OPEN ACCESS Freely available online

PLoS COMPUTATIONAL BIOLOGY

Education

# Integrating Bioinformatics Tools to Handle Glycosylation

**Yuliet Mazola*, Glay Chinea, Alexis Musacchio**

Department of Bioinformatics, Center for Genetic Engineering and Biotechnology, Havana, Cuba

This is an original *PLoS Computational Biology* tutorial.

## Introduction

This tutorial is planned for biologists and computational biologists interested in bioinformatics applications to study protein glycosylation. Glycosylation is a co- and post-translational modification that involves the selective attachment of carbohydrates to proteins. The enhancement of glycosylation by applying glycoengineering strategies has become widely used to improve properties for protein therapeutics. In this tutorial, the use of bioinformatics to assist the rational design and insertion of *N*-glycosylation sites in proteins is described.

## Background

Glycosylation is a co- and post-translational modification involving the covalent addition of carbohydrates to proteins. Carbohydrates (also referred to as glycans, sugars, or saccharides) are adopting linear and branched structures and are composed of monosaccharides, which are covalently linked by glycosidic bonds. There are four enzymatic glycosylation processes: *N*-glycosylation, *O*-glycosylation, *C*-glycosylation (or *C*-mannosylation), and glycosylphosphatidylinositol (GPI) anchor (Figure 1). Glycan acceptor sites for each glycosylation type are described in Table 1. Experimental detection of occupied glycosylation sites in proteins is an expensive and laborious process [1]. Instead, a number of glycosylation prediction methods as well as glycan and glycoprotein analysis tools have been developed (Table 2 and Table 3). For a detailed description of glycobiology-related databases and software, including glycosylation predictors, the reader is referred to nice reviews on the subject [2–5].

## The Attractiveness of Modifying Protein Glycosylation

Of particular interest is the role of carbohydrates in modulating physicochemical and biological properties of proteins. Several glycosylation parameters are involved, including the number of glycans attached, the position of the glycosylation sites, and the glycan features (such as the molecular size, sequence, and charge). Glycan can influence protein function [6]; the presence of a glycosyl chain pointing toward a binding pocket might block such a cavity and hence, influence the ligand binding mode and affect protein biological activity (Figure 2). Carbohydrates can also increase protein stability and solubility, as well as reduce immunogenicity and susceptibility to proteolysis [7]. This explains why the rational manipulation of glycosylation parameters (glycoengineering) is widely applied to obtain proteins suited for therapeutic applications [8]. Glycoengineering can enhance in vivo activity even in proteins that do not normally contain *N*-glycosylation sites [9]. Some protein instabilities prevented by applying glycosylation engineering include proteolytic degradation, formation of crosslinked species, unfolding processes, oxidation, low solubility, aggregation, and kinetic inactivation [10].

## Rational Design and Insertion of *N*-glycan Sites in Proteins

One of the strategies used in glycoengineering involves the introduction of *N*-glycosylation sequons to increase carbohydrate content in protein pharmaceuticals [7]. In this tutorial, a workflow for the rational design and insertion of *N*-glycan sites into a desirable protein (also referred to as a target protein) using bioinformatics is provided (Figure 3). A detailed description of the workflow is given below. General features and availability of non-glycobiology-related bioinformatics resources can be found in Table 4.

The target protein amino acid sequence is the starting point in this analysis. Additional information, such as post-translational modifications, site-directed mutagenesis studies, and three-dimensional (3D) structure, are also helpful. These data can be found in the protein annotation and literature databases UniProtKB [11] and PubMed [12], respectively.

Prior to performing any modification to the target protein sequence, one should know the residues involved in protein function and tertiary structure. These residues should not be modified. In general, functional and structural relevant residues tend to be more conserved within a protein family [13]. Conserved residues are identified by multiple sequence alignment using, for example, the CLUSTALW server [14], analyzing the sequence similarity among the target protein and its homologues. In particular, a multiple sequence alignment with diverse and divergent protein homologue sequences is suggested, since residues conserved over a longer period of time are under stronger evolutionary constraints. The homologue proteins are recognized via a pairwise alignment using, for instance, the BLASTp server [15]. A degree of conservation for each aligned position in the multiple sequence alignment is quantified. At this step, available tools for sequence conservation analysis could be applied, like the AL2CO server [16]. The amino acid frequencies for each aligned position are estimated and the conservation index is calculated from those frequencies. As input for the AL2CO server, the multiple sequence alignment file is required. Optionally, if a Protein Data Bank (PDB) file (atomic coordinates) of the target or any related homologue protein is also uploaded, the

**Citation:** Mazola Y, Chinea G, Musacchio A (2011) Integrating Bioinformatics Tools to Handle Glycosylation. PLoS Comput Biol 7(12): e1002285. doi:10.1371/journal.pcbi.1002285

**Editor:** Fran Lewitter, Whitehead Institute, United States of America

**Published** December 29, 2011



**Funding:** The authors received no specific funding for this work.

**Competing Interests:** The authors have declared that no competing interests exist.

* E-mail: yuliet.mazola@cigb.edu.cu

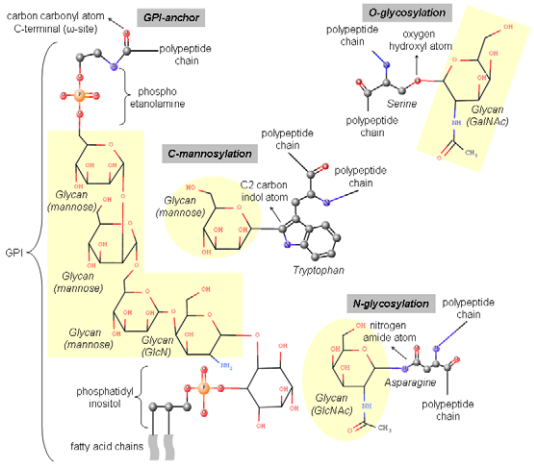


**Figure 1. Schematic representation of glycosylation forms.** For each glycosylation type, the amino acid acceptor site is illustrated in balls and sticks: *N*-glycosylation (asparagine residue), *O*-glycosylation (serine residue), *C*-mannosylation (tryptophan residue), and glycosylphosphatidylinositol (GPI) anchor (C-terminal protein residue). Small balls colored in grey, red, blue, and orange represent carbon, oxygen, nitrogen, and phosphorus atoms, respectively. Hydrogen atoms were not shown. The atoms involved in glycan linkage are indicated with rows. Glycan molecules are shown as sticks and highlighted with a yellow background color. The GPI molecule was divided into three parts: phosphoethanolamine, glycan core, and phosphatidylinositol. The glycan core is composed of one non-acetylated glucosamine (GlcN) and three mannose moieties. The long fatty acids contained in the phosphatidylinositol portion are indicated using waves.
doi:10.1371/journal.pcbi.1002285.g001

AL2CO server adds the calculated conservation indices into the output PDB file. Then, conserved motifs can be mapped onto the 3D structure and visualized with the Visual Molecular Dynamics (VMD) software [17].

We recommend the insertion of *N*-glycan sites, such as Asn-x-Ser/Thr, preferentially at positions where potential *N*-glycosylation sequons predominate in the homologue proteins. The prediction of *N*-glycosylation sites has to be done for the target and homologue proteins, and any of the available prediction servers, such as NetNGlyc, EnsembleGly, or GPP, can be used (Table 2). The GPP server input is the protein amino acid sequence and the output is sent by email. For NetNGlyc and EnsembleGly servers, the protein UniProtKB/Swiss-Prot accession number or primary amino acid sequences are accepted as input. Results are shown online and are easy to understand. Predicted *N*-glycan sites are mapped and scored onto the protein sequence representing the occurrence probability of *N*-glycosylation. In the case of NetNGlyc, the predicted Asn-x-Ser/Thr motifs are highlighted in red color, and a graph showing potential

**Table 1.** General features of different glycosylation types.

| Glycosylation Type | Glycosylation Sequences Motifs | Glycosylation Acceptor Site | Organism | Reference |
|---|---|---|---|---|
| *N*-glycosylation | In eukaryotes, glycan molecules are attached to the asparagine residue from sequons: Asn-x-Ser and Asn-x-Thr, or in some rare cases in Asn-x-Cys where x is not a proline residue. In prokaryotes, the sequon is extended to Asp/Glu-z-Asn-x-Ser and Asp/Glu-z-Asn-x-Thr, where x and z are not proline residues. | Nitrogen atom from the amide group in the asparagine residue | Eukaryotes and prokaryotes | [30,31] |
| *O*-glycosylation | No specific sequence motifs have been defined. Sugars are attached to serine and threonine residues usually found in a beta conformation and in close vicinity to proline residues. | Oxygen atom from the hydroxyl group in serine or threonine residues | Eukaryotes and prokaryotes | [32–34] |
| *C*-glycosylation | Carbohydrates are attached to the first tryptophan residue from the following motifs: Trp-x-x-Trp, Trp-x-x-Phe, Trp-x-x-Tyr, and Trp-x-x-Cys. Any amino acid could be placed at the x position, although small and/or polar residues are preferred, such as alanine, glycine, serine, and threonine. | Carbon atom (C2) from the indole group in the tryptophan residue | Eukaryotes except yeast | [35–39] |
| GPI anchor | A specific C-terminal signal sequence is recognized and cleaved, creating a new C-terminal protein end (ω-site). The GPI molecule is added to the ω-site. No consensus sequence for ω-site localization has been described. Typical residues in ω-site include: cysteine, aspartic acid, glycine, asparagine, and serine. | Carbon atom from the C-terminal carbonyl group at the ω-site | Eukaryotes and a reduced subset of archaea | [40–42] |

GPI, glycosylphosphatidylinositol.
doi:10.1371/journal.pcbi.1002285.t001

**Table 2.** Glycosylation prediction servers.

| Server | Portable Version | Method | Description | URL |
|---|---|---|---|---|
| NetNGlyc | Only for academics | ANN | *N*-glycosylation | http://www.cbs.dtu.dk/services/NetNGlyc/ |
| EnsembleGly | No | SVM | *N*-glycosylation | http://turing.cs.iastate.edu/EnsembleGly/ |
| EnsembleGly | No | SVM | *O*-glycosylation | http://turing.cs.iastate.edu/EnsembleGly/ |
| EnsembleGly | No | SVM | *C*-glycosylation | http://turing.cs.iastate.edu/EnsembleGly/ |
| GPP | No | RFM | *N*-glycosylation | http://comp.chem.nottingham.ac.uk/glyco/ |
| GPP | No | RFM | *O*-glycosylation | http://comp.chem.nottingham.ac.uk/glyco/ |
| NetOGlyc | Only for academics | ANN | *O*-glycosylation | http://www.cbs.dtu.dk/services/NetOGlyc/ |
| Oglyc | No | SVM | *O*-glycosylation | http://www.biosino.org/Oglyc/ |
| CKSAAP_OGlySite | No | SVM | *O*-glycosylation | http://bioinformatics.cau.edu.cn/zzd_lab/CKSAAP_OGlySite/ |
| YinOYang | Only for academics | ANN | *O*-glycosylation | http://www.cbs.dtu.dk/services/YinOYang/ |
| Big-PI | No | | GPI anchor | http://mendel.imp.ac.at/gpi/gpi_server.html |
| GPI-SOM | Yes | ANN | GPI anchor | http://gpi.unibe.ch/ |
| FragAnchor | No | ANN and HMM | GPI anchor | http://navet.ics.hawaii.edu/~fraganchor/NNHMM/NNHMM.html |
| PredGPI | Only for academics upon request | HMM and SVM | GPI anchor | http://gpcr.biocomp.unibo.it/predgpi/ |
| NetCGlyc | Only for academics | ANN | *C*-mannosylation | http://www.cbs.dtu.dk/services/NetCGlyc/ |

ANN, artificial neural network; SVM, support vector machine; RFM, random forest method; HMM, hidden Markov model.
doi:10.1371/journal.pcbi.1002285.t002

**Table 3.** Tools for glycan and glycoprotein analysis.

| Tools | Portable Version | Description | URL |
|---|---|---|---|
| GlyProt | No | Modeling 3D structure of glycoproteins with attached *N*-glycans | http://www.glycosciences.de/modeling/glyprot/php/main.php |
| SWEET-II | No | Building 3D carbohydrate models | http://www.glycosciences.de/modeling/sweet2/ |
| Glydict | No | Prediction of *N*-glycan 3D structures | http://www.glycosciences.de/modeling/glydict/ |
| Shape | Yes | Prediction of carbohydrate conformational space | http://sourceforge.net/projects/shapega/ |
| GlySeq | No | Statistical analysis of residues neighboring *N*-glycan sequons in protein sequence | http://www.glycosciences.de/tools/glyseq/ |
| GlyVicinity | No | Statistical analysis of residues surrounding carbohydrate chains in protein 3D structure | http://www.glycosciences.de/tools/glyvicinity/ |

3D, three-dimensional.
doi:10.1371/journal.pcbi.1002285.t003

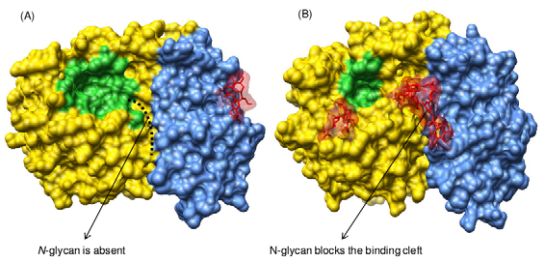


**Figure 2. Three-dimensional structures of two glycosyl hydrolase 32 (GH32) family enzymes.** Surface representation of the overall 3D structure of (A) *Arabidopsis thaliana* cell-wall invertase (PDB database accession code: 2AC1) and (B) *Cichorium intybus* fructan 1-exohydrolase IIa (PDB database accession code: 1ST8). The N- and C-terminal domains are colored in yellow and blue, respectively. The attached *N*-glycan molecules are represented as sticks in red color. The active site is shown in green. Another binding pocket that extends between N- and C-terminal domains is orange, highlighted in (A). This cleft is reserved for higher DP-inulin type fructans. An open conformation of the mentioned cavity is observed in GH32 enzymes capable of degrading inulin substrates, such as *C. intybus* fructan 1-exohydrolase IIa (A). However, the introduction of a glycosyl chain blocks the cleft and prevents inulin binding and degradation in some GH32 enzymes, such as in *A. thaliana* invertase (B).
doi:10.1371/journal.pcbi.1002285.g002

*N*-glycosylation versus amino acids position is also given.

Following the glycosylation prediction, three potential cases may emerge: (a) predicted *N*-glycan sites are found in both the target and the homologue proteins; (b) predicted *N*-glycan sites are found only in homologue proteins; and (c) no *N*-glycan sites are predicted either in the target protein or in homologue proteins. How to proceed?

In case (a), an optimization of Asn-x-Ser/Thr sequons replacing residues at position +1 (Asn occupies position 0) or surrounding the sequon is done. Statistical analysis of occupied and non-occupied *N*-glycosylation sites revealed that the amino acids at position +1 and nearby *N*-glycan sequons modulate the occurrence of *N*-glycosylation (Table 5). Some suggestions

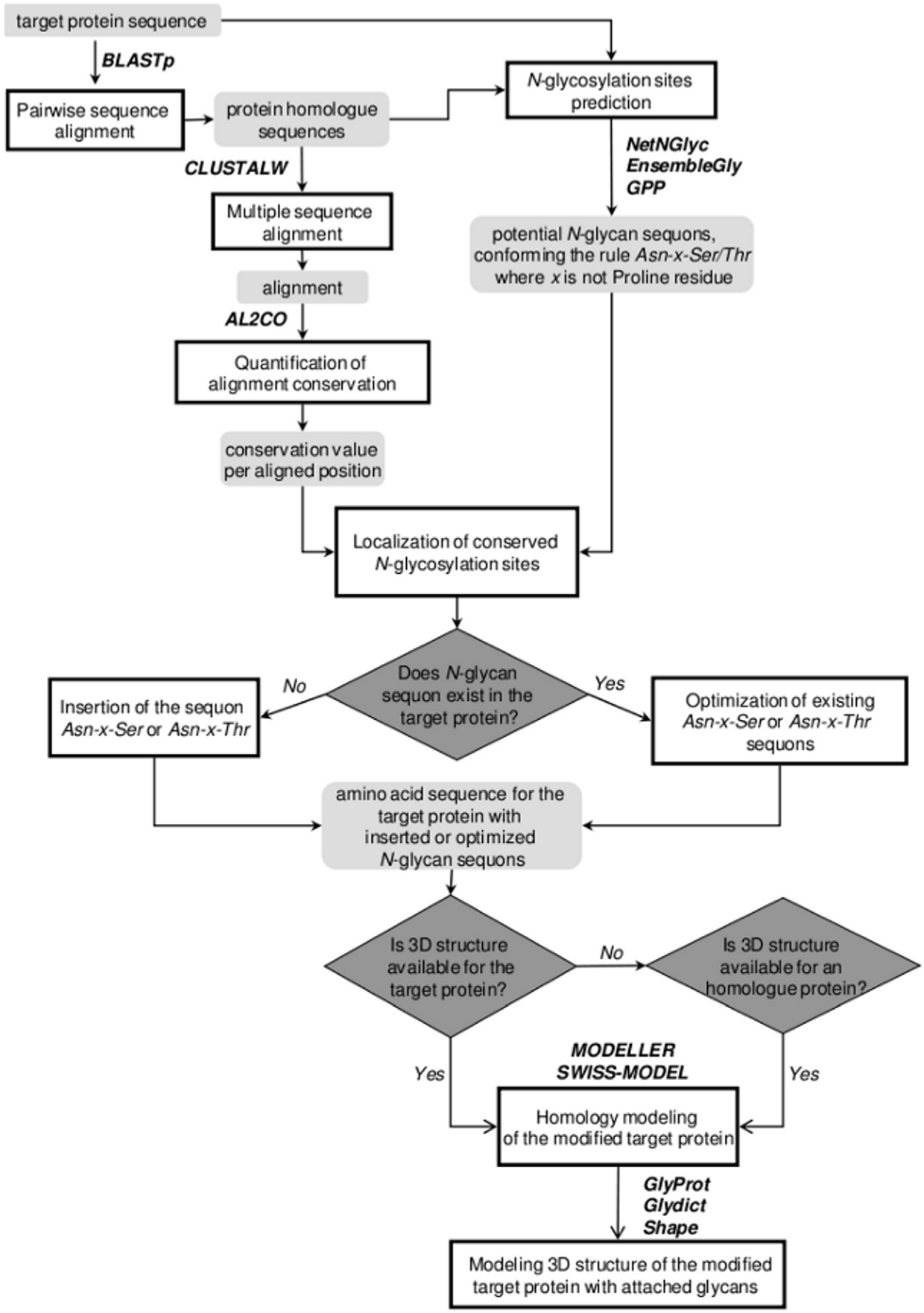


**Figure 3. Workflow for rational design and insertion of *N*-glycan sites in proteins.**
doi:10.1371/journal.pcbi.1002285.g003

**Table 4.** Software for protein sequence and tertiary structure analysis.

| Software | Portable Version | Description | URL |
|---|---|---|---|
| BLASTp | Yes | Pairwise sequence alignment comparing a query protein sequence with a database of protein sequences | http://blast.ncbi.nlm.nih.gov/Blast.cgi |
| CLUSTALW | Yes | Multiple sequence alignment comparing a number of protein sequences | http://www.ebi.ac.uk/Tools/msa/clustalw2/ |
| AL2CO | Yes | Quantification of conservation index at each aligned position in a multiple sequence alignment | http://prodata.swmed.edu/al2co/al2co.php |
| VMD | Yes | Molecular visualization for displaying, animating, and analyzing large biomolecular systems using 3D graphics | http://www.ks.uiuc.edu/Research/vmd/ |
| PSI-PRED | Yes | Prediction secondary structure from amino acid sequence | http://bioinf.cs.ucl.ac.uk/psipred/ |
| MODELLER | Free for academics, but commercial versions are also available. | Homology modeling of protein 3D structures. An option to introduce single point mutations in the target protein to obtain its 3D model is also included. | http://www.salilab.org/modeller/ |
| SWISS-MODEL | No | Automated comparative modeling of 3D protein structures | http://swissmodel.expasy.org/ |
| GROMACS | Yes | Molecular dynamics simulations | http://www.gromacs.org/ |

3D, three-dimensional.
doi:10.1371/journal.pcbi.1002285.t004

for amino acid substitutions: (a) aromatic amino acids (phenylalanine, tyrosine, and tryptophan) in position −2 and −1, (b) small nonpolar amino acids (glycine, alanine, and valine) in position +1, and (c) bulky hydrophobic amino acids (leucine, isoleucine, and methionine) in positions +3 to +5 (Figure 4). The statistical analysis of amino acids neighboring *N*-glycosylation sites in the protein primary sequence and tertiary structure can be conducted using the GlySeq and GlyVicinity software, respectively [18].

In case (b), a sequence pattern like Asn-x-Ser or Asn-x-Thr is inserted in the target protein. There is a large preference for threonine, as opposed to serine, in position +2. This is in agreement with the observation that replacing serine with threonine in the sequon results in an overall increase of the occupancy [19]. Some suggestions for amino acid substitution at position +1 are (a) highly conserved amino acids at the position +1 within the homologue proteins may be kept except proline, and (b) small nonpolar amino acids (glycine, alanine, and valine) at the position +1 increase the probability of sequon occupancy [20].

In case (c), the analysis of the secondary structure has to be performed to insert the *N*-glycan sites at or just after protein secondary structure changes. Glycosylation sites are frequently found in points of changes of secondary structure, with a bias toward turns and bends [19]. Protein secondary structure features are described in the PDB file. If no 3D structures are available, a prediction of the secondary structure can be solved using, for example, the PSI-PRED server [21]. Only the primary amino acid sequence is required as input.

With the insertion of *N*-glycosylation sites in the target protein primary structure, the attachment of *N*-glycan molecules is favored. Then, the analysis and visualization of the glycoprotein is also helpful. Tertiary glycoprotein structure having attached *N*-glycans can be modeled using the GlyProt server [22]. This facilitates the identification of spatially unfavorable *N*-glycosylation sites [6].

The 3D glycan structures are provided in the GlyProt server database; they can also be implemented using the SWEET-II [23], Glydict [24], and Shape [25] software. For the GlyProt server input 3D protein structure, the atomic coordinate file from the modified target protein is required. In this case, a 3D structure model has to be built, using the structure of the native target protein or related homologue as a template. The sequence used as input to build the 3D model has to contain the inserted *N*-glycan sequons, for which homology modeling software like MODELLER [26] and the online SWISS-MODEL server [27] can be used.

Finally, molecular dynamics simulations to explore protein backbone conformational changes could be applied using, for example, the GROMACS software [28]. This strategy allows for the refinement of the initial glycoprotein structure. All bioinformatics software previously mentioned are freely available. An example of the application of the workflow presented in this manuscript is available in Supporting Information (Text S1 and Figures S1, S2, S3, S4).

**Table 5.** Comparative studies for occupied and non-occupied *N*-glycan sites.

| Description | Reference |
|---|---|
| Influence of proline residue neighboring the Asn-x-Ser and the Asn-x-Thr sequons over *N*-glycosylation in the yeast invertase protein. | [43] |
| Relevance of certain amino acid substitutions at the position +1 in the Asn-x-Ser sequon for *N*-glycosylation efficiency in the rabies virus glycoprotein. | [44] |
| Relevance of certain amino acid substitutions at the position +1 in the Asn-x-Ser and the Asn-x-Thr sequons for *N*-glycosylation efficiency using different variants of rabies virus glycoprotein. | [45] |
| Influence of the 20 amino acids at the position following the Asn-x-Ser and Asn-x-Thr sequons for *N*-glycosylation efficiency using different variants of rabies virus glycoprotein. | [46] |
| Occurrence frequency analysis of some amino acid residues at position +1 in the Asn-x-Ser and Asn-x-Thr sequons studying glycoproteins from the PDB database [47]. | [48] |
| Influence of the 20 amino acids flanking the upstream and downstream of Asn-x-Ser and Asn-x-Thr sequons, using glycoproteins from the UniProtKB/Swiss-Prot database [11]. | [19] |
| Primary, secondary, and tertiary structures statistical analysis of occupied and non-occupied *N*-glycosylation sites using glycoproteins from the PDB database [47]. | [49] |

doi:10.1371/journal.pcbi.1002285.t005

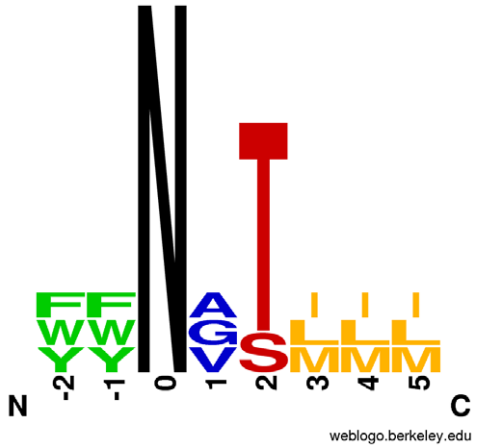


**Figure 4. Amino acid preferences in occupied *N*-glycan sites.** The sequence logo displays residues preferentially placed at occupied *N*-glycan sequons. Neighboring residues located downstream (positions +3 to +5) and upstream (positions −1 and −2) from the asparagine residue (position 0) are also shown. The size of each letter represents the residue prevalence at the putative position. For example, threonine residue is preferred over serine, at position +2. The WebLogo server [29] was used to generate the sequence logo.
doi:10.1371/journal.pcbi.1002285.g004

## Concluding Remarks

In a brief survey, a workflow integrating available bioinformatics resources to assist protein glycosylation was exposed. In particular, the rational manipulation of the native *N*-glycosylation pattern, including in silico tools, was given. The application of the bioinformatics strategy described in this tutorial, at the early stages of glycoengineering, can help the design and insertion of *N*-glycan sites in proteins, reducing time, effort, and cost.

## Supporting Information

**Figure S1** Protein tertiary structure.
(TIF)

**Figure S2** Multiple sequence alignment.
(PDF)

**Figure S3** Pairwise sequence alignment.
(PDF)

**Figure S4** Protein tertiary structure with modeled *N*-glycans.
(TIF)

**Text S1** Supporting information text.
(DOC)

## References

1. Zaia J (2008) Mass spectrometry and the emerging field of glycomics. Chem Biol 15: 881–892.
2. der Lieth CW, Bohne-Lang A, Lohmann KK, Frank M (2004) Bioinformatics for glycomics: status, methods, requirements and perspectives. Brief Bioinform 5: 164–178.
3. Mahal LK (2008) Glycomics: towards bioinformatic approaches to understanding glycosylation. Anticancer Agents Med Chem 8: 37–51.
4. Aoki-Kinoshita KF (2008) An introduction to bioinformatics for glycomics research. PLoS Comput Biol 4: e1000075. doi:10.1371/journal.pcbi.1000075.
5. Frank M, Schloissnig S (2010) Bioinformatics and molecular modeling in glycobiology. Cell Mol Life Sci 67: 2749–2772.
6. Le Roy K, Verhaest M, Rabijns A, Clerens S, Van Laere A, et al. (2007) N-glycosylation affects substrate specificity of chicory fructan 1-exohydrolase: evidence for the presence of an inulin binding cleft. New Phytol 176: 317–324.
7. Sinclair AM, Elliott S (2005) Glycoengineering: the effect of glycosylation on the properties of therapeutic proteins. J Pharm Sci 94: 1626–1635.
8. Sola RJ, Griebenow K (2010) Glycosylation of therapeutic proteins: an effective strategy to optimize efficacy. BioDrugs 24: 9–21.
9. Elliott S, Lorenzini T, Asher S, Aoki K, Brankow D, et al. (2003) Enhancement of therapeutic protein in vivo activities through glycoengineering. Nat Biotechnol 21: 414–421.
10. Sola RJ, Griebenow K (2009) Effects of glycosylation on the stability of protein pharmaceuticals. J Pharm Sci 98: 1223–1245.
11. The UniProt Consortium (2011) Ongoing and future developments at the Universal Protein Resource. Nucleic Acids Res 39: D214–D219.
12. National Center for Biotechnology Information (2011) PubMed database. Available: http://www.ncbi.nlm.nih.gov/pubmed. Accessed 15 April 2011.
13. Mirny LA, Shakhnovich EI (1999) Universally conserved positions in protein folds: reading evolutionary signals about stability, folding kinetics and function. J Mol Biol 291: 177–196.
14. Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22: 4673–4680.
15. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215: 403–410.
16. Pei J, Grishin NV (2001) AL2CO: calculation of positional conservation in a protein sequence alignment. Bioinformatics 17: 700–712.
17. Humphrey W, Dalke A, Schulten K (1996) VMD: visual molecular dynamics. J Mol Graph 14: 33–38.
18. Lutteke T, Frank M, der Lieth CW (2005) Carbohydrate Structure Suite (CSS): analysis of carbohydrate 3D structures derived from the PDB. Nucleic Acids Res 33: D242–D246.
19. Petrescu AJ, Milac AL, Petrescu SM, Dwek RA, Wormald MR (2004) Statistical analysis of the protein environment of N-glycosylation sites: implications for occupancy, structure, and folding. Glycobiology 14: 103–114.
20. Yurist-Doutsch S, Chaban B, VanDyke DJ, Jarrell KF, Eichler J (2008) Sweet to the extreme: protein glycosylation in Archaea. Mol Microbiol 68: 1079–1084.
21. McGuffin LJ, Bryson K, Jones DT (2000) The PSIPRED protein structure prediction server. Bioinformatics 16: 404–405.
22. Bohne-Lang A, der Lieth CW (2005) GlyProt: in silico glycosylation of proteins. Nucleic Acids Res 33: W214–W219.
23. Bohne A, Lang E, von der Lieth C-W (1998) W3-SWEET: Carbohydrate Modeling By Internet. J Mol Model 4: 33–43.
24. Frank M, Bohne-Lang A, Wetter T, Lieth CW (2002) Rapid generation of a representative ensemble of N-glycan conformations. In Silico Biol 2: 427–439.
25. Rosen J, Miguet L, Pérez S (2009) Shape: automatic conformation prediction of carbohydrates using a genetic algorithm. J Cheminf 1: 1–7.
26. Fiser A, Sali A (2003) Modeller: generation and refinement of homology-based protein structure models. Methods Enzymol 374: 461–491.
27. Schwede T, Kopp J, Guex N, Peitsch MC (2003) SWISS-MODEL: An automated protein homology-modeling server. Nucleic Acids Res 31: 3381–3385.
28. Van Der SD, Lindahl E, Hess B, Groenhof G, Mark AE, et al. (2005) GROMACS: fast, flexible, and free. J Comput Chem 26: 1701–1718.
29. Crooks GE, Hon G, Chandonia JM, Brenner SE (2004) WebLogo: a sequence logo generator. Genome Res 14: 1188–1190.
30. Kowarik M, Young NM, Numao S, Schulz BL, Hug I, et al. (2006) Definition of the bacterial N-glycosylation site consensus sequence. EMBO J 25: 1957–1966.
31. Schaffer C, Graninger M, Messner P (2001) Prokaryotic glycosylation. Proteomics 1: 248–261.
32. Gupta R, Brunak S (2002) Prediction of glycosylation across the human proteome and the correlation to protein function. Pac Symp Biocomput. pp 310–322.
33. Nothaft H, Szymanski CM (2010) Protein glycosylation in bacteria: sweeter than ever. Nat Rev Microbiol 8: 765–778.
34. Gentzsch M, Tanner W (1997) Protein-O-glycosylation in yeast: protein-specific mannosyltransferases. Glycobiology 7: 481–486.
35. Julenius K (2007) NetCGlyc 1.0: prediction of mammalian C-mannosylation sites. Glycobiology 17: 868–876.
36. Krieg J, Hartmann S, Vicentini A, Glasner W, Hess D, et al. (1998) Recognition signal for C-mannosylation of Trp-7 in RNase 2 consists of sequence Trp-x-x-Trp. Mol Biol Cell 9: 301–309.
37. Hofsteenge J, Blommers M, Hess D, Furmanek A, Miroshnichenko O (1999) The four terminal

components of the complement system are C-mannosylated on multiple tryptophan residues. J Biol Chem 274: 32786–32794.
38. Zanetta JP, Pons A, Richet C, Huet G, Timmerman P, et al. (2004) Quantitative gas chromatography/mass spectrometry determination of C-mannosylation of tryptophan residues in glycoproteins. Anal Biochem 329: 199–206.
39. Brazier-Hicks M, Evans KM, Gershater MC, Puschmann H, Steel PG, et al. (2009) The C-glycosylation of flavonoids in cereals. J Biol Chem 284: 17926–17934.
40. Kobayashi T, Nishizaki R, Ikezawa H (1997) The presence of GPI-linked protein(s) in an archaeobacterium, Sulfolobus acidocaldarius, closely related to eukaryotes. Biochim Biophys Acta 1334: 1–4.
41. Ikezawa H (2002) Glycosylphosphatidylinositol (GPI)-anchored proteins. Biol Pharm Bull 25: 409–417.
42. Orlean P, Menon AK (2007) Thematic review series: lipid posttranslational modifications. GPI anchoring of protein in yeast and mammalian cells, or: how we learned to stop worrying and love glycophospholipids. J Lipid Res 48: 993–1011.
43. Roitsch T, Lehle L (1989) Structural requirements for protein N-glycosylation. Influence of acceptor peptides on cotranslational glycosylation of yeast invertase and site-directed mutagenesis around a sequon sequence. Eur J Biochem 181: 525–529.
44. Shakin-Eshleman SH, Spitalnik SL, Kasturi L (1996) The amino acid at the X position of an Asn-X-Ser sequon is an important determinant of N-linked core-glycosylation efficiency. J Biol Chem 271: 6363–6366.
45. Kasturi L, Chen H, Shakin-Eshleman SH (1997) Regulation of N-linked core glycosylation: use of a site-directed mutagenesis approach to identify Asn-Xaa-Ser/Thr sequons that are poor oligosaccharide acceptors. Biochem J 323(Pt 2): 415–419.
46. Mellquist JL, Kasturi L, Spitalnik SL, Shakin-Eshleman SH (1998) The amino acid following an asn-X-Ser/Thr sequon is an important determinant of N-linked core glycosylation efficiency. Biochemistry 37: 6833–6837.
47. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, et al. (2000) The Protein Data Bank. Nucleic Acids Res 28: 235–242.
48. Christlet TH, Biswas M, Veluraja K (1999) A database analysis of potential glycosylating Asn-X-Ser/Thr consensus sequences. Acta Crystallogr D Biol Crystallogr 55: 1414–1420.
49. Ben Dor S, Esterman N, Rubin E, Sharon N (2004) Biases and complex patterns in the residues flanking protein N-glycosylation sites. Glycobiology 14: 95–101.