Article

Glycosylation Flux Analysis of Immunoglobulin G in Chinese Hamster Ovary Perfusion Cell Culture

Sandro Hutter[1,2,a], Moritz Wolf[1,a], Nan Papili Gao[1,2], Dario Lepori[1], Thea Schweigler[1], Massimo Morbidelli[1] and Rudiyanto Gunawan[1,2,3,*]

[1] Institute for Chemical and Bioengineering, Department of Chemistry and Applied Biosciences, ETH Zurich, Zurich 8093, Switzerland
[2] Swiss Institute of Bioinformatics, Lausanne 1015, Switzerland
[3] Department of Chemical and Biological Engineering, University at Buffalo, The State University of New York, Amherst, NY 14260, USA
[a] Contributed equally
* Correspondence: rudiyant@buffalo.edu

Abstract: A critical quality attribute of therapeutic monoclonal antibodies (mAbs) is the terminal sugar molecules of the N-linked glycan attached to the fragment crystalizable (Fc) region. There exists naturally-occurring heterogeneity in the N-linked glycan structure of mAbs, and such heterogeneity has a significant influence on the clinical safety and efficacy of mAb drugs. We previously proposed a constraint-based modeling method called glycosylation flux analysis (GFA) to characterize the rates (fluxes) of intracellular glycosylation reactions and applied the method to examine the N-linked glycosylation of immunoglobulin G (IgG) in fed-batch Chinese hamster ovary (CHO) fed-batch cultivations. In this work, we significantly improved the computational efficiency of the GFA, and employed the method to analyze the glycosylation of IgG in continuous perfusion CHO cultivations. Perfusion cell cultures have several advantages over the traditional (fed-)batch operation, including higher productivity per unit volume of reactor and more consistent product quality. The GFA showed that as in the fed-batch cultivation, the dynamical changes of IgG glycan heterogeneity in the perfusion culture are mainly attributed to alterations in the galactosylation flux activity. Furthermore, a regression analysis of the galactosylation flux activity using random forest regression linked the dynamics of galactosylation activity with the cell-specific productivity of IgG and the extracellular ammonia concentration.

Keywords: N-linked glycosylation; perfusion cell culture; CHO cells; constraint-based modeling; monoclonal antibody

1. Introduction

Therapeutic recombinant monoclonal antibodies constitute the most important class of drugs in the biopharmaceutical industry, making up approximately half of the total revenue of biopharmaceutical products in 2013 [1]. The production of mAb drugs typically employs (fed-)batch cultivations of mammalian cells. The state of the art (fed-)batch cell cultures are able to meet the large production volume requirement of mAbs with reactors of up to 25,000 L [2]. Nevertheless, the increasing number of mAb products entering different stages of clinical trials and the burgeoning market of biosimilars driven by impending patent expirations of blockbuster mAb drugs, give strong motivation for the development of new production technology that is more robust and cost effective [3]. The US Food and Drug Administration (FDA) initiative on quality by design and process analytical technology put further stress on implementing quantitative approaches for process improvements in biopharmaceutical manufacturing [4,5].

Continuous manufacturing technology offers an effective and flexible way for large scale and robust production of drug compounds [6]. In the biopharmaceutical industry, the application of continuous cell culture technology has thus far been limited to the production of unstable products that require constant recovery [7]. Nevertheless, continuous perfusion cell cultures have previously been demonstrated to be capable of producing antibodies at a volumetric rate that match or exceed

that of fed-batch cultures [8]. In addition to the high productivity, the stable steady state operation mode in conjunction with the short residence time of perfusion cultures translate to a tight maintenance of product quality.

Among the most important critical quality attribute (CQA) of therapeutic mAbs is the glycan structures of the Fc domain [9]. The N-linked glycosylation is a common post-translational modification of proteins, a process that occurs in the endoplasmic reticulum (ER) and Golgi apparatuses. The Fc glycan structure has been shown to impact protein folding [10], clearance [11], bioactivity [12], efficacy [13] as well as immunogenicity [14]. Moreover, there exists naturally-occurring heterogeneity in the glycosylation of mAbs. Thus, the FDA approval of mAb drugs is given for a particular composition of mAb glycoforms [5]. Because of the importance of N-linked glycosylation, the majority of recombinant mAb drugs are produced using mammalian host cells in order to achieve human-like glycan structures [15]. In particular, CHO cells have become the major expression host for the biopharmaceutical production of therapeutic mAbs [16].

The N-linked glycosylation of mAbs has been shown to depend on the host genetic background [17], expression vector [18], media composition [9], media supplements [19], and bioprocess parameters [20]. But, the mechanism of the above dependence remains to be established [21]. Toward closing this gap, mathematical models of the glycosylation network have previously been developed [22–28]. Many of these models have a large number of unknown and system-specific kinetic parameters that need to be fitted to experimental data [22–25]. A number of parameter-free models have also been proposed to study the N-linked glycosylation process based on the constraint-based modeling approach (i.e. stoichiometric models) [26–28]. Recently, we proposed a flux analysis method, called glycosylation flux analysis, which provides predictions of the rate or flux of intracellular glycosylation reactions using the stoichiometric model of the glycosylation network and the cell secretion rates of mAb glycoforms [28]. Similar to the well-known metabolic flux analysis (MFA), the GFA is based on the molar balances of glycoforms involved in the glycosylation reaction network under the pseudo steady state assumption [29]. The GFA further makes use of the relatively small number of enzymes involved in the glycosylation process to reduce the degrees of freedom in the flux estimation [30]. More specifically, we assumed that glycosylation fluxes vary with time according to a (global) cell-specific factor and a (local) enzyme specific factor. When applied to study the IgG production of CHO cells in fed-batch cultivations, GFA was able to give insights on how changes in the media affect the glycosylation reactions [28].

The formulation of the flux estimation in the GFA involves a nonlinear least-square regression, which requires computationally intensive global optimizations [28]. In this work, we improved the computational efficiency of the GFA by decomposing the flux estimation into two iterative linear regression problems. Subsequently, we applied the improved GFA to analyze the N-linked glycosylation of IgG in perfusion CHO cell cultures, to elucidate the key controlling factors and to identify differences in the IgG glycosylation– if any – between the continuous perfusion and fed-batch cultivations.

2. Materials and Methods

2.1. Continuous perfusion cell cultures

The detailed procedure of the perfusion cell culture is available elsewhere [31]. Here, we provide a brief summary of the experiments. A proprietary CHO cell line expressing IgG1 was cultured using a previously developed 1.5 L perfusion cell culture setup [32], as depicted in Figure 1. In total, four perfusion cell culture experiments were performed with different viable cell density (VCD) and perfusion rate (PR) set-point profiles, as shown in Figure 2. Each of the experiments was inoculated following the same procedure. CHO cells were held back in the reactor by a cell retention device so that only cell free reaction mixture left through the harvest stream. The feed flowrate (F) of fresh media into the reactor and the bleed (B) and harvest (H) flowrates out of the reactor were balanced

to keep the reactor volume constant. The perfusion rate (PR) is given by the flowrate through the reactor, as follows:

$$PR = H + B = F \quad (1)$$

The perfusion rate represents the rate of fresh media supplied to the cells. We define the cell-specific perfusion rate (CSPR) as the rate of fresh media per cell:

$$CSPR = \frac{PR}{VCD} \quad (2)$$

During the experiments, there following measurements were collected on a daily basis: cell counts and cell viability, concentrations of glucose, lactate and ammonia, IgG titer, and protein glycan distribution.

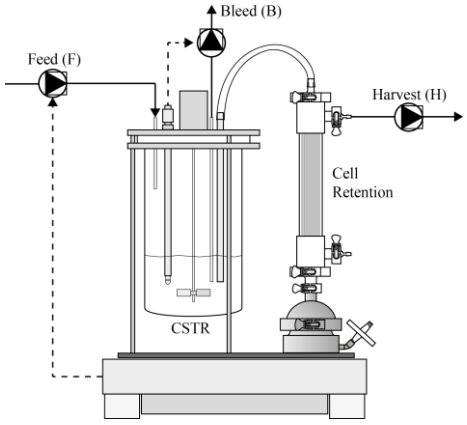

Figure 1. A schematic of perfusion cell culture reactor[1]. CHO cells are cultivated in suspension in a continuous stirred tank reactor with continuous feeding of fresh nutrients. Cell-free spent media is constantly collected in the harvest stream, while cells remain in the stirred tank reactor thanks to a cell retention device. A bleed stream removes a small fraction of the reactor mixture, including biomass, which is used to regulate the viable cell density.

[1] Figure adapted from [32] with the permission from the authors.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 30 May 2018 doi:10.20944/preprints201805.0456.v1

Peer-reviewed version available at Processes 2018, 6, 176; doi:10.3390/pr6100176

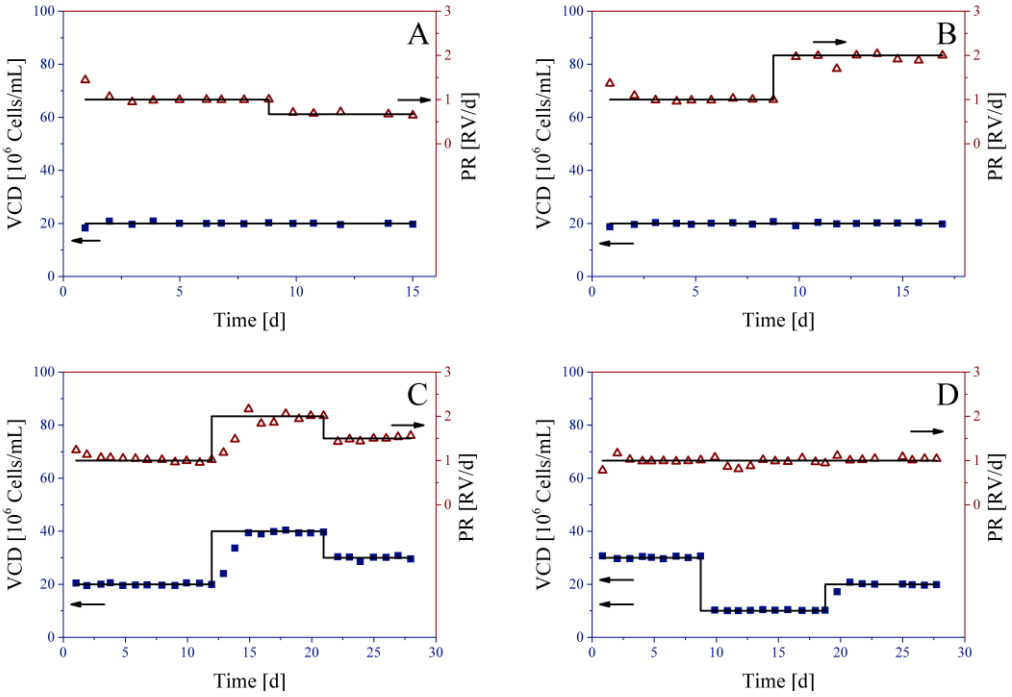

Figure 2. Perfusion cell culture experiments. Four perfusion cell culture experiments were conducted with varying VCD and PR given set-points (lines). The experimental VCD and PR are shown as blue filled squares and red empty triangles, respectively.

2.2 Estimation of IgG glycoforms secretion fluxes

As inputs to the GFA, the cell-specific secretion flux of each IgG glycoform ($v_{\mathrm{E},i}$) was determined based on the molar balance in the reactor, as follows

$$V\frac{dc_{\mathrm{E},i}(t)}{dt} = Vv_{\mathrm{E},i}(t)\mathrm{VCD}(t) - \big(\mathrm{H}(t) + \mathrm{B}(t)\big)c_{\mathrm{E},i}(t) \tag{3}$$

where $c_{\mathrm{E},i}$ denotes the concentration of the i-th IgG glycoform and V denotes the reactor volume. Since the reactor volume is kept (approximately) constant, Equation (3) can be rearranged to give:

$$v_{\mathrm{E},i}(t) = \frac{\left(\frac{dc_{\mathrm{E},i}(t)}{dt} + \big(\mathrm{h}(t) + \mathrm{b}(t)\big)c_{\mathrm{E},i}(t)\right)}{\mathrm{VCD}(t)} \tag{4}$$

The variables $\mathrm{h(t)} = \frac{\mathrm{H(t)}}{\mathrm{V}}$ and $\mathrm{b(t)} = \frac{\mathrm{B(t)}}{\mathrm{V}}$ represent the specific harvest and bleed rate, respectively. The concentration of the IgG glycoform ($c_{\mathrm{E},i}$) was calculated as the product of the measured glycoform fraction (f_i) and the IgG titer (T) according to:

$$\mathrm{c}_{\mathrm{E},i}(t_k) = \mathrm{f}_i(t_k)\mathrm{T}(t_k) \tag{5}$$

where t_k denotes the k-th measurement time point. For the estimation of the secretion fluxes in Equation (4), the computed concentration of IgG glycoforms $\mathrm{c}_{\mathrm{E},i}(t_k)$ was first smoothened (as a function of time) using spline fitting. The time derivative $\frac{dc_{\mathrm{E},i}(t)}{dt}$ was then evaluated by taking the first order derivative of the spline curve.

2.3 Glycosylation Flux Analysis

The glycosylation flux analysis is based on a constraint-based modeling of the protein glycosylation reaction network to evaluate the intracellular glycosylation fluxes using the secretion rates of mAb glycoforms. Like the MFA, the GFA uses a pseudo steady state assumption to derive the stoichiometric model of the glycosylation process, as follows:

$$\frac{d\mathbf{c}_\mathrm{I}(t)}{dt} = \mathbf{S}\mathbf{v}_\mathrm{I}(t) - \mathbf{v}_\mathrm{E}(t) = 0 \tag{6}$$

where $\mathbf{c}_\mathrm{I}$ denotes the vector of m intracellular IgG glycoform concentrations, $\mathbf{v}_\mathrm{I}$ denotes the vector of n intracellular IgG glycosylation reaction fluxes (rates), $\mathbf{v}_\mathrm{E}$ denotes the vector of secretion fluxes of IgG glycoforms estimated above, and $\mathbf{S}$ denotes the $m \times n$ stoichiometric matrix. The (i,j)-th element of $\mathbf{S}$ gives the number of the i-th glycoform molecule produced (if positive) or consumed (if negative) by the j-th glycosylation reaction. Since the number of reaction fluxes (i.e. the number of unknowns) typically exceeds that of glycoforms (i.e. the number of equations), the estimation of $\mathbf{v}_\mathrm{I}$ from $\mathbf{v}_\mathrm{E}$ in Equation (6) is underdetermined.

To address the issue of underdetermined regression, in the GFA each glycosylation flux $\mathrm{v}_{\mathrm{I},j}(t)$ is computed as the product of a reference flux value $\mathrm{v}_{\mathrm{I},j}{}^{\mathrm{ref}}$, an enzyme specific factor $\alpha_J(t)$ and a cell-specific factor $\beta(t)$, as follows:

$$\mathrm{v}_{\mathrm{I},j}(t) = \alpha_J(t)\beta(t)\mathrm{v}_{\mathrm{I},j}{}^{\mathrm{ref}} \tag{7}$$

The variables $\alpha_J(t)$ and $\beta(t)$ represent the fold-change amplification or attenuation, and can therefore be normalized to 1 at a chosen reference time point t_{ref}. The cell-specific factor $\beta(t)$ captures the (global) influence of the cell metabolism on the glycosylation process, more specifically the total amount of mAb entering/leaving the glycosylation network. For this reason, $\beta(t)$ is represented by the ratio of the cell-specific productivity ($\mathrm{q}_{\mathrm{mAb}}$) between time t and the reference time t^{ref}, as follows:

$$\beta(t) = \frac{\mathrm{q}_{\mathrm{mAb}}(t)}{\mathrm{q}_{\mathrm{mAb}}(t^{\mathrm{ref}})} \tag{8}$$

Meanwhile, the factor $\alpha_J(t)$ describes the (local) influence of enzymatic processing capacity, which captures the dependence of glycosylation on factors such as enzyme expression and activity, as well as co-factor and nucleotide sugar availability. Note that the number of enzymes involved in the glycosylation network is typically much smaller than the number of reactions, as an enzyme catalyzes multiple reactions. Thus, the estimation of $\mathbf{v}_\mathrm{I}(t)$ can be reformulated to fitting $\alpha_J(t)$ and $\mathrm{v}_{\mathrm{I},j}{}^{\mathrm{ref}}$ to the secretion fluxes, as follows:

$$\min_{\alpha_J(t),\mathrm{v}_{\mathrm{I},j}{}^{\mathrm{ref}}} \Phi = \|\mathbf{S}\mathbf{v}_\mathrm{I}(t) - \mathbf{v}_\mathrm{E}(t)\|_\mathbf{2}^\mathbf{2} \tag{9}$$

For a more detailed derivation of the GFA, we refer the interested readers to the original publication [28].

The formulation in Equation (9) is a nonlinear programming problem that requires a global optimization algorithm to solve. In the following, we describe an alternative and more computationally efficient procedure for solving the regression problem in the GFA. The procedure is based on decomposing the nonlinear least square regression above into two linear regression problems. First, for a given reference flux vector $\mathbf{v}_\mathrm{I}{}^{\mathrm{ref}}$, one can formulate the following linear regression problem to obtain the least square estimate of α_J:

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 30 May 2018 doi:10.20944/preprints201805.0456.v1

Peer-reviewed version available at Processes 2018, 6, 176; doi:10.3390/pr6100176

$$\mathbf{v}_{\mathrm{E}}(t_k) = \beta(t_k)\mathbf{S}\boldsymbol{\Psi}_{\mathrm{I}}\alpha(t_k) = \beta(t_k)\mathbf{S}\begin{bmatrix} \mathrm{v}_{\mathrm{I},1}^{\mathrm{ref}} & & & \\ \vdots & 0 & & \\ \mathrm{v}_{\mathrm{I},n_{J_1}}^{\mathrm{ref}} & & & \\ & \mathrm{v}_{\mathrm{I},n_{J_1+1}}^{\mathrm{ref}} & 0 & \\ 0 & \vdots & & \\ & \mathrm{v}_{\mathrm{I},n_{J_1+J_2}}^{\mathrm{ref}} & & \\ & & & \ddots \end{bmatrix}\begin{bmatrix} \alpha_{J_1}(t_k) \\ \alpha_{J_2}(t_k) \\ \vdots \end{bmatrix} \tag{10}$$

where $\boldsymbol{\Psi}_{\mathrm{I}}$ is an $n \times e$ matrix with e being the number of enzymes involved in the glycosylation network, and n_{J_l} is the number of fluxes catalyzed by the enzyme J_l. $\boldsymbol{\Psi}_{\mathrm{I}}$ is constructed by grouping the reference fluxes ($\mathrm{v}_{\mathrm{I},j}^{\mathrm{ref}}$) according the enzyme that catalyzes the reactions, and stacking each group (block-)diagonally. In this manner, each $\mathrm{v}_{\mathrm{I},j}^{\mathrm{ref}}$ is multiplied with the corresponding enzyme specific factor (α_J). On the other hand, given $\alpha_J(t_k)$, one can obtain the least square values of ${\mathrm{v}_{\mathrm{I},j}}^{\mathrm{ref}}$ using the following linear regression problem:

$$\mathbf{v}_{\mathrm{E}}(t_k) = \beta(t_k)\mathbf{S}\boldsymbol{\Omega}(t_k){\mathrm{v}_{\mathrm{I},j}}^{\mathrm{ref}} = \beta(t_k)\mathbf{S}\begin{bmatrix} \alpha_{J_1}(t_k) & & & & \\ & \ddots & & & \\ & & \alpha_{J_1}(t_k) & & 0 \\ & & & \alpha_{J_2}(t_k) & \\ & & & & \ddots \\ & & 0 & & \alpha_{J_2}(t_k) \\ & & & & & \ddots \end{bmatrix}\begin{bmatrix} {\mathrm{v}_{\mathrm{I},1}}^{\mathrm{ref}} \\ \vdots \\ {\mathrm{v}_{\mathrm{I},n_{J_1}}}^{\mathrm{ref}} \\ {\mathrm{v}_{\mathrm{I},n_{J_1}+1}}^{\mathrm{ref}} \\ \vdots \\ {\mathrm{v}_{\mathrm{I},n_{J_1}+n_{J_2}}}^{\mathrm{ref}} \\ \vdots \end{bmatrix} \tag{11}$$

where $\boldsymbol{\Omega}$ is an $n \times n$ diagonal matrix with the enzyme specific factor (α_J) as its diagonal elements.

Given the two linear regressions in Equations (10) and (11), we estimated $\boldsymbol{\alpha_J}$ and $\mathbf{v_I}^{\mathbf{ref}}$ following an iterative procedure as follows:

(1) generate a uniformly distributed random vector of $\mathbf{v}_{\mathrm{I}}{}^{\mathrm{ref}}$ within a biologically feasible range (${\mathrm{v}_{\mathrm{I},j}}^{\mathrm{ref}} \in [0, 25\mathrm{pg/Cell/day}]$),

(2) given $\mathbf{v_I}^{\mathbf{ref}}$ from (1), solve for or update $\boldsymbol{\alpha_J}$ using Equation (10),

(3) given $\boldsymbol{\alpha_J}$ from (2), solve for or update $\mathbf{v_I}^{\mathbf{ref}}$ using Equation (11), and

(4) repeat step (2) and (3) until the change of Φ as described in Equation (9) becomes smaller than a threshold (10^{-10}).

Equations (10) and (11) were solved using the MATLAB subroutine `lsqlin` constraining $\boldsymbol{\alpha_J}$ to values between 0 and 20 and $\mathbf{v_I}^{\mathbf{ref}}$ to be positive. In order to improve the chance of obtaining the global minimum Φ, we adopted a multi-start strategy and ran the aforementioned iterative estimation for multiple random initial vectors of $\mathbf{v}_{\mathrm{I}}{}^{\mathrm{ref}}$. Among the results of the multi-start runs, we took the best $\boldsymbol{\alpha_J}$ and $\mathbf{v_I}^{\mathbf{ref}}$ values that correspond to the minimum value of Φ. Note that the multi-start strategy above is embarrassingly parallel, and can be easily implemented on a multiprocessor or cluster computing platform.

We compared the computational speed of the iterative GFA above with that of the original formulation in analyzing the perfusion cell culture datasets. On a test computer (3.33GHz Intel Xeon W3680, 18GB RAM), the original formulation of GFA using the global optimization toolbox MEIGO [32] required at least 42 minutes to converge, and only 1 out 10 repeated runs of the original GFA converged to an optimal solution with the same minimum Φ value as the iterative GFA. Meanwhile, the iterative GFA converged to the optimal solution within 5 minutes using 250 random starting points. Of note, among the 250 random starts, 40% converged to the final minimum Φ value, implying that 100 runs would have been sufficient in this particular analysis.

2.4 Random forest for regression

For identifying explanatory variables of the dynamical changes in the galatosyltransferase activity $\alpha_{J,\mathrm{GalT}}$, we formulated a regression problem using the change in GalT enzyme-specific factors over time $d\alpha_{J,\mathrm{GalT}}/dt$ as the response variable, and experimental parameters of perfusion cell cultures as predictor variables. More specifically, we are interested the following regression problem:

$$\frac{d\alpha_{J,\mathrm{GalT}}}{dt} = g\left(t, \alpha_{J,\mathrm{GalT}}, \mathbf{p}\right) \tag{12}$$

which describes the dynamic change in $\alpha_{J,\mathrm{GalT}}$ using the function $g\left(t, \alpha_{J,\mathrm{GalT}}, \mathbf{p}\right)$ that depends on time, $\alpha_{J,\mathrm{GalT}}$, and process parameters $\mathbf{p}$. Note that the function $g\left(t, \alpha_{J,\mathrm{GalT}}, \mathbf{p}\right)$ is likely nonlinear in nature. Here, we employed Random Forest (RF) [34] to build the above regression model using data from all four perfusion cell culture experiments. RF regression involves building an ensemble of unpruned regression trees, in which each regression tree is created using a bootstrap sample of the original dataset. At each node of a tree, a subset of predictors is selected randomly to determine the best decision split of the samples. The final prediction of the regression trees in RF is obtained by averaging the predictions of the entire ensemble. Notably, RF regression is able to capture nonlinear dependencies of the response variable on the predictors.

In our work, we applied RF using normalized data of each variable, in which the data were centered and divided by the standard deviation. We created a RF regression model using an ensemble of 100 trees and employed one third of the total predictors for the decision split at each node. Predictor variables (features) were subsequently ranked based on their average impurity gain over all splits and all trees. Predictor variables with higher impurity gains contribute more to the variability in the prediction of the response variable, and hence are considered more important.

3. Results

3.1 Perfusion cell culture experiments

We performed four perfusion cell cultures (Experiment A, B, C and D) using different VCD and PR set-point profiles, as illustrated in Figure 2. In Experiments A and B, we kept the VCD set-point constant for the entire duration of the cell cultures but shifted the PR set-point between day 9 and 10 (down in Experiment A and up in Experiment B). In Experiment C, we changed the VCD and PR set-points in a manner that maintained the CSPR at the same value. Finally, in Experiment D, we varied the VCD set-point while leaving the PR set-point constant. Figure 2 shows that the actual VCD and PR follow the set-points very well with only minor deviations (also see Table S1). As the control of VCD was done only by bleeding (i.e., removal of cells), the VCD unsurprisingly tracked a decrease in the set-points better than an increase. Supplementary Figures S1-4 give a summary of the cell culture parameters, including VCD, PR, CSPR, the concentrations of glucose (Glc), lactose (Lac), ammonia (Amm) and IgG, and the cell-specific growth rate (μ), for Experiments A to D. As shown in Figure 3, the cell-specific productivities of IgG, i.e. the secretion rate of IgG divided by the VCD, in all experiments follow a decreasing trend over the course of the cell cultivation. Correspondingly, the estimated secretion rates of the IgG glycoforms decrease with time (see Figure 4 and Supplementary Figure S5).

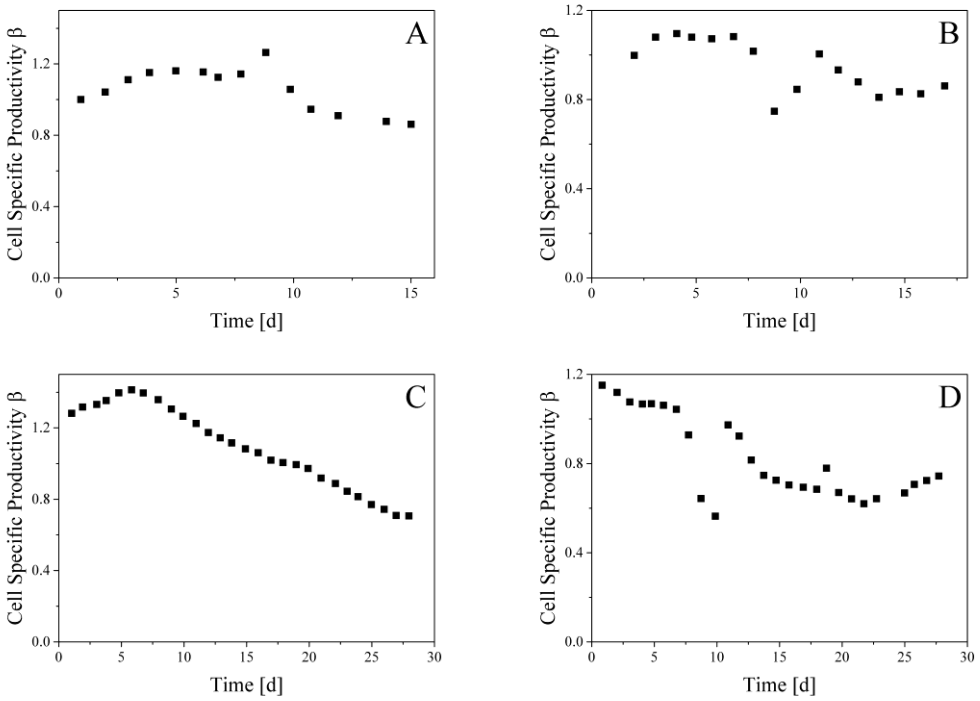

Figure 3. Cell-specific productivity. The cell-specific productivity generally decreases over the course of the perfusion cell cultures.

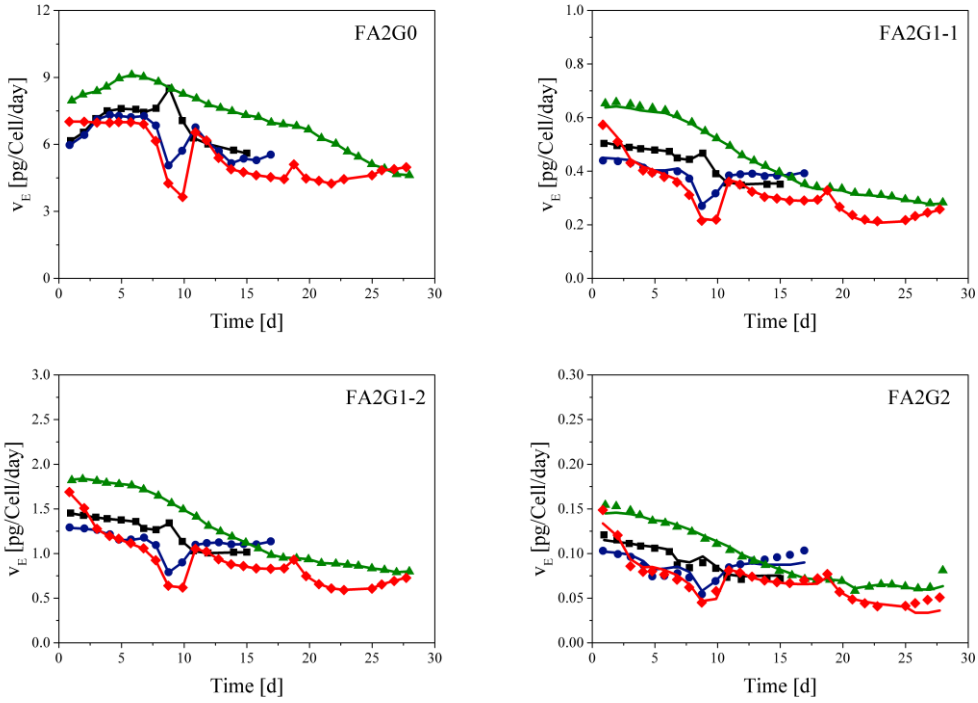

Figure 4. Secretion fluxes of the main IgG glycoforms. The solid symbols show the measured secretion fluxes (Experiment A: black squares, Experiment B: blue circles, Experiment C: green triangle, Experiment D: red diamonds) and the lines show the fitted secretion fluxes by the GFA.

3.2. Glycosylation flux analysis

For the GFA, we employed the IgG glycosylation reaction network depicted in Figure 5, which consists of 19 IgG glycoforms and 25 glycosylation reactions. The glycosylation reaction network was based on a previously published network [35], where we omitted glycosylation reactions and molecules corresponding to IgG glycoforms that are not detected in our experiments and do not participate as intermediate species. Since each perfusion cell culture was started in the same manner, we used the same reference glycosylation flux vector $\mathbf{v}_I^{ref}$ in the GFA of all experiments, more specifically using day 1 of experiment A as the reference time sample point (i.e. $\mathbf{v}_I^{ref}$ of day 1 in Experimental A is set to the vector of 1s). Figure 4 shows the GFA data fitting of secretion fluxes for the major IgG glycoforms (see Supplementary Figure S5 for other IgG glycoforms).

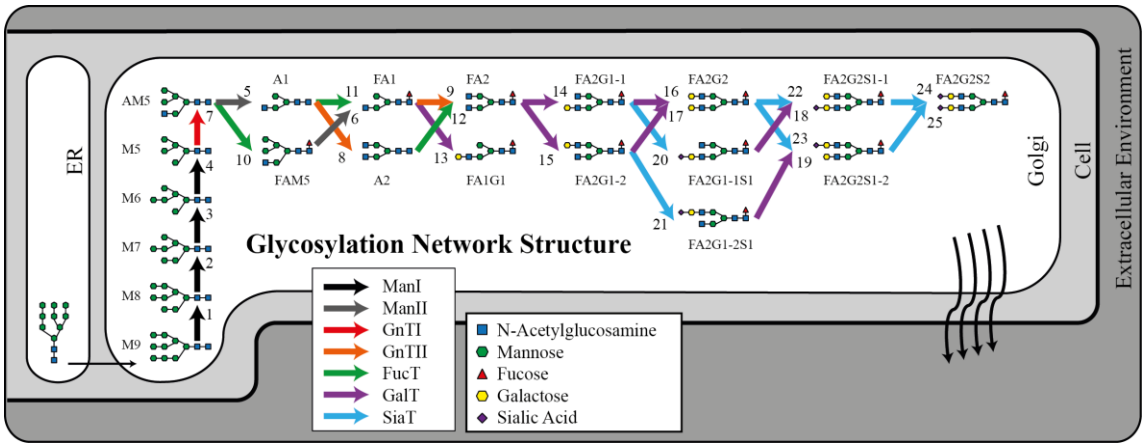

Figure 5. Glycosylation network for the GFA of immunoglobulin G in CHO-S. The enzyme names are abbreviated as follows: α-Mannosidase I and II (Man I/II), N-Acetylglucosaminyltransferase I and II (GnT I/II) and Fucosyltransferase (FucT), Galactosyltransferase (GalT) and Sialyltransferase (SiaT). The glycan labels are provided in Supplementary Table S2.

Figure 6 gives the time profiles of the enzyme specific factors $\alpha_J(t)$ for each of the enzymes in the glycosylation network. Most of the enzyme specific factors, particularly those of α-Mannosidase I and II (Man I/II), N-Acetylglucosaminyltransferase I and II (GnT I/II) and Fucosyltransferase (FucT), maintain a constant activity level throughout the cell cultivation (i.e. $\alpha_J(t) \approx 1$). Meanwhile, the galactosyltransferase (GalT) specific factor decreases during the beginning of the four experiments and varies with changes in the VCD and PR set-points. Since the fluxes catalyzed by sialyltransferase (SiaT) are close to zero, the estimate of the corresponding $\alpha_J(t)$ becomes unreliable due to high sensitivity to experimental noise and thus is omitted from further analysis.

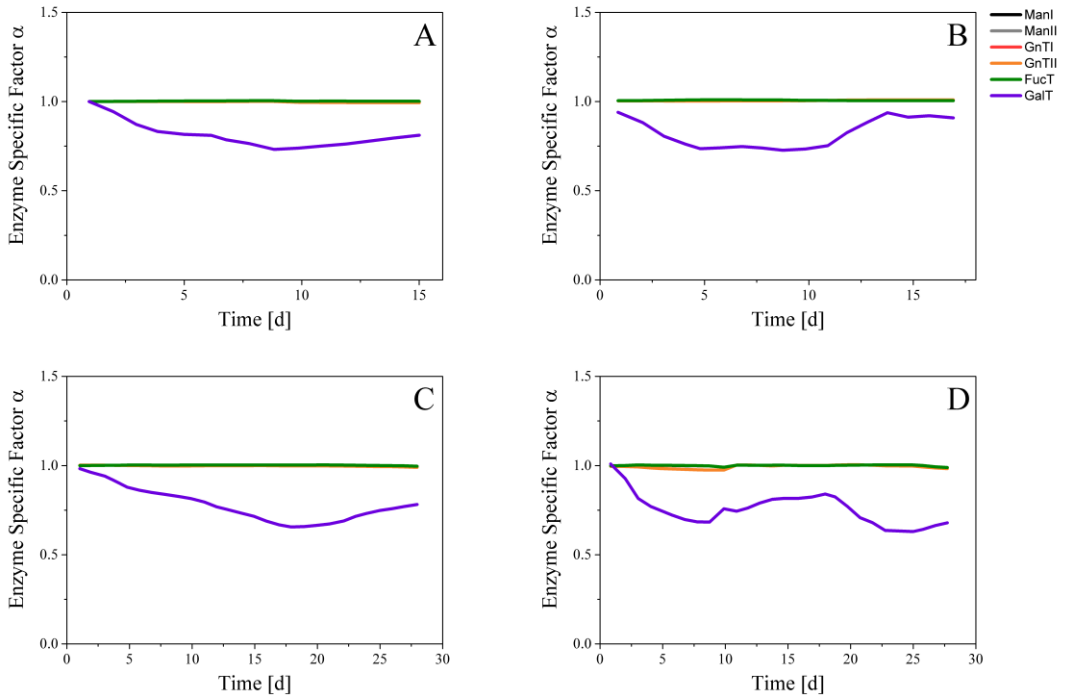

Figure 6. Predicted enzyme specific factors. The activity of fluxes catalyzed by Man I (black), Man II (grey), GnT I (red), GnT II (orange) and FucT (green) remain relatively constant in all experiments. However, the fluxes catalyzed by GalT (purple) shows significant variation over the course of the perfusion cell cultures.

3.3. Effects of process parameters on glycosylation

As mentioned above, the enzyme-specific factor of IgG galactosylation displays the most dynamical change during the perfusion cell culture. However, the relationship between the changes in the enzyme-specific factor $\alpha_{GalT}(t)$ and the other process parameters is difficult to discern by a simple observation of the experimental data. For this reason, we employed a random forest regression analysis using the change of galactosyltransferase specific factor over time $\frac{\mathrm{d}\alpha_{GalT}(t)}{\mathrm{d}t}$ as the response variable and using 14 process parameters as the predictor variables (see Materials and Methods). A RF regression model is able to capture nonlinear dependencies between the response and predictor variables. Here, we considered the following predictors: $\alpha_{GalT}(t)$, VCD, PR, B, CSPR, time and the concentrations of IgG, Glc, Lac and Amm, and the specific productivities of IgG, Glc, Lac and Amm (i.e. q_{IgG}, q_{Glc}, q_{Lac} and q_{Amm}, respectively). Furthermore, we excluded data from the startup period of the cell culture (i.e. days 1 to 3 of each experiment), as we were more interested in the regulation of IgG glycosylation during the steady state operations of perfusion cell culture. Finally, we ranked the predictor variables in decreasing magnitudes of the impurity gains. A higher impurity gain points to a predictor variable with higher importance in explaining the response variable.

Figure 7 gives the ranking of the predictor variables in decreasing impurity gains. The specific productivity of IgG (q_{IgG}) and the concentration of ammonia (Amm) are the two most important predictors of the dynamical changes in the galactosyltransferase specific activity. Indeed, when we repeated the RF regression using only q_{IgG} and Amm as the predictor variables, we observed a similar quality of data fitting to the response variables (see Supplementary Figure S6). Kolmogorov-Smirnov (KS) test and Wilcoxon rank sum test further confirmed that the residuals of the RF regression models

using all 14 predictors and those using only q_{IgG} and Amm, are not statistically different (KS test p-value = 0.857; Wilcoxon rank sum test p-value = 0.824).

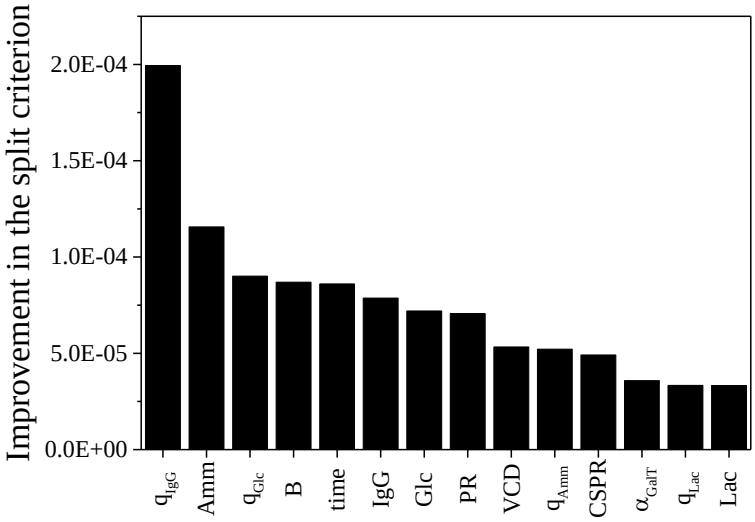

Figure 7. Ranking of predictors based on importance. The predictors are sorted in decreasing magnitude of the impurity gains.

4. Discussion

In this work, we improved the computational efficiency of the GFA and applied the GFA to analyze the IgG glycosylation of four perfusion CHO cell culture experiments. The GFA uses a stoichiometric model of the IgG glycosylation network to estimate intracellular glycosylation reaction fluxes using secretion fluxes of IgG glycoforms. The GFA is based on the assumption that the intracellular IgG glycosylation fluxes vary with time in proportion to α_J and β_J. The enzyme-specific factor α_J represents the relative change of the fraction of IgG that can processed by a specific enzyme, while the cell-specific factor β_J describes the relative change of the amount of IgG entering the glycosylation network. Note that the GFA requires data for computing the secretion fluxes of different glycoforms over time, which include viable cell density, IgG titer and glycoform fractions.

The four perfusion cell cultures in this study differed among each other in the VCD and PR set-point profiles, which were designed with the goal of understanding how IgG productivity and glycosylation vary with these set-points during steady-state operations. The process conditions (i.e. media composition, seeding density, pH, temperature) of the four experiments were kept to be as similar as possible, so that the effect of changes in operating set-points VCD and PR on the glycosylation process could be examined.

In our previous study using the same CHO cells, we reported that the cell-specific productivity of IgG in fed-batch cultivations increases with time, and that such an increase is associated with lower α_Js for the upstream enzymes ManI/II, GnTI/II and FucT [28]. We attributed the decreasing α_Js to the inability of these enzymes to process the increasing amount of IgG entering the intracellular glycosylation network, beyond the processing capacity of each enzyme. In contrast to fed-batch cultivations, the cell-specific productivity of IgG generally decreases over the course of the perfusion cell culture, as shown in Figure 3. The GFA shows that the enzyme-specific factors of the aforementioned upstream enzymes stay approximately constant at roughly 1 throughout the experiments. With α_J staying near 1, the intracellular glycosylation fluxes associated with these

enzymes thus vary proportionally with the cell-specific productivity $\beta(t)$. This trend is not surprising in consideration that the amount of IgG that needs to be processed decreases over time.

The enzyme-specific factor with the most dynamical changes in the perfusion cell cultures corresponds to the galactosyltransferase (GalT), which parallels that in the fed-batch cultivations [28]. A random forest regression analysis further points to the specific productivity of IgG and the concentration of ammonia as the two variables with the highest importance in explaining the dynamic changes of IgG galactosylation (see Figure 7). The accumulation of ammonia and its inhibitory activity on GalT have previously been reported as one of the main reasons for decreasing IgG galactosylation in fed-batch cultivations [28,36]. Unlike ammonia, the connection between the IgG specific productivity and galactosylation is not immediately clear. The partial correlation between $\frac{\mathrm{d}\alpha_{GalT}(t)}{\mathrm{d}t}$ and q_{IgG} is small and negative (partial correlation = −0.037), suggesting that the relationship between the two variables as revealed by RF analysis is likely to be nonlinear. The negative partial correlation further implies that keeping all other process parameters the same, a higher q_{IgG} is concomitant with decreasing α_{GalT} with time. Such a trend is in general agreement with how α_J of the upstream enzymes vary with the cell-specific productivity of IgG as explained earlier.

Taken together, the key process variables that are associated with IgG galactosylation and the relationship among them in the perfusion cell culture resemble those in the fed-batch cultivation. Extensive data and knowledge from traditional (fed-)batch production may thus prove to be informative for understanding IgG glycosylation in continuous perfusion cell cultures. More specifically, the above finding suggests the possibility of predicting IgG glycosylation dynamics in perfusion cell cultures, for example by creating a (soft) sensor based on fed-batch data, which is a topic of great interest and importance in continuous biomanufacturing.

5. Conclusion

The application of glycosylation flux analysis to IgG production data from four perfusion CHO cell culture experiments shed light on the possible controlling factors of IgG glycosylation. The activity of galactosyltransferase and thus intracellular IgG galactosylation fluxes displayed the most time-varying changes during the perfusion cell cultivation. The GFA, coupled with a random forest regression analysis, pointed to the cell-specific productivity of IgG and the concentration of ammonia in the media as the most important process parameters for explaining the dynamical alterations in GalT activity. These observations parallel our previous report on IgG glycosylation in fed-batch production. The similarity between fed-batch and perfusion cell culture, especially in how IgG galactosylation flux activity varies with the cell culture parameters, suggests the possibility of using extensive data and knowledge from (fed-)batch production to predict the dynamics of IgG glycosylation in the continuous perfusion cell cultures.

Supplementary Materials: The following are available online at www.mdpi.com/link, Table S1. Summary of the reactor set points and the corresponding measured values, Table S2. Glycan structures in glycosylation network, Figure S1. Process measurements of Experiment A, Figure S2. Process measurements of Experiment B, Figure S3. Process measurements of Experiment C, Figure S4. Process measurements of Experiment D, Figure S5. Secretion flux fitting of other IgG glycoforms, Figure S6. Residuals of random forest regression.

Acknowledgments: The authors would like to thank Ernesto Scibona and Thomas Villiger for their help and explanations regarding the glycosylation process.

Author Contributions: S.H., M.W., M.M. and R.G. conceived and designed the study; S. H. performed the computational analysis; M.W. performed the wet lab experiments; N.P.G. performed the random forest regression analysis; S.H., D.L., T.S. and R.G. developed the improved GFA; M.M. supervised the wet lab experiments; S.H., M.W and R.G. wrote the paper.

Conflicts of Interest: No conflicts of interest are declared

References

1. Ecker, D. M.; Jones, S. D.; Levine, H. L. The therapeutic monoclonal antibody market. *MAbs* **2015**, *7*, 9–14.
2. Kelley, B. Industrialization of MAb production technology. *MAbs* **2009**, *1*, 443–452.
3. Li, F.; Vijayasankaran, N.; Shen, A.; Kiss, R.; Amanullah, A. Cell culture processes for monoclonal antibody production. *MAbs* **2010**, *2*, 466–479.
4. Rathore, A. S. Roadmap for implementation of quality by design (QbD) for biotechnology products. *Trends Biotechnol.* **2009**, *27*, 546–553.
5. FDA Guidance for Industry. PAT - A Framework for Innovative Pharmaceutical Development, Manufacturing, and Quality Assurance. **2004**, 19.
6. Reay, D.; Ramshaw, C.; Harvey, A. *Process Intensification: Engineering for Efficiency, Sustainability and Flexibility: Second Edition*; 2013.
7. Boedeker, B. G. D. Recombinant factor VIII (Kogenate®) for the treatment of Hemophilia A: The first and only world-wide licensed recombinant protein produced in high-throughput perfusion culture. In *Modern Biopharmaceuticals*; Wiley-VCH Verlag GmbH & Co. KGaA, 2013; pp. 429–443.
8. Clincke, M. F.; Molleryd, C.; Zhang, Y.; Lindskog, E.; Walsh, K.; Chotteau, V. Very high density of CHO cells in perfusion by ATF or TFF in WAVE bioreactorTM, Part I: Effect of the cell density on the process. *Biotechnol. Prog.* **2013**, *29*, 754–767.
9. Berger, M.; Kaup, M.; Blanchard, V. Protein glycosylation and its impact on biotechnology. In *Genomics and Systems Biology of Mammalian Cell Culture*. Hu, W. S.; Zeng, A.-P., Eds.; Springer Berlin Heidelberg: Berlin, Heidelberg, 2012; pp. 165–185.
10. Aebi, M. N-linked protein glycosylation in the ER. *Biochim. Biophys. Acta - Mol. Cell Res.* **2013**, *1833*, 2430–2437.
11. Solá, R. J.; Griebenow, K. Glycosylation of therapeutic proteins: An effective strategy to optimize efficacy. *BioDrugs* **2010**, *24*, 9–21.
12. Jefferis, R. Glycosylation as a strategy to improve antibody-based therapeutics. *Nat. Rev.* **2009**, *8*, 226–34.
13. Goh, J. S. Y.; Liu, Y.; Liu, H.; Chan, K. F.; Wan, C.; Teo, G.; Zhou, X.; Xie, F.; Zhang, P.; Zhang, Y.; Song, Z. Highly sialylated recombinant human erythropoietin production in large-scale perfusion bioreactor utilizing CHO-gmt4 (JW152) with restored GnT I function. *Biotechnol. J.* **2014**, *9*, 100–109.
14. Harding, F. A.; Stickler, M. M.; Razo, J.; DuBridge, R. B. The immunogenicity of humanized and fully human antibodies: Residual immunogenicity resides in the CDR regions. *MAbs* **2010**, *2*, 256–265.
15. Matasci, M.; Hacker, D. L.; Baldi, L.; Wurm, F. M. Protein therapeutics Recombinant therapeutic protein production in cultivated mammalian cells : current status and future prospects. *Drug Discov. Today Technol.* **2008**, *5*, 37–42.
16. Jayapal, K. P.; Wlaschin, K. F.; Hu, W.-S.; Yap, M. G. S. Recombinant protein therapeutics from CHO cells — 20 years and counting. *Chem. Eng. Prog.* **2007**, *103*, 40–47.
17. Wright, A.; Morrison, S. L. Effect of glycosylation on antibody function : Implications for genetic engineering. *Trends Biotechnol.* **1997**, *15*, 26–32.
18. Moremen, K. W.; Ramiah, A.; Stuart, M.; Steel, J.; Meng, L.; Forouhar, F.; Moniz, H. A.; Gahlay, G.; Gao, Z.; Chapla, D.; Wang, S.; Yang, J.-Y.; Prabhakar, P. K.; Johnson, R.; Rosa, M. dela; Geisler, C.; Nairn, A. V; Seetharaman, J.; Wu, S.-C.; Tong, L.; Gilbert, H. J.; LaBaer, J.; Jarvis, D. L. Expression system for structural and functional studies of human glycosylation enzymes. *Nat. Chem. Biol.* **2017**, *14*, 156–162.
19. Radhakrishnan, D.; Robinson, A. S.; Ogunnaike, B. A. Controlling the glycosylation profile in mAbs using time-dependent media supplementation. *Antibodies*. **2018**, *7*, 1.
20. Ivarsson, M.; Villiger, T. K.; Morbidelli, M.; Soos, M. Evaluating the impact of cell culture process parameters on monoclonal antibody N-glycosylation. *J. Biotechnol.* **2014**, *188C*, 88–96.
21. Varki, A. Biological roles of oligosaccharides : all of the theories are correct. *Glyco* **1993**, *3*, 97–130.
22. Umaña, P.; Bailey, J. E. A Mathematical model of N-linked glycoform biosynthesis. *Biotechnol. Bioeng.* **1997**, *55*, 890–908.
23. Jimenez del Val, I.; Nagy, J. M.; 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.

24. Jiménez del Val, I.; Constantinou, A.; Dell, A.; Haslam, S.; Polizzi, K. M.; Kontoravdi, C. A quantitative and mechanistic model for monoclonal antibody glycosylation as a function of nutrient availability during cell culture. *BMC Proc.* **2013**, *7*, O10.
25. Jedrzejewski, P. M.; del Val, I. J.; Constantinou, A.; Dell, A.; Haslam, S. M.; Polizzi, K. M.; Kontoravdi, C. Towards controling the glycoform: A model framework linking extracellular metabolites to antibody glycosylation. *Int. J. Mol. Sci.* **2014**, *15*, 4492–4522.
26. Spahn, P. N.; Hansen, A. H.; Henning, G.; Arnsdorf, J.; Kildegaard, H. F.; Lewis, N. E.; Arnsdorf, J.; Kildegaard, H. F.; Lewis, N. E.; Markov, A. A markov chain model for N-linked protein glycosylation – towards a low-parameter tool for model-driven. *Metab. Eng.* **2015**, 1–15.
27. Spahn, P. N.; Hansen, A. H.; Kol, S.; Voldborg, B.; Lewis, N. E. Predictive glycoengineering of biosimilars using a Markov chain glycosylation model. *Biotechnol. J.* **2017**, *12*, 1–8.
28. Hutter, S.; Villiger, T. K.; Brühlmann, D.; Stettler, M.; Broly, H.; Soos, M.; Gunawan, R. Glycosylation flux analysis reveals dynamic changes of intracellular glycosylation flux distribution in Chinese hamster ovary fed-batch cultures. *Metab. Eng.* **2017**, *43*, 9–20.
29. Antoniewicz, M. R. Methods and advances in metabolic flux analysis: A mini-review. *J Ind Microbiol Biotechnol* **2015**, *42*.
30. Hossler, P.; Goh, L.-T.; Lee, M. M.; Hu, W.-S. GlycoVis: Visualizing glycan distribution in the protein N-glycosylation pathway in mammalian cells. *Biotechnol. Bioeng.* **2006**, *95*, 946–960.
31. Wolf, M. Development and optimization of mammalian cell perfusion cultures for continuous biomanufacturing. PhD Thesis. ETH Zurich, Zurich, Switzerland. (in press)
32. Karst, D. J.; Serra, E.; Villiger, T. K.; Soos, M.; Morbidelli, M. Characterization and comparison of ATF and TFF in stirred bioreactors for continuous mammalian cell culture processes. *Biochem. Eng. J.* **2016**, *110*, 17–26.
33. Egea, J. A.; Henriques, D.; Cokelaer, T.; Villaverde, A. F.; MacNamara, A.; Danciu, D.-P.; Banga, J. R.; Saez-Rodriguez, J. MEIGO: An open-source software suite based on metaheuristics for global optimization in systems biology and bioinformatics. *BMC Bioinformatics* **2014**, *15*, 136.
34. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32.
35. Villiger, T. K.; Scibona, E.; Stettler, M.; Broly, H.; Morbidelli, M.; Soos, M. Controlling the time evolution of mAb N-linkedglycosylation - Part II: Model-based predictions. *Biotechnol. Prog.* **2016**.
36. Gawlitzek, M.; Ryll, T.; Lofgren, J.; Sliwkowski, M. B. Ammonium alters N-glycan structures of recombinant TNFR-IgG: Degradative versus biosynthetic mechanisms. *Biotechnol. Bioeng.* **2000**, *68*, 637–646.