



## Adaptive nonlinear model predictive control of monoclonal antibody glycosylation in CHO cell culture

Yingjie Ma  <sup>a</sup>, Jing Guo  <sup>a,b</sup>, Alexis B. Dubs  <sup>a</sup>, Krystian Ganko  <sup>a</sup>, Richard D. Braatz  <sup>a,\*</sup>

<sup>a</sup> Massachusetts Institute of Technology, Cambridge, MA, 02139, USA

<sup>b</sup> Polytechnique Montréal, Montréal, QC, H3T 0A3, Canada

### ARTICLE INFO

#### Keywords:

Glycosylation  
Adaptive nonlinear model predictive control  
Multiscale model  
CHO cell culture

### ABSTRACT

N-glycosylation is a critical quality attribute of monoclonal antibodies (mAbs), the dominant class of biopharmaceuticals. Controlling glycosylation remains difficult due to intrinsic pathway complexity, limited online measurements, and a lack of tailored control strategies. This work applies an adaptive nonlinear model predictive control (ANMPC) framework to a fed-batch mAb production process, using a multiscale model that links extracellular conditions to intracellular Golgi reactions to predict glycan profiles. Model parameters are updated online as new measurements arrive, after which a shrinking-horizon optimization computes the control inputs; only the first control move is implemented each cycle. Case studies show that, with a minimal day-1 galactose excitation, ANMPC mitigates model-plant mismatch and achieves up to 130% and 96% higher performance than open-loop optimization and state NMPC, respectively. Under more realistic conditions (partial measurement availability and longer preparation time), ANMPC maintains comparable performance, indicating robustness to practical limitations. Overall, the results demonstrate that ANMPC can actively shape glycan distributions in silico and offers a viable path toward closed-loop control of mAb glycosylation.

### 1. Introduction

Monoclonal antibodies (mAbs) are among the highest-revenue biopharmaceuticals, representing US \$217 billion of the global biopharmaceutical market (US \$343 billion) in 2021 (Walsh & Walsh, 2022). They are widely administered to treat cancer, infectious diseases, and inflammatory disorders (Mullard, 2021). The N-linked glycans attached to the Fc region of a mAb profoundly influence its bioactivity and therapeutic efficacy (Jefferis, 2009; Majewska et al., 2020). Because the glycan distribution must remain within tight limits, regulatory guidelines classify glycosylation as a critical quality attribute (CQA) for mAb products (Batra & Rathore, 2016). Variations in core fucosylation and terminal galactosylation, for example, directly modulate antibody-dependent cellular cytotoxicity (ADCC) and complement-dependent cytotoxicity (CDC)—the primary mechanisms through which mAbs destroy target cells (James, 2005; Matsumiya et al., 2007; Sha et al., 2016). Conversely, an increased proportion of high-mannose glycans accelerates serum clearance and reduces therapeutic potency (Goetze et al., 2011).

Although essential for mAb bioactivity and therapeutic efficacy, achieving a consistent batch-to-batch glycan profile remains a formidable challenge for biopharmaceutical manufacturers (Federici et al., 2013; McCamish & Woollett, 2013; Michaela M. et al., 2014a).

Chinese hamster ovary (CHO) cells are the industry workhorse for mAb production because they perform complex post-translational modifications and secrete proteins that are both human-compatible and biologically active (Kotidis et al., 2019). The CHO-based production workflow comprises a sequence of template-driven steps—DNA transcription, mRNA translation—followed by protein folding and glycosylation, which, unlike the earlier stages, do not rely on a nucleic acid template (Michaela M. et al., 2014b; Spahn et al., 2016). Instead, glycosylation emerges from a non-template enzymatic network of thousands of coupled reactions in the endoplasmic reticulum and Golgi apparatus, all modulated by diverse intracellular factors (Kornfeld & Kornfeld, 1985). Consequently, glycan biosynthesis and conjugation to the antibody frequently yield a heterogeneous mixture of glycoforms (Michaela M. et al., 2014b). Overall, the inherent complexity of the glycosylation process, the lack of robust real-time glycan analytics, and the absence of dedicated control strategies have so far limited the implementation of online glycosylation control (Michaela M. et al., 2014b).

Most published control strategies for mAb bioprocesses have targeted the productivity of mAb or the concentrations of extracellular metabolites (e.g., glucose, glutamine, lactate) rather than the glycan profile itself (Dan et al., 2025; Dewasme et al., 2015; Lu et al., 2013; Sarna et al., 2023; Sauer et al., 2000). The emphasis on productivity reflects

\* Corresponding author.

E-mail addresses: [yingjima@mit.edu](mailto:yingjima@mit.edu) (Y. Ma), [jing.guo@polymtl.ca](mailto:jing.guo@polymtl.ca) (J. Guo), [adubs@mit.edu](mailto:adubs@mit.edu) (A.B. Dubs), [kkganko@mit.edu](mailto:kkganko@mit.edu) (K. Ganko), [braatz@mit.edu](mailto:braatz@mit.edu) (R.D. Braatz).

the historical absence of mechanistic models that couple extracellular conditions to the intracellular glycosylation network—yet maximizing titer (mAb concentration) can be at odds with meeting a specified glycan distribution. Seeking to reconcile this tradeoff, Wong et al. (2005) investigated how dynamic feed profiles influence N-glycan quality, but their policies were still framed in terms of holding nutrient concentrations at preset levels rather than directly steering the glycosylation state space. Reference (Zupke et al., 2015) introduced a model predictive control (MPC) scheme to maintain a target high-mannose (HM) fraction, but the underlying model only comprised a single manipulated variable and a single glycoform, limiting its ability to balance yield enhancement with multi-attribute quality specifications.

The emergence of multiscale mechanistic models that explicitly link extracellular metabolism to intracellular glycosylation has enabled high-fidelity open-loop optimizations capable of simultaneously boosting product titer and shaping the glycan profile (Kappatou et al., 2020; Kotidis et al., 2019). Reference (Kotidis et al., 2019) validated their optimization experimentally, achieving an over 90% increase in the desired glycoform and underscoring the promise of this approach. Nevertheless, both studies generated feed trajectories offline; such open-loop policies are vulnerable to model–plant mismatch and unanticipated disturbances (Rawlings et al., 2017). Embedding the resulting partial-differential-algebraic equation (PDAE) models directly in a nonlinear MPC (NMPC) framework remains computationally prohibitive—a single optimization may require 20 h (Ma et al., 2025) to 40 h (Kappatou et al., 2020). To reduce this burden, several groups have linearized the multiscale model into a process-gain matrix and applied the resulting controller in fed-batch (Luo et al., 2023) and perfusion bioreactors (Maloney, 2021), but such linearizations are accurate only near the operating point used for their construction (Michaela M. et al., 2014b), limiting their robustness for the wide excursions typical of industrial fed-batch or disturbance-prone perfusion processes.

A second challenge is parameter identification for the multiscale model. Adaptive MPC techniques—updating model parameters online from process data—have been used to address model uncertainty with promising results (Hajizadeh et al., 2018; Jabarivelisdeh et al., 2020; Pickhardt, 2000). Motivated by these gaps, we develop an adaptive NMPC (ANMPC) controller for a fed-batch mAb bioreactor using the multiscale PDAE glycosylation model of Kotidis et al. (2019). At each sampling time, ANMPC updates model parameters using all available measurements and then recomputes the control moves via model-based optimization. To solve the resulting dynamic optimization and parameter estimation problems efficiently and robustly, we employ control-vector parameterization (CVP) with embedded simulations accelerated by a parallel quasi-steady-state (QSS) approach proposed in our previous work (Ma et al., 2025). Case studies initialized from multiple parameter sets compare open-loop optimization, state NMPC, and ANMPC; ANMPC mitigates model–plant mismatch and achieves up to 130% higher penalized merit than the alternatives. We further evaluate ANMPC under more realistic measurement availability (no NSD measurements) and with a 4 h analytics/actuation delay, observing no material performance loss, which demonstrates robustness.

In what follows, Section 2 describes the multiscale glycosylation model; Section 3 formulates the control problem; Section 4 presents the ANMPC framework; Section 5 examines actuator choices and demonstrates the algorithm through case studies; and Section 6 concludes.

## 2. Glycosylation model

Fig. 1 illustrates the three-level multiscale glycosylation model employed in this work:

- (1) Bioreactor-level cell culture model – predicts viable cell density, extracellular metabolite concentrations, specific productivity, and the concentrations of secreted mAb.
- (2) Intracellular nucleotide sugar donor (NSD) synthesis model – converts the extracellular metabolite information from the culture model into intracellular NSD using a detailed synthesis pathway and associated rate laws.
- (3) Golgi-level glycosylation reaction model – receives the NSDs as sugar donors and computes the fractional distribution of glycoforms leaving the secretory pathway via a network of enzyme-catalyzed reactions coupled with glycoprotein transport equations.

The overall structure follows Kotidis et al. (2019) and is extended to incorporate manganese- and ammonia-dependent kinetics reported by Villiger et al. (2016). Information propagates downstream from the extracellular environment through the NSD synthesis layer and to the Golgi reaction network, and the resulting NSD consumption and glycoform flux then feeds back to the NSD synthesis model and cell culture model—closing the loop and capturing the two-way coupling between reactor conditions and intracellular glycosylation.

The multiscale model is formulated as a large-scale PDAE system that couples 30 ordinary differential equations (ODEs), 34 partial differential equations (PDEs), and numerous highly nonlinear algebraic relations. The remainder of this section introduces each submodel in turn.

### 2.1. Cell culture model

At the reactor scale, we employ an unstructured differential algebraic equation (DAE) model to describe cell growth, death, and metabolism (Kontoravdi et al., 2010; Kotidis et al., 2019). The dynamic balances for cell culture volume and cell populations are

$$\frac{dV}{dt} = F_{\text{in}} - F_{\text{out}} \quad (1)$$

$$\frac{d(VX)}{dt} = \mu V X_v - F_{\text{out}} X \quad (2)$$

$$\frac{d(VX_v)}{dt} = (\mu - \mu_{\text{death}}) V X_v - F_{\text{out}} X_v \quad (3)$$

where  $V$  (L) is the cell culture volume,  $F_{\text{in}}$  and  $F_{\text{out}}$  ( $\text{L h}^{-1}$ ) are the inlet and outlet flow rates,  $X$  and  $X_v$  ( $\text{cells L}^{-1}$ ) denote the total and viable cell densities, and  $\mu$  and  $\mu_{\text{death}}$  ( $\text{h}^{-1}$ ) are the specific cell growth and death rates, respectively.

The viability is defined as the fraction of living cells:

$$\text{viability} = \frac{X_v}{X} \quad (4)$$

The specific growth ( $\mu$ ) and death ( $\mu_{\text{death}}$ ) rates are determined by nutrient availability and by the buildup of inhibitory byproducts. Glucose and asparagine supply is growth-limiting, whereas lactate, ammonia, and uridine exert inhibitory effects—with ammonia and uridine also driving cell death. The rate expressions are given in Eqs. (A.1)–(A.4).

The extracellular mass balance for each metabolite is described by

$$\frac{d(V[\text{Met}])}{dt} = F_{\text{in}}[\text{Met}]_{\text{in}} - F_{\text{out}}[\text{Met}] + q_{\text{Met}} V X_v \quad (5)$$

where  $[\text{Met}]_{\text{in}}$  is the metabolite concentration in the feed stream,  $[\text{Met}]$  is its concentration in the culture, and their units are  $\text{mg L}^{-1}$  for mAb and  $\text{mmol L}^{-1}$  for the other metabolites.  $q_{\text{Met}}$  is the cell-specific production or consumption rate, which is expressed in  $\text{pg cell}^{-1} \text{h}^{-1}$  for mAb and in  $\text{mmol cell}^{-1} \text{h}^{-1}$  for the remaining metabolites. Metabolites considered in the model include ammonia (Amm), asparagine (Asn), aspartate (Asp), glucose (Glc), galactose (Gal), glutamine (Gln), glutamate (Glu), lactose (Lac), and uridine (Urd). Fig. 2 depicts the corresponding metabolic network, and the individual reaction rates are given in Eqs. (A.5)–(A.15).

The extracellular mass balance equation for each glycoform  $\text{GLY}_i$  is

$$\frac{d(V[\text{GLY}_i^{\text{extra}}])}{dt} = -F_{\text{out}}[\text{GLY}_i^{\text{extra}}] + V q_{\text{mAb}} X_v Y_i^{\text{intra}} \quad (6)$$

where  $[\text{GLY}_i^{\text{extra}}]$  ( $\text{mg L}^{-1}$ ) is the extracellular concentration of glycoform  $\text{GLY}_i$ , and  $Y_i^{\text{intra}} = [\text{GLY}_i^{\text{intra}}]/[\text{mAb}^{\text{intra}}]$  is its intracellular fractional abundance. Here, the intracellular glycoform concentration

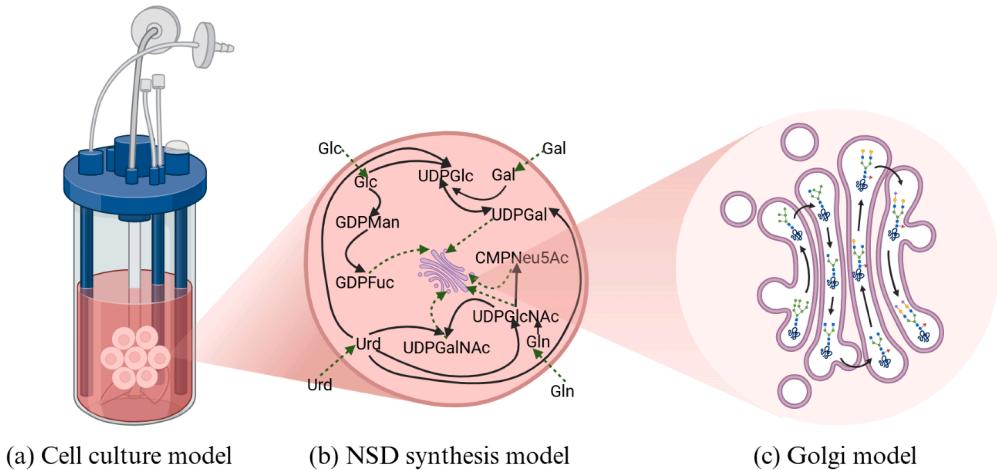


Fig. 1. Multiscale glycosylation model.

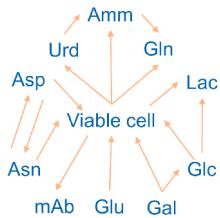


Fig. 2. The metabolic reaction network in the cell culture. The start of an arrow refers to the reactant, while the end of the arrow denotes the product.

$[GLY_i^{\text{intra}}]$  ( $\mu\text{mol L}^{-1}$ ) is obtained from the Golgi model, while the total intracellular mAb level  $[m\text{Ab}^{\text{intra}}]$  ( $\mu\text{mol L}^{-1}$ ) is approximately  $94 \mu\text{mol L}^{-1}$  (Kotidis, 2021). By analogy, the extracellular fractional abundance is  $Y_i^{\text{extra}} = [GLY_i^{\text{extra}}]/[m\text{Ab}]$ . Nine frequently observed glycans are considered in this work: HM, FA1G1, FA2G0, FA2G1, FA2G2, SIA, G0, G2, and Man5. Their full names are provided in Appendix B.

## 2.2. NSD synthesis model

NSDs provide monosaccharides for the glycosylation reactions, and they are synthesized inside the cells. Therefore, the NSD synthesis model consisting of a DAE system is used to track the intracellular concentrations of seven monosaccharide donors: guanosine diphosphate mannose (GDP-Man), guanosine diphosphate fucose (GDP-Fuc), uridine diphosphate galactose (UDP-Gal), uridine diphosphate glucose (UDP-Glc), uridine diphosphate *N*-acetylgalactosamine (UDP-GalNAc), uridine diphosphate *N*-acetylglucosamine (UDP-GlcNAc), and cytidine monophosphate *N*-acetylneurameric acid (CMP-Neu5Ac). The reaction network, depicted in Fig. 1(b), comprises  $N_{\text{NSD}}$  enzyme-catalyzed steps whose stoichiometry is encoded in the matrix  $v^{\text{nsd}}$ . For each donor  $i$  ( $i = 1, \dots, N_{\text{NSD}} = 7$ ), the mass balance is

$$\frac{d([NSD_i^{\text{intra}}])}{dt} = \sum_{j=1}^{N_{\text{R1}}} v_{i,j} r_j^{\text{nsd}} - f_{NSD_i}^{\text{hcp/lipid}} - f_{NSD_i}^{\text{precursor}} - f_{NSD_i}^{\text{glyc}} \quad (7)$$

where  $[NSD_i^{\text{intra}}]$  ( $\text{mmol L}^{-1}$ ) is the intracellular concentration of donor  $i$ ;  $r_j^{\text{nsd}}$  ( $\text{mmol L}^{-1} \text{h}^{-1}$ ) is the rate of a reaction  $j$ ; and the sink terms  $f_{NSD_i}^{\text{hcp/lipid}}$ ,  $f_{NSD_i}^{\text{precursor}}$ , and  $f_{NSD_i}^{\text{glyc}}$  ( $\text{mmol L}^{-1} \text{h}^{-1}$ ) are consumption rates of the donor for host cell protein and glycolipid synthesis, precursor oligosaccharide assembly, and N-linked glycosylation, respectively. All reaction rates follow Michaelis–Menten kinetics that depend on the intracellular NSD concentrations together with extracellular glucose,

galactose, and uridine concentrations and the intracellular glutamine level. The explicit rates are given in Eqs. (C.1)–(C.18).

## 2.3. Golgi model

The Golgi model predicts the intracellular glycoform distribution using a dynamic plug flow reactor (PFR) representation of the Golgi apparatus, formulated as a PDAE system (Jimenez del Val et al., 2011; Kotidis et al., 2019). The spatiotemporal balance for each oligosaccharide  $OS_i$  is

$$\frac{\partial [OS_i]}{\partial t} = -\text{vel}_{\text{golgi}} \frac{\partial [OS_i]}{\partial z} + \sum_{j=1}^{N_{\text{R2}}} v_{i,j} r_j, \quad i = 1, \dots, N_{\text{OS}} \quad (8)$$

where  $[OS_i]$  ( $\mu\text{mol L}^{-1}$ ) is the local concentration of  $OS_i$  in the Golgi apparatus,  $r_j$  ( $\mu\text{mol L}^{-1} \text{min}^{-1}$ ) is the rate of reaction  $j$ ,  $v_{i,j}$  is the stoichiometric coefficient of  $OS_i$  in reaction  $j$ , and  $N_{\text{R2}}$  is the total number of glycosylation reactions. The glycosylation reaction network involving 33 oligosaccharides and 43 reactions is depicted in Fig. 3 (Villiger et al., 2016).

The glycosylation reaction rates  $r_j$  fall into three mechanistic classes: Michaelis–Menten kinetics, sequential order Bi–Bi kinetics, and random-order Bi–Bi kinetics (Jimenez del Val et al., 2011; Villiger et al., 2016), and the detailed equations are shown in Eqs. (D.1)–(D.6).

## 3. Control problem

We focus on economic control of a fed-batch bioreactor—the mode most employed in industry because of its operational flexibility. The culture runs for a total duration  $T$ . Nutrient supplements are added for a short period  $\Delta t_{\text{feed}}$  once every sampling interval  $\tau = T/N$ , where  $N$  is the number of control intervals. No harvest stream is withdrawn until the batch ends at  $t = T$ .

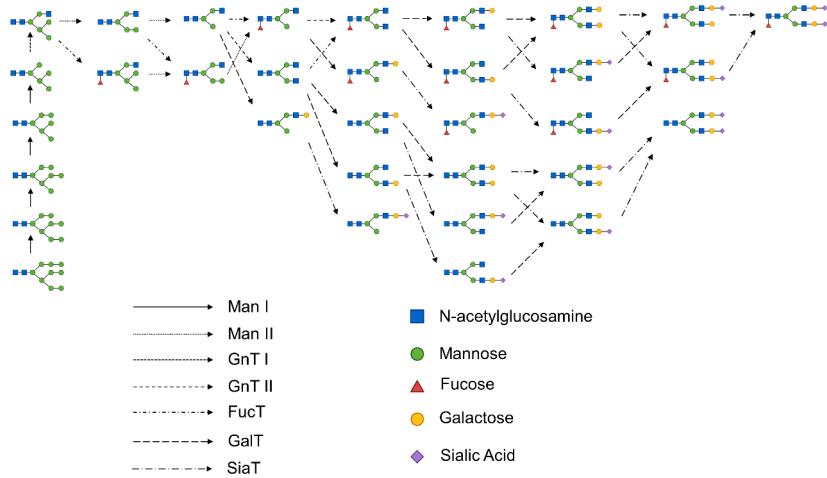
The goal is to maximize the harvest titer of the target glycoform while satisfying all process and product constraints. A widely used performance metric is the galactosylation index (GI) at harvest,

$$GI(T) = [FA2G1](T) + 2[FA2G2](T), \quad (9)$$

where the bracketed terms denote the extracellular concentrations of the indicated glycoforms. Manipulated variables  $u$  (feed flow rates) and state variables  $x$  (cell density, metabolite concentrations, glycoform percentages, etc.) are bounded by

$$u_{\text{lb}} \leq u \leq u_{\text{ub}}, \quad x_{\text{lb}} \leq x \leq x_{\text{ub}}. \quad (10)$$

The manipulated inputs are the feed stream flow rates  $F_s$  for  $s = 1, \dots, S$ . Up to three feeds can be employed—supplement medium,



**Fig. 3.** The glycosylation reaction network in the Golgi apparatus.

**Table 1**  
Compositions of the feed streams (all concentrations in mmol L<sup>-1</sup>).

Metabolite	Medium	Gal solution	Urd solution
Glc	144.37	0	0
Gln	0	0	0
Lac	0	0	0
Amm	0.06	0	0
Glu	12.19	0	0
Asn	26.99	0	0
Asp	51.95	0	0
Gal	0	3600	0
Urd	0	0	2000

galactose solution, and uridine solution—so  $0 \leq S \leq 3$ . The composition of each stream is listed in **Table 1**.

In an ideal scenario-perfect model, exact initial state, and no disturbances—the economic optimization of the fed-batch process reduces to an offline dynamic optimization (DO-offline),

#### (DO-offline)

$$\max_{x(t), u(t)} \text{GI}(T) \quad (11)$$

$$\text{s.t. } x(t) = \text{glyco}(u(t), p), \quad t \in [0, T], \quad (12)$$

$$x(0) = x_0, \quad (13)$$

$$x_{lb}^{\text{path}} \leq x(t) \leq x_{ub}^{\text{path}}, \quad t \in (0, T), \quad (14)$$

$$x_{lb}^{\text{terminal}} \leq x(T) \leq x_{ub}^{\text{terminal}}, \quad (15)$$

$$u_{lb}(t) \leq u(t) \leq u_{ub}(t), \quad t \in [0, T]. \quad (16)$$

where  $\text{glyco}(u, p)$  denotes the implicit PDAE model parameterized by  $p$  (the glycosylation model parameters);  $x_0$  represents the initial values of the state variables;  $x^{\text{path}}$  and  $x^{\text{terminal}}$  are path and terminal constraint bounds; and the input bounds  $u_{lb}(t)$  and  $u_{ub}(t)$  are time-dependent, allowing, for example, zero flow rates during hold periods.

In practice, the ideal assumptions rarely hold. We address model-plant mismatch by embedding parameter adaptation within the NMPC framework (ANMPC), while the remaining assumptions hold given careful pre-start measurements and carefully controlled bolus feeding.

To evaluate the closed-loop performance of ANMPC, we inject measurement noise into simulated data using the standard deviations (stds) computed from the dataset of [Kotidis et al. \(2019\)](#). Two measurement scenarios are considered:

(1) **Full measurement** – all state variables are available;

**Table 2**  
Measurement variables and standard deviations in the cell culture model.

Variables	Std	Variables	Std
$V$ (L)	0.001	$c_{\text{Glu}}$ (mmol L <sup>-1</sup> )	0.22
$X$ (cells $\mu\text{L}^{-1}$ )	494.66	$c_{\text{Gal}}$ (mmol L <sup>-1</sup> )	3.80
$X_v$ (cells $\mu\text{L}^{-1}$ )	494.66	$c_{\text{Urd}}$ (mmol L <sup>-1</sup> )	0.87
$c_{\text{Glc}}$ (mmol L <sup>-1</sup> )	1.45	$c_{\text{Asn}}$ (mmol L <sup>-1</sup> )	0.22
$c_{\text{Gln}}$ (mmol L <sup>-1</sup> )	0.11	$c_{\text{Asp}}$ (mmol L <sup>-1</sup> )	0.84
$c_{\text{Lac}}$ (mmol L <sup>-1</sup> )	0.48	$c_{\text{mAb}}$ (mmol L <sup>-1</sup> )	11.05
$c_{\text{Amm}}$ (mmol L <sup>-1</sup> )	0.17		

**Table 3**

Measurement variables and standard deviations in the NSD model (all concentrations/Std are in mmol L<sup>-1</sup>).

Variables	Std	Variables	Std
$c_{\text{UDP-GlcNAc}}$	0.25	$c_{\text{UDP-Glc}}$	0.22
$c_{\text{GDP-Fuc}}$	0.004	$c_{\text{GDP-Man}}$	0.003
$c_{\text{UDP-Gal}}$	0.18	$c_{\text{UDP-GalNAc}}$	0.03
$c_{\text{CMP-Neu5Ac}}$	0.005		

**Table 4**

Measurement variables and standard deviations in the Golgi model (all variables/Std are in %).

Variables	Std	Variables	Std
$y_{\text{HM}}^{\text{extra}}$	0.64	$y_{\text{SIA}}^{\text{extra}}$	0.12
$y_{\text{FA1G1}}^{\text{extra}}$	0.15	$y_{\text{G0}}^{\text{extra}}$	0.42
$y_{\text{FA2G0}}^{\text{extra}}$	1.05	$y_{\text{G2}}^{\text{extra}}$	0.49
$y_{\text{FA2G1}}^{\text{extra}}$	0.99	$y_{\text{Man5}}^{\text{extra}}$	0.90
$y_{\text{FA2G2}}^{\text{extra}}$	0.82		

(2) **Industry-realistic measurement** – NSDs and certain metabolites (Asn, Asp) and glycoforms (HM, FA1G1, SIA) remain unobserved, reflecting current analytical limitations.

The potential measurement variables and their standard deviations are summarized in **Tables 2–4**.

#### 4. Adaptive NMPC (ANMPC) framework

The adaptive NMPC loop iteratively updates model parameters, state estimates and future control actions using online measurements. [Algorithm 1](#) summarizes the procedure:

**Table 5**  
Parameters in the ANMPC algorithm for the glycosylation process control problem.

Parameter	Description	Value
$T$ (h)	Total culture time	288
$V_0$ (L)	Initial working volume	1.5
$\Delta t_{\text{feed}}$ (h)	Feed window length	0.01
$\Delta t_{\text{prep}}$ (h)	Preparation time	2 in Sections 5.2–5.3; 4 in Sections 5.4–5.5
$\tau$ (h)	Length of each control interval	24
$N$ (–)	Number of control intervals	12

### Algorithm 1 ANMPC algorithm.

#### Initialization

Set counter  $k = 0$ , current time  $t = 0$ . Specify: total batch duration  $T$ ; number of control intervals  $N$  and length  $\tau = T/N$ ; feed window  $\Delta t_{\text{feed}}$ ; preparation time  $\Delta t_{\text{prep}}$ .

Times:  $t_j = j\tau$  ( $j = 0, \dots, N$ ),  $t_j^{\text{end}} = j\tau + \Delta t_{\text{feed}}$  ( $j = 0, \dots, N-1$ ),  $t_j^{\text{sample}} = j\tau - \Delta t_{\text{prep}}$  ( $j = 1, \dots, N$ ).

Initialize parameters  $\hat{p}(0)$ .

#### Step 1: Parameter estimation

If  $k \geq 1$  and the  $k$ th sample has been analyzed, update  $\hat{p}(k)$  using all the available measurements  $\{y(j)\}_{j=1}^k$ ; otherwise skip to Step 3.

#### Step 2: State estimation

Simulate the process model from  $t = 0$  to  $t_k$  with  $p = \hat{p}(k)$  and obtain the current state estimate  $\hat{x}(k)$ .

#### Step 3: Dynamic optimization

Solve the dynamic optimization problem (soft-DO- $k$ ) over  $t_k \leq t \leq T$  with parameters  $\hat{p}(k)$  and initial condition  $\hat{x}(k)$ , and generate the optimal control sequence  $u = [u(k)^T, u(k+1)^T, \dots, u(N-1)^T]^T$ .

**Step 4: Bioreactor operation & sampling** Implement  $u(k)$  over  $t_k \leq t \leq t_k^{\text{end}}$ , and acquire a new sample at  $t_{k+1}^{\text{sample}}$  with the analysis results of  $y(k+1)$ .

If  $t_{k+1} < T$ , set  $k \leftarrow k + 1$  and return to Step 1; otherwise proceed to Step 5.

#### Step 5: Harvest

Shut down the bioreactor and harvest the product.

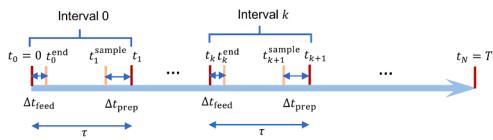


Fig. 4. Timeline of the ANMPC algorithm.

The timeline of the ANMPC algorithm is shown in Fig. 4. The preparation time  $\Delta t_{\text{prep}}$  in ANMPC must exceed the combined duration of analytics, parameter estimation, state estimation, and optimization to guarantee real-time feasibility. This relies on the efficient optimization algorithm introduced later. For the case studies considered here, the algorithm parameters are listed in Table 5.

The next subsections detail the dynamic optimization, parameter estimation, and state estimation components of the ANMPC.

#### 4.1. Dynamic optimization formulation

To generate the control actions at  $t_k$  in Step 3 of ANMPC, the controller solves the finite-horizon optimization (DO- $k$ ):

##### (DO- $k$ )

$$\max_{x(t), u(t)} \text{GI}(T) \quad (17)$$

$$\text{s.t. } x(t) = \text{glyco}(u(t), \hat{p}(k)), \quad t \in [0, T], \quad (18)$$

$$x(k) = \hat{x}(k), \quad (19)$$

$$x_{\text{lb}}^{\text{path}} \leq x(t) \leq x_{\text{ub}}^{\text{path}}, \quad t \in (0, T), \quad (20)$$

$$x_{\text{lb}}^{\text{terminal}} \leq x(T) \leq x_{\text{ub}}^{\text{terminal}}, \quad (21)$$

$$u_{\text{lb}}(t) \leq u(t) \leq u_{\text{ub}}(t), \quad t \in [t_k, T]. \quad (22)$$

Relative to the offline problem (DO-offline), the model parameters  $p$  are replaced by their current estimates  $\hat{p}(k)$  in the problem (DO- $k$ ), the initial conditions are substituted by the latest state estimate  $\hat{x}(k)$ , and the horizon is shortened to  $[t_k, T]$  because of the shrinking horizon in the fed-batch operation (Eaton & Rawlings, 1990; Nagy & Braatz, 2003). It is expected that as data accumulate,  $\hat{p}(k)$  and  $\hat{x}(k)$  move closer to the ground-truth values, the optimization problem (DO- $k$ ) will yield progressively better predictions than the static offline optimization that uses  $\hat{p}(0)$ .

Because parameter uncertainty can cause previously computed inputs to be overly aggressive, DO- $k$  may become infeasible even when DO-0 was feasible. To reduce infeasibility in the optimization, we employ a soft-constraint formulation (Scokaert & Rawlings, 1999),

##### soft-DO- $k$ :

$$\max_{\substack{x(t), u(t), \\ \bar{w}, \underline{w}, \bar{v}, \underline{v}}} \text{GI}(T) - \int_{t_k}^T \rho_w^T (\bar{w} + \underline{w}) dt - \rho_v^T (\bar{v} + \underline{v}) \quad (23)$$

s.t. Eqs. (18)–(19)

$$x_{\text{lb}}^{\text{hard}, \text{path}} \leq x^{\text{hard}}(t) \leq x_{\text{ub}}^{\text{hard}, \text{path}}, \quad t \in (0, T) \quad (24)$$

$$x_{\text{lb}}^{\text{hard}, \text{terminal}} \leq x^{\text{hard}}(T) \leq x_{\text{ub}}^{\text{hard}, \text{terminal}} \quad (25)$$

$$x_{\text{lb}}^{\text{soft}, \text{path}} - \underline{w} \leq x^{\text{soft}}(t) \leq x_{\text{ub}}^{\text{soft}, \text{path}} + \bar{w}, \quad t \in (0, T) \quad (26)$$

$$x_{\text{lb}}^{\text{soft}, \text{terminal}} - \underline{v} \leq x^{\text{soft}}(T) \leq x_{\text{ub}}^{\text{soft}, \text{terminal}} + \bar{v} \quad (27)$$

$$\bar{w}, \underline{w}, \bar{v}, \underline{v} \geq 0 \quad (28)$$

$$u_{\text{lb}}(t) \leq u(t) \leq u_{\text{ub}}(t), \quad t \in [t_k, T] \quad (29)$$

Here the state constraints are partitioned into hard constraints, which must always be respected, and soft constraints, which may be violated at the cost of penalty weights  $\rho_w$  (path) and  $\rho_v$  (terminal). The variables  $\bar{w}(t)$  and  $\underline{w}(t)$  are slack variables for the soft path constraints, whereas  $\bar{v}$  and  $\underline{v}$  are slack variables for the soft terminal constraints. Problem (soft-DO- $k$ ) is solved only if the original problem (DO- $k$ ) is infeasible.

Table 6 summarizes the constrained variables along with their permissible ranges and associated penalty weights. The lower bound of 5% for Man5 is based on Pacis et al. (2011). The penalty weights are chosen based on the orders of magnitude of the three variables in the merit function: GI is on the order of 100, viability (in %) is on the order of 10, and Man5 (in %) is on the order of 1. To ensure that each term has a comparable order of magnitude, the penalty weights of the viability and Man5 constraints are set to 10 and 100, respectively.

#### 4.2. Parameter estimation formulation

The multiscale model contains ~100 kinetic and transport parameters—far too many to calibrate accurately from the limited data available in most cases. Consequently, the parameters are adapted online using every measurement collected to date, i.e., Step 1 in

**Table 6**  
Constraint bounds, types, and penalty weights.

Variable	Description	Constraint type	Lower bound	Upper bound	Penalty weight
viability (%)	Cell viability	Terminal, soft	60	–	10
$Y_{\text{Man5}}^{\text{extra}}$ (%)	Extracellular Man5 percentage	Terminal, soft	–	5	100
$V$ (L)	Working volume	Terminal, hard	0.75	2.25	–
$F_s$ (L h <sup>-1</sup> ) $s = 1, 2, 3$	Feed flow rate	Input	0	100	–

ANMPC. Furthermore, the simulation with the updated parameters is used for state estimation, i.e., Step 2 in ANMPC.

We explore two windowing strategies for the online parameter estimation problem—expanding horizon estimation (EHE) and full horizon estimation (FHE). In the fed-batch process, the former is more natural as the horizon with available data becomes longer over time, so we are interested in the expanded horizon. In EHE, once the  $k$ th sample at  $t_k^{\text{sample}}$  ( $k \geq 1$ ) is available, we solve the problem

**EHE- $k$ :**

$$\min_{\hat{p}(k), x(t)} \sum_{j=1}^k (y(j) - \hat{y}(j))^T V_{\epsilon}(j)^{-1} (y(j) - \hat{y}(j)) + (\hat{p}(k) - \hat{p}(0))^T P (\hat{p}(k) - \hat{p}(0)) \quad (30)$$

$$\text{s.t. } x(t) = \text{glyco}(u(t), \hat{p}(k)), \quad t \in [0, t_k^{\text{sample}}], \quad (31)$$

$$x(0) = x_0, \quad (32)$$

$$p_{\text{lb}} \leq \hat{p}(k) \leq p_{\text{ub}}. \quad (33)$$

where  $y(j)$  and  $\hat{y}(j)$  are the measured and predicted outputs,  $V_{\epsilon}(j)$  is the measurement-noise covariance, and  $P$  penalizes deviation from the prior parameter vector  $\hat{p}(0)$ . When  $P = V_p^{-1}$  (the prior covariance matrix), we obtain a maximum a posteriori (MAP) estimator (e.g., [Gunawan et al., 2003](#)); when  $P = 0$ , we get a maximum likelihood estimator (MLE) (e.g., [Beck & Arnold, 1977](#)).

EHE may overfit early data and predict physically impossible trajectories (e.g., negative concentrations) in the later horizon, jeopardizing feasibility of the subsequent NMPC problem. To reduce overfitting, we instead fit the parameters over the entire horizon  $[0, T]$  even though data exist only up to  $t_k^{\text{sample}}$ , i.e., full-horizon estimation (FHE). In this way, parameters causing infeasible simulation beyond the measurement window are automatically discarded when solving the parameter estimation problem iteratively. To further protect against unphysical states, upper and lower bounds on the states can be incorporated into the estimation problem according to domain knowledge, resulting in the constrained FHE estimator

**cons-FHE- $k$ :**

$$\min_{\hat{p}(k), x(t)} \text{Eq. (30)} \quad (34)$$

$$\text{s.t. } x(t) = \text{glyco}(u(t), \hat{p}(k)), \quad t \in [0, T],$$

$$\text{Eqs. (32) and (33),}$$

$$x_{\text{lb}} \leq x(t) \leq x_{\text{ub}}, \quad t \in [0, T]. \quad (35)$$

There are two points to mention for the implementation of the online estimator:

- In the online parameter adaptation step, distinct estimators are applied to different submodels. Parameters in the cell culture submodel are updated with a MAP estimator ( $P = V_p^{-1}$  in Eq. (30)), whereas parameters in the NSD and Golgi submodels are fitted together by MLE ( $P = 0$  in Eq. (30)). Alternative combinations were tested but proved inferior: applying MLE to the cell culture block led to severe overfitting in the early phase of the run, which in turn drove the simulation and subsequent optimization to diverge; conversely, employing MAP for the NSD and Golgi blocks introduced an overly strong prior that resulted in underfitting and degraded ANMPC performance. Fitting the NSD and Golgi models separately also produced underfitting and was therefore abandoned.

- Because the parameters span several orders of magnitude, the optimization is performed in log-space for self-normalization; that is, we estimate  $\ln \hat{p}(k)$  rather than  $\hat{p}(k)$ . Without this approach, parameter estimation often fails due to divergence in the sensitivity-equation evaluations, where the Jacobian becomes highly ill-conditioned.

#### 4.3. Solution of the optimization problems

The above dynamic optimization and parameter estimation both require repeated solutions of optimization problems subject to the multiscale glycosylation model. We adopt the control vector parameterization (CVP) method for the optimizations ([Kraft, 1985; Vassiliadis et al., 1994](#)), which is more reliable than simultaneous methods for strongly nonlinear dynamics-constrained problems ([Hong et al., 2006](#)). In CVP, however, the large PDAE system described in [Sections 2](#) and its sensitivity equations must be solved tens or hundreds of times with an implicit time-stepping scheme, which dominates the computational cost.

To make the optimization tractable, we accelerate each model evaluation with the parallel QSS algorithm introduced in our earlier work, which reduces overall simulation/optimization time by two orders of magnitude ([Ma et al., 2025](#)). In the multiscale model, both the cell culture submodel and the NSD submodel consist solely of DAE equations, while the Golgi submodel involves PDEs in both the temporal and spatial domains (along the length of the Golgi apparatus), making the entire model a PDAE, which is computationally expensive. The QSS approach decouples the simulations in the temporal and spatial domains by exploiting the clear timescale separation: the Golgi reactions reach equilibrium much faster than the outer-layer dynamics. As a result, the Golgi submodel achieves a local steady state long before extracellular conditions change appreciably. Under steady-state conditions, the Golgi submodel simplifies to a DAE (in the spatial domain along the length of the Golgi apparatus), as opposed to the original PDAE. By connecting a series of quasi-steady states of the Golgi submodel at different time points, we can obtain the trajectories of its variables. Finally, the extracellular glycan percentages ( $Y_i^{\text{extra}}$ ) are derived by integrating the cell culture model, using the intracellular glycan percentages ( $Y_i^{\text{intra}}$ ) as time-varying inputs. The complete QSS algorithm is illustrated by [Fig. 5](#) and described as [Algorithm 2](#).

#### Algorithm 2 QSS algorithm.

**Initialization.** Choose a set of time points  $\mathcal{T} = \{t_s\}_{s=1}^{N_{\text{QSS}}}$ .

**Step 1: Get env.** Integrate the cell-culture model (omitting glycoprotein balances) and the NSD model to obtain the environmental states  $\text{env}(t_s)$  at each  $t_s$ .

**Step 2: Get  $Y_i^{\text{intra}}$ .** For every  $t_s$ , solve the steady-state Golgi DAE with  $\text{env}(t_s)$  as inputs, yielding intracellular glycoform fractions  $Y_i^{\text{intra}}(t_s)$ . The  $N_{\text{QSS}}$  Golgi simulations are fully multithread parallelizable.

**Step 3: Get  $Y_i^{\text{extra}}$ .** Reintegrate the complete cell culture model to produce the extracellular glycoform trajectories  $Y_i^{\text{extra}}(t_s)$ , where all the  $Y_i^{\text{intra}}(t_s)$  in the model are treated as time-varying parameters and are obtained from Step 2.

The number and allocation of the time points in  $\mathcal{T}$  determine the accuracy of the trajectories of the intracellular variables. More time points potentially increase the accuracy, but increasing the number of DAEs for

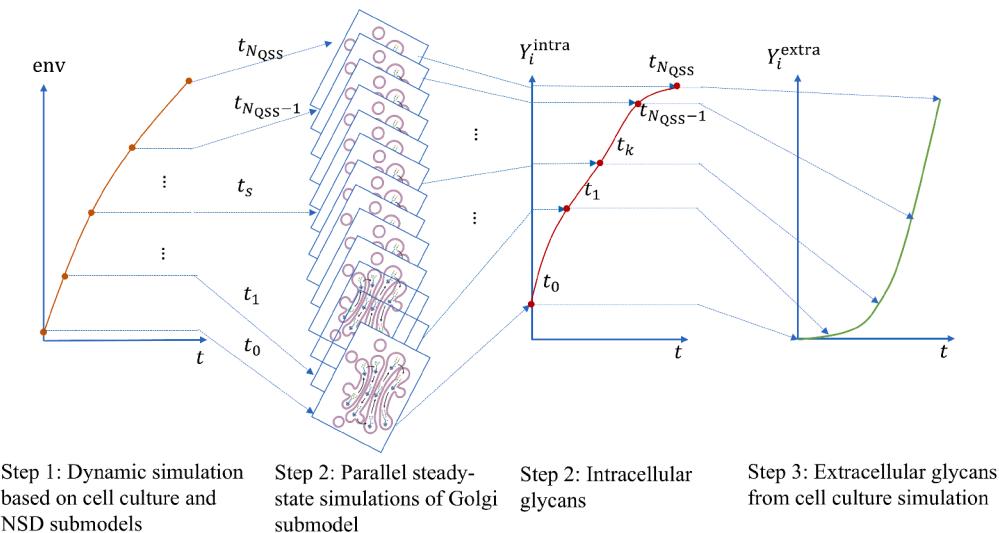


Fig. 5. Quasi-steady-state simulation method for the multiscale glycosylation model.

steady-state Golgi submodel to solve and increasing computational cost. Following Ma et al. (2025), we achieve satisfactory accuracy for  $Y_i^{\text{extra}}$  when the QSS time grid  $\mathcal{T}$  includes all event instants (feed start/stop, sampling, etc.) and is augmented with 100 uniformly spaced points over  $[0, T]$ .

## 5. Case study

To evaluate the proposed ANMPC framework, we first calibrate the multiscale model to the dataset of Kotidis et al. (2019), obtaining a MAP parameter vector  $\ln p^*$  and its covariance  $V_{\ln(p)}$ . We treat  $\ln p^*$  as the ground truth. Five distinct initial guesses  $\ln \hat{p}(0)$  are then generated on the boundary of the 50% confidence ellipsoid defined by  $V_{\ln(p)}$  so that the conclusions based on the case studies are more reliable. These  $\ln \hat{p}(0)$  vectors, together with  $V_{\ln(p)}$ , serve as the prior mean and covariance in the online estimation step. The ground-truth and initial parameter values are listed in Tables S2–S4, while the covariances for submodels are provided in the Excel file in Supplementary 2. For convenience, we denote the five initial parameter sets as `Init_0`, `Init_1`, `Init_2`, `Init_3`, and `Init_4`. The lower and upper bounds of the parameters are set to 1/1000 and 1000 times the ground-truth values, respectively.

All computations are performed on a Windows 11 machine with a 12th-Generation Intel Core i7-12700H CPU (14 cores/20 threads, 2.30 GHz) and 16 GB RAM. Python 3.9 (Van Rossum & Drake, 2009) and CasADI 3.6.5 (Andersson et al., 2019) were used as the computation platform. The nonlinear programs (NLPs) arising from the CVP formulation are solved with PySQP, an in-house sequential quadratic programming (SQP) solver based on the I-SQP algorithm (Ma et al., 2024) with a watchdog-technique enhancement. The DAE systems in the QSS algorithm are integrated with the IDAS solver (Gardner et al., 2022; Hindmarsh et al., 2005) via CasADI, using 15 threads for parallel computation. Under these settings, the total runtime per control move (parameter estimation + state estimation + dynamic optimization) is less than 2 h in all case studies. In Sections 5.2 and 5.3, we assume rapid analytics and set the preparation time  $\Delta t_{\text{prep}} = 2$  h; in Sections 5.4 and 5.5, we use a more realistic  $\Delta t_{\text{prep}} = 4$  h to account for assay time.

### 5.1. Open-loop optimization study

Because supplement medium is routinely fed whereas Gal and Urd are not, we first solve the open-loop optimization with different combinations of feed streams to study the influence of Gal and Urd feeding. Here, the ground-truth parameters are used, and the optimization results are shown in Table 7.

Table 7

Optimized fed-batch performance with different feeds.

Feed streams	Harvest volume (L)	GI (mg L <sup>-1</sup> )
None (batch)	1.38	174.27
Medium only	2.18	247.28
Medium + Gal	2.23	404.00
Medium + Urd	2.23	405.25
Medium + Gal + Urd	2.23	405.25

Adding only the supplement medium raises the galactosylation index (GI) by about 42% relative to the batch baseline. Supplementing the medium with either galactose or uridine produces a much larger benefit: the optimal GI is roughly 64% higher than with medium alone (and more than double the batch value). Adding uridine gives a slightly (0.3%) higher GI than adding galactose. Introducing both galactose and uridine simultaneously provides no further gain, because the optimizer drives the galactose feed rate to zero whenever uridine is available. Thus, only one of the two supplemental streams is required; the decision between galactose and uridine should be based on robustness and operational practicality rather than peak GI alone.

Fig. 6 compares the optimal time courses obtained with different feed strategies. Fig. 6(a) confirms the ranking in Table 7: every feeding strategy raises the harvest GI relative to the batch baseline, but the trajectories diverge markedly once galactose or uridine is introduced. When only supplement medium is added, the improvement in extracellular GI stems almost entirely from biomass growth—the viable cell density roughly doubles (Fig. 6(b))—even though the intracellular GI falls slightly (Fig. 6(d)). Adding galactose or uridine in addition to the medium changes the picture: both supplements elevate the intracellular GI by roughly 20% (Fig. 6(d)), on top of the higher viable cell density achieved with medium feeding, yielding the pronounced increase in extracellular GI shown in Fig. 6(a). The increase in intracellular GI is driven by an elevated UDPGal concentration, a key donor for galactosylation (Fig. 6(c)). Once the UDPGal concentration is sufficiently high, however, the incremental benefit plateaus; this is evident from the close overlap of the green dash-dot and red dotted curves in Fig. 6(d), which correspond to Gal and Urd supplement scenarios, respectively.

Although supplementing the reactor with Urd yields a slightly higher GI than Gal, the process is much more sensitive to Urd dosing. We quantify the response to the initial Urd pulse volume  $V_{\text{Urd}}$  by varying its value from 0 mL to 20 mL while holding the medium feed fixed. Fig. 7 shows a sharp optimum at 0.014 mL, giving a harvest GI of 405 mg L<sup>-1</sup>. Increasing

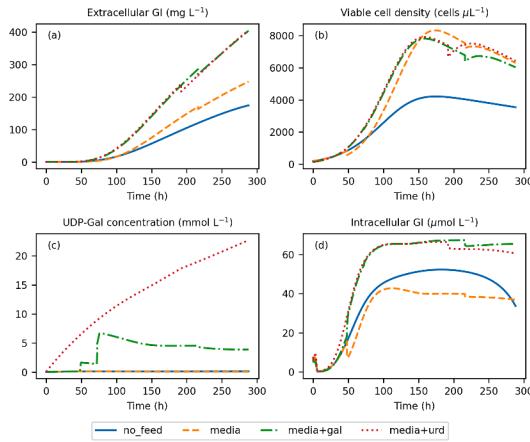


Fig. 6. Optimal trajectories of crucial variables when using different feed streams.

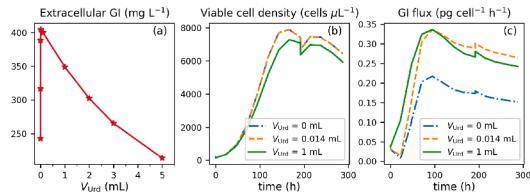


Fig. 7. a) Sensitivity of harvest GI to the Urd feed volume on day 1, b) viable cell density trajectories for three pulse feed volumes, and c) intracellular GI trajectories for the same three different three pulse feed volumes.

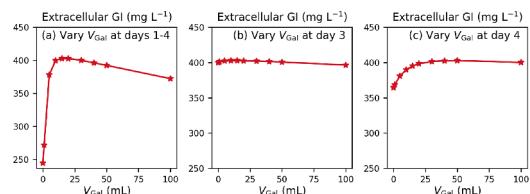


Fig. 8. Sensitivity of harvest GI to the Gal feed volume, a) vary cap V sub Gal simultaneously on days 1–4, b) vary cap V sub Gal on day 3, c) vary cap V sub Gal on day 4.

the pulse to just 1 mL lowers GI to about  $350 \text{ mg L}^{-1}$ —a 13.6% drop—demonstrating that small deviations from the optimum can markedly degrade product quality and, therefore, robustness in the presence of model–plant mismatch or input disturbances. Fig. 7(b)–(c) further indicate that the GI loss is driven primarily by reduced viable cell density throughout the run and by lower intracellular GI during the high cell density phase.

In contrast, the harvest GI is far less sensitive to the galactose pulse volume  $V_{\text{Gal}}$  near its optimum—15 mL on day 3 and 58 mL on day 4, with no Gal on the other days. Fig. 8(a)–(c) illustrate this robustness. When  $V_{\text{Gal}}$  is above 10 mL on each of the first four days (Fig. 8(a)), the harvest GI changes only slowly, and declines rapidly only when  $V_{\text{Gal}}$  falls below 5 mL. Moreover, varying the Gal dose on a single day while keeping the other days at their optimal values (Fig. 8(b)–(c)) produces only minor changes in GI. These results confirm that galactose feeding is a potent yet forgiving lever for glycosylation control, whereas uridine requires near-perfect tuning to realize its modest advantage.

Given the above analysis, the subsequent closed-loop studies employ medium + Gal as the manipulated streams and omit Urd.

Table 8  
Merits under different control algorithms.

	Init_0	Init_1	Init_2	Init_3	Init_4
Open loop	185.88	272.64	231.17	219.86	274.78
State NMPC	276.08	— <sup>a</sup>	— <sup>a</sup>	202.26	367.22
ANMPC	216.82	362.89	220.38	218.30	380.92
Cell culture			8.70		
Ground truth			404.00		

<sup>a</sup> State NMPC terminated prematurely for Init\_1 and Init\_2.

## 5.2. Open-loop optimization vs. state NMPC vs. ANMPC

To benchmark the ANMPC scheme, we compared five controllers, all initiated from the same inaccurate parameter vector:

- (1) Ground truth – dynamic optimization with the ground-truth parameters  $p^*$ .
- (2) Cell culture – dynamic optimization with the objective of maximizing titer, which is a common scenario when there is only the cell culture model. Here, Ground-truth parameters  $p^*$  are used.
- (3) Open loop – dynamic optimization with the inaccurate parameters  $\hat{p}(0)$ .
- (4) State NMPC – including feedback on measured states and using fixed parameters  $\hat{p}(0)$ .
- (5) ANMPC – adjust parameters online with the measured states.

To compare with state NMPC more easily, we assume full state measurement in this section, while the more realistic measurement scheme is considered in Sections 5.4 and 5.5.

Given the importance of the constraints on the cell viability and Man5 fraction in Table 6, the controllers are compared in terms of the penalized merit function

$$\text{Merit} = \text{GI}(T) - \rho_{v,1} \max(0, 60 - \text{viability}) - \rho_{v,2} \max(0, Y_{\text{Man5}}^{\text{extra}} - 5), \quad (36)$$

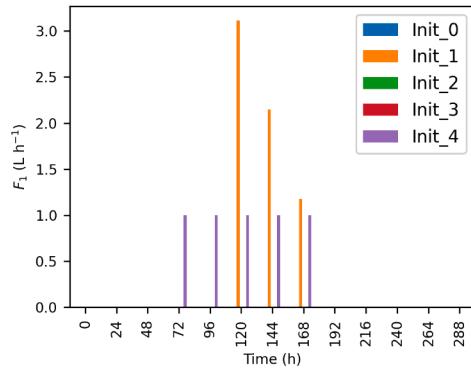
using the same penalty weights  $\rho_v$  as in the soft-constrained formulation (soft-DO- $k$ ; see Table 6). Table 8 reports the merits across the five parameter scenarios.

From Table 8, optimizing the cell culture model only will cause extremely low GI (50 times smaller than the ground-truth solution) because the operating condition of maximizing titer leads to very little FA2G2 and FA2G1 synthesized in the Golgi apparatus (less than 1% for most of the culture time). Open-loop optimization performs second worst for most initializations, producing merits roughly one-half of the ground-truth optima. ANMPC shows a substantial gain (over 30%) for Init\_1 and Init\_4, but offers only marginal or no improvement over open-loop optimization for the other runs. State NMPC terminates prematurely for Init\_1 and Init\_2 due to infeasible simulations during optimization. Overall, none of the controllers consistently achieve satisfactory performance.

A closer look at the ANMPC operation strategies shows that the low-merit cases (Init\_0, Init\_2, Init\_3) never feed Gal, whereas the high-merit cases (Init\_1, Init\_4) apply appropriate Gal dosing (Fig. 9). Consistent with Table 7, omitting Gal caps the ground-truth optimal GI at just  $247 \text{ mg L}^{-1}$ . This behavior occurs because the inaccurate initial parameters undervalue Gal's positive impact on galactosylation, so the optimizer chooses not to add Gal. Without Gal feeding, the Gal-related parameters are not excited and therefore cannot be updated, which masks the potential advantage of ANMPC.

## 5.3. Enforcing a minimum galactose dose

The previous section showed that, when ANMPC selects zero galactose feeding, the Gal-related parameters cannot be updated and performance stalls. To force at least minimal excitation, we impose a lower



**Fig. 9.** Gal feed rate profiles under ANMPC for different initial parameter sets. The feed window duration is 0.01 h.

**Table 9**

Merits after imposing a 1 mL Gal feeding lower bound on day 1.

	Init_0	Init_1	Init_2	Init_3	Init_4
Open loop	183.36	324.88	197.27	166.23	201.67
State NMPC	224.55	— <sup>a</sup>	— <sup>a</sup>	197.11	378.11
ANMPC	380.37	398.29	389.53	386.79	387.06
Ground truth				404.00	

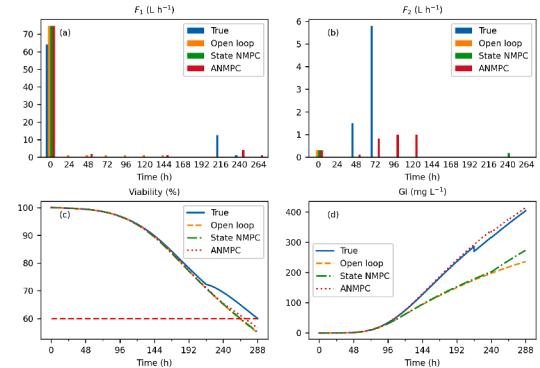
<sup>a</sup> State NMPC terminated prematurely for Init\_1 and Init\_2.

bound of 1 mL on the Gal feed volume during day 1—a negligible volume compared with over 1 L working volume. **Table 9** compares the resulting merits for the various control strategies under this specification.

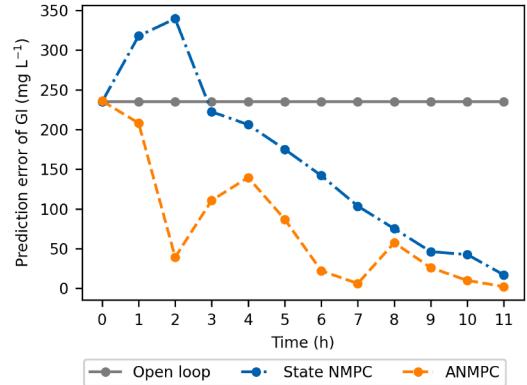
Comparing **Table 8** (no Gal feed lower bound) with **Table 9** (1 mL lower bound on day 1), the open-loop and state-NMPC controllers show almost no progress: their merits remain roughly half of the 404 benchmark, and state NMPC still fails to converge for two of the five initializations. By contrast, ANMPC improves markedly once the small day-1 galactose dose is enforced. Its merits lie within 6% of the ground-truth optimum for every initialization and exhibit little scatter, indicating substantially greater robustness. Relative to ANMPC without a Gal lower bound, the merit increases by more than 75% for Init\_0, Init\_2, and Init\_3. Compared with open-loop optimization and state NMPC under the same Gal feed constraint, ANMPC delivers gains of up to 130% and 96% (Init\_3), respectively. These results confirm that introducing even a minimal Gal feed early in the cell culture provides sufficient excitation to update Gal-related parameters and enables the adaptive controller to approach its full potential.

**Fig. 10** illustrates the input profiles and trajectories for Init\_0 along with the ground-truth policy. The ground-truth optimization splits the medium addition into two pulses—one at the start of culture and a second after nine days (**Fig. 10(a)**). This staged strategy moderates the late-phase viability drop and satisfies the terminal viability limit (**Fig. 10(c)**). In contrast, the other controllers deliver nearly the entire medium volume at the outset, pushing the broth quickly to the reactor's upper volume limit and leaving little capacity for a compensatory feed later; the resulting nutrient depletion leads the NMPC variants to violate the viability constraint. The second key difference appears in **Fig. 10(b)**: substantial Gal supplementation occurs only in the ground-truth and ANMPC policies, yielding a markedly higher GI in **Fig. 10(d)**. Like open-loop and state NMPC, ANMPC omits Gal during the first three days due to initial model mismatch. Once the parameter estimates improve, however, ANMPC initiates Gal feeding; from that point forward, its GI trajectory closely tracks the ground-truth benchmark and remains well above the curves from the other two controllers.

The contrasting control performances can be traced to how each algorithm handles model–plant mismatch, as illustrated by the evolution of the GI prediction errors at the harvest time in **Fig. 11**. At the start of the run (day 0) the three schemes—open-loop optimization, state NMPC,



**Fig. 10.** Control inputs and key state trajectories for the control schemes starting from Init\_0 and the ground-truth optimization: (a) medium feed flow rate, (b) Gal feed flow rate, (c) cell viability, and (d) GI.



**Fig. 11.** Evolution of the GI prediction error at harvest time when starting from Init\_0.

and ANMPC—share the same parameter set, so their GI prediction errors are identical, about 200 mg L<sup>-1</sup>. Thereafter, the open-loop error remains unchanged because no feedback is used to refine the model. In both state NMPC and ANMPC, the error decreases as new measurements arrive, but ANMPC's errors are consistently smaller than those of state NMPC. Within two days, the GI error under ANMPC drops below 50 mg L<sup>-1</sup> (except on days 3–5), whereas state NMPC does not reach this level until about day nine. The rapid improvement in ANMPC stems from its online parameter adaptation, whereas the slower drift in state NMPC reflects only the shrinking prediction horizon and the availability of full-state measurements, not improved model accuracy. By the time state NMPC finally achieves a similar error magnitude, only three days remain to influence the harvest GI, leaving ANMPC with the practical advantage.

#### 5.4. ANMPC with fewer measurements and longer preparation time

Here we evaluate ANMPC under more realistic conditions: (1) partial measurement availability—no readings for extracellular Asn-and Asp, the extracellular glycan fractions HM, FA1G1, and SIA, or any intracellular NSD concentrations; (2) a longer preparation time of 4 h to account for computation and assay turnaround. The results are summarized in **Table 10**.

**Table 10** shows that, for a given initial parameter set, the penalized merit varies by less than 2% across different measurement schemes and preparation times. In particular, using partial measurements and a 4 h preparation window does not necessarily degrade performance relative to the full measurements and 2 h baseline. Therefore, a leaner measurement set and a relaxed preparation window can be adopted as a more economical choice without sacrificing control quality.

**Table 10**

ANMPC under different measurements and preparation time.

	Init_0	Init_1	Init_2	Init_3	Init_4	Avg
Full <sup>a</sup> + $\Delta t_{\text{prep}} = 2$ h	380.37	398.29	389.53	386.79	387.06	388.41
Partial <sup>b</sup> + $\Delta t_{\text{prep}} = 2$ h	387.81	395.60	392.24	388.52	384.52	389.74
Partial <sup>b</sup> + $\Delta t_{\text{prep}} = 4$ h	386.18	395.25	387.89	386.62	379.40	387.07

<sup>a</sup> "Full" = all variables measured.<sup>b</sup> "Partial" = excludes extracellular Asn-and Asp, extracellular glycan fractions (HM, FA1G1, SIA), and all intracellular NSD concentrations.

Glycans are measured only at the end of a batch in many companies. Regarding employment of the multiscale model in that scenario, any company can take glycosylation experiments once a day by manual sampling during the fed-batch run, freeze the samples, and analyze all samples after the fed-batch experiment – to generate time-series data for use in estimating model parameters. Once the multiscale model has been validated offline, the multiscale model is still appropriate for predicting glycan trajectories (time series). The closed-loop performance obtained by the MPC algorithm will be lower if the glycosylation is only measured at the end of each batch run, since there would be less data for updating the glycosylation model parameter estimates. The MPC algorithm would still function, with the real-time update of the glycosylation model parameter estimates turned off since those parameters would only be updated after each batch, although with lower performance.

## 6. Conclusions

We applied an ANMPC framework to regulate the N-glycosylation in a fed-batch mAb bioreactor. The controller uses a high-fidelity multiscale PDAE model that links extracellular operating conditions to intracellular Golgi reactions to predict both mAb productivity and glycan profiles. To mitigate model–plant mismatch, parameters are updated online as new measurements arrive, after which a shrinking-horizon optimization recomputes the inputs. When the nominal dynamic optimization becomes infeasible, a soft-constraint formulation preserves solvability. To prevent unphysical predictions (e.g., negative concentrations), parameter estimation is performed over the full fed-batch horizon with state bounds. Computationally, parameter estimation and dynamic optimization yield large-scale, model-constrained problems that are challenging to solve. We therefore employ the control-vector parameterization (CVP) method for reliable convergence and accelerate embedded simulations via a parallel QSS algorithm.

Sensitivity studies show that both Gal and Urd can raise GI by about 64% relative to feeding medium alone to the bioreactor, but Urd is overly dosage-sensitive for control. Across case studies with multiple initial parameter sets, ANMPC delivers up to 130% and 96% higher penalized merit than open-loop optimization and state NMPC, respectively –

provided a small day-1 Gal lower bound is imposed to excite Gal-related dynamics and enable adaptation. Moreover, using fewer measurements and a longer preparation time (4 h) did not materially degrade ANMPC performance, supporting practical deployability.

Regarding implementation, experimental validation of the control algorithm requires the implementation of an automated glycan assay, which requires additional equipment and software beyond what is available at some companies. Future work will explore stochastic ANMPC (e.g., chance-constrained MPC) to balance constraint satisfaction and performance under uncertainty, and experimental validation of the proposed strategies.

## CRediT authorship contribution statement

**Yingjie Ma:** Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization; **Jing Guo:** Writing – review & editing, Visualization, Validation, Software, Investigation, Data curation; **Alexis B. Dubs:** Writing – review & editing, Visualization, Validation, Formal analysis, Data curation; **Krystian Ganko:** Writing – review & editing, Visualization, Software, Formal analysis, Data curation; **Richard D. Braatz:** Writing – review & editing, Supervision, Resources, Project administration, Methodology, Investigation, Funding acquisition, Conceptualization.

## Data availability

The code used for the paper is available at [https://github.com/eaglema/glycosylation\\_python/tree/ANMPC-paper](https://github.com/eaglema/glycosylation_python/tree/ANMPC-paper).

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgment

This work was supported by a Project Award Agreement from the National Institute for Innovation in Manufacturing Biopharmaceuticals (NIIMBL), with financial assistance from the U.S. Department of Commerce, National Institute of Standards and Technology (NIST), awards 70NANB17H002 and 70NANB20H037. We thank Professor Cleo Konoravdi for providing the experimental data reported in her publication.

## Supplementary material

Supplementary material associated with this article can be found in the online version at [10.1016/j.conengprac.2025.106731](https://doi.org/10.1016/j.conengprac.2025.106731).

## Appendix A. Kinetic equations in the cell culture model

The equations for computing the cell growth and death rates are

$$\mu = \mu_{\max} f_{\text{lim}} f_{\text{inh}} \quad (\text{A.1})$$

$$\mu_{\text{death}} = \mu_{\text{death,max}} \left( \frac{[\text{Amm}]}{[\text{Amm}] + K_{d,\text{Amm}}} + \frac{[\text{Urd}]}{[\text{Urd}] + K_{d,\text{Urd}}} \right) \quad (\text{A.2})$$

$$f_{\text{lim}} = \frac{[\text{Glc}]}{[\text{Glc}] + K_{\text{Glc}}} \frac{[\text{Asn}]}{[\text{Asn}] + K_{\text{Asn}}} \quad (\text{A.3})$$

$$f_{\text{inh}} = \frac{K_{\text{I,Amm}}}{[\text{Amm}] + K_{\text{I,Amm}}} \frac{K_{\text{I,Lac}}}{[\text{Lac}] + K_{\text{I,Lac}}} \frac{K_{\text{I,Urd}}}{[\text{Urd}] + K_{\text{I,Urd}}} \quad (\text{A.4})$$

where  $\mu_{\max}$  and  $\mu_{\text{death,max}}$  ( $\text{h}^{-1}$ ) are the maximum specific growth and death rates, respectively;  $f_{\text{lim}}$  and  $f_{\text{inh}}$  are the substrate-limiting and metabolite-inhibiting factors, respectively;  $K_{\text{Glc}}$  and  $K_{\text{Asn}}$  ( $\text{mmol L}^{-1}$ ) are the Monod half-saturation constants for glucose and asparagine;  $K_{d,\text{Amm}}$  and  $K_{d,\text{Urd}}$  ( $\text{mmol L}^{-1}$ ) are the death constants for ammonia and uridine;  $K_{\text{I,Amm}}$ ,  $K_{\text{I,Lac}}$ , and  $K_{\text{I,Urd}}$  ( $\text{mmol L}^{-1}$ ) are inhibition constants for ammonia, lactate, and uridine; and  $[\text{Glc}]$ ,  $[\text{Asn}]$ ,  $[\text{Amm}]$ ,  $[\text{Lac}]$ , and  $[\text{Urd}]$  ( $\text{mmol L}^{-1}$ ) denote the extracellular concentrations of glucose, asparagine, ammonia, lactose and uridine, respectively.

The kinetic equations for cell metabolism are

$$q_{\text{Glc}} = \left( -\frac{\mu}{Y_{X_{\text{Glc}}}} - m_{\text{Glc}} \right) \left( \frac{K_{C_{\text{Gal}}}}{K_{C_{\text{Gal}}} + [\text{Gal}]} \right)^{n_{\text{Gal}}} \quad (\text{A.5})$$

$$n_{\text{Gal}} = 1 - f_{\text{Gal}} \frac{q_{\text{Gal}}}{q_{\text{Glc}}} \quad (\text{A.6})$$

$$q_{\text{Gln}} = \frac{\mu}{Y_{X_{\text{Gln}}}} + q_{\text{Amm}} Y_{\text{Gln/Amm}} \quad (\text{A.7})$$

$$q_{\text{Lac}} = \left( \frac{\mu}{Y_{X_{\text{Lac}}}} - Y_{\text{Lac/Glc}} q_{\text{Glc}} \right) \frac{\text{Lac}_{\text{max1}} - [\text{Lac}]}{\text{Lac}_{\text{max1}}} + m_{\text{Lac}} \frac{\text{Lac}_{\text{max2}} - [\text{Lac}]}{\text{Lac}_{\text{max2}}} \quad (\text{A.8})$$

$$q_{\text{Amm}} = \frac{\mu}{Y_{X_{\text{Amm}}}} - Y_{\text{Amm/Urd}} q_{\text{Urd}} \quad (\text{A.9})$$

$$q_{\text{Glu}} = -\frac{\mu}{Y_{X_{\text{Glu}}}} \quad (\text{A.10})$$

$$q_{\text{Gal}} = -\frac{\mu}{Y_{X_{\text{Gal}}}} \frac{[\text{Gal}]}{[\text{Gal}] + K_{\text{Gal}}} \quad (\text{A.11})$$

$$q_{\text{Urd}} = \frac{\mu}{Y_{X_{\text{Urd}}}} \frac{[\text{Urd}]}{[\text{Urd}] + K_{\text{Urd}}} \quad (\text{A.12})$$

$$q_{\text{Asn}} = -\frac{\mu}{Y_{X_{\text{Asn}}}} - Y_{\text{Asn/Asp}} q_{\text{Asp}} \quad (\text{A.13})$$

$$q_{\text{Asp}} = -\frac{\mu}{Y_{X_{\text{Asp}}}} - Y_{\text{Asp/Asn}} q_{\text{Asn}} \quad (\text{A.14})$$

$$q_{\text{mAb}} = Y_{\text{mAb}_X} \mu + m_{\text{mAb}} \quad (\text{A.15})$$

In these expressions, the model parameters to be estimated include the yield coefficients  $Y_{X_{\text{Met}}}$  ( $\text{cell mmol}^{-1}$ ), maintenance terms  $m_{\text{Met}}$  ( $\text{mmol cell}^{-1} \text{h}^{-1}$ ) for metabolites excluding mAb, maintenance term  $m_{\text{mAb}}$  ( $\text{pg cell}^{-1} \text{h}^{-1}$ ) for mAb, the fraction  $f_{\text{Gal}}$  (-), cross-yield ratios  $Y_{\text{Met1/Met2}}$  ( $\text{mmol mmol}^{-1}$ ), saturation or inhibition constants  $K_{\text{Met}}$  ( $\text{mmol L}^{-1}$ ), the empirical limits  $\text{Lac}_{\text{max1}}$  and  $\text{Lac}_{\text{max2}}$  ( $\text{mmol L}^{-1}$ ), and the product formation yield  $Y_{\text{mAb}_X}$  ( $\text{pg cell}^{-1}$ ).

## Appendix B. Full names of glycans

The full names of the glycans in the current model are

- HM: High-mannose-type (oligomannose) N-glycan(s), i.e., the high-mannose family such as Man5–Man9.
- FA1G1: Core-fucosylated, mono-antennary, mono-galactosylated complex-type N-glycan.
- FA2G0: Core-fucosylated, bi-antennary, agalactosylated complex-type N-glycan.
- FA2G1: Core-fucosylated, bi-antennary, mono-galactosylated complex-type N-glycan.
- FA2G2: Core-fucosylated, bi-antennary, di-galactosylated (asialo) complex-type N-glycan.
- SIA: Sialylated N-glycan(s), i.e., glycans bearing terminal sialic acid residues.
- G0: Agalactosylated bi-antennary complex-type N-glycan (no core fucose).
- G2: Di-galactosylated bi-antennary complex-type N-glycan (no core fucose).
- Man5: Oligomannose-5 (high-mannose) N-glycan with five mannose residues.

## Appendix C. Kinetic equations for the NSD reactions

The kinetic equations for intracellular NSD synthesis reactions are

$$r_1 = V_{\max,1} \frac{[\text{Gln}_{\text{intra}}]}{K_{M1_{\text{Gln}}} + [\text{Gln}_{\text{intra}}]} \quad (\text{C.1})$$

$$r_{1_{\text{sink}}} = V_{\max,1_{\text{sink}}} \frac{[\text{UDPGlcNAc}]}{\left( K_{M1_{\text{sink}}} + [\text{UDPGlcNAc}] \right) \left( 1 + \frac{[\text{CMPNeu5Ac}]}{K_{I1_{\text{sink}}}} \right)} \quad (\text{C.2})$$

$$r_2 = V_{\max,2} \frac{[\text{Glc}]}{K_{M2_{\text{Glc}}} + [\text{Glc}]} \quad (\text{C.3})$$

$$r_{2b} = V_{\max,2b} \frac{[\text{UDPGal}]}{K_{M2b,\text{UDPGal}} \left( 1 + \frac{[\text{UDPGlcNAc}]}{K_{I2A}} + \frac{[\text{UDPGalNAc}]}{K_{I2B}} + \frac{[\text{UDPGlc}]}{K_{I2C}} + \frac{[\text{UDPGal}]}{K_{I2D}} \right) + [\text{UDPGal}]} \quad (\text{C.4})$$

$$r_3 = V_{\max,3} \frac{[\text{Glc}]}{K_{M3_{\text{Glc}}} + [\text{Glc}]} \quad (\text{C.5})$$

$$r_4 = V_{\max,4} \frac{[\text{UDPGlcNAc}]}{K_{M4_{\text{UDPGlcNAc}}} + [\text{UDPGlcNAc}]} \quad (\text{C.6})$$

$$r_5 = V_{\max,5} \frac{[\text{UDPGlcNAc}]}{K_{M5_{\text{UDPGlcNAc}}} \left( 1 + \frac{[\text{CMPNeu5Ac}]}{K_{I5}} \right) + [\text{UDPGlcNAc}]} \quad (\text{C.7})$$

$$r_6 = V_{\max,6} \frac{[\text{UDPGlc}]}{K_{M6,\text{UDPGlc}} \left( 1 + \frac{[\text{UDPGlcNAc}]}{K_{I6A}} + \frac{[\text{UDPGalNAc}]}{K_{I6B}} + \frac{[\text{UDPGal}]}{K_{I6C}} \right) + [\text{UDPGlc}]} \quad (\text{C.8})$$

$$r_{6_{\text{sink}}} = V_{\max,6_{\text{sink}}} \frac{[\text{UDPGal}]}{K_{M6_{\text{sink}}} \left( 1 + \frac{[\text{UDPGlc}]}{K_{I6_{\text{sink}}}} \right) + [\text{UDPGal}]} \frac{[\text{Gal}]}{[\text{Gal}] + K_{\text{regulator}}} \quad (\text{C.9})$$

$$r_7 = V_{\max,7} \frac{[\text{GDPMAn}]}{\left( K_{M7_{\text{GDPMAn}}} + [\text{GDPMAn}] \right) \left( 1 + \frac{[\text{GDPFuc}]}{K_{I7}} \right)} \quad (\text{C.10})$$

$$r_{7_{\text{sink}}} = V_{\max,7_{\text{sink}}} \frac{[\text{GDPFuc}]}{K_{M7_{\text{sink}}} + [\text{GDPFuc}]} \quad (\text{C.11})$$

where  $V_{\max,j}$  ( $\text{mmol L}^{-1} \text{ h}^{-1}$ ) is the maximum turnover rate of reaction  $j$ ;  $K_{Mj_{\text{comp}}}$  ( $\text{mmol L}^{-1}$ ) is the Michaelis constant of a component (metabolite or NSD) in reaction  $j$ ;  $K_{Ij}$  ( $\text{mmol L}^{-1}$ ) is the corresponding inhibition constant; and  $K_{\text{regulator}}$  ( $\text{mmol L}^{-1}$ ) is introduced so that the factor  $\frac{[\text{Gal}]}{[\text{Gal}] + K_{\text{regulator}}}$  becomes zero when extracellular Gal is absent.

The intracellular glutamine concentration ( $\text{mmol L}^{-1}$ ) in Eq. (C.1) is estimated through

$$\text{Gln}_{\text{intra}} = f_{\text{Gln}} [\text{Gln}],$$

where  $f_{\text{Gln}}$  is a constant parameter to connect intracellular Gln concentration with its extracellular counterpart.

The NSD synthesis reaction rates that depend on Urd or Gal are

$$r_{j_{\text{Urd}}} = V_{\max,j_{\text{Urd}}} \frac{[\text{Urd}]}{K_{M,j_{\text{Urd}}} + [\text{Urd}]}, \quad j \in \{1, 2, 4, 6\}, \quad (\text{C.12})$$

$$r_{6_{\text{Gal}}} = \frac{V_{\max,6_{\text{Gal}}} [\text{Gal}]}{K_{M,6_{\text{Gal}}} \left( 1 + \frac{[\text{UDPGal}]}{K_{\text{I6D}}} + \frac{[\text{Gal}]}{K_{\text{I6E}}} + \frac{[\text{Urd}]}{K_{\text{I6F}}} \right) + [\text{Gal}]} \quad (\text{C.13})$$

Additional intracellular fluxes for each NSD<sub>i</sub> are

$$f_i^{\text{hcp/lipid}} = \frac{[\text{NSD}_i]}{K_{\text{TP},i} + [\text{NSD}_i]} \frac{v_i^{\text{hcp/lipid}} \mu}{V_{\text{cell}}}, \quad (\text{C.14})$$

$$f_i^{\text{precursor}} = \frac{[\text{NSD}_i]}{K_{\text{TP},i} + [\text{NSD}_i]} \frac{v_i^{\text{precursor}} q_{\text{mAb}}}{V_{\text{cell}}}, \quad (\text{C.15})$$

$$f_i^{\text{glyc}} = \frac{[\text{NSD}_i]}{K_{\text{TP},i} + [\text{NSD}_i]} r_i^{\text{glyc}}, \quad (\text{C.16})$$

where  $K_{\text{TP},i}$  (mmol L<sup>-1</sup>) is the transport protein saturation constant for NSD<sub>i</sub>;  $v_i^{\text{hcp/lipid}}$  (mmol cell<sup>-1</sup>) and  $v_i^{\text{precursor}}$  (mmol pg<sup>-1</sup>) are the stoichiometric requirements for host cell protein/glycolipid synthesis and precursor oligosaccharide assembly, respectively;  $V_{\text{cell}}$  (L cell<sup>-1</sup>) is the specific cell volume and the NSD<sub>i</sub> consumption rate in glycosylation reactions  $r_i^{\text{glyc}}$  (mmol L<sup>-1</sup> h<sup>-1</sup>) is (Kotidis et al., 2019)

$$r_i^{\text{glyc}} = \text{vel}_{\text{golgi}} \left( \frac{V_{\text{golgi}}}{V_{\text{cell}}} \right) \sum_{j=1}^{N_{\text{OS}}} \{v_{i,j} [\text{OS}_j](z=1)\} \quad (\text{C.17})$$

$$\text{vel}_{\text{golgi}} = \frac{2 \times 10^{-6} q_{\text{mAb}}}{60 \text{ MW}_{\text{mAb}} V_{\text{golgi}} [\text{OS}_1](z=0)} \quad (\text{C.18})$$

where  $\text{vel}_{\text{golgi}}$  (Golgi length min<sup>-1</sup>) is the length-normalized transit velocity,  $V_{\text{golgi}}$  (L cell<sup>-1</sup>) is the Golgi volume,  $v_{i,j}$  is the number of NSD<sub>i</sub> molecules required for one oligosaccharide  $j$  molecule,  $[\text{OS}_j](z)$  (mmol L<sup>-1</sup>) is the oligosaccharide concentration at axial position  $z$  (with  $z=0$  at entry and  $z=1$  at exit), and  $\text{MW}_{\text{mAb}}$  is the molecular weight of mAb (165 × 10<sup>3</sup> g mol<sup>-1</sup>).

#### Appendix D. Kinetic equations for glycosylation reactions

The kinetic equations for the glycosylation reactions in the Golgi apparatus are shown in Eqs. (D.1)–(D.3).

(i) Michaelis-Menten kinetics – used for enzymes Man I and Man II-catalyzed reactions

$$r_j = \frac{k_{f,j} [E_j] [\text{OS}_i]}{K_{d,i}^{\text{enz}} \left( 1 + \sum_{k=1}^{N_{\text{C}}} \frac{[\text{OS}_k]}{K_{d,k}^{\text{enz}}} \right)} \quad (\text{D.1})$$

(ii) Sequential order Bi-Bi kinetics – used for enzymes GnT I, GnT II, and GalT-catalyzed reactions

$$r_j = \frac{k_{f,j} [E_j] [\text{Mn}] [\text{NSD}_z] [\text{OS}_i]}{K_{d,\text{Mn}}^{\text{enz}} K_{d,z}^{\text{enz}} K_{d,i}^{\text{enz}} \left( 1 + \frac{[\text{Mn}]}{K_{d,\text{Mn}}^{\text{enz}}} + \frac{[\text{Mn}]}{K_{d,z}^{\text{enz}}} \frac{[\text{NSD}_z]}{K_{d,z}^{\text{enz}}} + \frac{[\text{Mn}]}{K_{d,\text{Mn}}^{\text{enz}}} \frac{[\text{NSD}_z]}{K_{d,z}^{\text{enz}}} \sum_{k=1}^{N_{\text{C}}} \frac{[\text{OS}_k]}{K_{d,k}^{\text{enz}}} + \frac{[\text{Nuc}_n]}{K_{d,n}^{\text{enz}}} \frac{[\text{OS}_{i+1}]}{K_{d,i+1}^{\text{enz}}} + \frac{[\text{Nuc}_n]}{K_{d,n}^{\text{enz}}} \right)} \quad (\text{D.2})$$

(iii) Random-order Bi-Bi kinetics – used for enzymes FucT- and SiaT-catalyzed reactions

$$r_j = \frac{k_{f,j} [E_j] [\text{NSD}_z] [\text{OS}_i]}{K_{d,z}^{\text{enz}} K_{d,i}^{\text{enz}} \left( 1 + \frac{[\text{NSD}_z]}{K_{d,z}^{\text{enz}}} + \sum_{k=1}^{N_{\text{C}}} \frac{[\text{OS}_k]}{K_{d,k}^{\text{enz}}} + \frac{[\text{NSD}_z]}{K_{d,z}^{\text{enz}}} \sum_{k=1}^{N_{\text{C}}} \frac{[\text{OS}_k]}{K_{d,k}^{\text{enz}}} + \frac{[\text{Nuc}_n]}{K_{d,n}^{\text{enz}}} + \frac{[\text{Nuc}_n]}{K_{d,n}^{\text{enz}}} \frac{[\text{OS}_{i+1}]}{K_{d,i+1}^{\text{enz}}} \right)} \quad (\text{D.3})$$

In these expressions,  $\text{OS}_i$  and  $\text{OS}_{i+1}$  are the reactant and product of reaction  $j$ , respectively; the set  $\{\text{OS}_k \mid k = 1, 2, \dots, N_{\text{C}}\}$  contains every oligosaccharide that can bind to the enzyme  $E_j$ ;  $K_{d,i}^{\text{enz}}$  and  $K_{d,k}^{\text{enz}}$  (mmol L<sup>-1</sup>) are the dissociation constants for  $\text{OS}_i$  and  $\text{OS}_k$  bound to the enzyme  $\text{enz} \in \{\text{Man I, Man II, GnT I, GnT II, GalT, FucT, SiaT}\}$ , respectively;  $K_{d,z}^{\text{enz}}$  and  $K_{d,n}^{\text{enz}}$  (mmol L<sup>-1</sup>) are the dissociation constants for nucleotide sugar donor  $\text{NSD}_z$  and nucleotide  $\text{Nuc}_n$  bound to the enzyme  $\text{enz}$ , where  $n \in \{1, 2, 3, 4\}$ ;  $K_{d,\text{Mn}}^{\text{enz}}$  refers to the dissociation constants for manganese bound to enzymes  $\text{enz} \in \{\text{GnT I, GnT II, GalT}\}$ ;  $[\cdot]$  denotes the relevant intracellular concentration (mmol L<sup>-1</sup>); and  $k_{f,j}$  (min<sup>-1</sup>) is the kinetic rate constant of reaction  $j$ .

In Eqs. (D.2)–(D.3), the NSD concentration inside the Golgi apparatus,  $[\text{NSD}_z]$ , is assumed to equal 20 times of the corresponding cytosolic concentration,  $[\text{NSD}_i]$ , computed from the NSD synthesis model (Kotidis et al., 2019). The nucleotide concentration inside the Golgi apparatus,  $[\text{Nuc}_n]$ , is taken to be saturating and is therefore fixed at the constant values reported in Table S1 (Kotidis et al., 2019).

The catalytic rate constant  $k_{f,j}$  follows the pH-dependent expression (Villiger et al., 2016):

$$k_{f,j} = k_{f,j}^{\max} \exp \left( -\frac{1}{2} \left( \frac{\text{pH}_{\text{golgi}} - \text{pH}_{\text{opt}}}{\omega_{f,j}} \right)^2 \right), \quad (\text{D.4})$$

with the intra-Golgi pH estimated from the ammonia-based buffer relation

$$pH_{golgi} = pK_A^{golgi} + \log \left( \frac{[Amm]}{N_A^{golgi} - [Amm]} \right), \quad (D.5)$$

where  $k_{f,j}^{\max}$  ( $\text{min}^{-1}$ ) is the maximal kinetic rate,  $pH_{\text{opt}}^{golgi}$  is the optimal pH for enzyme  $j$ ,  $\omega_{f,j}$  is a dimensionless width parameter,  $pK_A^{golgi}$  is the apparent acid dissociation parameter, and  $N_A^{golgi}$  is a fitted buffer capacity.

Each enzyme concentration along the Golgi cisternae follows a Gaussian (bell-shaped) profile:

$$[E_j](z) = E_{j,\max} \exp \left[ -\frac{1}{2} \left( \frac{z - z_{j,\max}}{\sigma_j} \right)^2 \right], \quad (D.6)$$

where  $E_{j,\max}$  ( $\text{mmol L}^{-1}$ ) is the peak concentration,  $z_{j,\max}$  is the axial position of this peak (scaled 0–1 from cis to trans), and  $\sigma_j$  is the standard deviation that characterizes enzyme dispersion along the Golgi stack.

## References

Andersson, J. A. E., Gillis, J., Horn, G., Rawlings, J. B., & Diehl, M. (2019). CasADI: A software framework for nonlinear optimization and optimal control. *Mathematical Programming Computation*, 11(1), 1–36. <https://doi.org/10.1007/s12532-018-0139-4>

Batra, J., & Rathore, A. S. (2016). Glycosylation of monoclonal antibody products: current status and future prospects. *Biotechnology Progress*, 32(5), 1091–1102. <https://doi.org/10.1002/btpr.2366>

Beck, J. V., & Arnold, K. J. (1977). *Parameter estimation in engineering and science*. Wiley.

Dan, A., Liu, B., Patil, U., Manuraj, B. N. M., Gandhi, R., Buchel, J., Chundawat, S. P. S., Guo, W., & Ramachandran, R. (2025). Machine learning model-based design and model predictive control of a bioreactor for the improved production of mammalian cell-based bio-therapeutics. *Control Engineering Practice*, 156, 106198. <https://doi.org/10.1016/j.conengprac.2024.106198>

Dewasme, L., Fernandes, S., Amribit, Z., Santos, L. O., Bogaerts, P., & Vande Wouwer, A. (2015). State estimation and predictive control of fed-batch cultures of hybridoma cells. *Journal of Process Control*, 30, 50–57. <https://doi.org/10.1016/j.jprocont.2014.12.006>

Eaton, J. W., & Rawlings, J. B. (1990). Feedback control of chemical processes using online optimization techniques. *Computers & Chemical Engineering*, 14(4), 469–479. [https://doi.org/10.1016/0098-1354\(90\)87021-G](https://doi.org/10.1016/0098-1354(90)87021-G)

Federici, M., Lubiniecki, A., Manikwar, P., & Volkin, D. B. (2013). Analytical lessons learned from selected therapeutic protein drug comparability studies. *Biologicals*, 41(3), 131–147. <https://doi.org/10.1016/j.biologicals.2012.10.001>

Gardner, D. J., Reynolds, D. R., Woodward, C. S., & Balos, C. J. (2022). Enabling new flexibility in the SUNDIALS suite of nonlinear and differential/algebraic equation solvers. *ACM Transactions on Mathematical Software*, 48(3), 31. <https://doi.org/10.1145/3539801>

Goetze, A. M., Liu, Y. D., Zhang, Z., Shah, B., Lee, E., Bondarenko, P. V., & Flynn, G. C. (2011). High-mannose glycans on the Fc region of therapeutic IgG antibodies increase serum clearance in humans. *Glycobiology*, 21(7), 949–959. <https://doi.org/10.1093/glycob/cwr027>

Gunawan, R., Jung, M. Y. L., Seebauer, E. G., & Braatz, R. D. (2003). Maximum a posteriori estimation of transient enhanced diffusion energetics. *AIChE Journal*, 49(8), 2114–2123. <https://doi.org/10.1002/aic.690490819>

Hajizadeh, I., Rashid, M., Sevil, M., Brandt, R., Samadi, S., Hobbs, N., & Cinar, A. (2018). Adaptive model predictive control for nonlinearity in biomedical applications. In *6th IFAC conference on nonlinear model predictive control (NMPC 2018)* (pp. 368–373). (vol. 51). <https://doi.org/10.1016/j.ifacol.2018.11.061>

Holdmarch, A. C., Brown, P. N., Grant, K. E., Lee, S. L., Serban, R., Shumaker, D. E., & Woodward, C. S. (2005). SUNDIALS: Suite of nonlinear and differential/algebraic equation solvers. *ACM Transactions on Mathematical Software*, 31(3), 363–396. <https://doi.org/10.1145/1089014.1089020>

Hong, W., Wang, S., Li, P., Wozny, G., & Biegler, L. T. (2006). A quasi-sequential approach to large-scale dynamic optimization problems. *AIChE Journal*, 52(1), 255–268. <https://doi.org/10.1002/aic.10625>

Jabarivelisdeh, B., Carius, L., Findeisen, R., & Waldherr, S. (2020). Adaptive predictive control of bioprocesses with constraint-based modeling and estimation. *Computers & Chemical Engineering*, 135, 106744. <https://doi.org/10.1016/j.compchemeng.2020.106744>

James, D. C. (2005). Control of recombinant monoclonal antibody effector functions by Fc N-glycan remodeling in vitro. *Biotechnology Progress*, 21(6), 1644–1652. <https://doi.org/10.1002/btpr.50228>

Jefferis, R. (2009). Glycosylation as a strategy to improve antibody-based therapeutics. *Nature Reviews Drug Discovery*, 8(3), 226–234. <https://doi.org/10.1038/nrd2804>

Jimenez del Val, I., Nagy, J. M., & Kontoravdi, C. (2011). A dynamic mathematical model for monoclonal antibody N-linked glycosylation and nucleotide sugar donor transport within a maturing golgi apparatus. *Biotechnology Progress*, 27(6), 1730–1743. <https://doi.org/10.1002/btpr.688>

Kappatou, C. D., Ehsani, A., Niedenführ, S., Mhamdi, A., Schuppert, A., & Mitsos, A. (2020). Quality-targeting dynamic optimization of monoclonal antibody production. *Computers & Chemical Engineering*, 142, 107004. <https://doi.org/10.1016/j.compchemeng.2020.107004>

Kontoravdi, C., Pistikopoulos, E. N., & Mantalaris, A. (2010). Systematic development of predictive mathematical models for animal cell cultures. *Computers & Chemical Engineering*, 34(8), 1192–1198. <https://doi.org/10.1016/j.compchemeng.2010.03.012>

Kornfeld, R., & Kornfeld, S. (1985). Assembly of asparagine-linked oligosaccharides. *Annual Review of Biochemistry*, 54, 631–664. <https://doi.org/10.1146/annurev.bi.54.070185.003215>

Kotidis, P. (2021). Enhanced understanding of protein glycosylation in cho cells through computational tools and experimentation. Ph.D. thesis. Imperial College London. <https://doi.org/10.25560/100281>

Kotidis, P., Jedrzejewski, P. M., Sou, S. N., Sellick, C., Polizzi, K., Jimenez del Val, I., & Kontoravdi, C. (2019). Model-based optimization of antibody galactosylation in CHO cell culture. *Biotechnology and Bioengineering*, 116(7), 1612–1626. <https://doi.org/10.1002/bit.26960>

Kraft, D. (1985). On converting optimal control problems into nonlinear programming problems. In K. Schittkowski (Ed.), *Computational mathematical programming* (pp. 261–280). Springer Berlin Heidelberg.

Lu, F., Toh, P. C., Burnett, I., Li, F., Hudson, T., Amanullah, A., & Li, J. (2013). Automated dynamic fed-batch process and media optimization for high productivity cell culture process development. *Biotechnology and Bioengineering*, 110(1), 191–205. <https://doi.org/10.1002/bit.24602>

Luo, Y., Kurian, V., Song, L., Wells, E. A., Robinson, A. S., & Ogunnaike, B. A. (2023). Model-based control of titer and glycosylation in fed-batch mab production: Modeling and control system development. *AIChE Journal*, 69, e18075. <https://doi.org/10.1002/aic.18075>

Ma, Y., Gao, X., Liu, C., & Li, J. (2024). Improved SQP and SLSQP algorithms for feasible path-based process optimisation. *Computers & Chemical Engineering*, 188, 108751. <https://doi.org/10.1016/j.compchemeng.2024.108751>

Ma, Y., Guo, J., Maloney, A. J., & Braatz, R. D. (2025). Quasi-steady-state approach for efficient multiscale simulation and optimization of mab glycosylation in CHO cell culture. *Chemical Engineering Science*, 318, 122162. <https://doi.org/10.1016/j.ces.2025.122162>

Majewski, N. I., Tejada, M. L., Betenbaugh, M. J., & Agarwal, N. (2020). N-Glycosylation of IgG and IgG-like recombinant therapeutic proteins: Why is it important and how can we control it? *Annual Review of Chemical and Biomolecular Engineering*, 11(1), 311–338. <https://doi.org/10.1146/annurev-chembioeng-102419-010001>

Maloney, A. J. (2021). Case studies in the modeling and control of continuous pharmaceutical manufacturing processes. Ph.D. thesis. Massachusetts Institute of Technology. <https://dspace.mit.edu/handle/1721.1/158315>

Matsumiya, S., Yamaguchi, Y., Saito, J., Nagano, M., Sasakawa, H., Otaki, S., Satoh, M., Shitara, K., & Kato, K. (2007). Structural comparison of fucosylated and nonfucosylated Fc fragments of human immunoglobulin G1. *Journal of Molecular Biology*, 368(3), 767–779. <https://doi.org/10.1016/j.jmb.2007.02.034>

McCamish, M., & Woollett, G. (2013). The continuum of comparability extends to biosimilarity: How much is enough and what clinical data are necessary? *Clinical Pharmacology & Therapeutics*, 93(4), 315–317. <https://doi.org/10.1038/clpt.2013.17>

Michaela, S. A. M., Radhakrishnan, D., Robinson, A. S., & Ogunnaike, B. A. (2014a). Identification of manipulated variables for a glycosylation control strategy. *Biotechnology and Bioengineering*, 111(10), 1957–1970. <https://doi.org/10.1002/bit.25251>

Michaela, S. A. M., Tran, K., Radhakrishnan, D., Robinson, A. S., & Ogunnaike, B. A. (2014b). Controllability analysis of protein glycosylation in CHO cells. *PloS One*, 9(2), e87973. <https://doi.org/10.1371/journal.pone.0087973>

Mullard, A. (2021). FDA Approves 100th monoclonal antibody product. *Nature Reviews Drug Discovery*, 20(7), 491–495. <https://doi.org/10.1038/d41573-021-00079-7>

Nagy, Z. K., & Braatz, R. D. (2003). Robust nonlinear model predictive control of batch processes. *AIChE Journal*, 49(7), 1776–1786. <https://doi.org/10.1002/aic.690490715>

Pacis, E., Yu, M., Autzen, J., Bayer, R., & Li, F. (2011). Effects of cell culture conditions on antibody n-linked glycosylation—what affects high mannose 5 glycoform. *Biotechnology and Bioengineering*, 108(10), 2348–2358. <https://analyticalsciencejournals.onlinelibrary.wiley.com/doi/abs/10.1002/bit.23200>. <https://doi.org/10.1002/bit.23200>

Pickhardt, R. (2000). Nonlinear modelling and adaptive predictive control of a solar power plant. *Control Engineering Practice*, 8(8), 937–947. [https://doi.org/10.1016/S0967-0661\(00\)00009-5](https://doi.org/10.1016/S0967-0661(00)00009-5)

Rawlings, J. B., Mayne, D. Q., & Diehl, M. (2017). *Model Predictive Control: Theory, Computation, and Design* (vol. 2). (2nd ed.). Nob Hill Publishing.

Sarna, S., Patel, N., Corbett, B., McCready, C., & Mhaskar, P. (2023). Process-aware data-driven modelling and model predictive control of bioreactor for the production of monoclonal antibodies. *The Canadian Journal of Chemical Engineering*, 101(5), 2677–2692. <https://doi.org/10.1002/cjce.24752>

Sauer, P. W., Burky, J. E., Wesson, M. C., Sternard, H. D., & Qu, L. (2000). A high-yielding, generic fed-batch cell culture process for production of recombinant antibodies. *Biotechnology and Bioengineering*, 67(5), 585–597. [https://doi.org/10.1002/\(SICI\)1097-0290\(20000305\)67:5<585::AID-BIT9>3.0.CO;2-H](https://doi.org/10.1002/(SICI)1097-0290(20000305)67:5<585::AID-BIT9>3.0.CO;2-H)

Scokaert, P. O. M., & Rawlings, J. B. (1999). Feasibility issues in linear model predictive control. *AIChE Journal*, 45(8), 1649–1659. <https://doi.org/10.1002/aic.690450805>

Sha, S., Agarabi, C., Cordon, K., Lee, D.-Y., & Yoon, S. (2016). N-Glycosylation design and control of therapeutic monoclonal antibodies. *Trends in Biotechnology*, 34(10), 835–846. <https://doi.org/10.1016/j.tibtech.2016.02.013>

Spahn, P. N., Hansen, A. H., Hansen, H. G., Arnsdorf, J., Kildegaard, H. F., & Lewis, N. E. (2016). A Markov chain model for n-linked protein glycosylation – towards a low-parameter tool for model-driven glycoengineering. *Metabolic Engineering*, 33, 52–66. <https://doi.org/10.1016/j.ymben.2015.10.007>

Van Rossum, G., & Drake, F. L. (2009). *Python 3 Reference Manual*. CreateSpace.

Vassiliadis, V. S., Sargent, R. W. H., & Pantelides, C. C. (1994). Solution of a class of multistage dynamic optimization problems. 2. problems with path constraints. *Industrial & Engineering Chemistry Research*, 33(9), 2123–2133. <https://doi.org/10.1021/ie00033a015>

Villiger, T. K., Scibona, E., Stettler, M., Broly, H., Morbidelli, M., & Soos, M. (2016). Controlling the time evolution of mAb N-linked glycosylation—Part II: Model-based predictions. *Biotechnology Progress*, 32(5), 1135–1148. <https://doi.org/10.1002/btpr.2315>

Walsh, G., & Walsh, E. (2022). Biopharmaceutical benchmarks 2022. *Nature Biotechnology*, 40(12), 1722–1760. <https://doi.org/10.1038/s41587-022-01582-x>

Wong, D. C. F., Wong, K. T. K., Goh, L. T., Heng, C. K., & Yap, M. G. S. (2005). Impact of dynamic online fed-batch strategies on metabolism, productivity and N-glycosylation quality in CHO cell cultures. *Biotechnology and Bioengineering*, 89(2), 164–177. <https://doi.org/10.1002/bit.20317>

Zupke, C., Brady, L. J., Slade, P. G., Clark, P., Caspary, R. G., Livingston, B., Taylor, L., Bigham, K., Morris, A. E., & Bailey, R. W. (2015). Real-time product attribute control to manufacture antibodies with defined N-linked glycan levels. *Biotechnology Progress*, 31(5), 1433–1441. <https://doi.org/10.1002/btpr.2136>