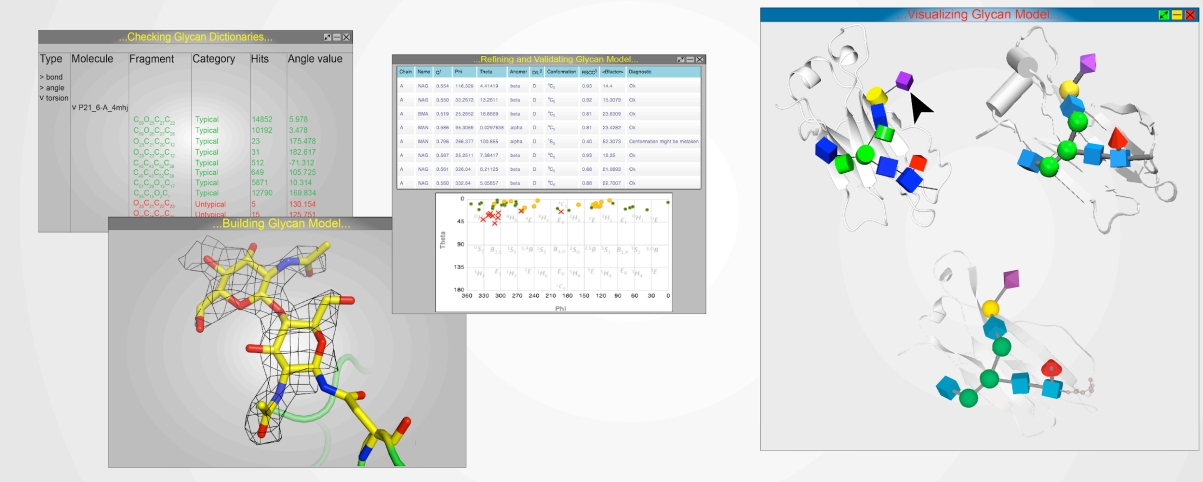
Structural glycobiology in the age of electron cryo-microscopy

Mihaela Atanasova, Haroldas Bagdonas and Jon Agirre[*]

York Structural Biology Laboratory, Department of Chemistry, University of York, UK

[*]Correspondence: jon.agirre@york.ac.uk

Graphical abstract



Highlights

- Progress is being made on providing automated model building and validation tools for structural glycobiology
- Electron cryo-microscopy (cryo-EM) can now be routinely used for resolving protein glycosylation
- High-resolution cryo-EM structures show fewer pyranose high-energy conformations than X-ray ones
- Re-refinement with the latest methods can produce better structures of glycoproteins automatically

Abstract

The methodology underpinning the construction, refinement, validation and analysis of atomic models of glycoproteins and protein-carbohydrate complexes has received a long-overdue boost in the last five years. This is a very timely development, as the resolution revolution in electron cryo-microscopy is now routinely delivering structures of key glycomedical importance, with a three-dimensional precision where X-ray crystallographic methods have traditionally floundered. This review will focus on the new software developments that have been introduced in the past two years, and their impact on the field of structural glycobiology in terms of published structures.

Introduction

Protein glycosylation plays a crucial role in recognition processes in e.g. viral infection, cancer, fertilisation, immunity and inflammation [1]. In this role, glycans are expected to provide stabilising contacts within the buried surface of a glycoprotein, while additionally playing a role as interaction partners on the surface, via hydrogen bonds or CH-π interactions. As independent entities, carbohydrates also have promising biotechnological applications, being a staple in the production of more eco-friendly second-generation biofuels from previously untractable crop waste. Assisting in this task, carbohydrate-active enzymes recognise, transfer and cut saccharide building blocks, often distorting individual rings to achieve catalysis.

Complicated stereochemistry, branching and unpredictable sequence/structure make protein glycosylation in particular harder to work with than pure protein, or even nucleic acid. Perhaps unsurprisingly, the software for handling structures of carbohydrate moieties is not yet as featureful as that for other biomolecules. This gap in capabilities becomes evident in both macromolecular crystallography (MX) and electron cryo-microscopy (cryo-EM) whenever the model fitting problem deviates from standard propositions. Indeed, at high-resolution it is possible to identify a monosaccharide and ascertain its ring conformation (Figure 1A) – to date, this has

only been possible with X-ray crystallography. Nevertheless, we fully expect cryo-EM to reach this level of precision in the near future. As resolution decreases, it becomes increasingly difficult to determine its ring conformation - thus requiring additional restraints for idealising ring puckering (Figure 1B-F) [2]. Finally, at low resolution, usually neither the monosaccharide nor its conformation can be identified (Figure 1C-F). It is in this particular case where the articulation of prior glyco-chemical knowledge must cross boundaries from the realm of validation, and play a central role in the structure building process: lowest energy ring conformations, a constant in pyranosides except in rare cases (catalysis is one of them), can be enforced using unimodal torsion restraints; the most probable linkage types, which should match the expression system's available glycosyltransferases, can be modelled using automated tools (*vide infra*); low energy glycosidic linkage orientations can be encouraged by using information from homologous structures via external restraints. As with protein methodology, whatever prior information is useful for validation at high resolution – e.g. the Ramachandran criterion – can be turned into restraints for refinement at low resolution – e.g. Ramachandran restraints. In becoming a target for refinement, validation metrics lose independence; yet as part of a balance between experimental and geometric terms, they are still useful as validation criteria – e.g. ideal bond lengths and angles are also used both as restraints in refinement, and as a measure of distortion particularly for ligands. It is ultimately the structural biologist's choice whether they want to produce the best possible structure, or have a measure of how correct it is.

Experimentally, it is clear that the mobility of the glycans poses a problem for both MX and cryo-EM, with Nuclear Magnetic Resonance (NMR) providing much of the insight into protein-carbohydrate interactions due to the degrading resolvability of the sugars down the glycans' branches [3] typically found with the two former techniques. On the other hand, most of the challenges present in software spring from the particularities of carbohydrate chemistry. Upon cyclisation, there are two choices for the orientation of the anomeric hydroxy group, which leads to two anomeric forms – alpha or beta (refer to [4] for a graphical description). Most D-sugar pyranoses adopt the $^{4}C_{1}$ conformation, while most L-sugar pyranoses adopt the $^{1}C_{4}$ conformation. Interconversion of pyranose rings between different conformations requires an itinerary, which can be described using the Cremer-Pople

sphere [5]. The two chair conformations, $^{4}C_{1}$ and $^{1}C_{4}$ are optimal because of the 60/-60 degree torsion angle between substituents, leaving them staggered instead of eclipsed. Conversion from $^{4}C_{1}$ to $^{1}C_{4}$ and vice versa requires jumping over a very high energy barrier, and normally would involve catalysis, which can be achieved with the help of a carbohydrate active enzyme [4,6].

Carbohydrate residue nomenclature is challenging for several reasons, including the two different types of glycosidic linkages (alpha or beta), branching and ring contortions. Lutteke *et al*., 2004 [7] first reported that about 30% of the deposited carbohydrate structures contain one or more nomenclature errors, a finding that gave rise to carbohydrate validation software, recently reviewed in [8,9]. A few years later, Crispin *et al*. also criticised the lack of methodological support for carbohydrates, singling out a deposited structure with a glycosidic linkage for which there were no available glycosyltransferases along its biosynthetic pathway [10,11]. More recently, Agirre *et al*. [2] performed an analysis on all N-glycan forming D-pyranosides found in the PDB using the Privateer software (CCP4 suite [12]): as data resolution decreases, more and more sugar monomers appear in high-energy conformations and/or have low real-space correlation. This indicated the need for using appropriate restraints during refinement.

In this review, we shall go through the latest software developments and their application to solving real-world structures, placing an emphasis on their impact on the recent evolution of electron cryo-microscopy into an all-around player in the structural glycobiology field. Aside from the growing access to automated, integrative model building and validation tools, a number of online support resources are available to the structural glycobiologist too: see [13,14] for a review of online resources, and Perez and De Sanctis [15] for a recent summary of the resources and techniques available where a synchrotron light source is available.

Dictionaries: the book of chemical knowledge

The model building process involves macromolecular refinement programs deriving geometric restraints from libraries of dictionaries, at least for most commonly

occurring monomers. Dictionaries are used to store prior chemical knowledge about compounds, including their composition, connectivity and stereochemistry. The CCP4 Monomer Library, one of the first examples of its kind, was based on the geometry proposed by Engh and Huber [16], which is now outdated particularly concerning sugars [4]. If a chemical compound does not have a library entry, or if it is incorrect, a new one needs to be generated. There are several programs that can be used for this, with irregular results for carbohydrates [4]. The CCP4 program ACEDRG [17,18] works by mining databases such as the Crystallography Open Database (COD) [18] to generate dictionaries from the data available there. It then uses RDKit (open source chemoinformatics; http://www.rdkit.org) to generate conformers which are ranked by free energy, and the minimal-energy one is chosen. ACEDRG/COD produces similar results to GRADE (Global Phasing Ltd.) and Phenix.eLBOW [19], which derive their restraints from Mogul [20], a tool that in turn mines the Cambridge Structural Database (CSD). Mogul is currently in use for geometry validation upon deposition with the Protein Data Bank, meaning that the use of old dictionaries during refinement with tight geometry targets – e.g. when refining against a cryo-EM map – can produce a disproportionate number of bond length and angle outliers. A modernisation effort is currently underway in CCP4, with hundreds of carbohydrate entries being marked for update through the combination of ACEDRG and Privateer [21]. The new dictionaries have an expected release date of 2020.

Model building

The improved N-glycosylation building module for Coot

Coot [22] has a carbohydrate-building tool [23] – earlier version reviewed in [9] – that can be used to build N-glycosylation into both crystallographic and cryo-EM maps. The module has three modes: manual, semi-automated and automated. The manual mode allows the user to choose a monosaccharide and a bond type from a selection of commonly available glycoforms. Coot chooses the best position, orientation and conformation for the selected monosaccharide, and refines the structure. In the semi-automated mode the user selects a glycan type and Coot returns possible

options for the monosaccharide and the glycosidic bond. The automated mode requires the user to simply choose the starting point and the glycosylation tree type, and Coot builds it automatically, interrupting the process when no more sugars can be built into clear density. An overview of results is presented in Figure 2 (adapted with permission from IUCr Journals). The tool has received positive adoption by the community, as evidenced by its use on several high-profile X-ray and cryo-EM structures with abundant protein N-glycosylation [24–27].

Its main limitation is the relatively narrow selection of glycoforms available. This is clearly a design decision rather than an oversight, as these represent the most common forms that can usually be resolved experimentally. Moreover, *Coot* does not include temperature-factor refinement, as all atoms are set to a fixed value. The authors suggested integrating the model-free B-factor refinement procedure described by Cowtan and Agirre [28] as an improvement.

PDB-REDO: Carbivore and carbonanza

Van Beusekom *et al.* [29] presented a set of tools that build on the Coot N-glycosylation building module to achieve a more automated behaviour; indeed, the software is meant to be part of their PDB-REDO [30] rebuilding and re-refinement pipeline. The first tool they presented is *Carbivore*, which can be used to rebuild and extend existing N-glycosylation trees automatically, or add new trees where they are missing. For the case glycosylation was not detected due to C1 not facing the asparagine side-chain, the authors introduced another program, named Carbonanza, to generate link records. The whole-tree addition method of Coot was extended to allow for building partial trees, i.e. extending existing trees. Moreover, a feature that finds N-glycosylation sites based on the consensus sequence Asn-X-Ser/Thr was implemented in *Carbivore*. In addition, an option for finding N-glycosylation sites based on homologous models was also presented, however this is not used by default as the search is likely to be slow.

ISOLDE

The ISOLDE plugin [31] for ChimeraX [32] offers a refreshing way of dealing with protein glycosylation, and supports both electron cryo-microscopy and X-ray crystallographic data. The graphical frontend connects to an interactive, GPU-accelerated molecular mechanics simulation, updating the model – and electron density maps, if working on crystallographic data – based on both the user's push-pull movements and the results of running the simulation on the updated coordinates. Technology-wise, this new tool makes use of the OpenMM toolkit [33] for simulations, and the Clipper-python module [34] for electron density calculations, which is heavily CPU-parallelised – using C++11-style threads – in the latest version available from the *ChimeraX toolshed* at the time of publication. Protein glycosylation is handled by an adapted version of the GLYCAM force field [35]. Although at present some unwanted effects such as ring inversions might appear as a result of the unrealistically high temperatures simulated by the user's push-pull movements, it is clear that this tool will be of great assistance when multiple overall glycan conformations need to be evaluated in a low resolution map; a combination with real-time validation at both the monosaccharide and glycan levels could further inform the fitting process and prevent errors too. The capabilities of ISOLDE are most effectively demonstrated in the supplementary video of [31].

Sails

Sails [36] can be used to build sugars automatically, either covalently linked to protein or as ligands. The software is currently in the middle of a major infrastructural change but is slated for general release in 2020 (with, or through an update to CCP4 7.1). It uses a method similar to that of Nautilus [37] and Buccaneer [38,39], using fingerprint-based detection of fragments, which account for both the target and its environment. The correlation function behind Sails has been proven to work with electron cryo-microscopy data, although adjustments may be needed if *e.g.* the scale of the EM map is not accurate or different map sharpening or blurring is required. Privateer and Refmac will be integrated with Sails in a pipeline for iterative building, refinement and validation.

Refinement and validation

Privateer

Privateer [21] is a carbohydrate-specific validation tool that can determine ring conformation of furanose and pyranose rings, anomeric form, absolute stereochemistry, real space correlation between model and omit density. In addition, Privateer generates other output such as SVG glycan diagrams in the Symbol Nomenclature For Glycans (SNFG) notation, and scripts for both Refmac5 [40] and Coot [22]. Like Sails, it is undergoing a change in infrastructure in order to future-proof its architecture.

Among the different checks that Privateer will do on carbohydrate models, a comparison of ring conformation and the ideal, minimal-energy conformation for each monosaccharide provides the fastest and most useful indication of potential mistakes in modelling and/or refinement: at high resolution, unjustified high-energy conformations - those without support of clear electron density - can reveal problems in the glycosidic bond (wrong anomer used, for instance) or wrong restraints (e.g. inverted chiralities). At low resolution, the problem can appear if the model is allowed to deviate from the ideal geometry due to providing insufficient restraints during refinement. Privateer generates dictionaries containing unimodal restraints upon detecting unjustified high-energy conformations. The validation and re-refinement process via these dictionaries is now completely automated via the CCP4i2 interface [41]. These developments were spearheaded after it was revealed that the PDB contained an unrealistically high number of non-chairs as part of N-glycosylation [2].

Many newer cryo-EM structures of glycoproteins are in the 2 Å to 6 Å resolution range due to improvement in electron sources, detectors, and image processing and 3D reconstruction algorithms. But the software for structure solution and validation have also improved, and perhaps as a result of that, high-resolution cryo-EM structures display fewer sugars in high-energy conformations than crystallographic ones. To illustrate this point, Privateer was run on all N-glycosylated structures in the PDB, solved with X-ray crystallography and cryo-EM. The decoupled results are shown in Figure 3. D-sugars are shown in blue, L-sugars are shown in yellow.

Ideally, in the particular case of N-glycosylation all D-sugars should be in $^4C_1$ conformation, and all L-sugars in $^1C_4$ conformation.

As previously highlighted elsewhere [4], pyranose higher-energy conformations are even more unusual than Ramachandran outliers, and should be reported alongside them in the refinement summary table.

Phenix, Rosetta and AMBER

Phenix uses a conformation-dependent library of restraints for the protein backbone [42] and homology refinement [43] for protein modeling. Rosetta can be used for carbohydrate refinement of both X-ray and cryo-EM structures using parameterisation derived from X-ray structures to approximate conformational energy [44]. Frenz *et al.*, [45] developed a protocol that can use either low-resolution crystallographic data, through Phenix-Rosetta integration [46] or cryo-EM data.

The RosettaCarbohydrate framework includes torsion-space refinement for glycans, which assumes ideal bond lengths and angles [47]. Frenz *et al.*, [45] build on previous work by expanding Rosetta's geometry term to include bond geometry deviations. These were derived from Phenix using eLBOW with AM1 optimization and added to the Rosetta database. Currently the sugar monomers included are alpha and beta glucose, N-acetyl glucosamine, alpha and beta mannose, and alpha and beta fucose.

The authors recommend using Privateer [21] before and after refinement to detect errors in the structure. For refinement of crystallography data, Rosetta's integration with Phenix can be used [48]. The protocols were modified to account for glycans, including steps for minimisation, increasing repulsive weights, and idealisation of anomeric hydrogens.

Phenix also offers integration with the AMBER molecular mechanics package, which is known for calculating torsion potentials accurately [49].

A word on legacy validation tools

While the tools outlined in this section are now sadly unsupported, it is worth mentioning them not just for the sake of completeness, but because there is no substitute tool yet for some of the key functions they provide. PDB-CARE (PDB CArbohydrate REsidue check; [50,51]) is a tool that can be used for bond and nomenclature validation. It is based on pdb2linucs, which is a software for carbohydrate detection based on atom types and their coordinates. The LINUCS notation [52] is used to normalise carbohydrate structures. This is done by comparing the carbohydrate structures' LINUCS notation to the PDB HET Group Dictionary, which contains sugar residues present in the coordinate file [50]. If a structure contains multiple anomers due to mutarotation at the reducing end of a saccharide, both forms need to have the correct PDB three-letter codes.

CARP (CArbohydrate Ramachandran Plot) is a tool that can be used to evaluate glycosidic linkage torsions. CARP also uses the pdb2linucs algorithm to analyse data, and compares it to data in GlyTorsionDB or GlycoMapsDB (for less common linkages). For each pair of monosaccharides and linkage combination, a separate torsional plot is created [7]. While these tools have been used mainly for validation purposes, they are a nice complement when examining the different linkage conformations in disaccharides [53].

Representation

While all-atom representations are the way to go for showing the interactions between protein and carbohydrate ligands, there is a case for using a simplified representation for glycans taking part in protein glycosylation; indeed, the sheer number of potential interactions occurring due to the size of the glycans – in optimal cases, 9 or more linked monosaccharides could be visible – and the particular relevance of their composition make all-atom figures difficult or near-impossible to follow. McNicholas and Agirre [54] introduced a representation (*Glycoblocks* for CCP4mg [55]) that, building on a 3D extension of the now standard Symbol Nomenclature For Glycans (SNFG) [6,56], added minimalistic dashed lines for

hydrogen bonds and CH-π interactions.

Not focusing on interactions, many 3D SNFG representations exist now either as plugins or as an integral part of wider-purpose graphics software, e.g. VMD [57], LiteMol [58], and UCSF Chimera [59] via the Tangram plugin [60]. These provide stand-out depictions of protein glycosylation using big regular polyhedra. A side-by-side comparison is shown in Figure 4. Finally, other software such as SweetUnityMol [61] and Pymol [62] combine the familiar colouring scheme with a more atomistic representation.

Future perspectives

It appears the gears are finally turning in the methodological machine towards implementing better support for carbohydrates. However, software still require expert knowledge of carbohydrate structure or very high resolution to work automatically. Work is currently being done on the Sails program to be able to overcome many of these limitations. In addition, based on encouraging early results [63–66], new carbohydrate dictionaries with more faithful model geometry and accurate torsion restraints will improve refinement, particularly for cryo-EM. Finally, sugars in active sites of enzymes might be distorted into high energy conformations, and thus may require further validation; work will need to be done in this respect in order to give users a confidence level on their conformational assignment.

We should like to emphasise that model building, refinement and validation will need to be further integrated together for maximum benefit of users. Recently, Van Beusekom, Lutteke and Joosten [8] used a set of tools, including PDB-REDO [30], Privateer [21] and CARP [51] to analyse 8,114 glycoproteins from the PDB. They succeeded in correctly re-annotating 3,620 carbohydrate residues, which were then re-refined and are now available for the community to use. Incorporating prior glyco-chemical knowledge into the structure solution process will, as exemplified by the aforementioned authors, extend the limits of resolvability further down our glycans.

Acknowledgements

Mihaela Atanasova is funded by the UK Engineering and Physical Sciences Research Council [EPSRC, grant number EP/R513386/1]. Haroldas Bagdonas is funded by The Royal Society [grant number RGF/R1/181006]. Jon Agirre is a Royal Society University Research Fellow [award number UF160039]. We should also like to acknowledge the support – by no means limited to financial backing – of the Department of Chemistry and the University of York.

Figures

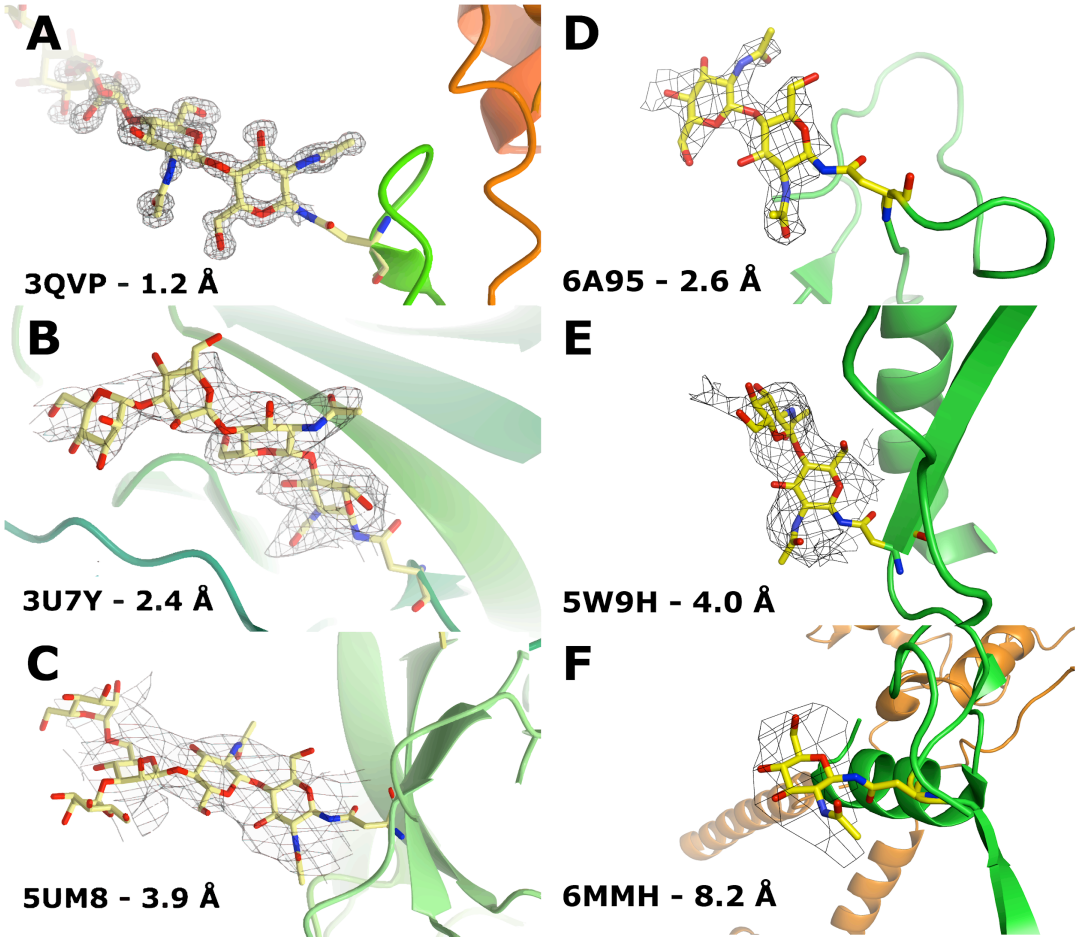


Figure 1. Comparison of N-glycan features in electron density maps over a range of resolutions. A-C: electron density maps obtained with X-Ray crystallography (MX). D-F: electronic potential maps obtained with cryo-EM; PDB codes and data resolution have been annotated directly on the figure. In the MX cases (A-C), at high resolution it is possible to identify monosaccharides and their ring conformation from the density map; at medium resolution, ring conformation becomes difficult to determine, whereas at low resolution, and indeed with many cryo-EM maps (D-F), a modelled N-glycan should always be backed by prior glyco-chemical knowledge: lowest energy ring conformations, most probable linkage types considering the expression system's available glycosyltransferases, and low energy glycosidic linkage orientations.

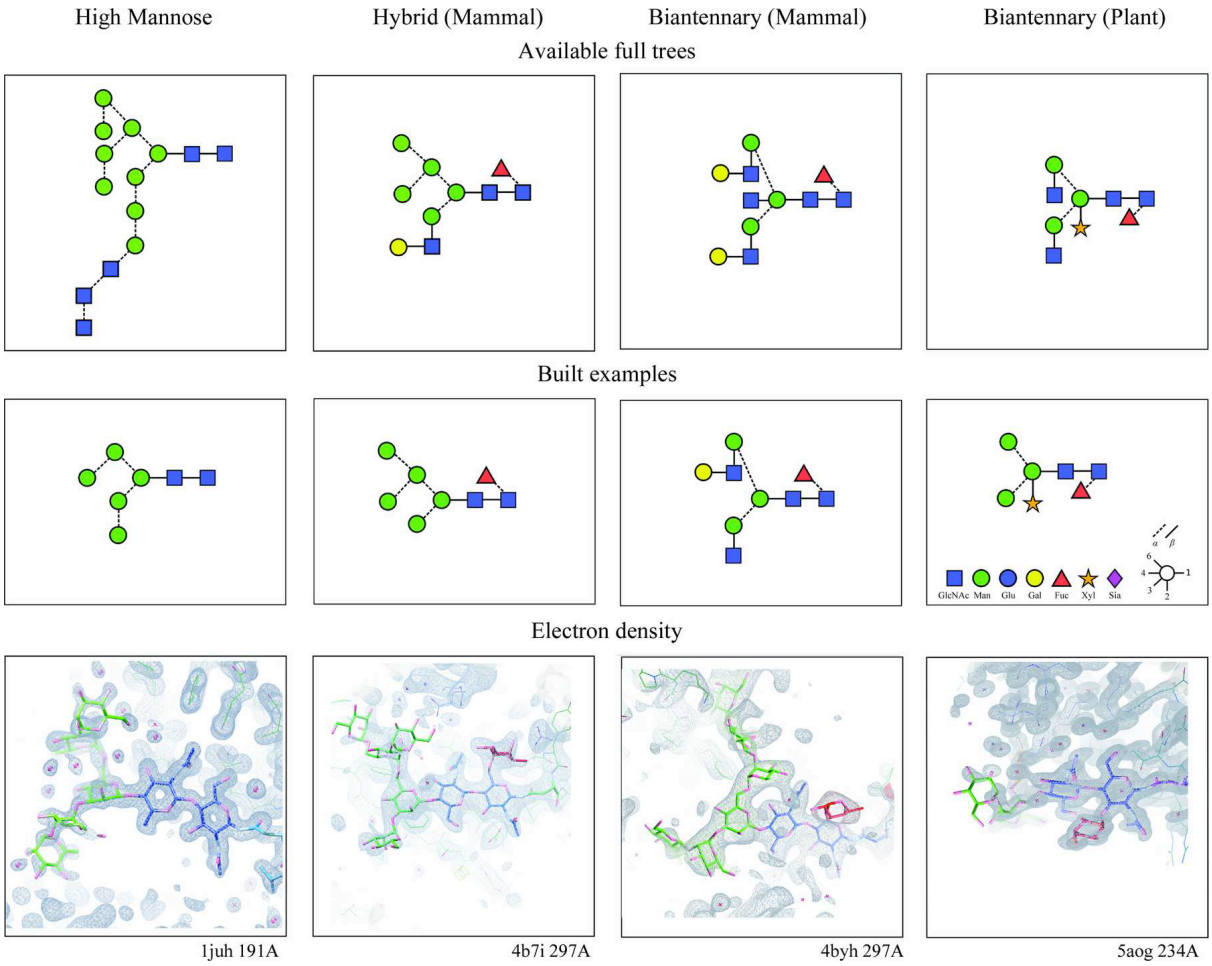


Figure 2. Results from a test of the N-glycosylation building tool in Coot [23]. The diagrams in SNFG format show the expected glycoforms and the subsets Coot was able to build automatically, while the third row of pictures shows how the maps looked like in each example. Reproduced from [23] with permission of the International Union of Crystallography.

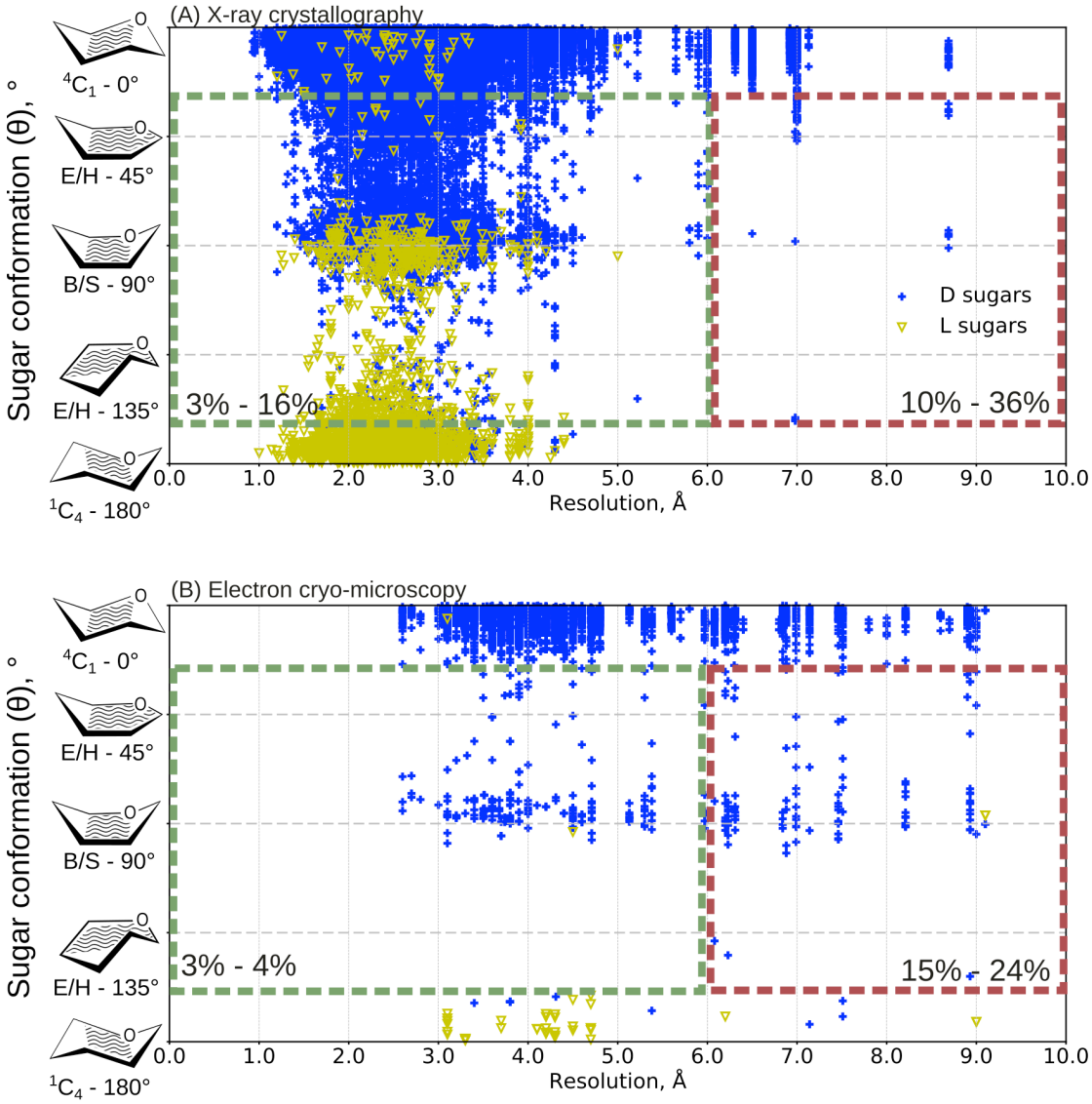


Figure 3. Pyranose ring conformations vs resolution for all sugars part of N-linked glycoproteins solved with (A) X-ray crystallography or (B) electron cryo-microscopy in the PDB by April 2019. E/H: Envelopes and Half-chairs, B/S: Boats and Skew-boats. Wavy lines denote the main ring plane. For reasons of clarity, half-chair, skew-boat and envelope were omitted from the axes at θ=45°, θ=90° and θ=135° respectively. Percentage of sugars in non-chair conformations is shown for resolution ranges 0.0-6.0 Å and 6.01-10.0 Å.

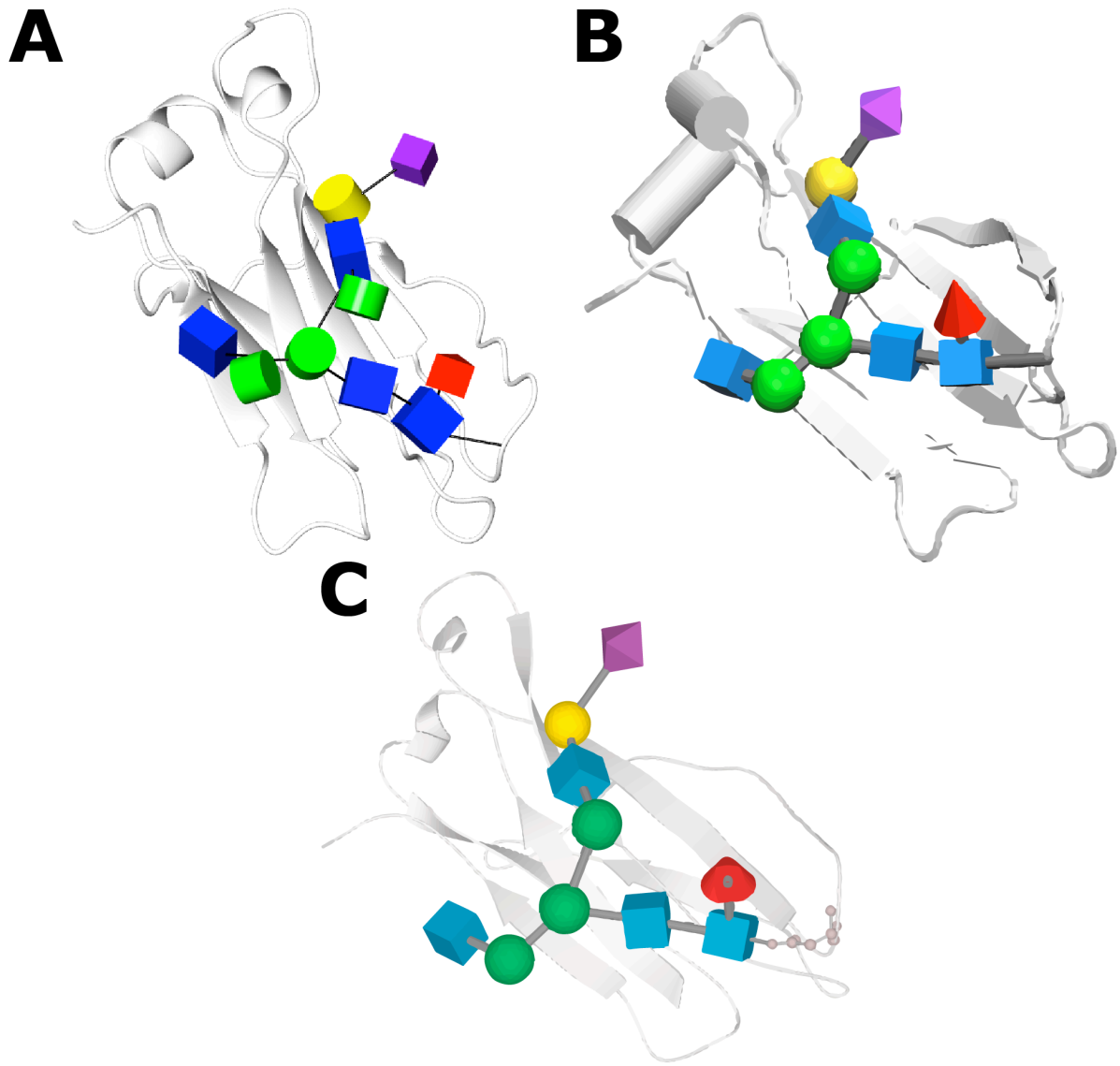


Figure 4. 3D SNFG glycan representation comparison of PDB code 4BYH in selected software: (A) CCP4mg [53] with Glycoblocks[54], (B) VMD [56] and (C) LiteMol [57].

Bibliography

1. Schnaar RL: **Glycobiology simplified: diverse roles of glycan recognition in inflammation**. *J Leukoc Biol* 2016, **99**:825–838.

2. Agirre J, Davies G, Wilson K, Cowtan K: **Carbohydrate anomalies in the PDB**. *Nature Chemical Biology* 2015, **11**:303.

3. Valverde P, Quintana JI, Santos JI, Ardá A, Jiménez-Barbero J: **Novel NMR Avenues to Explore the Conformation and Interactions of Glycans**. *ACS Omega* 2019, **4**:13618–13630.

4. Agirre J: **Strategies for carbohydrate model building, refinement and validation**. *Acta Crystallographica Section D: Structural Biology* 2017, **73**:171–186.

5. Cremer D, Pople JA: **A General Definition of Ring Puckering Coordinates**. *Journal of the American Chemical Society* 1975, **97**:1354–1358.

6. Varki A, Cummings RD, Aebi M, Packer NH, Seeberger PH, Esko JD, Stanley P, Hart G, Darvill A, Kinoshita T, et al.: **Symbol nomenclature for graphical representations of glycans**. *Glycobiology* 2015, **25**:1323–1324.

7. Lütteke T, Frank M, Von Der Lieth CW: **Data mining the protein data bank: Automatic detection and assignment of carbohydrate structures**. In *Carbohydrate Research*. . 2004:1015–1020.

8. Van Beusekom B, Lütteke T, Joosten RP: **Making glycoproteins a little bit sweeter with PDB-REDO**. *Acta Crystallogr Sect F Struct Biol Cryst Commun* 2018, **74**:463–472.

9. Agirre J, Davies GJ, Wilson KS, Cowtan KD: **Carbohydrate structure: the rocky road to automation**. *Current Opinion in Structural Biology* 2017, **44**:39–47.

10. Crispin M, Stuart DI, Yvonne Jones E: **Building meaningful models of glycoproteins**. *Nature Structural & Molecular Biology* 2007, **14**:354–354.

11. Berman HM, Henrick K, Nakamura H, Markley J: **Reply to: Building meaningful models of glycoproteins**. *Nature Structural & Molecular Biology* 2007, **14**:354–355.

12. Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, Evans PR, Keegan RM, Krissinel EB, Leslie AGW, McCoy A, et al.: **Overview of the CCP4 suite and current developments**. *Acta Crystallographica Section D: Biological Crystallography* 2011, **67**:235–242.

13. Yuriev E, Ramsland PA: **Carbohydrates in Cyberspace**. *Front Immunol* 2015, **6**:300.

14. Emsley P, Brunger AT, Lütteke T: **Tools to assist determination and validation of carbohydrate 3D structure data**. *Methods Mol Biol* 2015, **1273**:229–240.

15. Pérez S, de Sanctis D: **Glycoscience@Synchrotron: Synchrotron radiation applied to structural glycoscience**. *Beilstein Journal of Organic Chemistry* 2017, **13**:1145–1167.

16. Engh RA, Huber R: **Accurate bond and angle parameters for X-ray protein structure refinement**. *Acta Crystallographica Section A Foundations of Crystallography* 1991, **47**:392–400.

17. Long F, Nicholls RA, Emsley P, Gražulis S, Merkys A, Vaitkus A, Murshudov GN: **AceDRG: A stereochemical description generator for ligands**. *Acta Crystallographica Section D: Structural Biology* 2017, **73**:112–122.

18. Long F, Nicholls RA, Emsley P, Gražulis S, Merkys A, Vaitkus A, Murshudov GN: **Validation and extraction of molecular-geometry information from small-molecule databases**. *Acta Crystallographica Section D: Structural Biology* 2017, **73**:103–111.

19. Moriarty NW, Grosse-Kunstleve RW, Adams PD: **electronic Ligand Builder and Optimization Workbench (eLBOW): a tool for ligand coordinate and restraint generation**. *Acta Crystallogr D Biol Crystallogr* 2009, **65**:1074–1080.

20. Bruno IJ, Cole JC, Kessler M, Luo J, Momerwell WDS, Purkis LH, Smith BR, Taylor R, Cooper RI, Harris SE, et al.: **Retrieval of crystallographically-derived molecular geometry information**. *J Chem Inf Comput Sci* 2004, **44**:2133–2144.

21. Agirre J, Iglesias-Fernández J, Rovira C, Davies GJ, Wilson KS, Cowtan KD: **Privateer: Software for the conformational validation of carbohydrate structures**. *Nature Structural and Molecular Biology* 2015, **22**:833–834.

22. Emsley P, Lohkamp B, Scott WG, Cowtan K: **Features and development of Coot**. *Acta Crystallogr D Biol Crystallogr* 2010, **66**:486–501.

23. Emsley P, Crispin M: **Structural analysis of glycoproteins: Building N-linked glycans with coot**. *Acta Crystallographica Section D: Structural Biology* 2018, **74**:256–263.

24. Zhu S, Noviello CM, Teng J, Walsh RM, Kim JJ, Hibbs RE: **Structure of a human synaptic GABAA receptor**. *Nature* 2018, **559**:67–72.

25. Zhang B, Wang KB, Wang W, Wang X, Liu F, Zhu J, Shi J, Li LY, Han H, Xu K, et al.: **Enzyme-catalysed [6+4] cycloadditions in the biosynthesis of natural products**. *Nature* 2019, **568**:122–126.

26. Klünemann T, Preuß A, Adamczack J, Rosa LFM, Harnisch F, Layer G, Blankenfeldt W: **Crystal Structure of Dihydro-Heme d1 Dehydrogenase NirN from Pseudomonas aeruginosa Reveals Amino Acid Residues Essential for Catalysis**. *Journal of Molecular Biology* 2019, **431**:3246–3260.

27. Lee Y, Wiriyasermkul P, Jin C, Quan L, Ohgaki R, Okuda S, Kusakizako T, Nishizawa T, Oda K, Ishitani R, et al.: **Cryo-EM structure of the human L-type amino acid transporter 1 in complex with glycoprotein CD98hc**. *Nat Struct Mol Biol* 2019, **26**:510–517.

28. Cowtan K, Agirre J: **Macromolecular refinement by model morphing using non-atomic parameterizations**. *Acta Crystallogr D Struct Biol* 2018, **74**:125–131.

29. van Beusekom B, Wezel N, Hekkelman ML, Perrakis A, Emsley P, Joosten RP: **Building and rebuilding N-glycans in protein structure models**. *Acta Crystallographica Section D: Structural Biology* 2019, **75**:416–425.

30. Joosten RP, Lütteke T: **Carbohydrate 3D structure validation**. *Current Opinion in Structural Biology* 2017, **44**:9–17.

31. Croll TI: **ISOLDE: a physically realistic environment for model building into low-resolution electron-density maps**. *Acta Crystallogr D Struct Biol* 2018, **74**:519–530.

32. Goddard TD, Huang CC, Meng EC, Pettersen EF, Couch GS, Morris JH, Ferrin TE: **UCSF ChimeraX: Meeting modern challenges in visualization and analysis**. *Protein Sci* 2018, **27**:14–25.

33. Eastman P, Swails J, Chodera JD, McGibbon RT, Zhao Y, Beauchamp KA, Wang L-P, Simmonett AC, Harrigan MP, Stern CD, et al.: **OpenMM 7: Rapid development of high performance algorithms for molecular dynamics**. *PLoS Comput Biol* 2017, **13**:e1005659.

34. McNicholas S, Croll T, Burnley T, Palmer CM, Hoh SW, Jenkins HT, Dodson E, Cowtan K, Agirre J: **Automating tasks in protein structure determination with the clipper python module**. *Protein Sci* 2018, **27**:207–216.

35. Kirschner KN, Yongye AB, Tschampel SM, González-Outeiriño J, Daniels CR, Foley BL, Woods RJ: **GLYCAM06: a generalizable biomolecular force field. Carbohydrates**. *J Comput Chem* 2008, **29**:622–655.

36. glycojones: **glycojones/sails**. *GitHub* [date unknown],

37. Cowtan K: **Automated nucleic acid chain tracing in real time**. *IUCrJ* 2014, **1**:387–392.

38. Cowtan K: **The Buccaneer software for automated model building. 1. Tracing protein chains**. *Acta Crystallogr D Biol Crystallogr* 2006, **62**:1002–1011.

39. Cowtan K, IUCr: **Completion of autobuilt protein models using a database of protein fragments**. *Acta Crystallogr D Biol Crystallogr* 2012, **68**:328–335.

40. Murshudov GN, Skubák P, Lebedev AA, Pannu NS, Steiner RA, Nicholls RA, Winn MD, Long F, Vagin AA: **REFMAC5 for the refinement of macromolecular crystal structures**. *Acta Crystallogr D Biol Crystallogr* 2011, **67**:355–367.

41. Potterton L, Agirre J, Ballard C, Cowtan K, Dodson E, Evans PR, Jenkins HT, Keegan R, Krissinel E, Stevenson K, et al.: **CCP4i2: the new graphical user interface to the CCP4 program suite**. *Acta Crystallogr D Struct Biol* 2018, **74**:68–84.

42. Moriarty NW, Tronrud DE, Adams PD, Karplus PA: **Conformation-dependent backbone geometry restraints set a new standard for protein crystallographic refinement**. *FEBS J* 2014, **281**:4061–4071.

43. Park H, Ovchinnikov S, Kim DE, DiMaio F, Baker D: **Protein homology model refinement by large-scale energy optimization**. *Proceedings of the National Academy of Sciences* 2018, **115**:3054–3059.

44. Alford RF, Leaver-Fay A, Jeliazkov JR, O'Meara MJ, DiMaio FP, Park H, Shapovalov MV, Renfrew PD, Mulligan VK, Kappel K, et al.: **The Rosetta All-Atom Energy Function for Macromolecular Modeling and Design**. *J Chem Theory Comput* 2017, **13**:3031–3048.

45. Frenz B, Rämisch S, Borst AJ, Walls AC, Adolf-Bryfogle J, Schief WR, Veesler D, DiMaio F: **Automatically Fixing Errors in Glycoprotein Structures with Rosetta**. *Structure* 2019, **27**:134–139.e3.

46. Dimaio F, Echols N, Headd JJ, Terwilliger TC, Adams PD, Baker D: **Improved low-resolution crystallographic refinement with Phenix and Rosetta**. *Nat Methods* 2013, **10**:1102–1106.

47. Labonte JW, Adolf-Bryfogle J, Schief WR, Gray JJ: **Residue-centric modeling and design of saccharide and glycoconjugate structures**. *J Comput Chem* 2017, **38**:276–287.

48. Terwilliger TC, Di Maio F, Read RJ, Baker D, Bunkóczi G, Adams PD, Grosse-Kunstleve RW, Afonine PV, Echols N: **Phenix.mr-rosetta: Molecular replacement and model rebuilding with Phenix and Rosetta**. *J Struct Funct Genomics* 2012, **13**:81–90.

49. Case DA, Cheatham TE, Darden T, Gohlke H, Luo R, Merz KM, Onufriev A, Simmerling C, Wang B, Woods RJ: **The Amber biomolecular simulation programs**. *Journal of Computational Chemistry* 2005, **26**:1668–1688.

50. Lütteke T, Von Der Lieth C-W: **pdb-care (PDB CArbohydrate REsidue check): a program to support annotation of complex carbohydrate structures in PDB files**. *BMC Bioinformatics* 2004, **5**.

51. Lutteke T: **Carbohydrate Structure Suite (CSS): analysis of carbohydrate 3D structures derived from the PDB**. *Nucleic Acids Research* 2004, **33**:D242–D246.

52. Bohne-Lang A, Lang E, Förster T, Von der Lieth CW: **LINUCS: LInear Notation for Unique description of Carbohydrate Sequences**. *Carbohydr Res* 2001, **336**:1–11.

53. Fushinobu S: **Conformations of the type-1 lacto-N-biose I unit in protein complex structures**. *Acta Crystallogr Sect F Struct Biol Cryst Commun* 2018, **74**:473–479.

54. McNicholas S, Agirre J: **Glycoblocks: A schematic three-dimensional representation for glycans and their interactions**. *Acta Crystallographica Section D: Structural Biology* 2017, **73**:187–194.

55. McNicholas S, Potterton E, Wilson KS, Noble MEM: **Presenting your structures: the CCP4mg molecular-graphics software**. *Acta Crystallogr D Biol Crystallogr* 2011, **67**:386–394.

56. Varki A, Cummings RD, Esko JD, Freeze HH, Stanley P, Marth JD, Bertozzi CR, Hart GW, Etzler ME: **Symbol nomenclature for glycan representation**. *Proteomics* 2009, **9**:5398–5399.

57. Thieker DF, Hadden JA, Schulten K, Woods RJ: **3D implementation of the symbol nomenclature for graphical representation of glycans**. *Glycobiology* 2016, **26**:786–787.

58. Sehnal D, Grant OC: **Rapidly Display Glycan Symbols in 3D Structures: 3D-SNFG in LiteMol**. *J Proteome Res* 2019, **18**:770–774.

59. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE: **UCSF Chimera--a visualization system for exploratory research and analysis**. *J Comput Chem* 2004, **25**:1605–1612.

60. insilichem: **insilichem/tangram_snfg**. *GitHub* [date unknown],

61. Pérez S, Tubiana T, Imberty A, Baaden M: **Three-dimensional representations of complex carbohydrates and polysaccharides--SweetUnityMol: a video game-based computer graphic software**. *Glycobiology* 2015, **25**:483–491.

62. Arroyuelo A, Vila JA, Martin OA: **Azahar: a PyMOL plugin for construction, visualization and analysis of glycan molecules**. *J Comput Aided Mol Des* 2016, **30**:619–624.

63. Agirre J, Ariza A, Offen WA, Turkenburg JP, Roberts SM, McNicholas S, Harris PV, McBrayer B, Dohnalek J, Cowtan KD, et al.: **Three-dimensional structures of two heavily N-glycosylated Aspergillus sp. family GH3 β-D-glucosidases**. *Acta Crystallogr D Struct Biol* 2016, **72**:254–265.

64. Agirre J, Moroz O, Meier S, Brask J, Munch A, Hoff T, Andersen C, Wilson KS, Davies GJ: **The structure of the AliC GH13 α-amylase from Alicyclobacillus sp. reveals the accommodation of starch branching points in the α-amylase family**. *Acta Crystallogr D Struct Biol* 2019, **75**:1–7.

65. Schumann B, Malaker SA, Wisnovsky SP, Debets MF, Agbay AJ, Fernandez D, Wagner LJS, Lin L, Choi J, Fox DM, et al.: **Chemical precision glyco-mutagenesis by glycosyltransferase engineering in living cells**. *bioRxiv* 2019, doi:10.1101/669861.

66. Ji S, Dix SR, Aziz AA, Sedelnikova SE, Baker PJ, Rafferty JB, Bullough PA, Tzokov SB, Agirre J, Li F-L, et al.: **The molecular basis of endolytic activity of a multidomain alginate lyase from Defluviitalea phaphyphila, a representative of a new lyase family, PLxx**. *J Biol Chem* 2019, doi:10.1074/jbc.RA119.010716.

Reference annotations:

3. (**) An overview of the recent advances for analysing protein-glycan interactions with Nuclear Magnetic Resonance spectroscopy, a great alternative technique when neither crystallography nor electron cryo-microscopy can resolve them.

4. (*) An overview of the manual model building process for carbohydrates, including dictionary generation, refinement and validation.

8. (**) A study of how re-annotation and re-refinement of carbohydrate residues improves carbohydrate models.

9. (**) A review of the recent software developments for carbohydrate structure solution.

15. (*) Review of the use of synchrotron radiation experiments for structure determination of glycan-interacting proteins.

23. (**) A summary of the tools for building N-linked glycans available within Coot, with examples.

29. (*) A set of tools incorporated in the PDB-REDO pipeline for building, rebuilding and extending glycosylation trees.

45. (*) A Rosetta-based protocol for carbohydrate refinement of X-ray and cryoEM structures.

64. (*) A description of the structure solution of AliC GH13 alpha-amylase, including refinement and validation. Ad-hoc dictionaries were needed to refine two coexisting anomeric forms that were rotated with respect to each other.