Published in final edited form as:

Metab Eng. 2016 January ; 33: 52–66. doi:10.1016/j.ymben.2015.10.007.

A Markov chain model for N-linked protein glycosylation – towards a low-parameter tool for model-driven glycoengineering

Philipp N. Spahn^{1,3}, Anders H. Hansen⁴, Henning G. Hansen⁴, Johnny Arnsdorf⁴, Helene F. Kildegaard⁴, and Nathan E. Lewis^{2,3,*}

¹Department of Bioengineering, University of California, San Diego, La Jolla, CA 92093

²School of Medicine, Department of Pediatrics, University of California, San Diego, La Jolla, CA 92093

³The Novo Nordisk Foundation Center for Biosustainability at the University of California, San Diego School of Medicine, San Diego, La Jolla, CA 92093

⁴The Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Hørsholm, Denmark

Abstract

Glycosylation is a critical quality attribute of most recombinant biotherapeutics. Consequently, drug development requires careful control of glycoforms to meet bioactivity and biosafety requirements. However, glycoengineering can be extraordinarily difficult given the complex reaction networks underlying glycosylation and the vast number of different glycans that can be synthesized in a host cell. Computational modeling offers an intriguing option to rationally guide glycoengineering, but the high parametric demands of current modeling approaches pose challenges to their application. Here we present a novel low-parameter approach to describe glycosylation using flux-balance and Markov chain modeling. The model recapitulates the biological complexity of glycosylation, but does not require user-provided kinetic information. We use this method to predict and experimentally validate glycoprofiles on EPO, IgG as well as the endogenous secretome following glycosyltransferase knock-out in different Chinese hamster ovary (CHO) cell lines. Our approach offers a flexible and user-friendly platform that can serve as a basis for powerful computational engineering efforts in mammalian cell factories for biopharmaceutical production.

Keywords

glycosylation; glycoengineering; Markov chains; flux-balance analysis

*Corresponding author: nlewisres@ucsd.edu.

Author Contribution

P.S. conceptualized and programmed the model and prepared the manuscript, A.H. carried out the glycoprofiling and consulted in glycan analysis, H.H. characterized the Fut8 knock-out cell lines and prepared the secretome samples, J.A. generated the Fut8 cell lines, H.K. supervised the experimental part of the project and wrote the corresponding parts in the manuscript, N.L. gave advice on model development, edited text and figures and supervised the computational part of the project. All authors commented on and approved the final manuscript.

Introduction

Protein glycosylation plays a key role in numerous biological processes, affecting protein folding, protein stability and solubility as well as binding to other proteins, for instance in ligand-receptor signaling (Varki et al., 2009). In biopharmaceuticals, drug quality has been shown to depend on proper glycosylation (Hayes et al., 2014; Li and d'Anjou, 2009; Niwa and Satoh, 2015) since glycoforms of a recombinant protein can significantly impact drug half-life in the body (Sørensen et al., 2009), drug potency (Shields et al., 2002), inflammatory and immunological responses to therapeutics (Dalziel et al., 2014), and safety (Hermeling S et al., 2004; Van Beers and Bardor, 2012). The various ways in which glycans can influence a protein's interactions with its environment make glycosylation a critical property to control in the production of recombinant proteins.

Efforts to control glycosylation in recombinant protein production have been difficult because of the complexity of glycans and the factors through which they are produced. The type and quantity of glycans attached to glycosylation sites across a protein cannot be readily read out from the genetic code. Instead, glycans are synthesized through a complex sequence of reactions involving a multitude of enzymes (glycosyltransferases and glycosidases) whose expression level and kinetics will affect glycosylation, as will a diverse plethora of other influencing factors. Among them are metabolite concentrations, various cell culture parameters such as shearing forces and temperature (Hossler et al., 2009) as well as the enzymatic accessibility of the glycosylation site as determined by the protein's structure (Senger and Karim, 2008; Thaysen-Andersen and Packer, 2012). Consequently, the complex interplay of these influencing factors presents great challenges to its rational manipulation for glycoengineering purposes. This challenge has been a particular hurdle for biosimilar development, where their glycoprofiles are required to match those of the primary drugs, within bounds set by regulatory agencies.

Experimental control strategies, including genetic modifications and media additives, have been used to manipulate glycosylation (Brühlman et al., 2015; Shi and Goudar, 2014; Sou et al., 2014; St Amand et al., 2014), but so far a systematic understanding of these strategies is largely lacking. Therefore, computational models of glycosylation carry a great potential in elucidating control strategies in a rational and cost-efficient manner. In fact, a number of computational glycosylation models have emerged in the past 10 years which have greatly contributed to the increased understanding of the glycosylation process as well as to its application in glycoengineering efforts (Hossler, 2012; Neelamegham and Liu, 2011; Spahn and Lewis, 2014). However, while the most advanced models describe glycosylation in great detail, they require the specification of a large number of parameters. Thus, such models are more difficult to use since the wealth of kinetic and other information required to run these models is usually not readily available. In this work we present an alternative modeling approach that is not based on kinetics, but instead uses methods from Markov chain theory and flux balance analysis to describe glycosylation as a stochastic process. We demonstrate that glycoprofiles for glycoengineered cell lines can be predicted in the absence of kinetic information after required model parameters are obtained through training the model on measured glycoprofiles. The method is validated on a dataset comprising recombinant EPO, IgG, and the endogenous secretome from cell lines derived from the various CHO lineages.

Thus, we present a novel modeling framework that shows high potential for glycoengineering in biopharmaceutical and biosimilar development.

Results

Generation of the generic N-glycosylation reaction network

By adopting previously published reaction rules (Bennun et al., 2013) (Table 1) we constructed a generic N-glycan reaction network, capable of generating the glycoform complexity seen in CHO cell lines. We apply our analysis to N-glycosylation in CHO, since it is the primary host for biopharmaceutical production (Jayapal et al., 2007; Kildegaard et al., 2013; Xu et al., 2011). However, the approach itself can be applied to any set of reaction rules for glycan synthesis, regardless of glycan type or host cell. Based on the known enzyme specificities, the algorithm constructs the complete N-glycosylation reaction network required to generate all glycans of a user-defined complexity level. Specifically, the complexity level k refers to the set of glycans that can be produced from the initial Man₉GlcNAc₂ glycan in k reaction steps (iterations) or less following the reaction rules (Fig. 1A, Table 1, Suppl. Fig. 1). The total number of possible glycans rises exponentially after the first few iterations, but since the presence of terminal sialic acids does not allow further modification of individual glycan branches, this rapid increase levels off after ~16 iterations when terminal sialic acids start to reduce the number of new substrates for the next iteration (Fig. 1A). Glycosylation reactions are generally considered to take place in different compartments of the Golgi apparatus (Moremen et al., 2012), so in our model glycans are labeled with a localization tag to effectively describe transitions to other compartments (Fig. 1B). The network generated in this way is a generic network (rather than being specific for a particular cell line) since it represents a repository of all reactions that, theoretically, could occur in CHO. To make the network specific to a cell line or recombinant protein, it is tailored to an observed glycoprofile, as laid out in detail in the next section.

The size of these generic networks necessitates the use of a suitable modeling format. Constraints-based modeling (e.g. flux-balance analysis) is a highly scalable alternative to kinetic approaches to analyze these reaction networks since it requires few parameters and its predictive capabilities have been successfully demonstrated in numerous studies (Bordbar et al., 2014; Lewis et al., 2012). However, since these methods operate on the generic reaction network and do not account for the reaction rates specific to cell lines or individual glycoproteins, they may produce unsatisfying results when applied to glycosylation. Indeed, constraints-based methods gave a poor glycosylation prediction with knock-out profiles and did not qualitatively match the corresponding experiments (Fig. 1C). It is thus mandatory to develop alternative modeling frameworks that manage to integrate glycoprotein specificities while maintaining low parametric demands.

Glycosylation can be effectively described as a Markov chain

By omitting the sugar nucleotides involved in glycan reactions, one can regard each glycan as a state in a stochastic network that transitions to other states (glycans) with certain transition probabilities (Fig. 1B). Since the reactions possible on a certain glycan only

depend on its current structure and not on its chemical history, these transition probabilities naturally fulfill the Markov property stating independence of the past (Suppl. Fig. 2A). That way, glycosylation can be described as a time-discrete Markov chain with start distribution π_0 concentrated on the initial $\text{Man}_9\text{GlcNAc}_2$ glycan and a transition matrix Π comprising the probabilities of transitions between glycans (Figs. 1B, 2A,B; Suppl. Fig. 2B). In order to model glycan secretion from the trans-Golgi, artificial absorbing states are introduced to allow the chain to virtually stop in a particular glycan, thereby mimicking its secretion (Fig. 2B). If transition probabilities are known, Markov chain theory allows one to easily compute the probability of the chain being absorbed in a particular state, i.e. the probability of yielding a certain glycan in the secreted glycoprofile (Fig. 2C).

Glycosylation is a highly complex process depending on various factors whose quantitative interplay is incompletely understood. As a consequence, direct *a-priori* derivation of all transition probabilities in the entire reaction network for an arbitrary glycoprotein proves challenging. However, if a measured glycoprofile is used as a calibration standard, transition probabilities can be inferred by fitting the model to the measured frequencies in the profile (Fig. 2D). In a first step, flux variability analysis (Burgard et al., 2001) is used to identify all reactions in the generic glycosylation network that do not make a contribution to synthesize the observed glycoprofile. These reactions are then trimmed from the network (see Materials & Methods for details). This accounts for the fact that on a specific glycoprotein typically only a small subset of all possible reactions actually takes place. The stoichiometric information in this reduced reaction network is then used to set a large portion of transition probabilities to zero in the Markov transition matrix since non-zero (yet still unknown) transition probabilities only have to be defined if a corresponding reaction is present in the reduced network (Suppl. Fig. 2C,D).

Standard optimization techniques could be applied to find transition probabilities that reproduce the frequencies in simple glycoprofiles (Fig. 2D); however, most complex glycoprofiles require thousands of reactions, and at this size, standard optimization algorithms are computationally too demanding. Fortunately, the intrinsic interrelatedness of transition probabilities and fluxes (Fig. 1B) allows a canonical transformation, in which the unknown transition probabilities correspond to a steady-state flux solution in an equivalent flux-balance problem (Suppl. Fig. 3; Suppl. Text), enabling the use of the highly scalable toolset of constraint-based modeling. Since flux-balance problems are typically under-determined, we use Monte-Carlo sampling to obtain a representative set of steady-state flux solutions which are then back-transformed into transition probabilities. Due to the large space of possible steady-state fluxes, the inferred transition probabilities generally show a considerable degree of variance (Suppl. Fig. 4). Thus, rather than one transition matrix, an array of transition matrices is retained, each of which describes a Markov chain that reproduces the measured glycoprofile (Fig. 3).

Adjustments of the wildtype reaction network can be used to simulate enzyme knock-down and overexpression

The array of transition matrices obtained through this fitting process represents a “wildtype”-specific Markov model of glycosylation for the particular glycoprotein. To

simulate the quantitative effects of enzyme knock-downs on the wildtype glycoprofile in this stochastic framework, the wildtype transition matrices are modified to obtain a knock-down model. First, the algorithm identifies all reactions catalyzed by the affected enzyme, and the corresponding transition probabilities are scaled down by a factor ϵ ($\epsilon = 0$ for a complete knock-out) (Fig. 4: steps 1–2). Since downstream transition probabilities from any given glycan must add up to 1, probabilities for the remaining transitions must be adjusted accordingly (Fig. 4: step 3). If the wildtype network does not have another transition for a given glycan apart from the one obliterated by the knock-out, it is assumed that the glycan is passed on to the next Golgi compartment or, if the knock-out took place in the trans-Golgi, become secreted (Fig. 4: step 4). Apart from these modifications, all other transition probabilities remain unchanged.

One further modification may be necessary when an enzyme knock-out leads to the appearance of glycans that were not present in the wildtype network. Since the wildtype network is tailored to contain only reactions required to produce the wildtype glycoprofile, a glycan that is passed to another compartment as a consequence of a knock-out in the previous compartment (Fig. 4, step 4) may fail to be further modified in the new compartment since corresponding reactions do not occur in the wildtype (Fig. 5A). However, *in vivo*, such “alien” glycans could still be substrates for enzymes in the new compartment, so it is reasonable to expand the model appropriately. For this, alien glycans are processed based on information provided by the wildtype reaction network. It is assumed that the alien will behave similarly to the wildtype glycan with the highest chemical resemblance. Thus, the algorithm first identifies a “closest match” among all wildtype glycans in the same Golgi compartment, based on chemical and structural similarity (Fig. 5B, step 1; see Materials & Methods for details). The reactions present on the closest match are then copied over to the alien, given it has a corresponding reaction substrate. Since reactions would be carried out by the same enzymes, transition probabilities are assumed to be identical to their analogs on the closest match or assumed to be as close as possible, in case certain reactions do not have analogs on the glycan (for instance due to a missing branch) (Fig. 5B; Suppl. Fig. 5).

With these modifications, each transition matrix representing a wildtype Markov chain is transformed into a Markov chain modeling an enzyme knock-down. Hence, by running these chains the resulting absorption probabilities would correspond to predicted frequencies in the mutant glycoprofile, as exemplified in the next section.

Enzyme overexpression can be modelled in a similar fashion by scaling up rather than scaling down the corresponding transition probability and adjusting the remaining transitions, analogous to the knock-down scenario (Fig. 6).

Glycoprofiles from mutant cell lines can be predicted without kinetic data

We validated our modeling approach with test datasets both from the literature and our laboratory. Following the procedure above, the Markov chain was fit to experimentally derived glycoprofiles from wildtype cell lines as the only required input source. In particular, as laid out above, no kinetic information of any kind needed to be estimated and fed into the model. The first glycoprofile analyzed, as reported in (Imai-Nishiya et al., 2007) for an antibody producing CHO cell line, consisted of fucosylated, bi-antennary glycans. Fitting

this profile to the Markov model revealed that only a small-scale network comprising less than 50 reactions was needed to generate the profile. Experimental knock-down of Fut8 using siRNA had led to successful omission of the core fucosylation and the substitution of the three glycans with their non-fucosylated versions (Imai-Nishiya et al., 2007), a result very well predicted by our Markov model in simulations (Fig. 7A). Although the inferred Markov transition probabilities for the internal reactions showed a significant degree of variance (Suppl. Fig. 4A), variance in the predicted glycoprofile turned out to be negligible (see Discussion). Consequently, our model proves to yield high predictive power on small-scale glycosylation networks.

Aiming for validation on a more complex dataset, we analyzed recently published glycoprofiles from a CHO-GS cell line expressing erythropoietin (EPO) (Yang et al., 2015). The size of the reconstructed network exceeded 500 reactions, mostly due to the vast diversity of pathways by which the tetra-antennary glycans can be processed in the trans-Golgi (not shown). Despite the variance in inferred transition probabilities (Suppl. Fig. 4B), simulation of a GnTIV knock-out yielded good quantitative agreement with the experimental knock-out profile (Fig. 7B), demonstrating that the Markov model can serve as a helpful predictive tool even in high-dimensional networks where intuitive analysis of knock-outs is no longer possible due to the sheer number of possible pathways.

Finally, we tested if our model could predict mutant glycoprofiles even when applied to a mixture of proteins rather than one isolated protein. Glycoprofiling of the entirety of secreted proteins from a cell line (its “secretome”) is a reliable and efficient method to study perturbation of glycosylation in knock-outs or media changes. Thus, we tested if the Markov model would be capable of predicting secretome glycosylation. We generated Fut8 knock-out cell lines using the CRISPR/Cas9 genome editing technology, followed by FACS sorting (Grav et al., 2015). Released glycans from all harvested secreted proteins were labeled and analyzed using HPLC-MS. This sort of secretome glycoprofiling quantifies all major N-linked glycans attached to secreted proteins but does not differentiate between individual glycoproteins. This way, a pooled secretome glycoprofile of a CHO-S suspension culture featuring bi-, tri- and tetra-antennary structures was obtained and fit to the Markov model (Fig. 7C) which would now describe the average reaction rates over the entire secretome. Simulation of a Fut8 knock-out proved to closely resemble the measured knock-out profile when applying the “Equal Ratio” model variant while the “Minimum Adjustment” approach (see Materials & Methods) performed weaker (Fig. 7C; see Discussion). Nevertheless, these results show that even in a challenging setting with incomplete experimental resolution and on a moderately complex network (~400 reactions) our Markov modeling framework proves to be an effective predictive tool for glycoengineering purposes.

Implementation of Golgi compartmentalization is mandatory for glycosylation models

Glycosylation involves multiple glycosyltransferases in the Golgi. Furthermore, the non-random, sequential action of glycosyltransferases on the growing glycan chain occurs through differential localization of these enzymes to different compartments of the Golgi (Colley, 1997; Nilsson et al., 1994, 1993a; Rabouille et al., 1995; Tu and Banfield, 2010). Knowledge on enzyme localization is of relevance since co-localized enzymes compete for

the same substrates. Thus, different assumptions on enzyme localization considerably impact reaction network topology and, as a consequence, glycoengineering strategies. While the localizations assumed for this model rely on a widely accepted consensus (Table 1) (Moremen et al., 2012), there remain disagreements with how and to what extent Golgi compartmentalization should be implemented in glycosylation models. Given the availability of mutant glycoprofiles, our model can test the impact of Golgi compartmentalization by running simulations with different localization scenarios and comparing them to the experimental results. To demonstrate this, we ran knock-out simulations with all enzymes localized to the same compartment, thus effectively turning the Golgi into a non-compartmentalized organelle. As seen from our validation datasets, this leads to knock-out predictions not in congruence with the experimental glycoprofiles (Fig. 8). The discrepancies seen in this case stem from changes in network topology caused by the emergence of reactions that were absent in the compartmentalized case due to the confinement of enzymes to different compartments. For instance, in the case of the antibody glycosylation network (Fig. 8A), absence of compartmentalization would imply competition between GalT and Fut8 which was not present with both enzymes being localized to different compartments. As a consequence, simulation of a Fut8 knockout in this scenario significantly changes the predicted glycoprofile which would now only contain a single fully galactosylated glycan (glycan #6, Fig. 8A). Similar changes in network topology lead to the falsely predicted glycans in a GnTIV knock-out if GalT is assumed to directly compete with GnTIV in a non-compartmentalized Golgi (Fig. 8B). We therefore conclude that Golgi compartmentalization is a property being vital for computational glycosylation modeling efforts. In addition, by testing different localization scenarios, our model might serve as a tool to assess the likelihood of localization hypotheses based on the congruence between predicted and observed knock-out glycoprofiles.

Discussion

A non-kinetic approach to glycosylation modeling

Computational glycosylation models have greatly contributed to the understanding of this complex cellular process. Previous models have addressed diverse aspects of glycosylation, such as the dynamics governing the initial co-translational oligosaccharide attachment (Shelikoff et al., 1996), sequence motifs correlating with glycosylation site occupancy (Senger and Karim, 2005), and prediction of the total glycan diversity produced by a cell based on enzyme kinetics and expression (Hossler et al., 2007; Kawano et al., 2005; Krambeck and Betenbaugh, 2005). Other models have focused on predicting glycoprofiles on pharmaceutically relevant proteins, such as monoclonal antibodies (del Val et al., 2011; Jedrzejewski et al., 2014; Kaveh et al., 2013). In this work, we follow a similar objective as we aim to address problems arising in the glycoengineering of individual glycoproteins produced in cell culture. However, we present a more pragmatic approach by subsuming the various factors influencing glycosylation into transition probabilities. A probabilistic framework is especially effective for modeling glycosylation, since probabilities can be thought of as “containers” comprising a multitude of predictor variables whose quantitative link to the process under study are incompletely understood. While our model does not explicitly contain the mechanistic resolution of some biophysical properties influencing

glycosylation, the influences of these mechanisms are implied in each probability. For instance, the scaling factor E in the “overexpression” simulations (Fig. 6) refers to transition probability as a whole and not to the actual level of protein over-expression required to achieve the effect. However, sacrificing mechanistic resolution greatly unburdens model utilization since laborious *a-priori* parameter estimation is not necessary in this context. In fact, after being trained on a standard condition, such as a wildtype profile, our model precision is comparable to kinetic models (del Val et al., 2011). Furthermore, the requirement of a preceding training on an initial standard profile does not represent a shortcoming of the method; rather it allows flexibility to account for any specific process, recombinant protein, and cell line being studied. In addition, in many situations, the starting glycoprofiles are usually known (e.g. a wildtype host producing a recombinant protein) and glycoengineering efforts seek to modulate this profile for biosimilar development (da Silva et al., 2014; Tsuruta et al., 2015) or to obtain more desired branching (Fan et al., 2014; Gornik et al., 2012; Niwa and Satoh, 2015).

Possible factors affecting prediction accuracy

The variance observed in our predicted knock-out glycoprofiles is a direct consequence of the flux variability in the corresponding flux-balance problem, which in turn is implied by the underdetermined linear system (more reactions than metabolites). Thus, this variance is an inherent property of the reaction network rather than a shortcoming caused by insufficient sampling. It is interesting to see, though, that variance in probability for “internal” transitions within the wildtype network does not necessarily translate into variance in the predicted glycan frequencies (Fig. 7A; Suppl. Fig. 4A). The reason for this is the presence of certain “bottlenecks” in the reaction network topology, such as the transition of one key intermediate glycan (e.g. glycan #1 in Fig. 8A) to another compartment. Since these transitions have to occur for all possible pathways that lead to the secreted profile (and thus occur with a probability of 1), variability further upstream the reaction network will have no effect on the frequency of secreted glycans.

Apart from the flux variability, the adjustment of transition probabilities following a knock-down is another factor affecting prediction accuracy. The “Equal ratio” approach implemented in our model assumes conservation of reaction rate ratios, but other approaches are conceivable in theory. An alternative option would be, for instance, to require the adjusted transition probabilities to minimally deviate from the wildtype given the knock-down constraint - a rationale very similar to the MOMA algorithm in flux-balance analysis (Segrè et al., 2002). However, this alternative approach performed distinctly weaker than the “Equal ratio” option (Fig. 7C). Minimum adjustment (although maximizing parsimony) might yield a weaker performance because it does not necessarily have a biological correlate, whereas the “Equal Ratio” model reflects a genuine biochemical process by maintaining the natural competition between co-localized enzymes for the same glycan substrate.

It is important to note that, although our model describes “transitions” of glycans from one compartment into another, these do not necessarily refer to transport reactions; instead, the interpretation of these transitions depend on the assumed architectural model of the Golgi

(Moremen et al., 2012). For example, under the assumption of the “vesicular transport” model in which cargo migrates through static Golgi compartments, glycan transitions between compartments would indeed refer to glycan trafficking in vesicles. In contrast, in the “cisternal maturation” model Golgi compartments migrate themselves while changing their enzymatic composition. In this situation, a “transition” from one compartment to another would be interpreted as an event where the glycan remains unmodified until the enzyme composition in its compartment has changed. In any case, our model does not depend on any of these interpretations and does, thus, not depend on what architectural concept of the Golgi is proposed.

A major challenge in predicting glycoprofiles following a perturbation is the synthesis of glycans that were not possible in the wildtype network. The alien-processing algorithm implemented in our model (Fig. 5) aims to address this challenge by suggesting that new glycans are processed analogously to their closest chemical matches in the wildtype. In our analysis we demonstrated that this approach successfully predicts novel glycans. However, again, the assumptions concerning chemical resemblance are, given the absence of robust kinetic data, based on parsimony which may be inappropriate in certain cases. Our algorithm differs from previously published glycan alignment methods (Aoki et al., 2003; Lee et al., 2015) as it does not penalize gaps, i.e. an alien with a missing branch would not necessarily be mapped to the wildtype glycan with the least number of additional monosaccharides (Suppl. Fig. 5: step 1). Our algorithm is reasonable for our specific purpose, though, since a shorter branch will likely be subject to other enzymes than the neighboring branch, so the glycan would be more likely to display different chemical behavior. Future developments can build upon this framework to address these current uncertainties and better predict novel glycans following perturbations.

The enzymatic activities of many mammalian glycosyltransferases have been carefully characterized. However, frequently, the *in vivo* binding affinities are not fully detailed, and so predictions must be made without complete knowledge. In particular, since our approach relies on fitting, the provided wildtype glycan profile represents the basis for reconstruction of the wildtype reaction network and, in turn, prediction of its adjustments upon a knock-out. As laid out above, these adjustments mainly comprise of shutting down the reactions to be knocked out and funneling the reaction flux towards the alternative reactions for the particular substrate (Fig. 4: steps 3,4). This method proved to work well in our test dataset, but it may fail in situations where alternative reactions occurred at very low rates in the wildtype, so associated glycans were not detected in the glycoprofile. It is apparent, therefore, that the accuracy of knock-out prediction in our model depends on the accuracy by which the wildtype network is reconstructed which in turn depends on the comprehensiveness of the wildtype glycoprofile annotation. Improvements in analytical glycomics, such as higher resolution techniques and high-throughput glycoprofiling (Cummings and Pierce, 2014; Shubhakar et al., 2014) will help to reconstruct the reaction network in higher precision and, consequently, further improve knock-out predictions.

Future prospects for model enhancement

Our model constructs a generic N-glycosylation reaction network that is sufficiently large to be fit to nearly all possible glycoforms obtained from CHO. However, our list of enzymes (Table 1) does not include enzymes that show low expression or lack of activity in CHO, e.g. GnTIII or a6SiaT (Butler and Spearman, 2014; Yang et al., 2015). Although their inclusion would imply a higher computational burden, these additional enzymes could readily be integrated in the generating algorithm if a specific glycoprofile under study requires the corresponding reactions. Also, while we focused on N-glycosylation in this study due to its importance in biotherapeutics (Shriver et al., 2004) and the availability of data from the literature, the method is by no means restricted to it. In particular, since the modeling framework only encompasses glycosylation itself, not glycan attachment to the protein backbone, O-glycosylation can be analogously modeled with the methods described in this paper (Suppl. Fig. 7).

In addition to the enzyme diversity supported by the model, enzyme localization is another issue that is of primary importance to model development. As shown by our model, Golgi compartmentalization and enzyme localization greatly influence reaction network topology and, as a result, the prediction of knock-out profiles. Although the degree to which individual enzymes have been studied may vary, evidence of differential localization of glycosylation enzymes has been unambiguously demonstrated. Thus, in contrast to previous studies that attempted a non-compartmentalized view on Golgi reaction topology (McDonald et al., 2014), we find it imperative to carefully consider differential Golgi enzyme localization for modeling purposes. We are aware, though, that exclusive assignment of enzymes to distinct compartments likely represents an oversimplification since enzyme activity actually only peaks in a certain compartment but spans multiple compartments (Nilsson et al., 1993b; Rabouille et al., 1995). Future developments of our model could account for this by localizing enzymes to multiple compartments and weighing the transition probabilities for the catalyzed reactions with the enzyme's occurrence probability in the particular compartments, similar to efforts in kinetic models (Hossler et al., 2007).

Finally, numerous previous studies have demonstrated the impact of media on glycosylation (Brühlman et al., 2015; Hossler et al., 2014; Surve and Gadgil, 2015). Thus, it will be of great value to integrate our model with models of other metabolic pathways. Such coupled models would enable predictive glycoengineering using both genetic means (Lin et al., 2015; Meuris et al., 2014; Wu and Chan, 2014), and media alterations, such as alternative sugars and other media additives. Furthermore, such a coupled model could help identify the mechanisms underlying media effects on glycosylation. Thus, predictive models comprising a virtual representation of the cell culture setup including media composition, cell metabolism and glycosylation will be powerful tools for the fast and cheap development of glycoengineering strategies.

Conclusion

Here we have presented a powerful low-parameter framework for predicting glycoprofiles following genetic perturbations. Such a platform has great potential for glycoengineering as cell engineers aim to rationally design glycosylation on novel products and mimic

glycoprofiles on biosimilars. Furthermore, prospective applications could also be achieved through reversing the workflow and, instead of predicting a glycoprofile, use different profiles to infer specific changes in glycosyltransferase activity which could have potential use in the study of human health, including congenital disorders of glycosylation, the impact of genotype on milk oligosaccharide profiles, or cancer-related phenotypes.

Materials & Methods

Constraints-based methods

Methods from the COBRA toolbox v.2.0 (Schellenberger et al., 2011), as implemented in MATLAB (MathWorks), were used to write the algorithms that generate the glycosylation reaction network (up to a user-defined complexity level). The starting glycan list only contains the initial $\text{Man}_9\text{GlcNAc}_2$. In each iteration, the algorithm scans all glycans in the list to look for substrates for any of the 10 enzymes listed in Table 1. If a matching substrate is found and the localization of the glycan matches the localization of the enzyme, the corresponding reaction is added to the network and the resulting glycan is added to the glycan list. This procedure is repeated until the user-submitted iteration limit (the “complexity level”) is reached (See Suppl. Fig. 1 for an exemplary generation of complexity level 8). This algorithm is similar to previous models (Krambeck and Betenbaugh, 2005; Krambeck et al., 2009), and the reaction substrates and constraints were taken from the literature (Bennun et al., 2013). To account for Golgi compartmentalization, each glycan structure receives a localization label, depending on its localization in the cis-, medial or trans-Golgi. Transitions between these compartments are included in the network, as are secretion reactions that represent the exit of the glycan from the trans-Golgi. To initialize flux into the glycosylation network, the initial $\text{Man}_9\text{GlcNAc}_2$ glycan is assumed to be transported from the ER into the cis-Golgi at a normalized flux of 1.0. All remaining reactions are left open, i.e. no maximal or minimal flux bounds are imposed on them. In order to make this generic network specific, it is fit to a measured glycoprofile. For this, the secretion reactions for all glycans, except for the few appearing the profile, are shut off by setting their flux to 0. In contrast, the secretion flux for the glycans appearing in the profile is set to the experimentally observed frequencies (in cases where glycan structures have remained ambiguous, e.g. glycan #2 in Fig. 7A, the frequency was split evenly among all possible structures). Subsequently, flux variability analysis (Becker et al., 2007; Burgard et al., 2001; Gudmundsson and Thiele, 2010) is used to calculate the maximum and minimum possible flux through each reaction, given these constraints on the system’s glycan secretion. Reactions that cannot carry any flux under these circumstances are removed, leading to reduction of the network. That way, the generic network is tailored to contain only the minimal set of reactions required to generate the observed glycoprofile. The space of possible fluxes through this specific network (i.e. from the initial $\text{Man}_9\text{GlcNAc}_2$ to the secreted glycans) is investigated using Monte-Carlo sampling (Schellenberger and Palsson, 2009). For this, optGpSampler (Megchelenbrink et al., 2014) is run with four chains in parallel, saving 10,000 sample points with a step count of 100,000 between samples. After clipping the presumed burn-in period (50% of each sampling chain), the Gelman-Rubin-Diagnostic (Gelman and Rubin, 1992) is used to assess convergence of the sampling procedure with a standard deviation of the Gelman-Rubin statistic of 0.01 or less being

considered sufficiently indicative of convergence. A total of 5,000 samples of flux through the specific network is retained to represent the “wildtype”-specific generation of the observed glycoprofile. In order to predict alteration of this profile in response to an enzyme knock-down using a pure constraints-based method (Fig. 1C), all fluxes from this sample are adjusted following the MOMA (“Minimization of metabolic adjustment”) procedure (Segrè et al., 2002), thus, basically, by calculating a flux that minimally deviates from the wildtype flux while fulfilling the additional constraint of certain reactions being impeded due to the enzyme knock-out.

Construction of the Markov chain model

In the next step, the specific reaction network is transformed into a Markov chain, i.e. every glycan in the network is regarded as a state in a stochastic network that can transition to others with a certain transition probability (see (Rubino and Sericola, 2014) for methods and nomenclature from Markov chain theory used throughout this work). Secreted glycans are modeled as absorbing states, meaning states that transition to themselves with probability 1 (Fig. 2B). The Markov transition matrix Π describing this Markov chain is constructed by reading information from the reaction network: If a reaction is found that turns glycan 1 into glycan 2, a corresponding probability parameter is added in the transition matrix (Suppl. Fig. 2C). If a reaction is found that secretes a glycan, an absorbing state for that glycan is introduced. Then, a probability parameter is added to the transition into the absorbing state, and the state is made absorbing by adding a probability of 1 for its transition to itself (Suppl. Fig. 2D). The absorption probability describes the probability with which the chain reaches an absorbing state (starting in the initial $\text{Man}_9\text{GlcNAc}_2$) and is given by:

$$\pi_{\text{abs}} = \pi_{0|\text{T}} \cdot (\text{I} - \Pi_{\text{T}})^{-1} \cdot \Pi_{\text{TA}}$$

where Π_{T} and Π_{TA} are submatrices of Π obtained by partitioning the set of states into regular glycans (non-absorbing states) and secreted glycans (absorbing states) and ordering them according to Fig. 2D. $\pi_{0|\text{T}}$ is the start distribution vector, restricted to the regular glycans. I is the unity matrix. To avoid the high-dimensional computation required to find the transition probabilities in Π that best match the frequencies in the measured glycoprofile, the optimization problem in Fig. 2D is transformed into the equivalent flux-balance problem using the canonical transformation laid out in Suppl. Fig. 3A. The total of 5,000 fluxes obtained through sampling (see above) is back-transformed into probabilities (Fig. 3D) to obtain an array of 5,000 transition matrices each representing a Markov chain running into an absorbing profile that matches the submitted glycoprofile. At this point, a Markov chain model is obtained that describes the glycan reaction network at one particular glycosylation site on the protein. If separate experimental glycoprofiles are available for different glycosylation sites on the same protein, these can be used to fit individual Markov models. If, instead, the experimental glycoprofile is obtained by pooling multiple glycosylation sites (as it is the case for all datasets used in this study), the inferred transition probabilities for the model represent the average reaction rates from all glycosylation sites combined. In particular, if the experimental glycoprofile spans sites from a collection of glycoproteins

(such as the secretome), the model's transition probabilities describe reaction rates, as averaged over all these proteins.

Prediction of mutant glycan profiles

In order to simulate the knock-down of a glycosylation enzyme, the algorithm identifies all glycans in the network that transition to another glycan in a reaction that requires the specified enzyme. For each of these glycans, the transition probabilities have to be adjusted to simulate the knock-down situation. First, the probability of the affected reaction(s) is scaled down by a factor ε ($0 \leq \varepsilon < 1$) specified by the user, resulting in an adjusted transition probability $\pi' = \varepsilon \cdot \pi$, where $\pi' = 0$ would correspond to a full knock-out. Subsequently, the remaining transition probabilities have to be adjusted to maintain a proper probability structure with transition probabilities summing up to 1 in every row of the transition matrix. For this, the algorithm cycles through all affected glycans and considers the number of alternative transitions, i.e. transitions from that glycan that do not depend on the enzyme(s) being knocked down (Double, or multiple knock-downs can equally be simulated). If there are no alternative transitions, it is assumed that, as a consequence of the knock-down, the glycan transitions into the next compartment with a probability of $1 - \pi'$, or becomes secreted in case the knock-down took place in the trans Golgi.

If there are alternative transitions, their adjusted probabilities are calculated by solving an optimization problem using one of two approaches ("Equal Ratio" or "Minimum Adjustment") as chosen by the user. The "Equal Ratio" option models competition among enzymes for alternative reactions and relies on the assumption that the pairwise affinity ratio between two enzymes not affected by the knock-down should be maintained. That is, the function

$$\sum |\pi'_i / \pi'_j - \pi_i / \pi_j| \quad (i \neq j; i, j \neq 1)$$

is to be minimized under the constraints of $\pi'_1 = \varepsilon \cdot \pi_1$ and $\sum \pi'_i = 1$ with π'_1 representing the transition probability of the reaction to be knocked down (Fig. 4: step 3). Alternatively, the "Minimum Adjustment" option relies on the assumption of minimal change after a knock-down in order to maximize parsimony (in the same manner as the MOMA algorithm simulates gene knock-downs in flux-balance analysis (Segrè et al., 2002)). In particular, the function

$$\|\pi' - \pi\|$$

is minimized under the same constraints as above.

As explained above, a knock-down may cause a glycan to be passed on into a compartment where it is not located in the wildtype network. Thus, this glycan now represents an "alien" in the new compartment for which there are no reactions present to process it (Fig. 5A). For each alien glycan created this way, the script identifies its "closest match" among the wildtype glycans in that compartment. Based on the assumption that similar glycans will

likely be subject to similar reactions, the algorithm tries to copy reactions for the alien from the wildtype reaction set. This is done by finding reactions whose glycan substrate is closest to the alien glycan in morphology and chemical similarity. For this, all wildtype glycans from the compartment are identified that have the maximum number of branches in common with the alien. Among these, the algorithm compares each branch separately for its resemblance with the corresponding branch on the alien and calculates a measure representing the overall distance between the wildtype and the alien glycan by summing up the resemblance metric of all branches. Morphology is accounted for by comparing the branch lengths (i.e., the number of monosaccharide units) while chemical similarity is accounted for by comparing the monosaccharides at the branch tips since these represent the substrate for additional reactions. In total, the distance function is:

$$\sum_{\text{branches}} \alpha \cdot d_L + \beta \cdot d_C$$

with $d_L = |L_1 - L_2|$ being the difference in branch length and

$$d_C = \begin{cases} 0 & \text{if } G_1^* = G_2^* \\ 1 & \text{otherwise} \end{cases}$$

being the chemical discrepancy (G_i^* is the terminal monosaccharide on glycan i). Both parts of the distance function can be weighted with parameters chosen by the user ($\alpha=1, \beta=3$ used for default). If, as a consequence of a knock-down of a branching enzyme, the alien glycan misses a branch that is present on all wildtype glycans, the algorithm will give the lowest distance to the wildtype glycan whose corresponding branch equals the branch neighboring the missing one on the alien. In the theoretical case (not encountered in the validation datasets) in which the closest match turns out to be ambiguous, the algorithm proceeds with the first match in the list, and a warning message is displayed.

Finally, the algorithm considers all reactions taking place on the closest match and copies reactions to the alien glycan if the proper reaction substrate is present and the alien meets the structural reaction constraints for the particular enzyme (Table 1). If no such reaction can be copied, the alien is passed on to the next compartment or secreted, respectively. The probabilities for these copied reactions are calculated by the same rationale as in the knock-down, i.e. via either an “Equal Ratio” or a “Minimum Adjustment” procedure (see above), with the only difference that the probabilities π now represent the probabilities for the reactions on the closest match (Fig. 5: step 3; Suppl. Fig. 5: step 3).

After this processing has been carried out for all retained wildtype transition matrices, an array of transition matrices is obtained representing a Markov model for the knock-down situation. Running these Markov models, i.e. calculating their absorption probability, yields the predicted mutant glycan profile. Absorption probabilities are averaged over the entire array and given with their standard deviations.

Enzyme overexpression is simulated analogously to an enzyme knock-down. The user submits a wildtype Markov model together with an enzyme to be overexpressed as well as an overexpression parameter $E > 1$. The program first identifies all glycans in the network having reactions that depend on the enzyme to be overexpressed, and, subsequently, cycles through this glycan list to adjust transition probabilities. For this, the probability of the affected reaction is scaled up to

$$\pi' = \min(E \cdot \pi, 1)$$

and the probabilities for alternative reactions (if present) are calculated using either the “Equal Ratio” or the “Minimum Adjustment” approach as laid out before. Since transitions into new compartments are not created in this way, alien processing is not necessary.

Cell cultivation

CHO-S suspension cells (Life Technologies, Thermo Scientific, Rockford, IL) were grown as previously described (Grav et al., 2015). In summary, cells were grown in CD CHO medium supplemented with 8mM L-glutamine and 2 μ L/mL anti-clumping agent (Life Technologies). Cells were expanded in Corning vent cap shake flasks (Sigma-Aldrich, St.Louis, MO) in a humidified incubator at 120 rpm (25 mm orbit), 37°C and 5% CO₂. Viable cell densities were measured using the NucleoCounter NC-200 Cell Counter (ChemoMetec, Allerød, Denmark) and cells were passed into fresh medium every two to three days with seeding densities at 3–5 $\times$ 10⁵ cells/mL.

Construction of Fut8 knock-out cell lines using CRISPR/Cas9

The plasmids we used to generate the two Fut8 knock-out cell lines are expression vectors encoding GFP_2A_Cas9 (Grav et al., 2015) and sgRNAs targeting Fut8 (sgRNA targeting sgRNA2_F or sgRNA3_F, previously published in (Ronda et al., 2014)). However, an unintended mutation (Proline to Leucine mutation at amino acid position 28) in the amino acid sequence of Cas9 in the GFP_2A_Cas9 expression vector applied in this study was later detected. Apparently, this mutation does not abolish Cas9 activity and this plasmid was named GFP_2A_Cas9m. CHO-S cells were washed and seeded at 5–6 $\times$ 10⁵ cells/mL, without anti-clumping agents, one day prior to transfection. For each sample, 3 $\times$ 10⁶ cells at a density of 1 $\times$ 10⁶ cells/mL were transfected in a 6-well plate (BD Biosciences, San Jose, CA) with 1.9 μ g GFP_2A_Cas9m and 1.9 μ g sgRNA using FreeStyle™ MAX reagent together with OptiPRO SFM medium (Life Technologies) according to manufacturer's recommendations. Transfection with pmaxGFP® vector (Lonza, Basel, Switzerland) was applied as control for transfection efficiency. Two days after transfection, a FACSJazz (BD Biosciences) was used to single-sort transfected cells as previously published (Grav et al., 2015). Genomic DNA was extracted from monoclonal cell pools using QuickExtract DNA extraction solution (Epicentre, Illumina, Madison, WI) according to manufacturer's instructions. Targeted re-sequencing was performed using a MiSeq Benchtop Sequencer (Illumina, San Diego, CA) as previously described (Ronda et al., 2014). From the deep sequencing data, one Fut8 knockout cell line generated from each of the two sgRNAs were selected and expanded (FUT8_sgRNA2_F and FUT8_sgRNA3_F).

Secretome glycoprofiling

Glycoprofiling was performed as previously published (Grav et al., 2015). In short, exponentially growing cells were seeded at 1×10^6 cells/mL and supernatants harvested after 4 days by centrifugation. Supernatants were filtered and proteins contained in the sample were concentrated by centrifugation using Amicon Ultra columns (Merck Millipore, Merck KGaA, Darmstadt, Germany) with 3000 Da cutoff. N-glycans from retained proteins were released and fluorescently labeled with GlykoPrep Rapid N-Glycan kit (ProZyme Inc., Hayward, CA). Labeled N-glycans were analyzed by LC-MS on a Thermo Ultimate 3000 HPLC with fluorescence detector coupled on-line to a Thermo Velos Pro Iontrap MS. Glycan abundance was measured by integrating the areas under normalized fluorescence spectrum peaks with Xcalibur software (Thermo Fisher Scientific) giving the relative amount of the glycans. All annotated sugar structures are peaks with correct mass (Suppl. Fig. 6, Suppl. Table S2) and at least a signal to noise value of 20:1 as calculated with Xcalibur. In total, the analyzed secretome consisted of more than a thousand proteins.

Glycoprofiles obtained from the literature

The glycoprofiles of IgG1 (Figs. 7A, 8A) were taken from (Imai-Nishiya et al., 2007). Fucosyltransferase activity in a CHO/DG44 cell line had been knocked down using a combined Fut8/GMD (GDP-mannose-dehydrogenase) siRNA construct. The glycoprofiles of EPO (Figs. 1C, 7B, 8B) were taken from (Yang et al., 2015). For abolishment of GnTIV activity, a CHO-GS line had been mutagenized using zinc-finger-nucleases targeting mgat4A and mgat4B. For more details on cell line development and glycoprofiling, please refer to the cited articles.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We would like to express our gratitude to Mitsuo Satoh and Henrik Clausen and their laboratories for sharing their raw data used in this study. We also acknowledge NIH 1 R21 HD080682-01A1 and generous funding from the Novo Nordisk Foundation.

Abbreviations

GlcNAc	N-acetylglucosamine
GalT	β -1-4-galactosyltransferase
Fut8	α -1-6-fucosyltransferase 8
GnTIV	N-acetylglucosamine-transferase IV
CHO	Chinese hamster ovary
EPO	erythropoietin
mAb	monoclonal antibody

IgG immunoglobulin

References

- Aoki KF, Yamaguchi A, Okuno Y, Akutsu T, Ueda N, Kanehisa M, Mamitsuka H. Efficient tree-matching methods for accurate carbohydrate database queries. *Genome informatics*. 2003; 14:134–143. [PubMed: 15706528]
- Becker SA, Feist AM, Mo ML, Hannum G, Palsson BØ, Herrgard MJ. Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox. *Nat Protoc*. 2007; 2:727–738. [PubMed: 17406635]
- 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.doi: 10.1371/journal.pcbi.1002813 [PubMed: 23326219]
- Bordbar A, Monk JM, King Za, Palsson BO. Constraint-based models predict metabolic and associated cellular functions. *Nat Rev Genet*. 2014; 15:107–20. DOI: 10.1038/nrg3643 [PubMed: 24430943]
- Brühlman D, Jordan M, Hemberger J, Sauer M, Stettler M, Broly H. Tailoring recombinant protein quality by rational media design. *Biotechnol Prog*. 2015; 31:615–29. DOI: 10.1002/btpr.2089 [PubMed: 25864704]
- Burgard, aP; Vaidyaraman, S.; Maranas, CD. Minimal reaction sets for Escherichia coli metabolism under different growth requirements and uptake environments. *Biotechnol Prog*. 2001; 17:791–797. DOI: 10.1021/bp0100880 [PubMed: 11587566]
- Butler M, Spearman M. The choice of mammalian cell host and possibilities for glycosylation engineering. *Curr Opin Biotechnol*. 2014; 30C:107–112. DOI: 10.1016/j.copbio.2014.06.010 [PubMed: 25005678]
- Colley KJ. Golgi localization of glycosyltransferases: more questions than answers. *Glycobiology*. 1997; 7:1–13. DOI: 10.1093/Glycob/7.1.1-B [PubMed: 9061359]
- Cummings RD, Pierce JM. The challenge and promise of glycomics. *Chem Biol*. 2014; 21:1–15. DOI: 10.1016/j.chembiol.2013.12.010 [PubMed: 24439204]
- Da Silva A, Kronthaler U, Koppenburg V, Fink M, Meyer I, Papandrikopoulou A, Hofmann M, Stangler T, Visser J. Target-directed development and preclinical characterization of the proposed biosimilar rituximab GP2013. *Leuk Lymphoma*. 2014; 55:1–9. DOI: 10.3109/10428194.2013.843090 [PubMed: 23540343]
- Dalziel M, Crispin M, Scanlan CN, Zitzmann N, Dwek RA. Emerging principles for the therapeutic exploitation of glycosylation. *Science*. 2014; 343:1235681.doi: 10.1126/science.1235681 [PubMed: 24385630]
- Del Val IJ, 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. DOI: 10.1002/btpr.688 [PubMed: 21956887]
- Fan Y, Jimenez Del Val I, Müller C, Wagtberg Sen J, Rasmussen SK, Kontoravdi C, Weilguny D, Andersen MR. Amino acid and glucose metabolism in fed-batch CHO cell culture affects antibody production and glycosylation. *Biotechnol Bioeng*. 2014; 112:521–535. DOI: 10.1002/bit.25450 [PubMed: 25220616]
- Gelman A, Rubin DB. Inference from iterative simulation using multiple sequences. *Stat Sci*. 1992; 7:457–511.
- Gornik O, Pavi T, Lauc G. Alternative glycosylation modulates function of IgG and other proteins - Implications on evolution and disease. *Biochim Biophys Acta - Gen Subj*. 2012; 1820:1318–1326. DOI: 10.1016/j.bbagen.2011.12.004
- Grav LM, Lee JS, Gerling S, Kallehauge T, H Hansen A, Kol S, Lee GM, Pedersen L, Kildegaard H. One-step generation of triple knockout CHO cell lines using CRISPR Cas9 and fluorescent enrichment. *Biotechnol J*. 2015; 10:1446–56. DOI: 10.1002/biot.201500027 [PubMed: 25864574]
- Gudmundsson S, Thiele I. Computationally efficient flux variability analysis. *BMC Bioinformatics*. 2010; 11:489.doi: 10.1186/1471-2105-11-489 [PubMed: 20920235]

- Hayes, JM.; Cosgrave, EFJ.; Struwe, WB.; Wormald, M.; Davey, GP.; Jefferis, R.; Rudd, PM. Microbiology and Immunology. Springer International Publishing; 2014. Fc Receptors; p. 165-199.
- Hermeling S, C DJ, S H, J W. Structure-immunogenicity relationships of therapeutic proteins. Pharm Res. 2004; 21:897–903. [PubMed: 15212151]
- Hossler P. Protein Glycosylation Control in Mammalian Cell Culture: Past Precedents and Contemporary Prospects. Genomics and Systems Biology of Mammalian Cell Culture. 2012; : 187–219. DOI: 10.1007/10
- Hossler P, Khattak SF, Li ZJ. Optimal and consistent protein glycosylation in mammalian cell culture. Glycobiology. 2009; 19:936–49. DOI: 10.1093/glycob/cwp079 [PubMed: 19494347]
- Hossler P, McDermott S, Racicot C, Chumsae C, Raharimampionona H, Zhou Y, Ouellette D, Matuck J, Correia I, Fann J, Li J. Cell culture media supplementation of uncommonly used sugars sucrose and tagatose for the targeted shifting of protein glycosylation profiles of recombinant protein therapeutics. Biotechnol Prog. 2014; :20–24. DOI: 10.1002/btpr.1968 [PubMed: 25219977]
- Hossler P, Mulukutla BC, Hu WS. Systems analysis of N-glycan processing in mammalian cells. PLoS One. 2007; 2:e713.doi: 10.1371/journal.pone.0000713 [PubMed: 17684559]
- Imai-Nishiya H, Mori K, Inoue M, Wakitani M, Iida S, Shitara K, Satoh M. Double knockdown of alpha1,6-fucosyltransferase (FUT8) and GDP-mannose 4,6-dehydratase (GMD) in antibody-producing cells: a new strategy for generating fully non-fucosylated therapeutic antibodies with enhanced ADCC. BMC Biotechnol. 2007; 7:84.doi: 10.1186/1472-6750-7-84 [PubMed: 18047682]
- Jayapal KP, Lian W, Glod F, Sherman DH, Hu WS. Comparative genomic hybridizations reveal absence of large Streptomyces coelicolor genomic islands in Streptomyces lividans. BMC Genomics. 2007; 8:229.doi: 10.1186/1471-2164-8-229 [PubMed: 17623098]
- Jedrzejewski PM, del Val IJ, Constantinou A, Dell A, Haslam SM, Polizzi KM, Kontoravdi C. Towards controlling the glycoform: a model framework linking extracellular metabolites to antibody glycosylation. Int J Mol Sci. 2014; 15:4492–522. DOI: 10.3390/ijms15034492 [PubMed: 24637934]
- Kaveh O, Hengameh A, Johannes G, Legge MM, Jenö S, Hector B. Novel Dynamic Model to Predict the Glycosylation Pattern of Monoclonal Antibodies from Extracellular Cell Culture Conditions. 12th Symposium on Computer Applications in Biotechnology. 2013
- Kawano S, Hashimoto K, Miyama T, Goto S, Kanehisa M. Prediction of glycan structures from gene expression data based on glycosyltransferase reactions. Bioinformatics. 2005; 21:3976–82. DOI: 10.1093/bioinformatics/bti666 [PubMed: 16159923]
- Kildegaard HF, Baycin-Hizal D, Lewis NE, Betenbaugh MJ. The emerging CHO systems biology era: harnessing the 'omics revolution for biotechnology. Curr Opin Biotechnol. 2013; 24:1102–7. DOI: 10.1016/j.copbio.2013.02.007 [PubMed: 23523260]
- 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. DOI: 10.1093/glycob/cwp081 [PubMed: 19506293]
- Krambeck FJ, Betenbaugh MJ. A mathematical model of N-linked glycosylation. Biotechnol Bioeng. 2005; 92:711–28. DOI: 10.1002/bit.20645 [PubMed: 16247773]
- Lee HS, Jo S, Mukherjee S, Park SJ, Skolnick J, Lee J, Im W. GS-align for glycan structure alignment and similarity measurement. Bioinformatics. 2015; 31:2653–2659. DOI: 10.1093/bioinformatics/btv202 [PubMed: 25857669]
- Lewis NE, Nagarajan H, Palsson BO. Constraining the metabolic genotype-phenotype relationship using a phylogeny of in silico methods. Nat Rev Microbiol. 2012; 10:291–305. DOI: 10.1038/nrmicro2737 [PubMed: 22367118]
- Li H, d'Anjou M. Pharmacological significance of glycosylation in therapeutic proteins. Curr Opin Biotechnol. 2009; 20:678–684. DOI: 10.1016/j.copbio.2009.10.009 [PubMed: 19892545]
- Lin N, Mascarenhas J, Sealover NR, George HJ, Brooks J, Kayser KJ, Gau B, Yasa I, Azadi P, Archer-Hartmann S. Chinese hamster ovary (CHO) host cell engineering to increase sialylation of recombinant therapeutic proteins by modulating sialyltransferase expression. Biotechnol Prog. 2015; 31:334–346. DOI: 10.1002/btpr.2038 [PubMed: 25641927]

- Liu G, Neelamegham S. A computational framework for the automated construction of glycosylation reaction networks. *PLoS One*. 2014; 9doi: 10.1371/journal.pone.0100939
- McDonald AG, Hayes JM, Bezak T, Gluchowska Sa, Cosgrave EFJ, Struwe WB, Stroop CJM, Kok H, van de Laar T, Rudd PM, Tipton KF, Davey GP. Galactosyltransferase 4 is a major control point for glycan branching in N-linked glycosylation. *J Cell Sci*. 2014; 127:5014–26. DOI: 10.1242/jcs.151878 [PubMed: 25271059]
- Megchelenbrink W, Huynen M, Marchiori E. optGpSampler: an improved tool for uniformly sampling the solution-space of genome-scale metabolic networks. *PLoS One*. 2014; 9:e86587.doi: 10.1371/journal.pone.0086587 [PubMed: 24551039]
- Meuris L, Santens F, Elson G, Festjens N, Boone M, Dos Santos A, Devos S, Rousseau F, Plets E, Houthuys E, Malinge P, Magistrelli G, Cons L, Chatel L, Devreese B, Callewaert N. GlycoDelete engineering of mammalian cells simplifies N-glycosylation of recombinant proteins. *Nat Biotechnol*. 2014; 32:485–9. DOI: 10.1038/nbt.2885 [PubMed: 24752077]
- Moremen KW, Tiemeyer M, Nairn AV. Vertebrate protein glycosylation: diversity, synthesis and function. *Nat Rev Mol Cell Biol*. 2012; 13:448–62. DOI: 10.1038/nrm3383 [PubMed: 22722607]
- Neelamegham S, Liu G. Systems glycobiology: biochemical reaction networks regulating glycan structure and function. *Glycobiology*. 2011; 21:1541–53. DOI: 10.1093/glycob/cwr036 [PubMed: 21436236]
- Nilsson T, Hoe MH, Slusarewicz P, Rabouille C, Watson R, Hunte F, Watzele G, Berger EG, Warren G. Kin recognition between medial Golgi enzymes in HeLa cells. *EMBO J*. 1994; 13:562–574. [PubMed: 8313901]
- Nilsson T, Pypaert M, Hoe MH, Slusarewicz P, Berger EG, Warren G. Overlapping distribution of two glycosyltransferases in the Golgi apparatus of HeLa cells. *J Cell Biol*. 1993a; 120:5–13. DOI: 10.1083/jcb.120.1.5 [PubMed: 8416995]
- Nilsson T, Slusarewicz P, Hoe MH, Warren G. Kin recognition. A model for the retention of Golgi enzymes. *FEBS Lett*. 1993b; 330:1–4. DOI: 10.1016/0014-5793(93)80906-B [PubMed: 8370450]
- Niwa R, Satoh M. The Current Status and Prospects of Antibody Engineering for Therapeutic Use: Focus on Glycoengineering Technology. *J Pharm Sci*. 2015; :1–12. DOI: 10.1002/jps.24316 [PubMed: 25421602]
- Palsson, BØ. *Systems Biology: Constraints-based reconstruction and analysis*. 2nd. Cambridge University Press; 2015.
- Rabouille C, Hui N, Hunte F, Kieckbusch R, Berger EG, Warren G, Nilsson T. Mapping the distribution of Golgi enzymes involved in the construction of complex oligosaccharides. *J Cell Sci*. 1995; 108:1617–1627. [PubMed: 7615680]
- Ronda C, Pedersen LE, Hansen HG, Kallehauge TB, Betenbaugh MJ, Nielsen AT, Kildegaard HF. Accelerating genome editing in CHO cells using CRISPR Cas9 and CRISPy, a web-based target finding tool. *Biotechnol Bioeng*. 2014; 111:1604–1616. DOI: 10.1002/bit.25233 [PubMed: 24827782]
- Rubino, G.; Sericola, B. *Markov Chains and Dependability Theory*. Cambridge University Press; 2014. doi:<http://dx.doi.org/10.1017/CBO9781139051705>
- Schellenberger J, Palsson B. Use of randomized sampling for analysis of metabolic networks. *J Biol Chem*. 2009; 284:5457–5461. DOI: 10.1074/jbc.R800048200 [PubMed: 18940807]
- Schellenberger J, Que R, Fleming RMT, Thiele I, Orth JD, Feist AM, Zielinski DC, Bordbar A, Lewis NE, Rahmanian S, Kang J, Hyduke DR, Palsson BØ. Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0. *Nat Protoc*. 2011; 6:1290–1307. DOI: 10.1038/nprot.2011.308 [PubMed: 21886097]
- Segrè D, Vitkup D, Church GM. Analysis of optimality in natural and perturbed metabolic networks. *Proc Natl Acad Sci U S A*. 2002; 99:15112–15117. DOI: 10.1073/pnas.232349399 [PubMed: 12415116]
- Senger RS, Karim MN. Prediction of N-linked glycan branching patterns using artificial neural networks. *Math Biosci*. 2008; 211:89–104. DOI: 10.1016/j.mbs.2007.10.005 [PubMed: 18054050]
- Senger RS, Karim MN. Variable site-occupancy classification of N-linked glycosylation using artificial neural networks. *Biotechnol Prog*. 2005; 21:1653–62. DOI: 10.1021/bp0502375 [PubMed: 16321048]

- Shelikoff M, Sinskey aJ, Stephanopoulos G. A modeling framework for the study of protein glycosylation. *Biotechnol Bioeng.* 1996; 50:73–90. DOI: 10.1002/(SICI)1097-0290(19960405)50:1 [PubMed: 18626901]
- Shi HH, Goudar CT. Recent advances in the understanding of biological implications and modulation methodologies of monoclonal antibody N-linked high mannose glycans. *Biotechnol Bioeng.* 2014; 111:1907–1919. DOI: 10.1002/bit.25318 [PubMed: 24975601]
- Shields RL, Lai J, Keck R, O'Connell LY, Hong K, Gloria Meng Y, Weikert SHa, Presta LG. Lack of fucose on human IgG1 N-linked oligosaccharide improves binding to human Fc γ RIII and antibody-dependent cellular toxicity. *J Biol Chem.* 2002; 277:26733–26740. DOI: 10.1074/jbc.M202069200 [PubMed: 11986321]
- Shriver Z, Raguram S, Sasisekharan R. Glycomics: a pathway to a class of new and improved therapeutics. *Nat Rev Drug Discov.* 2004; 3:863–873. DOI: 10.1038/nrd1521 [PubMed: 15459677]
- Shubhakar A, Reiding KR, Gardner Ra, Spencer DIR, Fernandes DL, Wuhler M. High-Throughput Analysis and Automation for Glycomics Studies. *Chromatographia.* 2014;1–13. doi:10.1007/s10337-014-2803-9.
- Sørensen AL, Rumjantseva V, Nayeb-Hashemi S, Clausen H, Hartwig JH, Wandall HH, Hoffmeister KM. Role of sialic acid for platelet life span: Exposure of β -galactose results in the rapid clearance of platelets from the circulation by asialoglycoprotein receptor-expressing liver macrophages and hepatocytes. *Blood.* 2009; 114:1645–1654. DOI: 10.1182/blood-2009-01-199414 [PubMed: 19520807]
- Sou SN, Sellick C, Lee K, Mason A, Kyriakopoulos S, Polizzi KM, Kontoravdi C. How does mild hypothermia affect monoclonal antibody glycosylation? *Biotechnol Bioeng.* 2014; 112:1165–76. DOI: 10.1002/bit.25524 [PubMed: 25545631]
- Spahn PN, Lewis NE. Systems glycobiology for glycoengineering. *Curr Opin Biotechnol.* 2014; 30C: 218–224. DOI: 10.1016/j.copbio.2014.08.004 [PubMed: 25202878]
- St Amand MM, Radhakrishnan D, Robinson AS, Ogunnaike Ba. Identification of manipulated variables for a glycosylation control strategy. *Biotechnol Bioeng.* 2014; 9999:1–14. DOI: 10.1002/bit.25251
- Surve T, Gadgil M. Manganese increases high mannose glycoform on monoclonal antibody expressed in CHO when glucose is absent or limiting: Implications for use of alternate sugars. *Biotechnol Prog.* 2015; 31:460–467. DOI: 10.1002/btpr.2029 [PubMed: 25504679]
- Thaysen-Andersen M, Packer NH. Site-specific glycoproteomics confirms that protein structure dictates formation of N-glycan type, core fucosylation and branching. *Glycobiology.* 2012; 22:1440–52. DOI: 10.1093/glycob/cws110 [PubMed: 22798492]
- Tsuruta LR, Lopes dos Santos M, Moro AM. Effects of copper on cho cells: Insights from gene expression analyses. *Biotechnol Prog.* 2015; :1–34. DOI: 10.1002/btpr [PubMed: 25311397]
- Tu L, Banfield DK. Localization of Golgi-resident glycosyltransferases. *Cell Mol Life Sci.* 2010; 67:29–41. DOI: 10.1007/s00018-009-0126-z [PubMed: 19727557]
- Van Beers MMC, Bardor M. Minimizing immunogenicity of biopharmaceuticals by controlling critical quality attributes of proteins. *Biotechnol J.* 2012; 7:1473–1484. DOI: 10.1002/biot.201200065 [PubMed: 23027660]
- Varki, A.; Cummings, R.; Esko, J.; Al, E. *Essentials of glycobiology.* 2nd. Cold Spring Harbor Laboratory Press; Cold Spring Harbor (NY): 2009.
- Wu M, Chan C. Prediction of therapeutic microRNA based on the human metabolic network. *Bioinformatics.* 2014; :1–9. DOI: 10.1093/bioinformatics/btt751
- Xu X, Nagarajan H, Lewis NE, Pan S, Cai Z, Liu X, Chen W, Xie M, Wang W, Hammond S, Andersen MR, Neff N, Passarelli B, Koh W, Fan HC, Wang J, Gui Y, Lee KH, Betenbaugh MJ, Quake SR, Famili I, Palsson BO, Wang J. The genomic sequence of the Chinese hamster ovary (CHO)-K1 cell line. *Nat Biotechnol.* 2011; 29:735–41. DOI: 10.1038/nbt.1932 [PubMed: 21804562]
- Yang Z, Wang S, Halim A, Schulz MA, Frodin M, Rahman SH, Vester-Christensen MB, Behrens C, Kristensen C, Vakhrushev SY, Bennett EP, Wandall HH, Clausen H. Engineered CHO cells for production of diverse, homogeneous glycoproteins. *Nat Biotechnol.* 2015; 33:842–844. DOI: 10.1038/nbt.3280 [PubMed: 26192319]

Fig. 1. Generation of the N-glycosylation reaction network

(A): Starting in the initial Man₉GlcNAc₂ glycan, an iterative application of the reaction rules in Table 1 generates the N-glycosylation reaction network. The glycan complexity level refers to the set of glycans generated in a specific number of iterations. Structure refers to individual glycans ignoring their compartmental localization. Reactions comprise enzymatic reactions on glycans, glycan transitions to different compartments, and glycan secretion.

(B): Compartmentalization of the Golgi is implemented by giving glycans labels for cis-, medial or trans-Golgi, depending on their localization. Reaction flux (grey) through the network can be canonically transformed into transition probabilities (blue) by taking flux ratios, thereby turning the reaction network into a stochastic network.

(C): Experimental glycoprofiles of erythropoietin (EPO) taken from both wildtype (top) and an mgat4A/4B mutant cell line (middle) (obtained from (Yang et al., 2015)). Knockout prediction using a constraints-based method (“Minimization of metabolic adjustment”, MOMA) shows an excess of predicted glycans (#9 – #17, bottom) that do not appear in the experimental profile. Only major frequencies are shown.

Fig. 2. Markov chain model of glycosylation

(A): The transition matrix Π comprises the probabilities of transitioning from each glycan (rows) to any other glycan (columns) in one reaction step. Glycan ordering and numbering is arbitrary. (B): The start distribution π_0 of the chain mirrors the initialization of N-glycosylation and is, thus, concentrated on the M₉GlcNAc₂ glycan. Glycans that are to be secreted can transition into an artificial absorbing state (framed glycan) which will transition to itself with probability 1 in every step, leading to absorption of the chain. (C): To model the reaction path from the initial glycan attachment to secretion of the glycoprotein, the

Markov chain starts in the initial $M_9\text{GlcNAc}_2$ glycan, takes a random walk through the reaction network following probabilities given by the transition Matrix Π and eventually gets absorbed in one of the absorbing states. The probability of getting absorbed in a particular state, i.e. the probability to secrete a particular glycan, is given by π_{abs} . **(D):** In order to infer the unknown transition probabilities in Π , the model is fit to the frequencies in a measured glycoprofile obtained under standard conditions (e.g. a wildtype cell line). The frequencies in the predicted glycoprofile are given by π_{abs} which can be calculated using π_0 and Π (see Materials & Methods). Numerical optimization can be used to find the transition probabilities (entries in Π) that minimize the distance between the predicted profile from the Markov chain (π_{abs}) and the observed profile (w).

Fig. 3. Model solving workflow

(A): To infer the unknown transition probabilities, an optimization problem is solved, involving up to several hundred variables in case of a complex glycoprofile. (B): In order to improve computational scalability, the problem is transformed into an equivalent flux-balance problem using the canonical transformation laid out in Suppl. Fig. 3A. (C): Due to the underdetermined nature of the flux-balance problem, a large space of alternative steady-state solution is generated. A representative sample of solutions from this space is retrieved using Monte-Carlo sampling. (D): Every flux solution is back-transformed into a Markov transition matrix via the canonical transformation (Suppl. Fig. 3A). Each of these transition matrices describes a Markov chain that runs into the glycoprofile in (A).

1. Identify reactions affected by enzyme knock-down

Example:

2. Scale down affected transition probabilities

Example:

3. Adjust alternative transition probabilities

Example:

4. Add transport/secretion reaction in case of absence of alternative reactions

Example:

Fig. 4. Modeling enzyme knock-downs

Step 1: All reactions in the wildtype glycosylation reaction catalyzed by the enzyme(s) to be knocked down (e.g. the fucosyltransferase) are identified. Step 2: The transition probabilities of these reactions are then scaled down. To model a complete knock-out, they are set to 0 as illustrated here. Step 3: The transition probabilities of alternative reactions, i.e. those taking the same glycan substrate as the one being knocked down (black arrows), need to be adjusted to maintain a probability sum of 1 for each glycan in the network. For details on the adjustment calculations see Materials & Methods. Step 4: If no alternative reactions exist, it

is assumed that the glycan will leave the corresponding Golgi compartment unmodified and transition into the next compartment or becomes secreted (if already in the trans Golgi). For clarity, localization labels on glycans have been omitted in the remaining cartoons. All other transition probabilities remain identical to the wildtype network.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Fig. 5. Processing of “alien” glycans after an enzyme knock-out

(A): After Fut8 is knocked out, the non-fucosylated glycan is passed on into the trans Golgi where it lacks any further reactions since all trans-Golgi glycans in the wildtype network are fucosylated. As this non-fucosylated “alien” glycan closely resembles its fucosylated variant (the “closest match”), it will likely undergo analogous processing (galactosylation in this case), so the reactions on the closest match are assumed to occur likewise on the alien (dotted arrows). (B): Step 1: To build the reactions for the alien glycans, the closest match to an alien glycan is chosen from the pool of wildtype glycans available in the same compartment on the basis of chemical and structural similarity (see Materials & Methods). Step 2: The analogous reactions are linked to the alien glycan (in this case: galactosylation of both branches). Step 3: Corresponding transition probabilities are assumed to be identical to the analogs in the wildtype. Slight adjustments are required in more complicated cases where certain reactions on the closest match cannot be copied as they lack a reaction substrate in the alien (Suppl. Fig. 5). Since, typically, analogous reactions on the alien will themselves create new alien glycans, not present in the wildtype network (in this case

galactosylated glycans lacking the core fucose), this procedure is repeated until all aliens have been processed. Glycan localization labels are omitted for clarity.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Fig. 6. Modeling enzyme overexpression

(A): In analogy to the knock-out approach, overexpression is modeled by scaling up a certain transition probability by a factor E (>1), specified by the user (For details see Materials & Methods). (B): Simulation shows how the glycoprofile on EPO is gradually shifted to a fully sialylated glycan (#3) through upscaling of sialic acid transferase (SiaT) dependent reactions. As a consequence, the competing reaction on the same substrate, namely chain elongation through lactosamine (glycan #4), is down-regulated.

Fig. 7. Prediction of mutant glycoprofiles

Knock-out simulations ($\epsilon = 0$) using the Markov model were run either with the "Equal Ratio" variant (red bars) or the "Minimum Adjustment" variant (grey bars) (see Materials & Methods). Bars represent mean $\pm$ s.d. Glycans with 0% frequency are shown transparent. Experimental data is shown in black bars. **(A):** IgG1 glycoprofile from a wildtype and a Fut8 knock-out CHO/DG44 cell line with corresponding model predictions (Experimental data from (Imai-Nishiya et al., 2007)). **(B):** EPO glycoprofile from a wildtype and GnTIV knock-out CHO-GS line with corresponding predictions (Experimental data from (Yang et al., 2015)). **(C):** Whole secretome glycoprofiles from a wildtype and Fut8 knock-out CHO-S suspension culture with corresponding predictions. The full spectrum is shown in Suppl. Fig. 6. Mean and standard deviation data shown in A–C is documented in Suppl. Table S1.

Fig. 8. Impact of Golgi compartmentalization on glycoprofile prediction

(A): Left: Experimental Fut8 knock-out profile from Fig. 7A (black bars) along with knock-out simulations (“Equal Ratio” variant), run either with the Golgi assumed to be compartmentalized (like in Fig. 7A) (grey bars) or not compartmentalized (red bars). Right: A section of the network topologies implied by a compartmentalized and non-compartmentalized Golgi model, respectively. In case of compartmentalization, the knock-out of the fucosylation reaction leads to passage of the non-fucosylated glycan (#4) into the trans-Golgi where it is processed according to the reactions present on its closest match (#1). In case of no compartmentalization, the implied parallel action of Fut8 and GalT leads to a topology that would only secrete a single, fully galactosylated glycan (glycan #6) when fucosylation is knocked out. (B): Left: Experimental GnTIV knock-out profile from Fig. 7B (black bars) along with knock-out simulations (“Equal Ratio” variant), run either with the Golgi assumed to be compartmentalized (like in Fig. 7B) (grey bars) or not compartmentalized (red bars). Right: In case of compartmentalization, the knock-out of the GnTIV-dependent branching in the medial Golgi leads to passage of the non-bifurcated glycan (#11) into the trans-Golgi where it is processed according to the reactions present on

its closest match (#12), eventually leading to four tri-antennary structures (#5 – #8). In case of no compartmentalization, the enzymes GalT and GnTIV act in parallel. As a consequence of the implied network topology, only two rather than four glycans (#9, #10) would be secreted once GnTIV is knocked out.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 1

Enzymes and reaction rules implemented in the model. Due to the explicit inclusion of enzyme and substrate localization, constraints may read slightly different from previous formulations (Bennun et al., 2013), although the rules are identical. A graphical representation of the first eight iterations of the generating algorithm based on these rules is shown in Suppl. Fig. 1.

Enzyme	Substrate	Product	Constraint	Localization
ManI	(Ma2Ma	(Ma	–	cis
GnTI	(Ma3(Ma3(Ma6)Ma6)Mb4	(GNb4Ma3(Ma3(Ma6)Ma6)Mb4	–	cis
ManII	(Ma3(Ma6)Ma6	(Ma6Ma6	(GNb2 Ma3	medial
ManII	(Ma6Ma6	(Ma6	(GNb2 Ma3	medial
GnTII	(Gnb2 Ma3(Ma6)Mb4	(GNb2 Ma3(GNb2Ma6)Mb4	–	medial
a6FucT	GNb4GN	GNb4(Fa6)GN	GNb2 Ma3	medial
GnTIV	(GNb2Ma3	(GNb2(GNb4)Ma3	–	medial
GnTV	(GNb2Ma6	(GNb2(GNb6)Ma6	–	trans
b4GalT	(GN	(Ab4GN	–	trans
iGnT	(Ab4GN	(GNb3Ab4GN	~*...Ma3	trans
a3SiaT	(Ab4GN	(NNa3Ab4GN	–	trans

... = Continuation, i.e. any string (possibly empty) with all parentheses matched

| = Possible branch point, i.e. empty string or (...)