

## Tools to Assist Determination and Validation of Carbohydrate 3D Structure Data

Paul Emsley, Axel T. Brunger, and Thomas Lütteke

### Abstract

The frequency of glycosylated protein 3D structures in the Protein Data Bank (PDB) is significantly lower than the proportion of glycoproteins in nature, and if glycan 3D structures are present, then they often exhibit a large degree of errors. There are various reasons for this, one of which is a comparably low support of carbohydrates in software tools for 3D structure determination and validation. This chapter illustrates the current features that assist crystallographers with handling glycans during 3D structure determination in Coot and CNS and with validation of the results.

**Key words** Glycosylation, *N*-Glycan, X-ray structures, CNS, Coot, pdb-care

---

### 1 Introduction

Glycosylation is the most complex and also the most common protein modification [1, 2]. It is estimated that about 50 % of all proteins are glycoproteins [3]. However, only about 3–4 % of protein 3D structures in the Protein Data Bank (PDB) [4] carry glycan chains [5, 6]. There are various reasons for this low abundance of glycoprotein 3D structures in the PDB. Proteins used for crystallization are often recombinantly expressed in bacteria such as *Escherichia coli* that lack a glycosylation machinery that is comparable to the eukaryotic one, i.e., they either do not utilize *N*-glycosylation at all, or the recognition motif for enzymes attaching glycans to proteins differs from the eukaryotic Asn-Xxx-Ser/Thr (Xxx not proline) *N*-glycosylation sequon [7, 8]. *N*-Glycosylation, however, is often important for correct protein folding [9–11] and has a stabilizing effect on the protein fold [12, 13]. Therefore, eukaryotic expression systems are preferred for eukaryotic proteins to ensure that the proteins are able to fold under largely native conditions [14]. However, if glycans are present in a protein, they often exhibit a large microheterogeneity, i.e., glycan sequences may differ between individual copies of a glycoprotein

in a single cell resulting in a variety of different glycoforms of that protein. For example, more than 120 different glycans have been found to be associated with one *N*-glycosylation site of human CD59 protein [15]. Even if only one glycoform is present, the glycans exhibit a large flexibility [16, 17]. This chemical and conformational diversity may hamper crystal growth; therefore, glycan chains are frequently cleaved off enzymatically before crystallization [18]. If glycans are present in the crystal, the diversity of the glycans often results in insufficient electron density for resolving the glycan structures especially in areas distinct from the reducing end of the sugar chain [17], as small conformational changes of the linkage between the proximal GlcNAc and the Asn side chain may result in large changes of the positions of terminal glycan residues [14]. For these reasons, almost 80 % of *N*-glycan chains present in X-ray structures consist only of one or two residues [6].

Unfortunately, carbohydrate moieties in PDB entries exhibit a rather large error rate, such as mismatches between PDB residue names and residues actually present in the 3D structures, superfluous atoms in glycosidic linkages, or glycan chains that are not in accordance with biosynthesis pathways [5, 6, 19, 20]. The reasons for these errors are manifold. It is often difficult for nonspecialists to distinguish between various carbohydrates that only differ in their stereochemistry, and knowledge of glycans is less widespread among crystallographers than knowledge of proteins. Such knowledge, however, is of special importance when electron density is not sufficient to unambiguously assign the positions of carbohydrate atoms, which for the reasons stated above is a rather frequent case. Residue notation in PDB format can also be confusing. Residue definitions were improved during PDB remediation [21] by removing many ambiguities and redundancies that existed within the Ligand Depot database [22] by then. For example, there had been three residue names for *D*-galactopyranoses before the remediation: GLA was defined for the alpha anomer, GLB for the beta anomer, and GAL could be used for any anomer [23]. The latter one was most frequently used to encode galactose residues by then. During remediation, GAL was redefined to encode the beta anomer and GLB was marked as obsolete. The limitation of residue names to three characters in PDB format, however, still complicates handling of carbohydrates in PDB entries, as commonly used monosaccharide names are longer than three characters. Therefore, carbohydrate residue abbreviations used in PDB format look rather cryptic to many glycoscientists and make it difficult for them to spot inconsistencies in residue notation.

Furthermore, validation tools for protein 3D structures have a long history [24–27], while validation of carbohydrate coordinates emerged much later. Software tools for macromolecular structure determination such as Coot [28] or CNS [29, 30] only recently started to support glycans. Such support, however, is important to

correctly deal with glycan residues and the peculiarities of linkages between individual glycan monomers and between carbohydrate and protein moieties. This chapter gives an overview of the current state of carbohydrate support in Coot and CNS and illustrates validation of glycan moieties in PDB structures.

---

## 2 Materials

### 2.1 Coot

Coot is a molecular graphics application primarily focused on fitting protein models using X-ray data [28]. Coot provides tools such as rotamer fitting, peptide flipping, rigid-body refinement, terminal residue addition, and secondary structure placement. Using Coot, one can build, refine, and validate macromolecular structures. For real-space refinement, Coot uses the Refmac monomer library to describe the geometry of monomers and the ways in which they can be linked.

### 2.2 CNS: Crystallography and NMR System

CNS is a software system for structure determination with data from X-ray diffraction or nuclear magnetic resonance experiments [29, 30]. It uses a modular, multilevel approach to computing, utilizing a high-level symbolic structure-determination language. Its key advantages are the flexibility and generality of the system, allowing “computational experimentation” with new algorithms or applications to different types of experimental data without tedious software development. User-friendly task-oriented input files are available for nearly all aspects of macromolecular structure determination by X-ray crystallography and solution NMR.

### 2.3 pdb-care

PDB CArbohydrate REsidue check (pdb-care) [23] is the first validation tool focused on carbohydrates. The first version mainly checked residue notation and atom connectivities, whereas the current release also evaluates matches to biosynthesis pathways (currently limited to the *N*-glycan core structure) to examine biological correctness of the detected glycans. Moreover, the web interface was completely refurbished to improve the presentation of validation results, and results are also offered in XML format for access by automated validation routines.

### 2.4 CARP: Carbohydrate Ramachandran Plot

Similar to proteins, the overall shape of a glycan chain is mainly determined by the  $\phi$  and  $\psi$  torsions of the linkages between the monomeric building blocks. Plotting these torsions against each other in a Ramachandran plot [31] is a frequently used tool in protein structure validation [32]. This technique can also be used to evaluate glycan 3D structures. In contrast to proteins, however, the preferred combinations of  $\phi$  and  $\psi$  torsions strongly depend on the involved residues and the type of linkage by which they are connected [33, 34]. Therefore, it is not possible to collect all

torsions in one plot, but instead a separate plot for each disaccharide fragment has to be generated [6]. This task is performed by CARP (carbohydrate Ramachandran plot) [35]. To be able to use the plot for structure validation, it is necessary to have some knowledge of the  $\phi$  and  $\psi$  torsions that are usually observed for a specific linkage. This information can be obtained e.g., from the GlyTorsion database [35, 36], which stores glycosidic torsions observed in PDB entries, or from conformational maps derived from molecular dynamics (MD) simulations provided by GlycoMapsDB [37].

---

## 3 Methods

### 3.1 Carbohydrate Fitting in Coot

The modeling of carbohydrate in Coot [28] is based on the Refmac monomer library [38]. The format of the Refmac monomer library is mmCIF/PDBx [39] used by several macromolecular refinement packages. The library contains geometric descriptions of many monomers, some of which are carbohydrate monomers (saccharide parameters for which are mostly derived from [40]). Monomer descriptions typically contain entries for atom attributes, bonds, angles, torsions, planes, and chiral centers.

Using Coot, one can simple-mindedly import a carbohydrate monomer to the center of the screen using “*Get Monomer*.” This entails knowing the three-letter code of the carbohydrate monomer. If this is not known, then the relevant monomer can be found via a text search of the monomer names in the library.

The Refmac library also contains geometric descriptions of how monomers might be linked. These “chem\_links” provide a description of new geometry (bond, angles torsion, chiral centers, and planes) formed as a result of a chemical reaction. So, for example, for the linking of amino acids, we have TRANS, PTRANS, CIS, and PCIS links (for *cis*-, *trans*-, and proline links). In the case of carbohydrates, we have alpha and beta linkages as well as links to peptide (such as NAG-ASN, NAG-SER, and MAN-THR).

Using the multi-residue-torsion function of Coot, the atoms of, for example, an alpha-1-6-linked mannose (and any further downstream linked carbohydrate) can be moved by rotation about the linking phi, psi, and omega torsion angles.

When considering the links between carbohydrate monomers using the *refine-residues* function, Coot does not use LINK records as specified in the coordinates file (in the author’s experience, these are sometimes missing or incorrect). Instead, a distance criterion is used (bonding atoms need to be less than 3Å apart for the link to be activated) in conjunction with a search across all links types specified in the library.

More recently, the function *add-linked-residue* has been added to Coot. This allows the addition of a new residue type with a given link type to be added to a specified residue (so, e.g., one might

add a “NAG” with link type “NAG-ASN” to an “ASN” residue). The link names and the monomer names correspond to those in the monomer library. The new coordinates are based on template/reference structures for various link types.

As of this writing, the only supported protein-carbohydrate link (when using *add-linked-residue*) is the N-link. There are template structures for pyranose addition only. The addition of linked residues doesn't yet have a GUI. In due course, these issues will be addressed.

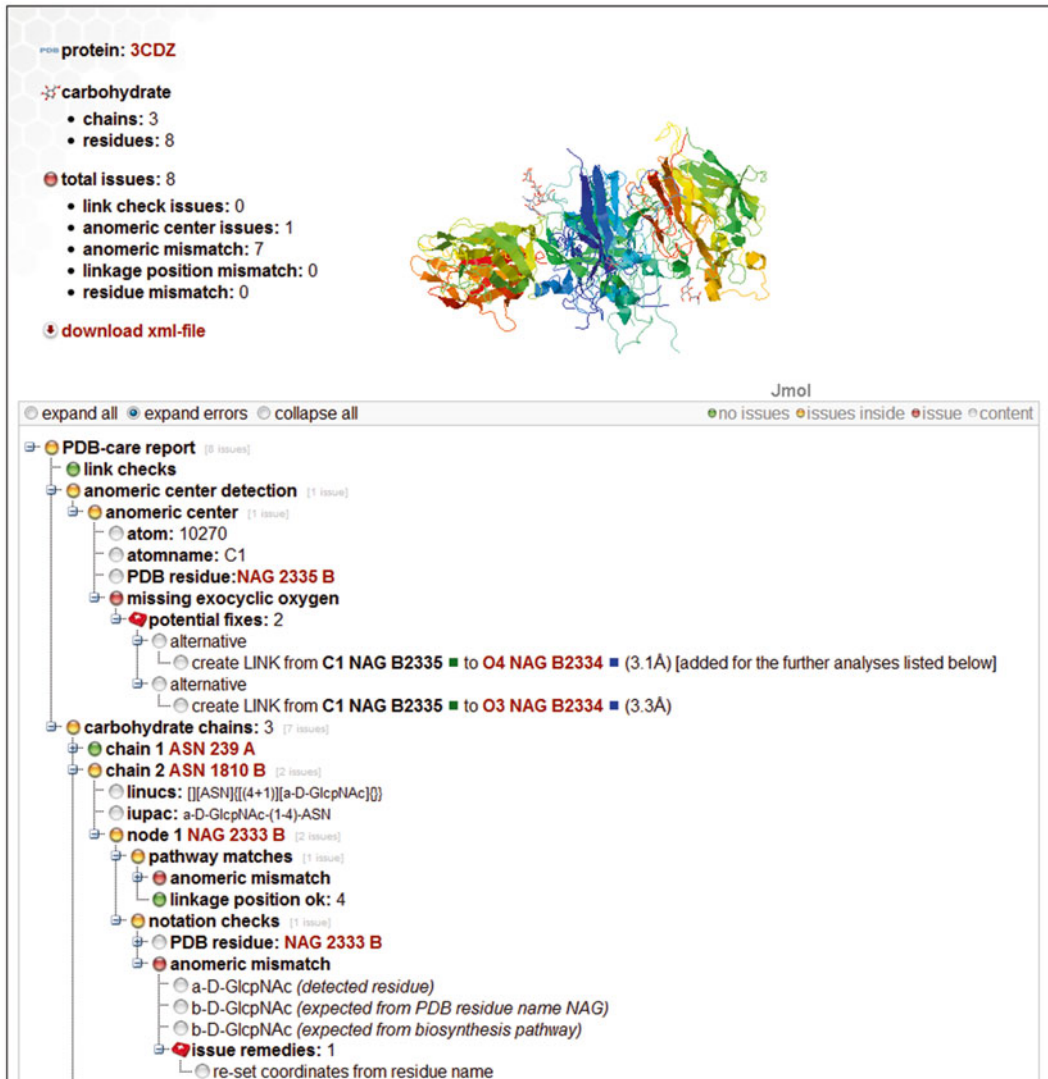
### 3.2 Stereochemical Parameters for Carbohydrates for CNS

The stereochemical parameters of carbohydrates that are used in CNS [29, 30] have a long and convoluted history. These parameters originated from carbohydrate parameters that were implemented in the program CHARMM [41] by John Brady and collaborators [42]. Subsequently, the parameterization for the sugar moieties was modified by William Weis to account for the anomeric effect [43, 44]. These modified parameters were used for the refinement of the influenza virus hemagglutinin [45] in order to account for the N-linked carbohydrates. Subsequently, the protein-specific portions of the parameters were adjusted to make them consistent with the protein parameters by Engh and Huber [46]. More recently, in version 1.3 of CNS [47], the nomenclature for the carbohydrates and their anomers was updated.

Different anomeric states of carbohydrates are defined by their respective names. For example, GLC is alpha-D-glucopyranose (i.e., the alpha-anomeric state), whereas BGC is beta-D-glucopyranose (i.e., the beta-anomeric state). CNS also requires the definition of link residues in order to produce covalent links between carbohydrates and to create the correct anomeric configuration. For example, the link B12 creates a link between BGC, GAL, BMA, NAG, FCB, and XYP to any other carbohydrate, whereas A12 creates a link between GLC, GLA, MAN, NDG, A2G, and FCA to any other carbohydrate. These links are created automatically in CNS version 1.3 when using task files that employ automatic molecular topology generation using distance-based criteria to create the covalent links (*see Note 1*). However, this method will fail if the atoms that should be connected by a covalent link are too far apart (the required distance is 2.5 Å or less).

### 3.3 Validation of Carbohydrate 3D Structures with pdb-care

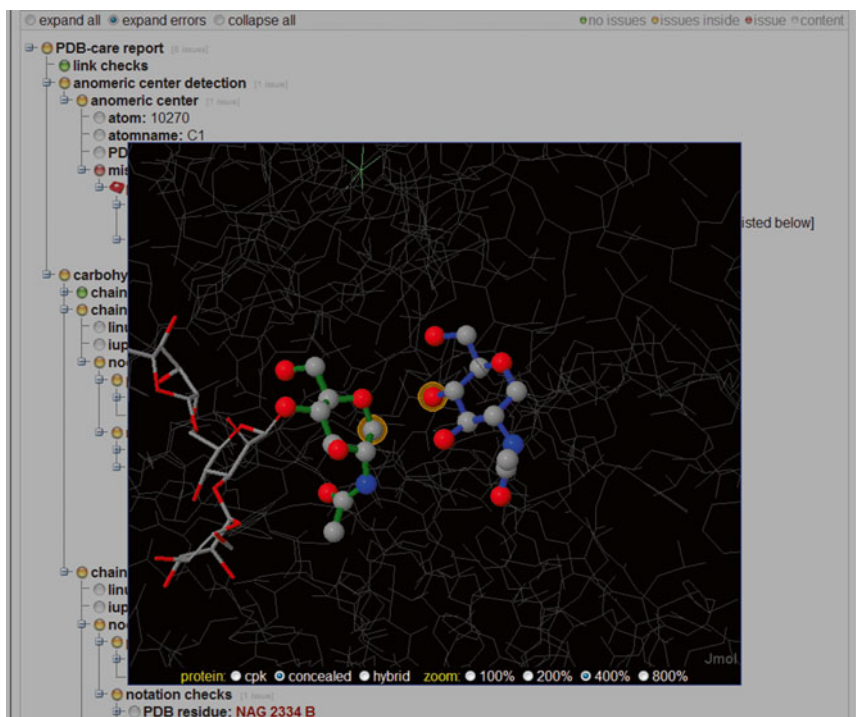
1. Go to <http://www.glycosciences.de/tools/pdb-care2/>.
2. You can either specify a PDB ID to check a published PDB entry or upload your own file in PDB format. For example, enter “3CDZ” into the “PDB ID” field.
3. Keep standard values for all other settings (*see Notes 2 and 3*) and click the “Submit” button.
4. In the upper part of the results page (*see Fig. 1*), a summary of detected glycans and of identified issues is given, and the



**Fig. 1** pdb-care output for PDB ID 3CDZ

analyzed structure is visualized in a Jmol applet [48]. Below, detailed results are presented in a treelike fashion. By default, only those branches that contain issues are opened. This behavior can be changed using the radio buttons “*expand all*,” “*expand errors*” (default), and “*collapse all*” on top of this section or individually by clicking on the “+” and “-” symbols in the results tree to open or close specific subparts.

- In the example of PDB entry 3CDZ, pdb-care observed a problem when assigning the anomeric centers: no exocyclic oxygen or equivalent atom is linked to the C1 of residue NAG 2335B, which is thus indicated as “*missing exocyclic oxygen*.” In such a case, the tool tries to find potentially linked atoms in



**Fig. 2** Detected issues are visualized in Jmol. In this example, the atoms that potentially form a missing linkage between two monosaccharides are highlighted

the vicinity of the C1. In this example, two potential linkages are detected, which are listed in the “*potential fixes*” section of this branch sorted by probability of correctness (on the basis of bond lengths and bond angles). As indicated in the web interface (*see* Fig. 1), the first linkage was added during detection, i.e., the further results are based on a structure that has this linkage. This is done to avoid identification of truncated glycan chains.

6. You can click on any of the linked PDB residue names in the interface to open a second Jmol window with the selected residue highlighted, so that you can directly examine the 3D structure to check the evaluated residue (*see* Note 4). If you click on a residue name in the list of potential links, the two residues that are involved in the linkage are displayed using blue and green bonds, and the potentially linked atoms are highlighted by orange halos (*see* Fig. 2). Click anywhere in your browser outside the Jmol area to close the Jmol view and return to the validation results.
7. Below this section, the detected glycan chains are listed (*see* Fig. 1). In this example three chains are found. Of each chain, the detected residues are listed, the identified residues

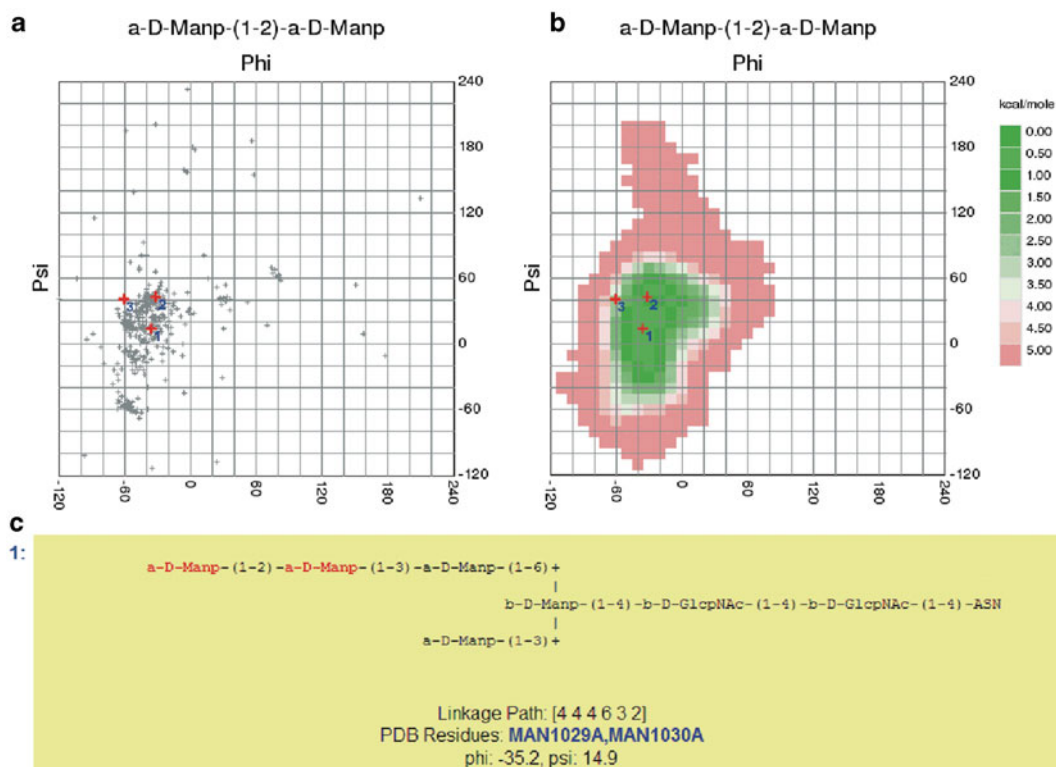
are compared with the ones that are defined by the PDB residues, and in the case of *N*-glycans, their accordance with biosynthesis pathways is also checked. Chain 2 of this example has an  $\alpha$ -D-GlcpNAc residue linked to an ASN side chain. In this position of the *N*-glycan core, however,  $\beta$ -D-GlcpNAc would be expected, i.e., the detected residue has the wrong anomeric state. The (correct) residue name NAG is used here which actually encodes a  $\beta$ -D-GlcpNAc residue. Due to this inconsistency, this residue is flagged both in the pathway and in the notation section.

8. pdb-care does not only point out inconsistencies but also tries to suggest ways how to fix these problems. Two potential reasons are principally able to cause mismatches between detected residue and PDB residue name. The residue name could be wrong, or the interpretation of electron density that revealed the given set of atom coordinates was incorrect. In the case of *N*-glycan core structures, it is often possible to use biosynthesis information to decide if the residue name or the atom coordinates have to be adjusted. This becomes obvious in chain 3 of this example: two  $\beta$ -D-Manp residues are encoded as MAN in this chain, which actually is defined for  $\alpha$ -D-Manp. While the first one (MAN 2336B) is expected to be the beta anomer, the second one (MAN 2338B) actually should be in alpha-anomeric form. Therefore, pdb-care suggests renaming the residue from MAN to BMA in the former case, while in the latter case (as well as in the case of the NAG residue mentioned in **step 7**), the atom coordinates have to be adjusted. If no such decision can be made automatically, both alternatives are listed. Whereas renaming of residues in PDB files can be done with any text editor, you have to obtain the electron density from EDS (electron density server, <http://eds.bmc.uu.se/eds/>) [49] and use X-ray tools such as Coot or CNS (*see* Subheadings **3.1** and **3.2**, respectively) to adjust the problematic coordinates.
9. If at least one of the detected chains consists of at least two monosaccharides (i.e., there is at least one glycosidic linkage present), a form to start a CARP analysis (*see* Subheading **3.4**) is displayed below the validation results.

### **3.4 Evaluating the Conformations of Glycosidic Linkages with CARP**

1. Go to <http://www.glycosciences.de/tools/carp/>.
2. As in pdb-care (*see* Subheading **3.3**), you can specify an existing PDB entry by its ID or upload a file in PDB format. Enter, e.g., “3OGS” into the “PDB ID” field.
3. Select “PDB (NMR-like definition)” as “plot background” (*see* **Note 5**).
4. Keep standard values for all other settings (*see* **Notes 2** and **3**) and click the “Submit” button.





**Fig. 3** Validation of glycosidic torsion angles with CARP. Torsions identified in the analyzed structure can be compared with those detected in other PDB entries (a) or with computed conformational maps (b). Below the plots, the primary sequences of the involved glycans are listed (c). A linkage path describing the linkages from the reducing end to the analyzed linkage is used to unambiguously identify a linkage within a sequence

5. When looking at the results, you will observe that most torsions of this structure are within frequently adopted areas of the plots (Fig. 3). The a-D-Glcp-(1, 2)-a-D-Manp linkage, however, is difficult to judge this way as there are very few data available for this torsion in the PDB.
6. Repeat **steps 1** and **2**, and select “*GlycoMapsDB*” as “*plot background*” this time to compare the observed torsions to computed conformational maps. These do not depend on the number of torsions available in the PDB, and energy values are computed from the frequency of  $\phi/\psi$  combinations observed in MD simulations, which makes it easy to quantify the results. For some linkages, however, no maps are available so far in GlycoMapsDB; thus, not all torsions can be evaluated this way.
7. The a-D-Glcp-(1-3)-a-D-Manp linkage is present in an area that was not occupied during the MD simulation that yielded the map. You can try to use a different map by clicking on the “*select map*” link below the glycan sequence and then on a map ID in the subsequent page, but the  $\phi$  torsion of this linkage is not occupied in any of the available maps.

8. Click on the corresponding PDB residue names in the carbohydrate sequence information area below the plots to open a Jmol view of this glycosidic linkage. “Unusual” torsion angles do not necessarily indicate errors in the structures, as glycans are more flexible than proteins. Furthermore, the computed conformational maps are derived from free glycans, whereas interactions with the protein part of a structure can induce conformations that are not occupied frequently by the free glycan [50]. In this example, however, no direct influence of the protein is observable. If you download the corresponding electron density map of PDB entry 3OGS from EDS [49] and have a look at GLC 1040A, e.g., in Coot, you will see that there is less electron density available for this residue than for the other monosaccharides of this chain, i.e., the unusual torsion might be a result of wrong interpretation of electron density data.

---

## 4 Notes

1. At present, only N-linked carbohydrates are supported with CNS.
2. If you upload a file in PDB format that does not originate from the PDB, carbohydrates might be present in ATOM records instead of HETATM. In that case, you have to adjust the “*Find carbohydrates in*” option.
3. By default pdb-care and CARP, read atom connectivities from LINK and CONECT records of the PDB file. If you upload a file that is lacking such records, you should select the “*Assign connections by atom distances*” option. However, this might result in superfluous connections if the structure has a distorted geometry that causes atoms to be close enough in space to be wrongly regarded as connected.
4. The Jmol scripts to achieve centering and highlighting of residues sometimes are not transmitted correctly. If your residue of interest is not shown, click on the area outside the Jmol window to close the current Jmol view, and then click on the residue name again to reopen the Jmol view.
5. There are several definitions of the atoms that form glycosidic  $\phi$  and  $\psi$  torsions [6]. A frequently used definition that makes use of hydrogen atoms is called “NMR definition.” If no hydrogen atoms are present in the analyzed structure, torsion values in NMR definitions can be computed from those in crystallographic definitions that are measured in the structure.

## References

1. Helenius A, Aebi M (2001) Intracellular functions of N-linked glycans. *Science* 291: 2364–2369
2. Moremen KW et al (2012) Vertebrate protein glycosylation: diversity, synthesis and function. *Nat Rev Mol Cell Biol* 13:448–462
3. Apweiler R et al (1999) On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database. *Biochim Biophys Acta* 1473:4–8
4. Berman HM et al (2000) The protein data bank. *Nucleic Acids Res* 28:235–242
5. Lütteke T et al (2004) Data mining the protein data bank: automatic detection and assignment of carbohydrate structures. *Carbohydr Res* 339:1015–1020
6. Lütteke T (2009) Analysis and validation of carbohydrate three-dimensional structures. *Acta Crystallogr D Biol Crystallogr* 65: 156–168
7. Szymanski CM, Wren BW (2005) Protein glycosylation in bacterial mucosal pathogens. *Nat Rev Microbiol* 3:225–237
8. Kowarik M et al (2006) Definition of the bacterial N-glycosylation site consensus sequence. *EMBO J* 25:1957–1966
9. Parodi AJ (2000) Protein glucosylation and its role in protein folding. *Annu Rev Biochem* 69: 69–93
10. Molinari M (2007) N-glycan structure dictates extension of protein folding or onset of disposal. *Nat Chem Biol* 3:313–320
11. Roth J et al (2010) Protein N-glycosylation, protein folding, and protein quality control. *Mol Cells* 30:497–506
12. Wormald MR, Dwek RA (1999) Glycoproteins: glycan presentation and protein-fold stability. *Struct Fold Des* 7:R155–R160
13. Bosques CJ et al (2004) Effects of glycosylation on peptide conformation: a synergistic experimental and computational study. *J Am Chem Soc* 126:8421–8425
14. Davis SJ, Crispin M (2011) Solutions to the glycosylation problem for low- and high-throughput structural glycoproteomics. In: Owens RJ, Nettleship JE (eds) *Functional and structural proteomics of glycoproteins*. Springer, Dordrecht, pp 127–158
15. Rudd PM et al (1997) The glycosylation of the complement regulatory protein, human erythrocyte CD59. *J Biol Chem* 272:7229–7244
16. Kozar T et al (1998) Studies on the conformational behaviour of GlcNAc-Man3-GlcNAc2 oligosaccharides using molecular dynamics simulations. *Glycoconj J* 15:187–191
17. Nage M, Yamaguchi Y (2012) Function and 3D structure of the N-glycans on glycoproteins. *Int J Mol Sci* 13:8398–8429
18. Chang VT et al (2007) Glycoprotein structural genomics: solving the glycosylation problem. *Structure* 15:267–273
19. Crispin M et al (2007) Building meaningful models of glycoproteins. *Nat Struct Mol Biol* 14:354
20. Nakahara T et al (2008) Glycoconjugate data bank: structures – an annotated glycan structure database and N-glycan primary structure verification service. *Nucleic Acids Res* 36: D368–D371
21. Henrick K et al (2008) Remediation of the protein data bank archive. *Nucleic Acids Res* 36:D426–D433
22. Feng Z et al (2004) Ligand depot: a data warehouse for ligands bound to macromolecules. *Bioinformatics* 20:2153–2155
23. Lütteke T, von der Lieth CW (2004) pdb-care (PDB carbohydrate residue check): a program to support annotation of complex carbohydrate structures in PDB files. *BMC Bioinformatics* 5:69
24. Laskowski RA et al (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *J Appl Crystallogr* 26: 283–291
25. Hooft RW et al (1996) Errors in protein structures. *Nature* 381:272
26. Davis IW et al (2007) MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res* 35: W375–W383
27. Read RJ et al (2011) A new generation of crystallographic validation tools for the protein data bank. *Structure* 19:1395–1412
28. Emsley P et al (2010) Features and development of coot. *Acta Crystallogr D Biol Crystallogr* 66:486–501
29. Brunger AT et al (1998) Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr D Biol Crystallogr* 54:905–921
30. Brunger AT (2007) Version 1.2 of the crystallography and NMR system. *Nat Protoc* 2: 2728–2733
31. Ramachandran GN et al (1963) Stereochemistry of polypeptide chain configurations. *J Mol Biol* 7:95–99

32. Hoofst RW et al (1997) Objectively judging the quality of a protein structure from a Ramachandran plot. *Comput Appl Biosci* 13: 425–430
33. Petrescu AJ et al (1999) A statistical analysis of N- and O-glycan linkage conformations from crystallographic data. *Glycobiology* 9:343–352
34. Wormald MR et al (2002) Conformational studies of oligosaccharides and glycopeptides: complementarity of NMR, X-ray crystallography, and molecular modeling. *Chem Rev* 102:371–386
35. Lütteke T et al (2005) Carbohydrate structure suite (CSS): analysis of carbohydrate 3D structures derived from the protein data bank. *Nucleic Acids Res* 33:D242–D246
36. Lütteke T, von der Lieth CW (2009) Data mining the PDB for glyco-related data. *Methods Mol Biol* 534:293–310
37. Frank M et al (2007) GlycoMapsDB: a database of the accessible conformational space of glycosidic linkages. *Nucleic Acids Res* 35:287–290
38. Vagin AA et al (2004) REFMAC5 dictionary: organization of prior chemical knowledge and guidelines for its use. *Acta Crystallogr D Biol Crystallogr* 60:2184–2195
39. Fitzgerald PMD et al (1996) The mmCIF dictionary: community review and final approval. *Acta Crystallogr A* 52:C575
40. Saenger W (1983) Principles of nucleic acid structure. Springer, Berlin
41. Brooks BR et al (1983) Charmm – a program for macromolecular energy, minimization, and dynamics calculations. *J Comput Chem* 4: 187–217
42. Ha SN et al (1988) A revised potential-energy surface for molecular mechanics studies of carbohydrates. *Carbohydr Res* 180:207–221
43. Jeffrey GA (1990) Crystallographic studies of carbohydrates. *Acta Crystallogr B Struct Sci* 46:89–103
44. Hirotsu K, Shimada A (1974) Crystal and molecular-structure of beta-lactose. *Bull Chem Soc Jpn* 47:1872–1879
45. Weis WI et al (1990) Refinement of the influenza-virus hemagglutinin by simulated annealing. *J Mol Biol* 212:737–761
46. Engh RA, Huber R (1991) Accurate bond and angle parameters for X-ray protein-structure refinement. *Acta Crystallogr A* 47:392–400
47. Schroder GF et al (2010) Super-resolution biomolecular crystallography with low-resolution data. *Nature* 464:1218–1222
48. Jmol: an open-source Java viewer for chemical structures in 3D. <http://www.jmol.org/>
49. Kleywegt GJ et al (2004) The Uppsala electron-density server. *Acta Crystallogr D Biol Crystallogr* 60:2240–2249
50. Raman R et al (2003) Structural specificity of heparin binding in the fibroblast growth factor family of proteins. *Proc Natl Acad Sci U S A* 100:2357–2362