CHEMICAL ENGINEERING TRANSACTIONS
VOL. 110, 2024
A publication of AIDIC
The Italian Association of Chemical Engineering
Online at www.cetjournal.it
Guest Editors: Marco Bravi, Antonio Marzocchella, Giuseppe Caputo

ISBN 979-12-81206-10-6; **ISSN** 2283-9216
DOI: 10.3303/CET24110026

Protein Glycosylation in the Golgi Apparatus: Agent-Based Modeling of Molecular Interactions

Peter Götz*, Christian Jetschni

Berliner Hochschule für Technik, Bioprocess Engineering, Seestrasse 64, 13347 Berlin, Germany
Peter.Goetz@BHT-Berlin.de

Glycosylation is one of the most common post-translational modifications of proteins, it has significant influence on stability, biological activity and clearance rate of protein therapeutics. However, glycoprotein production usually leads to a complex mixture of glycoforms, with variations in both the sites of glycosylation and oligosaccharide structures. This wide range of diversity in the resulting glycoproteins involving a large variety of glycosylation enzymes included in this process pose a challenging task in an attempt to simulate those processes *in silico*. Agent-based modeling (ABM) has emerged as an effective tool for simulating complex behaviour, especially in pattern formation and self-organizing systems. For glycosylation, ABM allows a detailed and dynamic representation of individual agents and their interactions within biological entities, making it possible to simulate glycan pattern formation. An available *in silico* representation of protein glycosylation by an ABM (Jetschni and Götz, 2023) captures the stochastic nature of the involved molecular diffusion, convection and biochemical reaction steps within the Golgi apparatus. Model verification towards reproducibility and statistical analysis of pattern formation shows, that the modeling approach is robust and suitable to predict the outcome of protein glycosylation.

1. Introduction

Glycosylation in eukaryotic cells is one of the most common post-translational modifications of proteins and determines final glycoprotein structure, function, and stability. The process of protein glycosylation occurs in different subcellular organelles such as the endoplasmic reticulum (ER) and the Golgi apparatus and involves a series of sequential enzymatic reaction steps. Unlike protein translation, which is based on mRNA providing the sequence of amino acids in the protein, glycosylation is not predetermined by a template which is carrying information on glycan structure. The pathway to the final glycoprotein is determined by diffusion and intracellular convection coupled with biochemical reactions catalyzed by a variety of membrane-bound enzymes in the Golgi apparatus. Presence of these enzymes in different compartments has a significant effect on the final glycan structure, in combination with the stochastic nature of diffusion and the intracellular convection events, the glycosylation process leads to a heterogeneous product spectrum for the glycoproteins. Nevertheless, the spatial structure of the Golgi together with physical and biochemical parameters will result in a product spectrum which is exhibiting pattern formation for a limited number of molecular species with relatively high abundance. Since the Golgi apparatus itself is a product of self-organization (Tachikawa and Mochizuki, 2017), there is no defined a priori information on its structure and enzyme localization. Therefore, pattern formation in the glycoprotein product spectrum, arising from a self-organized structure, the Golgi apparatus, constitutes two connected levels of complex behaviour resulting from the collective interactions of molecular components. The macroscopic outcome of self-organization is generally not predictable by a model, all steps of the self-organizing process must be performed in order to yield a forecast with high probability.
From an industrial viewpoint, process development for production of diagnostic and therapeutic glycoproteins in eukaryotic cells is a challenging task. Product quality is significantly dependent on the glycan structure, which influences physical parameters (solubility, thermal stability) as well as biological activity (e.g. immunogenicity, clearance rate after injection). The unpredictability of glycosylation, due to its stochastic nature, allows no quantification of effects of changes in the production process environment.

Paper Received: 02 February 2024; Revised: 27 April 2024; Accepted: 05 June 2024
Please cite this article as: Goetz P., Jetschni C., 2024, Protein Glycosylation in the Golgi Apparatus: Agent-based Modeling of Molecular Interactions, Chemical Engineering Transactions, 110, 151-156 DOI:10.3303/CET24110026

For example glucose starvation, presence of hormones and vitamins, intracellular pH and concentrations of cations are all known to change the glycan structure and thus product quality (Goochee and Monica, 1990). Although it has been shown experimentally, that certain process parameters (reactor type, cultivation scale, temperature, cell density) have no significant impact on the glycosylation pattern (Butler and Reichl, 2017), a quantitative model description of glycosylation would have considerable value for process control and the implementation of a "digital twin". Creating a general *in silico* model for glycosylation would also allow a systematic approach to glycoengineering (Dammen-Brower et al., 2022). As an example, glycoengineeering of yeast for production of humanized glycoproteins was commercialized by GlycoFi Inc. (Wildt and Gerngross, 2005), further development of applications of yeast as production platform would profit from a computer-based predictive model of glycosylation.

First modeling approaches by Shelikoff et al. (1996) and Umana and Bailey (1997), extended by Krambeck and Betenbaugh (2005), are representing the glycosylation process in the four compartments of the Golgi apparatus by mass balances and kinetic equations. The four compartments can be modeled as a series of four ideally mixed continuous reactors (CSTR) or four plug flow reactors (PFR) (Hossler et al., 2007). These two assumptions are related to the hypotheses for the intracellular processes of vesicle transport and Golgi maturation respectively. Other modeling approaches consider the stochastic nature of glycosylation, applying neural networks (Senger and Karim, 2005), Markov chains (Spahn et al., 2016) or stochastic simulation algorithms (Fisher et al., 2019). More detail on model development since 1996 is summarized in a review article (Galleguillos et al., 2017).

The agent-based approach, which will be analyzed in this study, uses the discretization of the Golgi apparatus into a grid, on which agents can move and interact to simulate the process of glycosylation (Jetschni and Götz, 2023). The Golgi apparatus consists of cisternae, which are the reaction volumes where the protein modifications take place. The small size of these disk-shaped organelles, with a diameter in the order of 500 nm, and the rather large size of enzymes and glycoproteins allow only a limited number of these molecules in the reaction volume. Enzymes are located on the membranes of the Golgi apparatus, they catalyze the assembly and maturation of glycoproteins as these traverse through the Golgi compartments. For discretization (Figure 1), the assumed average size of a glycoprotein requires a spatial resolution of 10 nm. Simplifying the lumen of one of the cisternae to a square with a side length of 500 nm, the reaction volume is represented by three layers of patches with a grid size of 50 x 50 patches. The whole internal space of one cisterna would therefore consist of 7500 Patches, while the entire Golgi structure of eight cisternae for the four compartments would comprise 60,000 Patches. For simulation, for discrete time steps (ticks) glycoproteins move on the grid, eventually encountering membrane-bound enzymes, which may catalyze a biochemical reaction with a given probability. Proof of concept was achieved for this model, simulations of this stochastic approach exhibited pattern formation within the variety of resulting glycan structures.

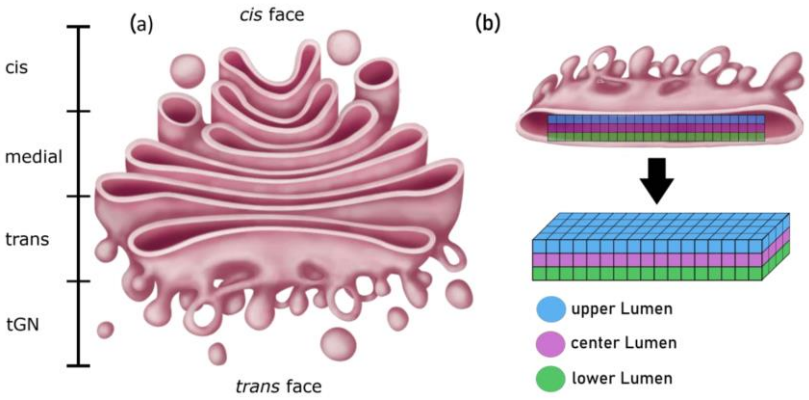

Figure 1: Discretization of the lumen (reaction volume) of the Golgi apparatus (a) into a grid (b) for ABM (Jetschni and Götz, 2023)

2. Materials and Methods

Considering the stochastic nature of the agent-based modeling approach, model verification requires an analysis of the fundamental behavior of the simulations: ● reproducibility of simulation results, ● parameter sensitivity, ● robustness towards variations in initial and boundary conditions. For all simulations, the computer model by Jetschni and Götz (2023) was used. The model implementation includes 13 different glycosylation enzymes leading to the observed formation of more than 430 different glycan structures. Their occurence is depending on parameters as the amount of enzyme utilized, the predefined reaction probabilities and the residence time within the reaction volumes. Investigating parameter influence on glycosylation, these structures are classified according to features related to biological function: ● glycan type (high-mannose, hybrid, complex),

● antennarity (count 0-4), ● sialylation (non, mono, bi, tri, tetra), ● fucosylation (core and antennary). Furthermore, the patterns in glycan spectra can be visualized, analyzed and compared after simulation by creating a corresponding mass spectrum, including glycan structures provided by GlycoWorkbench (Ceroni et al., 2008). This allows the comparison of simulations with different model parameter settings and with experimentally measured mass spectra. Simulations were carried out with 55 proteins and 950 enzymes in total (13 enzyme species in 4 compartments).

3. Results

In model verification, systematic exploration of initial condition and parameter space is used to confirm agreement with the predefined model specifications, assumptions and hypotheses. At first, variation of the initial location of enzymes within the membrane of the respective compartment and the initial location of the glycoproteins within the reaction volume was investigated. These initial conditions should have neglectable influence on the final glycan pattern, since they stem from the self-organizing Golgi apparatus with its random nature. For robust function of glycan synthesis, this is a necessary feature, which is correctly predicted by the model. The observed exception for simulations with short residence times and low reaction probabilities is outside the realistic range of operation and was expected from basic considerations on reaction/diffusion systems. The model parameters ● amount of enzyme, ● reaction probability, ● residence time in compartment, have similar effects on pattern formation. Their influence is illustrated here by the example of simulating two enzyme reaction probabilities. For both parameter values, the model was run 300 times, each run returned a total of 55 glycan structures on the 55 proteins. The frequency of occurrence for each structural feature was calculated by summing up the occurrences of the corresponding structural feature within these 55 glycans. Box plots show mean result and the variability between simulation runs (Figure 2).

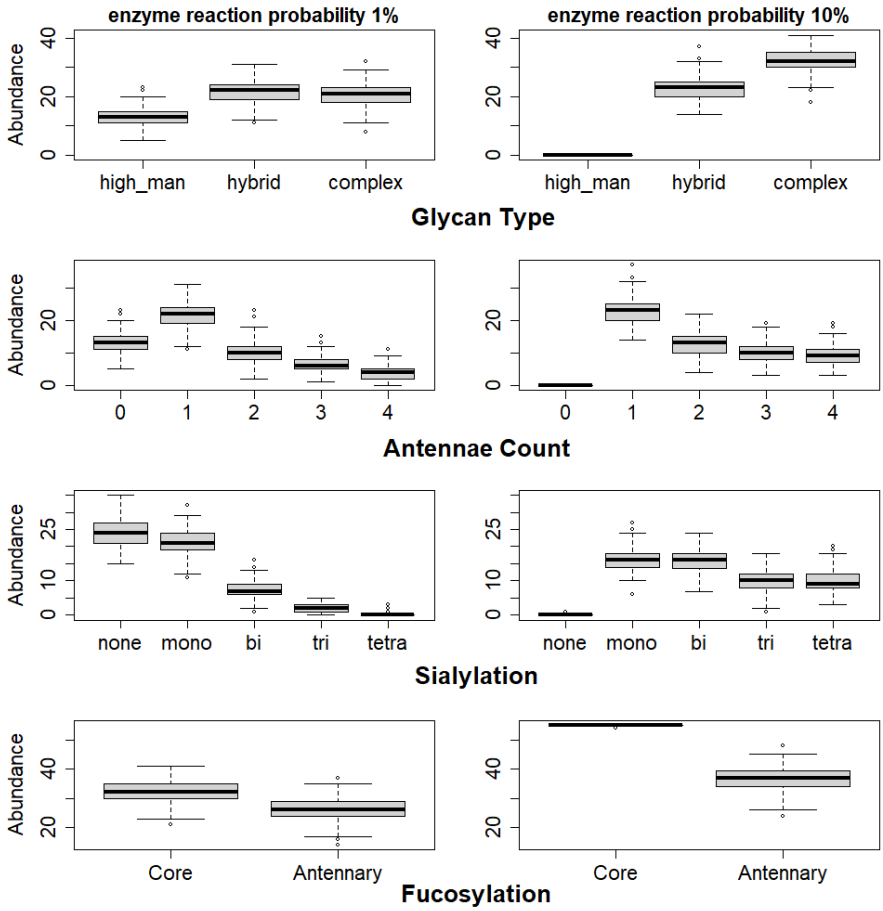

Figure 2: Differences in main structural features between an enzyme reaction probability of 1% (left column) and 10% (right column) over 300 simulation runs for 55 glycoproteins.

Regarding the glycan type between the enzyme reaction probability levels of 1% and 10% in Fig. 2, the complete disappearance of high-mannose structures, a slight increase in hybrid types, and a significant increase in complex types can be observed at 10%. This can be explained through the enhanced enzyme reactivity leading to increased processing of the high-mannose preglycan in the early stages of the N-glycosylation. The Golgi apparatus's compartmentalization and sequential nature provide additional glycoprotein substrates for the glycosyltransferases in later phases, which are responsible for the branching and elongation of glycan chains, resulting in the emergence of more hybrid and complex structures. At low reaction probability, available glycoprotein substrates are scarce for the glycosyltransferases in the later stages of the glycosylation process, resulting predominately in structures with a low number of antennae and subsequently in a lower number of sialylations. In general, the number of antennae increased with higher reactivity but plateaued by the different capping mechanisms of the N-glycosylation, e.g. the addition of a bisecting GlcNAc ensures a relatively even distribution of the number of antennae. For fucosylation, the sum of core and antennae fucosylations can be larger than 55, because both types can occur in one glycoprotein. At high reaction probability, almost all 55 cores are fucosylated and additionally more than half of the antennae are fucosylated as well.
Analyzing pattern formation for glycans within the sets of 300 simulations, adding occurrences of all identical glycan structures and normalizing to the maximum value exhibits distinct patterns in product spectra (Fig. 3, upper plots).

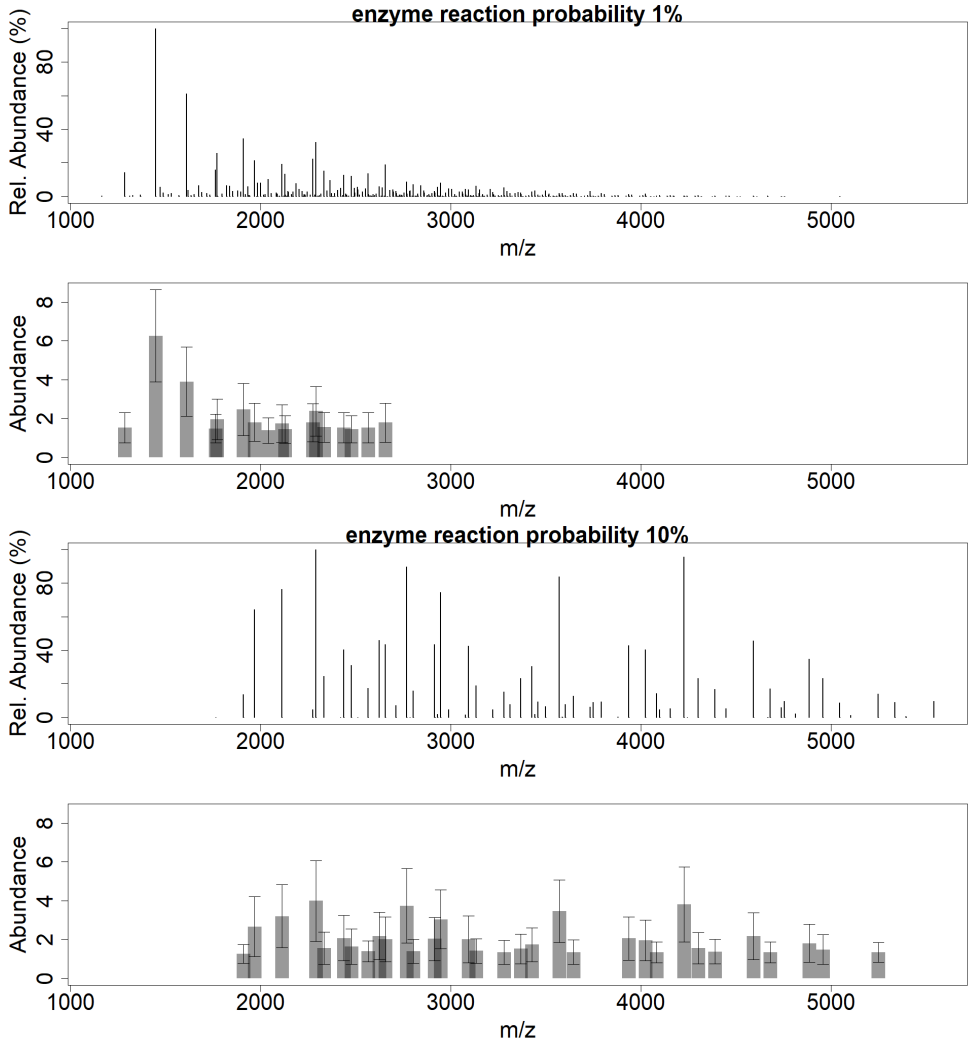

Figure 3: Simulated mass spectra from 300 simulation runs with an enzyme reaction probability of 1% and 10%. Upper plots: Relative abundance (cumulative over 300 runs) of all glycan structures. Lower plots: Mean absolute abundance and standard deviation of glycan structures (per simulation run) with a 5% cutoff threshold.

Higher m/z values indicate the hybrid and more complex structures, whereas lower m/z values usually correlate to the early stages in the N-glycosylation process, i.e. high mannose glycans with smaller and less complex structures. The 1% enzyme reaction probability spectrum shows a high abundance of lower weight structures and generally a high diversity in its resulting structures with a steady decline towards higher m/z values. In the 10% spectrum, less variety in structures can be observed and more complex structures accumulate.
Reproducibility of the pattern formation is shown by analyzing the occurrence of identical structures per run and calculating the standard deviation between 300 runs (Fig. 3, lower plots). Structures with a relative abundance below 5% are omitted in this analysis. Results show that more abundant structures are reproducibly favored in all simulation runs, confirming the hypothesis, that the structure of the Golgi apparatus leads to pattern formation.

4. Discussion

The granularity of molecular interactions on the intracellular scale is not adequately represented by reaction kinetics based on homogeneous concentrations. Spatial inhomogeneities on the molecular scale become important, making an agent-based approach the modeling tool of choice. Moreover, intricate spatial structures, the spatial distribution of molecules, diffusion and complex interactions can be included and easily implemented. Finally, capturing all these features may allow for the description of emergent behavior and properties characteristic of living systems.
Following the general pathway of glycosylation, N-glycans on glycoproteins in the cis-Golgi are of the high mannose type and usually contain eight or nine mannose (Man) residues. However, N-glycans are often processed in the Golgi by a set of α-mannosidases that remove Man residues to generate the Man_5 GlcNAc_2 Asn intermediate which is the substrate of the medial Golgi GT GlcNAcT-I. This structure plays a key role in the N-glycan synthesis because if a glycan is not broken down to this structure, it cannot be processed further and remains a high-mannose structure. The transfer of GlcNAc to Man_5 GlcNAc_2 Asn by GlcNAcT-I initiates the synthesis of hybrid and complex N-glycans. Following the initial trimming of mannoses by Man I, the number of possible N-glycan structures generated in the Golgi apparatus increases exponentially with each additional monosaccharide until the capping of antennae with sialic acid.
If one were to choose the ratio between reaction rate and residence time so that the glycans in a compartment could reach one of these final states above, only completely processed glycans would be produced and the variety of structures would be small, which agrees with the result of the respective simulations. In addition, since all enzymes are present in at least two different compartments, the enzymes that are located in the later compartments would not contribute to the pathway of glycosylation.
Opposed to this, finite residence times and high reaction probabilities lead to “immature” glycans, which may be a substrate for many different glycosyltransferases that compete. Also, a glycoconjugate may transit a Golgi compartment too quickly to be acted on by all glycosyltransferases capable of using it as a substrate. Such a glycan spectrum will include many glycoforms and these may vary in functional activity. This variety could be an evolutionary advantage and corresponds to experimental findings for glycan spectra as well as to the respective model simulations.

5. Conclusions

The assumption, that an agent-based model is able to describe pattern formation in glycan synthesis, was successfully verified. Variation of model parameters and initial conditions yielded reproducible results for the stochastic simulation approach. Independence of pattern formation from initial location of enzymes and initial location of glycoproteins within their compartment could be shown. According to the assumptions used for the model, dependence of the simulation results on biochemical reaction probabilities is observed and can be interpreted from a biological viewpoint. Next step in model development is model validation via parameter estimates to correlate simulation results with actual experimental data. Main parameters to be estimated in model validation will be enzyme distribution between the four compartments and activities of the enzymes involved, yet the respective parameter spaces are restricted due to prior knowledge from experimental research. After identifying suitable parameter sets, their uniqueness and robustness against perturbations will be investigated. During model validation, it may be necessary to refine the model regarding the simplification of spatial representation of the Golgi apparatus and the convective transport of glycoproteins between compartments. After successful model validation, applications in systematic glycoengineering will be possible.

Acknowledgments

We want to thank Stefan Hinderlich and Peter Neubauer for their support and advice.

References

Butler M., Reichl U., 2017, Animal Cell Expression Systems, Adv Biochem Eng Biotechnol., 175: 1–36 DOI: 10.1007/10_2017_31

Ceroni A., Maass K., Geyer H., Geyer R., Dell A., Haslam S. M., 2008, GlycoWorkbench: A Tool for the Computer-Assisted Annotation of Mass Spectra of Glycans, J. Proteome Res. 7:4, 1650-1659. DOI: 10.1021/pr7008252

Dammen-Brower K., Epler P., Zhu S., Bernstein Z. J., Stabach P. R., Braddock D. T., Spangler J. B., Yarema K. J., 2022, Strategies for Glycoengineering Therapeutic Proteins, Front. Chem. 10:863118. DOI: 10.3389/fchem.2022.863118

Galleguillos S.N., Ruckerbauer D., Gerstl M.P., Borth N., Hanscho M., Zanghellini J., 2017, What can mathematical modelling say about CHO metabolism and protein glycosylation?, Comput. Struct. Biotechnol. J., 15:212–21. DOI: 10.1016/j.csbj.2017.01.005

Hossler P., Mulukutla B.C., Hu W.-S., 2007, Systems Analysis of N-Glycan Processing in Mammalian Cells, PLoS ONE 2(8): e713. DOI:10.1371/journal.pone.0000713

Jetschni C., Götz P., 2023, Implementing an Agent-Based Modeling Approach for Protein Glycosylation in the Golgi Apparatus, Fermentation, 9(9):849. DOI: 10.3390/fermentation9090849

Krambeck, F. J., and M. J. Betenbaugh, 2005, A mathematical model of N-linked glycosylation, Biotechnol. Bioeng. 92:711–728.

Kulagina N., Besseau S., Godon C., Goldman G. H., Papon N., Courdavault V., 2021, Yeasts as Biopharmaceutical Production Platforms, Front. Fungal Biol. 2:733492. DOI: 10.3389/ffunb.2021.733492

Senger R.S., Karim M.N., 2005, Variable site-occupancy classification of N-linked glycosylation using artificial neural networks, Biotechnol. Prog. 21(6):1653–62. DOI: 10.1021/bp0502375

Shelikoff M., Sinskey A. J., Stephanopoulos G., 1996, A modeling framework for the study of protein glycosylation. Biotechnol. Bioeng. 50(1):73–90.

Spahn P.N., Hansen A.H., Hansen H.G., Arnsdorf J., Kildegaard H.F., Lewis N.E., 2016, A Markov chain model for N-linked protein glycosylation–towards a low-parameter tool for model-driven glycoengineering,l Metab Eng 33:52–66. DOI:10.1016/j.ymben.2015.10.007.

Wildt S., Gerngross T., 2005, The humanization of N-glycosylation pathways in yeast, Nat Rev Microbiol 3, 119–128, DOI: 10.1038/nrmicro1087