protein context, which offers a convenient way to validate ligand syntax and bonding logic prior to full complex modeling.

Specifying glycans through *bondedAtomPairs* (BAP) syntax requires the use of monosaccharide CCD codes with appropriate anomeric specification. For simplicity, a partial list of human-relevant monosaccharide CCD codes is presented in Fig. 2a. A comprehensive CCD code table, expanded to include L/D configurations, linear and cyclic (pyranose/furanose) forms, and α/β anomeric designations was compiled based on the SNFG (Varki et al. 2015) and is provided in Table S1. An example of *bondedAtomPairs* (BAP) syntax for the G2 *N*-glycan is illustrated in Fig. 2b. The order of the residues in the *ccdCodes* section of the JSON file corresponds directly to the residue numbering used in the *bondedAtomPairs* (BAP) field, as visualized in Fig. 2c. To support multiple glycan ligands or multiple glycosylation sites, unique identifiers can be defined in the *id* field; in this example, a single copy of G2 was assigned as “NG.”

To demonstrate bonding logic, we showcased the β 1,4 linkage between chitobiose and β -Man. The CCD codes for 2-acetamido-2-deoxy- β -D-glucopyranose (NAG) and β -D-mannopyranose (BMA) were assigned as residues 2 and 3, respectively, in the *ccdCodes* section (Fig. 2b, red). A β 1,4 linkage was defined by connecting the O4 atom of NAG (residue 2) with the C1 atom of BMA (residue 3) in the *bondedAtomPairs* (BAP) field (Fig. 2b, blue). In this case, the β -linkage is being specified by the CCD code for the β -anomer of mannose (BMA). Atom numbering for NAG and BMA is provided in Fig. 2d. By default, AF3 removes the O1 atom of the donor monosaccharide during glycosidic bond formation, as specified by the *.pdbx_leaving_atom_flag* in the official macromolecular Crystallographic Information File (mmCIF).

Extending our investigation to additional monosaccharides, we observed that not all leaving groups are automatically removed as expected. Sialic acid, specifically *N*-acetyl- α -neuraminic acid (CCD code SIA), a ketose unlike most aldose monosaccharides, links through its C2 position (Meng et al. 2013). In this case, the O2 atom of SIA is not removed by default (Fig. S1a), even when the *.pdbx_leaving_atom_flag* is correctly set. We also tested an alternative linkage by connecting the C6 atom of GAL to the O2 atom of SIA. In this configuration, both the O6 atom of GAL and the O2 atom of SIA were retained, generating unrealistic valence states and unfavorable torsion angles at the glycosidic junction (Fig. S1b). Manual deletion of leaving oxygens in the monosaccharide CCD entry, O2 in SIA (Fig. S1c and d), is necessary to maintain proper valency. The items removed were highlighted in the full standard SIA mmCIF (Supplementary Document 1). Similarly, CCDs converted into *userCCD* format without manual edits retain O1 atoms, leading to duplicate oxygen atoms at glycosidic junctions.

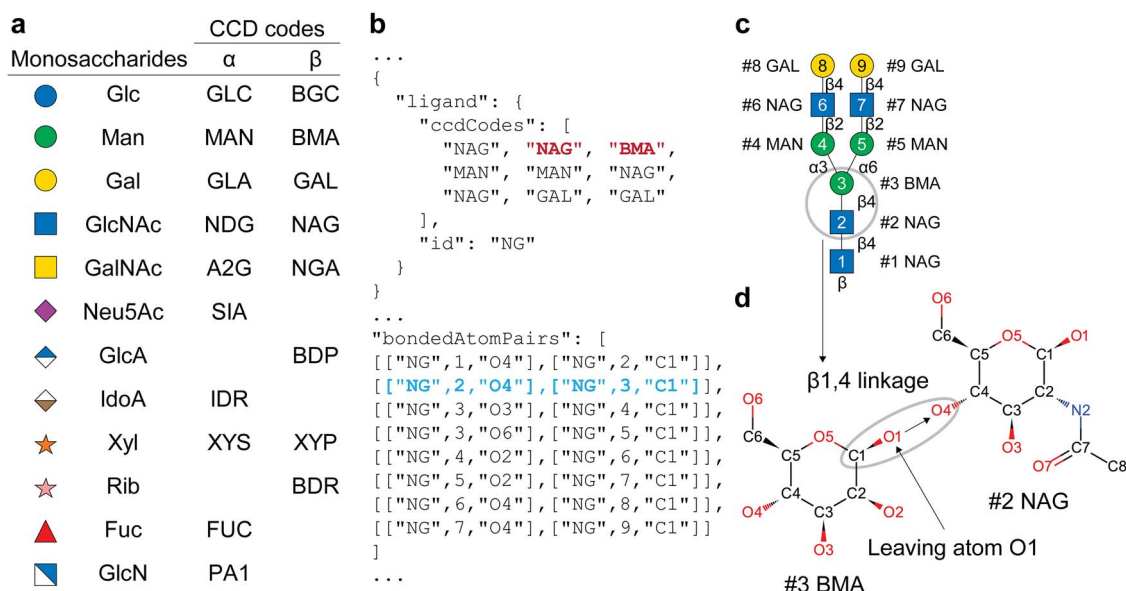


Fig. 2. Glycan specification using chemical component dictionary (CCD) and *bondedAtomPairs* (BAP) syntax in AlphaFold 3. **a)** Representative CCD codes for common human monosaccharides. **b)** Partial JSON script used to generate G2 *N*-linked glycan. **c)** SNFG representation of G2 *N*-linked glycan with residue numbering equivalent to what is used in the JSON modeling input file. The order of the monosaccharides in the JSON file *ccdCodes* list corresponds with the residue numbering in the SNFG representation. **d)** Demonstration of a β 1,4 linkage between GlcNAc and Man, specified by the second and third CCD entries (NAG and BMA shown in bold red font in panel **b**). The glycosidic bond is defined by the *bondedAtomPairs* (BAP) field (shown in bold blue font in panel **b**), specifying a bond between the O4 atom of NAG and the C1 atom of BMA (oval in panel **d**). The entry in the *bondedAtomPairs* (BAP) list leads to covalent bond formation between the respective hydroxyl oxygen (O4 of NAG) and the anomeric C1 (C1 of BMA) with the removal of the β -linked O1 (leaving atom) on the C1 of BMA. The resulting glycosidic linkage retains the anomeric configuration at C1 of the original CCD entry (β -linkage for BMA).

This behavior also applies to non-carbohydrate ligands. Glycan modifications such as phosphorylation (Li et al. 2022), sulfation (Wu et al. 2023), methylation (Urbanowicz et al. 2012), and acetylation (Wang et al. 2021) follow the same principle: manual removal of leaving oxygen atoms is required to ensure chemically plausible connectivity. In light of the error-prone nature of glycan linkage specification and the advanced domain knowledge it demands, we have curated a collection of validated JSON input files encompassing all common *N*-glycans (Stanley et al. 2022) and other glycans modeled herein as reference templates for reproducible implementation (Supplementary Document 2).

Benchmarking AlphaFold 3 for high-complexity glycan modeling

After establishing the syntax for consistent glycan modeling, we benchmarked the capacity of AF3 to model glycans within glycoprotein complexes. We began with the highly branched M9 *N*-glycan (Fig. 3a). As a free-reducing end ligand, M9 was appropriately modeled with expected conformations both as a free glycan (Fig. 3b) and when bound as ligand in the active site of an α -mannosidase involved in trimming M9 to M5 during *N*-glycan maturation (MAN1A1 (Lal et al. 1994), Fig. 3c). The α 1,3-Man branch of M9 (3-arm) extended deep into the catalytic pocket, with the terminal α 1,2-Man glycone residue (residue 9) coordinated to an enzyme-bound Ca^{2+} ion. Notably, residue 9 adopted a distorted $^3\text{O}B$ boat conformation (Fig. 3i) rather than its solution-phase 4C_1 chair ground state (Fig. 3h). Compared with the crystallographic structure (Xiang et al. 2016) (Fig. 3d), the 3-arm adopted a similar binding pose (Fig. 3e) (root-mean-square deviation (RMSD) 0.284 Å), whereas the core GlcNAc (residue 2) and

α 1,6-Man branch (6-arm) deviated. These latter differences from the empirical crystal structure are likely due to their solvent exposure and increased conformational flexibility. The distorted $^3\text{O}B$ boat conformation of the terminal Man (residue 9) in the AF3 model contrasts with an unfavorable, high-energy 1C_4 chair conformation for the glycone in the crystal structure (Fig. 3j). Although this may initially appear inconsistent, the AF3 prediction aligns with the Cremer-Pople ring puckering itinerary (Cremer and Pople 1975) proposed for a homologous GH47 α -mannosidase (Thompson et al. 2012) where glycone conformation transitions during catalysis from $^3\text{O}B/{}^3S_1$ through 3H_4 intermediate *en route* to a 1C_4 enzymatic product.

It is noteworthy that the crystal structure of MAN1A1 was solved in complex with a lanthanum ion (La^{3+}) as an inhibitor (Xiang et al. 2016). However, AF3 models incorporating either Ca^{2+} or La^{3+} generated the same $^3\text{O}B$ structure for the glycone residue rather than 1C_4 conformation (data not shown). This observation suggests that even if a MAN1A1 structure was included in AF3 training set, its internal representation does not rigidly duplicate the training input. Instead, AF3 accommodates alternative, yet chemically plausible, conformations including more flexible branches like the 6-arm and distinct ring-puckering states, presumably due to the diffusion-based architecture and learned representations of relevant complexes from the PDB such as the co-crystal structure of MAN1A1 with M9 and La^{3+} (Xiang et al. 2016) and glycoside hydrolase family 47 (GH47) members co-crystalized with a non-hydrolyzed thiomannobiose (Karaveg et al. 2005; Thompson et al. 2012) or glycone mimic inhibitors, 1-deoxymannojirimycin or kifunensine (Vallee et al. 2000; Lobsanov et al. 2002).

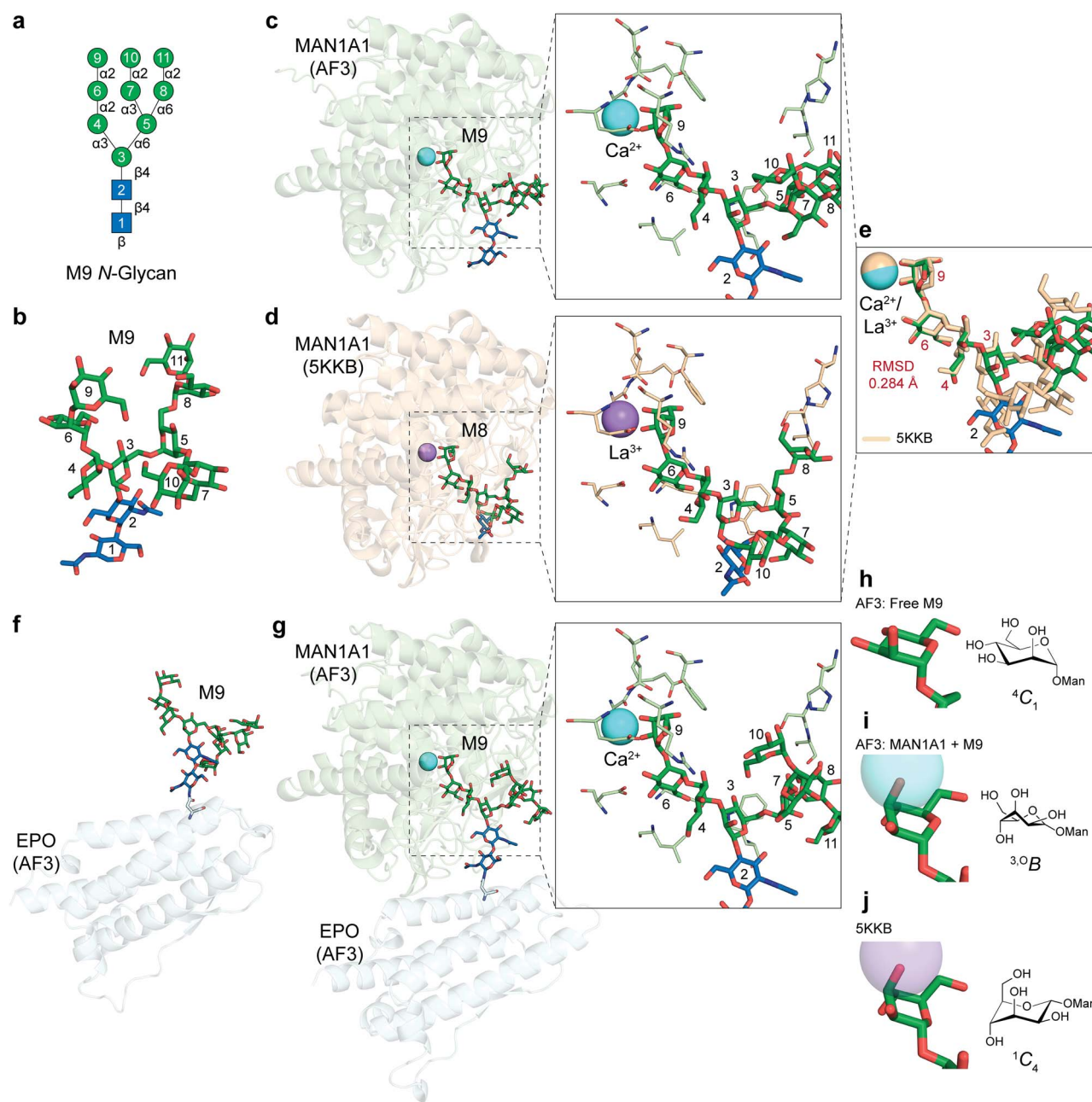


Fig. 3. Benchmarking AlphaFold 3 models with MAN1A1/M9/Ca²⁺ Michaelis complex. **a**) SNFG cartoon representation of M9 N-glycan; residue numbers are consistently annotated across panels. **b**) AF3-modeled free-reducing end M9 glycan. **c**) Model of the *Mus musculus* Golgi mannosidase MAN1A1 bound to M9 and Ca²⁺, representing a Michaelis complex; the active site is highlighted, showing the glycan binding pose and interacting residues. **d**) Crystallographic structure of MAN1A1 in complex with M8 and La³⁺ (PDB: 5KKB). **e**) Structural alignment of the AF3 MAN1A1/M9/Ca²⁺ model (colored) with the crystallographic structure (tan); mannose residues on the 3-arm (residues 4, 6, and 9) overlay well in the catalytic pocket, while solvent-exposed residues display greater divergence from the complex in the crystal structure. The root-mean-square deviation (RMSD) of residues 3, 4, 6, and 9 (colored in red) between the AF3 model and 5KKB is annotated. **f**) M9 glycan modeled as an N-glycan attached to *H. sapiens* erythropoietin (EPO). **g**) MAN1A1 (green) modeled in complex with Ca²⁺ and M9 carried on EPO (cyan). **h-j**) Cremer-Pople ring pucker conformations of α 1,2-mannose (residue 9) in different contexts: **h**) free-reducing end M9 (panel **b**) exhibiting a stable ⁴C₁ chair; **i**) MAN1A1-M9-Ca²⁺ model (panel **c**) showing a distorted ^{3,0}B boat; **j**) crystallographic 5KKB structure (panel **d**) adopting an unfavorable ¹C₄ chair.

To evaluate generalizability, we modeled M9 as a PTM on erythropoietin (EPO), which features solvent-exposed N-glycosylation sites with minimal steric hindrance (Adams et al. 2022) (Fig. 3f). When co-modeled with MAN1A1, the 3-arm adopted a binding pose (Fig. 3g) nearly identical to the complex that was observed using the free-reducing M9 (RMSD 0.264 Å). These results highlight the capacity of

AF3 to effectively model enzyme-glycoprotein complexes and recapitulate physiologically relevant intermediates.

We next evaluated a more complex case involving a glycosyltransferase bound to a divalent cation, a sugar nucleotide donor, and a glycan acceptor. The A1 N-glycan (Fig. S2a and b) was modeled in complex with MGAT2 (Kadirvelraj et al. 2018) (Fig. S2c), the GlcNAc transferase responsible

for converting hybrid to complex *N*-glycans. The 3-arm GlcNAc (residue 6) oriented away from the catalytic pocket toward an “exosite” identified in the MGAT2 crystal structure, while the 6-arm mannose (residue 5) extended into the active site and aligned with UDP-GlcNAc, which was coordinated with a manganese ion (Mn^{2+}). The modeled complex closely matched the crystallographic pose (Kadirvelraj et al. 2018), including the glycan (RMSD 0.241 Å), nucleotide, and metal ion arrangement (Fig. S2d and e). This model was also recapitulated on EPO as a glycoprotein scaffold (Fig. S2f) and in the ternary complex with MGAT2 (Fig. S2g).

To further test the performance of AF3, we benchmarked G2F, a fully extended, biantennary, core-fucosylated *N*-glycan (Fig. S3a and b), as the ligand for ST6GAL1 (Fig. S3c), an α 2,6-sialyltransferase. The 3-arm of G2F showed extensive interactions within the ST6GAL1 catalytic pocket, consistent with the known glycan branch preference for the enzyme (Barb et al. 2009), while the 6-arm remained more disordered due to solvent exposure and minimal interaction with the protein surface. The modeled conformation corresponded closely to the crystallographic structure (RMSD 0.236 Å) (Kuhn et al. 2013) (Fig. S3d and e), where the G2F glycan originated from the *N*-glycosylation of a crystallographic symmetry mate. Surprisingly, AF3 was able to model this substrate interaction as a ligand despite the complex not being explicitly present in the PDB file used for training (Kuhn et al. 2013). We also placed G2F on EPO as a PTM (Fig. S3f) and modeled an equivalent glycan-protein complex with ST6GAL1 (Fig. S3g).

Across all three high-complexity scenarios, M9 with MAN1A1, A1 with MGAT2, and G2F with ST6GAL1, AF3 generated stereochemically and conformationally plausible models that were in close agreement with crystallographic data. In several cases, it also captured subtleties such as ring puckering and flexible branching, offering valuable structural insight into glycan-protein interactions.

Modeling lipid-linked glycans for covalent glycan-lipid-protein assemblies

Building upon the optimized *bondedAtomPairs* (BAP) syntax and effective benchmarking of protein-glycan complexes, we next extended our evaluation to lipid-linked glycoconjugates. Glycosphingolipids (GSLs), composed of glycans linked to ceramide, represent a structurally diverse and biologically significant class of glycolipids (Schnaar et al. 2022). As the CCD database lacks a direct entry for ceramide, we modeled it by covalently linking sphingosine and stearic acid (CCD codes: SPH and STE). Complex GSLs such as the fully extended, sialylated ganglioside GP1c (Fig. 4a and 4f) and the highly fucosylated Lewis b antigen lactoside (Fig. 4b and 4g) were correctly modeled without notable conformational distortions. To assess receptor interactions, we modeled the globoside Globo H (Fig. 4c) in complex with Bc2L-C, a homotrimeric lectin from the opportunistic pathogen *Burkholderia cenocepacia* (Bermeo et al. 2020) (Fig. 4h). To reduce lipid tail disorder, ceramide-containing glycolipids were positioned on a mock lipid layer, resulting in more extended and physiologically plausible lipid tail conformations. Despite the inclusion of additional protein domains and full GSL components in the AF3 model, the binding pose remains consistent (RMSD 0.137 Å) with the known protein structure (PDB: 6TIG) (Bermeo et al. 2020).

We also modeled dolichol pyrophosphate (Dol-PP)-linked chitobiose, a key intermediate in *N*-glycan biosynthesis (Ramírez and Locher 2023) (Fig. 4d). Octaprenyl pyrophosphate (CCD code: OTP) was used as a Dol-PP analog. Unlike the β -linked chitobiose commonly found as a protein PTM, AF3 correctly modeled the α -linked Dol-PP glycan (Ramírez et al. 2019) (Fig. 4i). The model generated in complex with ALG1, Mn^{2+} , and GDP-mannose exhibited a plausible geometry for each ligand component, with the C4 hydroxyl of the terminal GlcNAc (residue 2) properly oriented toward the donor GDP-Man.

As the most structurally complex example, we modeled a glycosylphosphatidylinositol (GPI)-anchored protein (Kinoshita 2020) using glypican-1 (GPC1) as the scaffold (Fig. 4e and 4j). The phosphatidylinositol (PI) anchor was assembled from glycerol (GOL), palmityl alcohol (PL3), stearic acid (STE), and inositol-1-phosphate (IPD) (Kinoshita and Fujita 2016). The ethanolamine phosphate moiety was built using ETA and PO4. To establish a realistic covalent linkage between the GPI anchor and the C-terminal serine (Xu et al. 2023) of GPC1, we employed the *ptmType* function to define a more plausible amide bond length. While many top-ranked models exhibited conformational artifacts, reasonable models were obtained by increasing the number of random seeds. This challenging case illustrates the capacity of AF3 to accommodate a wide range of covalent bonding topologies across lipid, glycan, and protein domains.

Structural modeling of O-linked glycans in protein contexts

O-linked glycans attached to disordered or flexible regions of proteins pose significant obstacles for structural elucidation. Consequently, these glycoprotein architectures are underrepresented in experimental datasets and, by extension, likely underrepresented in the AF3 training set. To assess whether AF3 can generate feasible glycan conformations in such contexts, we systematically modeled several representative O-glycan structures.

Four common O-GalNAc-linked core structures (Wandall et al. 2021), including the branched core 2 and core 4 structures, were modeled on EPO and yielded stereochemically correct conformations (Fig. 5a-d). We further extended core 1 into a common disialylated configuration (Fig. 5e) and core 2 into a glycosaminoglycan (GAG) keratan sulfate (KS) (Wu et al. 2024) (Fig. 5f). Both extensions were modeled and exhibited chemically and conformationally plausible geometries.

We next examined the O-Xyl linker motif that initiates heparan sulfate (HS) and chondroitin sulfate (CS) biosynthesis (Lin et al. 2024). AF3 can model HS-modified GPC1 (Fig. 5g) and CS-modified bikunin (Fig. 5h), albeit with credible conformers emerging from lower-ranked predictions, similar to the behavior observed for GPI-anchored models. Another structurally complex example was the extended O-Man core M3 glycan (Fig. 5i), which serves as the glycan core for matriglycan biosynthesis (Sheikh et al. 2022). This structure incorporates a non-canonical, non-cyclic moiety, ribitol phosphate, within the glycan backbone (Praisman et al. 2016). AF3 was able to resolve both the linkages and overall topology of this structure.

Collectively, these findings suggest that AF3 can be a valuable tool in generating plausible conformations of O-linked glycans, despite the limited representation of such structures

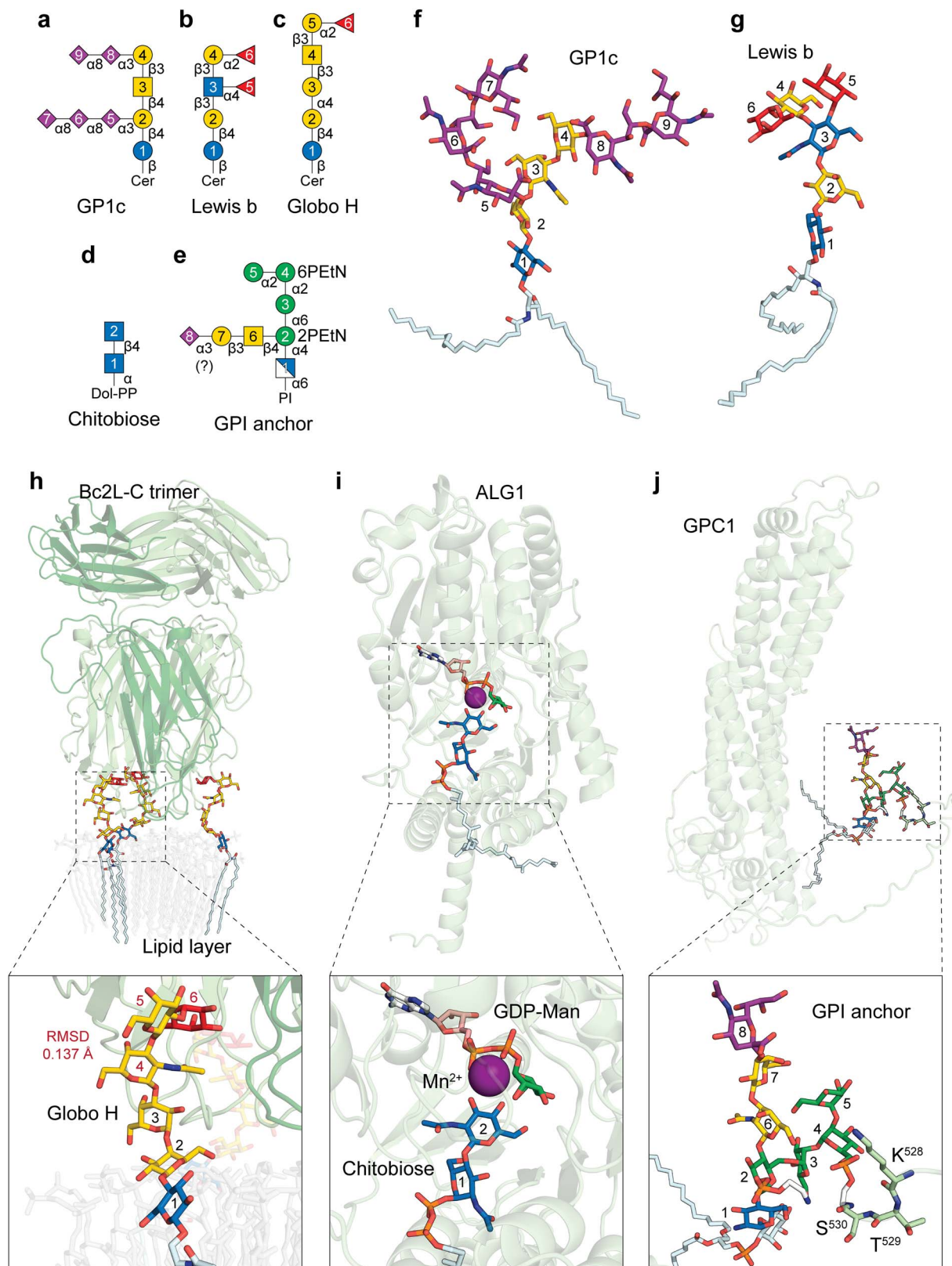


Fig. 4. Gallery of lipid-linked glycans modeled by AlphaFold 3. **a-e**) SNFG cartoon representations of lipid-linked glycans. **f**) Ganglio-series glycosphingolipid (GSL) GP1c (panel **a**) featuring a β -linkage between Glc and ceramide (Cer). **g**) Lacto-series GSL modified with a Lewis b (Le^b) blood group antigen (panel **b**). **h**) Globo-series GSL Globo H (panel **c**) modeled in complex with *B. cenocepacia* lectin Bc2Lc; Bc2Lc homotrimer binds to three Globo H molecules. A mock lipid layer was modeled to orient the lipid tail positioning. The root-mean-square deviation (RMSD) of residues 4–6 (colored in red) between the AF3 model and 6TIG is annotated. **i**) Dolichol pyrophosphate (dol-PP)-linked chitobiose (panel **d**) forming Michaelis complex with GDP-Man, Mn²⁺, and *H. sapiens* mannosyltransferase ALG1. Notably, chitobiose is linked to dol-PP with an α anomeric configuration. **j**) Extended glycosylphosphatidylinositol (GPI) anchor (panel **e**) linked to *H. sapiens* glypican-1 (GPC1) through phosphoethanolamine (PETN) and the C-terminal serine (S⁵³⁰). The sialyl linkage has not been determined; an α 2,3-sialic acid was modeled provisionally.

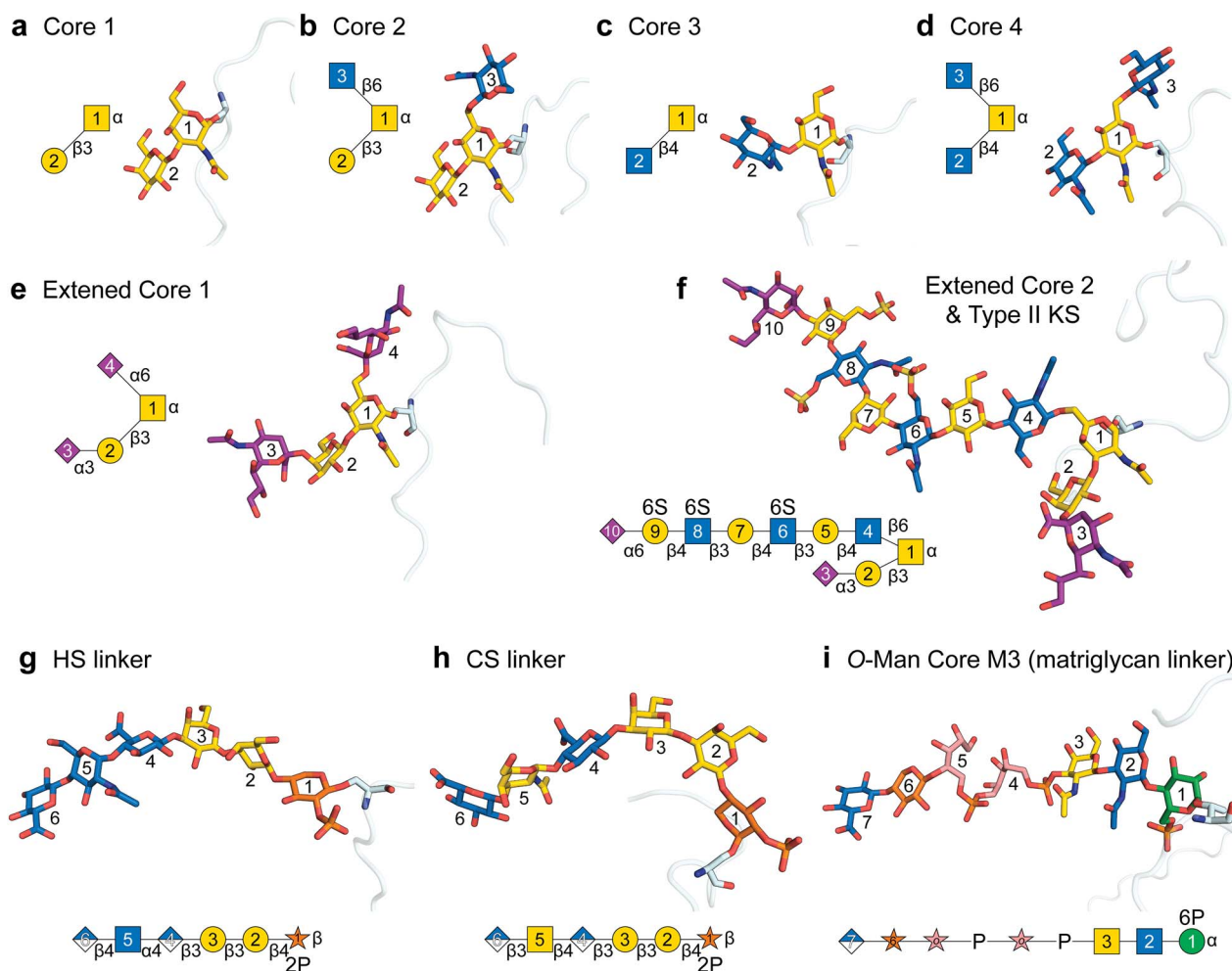


Fig. 5. AlphaFold 3 modeling of diverse O-linked glycans. **a-d**) Core 1–4 O-GalNAc glycans modeled on *Homo sapiens* erythropoietin (EPO); only the loop regions are displayed to highlight the glycans. **e**) Extended disialyl core 1 structure on EPO. **f**) Extended core 2 glycan featuring keratan sulfate (KS) on EPO, with 6-O-sulfation on GlcNAc and Gal. **g**) Heparan sulfate (HS) linker modeled on *H. sapiens* glypican-1 (GPC1), including 2-O-phosphorylation on Xyl. **h**) Chondroitin sulfate (CS) linker modeled on *H. sapiens* bikunin, also containing 2-O-phosphorylation on Xyl. **i**) Extended O-Man Core M3 glycan modeled on *H. sapiens* α -dystroglycan with 6-O-phosphorylation on Man.

in the training set, particularly when supplemented by input optimization and increased sampling.

Accurate syntax enables modeling of biologically complex glycoprotein architectures

AF3 can model protein-glycan complexes with low computational cost and rapid turnaround, making it feasible to investigate larger and more intricate glycoprotein assemblies. To demonstrate its broader applicability, we modeled CD22, also known as sialic acid-binding immunoglobulin-like lectin 2 (Siglec-2), a transmembrane lectin comprising six immunoglobulin constant (IgC)-like domains and a membrane-distal immunoglobulin variable (IgV)-like lectin domain that projects beyond the cell-surface glycocalyx (Macauley et al. 2014; Duan and Paulson 2020) (Fig. 6d). Siglec-2 contains four N-glycosylation sites on the IgV domain and seven additional sites distributed across the IgC domains (Wasim et al. 2019). While the IgV domain mediates recognition of α 2,6-sialylated glycans to activate downstream intracellular signaling (Duan and Paulson 2020), its binding pocket is frequently masked by cis-interactions with α 2,6-sialylated glycans on adjacent Siglec-2 molecules (Razi and

Varki 1998; Collins et al. 2004; Macauley et al. 2015; Wasim et al. 2019). This masking can be disrupted upon encountering high-avidity or high-affinity ligands, such as 6-O-sulfo-6'-sialyl LacNAc (Collins et al. 2006; Macauley et al. 2015; Wu et al. 2023; Ma et al. 2024; Jung et al. 2025).

To investigate this behavior, we modeled two Siglec-2 molecules, each fully modified with eleven G2S2 N-glycans (Fig. 6a). The resulting structure (Fig. 6c) recapitulated features consistent with prior negative-stain electron microscopy (EM) and small-angle X-ray scattering (SAXS) studies (Ereño-Orbea et al. 2017), including a $\sim 120^\circ$ angular displacement between the IgV-IgC2 and IgC3-IgC6 domains. The C-terminal intracellular tails were erroneously folded back toward IgC6, an artifact resulting from the lack of an explicit membrane boundary for the model.

The two IgV lectin domains displayed cis-interactions in which G2S2 glycans at Asn67 of one Siglec-2 interacted with the opposing IgV domain (Fig. 6e). The glycan-lectin binding pose aligned closely with the crystallographic structure (RMSD 0.292 Å, PDB: 5VKM); notably, the crystal structure includes only a terminal disaccharide and comprises just the IgV-IgC2 domains, with five N-glycosylation

structure of PmHS2 was solved by cryo-EM with bound 2-O-sulfated HS 5-mer, UDP, and Mn^{2+} (8VIW (Stancanelli et al. 2024), PDB release 2024 July 24, resolution 3.30 Å, Fig. S5a and b). AF3 modeled the dimeric complex with UDP-GlcNAc, Mn^{2+} , and HS as a Michaelis complex (Fig. S5c) showing good agreement with the empirical structure (RMSD: 0.360 Å; Fig. S5d).

The G1 domain of aggrecan shows hyaluronan (HA) binding activity and the crystal structure of the protein complex with bound HA 10-mer was determined (9DFE (Otsuka et al. 2025), PDB release 2025 April 30, resolution 2.59 Å, Fig. S6a and b). The non-reducing end unsaturated GlcA, a lyase-derived product, was excluded from the comparison. The AF3 model recapitulated the observed binding pose for the remainder of the bound HA product with good agreement (RMSD: 0.700 Å; Fig. S6c and d).

PP2 is a plant lectin recognizing GlcNAc moiety on a glycan chain. The structure of chitotriose-bound PP2 was characterized (7W4B (Bobbili et al. 2023), PDB release 2023 March 8, resolution 2.50 Å, Fig. S7a and b). The AF3 model replicated the binding pose with high accuracy (RMSD: 0.252 Å; Fig. S7c and d).

We further broadened our filtering criteria to include newly released protein structures that contain homologous sequences or are complexed with ligands not represented in the AlphaFold 3 training set. Examples include *Bacteroides thetaiotamicron* endoglycosidase BT1285 (Sastre et al. 2024), *Bacteroides ovatus* polysaccharide lyase family 38 (BoPL38), *Pedobacter terrae* glucanase (Tandrup et al. 2025), *Bacillus circulans* xylanase (PDB: 7WWC), *Escherichia coli* fimbrial adhesin FimH (Krammer et al. 2023), and botulinum neurotoxin A (PDB: 8RVG).

The endoglycosidase BT1285 was crystallized with an M9 N-glycan (8U48 (Sastre et al. 2024), PDB release 2024 May 29, resolution 1.90 Å, Fig. S8a and b), showing extensive interaction with glycan residues 1 through 7, where the cleavage site is between residues 1 and 2. AF3 reproduced the binding pose with high accuracy (RMSD: 0.324 Å; Fig. S8c and d).

The crystal structure of BoPL38 was determined with a bound alginate substrate (-4-D-mannuronic acid (ManA)- β -4) (9FHU (Tandrup et al. 2025), PDB release 2025 July 9, resolution 2.09 Å, Fig. S9a and b). We modeled BoPL38 with the same alginate 4-mer (Fig. S9b) and the AF3 model (Fig. S9c) aligns well with the binding pose of residues 2 to 4 within the active site, where the cleavage occurs between residues 3 and 4 (RMSD: 0.261 Å; Fig. S9d).

P. terrae glucanase was crystallized with laminarin (7WWC, PDB release 2023 February 15, resolution 2.20 Å, Fig. S10a), revealing a cleavage site between glycan residues 2 and 3 (Fig. S10b). As no accompanying publication was available, we highlighted side chains for potential protein-glycan interactions within 4 Å. The AF3 model demonstrated good agreement with the crystal structure (RMSD 0.214 Å; Fig. S10c and d).

The catalytically inactive *B. circulans* xylanase was co-crystallized with xylotriose (8QXY, PDB release 2024 August 21, resolution 1.41 Å, Fig. S11a and b). Extensive contacts were observed between the enzyme and all three glycan residues. AF3 successfully reproduced the binding configuration (RMSD 0.386 Å; Fig. S11c and d).

The crystal structure of *E. coli* fimbrial adhesin FimH-M3F N-glycan complex was characterized (7BHD (Krammer et al. 2023), PDB release 2022 July 20, resolution 1.40 Å, Fig. S12a

and b) showing specific interactions with the 3-arm of the glycan, while the solvent-exposed core and fucose residues remained flexible. The AF3 model reflected this binding pose well (RMSD 0.189 Å; Fig. S12c and d).

Botulinum neurotoxin A (BoNT/A) mutant was crystallized in complex with GM1b ganglioside (8RVG, PDB release 2025 February 12, resolution 1.9 Å, Fig. S13a and b), highlighting interactions with residues near the reducing end. Amino acid residues within 4 Å of the glycan were used for comparison. The AF3 model reproduced the binding pose closely (RMSD 0.223 Å; Fig. S13c and d).

Together, these examples span diverse protein classes, including glycosyltransferases, glycoside hydrolases, lectins, and polysaccharide lyases. In each case, the corresponding modeled complex closely matches the bound pose of the glycan ligand observed in the crystal structure, despite these specific complexes not being present in the AF3 training set. These results suggest a broader applicability of AF3 in modeling glycan-protein complexes than previously anticipated. Notably, the modeling of complex polysaccharides such as heparan sulfate (HS), hyaluronic acid (HA), and alginate highlights the capability of AF3 to model such challenging polymer ligands when bound to enzymes and lectins.

However, we emphasize that glycan structure modeling in AF3 is highly context dependent. We observed numerous instances where modeled glycan structures did not maintain appropriate glycan stereochemistry, or the modeled complexes failed to replicate interactions present in the corresponding crystal structures. Therefore, while this approach offers a valuable tool for hypothesis generation in biologically relevant glycan-protein interactions, each model must be carefully evaluated, and appropriate structural and functional validation remains essential.

Discussion

AF3 presents new opportunities for ligand modeling, and our evaluation across multiple glycan classes demonstrates its potential as a powerful tool for developing structural hypotheses that can be further tested empirically. While our study focused on human glycans, the modeling strategies, input syntax, and curated templates provided here can be generalized to glycan systems across all domains of life. Notably, we present pre-built input files for structurally complex ligands, including GPI anchors, featuring covalent linkages between glycan, lipid, and protein components.

Among the available ligand input formats, SMILES is the most straightforward, requiring minimal pre- or post-processing. However, when applied to stereochemically complex molecules such as glycans, SMILES frequently fails to reproduce correct conformations, even for relatively simple linear oligosaccharides like LNnT. Converting SMILES or structural files into *userCCD* using tools such as *rdkit_utils* offers more control over conformer generation and optimization, but this approach also lacks reliability in glycan modeling. This may be attributed to the reliance of AF3 on RDKit (<https://www.rdkit.org>) for internal ligand handling, which struggles with the conformational diversity and stereochemical precision required for glycans.

By contrast, idealized coordinates (Westbrook et al. 2015) generated through proprietary cheminformatics engines such as Corina (Molecular Networks) (Gasteiger et al. 1990;

Sadowski et al. 1994) or OMEGA (OpenEye) (Hawkins et al. 2010; Hawkins and Nicholls 2012), are routinely used in novel PDB ligand deposition and exhibit higher structural fidelity. These coordinates are incorporated into CCD entries that undergo rigorous biocuration under the Worldwide Protein Data Bank (wwPDB) (Young et al. 2018), ensuring consistency and accuracy. Accordingly, idealized CCD input for monosaccharide building blocks with appropriate anomeric configurations coupled with *bondedAtomPairs* (BAP) syntax yielded the most reliable and stereochemically valid glycan conformations in our evaluation of structural models.

Caveats and limitations

The use of *bondedAtomPairs* (BAP) syntax requires both glycobiological expertise to define the correct glycan types and linkages and a strong foundation in glycochemistry to interpret conformational plausibility. We addressed the former by compiling input templates for common glycan motifs, and the latter by systematically evaluating the structural plausibility of modeled glycans presented in this study. Analytic tools such as GlyProbity (Woods 2019) and Privateer (Dialpuri et al. 2023) are available for assessing ring puckering and glycosidic torsions based on CCD identifiers, which may increase the throughput of model quality assessment. However, these tools were originally developed for validating models derived from X-ray crystallography or cryo-electron microscopy, and their applicability to predictions from AF3 remains to be fully validated. A more thorough evaluation across diverse monosaccharides and glycan types is warranted, especially with an emphasis on comparison with empirically derived structures to validate the computational models. We anticipate that AF3 will expand access to modeling of glycan-protein complexes and generation of structural hypotheses for biological function. However, like any modeling approach, extreme care must be taken in overinterpretation of modeling results and all models should be further tested and validated by direct empirical experimentation.

If a glycan is modeled into an incorrect conformation, even a single anomeric or epimeric misassignment can significantly impact the interpretation of structural and biological function. However, AF3 currently lacks discernible metrics to penalize such conformational issues in glycan models. The predicted local-distance difference test (pLDDT) (Mariani et al. 2013), commonly used to assess protein confidence, is unreliable for ligands. We observed instances in which glycans displayed incorrect linkages, epimers, or conformations while retaining high pLDDT scores, particularly when located in catalytic pockets. Conversely, when glycans extend into solvent as post-translational modifications, pLDDT scores often decrease with distance from the protein core, even when conformations are plausible (Fig. S14). Other AF3 scoring metrics such as predicted aligned error (PAE) (Evans et al. 2022), predicted template modeling (pTM) (Zhang and Skolnick 2004), interface PTM (ipTM) (Evans et al. 2022), and the sample ranking score (Abramson et al. 2024) also fail to capture glycan-specific conformational errors. RMSD comparisons with crystallographic or energy-minimized structures are similarly insufficient, as glycan flexibility can obscure subtle but critical deviations. Although glycosidic torsion angles offer more direct insight, comprehensive reference data exist primarily for *N*-glycans (e.g. GlyProbity and Privateer), limiting broader evaluation of glycan model accuracy.

We also observed that certain glycan linkages, particularly those underrepresented in the PDB or the AF3 training set, required higher sampling (i.e. more seeds) to recover conformationally plausible models. Interestingly, false conformations were often found among top sample ranking score models, whereas effective geometries frequently appeared in lower-ranked outputs. It is unclear how chirality penalties are enforced in AF3, especially for stereochemically complicated glycans. As such, we recommend evaluating both high- and low-scoring models to identify structurally plausible predictions when encountering difficulties. This approach proved useful in modeling GPI anchors and HS/CS linkers. However, when models were extended for HS or CS polymers beyond the linker region, no models with correct configurations were recovered. Common artifacts included incorrect anomeric configurations, misassigned epimers, and high-energy ring conformations (data not shown). In many cases, only the regions of HS/CS chains directly interacting with target proteins yielded plausible structures. Modeling specific degrees of polymerization may mitigate this issue.

In the case of matriglycan, which consists of [Xyl- α 1,3-GlcA- β 1,3] disaccharide repeats (Sheikh et al. 2022), AF3 consistently misassigns α -xylose as β , even in contexts where structures of matriglycan-bound proteins exist in the PDB (e.g. laminin (Briggs et al. 2016), Lassa virus glycoprotein (Katz et al. 2022)). Given the limited natural occurrence and greater conformational diversity of α -xylose compared to β -xylose, it may be beneficial to test SMILES or *userCCD* input formats when ambiguity in stereochemical encoding exists. While idealized coordinates in the CCD offer necessary flexibility to accommodate electron density maps, this same flexibility can be a double-edged sword for AF3. A notable advantage of this adaptability was observed when modeling glycone interactions of an M9 glycan substrate with MAN1A1, where the bound α -Man residue appropriately adopted a high-energy transition state conformation. Such conformational plasticity offers opportunities for mechanistic insights into enzyme catalysis. Nonetheless, this conformational freedom can also obscure critical features, such as incorrect assignment of anomeric configurations and epimeric orientations, thereby complicating the accurate prediction of glycan structures.

We also attempted to model a more complex case of glycoenzyme interactions with the conserved Fc *N*-glycan on immunoglobulin G (IgG). There are numerous structures in PDB of IgG or Fc domains containing complex type biantennary *N*-glycan structures tightly packed between Fc chains, effectively inaccessible to glycan-processing enzymes. Attempts to model glycan processing enzymes such as MGAT1, MAN2A1, MGAT2, FUT8, and B4GALT1 in complex with the Fc domain *N*-glycan structure were unsuccessful. The positions of the glycan structures remained consistent with crystallographic observations within the interior of the Fc homodimer core. In each case, AF3 prioritized interactions between the glycan processing enzymes and the Fc polypeptide rather than the *N*-glycan structures (data not shown). This suggests that the AF3 modeling was strongly biased by the abundant Fc domain crystal structures in the PDB training set. As a result, AF3 presently lacks the capacity to capture the dynamic and transient positioning of glycans and glycan-bearing C'E loops in the Fc domain (Borrok et al. 2012), a feature that has been experimentally validated by nuclear magnetic resonance (NMR) (Barb and

Prestegard 2011) and in IgG Fc-endoglycosidase S (EndoS) co-crystal structures (Trastoy et al. 2018). Efforts to model an IgG Fc-endoglycosidase S (EndoS) complex with the glycan positioned in the active site were also unsuccessful, despite this complex presumably being present in the training dataset for modeling. In contrast, AF3 was able to recapitulate key aspects of the interaction between glycosylated IgG Fc and glycosylated Fc γ RIIIa (CD16a) (Falconer et al. 2018) (data not shown), consistent with known binding modes. These observations underscore the importance of critically evaluating AF3-generated models in a context-dependent manner.

AF3 does not incorporate an explicit energy function; therefore, the generated models do not represent thermodynamic equilibria nor do they explicitly represent energetically favored conformations. Furthermore, the predicted structures should be interpreted only as static snapshots, similar to those obtained from empirical methods such as X-ray crystallography, which contribute substantially to the PDB dataset used in AF3 training. The AF3 models of bound glycans commonly deviate from the equivalent crystal structure complexes when the glycan termini extend into solvent. These differences in modeled conformation correlate with lower pLDDT scores for the modeled ligands and higher B-factors for the same regions of the ligand in the crystal structures indicating greater disorder for solvent oriented ligand subregions (Fig. S14). Users interested in glycan dynamics or equilibrium conformational sampling are advised to employ complementary computational methods, such as molecular dynamics simulations or ensemble-based tools like GlycoShield (Tsai et al. 2024) or GlycoShape (Ives et al. 2024), which provide structural representations grounded in exhaustive conformational sampling.

Unanticipated and versatile applications in glycoscience and beyond

The relative scarcity of glycan structural data in the PDB, compared to protein structural data, would seemingly pose a challenge when training AF3 for glycan structural modeling. Thus, it would be reasonable to assume that effective glycan modeling with AF3 would be limited to “pre-trained” structures or simple models constructed from the fragmentary glycan information available in the PDB. Surprisingly, this assumption appears to be incorrect. AF3 is capable of modeling a wide range of glycan-binding proteins, including glycosyltransferases, glycoside hydrolases, lectins, and polysaccharide lyases, containing bound glycan structures that closely resemble crystallographic complexes not included in its training set. AF3 has clearly “learned” to generalize its predictions and generate plausible glycan models that extend well beyond the scope of its original training data. It remains unclear whether this capability stems from the structural constraints embedded in the monosaccharide building block CCD files, the partial glycan structures in the training data, or from patterns derived from non-carbohydrate chemical moieties included in its datasets. While this ability may seem counterintuitive given the limited scale of glycan training data, the resulting models offer a powerful tool for hypothesis generation.

Our investigations underscore the broad applicability of AF3 in modeling glycans, both as post-translational modifications and as free ligands. This capability opens new pathways for translational glycobiology applications. For instance, Bc2L-C, a dual-domain lectin from *B. cenocepacia*, can be

modeled in complex with both glycolipid and glycoprotein ligands to better understand its multivalent binding and role in infection (Sulák et al. 2011). Similarly, AF3-generated models of cis-interactions between CD22 molecules may offer insights into dynamic regulatory mechanisms that are difficult to capture through crystallography alone.

Although further improvements in accuracy are needed for antibody–antigen modeling, AF3 shows promise in the structural prediction of anti-glycan antibodies and glycan-containing epitopes. Given that most biologics are glycoproteins targeting glycoprotein antigens in vivo, AF3 offers a valuable framework for evaluating both the structural consequences of glycoform variation and the functional effects of glycosylation on therapeutic efficacy.

We have also noted that the AF3 modeling approach could be extended beyond glycoconjugates to other complex molecule types. This is made possible by the use of other molecular building blocks encoded with idealized CCD codes and linked via *bondedAtomPairs* (BAP) syntax. The simplicity of the JSON scripting strategy for input into AF3 enables an accessible, flexible, and universal workflow for assembling stereochemically complex macromolecules from the molecular elements found in the CCD library. Consequently, this approach significantly broadens future capabilities for modeling biologically important and structurally complex molecular interactions.

In conclusion, although structural validation using orthogonal methods remains essential, AF3 presents a transformative platform for modeling glycan-mediated interactions across proteins, lipids, and carbohydrates. With thoughtful syntax planning and critical evaluation, AF3 has the potential to dramatically streamline and accelerate research in the field of glycoscience.

Methods

AlphaFold 3 version

AlphaFold 3 Version 3.0.1 (available at <https://github.com/google-deepmind/alphafold3>) was run as a Singularity container on the Sapelo2 high-performance computing cluster at the Georgia Advanced Computing Resource Center (GACRC), University of Georgia. Modeling was performed using NVIDIA A100 or H100 GPUs. The AlphaFold 3 model parameter permission file was granted by the DeepMind team.

AlphaFold 3 input JSON file

A full description of the input JSON file structure is available on AlphaFold 3 GitHub repository (<https://github.com/google-deepmind/alphafold3/blob/main/docs/input.md>). The initial JSON template used in this study was adapted from AFusion (<https://github.com/Hanziwww/AlphaFold3-GUI>). All modeling was conducted using version 2 of the JSON input format.

SMILES input and glycan energy minimization

Energy minimized glycan PDBs were generated using GLYCAM Carbohydrate Builder web tool (<https://glycam.org/>) (Grant et al. 2025) and subsequently converted to SMILES format through Open Babel web-based converter (<https://www.cheminfo.org/Chemistry/Cheminformatics/FormatConverter/index.html>) (O’Boyle et al. 2011).

Conversion of SMILES and PDB to *userCCD* format

The *rdkit_utils* Python script from the AlphaFold 3 repository (https://github.com/google-deepmind/alphafold3/blob/main/src/alphafold3/data/tools/rdkit_utils.py) was used to convert SMILES strings or PDB files into CCD-like mmCIF files for use as *userCCD* inputs. *rdkit_utils* depends on RDKit (RDKit: Open-source cheminformatics. <https://www.rdkit.org>. DOI: 10.5281/zenodo.15286010).

CCD library for AlphaFold 3 input

Chemical Component Dictionary (CCD) codes and associated mmCIF files were retrieved and manually curated from PDBeChem database (<https://www.ebi.ac.uk/pdbe-srv/pdbechem/>) for compatibility with AlphaFold 3 input requirements.

Protein sequences for AlphaFold 3 input

Protein sequences were obtained from UniProt (<https://www.uniprot.org/>) using the following accession numbers: *M. musculus* MAN1A1 (P45700), *H. sapiens* MGAT2 (Q10469), *H. sapiens* ST6GAL1 (P15907), *H. sapiens* EPO (P01588), *B. cenocepacia* Bc2L-C (B4EH86), *H. sapiens* ALG1 (Q9BT22), *H. sapiens* GPC1 (P35052), *H. sapiens* Bikunin (P02760), *H. sapiens* α -dystroglycan (Q14118), *H. sapiens* Siglec-2 (P20273), *H. sapiens* B3GALT5 (Q9Y2C3), *P. multocida* heparosan synthase 2 (Q5SGE1), *H. sapiens* aggrecan (P16112), *C. sativus* phloem protein 2 (Q8LK69), *B. thetaiotamicron* BT1258 (Q8A889), *B. ovatus* polysaccharide lyase family 38 (A0A5M5BWR5), *P. terrae* glucanase (A0A1G7XNR7), *B. circulans* xylanase (P09850), *E. coli* fimbrial adhesin FimH (P08191), and *C. botulinum* neurotoxin A (P0DPI0).

Glycan and protein visualization

Glycan cartoon representations were generated using GlycoWorkbench 2.1 (stable build 146) (Ceroni et al. 2008), with custom modifications. Protein structures were visualized, aligned and rendered using PyMOL version 3.1.3.1. Root-mean-square deviation (RMSD) was calculated by PyMol *align* function.

Acknowledgments

We thank Shan-Ho Tsai and Jordan Utley at the Georgia Advanced Computing Resource Center (GACRC), University of Georgia, for their assistance with the installation and maintenance of AlphaFold 3.

Author contributions

Chin Huang (Conceptualization [equal], Data curation [equal], Formal analysis [equal], Investigation [equal], Methodology [equal], Resources [equal], Software [equal], Validation [equal], Visualization [equal], Writing—original draft [equal], Writing—review & editing [equal]), Natarajan Kannan (Writing—review & editing [equal]), Kelley W Moremen (Conceptualization [equal], Formal analysis [equal], Funding acquisition [equal], Investigation [equal], Project administration [equal], Supervision [equal], Writing—original draft [equal], Writing—review & editing [equal]).

Supplementary material

Supplementary material is available at *Glycobiology Journal* online.

Funding

This work was supported by U.S. National Science Foundation Bio-Foundation: Glycoscience Research, Education and Training [BioF:GREAT NSF: 2400220] and National Institutes of Health [GM154846 to K.W.M.].

Conflict of interest

None declared.

Data availability

All JSON input files and *userCCD* examples used in this study are provided in the Supplementary Document 2. All AF3 models and metadata are also available from ModelArchive using the following link: <https://www.modelarchive.org/doi/10.5452/ma-af3glycan>. Availability of predicted structural models is subject to the AlphaFold 3 Output Terms of Use (https://github.com/google-deepmind/alphafold3/blob/main/OUTPUT_TERMS_OF_USE.md).

References

- Abramson J et al. 2024. Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature*. 630:493–500. <https://doi.org/10.1038/s41586-024-07487-w>.
- Adams TM, Zhao P, Chapla D, Moremen KW, Wells L. 2022. Sequential in vitro enzymatic N-glycoprotein modification reveals site-specific rates of glycoenzyme processing. *J Biol Chem*. 298:102474. <https://doi.org/10.1016/j.jbc.2022.102474>.
- Agirre J. 2017. Strategies for carbohydrate model building, refinement and validation. *Acta Crystallogr D Struct Biol*. 73:171–186. <https://doi.org/10.1107/S2059798316016910>.
- Allain EP, Rouleau M, Lévesque E, Guillemette C. 2020. Emerging roles for UDP-glucuronosyltransferases in drug resistance and cancer progression. *Br J Cancer*. 122:1277–1287. <https://doi.org/10.1038/s41416-019-0722-0>.
- Barb AW, Prestegard JH. 2011. NMR analysis demonstrates immunoglobulin G N-glycans are accessible and dynamic. *Nat Chem Biol*. 7:147–153. <https://doi.org/10.1038/nchembio.511>.
- Barb AW, Brady EK, Prestegard JH. 2009. Branch-specific sialylation of IgG-fc glycans by ST6Gal-I. *Biochemistry*. 48:9705–9707. <https://doi.org/10.1021/bi901430h>.
- Bermeo R, Bernardi A, Varrot A. 2020. BC2L-C N-terminal lectin domain complexed with Histo blood group oligosaccharides provides new structural information. *Molecules*. 25:248. <https://doi.org/10.3390/molecules25020248>.
- Bobbili KB, Sivaji N, Priya B, Suguna K, Surolia A. 2023. Structure and interactions of the phloem lectin (phloem protein 2) Cus17 from Cucumis sativus. *Structure*. 31:464–479.e5. <https://doi.org/10.1016/j.str.2023.02.008>.
- Borrok MJ, Jung ST, Kang TH, Monzingo AF, Georgiou G. 2012. Revisiting the role of glycosylation in the structure of human IgG fc. *ACS Chem Biol*. 7:1596–1602. <https://doi.org/10.1021/cb300130k>.
- Briggs DC et al. 2016. Structural basis of laminin binding to the LARGE glycans on dystroglycan. *Nat Chem Biol*. 12:810–814. <https://doi.org/10.1038/nchembio.2146>.
- Case DA et al. 2005. The amber biomolecular simulation programs. *J Comput Chem*. 26:1668–1688. <https://doi.org/10.1002/jcc.20290>.
- Ceroni A et al. 2008. GlycoWorkbench: a tool for the computer-assisted annotation of mass spectra of glycans. *J Proteome Res*. 7:1650–1659. <https://doi.org/10.1021/pr7008252>.
- Collins BE et al. 2004. Masking of CD22 by cis ligands does not prevent redistribution of CD22 to sites of cell contact. *Proc Natl Acad Sci USA*. 101:6104–6109. <https://doi.org/10.1073/pnas.0400851101>.
- Collins BE et al. 2006. High-affinity ligand probes of CD22 overcome the threshold set by cis ligands to allow for binding, endocytosis,

- and killing of B cells. *J Immunol.* 177:2994–3003. <https://doi.org/10.4049/jimmunol.177.5.2994>.
- Cremer D, Pople JA. 1975. General definition of ring puckering coordinates. *J Am Chem Soc.* 97:1354–1358. <https://doi.org/10.1021/ja00839a011>.
- Dashti H et al. 2020. Probabilistic identification of saccharide moieties in biomolecules and their protein complexes. *Sci Data.* 7:210. <https://doi.org/10.1038/s41597-020-0547-y>.
- Dialpuri JS et al. 2023. Analysis and validation of overall N-glycan conformation in privateer. *Acta Crystallogr D Struct Biol.* 79:462–472. <https://doi.org/10.1107/S2059798323003510>.
- Discovery C et al. 2024. Chai-1: decoding the molecular interactions of life. *bioRxiv*:2024.2010.2010.615955.
- Duan S, Paulson JC. 2020. Siglecs as immune cell checkpoints in disease. *Annu Rev Immunol.* 38:365–395. <https://doi.org/10.1146/annurev-immunol-102419-035900>.
- Ereño-Orbea J et al. 2017. Molecular basis of human CD22 function and therapeutic targeting. *Nat Commun.* 8:764. <https://doi.org/10.1038/s41467-017-00836-6>.
- Evans R et al. 2022. Protein complex prediction with AlphaFold-multimer. *bioRxiv*:2021.2010.2004.463034.
- Fadda E. 2022. Molecular simulations of complex carbohydrates and glycoconjugates. *Curr Opin Chem Biol.* 69:102175. <https://doi.org/10.1016/j.cbpa.2022.102175>.
- Falconer DJ, Subedi GP, Marcella AM, Barb AW. 2018. Antibody Fucosylation lowers the Fc γ RIIIa/CD16a affinity by limiting the conformations sampled by the N162-glycan. *ACS Chem Biol.* 13:2179–2189. <https://doi.org/10.1021/acscchembio.8b00342>.
- Feng Z et al. 2004. Ligand depot: a data warehouse for ligands bound to macromolecules. *Bioinformatics.* 20:2153–2155. <https://doi.org/10.1093/bioinformatics/bth214>.
- Garron ML, Cygler M. 2010. Structural and mechanistic classification of uronic acid-containing polysaccharide lyases. *Glycobiology.* 20:1547–1573. <https://doi.org/10.1093/glycob/cwq122>.
- Gasteiger J, Rudolph C, Sadowski J. 1990. Automatic generation of 3D-atomic coordinates for organic molecules. *Tetrahedron Computer Methodology.* 3:537–547. [https://doi.org/10.1016/0898-5529\(90\)90156-3](https://doi.org/10.1016/0898-5529(90)90156-3).
- Grant OC, Woods RJ. 2014. Recent advances in employing molecular modelling to determine the specificity of glycan-binding proteins. *Curr Opin Struct Biol.* 28:47–55. <https://doi.org/10.1016/j.sbi.2014.07.001>.
- Grant OC et al. 2025. Generating 3D models of carbohydrates with GLYCAM-web. *bioRxiv*:2025.2005.2008.652828.
- Hawkins PC, Nicholls A. 2012. Conformer generation with OMEGA: learning from the data set and the analysis of failures. *J Chem Inf Model.* 52:2919–2936. <https://doi.org/10.1021/ci300314k>.
- Hawkins PC, Skillman AG, Warren GL, Ellingson BA, Stahl MT. 2010. Conformer generation with OMEGA: algorithm and validation using high quality structures from the protein databank and Cambridge structural database. *J Chem Inf Model.* 50:572–584. <https://doi.org/10.1021/ci100031x>.
- He X et al. 2024. Highly accurate carbohydrate-binding site prediction with DeepGlycanSite. *Nat Commun.* 15:5163. <https://doi.org/10.1038/s41467-024-49516-2>.
- Ives CM et al. 2024. Restoring protein glycosylation with GlycoShape. *Nat Methods.* 21:2117–2127. <https://doi.org/10.1038/s41592-024-02464-7>.
- Jumper J et al. 2021. Highly accurate protein structure prediction with AlphaFold. *Nature.* 596:583–589. <https://doi.org/10.1038/s41586-021-03819-2>.
- Jung J et al. 2025. Understanding the glycosylation pathways involved in the biosynthesis of the Sulfated glycan ligands for Siglecs. *ACS Chem Biol.* 20:386–400. <https://doi.org/10.1021/acscchembio.4c00677>.
- Kadirvelraj R et al. 2018. Human N-acetylglucosaminyltransferase II substrate recognition uses a modular architecture that includes a convergent exosite. *Proc Natl Acad Sci USA.* 115:4637–4642. <https://doi.org/10.1073/pnas.1716988115>.
- Karaveg K et al. 2005. Mechanism of class 1 (glycosylhydrolase family 47) {alpha}-mannosidases involved in N-glycan processing and endoplasmic reticulum quality control. *J Biol Chem.* 280:16197–16207.
- Katz M et al. 2022. Structure and receptor recognition by the Lassa virus spike complex. *Nature.* 603:174–179. <https://doi.org/10.1038/s41586-022-04429-2>.
- Kinoshita T. 2020. Biosynthesis and biology of mammalian GPI-anchored proteins. *Open Biol.* 10:190290. <https://doi.org/10.1098/rsob.190290>.
- Kinoshita T, Fujita M. 2016. Biosynthesis of GPI-anchored proteins: special emphasis on GPI lipid remodeling. *J Lipid Res.* 57:6–24. <https://doi.org/10.1194/jlr.R063313>.
- Kirschner KN et al. 2008. GLYCAM06: a generalizable biomolecular force field. *Carbohydrates. J Comput Chem.* 29:622–655.
- Krammer EM et al. 2023. Structural insights into a cooperative switch between one and two FimH bacterial adhesins binding pauci- and high-mannose type N-glycan receptors. *J Biol Chem.* 299:104627. <https://doi.org/10.1016/j.jbc.2023.104627>.
- Krishna R et al. 2024. Generalized biomolecular modeling and design with RoseTTAFold all-atom. *Science.* 384:ead12528.
- Kuhn B et al. 2013. The structure of human α -2,6-sialyltransferase reveals the binding mode of complex glycans. *Acta Crystallogr D Biol Crystallogr.* 69:1826–1838. <https://doi.org/10.1107/S0907444913015412>.
- Lal A, Schutzbach JS, Forsee WT, Neame PJ, Moremen KW. 1994. Isolation and expression of murine and rabbit cDNAs encoding an alpha 1,2-mannosidase involved in the processing of asparagine-linked oligosaccharides. *J Biol Chem.* 269:9872–9881. [https://doi.org/10.1016/S0021-9258\(17\)36964-8](https://doi.org/10.1016/S0021-9258(17)36964-8).
- Lebrilla CB, Liu J, Widmalm G, Prestegard JH. 2022. Oligosaccharides and polysaccharides. In: Varki A, et al., editors. *Essentials of glycobiology*. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press. pp. 33–42
- Li H et al. 2022. Structure of the human GlcNAc-1-phosphotransferase $\alpha\beta$ subunits reveals regulatory mechanism for lysosomal enzyme glycan phosphorylation. *Nat Struct Mol Biol.* 29:348–356. <https://doi.org/10.1038/s41594-022-00748-0>.
- Lin PH et al. 2024. Solid-phase-supported chemoenzymatic synthesis and analysis of chondroitin Sulfate proteoglycan glycopeptides. *Angew Chem Int Ed Engl.* 63:e202405671. <https://doi.org/10.1002/anie.202405671>.
- Lo JM, Kung CC, Cheng TR, Wong CH, Ma C. 2025. Structure-based mechanism and specificity of human galactosyltransferase β 3GalT5. *J Am Chem Soc.* 147:10875–10885. <https://doi.org/10.1021/jacs.4c11724>.
- Lobsanov YD et al. 2002. Structure of penicillium citrinum alpha 1,2-mannosidase reveals the basis for differences in specificity of the endoplasmic reticulum and Golgi class I enzymes. *J Biol Chem.* 277:5620–5630. <https://doi.org/10.1074/jbc.M110243200>.
- Long F et al. 2017. AceDRG: a stereochemical description generator for ligands. *Acta Crystallogr D Struct Biol.* 73:112–122. <https://doi.org/10.1107/S2059798317000067>.
- Luo Y, Parmeggiani F. 2025. CLIMBS: assessing carbohydrate-protein interactions through a graph neural network classifier using synthetic negative data. *bioRxiv*:2025.2002.2027.640667.
- Ma S et al. 2024. Enzyme-sialylation-controlled chemical sulfation of glycan epitopes for decoding the binding of Siglec ligands. *J Am Chem Soc.* 146:29469–29480. <https://doi.org/10.1021/jacs.4c08817>.
- Macauley MS, Crocker PR, Paulson JC. 2014. Siglec-mediated regulation of immune cell function in disease. *Nat Rev Immunol.* 14:653–666. <https://doi.org/10.1038/nri3737>.
- Macauley MS et al. 2015. Unmasking of CD22 Co-receptor on germinal Center B-cells occurs by alternative mechanisms in mouse and man. *J Biol Chem.* 290:30066–30077. <https://doi.org/10.1074/jbc.M115.691337>.
- Mariani V, Biasini M, Barbato A, Schwede T. 2013. IDDT: a local superposition-free score for comparing protein structures and

- models using distance difference tests. *Bioinformatics*. 29: 2722–2728. <https://doi.org/10.1093/bioinformatics/btt473>.
- Mendoza F, Masgrau L. 2021. Computational modeling of carbohydrate processing enzymes reactions. *Curr Opin Chem Biol*. 61: 203–213. <https://doi.org/10.1016/j.cbpa.2021.02.012>.
- Meng L et al. 2013. Enzymatic basis for N-glycan sialylation: structure of rat α 2,6-sialyltransferase (ST6GAL1) reveals conserved and unique features for glycan sialylation. *J Biol Chem*. 288: 34680–34698. <https://doi.org/10.1074/jbc.M113.519041>.
- Moremen KW, Haltiwanger RS. 2019. Emerging structural insights into glycosyltransferase-mediated synthesis of glycans. *Nat Chem Biol*. 15:853–864. <https://doi.org/10.1038/s41589-019-0350-2>.
- Moremen KW, Trimble RB, Herscovics A. 1994. Glycosidases of the asparagine-linked oligosaccharide processing pathway. *Glycobiology*. 4:113–125. <https://doi.org/10.1093/glycob/4.2.113>.
- Moremen KW, Tiemeyer M, Nairn AV. 2012. Vertebrate protein glycosylation: diversity, synthesis and function. *Nat Rev Mol Cell Biol*. 13:448–462. <https://doi.org/10.1038/nrm3383>.
- O’Boyle NM et al. 2011. Open babel: an open chemical toolbox. *J Cheminform*. 3:33.
- Otsuka MY et al. 2025. Aggrecan immobilizes to perineuronal nets through hyaluronan-dependent and hyaluronan-independent binding activities. *J Biol Chem*. 301:108525. <https://doi.org/10.1016/j.jbc.2025.108525>.
- Passaro S et al. 2025. Boltz-2: towards accurate and efficient binding affinity prediction. *bioRxiv*:2025.2006.2014.659707.
- Perez S, Makshakova O. 2022. Multifaceted computational Modeling in Glycoscience. *Chem Rev*. 122:15914–15970. <https://doi.org/10.1021/acs.chemrev.2c00060>.
- Pinho SS, Alves I, Gaifem J, Rabinovich GA. 2023. Immune regulatory networks coordinated by glycans and glycan-binding proteins in autoimmunity and infection. *Cell Mol Immunol*. 20:1101–1113.
- Praissman JL et al. 2016. The functional O-mannose glycan on α -dystroglycan contains a phospho-ribitol primed for matriglycan addition. *Elife*. 5:e14473. <https://doi.org/10.7554/eLife.14473>.
- Prestegard JH. 2021. A perspective on the PDB’s impact on the field of glycobiology. *J Biol Chem*. 296:100556. <https://doi.org/10.1016/j.jbc.2021.100556>.
- Ramírez AS, Locher KP. 2023. Structural and mechanistic studies of the N-glycosylation machinery: from lipid-linked oligosaccharide biosynthesis to glycan transfer. *Glycobiology*. 33:861–872. <https://doi.org/10.1093/glycob/cwad053>.
- Ramírez AS, Kowal J, Locher KP. 2019. Cryo-electron microscopy structures of human oligosaccharyltransferase complexes OST-A and OST-B. *Science*. 366:1372–1375. <https://doi.org/10.1126/science.aaz3505>.
- Razi N, Varki A. 1998. Masking and unmasking of the sialic acid-binding lectin activity of CD22 (Siglec-2) on B lymphocytes. *Proc Natl Acad Sci USA*. 95:7469–7474. <https://doi.org/10.1073/pnas.95.13.7469>.
- Sadowski J, Gasteiger J, Klebe G. 1994. Comparison of automatic three-dimensional model builders using 639 X-ray structures. *J Chem Inf Comput Sci*. 34:1000–1008. <https://doi.org/10.1021/ci00020a039>.
- Sastre DE et al. 2024. Human gut microbes express functionally distinct endoglycosidases to metabolize the same N-glycan substrate. *Nat Commun*. 15:5123. <https://doi.org/10.1038/s41467-024-48802-3>.
- Schnaar RL, Gerardy-Schahn R, Hildebrandt H. 2014. Sialic acids in the brain: gangliosides and polysialic acid in nervous system development, stability, disease, and regeneration. *Physiol Rev*. 94: 461–518. <https://doi.org/10.1152/physrev.00033.2013>.
- Schnaar RL, Sandhoff R, Tiemeyer M, Kinoshita T. 2022. Glycosphingolipids. In: Varki A, et al., editors. *Essentials of glycobiology*. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press. pp. 129–140
- Seeberger PH. 2022. Monosaccharide diversity. In: Varki A, et al., editors. *Essentials of glycobiology*. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press. pp. 21–32
- Senior AW et al. 2020. Improved protein structure prediction using potentials from deep learning. *Nature*. 577:706–710. <https://doi.org/10.1038/s41586-019-1923-7>.
- Shao C et al. 2021. Modernized uniform representation of carbohydrate molecules in the protein data Bank. *Glycobiology*. 31:1204–1218.
- Sheikh MO et al. 2022. Cell surface glycan engineering reveals that matriglycan alone can recapitulate dystroglycan binding and function. *Nat Commun*. 13:3617.
- Smith BAH, Bertozzi CR. 2021. The clinical impact of glycobiology: targeting selectins, Siglecs and mammalian glycans. *Nat Rev Drug Discov*. 20:217–243. <https://doi.org/10.1038/s41573-020-00093-1>.
- Stancanelli E et al. 2024. Structural and functional analysis of Heparosan synthase 2 from *Pasteurella multocida* (PmHS2) to improve the synthesis of heparin. *ACS Catal*. 14:6577–6588. <https://doi.org/10.1021/acscatal.4c00677>.
- Stanley P, Moremen KW, Lewis NE, Taniguchi N, Aebi M. 2022. N-glycans. In: Varki A, et al., editors. *Essentials of glycobiology*. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press. pp. 103–116
- Steiner RA, Tucker JA. 2017. Keep it together: restraints in crystallographic refinement of macromolecule-ligand complexes. *Acta Crystallogr D Struct Biol*. 73:93–102. <https://doi.org/10.1107/S2059798316017964>.
- Stowell SR et al. 2010. Innate immune lectins kill bacteria expressing blood group antigen. *Nat Med*. 16:295–301. <https://doi.org/10.1038/nm.2103>.
- Sulák O et al. 2011. Burkholderia cenocepacia BC2L-C is a super lectin with dual specificity and proinflammatory activity. *PLoS Pathog*. 7:e1002238. <https://doi.org/10.1371/journal.ppat.1002238>.
- Tandrup T et al. 2025. The Swiss Army knife of alginate metabolism: mechanistic analysis of a mixed-function polysaccharide lyase/epimerase of the human gut microbiota. *J Am Chem Soc*. 147:23594–23607. <https://doi.org/10.1021/jacs.5c03557>.
- Thompson AJ et al. 2012. The reaction coordinate of a bacterial GH47 α -mannosidase: a combined quantum mechanical and structural approach. *Angew Chem Int Ed Engl*. 51:10997–11001. <https://doi.org/10.1002/anie.201205338>.
- Trastoy B et al. 2018. Structural basis for the recognition of complex-type N-glycans by endoglycosidase S. *Nat Commun*. 9:1874.
- Tsai YX et al. 2024. Rapid simulation of glycoprotein structures by grafting and steric exclusion of glycan conformer libraries. *Cell*. 187:1296–1311.e26. <https://doi.org/10.1016/j.cell.2024.01.034>.
- Urbanowicz BR et al. 2012. 4-O-methylation of glucuronic acid in Arabidopsis glucuronoxylan is catalyzed by a domain of unknown function family 579 protein. *Proc Natl Acad Sci USA*. 109: 14253–14258. <https://doi.org/10.1073/pnas.1208097109>.
- Vallee F, Karaveg K, Herscovics A, Moremen KW, Howell PL. 2000. Structural basis for catalysis and inhibition of N-glycan processing class I alpha 1,2-mannosidases. *J Biol Chem*. 275:41287–41298. <https://doi.org/10.1074/jbc.M006927200>.
- Varki A. 2017. Biological roles of glycans. *Glycobiology*. 27:3–49. <https://doi.org/10.1093/glycob/cww086>.
- Varki A et al. 2015. Symbol nomenclature for graphical representations of glycans. *Glycobiology*. 25:1323–1324. <https://doi.org/10.1093/glycob/cwv091>.
- Wandall HH, Nielsen MAI, King-Smith S, de Haan N, Bagdonaite I. 2021. Global functions of O-glycosylation: promises and challenges in O-glycobiology. *FEBS J*. 288:7183–7212. <https://doi.org/10.1111/febs.16148>.
- Wang HT et al. 2021. Rational enzyme design for controlled functionalization of acetylated xylan for cell-free polymer biosynthesis. *Carbohydr Polym*. 273:118564. <https://doi.org/10.1016/j.carbpol.2021.118564>.
- Wasim L et al. 2019. N-linked glycosylation regulates CD22 organization and function. *Front Immunol*. 10:699. <https://doi.org/10.3389/fimmu.2019.00699>.
- Weininger D. 1988. SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. *J Chem Inf Comput Sci*. 28:31–36. <https://doi.org/10.1021/ci00057a005>.
- Westbrook JD et al. 2015. The chemical component dictionary: complete descriptions of constituent molecules in experimentally

- determined 3D macromolecules in the protein data Bank. *Bioinformatics*. 31:1274–1278. <https://doi.org/10.1093/bioinformatics/btu789>.
- Woods RJ. 2018. Predicting the structures of glycans, glycoproteins, and their complexes. *Chem Rev*. 118:8005–8024. <https://doi.org/10.1021/acs.chemrev.8b00032>.
- Woods RJ. 2019. Glyfinder and Glyprobit: New online tools for locating and curating carbohydrate structures in Wwpdb. In: *Time-proof perspectives on Glycoscience—Beilstein Glyco-bioinformatics symposium*. Limburg, Germany. pp. 82–83
- Wu Y et al. 2023. Exploiting substrate specificities of 6-O-sulfotransferases to enzymatically synthesize keratan Sulfate oligosaccharides. *JACS Au*. 3:3155–3164. <https://doi.org/10.1021/jacsau.3c00488>.
- Wu Y et al. 2024. A biomimetic synthetic strategy can provide keratan Sulfate I and II oligosaccharides with diverse Fucosylation and sulfation patterns. *J Am Chem Soc*. 146: 9230–9240. <https://doi.org/10.1021/jacs.4c00363>.
- Xiang Y, Karaveg K, Moremen KW. 2016. Substrate recognition and catalysis by GH47 α -mannosidases involved in Asn-linked glycan maturation in the mammalian secretory pathway. *Proc Natl Acad Sci USA*. 113:E7890–e7899.
- Xu Y et al. 2023. Structures of liganded glycosylphosphatidylinositol transamidase illuminate GPI-AP biogenesis. *Nat Commun*. 14:5520. <https://doi.org/10.1038/s41467-023-41281-y>.
- Young JY et al. 2018. Worldwide protein data Bank biocuration supporting open access to high-quality 3D structural biology data. *Database (Oxford)*. 2018: bay002. <https://doi.org/10.1093/database/bay002>.
- Zhang Y, Skolnick J. 2004. Scoring function for automated assessment of protein structure template quality. *Proteins*. 57:702–710. <https://doi.org/10.1002/prot.20264>.