

Azahar: a PyMOL plugin for construction, visualization and analysis of glycan molecules

Agustina Arroyuelo¹  · Jorge A. Vila¹ · Osvaldo A. Martin¹

Received: 10 June 2016 / Accepted: 17 August 2016
© Springer International Publishing Switzerland 2016

Abstract Glycans are key molecules in many physiological and pathological processes. As with other molecules, like proteins, visualization of the 3D structures of glycans adds valuable information for understanding their biological function. Hence, here we introduce Azahar, a computing environment for the creation, visualization and analysis of glycan molecules. Azahar is implemented in Python and works as a plugin for the well known PyMOL package (Schrodinger in The PyMOL molecular graphics system, version 1.3r1, 2010). Besides the already available visualization and analysis options provided by PyMOL, Azahar includes 3 cartoon-like representations and tools for 3D structure characterization such as a conformational search using a Monte Carlo with minimization routine and also tools to analyse single glycans or trajectories/ensembles including the calculation of radius of gyration, Ramachandran plots and hydrogen bonds. Azahar is freely available to download from <http://www.pymolwiki.org/index.php/Azahar> and the source code is available at <https://github.com/agustinaarroyuelo/Azahar>.

Keywords Glycans · Structural bioinformatics · Modelling

Introduction

There is an existing trend towards modelling complex molecular systems yet most of the molecular software packages are protein-centred [2]. Therefore, bioinformatics tools oriented to glycobiology are rare or scattered amongst diverse pieces of software [3]. This fact is not consistent with the evidence showing that glycoconjugates are involved in most disease processes like viral or bacterial infection, metastasis, autoimmunity or inflammation [4, 5], and that glycosylation is the most abundant post-translational modification both in terms of number of modified proteins and the associated molecular diversity [6]. Therefore, the understanding of glycan tertiary structure is tightly paired with the comprehension of the biological functions of glycoproteins. The lack of sufficient tools to deal with glycans may be due to the inherent complexity of these molecules compared to others, such as proteins. In fact, glycans have a greater number of stereocentres and are able to form branched-molecules in addition to the large variety of monomers available to form a polysaccharide [7]. In particular, for the construction and visualization of glycan molecules packages such as POLYS [8], GLYCAM-WEB [8], GlycanBuilder [9] and CarbBuilder [10], however, none of this software provides tools for the analysis of glycans nor the possibility to carry out conformational searches on glycan molecules, as Azahar does. Besides, Azahar has the advantageous feature of being a plugin for PyMOL a well-established molecular visualization piece of software. And hence most of the features already available at PyMOL, and through third party plugins, are instantly available for the Azahar user, thus providing a more integrated molecular visualization, construction and analysis experience to work with glycans and glyconconjugates, such as glyco-proteins.

✉ Agustina Arroyuelo
aarroyuelo@unsl.edu.ar

✉ Osvaldo A. Martin
omarti@unsl.edu.ar

¹ Instituto de Matemática Aplicada San Luis, IMASL, Universidad Nacional de San Luis and CONICET, D5700HHW San Luis, Argentina

Features

Code implementation

Coding was performed using the Python programming language [11] making use of scientific libraries, such as NumPy [12] and Matplotlib [13]. The Graphic User Interface (GUI) was written with Tkinter [14]. Azahar is implemented as a *plugin* for the PyMOL visualization software [1]. We chose PyMOL mainly because (a) it is a widely used open-source piece of software, and (b) it offers access, through the Python language, to several high-level functions, such as computation of torsional angles, a powerful selection algebra syntax, and high quality visualization options that facilitate the implementation of advanced features.

Glycan construction

Glycan construction can be done through the GUI by the following steps: (1) selecting residues from a list of 23 predefined monosaccharide templates, (2) specifying the number of residues to be added, (3) choosing glycosidic bond types and (4) choosing the positions of the branches. The topology of the molecule is stored in a human-readable file, in ASCII format, that can be edited. Alternatively this file can be provided by the user.

Azahar contains to date 23 monosaccharide templates (the full list is shown in Table 1) which are available to build glycan molecules through the plugin's GUI. Every template was written by following the guidelines set by the RCSB Protein Data Bank (PDB) regarding coordinate files, including atom and residue names [15]. For the three letter code required in the standard PDB format we used the one mentioned by Lütteke [16]. A priority was established to include the most frequently observed, and/or biologically relevant, saccharides in nature.

The existing template list in Azahar (see Table 1) is user-expandable by adding *.pdb* files into the template folder located in the plugin's installation directory. The newly added files should follow the guidelines mentioned above, otherwise the plugin could fail to work properly.

Glycan visualization

PyMOL does not currently include a *cartoon-like* representation for glycan molecules. To overcome this limitation Azahar includes 3 representations: *cartoon*, *wire* and *beads*. In each of them the individual atoms are hidden and the glycosidic rings and bonds are represented as simplified objects. The *cartoon* representation shows the glycan-molecule as a series of solid non-flat polygons (hexagons

Table 1 Full list of available monosaccharide templates

2-deoxy-2-fluoro- α -D-mannopyranose
5-N-acetyl- α -D-neuraminic-acid
α -D-desoxy-ribose
α -D-fructose
α -D-fucose
α -D-galactose-4-sulphate
α -D-galactose
α -D-glucose
α -D-mannose
α -D-N-ace-glucosamine
α -D-ribose
α -L-fucose
β -D-fructose
β -D-fucose
β -D-galactose-4-sulphate
β -D-galactose
β -D-glucosamine
β -D-glucose
β -D-mannose
β -L-fucose
octyl- β -D-glucopiranos
β -D-xylopiranos
β -D-N-acetylmuramic acid

for hexoses, pentagons for pentoses, etc) linked by rods (see Fig. 1); the *wire* representation is similar to the *cartoon* but the polygons are empty and the rods are thinner. Finally, in the *beads* representation each ring is replaced by a sphere. All 3 representations are built by using PyMOL's Compiled Graphic Objects (CGO), allowing the creation of

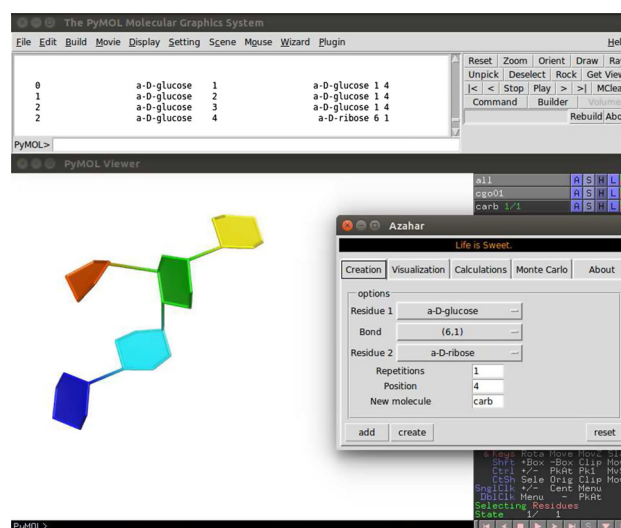


Fig. 1 A glycan molecule is shown using the *cartoon* representation

three-dimensional geometries which can be displayed at high-speed via OpenGL. CGO have some limitations that restrict their use, for example the colour of CGO objects has to be specified prior to the creation of the object. For that reason rewriting these *cartoon-like* representations in the C-source code of PyMOL will be done in the future.

Calculation of molecular properties

Since PyMOL can be used as a Python library, in principle, is possible to use it to compute any molecular property one wishes, of course this might be an overkill for users with no knowledge of programming or molecular biophysics. Therefore, Azahar provides convenient functions to compute and visualize molecular properties useful to study glycans: the radius of gyration, Ramachandran plots and hydrogen bonds. These properties can be computed on a single glycan or a trajectory, or *states* in PyMOL's parlance. For trajectories the user can compute these properties on sub-trajectories by choosing the starting and final states and also the step, i.e. to choose only every k states.

The radius of gyration is an important property for the description of polymers such as glycans. Its interpretation is related to molecule compactness [17]. Azahar offers this calculation in an automated way for any given glycan, protein and glycoprotein molecule. In addition, an optional visual representation of the results, as a semi-transparent sphere, is also available in Azahar.

Ramachandran plots are very useful for structural analysis but Ramachandran plots for glycans are not available "out-of-the-box" in PyMOL or other software, like VMD [18]. For the visualization of the Ramachandran plot we offer two options, a scatter and a *hex-bin* plot. When plotting a large number of points, scatter plots hide the structure of the data (like the localization of energy minima) because successive points are plotted one on top of the others. Alternatively, hex-bin plots are constructed by tessellating a 2D-space with hexagons and binning the data, and a colour gradient is used to display bins with a different number of binned data points.

The hydrogen bonds are visualized using dashed lines connecting the donor and acceptor atoms (this is the standard representation for hydrogen bonds used by PyMOL); additionally, a text summary is generated and saved on the user's working directory. This file contains information about the pairs of atoms involved and the occurrence of a hydrogen bond i.e. the percentage of the states a particular hydrogen bond is detected. The user can set a cut off to control which and hydrogen bonds are visualized and reported in the text file, hydrogen bonds with a occurrence lower than the cutoff will be excluded.

Conformational search

Azahar includes the possibility to carry out a conformational search of glycan molecules by using a Monte Carlo with Minimization (MCM) algorithm [19]. A MCM is like a regular Metropolis–Hastings algorithm [20] with the difference that an energy minimization is performed before comparing energies between the *new* and the *current* conformations. During the sampling of the conformational space the higher computational cost of the MCM algorithm, compared with a plain Metropolis–Hastings, is compensated by the higher acceptance/rejection ratio [21]. During the MCM simulations the internal energy of the molecule is computed with the MMFF94 force field [22] as implemented in OpenBabel [23]. As an option, during the conformational search, a solvation free-energy term can be included through a solvent-accessible surface area model [24, 25]. During the MCM the only variables are the ϕ , ψ , ω and χ torsional angles (See Fig. 2). In other words, the intracyclic angles, bond length and bond angles are considered fixed.

We test the MCM simulations on two simple glycan molecules: maltose and chitosan. Our MCM-results (see Figs. 4, 5) were compared against experimental data obtained from: (1) Glycosciences.de [27]; and (2) Molecular Dynamics simulations [28], respectively.

Results and discussion

Maltose conformational search

In order to test our implementation of the Monte Carlo with Minimization algorithm, we decided to study the conformational preferences of maltose (α -D-glucose-(1,4)- α -D-glucose). We chose this model for its simplicity and for the availability of experimental data, specifically a

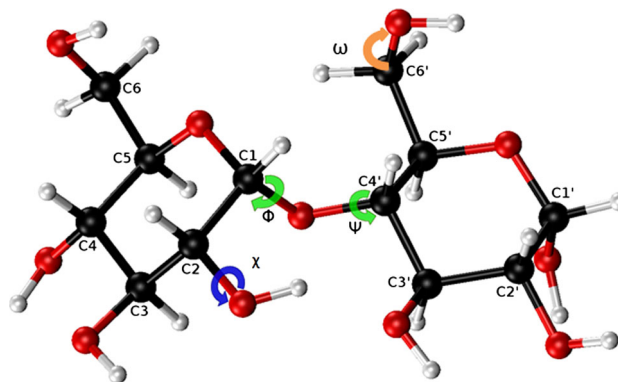


Fig. 2 The ϕ and ψ (green), ω (orange) and χ (blue) torsional angles shown in a maltose molecule [26]

Ramachandran plot for the molecule obtained from the database Glycoscience [29] (See Fig. 3). Four simulations for maltose, with and without solvation effects (Fig. 4, panels A and B respectively) were carried out and the results show that a single minimum exist located around the $(-175, 125)$ region of the Ramachandran map, in agreement with the observed data obtained from

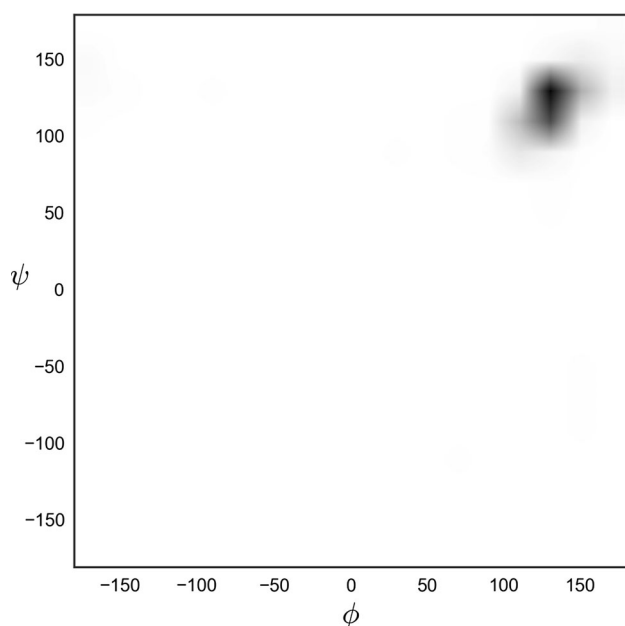


Fig. 3 Ramachandran plot for maltose retrieved from glycosciences.de [29]

glycoscience.de (see Fig. 3). Differences between simulations with and without solvent are negligible.

Chitosan conformational search

Chitosan is a natural linear polysaccharide, which consists of repetitions of β -(1,4)-glucosamine or β -(1,4)-N-acetylglucosamine. Its applications extend over several disciplines such as agriculture where its used as biopesticide and growth enhancer of plants [30] and medicine, where its ability to accelerate the clotting of blood allows its use as an antihemorrhagic agent [31]. It is commonly found in the lining of shrimp shells and other crustaceans as chitosan where is the major component. It is one of the known polymeric materials that biodegrade more easily and has been proposed as a bio-plastic [32].

We simulated a decamer of β -(1,4)-glucosamine in the same conditions to those applied to maltose. Four simulations were performed with and without solvent and the results are shown in panels A and B of Fig. 5, respectively.

The Ramachandran plot presented in Fig. 6 corresponds to Molecular Dynamics simulations for chitosan [28], by using Gromacs 4.0.3 [33] and applying the force field GROMOS 53a6 [34]. In these simulations, an explicit solvation model, a 0.25 M NaCl concentration and a temperature of 300 K was used during the simulations.

In Fig. 5 there is a single minima in the upper-right quadrant, but this is not observed in Fig. 6. One possible reason for this result is that the MCM simulation was carried out for a neutral molecule, while the MD simulation

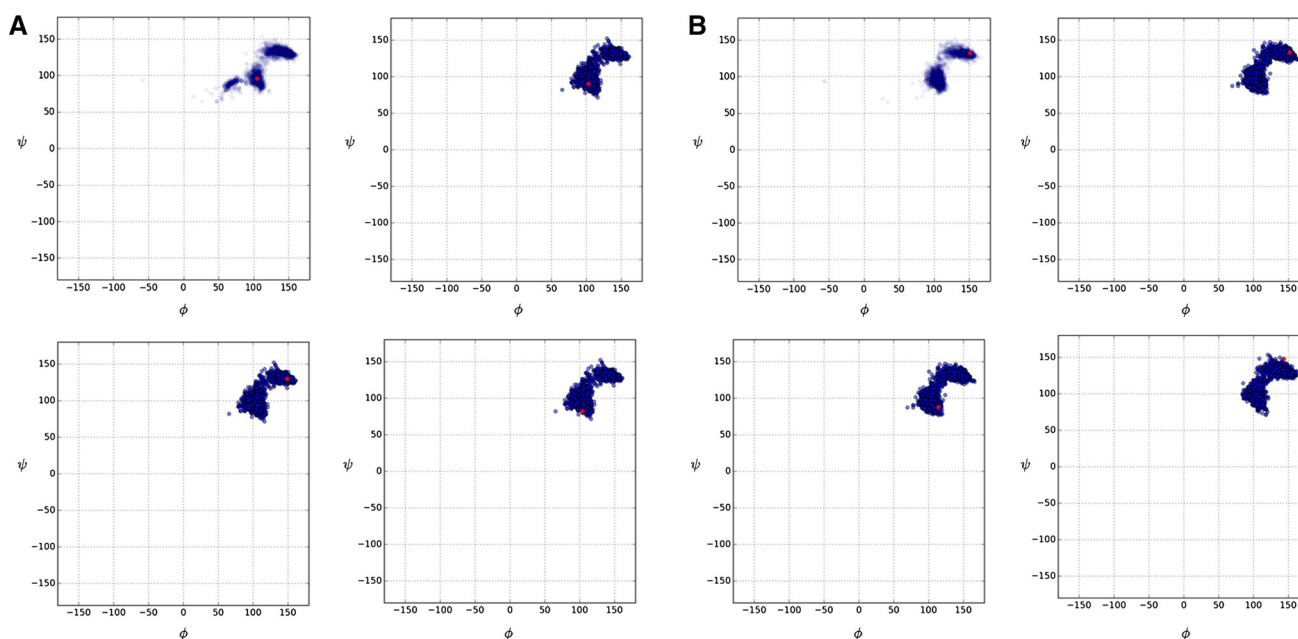


Fig. 4 Ramachandran plots obtained from four independent MCM simulations, for maltose with solvent (*panel A*) and without solvent (*panel B*). The *red dots* represent the lowest energy conformation for each run

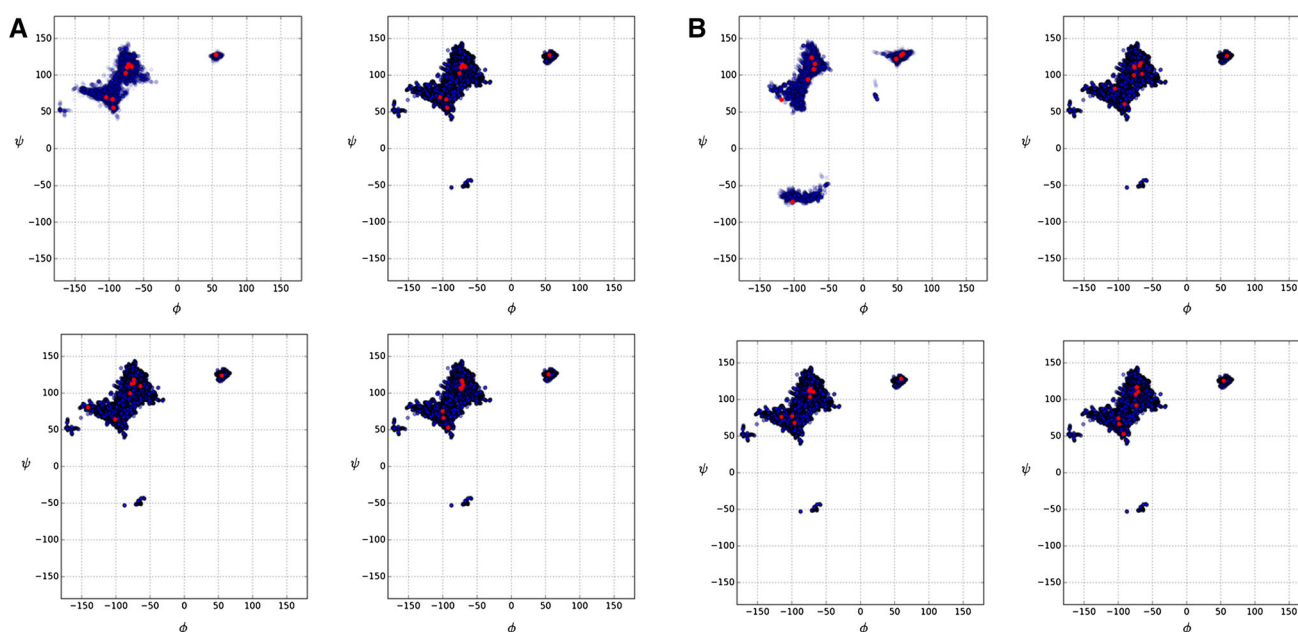


Fig. 5 Ramachandran plots obtained from four independent MCM simulations for a β -(1.4)-glucosamine decamer with solvent (*panel A*) and without solvent (*panel B*). The red dots represent the values of the 9 ϕ , ψ angles for the lowest energy conformation

was performed on a charged molecule. It is worth noting that consideration of charged groups during the MCM simulations would require to adopt a model of screening, otherwise there will be an over estimation of the electrostatic effect. Among several possible solutions to this problem, the Generalized Born model offers a fast (compared with the solution of the Poisson–Boltzmann equation) and accurate way to deal with solvent polarization effects when charged groups are considered. However, consideration of charged groups in Azahar, although a desirable feature, it is not possible with the current version of the code and, hence, out of the scope of the present work. Besides, the MD simulation at higher salt concentration, and hence, higher charge screening, shows the occurrence of a third minima in the upper-right quadrant (data not shown), consistent with our simulations. Additionally it is important to note that the MCM method is a global optimizer method, i.e. its principal aim is to find the global energy minima (and not to sample the correct distribution), while MD methods (and other types of Markov Chain Monte Carlo methods) are best suited to sample the whole conformational space.

Conclusion

In this work we introduce Azahar, a PyMOL plugin that provides an environment for glycan construction, visualization and analysis. Features currently implemented, and not available in the standard PyMOL package, include: (1)

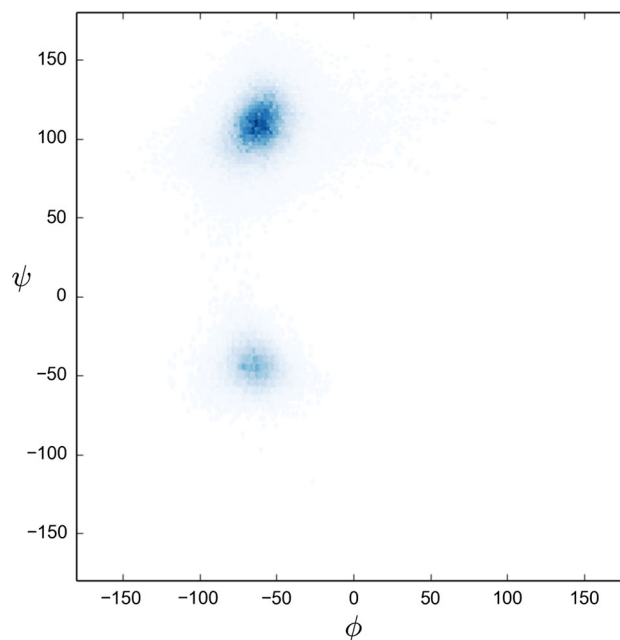


Fig. 6 Ramachandran plot for a charged decamer of β -(1.4)-glucosamine (using the hex-bin plot). Calculated for the conformation obtained from a molecular dynamics simulation as explained in the text. Other simulations gave similar results (data not shown)

a simple interface to build glycan molecules of arbitrary complexity based on a series of predefined monosaccharide templates; (2) the computation of the Radius of Gyration, Ramachandran plots and Hydrogen bonds (including an occurrence table for trajectories/ensembles) (3) cartoon

representations for glycans; and, (4) a conformational search routine, based on a MCM algorithm, with and without consideration of the solvent contribution to the total free-energy. The results of the performed MCM simulations for maltose and chitosan qualitatively reproduce the data obtained from: the site [glicosciences.de](http://www.glicosciences.de) and molecular dynamics simulations [28]. At this point it is worth noting that the contribution of the solvent to the total free-energy seems to be negligible indicating that the relative-weight between the internal energy and the solvent free-energy terms (assumed as 1.0 during the conformational search) might require optimization. The code is publicly available on github (<https://github.com/agusti-naarroyuelo/Azahar>), we encourage users to contribute to this project in the form of bug reports (using GitHub's issue tracker), requests of new features and of course new code to improve Azahar.

Acknowledgments This work was supported by: A Warren L. DeLano Memorial PyMOL Open-Source Fellowship; PIP-0030 from CONICET-Argentina; and PICT-2014-0556 from ANPCyT-Argentina.

References

- Schrodinger L (2010) The PyMOL molecular graphics system, version 1.3r1
- National Research Council (US) (2012) Transforming glycoscience: a roadmap for the future. The national academies collection: reports funded by National Institutes of Health. National Academies Press (US), Washington (DC). <http://www.ncbi.nlm.nih.gov/books/NBK109958/>
- Campbell MP, Ranzinger R, Lutteke T, Mariethoz J, Hayes CA, Zhang J, Akune Y, Aoki-Kinoshita KF, Damerell D, Carta G, York WS, Haslam SM, Narimatsu H, Rudd PM, Karlsson NG, Packer NH, Lisacek F (2014) BMC Bioinformatics 15(Suppl 1):S9. doi:10.1186/1471-2105-15-S1-S9. <http://www.biomedcentral.com/1471-2105/15/S1/S9/abstract>
- Ohtsubo K, Marth JD (2006) Cell 126(5):855. doi:10.1016/j.cell.2006.08.019
- Varki A, Cummings RD, Esko JD, Freeze HH, Stanley P, Bertozzi CR, Hart GW, Etzler ME (eds) (2009) Essentials of Glycobiology, 2nd edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor (NY). <http://www.ncbi.nlm.nih.gov/books/NBK1908/>
- Khoury GA, Baliban RC, Floudas CA (2011) Proteome-wide post-translational modification statistics: frequency analysis and curation of the swiss-prot database. Sci Rep 1:90. doi:10.1038/srep00090
- Ranzinger R, Frank M, Von der Lieth CW, Herget S (2009) Glycobiology 19(12):1563. doi:10.1093/glycob/cwp137. <http://glycob.oxfordjournals.org/content/19/12/1563>
- Engelsen SB, Hansen PI, Pérez S (2014) Biopolymers 101(7):733. doi:10.1002/bip.22449
- Ceroni A, Dell A, Haslam SM (2007) Source Code Biol Med 2(1):1. doi:10.1186/1751-0473-2-3
- Kuttel M, Mao Y, Widmalm G, Lundborg M (2011) CarbBuilder: an adjustable tool for building 3D molecular structures of carbohydrates for molecular simulation. In: Proceedings of the 7th IEEE International Conference e-Science, Dec 2011, pp 395–402. doi:10.1109/eScience.2011.61
- van Rossum G (1995) Python programming language version 2:7
- Walt SVD, Colbert SC, Varoquaux G (2011) Comput Sci Eng 13(2):22. doi:10.1109/MCSE.2011.37
- Hunter JD (2007) Comput Sci Eng 9(3):90. doi:10.1109/MCSE.2007.55
- Shipman JW (2013) Tkinter 8.5 reference: a gui for python. <http://infocost.nmt.edu/tcc/help/pubs/tkinter/tkinter.pdf>
- <http://www.wwpdb.org/documentation/file-format>
- Lütteke T, von der Lieth CW (2004) BMC Bioinform 5(1):1. doi:10.1186/1471-2105-5-69
- Lobanov MY, Bogatyreva NS, Galzitskaya OV (2008) Mol Biol 42(4):623. doi:10.1134/S0026893308040195
- Humphrey W, Dalke A, Schulten K (1996) J Mol Graph 14:33
- Li Z, Scheraga HA (1987) Proc Natl Acad Sci 84(19):6611. <http://www.pnas.org/content/84/19/6611.short>. Bibtext: li_monte_1987
- Hastings WK (1970) Biometrika 57(1):97. doi:10.1093/biomet/57.1.97
- Nayeem A, Vila J, Scheraga HA (1991) J Comput Chem 12(5):594. doi:10.1002/jcc.540120509. Bibtext: nayeem_comparative_1991
- Halgren TA (1996) J Comput Chem 17(5–6):490. doi:10.1002/(SICI)1096-987X(199604)17:5/6<490::AID-JCC1>3.0.CO;2-P
- O'Boyle N, Banck M, James C, Morley C, Vandermeersch T, Hutchison G (2011) J Cheminformatics 3(1):33. doi:10.1186/1758-2946-3-33. <http://www.jcheminf.com/content/3/1/33>
- Shrake A, Rupley J (1973) J Mol Biol 79(2):351–371
- Vila J, Williams RL, Vásquez M, Scheraga HA (1991) Proteins 10(3):199. doi:10.1002/prot.340100305
- Garay PG, Martin OA, Scheraga HA, Vila JA (2014) J Comput Chem 35(25):1854. doi:10.1002/jcc.23697
- Lutteke T, Frank M, von der Lieth CW (2005) Nucleic acids research 33(Database issue), D242. doi:10.1093/nar/gki013. Bibtext: lutteke_carbohydrate_2005
- Esteban C (2016) Doctoral thesis (personal communication)
- Frank TLM, von der Lieth CW (2007). <http://www.glycosciences.de/tools/glytorsion/>
- Hadwiger LA (2013) Multiple effects of chitosan on plant systems: solid science or hype. Plant Sci 208:42–49. doi:10.1016/j.plantsci.2013.03.007
- Lederle FA, Freischlag JA, Kyriakides TC et al (2009) Outcomes following endovascular vs open repair of abdominal aortic aneurysm: a randomized trial. JAMA 302(14):1535–1542. doi:10.1001/jama.2009.1426
- Fernandez JG, Ingber DE (2014) Macromol Mater Eng 299(8):932. doi:10.1002/mame.201300426
- Berendsen H, van der Spoel D, van Drunen R (1995) Comput Phys Commun 91(1–3):43. doi:10.1016/0010-4655(95)00042-E
- Oostenbrink C, Villa A, Mark AE, Van Gunsteren WF (2004) J Comput Chem 25(13):1656. doi:10.1002/jcc.20090