Published in final edited form as:
J Comput Chem. 2011 November 15; 32(14): 3135–3141. doi:10.1002/jcc.21886.

Glycan Reader: Automated Sugar Identification and Simulation Preparation for Carbohydrates and Glycoproteins

Sunhwan Jo[1], **Kevin C. Song**[1], **Heather Desaire**[2], **Alexander D. MacKerell Jr.**[3], and **Wonpil Im**[1,*]

[1]Department of Molecular Biosciences and Center for Bioinformatics, The University of Kansas, 2030 Becker Drive, Lawrence, KS

[2]Department of Chemistry, The University of Kansas, 2030 Becker Drive, Lawrence, KS

[3]Department of Pharmaceutical Sciences, School of Pharmacy, University of Maryland, Baltimore, MD

Abstract

Understanding how glycosylation affects protein structure, dynamics, and function is an emerging and challenging problem in biology. As a first step toward glycan modeling in the context of structural glycobiology, we have developed *Glycan Reader* and integrated it into the CHARMM-GUI, http://www.charmm-gui.org/input/glycan. *Glycan Reader* greatly simplifies the reading of PDB structure files containing glycans through (i) detection of carbohydrate molecules, (ii) automatic annotation of carbohydrates based on their three-dimensional structures, (iii) recognition of glycosidic linkages between carbohydrates as well as N-/O-glycosidic linkages to proteins, and (iv) generation of inputs for the biomolecular simulation program CHARMM with the proper glycosidic linkage setup. In addition, *Glycan Reader* is linked to other functional modules in CHARMM-GUI, allowing users to easily generate carbohydrate or glycoprotein molecular simulation systems in solution or membrane environments and visualize the electrostatic potential on glycoprotein surfaces. These tools are useful for studying the impact of glycosylation on protein structure and dynamics.

Keywords

Molecular Dynamics; Electrostatic Surface; Membrane; Visualization

INTRODUCTION

Glycosylation is the most common post-translational modification process in proteins, and over half of all secreted proteins are expected to be glycosylated.[1–3] In addition to being a common protein appendage, glycans are also important in that they may alter protein structure and dynamics, and thus modify enzyme activity, protein-protein interactions, and the *in vivo* circulation half-life of protein pharmaceuticals.[4–8] Glycans are also involved in specific interaction with glycan-binding proteins and play a role in molecular, including cellular, recognition.[7] At this time, however, it is difficult to understand, on a case-by-case basis, which glycans are important components of protein function and specific recognition, and how to modify those glycans to optimize the properties of interest. To be able to predict a glycan's impact on the glycosylated protein's function and specific interaction with other

*Corresponding author: Phone: (785) 864-1993; Fax: (785) 864-5558; wonpil@ku.edu.

proteins, it is critical to understand the structure and the dynamics of glycans and the glycosylated protein.

The Protein Data Bank (PDB) is the largest database of biomolecular structures,[9] and, as of January 2011, the database contains about 71,000 PDB entries. Among those entries, about 6% of entries contain carbohydrate structures. Any type of biomolecular simulation begins with reading a protein structure into the simulation program. However, a task as simple as reading a PDB structure file into a molecular simulation, often becomes non-trivial when carbohydrates are present due to the inconsistency and complexity in the PDB file format for this class of molecules. Despite efforts made to standardize nomenclature and data structure for representing carbohydrates in the PDB, the current naming convention does not unambiguously identify anomeric configurations; it also contains other limitations. For example, GLC and BGC refer to α-D-glucose and β-D-glucose for glucopyranose, but GAL and GLA refer to β-D-galactose and α-D-galactose in the case of galactopyranose.[10] Such inconsistency in the nomenclature potentially leads to errors in annotating carbohydrates from the PDB. This problem is confounded by about 30% of carbohydrate structures in the PDB containing at least one error regarding the carbohydrate-type assignment.[11] Furthermore, there are cases where the entire carbohydrate chain is treated as a single residue, e.g., PDB:1AGM.[11] Thus, it is necessary to develop an algorithm that is able to automatically annotate carbohydrate structures based on their three-dimensional (3D) structures instead of relying on the PDB annotation.

There are several web-based toolsets for structural glycobiology presently available. The GLYCOSCIENCES.de web portal (http://www.glycoscience.de)[12,13] and the Glycoconjugate Data Bank (http://www.glycostructures.jp)[14] offer convenient ways to automatically annotate carbohydrates in PDB files. In addition, a 3D model of a carbohydrate or glycoprotein structure can be generated through web-based tools such as SWEET (http://www.glycosciences.de/modeling/sweet2)[15], GlyProt (http://www.glycosciences.de/modeling/glyprot)[16], and Carbohydrate Builder (http://glycam.ccrc.uga.edu/ccrc/carbohydrates).[17] However, a significant effort is required to prepare glycoprotein or protein/glycan complex systems since the interfaces offer rather limited options for glycan structure generation and lack the ability to prepare the generated structures for biomolecular simulations.

Motivated by the above limitations and needs, we have developed *Glycan Reader* and its web-based interface (http://www.charmm-gui.org/input/glycan). *Glycan Reader* greatly simplifies the reading of PDB structure files with glycans by (i) detection of carbohydrates, (ii) automatic annotation of carbohydrates based on their 3D structures, (iii) recognition of glycosidic linkages between carbohydrates as well as N-/O-glycosidic linkages to proteins, and (iv) generation of inputs for the biomolecular simulation program CHARMM[18] with proper glycosidic linkage setup. In addition, *Glycan Reader* is linked to other functional modules in the CHARMM-GUI (http://www.charmm-gui.org),[19] allowing users to easily generate protein/carbohydrate complexes or glycoprotein molecular simulation systems in solution or membrane environments and visualize the electrostatic potential on glycoprotein surfaces. These tools are useful for studying the impact of glycosylation on protein structure and dynamics. *Glycan Reader* utilizes the recently developed CHARMM carbohydrate force field,[20–23] which includes a wide range of furanose and pyranose monosaccharides and glycosidic linkages including N-/O-glycosidic linkages to proteins and lipids.

In the next section, *Glycan Reader* and its web interface developments are described in detail. This is followed by some illustrations of *Glycan Reader*, such as PDB glycan statistics, electrostatic potential visualization on glycoprotein surfaces, and preparation of carbohydrate and glycoprotein simulation systems in both aqueous and membrane

environments. Future directions of the *Glycan Reader* development project are then discussed briefly.

METHODS

To annotate glycans in a given PDB file, *Glycan Reader* uses an algorithm that can detect carbohydrate-like molecules and assign correct carbohydrate types based on their molecular topology and 3D structures. The overall scheme in *Glycan Reader* is shown in Figure 1 and illustrated in Figure 2. Molecular topologies are built based on the HETATM records and CONECT records in a PDB file, and molecules that do not have carbohydrate-like topology are not considered. The chemical groups that are attached to the carbohydrate-like molecules are then examined to assign the correct carbohydrate type. Once the monomeric units are identified, glycosidic linkage types are determined.

Automatic Detection and Assignment of Sugar Types

In the first step (Figs. 1A and 2A), *Glycan Reader* builds topologies of molecules in a PDB file based on both HETATM and CONECT records. Carbohydrate-like structures are identified by the presence of six- or five-membered rings that are composed of one oxygen atom and five or four carbon atoms depending on the size of the ring. Each potential carbohydrate molecule is further examined to identify the anomeric carbon by checking the carbon atoms connected to the ring oxygen to see if one of them has oxygen or nitrogen atom attached to it. If such an atom is found, the atom is designated as the anomeric carbon (C-1) and the rest of the ring constituents are re-numbered accordingly. However, in the case that no apparent anomeric carbon is found due to a lack of electron density or an error in the PDB structure, a carbon atom that is connected to the ring oxygen and has an exocyclic carbon atom attached to it, is designated as C-5 (for six-membered rings) or C-4 (for five-membered rings), and the other carbon atom attached to the ring oxygen is assigned as C-1. This method will fail to properly detect the anomeric carbon if no exocyclic carbons are attached to the C-5 (for pyranose) or C-4 (furanose) atom, which would be the case for xylopyranose. However, this small failure is not expected to be encountered very often since this is a failsafe algorithm, which is only used when the oxygen atom attached to the anomeric carbon is not found.

Once the anomeric carbon is assigned, *Glycan Reader* determines the carbohydrate residue type. Carbohydrate monomers can be classified by examining the configuration of the hydroxyl group attached to each carbon atom in the ring. Therefore, *Glycan Reader* calculates the improper angle based on the angle difference between the C_n–O_n ($\mathbf{a}_3$) and a vector perpendicular to C_{n-1}–C_n ($\mathbf{a}_1$) and C_n–C_{n+1} ($\mathbf{a}_2$), as shown in Fig. 2B. The configuration is then compared with a pre-established look-up table to determine the carbohydrate residue type. Currently, *Glycan Reader* can recognize all pyranose and furanoses available in the recent version of CHARMM carbohydrate force field,[20] and a few more carbohydrates that are not available at the date of publication (see Table 1 for the current list of available carbohydrate types).

Derivatized carbohydrates, such as N-acetyl-glucosamine, N-acetyl-nuraminic acid, or iduronic acid, are recognized by comparing the exocyclic chemical groups. The CHARMM carbohydrate force field field[20,21] provides separate residue definitions for such modification (e.g., acetylation, oxidation or deoxidation), and the residue names of such derivatized carbohydrates are renamed to the corresponding CHARMM residue names. When there is no residue definition available, e.g., no definition available for N-acetyl-mannosamine, *Glycan Reader* simply considers the residue as non-carbohydrate residue. As additional carbohydrate definitions become available in the CHARMM force field, they will be implemented in *Glycan Reader*.

Glycosidic Linkage Detection and Assignment

The anomeric position of each carbohydrate monomer is examined to check if the residue is connected to another carbohydrate by the glycosidic linkage. In our scheme, the root residue of a carbohydrate chain is simply assigned to a residue that has a free reducing end: for example, α-D-N-acetyl-glucose in Fig. 2D. N- or O-glycosylation is determined by cross-referencing the connected protein residue on the reducing end of the glycan chain; N-glycosylated when the reducing end is connected to ASN and O-glycosylated when the reducing end is connected to THR or SER. During the implementation, we frequently found incorrectly assigned bonds in glycan chains, which interfere with glycosidic linkage detection. For example, Figure 3A and 3B show incorrectly assigned bonds between neighboring residues possibly due to close proximity between two atoms, which forms a small ring structure and hinders the correct glycosidic linkage assignment. To assign glycosidic linkages reliably, each glycosidic linkage is reexamined to remove any chemical bonds that do not make chemical sense, e.g., oxygen atoms having three covalent bonds. On the other hand, there are some glycan chains that have missing glycosidic linkages (Fig. 3C). In such cases, *Glycan Reader* examines the distance between the anomeric carbon and the exocyclic oxygen on the neighboring residue; if it is in close proximity (e.g., < 2.5 Å), a glycosidic linkage is generated between the two residues. In rare occasions, covalent bonds with extreme bond lengths are present in the PDB (Fig. 3D); any chemical bonds that are longer than 5 Å will be removed in *Glycan Reader*. While these error correction features have been tested on a number of internal test cases, users are always advised to make sure that the input structure is correct and the output from the *Glycan Reader* is as intended. In the case that a carbohydrate chain is connected to a non-carbohydrate molecule, the entire chain is ignored presently. For instance, PDB:1S0J contains a ligand molecule that is a derivative of sialic acid with a methylumbelliferyl moiety, and *Glycan Reader* classifies the molecule as a non-carbohydrate molecule. While currently not implemented in an automated fashion, the potential to treat such moieties using the CHARMM General Force Field is possible.[24]

The CHARMM carbohydrate force field[20,21] provides several linkage types for mixed pyranose and furanose compounds, such as sucrose, lactulose, melezitose, raffinose, kestose, 6-kestose, isomaltulose, planteose, and nystose. This is because it is not possible to use the same linkage type between pyranose and furanose due to different atom types. Therefore, *Glycan Reader* detects the presence of mixed pyranose and furanose compounds, and uses appropriate linkage types to make glycosidic linkages between the pyranose and furanose residues.

GUI Implementation of Glycan Reader and CHARMM Input Generation

Glycan Reader has been integrated into the CHARMM-GUI web interface.[19] The user can either specify the PDB ID or upload the PDB structure into the server to generate the carbohydrate or protein/carbohydrate complex structure. If a carbohydrate is detected, then the graphical representation of the carbohydrate chain sequence will be displayed and the user can select the carbohydrate chains that they want to initialize in CHARMM (see Fig. 4). CHARMM allows modification in chemical structures, e.g., disulfide bond formation or phosphorylation using patch residues, and glycosidic linkages are generated using specific patch residues in CHARMM. The *Glycan Reader* web interface assigns the proper patches for glycosidic linkages and generates the CHARMM protein structure file (PSF) and coordinate files in both the PDB format and the CHARMM-specific coordinate format (CRD).

Currently, there are various patch residues available in the CHARMM carbohydrate force field to cover a range of carbohydrates including the majority found in eukaryotes.[25] For

example, O-methyl-, octyl-, dodecyl-, phosphate, and sulfate groups can be added to the reducing end of a sugar, and those modifications are properly patched in the PSF generation step (see Table 2 for the complete list of patch residues available). However, other types of common derivatizations, such as deoxidation, are not available, and, in such cases, the basic form of the carbohydrate molecule is used without modification and *Glycan Reader* informs the user. For example, if a user uploaded a structure of 2-deoxy glucose, a glucose molecule will be generated instead.

RESULTS AND ILLUSTRATIONS

PDB Glycan Statistics

We have used *Glycan Reader* to analyze the entire PDB database to obtain statistics on the available glycan-containing structures out of the total of 70947 structures in the PDB as of January, 2011. The results are summarized in Figure 5. There are a total of 4,029 PDB structure files (6.0%) that have at least one glycan chain, yielding a total of 15,669 glycan chains. A total of 8,848 glycan chains (56%) are N-glycosylated, 688 glycan chains (4.3%) are O-glycosylated, and the rest (6,133 chains, 39%) exist as noncovalently-bound ligands. Figure 5A shows the number of PDB structures with glycans deposited each year into the PDB; despite only 6% of PDB structures containing carbohydrate segments, the trend shows a steady increase over time. Figure 5B shows the distribution of the number of monosaccharides in the glycan chains found in the PDB, illustrating that most glycan structures in the PDB contain only one or two monosaccharides. The large number of shorter glycan chain could be due to the removal of glycans prior to structural studies or due to crystallization conditions. Our survey showed β-N-acetyl glucosamine (GlcNAc) is the most abundant monosaccharide (4917 entries) and GlcNAc $\beta(1\rightarrow4)$ GlcNAc β is the most abundant disaccharide (1653 entries). The present survey is focused on the number of structures in the PDB with glycan chains and the composition of those glycans. As *Glycan Reader* allows one to conveniently and reliably recognize glycan chains, it can be used for further studies on protein-carbohydrate interactions. To this end, development of a PDB glycan database to retrieve any specific glycan structure is currently in progress.

Electrostatic Potential Visualization

Characterizing the electrostatic potential on a macromolecular surface is becoming a routine practice in structural biophysics.[26] Similarly, comparing the electrostatic representations with and without glycans could provide insights into the biological roles of glycans. For example, some carbohydrates, such as sialic acid or phosphorylated sugars, are negatively charged and their spatial distribution might be important for protein function. The CHARMM-GUI *PBEQ-Solver* (www.charmm-gui.org/input/pbeqsolver) calculates the electrostatic potential and solvation free energy of biomolecules by solving the Poisson-Boltzmann (PB) equation using the CHARMM PBEQ facility.[27–29] Using the web based electrostatic visualization interface, a user can quickly calculate the electrostatic potential of a glycoprotein or protein/glycan complex (Figure 4). Currently, a generic set of atomic radii (i.e., 2.3 Å for carbon, 1.8 Å for oxygen, and 2.3 Å for nitrogen) is used for carbohydrate molecules to calculate the electrostatic potential surface. Efforts are ongoing in our laboratory to fine-tune such atomic radii allowing for their use in the context of implicit solvent models.

Simulations of Carbohydrates and Glycoproteins in Aqueous/Membrane Environment

MD simulations of biomolecules have become a common tool in the study of structural, dynamical, and energetic aspects of biological mechanisms.[30] The methodology for such simulations is well established and thus simulation input generation can be greatly simplified and automated once a PDB structure has been successfully read; this capability

was the motivation for the widely-used CHARMM-GUI *MD Simulator* (www.charmm-gui.org/input/mdsetup) and CHARMM-GUI *Membrane Builder* (www.charmm-gui.org/input/membrane).[31,32] Unlike protein structures, each monosaccharide unit in a glycan chain is connected by different glycosidic linkages (i.e., specific patch residues), which needs to be correctly recognized by the simulation package. Moreover, glycan chains may be branched, which makes the residue numbering non-contiguous. Such complexity in a glycan structure complicates manual linkage building, making it susceptible to error. *Glycan Reader* is integrated with various modules in CHARMM-GUI, such as *MD Simulator* and *Membrane Builder*, which facilitate preparation of glycoproteins for simulations in solution or membrane embedded environments (Figure 5). In addition to CHARMM, files produced by *Glycan Reader* may be used directly to perform simulations in NAMD,[33] and the capabilities exist to perform simulations using the CHARMM force fields in GROMACS[34] and AMBER.[35]

CONCLUDING DISCUSSION

We have developed a web-based tool, *Glycan Reader* (http://www.charmm-gui.org/input/glycan) that can automatically identify and annotate carbohydrate based on atomic coordinates, atom types, and bonds in a PDB structure file. *Glycan Reader* reliably detects the carbohydrate molecules, assigns their configuration and identifies the glycosidic linkages between monosaccharides or monosaccharides and proteins. These capabilities will facilitate computational studies of glycoprotein or protein/glycan complexes. *Glycan Reader* may also be used during the determination of protein structures that contain carbohydrates to verify that the residue types and the chemical bonds are correctly assigned. It is integrated into the CHARMM-GUI website as the *Glycan Reader* module and cross-linked to various modules in CHARMM-GUI. For example, one can use *PBEQ Solver* (http://www.charmm-gui.org/input/pbeqsolver) to calculate and visualize the electrostatic potential surface of glycoprotein. In addition, one can use *MD Simulator* (http://www.charmm-gui.org/input/mdsetup) or *Membrane Builder* (http://www.charmm-gui.org/input/membrane) to quickly generate MD simulation systems of glycoproteins and protein-carbohydrate complexes in aqueous or lipid bilayer environment.

Acknowledgments

This work was supported by the University of Kansas General Research Fund allocation #2301388-003, Kanas-COBRE NIH P20 RR-17708, TeraGrid resources provided by Purdue University (NSF OCI-0503992) (to WI), NIH R01-GM070855 (to ADM), and NIH R01-RR026061 (to HD).

References

1. Apweiler R, Hermjakob H, Sharon N. Biochim Biophys Acta. 1999; 1473(1):4–8. [PubMed: 10580125]
2. Moens S, Vanderleyden J. Arch Microbiol. 1997; 168(3):169–175. [PubMed: 9382700]
3. Varki, A. Essentials of glycobiology. Cold Spring Harbor Laboratory Press; Cold Spring Harbor, N.Y: 2009.
4. Morell AG, Gregoriadis G, Scheinberg IH, Hickman J, Ashwell G. J Biol Chem. 1971; 246(5): 1461–1467. [PubMed: 5545089]
5. Baenziger JU. Am J Pathol. 1985; 121(3):382–391. [PubMed: 3934981]
6. Varki A. Glycobiology. 1993; 3(2):97–130. [PubMed: 8490246]
7. Rudd PM, Wormald MR, Dwek RA. Trends Biotechnol. 2004; 22(10):524–530. [PubMed: 15450746]
8. Morelle W, Michalski JC. Curr Pharm Des. 2005; 11(20):2615–2645. [PubMed: 16101462]

9. Berman HM, Battistuz T, Bhat TN, Bluhm WF, Bourne PE, Burkhardt K, Feng Z, Gilliland GL, Iype L, Jain S, Fagan P, Marvin J, Padilla D, Ravichandran V, Schneider B, Thanki N, Weissig H, Westbrook JD, Zardecki C. Acta Crystallogr D Biol Crystallogr. 2002; 58(Pt 6 No 1):899–907. [PubMed: 12037327]
10. Feng Z, Chen L, Maddula H, Akcan O, Oughtred R, Berman HM, Westbrook J. Bioinformatics. 2004; 20(13):2153–2155. [PubMed: 15059838]
11. Lutteke T, Frank M, von der Lieth CW. Carbohydr Res. 2004; 339(5):1015–1020. [PubMed: 15010309]
12. Lutteke T, Frank M, von der Lieth CW. Nucleic Acids Res. 2005; 33(Database issue):D242–246. [PubMed: 15608187]
13. Lutteke T, Bohne-Lang A, Loss A, Goetz T, Frank M, von der Lieth CW. Glycobiology. 2006; 16(5):71R–81R.
14. Nakahara T, Hashimoto R, Nakagawa H, Monde K, Miura N, Nishimura S. Nucleic Acids Res. 2008; 36(Database issue):D368–371. [PubMed: 17933765]
15. Bohne A, Lang E, von der Lieth CW. Bioinformatics. 1999; 15(9):767–768. [PubMed: 10498779]
16. Bohne-Lang A, von der Lieth CW. Nucleic Acids Res. 2005; 33(Web Server issue):W214–219. [PubMed: 15980456]
17. Woods Group. Complex Carbohydrate Research Center. University of Georgia; Athens, GA: p. 2005-2011.
18. Brooks BR, Brooks CL III, MacKerell AD Jr, Nilsson L, Petrella RJ, Roux B, Won Y, Archontis G, Bartels C, Boresch S, Caflisch A, Caves L, Cui Q, Dinner AR, Feig M, Fischer S, Gao J, Hodoscek M, Im W, Kuczera K, Lazaridis T, Ma J, Ovchinnikov V, Paci E, Pastor RW, Post CB, Pu JZ, Schaefer M, Tidor B, Venable RM, Woodcock HL, Wu X, Yang W, York DM, Karplus M. J of Comput Chem. 2009; 30(10):1545–1614. [PubMed: 19444816]
19. Jo S, Kim T, Iyer VG, Im W. J Comput Chem. 2008; 29(11):1859–1865. [PubMed: 18351591]
20. Guvench O, Greene SN, Kamath G, Brady JW, Venable RM, Pastor RW, MacKerell AD Jr. J Comput Chem. 2008; 29(15):2543–2564. [PubMed: 18470966]
21. Guvench O, Hatcher ER, Venable RM, Pastor RW, MacKerell AD Jr. J Chem Theory Comput. 2009; 5(9):2353–2370. [PubMed: 20161005]
22. Hatcher E, Guvench O, MacKerell AD Jr. J Phys Chem B. 2009; 113(37):12466–12476. [PubMed: 19694450]
23. Raman EP, Guvench O, MacKerell AD Jr. J Phys Chem B. 2010; 114(40):12981–12994. [PubMed: 20845956]
24. Vanommeslaeghe K, Hatcher E, Acharya C, Kundu S, Zhong S, Shim J, Darian E, Guvench O, Lopes P, Vorobyov I, MacKerell AD Jr. J Comput Chem. 2010; 31(4):671–690. [PubMed: 19575467]
25. Guvench O, Mallajosyula SS, Raman EP, Hatcher E, Vanommeslaeghe K, Foster TJ, Jamison FW II, Brady JW, Venable RM, Pastor RW, MacKerell AD Jr. Manuscript (submitted for publication).
26. Honig B, Nicholls A. Science. 1995; 268:1144–1149. [PubMed: 7761829]
27. Nina M, Beglov D, Roux B. J Phys Chem B. 1997; 101:5239–5248.
28. Roux B. Biophys J. 1997; 73:2980–2989. [PubMed: 9414213]
29. Im W, Beglov D, Roux B. Comput Phys Comm. 1998; 111:59–75.
30. Levy RM, Gallicchio E. Annu Rev Phys Chem. 1998; 49:531–567. [PubMed: 9933909]
31. Jo S, Kim T, Im W. PLoS ONE. 2007; 2(9):e880. [PubMed: 17849009]
32. Jo S, Lim JB, Klauda JB, Im W. Biophys J. 2009; 97:50–58. [PubMed: 19580743]
33. Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel RD, Kale L, Schulten KJ. Comput Chem. 2005; 26(16):1781–1802.
34. Van Der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE, Berendsen HJ. J Comput Chem. 2005; 26(16):1701–1718. [PubMed: 16211538]
35. Case DA, Cheatham TE III, Darden T, Gohlke H, Luo R, Merz KM Jr, Onufriev A, Simmerling C, Wang B, Woods RJ. J Comput Chem. 2005; 26(16):1668–1688. [PubMed: 16200636]
36. ChemAxon Ltd. 2007.

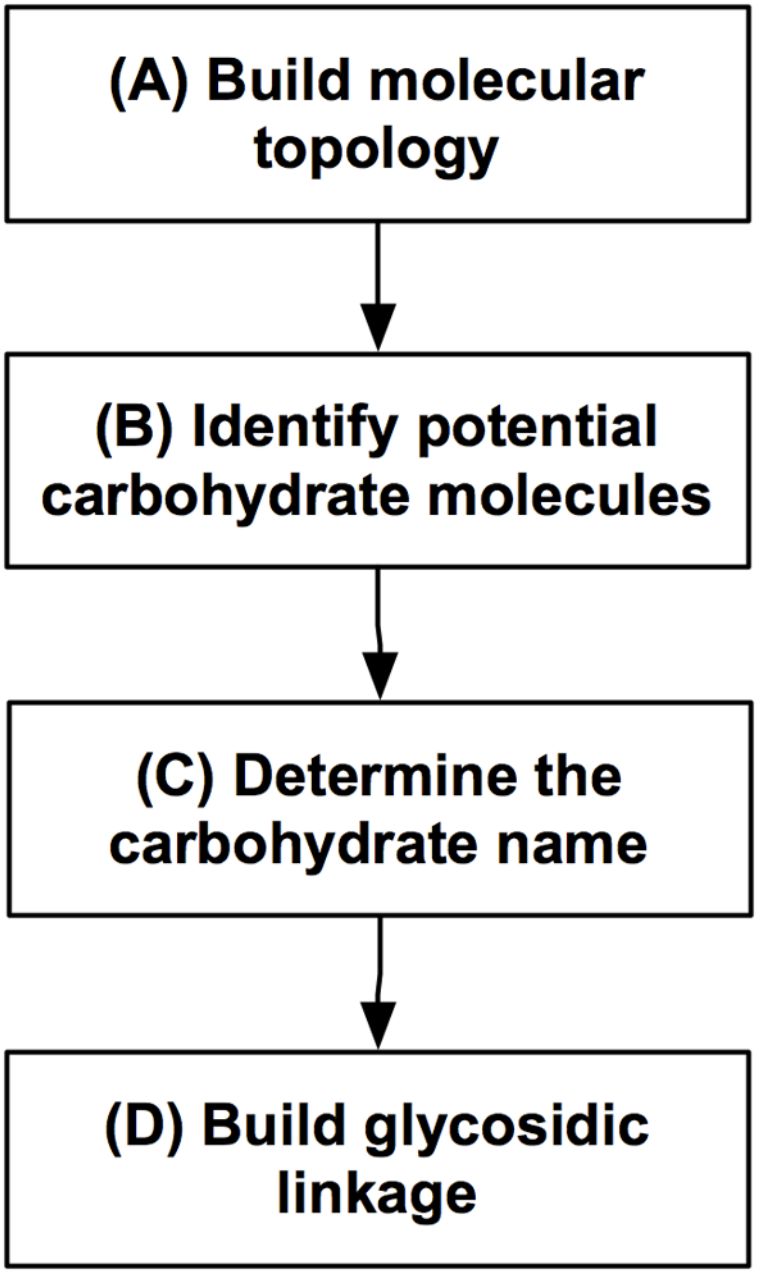

Figure 1.
Overview of carbohydrate annotation procedure in *Glycan Reader*.

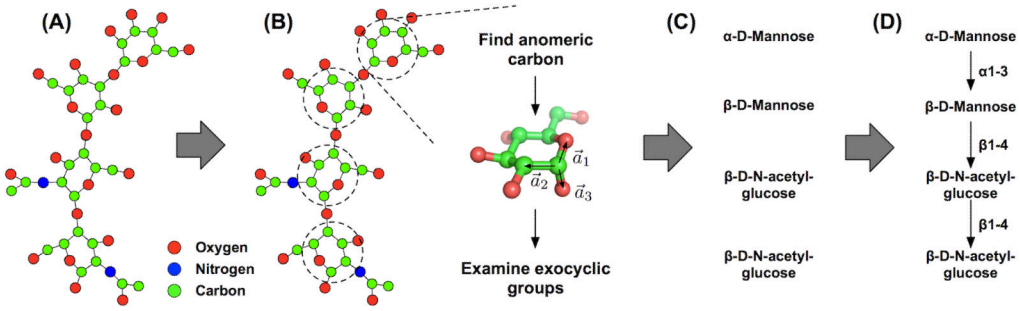

Figure 2.

Illustration of carbohydrate annotation procedure in *Glycan Reader*. (A) Molecular topology is built using HETATM and CONECT records in a PDB file. (B) Potential carbohydrate molecules are examined for anomeric carbon, stereochemistry of each ring carbon atoms, and exocyclic groups. (C) Carbohydrate type is annotated. (D) Glycosidic linkages are assigned between monosaccharides.

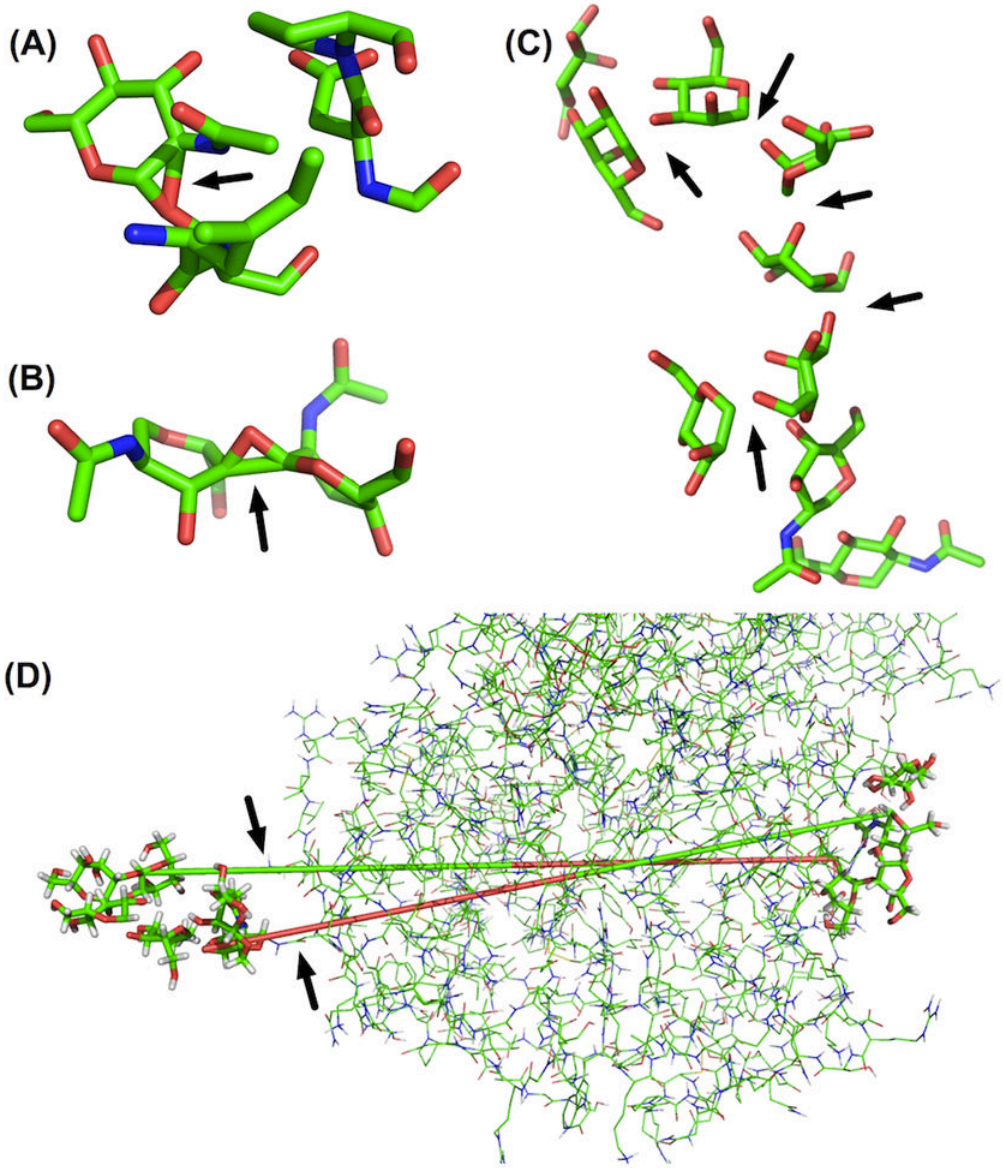

Figure 3.
Examples of erroneous glycan chains in PDB. (A) Additional bond between the glycosidic oxygen and the ring oxygen (PDB:1Q5C) (B) Additional carbon-carbon bond between two residues (PDB:1BCR). (C) Missing glycosidic linkages (PDB:2H6O) (D) Incorrect connectivity between two atoms (PDB:1INH). Erroneous bonds are marked by arrows.

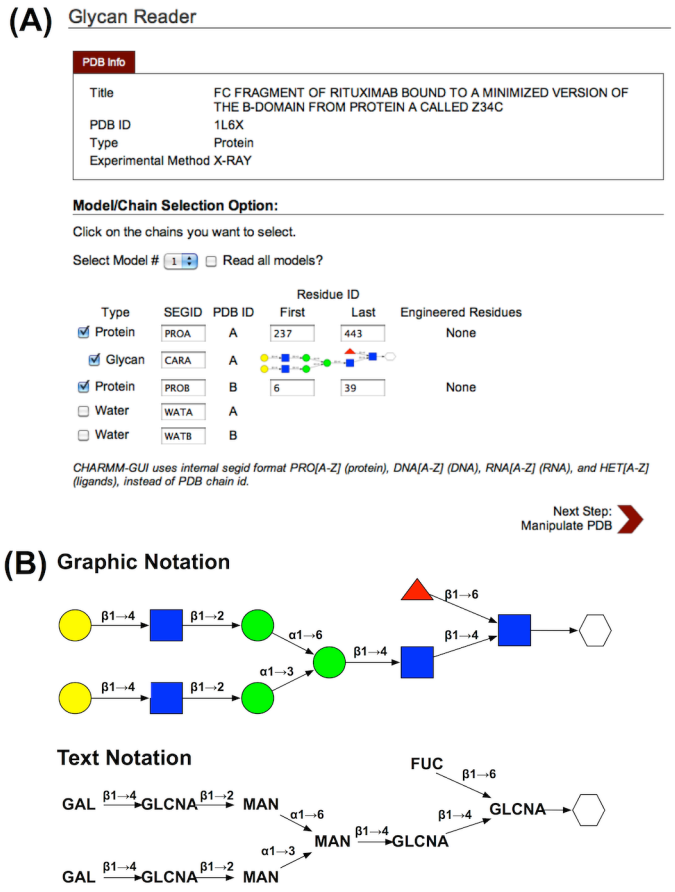

Figure 4.

Snapshot from CHARMM-GUI *Glycan Reader*. (A) When a glycan chain is found in a PDB, the sequence of the identified glycan chain is displayed. (B) When the sequence diagram is clicked, more detailed information on the glycan sequence is displayed in a popup window.

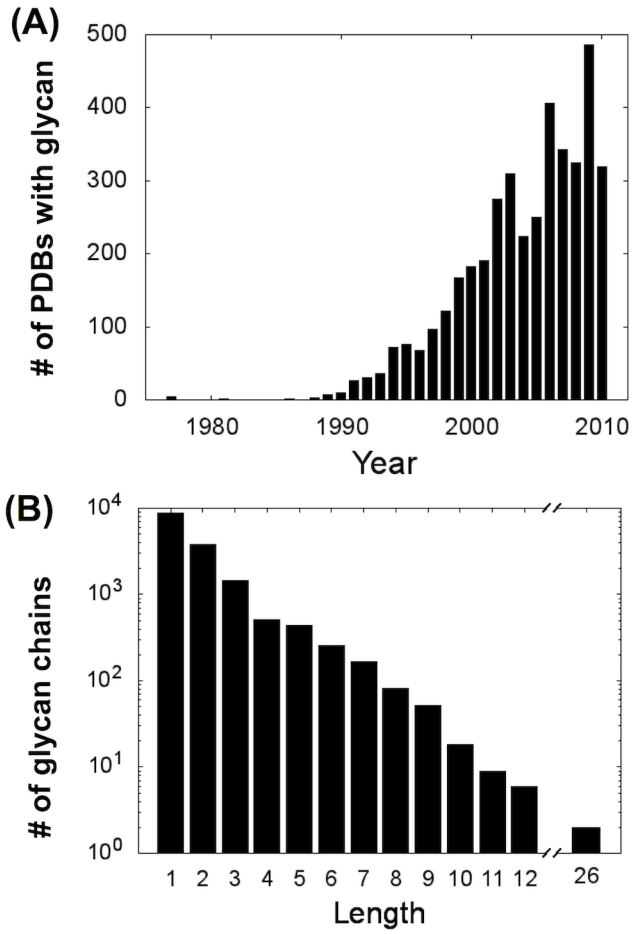

Figure 5.
PDB glycan statistics. (A) The number of structures with glycans added to the PDB each year. (B) The number of glycan chains with respect to the glycan chain length.

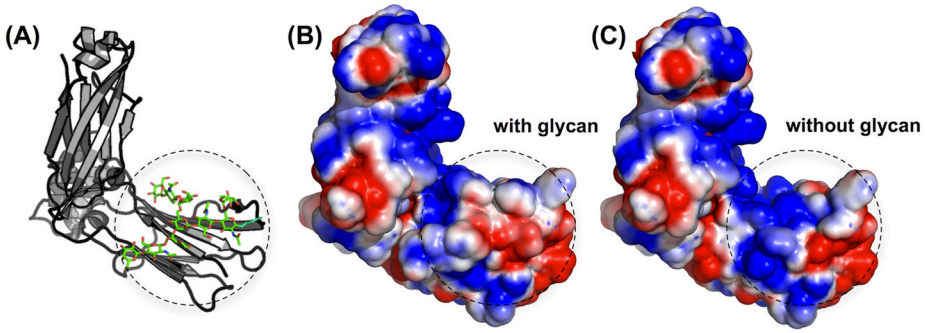

Figure 6.

Molecular images of a glycoprotein and its electrostatic potential surface with and without glycan. (A) A molecular image of PDB:1L6X (constant region of immunoglobulin G-1). (B) The electrostatic potential surface of the glycoprotein. (C) The electrostatic potential surface of the glycoprotein without glycan. The glycans on the surface are highlighted with dotted circles. Users can do various renderings of the images on the web using the *MarvinSpace* tools,[36] or using PyMOL.

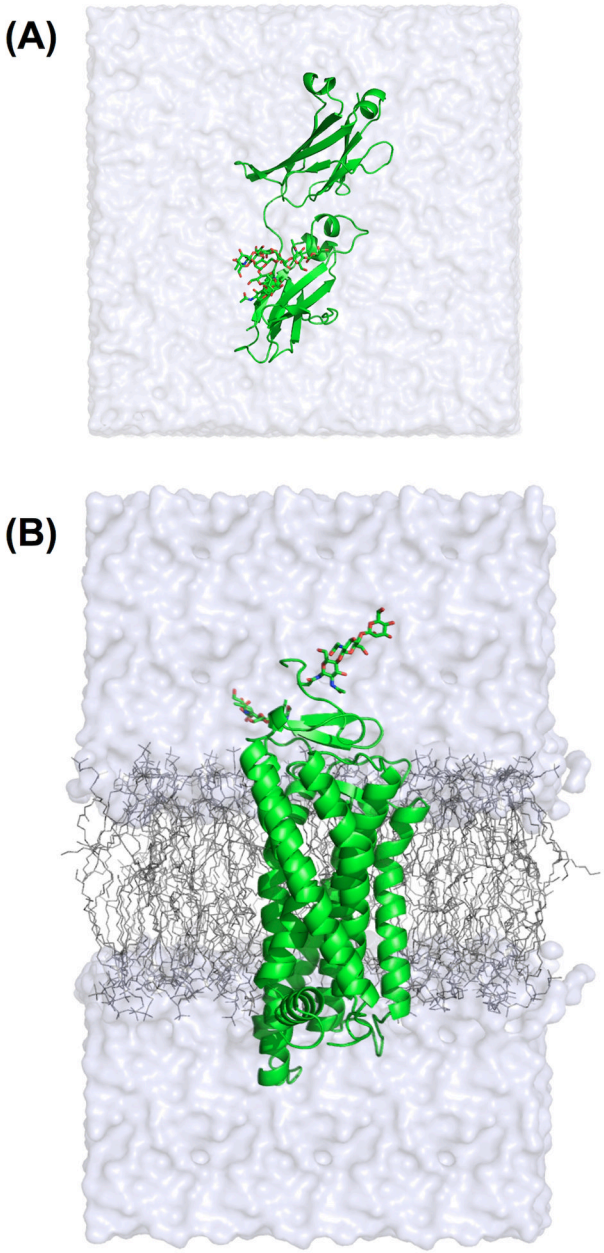

Figure 7.
Images of glycoprotein simulation systems. (A) Constant region of immunoglobulin-1 (PDB:1L6X) in aqueous environment prepared using the CHARMM-GUI *MD Simulator* (B) Bovine rhodopsin (PDB:1GZM) in a lipid bilayer environment prepared using the CHARMM-GUI *Membrane Builder*.

Table 1

List of carbohydrates recognized by *Glycan Reader*.

Chemical Name	CHARMM Residue Name
D-glucose	GLC
D-altrose	ALT
D-allose	ALL
D-galactose	GAL
D-gulose	GUL
D-idose	IDO
D-mannose	MAN
D-talose	TAL
D-xylose	XYL
L-fucose	FUC
L-rhamnose	RHM
D-glucoronic acid	GLCA
L-iduronic acid	IDOA
N-acetyl-D-glucosamine	GLCNA
N-acetyl-D-galactosamine	GALNA
N-acetyl-D-nuraminic acid	NE5AC
Tetrahydropyran (THP)	THP2
deoxy-ribose	DEO
Ribose	RIB
Arabinose	ARB
Lyxofuranose	LYF
Xylofuranose	XYF
Fructofuranose	FRU

Table 2

List of modifications recognized by *Glycan Reader*. The types of residues that are available for the modification are given in parenthesis.

Modification	CHARMM Patch Residue
O-methyl- at C1	OME (pyranose) FOME (furanose)
Octyl- at C1	OCT (pyranose)
Dodecyl- at C1	DDM (pyranose)
Phosphate at C1	PH (THP)
Phosphate at C2	PH2 (THP)
Sulfate at C1	SH (THP)