iGFA: Improved Glycosylation Flux Analysis

Shriramprasad Venkatesan*, Sriram Neelamegham**
Rudiyanto Gunawan***

University at Buffalo-SUNY, Buffalo, NY 14260 USA (e-mail: shriramp@buffalo.edu)
*** University at Buffalo-SUNY, Buffalo, NY 14260 USA (e-mail: neel@buffalo.edu)*
**** University at Buffalo-SUNY, Buffalo, NY 14260 USA (e-mail: rgunawan@buffalo.edu)*

Abstract: In biopharmaceutical manufacturing of monoclonal antibodies (mAbs), asparagine (N)-linked glycosylation profile of these proteins is a critical quality attribute. We introduce improved Glycosylation Flux Analysis (iGFA), enhancing our previous GFA by: (1) reformulating constraint-based modeling using enzymatic kinetics to obtain biologically interpretable factors; and (2) implementing the analysis using Python's Pyomo modeling language, which not only reduces computational costs significantly compared to the MATLAB-based GFA, but also makes the iGFA an open-source package. When applied to data from Chinese Hamster Ovary (CHO) cell culture production of mAb under varying pH conditions, the analysis revealed both common and distinct dynamic trends across different pH. We identified galactosylation as the most impacted glycosylation processing by pH. Further, the estimated enzyme-related factors correlated more strongly with gene expression levels than with nucleotide sugar availability, suggesting that glycosylation regulation is predominantly controlled at the transcriptional and/or translational level. Overall, the iGFA is a powerful tool for analyzing glycosylation dynamics in mAb production.

Keywords: Flux analysis, glycosylation, monoclonal antibodies, Chinese hamster ovary, modeling.

1. INTRODUCTION

Recombinant monoclonal antibodies (mAbs), such as immunoglobulin G (IgG), represent an important class of molecules for the biopharmaceutical industry. The market value of mAbs drugs in 2023 was estimated to be USD 300 billion, with projections indicating growth of >11% annually (Grand View Research [GVR], 2023). Mammalian cell culture systems using murine hybridoma or Chinese hamster ovary (CHO) cells are the most common cell platforms for producing these therapeutic antibodies to ensure human biocompatibility (Murakami, Matsumoto and Kanamori, 2019). Among the critical quality attributes in the production of therapeutic mAbs are the N-linked glycans—oligosaccharides attached to the asparagine residues in the fragment crystallizable region of these proteins. N-linked glycans influence several important properties of mAb drugs including their bioavailability, pharmacokinetics, cytotoxicity, and efficacy (Chen *et al.*, 2022). However, since glycosylation is a non-template driven process, there exists intrinsic heterogeneity in the glycan structures that are formed on mAbs (Wang, Zhu and Lu, 2020). Achieving the target distribution of glycans is of great importance in the biomanufacturing of recombinant mAbs.

The protein glycosylation process starts in the endoplasmic reticulum (ER) and proceeds through the Golgi apparatus (Neelamegham and Liu, 2011). The process occurs via an intricate enzymatic reaction network that produces an assortment of glycans. Protein glycoforms differ from each other in their attached glycan structures. The distribution of these glycoforms is often referred to as the glycosylation profile. In the cell culture production of mAb, the N-linked glycosylation profile is affected by various culture process parameters such as pH, osmolality, and temperature. (Alhuthali, Kotidis and Kontoravdi, 2021; Kiehl *et al.*, 2011; Pan *et al.*, 2017). However, the mechanism by which these parameters modify N-linked glycans is not well understood. Understanding this mechanism is crucial for optimizing bioprocess conditions to achieve target glycan profiles. Recent studies have focused on developing mathematical models that simulate glycosylation pathways, providing insights into how changes in cell culture parameters affect glycan synthesis.

Several approaches with varying complexity have been adopted for the mathematical modeling of glycosylation in IgG production. These approaches range from detailed kinetic modeling (Jimenez del Val, Nagy and Kontoravdi, 2011; Villiger *et al.*, 2016), to Bayesian modelling (Zhang *et al.*, 2021), to constraint-based modeling (CBM) (Aggarwal *et al.*, 2021; Hutter *et al.*, 2017; Hutter *et al.*, 2018; Spahn *et al.*, 2016). Notably, constraint-based models rely only on the reaction network stoichiometry to evaluate intracellular reaction rates (fluxes), bypassing the need for kinetic parameters that are often unknown or difficult to determine. Our previously published Glycosylation Flux Analysis (GFA) method employs CBM to estimate dynamic glycosylation reaction rates from cellular secretion rates of IgG glycoforms (Hutter *et al.*, 2017). The basic assumption in formulating GFA is that rates or fluxes vary with time because of two factors: dynamic changes in enzyme activity and in cell specific productivity. An extension of GFA, called compartmental GFA (cGFA), incorporates the segregation of the Golgi into cis, medial, and trans compartments (Aggarwal *et al.*, 2021). While these methods have advanced the analysis of glycosylation processes during cell cultivation, their formulation is based on *ad hoc* assumptions about flux

dynamics, resulting in estimated parameters that lack biological interpretability.

In this work, we present improved GFA (iGFA), which enhances the GFA in two important ways. Firstly, a rederivation of the CBM formulation using enzymatic kinetics enhances the interpretability of the estimated dynamic factors. Specifically, these factors are related to enzymatic processing capacity, intracellular mAb glycoform concentrations, and residence time in the Golgi. Secondly, in contrast to the MATLAB implementation of GFA and cGFA, the iGFA employs Python open-source Pyomo modeling language (Hart, Watson and Woodruff, 2011). The iGFA significantly outperforms the GFA in the computational cost for analysis. Here, we applied the iGFA to analyze N-linked glycosylation in CHO cell culture production of IgG and the effect of media pH on this process (Lee *et al.*, 2021).

2. METHODS

2.1 Experimental Data Processing

The iGFA uses time-series values of glycoform secretion fluxes as input. These fluxes are estimated from measurements of viable cell density (VCD), antibody titer $T(t)$, cell specific productivity $Q_{prod}(t)$, and glycoform fractions $f_i(t)$ by writing a mass balance for the mAb glycoform in the cell culture media. For brevity, we refer readers to the original GFA publication for the derivation of this balance equation (Hutter *et al.*, 2017). The following equation gives the cell-specific secretion rate of the i-th glycoform, $v_{E,i}^M(t)$:

$$v_{E,i}^M(t) = \frac{1}{VCD(t)} \frac{d(f_i(t) T(t))}{dt} + f_i(t) \left(Q_{prod}(t) - \frac{dT(t)}{dt} \frac{1}{VCD(t)} \right) \quad (1)$$

To reduce the impact of noise when taking time derivatives, the measurement data were smoothed using the curve_fit function from the package Scipy. The numerical time differentiation was performed using the function diff from the package Sympy.

2.2 Improved GFA (iGFA) formulation

The iGFA, like the original GFA, employs a CBM of the glycosylation network. By assuming (pseudo-)steady state condition, the glycosylation fluxes are written as:

$$\boldsymbol{S}\boldsymbol{v}_I(t) - \boldsymbol{v}_E(t) = \boldsymbol{0} \quad (2)$$

where $\boldsymbol{S}$ is the $m \times r$ stoichiometric matrix for the glycosylation reactions with m glycoforms and r glycosylation reactions, $\boldsymbol{v}_I(t)$ is the vector of (unknown) intracellular glycosylation fluxes, and $\boldsymbol{v}_E(t)$ is the vector of secretion fluxes for the glycoforms. Since the number of unknown reaction fluxes r exceeds the number of (detectable) glycoforms m, the estimation of intracellular fluxes $\boldsymbol{v}_I(t)$ given the values of $\boldsymbol{v}_E(t)$ is underdetermined.

The original GFA addressed the underdetermined issue by assuming that the intracellular glycosylation fluxes vary with time according to two multiplicative factors that are related to the specific enzyme associated with the reaction and the cell-specific productivity. Instead of using this *ad hoc* formulation,

the iGFA formulation begins with the Michaelis-Menten kinetics:

$$v_{I,j}(t) = \frac{v_{max}(t)c_{I,i}(t)}{K_m + c_{I,i}(t)} V(t) \quad (3)$$

where $c_{I,i}$ denotes the intracellular concentration of the i-th glycoform, v_{max} denotes the maximum reaction rate that can be achieved at saturating glycoform substrate concentration (*i.e.*, $c_{I,i} \gg K_m$), K_m is the Michaelis constant, and V is the volume of the Golgi. The variable v_{max} is related to enzyme concentration (expression), turnover number k_{cat} of the enzyme, and nucleotide sugar co-factor. Here we consider a scenario where v_{max}, $c_{I,i}$, and V depend on time t—that is, enzyme expression, enzyme activity, nucleotide sugar level, glycoform concentration, and Golgi volume may change during cell cultivation. In the scenario of $c_{I,i} \ll K_m$ (*i.e.*, when the glycoform substrate is limiting), the Michaelis-Menten kinetics reduces to a linear relationship:

$$v_{I,j}(t) = \frac{v_{max}(t)c_{I,i}(t)}{K_m} V(t) \quad (4)$$

Therefore, the intracellular fluxes at a given timepoint t can be evaluated from the values from a reference timepoint t_{ref} according to:

$$v_{I,j}(t) = \frac{v_{max}(t)}{v_{max}(t_{ref})} \frac{c_{I,i}(t)}{c_{I,i}(t_{ref})} \frac{V(t)}{V(t_{ref})} v_{I,j}(t_{ref}) \quad (5)$$

The secretion flux of the i-th glycoform can further be written in terms of the intracellular concentrations $c_{I,i}$ and the volumetric flow rate through the Golgi q as follows:

$$v_{E,i}(t) = q(t)C_{I,i}(t) \quad (6)$$

By substituting Eq. (6) to Eq. (5), we obtain:

$$v_{I,j}(t) = \frac{v_{max}(t)}{v_{max}(t_{ref})} \frac{v_{E,i}(t)}{v_{E,i}(t_{ref})} \frac{\frac{V(t)}{q(t)}}{\frac{V(t_{ref})}{q(t_{ref})}} v_{I,j}^{ref} \quad (7)$$

Therefore, the time-dependent intracellular glycosylation fluxes can be evaluated using the following equation:

$$v_{I,j}(t) = \alpha(t)\beta(t)\gamma(t)v_{I,j}(t_{ref}) \quad (8)$$

where

$$\alpha(t) = \frac{v_{max}(t)}{v_{max}(t_{ref})} \quad (9)$$

$$\beta(t) = \frac{v_{E,i}(t)}{v_{E,i}(t_{ref})} \quad (10)$$

$$\gamma(t) = \frac{\frac{V(t)}{q(t)}}{\frac{V(t_{ref})}{q(t_{ref})}} = \frac{t_r(t)}{t_r(t_{ref})} \quad (11)$$

where t_r denote the residence time of IgG in the Golgi.

In the formulation above, α captures the dynamic changes in enzymatic processing capacity (*i.e.* enzyme expression and activity) and nucleotide sugar availability. Glycosylation

reactions that are catalysed by the same enzyme share the same α factor. Meanwhile, the factor β captures the dynamic changes in the glycoform concentrations, and thus all reactions that have the same glycoform substrate share the same β. Lastly, the factor γ captures the dynamic changes in protein residence time in the Golgi. All glycoforms share the same γ.

The evaluation of intracellular glycosylation fluxes in the iGFA reduces to obtaining the intracellular glycosylation fluxes at the reference time of choice and the time-dependent factors: α—one for every enzyme, β—one for every reactant glycoform in the network, and γ. These unknowns are estimated using the following constrained optimization:

$$\min_{\alpha_J(t),\beta_i(t),\gamma(t),v_{I,j}^{ref}} \sum_{t \in K} \left[\left(Q_{prod}(t) - \sum_{i=1}^{m} v_{E,i}(t) \right)^2 + \right.$$
$$\left. \sum_{j=1}^{m} \left(v_{E,j}^M(t) - v_{E,j}(t) \right)^2 + \sum_{k \in Subs} \left(\beta_k(t) - \frac{v_{E,k}(t)}{v_{E,k}(t_{ref})} \right)^2 \right] \quad (12)$$

such that

$$\sum_{j=1}^{r} S_{i,j} v_{I,j}(t) - v_{E,i}(t) = 0 \quad (13)$$

$$\sum_{t \in K} \sum_{J \in Enz} \left(\alpha_J(t) - 1 \right)^2 + (\gamma(t) - 1)^2 \leq \lambda \quad (14)$$

$$0 \leq \alpha_J \leq \alpha_{J_{max}} \quad (15)$$

$$0 \leq \beta_i \leq \beta_{i_{max}} \quad (16)$$

$$0 \leq \gamma \leq \gamma_{max} \quad (17)$$

where K denotes the set of time points, $Subs$ denoting the set of glycoforms that act as substrates to glycosylation reactions, and Enz denotes the set of enzymes. The loss function includes deviations of (i) the total sum of secretion fluxes from the cell specific productivity, (ii) glycoform secretion fluxes calculated using Eq. (2) from those evaluated using measurements as described in Eq. (1), and (iii) the factor β from the ratio of secretion fluxes.

The regression problem in Eq. (12) contains $r + N_{Enz}(N_K - 1) + (N_K - 1) + N_{Subs}(N_K - 1)$ unknowns, where N_{Enz}, N_{Subs} and N_K are the number of enzymes, substrate glycoforms, and time points (i.e. the cardinality of the sets Enz, $Subs$ and K), respectively. In comparison to the GFA, the iGFA has additional unknowns that cause the estimation to become underdetermined — having more unknowns than data points. To avoid overfitting, $\alpha_J(t)$ and $\gamma(t)$ are regularized to stay near 1 using the regularization in Eq. (14). Here, L2-norm regularization aligns with biological assumption that changes associated with enzyme activities and residence times remain small. In this formulation, λ serves as a user-defined hyperparameter that dictates the strictness of the L2-norm regularization; a smaller λ value enforces $\alpha_J(t)$ and $\gamma(t)$ to be closer to 1, albeit at the potential cost of model fit.

The iGFA is implemented in the Pyomo modeling language (version 6.7.3) (Hart, Watson and Woodruff, 2011). Pyomo is a Python-based, open-source structured modelling language

for formulating, solving, and analyzing optimization problems. The optimization associated with the iGFA as described above is solved using the Interior Point Optimizer (IPOPT) (version 3.14.16), an open-source solver for solving large-scale non-linear constrained optimization problems (Wachter and Biegler, 2006).

2.3 N-linked glycosylation reaction network curation

For the case study (Lee *et al.*, 2021), N-linked glycosylation reaction network in CHO cells was manually curated. The initial reaction network was obtained by including glycoforms that were detected in the experiment and those that may act as intermediate molecules. Subsequently, the reaction network was pruned to produce a minimal number of reactions, keeping the glycoforms in the network connected. Fig. 1 depicts the curated network, comprising 13 glycoforms with 15 intracellular reactions. In total, seven glycoenzymes are accounted for in the network: ManI (Mannosidase), GnT I/II/IV (Glucosaminyltransferase), FucT (Fucosyltransferase), GalT (Galactosyltransferase), and SiaT (Sialyltransferase).
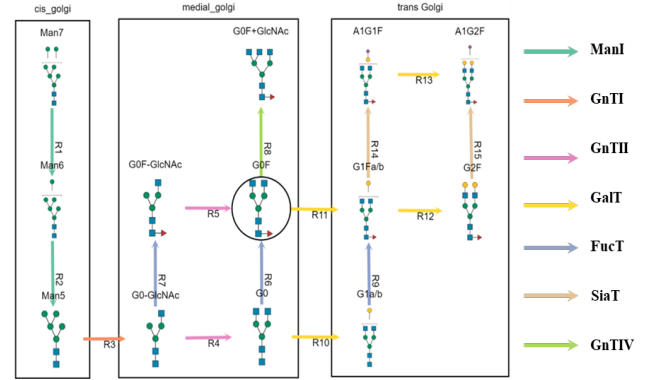

Figure 1. N-linked glycosylation reaction network. Glycoforms are shown by their glycan structures. Arrows depict glycosylation reactions where their colors refer to the glycoenzymes. The main glycoform G0F is circled.

3. RESULTS

3.1 Computational performance comparison: iGFA vs. GFA

The iGFA formulation was implemented using the open-source Pyomo modeling language in Python (Hart, Watson and Woodruff, 2011). The constrained optimization problem was solved the IPOPT coupled with a multistart strategy. To compare the computational performance of iGFA with the MATLAB-based GFA, we applied both methods to the CHO cultivation data in the original GFA publication (Hutter et al., 2017). To ensure a fair evaluation of the two methods, we ran the iGFA with the β factors fixed at 1, while fitting the factor γ to the specific productivity ratio $Q_{prod(t)}/Q_{prod(t_{ref})}$. These modifications translate the secretion flux formulation of the iGFA to be the same as that of the GFA. Both methods were ran using a multi-start strategy with *n*=100 random starting points. The fit quality was evaluated using the Sum of Squares of the Residuals (SSR). Here, the iGFA offered >15-fold reduction in computational times over the GFA, completing 100 multi-start optimizations in a total time of 786 seconds (~13 minutes) versus GFA's total time of 12,154 seconds

(~202 minutes) (Intel® Xeon® E-22146G 3.50 GHz with 16GB RAM). The SSR values were comparable (iGFA: 0.40 vs. GFA: 0.24), and the difference may have been caused by the L2-norm regularization enforced in the iGFA formulation.

3.2 pH effects on N-linked glycosylation fluxes in CHO cultivation of IgG production

Cell culture pH is known to affect N-linked glycosylation during mAb production. Lower pH has been shown to lead to higher galactosylated and sialylated glycoform fractions and a decrease in cell growth and antibody production (Aghamohseni *et al.*, 2014). However, pH effects on galactosylation appear to be cell line dependent, where some cell lines exhibit increased galactosylation with increased cell culture pH (Muthing *et al.*, 2003; Jiang, Chen and Xu, 2018; Lee *et al.*, 2021) and others showing the opposite effect galactosylation and sialylation rates (Ivarsson *et al.*, 2014). Base additions during fed-batch cell cultivation can induce pH excursions that correlate with an increase in lactate production, osmolality, and antibody galactosylation, and a decrease in specific productivity (Jiang, Chen and Xu, 2018).

Lee *et al.* carried out a comprehensive multiomics investigation of the effects of pH on N-linked glycans of IgG in CHO cell cultures (Lee *et al.*, 2021). Briefly, fed-batch cultures of GS-CHO were performed to express a proprietary IgG. Cell cultures were grown at 3 alternate pH conditions – 6.7 (low), 6.9 (medium), and 7.1 (high). The temperature during the culture was set at 36.5ºC with dissolved oxygen controlled at 40%. Daily samples were taken during the experiment to monitor cell viability, antibody titer, VCD, osmolality, and extracellular metabolite and salt concentrations. In addition, samples were collected on working day (WD) 1, 5, 8, 10, 12 and 14 for transcriptomics, metabolomics, proteomics and glycosylation profile analysis. We evaluated glycoform secretion fluxes for the time points: day 5, 8, 10, and 13, based on the availability of the cell specific productivity data.

Based on the glycoforms detected in the Lee *et al.* study, we curated the relevant N-linked glycosylation reaction network for our iGFA analysis (see Fig. 1 and Methods). We evaluated glycoform secretion fluxes using the time-series measurements of VCD, cell specific productivity, IgG titer, and glycoform fractions (see Methods). In the iGFA, we used day 5 as the reference time point and estimated the intracellular glycosylation fluxes for each pH condition separately. We conducted a grid search within the range of 0 to 200 to identify the optimal λ value ($\lambda = 100$) that gave the lowest loss function. We also employed a multistart strategy with randomized initialization of unknown parameters ($n = 500$). Among the feasible solutions, we selected the one with the lowest objective value. As shown in Fig. 2, the iGFA was able to produce flux values that fit the glycoform secretion values well in all pH conditions.

Fig. 3 depicts the results of the iGFA for the data from Lee *et al.* study. Fig 3a shows the estimated intracellular glycosylation fluxes and their dynamics in the three pH conditions. The estimated factors α and β are shown in Fig. 3c-d, respectively. As expected from the loss function of the
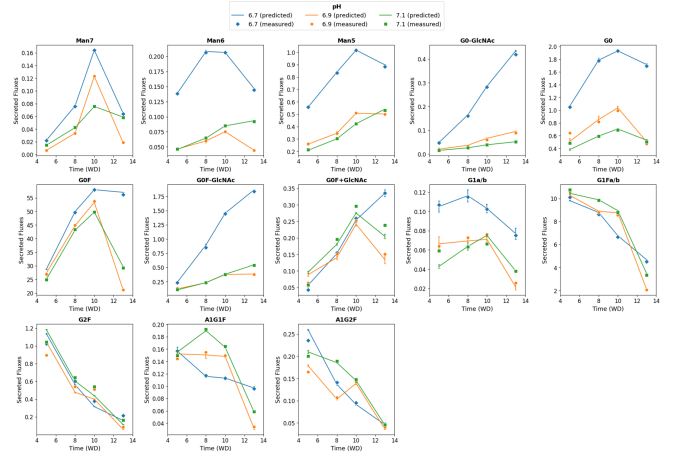

Figure 2. Fitting glycoform secretion fluxes in iGFA. Symbols shows secretion fluxes from time-series measurements. Lines show secretion fluxes computed in the iGFA.

iGFA, specifically the last term in Eq. (12), the trend of $\beta(t)$ follows closely the dynamics of the corresponding glycoform secretion fluxes. From Fig. 3a, we observed a consistent pattern across different pH conditions where the upstream glycosylation reactions related to the trimming of high mannose glycans (reactions R1-R3) increase with the cell cultivation time, but downstream reactions including galactosylation and sialylation that produce more complex glycan structures (reactions R10-R15) decrease with time. We noted that this trend correlates with a decrease in the factor γ (see Fig. 3b) that is shared among all glycosylation reactions. Here, a drop in γ indicates a lower residence time and correspondingly a shorter glycosylation processing time in the Golgi compartment. The shorter processing time thus implies a shift toward higher proportions of simpler glycoforms— there is less time for mAb products in the Golgi to generate
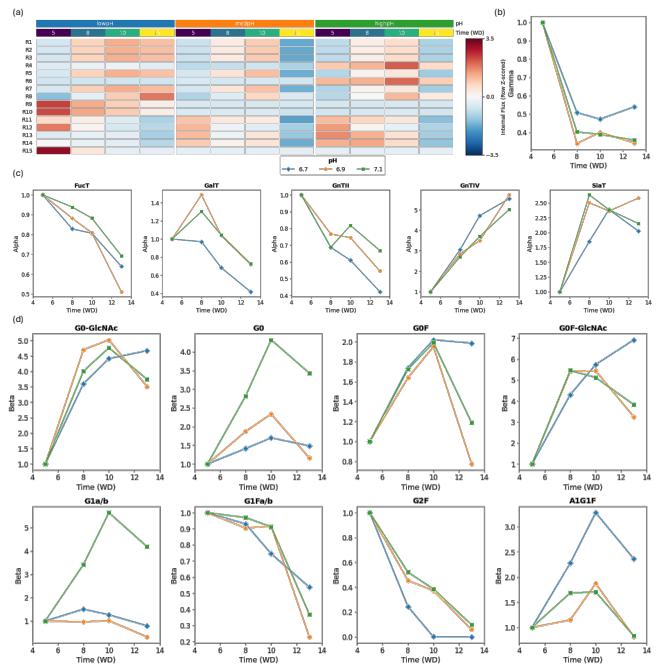

Figure 3. Results of iGFA (a) Estimated intracellular glycosylation fluxes over time points (day 5, 8, 10, 13) in different pH conditions. (b-d) Estimated factors $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, and $\boldsymbol{\gamma}$.

more complex glycoforms from simpler glycoforms. This trend is consistent with increased expression of proteins and metabolites involved in ER and Golgi vesicular transport with the cell cultivation time, indicating higher protein trafficking (see Supplementary Figure 6 in (Lee *et al.*, 2021)).

In addition, we noted a shift in the glycosylation reactions that generate the main IgG glycoform G0F. As depicted in Fig. 1, there are two routes to produce G0F from G0-GlcNAc, one through G0 (R4 and R6) and another through G0F-GlcNAc (R5 and R7). Specifically, glycosylation fluxes through G0 increases with higher pH, while the fluxes through G0F-GlcNAc decrease correspondingly. However, further research is needed to determine the biological mechanism behind this flux shift with pH.

3.3 Comparison of enzyme factors with gene expression and metabolite levels.

Based on the formulation of the iGFA (see Section 2.2), the enzyme factor α is an indicator of the processing capacity of specific glycoenzymes. This capacity is related to both glycoenzyme activity and associated nucleotide sugar availability. Fig. 3c shows that three enzymes, specifically fucosyltransferase (FucT), galactosyltransferase (GalT), and N-acetylglycosaminyltransferase II (GnT-II), have a decreasing trend with the cell cultivation. Meanwhile N-acetylglycosaminyltransferase IV (GnT-IV) and to a lesser degree, sialyltransferase (SiaT) have the opposite trend. Among these enzymatic processing, pH exerted the strongest effect on galactosylation, in agreement with the original study reporting a slight decrease in galactosylated products with decreasing culture pH (Lee et al., 2021).

We investigated more deeply the relationship between each enzyme factor α with the measured expression levels of related genes and nucleotide sugars. Fig. 4 shows the correlations between α and the expression of related gene(s) and associated nucleotide sugar. The correlations between the enzyme factors α and their corresponding nucleotide sugar levels were generally lesser than those between the factors α and the gene expressions. This suggests that the regulation of
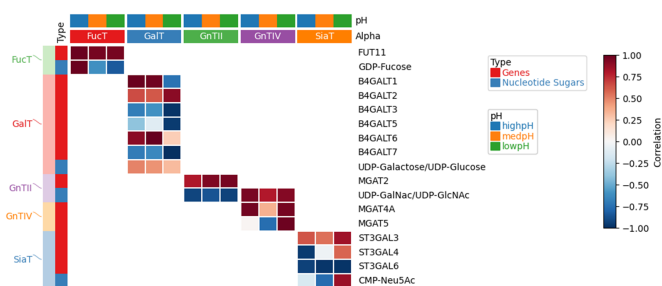

Figure 4. Correlations between the factors α and the gene expression and nucleotide sugar levels. The columns refer to the factors α computed for different enzymes and pH conditions. The rows refer to the genes and nucleotide sugars associated with different N-linked glycosylation processing enzymes: fucosylation (FucT), galactosylation (GalT), GlcNAcylation (GnTII and GnTIV), and sialylation (SiaT).

enzymatic processing capacity in the N-linked glycosylation network is more likely to be regulated at the level of enzyme expression and activity. This observation is in agreement with our previous study that demonstrated the robustness of N-linked glycosylation to changes in cellular nucleotide-sugar concentrations (Del Solar *et al.*, 2020).

Focusing further on the genes, we noted strong positive correlations between α and gene expression for enzymatic processing with a small number of isozymes, for example fucosylation (FucT) and GlcNAclyation (GnTII and GnTIV). Meanwhile, for enzymatic processing with a larger number of isozymes, such as galactosylation and sialylation, the correlations between α and the expression of these various genes are less clear—both positive and negative correlations are observed. Perhaps, the ambiguity of the correlations is expected since the enzymatic processing may rely on a subset of isozymes.

4. CONCLUSION

In this work, we present iGFA, an open-source, Python-based constraint-based modeling method for analyzing glycosylation processing in the cell culture production of therapeutic proteins. iGFA generates estimates of intracellular glycosylation fluxes using secretion fluxes of protein glycoforms computed from time-series measurements of viable cell density (VCD), cell-specific productivity, protein titer, and glycoform fractions. iGFA improves upon our previously published GFA by offering (1) significant computational efficiency with over an order of magnitude reduction in computational time, and (2) biologically interpretable parameters derived from Michaelis-Menten enzymatic kinetics. The application of iGFA to Lee *et al.*'s study on the effect of pH on N-linked glycosylation of IgG in CHO cell cultivation demonstrates the utility of iGFA in characterizing the dynamics of the N-linked glycosylation network and its changes with pH. Specifically, iGFA was able to characterize the decrease in Golgi residence time during cell cultivation. Furthermore, iGFA identified galactosylation as the glycosylation step most impacted by pH. Finally, correlation analysis of the iGFA results and multiomics measurements showed stronger support for the regulation of glycosylation at the transcriptional/translational level than at the metabolic level (nucleotide sugar).

The iGFA can be applied to larger or other glycosylation processes (*e.g.*, O-glycans). Note that the number of unknown model parameters increases linearly with the number of enzymes and glycoforms in the network, and parameter fitting may become computationally limiting for large networks. Another limitation is that although iGFA accounts for time-dependent factors, it assumes that the underlying enzymatic mechanisms remain unchanged throughout the process This assumption may not fully capture complex biological dynamics such as changes in enzyme conformation, allosteric regulation, or competitive inhibition that can alter reaction kinetics over time. Incorporating such dynamic behaviors would enhance iGFA's capabilities but would also increase model complexity. Also, uncertainty quantification for the estimated factors and glycosylation fluxes is not accounted for —a feature that will be added in the future.

5. DATA AND CODE AVAILABILITY

Source code and Python scripts for the iGFA and its application to Lee et al. study are available at the following website: https://github.com/CABSEL/iGFA.

6. ACKNOWLEDGEMENTS

REFERENCES

Aggarwal, S., Qi, X., Neelamegham, S. and Gunawan, R. (2021) 'Compartmental Glycosylation Flux Analysis', *Ifac Papersonline,* 54(3), pp. 287-293.

Aghamohseni, H., Ohadi, K., Spearman, M., Krahn, N., Moo-Young, M., Scharer, J. M., Butler, M. and Budman, H. M. (2014) 'Effects of nutrient levels and average culture pH on the glycosylation pattern of camelid-humanized monoclonal antibody', *J Biotechnol,* 186, pp. 98-109.

Alhuthali, S., Kotidis, P. and Kontoravdi, C. (2021) 'Osmolality Effects on CHO Cell Growth, Cell Volume, Antibody Productivity and Glycosylation', *Int J Mol Sci,* 22(7).

Chen, B., Liu, W., Li, Y., Ma, B., Shang, S. and Tan, Z. (2022) 'Impact of N-Linked Glycosylation on Therapeutic Proteins', *Molecules,* 27(24).

Del Solar, V., Gupta, R., Zhou, Y., Pawlowski, G., Matta, K. L. and Neelamegham, S. (2020) 'Robustness in glycosylation systems: effect of modified monosaccharides, acceptor decoys and azido sugars on cellular nucleotide-sugar levels and pattern of N-linked glycosylation', *Mol Omics,* 16(4), pp. 377-386.

Grand View Research (2023) *Monoclonal Antibodies Market Size, Share & Trends Analysis Report By Source Type (Chimeric, Murine, Humanized, Human), By Production Type (In Vivo, In Vitro), By Application, By End-use, By Region, And Segment Forecasts, 2023 - 2030.*: Grand View Research (GVR-1-68038-280-8.

Hart, W. E., Watson, J. P. and Woodruff, D. L. (2011) 'Pyomo: modeling and solving mathematical programs in Python', *Mathematical Programming Computation,* 3(3), pp. 219-260.

Hutter, S., Villiger, T. K., Bruhlmann, D., Stettler, M., Broly, H., Soos, M. and Gunawan, R. (2017) 'Glycosylation flux analysis reveals dynamic changes of intracellular glycosylation flux distribution in Chinese hamster ovary fed-batch cultures', *Metab Eng,* 43(Pt A), pp. 9-20.

Hutter, S., Wolf, M., Gao, N. P., Lepori, D., Schweigler, T., Morbidelli, M. and Gunawan, R. (2018) 'Glycosylation Flux Analysis of Immunoglobulin G in Chinese Hamster Ovary Perfusion Cell Culture', *Processes,* 6(10).

Ivarsson, M., Villiger, T. K., Morbidelli, M. and Soos, M. (2014) 'Evaluating the impact of cell culture process parameters on monoclonal antibody N-glycosylation', *J Biotechnol,* 188, pp. 88-96.

Jiang, R., Chen, H. and Xu, S. (2018) 'pH excursions impact CHO cell culture performance and antibody N-linked lycosylation', *Bioprocess Biosyst Eng,* 41(12), pp. 1731-1741.

Jimenez del Val, I., Nagy, J. M. and Kontoravdi, C. (2011) 'A dynamic mathematical model for monoclonal antibody N-linked glycosylation and nucleotide sugar donor transport within a maturing Golgi apparatus', *Biotechnol Prog,* 27(6), pp. 1730-43.

Kiehl, T. R., Shen, D., Khattak, S. F., Jian Li, Z. and Sharfstein, S. T. (2011) 'Observations of cell size dynamics under osmotic stress', *Cytometry A,* 79(7), pp. 560-9.

Lee, A. P., Kok, Y. J., Lakshmanan, M., Leong, D., Zheng, L., Lim, H. L., Chen, S., Mak, S. Y., Ang, K. S., Templeton, N., Salim, T., Wei, X., Gifford, E., Tan, A. H., Bi, X., Ng, S. K., Lee, D. Y., Ling, W. L. W. and Ho, Y. S. (2021) 'Multi-omics profiling of a CHO cell culture system unravels the effect of culture pH on cell growth, antibody titer, and product quality', *Biotechnol Bioeng,* 118(11), pp. 4305-4316.

Murakami, S., Matsumoto, R. and Kanamori, T. (2019) 'Constructive approach for synthesis of a functional IgG using a reconstituted cell-free protein synthesis system', *Sci Rep,* 9(1), pp. 671.

Muthing, J., Kemminer, S. E., Conradt, H. S., Sagi, D., Nimtz, M., Karst, U. and Peter-Katalinic, J. (2003) 'Effects of buffering conditions and culture pH on production rates and glycosylation of clinical phase I anti-melanoma mouse IgG3 monoclonal antibody R24', *Biotechnol Bioeng,* 83(3), pp. 321-34.

Neelamegham, S. and Liu, G. (2011) 'Systems glycobiology: biochemical reaction networks regulating glycan structure and function', *Glycobiology,* 21(12), pp. 1541-53.

Pan, X., Dalm, C., Wijffels, R. H. and Martens, D. E. (2017) 'Metabolic characterization of a CHO cell size increase phase in fed-batch cultures', *Appl Microbiol Biotechnol,* 101(22), pp. 8101-8113.

Spahn, P. N., Hansen, A. H., Hansen, H. G., Arnsdorf, J., Kildegaard, H. F. and Lewis, N. E. (2016) 'A Markov chain model for N-linked protein glycosylation--towards a low-parameter tool for model-driven glycoengineering', *Metab Eng,* 33, pp. 52-66.

Villiger, T. K., Scibona, E., Stettler, M., Broly, H., Morbidelli, M. and Soos, M. (2016) 'Controlling the time evolution of mAb N-linked glycosylation - Part II: Model-based predictions', *Biotechnology Progress,* 32(5), pp. 1135-1148.

Wachter, A. and Biegler, L. T. (2006) 'On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming', *Mathematical Programming,* 106(1), pp. 25-57.

Wang, Z., Zhu, J. and Lu, H. (2020) 'Antibody glycosylation: impact on antibody drug characteristics and quality control', *Appl Microbiol Biotechnol,* 104(5), pp. 1905-1914.

Zhang, L., Wang, M., Castan, A., Hjalmarsson, H. and Chotteau, V. (2021) 'Probabilistic model by Bayesian network for the prediction of antibody glycosylation in perfusion and fed-batch cell cultures', *Biotechnol Bioeng,* 118(9), pp. 3447-3459.