Workshop on Molecular Graphics and Visual Analysis of Molecular Data (MolVA) (2020)
J. Byška, M. Krone, and B. Sommer (Editors)

Improved *Umbrella Visualization* implemented in UnityMol gives valuable insight on sugar/protein interplay

C. Besançon[1*], H. Wong[1*], R. Rao[1], M. Dauchez[1,2], N. Belloy[1,2], J. Prévoteau-Jonquet[1,2] and S. Baud[1,2]

[*] These authors contributed equally to this work
[1]Université de Reims Champagne Ardenne, CNRS UMR 7369, Matrice Extracellulaire et Dynamique Cellulaire, Reims, France
[2]P3M, Multi-scale Molecular Modeling Plateform, Université de Reims Champagne Ardenne, Reims, France

Abstract

Among the various post-translational modifications, N-glycosylations are particularly important. They are linked to asparagine residues and their function as well as the one of the protein can be altered by modifications such as sialic acid hydrolysis. Since in vitro studies of N-glycans can be a challenging process (glycosylation chains have a great diversity and contain many reactive groups), in silico characterization using molecular dynamics simulation seems to be a good tool capable of overcoming experimental shortcomings thanks to exhaustive conformational samplings. In this paper, the Umbrella Visualization, a recent implementation into the molecular viewer UnityMol, is presented. This new and original visualization method is offering the possibility to follow and decipher the dynamics of very flexible sugar chains and enable the identification of the protein surface covered and potentially impacted by glycans. The latest module, described here and integrated within the Umbrella Visualization, complements the original statistical approach and allows for a meaningful description of glycan/protein interplay by combining, with shadow mapping, labelling, and hydrophobic properties of the surrounding aminoacids.

CCS Concepts

• ***Computing methodologies*** → ***Graphics file formats; Scientific visualization; Modeling methodologies; Texturing;***

1. Introduction

Post-translational modifications are enzymatic or chemical modifications of proteins. These changes have consequences on proteins activity, their degradation, their anchoring to membranes or other proteins, and thus on the cellular machineries both for normal and pathological cells. N- and O- glycosylations correspond to the addition of carbohydrate moieties to Asparagine and Serine residues, respectively. Glycans are very flexible molecular objects and their characterization at the experimental level is still difficult. Glycans can be branched or undergo modifications which contribute to enrich their structural and conformational diversity.

1.1. Related works

Many efforts have been made to allow the identification of glycosylation sites [HH08, GJB04], the analysis of experimental structures [CBJ*12] or the building of glycosylated systems [Woo05, JSD*11, CMR*18]. From a molecular graphics point of view, glycans visualization tools are implemented in software such as PyMol (*via* the plugin Azahar [AVM16]), Chimera or VMD (*via* tcl-scripting [THSW16]). Sweet UnityMol is an extension of the UnityMol platform, a molecular editor based on the Unity3D game engine, with representation modes dedicated to the visualization of sugar moieties [PTIB14]. Chimera, VMD, and Sweet UnityMol use the 3D-SNFG nomenclature [VCA*15, NAKB*19] which tends to standardize the representation of monosaccharides using graphic primitives and a color code.

Even if glycans simulation remains a challenging task, forcefields dedicated to the simulation of carbohydrates mechanics and dynamics are now available for the most commonly used simulation engines (Charmm, NAMD, Gromacs, Amber) at the atomic scale in Charmm [GMR*11] and Glycam [KYT*08] or as coarse-grained models in Martini [LRdV*09]. Standard and enhanced *in silico* methods generate thousands of frames that could be difficult to analyze or depict. In 1991, Mazurier et al [MDV*91] introduced the description of monofucosylated bi-antennary glycan as *bird*, *broken wing*, and *back folded* conformations. More generally, the description of glycans conformations relies on the analysis of torsion angle distributions of isolated individual linkages, pairwise RMSD calculations toward a reference structure, clustering or similarity measurements [JLSI13, LJM*15].

1.2. Motivation and Contribution

The conformational space sampled by N-glycans may be restrained by protein-carbohydrate interactions. The local protein topology

DOI: 10.2312/molva.20201097

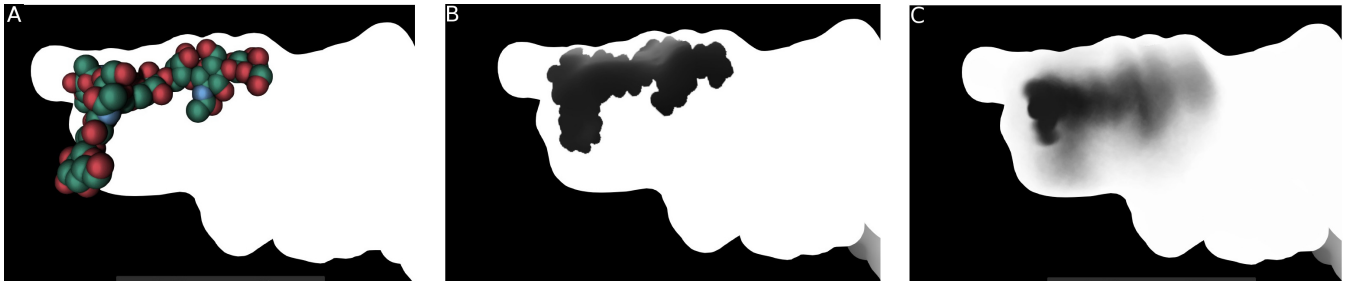

Figure 1: *From all-atoms sugar representation to Umbrella Visualization: the first implementation in UnityMol offers the possibility to highlight the position of a bi-antennary glycan on the surface of the insulin receptor (IR) (position 893) at the beginning of the molecular dynamics (MD) simulation (A), to project the corresponding shadow on the protein surface (B) and finally to integrate statistical information related to the coverage of the protein surface by the glycan during the MD simulation (C). For clarity, the surface of the IR was rendered using lighting that suppresses any effect of depth.*

may also sterically restrict the N-glycans to explore certain conformations. To rationalize this, the *Umbrella Visualization* framework was previously introduced in order to characterize how a N-glycan may act as an umbrella over the protein surface. Indeed, its ability to explore a covered area at the protein surface by using 2D density plots [GDB*16] was described. This 2D methodology was then extended using a UnityMol scene that combines a camera and a projector in a Unity's Prefab to map 3D density plots [BGB*18, BGB*20]. While analysis methods introduced in section 1.1 are focused on the geometry of the glycan itself, the approach described in the following part of the work integrates a graphical tool taking into account the macromolecule carrying the glycan, as well as the recent addition of a stereographic projection.

2. Developments and Implementation details

In this section, the principle and implementation of *3D Umbrella Visualization* described in the original paper by Besançon *et al* [BGB*20] are summarized, and then the latest added features are presented.

2.1. 3D *Umbrella Visualization*

Classical protein and glycan's representation modes include surface and van der Waals descriptions (Figure 1A).
The *3D Umbrella Visualization* Unity's Prefab is a template consisting of a plane, a directional light, a camera, and a projector, coded in C#. The α-carbon of the asparagine and δ-nitrogen engaged in the glycosidic bond define an Oz axis, normal to a plane (Ox, Oy). When the plane is instantiated, the normal to the plane is a vertical vector of coordinates (0,0,1). The normal to the center of the plane is used to compute the rotation angle applied to the Prefab. The plane is centered on the asparagine's α-carbon. The directional light casts the glycan's shadow on the plane. The camera records it in a Render Texture created and updated at run time. The Render Texture is finally used in the projector to display the glycan's shadow on the surface of the protein (Figure 1B).
The area covered during a molecular dynamics (MD) trajectory is computed as follows: for each frame, the Render Texture is recorded by the camera and converted into a Texture2D, then added to an array of pixels. This information is accumulated over the MD trajectory and normalized between 0 and 255. At the end of the procedure, the array represents the number of times a pixel has been darkened by the glycan's shadow. The array of pixels is eventually converted into a grayscale Render Texture and projected as previously described. The result is a statistically relevant representation of the glycan's coverage along the whole trajectory. The area covered reflects the spatial distribution of conformations (Figure 1C).

2.2. Contour map

The statistical distribution of conformations of glycans during a trajectory may seem complex to analyze. The very numerous nuances of gray gives an imprecise outline of the covered protein surface (Figure 1C and Figure 2A). This result can be simplified with the definition of a subset of eight intervals which clearly represent delimited areas in the form of density curves. For each interval, the assigned gray level is defined as detailed in the Table 1. Each point is therefore represented by the interval to which it corresponds. A contour representation mode or contour map is thus obtained; the delimitation between the zone impacted by the glycan and the rest of the protein appears distinctly, clearly defining the zone of influence of the glycan (Figure 2B).

Interval	0-31	32-63	64-95	96-127
RGB values	0	32	64	96
Color				
Interval	128-159	160-191	192-223	224-255
RGB values	128	160	192	255
Color				

Table 1: *Intervals for contour maps representation: the darker the color, the more the pixel (and thus the protein surface it represents) is darkened by the glycan during the MD trajectory.*

By presenting the results associated to the umbrella of glycans in the form of contour maps, the legibility of the results is greatly enhanced. In addition, the identification of the residues impacted by the presence of glycans on the surface of the protein is also improved and facilitated.

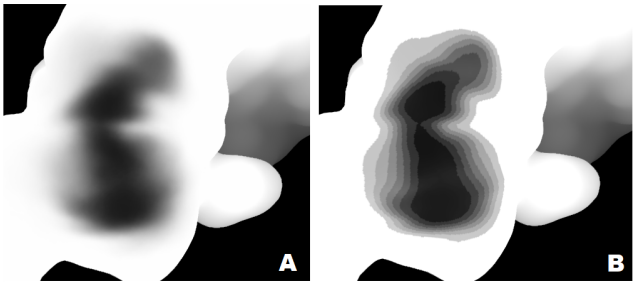

Figure 2: *Computation of contour maps improves Umbrella Visualization rendering: surface covered by a glycan either with the classical display using a scale of 256 shades of gray (A) or with the contour display mode that relies on a more restricted scale of 8 shades of gray (B). The darker parts correspond to the most covered area during the trajectory. Extraction and average of 12,500 frames related to glycosylated IR MD simulation.*

2.3. 2D density map and projection

Umbrella Visualization has been extended to implement new graphical methods allowing the user to gain deeper insight into the interactions between the glycan and the protein. Unity3D flexibility makes it easy to add interactive functionality for the end user. As such, a GUI button has been added to enable or disable the projection functionality.

2.3.1. Projection on a sphere and stereographic projection

As shown in Figure 3A, a sphere is used to detect the atoms of the molecule (they can belong either to the protein or to the glycan) within a given radius defined by the user. The sphere is centered on the α-carbon of the glycosylated asparagine, similarly to the *Umbrella Visualization*'s Prefab. For each atom, a vector is defined with the center of the sphere, normalized and scaled to the collision sphere radius. Once all the atoms are mapped to the sphere surface, a stereographic projection is applied. The α-carbon/δ-nitrogen axis is used to define the "north pole" of the sphere. This pole serves as point of origin to project any point P within the sphere to the corresponding P" point of a plane normal to the α-carbon/δ-nitrogen axis (Figure 3B).

The overlay of *Umbrella Visualization* and atoms projection is depicted in Figure 4. This representation allows the user to identify protein atoms at the immediate vicinity of the glycan, potentially indicating a region affected by the spatial proximity of the glycan, either for a specific conformation (Figure 4A) or for an ensemble of conformations (Figure 4B).

2.3.2. Physicochemical properties and atoms mapping

Umbrella Visualization gives the user the ability to discuss glycan's behavior on protein surface by combining, with shadow mapping, labelling, and physicochemical properties of the surrounding amino acids such as hydrophobicity. For instance, in Figure 3B, the color code used to represent the plane is associated to the Kyte-Doolittle scale of hydrophobicity [KD82]. Similar data representation could

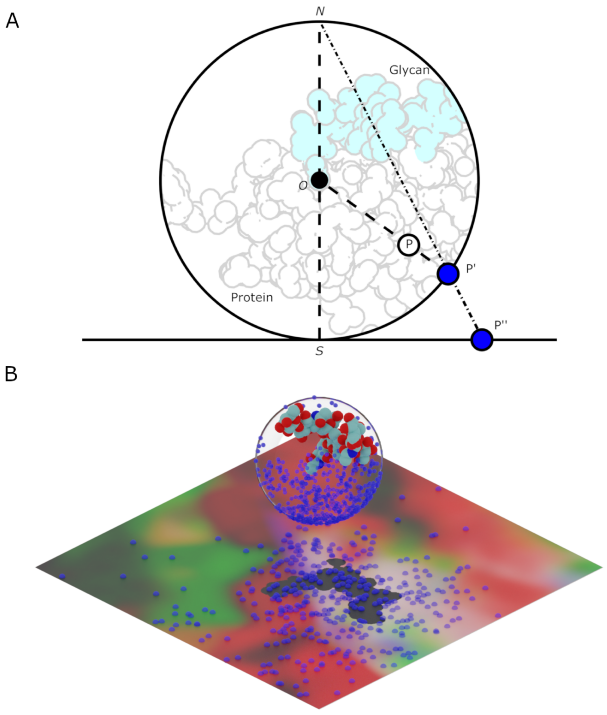

Figure 3: *Atoms projection and shadow mapping: within a sphere of a given radius and centered on the α-carbon of the glycosylated asparagin O, the position of each atom P is projected on the surface of the sphere (P'). Then a 2D map point P" is obtained using the position of the "north pole" N and projecting P' on a plane normal to the ON axis (A). The combination of the stereographic projection with the Umbrella Visualization glycan shadow projection allows for interesting 3D rendering where different types of information are superimposed (B). In the present case, the plane is colored according to the Kyte-Doolittle hydrophobocity scale of the underlying residues: white = hydrophobic, green = hydrophilic, red and blue = charged.*

be obtained for electrostatic isocontours/potentials calculated with software such as APBS [JES*18], Molecular Hydrophobicity Potential [ECP*07] or any other properties that can be mapped on the protein surface.

3. Results

The Insulin Receptor (IR) regulates glucose homeostasis. It has 18 potential N-glycosylation sites, mainly on the extracellular region: 14 on the α subunit and 4 on the β subunit. The essential role of glycosylations relates to the folding of the protein, its dimerization or its affinity for its ligand: insulin. The clustering performed on the MD trajectory of isolated glycan chains have demonstrated that in the presence of sialic acids, the bi-antennary glycans [GDB*16] (with and without fucose on the glycan core) preferentially adopt a *Broken Wing* conformation, with the α1-6 arm folded along the core. Without sialic acids, these glycans mainly adopt a *Bird* conformation, with the branches extended towards the solvent. For tri-

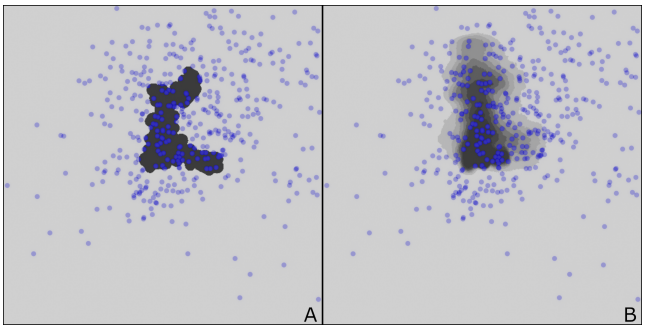

Figure 4: *The combination of Umbrella Visualization with stereographic projection allow the obtention of overlaid information on the plane: the position of the shadow corresponding to one frame of simulation (A) or corresponding to the statistics associated to 250 frames (B) can be compared to the stereographically projected atoms (blue dots) of the protein.*

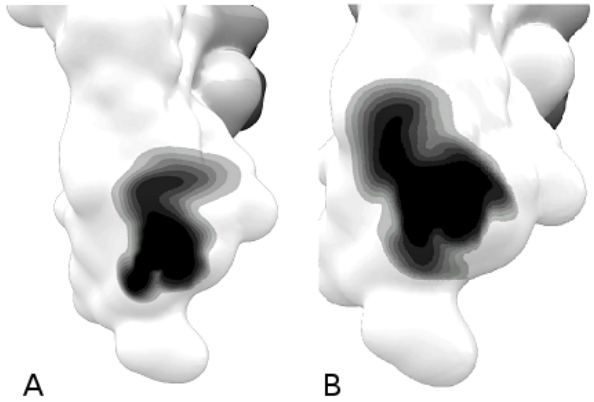

Figure 5: *Results obtained with Umbrella Visualization for fucosylated glycans without (A) and with sialic acids (B). Stabilization of the glycans in the presence of sialic acid results in a more extended zone of strong intensity compared to the case without sialic acid. Extraction and average of 12,500 frames related to glycosylated IR MD simulation.*

antennary glycans, the clustering suggested that the simultaneous presence of sialic acids and fucose leads to the presence of more compact conformations, with branches huddled together close to the core of the glycan.

250 ns long MD trajectories describing the behavior of the IR glycosylated on the asparagine 893 (Figure 5) have been used to validate the *Umbrella Visualization* approach. In each of the simulations, the IR (in an heterotetramer form) carries two glycans. We compared the most representative clusters extracted from the trajectories to the shade statistically obtained over the entire trajectory (12,500 frames). As it could be expected, in all cases, the cluster representing the most populated conformation corresponds to the shape of the shadow cast on the surface of the protein. However, the use of contour maps makes it possible to take into account the less represented conformations and the transitions between the conformations. *Umbrella Visualization* highlights the modulation of the surface statistically covered by a bi-antennary glycan depending on its sialilation state: as illustrated in Figure 5, sialic acids increase the amount of surface covered by the glycans. In addition the combined stereographic projection allows to colocalize atoms from negatively charge residues (glutamate and aspartate) with the inner core and branches of the glycans. These findings are key elements in deciphering the sugar/protein interactions and will help understand how the charge and steric hindrance linked to sialic acids could modify the interactions with the protein and, to a certain extent, the establishment of interactions with other protein partners.

4. Conclusions and Future works

We propose to extend the capabilities of *Umbrella Visualization* in the UnityMol molecular viewer. After migrating the code developed in an earlier version of the software, new rendering modes have been added. In terms of performance, the analysis of the trajectories (almost 30 Go for the full system and 1.2 Go for the protein and glycans in the case of the IR) can require up to 30 minutes of calculation on a desktop computer which is ultimately a relatively short time compared to the time devoted to molecular dynamics simulations which can extend over several weeks or even months on HPC resources. This extension of the UnityMol viewer offers a new visual way to analyze complex phenomena complementary to geometrical considerations. A link with experimental data could be adressed by dissecting accessibility to residues involved in protein/protein or in glycan-mediated interactions. The approach, as previously described, is efficient on globular domains of proteins. A drawback identified and not deeply investigated yet is the bias and the distortion of data induced by the spherical projection. Indeed, for systems where the glycan is in a cavity or close to a sinuous shape (in glycan-binding proteins, *i.e.* lectins, for example), nearby atoms will be detected and projected near the "north pole" of the sphere and consequently could not be projected on the plane. The first alternative would be to let the user choose among different primitives (other than a sphere) used for the projection in such a way that the shape of the selected primitive is more compatible with the topography of the considered domain. A second alternative would be to center the primitive in another way on individual carbohydrates moieties.

The approach developed in this paper is flexible and could be easily extended to other post-translational modifications since it only requires the detection of modified residues.

Acknowledgements

This work was supported by grants from the Region Grand Est and the European Regional Development Fund. The authors thank the ROMEO HPC Center, the Multiscale Molecular Modeling Platform (P3M) and the CINES (GENCI grant #c2016077569) for providing CPU-time and support. The chair MAgICS is acknowledged for financial and technical support. The authors thank Dr M. Baaden and Dr X. Martinez for fruitful discussions and technical support with UnityMol.

References

[AVM16] ARROYUELO A., VILA J. A., MARTIN O. A.: Azahar: a PyMOL plugin for construction, visualization and analysis of glycan molecules. *J. Comput. Aid. Mol. Des. 30*, 8 (2016), 619–624. doi:10.1007/s10822-016-9944-x. 1

[BGB*18] BESANÇON C., GUILLOT A., BLAISE S., DAUCHEZ M., BELLOY N., PRÉVOTEAU-JONQUET J., BAUD S.: New visualization of dynamical flexibility of n-glycans: Umbrella visualization in unitymol. In *IEEE International Conference on Bioinformatics and Biomedicine, BIBM 2018, Madrid, Spain, December 3-6, 2018* (2018), pp. 291–298. doi:10.1109/BIBM.2018.8621256.

[BGB*20] BESANÇON C., GUILLOT A., BLAISE S., DAUCHEZ M., BELLOY N., PRÉVOTEAU-JONQUET J., BAUD S.: Umbrella Visualization: A method of analysis dedicated to glycan flexibility with UnityMol. *Methods 173* (2020), 94–104. doi:10.1016/j.ymeth.2019.07.010. 2

[CBJ*12] CHUANG G.-Y., BOYINGTON J. C., JOYCE M. G., ZHU J., NABEL G. J., KWONG P. D., GEORGIEV I.: Computational prediction of N-linked glycosylation incorporating structural properties and patterns. *Bioinformatics 28*, 17 (Sept. 2012), 2249–2255. doi:10.1093/bioinformatics/bts426. 1

[CMR*18] CLERC O., MARIETHOZ J., RIVET A., LISACEK F., PÉREZ S., RICARD-BLUM S.: A pipeline to translate glycosaminoglycan sequences into 3D models. Application to the exploration of glycosaminoglycan conformational space. *Glycobiology 29*, 1 (09 2018), 36–44. doi:10.1093/glycob/cwy084. 1

[ECP*07] EFREMOV R. G., CHUGUNOV A. O., PYRKOV T. V., PRIESTLE J. P., ARSENIEV A. S., JACOBY E.: Molecular lipophilicity in protein modeling and drug design. *Curr Med Chem. 14*, 4 (2007), 393–415. doi:10.2174/092986707779941050. 3

[GDB*16] GUILLOT A., DAUCHEZ M., BELLOY N., JONQUET J., DUCA L., ROMIER B., MAURICE P., DEBELLE L., MARTINY L., DURLACH V., BAUD S., BLAISE S.: Impact of sialic acids on the molecular dynamic of bi-antennary and tri-antennary glycans. *Sci. Rep. 6* (Oct. 2016), 35666. doi:10.1038/srep35666. 2, 3

[GJB04] GUPTA R., JUNG E., BRUNAK S.: Prediction of n-glycosylation sites in human proteins. *In preparation 46* (Jan. 2004), 203–206. 1

[GMR*11] GUVENCH O., MALLAJOSYULA S. S., RAMAN E. P., HATCHER E., VANOMMESLAEGHE K., FOSTER T. J., JAMISON F. W., MACKERELL A. D.: Charmm additive all-atom force field for carbohydrate derivatives and its utility in polysaccharide and carbohydrate–protein modeling. *Journal of Chemical Theory and Computation 7*, 10 (2011), 3162–3180. doi:10.1021/ct200328p. 1

[HH08] HAMBY S. E., HIRST J. D.: Prediction of glycosylation sites using random forests. *BMC bioinformatics 9* (Nov. 2008), 500. doi:10.1186/1471-2105-9-500. 1

[JES*18] JURRUS E., ENGEL D., STAR K., MONSON K., BRANDI J., FELBERG L. E., BROOKES D. H., WILSON L., CHEN J., LILES K., CHUN M., LI P., GOHARA D. W., DOLINSKY T., KONECNY R., KOES D. R., NIELSEN J. E., HEAD-GORDON T., GENG W., KRASNY R., WEI G.-W., HOLST M. J., MCCAMMON J. A., BAKER N. A.: Improvements to the apbs biomolecular solvation software suite. *Protein Science 27*, 1 (2018), 112–128. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/pro.3280, arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1002/pro.3280, doi:10.1002/pro.3280. 3

[JLSI13] JO S., LEE H. S., SKOLNICK J., IM W.: Restricted n-glycan conformational space in the pdb and its implication in glycan structure modeling. *PLOS Computational Biology 9*, 3 (03 2013), 1–10. URL: https://doi.org/10.1371/journal.pcbi.1002946, doi:10.1371/journal.pcbi.1002946. 1

[JSD*11] JO S., SONG K. C., DESAIRE H., MACKERELL JR. A. D., IM W.: Glycan reader: Automated sugar identification and simulation preparation for carbohydrates and glycoproteins. *Journal of Computational Chemistry 32*, 14 (2011), 3135–3141. doi:10.1002/jcc.21886. 1

[KD82] KYTE J., DOOLITTLE R. F.: A simple method for displaying the hydropathic character of a protein. *J Mol Biol 157*, 1 (May 1982), 105–132. doi:10.1016/0022-2836(82)90515-0. 3

[KYT*08] KIRSCHNER K. N., YONGYE A. B., TSCHAMPEL S. M., GONZÁLEZ-OUTEIRIÑO J., DANIELS C. R., FOLEY B. L., WOODS R. J.: Glycam06: A generalizable biomolecular force field. carbohydrates. *Journal of Computational Chemistry 29*, 4 (2008), 622–655. doi:10.1002/jcc.20820. 1

[LJM*15] LEE H. S., JO S., MUKHERJEE S., PARK S.-J., SKOLNICK J., LEE J., IM W.: GS-align for glycan structure alignment and similarity measurement. *Bioinformatics 31*, 16 (04 2015), 2653–2659. doi:10.1093/bioinformatics/btv202. 1

[LRdV*09] LÓPEZ C. A., RZEPIELA A. J., DE VRIES A. H., DIJKHUIZEN L., HÜNENBERGER P. H., MARRINK S. J.: Martini coarse-grained force field: Extension to carbohydrates. *Journal of Chemical Theory and Computation 5*, 12 (2009), 3195–3210. doi:10.1021/ct900313w. 1

[MDV*91] MAZURIER J., DAUCHEZ M., VERGOTEN G., MONTREUIL J., SPIK G.: Molecular modeling of a disialylated monofucosylated biantennary glycan of the N-acetyllactosamine type. *Glycoconjugate Journal 8*, 5 (Oct. 1991), 390–9. 1

[NAKB*19] NEELAMEGHAM S., AOKI-KINOSHITA K., BOLTON E., FRANK M., LISACEK F., LÜTTEKE T., O'BOYLE N., PACKER N. H., STANLEY P., TOUKACH P., VARKI A., WOODS R. J., GROUP T. S. D.: Updates to the Symbol Nomenclature for Glycans guidelines. *Glycobiology 29*, 9 (06 2019), 620–624. doi:10.1093/glycob/cwz045. 1

[PTIB14] 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 25*, 5 (12 2014), 483–491. doi:10.1093/glycob/cwu133. 1

[THSW16] THIEKER D. F., HADDEN J. A., SCHULTEN K., WOODS R. J.: 3D implementation of the symbol nomenclature for graphical representation of glycans. *Glycobiology 26*, 8 (09 2016), 786–787. doi:10.1093/glycob/cww076. 1

[VCA*15] VARKI A., CUMMINGS R. D., AEBI M., PACKER N. H., SEEBERGER P. H., ESKO J. D., STANLEY P., HART G., DARVILL A., KINOSHITA T., PRESTEGARD J. J., SCHNAAR R. L., FREEZE H. H., MARTH J. D., BERTOZZI C. R., ETZLER M. E., FRANK M., VLIEGENTHART J. F., LÜTTEKE T., PEREZ S., BOLTON E., RUDD P., PAULSON J., KANEHISA M., TOUKACH P., AOKI-KINOSHITA K. F., DELL A., NARIMATSU H., YORK W., TANIGUCHI N., KORNFELD S.: Symbol Nomenclature for Graphical Representations of Glycans. *Glycobiology 25*, 12 (11 2015), 1323–1324. doi:10.1093/glycob/cwv091. 1

[Woo05] WOODS R.: Woods group. (2005-2020) glycam web. complex carbohydrate research center, university of georgia, athens, ga. (http://glycam.org), 2005. 1