



In silico simulation of glycosylation and related pathways

Yukie Akune-Taylor¹ · Akane Kon² · Kiyoko F. Aoki-Kinoshita^{1,2,3}

Received: 10 October 2023 / Revised: 30 April 2024 / Accepted: 2 May 2024
© The Author(s) 2024

Abstract

Glycans participate in a vast number of recognition systems in diverse organisms in health and in disease. However, glycans cannot be sequenced because there is no sequencer technology that can fully characterize them. There is no “template” for replicating glycans as there are for amino acids and nucleic acids. Instead, glycans are synthesized by a complicated orchestration of multitudes of glycosyltransferases and glycosidases. Thus glycans can vary greatly in structure, but they are not genetically reproducible and are usually isolated in minute amounts. To characterize (sequence) the glycome (defined as the glycans in a particular organism, tissue, cell, or protein), glycosylation pathway prediction using in silico methods based on glycome expression data, and glycosylation simulations have been attempted. Since many of the mammalian glycome genes have been identified and cloned, it has become possible to predict the glycan biosynthesis pathway in these systems. By then incorporating systems biology and bioprocessing technologies to these pathway models, given the right enzymatic parameters including enzyme and substrate concentrations and kinetic reaction parameters, it is possible to predict the potentially synthesized glycans in the pathway. This review presents information on the data resources that are currently available to enable in silico simulations of glycosylation and related pathways. Then some of the software tools that have been developed in the past to simulate and analyze glycosylation pathways will be described, followed by a summary and vision for the future developments and research directions in this area.

Keywords Glycosylation · Bioinformatics · Systems biology · Pathways · Prediction · Software

Introduction

Glycans, also called polysaccharides or carbohydrates, are chains of variously linked monosaccharides biosynthesized by glycosyltransferases. They can occur as free oligosaccharides or as parts of glycoproteins and glycolipids. They participate in a vast number of recognition systems in diverse organisms in health (development, cell differentiation, inflammation, signaling, and immunomodulation) and in disease (infectious and non-infectious including neoplasia) [1].

They are known to be involved in virus infection, including influenza [2] and even SARS-CoV-2 [3], as they cover the spike protein of this latter virus. The reason why they are so involved in many biological processes is that they are found on the cell surface of practically every cell in the body. They are attached to proteins and lipids on the cell surface, and sometimes, they are secreted outside of the cell. They are also involved in the extracellular matrix as proteoglycans; well-known proteoglycans are heparan sulfate, chondroitin sulfate, and keratan sulfate [4].

Glycans cannot be sequenced as proteins or DNA due to the absence of a fully characterizing sequencer technology and a replicating template. Instead, glycans are synthesized by a complicated orchestration of hundreds of glycosyltransferases, glycosidases, and other enzymes such as epimerases and sulfotransferases (often collectively termed “glycome genes”) localized widely from the endoplasmic reticulum, through the Golgi apparatus, and out to the trans-Golgi network. For example, an *N*-glycan is first extended in the cytoplasmic side of the ER to form Man9 with glucose-caps. It is then flipped to the luminal side of

Published in the topical collection *Advances in (Bio-)Analytical Chemistry: Reviews and Trends Collection 2024*.

✉ Kiyoko F. Aoki-Kinoshita
kkiyoko@soka.ac.jp

- ¹ Glycan and Life Systems Integration Center, Soka University, Tokyo, Japan
- ² Graduate School of Science and Engineering, Soka University, Tokyo, Japan
- ³ iGCORE, Nagoya University, Nagoya, Japan

the ER and transferred to an asparagine residue of a protein. The glucose-caps and mannoses are then trimmed by glucosidase I and II and mannosidase I (*N*-glycan precursor). When the proteins are transferred into the Golgi apparatus, the *N*-glycan precursor is further processed and modified by glycosyltransferases such as FucT, GalT, and SiaT (Fig. 1). The CAZy database encompasses more than 300 families of glycosidases and glycosyltransferases. Each of these enzyme families represents distinct modules with different substrate specificities and reaction conditions. Thus, glycans can vary greatly in structure. For example, bacterial glycans, such as those found on *N*-glycoproteins and lipopolysaccharides, are mainly found on the cell wall. They are involved in host-interactions and pathogenesis. Plant glycans, such as cellulose and pectin, play a key role in cell wall stabilization in terms of its strength and resistance to various environmental stresses. The differences in the functions and structures of glycans in each biological field are due to evolutionary and ecological factors, which have shaped the strategies of each organism to adapt to different environments. Therefore, it is crucial, when using software tools, to grasp these distinctions for a precise assessment of tool applicability.

To characterize (sequence) glycans, no next-generation sequencer exists, and so current technologies include mass spectrometry, liquid chromatography, and nuclear magnetic resonance. All these technologies require highly technical skills and sufficient amounts of samples to analyze. Therefore, other means of characterizing the glycome (defined as the glycans in a particular organism, tissue, cell, or protein) have been developed. This includes calculation algorithms to measure the physico-chemical similarities of glycan structures, glycosylation pathway prediction based on glycome expression data, and glycosylation simulation. In particular, many of the mammalian glycome genes have been identified, cloned, and their activities identified, so it has become possible to predict the glycan biosynthesis pathways in silico using these data. In the last decade, these in silico models have been rapidly improved by incorporating

multiomics data (genomics, transcriptomics, and proteomics), systems biology, and bioprocessing technologies to these pathway models. Furthermore, the addition of appropriate enzymatic parameters including enzyme and substrate concentrations and kinetic reaction parameters enable the prediction of the potentially biosynthesized glycans that are involved in specific pathways such those involved in health and disease.

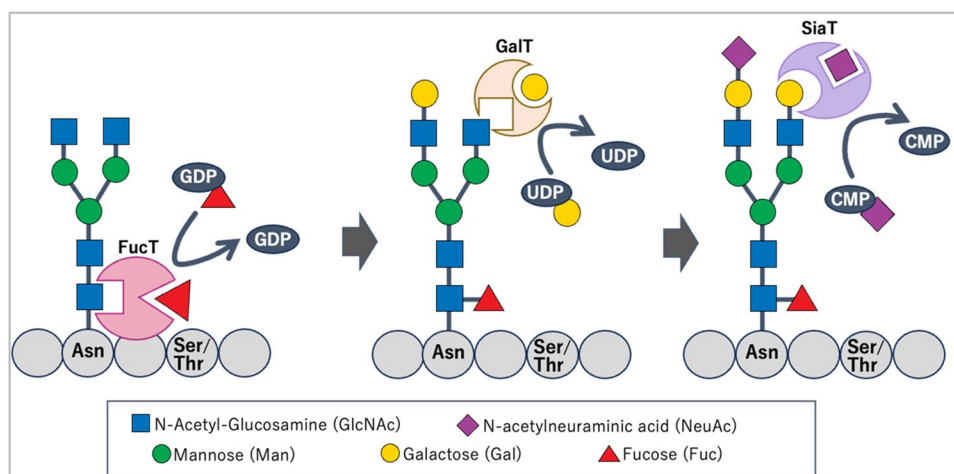
This article will present information on currently available data resources used for in silico simulations of glycosylation and related pathways. These include Web tools and programming libraries for pathway modeling that have been made available for users to utilize and or expand them for their own experimental data. There are also databases that have accumulated various information relevant to glycome genes. Then some of the software tools that have been developed to take a step further and perform simulations and analyze glycosylation pathways will be presented, followed by a summary and vision for the future developments and research directions in this area.

Data resources

There are several data resources that provide information on glycome genes. CAZy (Carbohydrate-Active enZymes) [5] is one of the oldest carbohydrate-related databases still running, having started circa 1998. It organizes information on carbohydrate enzymes involved in glycosylation, metabolism, and transportation of glycans into six classes, each of which are subdivided into families, the numbers of which continue to grow, but at the time of this writing number as follows:

1. Glycoside hydrolases (GH), 185 families
2. Glycosyltransferases (GT), 116 families
3. Polysaccharide lyases (PL), 42 families
4. Carbohydrate esterases (CE), 20 families

Fig. 1 An example scheme of a biosynthetic pathway of a glycan being processed in the Golgi apparatus. Various glycosidases and glycosyltransferases (here, FucT, GalT, and SiaT, which transfer a fucose residue, a galactose residue, and a sialic acid residue, respectively, from their corresponding nucleotide-sugar) remove or add monosaccharides to synthesize a glycan



5. Auxiliary activities (AA), 16 families
6. Carbohydrate-binding modules (CBM), 98 families

CAZy enzymes (often called CAZymes) are classified based on their amino acid sequence similarities as there are correlations between sequence and protein folding similarities. Many of the families are then classified based on the three-dimensional patterns of the protein structures. CAZy provides mutual links with KEGG, RCSB PDB, ExPasy, and other databases. Many of the CAZymes are automatically populated based on the sequences registered into NCBI GenBank.

The GlycoGene Database (GGDB) [6] was originally developed under the Japanese government-funded Glycogene Project (GG Project) in 2001. Over 180 genes of human glycosyltransferases and sulfotransferases were cloned and recorded into GGDB. Each entry page is manually curated and includes Gene ID, DNA sequences, tissue distribution of gene expression, substrate specificity, homologous genes, and external links to other databases such as GenBank and CAZy. The latest data in GGDB are now available in the ACGG-DB database (<https://acgg.asia/db/>), which are integrated into the GlyCosmos Glycoscience Portal [7].

KEGG is well-known as a Web resource for biological systems data including genomic, chemical, and health and disease information. It is a major provider of manually curated pathways, as well as databases for genes and genomes. It also includes a GLYCAN resource [8], which contains glycan structures that participate in the glycan-related pathways stored in KEGG, as well as glycogene information, often annotated with E.C. numbers. Most recently, disease-related information related to glycogenes have been incorporated into KEGG.

In the field of bioprocessing, Chinese Hamster Ovary (CHO) cells provide a standard platform for production of protein therapeutics because of their human-like glycosylation. Thus, CHOglycoNET [9] was developed as a comprehensive network encompassing glycosylation reactions that account for all experimentally observed glycans found in recombinant proteins and both intracellular, membrane, and secreted host cell proteins within two major CHO cell lineages, namely CHO-S and CHO-K1. This is the largest dataset of CHO cell glyco-profiles comprising 200 datasets sourced from seven different laboratories; it serves as the basis for uncovering potential latent reactions that could become active under a variety of genetic glycoengineering and metabolic perturbation scenarios, for a range of recombinant glycoproteins and CHO cell host cell proteins.

Software and tools

Many software tools have been developed in the past, even when there were few datasets readily available for analyzing glycogenes. Here, we describe software tools for

glycan biosynthesis analysis and for in silico simulation for the prediction of glycomes.

Glycan biosynthesis prediction

Glycologue (<https://glycologue.org/>) is a Web portal for glycosylation prediction tools for *N*- and *O*-glycans, human-milk oligosaccharides and gangliosides [10–12]. Users can simulate the glycosylation pathway by choosing a starting glycan structure and selecting glycosyltransferases from a predefined list. This list has been manually curated and includes the reaction pattern representing the substrate specificity of the given glycogene. The model then calculates the glycosylation pathway using the selected enzymes. It is also possible to calculate a minimal set of glycosyltransferases to biosynthesize the starting glycan structure. Using this model, the authors were able to predict a highly heterogeneous set of structures when all *O*-glycan-related enzymes (25 glycosyltransferase and sulfotransferase enzymes) were allowed to act, including many clinically important epitopes such as Sialyl-Lewis X. Moreover, in silico knockout experiments were performed, and they were able to achieve 98% coverage of glycans predicted for specific knockout cell lines.

GlycoVis is a visualization tool designed to illustrate the distribution of *N*-glycans within a reaction network, along with the potential pathways for reactions associated with each glycan [13]. The enzyme substrate specificities have been structured into a matrix of relationships. Upon inputting glycan distribution data, the program generates a pathway map that represents various glycans using distinct colors to indicate their relative abundance levels. Additionally, it identifies and traces all feasible reaction routes leading to each glycan on the map. To demonstrate GlycoVis's utility, it was applied to illustrate the glycoform distribution in Chinese Hamster Ovary (CHO) cell-derived tissue plasminogen activator (TPA), as well as human and mouse IgG.

Glycan Pathway Predictor (GPP) [14] is a Web tool to predict *N*-glycosylation pathways, given a starting glycan and a selected list of glycogenes; the computation is based on a mathematical model proposed earlier [15, 16]. It is available on the RINGS (<https://rings.glycoinfo.org>) resource [17]. A total of 19 glycosyltransferases are available by default, and users can also limit the size of the predicted pathway by specifying the maximum mass of glycans to predict. For example, Fig. 2 shows a snapshot of the results of predicting the biosynthetic pathway using just two genes iGnT and b4GalT starting from a single tetra-antennary *N*-glycan structure. The figure shows 19 glycans and 30 reactions that could potentially take place. This model does not take cellular localization into consideration, but constraints on the substrate specificity of each gene can be made to emulate such information.

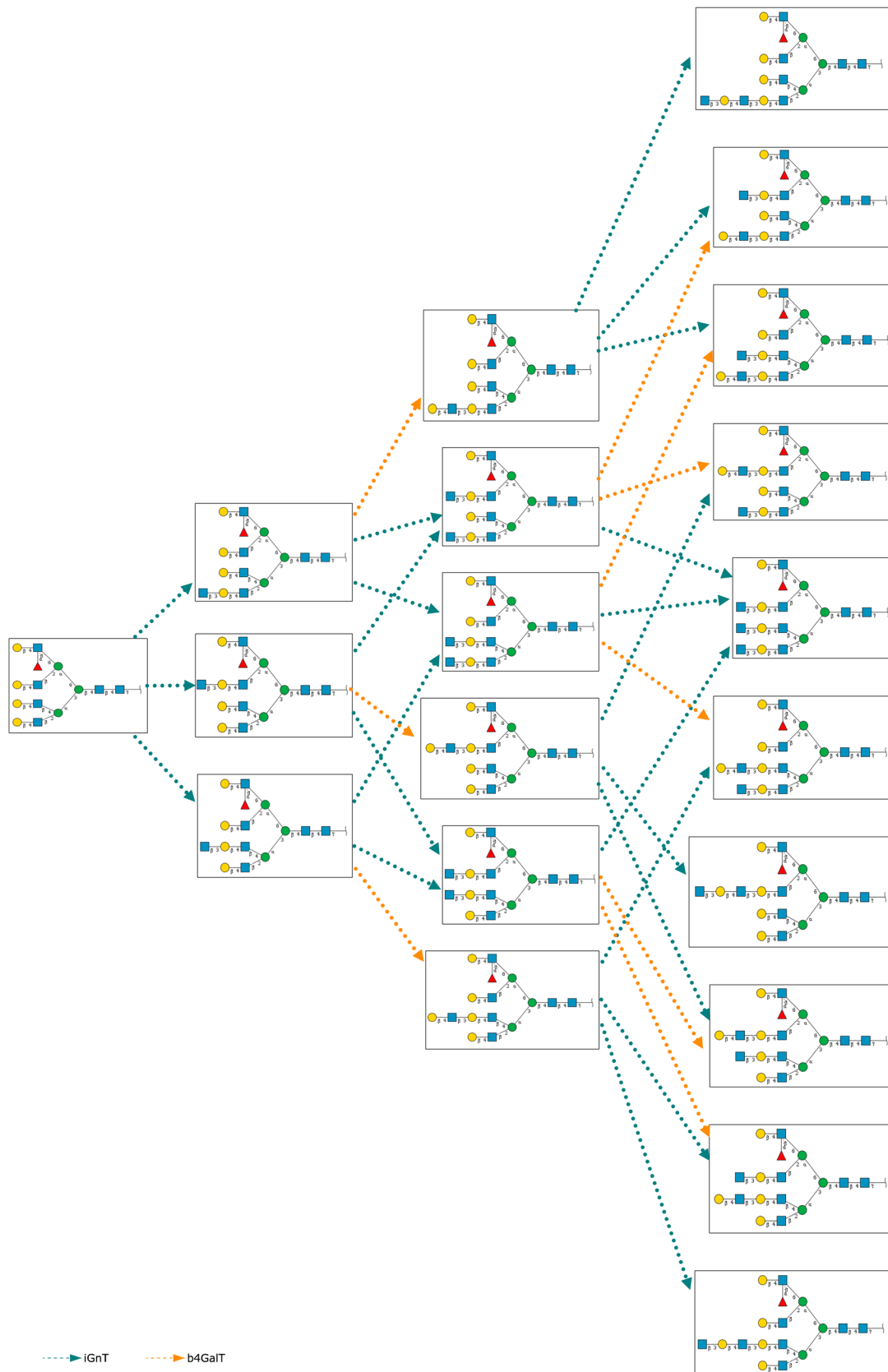


Fig. 2 An example of the resulting page from running the Glycan Pathway Predictor (GPP) tool in RINGS. All the potentially biosynthesized glycans using just two genes, iGnT and b4GalT, from a single tetra-antennary N-glycan structure are shown, indicating the complexity of glycan biosynthesis

UniCorn [18] is a database developed from the results of utilizing GPP to predict the glycosylation pathway of *N*-glycans (> 15 monosaccharide residues) using 45 human glycosyltransferases. Enzyme specificities were extracted from KEGG, CFG, CAZy, GGDB, and BRENDA [19]. As a result, more than 1.1 million theoretical structures and 4.7 million synthetic reactions were generated and stored in UniCorn, which was made available in UniCarbKB [20]. Similarly to Fig. 2, a tremendous amount of glycans can be potentially biosynthesized from a handful of glycans, but in reality, based on information deposited in glycan structure databases, only a few thousand have actually been identified. Therefore, more research into the cellular localization and structure of the Golgi apparatus, where most of the glycozymes reside, need to be made to better model these pathways.

VirtualGlycome (<https://virtualglycome.org>) is a Web portal developed by Neelamegham et al. to provide software tools and experimental resources for glycan structure analysis. Currently, five software tools are provided: DrawGlycan-SNFG [21], GNAT [22], and GlycoPAT [23], among others. GNAT, in particular, is a free open-source MATLAB toolbox for predicting glycosylation networks. It provides various functionality, including network prediction, similar to GPP, to reconstruct glycosylation networks given a set of reactions and/or products and a list of enzymes; prediction of networks from mass spectrometry data; and dynamic and steady-state simulations of reaction networks. While this software has not been updated in a while, it is still available for download, and the author is reachable for questions.

GlycoMME (Glycosylation Markov Model Evaluator) [24] is a toolkit for analyzing the effect of glycoengineering on the theoretical *N*-glycosylation biosynthesis. They facilitated *N*-glycosylation as a Markov model to quantify the specificity of isozymes and the interactions of glycosyltransferases, which helps users to predict the *N*-glycosylation process.

GlycoCompare [25] is a computational approach for the rapid and scalable analysis of comparing multiple glyco-profiles. It calculates the glycan intermediates, which are used as interpretable functional units, to address the hidden interdependencies between glycomics samples. The authors demonstrated the GlycoCompare method using recombinant erythropoietin (EPO) *N*-glycosylation, human milk oligosaccharides (HMOs), mucin-type O-glycans, gangliosides, and site-specific compositional data.

More recently, Glycowork [26] has been released as an open-source Python package for glycan-related data analysis

and machine learning algorithms. It provides ~50,000 glycan sequences with ~35,000 species-related, ~14,000 tissues-related, and ~1,000 disease-related annotations. It also provides over 550,000 glycan-protein binding data. Those data are used in the deep learning model. NSequonPred, which is one of the trained models in Glycowork, helps users to predict whether the given *N*-sequon is glycosylated. The latest version of the Glycowork framework has enriched the motif annotation and expanded the model for the multiple glycomics expression data sets [27].

Table 1 summarizes the glycan biosynthetic tools described here. The “Applications” column in this table indicates the biological applications that have been illustrated using these tools. Many of these have studied human milk oligosaccharides (HMO), their biosynthetic pathways, and relevant enzymes. Glycologue in particular has recently published work on HMOs to predict important enzymes in their biosynthesis [11] as well as on a study of glycoside hydrolases to conversely study the potential pathogens associated in human gut [28]. Others have shown that their tools are able to predict glycosylation pathways given glycomics profiles, usually from mass spectrometry experiments; these have been indicated as “MS.”

Most tools have focused on the more well-studied glycan types, but recently, there has also been a report on a theoretical model for glycosaminoglycan biosynthesis [29]. While this has not been implemented in any tool, it can potentially be incorporated into any of the pathway modeling tools incorporating kinetic parameters. Machine learning has also been used for predicting protein glycosylation [30] but has yet been implemented as a tool for practical use.

In silico simulations of glycosylation

In the field of glycoengineering, several attempts have been made to simulate glycosylation. As mentioned earlier, mathematical models have been proposed and further used to predict glycomes, comparing gene expression profiles and mass spectrometry glycomics datasets for validation [31, 32]. However, these models were based on Michaelis–Menten kinetics to model the reaction equations; many of the parameters for these reaction equations are often unknown, creating a bottleneck in simulating these models accurately. Various models have been proposed to emulate the Golgi apparatus, where the majority of glycozymes reside [30, 33]. Parameter estimation methods and sensitivity analysis tools have been applied to fill in these gaps, but validation has always been an issue.

Nevertheless, we have been developing the GlycoSim tool (<https://glycosim.rings.glycoinfo.org>) [34] to provide a means for non-computational scientists to access these models and use them with their own data. GlycoSim uses the same functionality as the GPP tool to predict the glycosylation pathway

Table 1 A summary of the selected software and tools described for glycan biosynthetic pathway predictions

Name	Glycan type	Function	Application type
Glycologue	<i>N</i> -, <i>O</i> -, HMO, gangliosides	Pathway prediction	Web tool
GlycoVis	<i>N</i> -	Glycan distribution and pathway prediction	Windows-based software tool (available by request)
Glycan Pathway Predictor (GPP)	<i>N</i> -	Pathway prediction	Web tool
UniCorn	<i>N</i> -	Glycosylation pathway database	Currently closed due to inaccessibility of UniCarbKB
GNAT	<i>N</i> -, <i>O</i> -	Pathway prediction	MATLAB (MathWorks)
GlyCompare	<i>N</i> -, <i>O</i> -, HMO, glycolipids	Pathway prediction and comparison	Python library
GlycoMME	<i>N</i> -	Pathway prediction based on glycomics data	MATLAB (MathWorks)
GlycoWork “network” module	Any linkage defined glycan data	Pathway prediction	Python library

given a substrate(s) and list of predefined glycogenes, but it also provides more flexibility in specifying substrate by using user-defined enzyme specificities. Figure 3 is a screenshot of the GlycoSim pathway prediction module, where step 1 for inputting enzyme specificities can be manually edited based on the LiCoRR rules [35]. Based on the predicted pathway (shown in the bottom of Fig. 3), a mathematical model is generated, where further parameters can be specified. We are also developing a database of predicted parameters for the reaction equations in these models. GlycoSim also has modules for parameter estimation of the missing parameters and sensitivity analysis of the parameters to determine the more sensitive parameters in the model. Such information can aid in determining the most important parameters in the model.

A large variety of parameter estimation methods are available in many software libraries, and optimization functionality is often used on top of the estimated parameters. In our work, we have previously attempted parameter estimation using the Particle Swarm and Simulated Annealing methods of COPASI [36] on glycomics and gene expression data from three types of mouse stem cells [37]. First, by using a normalized dataset of gene expression data for mouse ES cells, the *N*- and *O*-glycan biosynthesis pathways were predicted, and a mathematical model was generated using Systems Biology Markup Language (SBML) [38]. Then the model was imported into COPASI to test several parameter estimation methods. The estimated ranges were set initially to a very large range and, after repeated estimations, narrowed down to ranges to include the resulting values that were often estimated. This process was repeated 3–4 times. As a result, while the estimation process took an order of magnitude longer, Simulated Annealing was found to consistently produce stable values compared to Particle Swarm. The residual sum of squares (RSS) was used to estimate how well the model could reproduce the experimental results, and we found that when the RSS value was less than 1.0×10^{-10} , the simulation results were quite close to the experimental data. The ES model was then tested on the other stem cell data from mouse, namely

ExE and EB cells, using the estimated parameters from ES cells, where we could also obtain RSS values within 1.0×10^{-9} for *O*-glycans and 1.0×10^{-6} for *N*-glycans, indicating that we were successfully able to estimate these parameters well for *O*-glycans, but it was not as close for *N*-glycans. We found that in the glycomics data, there was a structure whose abundance could not be identified, thus resulting in a lower RSS value for the latter. However, conversely, we can claim that using RSS appears to be an effective method for scoring the fitness of a parameter set. In order to make this data available to the public, we are currently developing a database of these estimated parameters for others to test as well.

However, to avoid this missing parameter issue altogether, Boolean networks, Bayesian inference, Markov chain modeling, and other statistical methods have also been employed by others to mathematically calculate the parameters to reproduce glycan distributions without requiring kinetic information [25, 39, 40]. Flux analysis and multivariate data analysis are methods that attempt to capture bioprocesses more mechanistically, compared to the enzymatic methods which require many parameters [41, 42]. Moreover, a modeling framework based on genome reconstruction but using reaction flux flow stoichiometry, discretized variable state parameters, and mass balances has been developed, called DReaM-zyP, for discretized reaction network modeling using fuzzy parameters. This framework has been packaged into a tool called Glyco-Mapper which includes all CHO *N*-glycosylation genes, nucleotide sugar synthesis, transporter, and glycosylation-relevant metabolism genes [43]. It was shown to be able to model and predict many of the well-known CHO-engineered glycoforms published in the literature.

Future outlook

In this brief review, we have introduced databases and software tools to enable the *in silico* prediction of glycosylation, mainly for mammalian cells. Many databases in the

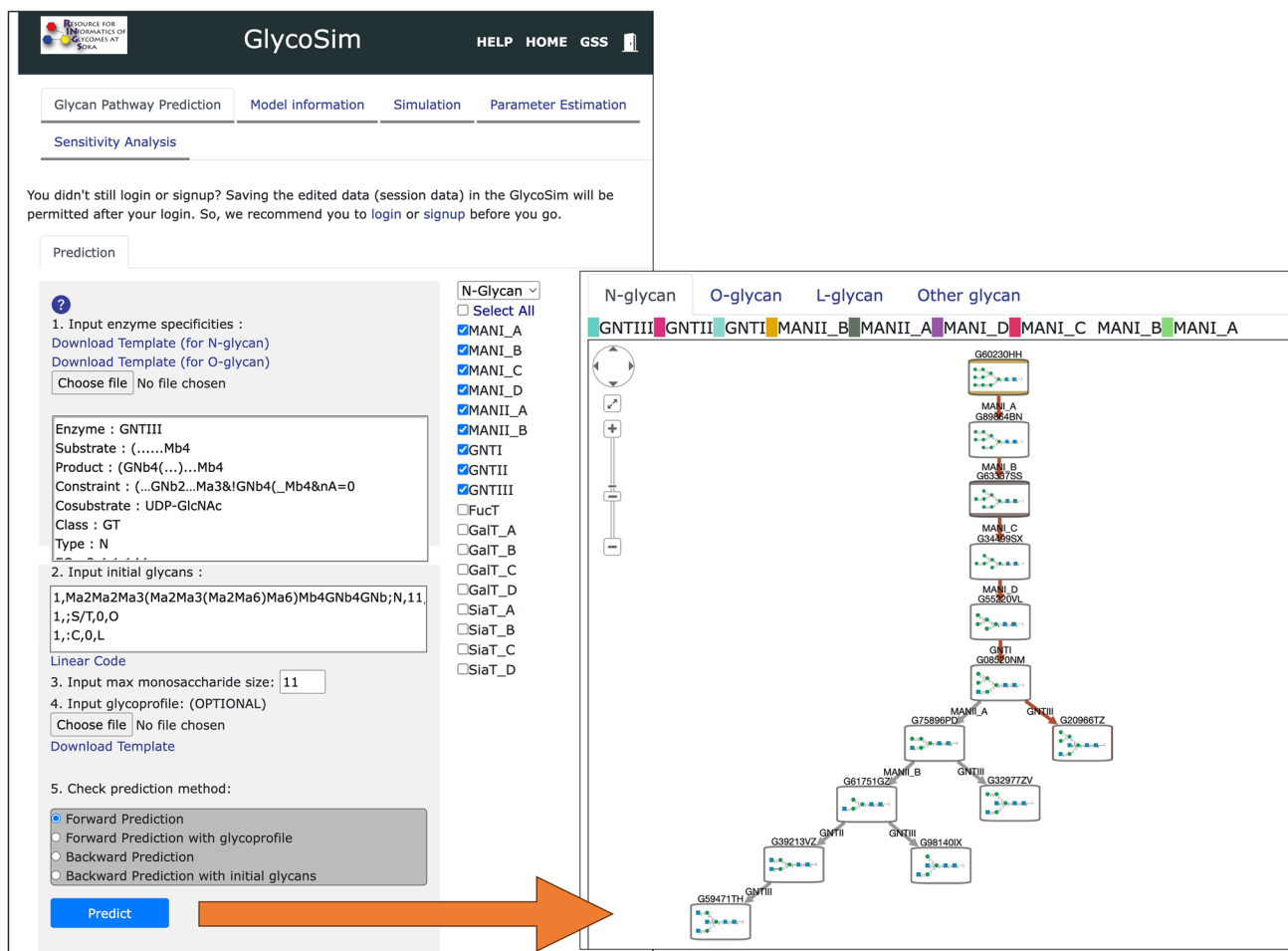


Fig. 3 A snapshot of the results of performing biosynthetic pathway prediction in GlycoSim. Enzyme specificities can be modified to add or remove constraints to each enzyme. The predicted pathway is then

stored as a mathematical model such that simulations can be performed on them

glycosciences have incorporated glycogene information to enable researchers better accessibility to such information. However, a centralized resource for such parameters is still a major need for the community. While more generalized databases for such parameters, such as BRENDA, exist, the substrate specificities are hard to define. The LiCoRR reaction rules based on LinearCode format, and the corresponding LiCoRRice format for IUPAC format, are considered the standards for such substrate specificities. We expect that databases providing such specificity information will be crucial to advance research in in silico glycosylation analysis.

Another issue is the lack of information on the localization of glycogenes especially at the compartmental level within the Golgi apparatus. Various models have been proposed in an attempt to simulate the Golgi, but research into the structure of the Golgi itself is still underway. Thus, the majority of the models presented in Table 1 have not considered multi-compartment models and are simply predictions of pathways disregarding localization. If localization

is taken into consideration, it is currently only possible to estimate where specific enzymes reside within the Golgi apparatus. Therefore, while there are multi-compartment models developed and shown to be relevant for bioprocessing specific glycosylation patterns [33, 44], they are not available freely for expanded use by the community. Further research in click-chemistry [45] and cryo-EM [46] are making headway to identify enzyme localization at the subcellular level, but the transfer of any new insights into the informatics side still requires much effort. Resources to store bioimaging data and 3D structures of glycoconjugates exist, but they are often not annotated sufficiently to identify the glycan-related components involved.

Moreover, this Trends has mainly focused on mammalian systems, but much research has also progressed in understanding glycosylation in bacteria [47] and plants [48] as well. However, the database integration and in silico tools for these systems are yet to be fully developed. Many of the tools and databases described here have started to accumulate

such data, but a user-friendly interface and infrastructure to enable plant and microbiologists to access and supplement these databases still needs to be constructed. Such developments would enable a better understanding of the roles of glycans in microbiomes and the environmental sciences. With the advancement of more high-throughput technologies and corresponding submission of high-quality data into databases and repositories, data-driven models can become more effective, especially with the remarkable development of large language models and AI technology.

In summary, there is much work to do in terms of bioinformatics, systems biology, microbiology, genome informatics, plant biology, etc. to better integrate the data produced and to develop user-friendly tools that allow researchers to access and analyze their data from a bird's eye view. The GlySpace Alliance [49], Glycoinformatics Consortium (<https://glic.glyco.info.org>), and Systems Glycobiology Consortium (<https://sysglyco.org>) are efforts to enable interactions between these heterogeneous fields aimed towards the same goal. The GlySpace Alliance consists of major glycan-based Web portals in the USA, Japan and Europe, where glycans, glycoproteins, and related metadata are shared freely. This alliance forms a basic informatics infrastructure for the glycosciences. The Glycoinformatics Consortium, or GLIC, is a group of glycoinformaticians who have developed databases and software for the glycosciences. Webinars and hackathons are held to enable interaction between glycoscience researchers and bioinformaticians, in an attempt to create synergy and more efficiently produce useful tools and data resources. Finally, the Systems Glycobiology Consortium, or SysGlyco, is a group of glyco-biologists and informaticians interested in developing systems biology tools for the glycosciences. Many of the products of this consortium were presented in this review. These groups are currently unfunded and number few, but time will tell when the fruits of their labor will contribute to the glycosciences and the life and environmental sciences as a whole.

Author contribution AK summarized the newest aspects of parameter estimation experiments and analysis, YAT and KFAK wrote the manuscript, and KFAK reviewed the entire manuscript.

Declarations

Ethics approval This work does not involve research on humans and/or animals.

Competing interests The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated

otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Pinho SS, Alves I, Gaifem J, Rabinovich GA. Immune regulatory networks coordinated by glycans and glycan-binding proteins in autoimmunity and infection. *Cell Mol Immunol.* 2023;20:1101–13. <https://doi.org/10.1038/s41423-023-01074-1>.
2. Broszeit F, Tzarum N, Zhu X, Nemanichvili N, Eggink D, Leenders T, Li Z, Liu L, Wolfert MA, Papanikolaou A, Martínez-Romero C, Gagarinov IA, Yu W, García-Sastre A, Wennekes T, Okamatsu M, Verheije MH, Wilson IA, Boons GJ, de Vries RP. N-glycolylneuraminic acid as a receptor for influenza A viruses. *Cell Rep.* 2019;27:3284–3294.e6. <https://doi.org/10.1016/j.celrep.2019.05.048>.
3. Casalino L, Gaieb Z, Goldsmith JA, Hjorth CK, Dommer AC, Harbison AM, Fogarty CA, Barros EP, Taylor BC, McLellan JS, Fadda E, Amaro RE. Beyond shielding: the roles of glycans in the SARS-CoV-2 spike protein. *ACS Cent Sci.* 2020;6:1722–34. <https://doi.org/10.1021/acscentsci.0c01056>.
4. Varki A, Cummings RD, Esko JD, Stanley P, Hart GW, Aebi M, Mohnen D, Kinoshita T, Packer NH, Prestegard JH, Schnaar RL, Seeberger PH. *Essentials of glycobiology*. 4th ed. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press; 2022.
5. Drula E, Garron M-L, Dogan S, Lombard V, Henriissat B, Terapon N. The carbohydrate-active enzyme database: functions and literature. *Nucleic Acids Res.* 2022;50:D571–7. <https://doi.org/10.1093/nar/gkab1045>.
6. Narimatsu H, Suzuki Y, Aoki-Kinoshita KF, Fujita N, Sawaki H, Shikanai T, Sato T, Togayachi A, Yoko-o T, Angata K, Kubota T, Noro E (2017) GlycoGene database (GGDB) on the semantic web. In: *A Practical Guide to Using Glycomics Databases*. Springer Japan, Tokyo, pp 163–175
7. Yamada I, Shiota M, Shinmachi D, Ono T, Tsuchiya S, Hosoda M, Fujita A, Aoki NP, Watanabe Y, Fujita N, Angata K, Kaji H, Narimatsu H, Okuda S, Aoki-Kinoshita KF. The GlyCosmos Portal: a unified and comprehensive web resource for the glycosciences. *Nat Methods.* 2020;17:649–50.
8. Kanehisa M (2017) KEGG glycan. In: *A practical guide to using glycomics databases*. Springer Japan, Tokyo, pp 177–193
9. Kotidis P, Donini R, Arnsdorf J, Hansen AH, Voldborg BGR, Chiang AWT, Haslam SM, Betenbaugh M, Jimenez del Val I, Lewis NE, Krambeck F, Kontoravdi C. CHOglycoNET: comprehensive glycosylation reaction network for CHO cells. *Metab Eng.* 2023;76:87–96. <https://doi.org/10.1016/j.ymben.2022.12.009>.
10. McDonald AG, Tipton KF, Davey GP. A knowledge-based system for display and prediction of o-glycosylation network behaviour in response to enzyme knockouts. *PLoS Comput Biol.* 2016;12:e1004844. <https://doi.org/10.1371/journal.pcbi.1004844>.
11. McDonald AG, Mariethoz J, Davey GP, Lisacek F. In silico analysis of the human milk oligosaccharide glycome reveals key enzymes of their biosynthesis. *Sci Rep.* 2022;12:10846. <https://doi.org/10.1038/s41598-022-14260-4>.
12. McDonald AG, Davey GP. Simulating the enzymes of ganglioside biosynthesis with glycologue. *Beilstein J Org Chem.* 2021;17:739–48. <https://doi.org/10.3762/bjoc.17.64>.

13. Hossler P, Goh L-T, Lee MM, Hu W-S. GlycoVis: visualizing glycan distribution in the protein lempN-glycosylation pathway in mammalian cells. *Biotechnol Bioeng.* 2006;95:946–60.
14. Aoki-Kinoshita KF. Analyzing glycan structure synthesis with the glycan pathway predictor (GPP) tool. *Methods Mol Biol.* 2015;1273:139–47.
15. Krambeck FJ, Betenbaugh MJ. A mathematical model of N-linked glycosylation. *Biotechnol Bioeng.* 2005;92:711–28. <https://doi.org/10.1002/bit.20645>.
16. Krambeck FJ, Bennun SV, Narang S, Choi S, Yarema KJ, Betenbaugh MJ. A mathematical model to derive N-glycan structures and cellular enzyme activities from mass spectrometric data. *Glycobiology.* 2009;19:1163–75. <https://doi.org/10.1093/glycob/cwp081>.
17. Akune Y, Hosoda M, Kaiya S, Shinmachi D, Aoki-Kinoshita KF. The RINGS resource for glycome informatics analysis and data mining on the Web. *OMICS.* 2010;14:475–86. <https://doi.org/10.1089/omi.2009.0129>.
18. Akune Y, Lin C-H, Abrahams JL, Zhang J, Packer NH, Aoki-Kinoshita KF, Campbell MP. Comprehensive analysis of the N-glycan biosynthetic pathway using bioinformatics to generate UniCorn: a theoretical N-glycan structure database. *Carbohydr Res.* 2016;431:56–63. <https://doi.org/10.1016/j.carres.2016.05.012>.
19. Chang A, Jeske L, Ulbrich S, Hofmann J, Koblit J, Schomburg I, Neumann-Schaal M, Jahn D, Schomburg D. BRENDA, the ELIXIR core data resource in 2021: new developments and updates. *Nucleic Acids Res.* 2021;49:D498–508. <https://doi.org/10.1093/nar/gkaa1025>.
20. Campbell MP, Packer NH. UniCarbKB: new database features for integrating glycan structure abundance, compositional glycoproteomics data, and disease associations. *Biochim Biophys Acta.* 2016;1860:1669–75.
21. Cheng K, Pawlowski G, Yu X, Zhou Y, Neelamegham S (2019) DrawGlycan-SNFG & gpAnnotate: rendering glycans and annotating glycopeptide mass spectra. *Bioinformatics (Oxford, England).* <https://doi.org/10.1093/bioinformatics/btz819>
22. Liu G, Neelamegham S (2014) A computational framework for the automated construction of glycosylation reaction networks. *PLoS ONE* 9. <https://doi.org/10.1371/journal.pone.0100939>
23. Liu G, Cheng K, Lo CY, Li J, Qu J, Neelamegham S. A comprehensive, open-source platform for mass spectrometry-based glycoproteomics data analysis. *Mol Cell Proteomics.* 2017;16:2032–47. <https://doi.org/10.1074/mcp.M117.068239>.
24. Liang C, Chiang AWT, Lewis NE. GlycoMME, a Markov modeling platform for studying N-glycosylation biosynthesis from glycomics data. *STAR Protocols.* 2023;4:102244. <https://doi.org/10.1016/j.xpro.2023.102244>.
25. Bao B, Kellman BP, Chiang AWT, Zhang Y, Sorrentino JT, York AK, Mohammad MA, Haymond MW, Bode L, Lewis NE. Correcting for sparsity and interdependence in glycomics by accounting for glycan biosynthesis. *Nature Communication.* 2021;12:4988. <https://doi.org/10.1038/s41467-021-25183-5>.
26. Thomès L, Burkholz R, Bojar D. Glycowork: a Python package for glycan data science and machine learning. *Glycobiology.* 2021;31:1240–4. <https://doi.org/10.1093/glycob/cwab067>.
27. Lundstrøm J, Urban J, Bojar D. Decoding glycomics with a suite of methods for differential expression analysis. *Cell Reports Methods.* 2023;3:100652. <https://doi.org/10.1016/j.crmeth.2023.100652>.
28. McDonald AG, Lisacek F. Simulated digestions of free oligosaccharides and mucin-type O-glycans reveal a potential role for Clostridium perfringens. *Sci Rep.* 2024;14:1649. <https://doi.org/10.1038/s41598-023-51012-4>.
29. Huang C-Y, Loo DM, Gu W. Modeling of glycosaminoglycan biosynthesis in intervertebral disc cells. *Comput Biol Med.* 2023;162:107039. <https://doi.org/10.1016/j.compbimed.2023.107039>.
30. Kotidis P, Kontoravdi C (2020) Harnessing the potential of artificial neural networks for predicting protein glycosylation. *Metab Eng Commun* 10. <https://doi.org/10.1016/j.mec.2020.e00131>
31. Krambeck FJ, Bennun SV, Andersen MR, Betenbaugh MJ. Model-based analysis of N-glycosylation in Chinese hamster ovary cells. *PLoS ONE.* 2017;12:e0175376. <https://doi.org/10.1371/journal.pone.0175376>.
32. Bennun SV, Yarema KJ, Betenbaugh MJ, Krambeck FJ. Integration of the transcriptome and glycome for identification of glycan cell signatures. *PLoS Comput Biol.* 2013;9:e1002813. <https://doi.org/10.1371/journal.pcbi.1002813>.
33. Yadav A, Vagne Q, Sens P, Iyengar G, Rao M. Glycan processing in the Golgi as optimal information coding that constrains cisternal number and enzyme specificity. *Elife.* 2022;11:e76757. <https://doi.org/10.7554/eLife.76757>.
34. Aoki-Kinoshita KF. Functions of glycosylation and related web resources for its prediction. *Methods Mol Biol.* 2022;2499:135–44.
35. Kellman BP, Zhang Y, Logomasini E, Meinhardt E, Godinez-Macias KP, Chiang AWT, Sorrentino JT, Liang C, Bao B, Zhou Y, Akase S, Sogabe I, Kouka T, Winzeler EA, Wilson IBH, Campbell MP, Neelamegham S, Krambeck FJ, Aoki-Kinoshita KF, Lewis NE. A consensus-based and readable extension of linear code for reaction rules (LiCoRR). *Beilstein J Org Chem.* 2020;16:2645–62. <https://doi.org/10.3762/BJOC.16.215>.
36. Hoops S, Sahle S, Gauges R, Lee C, Rgen Pahle J, Simus N, Singhal M, Xu L, Mendes P, Kummer U. COPASI—a complex pathway simulator. *Bioinformatics.* 2006;22:3067–74. <https://doi.org/10.1093/bioinformatics/btl485>.
37. Nairn AV, Aoki K, dela Rosa M, Porterfield M, Lim J-M, Kulik M, Pierce JM, Wells L, Dalton S, Tiemeyer M, Moremen KW. Regulation of glycan structures in murine embryonic stem cells: combined transcript profiling of glycan-related genes and glycan structural analysis. *J Biol Chem.* 2012;287:37835–56. <https://doi.org/10.1074/jbc.M112.405233>.
38. Hucka M, Finney A, Sauro HM, Bolouri H, Doyle JC, Kitano H, Arkin AP, Bornstein BJ, Bray D, Cornish-Bowden A, Cuellar AA, Dronov S, Gilles ED, Ginkel M, Gor V, Goryanin II, Hedley WJ, Hodgman TC, Hofmeyr JH, Hunter PJ, Juty NS, Kasberger JL, Kremling A, Kummer U, Le Novère N, Loew LM, Lucio D, Mendes P, Minch E, Mjolsness ED, Nakayama Y, Nelson MR, Nielsen PF, Sakurada T, Schaff JC, Shapiro BE, Shimizu TS, Spence HD, Stelling J, Takahashi K, Tomita M, Wagner J, Wang J. The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. *Bioinformatics.* 2003;19:524–31. <https://doi.org/10.1093/bioinformatics/btg015>.
39. Spahn PN, Hansen AH, Hansen HG, Arnsdorf J, Kildegaard HF, Lewis NE. A Markov chain model for N-linked protein glycosylation – towards a low-parameter tool for model-driven glycoengineering. *Metab Eng.* 2016;33:52–66. <https://doi.org/10.1016/j.ymben.2015.10.007>.
40. Tsopanoglou A, Jiménez del Val I. Moving towards an era of hybrid modelling: advantages and challenges of coupling mechanistic and data-driven models for upstream pharmaceutical bioprocesses. *Curr Opin Chem Eng.* 2021;32:100691. <https://doi.org/10.1016/j.coche.2021.100691>.
41. McDonald AG, Hayes JM, Davey GP. Metabolic flux control in glycosylation. *Curr Opin Struct Biol.* 2016;40:97–103. <https://doi.org/10.1016/j.SBI.2016.08.007>.
42. Fung Shek C, Kotidis P, Betenbaugh M. Mechanistic and data-driven modeling of protein glycosylation. *Curr Opin Chem Eng.* 2021;32:100690. <https://doi.org/10.1016/j.coche.2021.100690>.

43. Kremkow BG, Lee KH. Glyco-Mapper: A Chinese hamster ovary (CHO) genome-specific glycosylation prediction tool. *Metab Eng.* 2018;47:134–42. <https://doi.org/10.1016/j.ymben.2018.03.002>.
44. Jimenez del Val I, Nagy JM, Kontoravdi C. A dynamic mathematical model for monoclonal antibody N-linked glycosylation and nucleotide sugar donor transport within a maturing Golgi apparatus. *Biotechnol Prog.* 2011;27:1730–43. <https://doi.org/10.1002/btpr.688>.
45. Sørensen DM, Büll C, Madsen TD, Lira-Navarrete E, Clausen TM, Clark AE, Garretson AF, Karlsson R, Pijnenborg JFA, Yin X, Miller RL, Chanda SK, Boltje TJ, Schjoldager KT, Vakhrushev SY, Halim A, Esko JD, Carlin AF, Hurtado-Guerrero R, Weigert R, Clausen H, Narimatsu Y. Identification of global inhibitors of cellular glycosylation. *Nat Commun.* 2023;14:948. <https://doi.org/10.1038/s41467-023-36598-7>.
46. Katz M, Diskin R. Structural basis for matriglycan synthesis by the LARGE1 dual glycosyltransferase. *PLoS ONE.* 2022;17:e0278713. <https://doi.org/10.1371/journal.pone.0278713>.
47. Kawamoto A, Yamada T, Yoshida T, Sato Y, Kato T, Tsuge H. Cryo-EM structures of the translocational binary toxin complex CDTa-bound CDTb-pore from *Clostridioides difficile*. *Nat Commun.* 2022;13:6119. <https://doi.org/10.1038/s41467-022-33888-4>.
48. Kumar S, Wang Y, Zhou Y, Dillard L, Li F-W, Sciandra CA, Sui N, Zentella R, Zahn E, Shabanowitz J, Hunt DF, Borgnia MJ, Bartesaghi A, Sun T-P, Zhou P. Structure and dynamics of the Arabidopsis O-fucosyltransferase SPINDLY. *Nat Commun.* 2023;14:1538. <https://doi.org/10.1038/s41467-023-37279-1>.
49. Lisacek F, Tiemeyer M, Mazumder R, Aoki-Kinoshita KF. Worldwide glycoscience informatics infrastructure: the GlySpace alliance. *JACS Au.* 2023;3:4–12. <https://doi.org/10.1021/jacsau.2c00477>.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.