IMPERIAL COLLEGE LONDON

Enhanced understanding of protein glycosylation in CHO cells through computational tools and experimentation

Pavlos Kotidis

May 2021

Department of Chemical Engineering
Imperial College London
South Kensington Campus
London SW7 2AZ

A thesis submitted to Imperial College London in partial fulfilment of the requirements of the degree of Doctor of Philosophy

Declaration

I hereby certify that all material in this thesis that is not my own work has been appropriately acknowledged.

Pavlos Kotidis

London, U.K.

22/05/2021

Copyright

Abstract

Chinese hamster ovary (CHO) cells are the workhorse of the multibillion-dollar biopharmaceuticals industry. They have been extensively harnessed for recombinant protein synthesis, as they exhibit high titres and human-like post translational modifications (PTM), such as protein *N*-linked glycosylation. More specifically, *N*-linked glycosylation is a crucial PTM that includes the addition of an oligosaccharide in the backbone of the protein and strongly affects therapeutic efficacy and immunogenicity. In addition, the Quality by Design (QbD) paradigm that is broadly applied in academic research, necessitates a comprehensive understanding of the underlying biological relationships between the process parameters and the product quality attributes. To that end, computational tools have been vastly employed to elucidate cellular functions and predict the effect of process parameters on cell growth, product synthesis and quality.

This thesis reports several advancements in the use of mathematical models for describing and optimizing bioprocesses. Firstly, a kinetic mathematical model describing CHO cell growth, metabolism, antibody synthesis and *N*-linked glycosylation was proposed, in order to capture the effect of galactose and uridine supplementation on cell growth and monoclonal antibody (mAb) glycosylation. Subsequently, the model was utilized to optimize galactosylation, a desired quality attribute of therapeutic mAbs. Following the QbD paradigm for ensuring product titre and quality, the kinetic model was subsequently used to identify an *in silico* Design Space (DS) that was also experimentally verified. An elaborate parameter estimation methodology was also developed in order to adapt the existing model to data from a newly introduced CHO cell line, without altering model structure.

In an effort to reduce the burden of parameter estimation, the *N*-linked glycosylation submodel was replaced with an artificial neural network that was used as a standalone machine learning algorithm to predict the effect of feeding alterations and genetic engineering on the glycan distribution of several therapeutic proteins. In addition, a hybrid model configuration (HyGlycoM) incorporating the ANN-glycosylation model was also formulated to link extracellular process conditions to glycan distribution. The latter was found to outperform its fully kinetic equivalent when compared to experimental data.

Finally, a comprehensive investigation of mAb galactosylation bottlenecks was carried out. Five fed-batch experiments with different concentrations of galactose and uridine supplemented throughout the culturing period, were carried out and were found to present similar mAb galactosylation. In order to identify the bottlenecks that limit galactosylation, further experimental analysis, including the investigation of glycans microheterogeneity of CHO host cell proteins (HCPs), was conducted. The

experimental results were used to parameterize a novel and significant extension of the kinetic glycosylation model that simultaneously describes the *N*-linked glycosylation of both HCPs and the mAb product. Flux balance analysis was also used to analyse carbon and nitrogen metabolism using the experimental amino acid concentration profiles. In addition to the expression levels of the beta-1,4-galactosyltransferase enzyme, constraints imposed by the transport of the galactosylation sugar donor in the Golgi compartments and the consumption of resources towards HCPs glycosylation, were found to considerably influence mAb galactosylation.

Acknowledgments

I would like to start by thanking my supervisor, Dr. Cleo Kontoravdi, for all the support, guidance and the uncountable opportunities that she so generously offered to me throughout my studies. Apart from being a great supervisor and mentor, Cleo always made sure that all members of the group were doing well during the challenging times of their PhD studies; something that I very much appreciated.

I would also like to thank Dr. Karen Polizzi who welcomed me in her lab and made me feel part of her amazing group, while at the same time offering great advice and ideas. My PhD would have been far less exciting if I missed the opportunity to work with Karen and her group. In a similar note, I want to thank Dr. Ioscani Jimenez del Val, who spent a significant amount of time helping me understand the depth and magic of protein glycosylation and mammalian cell metabolism, especially during the first part of my studies. Dr. Masue Marbiah has been the person who very patiently taught me the molecular biology and mammalian cell line engineering techniques while at the same time being a very good friend, something that I will always be thankful for. I would also like to thank Dr. Cher Hui Goey for helping me in the beginning of my studies and re-introducing me to the lab. Moreover, I want to thank Prof. Mike Betenbaugh for giving me the chance to spend time in his lab and work with him during the entire duration of my studies. Similarly, I want to thank Dr. Oleksiy Klymenko and Dr. Fred Krambeck for our collaboration and our very fruitful conversations.

Another really important part of my studies was my relationship with the members of the group, who I consider great friends. I would like to thank Dr. Elli Makrydaki for her friendship and for her valuable help with western blots. I very much enjoyed the many hours we spent together in configuring the glycan analysis protocols. It goes without saying that Dr. Ignacio Moya Ramirez, Nacho, has been one of my dearest friends during my time at Imperial, but also an incomparable mentor and source of inspiration. Whilst our research paths only slightly crossed during these years, his unconditional love for science was the one of the strongest driving forces during my studies. I also want to specifically thank Dr. Sakhr Alhuthali for helping me in understanding the principles of mechanistic modelling and appreciating beer brewing. I would also like to thank Thanasi, Calum, Rod, Tom, Roberto, Zoltan, Sam and Chiara for making the everyday life in the office more enjoyable and amusing. I am also grateful for the help and patience from all members of Karen's lab.

My friends outside the college also deserve a big thank you, as they have been really helpful and understanding throughout these years. However, none of these would have been possible without the help, support and love of my family that has always been there for me. Their limitless support and love dates back to the first days that I remember myself and has only been growing through the years. Finally, I would like to thank Maria, who has suffered with me through the bad times of the PhD, has

laughed with me for all the successes and never thought it twice when leaving her home to move with me abroad.

I would like to dedicate this work to Eleni, that I dearly miss and that I am sure would have been extremely happy to witness the completion of my PhD studies.

Nomenclature

Symbol

mAb	Monoclonal antibody
CHO	Chinese hamster ovary
HEK	Human Embryonic Kidney
EMA	European Medicines Agency
FDA	Food and Drug Administration
ANN	Artificial neural network
PTM	Post translational modification
ER	Endoplasmic reticulum
GTase	Glycosyltransferase
Ig	Immunoglobulin
IgG	Immunoglobulin G
UTP	Uridine-triphosphate
UDP	Uridine-diphosphate
UMP	Uridine-monophosphate
CTP	Cytidine-triphosphate
CMP	Cytidine-monophosphate
GTP	Guanosine-triphosphate
GDP	Guanosine-diphosphate
ATP	Adenosine-triphosphate
ADP	Adenosine-diphosphate
AMP	Adenosine-monophosphate
PP-Dol	Dolichyl pyrophosphate
MSX	Methionine sulfoximine
Man	Mannose
NSD	Nucleotide sugar donor
Glc	Glucose
ERAD	Endoplasmic reticulum-associated degradation
CM	Cisternal maturation
VT	Vesicular transport
GalNAc	N-Acetylgalactosamine
GlcNAc	N-Acetylglucosamine

LacNAc	N-Acetyllactosamine
ManNAc	N-Acetylmannosamine
GlcN	Glucosamine
GDPMan	Guanosine-diphosphate mannose
GDPFuc	Guanosine-diphosphate fucose
UDPGlc	Uridine-diphosphate glucose
UDPGal	Uridine-diphosphate galactose
UDPGlcNAc	Uridine-diphosphate N-Acetylglucosamine
UDPGalNAc	Uridine-diphosphate N-Acetylgalactosamine
CMPNeu5Ac	Cytidine-monophosphate N-Acetylneuraminic
EPO	Erythropoietin
Fc	Crystallizable fragment
Fab	Antigen-binding fragment
CDR	Complementarity determining regions
CDC	Complement-Dependent Cellular Cytotoxicity
ADCC	Antibody-Dependent Cellular-Mediated Cytotoxicity
FcR	Fc receptor
FcγR	Fcγ receptor
NK	Natural-killer
PK	Pharmacokinetic
HM	High mannose
MR	Mannose receptor
ICH	International Conference on Harmonization
QTPP	Quality target product profile
CPP	Critical process parameters
QbD	Quality by Design
CQA	Critical quality attribute
DS	Design Space
TCA	Tricarboxylic acid
DO	Dissolved oxygen
CBM	Constraint-based models
GeM	Genome scale model
FBA	Flux balance analysis
MFA	Metabolic flux analysis

EFM	Elementary flux mode
KMM	Kinetic metabolic model
DDM	Data-driven model
ML	Machine learning
PCA	Principal component analysis
PLS	Partial least squares regression
PAT	Process analytical technology
CSTR	Continuous stirred-tank reactor
PFR	Plug and flow reactor
ODE	Ordinary differential equation
PDE	Partial differential equation
PCC	Pearson correlation coefficient
CCC	Concordance Correlation Coefficient
R^2	Coefficient of determination
PE	Parameter estimation
TSI	Total sensitivity index
SIT	Sensitivity index threshold
GSA	Global sensitivity analysis
cGSA	Constraint global sensitivity analysis
Gal	Galactose
Urd	Uridine
HL	Hidden layer
AAE	Average absolute error
DAO	Diamine-oxidase
PD	Parameter deviation
CI	Confidence interval
HCP	Host cell protein
IVCD	Integral viable cell density
GU	Galactose & uridine
MS	Mass spectrometry
HPLC	High-performance liquid chromatography
LC	Liquid chromatography
HPAEC	High-performance anion exchange chromatography
CE	Capillary electrophoresis

NMR	Nuclear magnetic resonance
MALDI-TOF	Matrix-assisted laser desorption/ionization
CGE-LIF	Capillary gel electrophoresis – laser induced fluorescence

Enzymes

GS	Glutamine synthetase
OST	Oligosaccharyltransferase
ERManI	Endoplasmic reticulum mannosidase I
UGGT	UDP-glucose/glycoprotein glucosyl transferase
GnTI	Alpha-1,3-mannosyl-glycoprotein 2-beta-N-acetylglucosaminyltransferase
GnTII	Alpha-1,6-mannosyl-glycoprotein 2-beta-N-acetylglucosaminyltransferase
ManI	Mannosyl-Oligosaccharide 1,2-Alpha-Mannosidase I
ManII	Mannosyl-Oligosaccharide 1,3-1,6-Alpha-Mannosidase
GnTIII	Beta-1,4-mannosyl-glycoprotein 4-beta-N-acetylglucosaminyltransferase
GnTIV	Alpha-1,3-mannosyl-glycoprotein 4-beta-N-acetylglucosaminyltransferase
GnTV	Alpha-1,6-mannosyl-glycoprotein 6-beta-N-acetylglucosaminyltransferase
a6FucT	Alpha-1,6-fucosyltransferase
a3FucT	Alpha-1,3-fucosyltransferase
b4GalT	Beta-N-Acetylglucosaminylglycopeptide beta-1,4-galactosyltransferase
a3GalT	N-Acetyllactosaminide alpha-1,3-galactosyltransferase
iGnT	N-Acetyllactosaminide beta-1,3-N-acetylglucosaminyltransferase
a3SiaT	N-Acetyllactosaminide alpha-2,3-sialyltransferase
a6SiaT	Beta-galactoside alpha-2,6-sialyltransferase
NST	Nucleotide sugar transporter
LDH	Lactate dehydrogenase

Model parameters and variables

V	Culture volume (L)
F_{in}	Feeding flowrate ($L \cdot h^{-1}$)
F_{out}	Sampling/outlet flowrate ($L \cdot h^{-1}$)
X_v	Viable cell density ($cell \cdot L^{-1}$)
μ	Specific cell growth rate (h^{-1})
μ_{death}	Specific cell death rate (h^{-1})
μ_{max}	Maximum specific cell growth rate (h^{-1})
$\mu_{death,max}$	Maximum specific cell death rate (h^{-1})
$[Metabolite]$	Extracellular concentration of metabolites (mM)

m_{Glc}	Maintenance coefficient of glucose (mmol·cell^{-1}·h^{-1})
c_{Gal}	Regulating concentration of galactose (mM)
n_{Gal}	Factor of glucose/galactose specific consumption rate (-)
f_{Gal}	Factor of glucose/galactose specific consumption rate (-)
Lac_{max1}	Kinetic constant for lactate consumption (mM)
Lac_{max2}	Kinetic constant for lactate consumption (mM)
f_{Gln}	Glutamine secretion factor (-)
$K_{regulator}$	Regulating factor for galactose consumption (mM)
$[Gln_{intra}]$	Intracellular concentration of glutamine (mM)
$[mAb]$	Concentration of secreted monoclonal antibody (mg·L^{-1})
$mGal$	Mab galactosylation (mg·L^{-1})
Vol_{DS}	Volume of the Design Space (-)
$[Metabolite_{feed}]$	Metabolite concentration in the feed (mM)
$q_{metabolite}$	Specific uptake/synthesis rate for each metabolite (mmol·cell^{-1}·h^{-1})
q_{mAb}	MAb specific productivity (mg·cell^{-1}·h^{-1})
$Y_{X_{Metabolite}}$	Yield of cell biomass for each metabolite (cell·mmol^{-1})
$Y_{Met1/Met2}$	Yield of one metabolite (Met1) from another (Met2) (mmol$_{Met1}$·mmol$_{Met2}^{-1}$)
K_{Met}	Monod-type saturation constant for each metabolite (mM)
KI_{Met}	Inhibition constant for each metabolite (mM)
$K_{d,Met}$	Monod-type constants for cell death (mM)
$V_{max,i}$	Maximum turnover rate of ith reaction (mmol$_{NSD}$·L$_{cell}^{-1}$·h^{-1})
$K_{Mi_{NSD}}$	Saturation constant for the examined NSD on ith reaction (mM)
$[NSD]$	Intracellular NSD concentration (mM)
r_i	Reaction in the NSD pathway (mmol·L$_{cell}^{-1}$·h^{-1})
$r_{i_{sink}}$	NSD sinking reactions (mmol·L$_{cell}^{-1}$·h^{-1})
Y_{mAb_X}	Yield of mAb from cell growth (mg·cell^{-1})
m_{mAb}	Non-growth associated term for mAb synthesis (mg·cell^{-1}·h^{-1})
KI_i	Inhibition constant in the ith reaction (mM)
$F_{out_{NSD}}$	Flux of each NSD towards the Golgi compartment (mmol$_{NSD}$·L^{-1}·h^{-1})
$K_{TP_{NSD}}$	Saturation constant for the examined NSD transport (mM)
V_{cell}	Cellular volume (L)
V_{Golgi}	Golgi volume (L)
$r_{NSD_k}^{mAb,glyc}$	Uptake rate of the k^{th} NSD towards mAb glycosylation (mmol$_{NSD}$·L^{-1}·h^{-1})
$r_{G,j}$	Rate of the jth glycosylation reaction (mmol$_{glycan}$·L^{-1}·h^{-1})

$[OS_i]$	Oligosaccharide concentration (µM)
Vel_{Golgi}	Velocity of protein transport in the Golgi (Golgi length·min^{-1})
$[Enz_k]$	kth glycosyltransferase concentration in the Golgi (µM)
$[NSD_{G,k}]$	NSD concentration in the Golgi (µM)
$K_{d,i}$	Dissociation constant of the complex between the OS and the enzyme (µM)
$K_{d,k}$	Dissociation constant between theyougttgt NSD and the enzyme (µM)
GI	Galactosylation Index
$Fin_{NSD}(c)$	Transport rate of each NSD in each Golgi compartment (c) (min^{-1})
k_f	Rate-limiting turnover rate for each enzyme (min^{-1})
$distr_{NST}(c)$	NST distribution in each Golgi compartment
$[NSD]_{cytosol}$	NSD concentration in the cytosol (µM)
K_m^{NST}	Monod-type saturation constant for each NST-catalysed reaction (µM)

Conventions

- Whilst capital letters are mainly used to describe human genes in literature, gene names were represented with capital letters throughout this thesis for all mammalian cells to avoid confusion.
- The terms glycoform, glycan and oligosaccharide are interchangeably used throughout the text.
- The terms GnGn/G0, GnGnF/G0F, AGn/G1, AGnF/G1F, AA/G2 and AAF/G2F are interchangeably used throughout the text.

Table of Contents

List of figures

Appendix

List of tables

Appendix

Introduction

The last two decades have witnessed a rapid growth of biopharmaceuticals development for tackling severe human diseases, such as cancer and autoimmune disorders. Biopharmaceuticals (also known as biotherapeutics or biologics) are complex natural molecules synthesized in living systems, such as proteins and nucleic acids, that exhibit *in vivo* therapeutic properties when administrated to patients. In contrast to the chemically synthesized therapeutics, biopharmaceuticals are usually large macromolecules. The reports of biopharmaceutical benchmarks published every four years by Gary Walsh, showcase the extraordinary growth of biologics and the various diseases that they are able to tackle[1-4]. As of 2018, there are 316 active licenses of biopharmaceuticals in the European Union (EU) and the United States (US)[4]. Interestingly, slightly more than 50% of the biopharmaceuticals approved between January 2014 and July 2018 (129 in total) were monoclonal antibodies (mAbs), nearly double the respective percentage for the previously examined period between 2010 and 2014[4]. The year of 2018 also marked the approval, both in the EU and US, of the first gene-silencing small interfering RNA - based (siRNA) biologic for the treatment of polyneuropathy caused by hereditary transthyretin (hATTR) amyloidosis[5-6].

Whilst the current thesis examines the synthesis and quality profile of several proteins, both recombinant and naturally produced, the main focus of the research presented herein is placed on mAbs. Monoclonal antibodies are large proteins with high molecular weight that have been found to exhibit therapeutic properties for several human diseases, such as cancer and rheumatoid arthritis[7]. As the approval rates demonstrate, mAbs currently consist the most popular group of biologics, with adalimumab (Humira®) constantly holding the position of the top-selling biotherapeutic and pembrolizumab (Keytruda®) exhibiting a rapid growth in sales[4, 8]. Another important element of the biopharmaceutical industry is the use of mammalian cells for the production of the recombinant products. The rationale behind the extensive use of mammalian hosts lies within their ability to achieve high titres and undertake human-like post translational modifications, ensuring that way the correct molecular structure and activity of the biologic, while at the same time minimizing immunogenicity[9]. Non-mammalian systems account only for a mere portion of biologics production, especially in the industrial level. While the use of bacteria (i.e. *Escherichia coli*), insect cells (i.e. baculovirus expression system) and yeast (i.e. *Saccharomyces cerevisiae*, *Pichia pastoris*) is gaining momentum in the area of biopharmaceuticals production, these cell lines are still a long way from becoming the dominant hosts[10-12].

Among the mammalian systems utilized for recombinant protein production, several hosts have been proposed, employed and studied through the years, such as murine myeloma cells (NS0 and Sp2/0)[13-

[14], human embryonic kidney cells (HEK293)[15], mouse hybridoma[16] and Chinese hamster ovary (CHO) cells [17]. Undoubtedly, CHO cells have been harnessed the most for biotherapeutics production, due to their remarkable ability of simultaneously achieving high titres (up to 10 $g \cdot L^{-1}$) and ensuring human-like protein quality attributes, such as *N*-linked glycosylation[18]. Recombinant tissue plasmogen activator (tPA) was the first biologic synthesized in CHO cells to be approved in 1987[19]. Since then, CHO cells have quickly become the major host cell line for the production of recombinant proteins and more specifically mAbs[20], employed as host platforms for approximately 84% of the antibodies approved between 2014 and 2018[4]. While several CHO clonal cell lines have been developed through the years, the three most common are CHO-K1, CHO-S and CHO-DG44, each one of which exhibits unique genotypic and phenotypic characteristics[21].

In addition, the genomic sequence of the CHO-K1 cell line was published in 2011[22], paving the way for a deeper understanding of the cell line mechanisms and offering many opportunities for harnessing the potential of CHO cells within the biopharmaceutical context. Follow-up studies elaborating on the sequence of individual chromosomes and the discrepancies between different CHO clones have further enriched our knowledge on the CHO genome[23-25]. The analysis of CHO cell proteome[26-27] and the reconstruction of its metabolic pathways[28], alongside with the several efforts to systematically identify differences between clones in the proteomic[29] and transcriptomic level[30-32] have led to the establishment of a hardly replaceable background knowledge for these hosts. In addition, all information around CHO genome, transcriptome and proteome is easily accessible on the CHOgenome.org platform[33]. Moreover, the development of selection systems, like the dihydrofolate reductase deficiency ($DHFR^{-/-}$) and glutamine synthetase (GS) systems, has enabled the establishment of robust frameworks for identifying high-producers during cell line development[34].

It is important to note that biotherapeutics are usually associated with high costs, expensive price-tags and long approval procedures due to the use of cellular platforms for drug production and the extensive clinical trials required for approval. However, with patents of several reference biotherapeutics expiring, there is a rapid increase in the use and development of biosimilars, meaning the new products that exhibit statistically similar therapeutic properties, safety and efficacy with the reference molecule[35]. Biosimilars require extensive analytical comparative studies for ensuring satisfactory biosimilarity with the original therapeutic, but at the same time undergo shorter clinical trials, reducing that way the overall cost of the final product and offering considerable economic benefits on the biopharma market[36]. The European Medicines Agency (EMA) has already approved 55 biosimilars, with the US Food and Drug Administration (FDA) following with 26 approvals, including that of several therapeutic proteins and targeting a range of severe diseases such as rheumatoid arthritis and cancer[37-38]. In addition, clinical studies have demonstrated the equivalent safety and

efficacy between biosimilars and the original products for both cancer[37, 39] and autoimmune diseases[40-42]. During biosimilars development, several protein and process parameters are considered in order to ensure the desired quality and potency of the product, with a major emphasis placed on the glycosylation profile of the new product and its similarity to the reference therapeutic[43-46].

However, both reference and similar biotherapeutics are produced in living organisms and therefore, variabilities in the quality attributes are expected. In an effort to ensure product safety and efficacy, both the industrial and academic sectors are actively exploring the Quality by Design (QbD) paradigm during product development and manufacturing. The QbD principle ensures the establishment of acceptable ranges within which the critical quality attributes (CQAs) of the product, such as protein *N*-linked glycosylation, can vary[47]. Additionally, QbD strongly encourages the comprehensive investigation of the underlying mechanisms that relate process and material parameters to the final product. As the complexity and the multi-interdependencies of cellular processes can result in unexpected product profiles, the deep understanding of such relations is of great worth for the biopharmaceutical industry. Another side benefit of this necessity for comprehension of pure biological functions that is often neglected, is the inevitable bridging between academic research and the industrial sector. Both parts are interested in improving their understanding on fundamental cellular processes, in order to, eventually, design safer, cheaper and more efficient pharmaceutical products.

Whilst QbD offers a structured framework for ensuring product quality, the experimentation required for an exhaustive investigation of the relationship between process variables and product targets is considerably costly and time-consuming. For that reason, several computational tools have emerged for approximating cellular behaviour and *in silico* predicting the cellular response to altered process conditions[48]. The aforementioned computational tools are based on several different forms of mathematical models, including kinetic, constraint-based and data-driven approaches, with all models aiming to describe cell metabolism. Kinetic models align well with the QbD principle as they are based on mechanistic principles and relationships between the examined variables, and have therefore been extensively utilized to optimize titre[49] and control protein *N*-linked glycosylation[50] by designing optimal feeding strategies. Constraint-based models are primarily employed for guiding cell line engineering as they successfully capture the detailed intracellular metabolism, and correctly evaluate the effect of gene modifications on cell growth and productivity[51-52]. Finally, data-driven, statistical and machine learning approaches are constantly gaining more and more attention for describing the cell metabolism, as the automation and utilization of deep omics analyses such as metabolomics, transcriptomics, glycomics and proteomics is becoming increasingly accessible, especially for microbial hosts[53-54]. Overall, the use of mathematical models to optimize bioprocess parameters, the genome

of the cell line in use or product quality attributes such as glycosylation, immensely contributes to expediting product and process development, while at the same time requiring minimal experimentation and reducing costs.

In this thesis we harness various mathematical tools in order to optimize the culturing process and, ultimately achieve the desired final titre and *N*-linked glycosylation profile. Firstly, the literature review sets the basis for the rest of the thesis, describing in detail key aspects of the thesis and demonstrating the importance of the work presented herein (Chapter 1). The materials and methods were integrated with the respective results in each chapter, to enable a better understanding of each study in a paper-like format. Initially, a mechanistic mathematical model was developed (Chapter 2) that holistically described CHO cell metabolism and mAb glycosylation in fed-batch experiments, in which galactose and uridine were supplemented in order to manipulate product glycosylation. Following training, the model was utilized for designing an optimal feeding strategy for achieving maximum galactosylation. Subsequently, the model was employed for *in silico* identifying a Design Space (DS) of feeding regimes within the QbD principle, that would ensure the desired final titre and glycosylation profile (Chapter 3). The model-based DS was successfully validated against experimental data, indicating that, indeed, kinetic mechanistic models can be utilized for applying the QbD paradigm, while at the same time minimizing experimentation. The metabolic part of the mechanistic model was then subjected to sensitivity analysis in order to identify the significant parameters that, when re-estimated, capture the variability between different batches. Utilizing the developed sensitivity analysis framework, the metabolic model was adapted to successfully describe the metabolism of a different CHO cell line that achieves considerably higher cell densities and titre, without introducing any changes to the equations and structure of the model. The model adaptation to the new cell line paves the way towards the development of a global mechanistic model for describing the cellular metabolism of various CHO cell lines with minor modifications in its parameter values.

In order to tackle the non-linearities that are encountered in the glycosylation process, a machine learning framework, based on feedforward artificial neural networks (ANNs), was developed (Chapter 4). The ANN model was validated against several in-house experiments and data from literature, and was found capable of describing the glycosylation of four different recombinant proteins and within different process conditions and gene engineering experiments. A hybrid model configuration, including a mechanistic metabolism and the ANN glycosylation model, was also found to outperform the respective holistically mechanistic model. Following the modelling work, five fed-batch experiments with feeding concentrations derived from the *in silico* Design Space were performed, in order to further investigate the effect of galactose and uridine addition on cell metabolism and

glycosylation (Chapter 5). More specifically, the experimental work, combined with an updated mechanistic glycosylation model including additional parts of the secretory pathway, and flux balance analysis, enabled the elucidation of the reasons behind the observed limitations in mAb galactosylation and the investigation of metabolic differences between the feeding and the control experiments. Finally, all results presented herein were summarized in the final chapter, alongside with the limitations and the future perspectives as derived from this thesis (Chapter 6).

Chapter 1 Literature review

1.1 Glycosylation

Glycosylation is a post-translational modification (PTM) that initiates in the Endoplasmic Reticulum (ER) and further occurs in the Golgi apparatus, holding a significant role in immune system efficiency and antigen recognition[55-57]. As a principle, glycosylation includes the enzymatic covalent attachment (glycosidic bond) of a carbohydrate (glycan or glycoconjugate) to the backbone of proteins or lipids, and mainly appears in eukaryotic cells. An early estimation has identified that approximately 50% of human proteins are glycosylated[58]. Glycosylation of mammalian small non-coding RNA has also been recently reported[59]. There is a variety of different types of glycosylation, including: *N*-linked, *O*-linked, phosphoglycosylation, *C*-mannosylation and glycosylphosphatidylinositol-anchor (GPI-anchor) synthesis. Additionally, it has been identified that *N*-linked and *O*-linked glycosylation occur in both bacteria and archaea[60-65]. Bacterial glycosylation follows a significantly different pathway compared to eukaryotic glycosylation and can considerably influence a microorganism's capability to invade cells[66]. Interestingly, human glycoproteins can be further modified by extracellular glycosyltransferases circulated in blood and platelets, through a process called extrinsic glycosylation[67-69].

Moreover, glycosylation is not template-driven. The extent of lipid and protein glycosylation depends on several factors; from the structure of the synthesized macromolecule, to the availability of substrates and the expression levels of glycosylation enzymes. The variety of enzymes catalysing glycan synthesis, conventionally and collectively referred to as glycosidases and glycosyltransferases (GTases), constitute only a small part of the enzymatic system that determines the extent of processing. Enzymes responsible for co-substrate synthesis and transport to the organelles' lumen, are an additional part of the extensive network that affects glycosylation. Once the strong dependence of sugar-donor synthesis (glycosylation co-substrates) on the central carbon metabolism is considered, the extent of glycosylation dependence on other cellular processes becomes immense. Due to the multiple dependencies of glycosylation and the different levels of cellular control over its extent, glycosylation is considered the most diverse category of PTMs[70-71].

Whilst deficiencies in any glycosylation type can have a significant impact on macromolecule structure and result in abnormal cellular behaviour[56], the major focus of this Literature Review is placed on protein *N*-linked glycosylation, due to its paramount importance for therapeutic protein synthesis, and relevance to the rest of this thesis.

1.1.1 Protein *N*-linked Glycosylation

Protein glycosylation includes the covalent attachment of a glycan in the nitrogen atom of an asparagine (*N*-linked) or in the oxygen atom of a serine/threonine (*O*-linked) amino acid residue of the polypeptide backbone. Extensive amounts of *O*-linked glycans are observed in specific proteins, such as mucins[72]. More specifically, mucin-type *O*-linked glycosylation initiates with the addition of a N-Acetylgalactosamine (GalNAc) molecule in the serine/threonine residue[73-74]. Although *O*-linked oligosaccharides have been observed in human immunoglobulin A (IgA) antibodies [75-80], this post-translational modification is not common in antibody molecules.

The effects and importance of *N*-linked glycosylation have been better characterized compared to *O*-linked glycosylation, partly because of the improved accessibility on *N*-linked glycans analysis[80]. The fundamental steps of *N*-linked glycosylation, which is part of the cellular secretory pathway[81], include: 1) the formation of a dolichyl pyrophosphate (PP-Dol) membrane precursor oligosaccharide ($Glc_3Man_9GlcNAc_2$); 2) the transfer and attachment of the PP-Dol oligosaccharide precursor, through enzymes called oligosaccharyltransferases (OSTs), in the polypeptide backbone of the protein; and 3) the further modification of the attached precursor in the ER and Golgi apparatus, through several glycosidases and glycosyltransferases[82]. Commonly encountered *N*-linked glycans and the adopted symbol nomenclature for graphical representation used in the entirety of this thesis can be found in Fig. 1.1[83]. In addition, throughout the glycosylation process, several nucleotide phosphates are utilized by the cells, including: uridine-triphosphate (UTP), uridine-diphosphate (UDP), uridine-monophosphate (UMP), cytidine-triphosphate (CTP), cytidine-monophosphate (CMP), guanosine-triphosphate (GTP), guanosine-diphosphate (GDP), adenosine-triphosphate (ATP), adenosine-diphosphate (ADP) and adenosine-monophosphate (AMP).

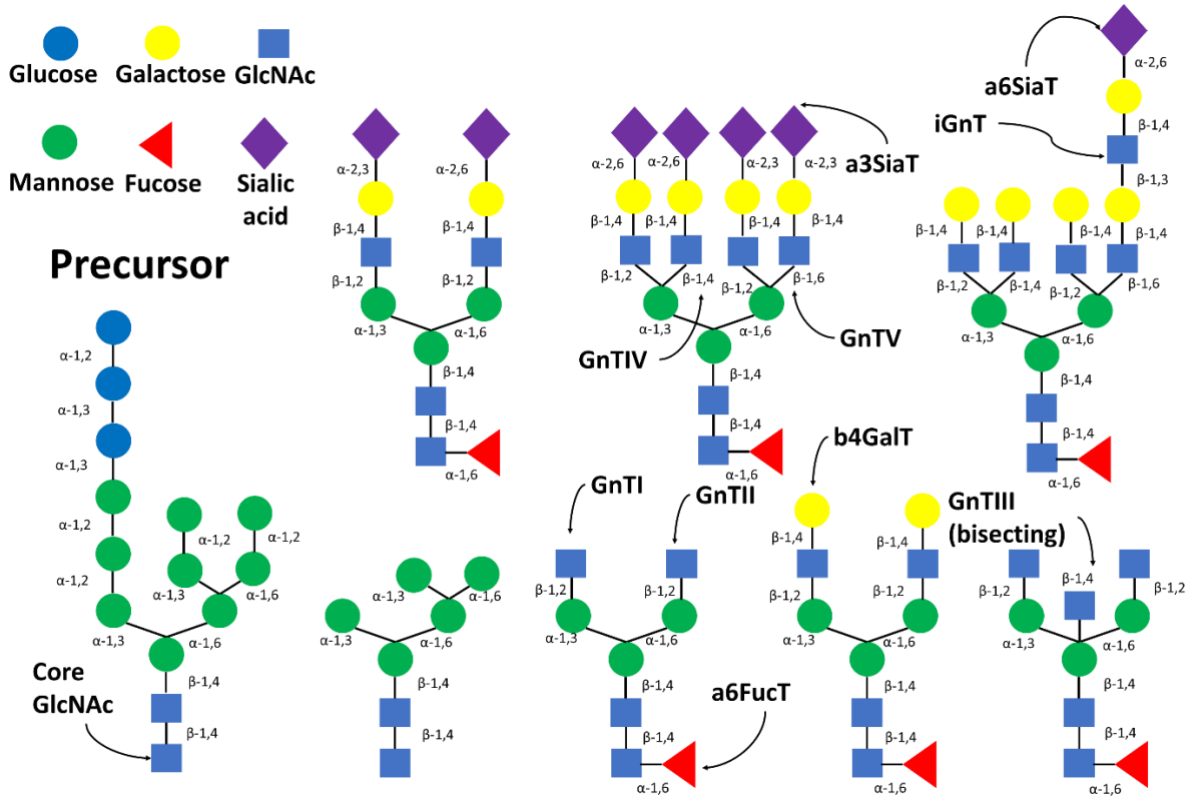

Figure 1.1: Graphical representation of glycan structures that are commonly observed in *N*-linked glycosylation, including the precursor oligosaccharide. The major glycosyltransferases involved in *N*-linked glycosylation are also reported. Note that sialic acid can be attached with both α-2,3 and α-2,6 bonds without any structural restrictions or preference. The nomenclature used throughout this thesis and presented in this graph for the graphical representation of glycan structures, can be found in Varki et al.[83].The abbreviation of the depicted glycosyltransferases can be found in Table 1.1.

The oligosaccharide precursor is synthesized through the sequential processing of the PP-Dol anchor. Initially, two N-Acetylglucosamine (GlcNAc) residues bind through a pyrophosphate linkage to PP-Dol that is located in the cytoplasmic side of the ER-membrane. The addition of the two GlcNAc molecules is followed by the further attachment of five mannose (Man) residues in the forming oligosaccharide. The energy for catalyzing the aforementioned reactions is provided by the saccharide-phosphate bond of the nucleotide sugar donors (NSDs) UDPGlcNAc and GDP-mannose (GDPMan). NSDs are co-substrates of glycosylation, offering the monosaccharide molecule and the required energy for the enzymatic reactions to take place. Subsequently, the formed dolichol-$Man_5GlcNAc_2$ residue is translocated, facing the luminal side of the ER, and becomes accessible to the ER-resident enzymes. Four additional mannose and three glucose (Glc) residues are added to the dolichol-$Man_5GlcNAc_2$, forming the dolichol-$Glc_3Man_9GlcNAc_2$ precursor.

Several, alpha-mannosyltransferases (ALG1, ALG2, ALG3, ALG9, ALG11 and ALG12) and alpha-glucosyltransferases (ALG6, ALG8 and ALG10) are involved in precursor synthesis, as shown in Fig. 1.2. The crystal structure of the yeast ALG6 glucosyltransferase, recently described in Bloch et al.[84],

revealed a catalytic asparagine residue conserved between all eukaryotic ER-luminal glycosyltransferases. Initially, the monosaccharide mannose and glucose residues are bound to a lipid dolichol in the cytoplasmic side of the ER-membrane. It is important to note that the mechanism behind the translocation (or flipping) for both the dolichol-monosaccharides and dolichol-$Man_5GlcNAc_2$ molecules is not yet well understood and characterized. ER- resident flippases have been attributed for carrying out the ATP-fueled transbilayer translocation of the dolichol-bound molecules[85-88]. Finally, OSTs transfer the $Glc_3Man_9GlcNAc_2$ oligosaccharide in the asparagine residue of the polypeptide backbone, initiating that way the core process of protein *N*-linked glycosylation. More specifically, the $Glc_3Man_9GlcNAc_2$ oligosaccharide covalently binds in an amino acid sequence of asparagine/X/serine or asparagine/X/threonine, with X being any amino acid apart from proline. PP-Dol, formed as a by-product of OST activity, is subsequently recycled for new precursor assembly. Identification of potential *N*-linked glycosylation sites through computational tools and the investigation of the extend of site occupation, have been subjects of several research studies[89-92]. Petrescu et al.[93] identified an approximate 65% occupation rate over 2592 sequons in 506 glycoproteins, 70% of which were of asparagine/X/threonine type and the remaining 30% of asparagine/X/serine.

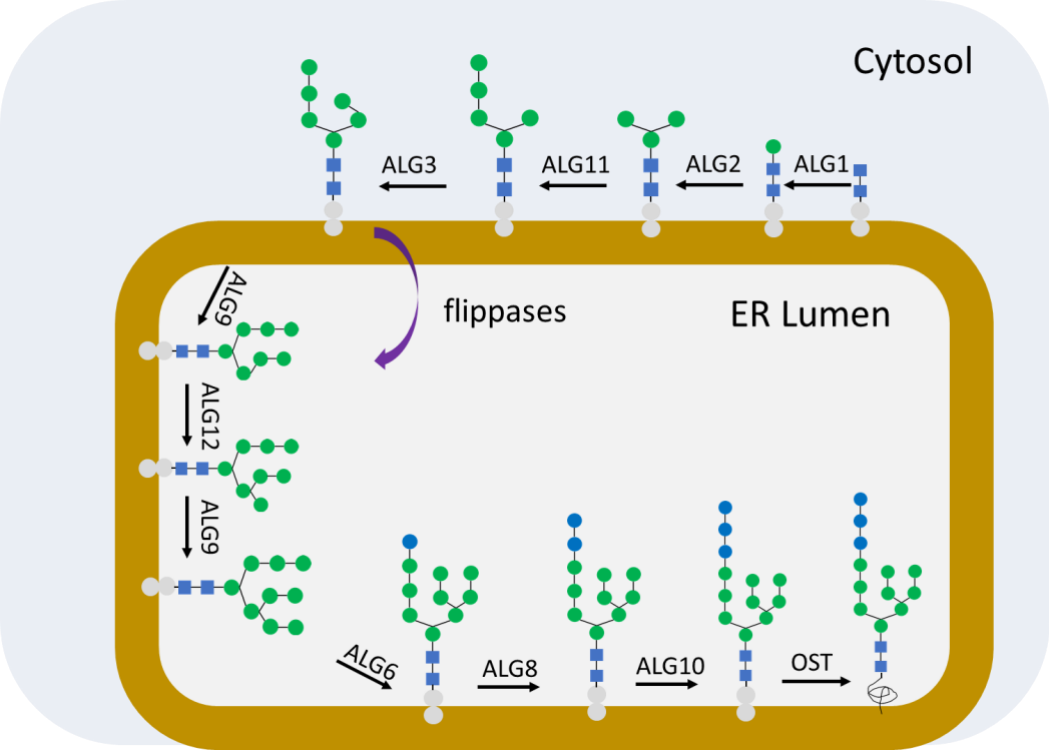

Figure 1.2: Precursor synthesis in the cytosol and the ER lumen. The precursor glycan ($Glc_3Man_9GlcNAc_2$) is synthesized while attached to PP-Dol and is then transferred to the polypeptide backbone through the OSTs activity.

The synthesis of the $Glc_3Man_9GlcNAc_2$-protein precursor initiates several protein quality control steps in the ER, part of the rigorous and complex ER quality control (ERQC) process[94-96]. Sequentially, ER-resident glucosidases I and II remove the glucose molecules, followed by the partial trimming of a single mannose molecule from the ER-alpha-1,2-mannosidase I (ERManI) enzyme[97]. $GlcMan_9GlcNAc_2$-protein, an intermediate of the aforementioned sequential steps, is recognized by ER-resident chaperones, namely binding immunoglobulin protein (BiP), calnexin (CNX) and calreticulin (CRT), that ensure proper oxidative folding and protein maturation[98-100]. Prior to ERManI activity and Golgi transport, UDP-glucose/glycoprotein glucosyl transferase (UGGT) evaluates protein conformation. If the glycoprotein is misfolded, UGGT triggers the re-glucosylation of the glycan, re-initiating that way the ERQC and the activity of BiP, CNX and CRT enzymes. Interestingly, ERManI has been identified as a critical enzyme for the regulation of the ER-associated degradation (ERAD) of misfolded proteins, further trimming the oligosaccharide for the formation of the degradation-signaling $Man_5GlcNAc_2$ or $Man_6GlcNAc_2$ glycans[101-105]. Researchers have recently partially reconstructed the *in vitro* synthesis of the precursor glycan, in an effort to *in vitro* assemble the $Man_9GlcNac_2$ structures[106].

Following the partial trimming from ERManI, both $Man_9GlcNAc_2$- and $Man_8GlcNAc_2$-protein complexes are transported from the ER to the Golgi environment. Whilst precursor synthesis and transit to the Golgi apparatus is a common process for all eukaryotic cells, glycosylation follows a remarkably different pathway in yeast, when compared to mammalian cells. Yeast glycosylation includes an extensive addition of mannose residues in the $Man_8GlcNAc_2$ sugar, leading to the synthesis of hyper-mannosylated glycans[107]. In mammalian cells, Golgi-residing alpha-1,2-mannosidase I (ManI) catabolizes several mannose molecules of the $Man_8GlcNAc_2$ glycan for the formation of $Man_5GlcNAc_2$. Subsequently, N-Acetylglucosaminyltransferase I (GnTI) initiates glycan rebuilding by attaching a GlcNAc molecule to the alpha-1,3-arm of the $Man_5GlcNAc_2$ glycan (Fig. 1.3). Following the first glycan elongation step, alpha-1,3-1,6-mannosidase II (ManII) further trims the glycan for the formation of $Man_3GlcNAc_3$. The latter consists the fundamental building block for the synthesis of more complex glycoforms. The remaining part of the glycosylation pathway includes the activity of several GTases, as shown in Table 1.1. GTases exhibit an unparallel enzymatic promiscuity, being able to catalyze various substrates during the elongation process. To that end, protein glycosylation carries high degrees of heterogeneity, with identical proteins carrying different glycans on the same *N*-linked glycosylation site (microheterogeneity). Therefore, glycoproteins display a diverse distribution of differentially processed glycans in their backbone. Upon glycosylation completion, proteins are secreted to the cytosol for appropriate localization.

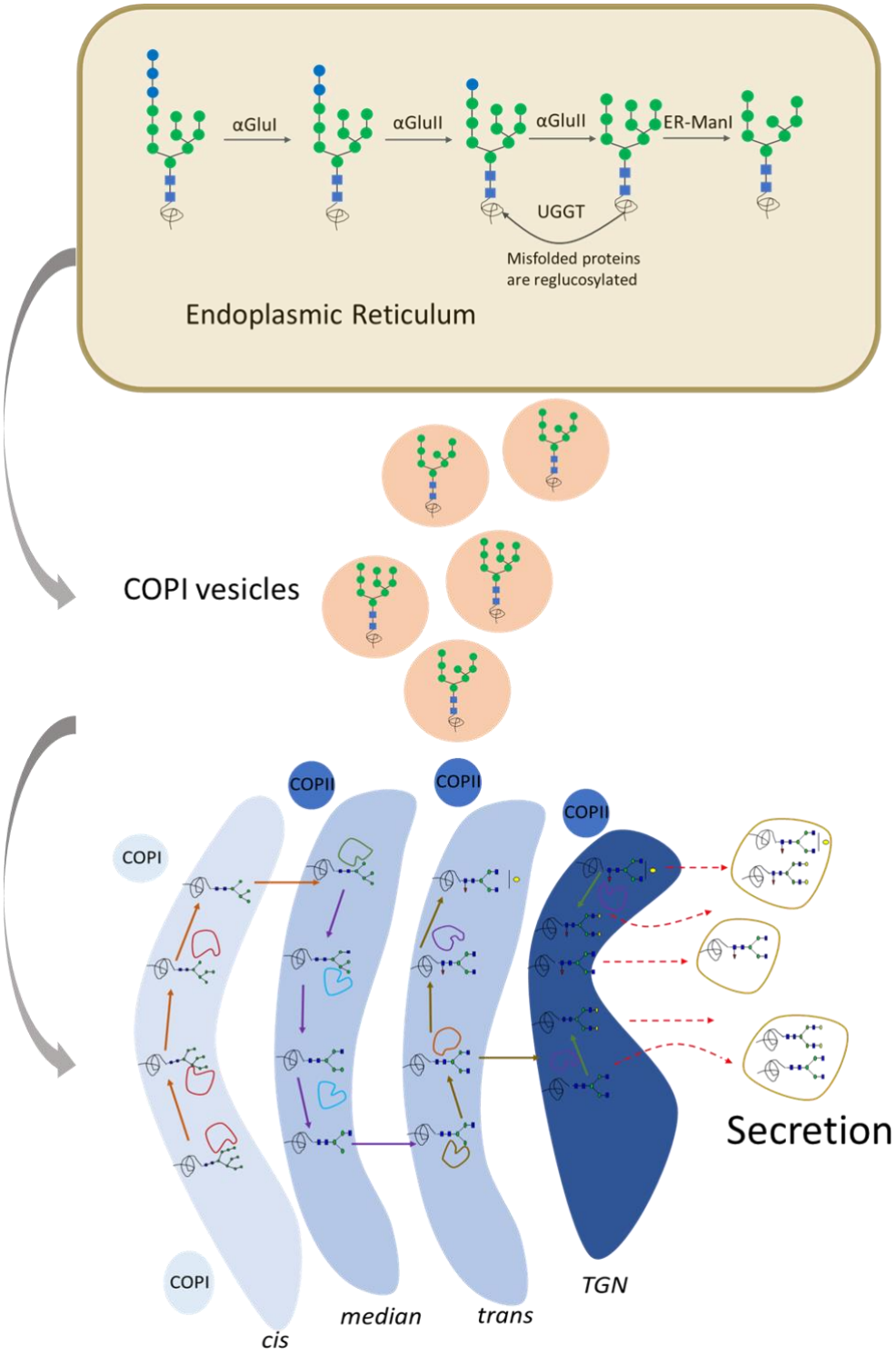

Figure 1.3: Schematic representation of *N*-linked glycosylation. Correctly folded and matured proteins carrying $Man_9GlcNAc_2$ and $Man_8GlcNAc_2$ glycans transverse towards the Golgi-cisternae where Golgi-resident glycosidases and glycosyltransferases further modify the *N*-linked oligosaccharide prior to secretion.

Table 1.1: Major enzymes and isoforms involved in protein *N*-linked glycosylation.

Enzyme	Symbol	Expasy entry (EC)	Function	Human Gene(s)	Substrate/ Donor
Mannosyl-oligosaccharide α-glucosidase I	αGluI	3.2.1.106	Hydrolysis of the first α-1,2-glucose molecule in the $Glc_3Man_9GlcNAc_2$ glycan	MOGS	H_2O
Glucan 1,3-alpha-glucosidase	αGluII	3.2.1.84	Sequential hydrolysis of two α-1,3-glucose molecules of the $Glc_2Man_9GlcNAc_2$ glycan	GANAB	H_2O
UDP-glucose:glycoprotein glucosyltransferase	UGGT	2.4.1.-	Reglucosylation of misfolded proteins in the ER	UGT1A, UGT1A1	UDPGlc
ER mannosyl-oligosaccharide 1,2-alpha-mannosidase	ER-ManI	3.2.1.209	Hydrolysis of a single α-1,2-mannose molecule from the $Man_9GlcNAc_2$ glycan	MAN1B1	H_2O
Mannosyl-Oligosaccharide 1,2-Alpha-Mannosidase I	ManI	3.2.1.113	Sequential hydrolysis of α-1,2-mannose molecules	MAN1A1, MAN1A2, MAN1C1	H_2O
Mannosyl-Oligosaccharide 1,3-1,6-Alpha-Mannosidase	ManII	3.2.1.114	Hydrolysis of α-1,3-mannose and α-1,6-mannose molecules	MAN2A1, MAN2A2	H_2O
Alpha-1,3-mannosyl-glycoprotein 2-beta-N-acetylglucosaminyltransferase	GnTI	2.4.1.101	Transfer of a β-1,2-GlcNAc residue in the α-1,3-branch of the glycan	MGAT1	UDPGlcNAc
Alpha-1,6-mannosyl-glycoprotein 2-beta-N-acetylglucosaminyltransferase	GnTII	2.4.1.143	Transfer of a β-1,2-GlcNAc residue in the α-1,6-branch of the glycan	MGAT2	UDPGlcNAc
Beta-1,4-mannosyl-glycoprotein 4-beta-N-acetylglucosaminyltransferase	GnTIII	2.4.1.144	Transfer of a β-1,4-GlcNAc residue in the bisecting β-1,4-branch of the glycan	MGAT3	UDPGlcNAc
Alpha-1,3-mannosyl-glycoprotein 4-beta-N-acetylglucosaminyltransferase	GnTIV	2.4.1.145	Transfer of a β-1,4-GlcNAc residue in the α-1,3-branch of the glycan	MGAT4A, MGAT4B	UDPGlcNAc
Alpha-1,6-mannosyl-glycoprotein 6-beta-N-acetylglucosaminyltransferase	GnTV	2.4.1.155	Transfer of a β-1,6-GlcNAc residue in the α-1,6-branch of the glycan	MGAT5, MGAT5B	UDPGlcNAc
N-Acetyllactosaminide beta-1,3-N-acetylglucosaminyltransferase	iGnT	2.4.1.149	Transfer of a β-1,3-GlcNAc residue on a terminal β-1,4-Gal molecule of the glycan	B3GNT2	UDPGlcNAc
Beta-N-Acetylglucosaminylglycopeptide beta-1,4-galactosyltransferase	b4GalT	2.4.1.38	Transfer of a β-1,4-Gal residue on any terminal GlcNAc glycan molecule	B4GALT1-7	UDPGal
N-Acetyllactosaminide alpha-1,3-galactosyltransferase	a3GalT	2.4.1.87	Transfer of an α-1,3-Gal residue on a β-1,4-Gal glycan molecule	GGTA1*	UDPGal
Alpha-1,6-fucosyltransferase	a6FucT	2.4.1.168	Transfer of an α-1,6-Fuc residue to the core GlcNAc glycan molecule	FUT8	GDPFuc
Alpha-1,3-fucosyltransferase	a3FucT	2.4.1.214	Transfer of an α-1,3-Fuc residue to a GlcNAc glycan molecule	FUT4, FUT7, FUT9	GDPFuc
Beta-galactoside alpha-2,6-sialyltransferase	a6SiaT	2.4.99.1	Transfer of an α-2,6-Neu5Ac residue to a terminal galactose molecule	ST6GAL1	CMPNeu5Ac
N-Acetyllactosaminide alpha-2,3-sialyltransferase	a3SiaT	2.4.99.6	Transfer of an α-2,3-Neu5Ac residue to a terminal galactose molecule	ST3GAL3, ST3GAL4, ST3GAL6	CMPNeu5Ac

*Inactive in humans but not all mammals.

Whilst mannose trimming mainly occurs in the *cis*-Golgi compartment (cisterna) and glycan elongation in the *median-*, *trans-* and *TGN*-Golgi compartments[108-110] mannosidases and GTases overlapping localization is evident and shows considerable diversity between different cell systems[111-116]. Recycling of the glycosylation enzymes between discrete Golgi compartments has also been reported[117-118]. Additionally, it is important to note that several complexes between different GTases and between GTases and nucleotide sugar transporters (NSTs) have been observed[119-120]. Conventionally, secreted $Man_{5-9}GlcNAc_2$ oligosaccharides are called high-mannose (HM) glycans, whilst sugars that carry five or four mannose molecules and are products of the GnTI activity and any downstream GTase, are called hybrid.

1.1.2 Nucleotide sugar donors

Nucleotide sugar donors are activated forms of monosaccharides, acting as co-substrates and monosaccharide-donors for GTase activity. NSDs consist of a nucleotide phosphate and a sugar molecule and are synthesised in the cytosolic environment, with the exception of CMP-sialic acid that is produced in the nucleus. The most common NSDs involved in *N*-linked glycosylation include: UDPGlc, UDPGal (UDP-galactose), UDPGlcNAc, CMPNeu5Ac (CMP-N-Acetylneuraminic or -sialic acid), GDPMan and GDPFuc (GDP-fucose). The NSD synthesis pathway is a complex metabolic network that includes several intermediate structures (Fig. 1.4) and is strongly dependent on the extracellular environment[121-122]. UDPGal and UDPGlc synthesis is also closely related to glycolysis. NSDs are translocated in the ER and Golgi lumens through the energy-independent activity of the nucleotide sugar transporters (NSTs). The NSTs act as antiporters for the exchange of nucleotide phosphates and NSDs between the cytosolic and lumen environments. NSTs can demonstrate a diverse specificity on the NSDs transported between the cytosolic and lumen environment[123-125]. In addition, NSD transport can be inhibited by the competitive transport of nucleotide-phosphates. NSD availability in the Golgi environment depends on the cytosolic NSD pools and the activity of the respective NSTs. Low NSD levels in the Golgi environment can directly affect the final structure of the *N*-linked oligosaccharide by becoming a limiting step towards GTase activity.

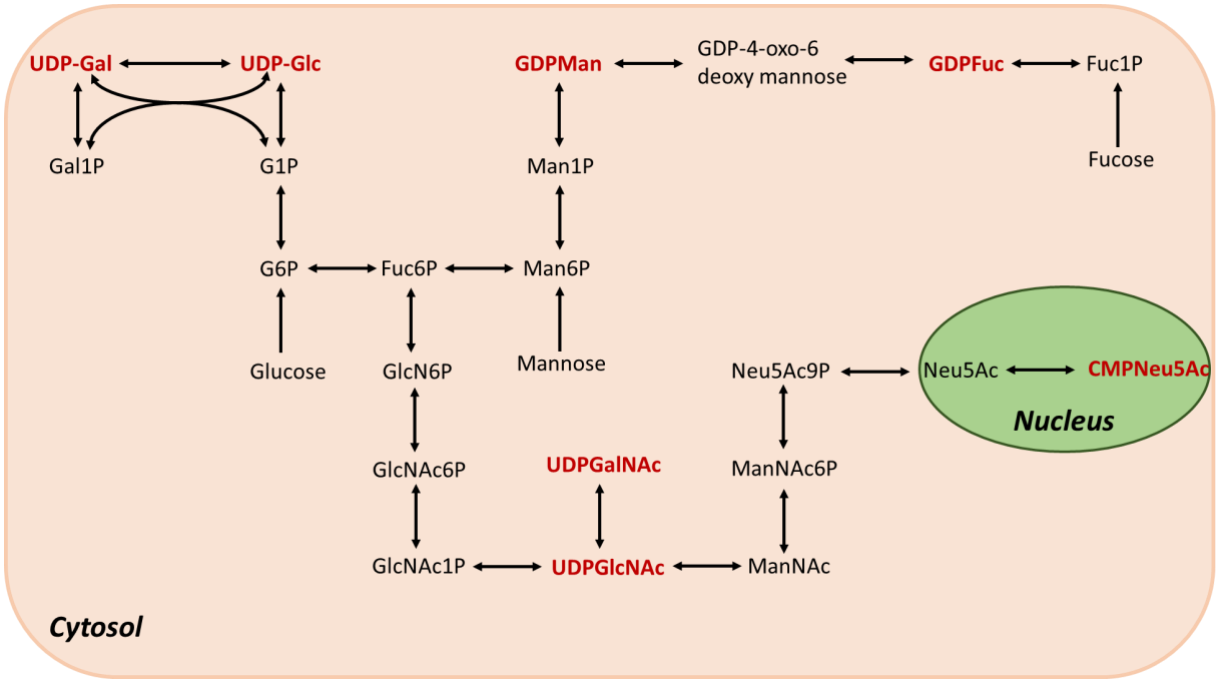

Figure 1.4: The pathway of NSD synthesis. Adapted from Jedrzejewski et al.[121]. G1P: glucose-1-phosphate, G6P: glucose-6-phosphate, Gal1P: galactose-1-phosphate, Fuc6P: fucose-6-phosphate, Fuc1P: fucose-6-phosphate, Man6P: mannose-6-phosphate, Man1P: mannose-1-phosphate, GlcN6P: glucosamine-6-phosphate, GlcNAc6P: GlcNAc-6-phosphate, GlcNAc1P: GlcNAc-1-phosphate, ManNAc6P: N-Acetylmannosamine-6-phosphate, Neu5Ac9P: Neu5Ac-9-phosphate.

1.1.3 Models for protein traffic in the Golgi

Whilst several aspects of protein glycosylation have been exhaustively studied, the mechanisms of glycan-protein complex transport within the Golgi apparatus, and its subsequent secretion, remain under intense debate[126-129]. Several theories around Golgi traffic have been developed over the years, with two prevailing models defining the fundamental principles for the rest: 1) the *cisternal maturation* (CM) and 2) the *vesicular transport* (VT) models[130]. In the CM model, Golgi morphology and structure change dynamically during protein trafficking. For that reason, Golgi compartmentalization in the CM model is blurry; the *cis*-cisterna gradually transforms into the *median*-cisterna that follows the same fate until the formation of the *TGN* compartment. The maturing cisterna carries all the necessary enzymes for protein PTM modifications, such as glycosylation. Finally, the *TGN*-cisterna scatters to smaller secretory vesicles to assist in the localization of the secreted proteins[128, 130-131]. COPII vesicles ensure the re-distribution of the Golgi enzymes to the newly formed cisternae. On the contrary, the VT model suggests that Golgi compartments are distinct and remain static throughout the secretion process. Protein trafficking is carried out by COPI and COPII vesicles that carry the modified proteins from one compartment to the next. Enzymes localization, i.e. glycosidases and GTases, are assumed to remain fixed to specific compartments[132]. Whilst both models present significant evidence that would accommodate their approval, there are several experimental observations that each of the two cannot sufficiently explain[112, 128, 133].

1.2 Importance of glycosylation in biopharmaceuticals production

Glycosylation has been established as one of the most crucial cellular processes, playing a convoluted role in several human disorders[56]. However, another scientific and research sector, within which glycosylation holds a pivotal role, is the synthesis of biopharmaceuticals. A previously estimated 40% of licensed biopharmaceuticals are reportedly recombinant glycoproteins[134], with the aforementioned number only expected to have risen during the last few years with the approval of several monoclonal antibodies as biotherapeutics[3-4]. As previously mentioned, the covalently attached glycan can significantly affect the conformation[135-136], stability[137], function and immunogenicity[138] of the synthesized therapeutic glycoprotein.

1.2.1 The role of monoclonal antibodies in the biopharmaceutical sector

It is important to note that there is a significant diversity on the number of glycosites and the structure of the attached glycans in each site, between the different therapeutic glycoproteins. For example, erythropoietin (EPO), a small in size but heavily glycosylated hormone that is regularly used for the treatment of anemia, contains three *N*-linked glycosylation sites on the Asn24, Asn38 and Asn83 residues of the backbone, and a single *O*-linked site in Ser126 residue[139-140]. Impressively, the occupied glycosylation sites contribute up to 40% of the total molecular weight of EPO (~34-38 kDa), with each site affecting *in vivo* protein synthesis and secretion in a unique way[139]. On the other hand, immunoglobulin G (IgG) antibodies, the most harnessed and common mAbs, are large molecules (~150 kDa) that carry two *N*-linked glycosylation sites in the crystallizable fragment (Fc) region (Fc-glycosylation) and potentially (~15-25% for IgG molecules) additional *N*-linked sites in the variable Fab (Fab-glycosylation) region[141]. The glycans occupying the Fc sites are normally less processed than the structures of the Fab region, and the glycans in both sites are considerably less complex when compared to the EPO glycans. The differences on the glycosylation profiles between the two glycoproteins and also between the sites of the same protein, showcase the strong influence that protein structure and amino acid sequence has on glycosylation. The following paragraphs present an overview of the importance of *N*-linked glycosylation on mAbs' properties and activity, in order to further demonstrate the significance of glycosylation in the biopharmaceutical sector.

Monoclonal antibody structure and function

Monoclonal antibodies are large protein molecules produced by B cells and that specifically target a single antigen. MAbs where first introduced by Paul Ehrlich early in the 20th century, as ''magic bullets'' that could selectively target an antigen while at the same time cause no harm to the exposed organism[142]. One of the first successful attempts to efficiently produce mAbs was presented in 1975, through the fusion of mouse spleen and myeloma cells[143]. Since the first approval of a mAb

biotherapeutic in 1986 from US-FDA, mAbs have been the leading treatment for severe human diseases[144-145]. Mammalian cells produce five different subclasses of antibodies: IgA, IgD, IgE, IgG and IgM. The structural differences between the five subclasses of the monoclonal antibodies occur due to the variation of the amino acid sequence in the heavy chains. Among the subclasses, IgGs are the most common mAbs, accounting for approximately 75% of the total human antibody secretion and 10-20% of the total plasma protein concentration[146]. In addition, fully human or humanized IgGs hold the lion's share in the licensed[147] and developing mAbs that are currently in late-stage clinical trials[4].

IgGs are further classified to four categories which show 95% homology: IgG1, IgG2, IgG3 and IgG4. IgGs consist of two identical light (L) and two identical heavy (H) polypeptide chains, forming a Y-shaped polypeptide backbone. Each IgG molecule includes one variable domain (V_H and V_L, respectively for the heavy and light chains) and three constant domains (C_H and C_L, respectively). The polypeptide chains are linked through disulphide bonds in the hinge region of the IgG. In addition, the differences between the IgG subclasses are mainly identified in the conformation of the hinge region and more specifically in the disulphide bonds structures[148]. The latter exhibit significant influence in the stability of the IgG molecules[149]. Moreover, the molecular weights of the chains usually range between 50-77 kDa for the heavy and approximately 25 kDa for the light chains[150]. The binding region of the molecules is all or part of the antigen-binding fragment (Fab) region. The antigen-specificity of each antibody is identified by the complementarity determining regions (CDR), parts of the Fab fragment that are hypervariable segments of 110 amino acids[151]. The rest of the IgG molecule, namely the Fc-region, is responsible for the function, safety, efficacy and the pharmacokinetic properties of the monoclonal antibody. Each light chain has a variable and a constant domain, while the heavy chains include three constant domains. All aforementioned structural conformations are shown in Fig 1.5.

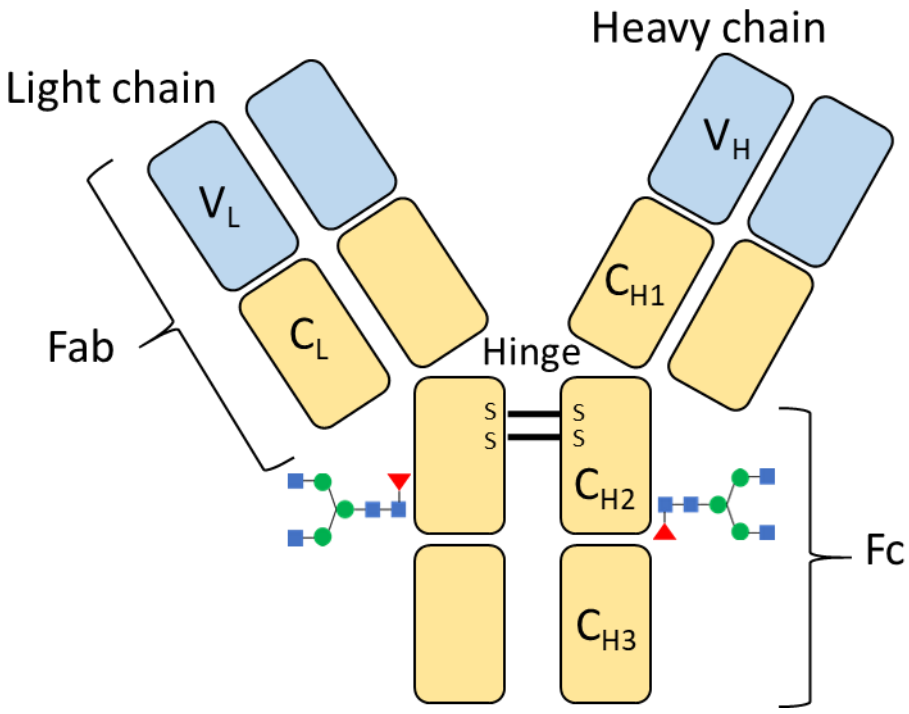

Figure 1.5: Schematic representation of the structure of an IgG1 molecule, including the *N*-linked glycans structures that are commonly found in the C_{H2} domain. Note that the *N*-linked glycans are found in the inner part of the Fc fragment but are shown on the external for illustrative reasons. Also, the glycan structures are not necessarily identical between the two chains as illustrated.

The IgG mechanism of action is a direct function of the molecule's structure. As previously described, the Fab-region is responsible for the antigen-binding, while the Fc-region initiates the cell response for the neutralization of the intruding cell. The antibody, through the Fc fragment, can either neutralize the target or provoke an immune response in order to protect the host against infection. Two important mechanisms of action (effector functions) for IgG antibodies have been described: Complement-Dependent Cellular Cytotoxicity (CDC) and Antibody-Dependent Cellular-Mediated Cytotoxicity (ADCC). In CDC, the antibody firstly binds to the antigen through the CDR fragments. Then, C1q proteins bind to the glycosylated Fc-region of the antibody and provoke a complementary action that leads to the disruption of the cell membrane and consequently to its death. On the other hand, in ADCC response, the antibody binds through the Fab-region to the antigen while also binding through the Fc-region with a Fcγ receptor (FcγR) of a natural-killer (NK) cell of the immune system that, in its turn, releases cytotoxic granules and consequently kills the intruding cell[152]. FcγRs are glycoproteins that belong to the immunoglobulin superfamily and are a vital part of the host immune system, expressed in the membrane of macrophages and NK cells, among others[153]. Major receptors in humans include FcγRI, FcγRIIa, FcγRIIc and FcγRIIIa[154].

Moreover, a highly conserved *N*-linked glycosylation site is normally identified in the Asn297 residue of each of the C_{H2} domains of the IgG molecules[155]. Due to steric hinderance that IgG structure imposes on the GTases activity, the glycosylation profile of the *N*-linked sites usually ranges over around ten possible structures, with the vast majority being core-fucosylated biantennary, non sialylated glycans. Interestingly, the glycans attached in the two distinct C_{H2} domains (Fc glycans) of the same molecule are not necessarily identical. Fig. 1.6 demonstrates the most commonly encountered glycan structures in IgG molecules. It is important to highlight that the nature of the glycan structures attached in the Fc-region significantly influences the folding and pharmacokinetic (PK) properties of the molecule[156-157]. To that end, the complete *in vitro* removal of Fc-glycosylation from IgG1 molecules has been found to result in reduced thermal stability[158] and lack of CDC and ADCC activity[159]. However, the extent to which deglycosylation affects antibody half-life and PK properties varies and might be dependent on the IgG subclass[43]. The effect of each designated glycosylation part is further discussed in the next few paragraphs, while Table 1.2 presents the outcome of several important studies in the field.

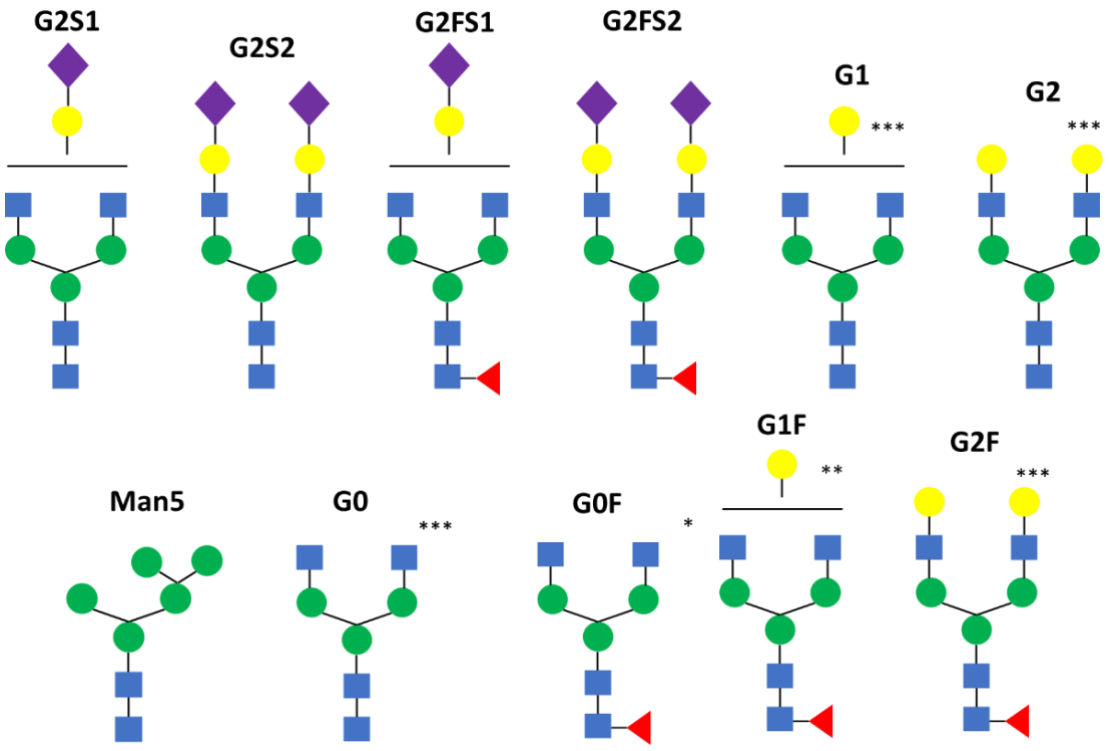

Figure 1.6: Schematic representation of the most common glycan structures encountered in therapeutic IgGs. (*) indicates the most abundant glycan (G0F), (**) the second most abundant glycan (G1F) and (***) the third most abundant glycans. The proposed abundancy does not apply to all IgGs. The shown abbreviation for glycans is used throughout the text, unless otherwise stated. G: stands for galactose, F: for fucose, S: for sialic acid and Man: for mannose. Bisecting glycosylation can also be observed in human serum IgG, but was not included in the current schematic representation, as a focus on the biopharmaceutical context of IgG glycosylation (where bisecting glycosylation is rare due to MGAT3 low expression levels in CHO cells) was placed.

Core-fucosylation

Moreover, specific parts of IgG Fc-glycosylation have been found to significantly influence the ADCC activity of the IgG molecules, through the regulation of the affinity between the Fc-region and the FcγRs[160]. Most importantly, core-fucosylation has been found to inhibit IgG binding with the FcγRIIIa receptor, as the removal of the core-fucose molecule leads to a 50-fold increase in the IgG-FcγRIIIa binding[161] and therefore to elevated ADCC activity[162]. However, lack of core-fucose has been found to result in no improvements in IgG binding to the neonatal Fc receptors, FcγRI or the C1q protein, while showing moderate improvements in the FcγRII binding, and therefore showcasing the specificity of glycosylation on designated parts of IgG function[161-162]. The effect of core-fucose on ADCC has been determined as independent of the IgG subclass, while core-fucosylation has been found to show no influence on the CDC activity of the IgG[163]. Interestingly, core-fucosylation has also been found to regulate the effect of sialylation on antibody ADCC efficiency, as terminal sialylation can decrease ADCC mediated IgG function in the presence of core-fucose, but without exhibiting any effects in non-fucosylated IgGs[164]. Finally, Kanda et al.[165] reported no impact of core-fucose on serum half-life and PK properties of IgG1.

Bisecting branching and terminal sialylation

Contrary to the adverse effect of core-fucose on ADCC activity that has been well characterized, the role of bisecting GlcNAcs and sialic acids on the aforementioned IgG function remains unclear. Early studies have identified that bisecting GlcNAc, the product of GnTIII activity, results in higher ADCC activity[166-168]. However, when the effect of bisecting branching was simultaneously studied with core-fucosylation, the latter was found to exhibit a significantly stronger effect on IgG activity, overshadowing the impact of bisected glycoforms[169]. In addition, the activity of the GnTIII enzyme has been found to inhibit core-fucosylation, indicating that possible improvements on ADCC mediated response could indirectly originate from reduced content in core-fucose caused by the increased levels of bisecting branching[170]. Elaborating on the effect of bisecting GlcNAc, Hodoniczky et al.[171] identified a ~10-fold increase in ADCC activity in the presence of bisected and core-fucosylated IgG, eventually demonstrating a direct effect of branching GlcNAc on ADCC.

Whilst the Fc sialylation levels of IgG are usually below 10%, terminal sialic acid has been found to increase serum half-life of several proteins[172-173] and, more specifically, of therapeutic antibodies for up to 9 times[174]. On the other hand, the decoration of the Fc glycans with sialic acid, has been strongly correlated to a decrease in the ADCC activity of the antibody molecule due to low binding affinity with the FcγRIIIa receptors[175-177]. In an interesting study, Zhang et al.[178] identified that the alpha-2,3-sialic acid in the 6-arm of IgG1, destabilizes the C_{H2} domain of the protein, while the alpha-2,3- and the

alpha-2,6-sialic acid in the 6-arm of the glycan cause no destabilization, demonstrating that way the linkage specific effects of sialylation. Moreover, Kaneko et al.[179] have identified that the presence of sialic acid is critical for the anti-inflammatory properties of tintravenous immunoglobulin G (IVIG). IVIG is administered in several autoimmune diseases and antibody-deficiency disorders. To that end, Anthony et al.[180] utilized a recombinant IgG to demonstrate that the anti-inflammatory properties of the protein are entirely dependent on alpha-2,6-sialylation. Contrary to the aforementioned reports, Boyd et al.[159] have reported no effect of terminal sialylation on CDC, ADCC or antigen binding activity.

Terminal galactosylation

Unlike most glycan attributes that exhibit a multi-level influence on ADCC activity through the regulation of FcγRIIIa receptor binding, the majority of studies around IgG beta-1,4-galactosylation of the Fc fragment, have identified a neutral or sometimes positive role in ADCC and antigen binding[159, 169, 171, 177, 181]. Thomann et al.[182] have argued that Fc glycan decoration with terminal galactose residues consistently improves ADCC activity of the IgG1 molecule but is unable to provide further improvements in the afucosylated variation of the antibody. However, antibody galactosylation has a more direct and explicit effect on CDC activity, due to an increase in IgG-C1q binding[159, 171, 183]. Whilst galactose content has been of great interest among the available therapeutic IgGs[184], a recent study has indicated that protein serum half-life can be negatively affected by the presence of galactose, which can only be inverted by the further addition of sialic acid[185]. Contrary to the aforementioned observations, Wright & Morrison[156] identified no changes in galactose deficient IgG1 regarding *in vivo* protein half-life or binding affinity towards the FcγRI receptor.

Cetuximab, a commercial therapeutic epidermal growth factor receptor (EGFR), exhibits high ADCC activity against several cancer types. However, the agalactosylated variation of the antibody showed dramatically reduced levels of ADCC activation due to weaker binding to both FcγRI and FcγRIIIa receptors[186]. Finally, the effect of terminal galactosylation on antibody clearance is probably dependent on the nature of the examined mAb[187-189]. Unlike beta-1,4-galactosylation, the extension of the glycan through the addition of an alpha-1,3-galactose has been characterized as highly immunogenic, probably due to the lack of GGTA1 (gene expressing the a3GalT enzyme) expression in humans[190].

Table 1.2: Effect of each glycosylation attribute on several functions of the IgG. Note that the effect of exposed terminal GlcNAcs is incorporated in the "Terminal galactose" section.

Glycosylation attribute	Study	Half-life	CDC	ADCC	C1q binding	FcγRI binding	FcγRII binding	FcγRIII binding	Other/Comments
Core-Fucose	Shields et al.[161]			↑	(-)	(-)		↑ (50x)	Neonatal FcR binding unaffected
	Shinkawa et al.[169]			↑(~100x)					Examined human IgG1 & chimeric anti-CD20 IgG1
	Niwa et al.[163]		(-)	↑					Examined all IgG subclasses
	Chung et al.[162]		(-)	↑ (8x)	(-)	(-)	↑	↑	Linear correlation between ADCC and FcγRIII binding
	Kanda et al.[165]	(-)	(-)	↑	(-)			↑	No effect on PK properties
Terminal sialic acid	Boyd et al.[159]		(-)	(-)					Antigen binding unaffected
	Scallon et al.[175]			↓			↓		Antigen binding reduced
	Naso et al.[176]			↓					Unaffected antigen-binding and PK properties
	Higel et al.[185]	↑							Examined receptor glycosylation of Fc-fusion drugs
	Millward et al.[191]	(-)							Examined Fab-sialylation
	Bas et al.[174]	↑							Sialylation increases half-life between 6-9x in IgG
Terminal galactose	Boyd et al.[159]		↑	(-)					Antigen binding unaffected
	Hodoniczky et al.[171]		↑(2x)	(-)	↑				Examined commeriacl Rituxan® and Herceptin®
	Thomann et al.[182]			↑					ADCC ↑only in fucosylated substrate
	Shinkawa et al.[169]			↑					The effect in ADCC is overshadowed by core-fucose
	Kumpel et al.[181]			↑		↑		↑	Examined an IgG3 molecule
	Patel et al.[186]			↑		↑		↑	Examined the IgG1 cetuximab
	Bheemareddy et al.[183]		↑						Examined anti CD-20 IgG1
	Newkirk et al.[188]	↓							Examined anti CD-20 IgG2

	Malhotra et al.[192]		↓						CDC ↓ due to reduced MRs binding
	Jones et al.[193]	↑							Result of reduced MRs binding
	Huang et al.[194]	↓							Examined Fab-glycosylation
Bisecting branching	Lifely et al.[166]			↑					First study to relate bisecting glycans with ADCC
	Umana et al.[167]			↑					Examined a chimeric IgG1
	Shinkawa et al.[169]			↑					ADCC effect observed on highly-fucosylated glycans
	Davies et al.[168]			↑				↑	Estimated 10-20x increase of both
	Hodoniczky et al.[171]		(-)	↑(10x)	(-)				No influence on antigen-binding
High-mannose structures	Wright & Morrison[156]	↓	↓			↓			Review reporting results on IgG1
	Goetze et al.[195]	↓							Examined Man5-decorated IgG1 & IgG2
	Alessandri et al.[196]	↓							40% increased clearance in Man5-7 glycans
	Zhou et al.[197]		↓	↑	↓			↑	Examined Man9- & Man8-decorated IgG1
	Kanda et al.[165]	↓	↓	↑	↓			↑	HM glycans are not core-fucosylated
	Millward et al.[191]	(-)							PK properties of IgG1 not affected
	Liu et al.[198]	↓							2-3x faster clearance
	Yu et al.[199]	↓(3x)	↓	↑(5-7x)		↓	↓	↑(4-8x)	Compared to normally glycosylated mAb

High-mannose structures

Whilst human serum IgG contains negligible traces of HM glycans, recombinant antibodies produced in CHO cells normally carry 1-5% of Man5 structures attached in the Fc-region. While IgG molecules carrying HM glycans are not native to humans, no cases of immunogenicity have been reported[200] despite earlier contradictive indications[201]. While Chen et al.[202] have reported minor effects of glycosylation on antibody clearance, HM structures can result in elevated serum clearance of IgG[195-196] and shorter half-life[198], due to the excessive specificity of mannose receptors (MRs) on glycans with exposed mannose residues[203]. The study performed by Wright & Morrison[156] was one of the first to identify the adverse effect of HM glycans on IgG1 CDC activity. Further investigations verified the negatively monotonous relationship between HM glycans and CDC efficiency, due to a reduction in the binding affinity between the IgG molecule and the C1q protein[165, 197]. Moreover, IgGs decorated with HM glycans exhibit enhanced FcγRIIIa binding and, consequently, elevated ADCC activity[165, 197, 199]. However, HM glycans, such as the Man5, Man8 and Man9 structures, are not core-fucosylated and therefore, any improvement in the ADCC activity could be an indirect effect of the lack of core-fucosylation.

Terminal N-Acetylglucosamine

In contrast to human serum IgG that presents considerable amounts of galactosylation and sialylation[204], recombinantly produced antibody normally exhibits high distributions of glycan structures decorated with terminal GlcNAc molecules, as shown in Fig. 1.7, mainly in the form of the bi-antennary G0, G0F and G1F (the agalactosylated branch of the glycan) structures. In general, terminal GlcNAc can be recognized by mannose receptors of the immune cells, indicating that reduced serum half-life of glycans exposed with terminal GlcNAc molecules is possible[205]. However, mannose receptors have been reported to resemble the structure and function the C1q proteins and initiate the CDC activity of IgGs[193, 206]. To that end, elevated levels of agalactosylated bi-antennary glycans in IgG molecules have been corelated with increased CDC activity in rheumatoid arthritis[192]. However, the evidence indicating that galactosylated mAb results in stronger C1q binding and therefore, in higher CDC activity, is significantly more abundant, and therefore, terminal GlcNAc molecules are not generally considered a desired attribute for recombinantly produced IgG.

N-linked glycosylation of the Fab fragment

To date, the impact of Fab-glycosylation on IgG activity has not been studied as comprehensively as the respective effect of Fc-glycosylation. However, Fab-glycosylation is currently of emerging interest due to its plausible impact on IgG conformation and Fab-antigen binding[207-208]. The glycans attached

to the Fab region (Fab glycans) of the antibody are normally less core-fucosylated, present higher heterogeneity and are normally more processed than the Fc glycans, presenting higher levels of bisecting branching and elevated terminal sialylation or galactosylation[209-210]. However, they can also present high-mannose structures and terminally fucosylated (a3FucT) or alpha-1,3-galactose (a3GalT) elongated glycans[211]. The observed discrepancies between Fc and Fab glycans are mostly attributed to the IgG conformation that exhibits a more exposed Fab region[208].

To that end, Chung et al.[212] reported that the presence of alpha-1,3-galactose in the Fab glycans of cetuximab resulted in severe immunogenic responses such as anaphylactic reactions. Fab glycans decorated with terminal beta-1,4-galactose molecules have been associated with marginally increased IgG clearance rates[194]. However, further processing with the addition of sialic acid has not been found to influence serum half-life[191]. Despite the fact that the binding between an Fc-receptor and the IgG molecule is strongly dependent on Fc conformation, Fab glycosylation has been found to influence the stability of the binding between the IgG and neonatal FcRs[213] and the PK properties of the antibody[214]. As elaborating on the importance of Fab glycosylation would fall out of the context of this thesis, an interested reader is encouraged to read the excellent reviews of Zhang et al.[138] and van de Bovenkamp[141] for more relevant information.

1.2.2 Glycosylation role within Quality by Design

The pharmaceutical industry is currently reconsidering the methods employed for determining product quality and controlling the manufacturing process. Quality by Testing (QbT) and Quality by Design are amongst the most popular quality-control and evaluation methods applied in the biopharmaceutical industry. QbT involves comprehensive testing of the effect that different configurations of raw materials and process conditions have on the end-product quality. Therefore, QbT usually results in costly and inefficient testing procedures[215] without providing any mechanistic insights between process parameters and the end-product. In order to compensate for the associated costs, the examined process conditions and input specifications are usually constrained within a narrow range, potentially limiting drug safety and in-process adjustability.

According to the International Conference on Harmonization (ICH) Q8(R2) document[216], QbD is a "systematic, risk-based and holistic approach to development that begins with predefined objectives and emphasizes on the product and process understanding with the knowledge that is gained over the product lifecycle". QbD is heavily based on several in-process quality measurements that, combined with the manufacturing process itself, are designed to ensure the quality target product profile (QTPP)[217-218]. A crucial attribute of QbD is the elucidation of the qualitative relationship between critical process parameters (CPPs) and CQAs. CQAs of the end-product are identified during

the early steps of product and process development, therefore directing the final steps of the development process[47].

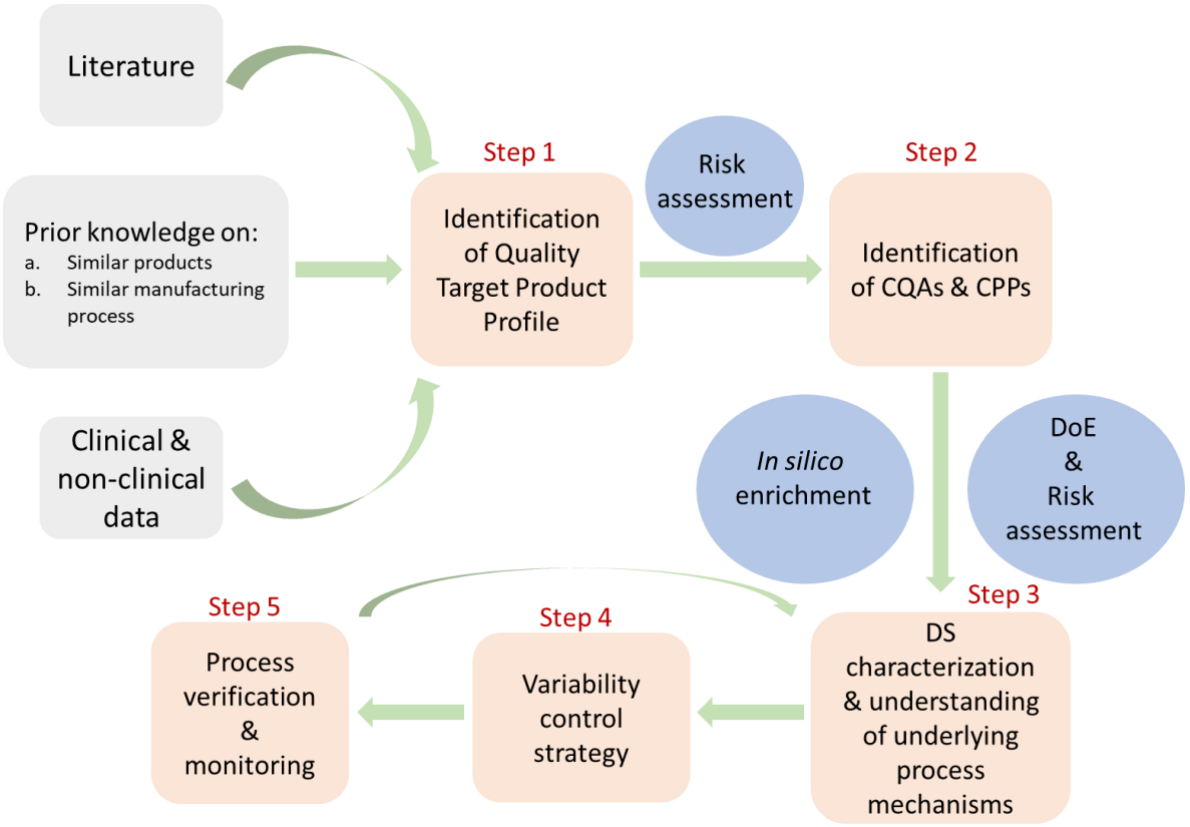

Figure 1.7: Fundamental steps of the Quality by Design paradigm as applied in the biopharmaceutical industry. Abbreviations: CQA for critical quality attribute, CPP for critical process parameter, DoE for design of experiments and DS for design space.

CQA is a property or a characteristic of the drug that is constrained within specific boundaries in order to ensure the desired product quality and thus, its safety and efficacy. Therefore, defining CQA boundaries demands extensive experimental characterizations of the drug. The identification of CQAs is performed through a risk assessment that incorporates data from similar molecules and identifies parameters that affect the end-product quality[216]. Following CQA identification, the investigation of the mechanistic relationships between process parameters and the designated CQAs is of vital importance within the QbD paradigm. CQAs that affect safety and immunogenicity usually lead to narrow boundaries that are strongly based on relevant clinical data[219]. On the other hand, the CQAs that influence product efficacy are acceptable within wider ranges, enabling a more flexible manufacturing process[220].

However, it is not only the CQAs that require acceptable ranges; the same holds true for the Design Space. DS is the multidimensional range within which the process variables and CPPs ensure that the CQA acceptance criteria are met[216, 221]. Interestingly, the DS can be based on mechanistic, empirical or

hybrid models[222-223] and can be distinguished into product- and process-related DS[47]. DS identification is considered to be the final stage of the QbD method for process understanding but can be re-informed and restructured if additional knowledge for the product is obtained through process validation[224]. The product-DS incorporates the acceptable ranges of the therapeutic physicochemical CQAs that lead to the desired product quality. On the other hand, the identification of the process-DS is achieved through risk analysis and design of experiments (DoE) for investigating the effect of each CPP and evaluating the importance of the chosen parameters[47]. An overview of the main steps followed in the QbD paradigm can be found in Fig. 1.7.

Within the biopharmaceutical context and the manufacturing of therapeutic mAbs, *N*-linked glycosylation is considered one of the most crucial CQAs for evaluating both the potency and the immunogenicity of the product due to its apparent heterogeneity and multi-dependencies[225-227]. According to the mAb-product effector function, the extent of Fc galactosylation or Fc core-fucosylation is commonly considered a glycosylation-related CQA (gCQA) for the CDC- and ADCC-mediated molecules, respectively[138]. With regards to immunogenicity, as mentioned in section 1.2.1, the attachment of alpha-1,3-galactose molecules on Fab glycans can result in severe anaphylactic shocks[212]. Therefore, the aforementioned glycan conformation should be considered a relevant gCQA.

In a seminal study, Schiestl et al.[228] examined the variations in the *N*-linked glycosylation profile of three licensed biopharmaceutical products, namely MabThera® (rituximab), Aranesp® (darbepoetin alpha) and Enbrel® (etanercept), over several batches manufactured between 2007 and 2011. All three biotherapeutics are glycosylated proteins, with rituximab being a chimeric IgG1, darbepoetin alpha being an engineered EPO to include two additional *N*-linked glycosylation sites that improve its *in vivo* half-life, and etanercept being a dimeric fusion protein, combining the receptor of the tumour necrosis factor (TNF) and the Fc fragment of an IgG1 antibody. In all three products, significant changes were observed in the proteins' glycoprofile at some point within the examined period, probably caused by alterations in the manufacturing process. The most notable changes included a reduction of the G2F distribution in the Enbrel® product from ~50% to ~30%, a 3-fold increase (albeit still remaining in low levels) of G0 distribution in MabThera® and a 10% reduction in the distribution of highly sialylated structures in Aranesp®. Similarly, monitoring of the Herceptin® (trastuzumab) over a course of 4.5 years and several lots in both the EU and the US, indicated that a drift towards increasing antibody fucosylation resulted in supressed ADCC activity of the drug[229]. The results of the studies showcase the importance of closely monitoring *N*-linked glycosylation during manufacturing, far beyond the product development process. Additionally, the continuous monitoring of the glycoprofile of biotherapeutics potentially enables the refinement of acceptable glycoprofile ranges for each

protein, when the altered glycosylation does not trigger immunogenic responses or results in reduced potency of the therapeutic.

N-linked glycosylation is also systematically considered during biosimilars development[230-231]. Glycosimilarity, meaning the chemically assessed comparability of the glycoprofile between the biosimilar and the reference therapeutic, is one of the most important biosimilarity characteristics for biosimilars development and for demonstrating adequate resemblance to the original drug[232]. Glycosimilarity is considered a Tier 1 biosimilarity CQA, meaning that a narrow acceptable range of ±1.5% with 95% confidence is required for discrepancies between the reference molecule and the similar[233]. The glycosimilarity index has therefore been introduced in order to evaluate both the disagreement between individual glycans of the glycoprofile and between several crucial gCQA groups[233].

Finally, it is important to note that the identification of a suitable DS for a bioprocess can be considerably time-consuming, even with the utilization of experimental scaled-down models[226]. Wet-lab approaches on configuring an appropriate DS of the culturing media composition for manipulating IgG glycosylation, have demonstrated the importance, but also the demanding nature, of experimental DS identification[234-237]. For that reason, the use of computational tools and models in order to assist and expedite DS characterization, especially within the biopharmaceutical sector, is highly encouraged[238]. Model-assisted DoE has been proposed for narrowing the boundaries of experimentation required for understanding CHO cell growth and IgG1 productivity. More specifically, Möller et al.[239] utilized an unstructured Monod-type mechanistic model within an iterative exercise, in order to describe the culturing process and, ultimately, simulate and constrain the suggested DoE experiments. In a similar manner, model-based optimizations of the DS boundaries have recently been proposed for additionally controlling the *N*-linked glycoprofile of recombinant proteins, by employing either fully mechanistic[240] or statistical[241] models.

1.3 Investigation of mammalian cell growth and recombinant protein glycosylation

The importance of protein *N*-linked glycosylation was demonstrated in sections 1.1 and 1.2. In the biopharmaceutical sector, glycoproteins are produced in mammalian hosts, with CHO cells being the longstanding workhorse of recombinant glycoprotein production in both academia and industry[4]. The main objective of recombinant protein production is the maximization of titre and the optimization of product quality attributes for achieving the desired efficacy. For that reason, one could identify three degrees of freedom for meeting both requirements: 1) the composition of the seeding media and feed, 2) the manipulation of the bioprocess operational conditions and 3) the employment of genetic

modifications. Cell metabolism has been found to directly affect the glycoprofile of the product[242] through the interlinked metabolic pathways of central carbon metabolism and NSD synthesis (Fig. 1.8). For that reason, the following paragraphs describe the main factors that affect mammalian cell growth, central carbon metabolism and, ultimately, recombinant protein *N*-linked glycosylation. Methods manipulating the aforementioned factors towards satisfying the desired objectives are also discussed. To that end, the studies that specifically target protein glycosylation through genetic modifications of the host cells are also presented, regardless of whether the effect of the genetic modification on cell growth was reported in the original studies. The main focus of the current section is placed on two out of the three approaches described earlier, namely media composition and genetic engineering techniques, as they are more relevant to this thesis.

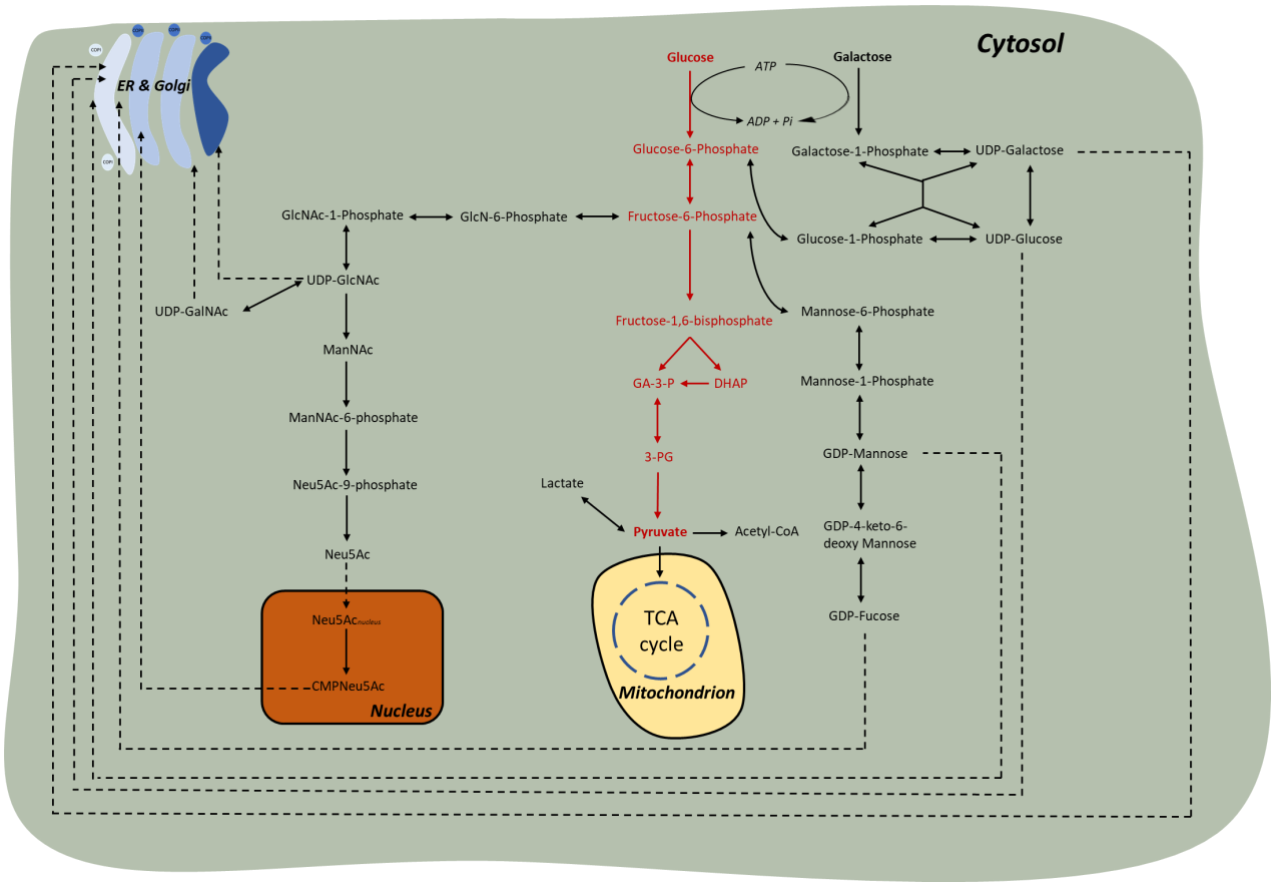

Figure 1.8: Interactions between the central carbon metabolism and the NSD synthesis pathway. The dashed lines are used to indicate transport from one cellular compartment to another. Abbreviations: GlcN for glucosamine, GalNAc: N-Acetylgalactosamine, ManNAc for N-Acetylmannosamine, GA-3-P for glyceraldehyde 3-phosphate, DHAP for dihydroxyacetone phosphate and 3-PG for 3-phosphoglycerate. Red colour is used to indicate the main reactions and metabolites involved in the glycolysis pathway. The transformation of galactose to galactose-1-phosphate can be ignored if galactose is not provided in the culturing media.

1.3.1 Metabolic perturbations

Metabolic by-products, such as ammonia and lactate, have been well-known inhibitors of mammalian cell growth and recombinant protein synthesis for several decades[243]. More specifically, ammonia, formed mainly by the glutamine/glutamate interconversion and the chemical decomposition of glutamine[244], is considered a toxic by-product of cellular metabolism as it can compromise cell growth through several mechanisms, including the perturbation of electrochemical gradients and initiation of pH shifts in several organelles[245]. Studies reporting either negligible[246] or significant[247-248] effects of elevated ammonia levels on CHO cell growth and specific productivity have been reported, indicating that the impact can be clone-dependent. On the contrary, hybridoma cells have been found to exhibit up to 50% reduced cell growth in ammonia concentrations higher than 4 mM due to pH drop[249]. However, ammonia can be utilized as an alternative nitrogen source when glutamine is depleted in a cell line-dependent manner[250].

Regarding glycosylation, early studies have indicated that increased ammonia concentrations can inhibit the activity of sialyltransferase enzymes (involved in both *N*-linked and *O*-linked glycosylation) by causing perturbations in the intracellular pH[251-253]. Additionally, ammonia has been associated with elevated pools of UDPGlcNAc that can potentially inhibit CMPNeu5Ac transport[253-255]. In a study of paramount importance, the use of ^{15}N-labelled ammonium chloride (NH_4Cl) indicated that ammonia is an important resource for UDPGlcNAc and CMPNeu5Ac synthesis. Ammonia transforms to fructose-6-phosphate through the activity of glucosamine-6-phosphate isomerase (GNPI), which in its turn is a substrate for UDPGlcNAc synthesis (Fig. 1.8)[256]. Moreover, increased ammonia concentration in CHO cell cultures can result in reduced expression of the GTases responsible for galactosylation/sialylation (B4GALT & ST3GAL genes) and the CMPNeu5Ac transporter (SLC35A1 gene)[257]. On the other hand, Brodsky et al.[258] reported no alterations in the expression levels of glycosylation-related genes for up to 5 mM of ammonia in NS0 cells, further supporting that the impact of ammonia is cell line-dependent. Gawlitzek et al.[259] have reported a negative correlation between ammonia concentration and TNF-IgG galactosylation and sialylation in CHO cells. While the mRNA levels of b4GalT and a3SiaT were not affected by ammonium addition, the activity of the enzymes was reduced by 50% in the presence of 10-15 mM ammonia, further supporting the theory of GTase inhibition through pH changes in the Golgi environment. Moreover, Yang & Butler[260] reported that 30 mM of NH_4Cl in CHO cell cultures could significantly alter secreted EPO glycosylation, by reducing both the degree of terminal sialylation and the distribution of tetra-antennary structures. In a similar manner, elevated levels of ammonia were found to substantially halt all galactosyltransferase activity in CHO cells[261].

Lactate, or its conjugate lactic acid, is considered an end-product of anaerobic glycolysis and is synthesized through the bidirectional conversion activity of the lactate dehydrogenase (LDH) enzyme

on cytosolic pyruvate, prior to pyruvate transport to mitochondria for the initiation of the tricarboxylic acid (TCA) cycle. In CHO cell cultures, lactate is not supplemented in the seeding media and is rapidly produced and secreted by the cells during the exponential phase. Interestingly, a metabolic shift usually occurs by the end of the exponential phase, where lactate consumption is initiated towards pyruvate synthesis. Lactate is considered an undesired metabolic by-product as it affects the pH and osmolality of the cell culture[262], consequently reducing the specific growth rate and recombinant protein synthesis in mammalian cell cultures[263].

While lactate production is considerably less energetically efficient, producing 2 ATP molecules instead of 36 that are synthesized through the complete oxidation of glucose in the TCA cycle[248], it has been correlated to cell proliferation. The production of lactate during the exponential phase strongly resembles the Warburg effect, in which cancer cells are found to over-produce lactate in order to proliferate. The mechanism behind the lactate metabolic shift is of great interest but still not fully understood[264-266]. However, the metabolic switch usually occurs under reduced glucose uptake and therefore, insufficient pyruvate concentration[266-267]. Whilst glucose depletion has been found to initiate lactate consumption[268], the lactate switch occurs even under glucose abundance in the culture[269]. Alternatively, the shift in lactate metabolism has been correlated with the levels of mitochondrial oxidative activity in CHO cells[270]. Interestingly, Templeton et al.[271] demonstrated that lactate consumption presented a positive correlation with increasing IgG specific production rate. The energy-expensive nature of recombinant protein synthesis and the inefficient energy generation during lactate synthesis were assumed as the underlying mechanisms that support the observation. Similar observations associating high IgG specific productivity with the initiation of the lactate consumption phase in CHO cell cultures have been proposed by a big-data study examining over 200 culture runs[272]. Martínez et al.[273] used stoichiometric modelling to identify that during lactate production phase in CHO cell cultures, a mere 3% of the acetyl-CoA formed by pyruvate was utilized in the TCA cycle. After the shift in lactate metabolism, the rate of pyruvate entering the TCA cycle through its transformation to acetyl-CoA increased to 68%.

As a result, the control of both ammonia and lactate is crucial for cell growth and protein synthesis. Early data suggested that the partial replacement of glutamine with glutamate in the media can reduce ammonia accumulation in mammalian cell cultures[274]. The generation of the GS-CHO cell line, overexpressing the glutamine synthetase enzyme, has established the conversion of glutamate to glutamine as the major source of the latter. Therefore, the genetically modified cells do not require glutamine in the seeding media reducing that way the production of ammonia from both the chemical decomposition of glutamine and the conversion of glutamine to glutamate[275]. In a similar manner, a common target for the reduction of lactate synthesis is the downregulation of the LDH-A gene and the

enhancement of recombinant protein production[276-277]. Introduction of the LDH-A downregulation in GS-CHO cells resulted in further reduction of ammonia accumulation and improved mAb galactosylation[278]. Overexpression of BCL2, an anti-apoptotic protein, in combination with LDH-A downregulation led to improved cell growth and longevity, in addition to lactate reduction[279]. The overexpression of anti-apoptotic genes has been previously shown to increase the demands on lactate consumption, and when combined with lactate feeding can result in enhanced cell growth and longevity[280]. Knocking down both pyruvate dehydrogenase (PDH), the enzyme responsible for acetyl-CoA synthesis from pyruvate, and LDH-A resulted in increased cell growth and antibody production in CHO cell cultures, by forcing the pyruvate to enter the TCA cycle[281]. Expression of the yeast pyruvate carboxylase (PYC2) enzyme in CHO cells led to enhanced cell growth and longevity, reduced lactate accumulation and improved IgG synthesis and galactosylation[282].

The use of alternative feeding strategies in order to regulate cell metabolism and more specifically lactate and ammonia concentration in mammalian cell cultures, has been one of the most popular methods for process control used for several decades now[263, 283]. To that end, the inclusion of lactate and pyruvate in the feed resulted in reduced ammonia accumulation but elevated lactate concentrations[284]. The replacement of glucose and glutamine with galactose and glutamate, respectively, resulted in significantly reduced levels of lactate and ammonia accumulation, but also halted growth and recombinant protein synthesis[285]. In a follow-up study, the partial replacement of glucose with galactose resulted in a moderately improved growth and final titre concentration in CHO cell cultures, while at the same time initiating an almost complete consumption of lactate produced in early-stage culture[268]. Building on previous studies that demonstrate the positive effect of lactate feeding on reducing lactate and ammonia accumulation[284, 286], Freund & Croughan[287] demonstrated that the supplementation of ~35 mM sodium lactate was adequate to eliminate lactate production in batch cultures. In addition, the authors demonstrated that the supplementation of lactate resulted in considerably reduced ammonia production, alongside with significantly increased cell growth.

Precursors of nucleotide sugar donors

The amino acid and metabolite composition of the seeding media is known to affect recombinant protein glycosylation by influencing cellular metabolism and, therefore, NSD synthesis[288-289]. One of the most widely applied practices in controlling *N*-linked glycosylation of recombinant glycoproteins is the manipulation of targeted intracellular NSD pools[290-291]. In order to achieve the desired regulation of NSD concentrations in the Golgi, one can supplement the metabolic precursors or inhibitors of the NSD synthesis reactions.

However, altering media composition for the optimisation of the glycosylation process could inhibit cell growth and recombinant protein synthesis[236]. Addition of galactose, uridine and manganese chloride is usually employed for improving and controlling the terminal mAb galactosylation[292-295]. The availability of uridine and galactose leads to enhanced UDPGal synthesis, the NSD responsible for the addition of galactose residues to the *N*-linked glycan. Manganese is a co-factor of several GTases and its supplementation has been found to enhance the activity of the b4GalT enzyme in mammalian cell cultures[296]. Wong et al.[297] conducted one of the earliest studies to examine the effect of multiple NSD precursors on human interferon-gamma (IFNγ) glycosylation produced in CHO cells. The authors elaborated on results from previous studies[259-260, 298-304] in order to design a series of experiments including the addition of galactose, uridine, N-Acetylmannosamine (ManNAc), cytidine and GlcN. ManNAc and cytidine feeding, targeting the increase of CMPNeu5Ac pools and, consequently, sialylation, did not inhibit growth and resulted in improved recombinant protein yields. The addition of ManNAc and cytidine resulted in up to a ~120-fold elevation of intracellular CMPNeu5Ac levels and a 37% increase in IFNγ sialylation. GlcN and uridine addition, targeting the UDPGlcNAc pools, resulted in reduced cell growth and productivity. The authors postulated that the halt of cell growth originated from reduced glucose uptake, as GlcN is a competitive substrate for glucose transport to the intracellular environment through the glucose transporters (GLUT1)[305]. As expected, the feed resulted in elevated UDP-HexNAc and UDP-Hex concentration and, ultimately, in a surprising 30% increase in sialylation. Galactose and uridine feeding showed negligible effects on growth and productivity, but resulted in a general reduction of IFNγ galactosylation, an observation that directly contradicts several similar studies[292-293, 306].

In a similar manner, Cha et al.[307] optimised the sialylation of albumin-EPO produced in CHO cells, by supplementing the nucleotide sugar precursors GlcN, galactose and ManNAc on day 0 and day 3 of the culture. Moreover, mannose is usually fed as an additional carbon source to increase the formation of HM glycans in mAbs[308-309]. However, replacement of glucose with mannose can additionally lead to a moderate increase in total sialylation, alongside with improved cell growth and titre concentration[310]. The increase in sialylation is attributed to the intervention of mannose to the UDPGlcNAc synthesis pathway, and therefore the pathway of CMPNeu5Ac, through its conversion to mannose-6-phosphate and subsequently to fructose-6-phosphate. On the basis of the synergistic effect of nucleotide and sugar precursor additions on NSD intracellular concentrations[297, 299], Jedrzejewski[311] studied the effects of combined guanosine and mannose addition on IgG production and glycosylation in GS-CHO cultures. The feeding strategy resulted in considerably reduced growth and viability, but significantly increased specific antibody productivity without presenting any major changes in the *N*-linked glycoprofile. In a comprehensive study investigating the effect of mannose,

galactose, fructose, uridine, GlcNAc, asparagine and manganese on IgG-producing CHO cells, Villiger et al.[261] reported the impact of several feeding strategies on cell growth, productivity, intracellular NSD levels and IgG glycosylation. Interestingly, addition of up to 8 mM GlcNAc did not affect the UDP-HexNAc and UDP-Hex concentration but resulted in increased G0F distribution and therefore reduced galactosylation and HM distribution. Similarly, in GlcNAc and uridine feeding experiments, GlcNAc addition exhibited a negative correlation with IgG galactosylation. Fucose and manganese addition resulted in enhanced GDPFuc levels, considerably elevated galactosylation, a moderate reduction in HM structures and, consequently, in a slight increase of total fucosylation. Galactose and manganese presented a synergistic effect, indicating that both b4GalT activity and intracellular UDPGal levels are limiting steps of IgG galactosylation for the utilized cell line. Asparagine addition exhibited no direct effect on NSD intracellular pools and IgG glycosylation. However, the conversion of asparagine to ammonia and the consequent pH increase can inhibit galactosylation[261, 295].

In another large-scale study, Kildegaard et al.[312] identified non-significant effects of GlcNAc, mannose, fucose, galactose, cytidine, uridine and Neu5Ac on cell growth and IgG productivity at the examined levels. As expected, galactose addition resulted in considerably elevated IgG galactosylation, while GlcNAc exhibited the opposite effect as previously reported. Mannose addition resulted in a slightly increased distribution HM glycans, while ManNAc supplementation was found to improve fucosylation. However, Neu5Ac feeding did not improve IgG sialylation. In an even more elaborate study including three CHO-K1-derived cell lines expressing different IgG1 mAbs and investigating feeding strategies of ten additives, Loebrich et al.[236] additionally screened the effect of 1 mM copper and 1% v/v glycerol on IgG glycosylation. Copper addition resulted in a decrease of the Man5 species and a respective increase in bi-antennary glycans, such as G0F. Whilst glycerol supplementation, which has been previously utilized to reduce IFNβ aggregation and induce sialylation[313], led to improved cell viability, it did not impact IgG glycosylation. Iron addition can result in improved OST activity and therefore increased glycosylation site occupancy[314]. Finally, the supplementation of 200-300 µM 1,3,4-O-Bu3ManNAc, a butyrate-derivatized analogue of ManNAc, was found to increase the sialic acid content of EPO by ~8 times in CHO cell cultures, with minor impact on cell growth[315-316].

1.3.2 Process parameters

Operating regime

Various cell culturing systems are used for the production of therapeutic proteins. Batch and fed-batch bioreactors are two of the most utilized operating regimes within academic and industrial research. Nutrient depletion or waste-product accumulation is strongly dependent on the chosen culturing

configuration. For example, glucose depletion is a common problem of batch cultures that leads to an early termination of the exponential phase and consequently, of the culture. Fed-batch cultures usually result in higher cell viability and growth, extended cultivation period and enhanced protein production compared to the batch systems[317]. Perfusion, a continuous bioprocessing regime, is another emerging operation mode that can significantly extend the fermentation period, achieving high cell densities and product cumulative amounts[318].

Extracellular pH and media osmolality

Much research has been conducted to elucidate the effects of pH on cell growth and protein glycosylation. As described in section 1.3.1, by-products of cell culturing, such as ammonia and lactate, can significantly perturb extracellular pH, halting cell growth and affecting protein glycosylation. The mechanism by which the extracellular pH affects *N*-linked glycosylation is through destabilizing the pH of the Golgi apparatus and hence reducing enzymatic activity[253, 319] and causing enzyme mislocalization within the Golgi[320]. Both the specific production rate and glycosylation of the recombinant protein have been found to be dependent on the extracellular pH in CHO cell cultures[321]. For example, pH variations between 6.8 and 7.8 have been shown to result in a 50% fluctuation of IgG1 galactosylation and sialylation[322]. Optimal pH ranges can also be identified depending on the desired glycoprofile and the recombinantly expressed protein[323-324]. Finally, a marginal 0.2 increase in Golgi pH was deemed enough to inhibit a3SiaT activity and cause its mislocalization in the Golgi apparatus. To that end, buffering or CO_2 supplementation can be employed for the regulation of extracellular pH in mammalian cell cultures[325].

In a similar manner, elevated levels of partial pressure of CO_2 (pCO_2) and osmolality have been shown to inhibit CHO[326] and hybridoma[327-328] cell specific growth rates. On the contrary, hyperosmotic conditions can increase productivity and final titre[329-331]. Elevated osmolality has also been linked to an elevation in Man5 levels across four different CHO cell lines[200] and a decrease in IgG fucosylation[332].

Mild hypothermia and dissolved oxygen

Other parameters affecting cell growth and productivity are the levels of dissolved oxygen[234, 333] and the culturing temperature[334]. Mild hypothermia, the switch of temperature to ~31-33 °C, has been found to increase CHO cell viability and antibody specific productivity, but also reducing galactosylation through the suppression of b4GalT expression and UDPGal synthesis[335]. Cultivation of CHO cells at 33 °C and 30 °C reduces EPO-Fc sialylation by 20 and 40%, respectively[336]. Similarly, introduction of mild hypothermia after the exponential phase can result in ~60% reduction of the potentially immunogenic N-glycolylneuraminic acid (Neu5Gc) in the glycoprotein B1[337]. Many

researchers have investigated the role of dissolved oxygen (DO) on cell growth and protein glycosylation, as different studies report mixed results for the effects of hypoxia[242, 338-339]. Increasing oxygen availability was found to strongly influence specific productivity and improve the sialylation of human follicle stimulating hormone (hFSH) in CHO cells, due to enhanced sialyltransferase activity[340]. Similarly, Kunkel et al.[341-342] reported that elevated DO results in improved IgG1 galactosylation. Moreover, the specific cell growth of EPO-expressing CHO cells was significantly reduced in cultures supplied with 200% DO but was not affected in DO levels below 50%[343]. Although specific EPO productivity was reduced at extreme hypoxia (3%) and hyperoxia (200%) levels, it remained relatively constant between 10% and 100% DO levels. The ratio between fucosylated and non-fucosylated glycans remained also constant during the experiments.

1.3.3 Engineering of the glycosylation pathway

Fucosylation

FUT8, the gene responsible for a6FucT expression, was one of the earliest targets of glycoengineering due to its strong effect on the ADCC activity of IgGs. Fig. 1.9 shows the reaction pathways that have been engineered for manipulating core-fucosylation. Following the discovery of the relationship between core-fucosylation and IgG potency, the complete knockout of the FUT8 gene in an IgG1-producing CHO-DG44 cell line led to a >100-fold increase of ADCC activity[344]. Prior to the FUT8 knockout, overexpression of the MGAT3 gene was employed for reducing IgG fucosylation, as the addition of bisecting GlcNAc molecules in the forming oligosaccharide can inhibit core-fucosylation[167]. In a follow-up study, the authors engineered chimeric versions of the GnTIII enzyme, bearing the localization sequences of other glycosylation enzymes, including GnTI, GnTII, ManII and a6FucT. Interestingly, the chimeric GnTIII enzyme carrying the ManII localization sequence led to the highest bisecting glycan levels. When both the aforementioned chimeric GnTIII and ManII were simultaneously overexpressed, the majority of the IgG1 glycoprofile consisted of complex bisected and non-fucosylated glycans[170].

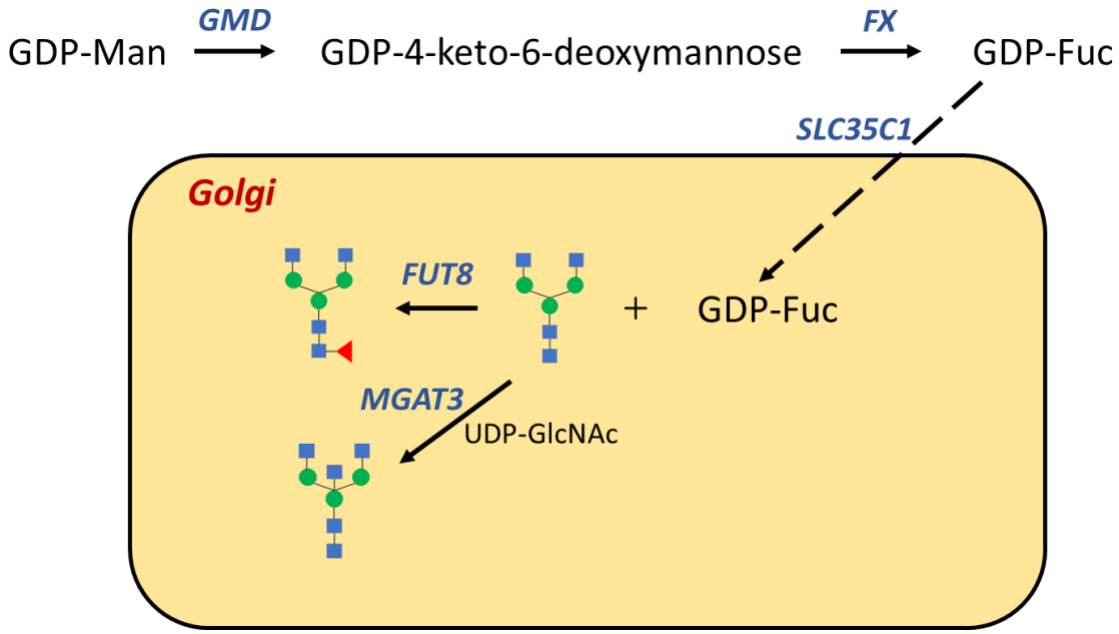

Figure 1.9: Different genes that have been manipulated towards inhibiting core-fucosylation and their part in the fucosylation pathway.

An alternative method to supress core-fucosylation is through the disruption of GDPFuc intracellular pools. To that end, the early development of the CHO-Lec13 cell line with a deficiency in the GDP-mannose-4,6-dehydratase (GMD) enzyme responsible for the synthesis of GDPFuc from GDPMan, resulted in significantly lower levels of core-fucosylation[345], while the knockout of the GMD gene led to complete silencing of fucosyltransferase activity[165]. To that end, double knockdown of both FUT8 and GMD has been proven an efficient method for disrupting core-fucosylation[346]. Similarly, the knockout of the GDP-4-keto-6-D-deoxymannose epimerase/GDP-4-keto-6-L-galactose reductase (FX) enzyme that is involved in the GDPMan to GDPFuc conversion too (Fig. 1.9), can significantly halt fucosylation[347]. Alternatively, silencing of the GDPFuc transporter (SLC35C1 gene), responsible for the molecule's translocation to the Golgi apparatus, has been found to effectively disrupt the core-fucosylation of a recombinant EPO-Fc fusion protein produced in engineered CHO cells[348].

Galactosylation

As described in section 1.2, terminal galactosylation is a desired IgG quality attribute for enhancing CDC activity. In mammalian cells, there are four genes that are primarily responsible for the majority of the b4GalT activity, namely B4GALT1, B4GALT2, B4GALT3 and B4GALT4. Whilst the knockout of the aforementioned genes has been well studied in order to evaluate the contribution of each individual gene towards *N*-linked galactosylation[349-351], the main interest lies in the overexpression of the enzyme. To that end, B4GALT1 upregulation has resulted in a homogenous IgG glycoprofile of fully galactosylated bi-antennary (G2F) structures[352]. The authors of the same study identified that the

knockout of the signal peptide peptidase-like 3 (SPPL3 gene), a protease responsible for the disruption of the bond between several GTases, including b4GalT, and their N-terminal membrane anchors[353], results in slightly improved IgG galactosylation. For recombinant proteins that present a more complex glycoprofile with several sialylated tri- and tetra-antennary glycans, the knockout of the ST3GAL genes can increase terminal galactosylation[349]. Interestingly, the triple knockout of FUT8, ST3GAL4 and ST3GAL6 alongside with the substitution of four amino acids in the IgG backbone leads to 80% G2 content[354]. The development of the CHO-Lec8 cell line that exhibits a strong deficiency of UDPGal transport from the cytosolic environment to the Golgi apparatus, led to a ~90% reduction of galactosylation[355]. However, the deficiency in UDPGal transporter levels (SLC35A2 gene) can be partially restored by the overexpression of the UDPGlcNAc transporter (SLC35A3 gene) that shows a weak affinity for UDPGal[356].

Sialylation, branching and polyLacNAc elongation

Unlike native human proteins, recombinant proteins synthesized in CHO cells do not carry alpha-2,6-sialic acid (a6SA) molecules due to the complete absence of ST6GAL gene expression. Instead, terminal sialylation is performed by the a3SiaT enzyme (ST3GAL genes). Whilst the alpha-2,3-sialic acid (a3SA) molecules are also present in human proteins and are not considered immunogenic[357], rituximab carrying a6SA has been found to exhibit increased FcγRIII binding and ADCC activity[358]. For that reason, Raymond et al.[359] overexpressed both the a6SiaT and b4GalT enzymes in order to favour a6SA addition over a3SA in IgG molecules (~25% total sialylation). Interestingly, the single overexpression of a6SiaT resulted in minor improvements on sialic acid content (~5% total sialylation), demonstrating the importance of galactosylation for the activity of the sialyltransferases. Moreover, engineering of the amino acid sequence in the IgG polypeptide backbone has been proven a reliable method for enhancing sialylation, galactosylation and even tri-antennary branching[360]. Substitution of the phenylalanine 243 amino acid in the backbone of an IgG3 antibody with an alanine molecule, has been found to result in conformational changes that increase the a3SiaT activity (~50% sialylation)[361]. Subsequent a6SiaT overexpression led to a further increase of sialylation at 60%, with an almost equal ratio between a3SA and a6SA molecules. Chung et al.[362] employed a double knockout of ST3GAL4 and ST3GAL6, followed by the overexpression of the ST6GAL1 gene. The authors achieved a depletion of a3SiaT activity and a considerable elevation the a6SA content in IgG molecules. Complete silencing or suppression of ST3GAL3, ST3GAL4 and ST3GAL6 has been also employed for investigating the contribution of each gene towards protein sialylation[349, 363]. Moreover, CHO-Lec2 cells that lack CMPNeu5Ac transporter (SLC35A1 gene) expression are unable to perform sialylation[364]. However, the phenotype can be completely reversed when the transporter is overexpressed[365]. Similarly, further

upregulation of CMPNeu5Ac transporter levels in CHO cells can result in ~15% increase in IFNγ sialylation[366].

Indirect methods, such as the knockout of sialidase genes[367] or suppression of glycosphingolipids synthesis[368], have been found to effectively improve recombinant EPO sialylation. Interestingly, CMPNeu5Ac synthesis is controlled by feedback inhibition as a function of the molecule concentration. When the function of the feedback inhibition was silenced, CHO cells exhibited considerably improved sialylation[369] that was significantly higher than the sialylation levels achieved through the overexpression of the CMPNeu5Ac synthetase enzyme[370]. In an effort to maximize EPO sialylation by increasing the complexity of the glycan structure, Yin et al.[371] investigated the effect of MGAT4, MGAT5 and ST6GAL1 overexpression. The authors identified that the aforementioned upregulation resulted in a considerably more complex glycoprofile with increased branching, sialylation and LacNAc (N-Acetyllactosamine) elongations (LacNAc is the Gal-GlcNAc disaccharide). In a recent study, the same group identified that sialyltransferase knockout or b4GalT1 overexpression can induce polyLacNAc elongation in an EPO molecule[372]. Surprisingly, upregulation of the iGnT enzyme (B3GNT2), responsible for LacNAc formation[349], resulted in reduced sialylation, galactosylation and polyLacNAc elongation of the EPO protein, indicating that a more elaborate investigation is required for further extending the glycan structures.

1.4 Metabolic and glycosylation mathematical modelling

A big part of this thesis evolves around the application of mathematical models for the representation, description and simulation of various metabolic parts of mammalian cells, including protein *N*-linked glycosylation. The emerging use of computational models for expanding our understanding of cellular metabolism and predicting cellular behaviour under alternative culture conditions or post the application of genetic modifications, has established mathematical modelling as a very powerful tool in metabolic engineering and process optimization[373]. The final section of the literature review includes a description of the main approaches employed for describing cellular metabolism, with an emphasis on mammalian and, more specifically, CHO cell metabolism. Finally, the evolution of *N*-linked glycosylation modelling through the years and its various applications on biotherapeutics development and control is described.

1.4.1 Constraint-based metabolic models

Development of constraint-based models (CBMs) is one of the most popular approaches on describing mammalian cell metabolism, especially for steady state scenarios[374]. Steady state assumes no changes in the system and, therefore, no changes in the value of each of the included variables (state variables) within the considered timeframe (Eq. 1.1).

$$\frac{df}{dt} = 0 \quad (1.1)$$

where, f is a state variable of the examined system.

Importantly, CBMs are strongly dependent on the vector of stoichiometric matrices (S) that represents the involvement of each metabolite (i) on each reaction (j) of the examined network. The fundamental variable in CBMs is the *flux* of the reaction (v is the vector of fluxes with units: mmol·gDCW^{-1}·h^{-1}), that essentially describes the flow of metabolites through the designated metabolic pathway. Fluxes are usually constrained between a maximum (max_j) and a minimum (min_j) value, based on the reversibility of the reaction and the available experimental data, i.e., transcriptomics[375-376], metabolomics[377-378], thermodynamics bounds[379-381] or results from supplementary computational tools[382-384]. The solution of CBMs usually involves two main steps: 1) the reconstruction of the metabolic network and the estimation of flux constraints and 2) the identification of the optimal solution based on an objective function[385]. The majority of CBMs use genome-scale models (GeMs) as the backbone for network reconstruction. GeMs include the entirety of known reactions, metabolites and genes involved in the metabolism of the examined cell line, in an effort to holistically describe the system. GeMs for CHO cells are systematically updated to include missing reactions[386] or pathways[52], and specific modules for each cell-line[28]. CBM-based estimation of the cellular metabolic profile (c) is usually achieved through linear programming (LP) techniques for solving the steady-state formulation of the problem, as described in Eq. 1.2 and Eq. 1.3. The differing point between the various categories of CBMs usually lies in the chosen objective function for the identification of the optimal solution. With the number of reactions in GeMs being higher than the respective number of metabolites, the degrees of freedom can considerably vary, resulting in a major dependence of the optimal solution on the chosen objective function.

$$\frac{dc}{dt} = S \cdot v = 0 \quad (1.2)$$

$$min_j \leq v_j \leq max_j \quad (1.3)$$

Flux balance analysis (FBA) is the most frequently applied category of CBMs in investigating intracellular metabolism[385]. For that reason, there have been numerous attempts to enrich and extend the list of objective functions that can be employed for solving an FBA problem[387-389]. The most common amongst them is the maximization of biomass growth, as it has been experimentally proven to capture the objective of bacteria cellular behaviour[390].

FBA is commonly employed in order to elucidate the effects of culturing perturbations, such as mild hypothermia[335], lactate switch[273] or cell-size increase[391], on intracellular CHO cell metabolism. In a

similar manner, flux variability analysis (FVA) is used for the estimation of flux boundaries in scenarios of suboptimal conditions and can be applied for the evaluation of model robustness[382, 392]. For gene knockout studies, where the engineered cells do not evolve towards maximal growth, the method of minimization of metabolic adjustment (MOMA) has been found to significantly outperform FBA. MOMA achieves convergence through the minimization of the distance between the new state and the previously (wild-type) identified optimal point[393].

Another application of CBMs is metabolic flux analysis (MFA). Unlike FBA problems, MFA's objective function is the minimization of the difference between experimental data and model simulations. Whilst MFA is usually focused on describing the fluxes of central carbon metabolism[394-395], it has also been used for investigating the *E. coli* GeM[396]. MFA can be performed either by using ^{13}C-labelled substrates (^{13}C-MFA), such as glucose, or through the calculation of the intracellular fluxes based on the agreement between the stoichiometric model and the extracellular concentrations of measured metabolites[397-398]. In CHO cells, MFA has been used, among other applications, to investigate the effects of mild hypothermia[399] and copper supplementation[400] on cellular metabolism and to demonstrate the divergence of fluxes between the growth and stationary phases[401].

Dynamic versions of both FBA (dFBA) and MFA (dMFA) have been developed in order to describe flux variation during the cell culturing period and overcome the limitations of the steady-state assumption[402-403]. Lastly, another approach on CBMs, namely elementary flux modes analysis (EFMA), identifies all the minimal pathways (elementary flux modes, EFMs) that collectively construct the considered metabolic network and within which all reactions are necessary for maintaining the steady-state assumption (non-decomposable)[404-405]. Importantly, the number of EFMs disproportionally increases with the complexity of the network, as several possible minimal pathways can be potentially identified for each set of additional reactions[406]. In order to reduce the number of plausible EFMs within a given network, several algorithms that incorporate criteria such as biological relevance and thermodynamic feasibility have been developed[407-409]. As EFMA identifies all the possible pathways through which a product of interest can be synthesized, EFMs are considered an important tool for metabolic engineering[405].

Whilst CBMs do not require the estimation of kinetic parameters, they are heavily based on the chosen objective function and solution approach, and they are usually constrained by the steady-state assumption. The high inherent variability of the analysis, the requirement for high-accuracy measurements and existence of multiple solutions for the same system, introduce additional limitations on CBMs applicability. The mathematical formulation for the aforementioned CBM

approaches is described in Table 1.3. An interested reader can find further information on CBMs in the excellent reviews of Llaneras & Picó[410] and Zomorrodi et al.[411].

Table 1.3: Main characteristics and formulations of the major CBM approaches. The mathematical formulations for each method were derived from: FBA[385], MOMA[393], MFA[412] and EFMA[413].

Method	Formulation	Comments
FBA	$S \cdot v = 0$ $min_j \leq v_j \leq max_j$ $Z = c^T \cdot v$ $\max(Z)$	Z is the objective function. c is the vector of the weights of each reaction towards Z. Note that when the maximization of biomass is set as the objective function then c is a vector of zeros apart from the reaction of biomass synthesis, where $c_j = 1$.
MOMA	$D(w,x) = \sqrt{\sum_{i=1}^{N}(w_i - x_i)^2}$ $\min(D)$	$D(w,x)$ is the Euclidian distance between the optimum flux vector of the wild-type (w) and the MOMA solution (x).
MFA	$\min(\|v^{exp} - v^{MFA}\|^2)$	The objective function of the MFA problem includes the minimization of the difference between the experimentally measured (v^{exp}) and calculated fluxes (v^{MFA}).
EFMA	$N \cdot e = 0$ $e_i \geq 0$ $if\ R(e') \subseteq R(e)\ then\ e'=0\ or$ $\|e'\|=\|e\|$	*N* represents the stoichiometric table of the EFMs' vector (e). Each reaction is considered irreversible. Reversible reactions are separated to two distinct reactions. *Elementarity* dictates that there is no subset of the *e* vector that is non-zero or unequal to the absolute value of *e*.

1.4.2 Kinetic metabolic models

Kinetic metabolic models (KMMs) attempt to describe cellular metabolism with the utilization of known underlying biological relationships that govern the system. The vast majority of KMMs employ Monod-type equations for describing the relationship between reaction rates and the considered metabolites[414]. Monod-type equations follow the structure of Eq. 1.4:

$$f(S) = \prod \frac{S_i}{S_i + K_{Si}} \quad (1.4)$$

where, $f(S)$ is the Monod-type function, S_i is the i^{th} substrate concentration (mM) and K_{Si} is the Monod constant (mM).

A popular alternative for Monod-type equations regularly employed for describing mammalian metabolism is the logistic function[415]. A common classification of KMMs distinguishes the models as structured or unstructured, segregated or non-segregated[414, 416]. Structured models examine cellular metabolism in each organelle of the cell in distinct submodules, in an effort to develop a detailed mechanistic representation of the system. On the contrary, unstructured models consider the cell as

one entity without investigating the differences in the intracellular level. Moreover, the difference between non-segregated and segregated models is that the latter consider the heterogeneity among cellular populations, i.e., depending on the phase of the cell cycle, cell age or DNA content. Population balance models (PBMs) are one of the most widely applied segregated approaches for modelling mammalian cell cultures[417-418]. Kyriakopoulos et al.[414] identified that unstructured and non-segregated models are the most frequently utilized kinetic tools for describing the cell culture process *in silico*. As shown in Fig. 1.10, model complexity and, therefore, the challenges in parameter estimation, increase as the model structure moves towards segregation and structural identification. Kontoravdi et al.[419] and Jedrzejewski et al.[121] developed a semi-structured and non-segregated approach on linking and integrating different models for simultaneously describing extracellular metabolism, intracellular synthesis of the recombinant antibody, intracellular NSD and nucleotide synthesis (only in Jedrzejewski et al.) and Golgi *N*-linked glycosylation. Due to the frequency that unstructured or semi-structured and non-segregated models are found in literature and because of their relevance to the thesis, the rest of the current section will mainly focus on the aforementioned sub-category.

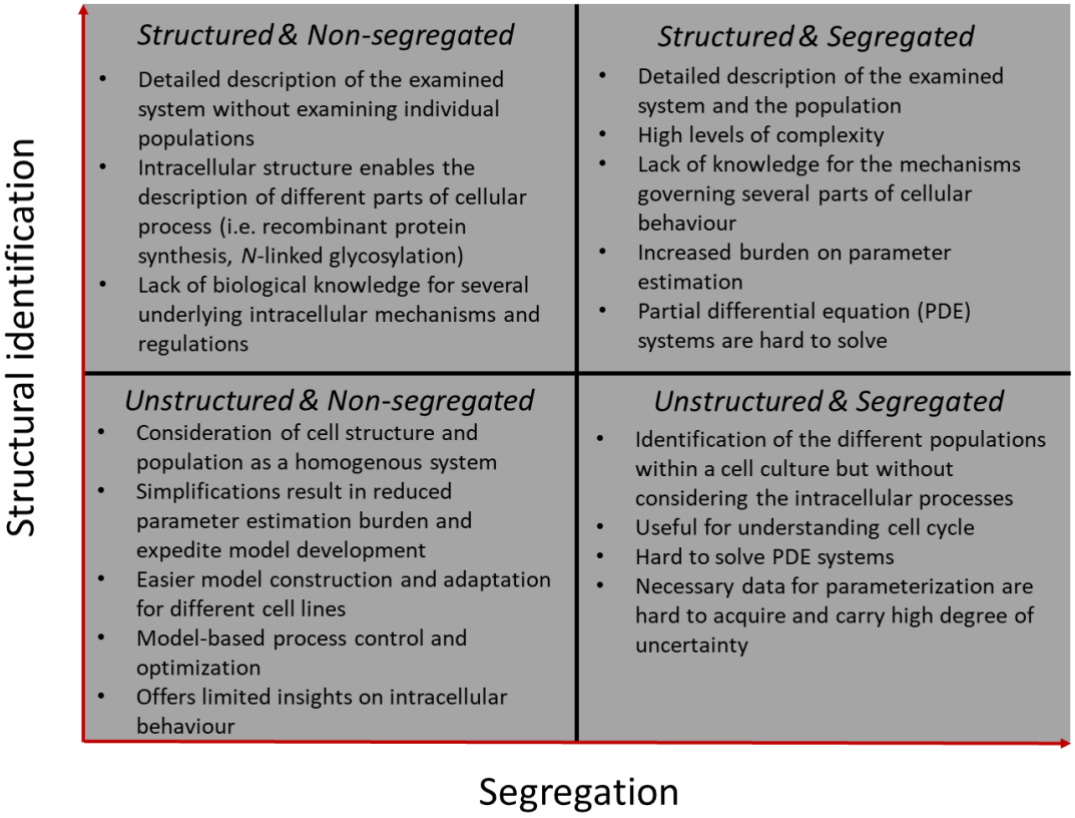

Figure 1.10: Main characteristics of the different combinations between the KMM categories.

The metabolic networks considered in KMMs are considerably reduced compared to the entirety of cellular metabolism, mainly examining the *limiting* and *inhibiting* substrates of cell growth. Limiting

are the metabolites or amino acids that are essential for cell metabolism and the depletion of which can significantly halt cell growth (i.e., glucose, glutamine). On the other hand, inhibiting are the metabolic by-products that hinder growth when accumulated (i.e., lactate). In addition, cell death is commonly expressed as a Monod-type function of the toxic by-products that directly induce cell death (i.e., ammonia). All aforementioned substrate-categories, alongside the structure of the network, are identified prior to model construction, either through the evaluation of experimental or CBM-based data[416, 420-421], or based on statistical analysis[422-423]. KMMs can also capture the entirety of cellular metabolism through the incorporation of GeM-networks[424-425]. Khodayari & Maranas[426] have developed a kinetic model describing the GeM of *E.coli* with 457 reactions and 337 metabolites. The authors found that the KMM-GeM approach considerably outperformed FBA and MOMA results. However, the development of KMM-GeMs and the expansion of the kinetic interpretation of genomic networks in mammalian cells have not been broadly demonstrated yet, due to the complexity of mammalian metabolism, the computational challenges imposed by parameter unidentifiability and the extent of experimental work required for model parameterization[427].

The most common building blocks of KMMs are the material balances that govern the examined system throughout the culture, commonly defined in the form of ordinary differential equations (ODEs). The differential dependence between substrates and time enables the dynamic simulation of the model throughout the culturing period and its adjustment for both batch and fed-batch systems[428-429]. The dynamic nature of KMMs is considered the primary difference from CBMs that are mostly simulated under the steady-state assumption[430]. However, the calculation of the KMM parameters values, a process termed as parameter estimation, is considerably more complex than the optimization algorithms employed in CBMs[431]. Therefore, the estimated parameters rarely represent the actual values and the global solution of the system. The aforementioned considerations on parameter estimation are attributed to the lack of parameter identifiability in KMMs, meaning that different combinations of the kinetic parameter values can result in the same model behaviour[432]. Despite the inability to tightly constrain parameter values of KMMs through model training on experimental data, kinetic models can demonstrate strong predictive powers[433].

Sensitivity analysis can be employed for evaluating the significance of model parameters towards the designated outputs of interest, or all measured outputs/states, therefore enabling the identification of parameters that necessitate precise estimation against relevant experimental data[434-436]. The non-significant parameters can be assigned to nominal values based on previous work and literature data. However, it is important to note that the assignment of parameters to their *in vitro* calculated values is not always recommended, as enzymatic behaviour can significantly deviate between the *in vitro* and *in vivo* conditions[437]. Approaches combining CBMs and KMMs can relieve some of the burden of

parameter estimation and incorporate more comprehensive datasets and cellular functionalities, i.e., the gene expression regulatory network[438-439]. Despite all the efforts on simplifying the parameter estimation process, the latter remains highly dependent on the structure of the model and cannot be easily generalized. In the case of mammalian cell cultures, where the culturing process is considerably time-consuming and the cells exhibit high degrees of variability between the different clones, the collection of an adequate dataset for fine-tuning the KMM remains a challenging task.

1.4.3 Data-driven approaches on modelling cellular metabolism

The use of data-driven models (DDMs) for simulating, controlling and optimizing bioprocesses is currently emerging[48]. Regarded as "black boxes", DDMs are usually integrated within a "hybrid" framework that additionally incorporates a CBM or a KMM module, providing that way a flexibility in determining the exact framework configuration based on data availability[440].

Due to the necessity for large and comprehensive datasets in training both DDMs and CBMs, the latter two are regularly combined for hybrid applications[441]. In the majority of applications, ML techniques are utilized for incorporating multiomics data towards assisting fluxomic analysis[54]. Unsupervised learning, meaning the training of an algorithm without providing the output values (labels), is mainly used for dimensionality reduction[442] and data interpretation[443], and has also been extensively applied in EFMs identification[444-447] and FBA assistance and results interpretation[448-449]. Supervised learning, meaning the construction of a regression or classification algorithm between the input (features) and output values, has found various applications in CBM-assisted predictions of cell growth and metabolism[450-451], due to the large availability of well-established supervised techniques. As a further elaboration on CBM-DDM applications would fall out of this thesis scope, an interested reader is encouraged to find further information in the comprehensive and recent reviews of Antonakoudis et al.[452], Rana et al.[441] and Zampieri et al.[54].

Data-driven approaches have been widely utilized for reinforcement learning and model predictive control (MPC) applications towards bioprocess control and optimization[453-455]. In an interesting study, an ANN configuration was found to overperform its mechanistic equivalent in the simulation and optimization of microalgal lutein production in a fed-batch regime[456]. In a second step, the authors developed a hybrid framework incorporating a kinetic module for correcting and imputing raw and missing data, that were subsequently fed to an ANN for predicting future system behaviour[457]. The kinetic model was then re-trained on the projected data-driven data, resulting in the robust control and optimization of the examined bioprocess. Alternatively, a convolutional neural network (CNN) has been successfully trained on the results of a computational fluid dynamics (CFD) and a kinetic model in order to optimize biomass growth and production[458].

Amongst mammalian cell cultures, the application of stand-alone DDMs has mainly focused on the statistical approaches and multivariate analysis. The applications of DDMs on controlling the cell culture process towards optimizing product quality attributes, such as the *N*-linked glycosylation profile of the product, will be discussed in the following section. Multivariate algorithms, such as principal component analysis (PCA) and partial least squares regression (PLS) can be used for identifying similar batches[459], interpreting and grouping early process-development data[460-463] and monitoring cell culture progress[461, 464-465]. In general, the use of such multivariate algorithms enables the better clustering, filtering, visualization and interpretation of the multidimensional culturing data.

Whilst DDMs are a valuable tool for process monitoring and implementation of the QbD and process analytical technology (PAT) paradigms, they have found limited applications, either as a standalone algorithm or within a KMM-hybrid configuration, for directly predicting mammalian cell metabolism. This observation can be attributed to the excessive demand for large datasets during DMMs' training, the lengthy and costly experimentation necessary for the generation of such datasets and the already satisfactory performance of KMMs in dynamic modelling. In an early effort to integrate ML modules within a kinetic framework, 6 parallel ANN models were employed for the estimation of the specific growth, product synthesis and metabolite uptake rates[466]. The parallel ANNs were trained on the same inputs, being process parameters such as pH, culture time, nitrogen flow etc., and each one was used for the calculation of a single specific rate. In a second step, the ANN-estimated rates were introduced to the material balances of a KMM, for the calculation of metabolites concentration, viable cell density and titre. However, it was only recently that the Morbidelli group re-established the use of ML for describing mammalian cell cultures. Initially, Narayanan et al.[467] developed a data-driven algorithm incorporating decision trees and PLS models for the prediction of product titre and several process CQAs. In a second step, the authors developed a hybrid model incorporating a single feed-forward ANN and a mechanistic model for describing CHO cell cultures[468]. The data-driven algorithm was used to calculate the specific growth rate that, in turn, was introduced to the kinetic model for predicting process characteristics, such as viable cell density, glucose, lactate concentration and titre. Finally, the hybrid model was found to outperform its equivalent PLS-oriented statistical model for all examined variables, further demonstrating the power of hybrid configurations towards describing and controlling mammalian cell cultures.

1.4.4 *N*-linked glycosylation modelling

Simulation and prediction of recombinant protein glycosylation has been of extensive interest during the last two decades, for improving both process control and design. Several frameworks have been developed in order to describe *N*-linked glycosylation and predict the effects of gene engineering and culturing alterations, with the vast majority of them being mechanistic models. However, data-driven

or statistical approaches are also emerging in order to alleviate the parameter estimation burden and expedite model development. The major discrepancy between the proposed mechanistic models lies in the choice of the Golgi protein-traffic model. From a model-designing perspective, the CM model is usually described through a single or a series of plug and flow reactors (PFRs) in order to represent the maturation of the Golgi and therefore, the gradual transformation of the *cis*-cisterna to the latter TGN compartment. On the other hand, the VT model is usually simulated through a series of continuous stirred-tank reactors (CSTRs), where each CSTR represents one of the Golgi compartments. However, the use of different reactor systems results in considerably altered model complexity. PFRs describe the glycan concentration as a function of both the culturing time and the distance in the Golgi apparatus, requiring that way partial differential equations (PDEs) for simulating the system. In addition, enzyme concentrations usually follow Gaussian distributions within the Golgi organelle. On the contrary, well mixed CSTRs assume a homogenous concentration for glycans and glycosylation enzymes in each compartment and therefore utilize ordinary differential equations (ODEs) that are considerably less complex and easier to solve. The following paragraphs attempt to comprehensively describe the gradual evolution of glycosylation modelling through the last, approximately, three decades, and showcase the applications of glycosylation models on predicting the effects of gene engineering and process manipulation. Fig. 1.11 summarizes and presents the relevant studies in a chronological order. It is important to note that the advances in *N*-linked glycosylation analysis and quantification are one of the major drivers for the development of more elaborate and detailed kinetic models in the recent years.

The work by Noe & Delenick[469] was one of the earliest to incorporate the glycosylation process in a kinetic model. However, the authors only examined the secretion process of newly synthesized proteins through the intracellular compartments, including the effect of PTMs on transport rates, and therefore, did not investigate the microheterogeneity of protein glycosylation. Shelikoff et al.[470], developed a mechanistic model to describe glycosylation site occupancy based on factors such as mRNA translation rate, protein synthesis rate and precursor oligosaccharide availability. Whilst the authors introduced mechanistic and kinetic concepts, such as material balances in the Golgi within a PFR representation, that would later be used to define the first complete glycosylation model, the study only examined site occupancy and did not simulate glycans distribution. Moving towards a mechanistic model describing glycans heterogeneity, Monica et al.[471] developed a kinetic framework for estimating the sialylation of glycoproteins recombinantly produced in CHO cells, such as CD4 and tPA. While the model did not examine the detailed glycosylation profile of the proteins, it introduced several considerations that were subsequently re-examined for developing a more holistic glycosylation framework, such as: the use of a CSTR for describing Golgi compartments, the NSD

availability in the Golgi, the GTase localization and the inhibition of enzymatic activity by competitive substrates.

Working in parallel with Monica et al., Umaña & Bailey[472] developed the first mechanistic kinetic model (UB1997 model) that could describe glycoforms distribution, based on a glycosylation network of 33 reactions and glycan structures. Importantly, the authors assumed that each glycan structure could act as substrate for a single enzyme, resulting in a network with an equal number of reactions and possible glycoforms. The kinetic parameters of the model were assigned values from literature. The reaction network did not include the activity of the a6FucT, iGnT and a3SiaT enzymes, in an effort to control the size of the system and following the assumption that the presence of core-fucose would not considerably affect the reaction kinetics. The authors employed four CSTRs to describe the different compartments of the Golgi and investigated the effect of glycoprotein productivity and gene overexpression on specific parts of the glycosylation process.

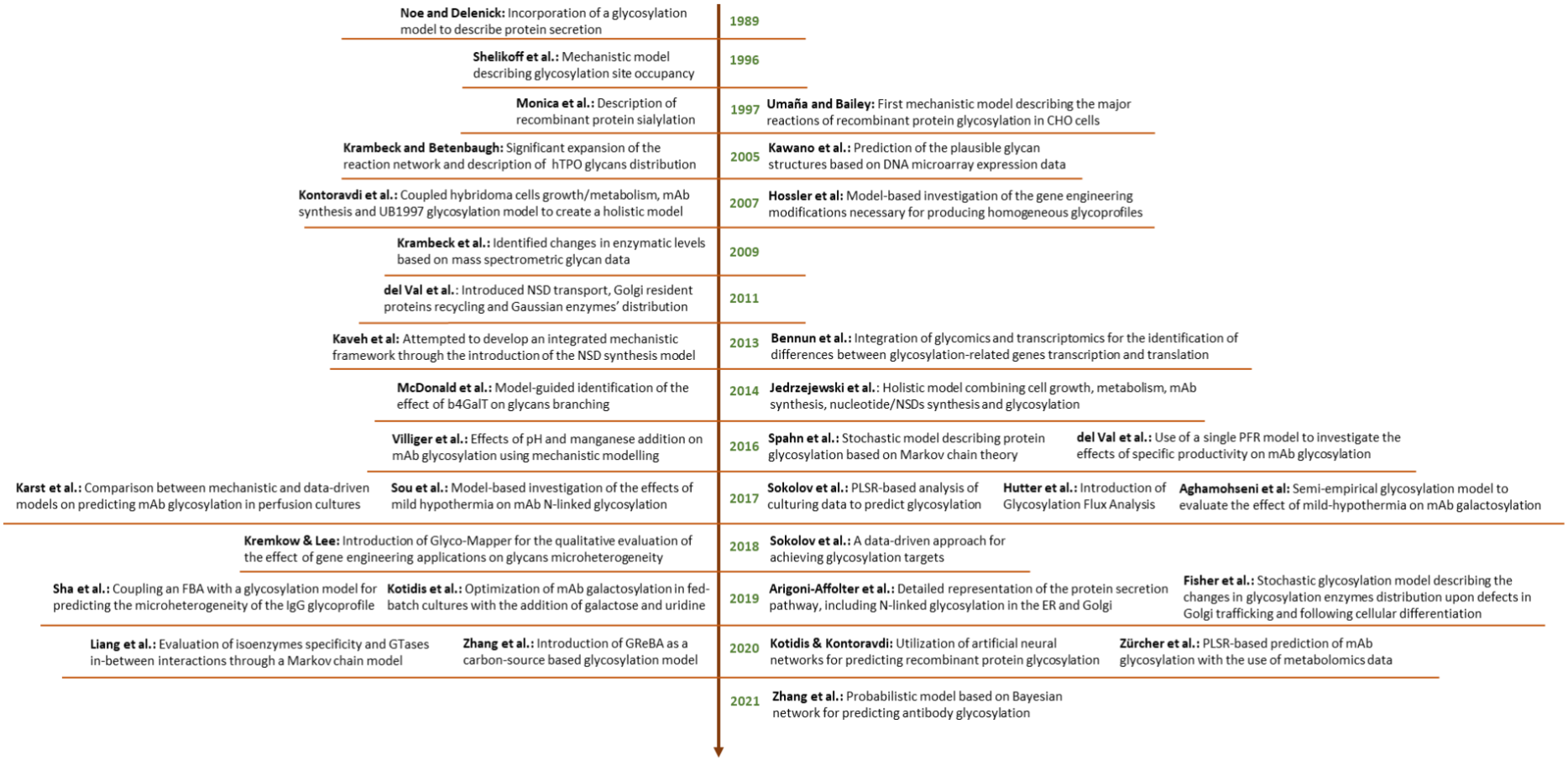

Figure 1.11: The evolution of *N*-linked glycosylation models and their applications.

In a work of major importance, Krambeck & Betenbaugh[473] further expanded the UB1997 network to account for the activity of the missing enzymes and incorporate 7,565 glycans structures and 22,871 reactions in total (KB2005 model). The KB2005 model included literature-derived rules to define the activity of the involved enzymes on each substrate, in an effort to mechanically reconstruct the glycosylation network and extract the values of the kinetic parameters. Importantly, the authors demonstrated that by *in silico* manipulating the concentration of the glycosylation enzymes, the model could closely describe the glycoprofile of recombinantly produced human thrombopoietin (hTPO). The adjustment of enzymes concentration is a necessary step towards the construction of protein-specific models, as it compensates for the discrepancies in the expression levels of the glycosylation enzymes between different cell lines and the steric hinderances that each glycoprotein imposes to the considered enzymes. However, it is important to note that the aforementioned steric hindrances are more accurately described by manipulating the inhibition constants of the reaction rates rather than enzyme concentrations. The results of the KB2005 study were the first to demonstrate that a kinetic model can accurately describe the detailed glycoform distribution of recombinantly produced glycoproteins and exemplify the use of such models for further controlling the glycosylation process. In a follow-up study, the authors developed a framework for adjusting the glycosylation model to mass spectrometric glyco-data of healthy and leukaemia monocytic cells, therefore, enabling the identification of differences in the variety of glycan structures and enzyme concentrations between the two cell lines[474]. The framework has also been utilized to describe the differences in expression levels of glycosidases and GTases between various CHO cell lines carrying genetic modifications, based on the glycosylation profile of the intracellular proteins[475].

In a very informative study, Hossler et al.[476] utilized a single PFR model (H2007 model) to evaluate the sensitivity of terminally processed glycans on enzymes concentration, identify the changes of glycans distribution within the Golgi and estimate the necessary regulations of enzyme profiles towards the maximization of terminal glycans synthesis. Interestingly, the authors assumed a uniform profile for enzymes concentration within the reactor, with only glycans distribution being a function of the distance within the Golgi. In an effort to achieve complete homogenization for each glycoform, the authors developed a 4-PFR model that was found to outperform its 4-CSTR equivalent. Model-based investigation indicated that residence time and enzymes localization were the two factors affecting glycoprofile homogeneity the most. Kontoravdi et al.[419] incorporated the UB1997 model within a kinetic framework describing hybridoma cells metabolism and mAb synthesis[477]. The proposed framework was the first modelling approach aiming to connect the extracellular metabolism with mAb synthesis and glycosylation, paving the way for the development of feeding strategies to simultaneously control and optimize recombinant protein concentration and glycosylation.

Del val et al.[478] further improved the mechanistic representation developed in Hossler et al.[476] by assuming a more realistic Gaussian distribution for the enzymes' concentration in a single PFR model, accounting for more detailed kinetic mechanisms for each enzyme, introducing Golgi proteins recycling and incorporating NSD transport from the cytosol to the Golgi apparatus (DV2011 model). The latter set the basis for the development of a holistic kinetic framework relating central carbon metabolism to protein glycosylation through the NSD synthesis and transport. The DV2011 model was found to outperform the KB2005 and H2007 equivalents in predicting the effect of a6FucT knockout on mAb glycosylation. Towards the development of a holistic framework, Kaveh et al.[479] combined the H2007 and DV2011 models with discrete cell metabolism and NSD synthesis models, in order to describe cell growth, glutamine and galactose concentration, alongside with mAb glycosylation. The rationale behind the model's development was that NSD synthesis depends on glucose and glutamine extracellular concentration. In its turn, the NSD profile can influence glycosylation and alter glycans distribution. However, the authors did not explicitly present the integration of extracellular glucose and glutamine concentrations in the NSD material balances.

Working towards the development of a holistic kinetic model that could mechanistically connect cell metabolism to protein glycosylation through the NSD synthesis pathway, Jedrzejewski et al.[121] integrated the DV2011 model within a framework that included a detailed description of the NSDs and nucleotides synthesis pathways. The model of extracellular metabolism was used for describing cell growth and metabolism and, more specifically, for the estimation of extracellular glucose, glutamine and the specific cell growth and antibody production rates. The aforementioned estimated variables were fed to the nucleotide/NSD synthesis model that, in its turn, calculated the fluxes of NSD towards the Golgi apparatus and the glycosylation module. The concept of holistically describing the cell culture process and relating protein glycosylation to extracellular metabolism through NSD metabolism, has subsequently been used to capture the effects of mild hypothermia[480], pH fluctuation and manganese feeding[235, 481] on IgG glycosylation, and optimization of IgG galactosylation through galactose and uridine supplementation[482]. Alternatively, a lumped and significantly reduced semi-empirical model has also been developed for capturing the effect of mild hypothermia specifically on mAb galactosylation without examining the detailed microheterogeneity of the glycans[483]. Sha et al.[484] employed an FBA model to estimate the NSD synthesis fluxes that were subsequently introduced to the DV2011 glycosylation model for calculating glycans distribution. In an effort to alternatively describe the dependence of IgG glycosylation to the extracellular levels of the NSD-precursor sugars, Zhang et al.[485] developed a framework of macro-reactions for calculating glycans distribution as a direct function of the respective fed carbon-sources. The authors subsequently applied the model to

control the consumption rates of the available carbon sources and manipulate IgG glycosylation in CHO cell perfusion cultures[50].

Whilst the major applications of kinetic models have been towards the prediction of recombinant product glycosylation, there have also been several efforts to utilize the available computational tools for elucidating cellular characteristics and mechanisms. For example, the DV2011 model has been used to investigate the effect of CHO cells specific productivity on Golgi volume and linear transport velocity of the protein through the Golgi, and, subsequently, on mAb glycosylation[486]. McDonald et al.[487] identified, through a model-based methodology, that b4GalT4 expression levels influence recombinant human chorionic gonadotropin (hCG) branching in CHO cells. Additionally, the KB2005 model has been employed for the identification of differences between the transcriptomic and proteomic levels of glycosylation-related genes, combining glycomics and transcriptomics data[488]. The KB2005 model has also been applied for glycosylation controllability, meaning the identification of input ranges (i.e., enzyme and NSD concentrations) within which the control of specific glycans distribution is possible[294]. Moreover, Kremkow & Lee[489] developed a flux-based framework for successfully identifying the qualitative glycoprofile of the targeted protein, following gene engineering modifications in CHO cells. Similar work has been performed for predicting plausible glycan structures based on gene expression data[490-491]. In a recent study of great importance, Arigoni-Affolter et al.[492] incorporated the DV2011 model within a larger kinetic framework that, in combination with extensive experimentation, was employed for exhaustively describing the glycoprotein secretion pathway. Application of MFA in the glycosylation reaction network has enabled the development of glycosylation flux analysis (GFA)[493], which has been used for identifying underlying links between IgG galactosylation and process parameters, such as specific IgG productivity rate and extracellular ammonia accumulation[494]. Additionally, site-specific flux analysis in the glycosylation network of a protein disulfide isomerase (PDI) molecule has demonstrated that the extent of glycan complexity in the designated sites of the glycoprotein is mainly limited from the interaction between specific amino acids of the backbone and the glycan, rather than from the steric hindrance imposed by the glycoprotein structure[495].

Apart from the kinetic and flux-based approaches on modelling protein *N*-linked glycosylation, there has been an increasing interest in employing data-driven and stochastic algorithms for replacing their kinetic counterparts and accelerating model development. One of the most representative works in the field has been the development of Markov chain models for predicting the effect of gene downregulations on IgG, EPO and secretome glycosylation[496]. Importantly, Markov models assume that the future state of the examined variable depends solely on its current state, ignoring the past states of the variable. The aforementioned assumption conveniently matches the glycosylation

process, where the future steps of the pathway only depend on the affinity between the available enzymes and the current structure of the glycan. However, model compartmentalization according to the Golgi structure was found to be of paramount importance for strengthening the predicting capabilities of the Markov model. The Markov chain model has also been employed for guiding cell line glycoengineering for biosimilar development[497] and identifying interactions between GTases and isoenzyme specificity[498]. In a similar manner, Fisher et al.[499] developed a stochastic glycosylation model for predicting the effects of Golgi trafficking defections and cellular differentiation on the distribution of glycosylation enzymes in the Golgi. The model has also been utilized to *in silico* guide the maximization of desired glycoforms[500]. A probabilistic glycosylation model utilizing the Bayesian theorem has also proven useful for predicting antibody glycosylation in both perfusion and fed-batch cultures[501]. Alternatively, Sokolov et al.[502] have employed PLS models to utilize designated process parameters, based on a genetic algorithm selection, for predicting mAb glycosylation under several process conditions. A similar methodology was also deployed for incorporating metabolomics data in predicting recombinant protein glycosylation[503]. Finally, as mentioned in section 1.2.2, glycosylation models have been successfully utilized for assisting the implementation of the QbD paradigm[240-241]. In a similar manner, multivariate frameworks combining PCA and decision trees algorithms have been extensively utilized for guiding and manipulating biosimilar glycosylation towards the desired target profile[460, 504-505].

Chapter 2 Model-based optimization of monoclonal antibody galactosylation

Chapter overview

Monoclonal antibodies normally carry two sites of *N*-linked glycosylation in the constant Fc-fragment. Whilst the *N*-linked glycosylation profile of an IgG molecule mainly constitutes of bi-antennary glycan structures, the glycoform distribution has been found to be considerably dependent on the bioprocess conditions, such as temperature[335], nutrient availability[288] and catabolite accumulation[247]. The influence of bioprocess conditions on glycoform distribution translates to an effect on the safety and potency of the recombinant antibody. More specifically, the presence of terminal galactose (galactosylation) has been found to enhance CDC[171] and ADCC[182] activity of monoclonal antibodies and is therefore a desired attribute for biopharmaceutical products. However, galactosylation has been identified as a source of mAb structural variability during scale-up and technology transfer processes[292] and therefore necessitates further control and optimization to avoid potential effects on protein structure and efficacy[169, 195, 199]. Increase of mAb galactosylation has been mainly achieved with the supplementation of galactose and uridine, the metabolic precursors of UDPGal, during the cell culture period[292-293, 297]. UDPGal acts as the co-substrate for mAb galactosylation that is catalysed by the b4GalT enzyme. However, the inclusion of galactose and uridine in the feed can induce metabolic perturbations and inhibit cell growth and productivity[293].

Several kinetic mathematical models have been developed in order to describe CHO cell metabolism[506-507] and *N*-linked glycosylation[472-473, 476, 478, 485]. In addition, a probability-based approach employing Markov-chain models has also been employed in order to describe protein glycosylation and elucidate isoenzyme specificity[496, 498]. In an effort to create a link between the extracellular concentrations of nutrients and mAb glycosylation, Jedrzejewski et al.[121] introduced a modelling framework for the description of NSD and nucleotide synthesis pathways as a channel between the extracellular and glycosylation profile of mammalian cells.

Following the concept of Quality by Design that dictates a detailed understanding of the interconnections between process conditions and product quality, a mathematical model that describes the effect of different feeding scenarios on titre levels and glycosylation would be a valuable tool for process monitoring and optimization. To that end, Chapter 2 presents the development of a mechanistic mathematical model, based on Monod-type equations, to describe CHO cell metabolism, antibody production, NSD synthesis and mAb glycosylation in fed-batch cultures. Particular effort was placed on predicting the effect of galactose and uridine addition on cell growth, cell productivity and

mAb galactosylation. Following model construction, development and parameterization, the model was utilized for the identification of an optimized feeding strategy that would improve mAb galactosylation while simultaneously avoiding compromises on cell growth and productivity. Model validation was performed through the comparison with the experimental results of the optimized feeding strategy.

The model developed in this chapter sets the basis for the following chapters and the rest of the experimental and computational work presented in this thesis. The model has been presented in Kotidis et al.[482]

Overall, Chapter 2 aims to investigate and provide insights on the following points:

- Can a Monod-type based kinetic model describe CHO cell metabolism and mAb glycosylation?
- Can a model-based feeding strategy lead to optimized mAb galactosylation, while maintaining cell growth and productivity?

2.1 Materials & Methods

2.1.1 Description of experimental data

The experimental data used for the construction, parameterization and validation of the model were taken from Jedrzejewski[311]. Five experimental datasets, incorporating the addition of galactose and uridine in the culture, were used for model development and parameterization: control, 10G, 10G5U, 10G20U and 50G5U, the details of which are described in Table 2.1. An independent feeding fed-batch experiment, the design of which is based on the developed model and targets the optimization of mAb galactosylation, was used for external validation (Table 2.1). Briefly, CHO-T127 cells producing an IgG antibody were maintained in suspension and passaged every 3 days in CD-CHO medium (Life Technologies, Paisley, UK) at 36.5 °C, 5% CO_2 and 150 rpm. 50 μM of methionine sulfoximine (MSX) was added for cell revival and in the first passage. The fed-batch experiments were performed in 500 mL vented Erlenmeyer flasks (Corning Inc.) with a working volume of 100 mL and a seeding density of 2 x 10^5 cell·mL^{-1}. The fed-batch cultures were harvested at 12 days post-seeding or when viability dropped below 60%. All fed-batch cultures were supplemented with 1 μM of manganese(II) chloride ($MnCl_2$) at seeding and 10% v/v CD EfficientFeed™ C AGT™ Nutrient supplement (Life Technologies, Paisley, UK) on days 2, 4, 6, 8 and 10 post-seeding. Manganese (Mn) is a well-known glycosyltransferase co-factor that is critical for b4GalT activity[508], the enzyme catalysing protein galactosylation. Its addition to the cell culture has been found to improve galactosylation rates[293]. Methods used for quantifying viable cell density, viability, metabolites concentration, IgG titre, NSD intracellular concentration and IgG glycoprofile can be found in Jedrzejewski[311].

Table 2.1: Concentrations of galactose and uridine in each experiment that was utilized for model training and validation. The (-) symbol indicates that feeding was not considered on that particular day.

Experiment	Galactose in Feed C (mM)				Uridine in Feed C (mM)			
Time point	**Day 4**	**Day 6**	**Day 8**	**Day 10**	**Day 4**	**Day 6**	**Day 8**	**Day 10**
Control	0	-	0	-	0	-	0	-
10G	100	-	100	-	0	-	0	-
10G5U	100	-	100	-	50	-	50	-
10G20U	100	-	100	-	200	-	200	-
50G5U	500	-	500	-	50	-	50	-
Optimization	65	93	90	87	8	13	28	100

2.1.2 Mathematical modelling

The modelling framework comprises three modules as shown in Fig. 2.1:

1. *Metabolism submodel*: A dynamic unstructured cell growth, death, antibody synthesis and metabolism model, based on Monod-type equations. The aim of the first module is to estimate the specific cell growth rate and the specific cell productivity by utilizing only the feeding schedule of the fed-batch experiments, the metabolite concentrations in the feed and the concentration of the metabolites in the seeding medium. The model also describes the extracellular profile of the examined metabolites and viable cell density through culturing time.
2. *NSD submodel*: A dynamic kinetic model describing NSD synthesis in the intracellular environment and based on Michaelis-Menten kinetics. Inputs to the *NSD submodel* are the specific cell growth and antibody production rates, as calculated by the *Metabolism submodel*. The goal of the module is to estimate the dynamic concentration profile of NSDs in the intracellular environment and the fluxes of the NSDs towards the Golgi apparatus.
3. *Glycomodel*: A dynamic *N*-linked glycosylation model describing the synthesis and distribution of glycoforms in the Golgi apparatus and the secretion of glycosylated mAb in the extracellular environment. The model was taken from del Val et al.[478] The model portrays the glycosylation process in the Golgi as a PFR reactor. Input to the *Glycomodel* is the intracellular NSDs concentration and the specific antibody production rate.

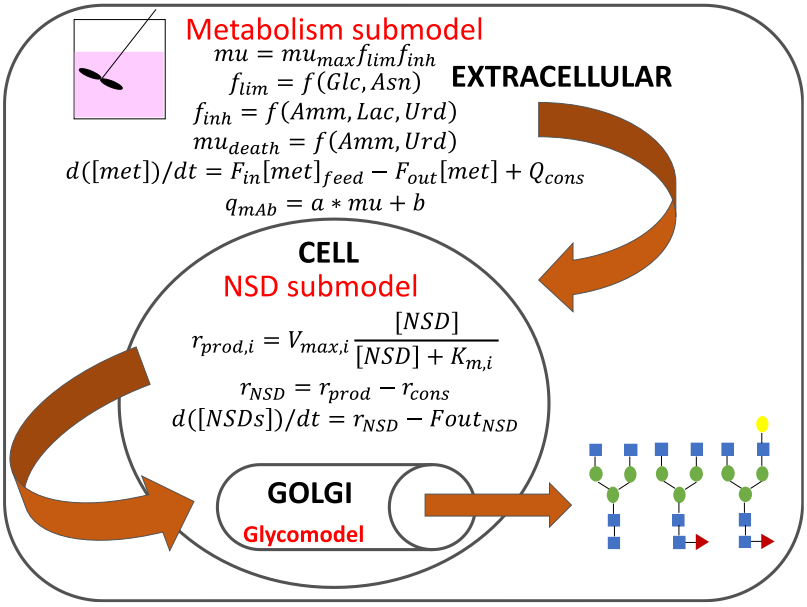

Figure 2.1: Model framework and configuration, consisting of three submodels: 1) *Metabolism submodel* estimating metabolites extracellular concentration, specific growth rate and specific antibody production rate, 2) *NSD submodel* estimating the intracellular concentration of NSDs and the NSD flux towards the Golgi apparatus and 3) *Glycomodel* estimating the glycoform distribution of the secreted mAb.

Model development and parameterization was performed in gPROMS ModelBuilder v.5.0.1 (Process Systems Enterprise, gPROMS, www.psenterprise.com/products/gproms, 1997-2020). The Maximum Likelihood formulation in gPROMS was used for parameter estimation. The Maximum Likelihood solver identifies the physical (model) and variance parameter values that would maximize the probability of correctly predicting the experimental data, assuming normally distributed measurement errors and standard deviations. The objective function of the formulation is shown in Eq. 2.1:

$$\Phi = \frac{N}{2}\ln(2\pi) + \frac{1}{2}\min_{\theta}\left(\sum_{i=1}^{NE}\sum_{j=1}^{NV_i}\sum_{k=1}^{NM_{ij}}\left[\ln\left(\sigma_{ijk}^2\right) + \frac{\left(\tilde{z}_{ijk} - z_{ijk}\right)^2}{\sigma_{ijk}^2}\right]\right) \qquad (2.1)$$

where:

- Φ is the Maximum Likelihood solver objective function
- θ are the physical parameters of the model subject to user-defined lower (θ_L) and upper (θ_U) bounds ($\theta_L < \theta < \theta_U$)
- N is the total number of experimental points
- NE is the number of experiments considered in the parameter estimation
- NV_i is the total number of variables measured in the i^{th} experiment
- NM_{ij} is the total number of measurements for the j^{th} variable in the i^{th} experiment
- σ_{ijk}^2 is the user-defined variance of the k^{th} measurement for the j^{th} variable in the i^{th} experiment

- $\tilde{z}_{ijk}$ is the k^{th} experimentally measured value of the j^{th} variable in the i^{th} experiment
- z_{ijk} is the k^{th} model predicted value of the j^{th} variable in the i^{th} experiment

2.1.3 Statistical analysis of model results

The MedCalc Statistical Software v.18.2.1 (MedCalc Software bvba, Ostend, Belgium) was used for the calculation of the Concordance Correlation Coefficient (CCC) and the OriginPro 2017 software (OriginLab, Northampton, MA) for the estimation of the Pearson Correlation Coefficient (PCC). A 95% confidence interval was set for the calculation of both the CCC and the PCC. The PCC was used to assess the correlation between the model predicted and the experimentally measured variables. CCC was used in addition to the coefficient of determination (R^2) to evaluate the agreement between the experimental data and model results. The CCC, PCC and R^2 values were used to evaluate the fitness of the model for viable cell density and extracellular monoclonal antibody concentration for the *Metabolism submodel* and for the intracellular concentrations of UDPGal and UDPGlcNAc for the *NSD submodel*. The aforementioned variables were chosen as representatives of the performance of the respective submodels for this set of experiments.

2.2 Model construction

Sections 2.2.1, 2.2.2 and 2.2.3 present the development of the three modules that consist the proposed modelling framework. The material balances and reaction rates expressions in the *Metabolism* and *NSD submodels* are based on the reduced networks presented in each section. Finally, section 2.2.4 presents the parameterization strategies followed for all three modules.

2.2.1 Metabolism submodel

The *Metabolism submodel* was adapted from previous work on fed-batch hybridoma cell cultures[506] and consists of three discrete but inter-dependent modules: 1) the CHO cell growth and death module, 2) the CHO cell metabolism module and 3) the mAb synthesis module. The *Metabolism submodel* is a reduced empirical model of CHO cell metabolism (Fig. 2.2) and considers the effect of the following metabolites and amino acids on the specific cell growth and antibody production rates: glucose, galactose, uridine, glutamine, lactate, ammonia, glutamate, asparagine and aspartate.

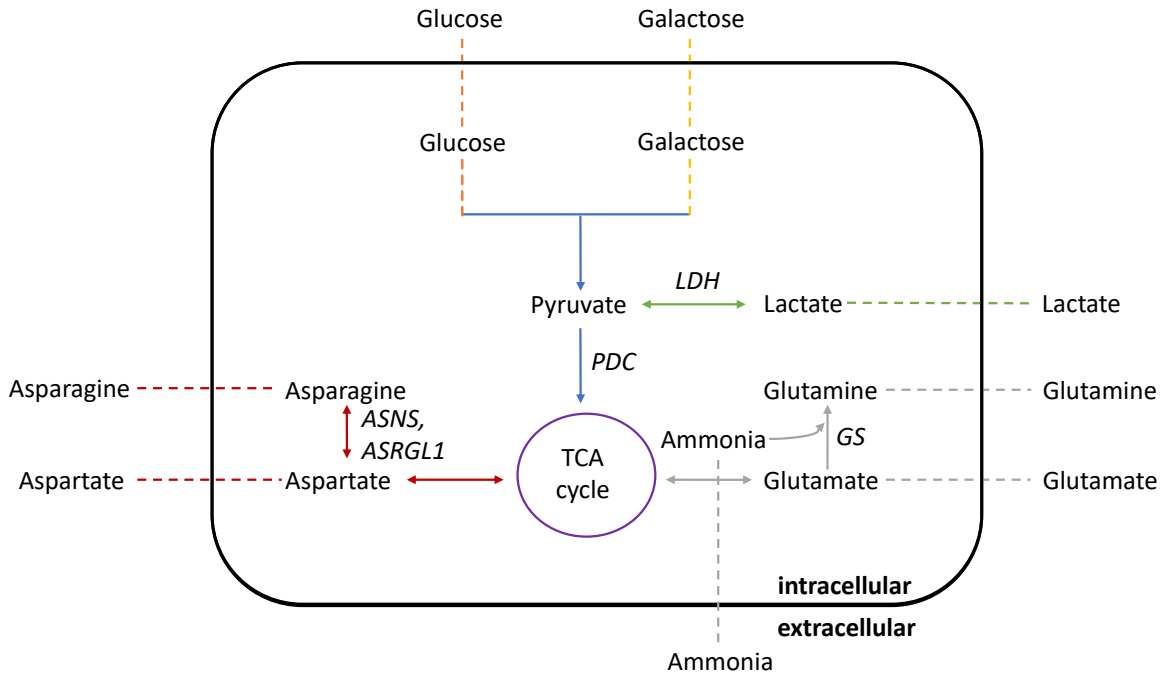

Figure 2.2: Reduced metabolic network considered in the *Metabolism submodel*. The enzymes catalysing the non-lumped reactions are also shown. The dashed lines show transport reactions. The acronyms used for the enzymes are: PDC for pyruvate dehydrogenase complex, LDH for lactate dehydrogenase, GS for glutamine synthetase, ASNS for asparagine synthetase and ASRGL1 for L-asparaginase.

CHO cell growth and death

The CHO cell growth and death module was based on Monod-type equations that were defined by the limiting substrates and the inhibiting metabolites and products of the culture. The material balances of the cell culture volume (Eq. 2.2) and the viable cell density (Eq. 2.3) are:

$$\frac{dV}{dt} = F_{in} - F_{out} \quad (2.2)$$

where, V (L) is the cell culture volume, t (h) is the culturing time, F_{in} (L·h^{-1}) is the flow rate of the feeding and F_{out} (L·h^{-1}) is the sampling flow rate.

$$\frac{d(VX_v)}{dt} = (\mu - \mu_{death})VX_v - F_{out}X_v \quad (2.3)$$

where, X_v (cell·L^{-1}) is the viable cell density and μ (h^{-1}) and μ_{death} (h^{-1}) indicate the specific cell growth and death rate, respectively.

Glucose and asparagine were identified as the limiting substrates of the cell culture, while ammonia, lactate and uridine were considered as the inhibiting substrates. More specifically, the specific cell growth rate is defined by Eq. 2.4, 2.5 and 2.6:

$$\mu = \mu_{max} f_{\lim} f_{inh} \quad (2.4)$$

where, μ_{max} (h^{-1}) is the maximum specific cell growth rate, f_{lim} (-) is the substrate limiting factor and f_{inh} (-) is the metabolite and products inhibiting factor.

$$f_{\mathrm{lim}} = \frac{[Glc]}{[Glc]+K_{Glc}} \frac{[Asn]}{[Asn]+K_{Asn}} \quad (2.5)$$

where, $[Glc]$ (mM) and $[Asn]$ (mM) are the extracellular concentrations of glucose and asparagine, respectively. K_{Glc} (mM) and K_{Asn} (mM) are the Monod constants for glucose and asparagine, respectively.

$$f_{\mathrm{inh}} = \frac{KI_{Amm}}{[Amm]+KI_{Amm}} \frac{KI_{Lac}}{[Lac]+KI_{Lac}} \frac{KI_{Urd}}{[Urd]+KI_{Urd}} \quad (2.6)$$

where, $[Amm]$ (mM), $[Lac]$ (mM) and $[Urd]$ (mM) are the extracellular concentrations of ammonia, lactate and uridine respectively. KI_{Amm} (mM), KI_{Lac} (mM), KI_{Urd} (mM) are the inhibiting constants for ammonia, lactate and uridine, respectively.

In a similar manner, ammonia and uridine were used for defining the specific cell death rate (Eq. 2.7)

$$\mu_{death} = \mu_{death,max} \left(\frac{[Amm]}{[Amm]+K_{d,Amm}} + \frac{[Urd]}{[Urd]+K_{d,Urd}} \right) \quad (2.7)$$

where, $\mu_{death,max}$ (h^{-1}) is the maximum specific cell death rate and $K_{d,Amm}$ (mM) and $K_{d,Urd}$ (mM) are the ammonia and uridine Monod constants for cell death, respectively.

CHO cell metabolism

CHO cell metabolism was described using the material balances of the metabolites and amino acids considered in the model. The general material balance is shown in Eq. 2.8:

$$\frac{d(V[Metabolite])}{dt} = F_{in}\left[Metabolite_{feed}\right] - F_{out}[Metabolite] + q_{metabolite} V X_v \quad (2.8)$$

where, $[Metabolite]$ (mM) is the extracellular concentration of the considered metabolite or amino acid, $\left[Metabolite_{feed}\right]$ (mM) is the concentration of the considered metabolite or amino acid in the feed and $q_{metabolite}$ (mmol·cell^{-1}·h^{-1}) is the specific production or consumption rate of the considered metabolite or amino acid. The $q_{metabolite}$ is based on the intracellular metabolic links of the metabolites as further discussed below. The sign convention used for the $q_{metabolite}$ is positive (+) for the production and negative (-) for the consumption rates.

The specific production/consumption rates are separately examined for each metabolite and amino acid. $Y_{X_{Metabolite}}$ (cell·mmol^{-1}) that describes the yield of cell biomass for each metabolite or amino acid, is assumed to remain constant during the cell culture to avoid overparameterization.

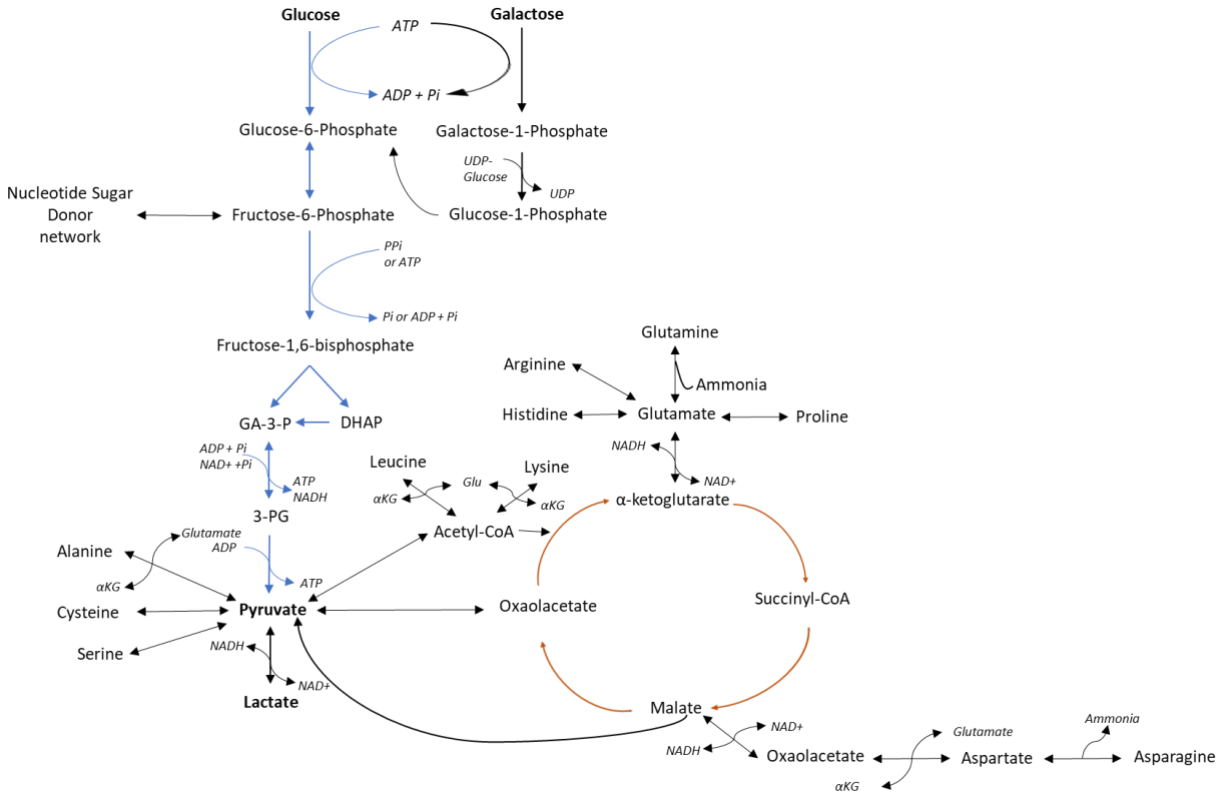

Figure 2.3: A reduced metabolic network that integrates various pathways and reactions like glycolysis and the TCA cycle. The proposed metabolic network accounts for only a few of the total reactions and aims to depict the fluxing relations between the metabolites and nutrients/amino acids. The blue lines indicate the reactions involved in glycolysis pathway, the orange lines indicate the reactions involved in the TCA cycle and the black lines indicate other reactions that interact with the previous pathways.

Glucose

Glucose consumption initiates the glycolysis pathway (Fig. 2.3) for the synthesis of cytosolic and subsequently mitochondrial pyruvate. However, the presence of galactose in the culture can lessen the flux of glucose towards the glycolytic pathway, as both glucose and galactose lead to the production of the glucose-6-phospate intermediate (Fig. 2.3). Eq. 2.9 and 2.10 attempt to describe this inter-dependency between the two metabolites:

$$q_{Glc} = \left(-\frac{\mu}{Y_{X_{Glc}}} - m_{Glc}\right) \cdot \left(\frac{Kc_{Gal}}{Kc_{Gal}+[Gal]}\right)^{n_{Gal}} \quad (2.9)$$

$$n_{Gal} = 1 - f_{Gal}\frac{q_{Gal}}{q_{Glc}} \quad (2.10)$$

where, q_{Glc} (mmol·cell^{-1}·h^{-1}) is the specific uptake rate of glucose, $Y_{X_{Glc}}$ (cell·mmol^{-1}) is the yield of cell biomass on glucose and m_{Glc} (mmol·cell^{-1}·h^{-1}) is the maintenance coefficient of glucose for its consumption towards other metabolic pathways of the cell. The $\frac{Kc_{Gal}}{Kc_{Gal}+[Gal]}$ term is used as a regulating factor that is only active when galactose is fed in the culture. The factor is equal to 1 when

no galactose is present and <1 when galactose is supplemented in the culturing medium. Therefore, when galactose is present the q_{Glc} is reduced due to the existence of the $\frac{Kc_{Gal}}{Kc_{Gal}+[Gal]}$ term, accounting that way for the effect of galactose presence on glucose consumption towards glycolysis. Regarding the regulating term: Kc_{Gal} (mM) is the regulating concentration of galactose, $[Gal]$ (mM) is the extracellular concentration of galactose and q_{Gal} (mmol·cell^{-1}·h^{-1}) is the specific uptake rate of galactose.

Finally, n_{Gal} (-) and f_{Gal} (-) are factors that participate in further regulating the specific consumption rate of glucose. Eq. 2.10 accounts for the balance between the specific consumption rates of glucose and galactose towards glycolysis. Moreover, the majority of mammalian glucose transporters (GLUTs) show affinity to both glucose and galactose and therefore a regulation of the respective uptake rates is possible[509-510].

Glutamine

$$q_{Gln} = \frac{\mu}{Y_{X_{Gln}}} + q_{Amm} Y_{Gln/Amm} \quad (2.11)$$

where, q_{Gln} (mmol·cell^{-1}·h^{-1}) is the specific production rate of glutamine, $Y_{X_{Gln}}$ (cell·mmol^{-1}) is the yield of cell biomass on glutamine, q_{Amm} (mmol·cell^{-1}·h^{-1}) is the specific production rate of ammonia and $Y_{Gln/Amm}$ (mmol$_{Gln}$·mmol$_{Amm}^{-1}$) is the yield of glutamine from ammonia due to the activity of the glutamine synthetase enzyme (Fig. 2.2).

Lactate

$$q_{Lac} = \left(\frac{\mu}{Y_{X_{Lac}}} - Y_{Lac/Glc} q_{Glc}\right) \frac{(Lac_{max1} - [Lac])}{Lac_{max1}} + m_{lac} \frac{Lac_{max2} - [Lac]}{Lac_{max2}} \quad (2.12)$$

where, q_{Lac} (mmol·cell^{-1}·h^{-1}) is the specific production/consumption rate of lactate, $Y_{X_{Lac}}$ (cell·mmol^{-1}) is the yield of cell biomass on lactate, $Y_{Lac/Glc}$ (mmol$_{Lac}$·mmol$_{Glc}^{-1}$) is the yield of lactate from glucose consumption towards the glycolysis pathway and through the synthesis of pyruvate, Lac_{max1} (mM) and Lac_{max2} (mM) are kinetic constants for the activation of lactate consumption during the stationary phase of the culture[270] and m_{lac} (mmol·cell^{-1}·h^{-1}) is the maintenance coefficient for the participation of lactate in other metabolic pathways of the cell.

Ammonia

$$q_{Amm} = \frac{\mu}{Y_{X_{Amm}}} - Y_{Amm/Urd}\, q_{Urd} \quad (2.13)$$

where, $Y_{X_{Amm}}$ (cell·mmol^{-1}) is the yield of cell biomass on ammonia, $Y_{Amm/Urd}$ (mmol$_{Amm}$·mmol$_{Urd}^{-1}$) is the yield of ammonia from uridine and q_{Urd} (mmol·cell^{-1}·h^{-1}) is the specific consumption rate of

uridine. In Eq. 2.13 the term $Y_{Amm/Urd}\ q_{Urd}$ is deduced from the $\frac{\mu}{Y_{X_{Amm}}}$ as q_{Urd} is a consumption rate and has a negative value. Thus, the $-Y_{Amm/Urd}\ q_{Urd}$ term is always positive, and Eq. 2.13 is practically a summation of the two terms.

Glutamate

$$q_{Glu} = -\frac{\mu}{Y_{X_{Glu}}} \quad (2.14)$$

where, q_{Glu} (mmol·cell^{-1}·h^{-1}) is the specific uptake rate of glutamate and $Y_{X_{Glu}}$ (cell·mmol^{-1}) is the yield of cell biomass on glutamate.

Galactose

$$q_{Gal} = -\frac{\mu}{Y_{X_{Gal}}}\frac{[Gal]}{[Gal]+K_{Gal}} \quad (2.15)$$

where, $Y_{X_{Gal}}$ (cell·mmol^{-1}) is the yield of cell biomass on galactose and K_{Gal} (mM) is a Monod constant for galactose consumption. The $\frac{[Gal]}{[Gal]+K_{Gal}}$ functions as a regulating factor for the consumption of galactose based on the availability of the metabolite in the extracellular environment.

Uridine

$$q_{Urd} = \frac{\mu}{Y_{X_{Urd}}}\frac{[Urd]}{[Urd]+K_{Urd}} \quad (2.16)$$

where, $Y_{X_{Urd}}$ (cell·mmol^{-1}) is the yield of cell biomass on uridine and K_{Urd} (mM) is a Monod constant for uridine consumption. Similarly to Eq. 2.15, the $\frac{[Urd]}{[Urd]+K_{Urd}}$ term is a regulating factor for the uptake of uridine based on its extracellular concentration.

Asparagine

$$q_{Asn} = -\frac{\mu}{Y_{X_{Asn}}} - Y_{Asn/Asp}q_{Asp} \quad (2.17)$$

where, q_{Asn} (mmol·cell^{-1}·h^{-1}) is the specific uptake rate of asparagine, $Y_{X_{Asn}}$ (cell·mmol^{-1}) is the yield of cell biomass on asparagine, $Y_{Asn/Asp}$ (mmol$_{Asn}$·mmol$_{Asp}^{-1}$) is the yield of asparagine from aspartate and q_{Asp} (mmol·cell^{-1}·h^{-1}) is the specific uptake rate of aspartate. Asparagine is synthesized from aspartate through the activity of the asparagine synthetase enzyme, encoded by the ASNS gene.

Aspartate

$$q_{Asp} = -\frac{\mu}{Y_{X_{Asp}}} - Y_{Asp/Asn}q_{Asn} \quad (2.18)$$

where, $Y_{X_{Asp}}$ (cell·mmol^{-1}) is the yield of cell biomass on aspartate and $Y_{Asp/Asn}$ (mmol$_{Asp}$·mmol$_{Asn}^{-1}$) is the yield of aspartate from asparagine.

mAb synthesis

The module of mAb synthesis includes the material balance of the antibody concentration and the respective specific antibody production rate:

$$\frac{d(V[mAb])}{dt} = -F_{out}[mAb] + q_{mAb}VX_v \quad (2.19)$$

$$q_{mAb} = Y_{mAb_X}\mu + m_{mAb} \quad (2.20)$$

where, $[mAb]$ (mg·L^{-1}) is the concentration of the secreted mAb in the extracellular environment, q_{mAb} (mg·cell^{-1}·h^{-1}) is the specific antibody production rate, Y_{mAb_X} (mg·cell^{-1}) is the yield of mAb from cell growth and m_{mAb} (mg·cell^{-1}·h^{-1}) is a non-growth associated term. As shown in the Eq. 2.20, q_{mAb} was considered to be linearly dependent on the specific cell growth rate. Specific production and cell growth rates were assumed linearly dependent as the PCC between q_{mAb} - μ, that describes the linear correlation between two variables, for the experimental data of the control experiment was found equal to 0.832.

2.2.2 NSD submodel

The *NSD submodel* describes the intracellular metabolism that leads to the synthesis of the NSDs, that act as co-substrates in *N*-linked glycosylation. The module includes 7 material balances and 16 reactions. The dependencies between the *NSD* and the *Metabolic submodels* are the specific cell growth rate (Eq. 2.4), the specific antibody production rate (Eq. 2.20) and the extracellular concentrations of glucose, galactose, glutamine and uridine. The NSD synthesis and consumption reaction rates are based on simple Monod type saturation kinetics. The nucleotides are assumed to be in excess based on our experimental data (Appendix – Fig. A1) and are therefore not included in the reaction rates. With the exception of glutamine, the intracellular concentrations of the remaining metabolites and amino acids were not calculated in order to avoid unnecessary overparameterization and oversimplifications. More specifically, the intracellular concentration of glutamine was calculated based on Eq. 2.21 assuming linear dependency between the intra- and extracellular concentrations of the amino acid:

$$[Gln_{intra}] = f_{Gln}[Gln] \quad (2.21)$$

where, $[Gln_{intra}]$ (mM) is the intracellular concentration of glutamine and f_{Gln} (-) is the secretion factor of glutamine to the extracellular environment.

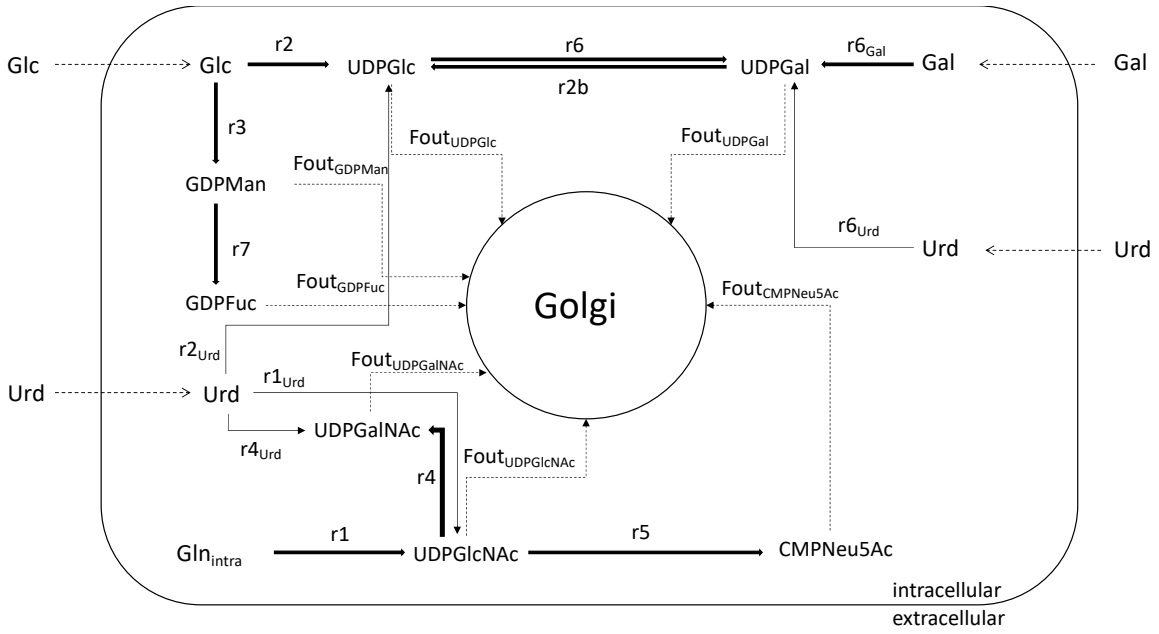

Figure 2.4: Simplified NSD metabolic pathway considered in the *NSD submodel*. The reactions are indicated by solid lines and the transport rates by dotted lines. Taken with permission from Kotidis et al.[482]

Whilst a detailed network of the NSD synthesis pathway has been reported for mouse hybridoma HNF7.1 cells[121], a reduced version of the network is considered in this study, as shown in Fig. 2.4. The material balances describe the inter-dependencies between the NSD species as shown in Fig. 2.4, and the rates are based on the reported mechanisms of the enzymes[511-512].

Eq. 2.22 to 2.32 describe the rates of the reactions shown in Fig. 2.4:

$$r_1 = V_{max1}\frac{[Gln_{intra}]}{K_{M1_{Gln}}+[Gln_{intra}]} \quad (2.22)$$

$$r_{1_{sink}} = V_{max1_{sink}}\frac{[UDPGlcNAc]}{\left(K_{M1_{sink}}+[UDPGlcNAc]\right)\left(1+\frac{[CMPNeu5Ac]}{KI1_{sink}}\right)} \quad (2.23)$$

$$r_2 = V_{max2}\frac{[Glc]}{K_{M2_{Glc}}+[Glc]} \quad (2.24)$$

$$r_{2b} = V_{max2b}\frac{[UDPGal]}{K_{M2b_{UDPGal}}\left(1+\frac{[UDPGlcNAc]}{KI_{2A}}+\frac{[UDPGalNAc]}{KI_{2B}}+\frac{[UDPGlc]}{KI_{2C}}+\frac{[UDPGal]}{KI_{2D}}\right)+[UDPGal]} \quad (2.25)$$

$$r_3 = V_{max3}\frac{[Glc]}{K_{M3_{Glc}}+[Glc]} \quad (2.26)$$

$$r_4 = V_{max4}\frac{[UDPGlcNAc]}{K_{M4_{UDPGlcNAc}}+[UDPGlcNAc]} \quad (2.27)$$

$$r_5 = V_{max5}\frac{[UDPGlcNAc]}{K_{M5_{UDPGlcNAc}}\left(1+\frac{[CMPNeu5Ac]}{KI_5}\right)+[UDPGlcNAc]} \quad (2.28)$$

$$r_6 = V_{max6}\frac{[UDPGlc]}{K_{M6_{UDPGlc}}\left(1+\frac{[UDPGlcNAc]}{KI_{6A}}+\frac{[UDPGalNAc]}{KI_{6B}}+\frac{[UDPGal]}{KI_{6C}}\right)+[UDPGlc]} \quad (2.29)$$

$$r_{6_{sink}} = V_{max6_{sink}}\frac{[UDPGal]}{K_{6_{sink}}\left(1+\frac{[UDPGlc]}{KI_{6_{sink}}}\right)+[UDPGal]}\frac{[Gal]}{[Gal]+K_{regulator}} \quad (2.30)$$

$$r_7 = V_{max7}\frac{[GDPMan]}{(K_{M7_{GDPMan}}+[GDPMan])\left(1+\frac{[GDPFuc]}{KI_7}\right)} \quad (2.31)$$

$$r_{7_{sink}} = V_{max7_{sink}}\frac{[GDPFuc]}{K_{M7_{sink}}+[GDPFuc]} \quad (2.32)$$

where, V_{max}, i (mmol$_{NSD}$·L$_{cell}^{-1}$·h^{-1}) is the maximum turnover rate of reaction i. Moreover, $K_{Mi_{NSD}}$ (mM) is the saturation constant of the examined NSD in reaction i, KI_i (mM) is the inhibition constant of reaction i and $K_{regulator}$ (mM) is a fixed parameter for the activation of the $\frac{[Gal]}{[Gal]+K_{regulator}}$ regulation term when galactose is present.

Moreover, the $r_{i_{sink}}$ terms (not shown in Fig. 2.4) demonstrate the fluxes towards pathways that were not considered in the designated reduced network and were found to affect the cytosolic levels of UDPGlcNAc ($r_{1_{sink}}$), UDPGal ($r_{6_{sink}}$) and GDPFuc ($r_{7_{sink}}$). In addition to the reaction rates described in Eq. 2.22 to 2.32, the contribution of galactose and uridine towards the NSDs concentration was quantified with separate reaction rates as shown in Eq. 2.33 and 2.34:

$$r_{i_{Urd}} = V_{max,i_{Urd}}\frac{[Urd]}{K_{M,i_{Urd}}+[Urd]} \quad (2.33)$$

$$r_{6_{Gal}} = V_{max6_{Gal}}\frac{[Gal]}{K_{M6_{Gal}}\left(1+\frac{[UDPGal]}{KI_{6D}}+\frac{[Gal]}{KI_{6E}}+\frac{[Urd]}{KI_{6F}}\right)+[Gal]} \quad (2.34)$$

where, $r_{i_{Urd}}$ (mmol·L$_{cell}^{-1}$·h^{-1}) is the rate of reaction i due to uridine addition, $V_{max,i_{Urd}}$(mmol·L$_{cell}^{-1}$·h^{-1}) is the maximum turnover rate of reaction i and $K_{M,i_{Urd}}$ (mM) are the saturation constants of reaction i. As shown in Fig. 2.4, an effect of uridine addition on the intracellular levels of the NSDs is observed, the synthesis of which requires UTP as a co-substrate: UDPGlcNAc ($r_{1_{Urd}}$), UDPGlc ($r_{2_{Urd}}$), UDPGalNAc ($r_{4_{Urd}}$) and UDPGal ($r_{6_{Urd}}$). In a similar manner, galactose addition enhances the synthesis of UDPGal, as shown in Eq. 2.34 where, $r_{6_{Gal}}$ (mmol·L$_{cell}^{-1}$·h^{-1}) is the reaction rate, $V_{max6_{Gal}}$ (mmol·L$_{cell}^{-1}$·h^{-1}) is the maximum turnover rate of $r_{6_{Gal}}$, $K_{M6_{Gal}}$ (mM) is a saturation constant and KI_{6D} (mM), KI_{6E} (mM) and KI_{6F} (mM) are the inhibition constants of UDPGal, galactose and uridine, respectively.

The material balance of each NSD is based on Eq. 2.35:

$$\frac{d([NSD])}{dt} = \sum_i v_i r_i - F_{out_{NSD}} \quad (2.35)$$

where, $[NSD]$ (mM) are the intracellular NSD concentrations for UDPGal, UDPGlc, UDPGalNAc, UDPGlcNAc, GDPFuc, GDPMan and CMPNeu5Ac. Moreover, v_i (-) is the stoichiometric coefficient for reaction i for the examined NSD that is equal to -1 if the NSD is consumed, equal to 0 if the NSD is not participating in the reaction and equal to +1 if the NSD is a product of the reaction. $F_{out_{NSD}}$ ($mmol_{NSD} \cdot L^{-1} \cdot h^{-1}$) represents the flux of each cytosolic NSD towards the Golgi apparatus where *N*-linked glycosylation occurs. The flux describes the transport rate of the NSDs that is necessary for the glycosylation of host cell proteins (HCPs), glycolipids and the recombinant IgG, as shown in Eq. 2.36:

$$F_{out_{NSD}} = \frac{[NSD]}{K_{TP_{NSD}}+[NSD]}\left(\frac{N_{HCP/Lipids_{NSD}}\mu}{V_{cell}} + \frac{N_{precursor_{NSD}}q_{mAb}}{V_{cell}} + r_{NSD_k}^{mAb,glyc}\right) \quad (2.36)$$

where, $K_{TP_{NSD}}$ (mM) is a saturation constant for the examined NSD transport and V_{cell} (L) is the cellular volume. The first term of the summation of Eq. 2.36, represents the flux of the examined NSD towards the *O*- and *N*-linked glycosylation of HCPs and the synthesis of glycolipids. More specifically, $N_{HCP/Lipids_{NSD}}$ ($mmol_{NSD} \cdot cell^{-1}$) is the stoichiometric coefficient of NSD consumption for HCPs and lipids glycosylation, as described in del Val et al.[513] The second term describes the demands of each NSD for the synthesis of the oligosaccharide precursor molecule ($Glc_3Man_9GlcNAc_2$) for the antibody glycosylation, through the $N_{precursor_{NSD}}$ ($mmol_{NSD} \cdot cell^{-1}$) stoichiometric coefficient. Finally, $r_{NSD_k}^{mAb,glyc}$ ($mmol_{NSD} \cdot L^{-1} \cdot h^{-1}$) is the uptake rate of the k^{th} NSD towards mAb glycosylation and is calculated through Eq. 2.37:

$$r_{NSD_k}^{mAb,glyc} = \left(\frac{V_{Golgi}}{V_{cell}}\right)\int_{z=0}^{z=1}\sum_{j=1}^{N.R.} v_{k,j} \cdot r_{G,j}(z)\, dz \quad (2.37)$$

where, V_{Golgi} (L) is the volume of the Golgi apparatus, $v_{k,j}$ ($mmol_{NSD} \cdot mmol_{glycan}^{-1}$) is the stoichiometric coefficient of the k^{th} NSD in the j^{th} glycosylation reaction, $r_{G,j}(z)$ ($mmol_{glycan} \cdot L^{-1} \cdot h^{-1}$) is the rate of the j^{th} glycosylation reaction along the normalized length of the Golgi apparatus (z) that is calculated in the *Glycomodel*.

However, the inclusion of Eq. 2.37 in the calculation of $F_{out_{NSD}}$ requires the convergence between the $F_{out_{NSD}}$ and the integral of Eq. 2.37 over the normalized Golgi length, and therefore substantially increases the simulation time (~8 min). For that reason, a simplified version of Eq. 2.37 is proposed in order to reduce the simulation time (~1.5 min) while maintaining the mechanistic connection between the *NSD submodel* and the *Glycomodel*. In the simplified version, the flux of the NSDs towards mAb glycosylation is assumed to be a function of the glycans concentration in the end of the Golgi apparatus (z=1), prior to their secretion to the cytosolic and subsequently to the extracellular environment. The validity of the assumption is based on the substantially lower residence time of the

glycoprotein in the Golgi apparatus (22 min)[472, 478], compared to the time scale within which changes in the NSD concentration are experimentally observed. The simplified function is shown in Eq. 2.38:

$$r_{NSD_k}^{mAb,glyc} = Vel_{Golgi}\left(\frac{V_{Golgi}}{V_{cell}}\right)\sum_{i=1}^{N.OS}\nu_{k,i}\cdot[OS_i](1) \quad (2.38)$$

where, Vel_{Golgi} (Golgi length·min^{-1}) is the linear velocity with which the glycoprotein molecules travel through the Golgi apparatus, $\nu_{k,i}$ (mmol$_{NSD}$·mmol$_{glycan}^{-1}$) is the stoichiometric coefficient of the k^{th} NSD required for the synthesis of the i^{th} oligosaccharide and $[OS_i](1)$ (mmol$_{glycan}$ · L^{-1}) is the concentration of the i^{th} oligosaccharide in the end of the Golgi apparatus (z=1), as calculated by the *Glycomodel*. Finally, $N.OS$ (-) is the number of oligosaccharides considered in the model, which in the current model is 77.

2.2.3 Glycomodel

The *Glycomodel*, adapted from del Val et al.[478], assumes that protein trafficking in the Golgi apparatus occurs through the cisternal maturation model. The reaction network of the model includes 95 reactions and 77 oligosaccharides. Therefore, the Golgi apparatus is represented as a single PFR, where the concentration profile of mannosidases and glycosyltransferases follows a Gaussian distribution. The NSD concentrations, as calculated in the *NSD submodel*, and the specific antibody production rate, as calculated in the *Metabolism submodel*, are the inputs in the *Glycomodel* that is employed for the estimation of the glycans distribution in the secreted mAb. More specifically, the *Glycomodel* assumes that the intra-Golgi concentration of the NSDs is 20 times greater than the concentration in the cytosol[473, 514]. The volume of the Golgi apparatus is assumed to remain constant during the cell culturing process and the linear velocity of protein trafficking within the Golgi apparatus is considered a function of the specific antibody production rate[486, 515]. Eq. 2.39 presents the material balance of the oligosaccharides in the Golgi, defined by partial differential equations due to the nature of the PFR model:

$$\frac{\partial[OS_i](t,z)}{\partial t} = -Vel_{Golgi}\frac{\partial[OS_i](t,z)}{\partial z} - \sum_{i=1}^{N.OS}\nu_{i,j}\cdot r_{G,j} \quad (2.39)$$

where, $\nu_{i,j}$ (-) is the stoichiometric coefficient of the i^{th} oligosaccharide in the j^{th} reaction of glycosylation. The variable linear velocity of the Golgi is further defined by Eq. 2.40:

$$Vel_{Golgi} = \frac{q_{mAb}\,Glycosites}{MW_{mAb}\,V_{Golgi}\,[OS_1^{z=0}]} \quad (2.40)$$

where, MW_{mAb} (Da) is the molecular weight of the IgG that is assumed to be equal to 150kDa, $Glycosites$ (μmol$_{glycan}$·μmol$_{mAb}$ $^{-1}$) is the number of *N*-linked glycosylation sites available in the glycoprotein, which for the examined IgG is equal to 2, and $[OS_1^{z=0}]$ (μM) is the concentration of the initial glycoform (Man9) in the entering point of the Golgi apparatus, after transferring from the ER.

The $[OS_1^{z=0}]$ is assumed to remain constant throughout the culturing process and for its calculation the Eq. 2.40 was appropriately rearranged. For $[OS_1^{z=0}]$ calculation, q_{mAb} was set equal to the highest experimentally observed value and Vel_{Golgi} was set equal to 0.045 Golgi length·min^{-1}, calculated from the reverse average residence time in the Golgi (22 min), as mentioned earlier. Therefore, $[OS_1^{z=0}]$ was calculated as 94 μM. The distribution of the enzymes in the Golgi apparatus is calculated from Eq. 2.41:

$$[Enz_k](z) = Enz_k^{max} e^{-\frac{1}{2}\left(\frac{z - z_k^{max}}{w_k}\right)^2} \quad (2.41)$$

where, $[Enz_k](z)$ (μM) is the concentration of the k^{th} enzyme in the Golgi apparatus, Enz_k^{max} (μM) is the maximum concentration of the k^{th} enzyme in the Golgi apparatus, z_k^{max} (Golgi length) is the normalized Golgi length at which the Enz_k^{max} is observed and w_k (Golgi length) is the width of the Gaussian distribution of the k^{th} enzyme.

Eq. 2.42 is an example of the reaction rates kinetics (sequential-order Bi-Bi) used in the *Glycomodel* as reported in del Val et al.[478]

$$r_{G,j}(t,z) = \frac{k_{f,z}\,[Enz_z]\,[NSD_{G,k}][OS_i]}{K_{d,i}K_{d,k}\left(1+\frac{[OS_i][NSD_{G,k}]}{K_{d,i}\,K_{d,k}}+\frac{[NSD_{G,k}]}{K_{d,k}}+\frac{[NSD_{G,k}]}{K_{d,k}}\sum_{n=1}^{n=NS}\frac{[OS_n]}{K_{d,n}}\right)} \quad (2.42)$$

where, $k_{f,z}$ (min^{-1}) is the rate-limiting turnover rate of the reaction as defined by the z^{th} enzyme involved in the j^{th} reaction, $[NSD_{G,j}]$ (μM) is the intra-Golgi concentration of the k^{th} NSD utilized in the j^{th} reaction, $K_{d,i}$ (μM) is the dissociation constant of the complex between the i^{th} substrate and the enzyme, $[OS_n]$ (μM) is any plausible alternative substrate of the enzyme utilized in the j^{th} reaction, $K_{d,n}$ (μM) is the dissociation constant of the complex between the $[OS_n]$ substrate and the enzyme, $K_{d,k}$ (μM) is the dissociation constant of the complex between the k^{th} NSD and the enzyme, NS (-) is the number of substrates in the reaction network and n (-) represents the different substrates of the reaction network and can take all values apart from i.

2.2.4 Parameter estimation

Metabolism and NSD submodels

A parameter estimation (PE) process, common for both the *Metabolism* and *NSD submodels*, was utilized. The two submodels combined include 102 unknown parameters, 17 of which were assigned to fixed values according to literature (Appendix-Table A1). The remaining 85 parameters were estimated using the Parameter Estimation (now Model Validation) entity in gPROMS v.5.0.1. PE was based on the minimization of the Maximum Likelihood solver objective function, as shown in Eq. 2.1.

For that reason, the 85 parameters were estimated by fitting the model equations to the available experimental data (~900 data points in total). As a first step and in order to achieve reasonable confidence intervals for the estimated parameters, the parameters that were not involved in the galactose and uridine metabolism were estimated based solely on the control experiment. Using the estimated values for the aforementioned parameters, the parameters that were part of equations relevant to galactose and uridine metabolism were estimated based on the feeding experiments (10G, 10G5U, 10G20U and 50G5U). The variance model necessary for the PE, was calculated based on the standard deviation of each experimental measurement used in the estimation of the unknown parameters.

Regarding the reaction rates, both the maximum turnover rates (V_{max}) and the Monod saturation constants (K_m) of each reaction were estimated during PE. Therefore, the problem was structurally unidentifiable due to the linear dependence between the reaction rate and the substrate concentration, when the latter is below the saturation point. As the linear dependence only allows for an estimation of the ratio between the V_{max} and K_m and in an effort to estimate more realistic values for the parameters, the majority of the K_m values were hypothesized to not differ more than one order of magnitude from the experimentally observed range of the substrate concentration. However, the estimated values should not be regarded as an accurate representation of the physical parameters due to the aforementioned problems of structural and numerical identifiability.

Glycomodel

As a final step, 18 unknown parameters of the *Glycomodel* were estimated by adapting a strategy previously presented in del Val et al.[486] The objective function in all estimations was the minimization of the enzyme total concentration. The parameters of the enzyme distributions were estimated only based on the control experiment, while the feeding experiments were used for the estimation of the $K_{d,i}$ and $K_{d,k}$ parameters. The parameterization strategy included the following steps:

- Step 1: Calculation of the secreted mAb fraction for each glycoform and for each time interval, as shown in Eq. 2.43. Eq. 2.43 was taken from Fan et al.[289]:

$$f_i^{\Delta t} = \frac{([OS_i]_{t_{n+1}} - [OS_i]_{t_n})}{([mAb]_{t_{n+1}} - [mAb]_{t_n})} \quad (2.43)$$

- Step 2: The parameters of ManI distribution (Enz_k^{max} and w_k) were estimated by knocking-out all the downstream enzymes and assigning the distribution to the high-mannose glycans (Man9 – Man5) based on the expected experimental profile. For example and for the IgG examined in this study, no high-mannose glycans apart from Man5 were experimentally

observed in the wild type. Therefore, a knockout of all the enzymes downstream the ManI would result in a glycoprofile with 100% Man5.

- Step 3: The parameters of GnTI, GnTII, ManII and a6FucT distribution were estimated while the parameters of ManI were fixed to the estimated values from Step 2. The b4GalT enzyme remained knocked out. Similarly to step 1, the distribution of the galactosylated substrates was redirected to the respective fucosylated and non-fucosylated substrates.
- Step 4: The b4GalT distribution parameters were estimated while the parameters of the rest of the enzymes were fixed to the values calculated in Step 2 and Step 3. The estimation of the b4GalT parameters was based on achieving the experimentally observed Galactosylation Index (GI) as described in Eq. 2.44:

$$GI = \frac{0.5*\Delta[monogalactosylated\ OS]+1*\Delta[digalactosylated\ OS]}{\Delta[mAb]} \quad (2.44)$$

where, $\Delta[monogalactosylated\ OS]$ (mg·L^{-1}) and $\Delta[digalactosylated\ OS]$ (mg·L^{-1}) are the mAb glycoforms bearing one and two galactose molecules, respectively, that are secreted during the examined interval and $\Delta[mAb]$ (mg·L^{-1}) is the total secreted antibody concentration during the same interval.

- Step 5: Finally, the data points from the interval of each experiment that presented the highest q_{mAb} were employed for the estimation of the six $K_{d,i}$ and $K_{d,k}$ values. The estimated parameters included: two $K_{d,i}$ and two $K_{d,k}$ values for b4GalT and one $K_{d,i}$ parameter for a6FucT and GnTI, respectively. The aforementioned dissociation constants were chosen due to the variable levels of galactosylation and the high accumulation of UDPGal in the experimental data, the very low concentrations of Man5 and the very high fucosylation levels of the IgG.

Overall, the estimated parameters included the sets of Enz_k^{max} and w_k for each of the considered enzymes: ManI, ManII, GnTI, GnTII, a6FucT and b4GalT. Six $K_{d,i}$ and $K_{d,k}$ parameters were also estimated as described in Step 5. The remaining $K_{d,i}$ and $K_{d,k}$ parameters and the z_k^{max} values were fixed to their nominal values[478]. The glycosyltransferases: GnTIII, GnTIV, GnTV, iGnT and a3SiaT were not considered in the estimation (GnTIII and a3SiaT) or the model (GnTIV, GnTV and iGnT) as no relevant glycans were observed. For the tuning of enzymes distribution parameters, the Optimisations entity of gPROMS v.5.0.1 was utilized, while a Parameter Estimation (now Model Validation) entity was utilized for the estimation of the $K_{d,i}$ and $K_{d,k}$ parameters.

2.3 Results

This section includes the results of model fitting to the experimental data used for parameter estimation in all three modules, and also showcases the predictive capabilities of the model against an independent feeding experiment that was not used during parameter estimation. In order to enable an easier understanding of the results, the model fitting of each experiment was examined separately, and the results are sequentially presented. The conditions of the independent experiment were determined by dynamic model-based optimisation, where mAb galactosylation was set as the objective function, and the concentration of galactose and uridine on feeding days 2, 4, 6, 8 and 10 was allowed to vary (decision variables).

2.3.1 Evaluation of *Metabolism submodel* fitting

A number of selected estimated parameters for the *Metabolism submodel* are presented in Table 2.2. The rest of estimated submodel parameters can be found in the Appendix-Table A1.

Table 2.2: Selected estimated parameters for the *Metabolism submodel*. Taken with permission from Kotidis et al.[482].

Estimated parameter	Value	Units	95% Confidence Interval
μ_{max}	6.50×10^{-2}	h^{-1}	7.44×10^{-4}
$\mu_{death,max}$	1.50×10^{-2}	h^{-1}	2.56×10^{-3}
K_{Glc}	14.04	mM	5.05×10^{-1}
K_{Asn}	2.62	mM	7.18×10^{-2}
KI_{Amm}	3.17	mM	2.49×10^{-1}
KI_{Lac}	1×10^{3}	mM	N/A
KI_{Urd}	41.09	mM	2.42
$K_{d,Amm}$	14.28	mM	3.00
$K_{d,Urd}$	27.86	mM	3.71
$Y_{mAb,X}$	3.39	$(pg\cdot cell^{-1})$	3.21×10^{-1}
m_{mAb}	4.10×10^{-1}	$(pg\cdot cell^{-1}\cdot h^{-1})$	6.57×10^{-3}

Control experiment

As presented in Fig. 2.5, the model achieves a very good fitting to the measured variables. Additionally, the high CCC, PCC and R^2 values shown in Table 2.3 indicate that the simulation results are very close to the experimentally measured values.

Table 2.3: Concordance correlation coefficients, Pearson correlation coefficients and R^2 for the evaluation of the agreement between experimental measurements and model results for viable cell density and extracellular monoclonal antibody concentration. Adapted from Kotidis et al.[482].

Experiment	CCC_{Xv}	CCC_{mAb}	PCC_{Xv}	PCC_{mAb}	R^2_{Xv}	R^2_{mAb}
Control	0.986	0.998	0.990	0.999	0.992	0.999
10G	0.892	0.958	0.971	0.970	0.979	0.968
10G5U	0.960	0.990	0.978	0.994	0.984	0.995
10G20U	0.741	0.974	0.946	0.988	0.970	0.988
50G5U	0.971	0.986	0.975	0.991	0.980	0.992

Viable cell density is accurately described throughout the cell culturing period, indicating a correct identification of the limiting and inhibiting substrates. Moreover, the experimentally measured mAb concentration is closely matched from the simulation results as suggested both by Fig. 2.5 and the high CCC_{mAb} and R^2_{mAb} values that are particularly close to 1 (Table 2.3). Glucose follows a similar trend, with only the points at 48h and 288h slightly deviating from the experimental measurements. Glutamine, glutamate and ammonia are also closely described by the model, with a few points presenting slight deviation from the experimental measurements towards the end of the cell culture. Importantly, the model captures the metabolic shift from the production to the consumption phase of lactate. Whilst the model closely monitors the lactate concentration, the point of metabolic shift is simulated earlier when compared to the experimental measurements, that additionally present a steeper decrease in lactate concentration.

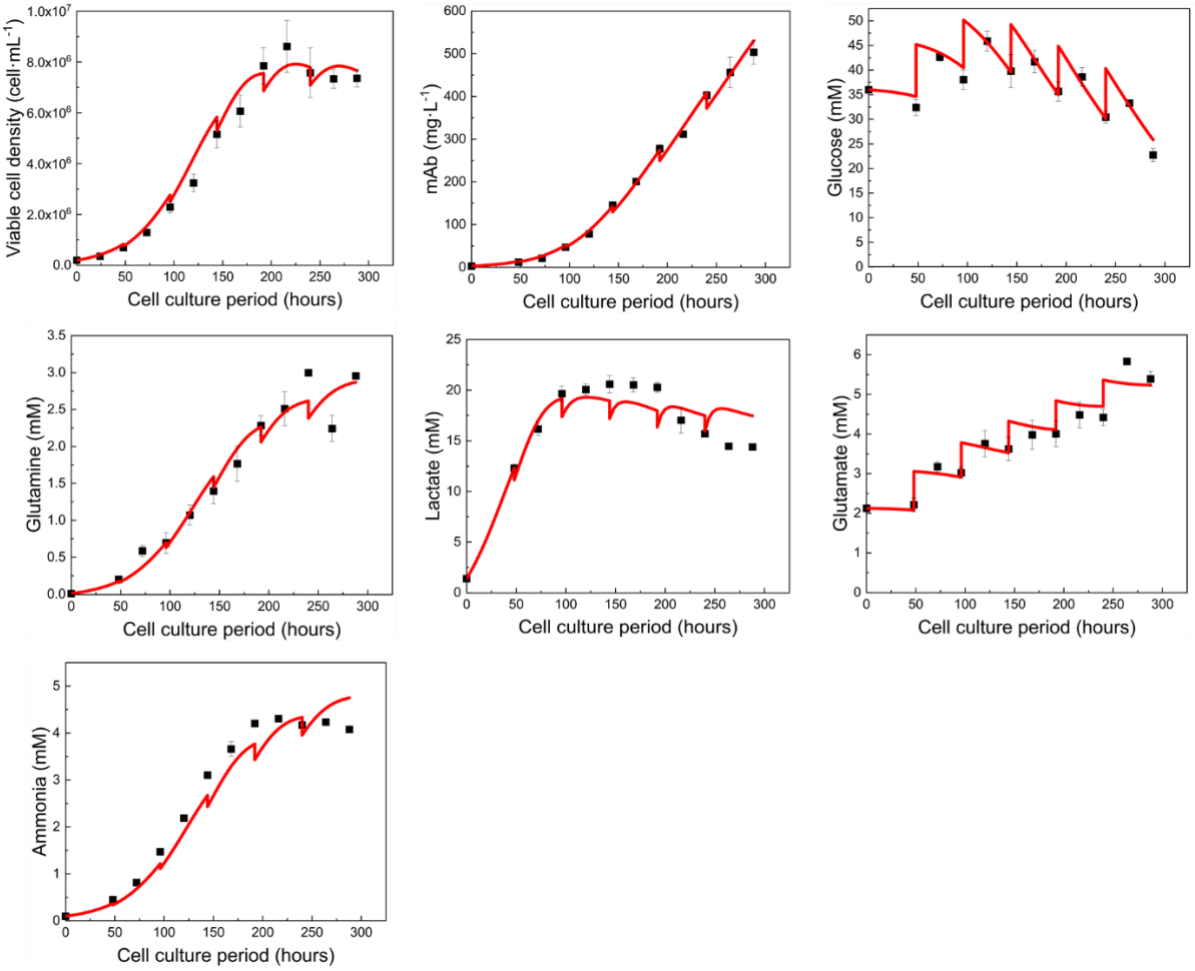

Figure 2.5: *Metabolism submodel* fitting to the experimental data for the control experiment following model parameterization. Galactose and uridine are not reported as they were not included in the feeding of the control experiment. The red lines show the model simulation results while the black squares present the experimental data. Taken with permission from Kotidis et al.[482].

10G experiment

The 10G experiment includes the addition of 10 mM galactose on days 4 and 8 of the cell culture. As discussed in section 2.2.1 and shown in Fig. 2.3, galactose can replace glucose towards the synthesis of glucose-6-phosphate in the glycolysis pathway and therefore can result in a reduced glucose uptake rate. Considering the aforementioned relationship between glucose and galactose, the model accurately describes the halted consumption rate of glucose while closely describing the dynamic profile of galactose, as shown in Fig. 2.6. The capture of the reduction of glucose uptake is attributed to Eq. 2.9 and 2.10 as discussed earlier in section 2.2.1. Whilst viable cell density and mAb concentration are closely monitored, the model deviates from the experimental results towards the end of the culture, indicating that the effect of the second galactose feeding (day 8) caused cell toxicity that was not expected and captured by the model. Viable cell density accounts for a CCC of 0.892 and R^2 of 0.979 (Table 2.3) that are considerably lower when compared to the control experiment.

Similarly, the CCC and R^2 values for mAb concentration are 0.958 and 0.968, respectively, indicating a good fitting but less accurate than the fitting of the control experiment.

In a similar trend, the model-simulated glutamine profile shows a substantial deviation from the experimental data towards the death phase of the cell culture. In the current modelling configuration, glutamine is assumed to be produced only from the activity of the GS enzyme and the utilization of ammonia and glutamate. The ammonia and glutamate levels of the 10G experiment were very close to the control experiment, indicating that glutamine was produced through a secondary pathway not considered in the model. On the other hand, model simulation of lactate, glutamate and ammonia present good fitting with the experimental data throughout the cell culture, as shown in Fig. 2.6.

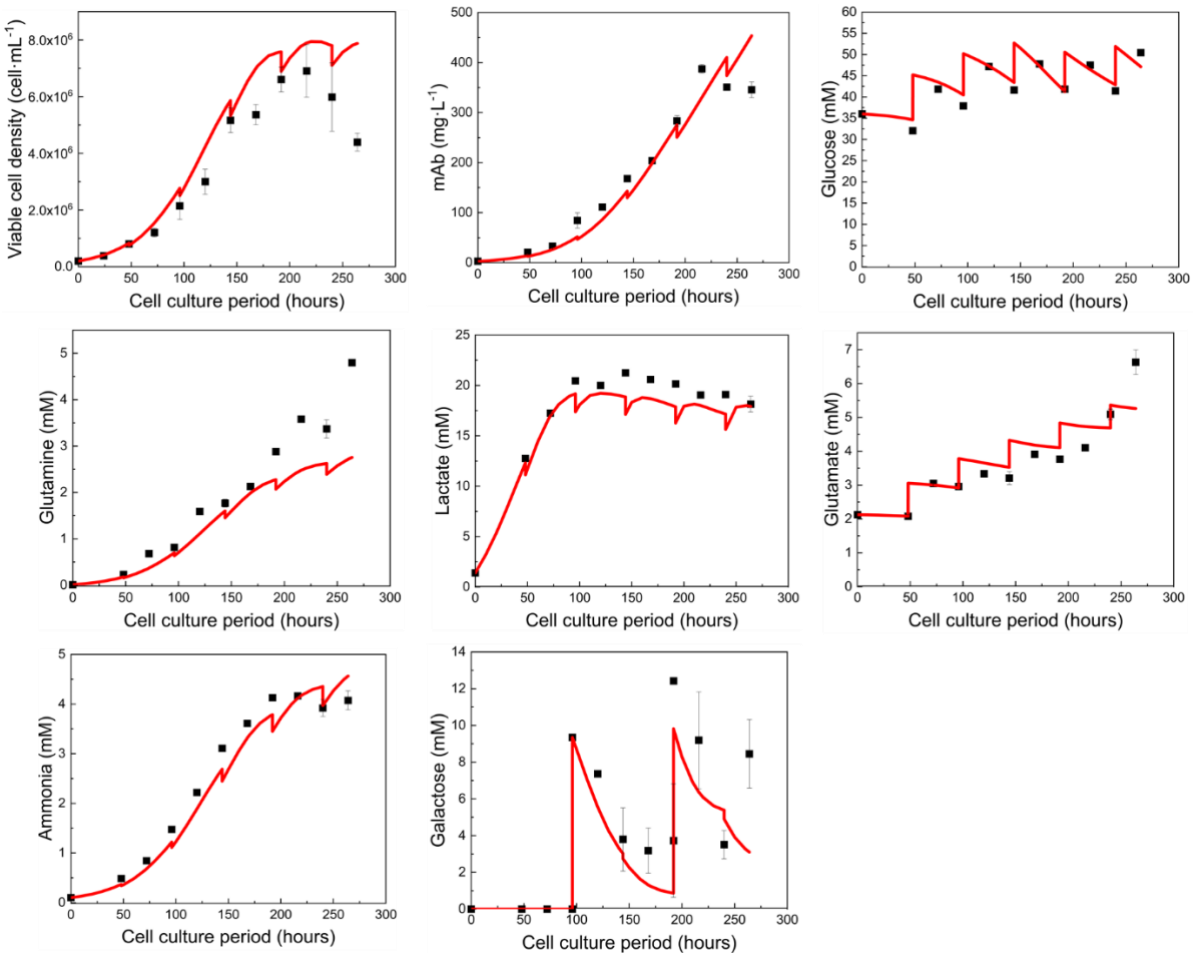

Figure 2.6: *Metabolism submodel* fitting to the experimental data for the 10G experiment following model parameterization. Uridine is not reported as it was not included in the feeding of the 10G experiment. The red lines show the model simulation results while the black squares present the experimental data. Taken with permission from Kotidis et al.[482].

10G5U experiment

The 10G5U experiment includes the feeding of 10mM galactose and 5mM uridine on days 4 and 8 of the cell culture. Whilst the inclusion of uridine in the feed is expected to increase the galactosylation

levels of the produced mAb, it is also expected to halt cell growth and productivity. The reduction of viable cell density and mAb concentration is closely matched by the model simulation, as shown in Fig. 2.7, and the CCC and R^2 values of 0.960 and 0.984 for viable cell density and 0.990 and 0.995 for mAb concentration, respectively, as shown in Table 2.3. Glucose concentration is well described for the first 192h of the culture, and then is progressively overpredicted until the end of the culturing period. The 10G5U experiment showed an unexpectedly higher consumption (both in terms of concentration levels and uptake rates) of glucose when compared to the 10G experiment, in which the same amount of galactose was added. Such an effect of uridine addition on glucose uptake is not included in the model. On the other hand, galactose consumption was similar between the two experiments and is well captured by the model. Uridine concentration is also well monitored apart from the measurements at 168h and 192h. Glutamine, lactate, glutamate and ammonia are closely matched from the model simulation and indicate a good overall fitting of the model.

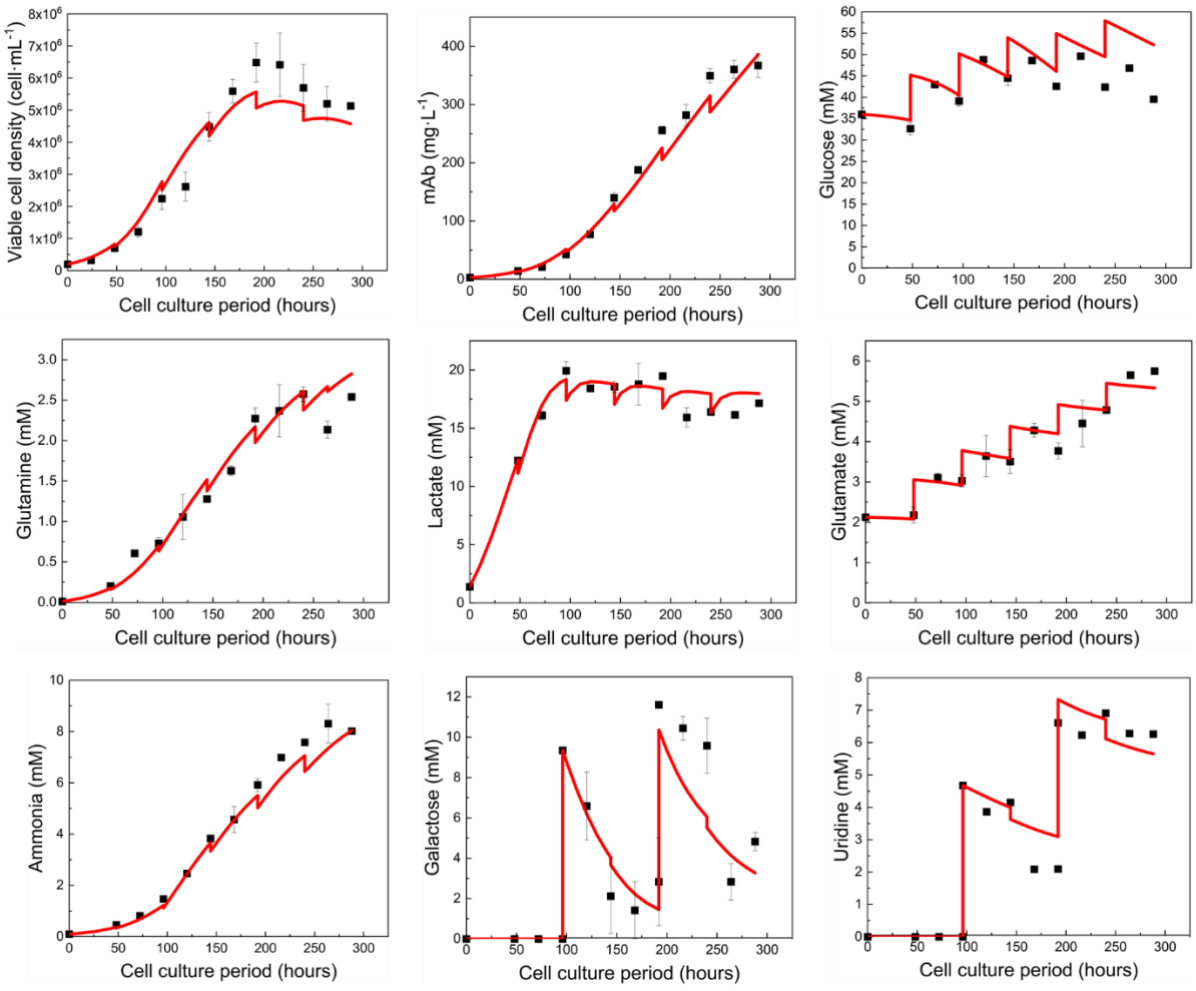

Figure 2.7: *Metabolism submodel* fitting to the experimental data for the 10G5U experiment following model parameterization. The red lines show the model simulation results while the black squares present the experimental data. Taken with permission from Kotidis et al.[482].

10G20U experiment

The 10G20U experiment includes the addition of 10 mM galactose and 20 mM uridine on day 4 of the culture. As the viability of the cell culture dropped below the 60% threshold at 192h, the culture was terminated and the feeding of galactose and uridine of day 8 did not occur. The increased concentration of uridine addition results in reduced cell growth and induced cell toxicity. Whilst the model correctly captures the effect of the elevated uridine concentration on viable cell density, the simulated values of viable cell density are considerably overpredicted at 168h and 192h of the cell culture period. The CCC and R^2 values (Table 2.3) for viable cell density, 0.741 and 0.970, respectively, confirm that the model overestimates the viable cell density when compared to the experimental data. On the other hand, the CCC and R^2 values of 0.974 and 0.988 (Table 2.3) confirm the trend presented in Fig. 2.8 indicating that the simulated mAb concentration shows only marginal differences from the experimentally measured values. Glucose, lactate, glutamate and ammonia are accurately matched by the model, with only slight deviations observed in the cell death phase. Glutamine was slightly underestimated, while both galactose and uridine showed a very good matching with the experimental data, with the exception of the measurement at 144h.

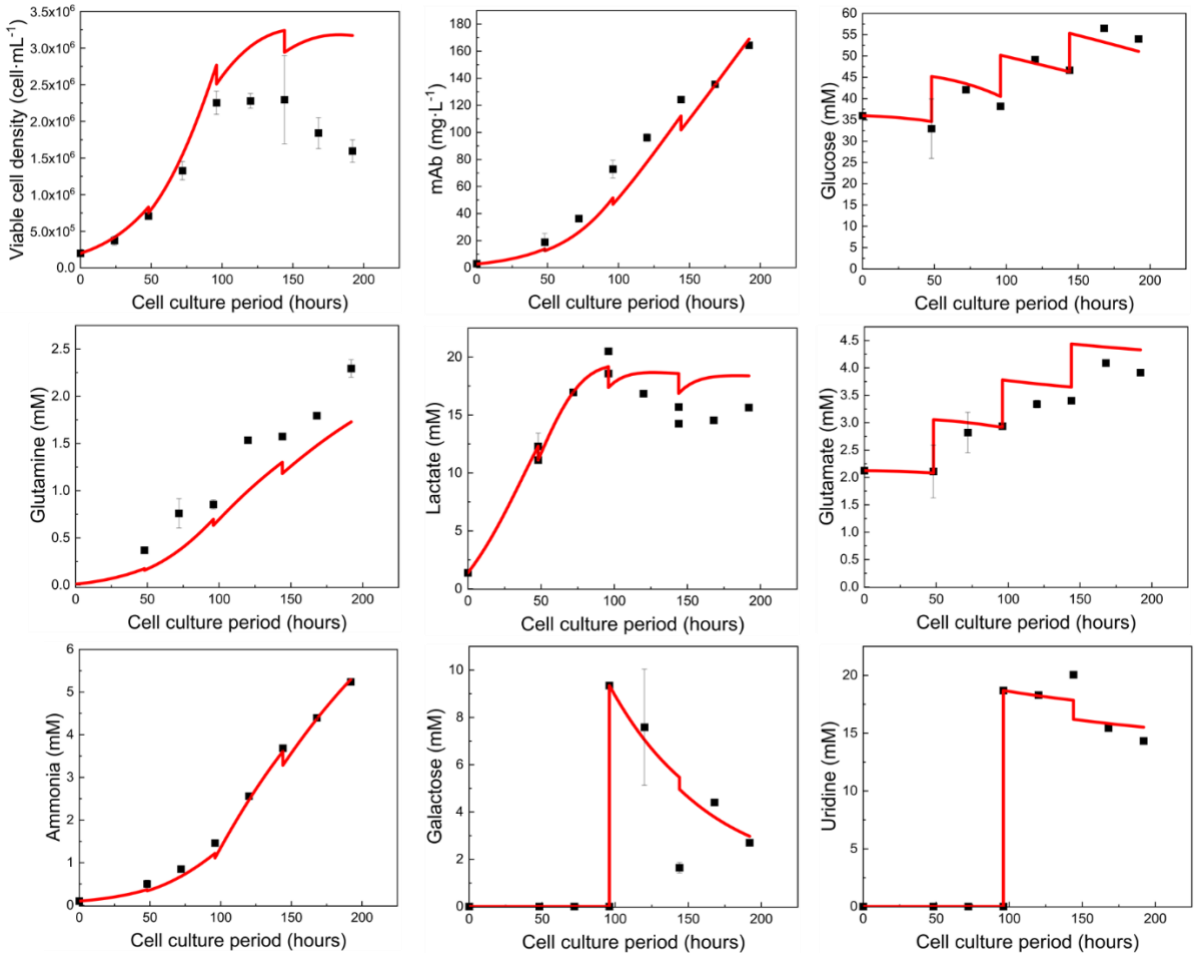

Figure 2.8: *Metabolism submodel* fitting to the experimental data for the 10G20U experiment following model parameterization. The red lines show the model simulation results while the black squares present the experimental data. Taken with permission from Kotidis et al.[482].

50G5U experiment

The 50G5U experiment includes the addition of 50 mM galactose and 5 mM uridine on days 4 and 8 of the culture. As shown in Fig. 2.9 and Table 2.3, the model accurately simulates the dynamic profile of both viable cell density and mAb concentration. The CCC and R^2 values for viable cell density are 0.971 and 0.980, respectively, and for mAb concentration are 0.986 and 0.992, respectively, confirming the trend observed in Fig. 2.9. Moreover, the model accurately matches the dynamic profiles of glucose, glutamine, lactate, glutamate and ammonia, with only a slight deviation observed in the glutamate concentration. Galactose and uridine concentrations were also, overall, closely described by the model simulations.

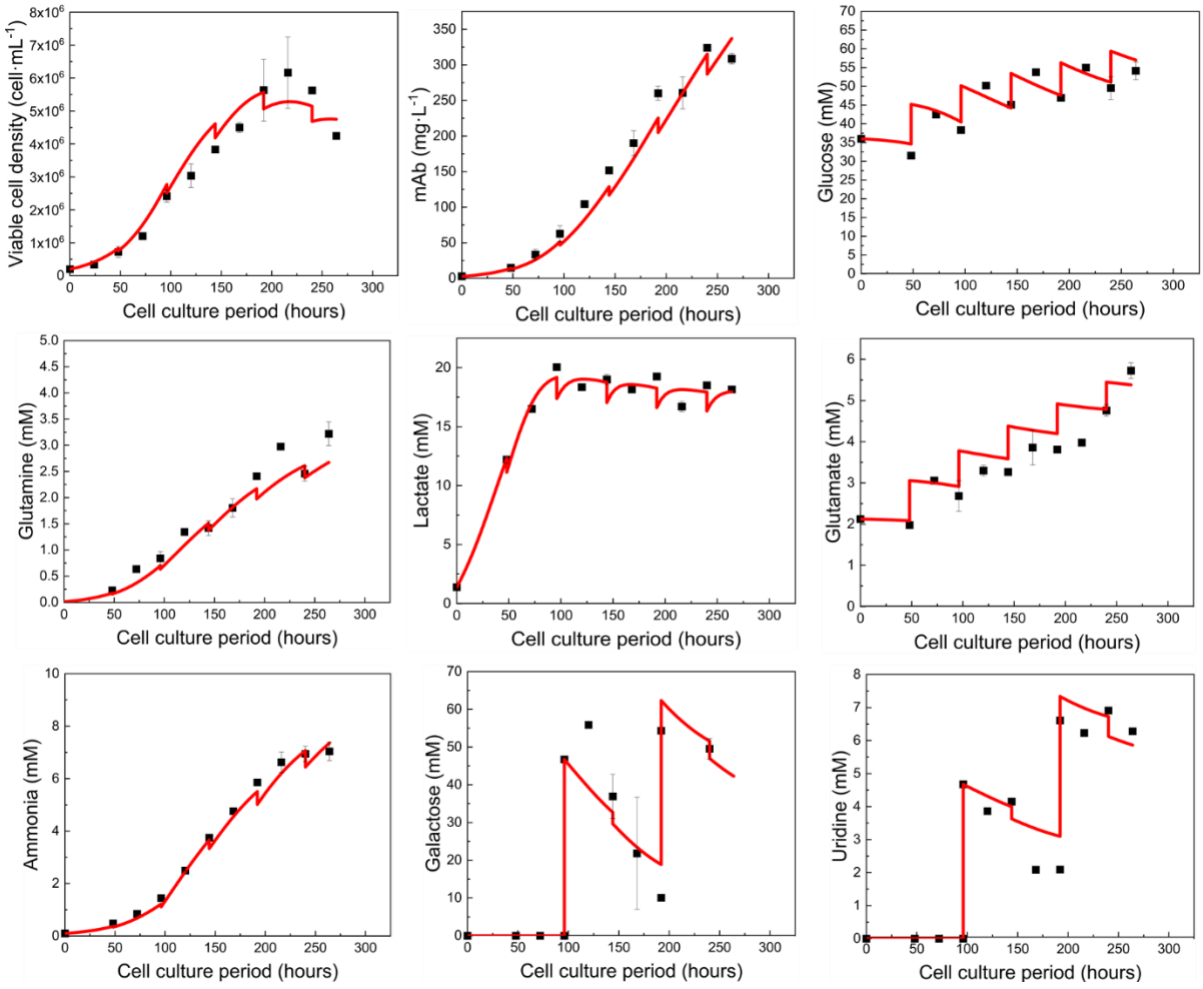

Figure 2.9: *Metabolism submodel* fitting to the experimental data for the 50G5U experiment following model parameterization. The red lines show the model simulation results while the black squares present the experimental data. Taken with permission from Kotidis et al.[482].

2.3.2 Evaluation of *NSD submodel* fitting

The aim of galactose and uridine addition is the build-up of the UDPGal intracellular concentration that can eventually lead to elevated mAb galactosylation levels. However, the inclusion of uridine in the feed also leads to increased intracellular UTP concentrations (Appendix-Fig. A1) and consequently to higher levels of all the UDP-sugars. For that reason, the main focus of the *NSD submodel* parameterization is placed on the fitting of primarily the UDPGal and secondarily the UDPGalNAc, UDPGlcNAc and UDPGlc profiles that are directly dependent to changes in the feed. GDPMan, GDPFuc and CMPNeu5Ac are not significantly affected from the galactose and uridine addition and are therefore less relevant to the performance of the model in the current context. However, as the latter NSDs are significant for the first steps of *N*-linked glycosylation (GDPMan and GDPFuc) and for HCPs or more complex recombinant proteins glycosylation (CMPNeu5Ac), a good fitting of the model for the respective concentration profiles was ensured (Fig. 2.10-2.14) in order to broaden the applicability of the model. Selected estimated parameters (maximum turnover rates) for the *NSD submodel* are

presented in Table 2.4. The rest of the estimated submodel parameters can be found in the Appendix-Table A1.

Table 2.4: Maximum turnover rates as estimated for the *NSD submodel*. Taken with permission from Kotidis et al.[482].

Estimated parameter	Value	Units	95% Confidence Interval
V_{max1}	9.22×10^{-1}	$(mmol_{NSD}\cdot L_{cell}^{-1}\cdot h^{-1})$	1.24×10^{-2}
V_{max2}	1.70×10^{-2}	$(mmol_{NSD}\cdot L_{cell}^{-1}\cdot h^{-1})$	9.78×10^{-4}
V_{max2b}	59.48	$(mmol_{NSD}\cdot L_{cell}^{-1}\cdot h^{-1})$	6.10
V_{max3}	5.50×10^{-2}	$(mmol_{NSD}\cdot L_{cell}^{-1}\cdot h^{-1})$	2.60×10^{-3}
V_{max4}	2.65×10^{-2}	$(mmol_{NSD}\cdot L_{cell}^{-1}\cdot h^{-1})$	2.12×10^{-3}
V_{max5}	1×10^{-4}	$(mmol_{NSD}\cdot L_{cell}^{-1}\cdot h^{-1})$	7.80×10^{-6}
V_{max6}	5.13	$(mmol_{NSD}\cdot L_{cell}^{-1}\cdot h^{-1})$	1.95×10^{-1}
V_{max7}	4.60	$(mmol_{NSD}\cdot L_{cell}^{-1}\cdot h^{-1})$	4.78×10^{-1}

Control Experiment

As described in section 2.2.2 the model parameters that were not related to the galactose and uridine addition were trained on the control experiment. Therefore, the fitting of the model to the control experimental data sets the basis for the model performance.

Table 2.5: Concordance correlation coefficients, Pearson correlation coefficients and R^2 values for the evaluation of the agreement between experimental measurements and model results for the intracellular concentration of UDPGal and UDPGlcNAc. Adapted from Kotidis et al.[482].

Experiment	CCC_{UDPGal}	$CCC_{UDPGlcNAc}$	PCC_{UDPGal}	$PCC_{UDPGlcNAc}$	R^2_{UDPGal}	$R^2_{UDPGlcNAc}$
Control	0.965	0.973	0.971	0.985	0.998	0.968
10G	0.897	0.919	0.930	0.983	0.963	0.922
10G5U	0.942	0.968	0.949	0.982	0.961	0.976
10G20U	0.982	0.941	0.989	0.965	0.998	0.952
50G5U	0.995	0.990	0.997	0.992	0.990	0.985

As shown in Table 2.5, the CCC and R^2 values for UDPGal and UDPGlcNAc are high enough to indicate a good fitting of the model. Indeed, the model achieves to closely describe all the intracellular profiles of the examined NSDs, as shown in Fig. 2.10. Whilst the model successfully tracks the steep increase of UDPGalNAc and UDPGlcNAc from 168h and until the harvest of the cell culture, the simulation

results present a significant deviation from the experimental data of the UDPGlcNAc up to the 120h time point. The discrepancy can be attributed to the imbalance between the synthesis and uptake reactions of UDPGlcNAc considered in the reduced NSD synthesis network, as presented in Fig. 2.4. As the fitting of the control experiment sets the basis for the fitting of feeding experiments, this phenomenon is encountered in the feeding experiments as well and its discussion will not be repeated.

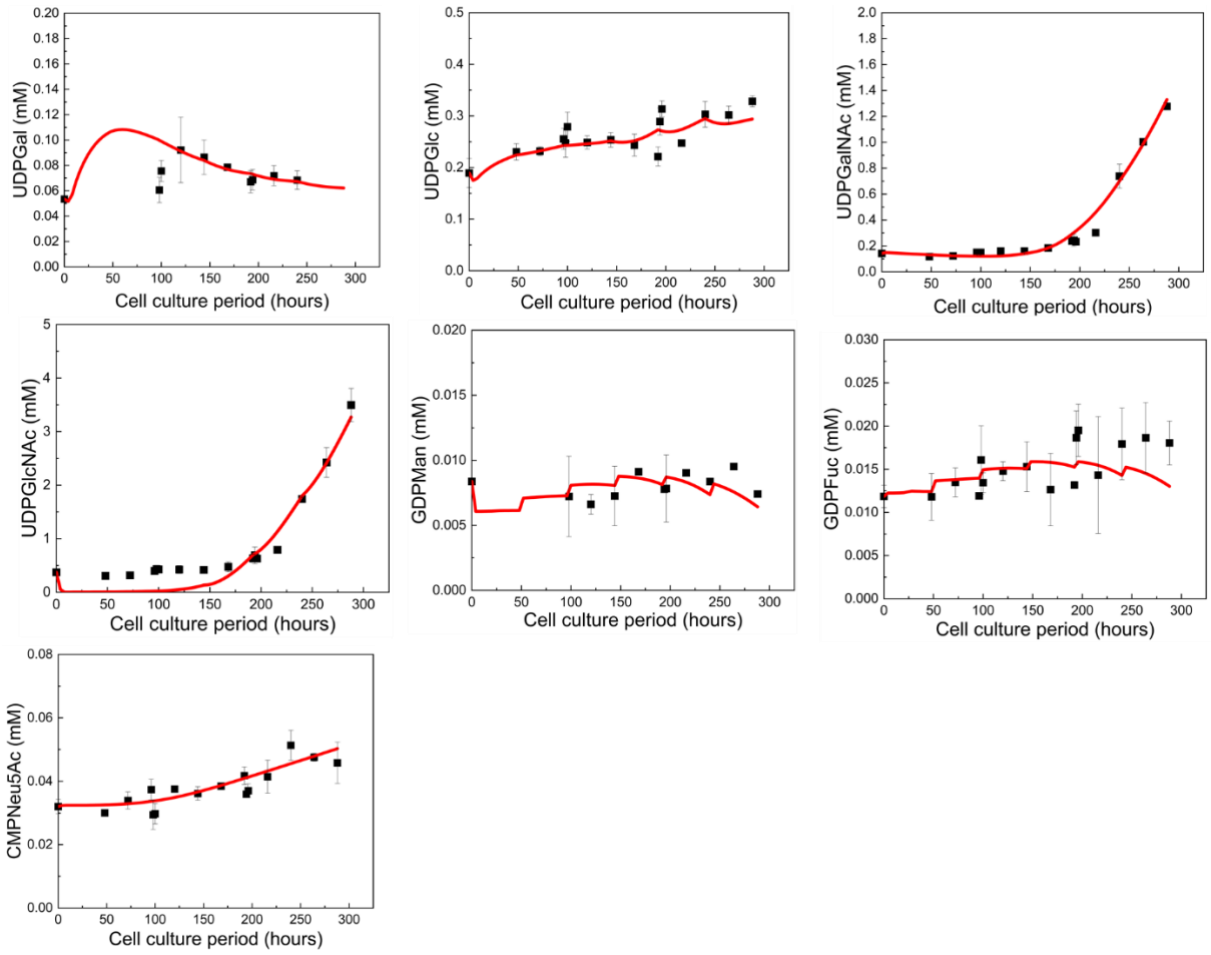

Figure 2.10: *NSD submodel* fitting to the experimental data for the control experiment following model parameterization. The red lines show the model simulation results while the black squares present the experimental data. Taken with permission from Kotidis et al.[482].

10G experiment

The inclusion of 10 mM galactose in the feed on days 4 and 8 of the culture results in the accumulation of UDPGal in the intracellular environment, as shown in Fig. 2.11. The CCC and R^2 values for UDPGal are 0.897 and 0.963, respectively, while the values for UDPGlcNAc are 0.919 and 0.922, respectively (Table 2.5). As presented in Fig. 2.11, the model accurately describes all the NSD dynamic profiles. Notably, the model is found to overestimate the concentration of UDPGlc after the 144h time point

of the cell culture period, possibly due to the elevated levels of UDPGal that acts as a substrate in the r2b epimerase reaction, catalysed by the UDP-glucose-4-epimerase enzyme (GALE gene). Within the modelling framework, the accumulation of UDPGal is expected to increase the flux of the reaction towards the synthesis of UDPGlc, a phenomenon that is not observed in the experimental data.

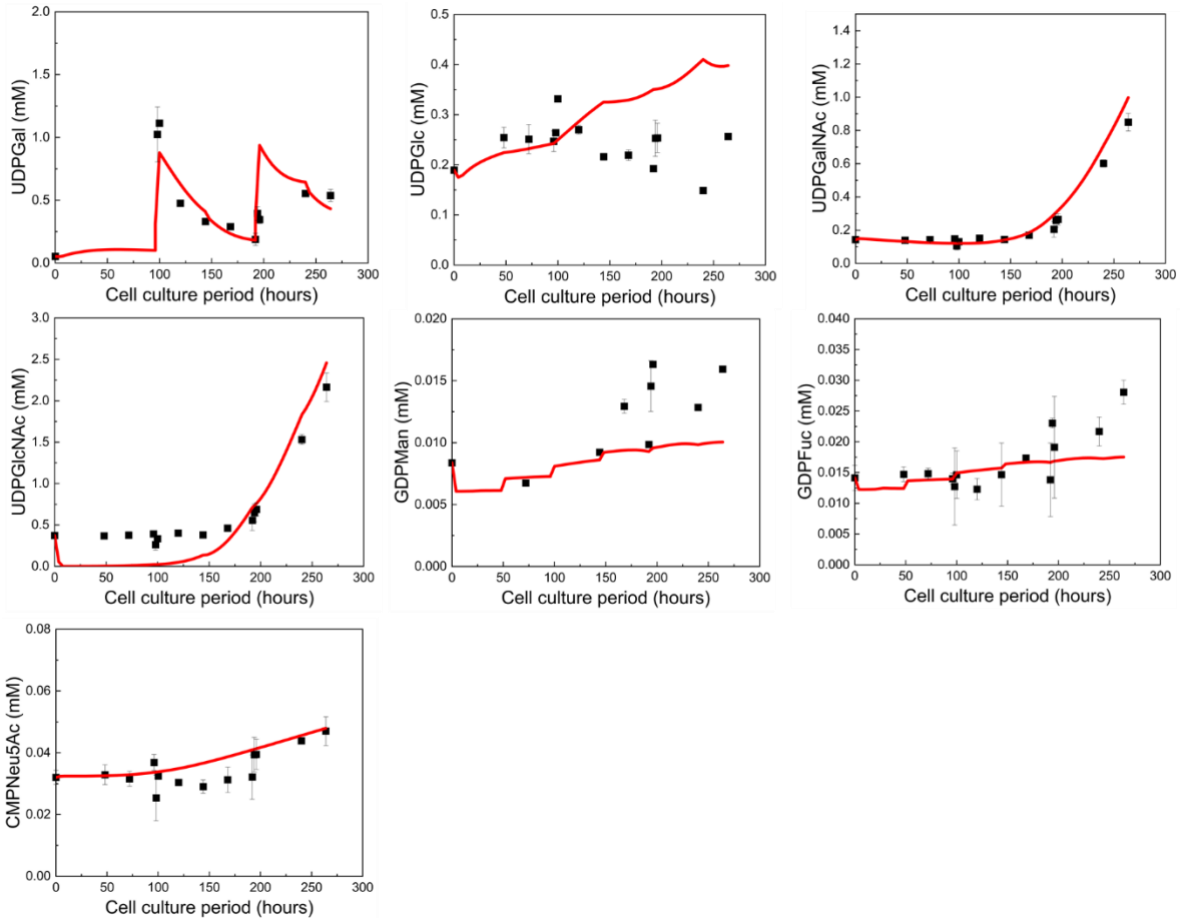

Figure 2.11: *NSD submodel* fitting to the experimental data for the 10G experiment following model parameterization. The red lines show the model simulation results while the black squares present the experimental data. Taken with permission from Kotidis et al.[482].

10G5U experiment

The addition of 10 mM galactose and 5 mM uridine is expected to increase the levels of all the UDP-sugars. As shown in Fig. 2.12, the model correctly captures the trend of increasing intracellular levels for all UDP-sugars. More specifically, the CCC and R^2 values for UDPGal are 0.942 and 0.961 (Table 2.5), respectively, showcasing the high levels of accuracy achieved through model fitting, with the exception of the measurement at 288h in which the model underestimates the UDPGal concentration. Moreover, the CCC and R^2 values for UDPGlcNAc are 0.968 and 0.976 (Table 2.5), respectively, supporting the close description of the UDPGlcNAc profile, as shown in Fig. 2.12. The model shows marginal differences from the experimental data, with the exception of UDPGlc that is underestimated

throughout the cell culture. The underestimation of the UDPGlc profile was also observed in the simulation of the remaining feeding experiments that included uridine addition, and its discussion is therefore not repeated. Whilst UDPGlc is an important NSD for the synthesis of the *N*-linked oligosaccharide precursor ($Glc_3Man_9GlcNAc_2$), glucose is trimmed from the precursor oligosaccharide immediately after its attachment to the glycoprotein backbone and is not added to the forming glycan in any of the steps of *N*-linked glycosylation that occur in the Golgi apparatus.

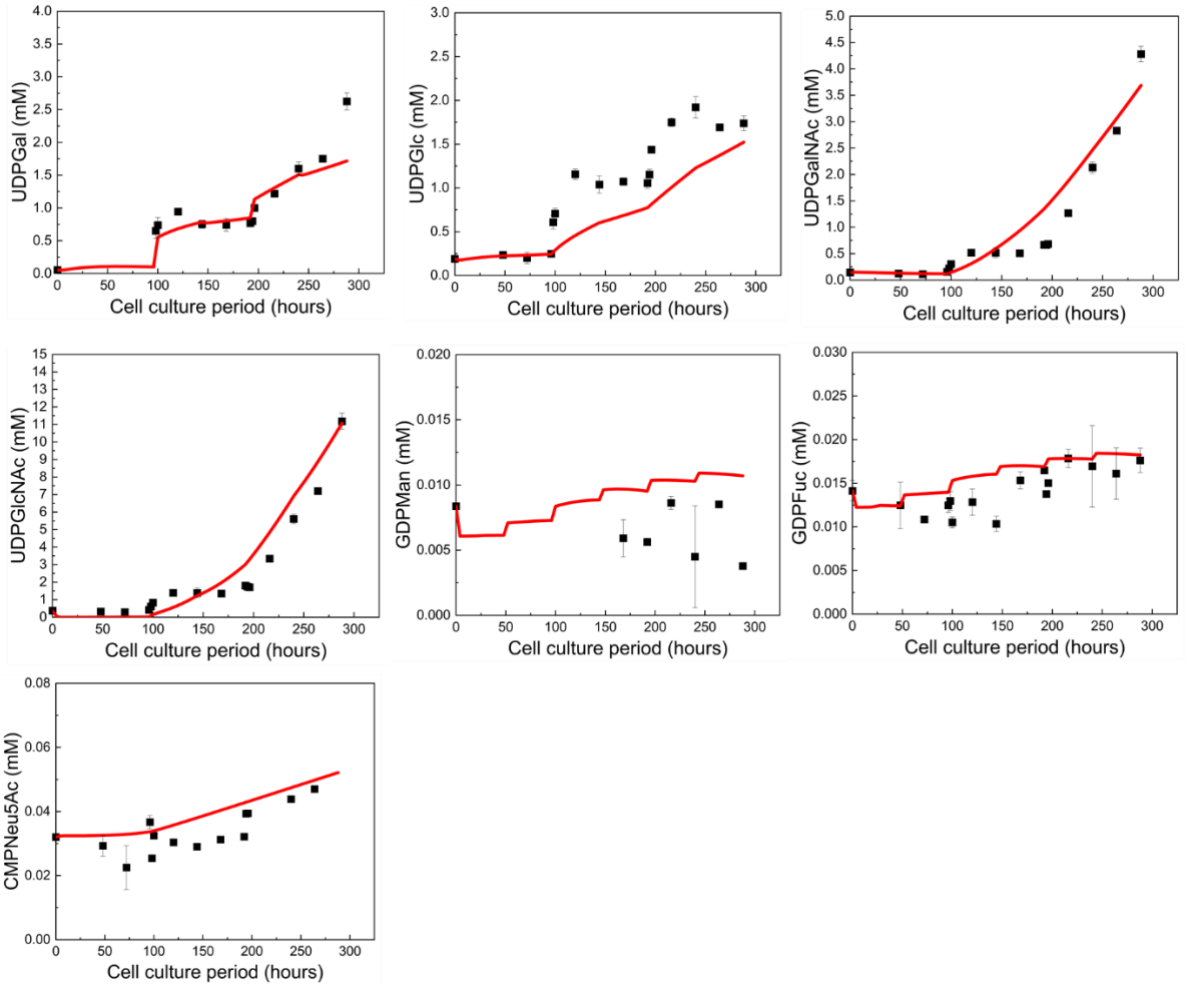

Figure 2.12: *NSD submodel* fitting to the experimental data for the 10G5U experiment following model parameterization. The red lines show the model simulation results while the black squares present the experimental data. Taken with permission from Kotidis et al.[482].

10G20U experiment

The inclusion of 10 mM galactose and 20 mM uridine results to even higher, when compared to the 10G5U experiment, UDP-sugar levels due to the increased uridine concentration in the feed. The model manages to accurately capture (Fig. 2.13) the effect of the elevated uridine levels on the UDP-sugars and for all time-points examined prior to cell culture harvest. As shown in Table 2.5, the CCC and R^2 values for UDPGal are 0.982 and 0.998, respectively, and for UDPGlcNAc are 0.941 and 0.952,

respectively, confirming the accurate fitting of the model to the experimental data, as suggested in Fig. 2.13.

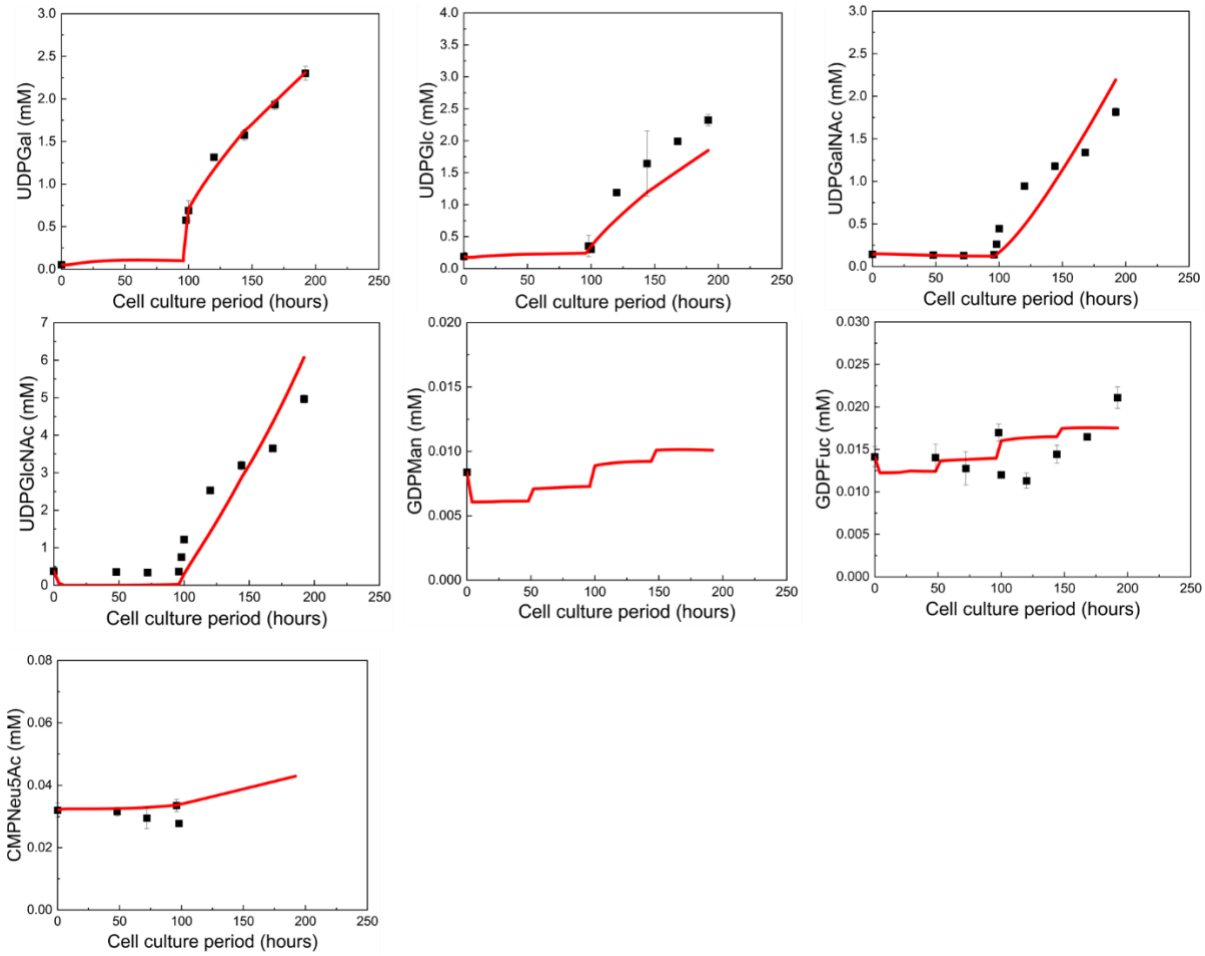

Figure 2.13: *NSD submodel* fitting to the experimental data for the 10G20U experiment following model parameterization. The red lines show the model simulation results while the black squares present the experimental data. Taken with permission from Kotidis et al.[482].

50G5U experiment

Similarly to the previous experiments, the model very accurately matches the UDPGal intracellular profile (Fig. 2.14) with the addition of 50 mM galactose and 5 mM uridine in the feed. The good fitting to the UDPGal profile is also confirmed from the CCC and R^2 values of 0.995 and 0.990 (Table 2.5), respectively. The CCC and R^2 values for UDPGlcNAc are 0.990 and 0.985 (Table 2.5), respectively, suggesting a successful fitting to the UDPGlcNAc profile.

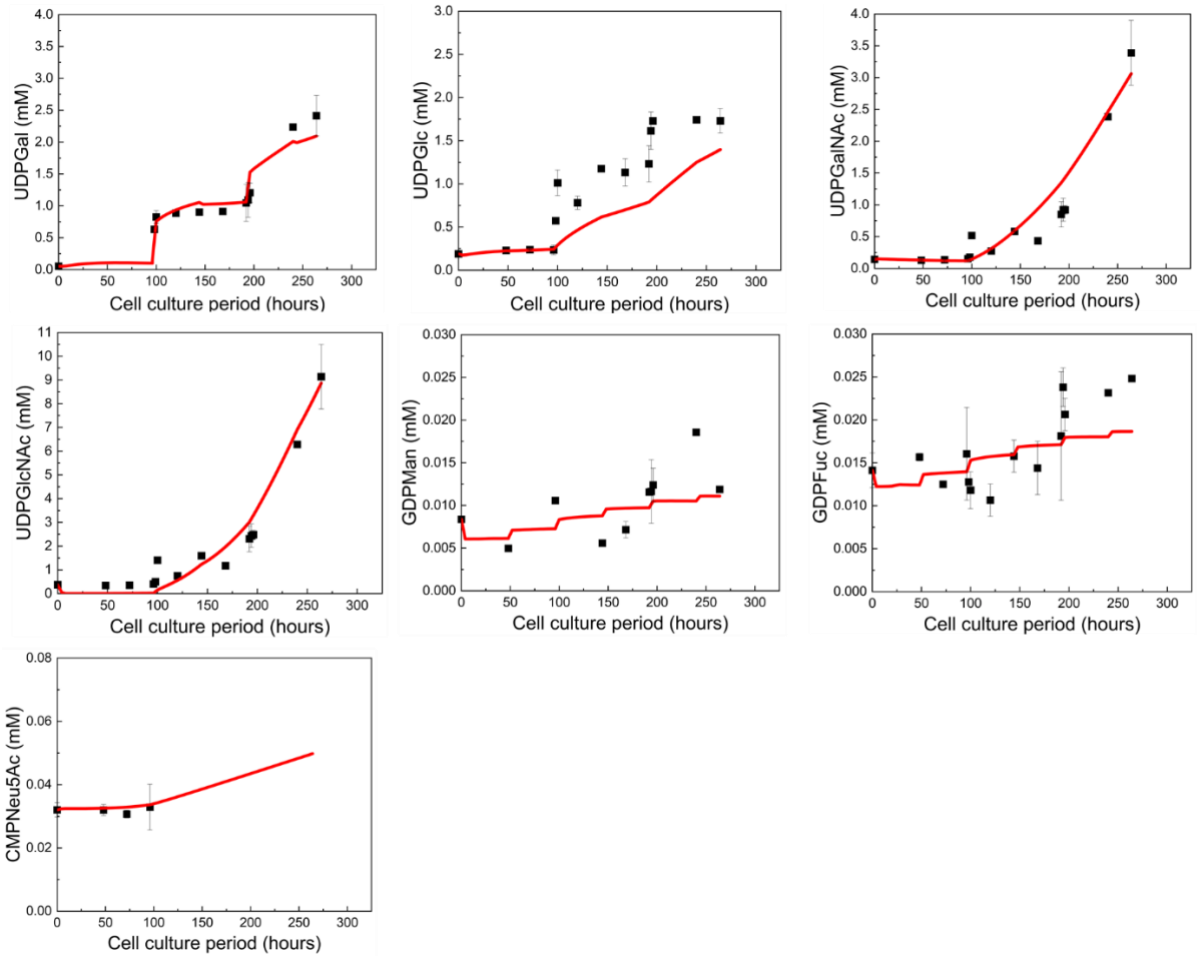

Figure 2.14: *NSD submodel* fitting to the experimental data for the 50G5U experiment following model parameterization. The red lines show the model simulation results while the black squares present the experimental data. Taken with permission from Kotidis et al.[482].

2.3.3 Evaluation of the *Glycomodel* fitting

The glycans that were experimentally measured and considered in the analysis are G0, G0F, G1F, G2, G2F and Man5 and were quantified at 168h, 216h, 264h and 288h of the cell culture period. For the cell cultures that were harvested earlier due to a viability below the 60% threshold, the model was used to predict the mAb glycoprofile for the remaining time points. As shown in Fig. 2.15 the vast majority of glycans distribution, estimated by the *Glycomodel*, is within a ±5% range of the experimental data. The only exception is the distribution of the G1F glycoform at 168h in the 50G5U experiment, in which the model underestimates the experimental value by 6.8%.

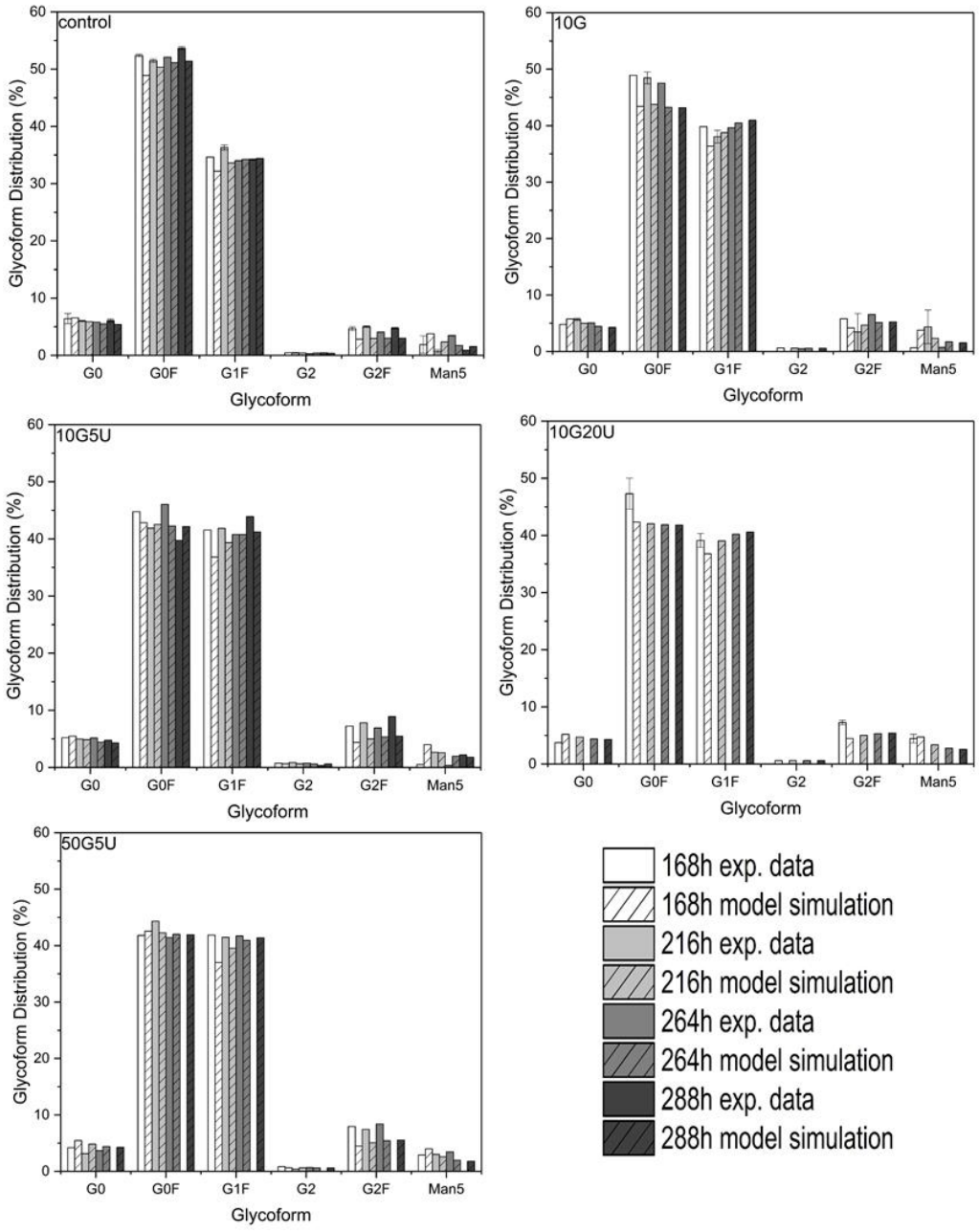

Figure 2.15: *Glycomodel* fitting to the experimental data for all the experiments considered during model parameterization. Taken with permission from Kotidis et al.[482].

In most cases of disagreement between model results and the experimental measurements, the model simulation presented an underestimation of the experimental data. The underestimation can be attributed to three reasons: 1) to insufficient estimated concentration of glycosyltransferases, 2) to the assumption that enzyme concentrations remain constant throughout the culturing process and 3) to the experimental quantification of only specific glycans. The enzyme concentrations of the *Glycomodel* were estimated based solely on the control experiment and therefore regulation effects of the glycosyltransferases due to the addition of galactose and uridine, as reported in literature[297,

[304], could not be captured. Additionally, the experimental analysis included the quantification of glycan distribution for certain targeted glycans and their normalization to account for 100% of the distribution. However, the model considers and calculates the formation of, in total, 77 potential glycan structures, the existence of which, even in small levels, can lead to reduced distribution of the targeted glycans. For example, the formation of G1 in detectable levels, the non-fucosylated version of G1F, is inevitable according to the suggested kinetics and mechanisms of the b4GalT enzyme and is calculated (by the model) to account for ~3.5% of the total distribution for the control and ~4.1% for the feeding experiments at harvest. The model-estimated distribution of the glycans was not normalized to only account for the G0, G0F, G1F, G2, G2F and Man5 glycans, as the model description is considered more realistic than the experimental analysis. Importantly, the simulated depletion of the intracellular UDPGlcNAc (discussed in section 2.3.2) in the early steps of the cell culture is not considered a possible factor for causing the observed underestimations, as the Man5 levels are accurately matched by the *Glycomodel*. Man5 is the substrate of GnTI, the first enzyme that utilizes UDPGlcNAc, and therefore the effect of low UDPGlcNAc levels would result in Man5 accumulation, something that is not observed in the model simulations. Finally, another contributing factor could be the effect of galactose and uridine addition in HCPs glycosylation that is not included in the current model configuration, and can indirectly, through the consumption of NSDs and as competitive substrates, influence mAb glycoform distribution.

Overall, the *Glycomodel* correctly identifies the increase of the galactosylated species (G1F, G2 and G2F) due to the addition of galactose and uridine in the feed, and closely describes most of the experimental data points. Glycoform distribution was not subjected to statistical analysis due to the relatively small number (4) of measurements for each experiment. The estimated parameters for the enzyme distributions and dissociation constants are reported in Table 2.6.

Table 2.6: Obtained enzyme distribution parameters and dissociation constants for the *Glycomodel*. Taken with permission from Kotidis et al.[482].

Estimated parameter	Value	Units	95% Confidence Interval[a]
$ManI_{max}$	1.034	μM	N/A
w_{ManI}	1.160	-	N/A
$ManII_{max}$	0.854	μM	N/A
w_{ManII}	0.953	-	N/A
$GnTI_{max}$	1.870	μM	N/A
w_{GnTI}	0.300	-	N/A

$GnTII_{max}$	1.566	μM	N/A
w_{GnTII}	0.912	-	N/A
$GalT_{max}$	0.464	μM	N/A
w_{GalT}	0.109	-	N/A
$FucT_{max}$	0.349	μM	N/A
w_{FucT}	0.470	-	N/A
Kdi_{FucTB}	11.2	μM	0.404
$Kdi_{GalTa1A}$	96.4	μM	2.653
$Kdi_{GalTa1B}$	345.8	μM	37.75
Kdi_{GnTI}	234.4	μM	7.15
$Kdk_{GalTa1A}$	1805.0	μM	213.2
$Kdk_{GalTa1B}$	1827.1	μM	1591

[a] The values have been obtained via the gPROMS Optimisation entity that does not report 95% confidence intervals.

2.3.4 Evaluation of model predictive capabilities

The validation of the model and the assessment of its predictive capabilities was performed against a model-based independent experiment aiming to maximize mAb galactosylation (mg·L^{-1}), as described in Eq. 2.45. The control variables of the optimization problem were the time points at which the model could include galactose and uridine in the feed and the concentrations of galactose and uridine in the feedings. The possible time points of galactose and uridine feeding started 48h post culture inoculation and was allowed every 48h thereafter.

$$galactosylation = 1 \times [monogalactosylated\ mAb] + 2 \times [digalactosylated\ mAb] \quad (2.45)$$

where, $[monogalactosylated\ mAb]$ (mg·L^{-1}) and $[digalactosylated\ mAb]$ (mg·L^{-1}) is the concentration of the secreted mAb that carry glycans with one or two galactose molecules, respectively.

Fig. 2.16 and Table 2.7 show a good fitting between model prediction data and the experimental measurements for both the viable cell density and the extracellular mAb concentration. Whilst the model accurately captures the effect of the added galactose and uridine concentrations on viable cell density and for all time points, the experimental measurements for the secreted mAb concentration are underestimated by the model prediction for the last three time points of the culture period. The reported elevated mAb concentration is not expected, as the final mAb concentration of the optimized experiment is higher than the respective concentration in the control (Fig. 2.5). Galactose and uridine are expected (both experimentally and in the model) to, even slightly, reduce the growth and

productivity of the cells, based on the trends of the training data. The aforementioned discrepancy can be attributed to batch-to-batch variations and to unknown metabolic effects that galactose and uridine can potentially show on cell growth and productivity. The CCC, PCC and R^2 values for both variables are above 0.95, indicating an, overall, accurate description of both viable cell density and mAb concentration.

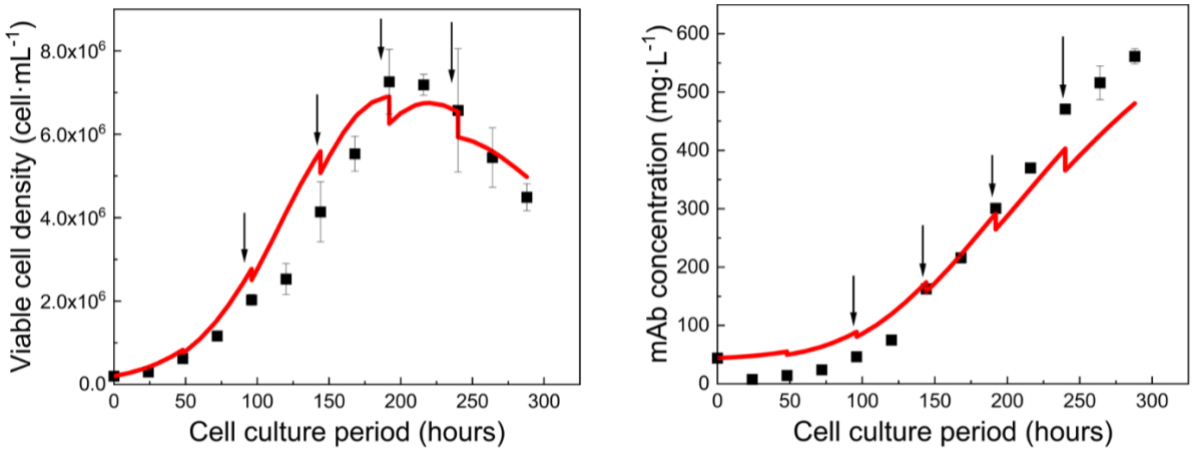

Figure 2.16: Model predictions (red line) compared to the experimental data (black squares) for viable cell density and secreted mAb concentration. Taken with permission from Kotidis et al.[482].

Similarly to the experimental datasets used for model training, the glycan analysis was performed at 168h, 216h, 264h and 288h of the cell culture period. As shown in Fig. 2.17, model accuracy was within a ±3.5% error range when compared to the experimental data, apart from the G0F distribution at 288h that was overpredicted by 5.3%. Interestingly, the experimentally observed mAb galactosylation (Eq. 2.45) exceeds by 33% the respective model-predicted value, mainly due to the underestimation of the final mAb concentration and G1F distribution from the model.

Table 2.7: Statistics for the evaluation of model predictions compared to the experimental data. Adapted from Kotidis et al.[482].

Concordance correlation coefficients		
Experiment	CCC_{Xv}	CCC_{mAb}
Optimization experiment	0.959	0.968
Pearson correlation coefficients		
Experiment	PCC_{Xv}	PCC_{mAb}
Optimization experiment	0.972	0.996
R squared (R^2)		
Experiment	R^2_{Xv}	R^2_{mAb}
Optimization experiment	0.984	0.984

As discussed in section 2.3.3, a regulation of the b4GalT enzyme due to galactose and uridine addition in the culture is highly expected but cannot be monitored through the current model configuration. Indeed, if the b4GalT concentration in the model is upregulated by 15% (Fig. 2.17), the agreement between the model and the experimental data is notably improved, specifically for the most abundant G0F and G1F glycans.

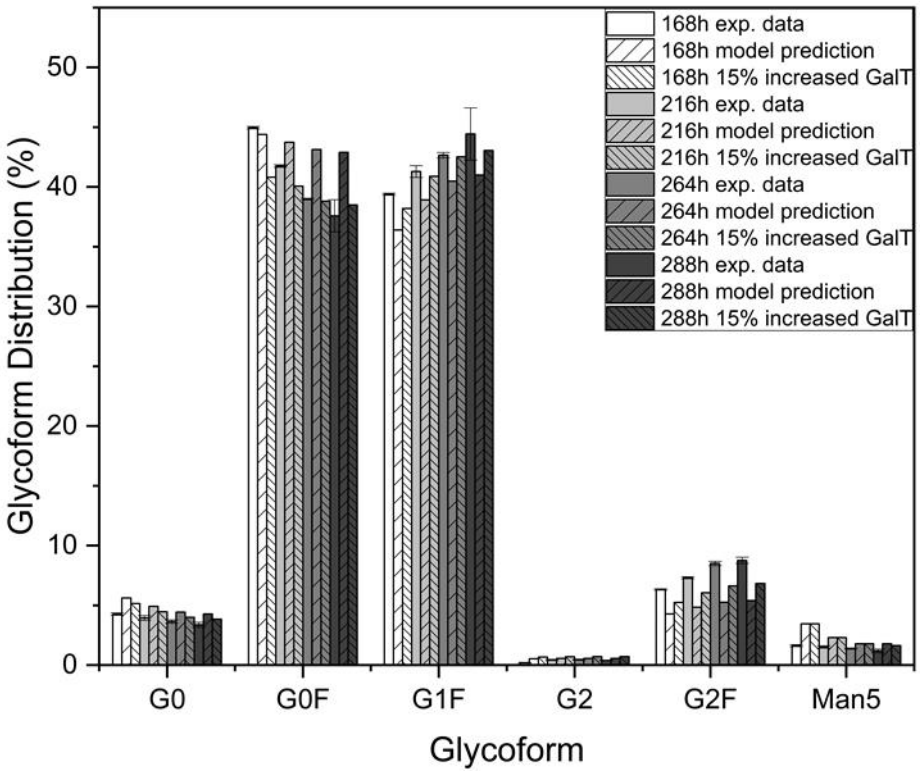

Figure 2.17: Model predictions for mAb glycosylation during the examined points of the cell culture period compared to the experimental data. A scenario of b4GalT upregulation by 15% is also examined. Taken with permission from Kotidis et al.[482].

The model-based optimized feeding experiment resulted in up to 64% increase of the absolute galactosylation (%) (Eq. 2.46), as shown in Fig. 2.18A. Moreover, the mAb galactosylation (Eq. 2.45) showed an even greater level of increase (Fig. 2.18B) due to the simultaneous elevation of the secreted mAb concentration when compared to the control experiment, overperforming the latter by 93% at the harvest point.

$$absolute\ galactosylation =$$

$$1 \times \%monogalactosylated\ mAb + 2 \times \%digalactosylated\ mAb \quad (2.46)$$

where, the right term of the equation describes the percentages of the distribution for the mAb that carries one or two glycans respectively. The difference between Eq. 2.45 and 2.46 is that the concentration of the secreted mAb is only included in the first. Qualitatively, Eq. 2.45 describes the concentration of the secreted mAb that is galactosylated (taking also into account the molecules of

galactose that are attached to the glycans), while Eq. 2.46 calculates the total and absolute value of the distribution of the secreted mAb that carries at least one galactose molecule in the *N*-linked glycans.

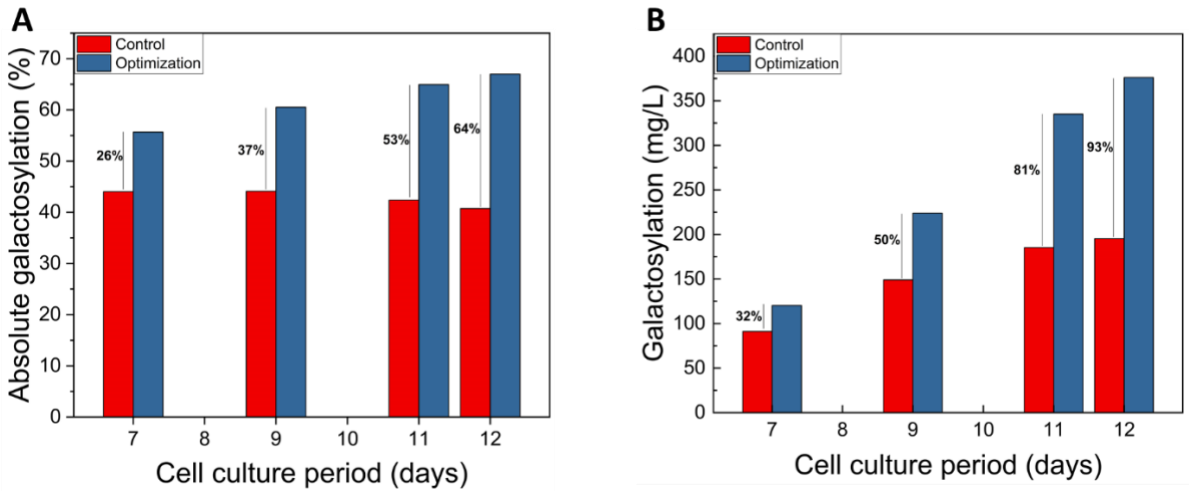

Figure 2.18: Comparison between the model-based optimization experiment and the control for the: (A) absolute galactosylation (%) and (B) mAb galactosylation (mg·L^{-1}). Taken with permission from Kotidis et al.[482].

2.4 Discussion

In Chapter 2, the development of a mechanistic mathematical modelling framework was described. The proposed framework includes three interconnected modules, namely: *Metabolism*, *NSD* and the *Glycomodel* (Fig. 2.1). The *Metabolism submodel* is a dynamic system that describes CHO cell growth, extracellular metabolism, cell death and antibody synthesis. The *NSD submodel* uses selected variables estimated in the *Metabolism submodel* in order to calculate the intracellular concentrations of the NSDs and their flux towards the Golgi apparatus where *N*-linked glycosylation occurs. Finally, the *Glycomodel* estimates the glycan distribution in the secreted mAb, using the NSD concentrations and fluxes from the *NSD submodel*.

The modelling framework was trained in a wide range of data from fed-batch experiments with various concentrations of galactose and uridine in the feed, and attempted to capture the effect of the aforementioned additions on the examined extracellular and intracellular pathways. Overall, model training achieved a very good agreement between the simulations and the experimental data, as expressed from CCC, PCC and R^2 values presented in the Table 2.3 and Table 2.5. Moreover, the model was able to describe the response of mAb glycosylation to the different feeding scenarios within a ±5% error range. Importantly, the structure of the proposed modelling framework enabled the prediction of mAb glycosylation solely based on the concentration of metabolites the culture media and feed.

In order to accurately capture the effect of galactose on CHO cell metabolism, the model assumes a regulation of the glucose uptake rate towards the glycolysis pathway, based on the ratio between the galactose and glucose specific consumption rates (Eq. 2.10). Galactose intervenes in the glycolysis pathway, as shown in Fig. 2.3, and can lead to reduced consumption levels of glucose[292, 297]. In order to avoid overparameterization, the model assumes a linear relationship between the specific growth and antibody production rates (Eq. 2.20) and manages to closely describe the secreted mAb concentration throughout the cell culture and for all training experiments. The *Metabolism submodel* is limited to a small range of metabolites and amino acids that can be easily monitored and controlled but can also be expanded to incorporate more amino acids and intermediates of the TCA cycle.

Similarly, the *NSD submodel* considers a reduced synthesis network in order to calculate the intracellular dynamic profiles of the NSDs (Fig. 2.4). When compared to a detailed NSD network that includes 34 material balances and 60 reactions[121], the current *NSD submodel* attempts to describe the dynamic profile of the same variables with a network reduced by ~75%, accounting for 7 material balances and 16 reactions. Additionally, the *NSD submodel* describes the effect of galactose and uridine addition on the directly affected NSDs, such as UDPGal, UDPGlcNAc, UDPGlc and UDPGalNAc, achieving a very good agreement with the experimental data, as shown in Table 2.5. However, due to the complex nature of the NSD network and the necessity for the estimation of both the maximum turnover rates and the saturation constants, the system suffers from structural unidentifiability. Whilst an effort to control the values of the saturation K_m constants and appropriately constrain them was undertaken, the estimated values of both parameters do not attempt to represent their actual physical values.

The *Glycomodel* (adapted from del Val et al.[478]) uses a PFR representation of the Golgi apparatus in order to describe the effects of Golgi maturation on protein *N*-linked glycosylation. The estimation of the *Glycomodel* parameters is based on the sequential steps proposed in section 2.2.3 and results in an overall good agreement with the experimental data. The limitations of the *Glycomodel* arise from the consideration of a constant concentration for the mannosidases, glycosyltransferases and NSD-transport proteins throughout the cell culturing period, not accounting the effect of different feeds, such as galactose and uridine, on the levels of the aforementioned enzymes[297, 304].

Following training completion, the model was validated against an independent experiment that was designed in a model-based manner in order to maximize mAb galactosylation. The model showed a very good agreement with the experimental data, closely describing the viable cell density and secreted mAb concentration (Table 2.7, Fig. 2.16). Importantly, the model-based experiment minimized the effect of galactose and uridine addition on viable cell density and mAb concentration.

The majority of reported glycans were described within a ±3.5% error range, while a 15% upregulation of the b4GalT concentration was proven enough to cover the differences between the experimental and model-predicted data (Fig. 2.17). Therefore, the model can be successfully used to describe both the intra- and extra-cellular dynamics and can be utilized as a designing tool for the identification of optimal feeding strategies, such as the optimal experiment that led to the 93% increase of mAb galactosylation reported in the model-based designed experiment of this study (Fig. 2.18).

Whilst the model can successfully describe the effect of galactose and uridine feeding on CHO cell metabolism and be used for designing optimal experiments, it is specific to the cell line used for training. A sensitivity analysis for the identification of the critical model parameters and their re-estimation based on data from a different cell line can broaden the applicability of the model and enable its use in a different system.

Chapter 3 Model applications within the Quality by Design paradigm

Chapter overview

A key challenge within the manufacturing context of biopharmaceuticals production is the control of protein glycosylation and the optimization towards a homogenous glycoform profile. Currently, the vast majority of bioprocesses lead to a vastly heterogeneous mix of glycosylated proteins[516]. As described in Chapter 2, mAb galactosylation is a desired attribute that enhances antibody CDC and ADCC activity[171, 182] and can be set as an objective for process optimization. However, the specific growth and protein production rates are two additional factors that are of paramount interest for evaluating cells' performance and bioprocess efficiency. It has been previously shown that both aforementioned variables affect recombinant protein glycosylation. The specific growth rate influences HCPs synthesis and therefore the NSD resources that are consumed towards HCPs glycosylation. Thus, a significant part of NSD pools is channelled to cellular glycosylation, and especially at high specific growth rates[513], indirectly affecting mAb glycosylation. On the other hand, specific protein production rate determines the residence time of the recombinant protein in the Golgi and, therefore, its availability for further processing by the GTases[486].

Adding another level of complexity between the aforementioned variables, and as described in Chapter 2, an attempt to manipulate protein glycosylation can hinder cell growth and productivity. Whilst several efforts have been made in order to assist the implementation of Quality of Design in biopharmaceuticals production using statistical tools[460, 467, 502], a mechanistic model can offer valuable insights on the underlying dependencies between process inputs and outputs[235, 240]. The utilization of mechanistic models enables the greedy exploration of optimal bioprocess conditions, such as feeding scenarios, that can ensure the implementation of the QbD paradigm and satisfy the CQA and KPIs constraints, such as mAb glycosylation and titre. The necessity for mechanistic models is strengthened by the fact that recombinant protein glycosylation shows a dynamic profile during the cell culture period, with significant discrepancies observed in the glycoform distribution between the early and the late stages of the culture. Thus, mechanistic models can assist in the understanding of how perturbations occurring early in the culture period can have an effect on growth, productivity and protein glycosylation towards the harvest point. Finally, a model-assisted implementation of Quality by Design can accelerate process development and reduce costs.

However, the tuning of mechanistic models includes the estimation of several kinetic parameters; a time-consuming process that is inextricably linked to model configuration and demands a

considerable amount of work that can significantly outweigh the effort put on the construction of model equations. For that reason, Global Sensitivity Analysis is commonly employed, for the identification and designation of parameters that significantly contribute towards the determination of the model variables, and the subsequent reduction of the parameter estimation burden[435-436]. Additionally, mechanistic models are usually specific to the training cell line and are sensitive to process variabilities. Thus, it is essential to develop a framework for the adaption of the mechanistic models to new process conditions and cell lines, enhancing that way their robustness for a use within the QbD methodology.

In Chapter 3 we examine two applications of the kinetic model proposed in Chapter 2, that fit within the Quality by Design paradigm. Firstly, we examine whether the model can be used as a trustworthy tool for identifying a Design Space of galactose and uridine feeding in fed-batch cultures, in order to satisfy inequality constraints, set for both the mAb concentration at harvest (quantity constraint) and the levels of mAb galactosylation (quality constraint). In order to examine the aforementioned scenario, we use the modelling framework in combination with constrained Global Sensitivity Analysis (cGSA) for the identification and dimensionality reduction of the Design Space, and then validate the results against the experimental data of a number of different feeding scenarios in fed-batch experiments. Secondly, using GSA we perform a strategic search for the identification of the minimum number of the *Metabolism submodel* parameters that are essential for ensuring model performance, and that, if correctly calibrated using data from a new cell line, can efficiently capture the metabolic activity using the same mechanistic equations and relationships. The proposed adaptation of the model to a different cell line widens the applicability of the model and enables the introduction of reduced kinetic models in describing multiple CHO cell lines metabolism.

The results of this work have been presented in two manuscripts: 1) Design Space identification in Kotidis et al.[517] and 2) *Metabolsim submodel* sensitivity analysis and adaptation in Kotidis & Kontoravdi[518].

The aim of this chapter is to provide answers to the following questions:

- Can the modelling framework proposed in Chapter 2 be utilized as a tool for the identification of a process Design Space, within which we can ensure product quantity and quality?
- Can the *Metabolism submodel* be calibrated to account for batch-to-batch variability?
- What are the significant parameters of the *Metabolism submodel*?
- Is it possible to adapt the *Metabolism submodel* to a new CHO cell line by only tuning the significant parameters?

3.1 Materials & Methods

3.1.1 Experimental data

Several experimental data of fed-batch cultures, including galactose and uridine addition, were utilized in Chapter 3. The experimental datasets are split between those used for: 1) Design Space identification and 2) *Metabolism submodel* sensitivity analysis and adaptation.

Design Space identification

The experiments used for the Design Space identification are presented in Kotidis et al.[517] and were performed by the members of the group: Dr. Cher H. Goey and Ms. Elisa Correa. The cultures followed an identical schedule to the experimental setup described in section 2.1.1 and 5.1.2. The viable cell density and viability of the cultures were monitored daily, whist mAb quantification and glycan analysis were performed on day 12 (harvest). Glycan analysis was performed by Dr. Calum McIntosh, former member of the group, as described in Kotidis et al.[517]. Briefly, mAb samples were diluted at 3 $g \cdot L^{-1}$ in 100 mM Tris pH 7.6 and 8 M urea and denatured with 10 mM DTT (Sigma-Aldrich, Dorset, UK) at 50 °C for 10 min. The samples were incubated at 37 °C for 4 h with mass spectrometry grade trypsin (Sigma-Aldrich, Dorset, UK) in 100 mM Tris pH 7.6, at a 1:20 trypsin to antibody ratio. Following trypsin digestion, samples were analysed using a triple quadrupole mass spectrometer (Waters, Milford, Massachusetts, USA). The only difference when compared to the experiments presented in Chapter 2 is regarding the concentration of galactose and uridine added in the culture, as shown in Table 3.1. Experiments were carried out in biological triplicates. The feeding strategies were chosen in order to explore the model-generated Design Space (Fig. 3.5).

Table 3.1: Concentration of galactose and uridine added in the feed for each feeding strategy and for each time point. FS1 is the control experiment where no galactose and uridine were added in the feed. Taken with permission from Kotidis et al.[517].

Feeding Strategy	**Galactose concentration in Feed C (mM)**				**Uridine concentration in Feed C (mM)**			
Time point	Day 4	Day 6	Day 8	Day 10	Day 4	Day 6	Day 8	Day 10
FS1	0	0	0	0	0	0	0	0
FS2	189.3	219.9	198.4	190.6	33.9	1.7	18.8	12.8
FS3	267.5	47.0	89.0	487.4	30.8	17.4	40.6	28.4
FS4	10.1	194.0	67.0	172.5	7.3	13.0	71.4	13.7
FS5	446.9	306.0	233.3	45.8	23.3	0.1	21.5	55.1
FS6	22.8	27.0	407.8	230.1	1.6	9.3	44.7	66.2
FS7	203.9	266.4	52.0	374.3	1.8	6.6	6.1	23.6

Metabolism submodel sensitivity analysis and adaptation

For the identification of the optimal sensitivity analysis, a set of five fed-batch experiments using the CHO-T127 cell line and each one following a different feeding strategy were performed. The detailed methodology used for cell culturing and analysis can be found in section 5.1.2 and is similar to the one described in section 2.1.1. The feeding strategies followed for these experiments were extracted from the boundaries of the Design Space and attempted to test model performance in the limits of the identified DS (Fig. 3.9).

The data for adapting the *Metabolism submodel* to a different cell line were taken from Kyriakopoulos & Kontoravdi[519]. Briefly, a CHO-GS46 cell line was cultured in fed-batch experiments using the same basal medium (CD-CHO) and fed with 10% v/v CD EfficientFeed™ C AGT™ Nutrient Supplement (same with the T127 cell line) every other day starting on day 2. No galactose or uridine were added to the culture. Cells were cultured at 36.5 °C, 140 rpm and 8% CO_2.

3.1.2 Model simulation and parameter estimation

All model simulations were performed using gPROMS ModelBuilder v.5.1.1 (Process Systems Enterprise, gPROMS, www.psenterprise.com/products/gproms, 1997-2020). For the re-estimation of parameter values, the Maximum Likelihood formulation in gPROMS was employed, as described in Eq. 2.1, utilizing the Parameter Estimation (now Model Validation) entity of gPROMS. R^2 values were estimated using the Regression function in the Data Analysis toolkit of Microsoft Excel for Office 365.

3.1.3 Global Sensitivity Analysis

The SobolGSA software[520] was used for performing Global Sensitivity Analysis. The Random Sampling-High Dimensional Model Representation (RS-HDMR) technique was used for metamodel construction[521], while three different sampling strategies were utilized for input generation and sensitivity analysis: Sobol', Pseudo-random and Scrambled-Sobol'. The number of generated inputs was set to 16,384 or 2^{14}. Moreover, the number of generated samples used for constructing the metamodel was set to 4,096 or 2^{12}, which was the maximum that the software could utilize considering the complexity of the *Metabolism submodel*. The maximum number of alphas (α) and betas (β) was set to 4 and 2, respectively, and the examined significance index thresholds (SIT) for evaluating model parameter significance were set as 0.05, 0.1 or 0.2.

3.1.4 Workflow for significant parameter identification

The strategic framework that was followed for identifying the significant parameters of the *Metabolism submodel* is shown in Fig. 3.1 and includes four main steps:

Step 1: GSA is performed for the model under consideration. The examined parameters are: 1) the sampling strategy for metamodel construction and parameter space (Sobol', Pseudo-random, Scrambled-Sobol') and 2) the deviation of parameters' value within which sampling occurs (10%, 30% and 50%).

Step 2: The constructed metamodels that exhibit high agreement ($R^2 > 0.9$) are used in the post-sensitivity analysis where three different SITs (0.05, 0.1 and 0.2) are examined in order to identify the significant parameters of the model.

Step 3: Each of the designated significant parameters is then included in a parameter estimation for adapting the model in use to new experimental data from the same cell line.

Step 4: The results of parameter estimation that indirectly indicate the efficiency of the sensitivity analysis, are subsequently evaluated in terms of goodness of fit with the experimental data and the optimum method is chosen.

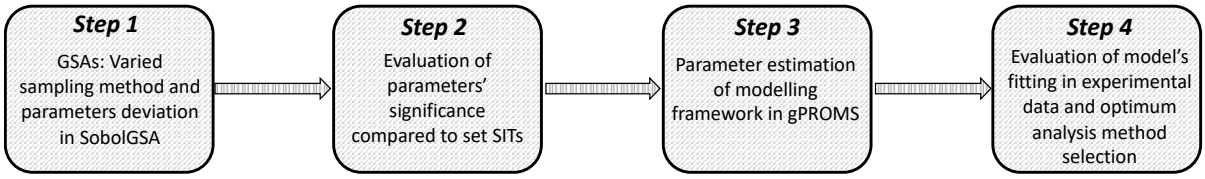

Figure 3.1: The strategic framework that was followed for the evaluation of different sensitivity analysis methods and identification of significant parameters. Taken with permission from Kotidis & Kontoravdi[518].

Fig. 3.2 shows the strategy used for examining the effect of: 1) the sampling method for input generation, 2) the chosen parameter deviation range and 3) the selected SIT. As a first step, any of the three sampling methods can be utilized and any of the parameter deviation (PD) ranges can be examined. The sampling methods are used in order to generate 16,384 (2^{14}) groups of parameter values within the constrained – by the PD range – values space, as shown in Eq. 3.1. Following the generation of the groups of parameter values, the model is simulated 16,384 (2^{14}) times using each time a different group of the generated parameter values.

$$x \in [(100\% - PD)Pv_{nominal},\ (100\% + PD)Pv_{nominal}] \quad (3.1)$$

where, x is any of the values that the examined parameter can take, PD is parameters' deviation and $Pv_{nominal}$ is the nominal value of the parameter.

The results of the multiple model simulations are used for sensitivity analysis, including the construction of the individual RS-HDMR metamodels. The same sampling methods are used for the construction of the metamodel and the sensitivity analysis. Finally, the significant parameters for each SIT and for metamodels that achieve a $R^2 > 0.9$ are re-estimated based on experimental data, and the

model fitting results are evaluated. The significant parameters identified using the optimum method that results from the strategic framework described in Fig. 3.1 and Fig. 3.2, are finally used for the adaptation of the model to experimental data of a new cell line (GS46).

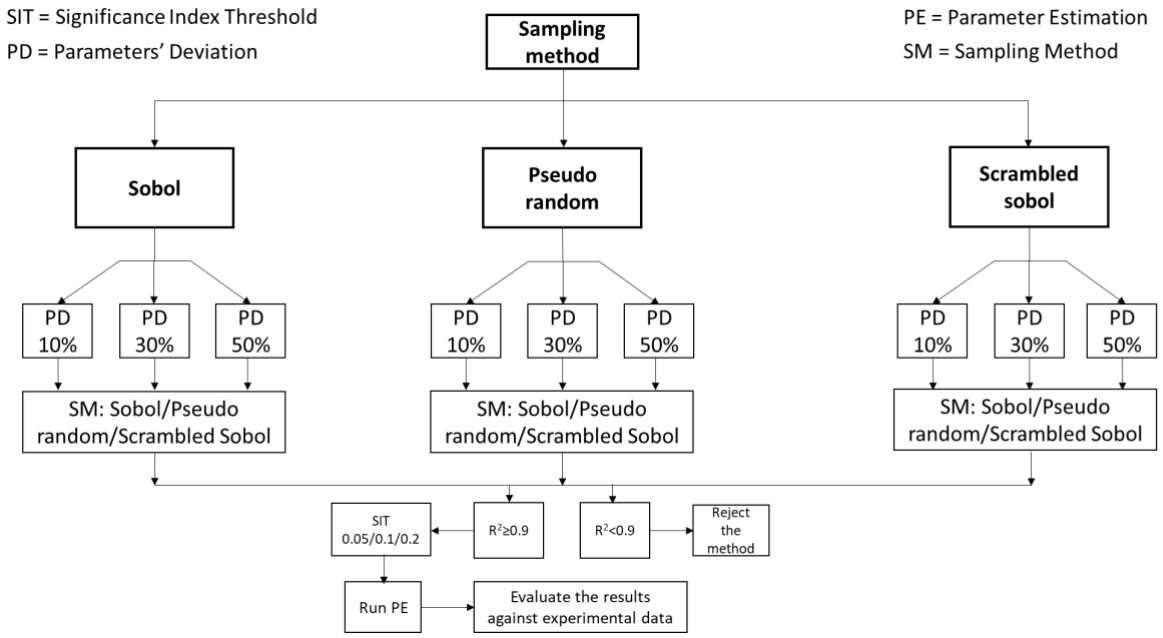

Figure 3.2: Workflow showcasing the methods used for evaluating the significant parameters of the *Metabolism submodel*. Taken with permission from Kotidis & Kontoravdi[518].

Importantly, sensitivity analysis was performed for each day of the culture. However, the importance of the parameters dynamically changes as the culture period progresses, with parameters related to cell growth accounting for the highest importance during the lag and exponential phase, and parameters related to cell death presenting a higher total sensitivity index (TSI) during the stationary and death phases of the cell culture. Therefore, due to the large aforementioned discrepancies of the TSI value for each parameter within the cell culturing period and in order to capture the importance of each parameter for the entire duration of the cell culture, the TSI significance was based on the inequality shown in Eq. 3.2:

$$\int_{0h}^{288h} f_P(t)dt \geq \int_{0h}^{288h} SIT\ dt \quad (3.2)$$

where, $f_P(t)$ is the function that describes the TSI of the parameter *P* as a function of cell culture time and SIT is the threshold chosen by the workflow shown in Fig. 3.2. When Eq. 3.2 is satisfied, the examined parameter *P* is considered significant for the chosen SIT.

An additional factor that could significantly influence the results of sensitivity analysis is the considered outputs, meaning the variables that are affected by the variation of the examined parameters and the response of which is the criterion for calculating the SIT. In the study presented

herein and for the CHO-T127 cell line, the viable cell density and the mAb concentration were considered as the set of outputs for all the analyses presented in Fig. 3.2. The chosen output variables are, to a great extent, affected by the rest of the model variables, such as the extracellular concentration of metabolites and amino acids, and therefore globally capture the effect of each parameter variation. Viable cell density was chosen over specific cell growth and cell death rates, as it captures the effect of both cell culture phases. On the other hand, all measured variables were included in the sensitivity analysis process used for model calibration to the GS46 cell line, in order to ensure the appropriate adaptation of all important model parameters.

3.1.5 Constrained Global Sensitivity Analysis

When compared to unconstrained GSA, the cGSA algorithm[522] utilized for the study presented in this chapter, accounts for inequality constraints, such as those imposed for product quality. The product quality constraints are defined by minimum thresholds for the values of mAb concentration and galactosylation at harvest. The examined process parameters in the cGSA are the concentrations of galactose and uridine in the feed on days 4, 6, 8 and 10 of the culture. The imposed quality constraints can be utilized for identifying parameters of the process (in this case the feeding concentrations of galactose and uridine) that significantly affect the constrained outputs (mAb concentration and galactosylation), and for identifying a model-generated Design Space.

Dimensionality reduction and Design Space identification through cGSA

The process examined herein accounts for 8 feeding concentrations of galactose and uridine in 4 feeding intervals. In order to reduce the dimensionality of the system and therefore expedite downstream analysis (DS identification) and enable the visualization of the results, the total sensitivity indices resulting from cGSA were used for the designation of the significant feeding parameters based on a predefined minimum threshold.

cGSA was also employed for the *in silico* identification of the DS, based on model prediction results. The code for DS identification has been developed by Dr. Oleksiy Klymenko and Mr. Panagiotis Demis and the equations of the system can be found in Kotidis et al. 2019[517]. Briefly, the code comprises an optimization problem which identifies the minimum amount and the nature of linear constraints that are adequate for constructing a DS that accounts for the maximum number of model-predicted feasible points while ensuring the exclusion of infeasible experiments.

Setting the constraints for cGSA

Secreted mAb concentration and mAb galactosylation were constrained based on the control experiment (FS1) where no galactose and uridine were added. The inclusion of the two compounds in

the feed is expected to lead to a reduction in cell growth and productivity. Thus, a reasonable 5% reduction of the final mAb concentration, when compared to the control experiment, was assumed as the minimum acceptable mAb concentration for the feeding experiments that would satisfy the requirements for inclusion in the DS. Therefore, the lower constraint for mAb at harvest day was set at 430 mg·L^{-1}. $mGal$, as calculated by Eq. 2.45, was constrained based on the maximum observed absolute galactosylation, as calculated by Eq. 2.46, in the control experiments and as shown in Eq. 3.3:

$$mGal_{constraint} = MAX(absolute\ galactosylation_{control})\ mAb_{constraint} \quad (3.3)$$

where, $MAX(absolute\ galactosylation_{control})$ is the maximum absolute galactosylation observed in the control experiment and that was 42%, and $mAb_{constraint}$ is the previously defined lower bound for the mAb concentration, set at 430 mg·L^{-1}. Therefore, the minimum $mGal_{constraint}$ is set at 180 mg·L^{-1}. No upper bound was set for both mAb concentration and $mGal$ as they are both desired attributes.

3.2 Results

This section includes the results of the cGSA for Design Space identification. The model-based predictions that are supposed to fall within the DS are compared with the respective experimental data, while two cases of experiments violating at least one of the DS constraints are also experimentally examined. Moreover, the results of various GSAs on the *Metabolism submodel* are evaluated in order to, eventually, choose the optimal method for adapting the model to experimental data from another cell line.

3.2.1 Design Space description

Collectively, the quality constraints for both model inputs (galactose and uridine concentration on days 4, 6, 8 and 10) and outputs (mAb galactosylation and mAb concentration) considered in the cGSA, are shown in Eq. 3.4-3.7:

$$[mAb] \geq 430\ mg \cdot L^{-1} \quad (3.4)$$

$$mGal \geq 180\ mg \cdot L^{-1} \quad (3.5)$$

$$0 \leq [Urd]_i \leq [Urd]_{max} \quad (3.6)$$

$$0 \leq [Gal]_i \leq [Gal]_{max} \quad (3.7)$$

where, $[Urd]_i$ and $[Gal]_i$ are the concentrations (mM) of uridine and galactose on days i = 4, 6, 8 or 10 in the feed and $[Urd]_{max}$ and $[Gal]_{max}$ are the maximum possible concentrations (mM) of uridine and galactose in the feed, respectively.

The constraints for $[mAb]$ and $mGal$ were set based on the analysis presented in section 3.1.5. The upper limits for the feeding concentrations were set at $[Urd]_{max} = 100$ mM and $[Gal]_{max} = 500$ mM. The maximum concentration for uridine was based on previous reports identifying the growth inhibitory levels of uridine[293], while for galactose the threshold was based on the compound's solubility in aqueous media.

The representation and identification of the Design Space suffers from the *curse of dimensionality*. The examined inputs of the system are eight, making the representation of the DS an *8D* problem. The inclusion of minimum constraints in the sensitivity analysis accommodates the identification of inputs that most significantly affect the targeted outputs. In order to reduce the dimensionality of the DS, a sample of 8,192 (2^{13}) points following uniform distribution in the *8D* dimension, was used for the implementation of cGSA[522-523]. The model was therefore simulated for these 8,192 different feeding scenarios and the results were used for cGSA. The implementation of cGSA was performed by the developers of the code Dr. Oleksiy Klymenko and Mr. Panagiotis Demis. The output of cGSA is similar to the SobolGSA software (that is used for unconstrained GSA) and calculates the main and total sensitivity index of each parameter (MSI and TSI, respectively). As presented in Fig. 3.3A, the unconstrained GSA (no inequalities imposed on the outputs) results in particularly high sensitivity indices for uridine concentration on day 4 ($[Urd]_4$) and day 6 ($[Urd]_6$) of the culture. All galactose feedings and the remaining uridine additions present low values of both MSI and TSI. The results clearly indicate that uridine concentration on the early days of the culturing period has a major effect on the unconstrained values of mAb galactosylation and concentration. Also, the almost identical values between MSI and TSI for each of the $[Urd]_4$ and $[Urd]_6$ indicate minimal synergistic effect between the inputs on the values of the outputs.

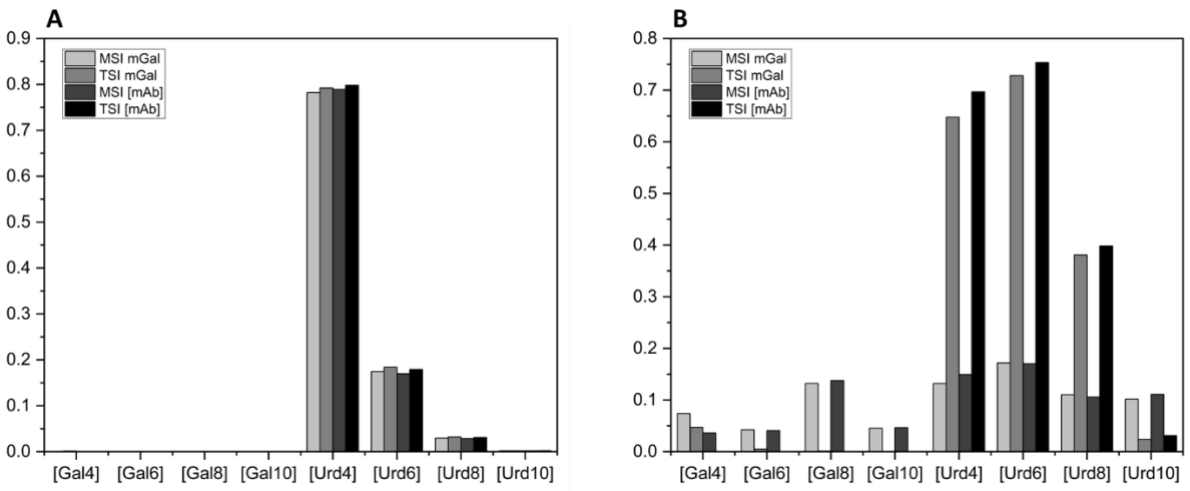

Figure 3.3: Main and total sensitivity index for each feeding concentration examining the effect on mAb concentration and galactosylation at harvest. (A) Results of the unconstrained GSA and (B) results of the cGSA, considering the constraints defined in Eq. 3.4 and 3.5. Taken with permission from Kotidis et al.[517].

With the introduction of the output constraints defined in Eq. 3.4 and 3.5, cGSA calculates the updated MSI and TSI values, as shown in Fig. 3.3B. Whilst $[Urd]_4$ and $[Urd]_6$ remain the most important feeding parameters, $[Urd]_8$ shows a higher TSI value for both $[mAb]$ and $mGal$. Therefore, the dimensionality of DS description can be reduced to a *3D* problem with the use of uridine feedings on days 4, 6 and 8. Additionally, cGSA revealed that out of the 8,192 samples, only 556 satisfied both constraints (feasible points). Notably, cGSA showed that the $mGal$-constraint, shown in Eq. 3.5, was less stringent than the $[mAb]$-constraint (Eq. 3.4) and was satisfied whenever $[mAb]$ was above the minimum accepted concentration. The normalised volume of the DS can be calculated through Eq. 3.8:

$$Vol_{DS} = \frac{Points\ satisfying\ both\ constraints}{Total\ number\ of\ points} = \frac{556}{8{,}192} = 0.068 \quad (3.8)$$

where, Vol_{DS} (-) is the normalized volume of the DS.

Normalizing the concentrations of $[Urd]_4$, $[Urd]_6$ and $[Urd]_8$ over $[Urd]_{max}$ results in the representation of all feeding concentrations in the [0,1] range and enables the calculation of the normalized volume for the DS. Fig. 3.4A shows the clustering of the points that fall within the DS (blue dots) against the points that fail to satisfy at least one of the output constraints (red dots). The resulting clustering of the successful feeding regimes indicates the accurate representation of the *8D* problem with the use of only the normalized $[Urd]_4$, $[Urd]_6$ and $[Urd]_8$.

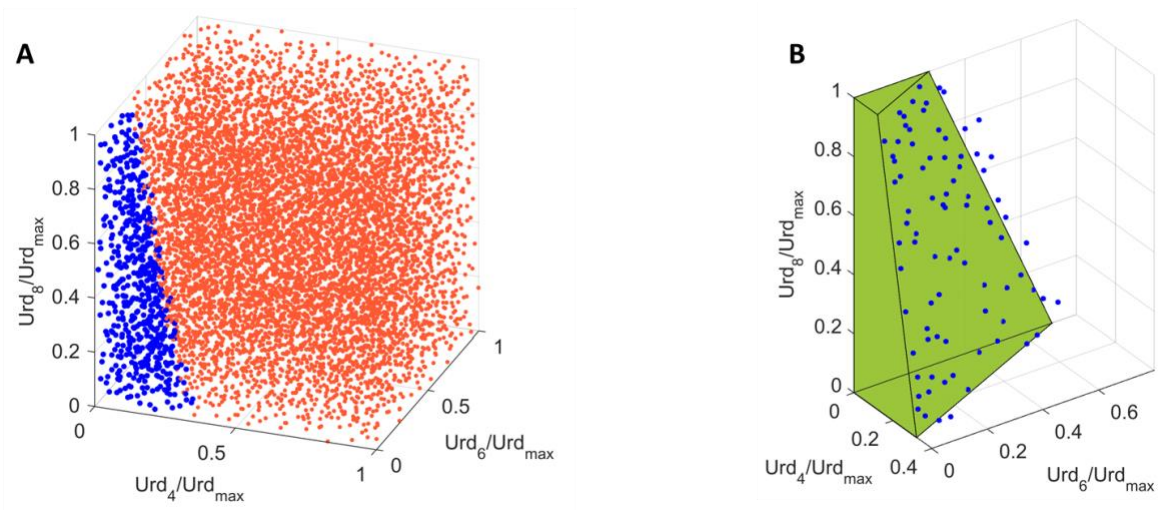

Figure 3.4: (A) A 3D representation of the DS using the normalised values of the most significant feeding concentrations as identified by the cGSA. Blue dots show the points that satisfy the output criteria of the DS (feasible points) and red dots the points that fail to satisfy the constraints. (B) The resulting DS representation after solving the optimization problem of a single linear inequality constraint. The blue dots show the feasible points that the DS representation is unable to capture. Graphs were plotted by Dr. Oleksiy Klymenko and are taken with permission from Kotidis et al.[517].

In order to represent the DS, the assumption of an additional linear inequality constraint and the solution of the optimization problem was found to closely describe the DS, as shown in Fig. 3.4B. Eq. 3.9 shows the result of the optimization problem for the identification of the linear constraint that leads to the DS representation shown in Fig. 3.4B. The resulting DS, defined by Eq. 3.9, exhibits a volume of 0.058, which is approximately 14.7% lower than the model-generated DS, as it rejects 83 out of the 556 feasible feeding regimes.

$$0.897[Urd]_4 + 0.405[Urd]_6 + 0.179[Urd]_8 \leq 0.288[Urd]_{max} \quad (3.9)$$

Therefore, the single linear inequality constraint approach successfully describes ~85% of the DS, whilst remaining a time and computationally efficient optimization problem. Moreover, the aforementioned approach outperforms more complex representations of the DS and leads to significantly higher accuracy than the comparably simplified 'box' shaped approach, in which each of the normalized feeding concentrations follows a separate inequality constraint[517].

3.2.2 Experimental validation of model-generated Design Space

As described in section 3.2.1 the model was used for simulating 8,192 feeding scenarios, 556 of which were found to satisfy both constraints defined in Eq. 3.4 and 3.5. In order to evaluate the DS that was identified based on model predictions for $[mAb]$ and $mGal$, the model was compared against seven experiments that were suggested by the model. Five of the experiments (FS1, FS4, FS5, FS6 and FS7) were expected to fall within the DS and two (FS2 and FS3) were expected to violate at least one of the constraints and therefore fall out of the DS. All feeding concentrations can be found in Table 3.1. As

discussed in section 3.2.1, no points violated the $mGal$-constraint and simultaneously satisfied the $[mAb]$-constraint, indicating that the lower acceptable value set for $mGal$ was not stringent enough. Fig. 3.5 shows the location of the experiments around and within the DS, according to their normalised values for $[Urd]_4$, $[Urd]_6$ and $[Urd]_8$.

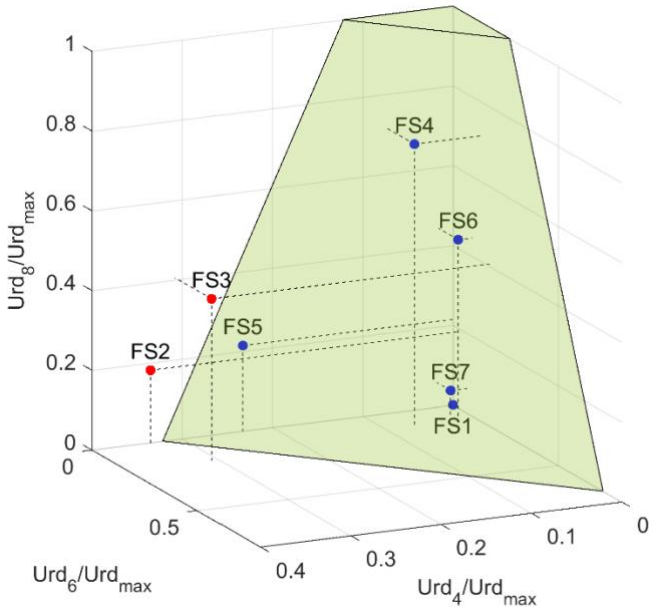

Figure 3.5: The position of the experiments against the Design Space. The blue dots represent the feasible points that satisfy both constraints and therefore fall within the Design Space. The red dots show the samples for which at least one of the constraints was violated. The graph was plotted by Dr. Oleksiy Klymenko and is taken with permission from Kotidis et al.[517].

Fig. 3.6 presents the agreement between model predictions and experimental data for mAb concentration and galactosylation on harvest day (day 12). For an experiment to violate a constraint, the average measured value, accounting also for the standard deviation derived from the biological triplicates, should not exceed the lower threshold. Model predictions closely describe the mAb concentration for all experiments, but FS7. Importantly, the model successfully identified the experiments that satisfy the constraint imposed by Eq. 3.4. The experimental measurement for mAb concentration at harvest and for the FS7 strategy considerably outperformed model-based predictions, as shown in Fig. 3.6A. Notably, the FS7 feeding experiment presented a mAb concentration that was notably higher than the respective concentration in the control (FS1). As discussed in Chapter 2, the formulation of the proposed modelling framework accounts for the inhibiting and toxic effects of galactose and uridine addition to the culture and therefore, an increased productivity and growth for a feeding experiment is not expected when compared to the control. Whilst the model underestimates the concentration of the produced mAb, FS7 undoubtebly satisfies the mAb-constraint as suggested by the model.

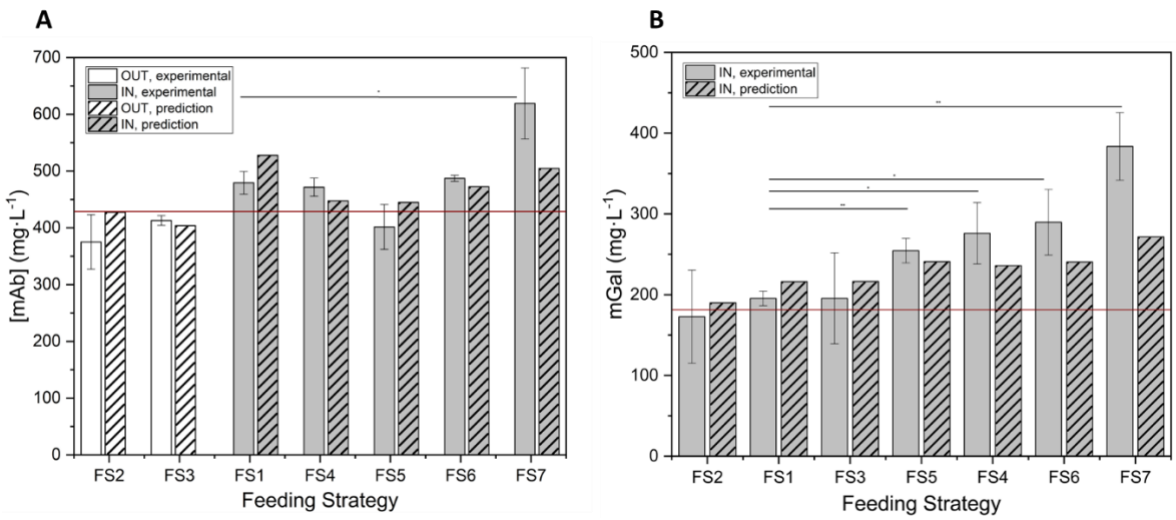

Figure 3.6: Model predictions (pattern bars) against the experimental data (solid bars) for (A) mAb concentration and (B) mAb galactosylation at harvest. The red line shows the minimum threshold that the variable needs to satisfy in order to fall within the Design Space. The white bars indicate the experiments that do not satisfy the imposed constraint for the depicted variable, while the grey bars show the experiments that exceed the minimum threshold value, when also accounting for the standard deviation of the experimental data. Taken with permission from Kotidis et al.[517].

As shown in Fig. 3.6B, all experimentally measured values of $mGal$ satisfy the inequality constraint (Eq. 3.5). Although FS2 resulted in a $mGal$ below the imposed threshold, the standard deviation of the experimental measurements indicates that there is no significant difference with the lowest accepted value. Notably, all feeding strategies resulted in higher values of $mGal$ compared to the control experiment, showcasing the improvement of both absolute galactosylation (percentage of glycans with at least one galactose molecule) and galactosylated mAb concentration. The improvement of absolute mAb galactosylation can be also seen in Fig. 3.7A and 3.7B, which show the agreement between model predictions and the experimentally measured mAb glycoform distribution.

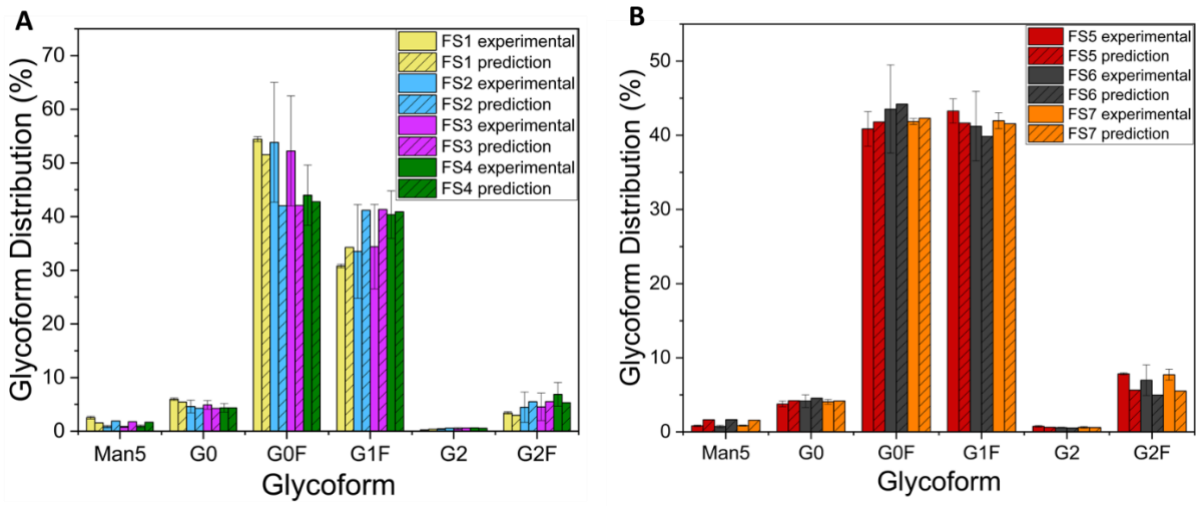

Figure 3.7: Model predictions (pattern bars) against experimental data (solid bars) for all the experiments and for each measured *N*-linked glycan. (A) Includes the FS1-FS4 experiments and (B) the FS5-FS7. Taken with permission from Kotidis et al.[517].

As shown in Fig. 3.7, model predictions very closely follow the experimentally measured glycoform distributions for FS1, FS4, FS5, FS6 and FS7 and for all examined glycans. The experimental measurements of feeding regimes FS2 and FS3 show considerably high standard deviations for the most abundant glycans, such as G0F and G1F. Whilst model predictions fall close to the standard deviation range of FS2 and FS3 experimental data, the model overpredicted the distribution of G1F in the aforementioned experiments. For that reason, G0F was considerably underpredicted as the distributions of G0F and G1F are strongly connected (G0F is the major substrate for G1F synthesis through the activity of b4GalT). Whilst the issue of b4GalT upregulation due to galactose and uridine addition has been discussed in Chapter 2, the underprediction of $mGal$ in the FS7 experiment is not expected to originate from inaccuracies in the description of the galactosylated glycan distribution (Fig. 3.7). In fact, the difference between the experimentally measured and model predicted $mGal$ value for the FS7 experiment is attributed to the respective disagreement in mAb concentration in this case.

As shown in Fig. 3.8, model predictions of viable cell density present an overprediction of the experimental values in all feeding regimes, including the control experiment. Surprisingly, the control condition in this set of feeding experiments presented a considerably suppressed growth profile compared to the results for the same control experiment used for model construction and training as presented in Chapter 2. Specifically, the new control experiment (FS1) achieved ~25% lower maximum viable cell density compared to the training control experiment. Considering that both control experiments were performed under exactly the same culture conditions and with the same cell line, the observed discrepancies can only be attributed to the batch-to-batch variability of cell growth, to

the inconsistency of Feed C performance and to the accumulation of errors during cell culturing and data analysis. Thus, model predictions resulted to an overestimation of the viable cell density for all experiments and especially for the exponential phase of the cell culture period.

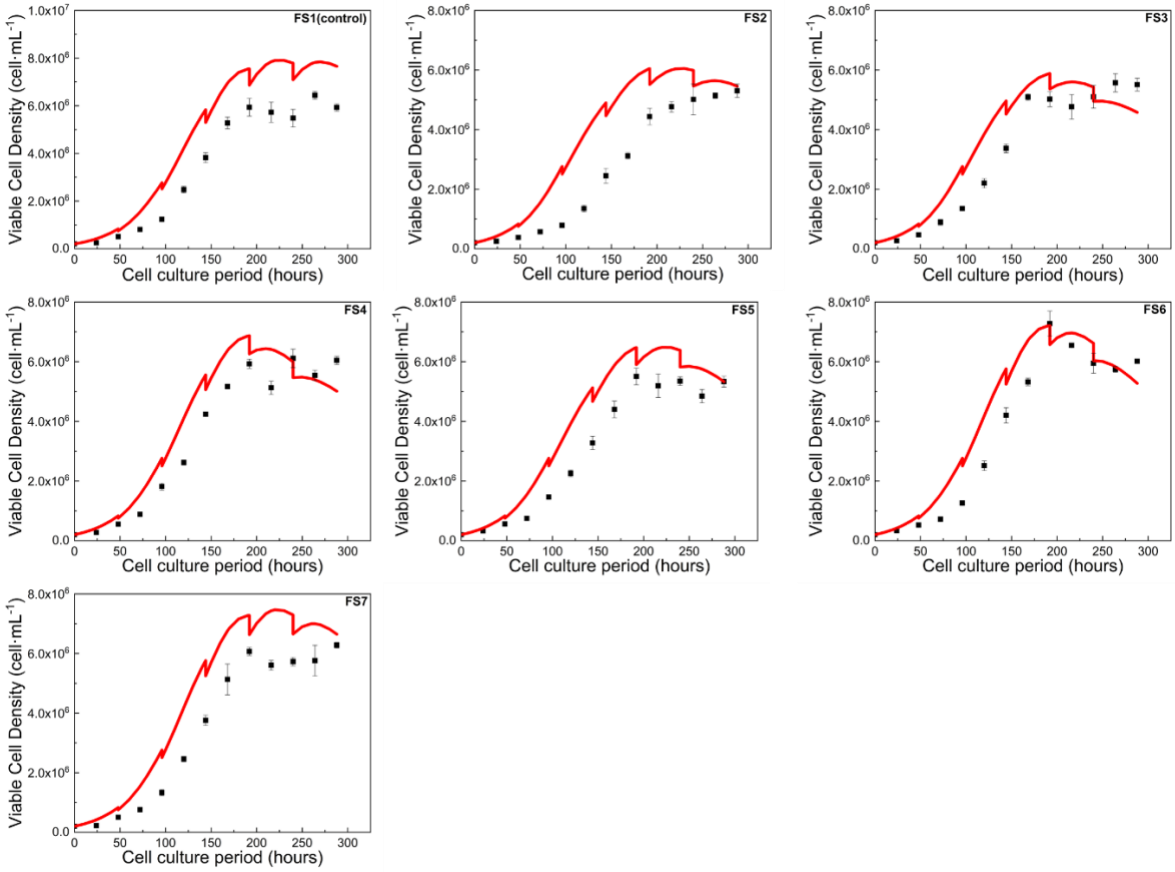

Figure 3.8: Model predictions (red line) for viable cell density against experimental data (black squares) and for all feeding regimes. Taken with permission from Kotidis et al.[517].

3.2.3 Model calibration to updated CHO-T127 experimental data

As shown in Fig. 3.8 and discussed in section 3.2.2, the CHO-T127 cell line exhibited a different dynamic profile of viable cell density, starting from the exponential growth phase, when compared to identical experiments that were performed from previous lab members. For that reason and in order to identify the parameters that can be quickly tuned for the adaption of the modelling framework to metabolic alterations of the cell line or to process variability, we performed an exhaustive GSA with an emphasis on viable cell density and mAb concentration. Five new feeding regimes of galactose and uridine (including the control experiment) that were extracted from the boundaries of the previously identified DS (Fig. 3.9), were carried out experimentally. The experimental performance of these experiments is presented in-detail in Chapter 5 of this thesis. The significant parameters were re-estimated according to the feeding experiment that showed the lowest experimental standard

deviation for viable cell density and mAb concentration (P1), and their predictive capabilities were evaluated against the four remaining experiments.

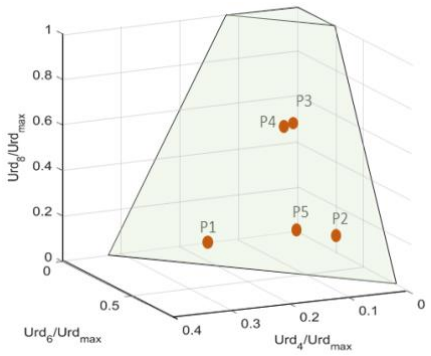

Figure 3.9: Coordinates of the new feeding experiments (P1-P5) in the DS. P5 is the control experiment. Note that whilst P3 and P4 show similar Urd_4, Urd_6 and Urd_8 concentrations, the Urd_{10} is notably different.

GSA results for the Metabolism submodel

Several different configurations of GSA were examined, as shown Fig. 3.2, investigating each time a different sampling method (Sobol', Scrambled-Sobol' and Pseudo-random) and degree of parameter deviation (10%, 30% and 50%). A deviation of parameter values over 50% was proven infeasible, as it resulted in multiple crashes (negative values for positive-only variables such as metabolite concentrations) during the 16,384 simulations performed for evaluation of parameter significance. The Pseudo-random sampling method resulted in $R^2 << 0.9$ of metamodel fitting for all time points of the cell culture period and for both variables examined, and was therefore rejected. However, both Sobol' and Scrambled-Sobol' methods resulted in a fitting of viable cell density and mAb concentration with $R^2 >> 0.95$. The accurate metamodel fitting was common for all examined parameter deviation values and resulted in a similar calculated TSI for each parameter and for both variables. Moreover, the different sampling methods for evaluating metamodel fitting had a minor impact on metamodel performance and TSI values.

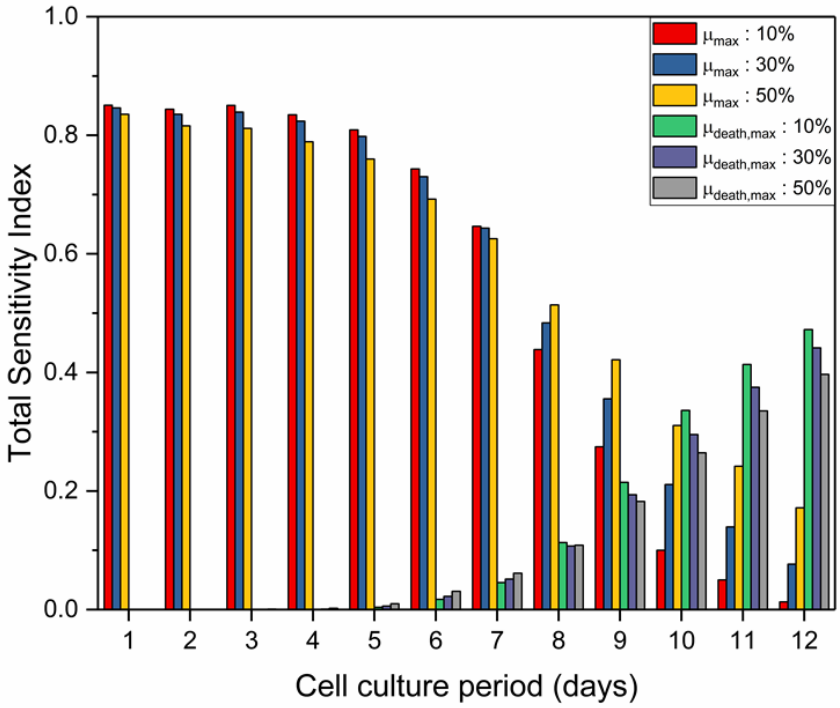

Figure 3.10: TSI values for μ_{max} and $\mu_{death,max}$ on viable cell density, during the cell culture period and for different parameter deviation ranges. Sobol' sequence was used as a sampling method. Taken with permission from Kotidis and Kontoravdi[518].

Interestingly, TSI values presented a highly dynamic profile during the cell culture period. Fig. 3.10 shows the TSI value variation of maximum specific cell growth rate (μ_{max}) and maximum specific cell death rate ($\mu_{death,max}$) for viable cell density, during the entire culturing period and every 24h. Similarly, Fig. 3.11 presents the variation of several additional parameters associated with cell growth and death in three different time points of the cell culture. More specifically, μ_{max} retained a relatively constant TSI value until day 4 of the cell culture, slightly decreased until day 7 and finally presented a steep decline until the termination of the cell culture, with a value between 0.013-0.017 for the different configurations. On the contrary, $\mu_{death,max}$ presented an entirely opposite profile with negligible TSI values until day 5 and then a rapidly increasing trend until the harvest of the culture, at which point the calculated TSI values were between 0.40-0.47. Interestingly, the lower parameter deviation range resulted in more extreme values and steep changes for both μ_{max} and $\mu_{death,max}$ during the cell culture period.

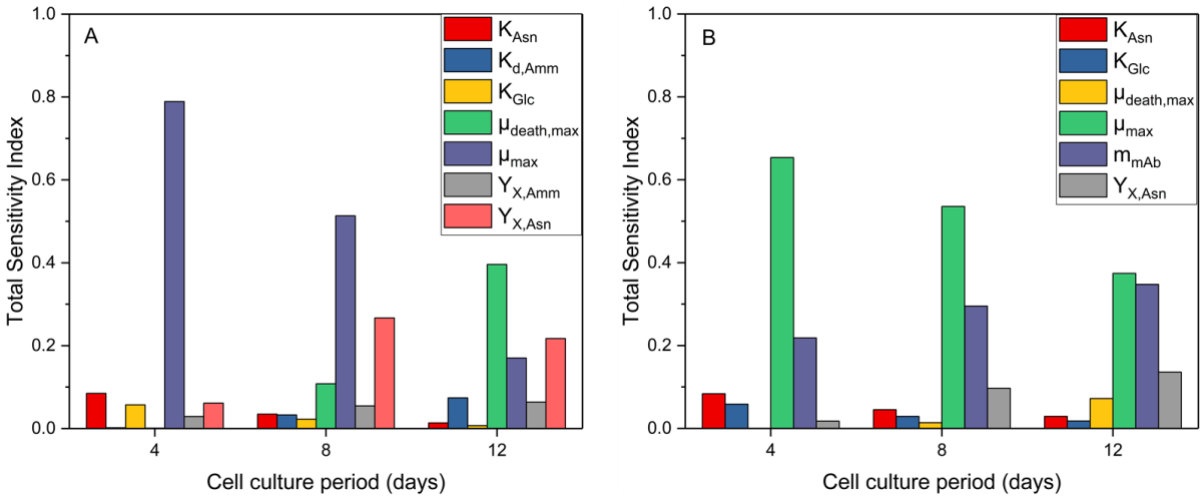

Figure 3.11: TSI values for different parameters with a TSI > 0.05 for at least one of the examined periods and for the effect of the parameters on (A) viable cell density and (B) mAb concentration. The parameters' deviation was set at 10% and the Scrambled Sobol' method was used for sampling. Taken with permission from Kotidis and Kontoravdi[518].

As shown in Fig. 3.11, parameters that are related to cell death, such as $K_{d,Amm}$ and $Y_{X_{Amm}}$, presented a steadily increasing TSI value for viable cell density during the evolvement of the cell culture. On the other hand, the TSI of K_{Asn} and K_{Glc} decreased over time for both viable cell density and mAb concentration, probably due to the decreased net cell growth in the late phases of the culture and therefore the reduced dependence of cells on asparagine and glucose. Interestingly, the TSI value of $Y_{X_{Asn}}$ for viable cell density increased during the end of the exponential phase and then slightly declined towards the end of the culture. The slight decrease of the TSI value during the latter stages of cell culture can be attributed to the stabilization of q_{Asn}, in combination with the increased ammonia contribution ($Y_{X_{Amm}}$). However, the TSI of $Y_{X_{Asn}}$ for mAb concentration presented a constantly increasing trend during the entire cell culturing period. Similarly to the case of viable cell density, the TSI values of μ_{max} and $\mu_{death,max}$ presented a declining and increasing trend, respectively, with the evolvement of the cell culture. Notably, $\mu_{death,max}$ presented a considerably lower TSI value for mAb concentration compared to viable cell density, probably due to the indirect effect that cell death has on mAb synthesis and concentration.

Post-analysis of GSA results and model calibration

As shown in Fig. 3.10 and 3.11, the TSI values of model parameters were found to significantly fluctuate during the cell culturing period. For that reason, Eq. 3.2 was employed for the evaluation of parameter significance during the entire experiment. Furthermore, as the identification of significant parameters is heavily dependent on the considered SIT and in order to examine the effect of the chosen SIT on the identification of significant parameters, we examined three commonly employed SIT values: 0.05, 0.1 and 0.2. As mentioned earlier, TSI values between Sobol' and Scrambled-Sobol'

methods were found to yield comparable results (<1% discrepancy) and therefore, analysis was performed based on the Sobol' sequence. Table 3.2 includes all the parameters that were found to exceed the SIT examined in each scenario. As expected, higher SITs resulted in fewer significant parameters. Interestingly, despite the influence of parameter deviation on TSI and the metamodeling fitting, all examined parameter deviation values eventually resulted to identical sets of significant parameters for each examined SIT.

Table 3.2: Results of parameter estimation for all parameters that were considered significant for each SIT. The dash indicates that the parameter was not found to significantly affect the outputs for the examined SIT. CI stands for confidence interval. Taken with permission from Kotidis & Kontoravdi[518].

	Set A (SIT=0.05)		Set B (SIT=0.1)		Set C (SIT=0.2)		
Parameter	**value**	**95% CI**	**value**	**95% CI**	**value**	**95% CI**	**Units**
K_{Asn}	2.66	0.24	-	-	-	-	mM
$\mu_{death,max}$	1.46×10^{-2}	2.92×10^{-3}	1.41×10^{-2}	2.46×10^{-3}	-	-	h^{-1}
μ_{max}	3.89×10^{-2}	1.14×10^{-3}	3.89×10^{-2}	1.14×10^{-3}	3.41×10^{-2}	7.24×10^{-4}	h^{-1}
m_{mAb}	1.07	5.10×10^{-2}	1.07	5.10×10^{-2}	1.13	5.46×10^{-2}	$pg\cdot cell^{-1}\cdot h^{-1}$
$Y_{X_{Asn}}$	3.46×10^{8}	2.79×10^{7}	3.46×10^{8}	2.79×10^{7}	-	-	$cell\cdot mmol^{-1}$

For each exercise, the significant parameters were re-estimated to fit the experiment with the lowest overall standard deviation for both viable cell density and mAb concentration (P1). Low standard deviation was chosen as a criterion because it accommodates more accurate fitting and better confidence. The estimated values for each parameter and the respective 95% confidence intervals (95% CIs) can be found in Table 3.2. The results were evaluated considering the fitting to the training experimental data and the calculated 95% CIs, with better fitting and narrower confidence intervals being the desired targets. The use of Set B for model training presented the lowest R^2 values (Fig. 3.12) and the narrowest 95% CIs overall (Table 3.2), apart from the 95% CI of μ_{max} that presented a lower CI in Set C. The strategy for parameter re-estimation included two steps: 1) estimation of μ_{max}, $Y_{X_{Asn}}$ and m_{mAb} with K_{Asn} and $\mu_{death,max}$ assigned to their nominal values and 2) estimation of K_{Asn} and $\mu_{death,max}$.

The aforementioned strategy was adapted depending on the set of parameters that were included in each scenario. As shown in Table 3.2, the absence of K_{Asn} in Set B resulted in almost identical values compared to Set A, with only $\mu_{death,max}$ presenting marginal differences. On the other hand, the considerably reduced parameter Set C resulted in a lower value for μ_{max} and slightly higher estimated value for m_{mAb}. Also, μ_{max} was calculated with the highest confidence in Set C, when compared with the other parameter estimation scenarios. However, the re-estimation of only μ_{max} and m_{mAb} was

not found adequate to correctly calibrate the model in the experimental data of the P1 experiment (Fig. 3.12). In all parameter estimations and for all examined parameters, the lower bound was set equal to zero and the upper bound equal to 3-fold the nominal parameter value, to enable a more extended search of the parameter value.

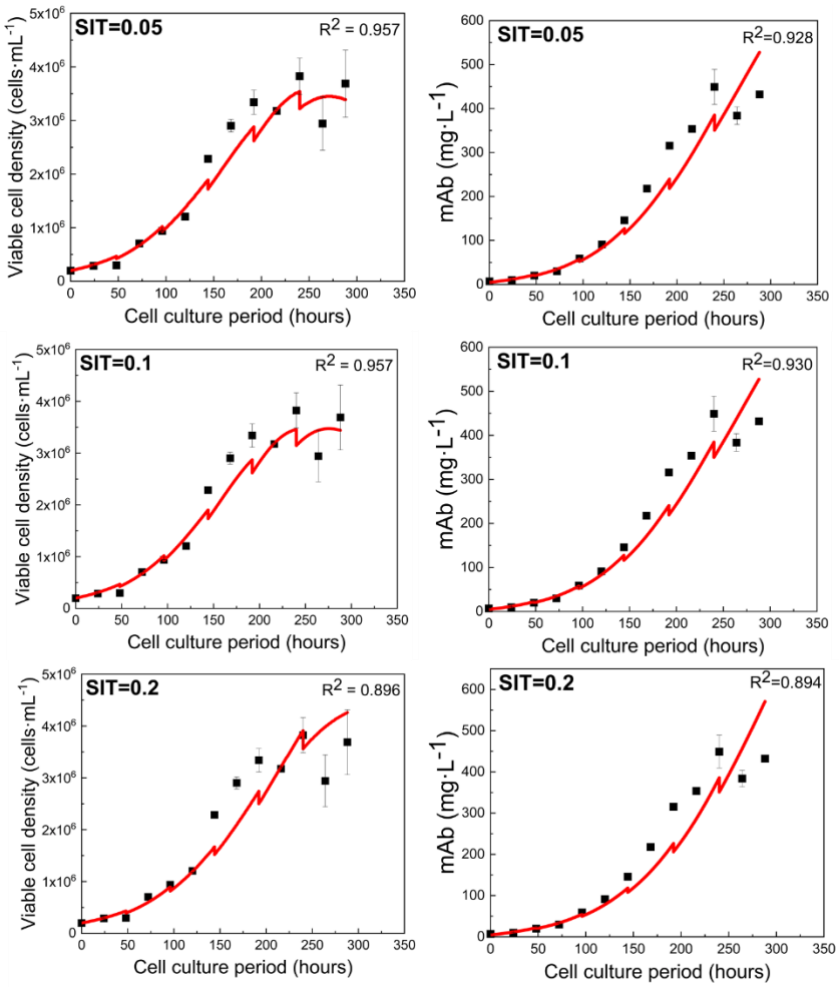

Figure 3.12: Fitting of the calibrated model (red line) to the experimental data (black squares) of the P1 experiment and for each SIT, over the entire cell culture period. The R^2 values of the fitting for viable cell density and mAb concentration are also reported.

Model predictive capabilities post re-calibration

The parameter values of Set B (Table 3.2) were utilized in order to examine the predictive capabilities of the model to the rest of the fed-batch experiments (P2-P5), that were not considered in model calibration. As described in section 3.1.1, P2-P5 are four fed-batch experiments that include the addition of different galactose and uridine concentrations in the culture from day 4 and every other day, with P5 being the control experiment with no feeding of galactose and uridine. In addition to galactose and uridine, all fed-batch experiments were supplemented with 10% v/v CD EfficientFeed™ C AGT™ Nutrient Supplement on day 2 and every other day.

Fig. 3.13 presents the agreement (R^2) between the re-calibrated model predictions and experimental data for the P1-P5 experiments. All experiments, and for both viable cell density and mAb concentration, were closely matched by model predictions that presented an agreement of $R^2 > 0.9$ in all cases. Viable cell density of P2 and P3 was slightly underpredicted but fell close to the standard deviation of the experimental data. The high R^2 values indicate a successful re-calibration of the model to the altered metabolic behaviour of the CHO-T127 cells.

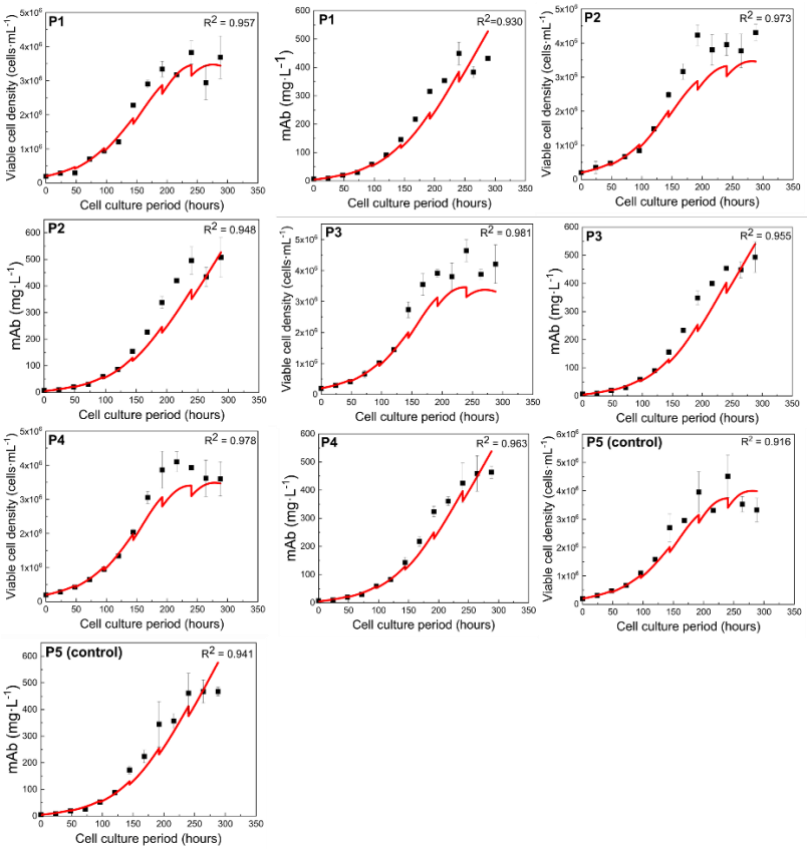

Figure 3.13: Accuracy of model predictions (red line) against the experimental data (black squares) for experiments that were not considered during model calibration (P2-P5). The fitting of the model to the training experimental data (P1) is also shown. The model performance is evaluated for the prediction of viable cell density and mAb concentration over the course of the cell culture period. Taken with permission from Kotidis & Kontoravdi[518].

3.2.4 Model training to a different CHO cell line (GS46)

The *Metabolism submodel* was subsequently trained on experimental data from the CHO-GS46 cell line that expresses a different IgG_4 monoclonal antibody. Whilst both cell lines are derived from CHO-K1 parental cells, the GS46 cells exhibit a significantly higher growth and productivity, achieving a mAb concentration of up to ~2.5 $g \cdot L^{-1}$ at harvest.

Specifically, the experimental dataset from a fed-batch culture of the GS46 cell line, with the addition of 10% v/v CD EfficientFeed™ C AGT™ Nutrient Supplement on day 2 and every other day, was used for applying the framework presented in Fig. 3.2 and performing the parameter estimation[519].

Galactose and uridine were not supplemented during the fed-batch culture. The basal media (CD-CHO) and the feed used for the fed-batch experiment were the same with the respective products that were used for CHO-T127 culturing. The availability of the complete experimental profile of amino acids and metabolites enabled the extension of GSA for the additional inclusion of glucose, glutamine, ammonia, lactate, asparagine, aspartate and glutamate as targeted outputs of the analysis. The incorporation of metabolites and amino acids led to the exclusion of viable cell density from the analysis due to the high correlation that it presents with the latter.

The optimal GSA settings for model calibration, as identified in section 3.2.3, were used for the identification of the significant parameters that are necessary for model calibration to the GS46 experimental data: Sobol' sampling method, 50% parameter deviation around the nominal value and SIT equal to 0.1. In addition, the GSA results were also processed with SIT equal to 0.05 and 0.2, respectively, in order to further explore the dependencies between the metabolites/amino acids and the model parameters (Fig. 3.14). Surprisingly, glucose only showed a weak dependency on $Y_{X_{Glc}}$ that was identified as a significant parameter only when SIT was set equal to 0.05. Moreover, aspartate was found to show relatively weak dependencies on parameters such as K_{Asn}, $Y_{Asp/Asn}$ and $Y_{Gln/Amm}$ that were labelled as significant only in the SIT=0.05 level. All metabolites/amino acids apart from lactate were found to be significantly dependent on μ_{max} up to the SIT=0.2 threshold. Interestingly and as a result of the exclusion of viable cell density from the analysis, $\mu_{death,max}$ was not labelled as a significant parameter for any of the examined outputs (including the monoclonal antibody concentration). The chosen sensitivity analysis identified 10 significant parameters for SIT equal to 0.1 and reduced the burden of parameter estimation from the original set of 33 parameters. The number of significant parameters identified in the current study are more than twice compared to the 4 significant parameters identified in section 3.2.3 due to the inclusion of more targeted outputs. The important parameters, as identified from the sensitivity analysis, can be found in Table 3.3.

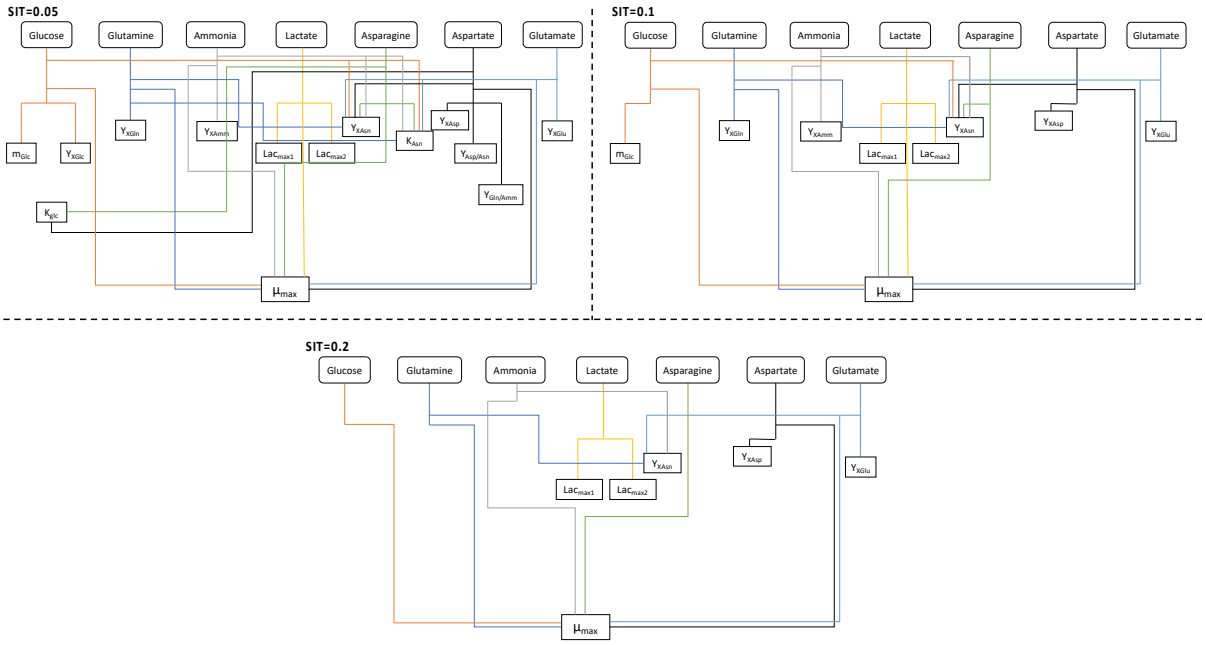

Figure 3.14: Parameters that were labelled as significant for each SIT and for each examined metabolite and amino acid. Each colour identifies the connections between a unique metabolite/amino acid and the identified significant parameters. Orange: glucose, blue: glutamine, grey: ammonia, yellow: lactate, green: asparagine, black: aspartate, light blue: glutamate. Taken with permission from Kotidis & Kontoravdi[518].

Importantly, due to the identified dependencies between the parameters and the examined model variables (outputs), a more structured and targeted parameter estimation strategy was employed. More specifically, each parameter was estimated using the experimental data of the variables for which it was labelled as significant. Briefly, the steps of parameter estimation are described below:

- *Step 1:* μ_{max} , $Y_{X_{Asn}}$, $Y_{X_{Amm}}$ and m_{Glc} were estimated against the experimental measurements of asparagine, ammonia and glucose.
- *Step 2:* $Y_{X_{Gln}}$ was estimated using the experimental measurements of glutamine.
- *Step 3:* $Lac_{max,1}$ and $Lac_{max,2}$ were estimated against the experimental data of lactate.
- *Step 4:* In a second round μ_{max} and $Y_{X_{Asn}}$ were re-estimated alongside with the estimation of $Y_{X_{Glu}}$ against the experimental data of glutamate. The estimations from Step 1 were used as initial guesses.
- *Step 5:* $Y_{X_{Asp}}$ was estimated against the experimental data of aspartate.
- *Step 6:* Steps 1-6 were repeated using the previously estimated values as initial guesses. In Step 1, in addition asparagine, ammonia and glucose, the experimental measurements of viable cell density were also included in the analysis.
- *Step 7:* Finally, m_{mAb} was estimated against the experimental data of the secreted monoclonal antibody concentration.

Whilst the proposed parameter estimation strategy includes several steps that are necessary in order to achieve the desirable fitting convergence, the extracted methodology is based on the GSA results and offers a well-structured and reproducible framework for estimating the significant parameters and adapting the model to a new cell line.

During parameter estimation, the non-significant parameters were assigned to their nominal values as calculated in Chapter 2. Table 3.3 shows the results of parameter estimation including the estimated values of the significant parameters and the respective 95% CI. The majority of significant parameters are the yields of cell biomass on the examined metabolite/amino acid ($Y_{X_{Metabolite}}$), with the exception of $Y_{X_{Glc}}$ and $Y_{X_{Lac}}$ that were not found to significantly influence the respective metabolites. Conversely, glucose presented a strong dependence on μ_{max} and m_{Glc}, while lactate was found to be strongly affected by the values of $Lac_{max,1}$ and $Lac_{max,2}$. Surprisingly, $Y_{X_{Asn}}$ was designated as an important parameter for asparagine, glucose, glutamate, ammonia and glutamine concentration, probably due to a correlation between the overall cell metabolism and asparagine levels. The latter was rapidly consumed by the cells during all culture phases, indicating a direct influence on cell growth and an indirect effect on other metabolites uptake or production. Finally, $\mu_{death,max}$ and K_{Asn} were not identified as significant parameters for any of the outputs, in contrast to the analysis presented in section 3.2.3, indicating that the effect of the aforementioned parameters lies directly on viable cell density.

Table 3.3: Results of parameter estimation for model adaption to the CHO-GS46 cell line. The 95% confidence intervals and the nominal values of the parameters for the CHO-T127 cell line, that were used as initial guesses for the GS46 parameter estimation, are also reported. Adapted from Kotidis & Kontoravdi[518].

Estimated parameter	CHO-T127 Nominal Value	GS46 Estimated Value	95% Confidence Interval	Units
$\boldsymbol{m_{mAb}}$	$4.10\text{x}10^{-1}$	1.31	$1.15\text{x}10^{-1}$	pg·cell^{-1}·h^{-1}
$\boldsymbol{Y_{X_{Asp}}}$	$1.46\text{x}10^{10}$	$1.15\text{x}10^{9}$	$6.44\text{x}10^{6}$	cell·mmol^{-1}
$\boldsymbol{Y_{X_{Glu}}}$	$8.25\cdot10^{7}$	$5.68\text{x}10^{9}$	$8.25\text{x}10^{7}$	cell·mmol^{-1}
$\boldsymbol{Lac_{max1}}$	21.20	18.55	4.02	mM
$\boldsymbol{Lac_{max2}}$	16	7.14	0.64	mM
$\boldsymbol{Y_{X_{Gln}}}$	$4.64\text{x}10^{9}$	$1.85\text{x}10^{10}$	$3.53\text{x}10^{8}$	cell·mmol^{-1}
$\boldsymbol{m_{Glc}}$	$3.43\cdot10^{-11}$	$3.35\text{x}10^{-11}$	$3.35\text{x}10^{-12}$	mmol·cell^{-1}·h^{-1}
$\boldsymbol{\mu_{max}}$	$6.50\cdot10^{-2}$	$6.96\text{x}10^{-2}$	$6.67\text{x}10^{-4}$	h^{-1}
$\boldsymbol{Y_{X_{Amm}}}$	$2.36\text{x}10^{9}$	$4.66\text{x}10^{9}$	$1.67\text{x}10^{8}$	cell·mmol^{-1}
$\boldsymbol{Y_{X_{Asn}}}$	$7.68\text{x}10^{8}$	$8.69\text{x}10^{8}$	$2.75\text{x}10^{7}$	cell·mmol^{-1}

The use of a common modelling framework for simulating the metabolism of two different CHO cell lines, enables the direct comparison between the obtained parameter values, and subsequently, the qualitative and comparative evaluation of the metabolic fluxes. Fig. 3.15 shows the fold change for the value of each of the parameters between the CHO-T127 and GS46 cell lines. Whilst a further statistical analysis (t-test) of the differences between the cell lines is not possible due to the nature of the comparison (estimated parameters) and lack of experimental replicates, the 95% confidence intervals, as calculated for each of the parameters, are included in Fig. 3.15. Accounting for the 95% CI, the cell lines show minor differences for the values of $Lac_{max,1}$, m_{Glc}, μ_{max} and $Y_{X_{Asn}}$. Interestingly, an approximately 3-fold higher m_{mAb} value is calculated for the GS46 cell line that depicts the higher specific productivity rates observed when compared to CHO-T127. Also, $Y_{X_{Amm}}$ and $Y_{X_{Gln}}$ are approximately 2-fold and 4-fold higher when compared to the respective values of CHO-T127, indicating that GS46 cells have a higher yield towards glutamine synthesis and ammonia production. On the other hand, $Y_{X_{Asp}}$, $Y_{X_{Glu}}$ and $Lac_{max,2}$ were estimated in lower values for the GS46 cell line indicating a lower flux of aspartate and glutamate towards cellular processes and a shorter period of lactate production from the CHO-T127 cells, as also experimentally observed.

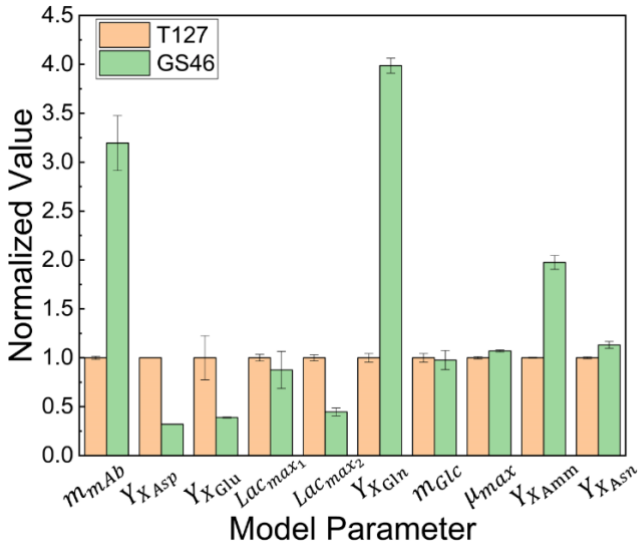

Figure 3.15: Comparison between the estimated values for the CHO-T127 and the GS46 cell lines. The values have been normalized against the nominal parameter values for the CHO-T127 cells. The error bars represent the 95% confidence intervals for each estimated parameter.

Model fitting to the experimental data for viable cell density, secreted mAb concentration, glucose, ammonia and asparagine is presented in Fig. 3.16. Apart from ammonia that showed an R^2 equal to 0.891, all variables presented an $R^2 > 0.95$ indicating a very good agreement between the model simulation and the experimental data. Notably, the model was able to adapt to the increased growth and productivity of the GS46 cell line. More specifically, the final titre for the GS46 cells was

approximately 2.5 g·L^{-1}, significantly higher compared to the ~0.5 g·L^{-1} achieved with the CHO-T127 cell line. In addition and in agreement with the trend indicated by the estimated value of m_{mAb}, the specific antibody productivity (q_{mAb}) for CHO-T127 and GS46 was experimentally measured at 17.4 pg·cell^{-1}·day^{-1} and 27.7 pg·cell^{-1}·day^{-1}, respectively. The model slightly underestimates the viable cell density at the time points of 192h, 216h and 240h of the cell culture period, and the ammonia concentration over the course of the experiment. Secreted mAb, glucose and asparagine concentrations are closely matched by the model over the entire cell culture period. More specifically, the ~3-fold higher value of m_{mAb} when compared to the concentration for CHO-T127 cells, in combination with the experimentally observed elevated levels of viable cell density, led to the accurate description of secreted mAb concentration over the course of the fed-batch culture.

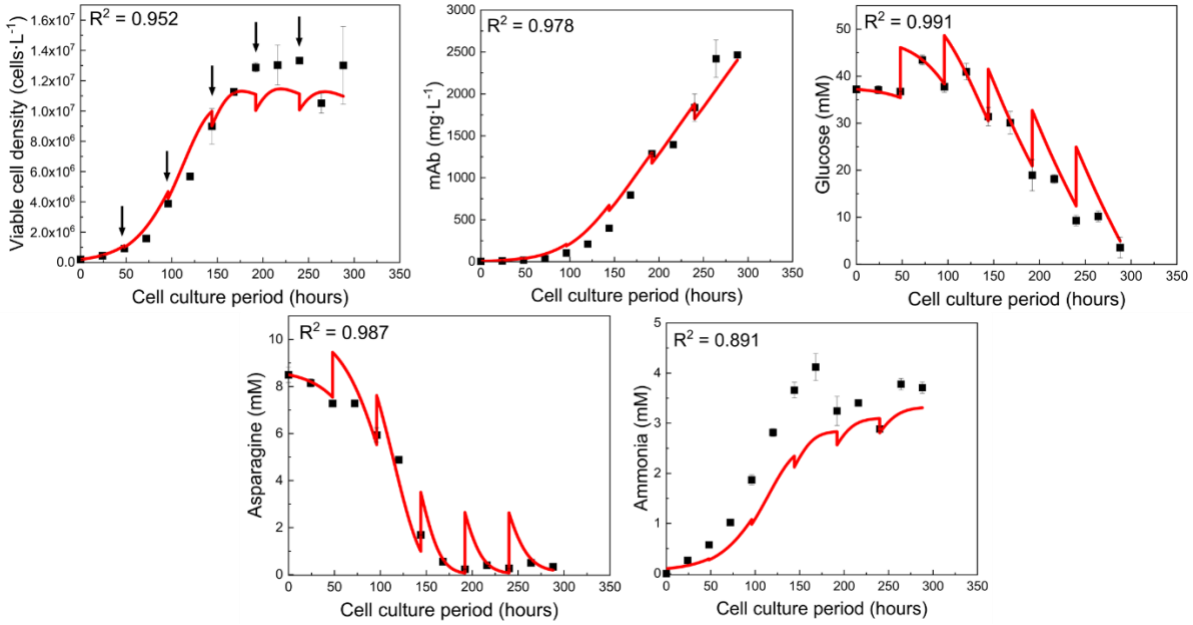

Figure 3.16: Model fitting to the experimental data of the GS46 cell for viable cell density, mAb concentration, glucose, asparagine and ammonia. Red line indicates model simulation data and black squares the experimental measurements. The black arrows in the graph of viable cell density indicate the time intervals at which 10% v/v CD EfficientFeed™ C AGT™ was added to the culture. Taken with permission from Kotidis & Kontoravdi[518].

3.3 Discussion

In Chapter 3 we attempted to showcase two applications of the proposed modelling framework that strengthen the importance of developing mathematical tools for describing bioprocesses, i.e., the production of recombinant proteins as described herein, in order to expedite process development, ensure that the quality attributes and key performance indicators meet the pre-determined requirements and, ultimately, reduce the costs of production and subsequently the price of the product.

The Quality by Design paradigm requires a deep understanding of the relationship between the inputs and outputs of the process. To that end, an integrated computational framework that utilizes constraint Global Sensitivity Analysis in combination with mechanistic model simulations, was proposed in sections 3.2.1 and 3.2.2, in order to identify an *in silico* Design Space of galactose and uridine concentrations in the feed that satisfy the desired quantity and quality constraints of the process outputs. cGSA was employed in order to reduce the dimensionality of the problem from an *8D* to a *3D* system by only accounting for the feeding parameters that significantly influence the constrained (Eq. 3.4 and 3.5) secreted mAb concentration and mAb galactosylation. The use of cGSA was found to outperform the unconstrained GSA with respect to the information provided for the effect of each feeding parameter on the outputs (Fig. 3.3). In order to generate data for employing the cGSA algorithm, the mechanistic model was simulated for 8,192 (2^{13}) scenarios of different galactose and uridine feeding concentrations in each interval. The integration of dimensionality reduction through cGSA and model simulations enabled the visualization of the process' Design Space (Fig. 3.4) and the further representation of the latter through a linear inequality constraint (Eq. 3.9). Five model-based experiments that fall within the DS and two experiments that fail to satisfy at least one of the two constraints (Eq. 3.4 and 3.5) were experimentally performed in order to validate the *in silico* DS. Finally, model predictions were verified based on the acquired experimental data, as shown in Fig. 3.6, indicating that the proposed mechanistic model can be used in order to define a Design Space that can ensure the satisfaction of the CQAs and KPIs, such as mAb concentration and galactosylation. The developed methodology contributes towards the utilization of computational tools with a minimal experimental dataset for identifying a suitable DS. A hybrid approach combining experimental and computational work is of great importance in problems that are characterized by multiple degrees of freedom (in this case the feeding concentrations at each interval) and that affect several outputs (i.e. cell growth and mAb concentration) in a time-dependent manner. An exhaustive exploration of the DS through solely experimental means would significantly increase the time and the associated costs. The inclusion of model-based data accelerates the process and offers mechanistic insights on the correlations between inputs and outputs.

Surprisingly, the $mGal$-constraint was satisfied in all experiments, indicating that the chosen threshold was not stringent enough and should be increased for future studies. More specifically, the considered $mGal$ is calculated as the product of mAb concentration and the absolute galactosylation and therefore incorporates the effect of process conditions on both variables. However, as mAb concentration is already constrained in Eq. 3.4, the $mGal$-constraint can be potentially replaced with a constraint on the absolute mAb galactosyation, as calculated by Eq. 2.46. Moreover, the discrepancies between model predictions and experimental measurements for $mGal$ are mainly

attributed to the underpredictions of the model for the mAb concentration (Fig. 3.6), as the glycoform distribution was accurately predicted for the majority of the experiments (Fig. 3.7). The latter observation further supports the replacement of the $mGal$-constraint with the absolute mAb galactosylation in relevant future experiments and studies. Another factor that could contribute to the discrepancies between the experimental data and model predictions is the regulation of glycosyltransferases at a cellular level and during the cell culture process. Whilst model predictions show a very good agreement with respect to the experimental mAb glycoform distribution at harvest, no measurements are available for the rest of the cell culture period to assess model agreement. However, the model has been proven (Chapter 2) to show very good agreement with the experimental mAb glycoform distribution for earlier days of the culture and therefore the aforementioned factor is not expected to significantly contribute to the observed discrepancies. Finally, mAb concentration is closely matched by model predictions with the exception of FS7 that, unexpectedly, considerably outperformed the control experiment. As discussed earlier, the model accounts for the inhibiting and toxic effects of galactose and uridine addition on cell metabolism and growth and therefore, an increase in mAb concentration is not expected. Overall, the model correctly identifies all the experiments that satisfy the constraints and fall within and out of the DS.

Further investigation of model performance against experimental measurements, revealed a disagreement with respect to the experimentally measured viable cell density (Fig. 3.8). The experimental results showed a considerable deviation between the training dataset (presented in Chapter 2) and the dataset used for DS validation (Chapter 3), even for the control experiment that was common for both sets of experiments. In order to calibrate the *Metabolism submodel* that simulates the viable cell density, a new set of fed-batch cultures was conducted, with varying concentrations of galactose and uridine in the feed. The experiments (P1-P5) were chosen in order to investigate the boundaries of the *in silico* DS.

The work presented in sections 3.2.3 and 3.2.4 attempts to (a) tackle problems caused in model calibration from high deviations in experimental variables of the cell culture and (b) enable the adaptation of the modelling framework to new datasets from different cell lines, products or process conditions. Initially, a strategic framework for the designation of the model's significant parameters, with respect to viable cell density and mAb concentration, was constructed in section 3.2.3 and shown in Fig. 3.1. Initially, the proposed workflow aims to identify the optimal settings for Global Sensitivity Analysis and verifies the results through parameter re-estimation against the experiment with the lowest experimental variation (P1). The effect of Sobol', Scrambled-Sobol' and Pseudo-random sampling methods and the deviation in parameter values by 10%, 30% and 50% on the targeted

outputs was evaluated. In addition, three values (0.05, 0.1 and 0.2) for the sensitivity index threshold, which determines the cut-off between significant and non-significant parameters, were investigated.

The results indicate that the Sobol' and Scrambled-Sobol' methods lead to similar (<1% discrepancy) metamodeling fitting and TSI values for the targeted outputs. Whilst parameter deviation was found to considerably influence the TSI values, it had minor effect on the resulting sets of significant parameters. The GSA settings of Sobol' sampling, 50% parameter deviation and SIT equal to 0.1 were found to identify a set of significant parameters (Table 3.2) that, when re-estimated, result in the highest agreement between training data and model simulations (Fig. 3.12). The re-estimated significant parameters for the optimal GSA settings were μ_{max}, $\mu_{death,max}$, m_{mAb} and $Y_{X_{Asn}}$. Model predictions for P2-P5 experiments using the re-estimated parameters, resulted in a very good agreement for both viable cell density and mAb concentration with $R^2 > 0.91$ for all experimental datasets (Fig. 3.13).

Following the successful calibration to new experimental profiles of the CHO-T127 cell line, the optimal settings for GSA were then applied for model adaptation to experimental data of the GS46 cell line (presented in Kyriakopoulos & Kontoravdi[519]). The GS46-dataset included a detailed profiling of all metabolites and amino acids of cellular metabolism and, therefore, enabled the extension of the targeted outputs and the inclusion of several additional model variables. GSA identified 10 parameters (Table 3.3) that significantly influence the targeted outputs, and consequently reduced the burden of parameter estimation for model adaptation, as the *Metabolism submodel* has 33 parameters in total. The re-estimation of the designated parameters resulted in a very good overall fitting (Fig. 3.16) scoring $R^2 > 0.95$ for both viable cell density and secreted mAb concentration. Moreover, the use of the same mechanistic model for simulating both cell lines' metabolism (CHO-T127 and GS46) enabled a more comprehensive understanding of the cellular discrepancies with respect to metabolism, based on the estimated parameters (Fig. 3.15). Finally, the insights from GSA also enabled the construction of a well-structured methodology for systematizing the parameter estimation process for model adaptation to a different cell line. Overall, the work presented in the current chapter successfully demonstrates that a single mechanistic modelling framework can be utilized, at least as a common platform, for different applications within the QbD and Industry 4.0 paradigms and for describing the metabolism of several cell lines and products. Finally, it is important to note that the model could require structural modifications if the growth-limiting substrates of the employed cell line are different.

Chapter 4 Artificial neural networks for predicting protein glycosylation

Chapter overview

As has been extensively discussed in Chapter 1 of this thesis, *N*-linked glycosylation is a post-translational modification of paramount importance for protein folding and function[135-137, 524], directly affecting the CDC and ADCC activity of therapeutic monoclonal antibodies[161, 169, 182, 525]. However, glycosylation is also a process of immense complexity, as it does not follow any genetic template. More specifically, the regulation of glycosylation in the cellular environment is dependent on two major factors: 1) the glycosylation machinery and 2) the glycoprotein structure. Glycosylation machinery includes all the enzymes and proteins associated with the process itself, such as glycosidases, GTases and NSTs. On the other hand, glycoprotein structure accounts for the steric hindrance that each specific protein imposes on the glycosylation enzymes. Therefore, both the cellular environment and the protein itself can strongly influence the resulting glycoprofile and glycans distribution.

The aforementioned multilevel control of the process increases the complexity of glycosylation modelling, especially when it is attempted through the use of mechanistic kinetic models. Several processes involved in glycosylation, such as the regulation of enzymes' expression and NSDs transport from cytosol to the Golgi compartments, exhibit a highly non-linear profile and complex mechanistic relationships. For example, NSTs and glycosyltransferases have been found to form complexes that introduce an additional level of complexity regarding the function and localization of these enzymes[123, 526]. To that end, low parametric approaches have been proposed in order to reduce the mechanistic knowledge necessary for describing the system and simultaneously reduce simulation times, either through probabilistic Markov Chain models[28, 497, 527], flux balance analysis[493] or statistical regression methods[502]. Model advances have been supported by the development of analytical methods for sensitive and high-throughput analysis of protein glycosylation using techniques such as nuclear magnetic resonance (NMR), liquid chromatography-mass spectrometry (LC-MS), matrix-assisted laser desorption/ionization-MS (MALDI-TOF-MS), MS/MS, high-performance liquid chromatography (HPLC) and capillary electrophoresis (CE)[528-530].

However, all the aforementioned modelling approaches, including mechanistic models, demand a considerable knowledge of both the computational tools and the biological background of glycosylation. The use of mechanistic models usually results in extended simulation times[531] and assumptions for model parameters, such as enzymes' concentration and distribution in the Golgi or

inhibition constants for the considered enzymatic reactions. Nominal values of the aforementioned parameters are usually derived from relevant *in vitro* kinetic studies of the glycosyltransferases. However, the *in vitro* enzymatic behaviour can be proven misleading when compared to the *in vivo* performance of the enzyme[532]. In addition, the results of parameter estimation during model's tuning are strongly dependent on the chosen initial parameter values and therefore, do not represent the global solution of the optimization problem, but just one of the several plausible solutions of the system. Consequently, the estimated values do not resemble the actual value of the parameters and therefore partially lose their mechanistic meaning. As shown in Chapter 2, an upregulation of the b4GalT levels in the optimized galactose and uridine experiment can significantly enhance model performance. Mechanistic models use a fixed concentration of the glycosylation enzymes and are therefore unable to capture the dynamic regulations of glycosyltransferases levels that are known to occur both in the proteomic[335] and transcriptomic levels[297, 304, 533] during the cell culture period. Additionally, the use of mechanistic, flux balance analysis and probabilistic models requires a detailed knowledge of the reaction rules and constraints, for the construction of the protein-specific glycosylation network. The considered glycosylation network can have a significant effect on the predictive performance of the model, especially on genetic engineering experiments, and the estimated values of model parameters.

In this chapter, Artificial Neural Networks are utilized in order to replace the kinetic *Glycomodel* and overcome all the limitations as previously described. In contrast to the previously proposed mechanistic[473] and probabilistic glycosylation models[496], ANNs require minimum knowledge of the biological background, no construction of a reaction network and reduced parameterization. Whilst the number of parameters remains high due to the existence of weights and biases, the only user-tunable parameter in the ANN configuration as used herein is the number of hidden layers and neurons. There is a limited level of control over the estimated values of weights and biases for the user, as the parameters are estimated during training. ANNs have been widely used in order to describe biological processes[534-536] and can readily adapt to the desired application. Up to date, ANNs have been used in order to predict glycosite location based on the amino acid sequence of the protein[537-539] and to describe cell cultures of mammalian[468, 540] and algal cells[456-457]. However, to the best of our knowledge, there have not been any efforts for utilizing ANNs, or alternative machine learning (ML) algorithms, for the prediction of protein glycosylation profiles.

Chapter 4 includes several studies that were carried out in order to identify the applicability and limitations of ANNs for predicting the glycoform distribution of several recombinant proteins that present increased structural complexity. Initially, ANNs were utilized for reliably describing the glycoprofile of a monoclonal antibody, subject to perturbations in cellular metabolism and using the

experimentally measured intracellular concentration of NSDs as inputs. The experiments used for this study were the same with the feeding regimes presented in Chapter 2, in order to enable the direct comparison between the two modelling approaches. Subsequently, an ANN was employed in order to describe the effect of manganese addition, a well known co-factor of several glycosyltransferases and more specifically of b4GalT, in combination with fucose and galactose on mAb glycosylation. The ANN was then incorporated in the mechanistic cell culture framework presented in Chapter 2, replacing that way the kinetic *Glycomodel*. The resulting hybrid (kinetic/ANN) model (HyGlycoM) was found to overperform the respective fully kinetic framework, additionally presenting a considerably reduced development and parameterization time. Similar to the fully kinetic framework, HyGlycoM utilizes only data from the feeding and initial concentration of specific metabolites and amino acids. Finally, a tunable configuration of ANNs was used in order to describe the glycoengineering effect of beta-1,4-galactosyltransferases' isoforms on the site-specific glycoprofile of complex fusion glycoproteins that carry 4-5 *N*-linked glycosylation sites. Trained on triple knockout experiments, the ANN model was able to calculate the effect of quadruple knockouts using the mRNA expression levels of the considered glycosyltransferases as inputs. Therefore, either as a standalone model or part of a hybrid configuration, ANNs can be proven a useful tool for the utilization of a vast range of inputs for accurately predicting protein glycosylation. The results of this chapter have been presented in Kotidis and Kontoravdi[541].

Overall, the aim of Chapter 4 is to provide answers to the following questions:

- Can machine learning algorithms, such as Artificial Neural Networks, be utilized to overcome the limitations of kinetic models in protein glycosylation?
- What are the applications of a machine learning model describing protein glycosylation?
- Can a hybrid (kinetic/data driven) model predict recombinant protein glycosylation?
- What are the advantages and disadvantages of using machine learning for describing cellular processes according to the results presented herein?

4.1 Materials & Methods

4.1.1. Cell culturing and protein glycosylation data

All experimental data used for model construction, training and validation were taken from literature [261, 350, 482]. The datasets for the initial ANN development in section 4.2.1, linking NSDs to the glycoprofile for an IgG product, are the same as in Chapter 2[482]. These datasets were also used to compare the performance of the ANN and hybrid configurations against the mechanistic model. Additionally, a dataset from Villiger et al.[261] evaluating the effect of manganese chloride, galactose and fucose addition on the NSD pools and the IgG glycoprofile, was utilized in order to evaluate the performance

of the ANN model on an alternative cell line and process regime. Finally, the Bydlinski et al.[350] dataset was employed in order to examine the applicability of ANN configurations on glycoproteins that demonstrate multiple glycosylation sites and increased complexity and versatility in their glycoprofile. The aforementioned dataset includes the glycoprofile of two fusion proteins: EPO-Fc and Fc-diamine-oxidase (Fc-DAO), in all glycosylation sites of the recombinant proteins and in several knockout scenarios of the b4GalT isoforms. Overall, four different glycoproteins (two different IgG, EPO-Fc and Fc-DAO) were considered for the studies presented herein.

4.1.2 Mechanistic-kinetic mathematical model

The kinetic model used in this chapter is the same with the one presented in Chapter 2 of this thesis. The parameters of the model were set to their nominal values, as estimated in Chapter 2. The model was simulated using the gPROMS ModelBuilder v.5.1.1 (Process Systems Enterprise, gPROMS, www.psenterprise.com/products/gproms, 1997-2020) and consists of three submodules: 1) *Metabolism submodel*, 2) *NSD submodel* and 3) *Glycomodel*. The latter was originally presented in del Val et al.[478].

4.1.3 Artificial Neural Network model construction

Python 3.7 was used for ANN construction, training, validation and testing. A general representation of the feedforward ANN used in this study is presented in Fig. 4.1. A typical neural network [542] consists of one or more hidden layers (HLs), each of which contains a number of nodes, or alternatively neurons. The HLs are normally placed between the layer of inputs and the respective layer of outputs. In order to follow the convention used within the machine learning community, the inputs will be referred to as *features* and the outputs as *labels*. Whilst the number of features and labels depends on the problem at hand, the number of HLs and neurons is subject to optimization. The number of HLs and neurons are the *hyperparameters* of the algorithm and the estimation of their optimal number is usually referred to as *tuning*. Each neuron is connected with all the neurons from the previous and next HL through the weights. Similarly, the features are connected to the neurons of the first and the labels to the neurons of the last HL. Each neuron carries a bias that together with the respective weight of each connection is used in order to calculate the contribution from each neuron of the previous HL through the *activation function*.

The feedforward ANN configuration utilized in the studies of this chapter, optimizes the value of the weights and biases through *backpropagation*. In backpropagation, the value of labels is estimated according to the initial value of weights and biases and is compared to the provided actual values. The difference between the estimated and actual values is then calculated (*loss function*) and is used in order to update the value of each weight and bias of the system. Each iteration is conventionally called

epoch, and the aforementioned iterative process (training) stops when either enough convergence has been achieved or a specific number of epochs has been simulated.

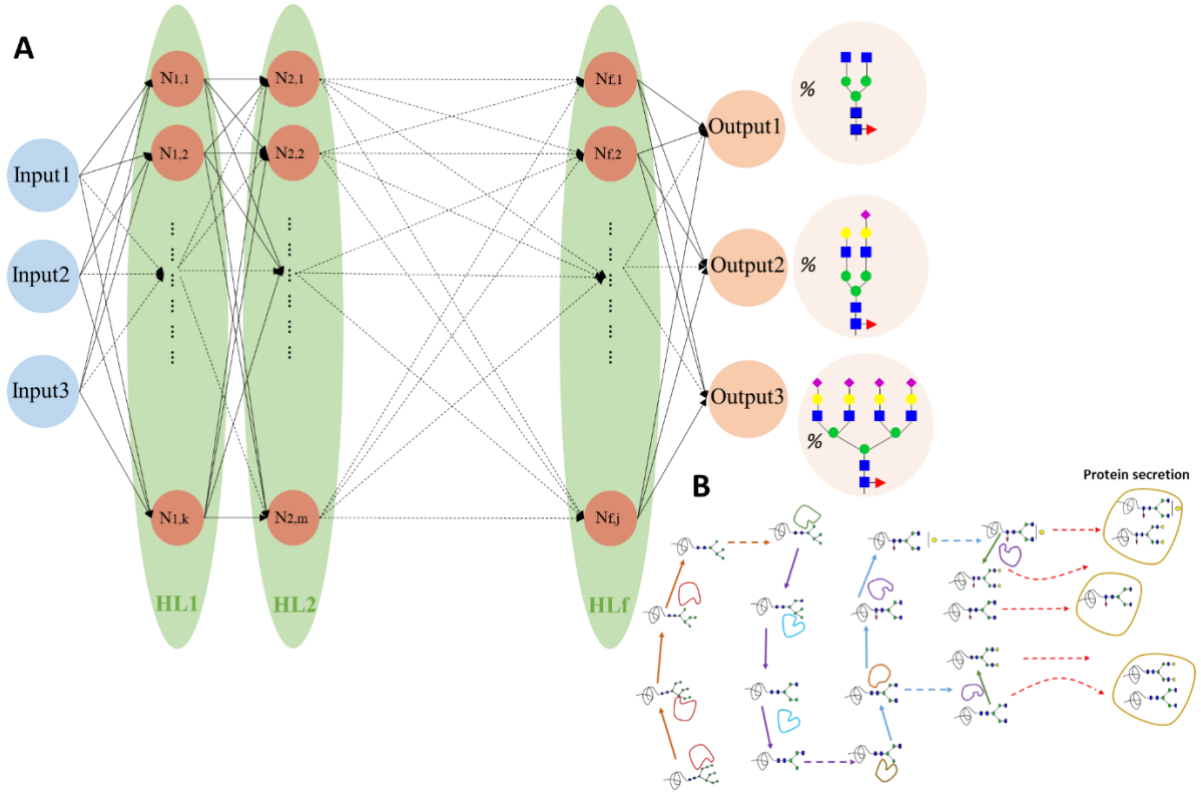

Figure 4.1: (A) Schematic representation of a feedforward Artificial Neural Network as used in this study. The depicted ANN consists of 3 inputs, 3 outputs, f hidden layers, and a variable number of nodes (neurons) in each of the layers. The outputs of the ANN are the distributions of each glycoform that are attached in the examined glycoprotein. The dashed lines are used to represent the connections between the nodes that are not shown in the graph. (B) Graphical representation of the *N*-linked glycosylation process in the Golgi apparatus. Arrows of different colour indicate the reactions taking place in each of the four Golgi compartments: cis (orange), medial (purple), trans (blue) and TGN (green). The dashed lines represent the transport of a glycoform from one compartment to another. Note that neither the structure of the ANN nor the reactions shown are the actual configurations considered in the current study. The graphs are shown just as a representation of the respective systems that were used herein. It is also important to note that the hidden layers of the ANN do not represent the Golgi compartments. Taken with permission from Kotidis & Kontoravdi[541].

For the ANN presented herein, the features utilized for each study depend on the dataset in use. The examined features in each study are presented in Table 4.1. In all cases, the labels of the ANN configuration were the distributions of the glycoforms considered for each examined glycoprotein.

Table 4.1: Utilized features for each study.

Study	Features (inputs)
Galactose and uridine effect on IgG glycosylation	Experimentally measured NSDs and nucleotides
HyGlycoM	Estimations of NSD levels based on mechanistic model simulations
Manganese, galactose and fucose effect on IgG glycosylation	Experimentally measured NSDs and cumulative extracellular manganese amount added
Effect of b4GalT isoforms knockouts on EPO-Fc and Fc-DAO glycosylation	b4GalT isoforms mRNA expression levels

Moreover, the list of glycoforms considered is *de novo* defined by the user. Regarding the configuration of the ANN, the sigmoid function was chosen for all layer connections as it has been successfully applied to relevant works for bioprocess modelling[543]. The Min-Max method was used for normalizing the features. The number of epochs was set to 20,000, apart from the configuration used for the manganese feeding experiments, where 2,000 epochs were found to be enough. The examined ANN configurations included two or three HLs.

The more the hidden layers and neurons included in the algorithm, the higher the probability of model overfitting. In addition, the inclusion of more than three HLs was found to significantly increase the network optimization time (hyperparameter tuning) without significantly improving model performance. Apart from controlling the number of neurons and HLs, alternative methods to avoid overfitting include the dropout, noise introduction, early stopping and weight constraining methods (regularization). More specifically, inclusion of dropout in the ANN configuration leads to neurons removal in a probabilistic manner during training in order to test the robustness of the model, while in noise introduction the user assigns an error distribution to the features and either creates additional points or deviates the original value of the existing variables. In early stopping, the model training stops when the desired convergence is achieved instead of running through the total number of epochs, and in regularization the weights of the configuration are constrained through a penalization of the loss function.

In addition, bias was set to zero for all the configurations presented herein. It is important to note that the hidden layers of the ANN do not represent the compartments of the Golgi and a relationship of that kind should not be assumed. During training, the hyperparameters of the ANN were tuned in order to minimize the error between model simulations and experimental data for the dataset of

interest (validation set), based on the strategy proposed in Del-Rio Chanona et al.[456]. The objective function of validation was set as the minimization of the sum of the absolute difference between simulation results and experimental measurements for the labels, as shown in Eq. 4.1.

$$OF = min \sum_i \left| EG_i - NG_i^{m,h1,h2,\dots,hm} \right| \quad (4.1)$$

where, OF is the objective function, i are the different glycans examined in the dataset, EG_i is the experimentally measured distribution value of each glycoform and $NG_i^{m,h1,h2,\dots hm}$ is the simulated distribution value of the i^{th} glycoform, for an ANN with m hidden layers and with $h1, h2, \dots, hm$ number of neurons for hidden layers $1, 2, \dots, m$ respectively.

The average absolute error (AAE) for each set of model predictions was calculated using Eq. 4.2.

$$AAE = \frac{\sum_i \left| EG_{i,k} - NG_{i,k} \right|}{n} \quad (4.2)$$

where, $EG_{i,k}$ is the experimentally measured distribution value of the i^{th} glycoform in the k^{th} set considered for training or prediction, $NG_{i,k}$ is the simulated or predicted distribution value of the model for the i^{th} glycoform in the k^{th} point, and n is the total number of points considered, calculated as the product of the total number of glycans and the total number of sets.

The predictive capabilities of each configuration were verified against independent experiments that were not used in either training or validation as described in each section of the Results. All the data and models that support this study can be found in: https://github.com/PK1617/ANN-glycosylation.

4.1.4 Multivariate analysis methods

OriginPro 2020 (OriginLab, Northampton, MA, USA) was used for the implementation of PCA and PLS.

4.1.5 Convention of glycans nomenclature

In contrast to the results presented in Chapter 2 and 3 of this thesis, Chapter 4 includes several complex glycans found in the examined fusion proteins. For that reason, a more comprehensive glycans symbology was applied for this chapter. The convention was adapted from Bydlinski et al.[350], where the fusion protein datasets were presented. The nomenclature for the involved sugars is presented in Table 4.2. Briefly, the abbreviation of each glycan includes the terminal sugars for each antenna and the presence or absence of fucose. Excluding fucose, the number of sugars in the glycan abbreviation is also an indicator of the antennae number. For example, the “NaGnGnF” glycan is a tri-antennary fucosylated glycan that carries a terminal N-Acetylneuramic acid and two terminal N-Acetylglucsamine molecules. The location of each terminal sugar on specific antennae was not included in the original study, neither the existence of polyLacNAc elongations, and was therefore not

considered in the symbology of this study unless explicitly mentioned. Finally, the conventional symbology does not specify the nature of the glycosidic linkage between sugars.

Table 4.2: Symbols for each sugar in the conventional symbology used in the study presented in Chapter 4.

Sugar	Conventional Symbol
Mannose	M
N-Acetylglucosamine (GlcNac)	Gn
Galactose	A
Fucose	F
N-Acetylneuramic acid (Neu5Ac)	Na

4.2 Results

The ANN approach was applied to data from three different studies and four recombinant proteins in total. All proteins were produced in CHO cells. The first dataset for IgG-producing cells presented in section 4.2.1 was taken from Chapter 2 of this thesis (can be also found in Kotidis et al.[482]) and was used for the construction of the hybrid model. The second dataset presented in section 4.2.2 includes the addition of manganese, fucose and galactose in CHO cell culture for the manipulation of IgG glycosylation, and was taken from Villiger et al.[261] Fucose and galactose are precursors of GDPFuc and UDPGal, respectively, that act as co-substrates in protein glycosylation. Manganese is a co-factor of several glycosyltransferases and is expected to enhance the activity of b4GalT. The last dataset presented in section 4.2.3 was taken from Bydlinski et al.[350] and includes the results of several knockouts of the four b4GalT isoforms in the glycosylation profile of two fusion proteins: Fc-DAO and EPO-Fc. Notably, Fc-DAO carries five *N*-linked glycosylation sites, whilst EPO-Fc four. Site-specific ANNs were developed for each examined glycosylation site.

4.2.1 A hybrid model describing cell metabolism and *N*-linked glycosylation

The hybrid model presented herein, consists of three submodules, similar to the fully mechanistic model constructed in Chapter 2. In order to evaluate the performance of ANNs for predicting protein *N*-linked glycosylation, a standalone ANN was first developed. Following ANN validation, the data-driven model was integrated with the *Metabolism* and *NSD submodels* presented in Chapter 2.

Establishing an ANN model to describe IgG N-linked glycosylation

The experiments used in this study are described in Table 2.1 and include the addition of galactose and uridine to fed-batch CHO cell cultures for the increase of targeted NSD pools, such as UDPGal, in the intracellular environment, and consequently the enhancement of IgG galactosylation. Nucleotide

and NSD concentrations were considered as features of the ANN. All available time points (days 7, 9, 11 and 12) were used for model training. The training set includes the control, 10G, 10G20U and 50G5U experiments. The 10G5U experiment was left out of the training set and was used for validation. The latter was chosen for validation as it was the only dataset found within the training space defined by the rest of the experiments with respect to the concentration of both galactose and uridine. A total of 11 datasets with 12 features (AMP, ADP, ATP, CTP, UTP, GTP, UDPGalNAc, UDPGlcNAc, UDPGal, UDPGlc, GDPMan and GDPFuc) was used for model training. CMPNeu5Ac was not included as no sialylation was reported for the IgG of this study. The aforementioned variables were chosen as features in order to resemble the underlying principles of the mechanistic *Glycomodel* that utilizes the outputs of the *NSD submodel* to predict glycans distribution.

Model validation results for all time points of the 10G5U experiment are shown in Fig. 4.2. ANN tuning through the validation set resulted in a configuration with 2 HLs and with 22 and 18 neurons in the first and second layer, respectively. The inclusion of a third HL was found to only marginally improve ANN performance and was therefore dismissed to avoid the risk of overfitting. The configured ANN closely describes the validation experiment, with the value of GnGnF on day 11 of the culture presenting the highest absolute error of ~4.1%. The AAE, described in Eq. 4.2, was calculated at 0.87%, showcasing the close agreement between experimental measurements and ANN results.

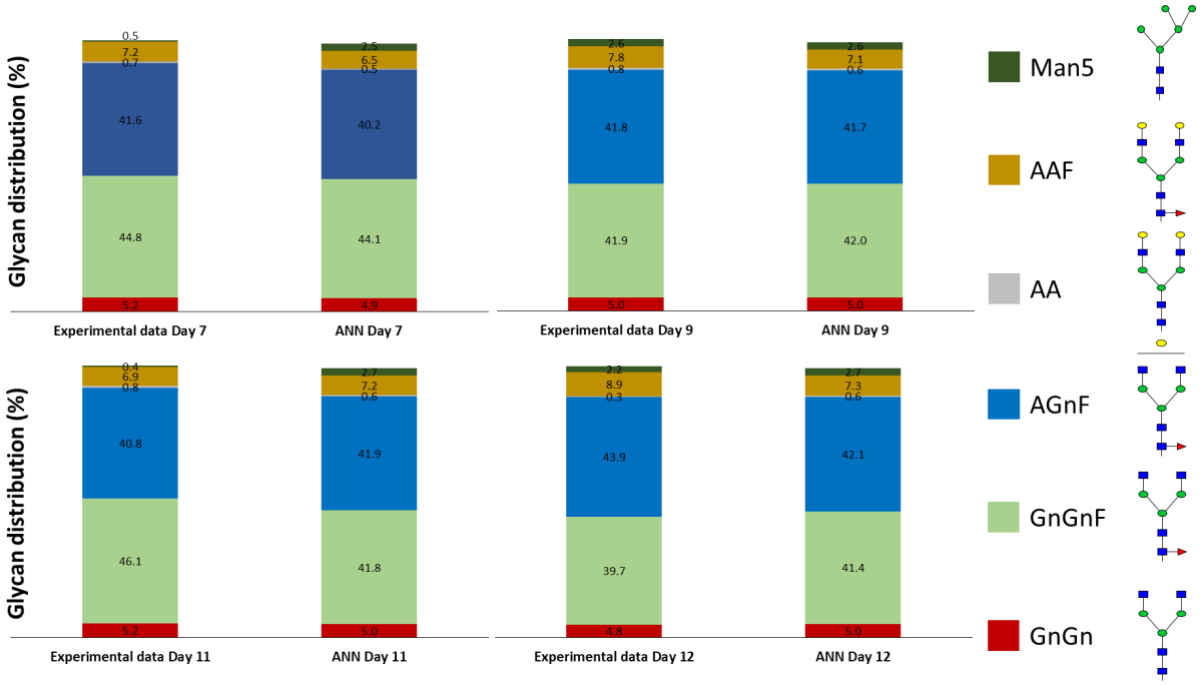

Figure 4.2: Comparison between the experimental measurements of the validation dataset (10G5U) and the experimental measurements for each day of the cell culture period. Taken with permission from Kotidis and Kontoravdi[541].

In order to further explore the robustness of the ANN configuration, the ANN was trained on different combinations of training sets, following an adapted leave-one-out-cross-validation (LOOCV) methodology, and was validated (tuned) against the experiment that was, each time, held out from training. The alternative ANN training resulted in good agreement with the independent experiments in each case, as shown in Fig. 4.3.

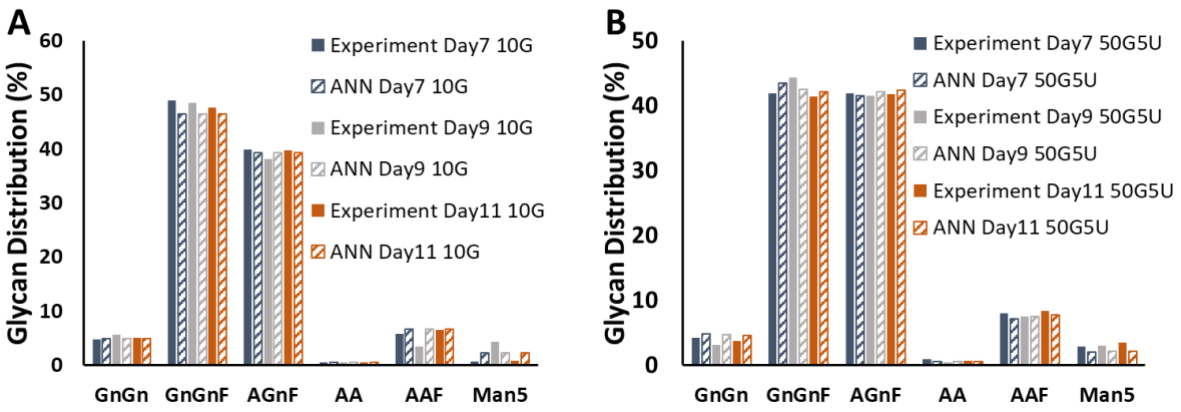

Figure 4.3: Results of ANN training to alternative training sets. The ANN results are compared with the validation experiment that was held out of the training set in each case: (A) 10G and (B) 50G5U. The experiments used for validation for each scenario were replaced with the 10G5U dataset in the training set. Taken with permission from Kotidis and Kontoravdi[541]. The ANN configuration for each case is: (A) 26/5 and (B) 20/7, for the first and second layer respectively.

In an additional effort to rigorously investigate the effect of each feature on ANN performance and whether the algorithm developed strong inadvertent dependencies on specific variables, the features were, one-by-one, excluded from the model. The number of neurons remained at 22 and 8 in the first and second HL, respectively, for all simulations. The effect of each exclusion on model performance against the 10G5U experiment was subsequently recorded. The AAE ranged between 0.87% for the full dataset, up to 1.25% when ADP was excluded from the feature-set (Fig. 4.4), indicating no excessive dependency of the ANN to any of the utilized features.

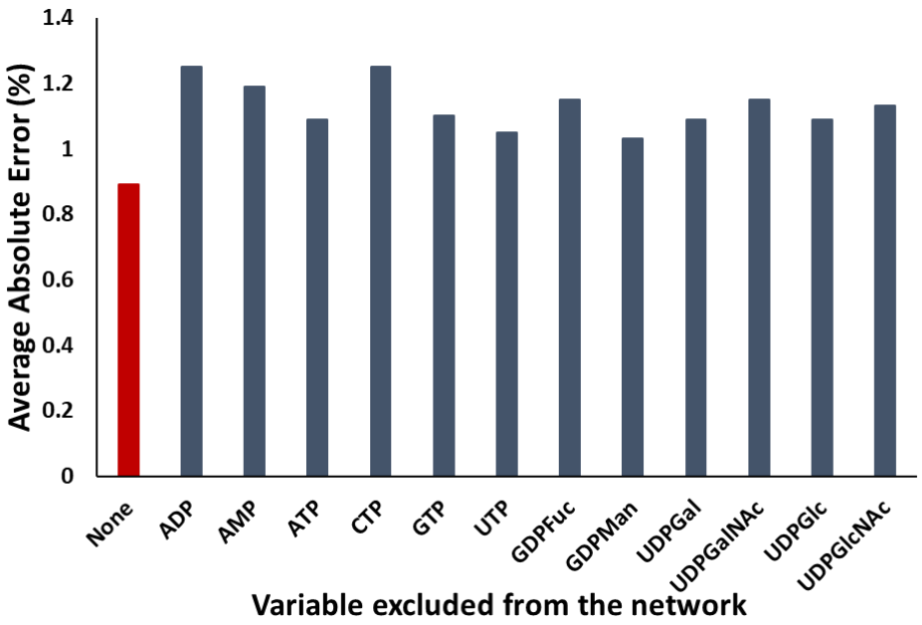

Figure 4.4: Effect of the one-at-a-time feature exclusion on the AAE of the ANN. The red bar shows the AAE of the scenario where no variable was excluded from the ANN. Taken with permission from Kotidis and Kontoravdi[541].

In addition, PLS was employed in order to compare ANN performance with alternative data-driven approaches. PLS is a computationally efficient multivariate method that requires minimal parameter tuning from the user and has been previously utilized to successfully describe mAb glycosylation based on several process parameters[502]. For that reason, a PLS model was trained on the control, 10G, 10G20U and 50G5U experiments, similar to the ANN. Fig. 4.5 presents the results of the ANN-PLS comparison against the experimental data for the 10G5U. The ANN was found to outperform the PLS prediction in the vast majority of data points and for all glycan structures. Exception to the ANN outperformance was the distribution of GnGnF on day 11 of the culture, where the PLS presented a closer value to the experimental data. The AAE for the PLS model was 1.66%, almost twice compared to the respective value for the ANN.

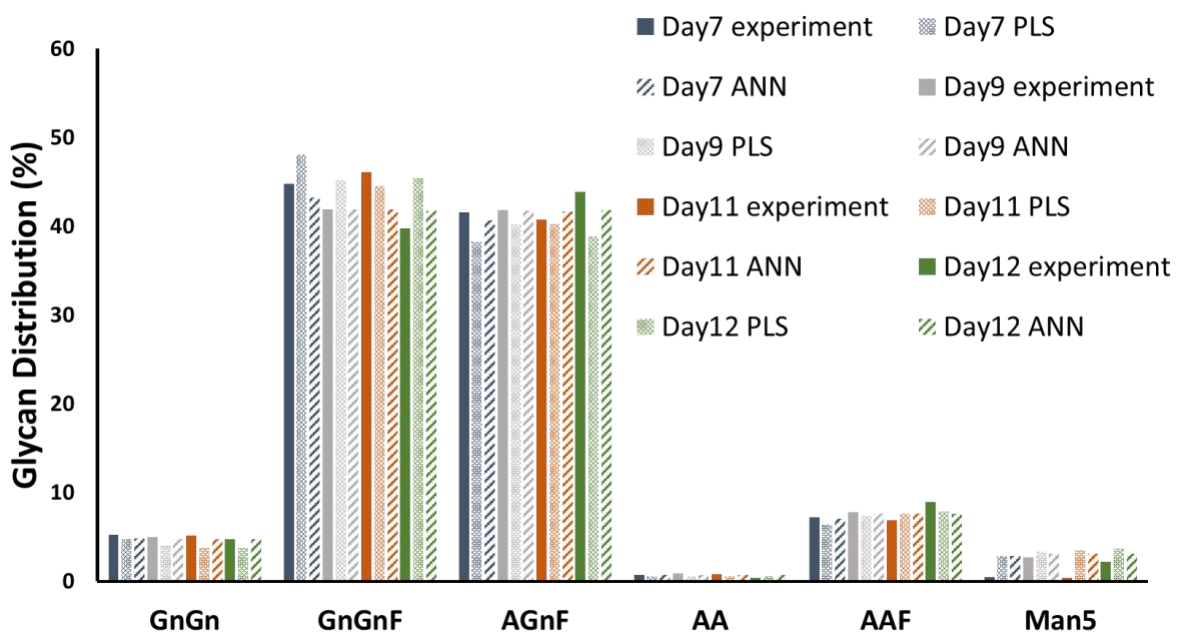

Figure 4.5: ANN and PLS performance for the 10G5U dataset compared to the experimental data. Taken with permission from Kotidis and Kontoravdi[541].

Hybrid Glycosylation Model (HyGlycoM)

Following the configuration and validation of the ANN model, the latter was coupled with the kinetic *Metabolism* and *NSD submodels*. Essentially, the ANN was used in order to replace the *Glycomodel* from the fully mechanistic model. A schematic representation of HyGlycoM is shown in Fig. 4.6. The hybrid nature of the model enables the utilization of the metabolites and amino acids concentration in the feed and seeding media through the mechanistic submodels that, in turn, calculate the intracellular NSD concentrations that are used as inputs for the ANN algorithm.

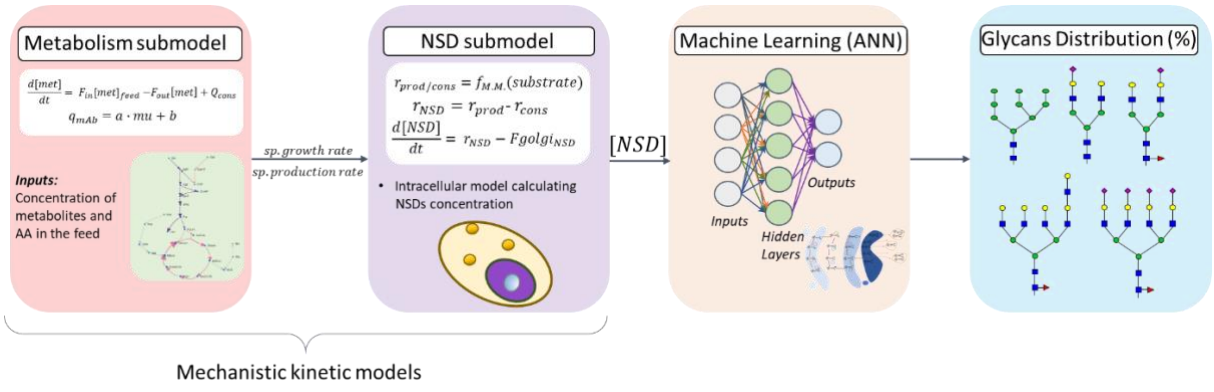

Figure 4.6: Schematic representation of HyGlycoM that is composed from the *Metabolism submodel*, *NSD submodel* and the ANN. Adapted with permission from Kotidis and Kontoravdi[541].

Similar to the previous study, the ANN part of HyGlycoM was trained using the estimated NSD concentrations for the control, 10G, 10G20U and 50G5U experiments, based on the simulation results

of the *Metabolism* and *NSD submodels*. The 10G5U experiment was used for ANN tuning during validation. The resulting ANN configuration consists of 2 HLs with 18 and 34 neurons in the first and second layer, respectively. Fig. 4.7A shows the comparison between the simulated and experimental values for the NSDs concentration, as calculated by the mechanistic submodels. HyGlycoM was able to absorb the inaccuracies of the mechanistic submodels in the estimation of the NSD concentration due to the adaptability of the ANN, therefore achieving a good fitting for the validation experiment (10G5U), as shown in Fig. 4.7B. Importantly, due to the nature of data-driven techniques, ANNs can show increased tolerance of inaccuracy in features' values as long as the qualitative differences between the data points are correctly described (Fig. 4.7B). The AAE of HyGlycoM was 0.98%, slightly higher compared to the standalone ANN (0.87%). It is important to note that the experimental quantification of the variables used as inputs for the HyGlycoM model (metabolites and amino acids concentrations in the media and feed) is considerably simpler and more high-throughput when compared to the inputs of the standalone ANN (intracellular NSD and nucleotide concentrations). Therefore, the ease of access to the necessary experimental inputs can compensate for the increased AAE of HyGlycoM when compared to the standalone ANN.

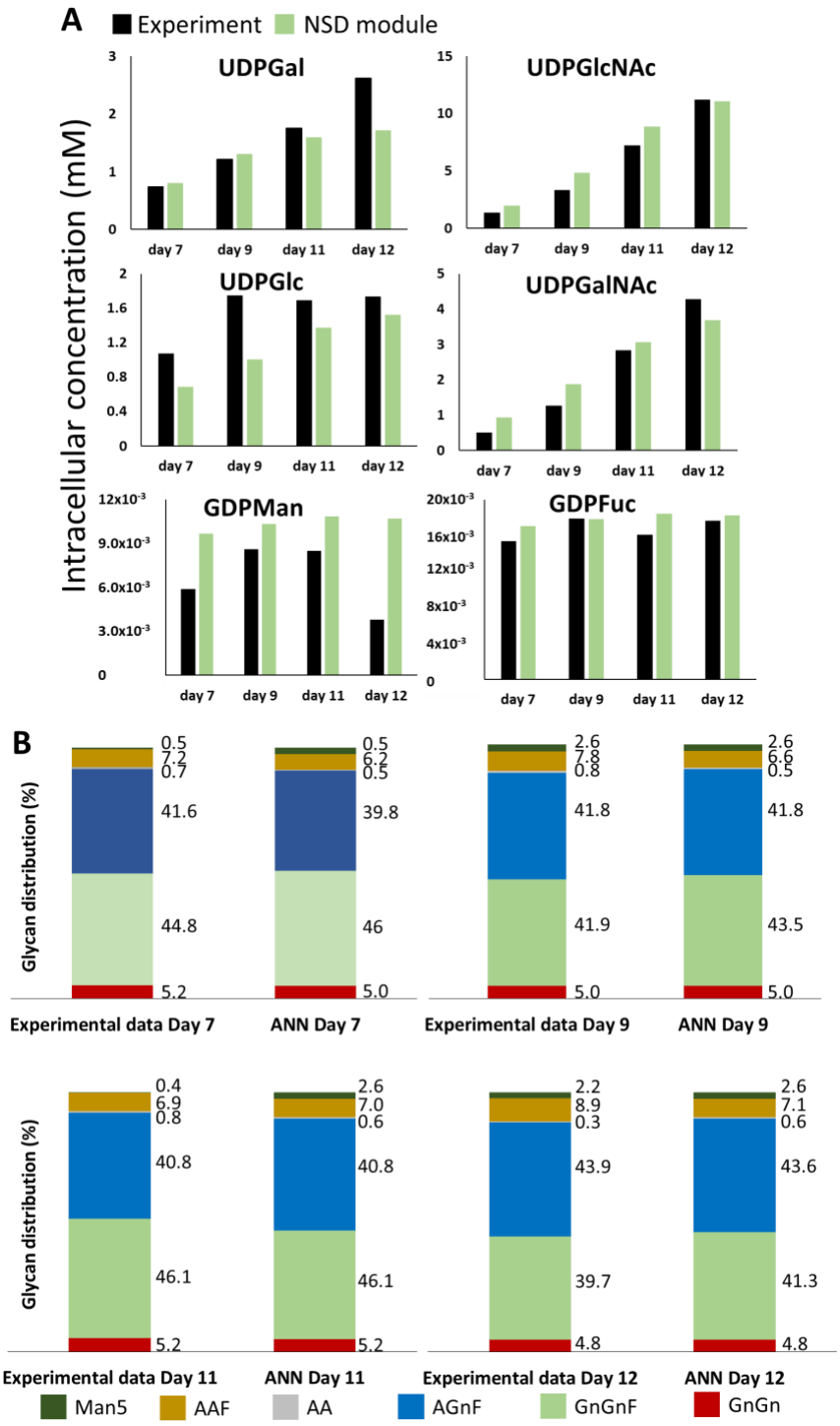

Figure 4.7: (A) Comparison of the mechanistic submodels simulations with the experimental data for the concentration of intracellular NSDs. (B) Comparison between HyGlycoM and experimental measurements for the glycans distribution of the 10G5U experiment. Taken with permission from Kotidis and Kontoravdi[541].

Comparison between HyGlycoM and the fully mechanistic model

As a next step, HyGlycoM was compared to the fully mechanistic model. The comparison of model performance was based on the predictions for the glycoform distribution of the model-based

optimization experiment, presented in Chapter 2 and Table 2.1. Fig. 4.8A demonstrates the results of the comparison between the fully kinetic model and HyGlycoM against the experimental measurements. With the exception of GnGnF distribution for days 7 and 9 of the cell culture, HyGlycoM outperforms the respective kinetic model for the vast majority of glycans species and in all time points. Importantly, HyGlycoM describes the galactosylated glycans (AGnF, AA, AAF) with considerably higher accuracy when compared to the fully kinetic model. The latter observation is attributed to the capability of HyGlycoM to capture, through the ANN part of the model, the non-linear regulation of glycosyltransferase expression and NSD transport to the intra-Golgi environment, which are not explicitly considered in the fully kinetic model. Monoclonal antibody galactosylation is of particular interest within this thesis and the aforementioned observations can be considered a significant improvement towards overcoming the limitations of kinetic models and extending model applicability, as described in Chapter 3. The IgG glycoprofile consists mainly of the fucosylated GnGnF and AGnF glycans, as shown in Fig. 4.8A, the abundance of which ranges between 37.6% to 53.7% and 34% to 44.4%, respectively, within the training set. Overall, HyGlycoM presented an AAE of ~1.25% for the independent optimization experiment, improving the results by 30% when compared to the fully mechanistic model. HyGlycoM was finally compared to a PLS model trained with the same datasets. Comparative PLS predictions for the independent optimization experiment are shown in Fig. 4.8B. HyGlycoM was found to considerably outperform PLS predictions for all glycan species and available time-points.

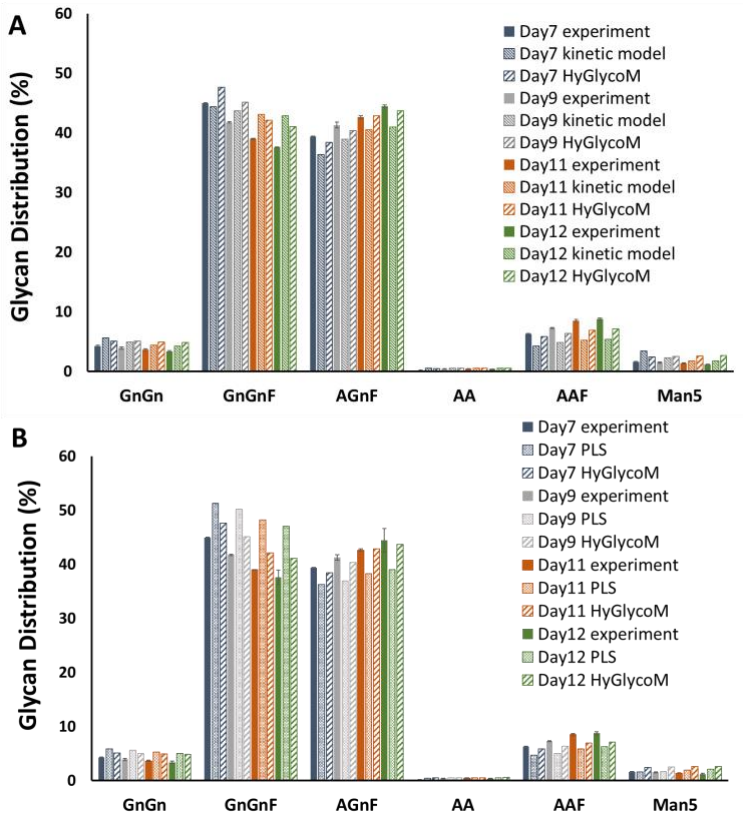

Figure 4.8: Comparison of HyGlycoM with (A) the fully kinetic model and (B) the PLS model, against the experimental measurements for the independent optimization experiment. Taken with permission from Kotidis and Kontoravdi[541].

4.2.2 Extending the ANN for predicting the effect of metal ion supplementation on IgG glycosylation

Metal ions are well known co-factors of glycosyltransferases, critically affecting the activity of the enzymes [544]. Several studies have investigated the effect of manganese chloride ($MnCl_2$) in CHO cell culture, identifying a positive correlation with recombinant protein galactosylation [261, 292-293]. To that end, the extracellular concentration of manganese has been incorporated in mechanistic models for the description of protein glycosylation in CHO cell cultures [235, 481]. Following a similar approach, the cumulative amount of manganese supplemented to the culture was integrated in the standalone ANN model proposed in section 4.2.1, in order to capture the effect of manganese on IgG-glycosylation. The experimental data utilized for this study are presented in Villiger et al.[261]. The authors evaluated the effect of $MnCl_2$, fucose and galactose addition on an IgG producing CHO-S cell line cultured in 10mL bioreactors, with a downwards shift in pH and temperature introduced on day 5 of the cell culture period. The IgG glycosylation profile of the Fc-region was analysed on day 17 (harvest point) of the culture. More details on the $MnCl_2$, galactose and fucose concentrations added in the culture can be found in Table 4.3. The experiments in the current section were performed with CHO-S cells, whereas the dataset employed in section 4.2.1 was generated using CHO-T217 cells, which were derived from a CHO-K1 parental cell line. In addition, the CHO-S cells were grown under considerably different

conditions: 1) pH (from 7.1 to 6.9) and temperature shift (from 36.5 °C to 33 °C) on day 5, 2) total cultivation duration of 17 days, 3) total volume of 10mL in bioreactors, 4) pH control through CO_2 supplementation and 5) feed addition on days 3, 5, 7, 10 and 14. Despite the efforts presented in Chapter 3 for adapting an established mechanistic model to a different cell line, such discrepancies in process conditions could lead to inevitable alterations in the structural backbone of the mechanistic model. However, the data-driven ANN enables a more high-throughput and equally reliable approach for simulating the new process conditions, as model calibration is automatically performed during training of the algorithm and minimal intervention is required from the user.

Table 4.3: Experimental details as presented in Villiger et al.[261]. The reported concentrations were supplemented to the cell culture on days 3, 5, 7, 10 and 14. Abbreviation: M: manganese; G: galactose; F: fucose. Taken with permission from Kotidis and Kontoravdi[541].

Experiment	$MnCl_2$ (μM)	Galactose (mM)	Fucose (mM)
M0G0F0	0	0	0
M2.5G0F0	2.5	0	0
M0G6F0	0	6	0
M0.25G6F0	0.25	6	0
M0G12F0	0	12	0
M0G18F0	0	18	0
M2.5G18F0	2.5	18	0
M0G6F1	0	6	1
M0.25G6F4	0.25	6	4
M2.5G6F8	2.5	6	8

The ANN configuration was adapted to include the cumulative manganese amount (in the form of $MnCl_2$) supplemented to the cell culture, in addition to the intracellular concentration of nucleotides and NSDs. The effect of galactose and fucose addition was reflected on UDPGal and GDPFuc intracellular levels, respectively, and the sugars' extracellular concentration was therefore not included in the list of features. The M2.5G6F8 experiment was excluded from the training and validation analyses and was used for the evaluation of model predictive capabilities (test set). A PCA performed on the available data indicated that the M0G6F1 dataset did not cluster with any other experiment, based on the measured values of the features (Fig. 4.9D). Thus, M0G6F1 was not used in the training process and was used for ANN validation and tuning. ANN tuning resulted in a configuration with 2HLs and 8 and 4 neurons in the first and second layer, respectively.

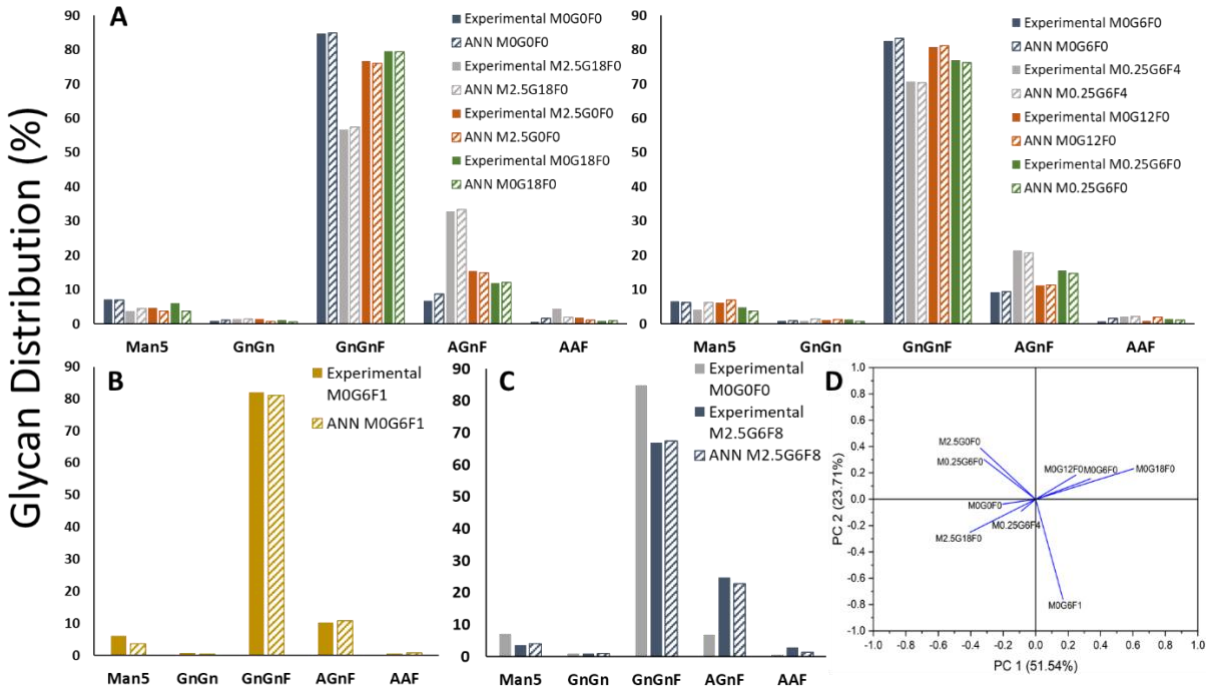

Figure 4.9: (A) ANN training results. (B) ANN fitting to the experiment used for model validation. (C) Comparison between ANN predictions and the experimental data of the M2.5G6F8 dataset. The experimental data of the control (M0G0F0) dataset are also displayed for comparison. (D) PCA performed in the available experiments for the identification of correlations between datasets. Taken with permission from Kotidis and Kontoravdi[541].

As shown in Fig. 4.9A, the ANN closely matches the glycosylation profile of all experiments used for training. Additionally, the model presents good agreement with both experiments employed for validation and prediction, respectively (Fig. 4.9B & 4.9C). Importantly, the effect of manganese and galactose addition on GnGnF and AGnF levels is correctly captured by the ANN, as presented in Fig. 4.9C and the AAE that is calculated at ~0.89%. Moreover, the model correctly identifies the underlying relationship between manganese and galactose. Similar to fucose addition, where no significant differences between the control and feeding experiments are observed, galactose and manganese have minor impact on the formation of AGnF when separately supplemented at the concentration of 12mM (M0G12F0) and 2.5 μM (M2.5G0F0), respectively. However, the experiment (M2.5G12F8) used for the evaluation of ANN predictive capabilities, results in considerably higher AGnF levels when compared to the combined increase of the M2.5G0F0 and M0G12F0 experiments. Galactose supplementation is expected to build up the intracellular UDPGal concentration and enhance AGnF synthesis. Thus, the - even in low levels - increase of AGnF distribution in the M0G12F0 and M0G18F0 experiments showcases the limitations of IgG galactosylation due to substrate availability. On the other hand, manganese acts as a co-factor of b4GalT and positively affects its enzymatic activity. The elevation of AGnF distribution in the M2.5G0F0 indicates a further limitation of IgG galactosylation on the enzymatic level. Finally, the synergistic effect between manganese and galactose on IgG

galactosylation showcases the complex and fine balance that governs protein glycosylation. The ANN successfully identifies the aforementioned complex relationship between the two compounds and closely describes the IgG glycan distribution for all training, validation and predictive experiments.

4.2.3 Utilization of ANN for predicting the effect of glycosyltransferases isoforms knockout

Whilst the manipulation of process parameters can lead to improved product quality, radical changes in protein characteristics such as glycosylation are usually achieved through gene engineering[349, 371, 545]. Bydlinski et al.[350] have performed a series of triple and quadruple knockouts of the four isoforms of b4GalT in CHO cells transiently producing EPO-Fc or Fc-DAO. In contrast to IgG, both aforementioned fusion proteins present multiple *N*-linked glycosylation sites. The authors investigated the profile of each *N*-linked glycosylation site separately, enabling a more comprehensive evaluation of each isoform contribution.

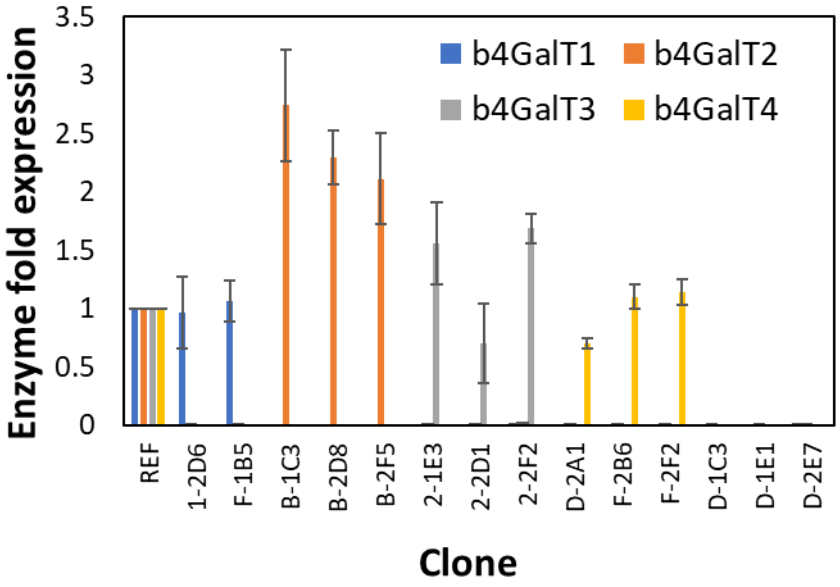

Figure 4.10: Expression levels of each isoform and for each clone, normalized according to the mRNA expression levels of the control experiment (REF). Adapted from Bydlinski et al.[350]

For that reason, the developed ANN employs the isoforms' mRNA expression levels as features (Fig. 4.10), in order to describe the site-specific glycosylation profile (labels) of both fusion proteins. The acquired mRNA expression levels are normalized against the relative transcript abundance of the GAPDH housekeeping gene[350]. In order to rigorously evaluate both the fitting and predictive capabilities of the ANN model, two studies were conducted: 1) in the fitting study, the triple knockouts were used for ANN training and the quadruple knockout for validation and 2) in the predictive study, two of the three triple knockouts were used for model training, the third triple knockout for validation and the quadruple knockout for the evaluation of ANN predictive capabilities (test set). The fitting

study was performed for the glycosites that presented > 18 glycans and the predictive study for the profiles with up to 18 glycans. A 3% error following Gaussian distribution was introduced in the features to generate an additional 16 artificial point for each experiment and strengthen the robustness of the ANN training[457, 546]. Experimental replicates were considered separately in the ANN, in order to account for enzymatic expression variability in different clones. The resulting configurations for all site-specific ANNs in both the fitting and predictive studies can be found in Table 4.4.

Table 4.4: Resulting configurations for all site-specific ANNs presented in section 4.2.3. The (-) symbol is used when the configuration was not examined. The abbreviation used in the glycosylation site indicates whether the site-specific ANN was configured for the fitting (F) or predictive (P) study. The x/y/z symbolism of the neurons represent the number of neurons in the first (x), second (y) and third (z) hidden layer, respectively. The z value is not included for the 2HLs configurations. For the AAE, the values for both the 2HLs and 3HLs configurations are reported when applicable.

***N*-linked glycosylation site**	**Neurons in the 2HLs configuration**	**Neurons in the 3HLs configuration**	**AAE for 2HLs/3HLs (%)**
Asn168 Fc-DAO (F)	6/16	7/3/19	1.35/1.41
Asn745 Fc-DAO (F)	20/11	2/7/19	0.75/0.63
Asn24 EPO-Fc (F)	17/23	2/10/20	0.96/0.83
Asn38 EPO-Fc (F)	14/23	6/9/10	0.81/0.66
Asn83 EPO-Fc (F)	17/21	2/11/14	0.92/0.88
Asn538 Fc-DAO (F)	12/10	4/11/18	1.07/1.06
Asn538 Fc-DAO (P)	1/5	-	1.39/-
Fc-site EPO-Fc (P)	17/19	-	0.42/-
Fc-site Fc-DAO (P)	23/16	-	0.71/-
Asn110 Fc-DAO (P)	9/7	-	2.02/-

ANN fitting study

The results of ANN training and validation in the fitting study are presented in Fig. 4.11. With the exception of the Asn538 Fc-DAO site (Fig. 4.11F), the rest of the glycosites show a rather complex distribution with more than 25 glycans. Despite the wide variety of identified glycans, the ANN is able to capture and correctly track the majority of glycan distributions. The highest discrepancies between model simulations and experimental measurements are observed for the Asn168 Fc-DAO (Fig. 4.11A) and Asn24 EPO-Fc (Fig. 4.11C) sites, and more specifically for the most abundant GnGnF glycan. In some cases, the model is unable to capture the depletion of specific structures that are not present in

the quadruple knockout experiment but are identified in the training and validation sets. The aforementioned inaccuracy for the GnGnF glycan in the Asn168 Fc-DAO and Asn24 EPO-Fc sites can be potentially attributed to the overestimation of the depleted glycans. Overall, the inclusion of a third HL was not found to significantly improve model results (Table 4.4). For that reason, a third hidden layer was not considered for the predictive analysis.

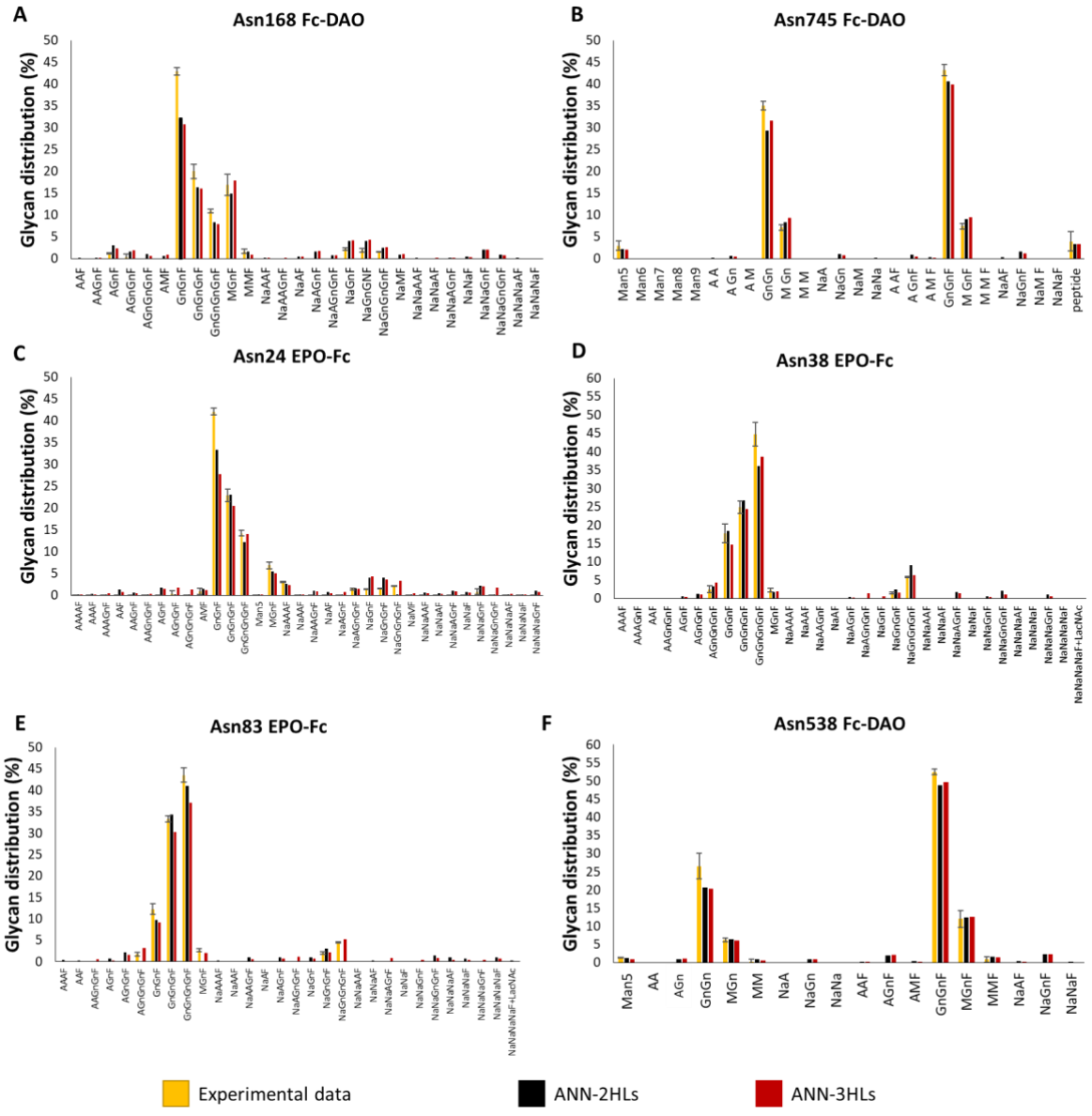

Figure 4.11: (A-F) ANN validation results (fitting) for each *N*-linked glycosylation site and for both fusion proteins. Both scenarios with two and three HLs were examined. The glycans included in the graph and the analysis were present in at least three of the knockout clones. Glycans measured in low abundances (<1% distribution) in up to two clones were excluded from the analysis. The presented experimental results are the average of the quadruple b4GalT1/2/3/4 knockout (clones: D-1C3, D-1E1 and D-2E7). Taken with permission from Kotidis and Kontoravdi[541].

ANN predictive study

In order to evaluate the predictive capabilities of the ANN configuration, an alternative approach on model training, validation and testing was proposed. The ANN was trained on part of the triple knockout scenarios, was validated against the remaining triple knockout experiments and tested against the quadruple knockouts. ANN predictions for the considered glycosites are shown in Fig. 4.12. Model predictions presented an overall AAE of 1.14% for the quadruple experiments, indicating a close match between the experimental measurements and ANN results. The Asn110 Fc-DAO glycoprofile presented minor differences between the quadruple and control experiments, due to negligible galactosylation levels in the latter. On the other hand, all remaining glycosylation sites exhibited extensive alterations, compared to the wild type, that were successfully captured by the ANN. Whilst the prediction of the quadruple clones glycosylation profile could be considered an extrapolation, the ANN was provided with enough data to correctly weigh the contribution of each b4GalT isoform.

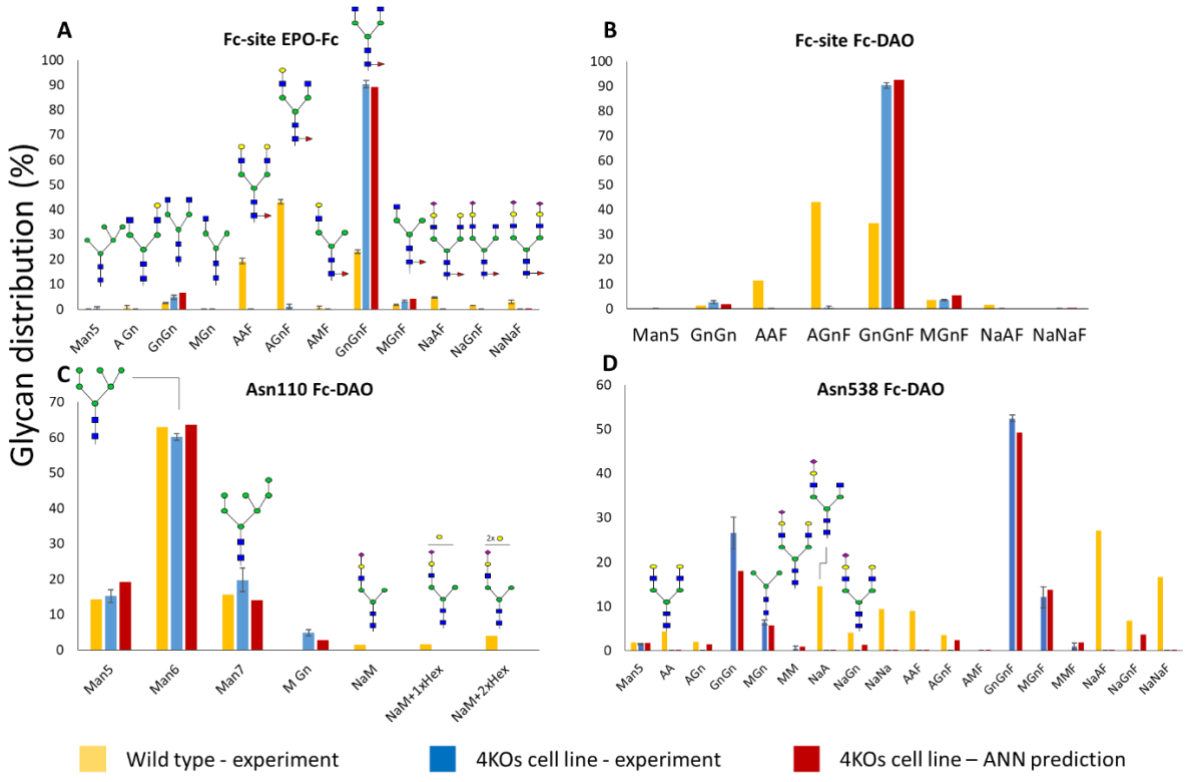

Figure 4.12: (A-D) ANN model predictions for the quadruple knockout clones and for the depicted fusion proteins and glycosites. The experimental data for the wild type cell lines (REF) are displayed for reference. Taken with permission from Kotidis and Kontoravdi[541].

4.3 Discussion

In Chapter 4, an ANN-based configuration for predicting recombinant protein glycosylation was presented. The aim of the current chapter was to showcase that data-driven and hybrid approaches can be employed in order to improve predictions, reduce user interference and simulation times and

expedite model development. Glycosylation is a highly non-linear cellular function that does not follow any known genetic template. Therefore, the use of machine learning tools, such as ANNs, can alleviate model parameterization and reduce the burden of mechanistically describing a complex set of thousands of reaction rules.

Initially, an ANN is proposed for reporting IgG glycosylation based on experimentally measured NSD and nucleotides intracellular concentrations. The model achieves an AAE of 0.87% during validation, as shown in Fig. 4.2. The network of NSDs and nucleotides synthesis has been previously described with the use of kinetic models, either in its entirety[121] or in reduced versions[480, 482]. Alternatively, Sha et al.[484] have attempted to calculate the fluxes of NSDs towards the Golgi apparatus where glycosylation occurs. Whilst kinetic models offer informative insights on the underlying mechanisms of the process, the Monod-type equations usually assumed, do not account for more complex phenomena, such as the regulation of transporters and glycosyltransferases expression that could significantly affect protein glycoprofile[293, 297]. In addition, the assumption of linear correlation between the NSDs flux towards the Golgi[484] or the cytosolic NSD levels[473] and the intra-Golgi NSD concentration, oversights the regulation of NSD transport from NSTs (i.e., SLC35A2). On the other hand, the proposed ANN configuration considers a non-linear relationship between the features (NSD and nucleotide concentrations) and labels (glycan distribution) with no further assumptions on the transport or regulation of the former. Kinetic studies that include NSD transport from cytosol to the Golgi demand excessive simulation times[478]. The low levels of AAE in the validation set and the accurate description of glycans distribution, confirm that NSDs and nucleotides structure an informative and robust feature set for describing IgG glycosylation. The ANN additionally presented a superior performance compared to PLS.

In contrast to kinetic models that require a major involvement of the user during training (parameter estimation), ANNs automatically calculate the weights and biases. Greedy search algorithms can also be employed for an automated hyperparameters estimation. Moreover, ANN training and validation usually last from minutes to a few hours, contrary to kinetic model parameterization, that additionally requires concise understanding of the biological background and the available computational techniques and tools. Several elegant and thorough parameterization strategies have been proposed for accelerating and standardizing the parameter estimation methods in kinetic models[476, 478, 482, 486]. Unlike ANNs that only require the identification of features and labels, mechanistic and probabilistic glycosylation models necessitate the construction of a protein-specific reaction network, tailored to specific characteristics (i.e., glycosyltransferase knockout) of the host cell line. Whilst efforts to automatically generate glycosylation reaction networks have been undertaken, the definition of enzymatic and reaction rules is still necessary[28, 474].

Following the successful establishment of NSDs and nucleotides concentrations as ANN predictors for describing IgG glycosylation, the data-driven model was integrated to the mechanistic framework, replacing the function of the *Glycomodel*. The proposed hybrid model, namely HyGlycoM, was employed for predicting glycans distribution utilizing the metabolites and amino acids concentration in the feed and seeding media, similar to the fully mechanistic. As shown in Fig. 4.6, the structure of the hybrid model consists of two mechanistic submodels describing CHO cell extracellular metabolism and intracellular NSD synthesis, and the ANN that utilizes the outcomes of the latter for predicting IgG glycosylation. HyGlycoM was configured during validation and was found to result in 30% improved predictions when compared to the fully mechanistic model (Fig. 4.8A). HyGlycoM was also found to significantly outperform the PLS equivalent (Fig. 4.8B). The use of the kinetic *Metabolism* and *NSD submodels* for describing the metabolic part of HyGlycoM, extends the applicability of the hybrid model to alternative process scenarios, such as altered amino acids concentration in the seeding media and feed, where a data-driven approach would require a comprehensive and extended experimental dataset. In addition, the reduced sensitivity to process perturbations exhibited by mechanistic models, can be proven useful for reliably describing cellular processes, such as metabolism, that inherently carry a high degree of variability. Finally, HyGlycoM presented an AAE of 1.25%, slightly higher than the average experimental deviation which was 0.93%.

Subsequently, the standalone ANN was reconfigured in order to account for the cumulative $MnCl_2$ amount added in CHO cell cultures. Manganese is a co-factor of the N-Acetylglucosaminyltransferases and beta-1,4-galactosyltransferase. The addition of $MnCl_2$ in the culture, in combination with galactose and fucose, is expected to improve IgG galactosylation. Following training and validation, the ANN successfully predicts the effect of manganese feeding on glycans distribution of an independent experiment (Fig. 4.9C). Finally, an alternative ANN-based approach is proposed for predicting the effect of triple and quadruple isoforms knockouts of the b4GalT enzyme. Utilizing the mRNA expression levels of isoforms as predictors, the standalone ANN achieves an accurate prediction of the quadruple knockout effect on recombinant EPO-Fc and Fc-DAO glycosylation sites, with up to 18 different glycans attached (Fig. 4.12). The overall AAE for ANN predictions is reported at 1.14%.

It is important to note that the description of complex glycoprofiles, such as the ones presented in the Asn24 and Asn83 glycosites of EPO-Fc, requires the construction of vast reaction networks. The complexity of the glycoprofiles presented in Fig. 4.11 approximates the variety of glycans reported for the HCPs of CHO cells [547]. Reaction networks with up to 15,000 structures and 50,000 reactions have been constructed in order to account for the complexity of such profiles[475]. In Krambeck et al.[475], the authors coupled the reconstructed network with an underlying mechanistic glycosylation model, which was subsequently applied on several datasets of mutant CHO cell lines in order to describe the

glycoprofile of intracellular and membrane HCPs. In order to achieve good fitting with the experimental measurements, the adjustment of several enzymatic concentrations, for each dataset, was found necessary. On the contrary, the ANN application presented herein, required no further tuning of the rest of the glycosyltransferases involved in the process.

ANN limitations and future perspectives

Whilst the ANN-based configurations presented in the current chapter highlight the wide applicability of the ANNs, it is important to also note the limitations of this powerful tool. Being data-driven machine learning algorithms, ANNs require an extended dataset for proper training in order to avoid overfitting scenarios, in which the model is unable to act as a reliable predictor despite presenting excellent training performance. Moreover, when extrapolating outside the training space, ANNs can prove to be sensitive to slight changes in the features, resulting in a considerable loss of accuracy. Thus, users are advised to be cautious when testing ANNs in unexplored spaces.

When employing data-driven tools for describing biological processes, it is crucial to carefully consider the predictors used as features of the model. The labels need to demonstrate a well-grounded dependence on the assigned features in order to improve the applicability and robustness of the model. Non-linear regression algorithms, such as the feedforward ANNs used in the study presented herein, can achieve a satisfying fitting even between weakly related variables. Thus, testing the ANN in independent experiments and evaluating its predictive performance is of paramount importance. Similar to mechanistic models, it is important to ensure the consistency of the analytical methods used for the labels and features quantification. For example, NSDs/nucleotides can be quantified through high-performance anion exchange chromatography (HPAEC) or MALDI-TOF-MS, enzyme expression levels through RNA-seq, microarrays, qRT-PCR or western blot (in the protein level) and glycans distribution through LC-MS, MALDI-TOF-MS or capillary electrophoresis. Each method introduces a unique variability to the measurements and, therefore, the consistency of applied methods between datasets ensures the minimization of the analysis effect on algorithm performance.

The ANN configurations proposed in Chapter 4 of this thesis, demonstrate the importance of integrating machine learning tools in bioprocess control and performance prediction. Comprehensive analyses routinely performed during cell line development, such as transcriptomics and proteomics, can enrich the list of predictors and improve ANN performance and applicability. More specifically, a combination of NSD concentrations with quantitative proteomics of glycosyltransferases, NSTs and other enzymes relevant to glycosylation, can significantly improve the performance of the models presented herein. Deep learning algorithms utilizing complex datasets that can offer unique insights

in glycosylation have already become available [548-549] and are paving the way for relevant applications to bioprocessing.

Chapter 5 Antibody galactosylation bottlenecks

Chapter overview

The mechanistic model described in Chapter 2, was utilized for the *in silico* development of a Design Space and for ensuring product quality and quantity by manipulating the feeding schedule of galactose and uridine (Chapter 3). The identified Design Space was then validated against experimental data from seven galactose and uridine feeding experiments. Interestingly, the feeding experiments presented a comparable or even improved productivity and growth profile compared to the control. In addition, the feeding experiments falling within the Design Space showed a particularly similar glycosylation profile (Fig. 3.7), despite the different galactose and uridine amounts added to each experiment during the cell culturing period. Whilst all feeding experiments presented higher levels of terminally galactosylated glycoforms compared to the control, the distribution of the aforementioned glycoforms was surprisingly similar between the feeding experiments, indicating an upper threshold in antibody galactosylation, regardless of the amount of galactose and uridine supplemented to the culture and, therefore, UDPGal synthesized. Overall, the highest observed antibody galactosylation at harvest, for all the experiments presented in the previous chapters of this thesis, was 67%, with the galactosylation values of the majority of galactose and uridine feeding experiments varying between 58% and 67%.

Five new fed-batch experiments (including the control) with various concentrations of galactose and uridine in the feed were performed to determine whether there exists an intracellular bottleneck that prevents a further increase in galactosylation levels using precursor feeding alone. The experiments were chosen from the edges of the Design Space, in an effort to also examine the performance of the Design Space in borderline scenarios. The analysis of the cell cultures presented in the current Chapter was considerably more elaborate compared to that in Chapter 3. More specifically, experimental analysis included: 1) monitoring of viable cell density and culture viability on a daily basis, 2) quantification of extracellular metabolites, including pyruvate, 3) quantification of extracellular amino acid concentrations, 4) quantification of intracellular nucleotide and NSD concentrations, 5) quantification of antibody titre on a daily basis, 6) IgG glycan analysis and 7) intracellular HCPs glycan analysis. In addition, all experimental results were utilized for performing a flux balance analysis for several intervals and for each cell culture, in an effort to identify differences between the experiments.

In order to identify the bottlenecks of antibody galactosylation, the mechanistic model was adapted to more holistically describe the protein secretory pathway. Importantly, the adapted model is capable of simultaneously describing intracellular HCPs and IgG *N*-linked glycosylation for the utilized

CHO cell line, in both the ER and the Golgi compartments. The model was parameterized against the experimental data acquired from the new feeding experiments and was used in order to investigate several plausible factors that could limit IgG galactosylation during galactose and uridine addition.

Overall, the aim of the current chapter is to provide answers to the following questions:

- What are the bottlenecks of antibody galactosylation?
- What are the similarities and discrepancies between the different feeding experiments and between the feeding and control experiments, in both the extracellular and intracellular level?
- Is *N*-linked glycosylation of HCPs affected by galactose and uridine addition?

5.1 Materials & Methods

5.1.1 Cell culture maintenance

CHO-T127 cells (kindly donated by MedImmune, Cambridge, UK) producing an IgG1 antibody were used for this study. The cells were maintained in CD-CHO medium (Life Technologies, Paisley, UK) at 36.5 °C ± 0.5 °C, 150 rpm and 5% CO_2 and were passaged every 3 days at a seeding density of 3 x 10^5 cell·mL^{-1}. 50 μM of MSX was supplemented on cell revival and on the first passage.

5.1.2 Fed-batch cultures

Following cell revival and two subsequent passages, fed-batch experiments were set at a working volume of 100 mL in 500 mL Erlenmeyer flasks. The seeding density was set at 2 x 10^5 cell·mL^{-1}. All cultures (including the control) were supplemented with 1 μM $MnCl_2$ (Sigma-Aldrich, Dorset, UK) at seeding and with CD EfficientFeed™ C AGT™ Nutrient Supplement (Life Technologies, Paisley, UK) on day 2 (48h) and every other day at 10% of the working volume. The feed is referred to as Feed C throughout this section. For the P1-P4 experiments, the cultures were supplemented with varying concentrations of D-(+)-galactose and uridine (both Sigma-Aldrich, Dorset, UK) on day 4 (96h) and every other day (Table 5.1). Appropriate galactose and uridine (GU) amounts were diluted in Feed C and added as part of its recommended feeding schedule, in an effort to avoid large volume perturbations. Experiments were carried out in biological duplicates, with the exception of the control that was performed in quadruplicates, and the cell cultures were terminated on day 12. It is important to note that the feeding strategies were selected in order to examine the edges of the Design Space presented in Chapter 3. The aforementioned DS was developed to ensure the desired levels of mAb titre and galactosylation at harvest were achieved, through the feeding of GU.

Table 5.1: Concentrations of galactose and uridine in each experiment. The (-) symbol indicates no feeding. Taken from Kotidis & Kontoravdi[518].

Experiment	Galactose in the feed (mM)				Uridine in the feed (mM)			
Feeding Point (h)	**96 (day 4)**	**144 (day 6)**	**192 (day 8)**	**240 (day 10)**	**96**	**144**	**192**	**240**
P1	79.35	15.38	10.99	248.29	15.87	3.08	2.20	49.66
P2	4.27	168.34	37.72	11.35	0.85	33.67	7.54	2.27
P3	5.19	3.11	235.29	249.94	1.04	0.62	47.06	49.99
P4	21.91	6.41	233.46	3.97	4.38	1.28	46.69	0.79
P5 (control)	-	-	-	-	-	-	-	-

5.1.3 Quantification of viable cell density and antibody titre

Cell culture viability and viable cell density were determined using the "Viability and Cell Count" assay on NucleoCounter NC-250™ (ChemoMetec A/S, Allerod, Denmark). Prior to analysis, cells were stained with the DAPI-containing solution 18 (ChemoMetec A/S, Allerod, Denmark). Both variables were monitored daily for all experiments and replicates. Extracellular antibody concentration was measured daily from 4 µL of the cell culture supernatant using the BLItz system (Pall ForteBio Europe, Portsmouth, UK) and the Dip and Read™ Protein A (ProA) Biosensors (Pall ForteBio, Portsmouth, UK).

5.1.4 Amino acid analysis and metabolites quantification

The amino acid analysis was performed with HPLC (Alliance, Waters, UK) using the AccQ-Tag method (Waters, UK). Cell culture supernatant was isolated through centrifugation for 5 mins at 100 *g* and was further filtered in 0.22 µm centrifuge tube filters (Sigma-Aldrich, Dorset, UK) to remove residual cells and debris (5 min, 100 g). Norleucine (Sigma-Aldrich, Dorset, UK), an amino acid that is not present in the culturing media and not synthesized through CHO metabolism, was used as an internal standard (2.5 mM solution in 0.1M HCl). Filtered samples were spiked with an equal volume of the internal standard solution and were diluted 25 times (final dilution) prior to derivatization, using Milli-Q grade water. 10 µL of the diluted spiked samples were added to 70 µL of the AccQ Fluor Borate buffer (Waters, UK) and 20 µL AccQ Derivatization Reagent (Waters, UK) and mixed immediately. Following mixing, the samples were heated at 55 °C for 10 min. The samples were left to reach room temperature and were filtered with 0.22 µm centrifuge tube filters for 5 min at 100 g. The Amino Acid Standard (Waters, UK) solution (AAS) was used for building the standard curves. Asparagine, tryptophan and glutamine (Sigma-Aldrich, Dorset, UK) were separately added to the AAS solution prior to analysis. A fluorescence detector (Alliance, Waters, UK) with excitation at 250 nm and emission at 395 nm was used for amino acids detection. Table 5.2 presents the flow-program for the analysis

according to the manufacturer's instructions, where A is the 1:20 Elution A (Waters, UK) diluted with Milli-Q grade water, B is 100% acetonitrile of HPLC-grade (Sigma-Aldrich, Dorset, UK) and C is Milli-Q grade water. Flowrate was set constant at 1 $mL \cdot min^{-1}$. Ammonia was also quantified through the AAA (included in the AAS). Amino acids concentration was measured daily for all cultures.

Table 5.2: The separation flow-schedule used for AAA.

Time (min)	A (%)	B (%)	C (%)
0	100	0	0
0.5	99	1	0
18	95	5	0
19	91	9	0
19.5	83	17	0
33	0	60	40
36	100	0	0
45	69	19	12
54	100	0	0
65	100	0	0

The extracellular concentration of glucose and lactate was monitored daily and quantified using the BioProfile FLEX (NOVA Biomedical, USA). The extracellular concentration of galactose was quantified using the Amplex™ Red Galactose/Galactose Oxidase Assay Kit (Life Technologies, Paisley, UK) and the concentration of pyruvate through a pyruvate oxidase-based assay kit (Abcam, UK). Extracellular uridine concentration was measured using the method described in section 5.1.5.

5.1.5 Nucleotide and nucleotide sugar donor analysis

A cell pellet of 2 x 10^6 cells was acquired through centrifugation at 100 *g* for 5 min. The cell pellet was washed with 2 mL ice cold 0.9% w/v sodium chloride and then centrifuged at 100 *g* for 5 min. Then, 400 µL of ice cold 50% v/v acetonitrile (Sigma-Aldrich, Dorset, UK) were added to the cell pellet and the mixture was incubated for 10 min on ice. Following the incubation step, the samples were centrifuged at 4 °C, for 5 min and at 10,000 g. The debris was discarded while the supernatant was dried using a SpeedVac (Savant Inc. Laboratory, MI). Then, the samples were resuspended in 150 µL distilled water and were filtered through 0.22 µm centrifuge tube filters (Sigma-Aldrich, Dorset, UK) for 5 min at 100 g. If not immediately analysed, the samples were stored at -80 °C to avoid nucleotides and NSDs degradation. The method used for samples analysis has been previously described in del Val

et al.[550]. Briefly, HPAEC using the Dionex CarboPac PA1 column (Dionex, USA), was employed for the analysis of NSDs and nucleotides concentration. The eluents of the HPAEC analysis were: 3 mM sodium hydroxide (E1) and 1.5 M sodium acetate in 3 mM sodium hydroxide (E2).

5.1.6 IgG and host cell proteins glycan analysis

N-linked glycosylation on the recombinantly produced IgG was analysed using the C100HT capillary gel electrophoresis – laser induced fluorescence (CGE-LIF) instrument (SCIEX, NL). For the analysis, 1 mL of cell culture supernatant (centrifuged at 100 *g* for 5 min) was purified with the use of Pierce™ Protein-A IgG Purification Kit (ThermoFischer Scientific, UK) for each experiment and for each examined time point. The eluted IgG was buffer-exchanged to 1 x PBS (ThermoFischer Scientific, UK) through multiple centrifugations in filter microcentrifuge tubes with 3 kDa cut-off (Thermo Fisher Scientific, UK) and at 14,000 *g* and 30 min (for each run). The buffer exchange was terminated after the samples were diluted 1000x in PBS. The buffer exchange from the elution buffer to PBS is necessary to enable optimum performance of the PNGase F enzyme. The final concentration of purified IgG in 1 x PBS was 2.5 $g \cdot L^{-1}$. Glycans were further modified using the C100HT Glycan kit (SCIEX, NL). Briefly, 5 µL of the kit's denaturing solution were added to 40 µL of 2.5 $g \cdot L^{-1}$ IgG in PBS. The samples were then incubated for 8 min at 60 °C for complete protein denaturation. Subsequently, 500 units of PNGase F (New England Biolabs, Hertfordshire, UK) were added and the samples were incubated for 20 min at 60 °C for the complete release of *N*-linked glycans from the Fc fragment. The samples were labelled using 1-aminopyrene-3,6,8-trisulfonic acid (APTS) for 20 min at 60 °C and were washed three times with acetonitrile (Sigma Aldrich, Dorset, UK). Magnetic beads supplemented with the kit were used for labelling and washing. The labelled-glycans were eluted with 50 µL of distilled water and were used for analysis. An injection of 3 kV and for 5 s was used for each measurement. The total analysis time was 20 min for each sample. The results were analysed using the 32 Karat software (SCIEX, NL).

The HCP glycan analysis was performed on samples from the harvest day. For the analysis, a pellet of $1x10^7$ cells was washed with 1 mL PBS and was given to Professor Haslam's Lab for further analysis. The HCPs analysis was kindly performed by PhD student Roberto Donini using an existing protocol for glycan analysis of intracellular CHO cell HCPs established by the Haslam group, using MALDI-TOF-MS[547]. Peaks were annotated using the GlycoWorkbench software[551].

For measuring the total protein concentration, cells were lysed as previously presented[335]. Briefly, 1 x 10^7 viable cells of the day prior to harvest (day 11) were washed with 1 mL 1 x PBS and were subsequently lysed with 200 µL of M-PER™ lysis buffer (ThermoFischer Scientific, UK) supplemented with 1% v/v protease inhibitor cocktail (Sigma Aldrich, Dorset, UK). Cells were incubated in the lysis

buffer for 10 min with gentle shaking. Following incubation, cells were sonicated on ice, using 3 bursts of 5 s, aptitude set at 20 and with intervals of 25 s between each burst. The mixture was centrifuged for 15 min at 14,000 *g* and the cell debris was discarded. The protein concentration was measured on the supernatant using the Coomassie Plus™ Protein Assay Reagent, 950 mL (ThermoFischer Scientific, UK).

5.1.7 Flux balance analysis

The FBA was simulated for all experiments and for three different intervals, namely interval 1: 96h – 144h, interval 2: 144h - 192h, and interval 3: 192h – 240h. These intervals were chosen for FBA as different GU concentrations were added in each experiment and in each feeding point (days 4, 6, 8 and 10). The interval between 240h – 288h was excluded from the analysis as the cultures were approaching the death phase.

The code was developed and kindly provided by Dr. Ioscani Jimenez del Val (University College Dublin) and the reactions included in the FBA model can be found in Appendix-Table A2. Briefly, the FBA is based on the minimization of multiple objective functions including: 1) the difference between the estimated and the experimental fluxes of the specific production/consumption rates and growth rate, 2) the conversion of malate to pyruvate that will therefore induce the conversion of malate to oxaloacetate and 3) the reverse ATP generation. For all reactions, the upper constraint for the estimated fluxes was set at 600 nmol·10^6cell^{-1}·h^{-1}, while the lower limit was set at 0 for the irreversible, and -600 nmol·10^6cell^{-1}·h^{-1} for the reversible reactions.

The backbone of the code was modified to include the following six reactions (F1-F6) that are related to GU supplementation in the cell culture:

$$Gal_{extr} \rightarrow Gal_{intr} \quad \text{(F1)}$$

$$Urd_{extr} \rightarrow Urd_{intr} \quad \text{(F2)}$$

$$Gal_{intr} + ATP \rightarrow Gal1P + ADP \quad \text{(F3)}$$

$$Gal1P + UDPGlc \leftrightarrow G1P + UDPGal \quad \text{(F4)}$$

$$G1P \leftrightarrow G6P \quad \text{(F5)}$$

$$Urd_{intr} + ATP \rightarrow UMP + ADP \quad \text{(F6)}$$

where, $intr$ and $extr$ indicate the intracellular and extracellular concentrations, respectively. $Gal1P$: galactose-1-phosphate, $G1P$: glucose-1-phosphate and $G6P$: glucose-6-phosphate.

F1 and F2 describe the transport of extracellular galactose and uridine to the intracellular environment, F3-F5 describe the intervention of galactose in the glycolysis pathway and F6 the introduction of uridine to the pyrimidines biosynthesis pathway. It is important to note that the

reaction of UDPGal synthesis from Gal1P and UTP, potentially catalyzed by the EC 2.7.7.64 and 2.7.7.10 enzymes (uridylyltransferases), was not added to the FBA system as the reaction is not included in the iCHOv1 genome scale model[28, 552]. Therefore, all UDPGal synthesis is channelled through reaction F4.

The inputs of the FBA include the experimentally measured specific uptake/synthesis rates of all amino acids and metabolites, including the specific growth rate and productivity. The demand of NSDs for the average oligosaccharide synthesis for the glycosylation of both the recombinant IgG and the HCPs was also included in the FBA inputs, and was calculated using the experimental data of IgG and HCP glycosylation using the framework presented in del Val et al.[513]. The FBA was performed in gPROMS ModelBuilder v.5.0.1 (Process Systems Enterprise, gPROMS, www.psenterprise.com/products/gproms, 1997-2020).

Specific uptake/synthesis rates (nmol·10^6cell^{-1}·h^{-1}) for all metabolites and amino acids were calculated for each feeding interval using linear regression for the three points of each interval. The regression was performed between the cumulative metabolite or amino acid concentration and the corresponding IVCD for each time point. Linear regression was performed in Microsoft Excel for Office 365. Finally, the Escher software was used for network and flux visualization[553], using the iCHOv1 genome scale model as a backbone[28, 552].

5.1.8 Formulas for the calculation of process variables

The values of integral viable cell density (IVCD in 10^5cell·h·mL^{-1}), specific mAb productivity (µg·10^6cell^{-1}·day^{-1}) and specific metabolites/amino acids uptake or synthesis rate (µmol·10^6cell^{-1}·day^{-1}) for each interval were calculated based on Eq. 5.1-5.3:

$$IVCD_{t_2} = \frac{(Xv_{t_2} + Xv_{t_1}) * (t_2 - t_1)}{2} + IVCD_{t_1} \quad (5.1)$$

$$q_{mAb} = \frac{[mAb]_{cumm}}{IVCD_{harvest}} \quad (5.2)$$

$$q_{met} = \frac{[Met]_{cumm}}{IVCD_{harvest}} \quad (5.3)$$

where, t_1 and t_2 are two time points of the culture period. The symbols used for each variable are presented in Chapter 2. Both q_{mAb} and q_{met} are calculated based on the cumulative concentration being consumed or produced during the entire cell culturing period.

The calculation of the HCPs and IgG demand on NSDs, used as input for the FBA, was calculated based on del Val et al.[513]:

$$f_{HCP}^{NSD} = StC_{HCP}^{NSD} \cdot NGS_{HCP} \cdot \frac{[HCP]}{MW_{HCP}} \quad (5.4)$$

$$f_{IgG}^{NSD} = StC_{IgG}^{NSD} \cdot NGS_{IgG} \cdot \frac{[mAb]}{MW_{IgG}} \quad (5.5)$$

where, f_{HCP}^{NSD} ($nmol_{NSD} \cdot 10^6 cell^{-1}$) and f_{IgG}^{NSD} ($nmol_{NSD} \cdot \mu g_{IgG}^{-1}$) are the HCPs and IgG demands for each NSD, StC_{HCP}^{NSD} ($mol_{NSD} \cdot mol_{HCP}^{-1}$) and StC_{IgG}^{NSD} ($molNSD \cdot mol_{IgG}^{-1}$) are the stoichiometric coefficients for each NSD based on the experimentally determined glycans on each protein, NGS_{HCP} ($molecule_{HCP}^{-1}$) and NGS_{IgG} ($molecule_{IgG}^{-1}$) are the number of *N*-linked glycosites per HCP and IgG protein molecule, respectively. The value for the IgG is equal to 2 $molecule_{IgG}^{-1}$ (one glycan site in each Fc fraction) and that for the HCPs is estimated[513] as 0.0809 $molecule_{HCP}^{-1}$. $\overline{MW}_{HCP}$ ($g_{HCP} \cdot mol^{-1}$) is the average molecular weight of HCPs estimated[513] as 46,167 $g_{HCP} \cdot mol^{-1}$ and MW_{IgG} ($molecule_{IgG}^{-1}$) is the molecular weight of the IgG molecule equal to ~150,000 $g_{IgG} \cdot mol^{-1}$.

Throughout the chapter, the terms galactosylation (Eq. 2.46), sialylation (Eq. 5.6), high mannose structures (Eq. 5.7), fucosylation and highly branched (> bi-antennary) glycans are used to describe the distribution of the respective sugar molecules on the glycans. It is important to note that, the contribution of each glycan structure towards galactosylation and sialylation depends on the number of galactose or sialic acid molecules present in the glycan. For example, the contribution of the NaAF glycan that carries 1 sialic acid and 2 galactose molecules is 1x to sialylation and 2x to galactosylation.

$$sialylation = 1\,\%monosialylated\ glycan + 2\,\%disialylated\ glycan + 3\,\%trisialylated\ glycan + 4\,\%tetrasialylated\ glycan \quad (5.6)$$

$$high\ mannose = \%Man9 + \%Man8 + \%Man7 + \%Man6 + \%Man5 \quad (5.7)$$

5.1.9 Statistical analysis

Where suitable, a two samples t-test assuming equal variance for $p < 0.05$ was performed for the evaluation of significant differences between samples and experimental conditions. Microsoft Excel for Office 365 was used for performing the t-tests. Calculation of the Pearson correlation coefficient and principal component analysis were performed in OriginPro 2020 (OriginLab, Northampton, MA). Prior to PCA the datasets were standardized to have a mean of 0 and a standard deviation of 1. Least-square feature selection was performed in Python 3.7 using the *linselect* package (Github repository: https://github.com/EFavDB/linselect).

5.1.10 Model modification and parameter estimation

The model presented in Chapter 2 was adapted to include various additional parts of the secretory pathway. These include *N*-linked glycosylation of intracellular HCPs in the ER and the Golgi, *N*-linked IgG glycosylation in the ER (in addition to the already existing *N*-linked glycosylation in the Golgi) and protein transport from the ER to the *cis*-Golgi and from the TGN to cell membrane. These additional features enable us to simulate HCP glycosylation in parallel with IgG glycosylation. Model simulations

and parameter estimations were performed in gPROMS ModelBuilder v.5.1.1 (Process Systems Enterprise, gPROMS, www.psenterprise.com/products/gproms, 1997-2020). Parameter estimation was performed either through the Model Validation or the Optimization entity in gPROMS. Importantly, the Optimization entity does not calculate a 95% confidence interval for the estimated parameter but enables the use of various constraints and the assignment of custom objective functions.

ER glycosylation

An additional CSTR was employed for describing the *N*-linked glycosylation process in the ER. Processes occurring in the ER for assessing protein folding and initiating the ERAD pathway usually result in extended protein residence times in the organelle, exceeding 30 mins[554]. For that reason, the residence time was set at 45 mins, twice the residence time in the Golgi as assumed in this model. The ER enzymes included in the model are: ER-α-glucosidase I (αGluI), ER-a-glucosidase II (αGluII), ER-α-mannosidase I (ER-ManI) and the glucosyltransferase UGGT. αGluI and αGluII are responsible for trimming the glucose residues from the precursor oligosaccharide. The product is further modified by ER-ManI that cuts one mannose molecule from the glycan, resulting to the final Man8 oligosaccharide. However, not all molecules are trimmed by ER-ManI and, therefore, both Man9 and Man8 structures enter the Golgi. Moreover, UGGT is responsible for adding glucose on the misfolded products of αGluI and αGluII, initiating again in that way the ER-glycosylation process and recycling the misfolded proteins. Enzyme concentrations were estimated based only on the glycoprofile of HCPs, assuming that cellular demands for the expression of glycosidases and GTases is dependent on the resources required for the glycosylation of the natural products of the cells that are the HCPs. The kinetics of enzymatic reactions involved in this part of the model are described in Eq. 5.8-5.12.

$$r_{\alpha GluI} = k_{f,\alpha GluI}\,[\alpha GluI]\frac{[Glc_3Man_9Glc_2]}{([Glc_3Man_9Glc_2]+K_{di,\alpha GluI})} \quad (5.8)$$

$$r_{\alpha GluII,A} = k_{f,\alpha GluI,A}\,[\alpha GluII]\frac{[Glc_2Man_9Glc_2]}{(\frac{[Glc_2Man_9Glc_2]}{K_{di,\alpha GluII,A}}+\frac{[Glc_1Man_9Glc_2]}{K_{di,\alpha GluII,B}}+1)\,K_{di,\alpha GluII,A}} \quad (5.9)$$

$$r_{\alpha GluII,B} = k_{f,\alpha GluI,B}\,[\alpha GluII]\frac{[Glc_1Man_9Glc_2]}{(\frac{[Glc_2Man_9Glc_2]}{K_{di,\alpha GluII,A}}+\frac{[Glc_1Man_9Glc_2]}{K_{di,\alpha GluII,B}}+1)\,K_{di,\alpha GluII,B}} \quad (5.10)$$

$$r_{UGGT} = k_{f,UGGT}\,[UGGT]\frac{[Man9]\,[UDPGlc]_{ER}}{K_{di,UGGT}\,K_{dk,UGGT}\,(\frac{[Man9]\,[UDPGlc]_{ER}}{K_{di,UGGT}\,K_{dk,UGGT}}+\frac{[UDPGlc]_{ER}}{K_{dk,UGGT}}+1)} \quad (5.11)$$

$$r_{ERManI} = k_{f,ERManI}\,[ERManI]\frac{[Man9]}{([Man9]+K_{di,ERManI})} \quad (5.12)$$

with, r_{enz} (μM· min^{-1}) are the reaction rates of each enzyme, $k_{f,enz}$ (min^{-1}) is the rate-limiting turnover rate for each enzyme, $[enz]$ (μM) is the concentration of each enzyme in the ER, $K_{di,enz}$ (μM) is the dissociation constant of the enzyme-glycan complex, $K_{dk,enz}$ (μM) is the dissociation constant of the enzyme-NSD complex. The UDPGlc concentration was assumed to not be limiting and equal to 10x the intracellular concentration of UDPGlc, following a similar approach with Krambeck & Betenbaugh[473], adjusted for the ER volume. With the exception of UGGT, for which sequential order Bi-Bi kinetics were assumed, a simple Michaelis-Menten type equation was applied for the rest of the enzymes. ER volume was considered as 10% of the total cellular volume[555] and equal to 1.1·10^{-13} L.

Transport between compartments

COPII and COPI vesicles are responsible for ER → Golgi and retrograde Golgi → ER transport of proteins, respectively. Transport between the ER and *cis*-Golgi, and between *trans*-Golgi and the cytosolic membrane (CM) was included in the model, but the retrograde protein transport was ignored for the sake of simplicity of the parameter estimation task. The velocity with which the protein travels between the compartments was set equal to the exiting velocity from the prior compartment, in order to avoid protein accumulation or shortage. The shape of the cellular compartment that covers both distances was assumed similar to a cylinder (tube), and the volume of both compartments was considered to be the same. For uniformity, the diameter was set equal to the diameter of the Golgi (7.82 μm)[478], while the length of each compartment was set at 0.15 μm [556]. The compartment length was based on data for the ER to Golgi distance. Spang[556] has reported that the distance between the ER and the *cis*-Golgi in yeast ranges between 100-150 nm. It is important to note that yeast lacks the ER-Golgi intermediate compartment (ERGIC) that is found in mammalian cells, and are therefore expected to exhibit a smaller distance between the two compartments[557]. However, immunoelectron microscopy of rat cells has indicated that the distance between the organelles is less than 200 nm [558-559]. Therefore, the reported 150 nm (0.15 μm) was regarded as a good approximation of the ER to Golgi distance and, therefore, the length of each compartment. Finally, while the ER to Golgi transport was considered for both the intracellular HCPs and the secreted IgG, the transport from the Golgi to the CM was only considered for the antibody, as the HCPs are distributed to various locations in the cell when secreted from the Golgi.

HCP and IgG N-linked glycosylation in the Golgi

The IgG glycosylation network of the updated model was identical to that presented in del Val et al.[478] and used in Chapter 2. The single PFR was replaced with four CSTRs in series to describe the different Golgi compartments (*cis*, *medi*, *trans* and *TGN*). Therefore, fixed distributions for each enzyme were considered in each compartment, and the partial differential equations describing glycan

concentration as a function of both time and distance within the Golgi, were replaced by ordinary differential equations that demonstrate the dependency of glycan concentration only on cell culture time. Residence times were assumed equal for both HCPs and the mAb. The general material balance followed in the Golgi and ER compartments is described in Eq. 5.13:

$$\frac{d[OS]_{j,c}}{dt} = V_{el}\left([OS]_{j,c-1} - [OS]_{j,c}\right) + \sum_{i=1}^{n} \nu_{i,j}\, r_i \quad (5.13)$$

where, $[OS]_{j,c}$ (μM) is the concentration of the j^{th} oligosaccharide at the compartment *c*, *n* is the number of reactions for the examined network (either HCPs or mAb), $\nu_{i,j}$ (-) is the stoichiometric coefficient that expresses the participation of the $[OS]_j$ oligosaccharide on the r_i (μM·min^{-1}) reaction and V_{el} (Golgi length·min^{-1}) is the linear velocity (reverse residence time) with which the protein molecule travels through the ER and the Golgi. It is important to note that Eq. 5.13 is adapted accordingly to replace the $[OS]_{j,c-1}$ term with either the initial concentration of the glycans for the ER compartments or the concentration from the ER → Golgi space for the *cis*-Golgi compartment.

The approach of employing four CSTRs to describe Golgi glycosylation is based on the vesicular transport Golgi model and has been previously used in several studies to describe the glycosylation process with high accuracy[472-473, 476]. Based on previous reports[472-473], a slightly modified distribution of the enzymes in each compartment was considered. Enzyme distributions alongside the $k_{f,enz}$ values are presented in Table 5.3. The a6SiaT was not considered as it is not expressed in CHO cells. The a3FucT and a3GalT enzymes were not assigned distributions as no glycans carrying the respective sugar products were identified in the experimental analysis and their concentration was set equal to zero. The $K_{di,enz}$ and $K_{dk,enz}$ were specific for either the HCPs or the IgG and can be found in the Appendix-Table A3 & A4. If not estimated, the K_d parameters were assigned to their nominal values according to literature. The enzyme concentrations were estimated in order to fit the experimental data of the HCPs glycoprofile.

Table 5.3: Enzymes distribution and rate-limiting turnover rate.

Enzyme	Distribution cis-Golgi	Distribution medi-Golgi	Distribution trans-Golgi	Distribution TGN	$k_{f,enz}$ (min^{-1})	Reference for $k_{f,enz}$
ManI	0.15	0.40	0.30	0.15	888	DV2011[478]
ManII	0.15	0.40	0.30	0.15	1924	DV2011[478]
GnTI	0.20	0.45	0.20	0.15	1022	DV2011[478]
GnTII	0.20	0.45	0.20	0.15	1406	DV2011[478]
GnTIII	0.20	0.45	0.20	0.15	629	DV2011[478]
GnTIV	0.20	0.45	0.20	0.15	187	KB2005[473]

GnTV	0.20	0.45	0.20	0.15	1410	KB2005[473]
iGnT	0.20	0.45	0.20	0.15	25	KB2005[473]
a6FucT	0.20	0.45	0.20	0.15	291	DV2011[478]
b4GalT	0	0.05	0.20	0.75	872	DV2011[478]
a3SiaT	0	0.05	0.20	0.75	491	DV2011[478]

Following the transformation of the model to account for a series of well mixed reactors instead of a PFR, the model was expanded to include the HCP glycosylation network. For that purpose, the GLYMMER software[474] was used for generating the complete reaction network to describe the glycome of CHO cells, as previously reported[475]. The generated network consists of approximately 20,000 reactions and 50,000 substrates. In order to reduce the size of the network and expedite model development and simulation, the *Lumping* function of the GLYMMER software was utilized for reducing the network based on the experimentally observed glycoprofile. A significantly reduced network with 180 reactions and 282 glycan structures was generated. Following its generation, the network was transferred to the gPROMS environment using stoichiometric tables. The conversion of the GLYMMER reaction network to a stoichiometric table was achieved through the use of Python 3.7 and expedites the transfer of any new network to the simulation software of choice. At this point, the IgG network was also adapted to a stoichiometric table-based system, in order to facilitate future expansions of the model to include alternative and more complex glycan profiles.

Finally, apart from the mannosidases (ManI and ManII) that were assumed to follow Michaelis-Menten type kinetics, all glycosylatransferases were assigned sequential-order Bi-Bi kinetics, similar to Eq. 5.11 that described the UGGT activity in the ER. The assignment of a sequential-order Bi-Bi mechanism is a simplification for the kinetics of a6FucT, GnTIII and a3SiaT enzymes that have been previously found to follow random-order kinetics[478]. The use of a single mechanism for describing glycosyltransferases kinetics was employed for reducing model complexity. The general formula for sequential-order Bi-Bi kinetics is also shown in Eq. 2.42. Importantly, the inhibition of HCP glycans on each enzyme was considered in the kinetics of IgG glycosylation, and vice versa. Thus, a strong dependency between the glycosylation of the two proteins was ensured, further enabling the investigation of their in-between relationship within the glycosylation context.

Nucleotide sugar donor transport

The regulation of NSD transport was considered only for the Golgi compartment, as the structure of the ER glycosylation system requires only the concentration of UPDGlc for a single reaction. Therefore, the UDPGlc concentration in the ER was calculated based on a simplifying assumption, as mentioned

earlier, that alleviates a considerable burden and complexity from the model. However, the transport of the NSDs (UDPGlcNAc, GDPFuc, UDPGal and CMPNeu5Ac) to the Golgi compartments was assumed to follow a Michaelis-Menten type transport rate without inhibition from the NSD concentration and based on the cytosolic concentration of each NSD (Eq. 5.14). The cytosolic NSD concentration was calculated directly from the respective intracellular NSD levels (*NSD submodel*) based on the assumption that the cytosolic volume is equal to ~70% of the total cell volume and therefore the cytosolic concentration of each NSD is ~1.4x the intracellular value[560]. The transport of NSDs was assumed to occur in each compartment independently; no NSD transfer between compartments was considered.

$$Fin_{NSD}(c) = \frac{A_{G,c}}{V_{G,c}} k_f^{NST} [NST]\ distr_{NST}(c) \frac{[NSD]_{cytosol}}{K_m^{NST} + [NSD]_{cytosol}} \quad (5.14)$$

where, $Fin_{NSD}(c)$ (min^{-1}) is the transport rate of each NSD in each compartment (c), $A_{G,c}$ (μm^2) and $V_{G,c}$ (μm^3) are the area and the volume of each Golgi compartment, respectively, $[NST]$ (μM) is the nucleotide sugar transporter concentration, $distr_{NST}(c)$ (-) is the distribution of the NST in each Golgi compartment, $[NSD]_{cytosol}$ (μM) is the cytosolic NSD concentration and K_m^{NST} (μM) is the saturation constant each NST. The $A_{G,c}$ and $V_{G,c}$ were common for all compartments and were set at 99 μm^2 and 6.25 μm^3, respectively[478]. Thus, the material balance for each intra-Golgi NSD and in each compartment is given by Eq. 5.15:

$$\frac{d[NSD](c)}{dt} = Fin_{NSD}(c) - \sum_{i=1}^{95} \nu_{NSD,i}^{IgG}\, r_i^{IgG} - \sum_{j=1}^{282} \nu_{NSD,j}^{HCPs}\, r_j^{HCPs} \quad (5.15)$$

where, $[NSD](c)$ (μM) is the NSD concentration in each compartment of the Golgi, $\nu_{NSD,i}^{IgG}$ (-) and $\nu_{NSD,j}^{HCPs}$ (-) are the stoichiometric coefficients that describe the usage of each NSD in the i^{th} and j^{th} reactions of the IgG and HCPs network, respectively, and r_i^{IgG} (μM·min^{-1}) and r_j^{HCPs} (μM·min^{-1}) are the reaction rates in the IgG and HCPs network, respectively. With the exception of the UDPGal transporter (SLC35A2 gene), the distribution of which was estimated, the rest of the transporters were assigned a similar distribution with the respective glycosyltransferases (Appendix-Table A5 & A6). Similarly, the K_m^{NST} value for each NST was assigned to its nominal values[478], with the exception of K_m^{UDPGal} that was estimated. The concentration of each NST was also estimated.

Parameter estimation

For all model simulations and parameter estimations, the parameters of the *Metabolism submodel* were assigned to their re-estimated values, as calculated in Chapter 3 and presented in Table 3.2 (Set B). The objective function of the ER-related parameter estimation was to maximize the amount of Man8 and Man9 glycans that leave the ER, while retaining a ratio of 1:4 (Man8:Man9). However, the

actual Man8:Man9 ratio has not been sufficiently studied and the latter constraint was reasonably relaxed. Estimated parameters for the ER glycosylation model can be found in Appendix-Table A7. The Golgi enzyme concentrations were estimated solely based on the native HCP glycoprofile. Further adjustment for both the HCPs and IgG kinetics was performed through the estimation of appropriate $K_{di,enz}$ and $K_{dk,enz}$ values and against each glycoproteins profile. All estimated parameters are shown in Appendix-Table A3 & A8. For estimating enzyme concentrations and inhibition constants, the approach presented in Chapter 2 was followed.

Regarding the estimation of the NSD transport constants and NST concentrations, a slightly different parameter estimation strategy was chosen for each NSD. The objective function for each parameter estimation was the minimisation of the difference between the cytosolic and intra-Golgi NSD concentrations. As CMPNeu5Ac is not substantially consumed towards IgG and HCP *N*-linked glycosylation, and is mainly involved in glycolipid synthesis and HCP *O*-linked glycosylation that are not considered in the current modelling framework, the objective function could not be further constrained based on the available experimental data, and a lower threshold at 40x the cytosolic concentration[473] was applied. On the other hand, the rest of the NSDs are considerably consumed for both IgG and HCP *N*-linked glycosylation and could be, therefore, further constrained. Each NSD was constrained within a different range of acceptable intra-Golgi concentrations based on previously reported values[473, 478]. For UDPGal the range was set at [50 μM, 1500 μM], for GDPFuc at [10 μM, 344 μM] and for UDPGlcNAc at [500 μM, 1500 μM].

Finally, when required, the intra-Golgi concentrations of UDPGal were estimated against the experimentally measured distributions of the major IgG glycans for each considered time point.

5.2 Results

5.2.1 Cell growth and antibody synthesis

For viable cell density, all feeding scenarios show similar profiles up to the 168h time point, with the experiments starting to deviate from the 192h time point onwards (Fig. 5.1A). The aforementioned observation is expected, as the feeding concentrations of all experiments are at low levels at 96h, starting to considerably increase after the feeding at 144h (Table 5.1). Notably, all cell cultures exhibit a moderate death phase during the late hours of the culture period due to the addition of Feed C. P1 and P4 include the highest concentrations of the first GU feeding at the 96h time point (with those in P1 being the highest overall at 79.35 mM galactose and 15.87 mM uridine), which, however, are considerably lower compared to the highest overall feeding concentrations in each experiment. Interestingly, while the elevated early feeding does not affect cell growth during the early exponential phase, the viable cell density of the P1 experiment presents the lowest levels during the late

exponential and stationary/death phase. The remaining cultures, P2-4, exhibit very similar viable cell density profiles across the culture duration, with that of P4 declining earlier than those of P2 and P3, from day 10 onwards, again pointing to a potentially negative effect of early supplementation.

The IVCD exhibits a more discrete profile compared to the rest of the examined variables (Fig. 5.1B). P1 and P4 show consistently lower levels of IVCD compared to the control experiment, while P3 presents an elevated IVCD profile compared to the control. P2 and P5 are almost identical throughout the culturing period. P1, that incorporates the highest feeding concentrations on day 4 as mentioned earlier, results in the lowest final IVCD, thus, indicating a plausible negative correlation between the early supplementation of galactose and uridine and IVCD. However, the values of IVCD at harvest were not found to be significantly different ($p > 0.05$) between the control and each of the feeding experiments. On the other hand, P3 and P1 presented a significantly different IVCD at harvest ($p < 0.05$).

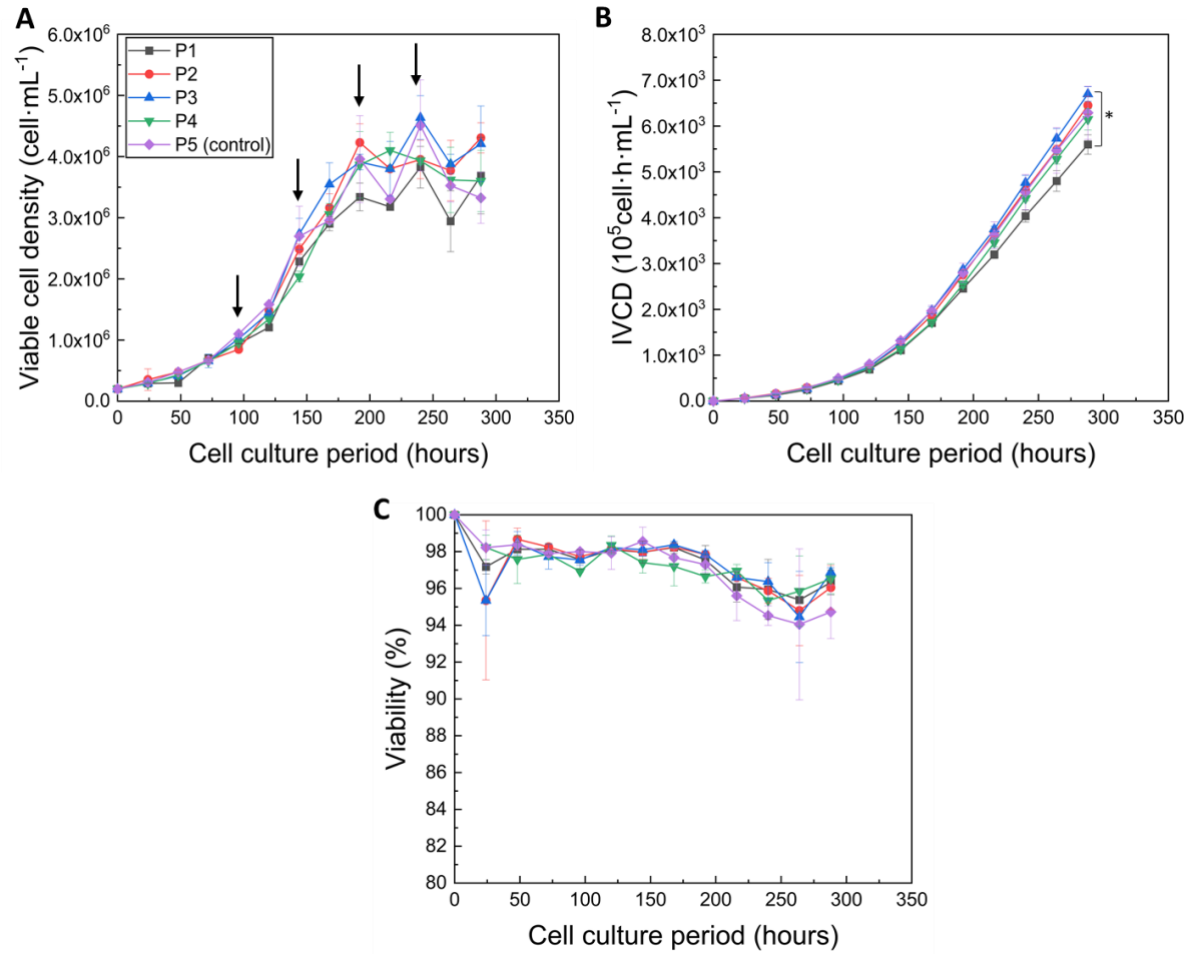

Figure 5.1: (A) Viable cell density, (B) integral viable cell density and (C) The arrows indicate the galactose and uridine feeding points. Note that the specific growth rate could not be calculated for 0h due to the structure of the formula used for its calculation. The legend presented in the A panel is common for all panels. The asterisk in the B panel indicates significant difference for p set at 0.05.

Culture viability remains at high levels during the entire cell culture period, with all values being > 94% at harvest, thanks to the addition of Feed C every two days and, therefore, replenishment of nutrients (Fig. 5.1C). Interestingly, the control presents the lowest viability from 216h onwards. The aforementioned observation demonstrates a plausible positive effect of the GU supplementation on culture viability during the stationary phase of the culture, the mechanism of which could be related to the utilization of galactose for fuelling the glycolysis pathway. However, the viability profiles present no significant differences between the different feeding experiments and the control ($p > 0.05$), and the absolute difference between the control and feeding points is rather small (~1.7%) and within the margin of experimental error.

Similar to cell growth, antibody titre and specific productivity present considerable similarities between the cultures (Fig. 5.2). During the first 168h all experiments exhibit similar mAb titre profiles, with the values starting to deviate during the late exponential and stationary/death phases. P2 and P3 show the highest final mAb titre, with P1 resulting in the lowest antibody concentration at harvest (432 ± 6 mg·L^{-1}). In addition, all experiments were found to satisfy the mAb concentration constraint for the Design Space, as defined in Eq. 3.4 (430 mg·L^{-1}). The decrease in mAb titre between a small number of intervals is probably a result of dilution (due to sampling and feed addition) and experimental error during titre quantification. Whilst the GU feeding experiments presented a slightly higher specific antibody production rate, no significant difference between the experiments was identified (Fig. 5.2B). Finally, the PCC between mAb titre and IVCD is calculated as ~0.96 over all experiments and a total of 130 points, showcasing the strong correlation between cell growth and antibody synthesis.

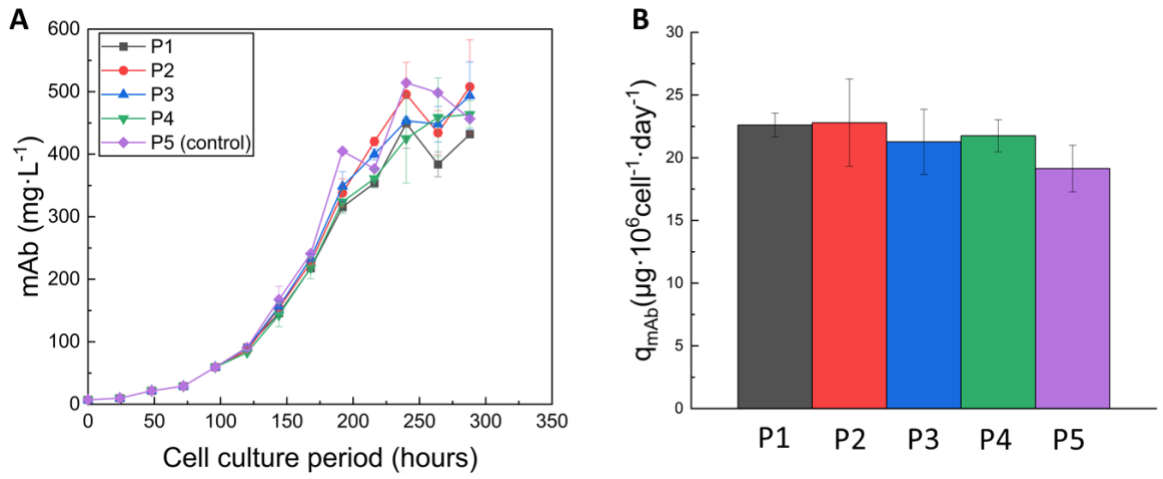

Figure 5.2: (A) Antibody concentration and (B) specific mAb productivity at harvest, accounting for the mAb cumulatively produced throughout the cell culturing period.

5.2.2 Extracellular metabolism

As shown in section 5.2.1, all cultures present similar profiles for growth and antibody synthesis, following an expected behaviour based on the *in silico* kinetic model simulations that guided the design of the feeding strategies. In order to further investigate the effects of the variable GU additions on cell metabolism, the extracellular profiles of all metabolites and amino acids were analysed. Regarding glucose, P5 presents the lowest levels of glucose throughout the entire cell culturing period (Fig. 5.3A). However, the glucose levels start considerably differing between the control and the feeding experiments following the 192h time point, with P5 presenting a final glucose concentration of ~47.8 mM. The elevated accumulation of glucose in the feeding experiments is attributed to the alternative consumption of galactose towards the synthesis of pyruvate and the fuelling of the TCA cycle. The intervention of galactose in the glycolysis pathway and the subsequent reduction of glucose consumption have been previously reported for both the specific cell line in-use[482] and for different CHO cell lines[292]. Regarding the GU feeding experiments, P2 presents the lowest final glucose concentration (~56.6 mM) and P1 the highest (~64 mM), with the profiles of all feeding experiments presenting similar values up to the 240h time point. Interestingly, the considerably different concentrations of galactose supplemented to each culture at 96h, 144h and 192h show no significant effect on glucose concentration up to the 240h of culture.

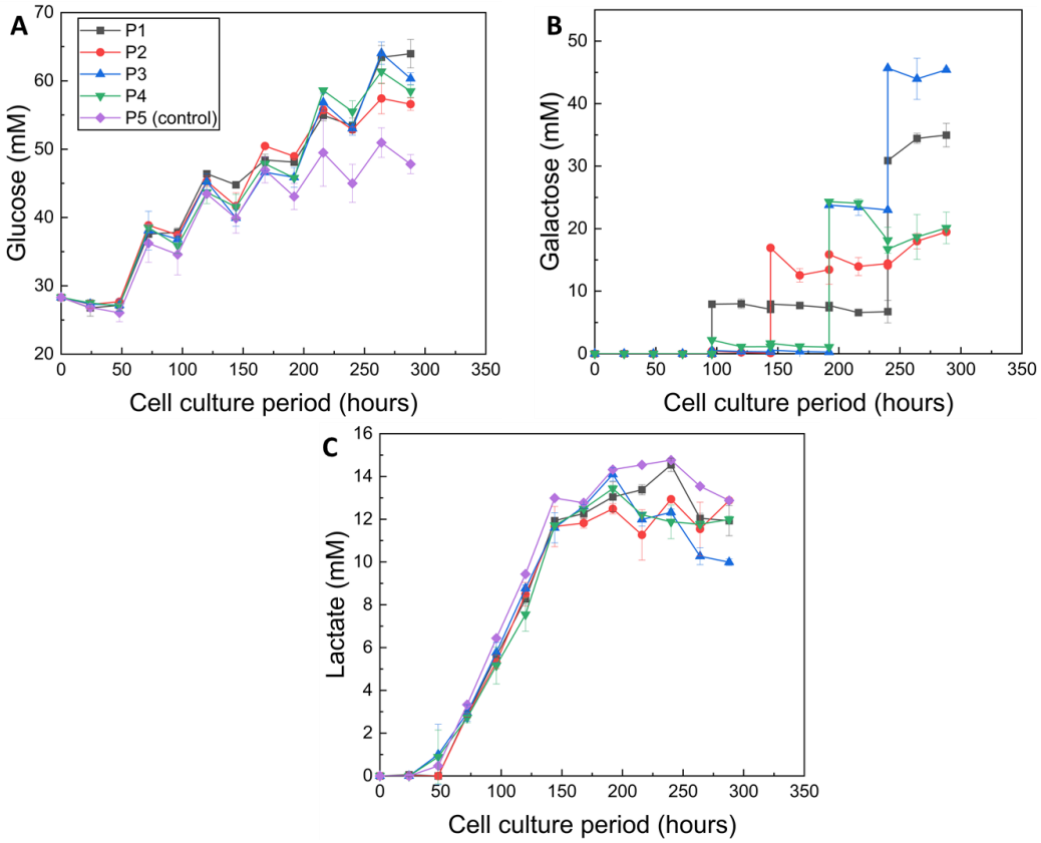

Figure 5.3: Extracellular concentration of (A) glucose, (B) galactose and (C) lactate over the entire cell culture period. The legend of panel A is common for all panels. Note that in panel B, the calculated galactose concentration immediately after feeding has been included, to enable a better understanding of the metabolite's profile. Also, the control (P5) is not included in panel B as no galactose was supplemented to the experiment.

Galactose is consumed throughout the cell culturing period at lower rates compared to glucose for the majority of the intervals (Fig. 5.3B). An increase in galactose concentration during the last two days of the cell culture for the P1, P2 and P4 experiments is attributed to measurement errors in galactose quantification, rather than to a metabolic shift towards galactose synthesis. The enzymatic assay used for galactose quantification is based on fluorescence detection that can introduce considerable measurement errors. The main discrepancies in the galactose profiles are attributed to the different GU concentrations supplemented to the cultures at each time point. The lactate profile is similar between all cultures up to the 144h time point, with the control exhibiting slightly higher concentrations (Fig. 5.3C). Following the 144h time point, the control presents, overall, the highest lactate concentrations compared to all feeding experiments, with the exception of P3 that demonstrates a late lactate metabolic shift and reaches similar levels to the control at 240h. However, the lactate concentration of P3 is gradually reduced at 264h and 288h in lower levels than the control. Interestingly, P2 presents a repetitive shift of lactate metabolism from 192h onwards, with lactate being consumed after feed supplementation, followed by lactate production for the next interval.

Overall, P3 presents the lowest lactate concentration at harvest (~10 mM) and P5 the highest (~13 mM). The elevated levels of lactate in the control experiment are in agreement with previous studies suggesting that galactose addition in CHO cell cultures can lead to reduced lactate accumulation[268, 561].

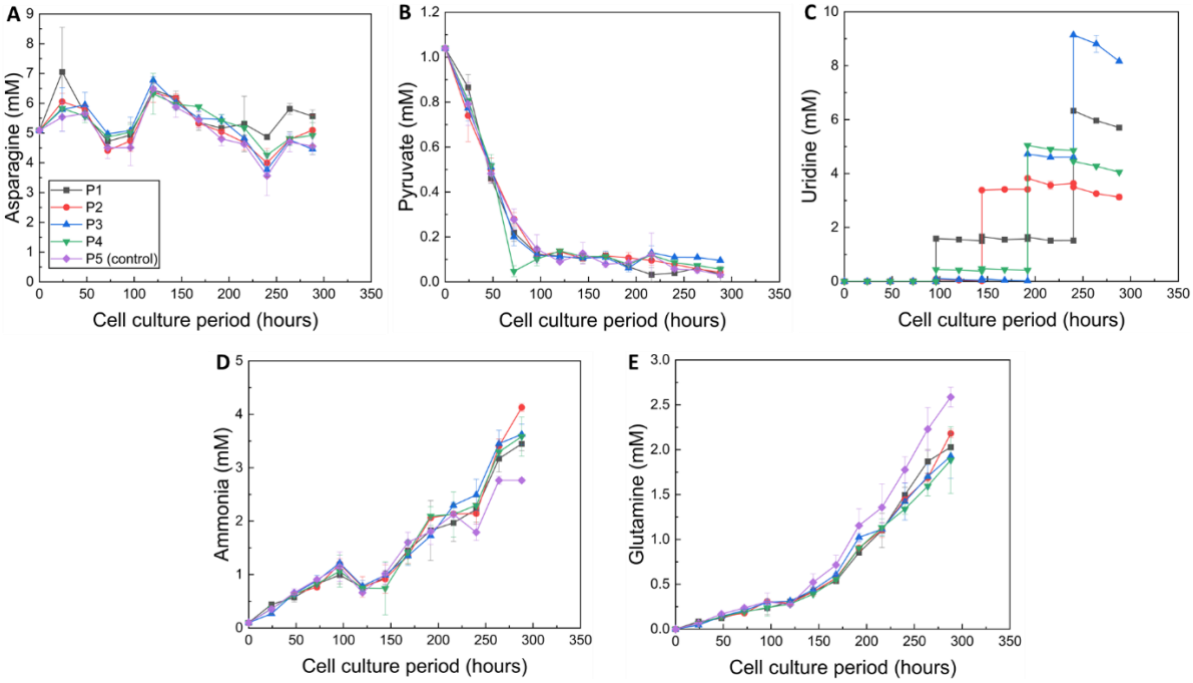

Figure 5.4: Extracellular concentration of (A) asparagine, (B) pyruvate, (C) uridine, (D) ammonia and (C) glutamine over the entire cell culture period. The legend of panel A is common for all panels. Note that in panel C, the calculated uridine concentration immediately after feeding has been included, to enable a better understanding of uridine's profile. Also, the control (P5) is not included in panel C as no uridine was supplemented to the experiment.

The asparagine profile of all experiments is almost identical, with the exception of P1 that deviates in the last three time points, resulting in a slightly higher final concentration (Fig. 5.4A). Similarly, pyruvate concentration follows a common trend for all cell cultures (Fig. 5.4B). Interestingly, extracellular pyruvate (supplemented through CD-CHO media) is rapidly consumed during the first 96h of culture, with the exception of P4 that exhibits a higher consumption rate. Following the 96h time point, pyruvate seems to reach a steady-state, as its concentration remains relatively constant for all cell cultures. Pyruvate is a product of the glycolysis pathway and acts as a substrate for the initiation of the TCA cycle and the synthesis of lactate. The rapid initial consumption of pyruvate, alongside the consumption of glucose and the synthesis of lactate (Fig. 5.3), demonstrates the vital role of pyruvate for the regulation of the fluxes towards the TCA cycle. Moreover, uridine is consumed for the synthesis of the uridine-phosphates nucleotides and the UDP-sugars, immediately after being added to the cell culture, with a more prominent reduction of uridine concentration for all feeding experiments during the last interval of the culture (Fig. 5.4C). The addition of galactose and uridine results in elevated ammonia accumulation and reduced glutamine synthesis, compared to the control

(Fig. 5.4D & E). Based on the data presented in Chapter 2 where the same amount of galactose was fed with and without uridine (10G and 10G20U experiments), the observed increase of ammonia levels in the cultures of the current chapter is attributed to the uridine addition, without however being able to identify the pathway through which uridine results to elevated ammonia and reduced glutamine concentration. As ammonia is consumed towards glutamine synthesis from glutamate, the reduced glutamine levels in the GU experiments can justify the observed ammonia accumulation and vice versa.

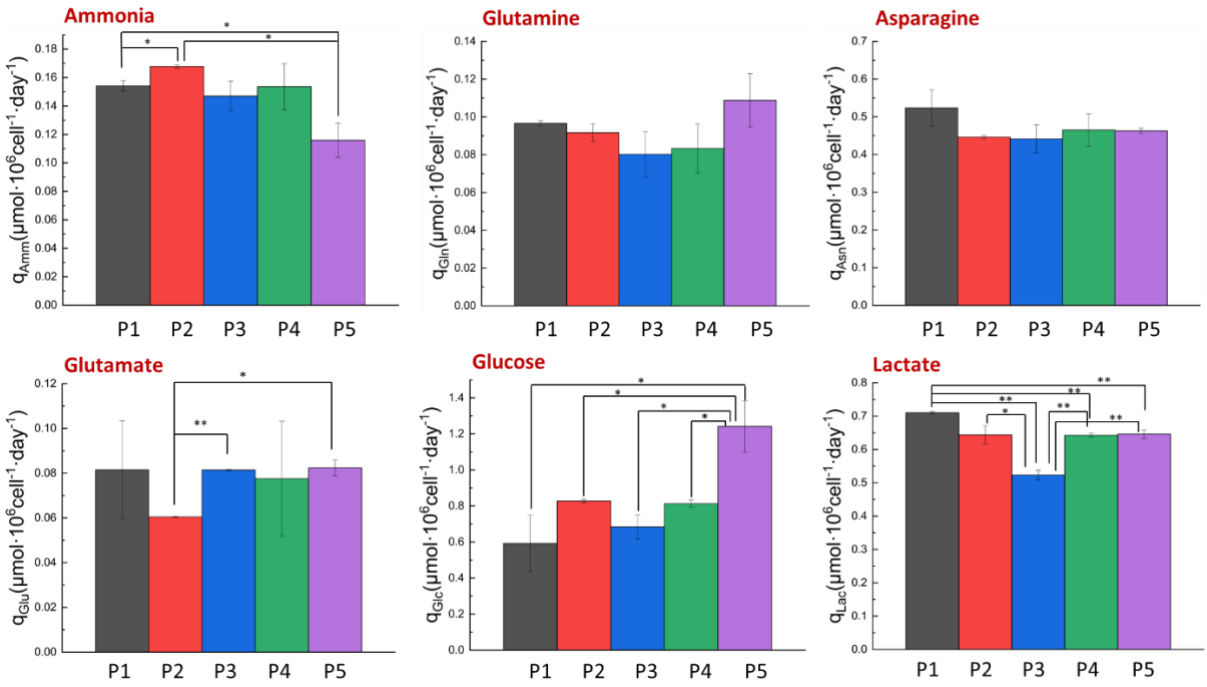

Figure 5.5: Specific synthesis rate for ammonia, glutamine and lactate. Specific uptake rates for asparagine, glutamate and glucose. The specific rates are calculated at harvest, accounting for the cumulatively produced/consumed metabolites. " * " indicates significant different at $p < 0.05$ and " ** " at $p < 0.01$.

Similarly, the control experiment presents the lowest specific production rate of ammonia and the highest rate for glutamine synthesis. Whilst P1 exhibits the highest specific uptake rate of asparagine, no significant differences are detected. P2 shows the lowest specific uptake rate of glutamate, with significant differences observed with the P3 and P5 experiments. As expected, the glucose consumption rate is higher for the control experiment, while P1 presents the highest lactate synthesis rate and P3 the lowest. Overall, the results indicate that the addition of galactose and uridine mainly affects the uptake or synthesis of glucose, ammonia and glutamine, with minor differences observed for the rest of amino acids and metabolites.

5.2.3 Intracellular levels of nucleotides and nucleotide sugar donors

Time-course measurements of intracellular NSD concentrations are presented in Fig. 5.6. CMPNeu5Ac co-elutes with intracellular uridine and therefore could not be accurately quantified for the GU feeding experiments. However, CMPNeu5Ac was quantified for the control experiment where no uridine was supplemented, and the results are displayed in Fig. 5.6G.

As expected, the levels of UDPGalNAc, UDPGlcNAc, UDPGal and UDPGlc are elevated in the feeding experiments due to the addition of uridine in the cell culture (Fig. 5.6A - D). Uridine transforms to UMP and, therefore, sequentially to UDP and UTP. Thus, the supplementation of uridine leads to increased UTP levels and consequently to increased UDP-sugars concentrations. More specifically, UDPGal, which is also dependent on galactose concentration, is measured between ~3-4x higher for the feeding experiments when compared to the control at harvest. The UDPGal profiles clearly demonstrate the dependence of intracellular UDPGal concentration on galactose and uridine addition. To elaborate, P1 shows the highest concentration between 96h and 144h, followed by P4. The two aforementioned experiments are supplemented with the two highest GU concentrations at 96h, with P1 receiving the highest of the two, and therefore, reasonably exhibiting higher UDPGal levels compared to P2 and P3. A similar trend between the UDPGal levels and the feeding concentrations of galactose and uridine is observed for all feeding experiments and time points. Interestingly, unless supplemented in sufficient levels, UDPGal is swiftly consumed in all feeding experiments. In addition, all feeding experiments present similar profiles after 240h, with P1 slightly deviating from the rest of the GU feeding experiments, reaching a similar UDPGal concentration at harvest. However, the amount of galactose and uridine added to the cultures throughout the culturing period is largely different (Table 5.1), indicating a plausible bottleneck in UDPGal synthesis rate. Given that the specific antibody production rate (Fig. 5.2A) and the specific growth rate are similar for all experiments, the uptake of UDPGal towards protein glycosylation is expected to be in the same range of values. In addition, whilst P1 and P3 are supplemented with considerably higher GU amounts at 240h (day 10), the experiments present similar UDPGal levels during the last three days of the culture with the rest of the feeding experiments, further supporting the possibility of saturation in UDPGal synthesis pathway. It is important to note, that, unlike uridine, galactose was not consumed in the 240h - 288h interval, suggesting that the synthesis of UDPGal during the last hours of the culture results from the transformation of UDPGlc (reaction F4) and not from the plausible reaction between galactose-1-phosphate and UTP (that was also not considered in the FBA). However, the uptake of uridine during the last three days of the culture was considerably different between the experiments, ranging between 0.06 $\mu mol \cdot 10^6 cell^{-1} \cdot day^{-1}$ (P4) and 0.12 $\mu mol \cdot 10^6 cell^{-1} \cdot day^{-1}$ (P3).

UDPGlc follows a similar pattern with UDPGal, with major discrepancies between the control and the feeding experiments observed after the 216h time point. Interestingly, the control experiment presents a reduction in UDPGlc concentration in the late stages of the culture, that in combination with the sudden increase of UDPGal for the control experiment, indicates an increase in the flux of UDPGal synthesis from UDPGlc (reaction F4) and a reduction in UDPGlc synthesis from glucose-1-phosphate. This shift is also observed for the GU feeding experiments, all of which present a rapid increase in UDPGal concentration in the 240h - 288h period, regardless of the amount of supplemented galactose and uridine. Unlike the control, UDPGlc concentration increases in the 240h - 264h period due to the addition of galactose and uridine, and is subsequently reduced at harvest, further supporting UDPGlc swift transformation to UDPGal.

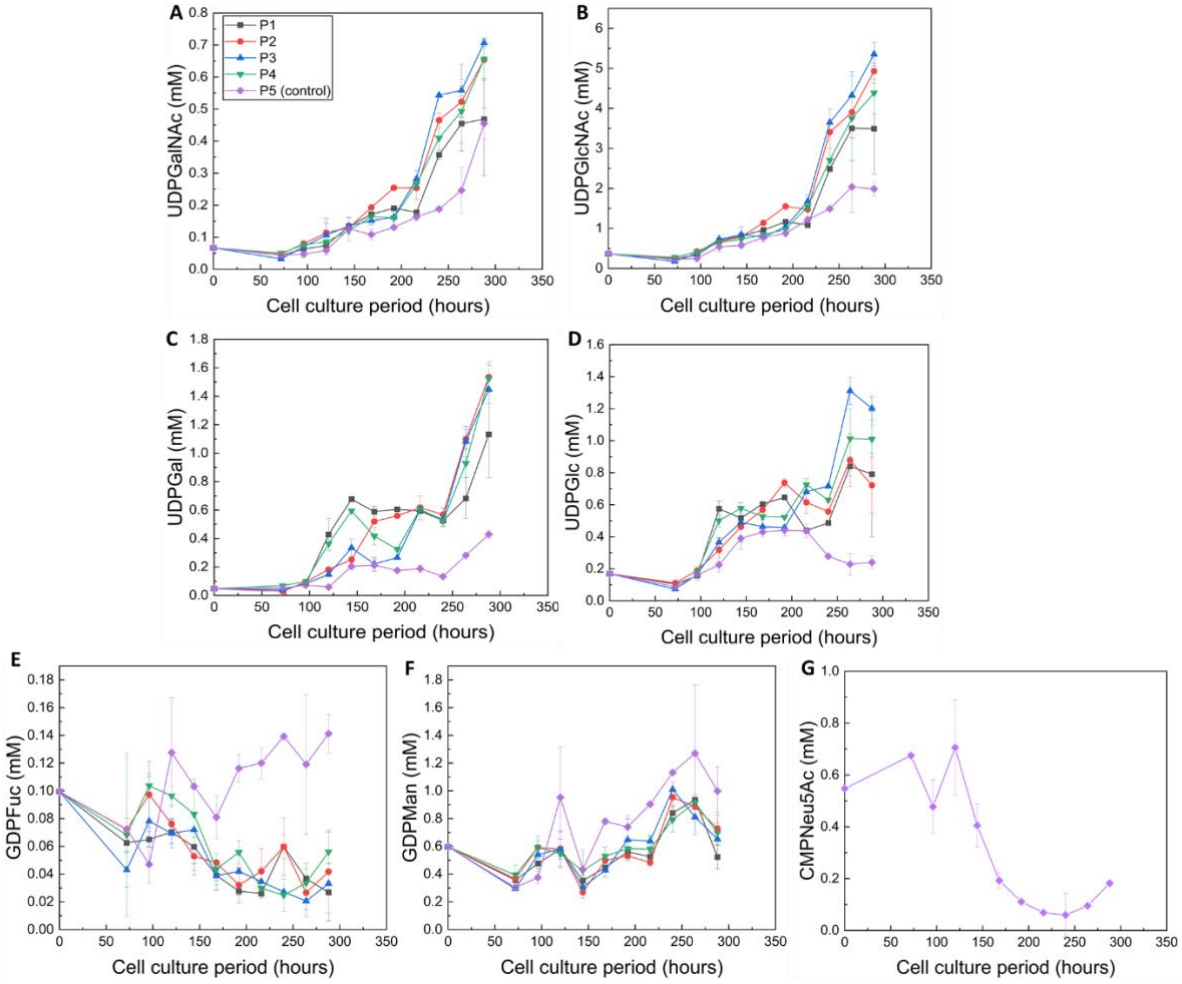

Figure 5.6: Intracellular concentrations of (A) UDPGalNAc, (B) UDPGlcNAc, (C) UDPGal, (D) UDPGlc, (E) GDPFuc, (F) GDPMan and (G) CMPNeu5Ac for all cultures. The legend of panel A is common for all panels.

The control and feeding experiments present similar UDPGalNAc and UDPGlcNAc profiles during the first 144h of the cell culture period. Gradually the profiles deviate for each experiment, but all experiments exhibit increasing concentrations towards the end of the culture. The increase of

UDPGlcNAc levels, and consequently UDPGalNAc through the epimerase reaction that connects the two NSDs, towards the late stages of cell culture has been related to ammonia accumulation[256]. Ammonia transforms to fructose-6-phosphate which acts as a substrate for UDPGlcNAc synthesis.

Surprisingly enough, the control presents higher GDPMan and GDPFuc concentrations compared to the feeding experiments, with the difference becoming more prominent after 120h of culture. The aforementioned discrepancy between the GU feeding and the control experiment can be partially attributed to the different necessities for cellular glycosylation, as examined later in this chapter. Higher levels of fucosylation can result to reduced GDPFuc concentrations. Finally, Fig. 5.6D shows the CMPNeu5Ac concentration of the control experiment that reduces approximately 10x from the start of the culture, to ~0.06 mM, followed by a slight increase of the CMPNeu5Ac levels during the last two days. Unlike IgG, HCPs exhibit high degrees of sialylation in both the *N*-linked and *O*-linked glycosites and therefore, the reduction or depletion of CMPNeu5Ac levels can have an adverse effect on HCP sialylation.

Moreover, UTP and CTP show noticeably higher levels in the GU feeding experiments compared to the control (Fig. 5.7A & C). The concentration profile closely corresponds to the time points of uridine supplementation, demonstrating the direct dependency of both nucleotide triphosphates on uridine concentration. As shown in Fig. 5.8, uridine transforms to UMP and gradually to UTP. However, UTP is further converted to CTP through the activity of the cytidine triphosphate synthetase, and therefore the latter develops an indirect positive correlation with uridine supplementation.

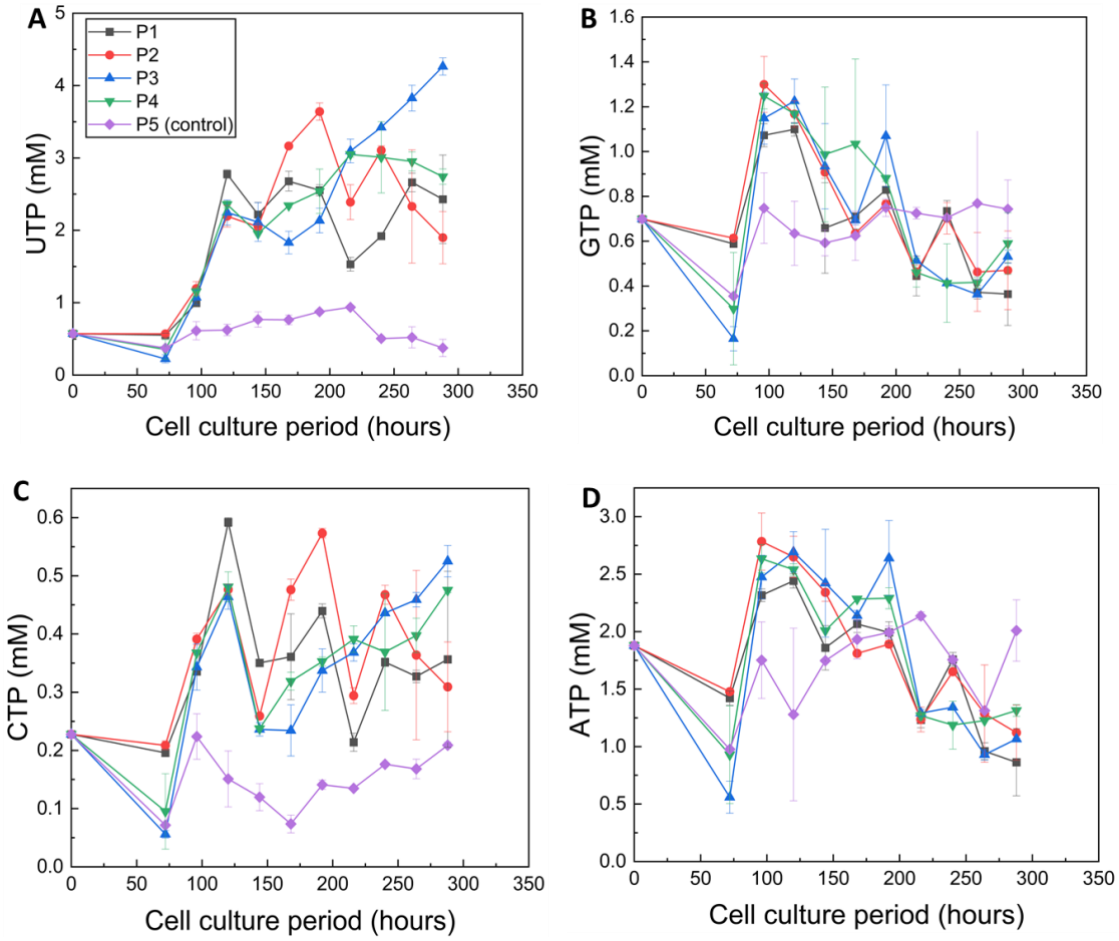

Figure 5.7: Intracellular concentrations of (A) UTP, (B) GTP, (C) CTP, and (D) ATP for all cultures. The legend of panel A is common for all panels.

Both GTP and ATP are reduced after the 96h time point for all GU feeding experiments, while the concentration of GTP remains practically stable and ATP fluctuates within a more narrow range in the control experiment. Whilst the concentration of both nucleotides is higher at the GU feeding experiments and the 96h time point, the control culture presents the higher values at harvest.

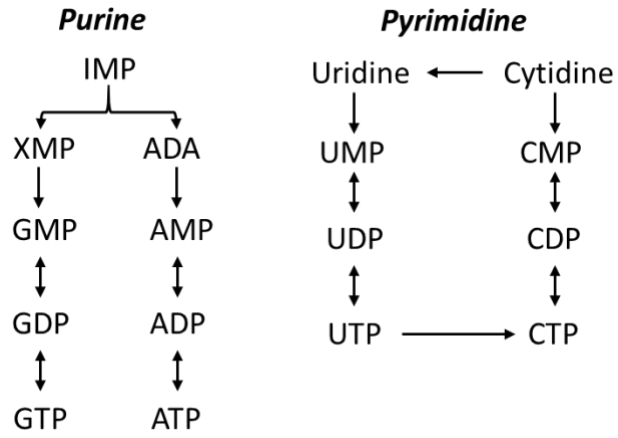

Figure 5.8: De novo purine and pyrimidine synthesis pathways. IMP: inosine monophosphate, XMP: xanthosine 5'-phosphate and ADA: adenylosuccinate. The pathways were extracted from the KEGG database resource[511].

In order to evaluate the intracellular data and identify common cellular behaviours in the synthesis of nucleotides and NSDs, a PCA was performed using the consumption/production rates, as proposed in Kyriakopoulos & Kontoravdi[519], of each variable and for the intervals between 96h - 144h, 144h - 192h, 192h - 240h and 240h - 288h. Intervals prior to 96h were not considered as no GU supplementation was included during the first days of the culture. NSDs are utilized towards IgG and cellular glycosylation, and, therefore, the consumption rates are an indicator of the NSD fluxes towards the glycosylation process.

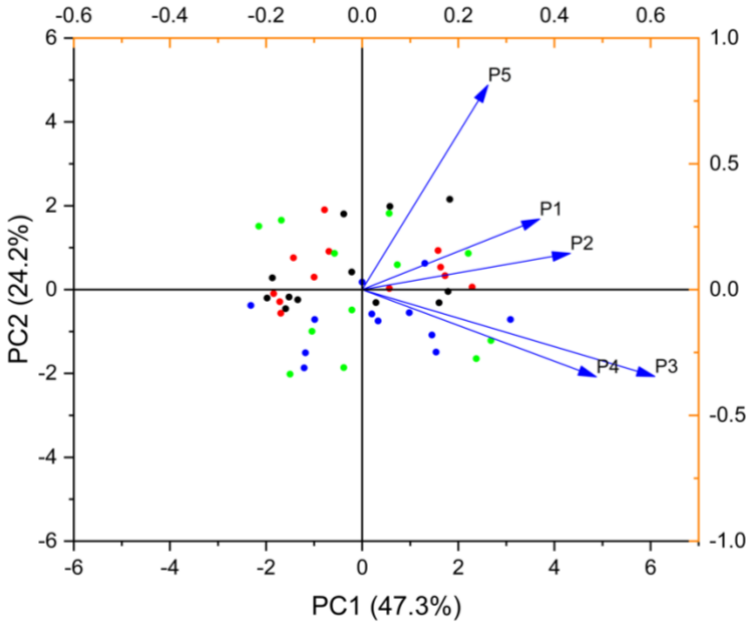

Figure 5.9: PCA on the consumption/production rates of all measured nucleotides and NSDs, excluding CMPNeu5Ac. The black dots indicate the interval between 96h - 144h, red dots between 144h - 192h, green dots between 192h - 240h and blue dots between 240h - 288h. The blue lines indicate the loadings of each experiment, expressing the coordinates for each principal component.

As expected, P5 (control) exhibits a unique behaviour when compared to the rest of the experiments (Fig. 5.9). Reasonably, PCA indicates that P3 and P4 follow a similar nucleotide and NSD metabolism. The GU concentrations supplemented to both cultures and up to the 240h time point are almost identical, and, therefore, a similar intracellular metabolism is expected. The GU amounts supplemented at 240h exhibit minor influence on the PCA results, probably due to the late stage of the culture. It is important to note that the highest GU amounts, excluding the 240h time point, for the P3 and P4 cultures are supplemented at 192h, for P2 at 144h and for P1 at 96h. Interestingly, the PCA identifies similar nucleotide and NSD metabolism for P1 and P2, despite the fact that the two feeding cultures follow a considerably different schedule, with P2 receiving approximately twice the GU concentration at 144h compared to P1 at 96h. Therefore, the similar loadings indicate that feeding higher GU concentrations at 144h could replace the 96h feeding without affecting the overall NSD metabolism.

5.2.4 Investigating intracellular metabolism through flux balance analysis

The metabolites and amino acids presented in section 5.2.2, alongside with the remaining extracellular amino acid profiles and NSD demands, as calculated by the experimental data and Eq. 5.4 and 5.5, were utilized for the FBA-based estimation of the intracellular fluxes. The FBA was adapted to incorporate the effect of galactose and uridine addition in CHO cell metabolism. Overall, the differences are mainly time (interval) dependent, with the results for the second interval (day 6-8) that describes the late exponential phase, exhibiting supressed fluxes towards glycolysis and the TCA cycle for all experiments. However, differences between the experiments are also reported.

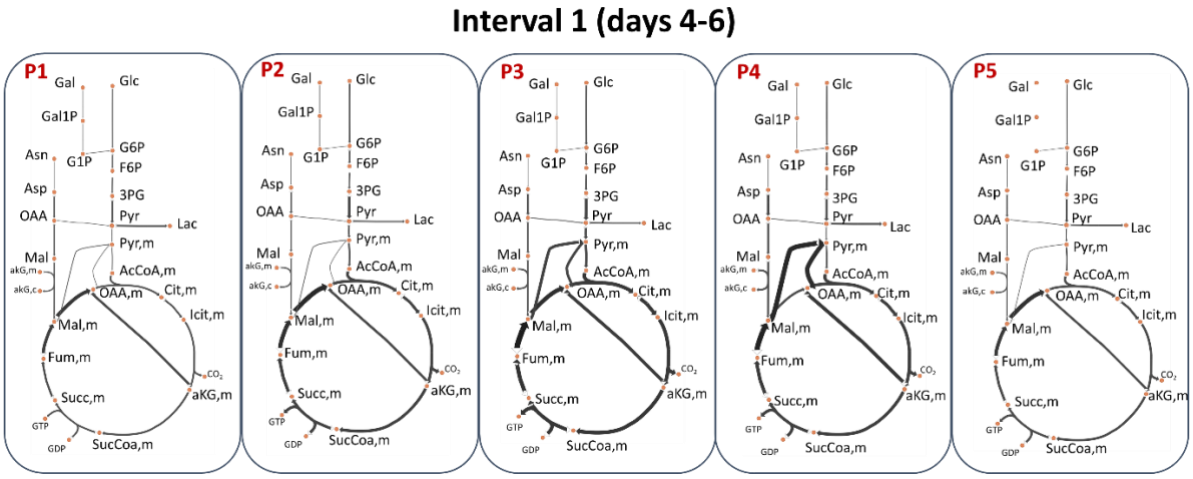

Figure 5.10: FBA results for all experiments and for the first interval. The "m" designation indicates mitochondrial concentration. G6P: glucose-6-phosphate, F6P: fructose-6-phosphate, Gal1P: galactose-1-phosphate, G1P: glucose-1-phosphate, 3PG: 3-phosphoglycerate, AcCoA: acetyl-coenzyme A, Cit: citrate, Icit: isocitrate, aKG: a-ketoglutarate, SucCoa: succinyl-CoA, Succ: succinate, Fum: fumarate, Mal: malate, OAA: oxaloacetate. Thickness of the lines denotes flux magnitude.

During the first interval, P1, that receives the highest galactose concentration on day 4, exhibits a considerable consumption of galactose towards glycolysis, approximately 24% of the total glycolytic flux, with the rest 76% supplemented by glucose uptake (Fig. 5.10). P4, that receives the second highest galactose concentration, shows a ~18% contribution towards glycolysis but with similar galactose flux compared to P1, while P2 and P3 a mere ~5% and ~3%, respectively. It is important to note that despite the 6% difference in galactose contribution between P1 and P4, the latter received > 3.5x lower galactose though the feed and exhibited similar galactose conversion rate towards G1P. The fact that not all galactose supplied was internalised by the cells, could indicate a saturation of galactose consumption in extracellular concentrations > 2.2 mM, at least for the early exponential phase.

Interestingly, modest amounts of galactose addition, like in P2 and P3, appear to be stimulating glucose uptake. The control, P5, exhibits the lowest glycolytic flux, while P2 and P3 exhibit the highest glycolytic flux, with up to ~80% higher values when compared to P1. P4 presents a moderate glycolytic flux, followed by P1 and P5. More specifically, P1 and P5 present a similar flux, 106.2 $nmol \cdot 10^6 cell^{-1} \cdot h^{-1}$ for P1 and 101.4 $nmol \cdot 10^6 cell^{-1} \cdot h^{-1}$ for P5, indicating that the addition of galactose in low levels, as in P2 and P3, stimulates glucose uptake in a more effective manner than the consumption of galactose itself. Indeed, P4 exhibits a 48% higher glucose intake compared to P1. Collectively, the results indicate that higher extracellular galactose concentrations result in elevated galactose intake, with a partial saturation observed at > 2.2 mM concentrations, and reduced glucose consumption. On the other hand, low galactose concentrations in the extracellular environment (~0.5 mM based on P2 and P3) lead to increased glucose consumption and, therefore, increased glycolysis metabolism. In addition, the levels of added uridine were not found to halt cell metabolism in the first interval.

Notably, P3 and P4 present an increased flux for the conversion of mitochondrial malate to pyruvate and subsequently to oxaloacetate. The synthesis of oxaloacetate from pyruvate is catalysed by the pyruvate carboxylase enzyme and exhibits an anaplerotic function for TCA cycle fuelling[562]. The rest of the experiments present significantly lower fluxes for these two reactions, with the control exhibiting no conversion of pyruvate to oxaloacetate. The elevated activity of pyruvate carboxylase in P3 and P4 indicates high levels of acetyl coenzyme A (AcCoA) that regulates the enzyme's activity[563]. It is important to note that the conversion of pyruvate to oxaloacetate seems to be stimulated by the conversion of malate to pyruvate. Apart from succinate conversion within the TCA cycle, fumarate is also produced as a by-product of arginine synthesis from aspartate and citrulline (reaction not shown in the network) and this, in its turn, leads to increased malate levels in the TCA cycle. All experiments exhibit a relatively similar flux for the conversion of malate to oxaloacetate within the TCA cycle, apart from P4 that shows a suppressed flux, as the majority of malate is transformed to mitochondrial

pyruvate. In the P1, P2 and P5 experiments, the feeding of the TCA cycle from pyruvate occurs mainly through its conversion to AcCoA. Additionally, each experiment exhibits a different conversion of cytosolic pyruvate towards lactate synthesis, ranging from ~80% for P1 to ~15% for P3, with the control showing ~40% conversion.

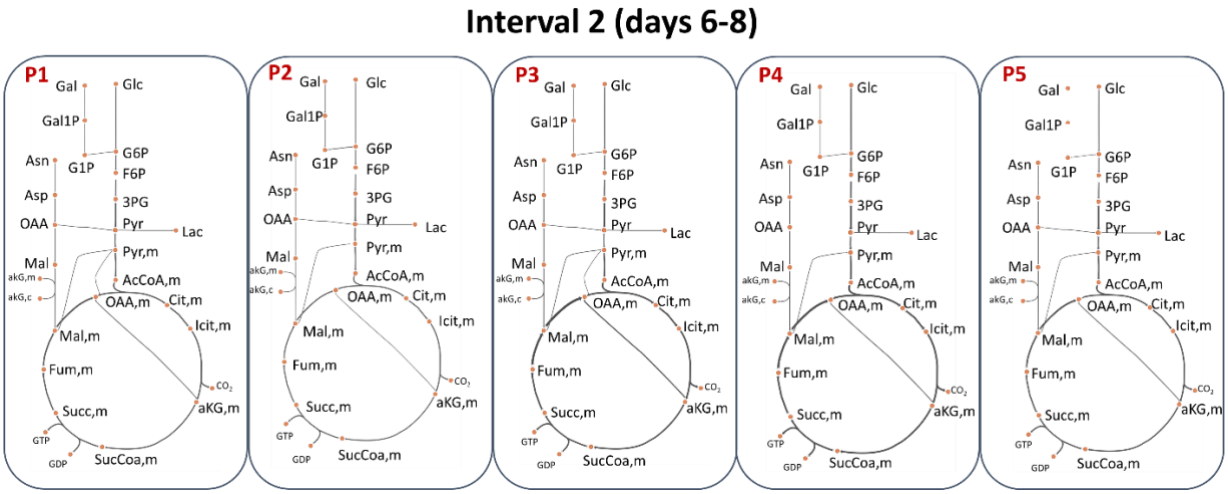

Figure 5.11: FBA results for all experiments and for the second interval. Thickness of the lines denotes flux magnitude.

During the second interval, P3 and P4 show the highest TCA fluxes, followed by P5 and, subsequently, by P1 and P2 (Fig. 5.11). Overall, the cells experience supressed fluxes during the examined timeframe. Compared to the first interval, all cultures present minor fumarate synthesis from aspartate and citrulline. The difference between P3-P4 and the remaining experiments is identified in the transformation of mitochondrial fumarate to malate within the TCA cycle, that exhibits a > 50% higher flux in the two experiments. All experiments exhibit reduced pyruvate to lactate conversion (< 13%) indicating that all cultures simultaneously approach the lactate metabolic shift point. During the second interval, galactose contribution towards G6P synthesis presents the highest value in P2, accounting for 56% of the total flux. The respective value for P1 is ~12%, while for P3 and P4 the flux is 4% and 6%, respectively. When compared to the first interval, the elevation in galactose contribution for P2, which is more than twice compared to the highest contribution observed during the first interval, is mainly attributed to the considerably lower glucose consumption. To elaborate, the G1P to G6P flux in P1 for the first interval, that presented a 23% (the highest) contribution, is 13.4 $nmol \cdot 10^6 cell^{-1} \cdot h^{-1}$, while the Glc to G6P flux is 44.4 $nmol \cdot 10^6 cell^{-1} \cdot h^{-1}$. On the other hand, the respective values for P2 in the second interval are 20.7 $nmol \cdot 10^6 cell^{-1} \cdot h^{-1}$ and 16.3 $nmol \cdot 10^6 cell^{-1} \cdot h^{-1}$. The results follow a reasonable pattern based on the supplemented galactose concentration for each experiment. Unlike in the first interval, the control presents the second highest glycolytic flux, right after P4, as the GU feeding experiments present a declining glycolytic flux (reduced by 25-60%), while the control remains more stable (15% reduction).

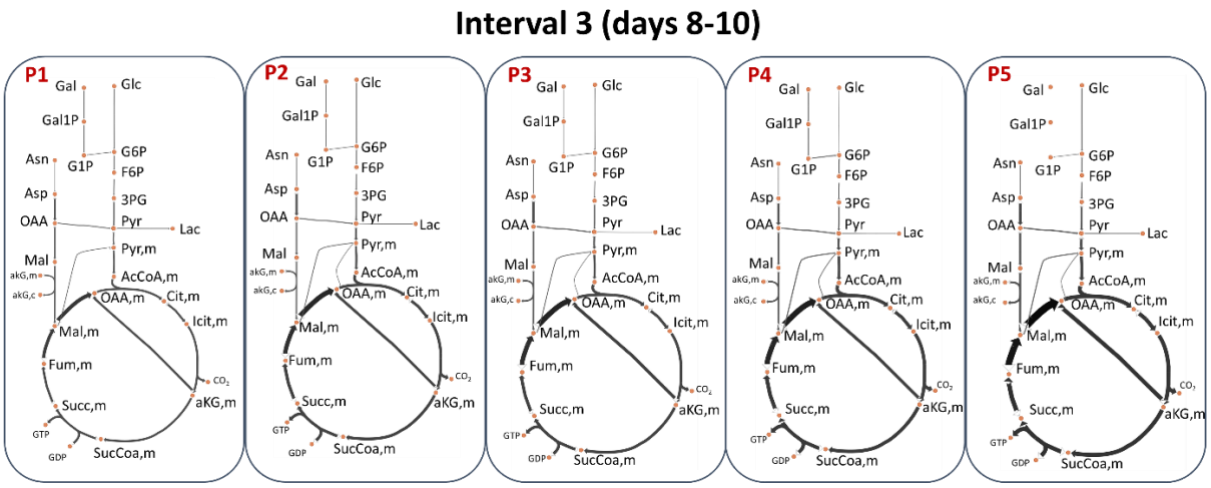

Figure 5.12: FBA results for all experiments and for the third interval. Thickness of the lines denotes flux magnitude.

During the third interval (days 8-10), all experiments exhibit an elevated total TCA flux, while lactate is consumed towards pyruvate synthesis (Fig. 5.12). Interestingly, the control experiment presents the highest TCA flux, ranging between 1.5x to 4.5x higher when compared to the GU feeding experiments. The control also presents the highest glycolytic flux compared to the rest of the experiments. The increased metabolic activity for the control experiment (specific growth on day 10 at 0.0125 h^{-1}) can be potentially attributed to the toxic effect of the accumulated uridine on cell growth for all GU feeding experiments (maximum specific growth on day 10 for the GU feeding experiments at 0.0084 h^{-1}). Unlike the first interval, the elevated mitochondrial malate levels do not lead to pyruvate synthesis. The majority of malate, synthesized from fumarate or from aspartate/oxaloacetate in the cytosol, directly converts to mitochondrial oxaloacetate. In its turn, the flux of oxaloacetate synthesis splits almost equally to citrate and alpha-ketoglutarate (αKG) synthesis. Surprisingly, galactose contribution towards glycolysis exhibits a significant difference between P3 (17%) and P4 (82%), despite the fact that the two cell cultures follow a similar feeding pattern up to that time point. The respective value in both P1 and P2 is ~20%.

5.2.5 Intracellular characteristics differentiating galactose and uridine feeding cultures from the control

To further elaborate on the differences between the GU experiments and the control, all fluxes were subjected to feature selection, a machine learning algorithm for identifying a representative subset of reactions that could efficiently capture the majority of data variation. Unlike PCA that creates latent variables, the least-squares feature selection enables the identification of a subset of the original features in an unsupervised manner, improving that way the results interpretability. A subset of 14 reactions (Table 5.4) was designated by the feature selection method.

Table 5.4: Representative subset of metabolic reactions identified by the least-square feature selection method. G6P: glucose-6-phosphate, F6P: fructose-6-phosphate, Gln: glutamine, Glu_{cyt}: cytosolic glutamate, Fum: fumarate, G1P: glucose-1-phosphate, OAA: oxaloacetate, Met: methionine, Ser: serine, αKb: alpha-ketobutyrate, $NH4^+$: ammonium, Pi: phosphate ion, GAP: glyceraldehyde 3-phosphate, 3PG: 3-phosphoglycerate, Trp: tryptophan, Ala: alanine, AcCoA: acetyl-coenzyme A, CoASH: coenzyme-A, DHAP: dihydroxyacetone phosphate, SucCoa: succinyl-CoA, HCO_3: bicarbonate.

#Reaction	Reaction
R68	UDPGlc → Glycogen + UDP
R2	G6P ↔ F6P
R62	Citrulline + ATP + Asp →Fum + Arg + AMP + 2Pi
R67	G1P + H^+ + UTP → UDPGlc + Pi
R24	OAA_{cyt} + $NADH_{cyt}$ ↔ Mal_{cyt} + NAD^+
R95	F6P + Gln + AcCoA + UTP --> UDPGlcNAc + Glu_{cyt} + CoASH
R53	Met + Ser + ATP → αKb + $NH4^+$ + AMP
R103	→ IgG
R5	GAP + NAD_{cyt} + ADP ↔ 3PG + $NADH_{cyt}$ + ATP
R6	3PG + ADP → Pyr_{cyt} + ATP
R41	Trp + 2CoASH + 2NAD → Ala + 2AcCoA + 2NADH + $4CO_2$
R3	F6P + ATP → DHAP + GAP + ADP
R40	αKb + CoASH + NAD^+ + HCO_3 + ATP → SucCoA + ADP + $NADH_{cyt}$ + CO_2
R90	18AcCoA + 18ATP + 14NADPH → Cholesterol + $9CO_2$ + 18ADP + $14NADP^+$

As expected, the designated subset includes reactions, the flux of which, differs significantly among the experiments and examined intervals. Interestingly, the subset includes three reactions (R67, R68 and R98) directly related to NSD synthesis and therefore to protein glycosylation. Overall, R64 maintained higher levels in the GU feeding experiments when compared to the control, due to both the elevated G1P levels from galactose uptake and increased UTP concentrations from uridine supplementation. Similarly, R65 that describes the synthesis of glycogen, a glucose polymer, from UDPGlc, was upregulated in the GU feeding experiments as higher UDPGlc concentrations were achieved due to uridine supplementation. Interestingly, glycogen was recently found to induce human ovarian cancer cell proliferation and metastasis due to its intervention in the glycolysis pathway[564]. Through the activity of glycogen phosphorylase, glycogen can release G1P molecules that are fed back to the glycolysis pathway. Glycogen was also recently identified as the major glycolysis fuel in the bulk body of mice[565]. Therefore, glycogen can supplement glycolytic flux in the GU experiments, during the reduced glucose and insufficient galactose consumption periods. Moreover, glycogen is involved in

biomass production and therefore the excessive levels in the GU feeding experiments can assist towards ensuring growth. Unlike R67 and R68, the flux of R98 was found in similar levels for all experiments, partially due to the similar HCPs and mAb glycans mannosylation.

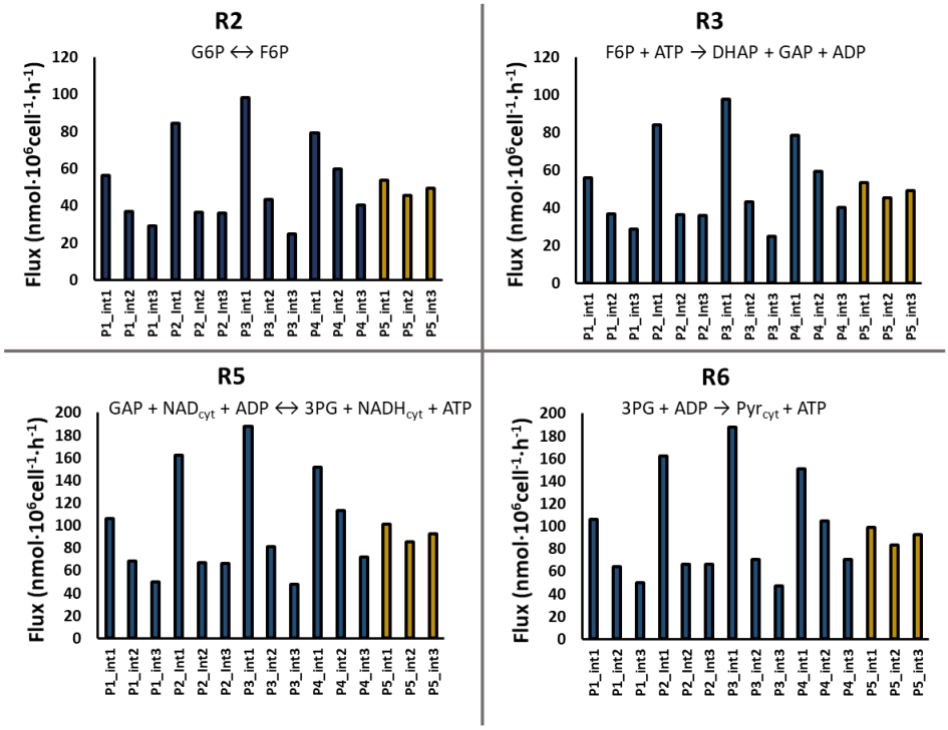

Figure 5.13: Estimated fluxes for glycolysis reactions for all experiments and intervals. The bars with the blue colour are for the GU feeding experiments and with the yellow are for the control (P5).

In addition, when comparing the fluxes between the control and GU feeding experiments, the glycolysis reactions R2, R3, R5 and R6 present a significantly different profile between the two sets of experiments (Fig. 5.13). The control maintains a relatively constant glycolytic flux throughout the culturing period, unlike the GU feeding experiments that present a high initial flux that is quickly reduced to levels lower than the control. As mentioned earlier, P2 and P3 that received the lowest initial GU concentration, present the highest glycolytic fluxes for the first interval, indicating that a moderate addition of galactose during the early culturing period can lead to increased glycolytic flux. On the other hand, high initial galactose concentrations (P1) do not positively affect the total glycolytic flux. Through the rapid downregulation of glycolysis, the cells in the GU experiments manage to reduce the discrepancies from the control and therefore, follow a reasonably similar growth and productivity, while achieving the desired IgG glycosylation profile.

5.2.6 IgG *N*-linked glycosylation

The *N*-linked glycosylation analysis of the produced IgG indicates a significant difference ($p < 0.05$) between the control and the GU feeding experiments (Fig. 5.14D). All GU feeding experiments present

similar glycoform distributions in each day of the cell culture, with considerably lower levels of GnGnF compared to the control (Fig. 5.14 A, B, C). Man5 is not affected by the GU addition, while GnGn is moderately supressed by the feeding. All galactosylated forms, including Aa6Gn, Aa3Gn, Aa6GnF, Aa3GnF, AA and AAF, reasonably exhibit higher distributions in the GU feeding experiments. The non-fucosylated AGn and AA glycans presented the most subtle increase, while the respective fucosylated variants showed a considerably elevated distribution. Whilst core-fucosylation presented a slightly increasing trend through the culturing period, no significant differences were observed between the different time points and experiments. Despite the differences in the UPDGal levels observed for each experiment at each time point, and the P1-P2 and P3-P4 similarities in the NSD metabolism (Fig. 5.9), no significant differences were observed between the glycans of the GU feeding experiments at all time points. Moreover, galactosylation was found to decrease during the cell culturing period in all experiments. The gradual reduction in IgG galactosylation can be attributed to ammonia accumulation that has been found to inhibit galactosyltransferase expression and activity[247, 257, 259, 261]. Finally, as expected, all GU feeding experiments were found to satisfy the mAb galactosylation constraint set for the Design Space, as defined in Eq. 3.5, ranging between ~258 $mg \cdot L^{-1}$ for P4 and 292 $mg \cdot L^{-1}$ for P2 and P3, with the control (P5) presenting a value of 207 $mg \cdot L^{-1}$. Therefore, considering the satisfaction of the mAb titre constraint that was reported in section 5.2.1, the model was found to result in trustworthy feeding experiments for achieving the desired titre and glycosylation profile.

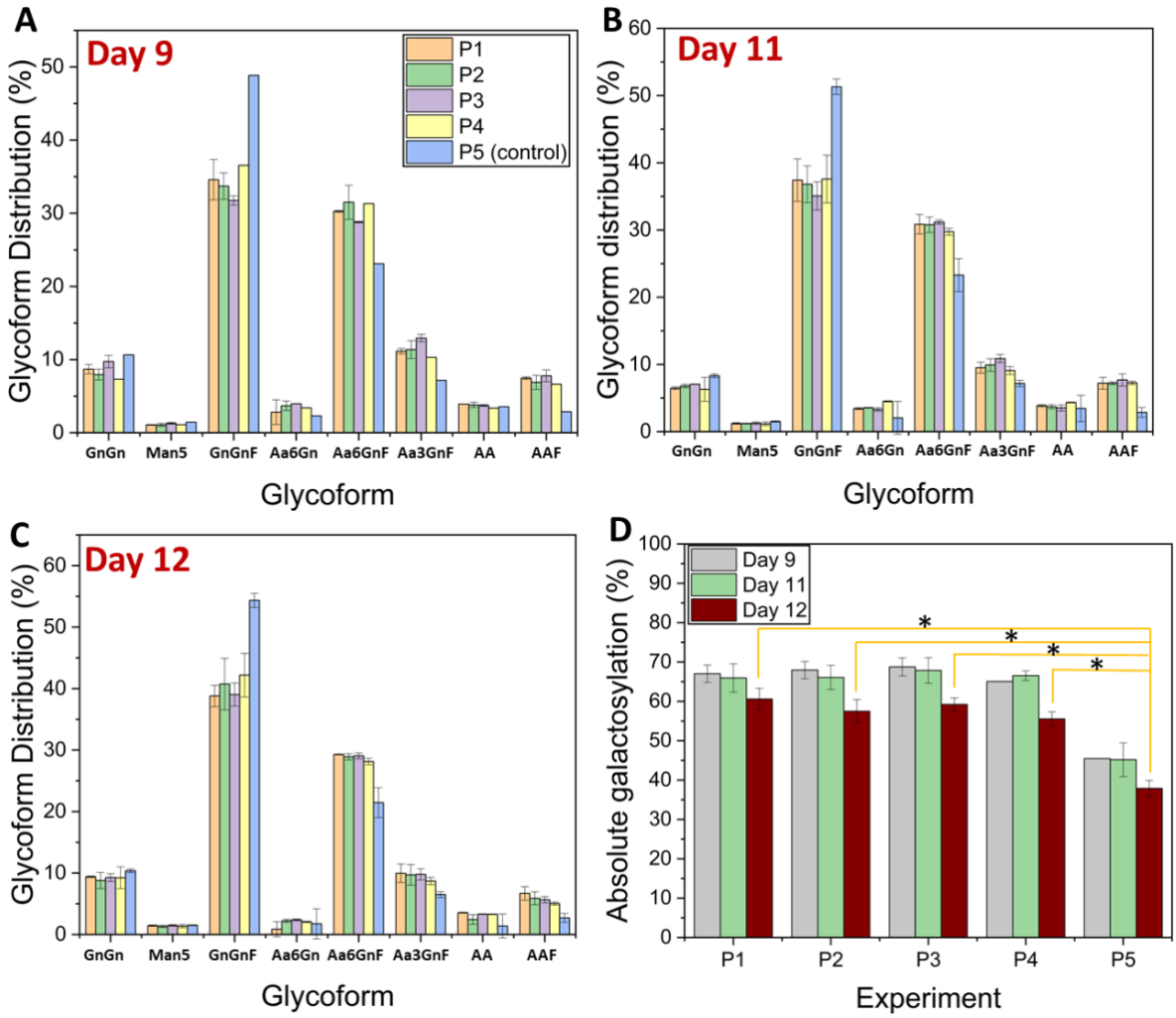

Figure 5.14: IgG *N*-linked glycosylation analysis of all experiments for days (A) 9, (B) 11 and (C) 12 of the cell culturing period (corresponding hours: 216h, 264h and 288h). The legend shown in panel A is common for panels B and C as well. (D) Absolute galactosylation for each experiment and for days 9, 11 and 12. Significant differences ($p < 0.05$) between the absolute galactosylation for day 12 are also denoted by the asterisk.

5.2.7 Does UDPGal synthesis limit antibody galactosylation?

As shown in Fig. 5.14, all GU feeding experiments presented similar mAb galactosylation, that was found to be significantly higher compared to the control. Despite the different amounts of galactose and uridine supplemented in each culture, UDPGal concentration was measured in comparable levels between all GU feeding experiments in the last three days of the cultures, suggesting a possible saturation in the reaction of UDPGlc conversion to UDPGal. The enzyme catalysing the UDPGlc to UDPGal conversion, namely galactose-1-phosphate uridylyltrasferase, presents various reported *in vitro* Michaelis-Menten saturation constants, with reported values for the human homolog varying between 0.08 mM [566] to 0.6 mM [567] for UDPGlc as a substrate. Therefore, a regulation of the reaction is possible if the *in vivo* kinetic constant lies in the lower range of reported values, as UDPGlc concentration is > 0.5 mM towards the end of the culture for all GU feeding experiments.

The uniform galactosylation levels of all experiments indicate that the upper limit in antibody galactosylation is unrelated to the differences in the cytosolic levels of UDPGal. More specifically and for the 216h (day 9) time point when the highest galactosylation levels are observed, and up to which the intracellular UDPGal concentration is considerably different between the experiments, the glycans distribution of the accumulated mAb among the different experiments was surprisingly similar.

To that end, plausible bottlenecks for the further increase in IgG galactosylation include: 1) the UDPGal transport rate in the Golgi apparatus from the SLC35A2 transporter that could have reached its maximum velocity, 2) the insufficient b4GalT availability for galactosylation, with the bottleneck being on the transcription or translation level, 3) the consumption of the excess intra-Golgi UDPGal for cellular glycosylation and 4) the combination of two or more of the aforementioned reasons. Whilst insufficient concentration of b4GalT in the Golgi is the most likely reason for the observed IgG galactosylation limit, the low K_m values for UDPGal transport (~2.4 μM) reported in literature[568] compared to the considerably higher intracellular UDPGal concentrations reported herein (100 – 1,600 μM), cannot rule out the possibility of transport saturation. Similarly, the high degree of *O*-linked and *N*-linked galactosylation in the CHO glycome [513] can significantly influence the intra-Golgi UDPGal levels.

5.2.8 *N*-linked glycosylation of the intracellular HCPs

The MALDI-TOF-MS analysis (Fig. 5.15A) of intracellular HCPs resulted in the identification of 49 *N*-linked glycans (without considering the potential isomeric structures) throughout the five experiments. An elaboration on the plausible isomers for each identified *m/z* would require further fragmentation through the use of MS/MS or a prior step of glycans separation, preferably through capillary electrophoresis that exhibits high efficiency in isomer separation[569].

Table 5.5: Specific parts of HCPs glycosylation for all experiments. Galactosylation and sialylation account for the total amount of molecules present in each glycan. High mannose glycans refer to structures with equal to or more than 5 mannose molecules.

Experiment	Galactosylation (%)	Sialylation (%)	Fucosylation (%)	High Mannose (%)	> Bi-antennary (%)
P1	52.66	10.06	26.35	68.99	9.22
P2	99.23	46.64	34.52	61.11	21.50
P3	72.95	22.28	28.55	65.64	14.13
P4	77.56	21.62	26.20	65.22	14.71
P5 (control)	39.80	16.96	16.96	70.53	9.33

Fig. 5.15B presents the profile of the 10 glycans with the highest % distribution among the identified structures. Immature high mannose glycan structures (Man ≥ 5) exhibit high levels in all experiments, accounting for 61 - 70% of the total distribution (Table 5.5). Man6 is the most abundant glycan for all experiments, with the exception of P2 in which the Man8 and Man9 glycans are the most prominent. Overall, feeding experiments present lower distributions for Man5, Man6 and Man7. Similarly, the distribution of non-galactosylated structures, including GnF, GnGnF and GnGnGnF, is suppressed in the GU feeding experiments, with the total galactosylation exhibiting up to 2.5-fold higher levels in the GU feeding experiments when compared to the control (Table 5.5). Thus, the supplementation of galactose and uridine and the accumulation of higher UDPGal pools in the intracellular environment leads to elevated HCP galactosylation levels, an observation that could potentially inhibit the further increase of antibody galactosylation.

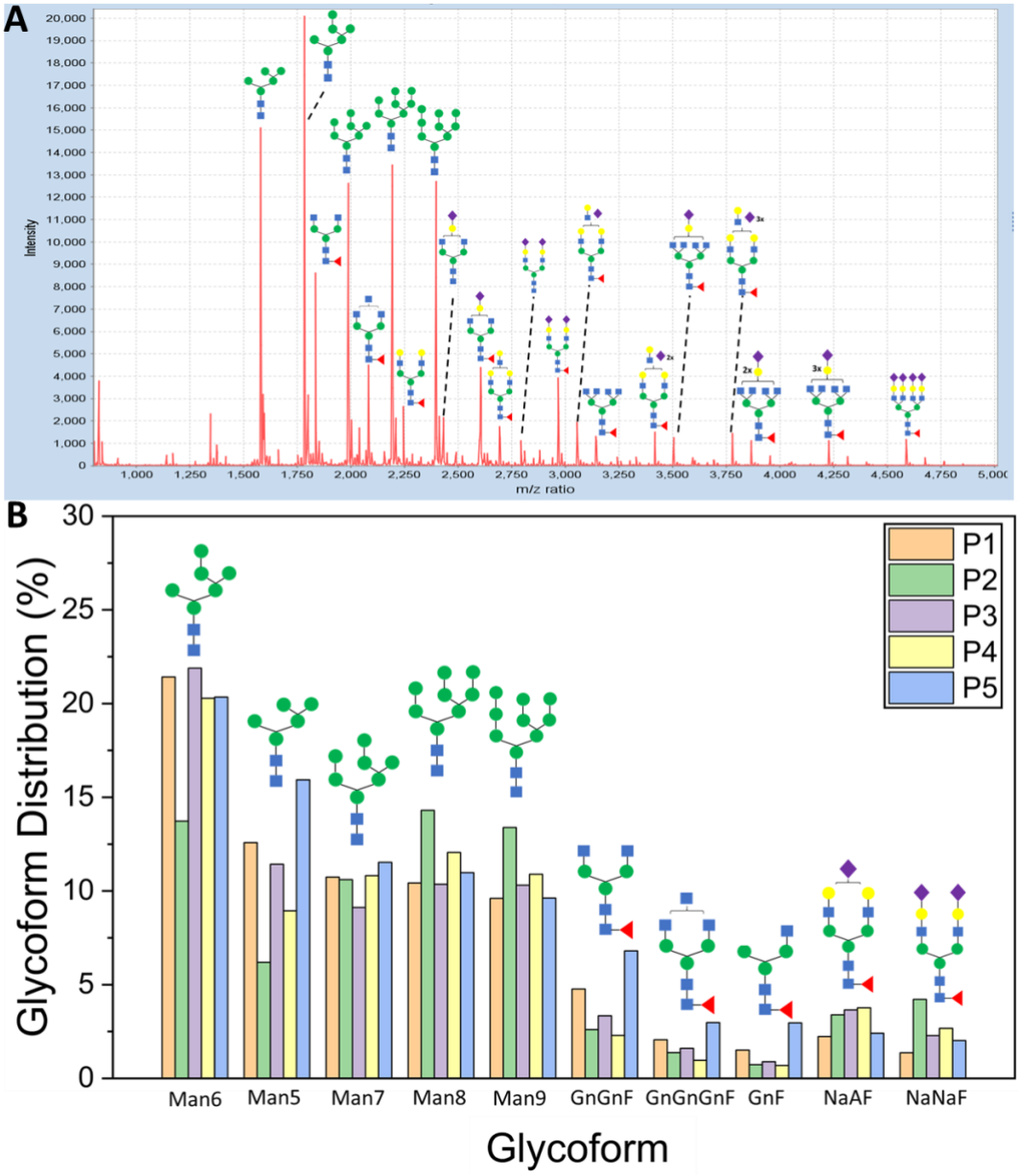

Figure 5.15: (A) MALDI-TOF-MS analysis of the intracellular HCPs of CHO cells in the P5 (control) experiment and at harvest. Selected glycan structures are also shown. (B) HCPs glycoform distribution for the 10 glycan structures that present the highest abundance at harvest.

Sialylation presents a similar profile with galactosylation in all experiments, except for P1 where the sialic acid content is reduced compared to the control (P5) despite the increase in galactosylation. Whilst a strong dependency of sialylation on galactosylation would be expected due to the former enzyme acting only on galactosylated substrates, overall, HCP galactosylation and sialylation exhibit a weak correlation with a PCC ~0.65 (Fig. 5.16). The aforementioned observation could be attributed to the considerable number of identified glycans in the GU feeding experiments that present several terminally galactosylated molecules that lack of further sialylation processing. Whilst the control experiment presented the highest GDPFuc concentrations, HCP fucosylation is elevated in the GU

feeding experiments, partially due to the reduction of the high mannose structures and consequently, the further processing of the glycans. The increased fucosylation observed in the feeding experiments can also explain the reduced GDPFuc concentrations compared to the control. Fucosylation is strongly correlated to galactosylation (PCC ~0.91), while all considered parts of glycosylation present a negative correlation to high mannose structures, as expected. Finally, highly branched glycans (> bi-antennary) present a strong positive correlation to galactosylation and a slightly weaker interdependence with fucosylation and sialylation. The positive association between highly branched and sialylated glycans is attributed to the monotonic dependency of sialylation on the number of branches in the glycan molecule.

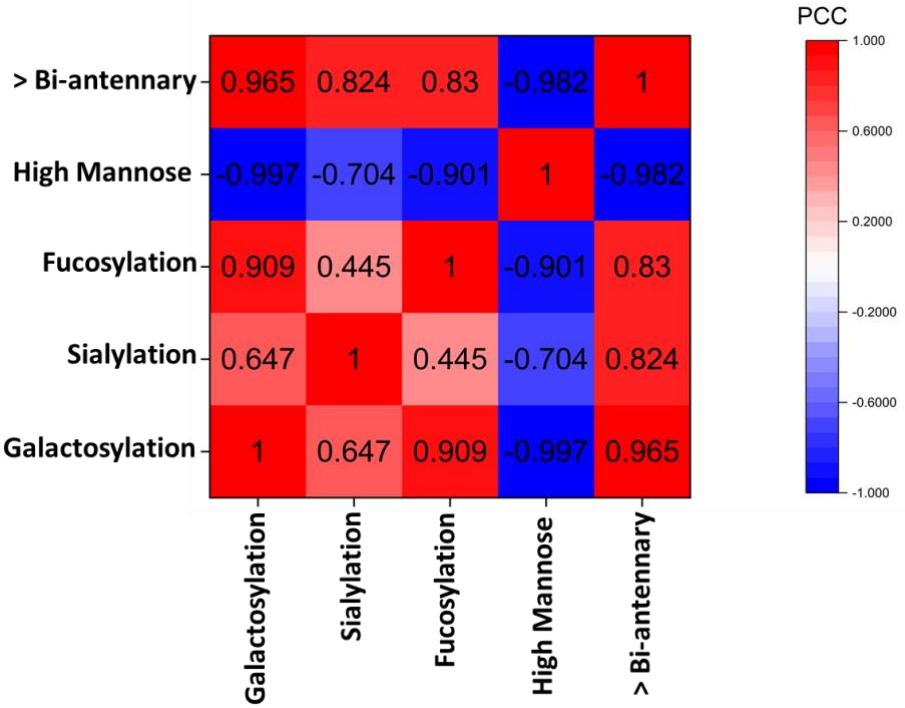

Figure 5.16: Heatmap of the correlation (PCC) between different parts of HCPs *N*-linked glycosylation.

5.2.9 Model-based investigation of galactosylation bottlenecks

A glycosylation model that enables the evaluation of HCPs-IgG interaction

In order to investigate the previously proposed bottlenecks of galactosylation *in silico*, the kinetic glycosylation model was adapted to more holistically describe the secretion process and include the following steps: 1) Golgi *N*-linked glycosylation of intracellular HCPs (simultaneously with IgG glycosylation), 2) *N*-linked glycosylation in the ER for both HCPs and IgG, 3) transport from ER to Golgi and 4) transport from Golgi to the cytosolic membrane for IgG. The model was also modified to account for all four compartments of the Golgi, assuming the vesicular transport Golgi model. No

transport from the Golgi to the membrane or to other cellular compartments was considered for HCPs, as they represent the entire intracellular proteome and they exhibit various localizations within the cell. The reaction network for HCP glycosylation was based on the previously identified glycans in all feeding experiments, including the control.

Substrate competition between HCPs and the IgG product was also considered as a single model and was used to simulate the glycosylation of both proteins. The model was fitted against the experimental data of both HCPs and IgG of the control experiment, with the enzymatic concentrations estimated against only the HCP profile, since they are the natural products of the cells. Model simulation results compared to experimental data are shown in Fig. 5.17. The model exhibits a good agreement with the data for both the intracellular HCPs and the IgG. More specifically, the model closely matches the distributions of the most abundant glycans, such as the high mannose structures (Man5-Man9) for the HCPs, while the fitting is less accurate for the lower abundance HCP glycans. To the best of our knowledge, this is the first modelling effort that successfully and simultaneously captures the *N*-linked glycosylation with the entire intracellular CHO glycome alongside with the glycoprofile of the secreted recombinant protein. The inclusion of HCP glycosylation enables the investigation of HCP-IgG interaction within the glycosylation context, as the two protein classes act as competitive substrates for the same glycosyltransferases and consume NSDs towards their glycosylation from the same intra-Golgi pool.

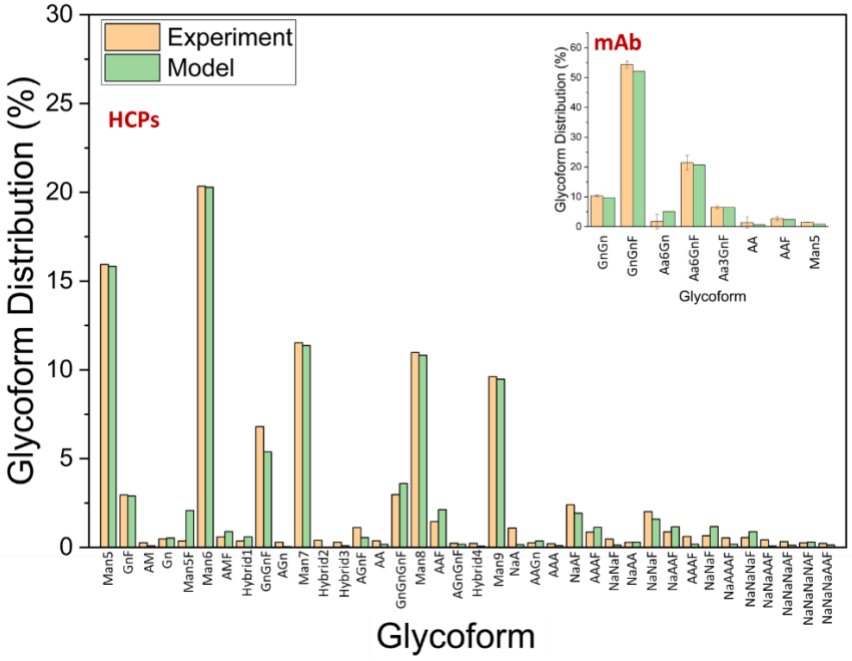

Figure 5.17: Model agreement between the experimental data of the control experiment at harvest for the intracellular HCPs and the secreted IgG.

UDPGal transport limits antibody galactosylation

The inclusion of the different steps within the secretion pathway (Fig. 5.18D) enables a more detailed representation of the intracellular distributions for each glycoform. A snapshot of the IgG glycan distributions for the different cellular compartments and for the control experiment at harvest is shown in Fig. 5.18A. As expected, the distribution of each glycan differs considerably between the examined compartments. GnF and Aa3GnF glycoforms of the IgG present a significantly different distribution compared to the last compartment of the Golgi (TGN). More specifically, GnF distribution in the TGN compartment at harvest is ~78% (Fig. 5.18B), considerably higher than that measured on secreted IgG (~52%). The relationship between the secreted distribution of each glycan and the respective distribution in the various compartments of the Golgi is important for developing a better understanding of the cellular state.

Given the spatial evolution of the glycoform that can be simulated using the model, it became possible to back calculate the required UDPGal concentration in the Golgi for acquiring the experimentally measured IgG glycoprofile when galactose and uridine are supplemented to the culture. As shown in Fig. 5.18C, the experimentally identified increase of intracellular UDPGal concentration in the GU feeding experiments surpasses the model-estimated concentration in the Golgi compartments where galactosylation occurs. Whilst a discrepancy between the increase in the intracellular and the estimated intra-Golgi concentrations is observed for all Golgi compartments and all time points, the large standard deviations (calculated based on the 95% confidence intervals of the estimated concentrations) moderate the significance in the difference between the two concentrations. Whilst a regulation of UDPGal transport is expected, the results clearly indicate that the intra-Golgi UDPGal concentration does not reflect the changes in the NSD intracellular pools and, therefore, the transport limits antibody galactosylation.

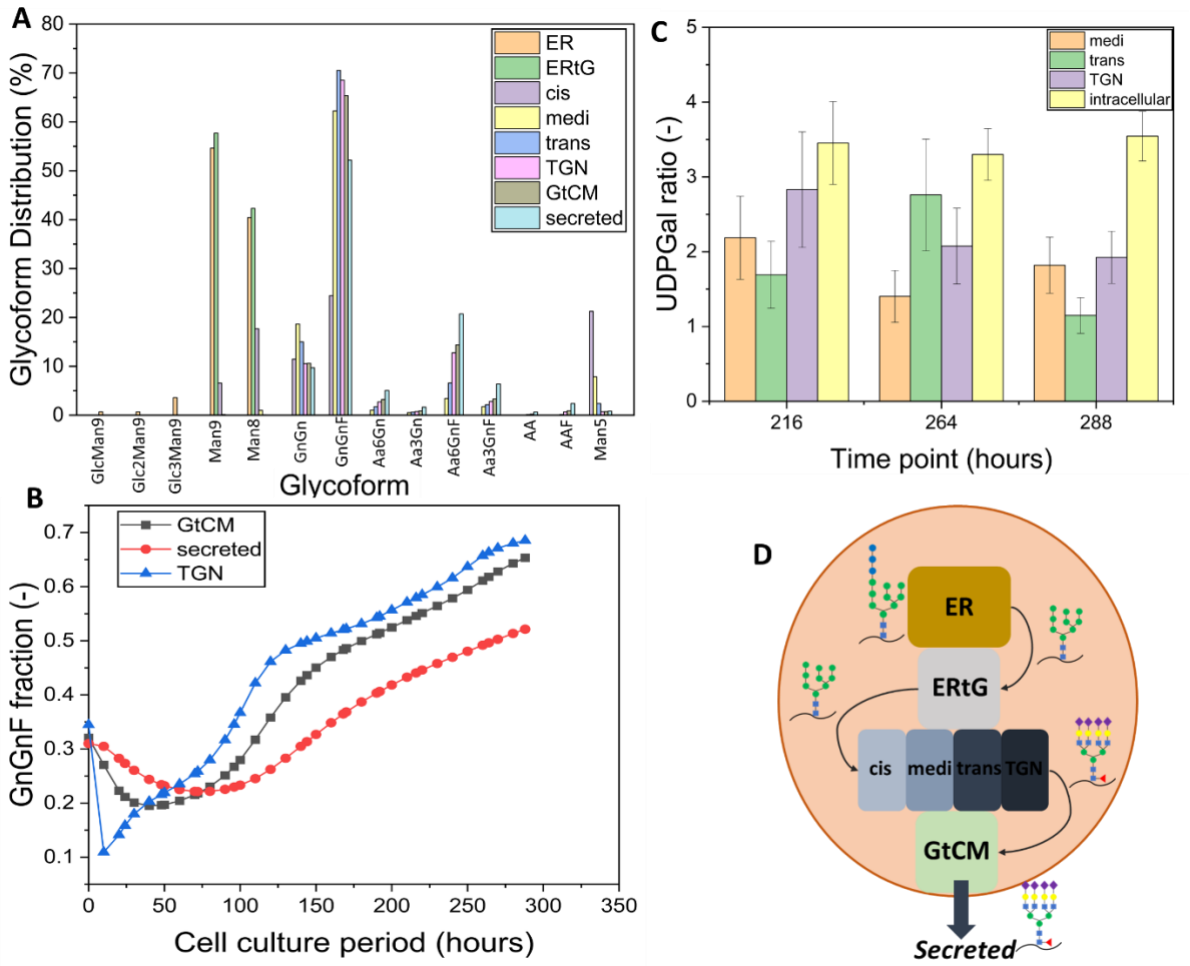

Figure 5.18: (A) Snapshot of the major glycans distribution at each cellular compartment for the secreted mAb in the control experiment at harvest, based on model simulation. (B) The fraction of GnGnF in three subcellular compartments during the entire cell culture period and for the control experiment, based on model simulation. (C) Ratio between the UDPGal concentration of the P4 and P5 experiments (P4:P5) in different compartments of the Golgi where galactosylation occurs, and in the intracellular environment. The intra-Golgi concentrations were calculated based on model parameterization in order to fit the experimentally observed glycan profile. Intracellular UDPGal concentrations were determined experimentally as reported in section 5.2.3. For the model-estimated concentrations the errors are based on the 95% confidence intervals and for the experimentally measured concentrations on the standard deviation. As glycan distributions and UDPGal levels are similar between the GU feeding experiments, only minor discrepancies are expected for P1, P2 and P3. (D) Parts of the secretory pathway examined by the model. Abbreviations: ERtG represents the ER to Golgi transport and GtCM represents the transport between the Golgi and the cellular membrane.

If no saturation of NSD transport between the cytosol and the Golgi occurred, the intra-Golgi concentration of UDPGal would present the same increase in the feeding experiments as the increase in intracellular concentration, when compared to the control. This would increase mAb galactosylation by ~ 43% (from 30% to ~ 44%) from the levels calculated based on the experimental data for P4 at harvest, for example, as shown in Fig. 5.19A. This demonstrates that regulation and saturation of NSD transport can have a profound effect on antibody galactosylation. The model further enables the

simulation of the potential effect of b4GalT overexpression by glycoengineering. Fig. 5.19B compares the effect of enzyme overexpression to that of increased UDPGal synthesis and clearly demonstrates that an increase in b4GalT levels has a more significant effect on galactosylation levels.

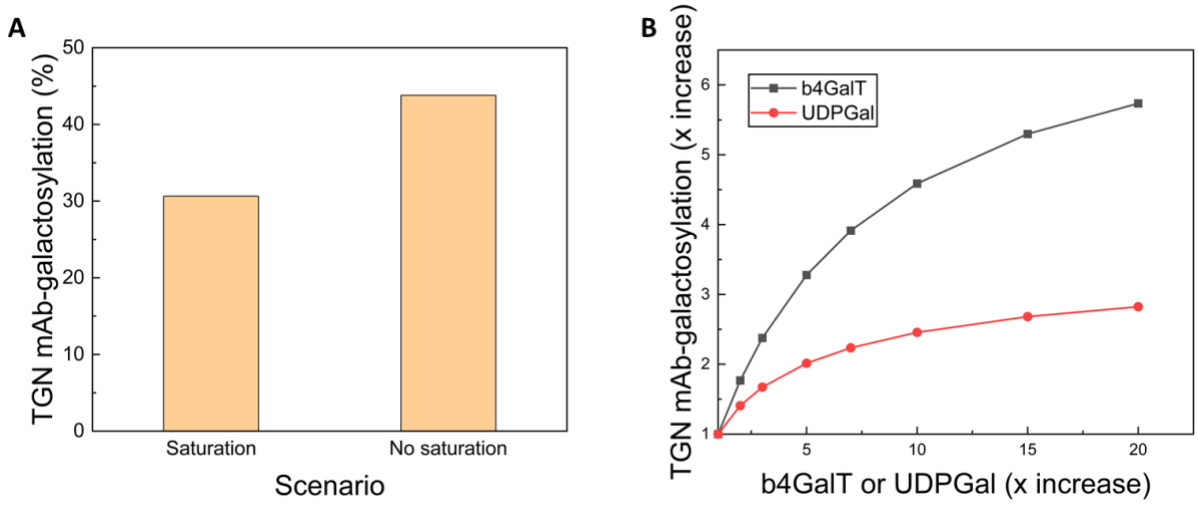

Figure 5.19: (A) Comparison of mAb galactosylation in the TGN compartment of the Golgi for the P4 experiment at harvest, when the UDPGal transport the intra-Golgi concentration is saturated and when the fold change in the intracellular NSD concentration between the control (P5) and the P4 experiment is reflected to the intra-Golgi UDPGal concentration (no saturation). (B) Increase in mAb galactosylation compared to its nominal value with the increase of either b4GalT concentration or UDPGal concentration.

HCP galactosylation disproportionally increases compared to antibody galactosylation

The consumption of UDPGal towards cellular *N*-linked glycosylation was further investigated. As shown in Fig. 5.20, the IgG galactosylation rate is constantly higher compared to the respective HCP rate through the entire cell culture period and for all three compartments where galactosylation occurs, with the exception of the last intervals in the TGN compartment in which the HCP rate slightly surpasses the respective value for the IgG. However, the HCP galactosylation rate is within the same order of magnitude as the IgG galactosylation rate, indicating that, indeed, HCPs are influencing IgG galactosylation. The model-based results are in agreement with the experimentally observed increase of HCP galactosylation in the GU feeding experiments. To that end, the UDPGal consumption rate for both HCP and mAb *N*-linked glycosylation was calculated based on the experimentally quantified glycan distributions at harvest. In order to further investigate the impact of HCP galactosylation on IgG glycans distribution, HCP galactosylation was blocked in the model through the knockout of the respective reactions. The *in silico* blocking of HCP galactosylation, resulted in a ~ 33% increase of IgG galactosylation, demonstrating the dependence of IgG on the cellular requirements for *N*-linked glycosylation.

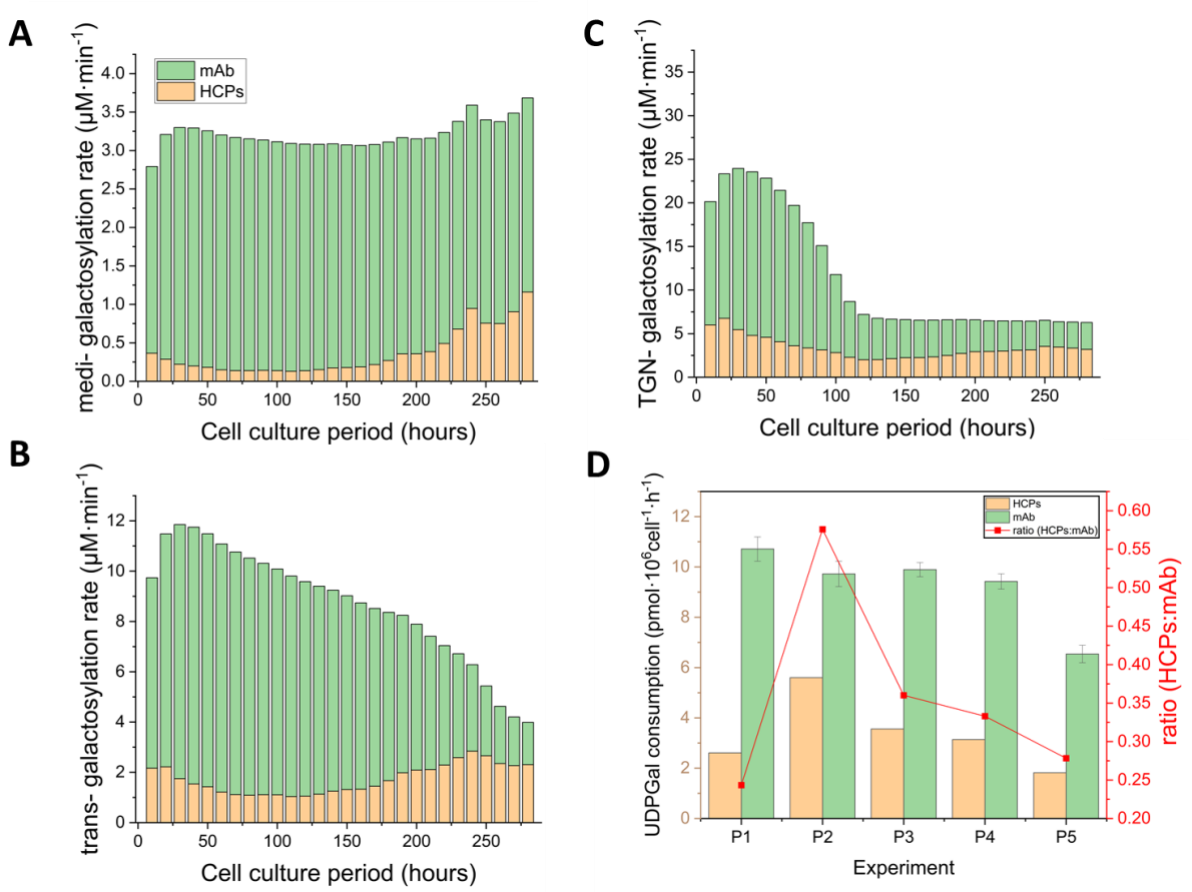

Figure 5.20: *In silico* galactosylation rates for both the IgG and the HCPs for the entire cell culture period of the control experiment and for (A) the medi, (B) the trans and (C) the TGN compartments. (D) UDPGal consumption for each experiment at harvest, based on HCPs and IgG glycan analysis. The ratio between the two consumption rates is also reported.

As presented in Fig. 5.20D, while IgG galactosylation maintains a higher consumption rate of UDPGal for all experiments, the rate towards HCP *N*-linked glycosylation presents a larger increase in the feeding experiments compared to the mAb. For that reason, the HCP:mAb ratio for UDPGal consumption is up to ~ 2x higher in the P2 experiment compared to the control, indicating that HCPs utilize any excess of UDPGal synthesised in the feeding experiments in a better way than the IgG, with the exception of P1 that presents a similar ratio with the control. However, it is important to note that despite the experimentally and model-based identified dependency between cellular and recombinant protein glycosylation, the consumption of NSDs towards HCPs galactosylation is not expected to act as the main bottleneck of IgG galactosylation, rather than weaken the effect of GU feeding on IgG galactosylation.

5.3 Discussion

In this chapter, an integrated experimental and computational analysis was carried out in order to evaluate and determine the reasons behind the experimentally observed upper threshold in mAb galactosylation. In addition, the effects of galactose and uridine addition on intracellular metabolism

were also examined. For that reason, five fed-batch experiments with different concentrations of galactose and uridine supplemented during the cell culture period were performed. The glycosylation profile of both the recombinantly produced IgG and the intracellular HCPs of the cell was quantified. The IgG glycoprofile verified the observation that an upper threshold in IgG galactosylation exists, regardless of the supplemented galactose and uridine concentrations. The analysis of the HCPs glycoprofile enabled the investigation of the resources that cells take up for the glycosylation of the intracellular proteins. Amino acid analysis of all experiments was utilized in order to identify discrepancies between the feeding and the control experiments through flux balance analysis. In addition, the kinetic glycosylation model was modified in order to account for additional parts of the secretory pathway and importantly to simultaneously simulate both HCP and antibody glycosylation.

The GU feeding experiments presented similar growth compared to the control, with the only statistically significant difference for the IVCD observed between the P1 and P3 experiments (Fig. 5.1B). Overall, the feeding experiments presented a slightly improved mAb productivity compared to the control, without the difference being statistically significant (Fig. 5.2B). As expected, the GU feeding experiments presented increased glucose concentrations (Fig. 5.3A) and therefore reduced glucose specific consumption rate (Fig. 5.5). Pyruvate was consumed within the first four days of the culture in all experiments (Fig. 5.4B), while the feeding of galactose and uridine resulted in increased ammonia (Fig. 4D) and reduced glutamine levels (Fig. 4E). In a similar manner, the GU feeding experiments presented higher ammonia synthesis rates, with only P1 and P2 being significantly different compared to the control (Fig. 5.5). Overall, P1 presented the lowest IVCD and antibody titre of all experiments at harvest, indicating that an elevated GU feeding in the early culturing period (day 4) can halt cell growth. On the contrary, P2 that incorporated low GU concentrations on the feeding at day 4, but higher concentrations at day 6, showed a similar growth profile with the rest of the experiments that included elevated feedings only on days 8 and/or day 10 of the culture period.

As expected, UDP-sugars presented elevated profiles in the GU feeding experiments (Fig. 5.6). UDPGal was measured in similar levels for all GU feeding experiments towards the end of the culture. Therefore, a saturation in UDPGal synthesis rate was considered as a possible scenario but was not found to affect IgG glycosylation based on the glycan analysis results. Surprisingly, GDPMan and GDPFuc concentration were measured in lower levels in the GU feeding experiments (Fig. 5.6) and the aforementioned results were partially attributed to the different requirements for cellular glycosylation. UTP and CTP concentrations were increased due to uridine addition, while GTP and ATP levels followed an overall declining profile for the feeding experiments after the 96h time point, while both nucleotide triphosphates were maintained largely stable in the control culture (Fig. 5.7). In addition, a PCA on the NSD and nucleotides data identified that the intracellular profiles of P1/P2 and

P3/P4 clustered closely together (Fig. 5.9). The inclusion of P1/P2 in the same cluster indicates that the elevated GU feeding at day 4 (P1) can be sufficiently replaced from a GU feeding at day 6 of the culturing period, without affecting NSD metabolism, and also ensuring an improved cell growth.

FBA was used in order to further study the intracellular metabolism of the cultures. In the first examined interval (day 4-6), galactose feeding was found to proportionally increase galactose consumption (Fig. 5.10). However, for concentrations higher than 2.2 mM, galactose uptake was only marginally increased. Additionally, the elevated galactose consumption was found to supress the total glycolytic flux, while the P3 and P4 experiments exhibited elevated activity of the pyruvate carboxylase enzyme. P1 that received the higher GU concentrations at day 4 of the culture showed the highest pyruvate to lactate conversion (80%). The second interval (day 6-8) showed a reduced total TCA flux and a suppressed pyruvate to lactate conversion, indicating that all cultures were approaching the lactate metabolic shift point (Fig. 5.11). In the third interval (day 8- 10), lactate was consumed towards pyruvate synthesis, the total TCA flux was increased compared to the second interval, while the pyruvate carboxylase activity was considerably lower compared to the first interval (Fig. 5.12). Finally, the unsupervised least-squares feature selection technique identified a subset of 14 reactions showing the highest variability between the experiments, with 4 of the main glycolysis reactions exhibiting a considerably different profile between the control and the GU feeding experiments throughout the cell culturing period. Finally, the GU feeding experiments were found to produce higher amounts of glycogen.

The IgG glycan analysis revealed significantly higher levels of galactosylation in all GU feeding experiments compared to the control, but similar between the feeding experiments (Fig. 5.14). Galactosylation was also reduced during the cell culturing period, possibly due to ammonia accumulation that has been found to inhibit b4GalT expression and activity[257, 259-261]. The resulting IgG glycoprofile indicated that the plausible saturation in the synthesis rate of intracellular UDPGal should not have an effect on antibody galactosylation. Similarly, HCPs glyoprofile showed higher galactosylation levels for all feeding experiments, while sialylation, fucosylation and highly branched glycans were also generally increased due to galactose and uridine addition (Table 5.5). Interestingly, HCPs galactosylation presented a strong positive correlation with fucosylation and highly branched glycans (Fig. 5.16).

Subsequently, the kinetic glycosylation model was modified in order to approximate the *N*-linked glycosylation profile of both the intracellular HCPs and the secreted IgG. In addition, the updated model incorporated ER glycosylation and protein transport between cellular compartments. The model presented good agreement with the experimental data of the control experiment (Fig. 5.17)

and was used in order to showcase that the identified saturation of NSD transport to the Golgi results in reduced IgG galactosylation (Fig. 5.18C, Fig. 5.19A). HCPs *N*-linked glycosylation was also found to consume considerable UDPGal resources, and also disproportionally more compared to the IgG in the GU feeding experiments (Fig. 5.20). Combined with the elevated amounts of UDPGal utilized towards HCPs *O*-linked glycosylation in CHO cells[513], HCPs glycosylation was found to considerably influence the glycoprofile of the recombinantly produced IgG. Overall, NSD transport saturation and HCPs glycosylation were found to halt the increase of IgG galactosylation by 33% and 43%, respectively.

Designing genetic engineering experiments based on the model-based analysis

Throughout this chapter, several insights in the intracellular metabolism and glycosylation of CHO cells were provided. The observed upper threshold in antibody galactosylation was partially attributed to NSD transport to the Golgi and to the amount of NSD resources consumed towards cellular glycosylation. The results are based on both experimental measurements and model-based simulations. In addition, the expression levels of b4GalT were found to be the major bottleneck in IgG galactosylation. To that end, the overexpression of i) the SLC35A2 gene responsible for the expression of the UDPGal transporter, and ii) the B4GALT1 gene, the main isoform responsible for b4GalT expression in CHO cells[350], could reveal further insights in the relationship between NSD transport and b4GalT availability. When combined with galactose and uridine supplementation in the culture and therefore elevated UDPGal levels in the cytosol, the overexpression scenarios can reveal the exact hierarchy of galactosylation bottlenecks.

As the manipulation of HCPs *N*-linked galactosylation requires the modification of b4GalT expression levels that would also affect IgG galactosylation, a plausible target is the beta-1,3-galactosyltransferase (B3GALT5 gene) that is responsible for HCPs *O*-linked galactosylation. The knockout of the aforementioned enzyme could release UDPGal resources that will subsequently be available for IgG galactosylation. Alternatively, the high positive correlation between HCPs galactosylation and the > bi-antennary glycans (Fig. 5.16) indicates that the knockout of the GnTIV, GnTV and iGnT enzymes could potentially lead to decreased HCP galactosylation and therefore to increased intra-Golgi levels of UDPGal that could be consumed towards IgG galactosylation. Importantly, the aforementioned enzymes are not involved in IgG glycosylation, and therefore such genetic engineering experiments will not compromise antibody glycosylation. However, transient transfection (electroporation) with siRNA molecules targeting the MGAT4A, MGAT4B and MGAT5 genes, resulted in no improvements in IgG galactosylation (data not shown), indicating that a complete knockout is possibly necessary to evaluate the effect of such genetic modifications on IgG glycosylation.

Chapter 6 Conclusions and future work

6.1 Summary of results

Mammalian cell culture is a complicated yet critical process for therapeutics production. The genome, metabolism and *N*-linked glycosylation pathways of CHO cells, the main hosts for recombinant protein synthesis, have been extensively, but not exhaustively, studied. The optimization of CHO cell-based systems remains a challenging task due to the complexity of cellular metabolism that is inherently related to product quality. Following the Quality by Design framework, the understanding of the relationship between process parameters and therapeutic protein synthesis and glycosylation is of paramount importance.

Overall, this thesis endeavoured to demonstrate the power of various computational tools in the optimization of CHO cell cultures and the implementation of the Quality by Design paradigm. A particular focus throughout the thesis was placed on the effect of galactose and uridine, the precursors for UDPGal synthesis, on CHO cell growth, productivity and mAb glycosylation. However, the applications of the developed models were not limited to galactose and uridine feeding and also included different therapeutic glycoproteins.

The literature review presented in Chapter 1 was mainly focused on showcasing the importance of protein glycosylation within the bioprocess context. Specific interest was placed on the effect of different glycosylation profiles, such as galactosylation, on the efficacy and immunogenicity of mAbs. In addition, the role of protein glycosylation within the Quality by Design paradigm was further analysed in an effort to set the basis for the rest of the thesis and demonstrate the importance of this cellular function. Subsequently, the effect of process parameters on mammalian cell growth and protein glycosylation were discussed, and methods from literature to control glycosylation were described. The last part of the literature review described the main categories of computational tools, including kinetic, stoichiometric and data-driven models and their applications on simulation, controlling and optimizing cell growth and protein glycosylation.

In Chapter 2 of the thesis, a holistic model describing CHO cell growth and mAb glycosylation was developed. In total, the model comprises three kinetic submodules, namely the *Metabolism submodel* that captures CHO cell growth, metabolism and antibody synthesis, the *NSD submodel* that utilizes a lumped reaction pathway for NSD synthesis and aims to estimate intracellular NSD concentrations, and finally the *Glycomodel* that is employed for simulating mAb glycosylation based on the NSD concentrations and specific antibody production rate. Importantly, the *Metabolism submodel* utilized a reduced reaction network, including a lumped glycolysis pathway, in order to minimize the number

of parameters requiring estimation. The *Glycomodel*, adapted from del Val et al.[478], adapts the Golgi maturation theory and uses a single PFR system in order to capture the dynamics within the compartment. The model was tailored to predict the effect of galactose and uridine addition on antibody titre and glycosylation, exhibiting an $R^2 > 0.95$ for viable cell density and mAb titre, $R^2 > 0.96$ for UDPGal concentration and a 5% error range against the experimental data of glycan analysis. Subsequently, the model was successfully utilized for designing a galactose and uridine feeding strategy in order to maximize mAb galactosylation without compromising cell growth and product titre. The optimized experiment achieved a 93% increase in antibody galactosylation at harvest, showcasing the power of model-based process optimization.

Following the development of the holistic model and the validation of its predictive capabilities, an *in silico* Design Space was constructed. Constrained GSA was employed for the identification of the Design Space, based on 8,192 (2^{13}) model simulations with varying galactose and uridine concentrations added on days 4, 6, 8 and 10 of the culture. In addition, the lowest acceptable values for both the antibody concentration and galactosylation were defined, ensuring that the feeding regimes falling within the Design Space would meet the pre-determined quantitative and qualitative targets, respectively. Only 556 out of the 8,192 simulations were found to satisfy both constraints, while the identified Design Space was validated against experimental data from 7 fed-batch experiments including galactose and uridine addition.

Subsequently, a strategic framework for the parameterization of the *Metabolism submodel* was developed. The framework was based on GSA and aimed to identify the model parameters that significantly influence the designated outputs, such as viable cell density and mAb concentration, and require tuning for model adaptation to new experimental data. The validity of the framework was initially verified against experimental data that presented a different profile compared to the experiments used for model training. The adapted model presented an $R^2 > 0.91$ for model predictions for 4 fed-batch feeding experiments, demonstrating the correct identification of the significant model parameters. In addition, the framework was applied to data from another CHO cell line. The aim was to examine whether the same model structure could be applied to describe the growth and metabolism of a newly introduced cell line. Following the re-estimation of 10 parameters, the model presented an $R^2 > 0.95$ in the simulation of viable cell density, mAb concentration, glucose and asparagine extracellular concentration and a R^2 ~0.89 for ammonia.

Whilst the holistic kinetic model was proven a powerful tool for predicting mAb glycosylation, the parameterization and development of such models remains a challenging task. For that reason, artificial neural networks were used in order to replace the kinetic *Glycomodel*. As a standalone

algorithm the ANN-based glycosylation model was found to describe the experimentally measured mAb glycoprofile with an AAE of 0.87%, outperforming its equivalent data-driven PLS model. The ANN-based model was also configured to describe the effect of manganese, galactose and fucose addition on mAb glycosylation, presenting an AAE of 0.89%. In a next step, the ANN-based model was tuned to predict the effect of knockouts of b4GalT isoforms on site-specific glycosylation of two fusion proteins, namely EPO-Fc and Fc-DAO. Finally, a hybrid model consisting of the kinetic *Metabolism* and *NSD submodels* and the ANN-based glycosylation model, namely HyGlycoM, was found to outperform its fully mechanistic equivalent by ~30% and simultaneously, significantly reducing the required simulation time.

The galactose and uridine feeding experiments used throughout this thesis revealed that IgG galactosylation reached an upper threshold, regardless the supplemented concentration of galactose and uridine and consequently, the intracellular levels of UDPGal. To that end, five fed-batch experiments with different concentrations of galactose and uridine (including the control) were performed, in order to examine the possible bottlenecks in antibody galactosylation and distinctive characteristics of intracellular metabolism in the feeding experiments. The experimental results were also used to inform an updated kinetic glycosylation model, describing both HCPs and IgG glycosylation in the ER and Golgi compartments, and protein transport between cellular compartments. The integrated experimental and computational framework demonstrated that the limitations in UDPGal transport towards the Golgi and the utilization of UDPGal resources towards HCPs glycosylation considerably halt IgG galactosylation (43% and 33%, respectively). In addition, expression levels of b4GalT were identified as the major bottleneck in IgG galactosylation. Finally, cell growth and productivity, metabolites profile and amino acid analysis were used for flux balance analysis. An unsupervised feature selection technique was applied in the FBA results, identifying the glycolysis reactions as the main difference between the control and the feeding experiments.

6.2 Main contributions and conclusions

The development of computational tools for describing mammalian cell metabolism and recombinant protein production has never been more relevant. The transition to QbD-based process development has triggered a growing interest in the use of model-based platforms for bioprocess control and optimization, in both academia and industry. To that end, mathematical models have proven to be valuable tools for expediting process development and reducing the relevant costs, while maintaining high levels of prediction confidence. However, it is important to note that model development is inextricably linked to relevant experimentation that necessitates careful designing in order to maximize the potential of the model under development. The motivating power behind this thesis was the exploration of different computational tools for describing mammalian bioprocesses.

Throughout this thesis, kinetic, constraint-based and machine learning models have altogether been used for different or similar purposes, demonstrating the importance and necessity of applying hybrid configurations for describing cellular processes. Indeed, this thesis includes the development of the first holistic hybrid model (HyGlycoM) that incorporates both kinetic and machine learning submodules for describing CHO cell growth and recombinant protein glycosylation.

Whilst the simulation and experimental investigation of CHO cell metabolism was part of this thesis' objectives, the major focus of the worked presented herein was placed on glycosylation, due to its paramount importance for recombinant protein activity and immunogenicity. In addition, glycosylation is a rather intriguing process; the non-template driven nature of glycosylation results in high degrees of variability between different cell lines and recombinant proteins, further increasing the difficulty in developing a predictive tool that can be seamlessly applied in several bioprocesses. In addition, this thesis was focused on the manipulation of antibody galactosylation through the supplementation of galactose and uridine during the cell culture period. We strongly believe that the mathematical models and experimental approaches developed herein, to describe and optimize the examined bioprocess or investigate relevant intracellular bottlenecks, can be applied to other bioprocesses of interest and through different objectives, with the introduction of minor modifications.

We believe that each chapter of this thesis presented a novel and important contribution in the field of bioprocess engineering. The main contributions of the work presented herein, and the relevant conclusions are listed below:

1. A kinetic model holistically describing CHO cell growth, metabolism, NSD synthesis, antibody production and glycosylation was developed. The model was found to successfully evaluate the effect of galactose and uridine (NSD precursors) addition on cell growth, mAb concentration and glycosylation. A modelling framework with such capabilities can be a valuable tool for process control and the design of optimal feeding regimes.
2. The developed holistic model was successfully used in order to design a fed-batch experiment that would maximize mAb galactosylation through the adjustment of galactose and uridine concentrations in the feed. The results of the optimal scenario were experimentally verified, achieving a > 90% increase in antibody galactosylation compared to the control, minimizing the costs of identifying an optimal feeding strategy.
3. For the first time in literature, a Design Space of galactose and uridine concentrations in the feed was constructed, based solely on model simulations and cGSA. The Design Space ensured that both the desired quality and quantity of the product could be achieved. Such a model-

based framework can significantly expedite bioprocess development, reduce relevant costs and increase the flexibility of the process, while simultaneously contributing to the implementation of the Quality by Design paradigm.

4. A comprehensive framework for successfully adapting the CHO metabolism model to data of the same cell line that deviated considerably from the training dataset, and to data from a new CHO cell line, was proposed. To our knowledge this is one of the first studies to develop a detailed workflow for adapting an existing metabolic model to a new mammalian cell line without changing model's configuration. Such frameworks can significantly accelerate the development of kinetic models for new cell lines and contribute towards the establishment of a more global kinetic model for describing CHO cell metabolism.
5. For the first time in literature, an ANN-based glycosylation model configuration was proposed and was verified against several bioprocess and genetic engineering scenarios. Machine learning models are simpler to build and parameterize, especially in processes like glycosylation that require the identification of product-specific reaction networks and the development of sophisticated parameterization methodologies. For that reason, the ANN-based glycosylation model can be particularly useful for industrial applications were hundreds of cell lines and several different molecules are examined during cell line and product development.
6. A hybrid model, namely HyGlycoM, was proposed. HyGlycoM is, to our knowledge, the first hybrid model to predict antibody glycosylation based on nutrients concentration in the media and feed. Importantly, HyGlycoM was found to outperform its fully mechanistic equivalent, while requiring significantly less time for development. The superior performance of HyGlycoM reflects the promising future of hybrid configurations for describing and optimizing new bioprocesses.
7. The comprehensive experimental analysis and the applied modifications in the kinetic glycosylation model enabled the identification of bottlenecks in antibody galactosylation. The identified bottlenecks can be manipulated to improve galactosylation, while a similar framework can be applied for the identification of other bottlenecks in protein glycosylation.
8. Importantly, the updated glycosylation model describes both HCPs and IgG glycosylation, being the first model to simultaneously describe both the cellular and the recombinant protein glycoprofile. The updated model can be utilized to design genetic engineering studies and feeding scenarios that would harness the captured balance between HCPs and IgG glycosylation.

9. Through the use of FBA and unsupervised machine learning, the major differences in the intracellular metabolism between the control and the GU feeding experiments were identified. The hybrid configuration provided valuable insights into CHO cell metabolism under galactose and uridine supplementation, that could be utilized for further refining the bioprocess.

6.3 Limitations and future work

Similar to every research project, the work presented herein exhibits certain limitations. The identified limitations, in combination with the new research avenues that are created based on the results of this thesis, are summarized below, in an effort to describe the future work that can further improve the current research.

6.3.1 Improvements in the holistic kinetic model

The kinetic glycosylation submodel (*Glycomodel*) does not account for enzymes regulation through the culture. The lack of a dynamic description of glycosyltransferases transcription/translation was found to constraint the predictive capabilities of the model in the validation experiment. A 15% upregulation in b4GalT concentration led to better predictions for the optimal experiment. As the mechanistic description of glycosyltransferase expression would be challenging, RNA-seq, microarray or proteomics data can be utilized from a data-driven algorithm to inform the *Glycomodel* for changes in enzymes concentrations during the cell culture period. In addition, several kinetic constants involved in the glycosylation reaction rates calculation were assigned to their nominal values due to lack of experimental data. Studies, either using transient or stable gene knockdown/knockout or inhibitors of glycosyltransferases can be performed, in order to generate relevant data that would accurately inform the *Glycomodel*. For glycoproteins, such as mAbs, that present a rather simple glycoprofile, the reaction network can be further reduced in order to accelerate model simulation and reduce the parameters requiring estimation.

The holistic kinetic model developed in this thesis was only trained and tested in fed-batch cultures, in which the variability was introduced from the supplementation of different galactose and uridine concentrations. In addition, all experimental data were derived from small-scale cultures (100 mL) in flasks. Incorporation of other feeding regimes, including NSD precursors (i.e., ManNAc, fucose, guanosine) or glycosylation inhibitors (e.g., kifunensine) and co-factors (e.g., manganese), or different process parameters, such as mild-hypothermia or pH shift, is required. Model training to data from scaled up experiments in bioreactors is necessary in order to ensure model applicability within the industrial context. In addition, model predictive capabilities can be examined against more CHO cell lines that i) exhibit considerably different specific cell growth and protein productivity rates, ii) are

derived from CHO-S or CHO-DG44 parental cell lines, or iii) express different glycoproteins. Model adaptation to a CHO cell line with a considerably different growth profile was partially covered (only the *Metabolism submodel* was adapted) in Chapter 3 of this thesis.

Finally, based on model performance in successfully maximizing antibody galactosylation, the model can be used in MPC-based applications for the online control of protein glycosylation by adjusting feeding composition during the cell culture period. It is important to note that the PFR-based system of the *Glycomodel* could be replaced with a 4-CSTR kinetic system (Chapter 5) or with the ANN-based glycosylation model (Chapter 4) in order to significantly reduce the simulation time and ensure the swift optimization convergence that is necessary for online applications.

6.3.2 Quality by Design applications

Whilst the model successfully predicted which experiments would fall in the *in silico* Design Space and which would fail to satisfy the constraints, discrepancies between the experimentally measured mAb titre and the model predicted values were observed. As the model assumes that the addition of galactose and uridine will halt cell growth and productivity, it is important to identify the reasons behind the observed increase in the experimentally measured titre in the feeding experiments and introduce this underlying relationship to the model. In addition, the proposed framework for DS identification can be applied on different feeding regimes, cell lines and products to examine its adaptability to new bioprocess conditions.

Whilst the adapted *Metabolism submodel* was successfully adapted to the GS46 cell line data, it is necessary to validate model performance in unrelated experimental data of the GS46 cell line in order to evaluate its predictive capabilities. Experiments including galactose and uridine supplementation should be prioritized in order to compare model performance between the T127 and GS46 cell line. The adaptation to experimental data from additional CHO cell lines, as also described in section 6.3.1, is necessary to investigate the extent at which a single kinetic model can be utilized for several cell lines. A similar framework can be developed for adapting the *NSD* and the *Glycomodel submodels*. Finally, the results of the GSA can be used for reducing the size of the model by removing parameters that do not affect model outputs. However, model reduction should be performed cautiously, considering that parameters not significant for the examined bioprocess (process parameters, cell line, product) might influence the results of an alternative bioprocess.

6.3.3 Machine learning and hybrid applications

Hybrid models combining mechanistic and machine learning modules are expected to dominate the field of bioprocess modelling, due to their flexibility, reduced simulation times and superior performance. However, hybrid and machine learning configurations, like the ANN-based glycosylation

model and HyGlycoM, require large datasets for appropriate training. The datasets used in the work presented herein were rather small; however, the introduction of Gaussian noise and the restrictions imposed in the complexity of the networks can reduce the risk of overfitting. Extensive and elaborate datasets are difficult to acquire in processes like mammalian cell culturing and especially in academic research, transforming the machine learning and hybrid applications mostly fit for industrial research. The introduction of omics data, such as transcriptomics or proteomics for identifying the mRNA or protein levels of the glycosylatransferases, metabolomics for the quantifying several crucial metabolites such as the NSDs, and glycomics for the comprehensive analysis of the glycoprofile, can significantly improve the applicability and performance of the ANN-based models presented in Chapter 4 of this thesis.

6.3.4 Glycosylation bottlenecks

In order to further evaluate and experimentally verify the mAb galactosylation bottlenecks identified in Chapter 5 of this thesis, several genetic engineering studies can be performed. Overexpression of the SLC35A2 gene will reveal the extent of limitations imposed by the saturation of UDPGal transport, while upregulation of the b4GalT expression will provide insights in enzyme's availability in the Golgi. In addition, downregulation or complete knockout of the B3GALT5, MGAT4A, MGAT4B and MGAT5 genes were found as potential strategies for manipulating HCPs glycosylation towards improving IgG galactosylation. Whilst preliminary results (data not shown) on MGAT4A/4B/5 transient siRNA-mediated knockdown, did not show increased IgG galactosylation, genes knockout might be necessary in order to correctly evaluate the effect of this genetic modification. Another potential target is the B3GNT2 gene expressing the iGnT enzyme that is responsible for polyLacNAcs formation. Whilst the CHO-T127 cell line did not exhibit high levels of polyLacNAc-extended glycans, silencing of iGnT could be proven useful for other cell lines, in order to reduce UDPGal consumption towards HCPs glycosylation.

The updated glycosylation model describing both HCPs and IgG glycosylation can be easily adapted to the glycoprofile of any cell line and glycoprotein. Due to the large size of the model, a GSA can be useful in reducing the size of the model and alleviating the burden imposed by parameter estimation. Model reduction will also considerably improve the confidence in the estimated parameter values (lower 95% CIs). Future applications should also include training and validation of the updated model in the GU feeding experiments. Finally, the insights provided by the FBA studies and the NSD analysis can be used to refine and further improve the design of GU feeding experiments.

6.4 Concluding remarks

Overall, the results of this thesis have considerably contributed to model-based prediction and optimization of antibody synthesis and glycosylation, while simultaneously improving our understanding of the process itself. The combination of experimental and computational work, including both mechanistic and machine learning models, was found to lead to trustworthy results within the Quality by Design context. The frameworks proposed in this thesis can significantly expedite process and product development, reduce the development and manufacturing costs and, ultimately, create more affordable biotherapeutics for the people in need.

List of publications

Articles

Kotidis, P.; Pappas, I.; Avraamidou, S.; Pistikopoulos, E.N.; Kontoravdi, C.; Papathanasiou, M., DigiGlyc: a hybrid tool for reactive scheduling in cell culture systems, *Computers & Chemical Engineering* 2021, 154, 107460.

Alhuthali, S.; Kotidis, P.; Kontoravdi, C., Osmolality effects on CHO cell growth, cell volume, antibody productivity and glycosylation, *International Journal of Molecular Sciences* 2021, 22 (7), 3290.

Kotidis, P.; Kontoravdi, C., Harnessing the potential of artificial neural networks for predicting protein glycosylation. *Metabolic Engineering Communications* 2020, 10, e00131.

Kotidis, P.; Kontoravdi, C., Strategic Framework for Parameterization of Cell Culture Models. *Processes* 2019, 7 (3), 174.

Kotidis, P.; Jedrzejewski, P.; Sou, S. N.; Sellick, C.; Polizzi, K.; del Val, I. J.; Kontoravdi, C., Model-based optimization of antibody galactosylation in CHO cell culture. *Biotechnology and bioengineering* 2019, 116 (7), 1612-1626.

Kotidis, P.; Demis, P.; Goey, C. H.; Correa, E.; McIntosh, C.; Trepekli, S.; Shah, N.; Klymenko, O. V.; Kontoravdi, C., Constrained global sensitivity analysis for bioprocess design space identification. *Computers & Chemical Engineering* 2019, 125, 558-568.

Reviews

Makrydaki, E.; Kotidis, P.; Kontoravdi, C., Hitting the Sweet Spot with Capillary Electrophoresis: Advances in N- Glycomics and Glycoproteomics, *Current Opinion in Biotechnology*, *accepted for publication, in-press*.

Fung Shek, C.; Kotidis, P.; Betenbaugh M.J., Mechanistic and Data-Driven Modeling of Glycoproteins, *Current Opinion in Chemical Engineering* 2021, 32, 100690.

Antonakoudis, A.; Barbosa, R.; Kotidis, P.; Kontoravdi, C., The era of big data: Genome-scale modelling meets machine learning. *Computational and Structural Biotechnology Journal* 2020, 18, 3287-3300.

Book Chapters & Proceedings

Kotidis, P., Marbiah, M., Donini, R., Gómez, I.A., del Val, I.J., Haslam, S.; Polizzi. K.M.; Kontoravdi, C., Rapid antibody glycoengineering in CHO cells via RNA interference and CGE-LIF N-glycomics, *Methods in Molecular Biology*, Editor: Gavin Davey, *in press*.

Kotidis, P.; del Val, I. J.; Krambeck F.J.; Betenbaugh M.J.; Kontoravdi, C., Novel approach to account for enzyme regulation in kinetic models of protein glycosylation, *BMC Proceedings* 2020, 15.

Antonakoudis A.; Kis Z.; Kontoravdi C.; Kotidis P.; Papathanasiou M.; Shah N.; Tomba E.; Varsakelis C.; von Stock M., Accelerating product and process development through a model centric approach, In: Quality by Design—An Indispensable Approach to Accelerate Biopharmaceutical Product Development, Editors: Francesca Campa, Amin Khan, *Parenteral Drug Association*, USA, 2021, ISBN Number: 9781945584220.

Kis Z.; Papathanasiou M.; Kotidis P.; Antonakoudis A.; Kontoravdi C.; Shah N., Stability modelling for biopharmaceutical process intermediates, In: Quality by Design—An Indispensable Approach to Accelerate Biopharmaceutical Product Development, Editors: Francesca Campa, Amin Khan, *Parenteral Drug Association*, USA, 2021, ISBN Number: 9781945584220.

Bibliography

1. Walsh, G., Biopharmaceutical benchmarks 2006. *Nature Biotechnology* 2006, *24* (7), 769-776.

2. Walsh, G., Biopharmaceutical benchmarks 2010. *Nature Biotechnology* 2010, *28* (9), 917-924.

3. Walsh, G., Biopharmaceutical benchmarks 2014. *Nature Biotechnology* 2014, *32* (10), 992-1000.

4. Walsh, G., Biopharmaceutical benchmarks 2018. *Nature Biotechnology* 2018, *36* (12), 1136-1145.

5. Kristen, A. V.; Ajroud-Driss, S.; Conceição, I.; Gorevic, P.; Kyriakides, T.; Obici, L., Patisiran, an RNAi therapeutic for the treatment of hereditary transthyretin-mediated amyloidosis. *Neurodegenerative Disease Management* 2018, *9* (1), 5-23.

6. Akinc, A.; Maier, M. A.; Manoharan, M.; Fitzgerald, K.; Jayaraman, M.; Barros, S.; Ansell, S.; Du, X.; Hope, M. J.; Madden, T. D.; Mui, B. L.; Semple, S. C.; Tam, Y. K.; Ciufolini, M.; Witzigmann, D.; Kulkarni, J. A.; van der Meel, R.; Cullis, P. R., The Onpattro story and the clinical translation of nanomedicines containing nucleic acid-based drugs. *Nature Nanotechnology* 2019, *14* (12), 1084-1087.

7. Zahavi, D.; Weiner, L., Monoclonal Antibodies in Cancer Therapy. *Antibodies (Basel)* 2020, *9* (3), 34.

8. EvaluatePharma Evaluate Pharma World Preview 2020, Outlook to 2026. https://www.evaluate.com/thought-leadership/pharma/evaluatepharma-world-preview-2020-outlook-2026 (accessed March 12, 2021).

9. Zhu, J., Mammalian cell protein expression for biopharmaceutical production. *Biotechnology Advances* 2012, *30* (5), 1158-1170.

10. Demain, A. L.; Vaishnav, P., Production of recombinant proteins by microbes and higher organisms. *Biotechnology Advances* 2009, *27* (3), 297-306.

11. Love, K. R.; Dalvie, N. C.; Love, J. C., The yeast stands alone: the future of protein biologic production. *Current Opinion in Biotechnology* 2018, *53*, 50-58.

12. McKenzie, E. A.; Abbott, W. M., Expression of recombinant proteins in insect and mammalian cells. *Methods* 2018, *147*, 40-49.

13. Dhara, V. G.; Naik, H. M.; Majewska, N. I.; Betenbaugh, M. J., Recombinant Antibody Production in CHO and NS0 Cells: Differences and Similarities. *BioDrugs* 2018, *32* (6), 571-584.

14. Dumont, J.; Euwart, D.; Mei, B.; Estes, S.; Kshirsagar, R., Human cell lines for biopharmaceutical manufacturing: history, status, and future perspectives. *Crit Rev Biotechnol* 2016, *36* (6), 1110-1122.

15. Chin, C. L.; Goh, J. B.; Srinivasan, H.; Liu, K. I.; Gowher, A.; Shanmugam, R.; Lim, H. L.; Choo, M.; Tang, W. Q.; Tan, A. H.-M.; Nguyen-Khuong, T.; Tan, M. H.; Ng, S. K., A human expression system based on HEK293 for the stable production of recombinant erythropoietin. *Scientific Reports* 2019, *9* (1), 16768.

16. Little, M.; Kipriyanov, S. M.; Le Gall, F.; Moldenhauer, G., Of mice and men: hybridoma and recombinant antibodies. *Immunology Today* 2000, *21* (8), 364-370.

17. Kim, J. Y.; Kim, Y.-G.; Lee, G. M., CHO cells in biotechnology for production of recombinant proteins: current state and further potential. *Applied Microbiology and Biotechnology* 2012, *93* (3), 917-930.

18. Puetz, J.; Wurm, F. M., Recombinant Proteins for Industrial versus Pharmaceutical Purposes: A Review of Process and Pricing. *Processes* 2019, *7* (8).

19. Jayapal, K. P.; Wlaschin, K. F.; Hu, W. S.; Yap, M. G. S., Recombinant protein therapeutics from CHO Cells - 20 years and counting. *Chemical Engineering Progress* 2007, *103* (10), 40-47.

20. Grilo, A. L.; Mantalaris, A., The Increasingly Human and Profitable Monoclonal Antibody Market. *Trends in Biotechnology* 2019, *37* (1), 9-16.

21. Reinhart, D.; Damjanovic, L.; Kaisermayer, C.; Sommeregger, W.; Gili, A.; Gasselhuber, B.; Castan, A.; Mayrhofer, P.; Grünwald-Gruber, C.; Kunert, R., Bioprocessing of Recombinant CHO-K1, CHO-DG44, and CHO-S: CHO Expression Hosts Favor Either mAb Production or Biomass Synthesis. *Biotechnology journal* 2019, *14* (3), 1700686.

22. Xu, X.; Nagarajan, H.; Lewis, N. E.; Pan, S.; Cai, Z.; Liu, X.; Chen, W.; Xie, M.; Wang, W.; Hammond, S.; Andersen, M. R.; Neff, N.; Passarelli, B.; Koh, W.; Fan, H. C.; Wang, J.; Gui, Y.; Lee, K. H.; Betenbaugh, M. J.; Quake, S. R.; Famili, I.; Palsson, B. O.; Wang, J., The genomic sequence of the Chinese hamster ovary (CHO)-K1 cell line. *Nature Biotechnology* 2011, *29* (8), 735-741.

23. Brinkrolf, K.; Rupp, O.; Laux, H.; Kollin, F.; Ernst, W.; Linke, B.; Kofler, R.; Romand, S.; Hesse, F.; Budach, W. E.; Galosy, S.; Müller, D.; Noll, T.; Wienberg, J.; Jostock, T.; Leonard, M.; Grillari, J.; Tauch, A.; Goesmann, A.; Helk, B.; Mott, J. E.; Pühler, A.; Borth, N., Chinese hamster genome sequenced from sorted chromosomes. *Nature Biotechnology* 2013, *31* (8), 694-695.

24. Lewis, N. E.; Liu, X.; Li, Y.; Nagarajan, H.; Yerganian, G.; O'Brien, E.; Bordbar, A.; Roth, A. M.; Rosenbloom, J.; Bian, C.; Xie, M.; Chen, W.; Li, N.; Baycin-Hizal, D.; Latif, H.; Forster, J.; Betenbaugh, M. J.; Famili, I.; Xu, X.; Wang, J.; Palsson, B. O., Genomic landscapes of Chinese hamster ovary cell lines as revealed by the Cricetulus griseus draft genome. *Nature Biotechnology* 2013, *31* (8), 759-765.

25. Cao, Y.; Kimura, S.; Itoi, T.; Honda, K.; Ohtake, H.; Omasa, T., Construction of BAC-based physical map and analysis of chromosome rearrangement in chinese hamster ovary cell lines. *Biotechnology and bioengineering* 2012, *109* (6), 1357-1367.

26. Baycin-Hizal, D.; Tabb, D. L.; Chaerkady, R.; Chen, L.; Lewis, N. E.; Nagarajan, H.; Sarkaria, V.; Kumar, A.; Wolozny, D.; Colao, J.; Jacobson, E.; Tian, Y.; O'Meally, R. N.; Krag, S. S.; Cole, R. N.; Palsson, B. O.; Zhang, H.; Betenbaugh, M., Proteomic Analysis of Chinese Hamster Ovary Cells. *Journal of Proteome Research* 2012, *11* (11), 5265-5276.

27. Park, J. H.; Jin, J. H.; Lim, M. S.; An, H. J.; Kim, J. W.; Lee, G. M., Proteomic Analysis of Host Cell Protein Dynamics in the Culture Supernatants of Antibody-Producing CHO Cells. *Scientific Reports* 2017, *7* (1), 44246.

28. Hefzi, H.; Ang, K. S.; Hanscho, M.; Bordbar, A.; Ruckerbauer, D.; Lakshmanan, M.; Orellana, C. A.; Baycin-Hizal, D.; Huang, Y.; Ley, D.; Martinez, V. S.; Kyriakopoulos, S.; Jiménez, N. E.; Zielinski, D. C.; Quek, L. E.; Wulff, T.; Arnsdorf, J.; Li, S.; Lee, J. S.; Paglia, G.; Loira, N.; Spahn, P. N.; Pedersen, L. E.; Gutierrez, J. M.; King, Z. A.; Lund, A. M.; Nagarajan, H.; Thomas, A.; Abdel-Haleem, A. M.; Zanghellini, J.; Kildegaard, H. F.; Voldborg, B. G.; Gerdtzen, Z. P.; Betenbaugh, M. J.; Palsson, B. O.; Andersen, M. R.; Nielsen, L. K.; Borth, N.; Lee, D. Y.; Lewis, N. E., A Consensus Genome-scale Reconstruction of Chinese Hamster Ovary Cell Metabolism. *Cell systems* 2016, *3* (5), 434-443.e8.

29. Sommeregger, W.; Mayrhofer, P.; Steinfellner, W.; Reinhart, D.; Henry, M.; Clynes, M.; Meleady, P.; Kunert, R., Proteomic differences in recombinant CHO cells producing two similar antibody fragments. *Biotechnology and bioengineering* 2016, *113* (9), 1902-1912.

30. Birzele, F.; Schaub, J.; Rust, W.; Clemens, C.; Baum, P.; Kaufmann, H.; Weith, A.; Schulz, T. W.; Hildebrandt, T., Into the unknown: expression profiling without genome sequence information in CHO by next generation sequencing. *Nucleic Acids Research* 2010, *38* (12), 3999-4010.

31. Clarke, C.; Doolan, P.; Barron, N.; Meleady, P.; O'Sullivan, F.; Gammell, P.; Melville, M.; Leonard, M.; Clynes, M., Large scale microarray profiling and coexpression network analysis of CHO cells identifies transcriptional modules associated with growth and productivity. *Journal of biotechnology* 2011, *155* (3), 350-359.

32. Becker, J.; Hackl, M.; Rupp, O.; Jakobi, T.; Schneider, J.; Szczepanowski, R.; Bekel, T.; Borth, N.; Goesmann, A.; Grillari, J.; Kaltschmidt, C.; Noll, T.; Pühler, A.; Tauch, A.; Brinkrolf, K., Unraveling the Chinese hamster ovary cell line transcriptome by next-generation sequencing. *Journal of biotechnology* 2011, *156* (3), 227-235.

33. Hammond, S.; Kaplarevic, M.; Borth, N.; Betenbaugh, M. J.; Lee, K. H., Chinese hamster genome database: An online resource for the CHO community at www.CHOgenome.org. *Biotechnology and bioengineering* 2012, *109* (6), 1353-1356.

34. Rita Costa, A.; Elisa Rodrigues, M.; Henriques, M.; Azeredo, J.; Oliveira, R., Guidelines to cell engineering for monoclonal antibody production. *European Journal of Pharmaceutics and Biopharmaceutics* 2010, *74* (2), 127-138.

35. EMA Biosimilar medicines: Overview. https://www.ema.europa.eu/en/human-regulatory/overview/biosimilar-medicines-overview (accessed March 12 2021).

36. Gulácsi, L.; Brodszky, V.; Baji, P.; Kim, H.; Kim, S. Y.; Cho, Y. Y.; Péntek, M., Biosimilars for the management of rheumatoid arthritis: economic considerations. *Expert Review of Clinical Immunology* 2015, *11* (sup1), 43-52.

37. Rugo, H. S.; Barve, A.; Waller, C. F.; Hernandez-Bronchud, M.; Herson, J.; Yuan, J.; Sharma, R.; Baczkowski, M.; Kothekar, M.; Loganathan, S.; Manikhas, A.; Bondarenko, I.; Mukhametshina, G.; Nemsadze, G.; Parra, J. D.; Abesamis-Tiambeng, M. L. T.; Baramidze, K.; Akewanlop, C.; Vynnychenko, I.; Sriuranpong, V.; Mamillapalli, G.; Ray, S.; Yanez Ruiz, E. P.; Pennella, E.; for the Heritage Study, I., Effect of a Proposed Trastuzumab Biosimilar Compared With Trastuzumab on Overall Response Rate in Patients With ERBB2 (HER2)–Positive Metastatic Breast Cancer: A Randomized Clinical Trial. *JAMA* 2017, *317* (1), 37-47.

38. Akram, M. S.; Pery, N.; Butler, L.; Shafiq, M. I.; Batool, N.; Rehman, M. F. u.; Grahame-Dunn, L. G.; Yetisen, A. K., Challenges for biosimilars: focus on rheumatoid arthritis. *Crit Rev Biotechnol* 2021, *41* (1), 121-153.

39. Blackwell, K.; Semiglazov, V.; Krasnozhon, D.; Davidenko, I.; Nelyubina, L.; Nakov, R.; Stiegler, G.; Singh, P.; Schwebig, A.; Kramer, S.; Harbeck, N., Comparison of EP2006, a filgrastim biosimilar, to the reference: a phase III, randomized, double-blind clinical study in the prevention of severe neutropenia in patients with breast cancer receiving myelosuppressive chemotherapy. *Annals of Oncology* 2015, *26* (9), 1948-1953.

40. Yoo, D. H.; Hrycaj, P.; Miranda, P.; Ramiterre, E.; Piotrowski, M.; Shevchuk, S.; Kovalenko, V.; Prodanovic, N.; Abello-Banfi, M.; Gutierrez-Ureña, S.; Morales-Olazabal, L.; Tee, M.; Jimenez, R.; Zamani, O.; Lee, S. J.; Kim, H.; Park, W.; Müller-Ladner, U., A randomised, double-blind, parallel-group study to demonstrate equivalence in efficacy and safety of CT-P13 compared with innovator infliximab when coadministered with methotrexate in patients with active rheumatoid arthritis: the PLANETRA study. *Annals of the Rheumatic Diseases* 2013, *72* (10), 1613.

41. Park, W.; Hrycaj, P.; Jeka, S.; Kovalenko, V.; Lysenko, G.; Miranda, P.; Mikazane, H.; Gutierrez-Ureña, S.; Lim, M.; Lee, Y.-A.; Lee, S. J.; Kim, H.; Yoo, D. H.; Braun, J., A randomised, double-blind, multicentre, parallel-group, prospective study comparing the pharmacokinetics, safety, and efficacy of CT-P13 and innovator infliximab in patients with ankylosing spondylitis: the PLANETAS study. *Annals of the Rheumatic Diseases* 2013, *72* (10), 1605.

42. Jørgensen, K. K.; Olsen, I. C.; Goll, G. L.; Lorentzen, M.; Bolstad, N.; Haavardsholm, E. A.; Lundin, K. E. A.; Mørk, C.; Jahnsen, J.; Kvien, T. K.; Berset, I. P.; Fevang, B. T. S.; Florholmen, J.; Kalstad, S.; Mørk, N. J.; Ryggen, K.; Tveit, K. S.; Sæther, S. K.; Gulbrandsen, B.; Hagfors, J.; Waksvik, K.; Warren,

D.; Henanger, K. J.; Asak, Ø.; Baigh, S.; Blomgren, I. M.; Bruun, T. J.; Dvergsnes, K.; Frigstad, S. O.; Gjesdal, C. G.; Grandaunet, B. H. J.; Hansen, I. M.; Hatten, I. S. H.; Huppertz-Hauss, G.; Henriksen, M.; Hoie, S. S.; Krogh, J.; Kruse, J. R.; Ljoså, M.-K. A.; Midtgard, I. P.; Mielnik, P.; Moum, B.; Noraberg, G.; Poyan, A.; Prestegård, U.; Rashid, H. U.; Rydning, J. H.; Sagatun, L.; Seeberg, K. A.; Skjetne, K.; Strand, E. K.; Stray, H.; Stray, N.; Torp, R.; Vold, C.; Ystrøm, C. M.; Zettel, C. C., Switching from originator infliximab to biosimilar CT-P13 compared with maintained treatment with originator infliximab (NOR-SWITCH): a 52-week, randomised, double-blind, non-inferiority trial. *The Lancet* 2017, *389* (10086), 2304-2316.

43. Liu, L., Antibody Glycosylation and Its Impact on the Pharmacokinetics and Pharmacodynamics of Monoclonal Antibodies and Fc-Fusion Proteins. *Journal of Pharmaceutical Sciences* 2015, *104* (6), 1866-1884.

44. Bui, L. A.; Hurst, S.; Finch, G. L.; Ingram, B.; Jacobs, I. A.; Kirchhoff, C. F.; Ng, C.-K.; Ryan, A. M., Key considerations in the preclinical development of biosimilars. *Drug Discovery Today* 2015, *20*, 3-15.

45. Kwon, O.; Joung, J.; Park, Y.; Kim, C. W.; Hong, S. H., Considerations of critical quality attributes in the analytical comparability assessment of biosimilar products. *Biologicals* 2017, *48*, 101-108.

46. Kaur, H., Characterization of glycosylation in monoclonal antibodies and its importance in therapeutic antibody development. *Crit Rev Biotechnol* 2021, *41* (2), 300-315.

47. Rathore, A. S.; Winkle, H., Quality by design for biopharmaceuticals. *Nature Biotechnology* 2009, *27* (1), 26-34.

48. Narayanan, H.; Luna, M. F.; von Stosch, M.; Cruz Bournazou, M. N.; Polotti, G.; Morbidelli, M.; Butté, A.; Sokolov, M., Bioprocessing in the Digital Age: The Role of Process Models. *Biotechnology journal* 2020, *15* (1), 1900172.

49. Kiparissides, A.; Pistikopoulos, E. N.; Mantalaris, A., On the model-based optimization of secreting mammalian cell (GS-NS0) cultures. *Biotechnology and bioengineering* 2015, *112* (3), 536-548.

50. Zhang, L.; Schwarz, H.; Wang, M.; Castan, A.; Hjalmarsson, H.; Chotteau, V., Control of IgG glycosylation in CHO cell perfusion cultures by GReBA mathematical model supported by a novel targeted feed, TAFE. *Metabolic Engineering* 2020.

51. Kol, S.; Ley, D.; Wulff, T.; Decker, M.; Arnsdorf, J.; Schoffelen, S.; Hansen, A. H.; Jensen, T. L.; Gutierrez, J. M.; Chiang, A. W. T.; Masson, H. O.; Palsson, B. O.; Voldborg, B. G.; Pedersen, L. E.; Kildegaard, H. F.; Lee, G. M.; Lewis, N. E., Multiplex secretome engineering enhances recombinant protein production and purity. *Nature Communications* 2020, *11* (1), 1908.

52. Gutierrez, J. M.; Feizi, A.; Li, S.; Kallehauge, T. B.; Hefzi, H.; Grav, L. M.; Ley, D.; Baycin Hizal, D.; Betenbaugh, M. J.; Voldborg, B.; Faustrup Kildegaard, H.; Min Lee, G.; Palsson, B. O.; Nielsen, J.;

Lewis, N. E., Genome-scale reconstructions of the mammalian secretory pathway predict metabolic costs and limitations of protein secretion. *Nature Communications* 2020, *11* (1), 68.

53. Presnell, K. V.; Alper, H. S., Systems Metabolic Engineering Meets Machine Learning: A New Era for Data-Driven Metabolic Engineering. *Biotechnology journal* 2019, *14* (9), 1800416.

54. Zampieri, G.; Vijayakumar, S.; Yaneske, E.; Angione, C., Machine and deep learning meet genome-scale metabolic modeling. *PLOS Computational Biology* 2019, *15* (7), e1007084.

55. Hayes, J. M.; Cosgrave, E. F. J.; Struwe, W. B.; Wormald, M.; Davey, G. P.; Jefferis, R.; Rudd, P. M., Glycosylation and Fc Receptors. In *Fc Receptors*, Daeron, M.; Nimmerjahn, F., Eds. Springer International Publishing: Cham, 2014; pp 165-199.

56. Reily, C.; Stewart, T. J.; Renfrow, M. B.; Novak, J., Glycosylation in health and disease. *Nature Reviews Nephrology* 2019, *15* (6), 346-366.

57. Marth, J. D.; Grewal, P. K., Mammalian glycosylation in immunity. *Nat Rev Immunol* 2008, *8* (11), 874-887.

58. Wong, C.-H., Protein Glycosylation: New Challenges and Opportunities. *The Journal of Organic Chemistry* 2005, *70* (11), 4219-4225.

59. Flynn, R. A.; Pedram, K.; Malaker, S. A.; Batista, P. J.; Smith, B. A. H.; Johnson, A. G.; George, B. M.; Majzoub, K.; Villalta, P. W.; Carette, J. E.; Bertozzi, C. R., Small RNAs are modified with N-glycans and displayed on the surface of living cells. *Cell* 2021.

60. Eichler, J.; Koomey, M., Sweet New Roles for Protein Glycosylation in Prokaryotes. *Trends in Microbiology* 2017, *25* (8), 662-672.

61. Zarschler, K.; Janesch, B.; Pabst, M.; Altmann, F.; Messner, P.; Schäffer, C., Protein tyrosine O-glycosylation—A rather unexplored prokaryotic glycosylation system. *Glycobiology* 2010, *20* (6), 787-798.

62. Iwashkiw, J. A.; Vozza, N. F.; Kinsella, R. L.; Feldman, M. F., Pour some sugar on it: the expanding world of bacterial protein O-linked glycosylation. *Molecular Microbiology* 2013, *89* (1), 14-28.

63. Nothaft, H.; Szymanski, C. M., Bacterial Protein N-Glycosylation: New Perspectives and Applications. *The Journal of Biological Chemistry* 2013, *288* (10), 6912-6920.

64. Abouelhadid, S.; North, S. J.; Hitchen, P.; Vohra, P.; Chintoan-Uta, C.; Stevens, M.; Dell, A.; Cuccui, J.; Wren, B. W., Quantitative Analyses Reveal Novel Roles for *N*-Glycosylation in a Major Enteric Bacterial Pathogen. *mBio* 2019, *10* (2), e00297-19.

65. Lu, Q.; Li, S.; Shao, F., Sweet Talk: Protein Glycosylation in Bacterial Interaction With the Host. *Trends in Microbiology* 2015, *23* (10), 630-641.

66. Nothaft, H.; Szymanski, C. M., New discoveries in bacterial N-glycosylation to expand the synthetic biology toolbox. *Current Opinion in Chemical Biology* 2019, *53*, 16-24.

67. Wandall, H. H.; Rumjantseva, V.; Sørensen, A. L. T.; Patel-Hett, S.; Josefsson, E. C.; Bennett, E. P.; Italiano, J. E., Jr.; Clausen, H.; Hartwig, J. H.; Hoffmeister, K. M., The origin and function of platelet glycosyltransferases. *Blood* 2012, *120* (3), 626-635.

68. Manhardt, C. T.; Punch, P. R.; Dougher, C. W. L.; Lau, J. T. Y., Extrinsic sialylation is dynamically regulated by systemic triggers in vivo. *Journal of Biological Chemistry* 2017, *292* (33), 13514-13520.

69. Lee-Sundlov, M. M.; Ashline, D. J.; Hanneman, A. J.; Grozovsky, R.; Reinhold, V. N.; Hoffmeister, K. M.; Lau, J. T. Y., Circulating blood and platelets supply glycosyltransferases that enable extrinsic extracellular glycosylation. *Glycobiology* 2017, *27* (2), 188-198.

70. Stanley, P., Golgi Glycosylation. *Cold Spring Harbor Perspectives in Biology* 2011, *3* (4).

71. Moremen, K. W.; Tiemeyer, M.; Nairn, A. V., Vertebrate protein glycosylation: diversity, synthesis and function. *Nature Reviews Molecular Cell Biology* 2012, *13* (7), 448-462.

72. Dube, D. H.; Prescher, J. A.; Quang, C. N.; Bertozzi, C. R., Probing mucin-type O-linked glycosylation in living animals. *Proceedings of the National Academy of Sciences of the United States of America* 2006, *103* (13), 4819.

73. Hang, H. C.; Bertozzi, C. R., The chemistry and biology of mucin-type O-linked glycosylation. *Bioorganic & Medicinal Chemistry* 2005, *13* (17), 5021-5034.

74. Hanisch, F.-G., O-Glycosylation of the Mucin Type. *Biological Chemistry* 2001, *382* (2), 143-149.

75. Allen, A. C.; Harper, S. J.; Feehally, J., Galactosylation of N- and O-linked carbohydrate moieties of IgA1 and IgG in IgA nephropathy. *Clinical and Experimental Immunology* 1995, *100* (3), 470-474.

76. Strasser, R., Challenges in O-glycan engineering of plants. *Frontiers in Plant Science* 2012, *3*, 218.

77. Hirayama, K.; Yuji, R.; Yamada, N.; Kato, K.; Arata, Y.; Shimada, I., Complete and Rapid Peptide and Glycopeptide Mapping of Mouse Monoclonal Antibody by LC/MS/MS Using Ion Trap Mass Spectrometry. *Analytical Chemistry* 1998, *70* (13), 2718-2725.

78. Leibiger, H.; Wüstner, D.; Stigler, R. D.; Marx, U., Variable domain-linked oligosaccharides of a human monoclonal IgG: structure and influence on antigen binding. *Biochemical Journal* 1999, *338* (Pt 2), 529-538.

79. Lehoux, S.; Ju, T., Chapter Five - Separation of Two Distinct O-Glycoforms of Human IgA1 by Serial Lectin Chromatography Followed by Mass Spectrometry O-Glycan Analysis. In *Methods in Enzymology*, Shukla, A. K., Ed. Academic Press: 2017; Vol. 585, pp 61-75.

80. Sjögren, J.; Lood, R.; Nägeli, A., On enzymatic remodeling of IgG glycosylation; unique tools with broad applications. *Glycobiology* 2020, *30* (4), 254-267.

81. Lippincott-Schwartz, J.; Roberts, T. H.; Hirschberg, K., Secretory Protein Trafficking and Organelle Dynamics in Living Cells. *Annual Review of Cell and Developmental Biology* 2000, *16* (1), 557-589.

82. Hubbard, S. C.; Ivatt, R. J., Synthesis and Processing of Asparagine-Linked Oligosaccharides. *Annual Review of Biochemistry* 1981, *50* (1), 555-583.

83. Varki, A.; Cummings, R. D.; Aebi, M.; Packer, N. H.; Seeberger, P. H.; Esko, J. D.; Stanley, P.; Hart, G.; Darvill, A.; Kinoshita, T.; Prestegard, J. J.; Schnaar, R. L.; Freeze, H. H.; Marth, J. D.; Bertozzi, C. R.; Etzler, M. E.; Frank, M.; Vliegenthart, J. F. G.; Lütteke, T.; Perez, S.; Bolton, E.; Rudd, P.; Paulson, J.; Kanehisa, M.; Toukach, P.; Aoki-Kinoshita, K. F.; Dell, A.; Narimatsu, H.; York, W.; Taniguchi, N.; Kornfeld, S., Symbol Nomenclature for Graphical Representations of Glycans. *Glycobiology* 2015, *25* (12), 1323-1324.

84. Bloch, J. S.; Pesciullesi, G.; Boilevin, J.; Nosol, K.; Irobalieva, R. N.; Darbre, T.; Aebi, M.; Kossiakoff, A. A.; Reymond, J.-L.; Locher, K. P., Structure and mechanism of the ER-based glucosyltransferase ALG6. *Nature* 2020, *579* (7799), 443-447.

85. Sanyal, S.; Frank, C. G.; Menon, A. K., Distinct flippases translocate glycerophospholipids and oligosaccharide diphosphate dolichols across the endoplasmic reticulum. *Biochemistry* 2008, *47* (30), 7937-7946.

86. Sanyal, S.; Menon, A. K., Specific transbilayer translocation of dolichol-linked oligosaccharides by an endoplasmic reticulum flippase. *Proceedings of the National Academy of Sciences of the United States of America* 2009, *106* (3), 767-772.

87. Sanyal, S.; Menon, A. K., Stereoselective transbilayer translocation of mannosyl phosphoryl dolichol by an endoplasmic reticulum flippase. *Proceedings of the National Academy of Sciences* 2010, *107* (25), 11289.

88. Rush, J. S., Role of Flippases in Protein Glycosylation in the Endoplasmic Reticulum. *Lipid Insights* 2016, *8* (Suppl 1), 45-53.

89. Caragea, C.; Sinapov, J.; Silvescu, A.; Dobbs, D.; Honavar, V., Glycosylation site prediction using ensembles of Support Vector Machine classifiers. *BMC Bioinformatics* 2007, *8* (1), 438.

90. Taherzadeh, G.; Dehzangi, A.; Golchin, M.; Zhou, Y.; Campbell, M. P., SPRINT-Gly: predicting N- and O-linked glycosylation sites of human and mouse proteins by using sequence and predicted structural properties. *Bioinformatics* 2019, *35* (20), 4140-4146.

91. Pitti, T.; Chen, C.-T.; Lin, H.-N.; Choong, W.-K.; Hsu, W.-L.; Sung, T.-Y., N-GlyDE: a two-stage N-linked glycosylation site prediction incorporating gapped dipeptides and pattern-based encoding. *Scientific Reports* 2019, *9* (1), 15975.

92. Chien, C. H.; Chang, C. C.; Lin, S. H.; Chen, C. W.; Chang, Z. H.; Chu, Y. W., N-GlycoGo: Predicting Protein N-Glycosylation Sites on Imbalanced Data Sets by Using Heterogeneous and Comprehensive Strategy. *IEEE Access* 2020, *8*, 165944-165950.

93. Petrescu, A.-J.; Milac, A.-L.; Petrescu, S. M.; Dwek, R. A.; Wormald, M. R., Statistical analysis of the protein environment of N-glycosylation sites: implications for occupancy, structure, and folding. *Glycobiology* 2004, *14* (2), 103-114.

94. de Silva, A. M.; Balch, W. E.; Helenius, A., Quality control in the endoplasmic reticulum: folding and misfolding of vesicular stomatitis virus G protein in cells and in vitro. *Journal of Cell Biology* 1990, *111* (3), 857-866.

95. Araki, K.; Nagata, K., Protein folding and quality control in the ER. *Cold Spring Harbor perspectives in biology* 2011, *3* (11), a007526-a007526.

96. Ferris, S. P.; Kodali, V. K.; Kaufman, R. J., Glycoprotein folding and quality-control mechanisms in protein-folding diseases. *Disease Models & Mechanisms* 2014, *7* (3), 331-341.

97. Benyair, R.; Ogen-Shtern, N.; Mazkereth, N.; Shai, B.; Ehrlich, M.; Lederkremer, G. Z., Mammalian ER mannosidase I resides in quality control vesicles, where it encounters its glycoprotein substrates. *Molecular biology of the cell* 2015, *26* (2), 172-184.

98. Molinari, M.; Helenius, A., Chaperone Selection During Glycoprotein Translocation into the Endoplasmic Reticulum. *Science* 2000, *288* (5464), 331.

99. Williams, D. B., Beyond lectins: the calnexin/calreticulin chaperone system of the endoplasmic reticulum. *Journal of Cell Science* 2006, *119* (4), 615.

100. Pobre, K. F. R.; Poet, G. J.; Hendershot, L. M., The endoplasmic reticulum (ER) chaperone BiP is a master regulator of ER functions: Getting by with a little help from ERdj friends. *The Journal of biological chemistry* 2019, *294* (6), 2098-2108.

101. Ermonval, M.; Kitzmüller, C.; Mir, A. M.; Cacan, R.; Ivessa, N. E., N-glycan structure of a short-lived variant of ribophorin I expressed in the MadIA214 glycosylation-defective cell line reveals the role of a mannosidase that is not ER mannosidase I in the process of glycoprotein degradation. *Glycobiology* 2001, *11* (7), 565-576.

102. Frenkel, Z.; Gregory, W.; Kornfeld, S.; Lederkremer, G. Z., Endoplasmic Reticulum-associated Degradation of Mammalian Glycoproteins Involves Sugar Chain Trimming to Man6–5GlcNAc2. *Journal of Biological Chemistry* 2003, *278* (36), 34119-34124.

103. Hosokawa, N.; Tremblay, L. O.; You, Z.; Herscovics, A.; Wada, I.; Nagata, K., Enhancement of Endoplasmic Reticulum (ER) Degradation of Misfolded Null Hong Kong α1-Antitrypsin by Human ER Mannosidase I. *Journal of Biological Chemistry* 2003, *278* (28), 26287-26294.

104. Avezov, E.; Frenkel, Z.; Ehrlich, M.; Herscovics, A.; Lederkremer, G. Z., Endoplasmic Reticulum (ER) Mannosidase I Is Compartmentalized and Required for N-Glycan Trimming to Man5–6GlcNAc2 in Glycoprotein ER-associated Degradation. *Molecular Biology of the Cell* 2007, *19* (1), 216-225.

105. Shenkman, M.; Ron, E.; Yehuda, R.; Benyair, R.; Khalaila, I.; Lederkremer, G. Z., Mannosidase activity of EDEM1 and EDEM2 depends on an unfolded state of their glycoprotein substrates. *Communications Biology* 2018, *1* (1), 172.

106. Li, S.-T.; Lu, T.-T.; Xu, X.-X.; Ding, Y.; Li, Z.; Kitajima, T.; Dean, N.; Wang, N.; Gao, X.-D., Reconstitution of the lipid-linked oligosaccharide pathway for assembly of high-mannose N-glycans. *Nature Communications* 2019, *10* (1), 1813.

107. Gemmill, T. R.; Trimble, R. B., Overview of N- and O-linked oligosaccharide structures found in various yeast species. *Biochimica et Biophysica Acta (BBA) - General Subjects* 1999, *1426* (2), 227-237.

108. Dunphy, W. G.; Fries, E.; Urbani, L. J.; Rothman, J. E., Early and late functions associated with the Golgi apparatus reside in distinct compartments. *Proceedings of the National Academy of Sciences of the United States of America* 1981, *78* (12), 7453-7457.

109. Dunphy, W. G.; Rothman, J. E., Compartmental organization of the golgi stack. *Cell* 1985, *42* (1), 13-21.

110. Farquhar, M. G., Progress in Unraveling Pathways of Golgi Traffic. *Annual Review of Cell Biology* 1985, *1* (1), 447-488.

111. Kleene, R.; Berger, E. G., The molecular and cell biology of glycosyltransferases. *Biochimica et Biophysica Acta (BBA) - Reviews on Biomembranes* 1993, *1154* (3), 283-325.

112. Nilsson, T.; Au, C. E.; Bergeron, J. J. M., Sorting out glycosylation enzymes in the Golgi apparatus. *FEBS Letters* 2009, *583* (23), 3764-3769.

113. Rabouille, C.; Hui, N.; Hunte, F.; Kieckbusch, R.; Berger, E. G.; Warren, G.; Nilsson, T., Mapping the distribution of Golgi enzymes involved in the construction of complex oligosaccharides. *Journal of Cell Science* 1995, *108* (4), 1617.

114. Nilsson, T.; Pypaert, M.; Hoe, M. H.; Slusarewicz, P.; Berger, E. G.; Warren, G., Overlapping distribution of two glycosyltransferases in the Golgi apparatus of HeLa cells. *Journal of Cell Biology* 1993, *120* (1), 5-13.

115. Colley, K. J., Golgi localization of glycosyltransferases: more questions than answers. *Glycobiology* 1997, *7* (1), 1-13.

116. Staehelin, L. A.; Kang, B.-H., Nanoscale Architecture of Endoplasmic Reticulum Export Sites and of Golgi Membranes as Determined by Electron Tomography. *Plant Physiology* 2008, *147* (4), 1454.

117. Storrie, B.; White, J.; Röttger, S.; Stelzer, E. H.; Suganuma, T.; Nilsson, T., Recycling of golgi-resident glycosyltransferases through the ER reveals a novel pathway and provides an explanation for nocodazole-induced Golgi scattering. *J Cell Biol* 1998, *143* (6), 1505-1521.

118. Liu, L.; Doray, B.; Kornfeld, S., Recycling of Golgi glycosyltransferases requires direct binding to coatomer. *Proceedings of the National Academy of Sciences* 2018, *115* (36), 8984.

119. Kellokumpu, S.; Hassinen, A.; Glumoff, T., Glycosyltransferase complexes in eukaryotes: long-known, prevalent but still unrecognized. *Cell Mol Life Sci* 2016, *73* (2), 305-325.

120. Hadley, B.; Litfin, T.; Day, C. J.; Haselhorst, T.; Zhou, Y.; Tiralongo, J., Nucleotide Sugar Transporter SLC35 Family Structure and Function. *Computational and Structural Biotechnology Journal* 2019, *17*, 1123-1134.

121. Jedrzejewski, P. M.; Del Val, I. J.; Constantinou, A.; Dell, A.; Haslam, S. M.; Polizzi, K. M.; Kontoravdi, C., Towards Controlling the Glycoform: A Model Framework Linking Extracellular Metabolites to Antibody Glycosylation. *Int J Mol Sci* 2014, *15* (3).

122. Murrell, M. P.; Yarema, K. J.; Levchenko, A., The Systems Biology of Glycosylation. *ChemBioChem* 2004, *5* (10), 1334-1347.

123. Hadley, B.; Maggioni, A.; Ashikov, A.; Day, C. J.; Haselhorst, T.; Tiralongo, J., Structure and function of nucleotide sugar transporters: Current progress. *Computational and structural biotechnology journal* 2014, *10* (16), 23-32.

124. Parker, J. L.; Newstead, S., Gateway to the Golgi: molecular mechanisms of nucleotide sugar transporters. *Current Opinion in Structural Biology* 2019, *57*, 127-134.

125. Parker, J. L.; Corey, R. A.; Stansfeld, P. J.; Newstead, S., Structural basis for substrate specificity and regulation of nucleotide sugar transporters in the lipid bilayer. *Nature Communications* 2019, *10* (1), 4657.

126. Pelham, H. R. B.; Rothman, J. E., The Debate about Transport in the Golgi—Two Sides of the Same Coin? *Cell* 2000, *102* (6), 713-719.

127. Stalder, D.; Gershlick, D. C., Direct trafficking pathways from the Golgi apparatus to the plasma membrane. *Seminars in Cell & Developmental Biology* 2020, *107*, 112-125.

128. Glick, B. S.; Luini, A., Models for Golgi traffic: a critical assessment. *Cold Spring Harbor perspectives in biology* 2011, *3* (11), a005215-a005215.

129. Beznoussenko, G. V.; Mironov, A. A., Models of intracellular transport and evolution of the Golgi complex. *The Anatomical Record* 2002, *268* (3), 226-238.

130. Pfeffer, S. R., How the Golgi works: A cisternal progenitor model. *Proceedings of the National Academy of Sciences* 2010, *107* (46), 19614.

131. Rothman, J. E.; Wieland, F. T., Protein Sorting by Transport Vesicles. *Science* 1996, *272* (5259), 227.

132. Mironov, A. A.; Weidman, P.; Luini, A., Variations on the Intracellular Transport Theme: Maturing Cisternae and Trafficking Tubules. *Journal of Cell Biology* 1997, *138* (3), 481-484.

133. Mironov, A. A.; Beznoussenko, G. V., Models of Intracellular Transport: Pros and Cons. *Frontiers in Cell and Developmental Biology* 2019, *7* (146).

134. Higgins, E., Carbohydrate analysis throughout the development of a protein therapeutic. *Glycoconjugate Journal* 2010, *27* (2), 211-225.

135. Lee, H. S.; Qi, Y.; Im, W., Effects of N-glycosylation on protein conformation and dynamics: Protein Data Bank analysis and molecular dynamics simulation study. *Scientific Reports* 2015, *5* (1), 8926.

136. Shental-Bechor, D.; Levy, Y., Effect of glycosylation on protein folding: A close look at thermodynamic stabilization. *Proceedings of the National Academy of Sciences* 2008, *105* (24), 8256.

137. Solá, R. J.; Griebenow, K., Effects of glycosylation on the stability of protein pharmaceuticals. *Journal of Pharmaceutical Sciences* 2009, *98* (4), 1223-1245.

138. Zhang, P.; Woen, S.; Wang, T.; Liau, B.; Zhao, S.; Chen, C.; Yang, Y.; Song, Z.; Wormald, M. R.; Yu, C.; Rudd, P. M., Challenges of glycosylation analysis and control: an integrated approach to producing optimal and consistent therapeutic drugs. *Drug Discovery Today* 2016, *21* (5), 740-765.

139. Dubé, S.; Fisher, J. W.; Powell, J. S., Glycosylation at specific sites of erythropoietin is essential for biosynthesis, secretion, and biological function. *Journal of Biological Chemistry* 1988, *263* (33), 17516-17521.

140. Jiang, J.; Tian, F.; Cai, Y.; Qian, X.; Costello, C. E.; Ying, W., Site-specific qualitative and quantitative analysis of the N- and O-glycoforms in recombinant human erythropoietin. *Anal Bioanal Chem* 2014, *406* (25), 6265-6274.

141. van de Bovenkamp, F. S.; Hafkenscheid, L.; Rispens, T.; Rombouts, Y., The Emerging Importance of IgG Fab Glycosylation in Immunity. *The Journal of Immunology* 2016, *196* (4), 1435.

142. Strebhardt, K.; Ullrich, A., Paul Ehrlich's magic bullet concept: 100 years of progress. *Nature Reviews Cancer* 2008, *8* (6), 473-480.

143. Köhler, G.; Milstein, C., Continuous cultures of fused cells secreting antibody of predefined specificity. *Nature* 1975, *256* (5517), 495-497.

144. Lu, R.-M.; Hwang, Y.-C.; Liu, I. J.; Lee, C.-C.; Tsai, H.-Z.; Li, H.-J.; Wu, H.-C., Development of therapeutic antibodies for the treatment of diseases. *Journal of Biomedical Science* 2020, *27* (1), 1.

145. Rodgers, K. R.; Chou, R. C., Therapeutic monoclonal antibodies and derivatives: Historical perspectives and future directions. *Biotechnology Advances* 2016, *34* (6), 1149-1158.

146. Vidarsson, G.; Dekkers, G.; Rispens, T., IgG subclasses and allotypes: from structure to effector functions. *Front Immunol* 2014, *5*, 520-520.

147. Ryman, J. T.; Meibohm, B., Pharmacokinetics of Monoclonal Antibodies. *CPT: Pharmacometrics & Systems Pharmacology* 2017, *6* (9), 576-588.

148. Liu, H.; May, K., Disulfide bond structures of IgG molecules: structural variations, chemical modifications and possible impacts to stability and biological function. *MAbs* 2012, *4* (1), 17-23.

149. McAuley, A.; Jacob, J.; Kolvenbach, C. G.; Westland, K.; Lee, H. J.; Brych, S. R.; Rehder, D.; Kleemann, G. R.; Brems, D. N.; Matsumura, M., Contributions of a disulfide bond to the structure, stability, and dimerization of human IgG1 antibody CH3 domain. *Protein Sci* 2008, *17* (1), 95-106.

150. Maverakis, E.; Kim, K.; Shimoda, M.; Gershwin, M. E.; Patel, F.; Wilken, R.; Raychaudhuri, S.; Ruhaak, L. R.; Lebrilla, C. B., Glycans in the immune system and The Altered Glycan Theory of Autoimmunity: A critical review. *Journal of Autoimmunity* 2015, *57*, 1-13.

151. Conroy, P. J.; Law, R. H. P.; Caradoc-Davies, T. T.; Whisstock, J. C., Antibodies: From novel repertoires to defining and refining the structure of biologically important targets. *Methods* 2017, *116* (Supplement C), 12-22.

152. Seidel, U.; Schlegel, P.; Lang, P., Natural Killer Cell Mediated Antibody-Dependent Cellular Cytotoxicity in Tumor Immunotherapy with Therapeutic Antibodies. *Front Immunol* 2013, *4* (76).

153. Chames, P.; Van Regenmortel, M.; Weiss, E.; Baty, D., Therapeutic antibodies: successes, limitations and hopes for the future. *Br J Pharmacol* 2009, *157* (2), 220-233.

154. Bournazos, S.; Gupta, A.; Ravetch, J. V., The role of IgG Fc receptors in antibody-dependent enhancement. *Nature Reviews Immunology* 2020, *20* (10), 633-643.

155. Cobb, B. A., The history of IgG glycosylation and where we are now. *Glycobiology* 2020, *30* (4), 202-213.

156. Wright, A.; Morrison, S. L., Effect of glycosylation on antibody function: implications for genetic engineering. *Trends in Biotechnology* 1997, *15* (1), 26-32.

157. Boune, S.; Hu, P.; Epstein, A. L.; Khawli, L. A., Principles of N-Linked Glycosylation Variations of IgG-Based Therapeutics: Pharmacokinetic and Functional Considerations. *Antibodies (Basel)* 2020, *9* (2), 22.

158. Zheng, K.; Bantog, C.; Bayer, R., The impact of glycosylation on monoclonal antibody conformation and stability. *MAbs* 2011, *3* (6), 568-576.

159. Boyd, P. N.; Lines, A. C.; Patel, A. K., The effect of the removal of sialic acid, galactose and total carbohydrate on the functional activity of Campath-1H. *Molecular Immunology* 1995, *32* (17), 1311-1318.

160. Cambay, F.; Raymond, C.; Brochu, D.; Gilbert, M.; Tu, T. M.; Cantin, C.; Lenferink, A.; Grail, M.; Henry, O.; De Crescenzo, G.; Durocher, Y., Impact of IgG1 N-glycosylation on their interaction with Fc gamma receptors. *Current Research in Immunology* 2020, *1*, 23-37.

161. Shields, R. L.; Lai, J.; Keck, R.; O'Connell, L. Y.; Hong, K.; Meng, Y. G.; Weikert, S. H. A.; Presta, L. G., Lack of Fucose on Human IgG1 N-Linked Oligosaccharide Improves Binding to Human FcγRIII and Antibody-dependent Cellular Toxicity. *Journal of Biological Chemistry* 2002, *277* (30), 26733-26740.

162. Chung, S.; Quarmby, V.; Gao, X.; Ying, Y.; Lin, L.; Reed, C.; Fong, C.; Lau, W.; Qiu, Z. J.; Shen, A.; Vanderlaan, M.; Song, A., Quantitative evaluation of fucose reducing effects in a humanized antibody on Fcγ receptor binding and antibody-dependent cell-mediated cytotoxicity activities. *MAbs* 2012, *4* (3), 326-340.

163. Niwa, R.; Natsume, A.; Uehara, A.; Wakitani, M.; Iida, S.; Uchida, K.; Satoh, M.; Shitara, K., IgG subclass-independent improvement of antibody-dependent cellular cytotoxicity by fucose removal from Asn297-linked oligosaccharides. *Journal of Immunological Methods* 2005, *306* (1), 151-160.

164. Li, T.; DiLillo, D. J.; Bournazos, S.; Giddens, J. P.; Ravetch, J. V.; Wang, L.-X., Modulating IgG effector function by Fc glycan engineering. *Proceedings of the National Academy of Sciences* 2017, *114* (13), 3485.

165. Kanda, Y.; Imai-Nishiya, H.; Kuni-Kamochi, R.; Mori, K.; Inoue, M.; Kitajima-Miyama, K.; Okazaki, A.; Iida, S.; Shitara, K.; Satoh, M., Establishment of a GDP-mannose 4,6-dehydratase (GMD) knockout host cell line: A new strategy for generating completely non-fucosylated recombinant therapeutics. *Journal of biotechnology* 2007, *130* (3), 300-310.

166. Lifely, M. R.; Hale, C.; Boyce, S.; Keen, M. J.; Phillips, J., Glycosylation and biological activity of CAMPATH-1H expressed in different cell lines and grown under different culture conditions. *Glycobiology* 1995, *5* (8), 813-822.

167. Umaña, P.; Jean–Mairet, J.; Moudry, R.; Amstutz, H.; Bailey, J. E., Engineered glycoforms of an antineuroblastoma IgG1 with optimized antibody-dependent cellular cytotoxic activity. *Nature Biotechnology* 1999, *17* (2), 176-180.

168. Davies, J.; Jiang, L.; Pan, L.-Z.; LaBarre, M. J.; Anderson, D.; Reff, M., Expression of GnTIII in a recombinant anti-CD20 CHO production cell line: Expression of antibodies with altered glycoforms leads to an increase in ADCC through higher affinity for FCγRIII. *Biotechnology and bioengineering* 2001, *74* (4), 288-294.

169. Shinkawa, T.; Nakamura, K.; Yamane, N.; Shoji-Hosaka, E.; Kanda, Y.; Sakurada, M.; Uchida, K.; Anazawa, H.; Satoh, M.; Yamasaki, M.; Hanai, N.; Shitara, K., The Absence of Fucose but Not the Presence of Galactose or Bisecting N-Acetylglucosamine of Human IgG1 Complex-type Oligosaccharides Shows the Critical Role of Enhancing Antibody-dependent Cellular Cytotoxicity. *Journal of Biological Chemistry* 2003, *278* (5), 3466-3473.

170. Ferrara, C.; Brünker, P.; Suter, T.; Moser, S.; Püntener, U.; Umaña, P., Modulation of therapeutic antibody effector functions by glycosylation engineering: Influence of Golgi enzyme localization domain and co-expression of heterologous β1, 4-N-acetylglucosaminyltransferase III and Golgi α-mannosidase II. *Biotechnology and bioengineering* 2006, *93* (5), 851-861.

171. Hodoniczky, J.; Zheng, Y. Z.; James, D. C., Control of Recombinant Monoclonal Antibody Effector Functions by Fc N-Glycan Remodeling in Vitro. *Biotechnology Progress* 2005, *21* (6), 1644-1652.

172. Morell, A. G.; Gregoriadis, G.; Scheinberg, I. H.; Hickman, J.; Ashwell, G., The Role of Sialic Acid in Determining the Survival of Glycoproteins in the Circulation. *Journal of Biological Chemistry* 1971, *246* (5), 1461-1467.

173. Wide, L.; Eriksson, K.; Sluss, P. M.; Hall, J. E., Serum Half-Life of Pituitary Gonadotropins Is Decreased by Sulfonation and Increased by Sialylation in Women. *The Journal of Clinical Endocrinology & Metabolism* 2009, *94* (3), 958-964.

174. Bas, M.; Terrier, A.; Jacque, E.; Dehenne, A.; Pochet-Béghin, V.; Beghin, C.; Dezetter, A.-S.; Dupont, G.; Engrand, A.; Beaufils, B.; Mondon, P.; Fournier, N.; de Romeuf, C.; Jorieux, S.; Fontayne, A.; Mars, L. T.; Monnet, C., Fc Sialylation Prolongs Serum Half-Life of Therapeutic Antibodies. *The Journal of Immunology* 2019, ji1800896.

175. Scallon, B. J.; Tam, S. H.; McCarthy, S. G.; Cai, A. N.; Raju, T. S., Higher levels of sialylated Fc glycans in immunoglobulin G molecules can adversely impact functionality. *Molecular Immunology* 2007, *44* (7), 1524-1534.

176. Naso, M. F.; Tam, S. H.; Scallon, B. J.; Raju, T. S., Engineering host cell lines to reduce terminal sialylation of secreted antibodies. *MAbs* 2010, *2* (5), 519-527.

177. Wada, R.; Matsui, M.; Kawasaki, N., Influence of N-glycosylation on effector functions and thermal stability of glycoengineered IgG1 monoclonal antibody with homogeneous glycoforms. *MAbs* 2019, *11* (2), 350-372.

178. Zhang, Z.; Shah, B.; Richardson, J., Impact of Fc N-glycan sialylation on IgG structure. *MAbs* 2019, *11* (8), 1381-1390.

179. Kaneko, Y.; Nimmerjahn, F.; Ravetch, J. V., Anti-Inflammatory Activity of Immunoglobulin G Resulting from Fc Sialylation. *Science* 2006, *313* (5787), 670.

180. Anthony, R. M.; Nimmerjahn, F.; Ashline, D. J.; Reinhold, V. N.; Paulson, J. C.; Ravetch, J. V., Recapitulation of IVIG Anti-Inflammatory Activity with a Recombinant IgG Fc. *Science* 2008, *320* (5874), 373.

181. Kumpel, B. M.; Rademacher, T. W.; Rook, G. A.; Williams, P. J.; Wilson, I. B., Galactosylation of human IgG monoclonal anti-D produced by EBV-transformed B-lymphoblastoid cell lines is dependent on culture method and affects Fc receptor-mediated functional activity. *Human antibodies and hybridomas* 1994, *5* (3-4), 143-51.

182. Thomann, M.; Reckermann, K.; Reusch, D.; Prasser, J.; Tejada, M. L., Fc-galactosylation modulates antibody-dependent cellular cytotoxicity of therapeutic antibodies. *Molecular Immunology* 2016, *73*, 69-75.

183. Bheemareddy, B. R.; Pulipeta, M.; Iyer, P.; Dirisala, V. R., Effect of the total galactose content on complement-dependent cytotoxicity of the therapeutic anti-CD20 IgG1 antibodies under temperature stress conditions. *Journal of Carbohydrate Chemistry* 2019, *38* (1), 1-19.

184. Raju, T. S.; Jordan, R. E., Galactosylation variations in marketed therapeutic antibodies. *MAbs* 2012, *4* (3), 385-391.

185. Higel, F.; Sandl, T.; Kao, C.-Y.; Pechinger, N.; Sörgel, F.; Friess, W.; Wolschin, F.; Seidl, A., N-glycans of complex glycosylated biopharmaceuticals and their impact on protein clearance. *European Journal of Pharmaceutics and Biopharmaceutics* 2019, *139*, 123-131.

186. Patel, D.; Guo, X.; Ng, S.; Melchior, M.; Balderes, P.; Burtrum, D.; Persaud, K.; Luna, X.; Ludwig, D. L.; Kang, X., IgG isotype, glycosylation, and EGFR expression determine the induction of antibody-dependent cellular cytotoxicity in vitro by cetuximab. *Human Antibodies* 2010, *19*, 89-99.

187. Eon-Duval, A.; Broly, H.; Gleixner, R., Quality attributes of recombinant therapeutic proteins: An assessment of impact on safety and efficacy as part of a quality by design development approach. *Biotechnology Progress* 2012, *28* (3), 608-622.

188. Newkirk, M. M.; Novick, J.; Stevenson, M. M.; Fournier, M. J.; Apostolakos, P., Differential clearance of glycoforms of IgG in normal and autoimmune-prone mice. *Clinical & Experimental Immunology* 1996, *106* (2), 259-264.

189. Datta-Mannan, A., Mechanisms Influencing the Pharmacokinetics and Disposition of Monoclonal Antibodies and Peptides. *Drug Metabolism and Disposition* 2019, *47* (10), 1100.

190. Galili, U., The α-gal epitope (Galα1-3Galβ1-4GlcNAc-R) in xenotransplantation. *Biochimie* 2001, *83* (7), 557-563.

191. Millward, T. A.; Heitzmann, M.; Bill, K.; Längle, U.; Schumacher, P.; Forrer, K., Effect of constant and variable domain glycosylation on pharmacokinetics of therapeutic antibodies in mice. *Biologicals* 2008, *36* (1), 41-47.

192. Malhotra, R.; Wormald, M. R.; Rudd, P. M.; Fischer, P. B.; Dwek, R. A.; Sim, R. B., Glycosylation changes of IgG associated with rheumatooid arthritis can activate complement via the mannose-binding protein. *Nature Medicine* 1995, *1* (3), 237-243.

193. Jones, A. J. S.; Papac, D. I.; Chin, E. H.; Keck, R.; Baughman, S. A.; Lin, Y. S.; Kneer, J.; Battersby, J. E., Selective clearance of glycoforms of a complex glycoprotein pharmaceutical caused by terminal N-acetylglucosamine is similar in humans and cynomolgus monkeys. *Glycobiology* 2007, *17* (5), 529-540.

194. Huang, L.; Biolsi, S.; Bales, K. R.; Kuchibhotla, U., Impact of variable domain glycosylation on antibody clearance: An LC/MS characterization. *Analytical Biochemistry* 2006, *349* (2), 197-207.

195. Goetze, A. M.; Liu, Y. D.; Zhang, Z.; Shah, B.; Lee, E.; Bondarenko, P. V.; Flynn, G. C., High-mannose glycans on the Fc region of therapeutic IgG antibodies increase serum clearance in humans. *Glycobiology* 2011, *21* (7), 949-959.

196. Alessandri, L.; Ouellette, D.; Acquah, A.; Rieser, M.; Leblond, D.; Saltarelli, M.; Radziejewski, C.; Fujimori, T.; Correia, I., Increased serum clearance of oligomannose species present on a human IgG1 molecule. *MAbs* 2012, *4* (4), 509-520.

197. Zhou, Q.; Shankara, S.; Roy, A.; Qiu, H.; Estes, S.; McVie-Wylie, A.; Culm-Merdek, K.; Park, A.; Pan, C.; Edmunds, T., Development of a simple and rapid method for producing non-fucosylated oligomannose containing antibodies with increased effector function. *Biotechnology and bioengineering* 2008, *99* (3), 652-65.

198. Liu, L.; Stadheim, A.; Hamuro, L.; Pittman, T.; Wang, W.; Zha, D.; Hochman, J.; Prueksaritanont, T., Pharmacokinetics of IgG1 monoclonal antibodies produced in humanized Pichia pastoris with specific glycoforms: A comparative study with CHO produced materials. *Biologicals* 2011, *39* (4), 205-210.

199. Yu, M.; Brown, D.; Reed, C.; Chung, S.; Lutman, J.; Stefanich, E.; Wong, A.; Stephan, J.-P.; Bayer, R., Production, characterization and pharmacokinetic properties of antibodies with N-linked Mannose-5 glycans. *MAbs* 2012, *4* (4), 475-487.

200. Pacis, E.; Yu, M.; Autsen, J.; Bayer, R.; Li, F., Effects of cell culture conditions on antibody N-linked glycosylation--what affects high mannose 5 glycoform. *Biotechnology and bioengineering* 2011, *108* (10), 2348-58.

201. Dasgupta, S.; Navarrete, A.-M.; Bayry, J.; Delignat, S.; Wootla, B.; André, S.; Christophe, O.; Nascimbeni, M.; Jacquemin, M.; Martinez-Pomares, L.; Geijtenbeek, T. B. H.; Moris, A.; Saint-Remy, J.-M.; Kazatchkine, M. D.; Kaveri, S. V.; Lacroix-Desmazes, S., A role for exposed mannosylations in presentation of human therapeutic self-proteins to CD4+ T lymphocytes. *Proceedings of the National Academy of Sciences* 2007, *104* (21), 8965.

202. Chen, X.; Liu, Y. D.; Flynn, G. C., The effect of Fc glycan forms on human IgG2 antibody clearance in humans. *Glycobiology* 2009, *19* (3), 240-249.

203. Schlesinger, P. H.; Doebber, T. W.; Mandell, B. F.; White, R.; DeSchryver, C.; Rodman, J. S.; Miller, M. J.; Stahl, P., Plasma clearance of glycoproteins with terminal mannose and N-acetylglucosamine by liver non-parenchymal cells. Studies with beta-glucuronidase, N-acetyl-beta-D-glucosaminidase, ribonuclease B and agalacto-orosomucoid. *Biochem J* 1978, *176* (1), 103-109.

204. Flynn, G. C.; Chen, X.; Liu, Y. D.; Shah, B.; Zhang, Z., Naturally occurring glycan forms of human immunoglobulins G1 and G2. *Molecular Immunology* 2010, *47* (11), 2074-2082.

205. Lee, S. J.; Evers, S.; Roeder, D.; Parlow, A. F.; Risteli, J.; Risteli, L.; Lee, Y. C.; Feizi, T.; Langen, H.; Nussenzweig, M. C., Mannose Receptor-Mediated Regulation of Serum Glycoprotein Homeostasis. *Science* 2002, *295* (5561), 1898.

206. Malhotra, R.; Lu, J.; Holmskov, U.; Sim, R. B., Collectins, collectin receptors and the lectin pathway of complement activation. *Clin Exp Immunol* 1994, *97 Suppl 2* (Suppl 2), 4-9.

207. Jefferis, R., Glycosylation of Recombinant Antibody Therapeutics. *Biotechnology Progress* 2005, *21* (1), 11-16.

208. Mimura, Y.; Ashton, P. R.; Takahashi, N.; Harvey, D. J.; Jefferis, R., Contrasting glycosylation profiles between Fab and Fc of a human IgG protein studied by electrospray ionization mass spectrometry. *Journal of Immunological Methods* 2007, *326* (1), 116-126.

209. Bondt, A.; Rombouts, Y.; Selman, M. H. J.; Hensbergen, P. J.; Reiding, K. R.; Hazes, J. M. W.; Dolhain, R. J. E. M.; Wuhrer, M., Immunoglobulin G (IgG) Fab Glycosylation Analysis Using a New Mass Spectrometric High-throughput Profiling Method Reveals Pregnancy-associated Changes. *Molecular & Cellular Proteomics* 2014, *13* (11), 3029.

210. Anumula, K. R., Quantitative glycan profiling of normal human plasma derived immunoglobulin and its fragments Fab and Fc. *Journal of Immunological Methods* 2012, *382* (1), 167-176.

211. Endo, T.; Wright, A.; Morrison, S. L.; Kobata, A., Glycosylation of the variable region of immunoglobulin G—site specific maturation of the sugar chains. *Molecular Immunology* 1995, *32* (13), 931-940.

212. Chung, C. H.; Mirakhur, B.; Chan, E.; Le, Q.-T.; Berlin, J.; Morse, M.; Murphy, B. A.; Satinover, S. M.; Hosen, J.; Mauro, D.; Slebos, R. J.; Zhou, Q.; Gold, D.; Hatley, T.; Hicklin, D. J.; Platts-Mills, T. A. E., Cetuximab-induced anaphylaxis and IgE specific for galactose-alpha-1,3-galactose. *N Engl J Med* 2008, *358* (11), 1109-1117.

213. Schoch, A.; Kettenberger, H.; Mundigl, O.; Winter, G.; Engert, J.; Heinrich, J.; Emrich, T., Charge-mediated influence of the antibody variable domain on FcRn-dependent pharmacokinetics. *Proceedings of the National Academy of Sciences* 2015, *112* (19), 5997.

214. Wang, W.; Lu, P.; Fang, Y.; Hamuro, L.; Pittman, T.; Carr, B.; Hochman, J.; Prueksaritanont, T., Monoclonal Antibodies with Identical Fc Sequences Can Bind to FcRn Differentially with Pharmacokinetic Consequences. *Drug Metabolism and Disposition* 2011, *39* (9), 1469.

215. Zhang, L.; Mao, S., Application of quality by design in the current drug development. *Asian Journal of Pharmaceutical Sciences* 2017, *12* (1), 1-8.

216. ICH *ICH Q8 (R2) Pharmaceutical development*; 2014.

217. Yu, L. X., Pharmaceutical Quality by Design: Product and Process Development, Understanding, and Control. *Pharmaceutical Research* 2008, *25* (4), 781-791.

218. del Val, I. J.; Kontoravdi, C.; Nagy, J. M., Towards the implementation of quality by design to the production of therapeutic monoclonal antibodies with desired glycosylation patterns. *Biotechnology Progress* 2010, *26* (6), 1505-1527.

219. Luciani, F.; Galluzzo, S.; Gaggioli, A.; Kruse, N. A.; Venneugues, P.; Schneider, C. K.; Pini, C.; Melchiorri, D., Implementing quality by design for biotech products: Are regulators on track? *MAbs* 2015, *7* (3), 451-455.

220. Kepert, J. F.; Cromwell, M.; Engler, N.; Finkler, C.; Gellermann, G.; Gennaro, L.; Harris, R.; Iverson, R.; Kelley, B.; Krummen, L.; McKnight, N.; Motchnik, P.; Schnaible, V.; Taticek, R., Establishing a control system using QbD principles. *Biologicals* 2016, *44* (5), 319-331.

221. Hakemeyer, C.; McKnight, N.; St. John, R.; Meier, S.; Trexler-Schmidt, M.; Kelley, B.; Zettl, F.; Puskeiler, R.; Kleinjans, A.; Lim, F.; Wurth, C., Process characterization and Design Space definition. *Biologicals* 2016, *44* (5), 306-318.

222. Lourenço, V.; Lochmann, D.; Reich, G.; Menezes, J. C.; Herdling, T.; Schewitz, J., A quality by design study applied to an industrial pharmaceutical fluid bed granulation. *European Journal of Pharmaceutics and Biopharmaceutics* 2012, *81* (2), 438-447.

223. Djuris, J.; Djuric, Z., Modeling in the quality by design environment: Regulatory requirements and recommendations for design space and control strategy appointment. *International Journal of Pharmaceutics* 2017, *533* (2), 346-356.

224. Tomba, E.; Facco, P.; Bezzo, F.; Barolo, M., Latent variable modeling to assist the implementation of Quality-by-Design paradigms in pharmaceutical development and manufacturing: A review. *International Journal of Pharmaceutics* 2013, *457* (1), 283-297.

225. Mandenius, C.-F.; Graumann, K.; Schultz, T. W.; Premstaller, A.; Olsson, I.-M.; Petiot, E.; Clemens, C.; Welin, M., Quality-by-Design for biotechnology-related pharmaceuticals. *Biotechnology journal* 2009, *4* (5), 600-609.

226. Kelley, B.; Cromwell, M.; Jerkins, J., Integration of QbD risk assessment tools and overall risk management. *Biologicals* 2016, *44* (5), 341-351.

227. Alt, N.; Zhang, T. Y.; Motchnik, P.; Taticek, R.; Quarmby, V.; Schlothauer, T.; Beck, H.; Emrich, T.; Harris, R. J., Determination of critical quality attributes for monoclonal antibodies using quality by design principles. *Biologicals* 2016, *44* (5), 291-305.

228. Schiestl, M.; Stangler, T.; Torella, C.; Čepeljnik, T.; Toll, H.; Grau, R., Acceptable changes in quality attributes of glycosylated biopharmaceuticals. *Nature Biotechnology* 2011, *29* (4), 310-312.

229. Kim, S.; Song, J.; Park, S.; Ham, S.; Paek, K.; Kang, M.; Chae, Y.; Seo, H.; Kim, H.-C.; Flores, M., Drifts in ADCC-related quality attributes of Herceptin®: Impact on development of a trastuzumab biosimilar. *MAbs* 2017, *9* (4), 704-714.

230. Schiestl, M.; Li, J.; Abas, A.; Vallin, A.; Millband, J.; Gao, K.; Joung, J.; Pluschkell, S.; Go, T.; Kang, H.-N., The role of the quality assessment in the determination of overall biosimilarity: A simulated case study exercise. *Biologicals* 2014, *42* (2), 128-132.

231. Kirchhoff, C. F.; Wang, X.-Z. M.; Conlon, H. D.; Anderson, S.; Ryan, A. M.; Bose, A., Biosimilars: Key regulatory considerations and similarity assessment tools. *Biotechnology and bioengineering* 2017, *114* (12), 2696-2705.

232. Hajba, L.; Szekrényes, Á.; Borza, B.; Guttman, A., On the glycosylation aspects of biosimilarity. *Drug Discovery Today* 2018, *23* (3), 616-625.

233. Szekrenyes, A.; Szigeti, M.; Dvorakova, V.; Jarvas, G.; Guttman, A., Quantitative comparison of the N-glycosylation of therapeutic glycoproteins using the Glycosimilarity Index. A tutorial. *TrAC Trends in Analytical Chemistry* 2020, *122*, 115728.

234. St. Amand, M. M.; Hayes, J.; Radhakrishnan, D.; Fernandez, J.; Meyer, B.; Robinson, A. S.; Ogunnaike, B. A., Identifying a robust design space for glycosylation during monoclonal antibody production. *Biotechnology Progress* 2016, *32* (5), 1149-1162.

235. Karst, D. J.; Scibona, E.; Serra, E.; Bielser, J.-M.; Souquet, J.; Stettler, M.; Broly, H.; Soos, M.; Morbidelli, M.; Villiger, T. K., Modulation and modeling of monoclonal antibody N-linked glycosylation in mammalian cell perfusion reactors. *Biotechnology and bioengineering* 2017, *114* (9), 1978-1990.

236. Loebrich, S.; Clark, E.; Ladd, K.; Takahashi, S.; Brousseau, A.; Kitchener, S.; Herbst, R.; Ryll, T., Comprehensive manipulation of glycosylation profiles across development scales. *MAbs* 2019, *11* (2), 335-349.

237. Mayrhofer, P.; Reinhart, D.; Castan, A.; Kunert, R., Rapid development of clone-specific, high-performing perfusion media from established feed supplements. *Biotechnology Progress* 2020, *36* (2), e2933.

238. Abt, V.; Barz, T.; Cruz-Bournazou, M. N.; Herwig, C.; Kroll, P.; Möller, J.; Pörtner, R.; Schenkendorf, R., Model-based tools for optimal experiments in bioprocess engineering. *Current Opinion in Chemical Engineering* 2018, *22*, 244-252.

239. Möller, J.; Kuchemüller, K. B.; Steinmetz, T.; Koopmann, K. S.; Pörtner, R., Model-assisted Design of Experiments as a concept for knowledge-based bioprocess development. *Bioprocess and Biosystems Engineering* 2019, *42* (5), 867-882.

240. Kappatou, C. D.; Ehsani, A.; Niedenführ, S.; Mhamdi, A.; Schuppert, A.; Mitsos, A., Quality-targeting dynamic optimization of monoclonal antibody production. *Computers & Chemical Engineering* 2020, *142*, 107004.

241. Wang, J.; Chen, X.; Nan, Y.; Zhou, J.; Xue, T., Narrow Operating Space Based on the Inversion of Latent Structures Model for Glycosylation Process. *IEEE Access* 2020, *8*, 190504-190515.

242. Hossler, P.; Khattak, S. F.; Li, Z. J., Optimal and consistent protein glycosylation in mammalian cell culture. *Glycobiology* 2009, *19* (9), 936-949.

243. Ozturk, S. S.; Riley, M. R.; Palsson, B. O., Effects of ammonia and lactate on hybridoma growth, metabolism, and antibody production. *Biotechnology and bioengineering* 1992, *39* (4), 418-431.

244. Ozturk, S. S.; Palsson, B. O., Chemical Decomposition of Glutamine in Cell Culture Media: Effect of Media Type, pH, and Serum Concentration. *Biotechnology Progress* 1990, *6* (2), 121-128.

245. Martinelle, K.; Häggström, L., Mechanisms of ammonia and ammonium ion toxicity in animal cells: transport across cell membranes. *Journal of biotechnology* 1993, *30* (3), 339-50.

246. Lao, M.-S.; Toth, D., Effects of Ammonium and Lactate on Growth and Metabolism of a Recombinant Chinese Hamster Ovary Cell Culture. *Biotechnology Progress* 1997, *13* (5), 688-691.

247. Yang, M.; Butler, M., Effects of ammonia on CHO cell growth, erythropoietin production, and glycosylation. *Biotechnology and bioengineering* 2000, *68* (4), 370-80.

248. Kurano, N.; Leist, C.; Messi, F.; Kurano, S.; Fiechter, A., Growth behavior of Chinese hamster ovary cells in a compact loop bioreactor. 2. Effects of medium components and waste products. *Journal of biotechnology* 1990, *15* (1), 113-128.

249. Doyle, C.; Butler, M., The effect of pH on the toxicity of ammonia to a murine hybridoma. *Journal of biotechnology* 1990, *15* (1), 91-100.

250. Lie, S.; Wang, T.; Forbes, B.; Proud, C. G.; Petersen, J., The ability to utilise ammonia as nitrogen source is cell type specific and intricately linked to GDH, AMPK and mTORC1. *Scientific Reports* 2019, *9* (1), 1461.

251. Andersen, D. C.; Goochee, C. F., The effect of ammonia on the O-linked glycosylation of granulocyte colony-stimulating factor produced by chinese hamster ovary cells. *Biotechnology and bioengineering* 1995, *47* (1), 96-105.

252. Thorens, B.; Vassalli, P., Chloroquine and ammonium chloride prevent terminal glycosylation of immunoglobulins in plasma cells without affecting secretion. *Nature* 1986, *321* (6070), 618-620.

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

254. Ryll, T.; Valley, U.; Wagner, R., Biochemistry of growth inhibition by ammonium ions in mammalian cells. *Biotechnology and bioengineering* 1994, *44* (2), 184-193.

255. Valley, U.; Nimtz, M.; Conradt, H. S.; Wagner, R., Incorporation of ammonium into intracellular UDP-activated N-acetylhexosamines and into carbohydrate structures in glycoproteins. *Biotechnology and bioengineering* 1999, *64* (4), 401-417.

256. Gawlitzek, M.; Papac, D. I.; Sliwkowski, M. B.; Ryll, T., Incorporation of 15 N from ammonium into the N-linked oligosaccharides of an immunoadhesin glycoprotein expressed in Chinese hamster ovary cells. *Glycobiology* 1999, *9* (2), 125-131.

257. Chen, P.; Harcum, S. W., Effects of elevated ammonium on glycosylation gene expression in CHO cells. *Metabolic Engineering* 2006, *8* (2), 123-132.

258. Brodsky, A. N.; Caldwell, M.; Bae, S.; Harcum, S. W., Glycosylation-related genes in NS0 cells are insensitive to moderately elevated ammonium concentrations. *Journal of biotechnology* 2014, *187*, 78-86.

259. Gawlitzek, M.; Ryll, T.; Lofgren, J.; Sliwkowski, M. B., Ammonium alters N-glycan structures of recombinant TNFR-IgG: degradative versus biosynthetic mechanisms. *Biotechnology and bioengineering* 2000, *68* (6), 637-46.

260. Yang, M.; Butler, M., Effects of Ammonia and Glucosamine on the Heterogeneity of Erythropoietin Glycoforms. *Biotechnology Progress* 2002, *18* (1), 129-138.

261. Villiger, T. K.; Steinhoff, R. F.; Ivarsson, M.; Solacroup, T.; Stettler, M.; Broly, H.; Krismer, J.; Pabst, M.; Zenobi, R.; Morbidelli, M.; Soos, M., High-throughput profiling of nucleotides and nucleotide sugars to evaluate their impact on antibody N-glycosylation. *Journal of biotechnology* 2016, *229*, 3-12.

262. Li, F.; Vijayasankaran, N.; Shen, A. Y.; Kiss, R.; Amanullah, A., Cell culture processes for monoclonal antibody production. *MAbs* 2010, *2* (5), 466-479.

263. Glacken, M. W.; Fleischaker, R. J.; Sinskey, A. J., Reduction of waste product excretion via nutrient control: Possible strategies for maximizing product and cell yields on serum in cultures of mammalian cells. *Biotechnology and bioengineering* 1986, *28* (9), 1376-89.

264. Hartley, F.; Walker, T.; Chung, V.; Morten, K., Mechanisms driving the lactate switch in Chinese hamster ovary cells. *Biotechnology and bioengineering* 2018, *115* (8), 1890-1903.

265. Mulukutla, B. C.; Gramer, M.; Hu, W.-S., On metabolic shift to lactate consumption in fed-batch culture of mammalian cells. *Metabolic Engineering* 2012, *14* (2), 138-149.

266. Luo, J.; Vijayasankaran, N.; Autsen, J.; Santuray, R.; Hudson, T.; Amanullah, A.; Li, F., Comparative metabolite analysis to understand lactate metabolism shift in Chinese hamster ovary cell culture process. *Biotechnology and bioengineering* 2012, *109* (1), 146-156.

267. Handlogten, M. W.; Lee-O'Brien, A.; Roy, G.; Levitskaya, S. V.; Venkat, R.; Singh, S.; Ahuja, S., Intracellular response to process optimization and impact on productivity and product aggregates for a high-titer CHO cell process. *Biotechnology and bioengineering* 2018, *115* (1), 126-138.

268. Altamirano, C.; Illanes, A.; Becerra, S.; Cairó, J. J.; Gòdia, F., Considerations on the lactate consumption by CHO cells in the presence of galactose. *Journal of biotechnology* 2006, *125* (4), 547-556.

269. Carinhas, N.; Duarte, T. M.; Barreiro, L. C.; Carrondo, M. J. T.; Alves, P. M.; Teixeira, A. P., Metabolic signatures of GS-CHO cell clones associated with butyrate treatment and culture phase transition. *Biotechnology and bioengineering* 2013, *110* (12), 3244-3257.

270. Zagari, F.; Jordan, M.; Stettler, M.; Broly, H.; Wurm, F. M., Lactate metabolism shift in CHO cell culture: the role of mitochondrial oxidative activity. *New Biotechnology* 2013, *30* (2), 238-245.

271. Templeton, N.; Smith, K. D.; McAtee-Pereira, A. G.; Dorai, H.; Betenbaugh, M. J.; Lang, S. E.; Young, J. D., Application of 13C flux analysis to identify high-productivity CHO metabolic phenotypes. *Metabolic Engineering* 2017, *43*, 218-225.

272. Le, H.; Kabbur, S.; Pollastrini, L.; Sun, Z.; Mills, K.; Johnson, K.; Karypis, G.; Hu, W.-S., Multivariate analysis of cell culture bioprocess data—Lactate consumption as process indicator. *Journal of biotechnology* 2012, *162* (2), 210-223.

273. Martínez, V. S.; Dietmair, S.; Quek, L.-E.; Hodson, M. P.; Gray, P.; Nielsen, L. K., Flux balance analysis of CHO cells before and after a metabolic switch from lactate production to consumption. *Biotechnology and bioengineering* 2013, *110* (2), 660-666.

274. Capiaumont, J.; Legrand, C.; Carbonell, D.; Dousset, B.; Belleville, F.; Nabet, P., Methods for reducing the ammonia in hybridoma cell cultures. *Journal of biotechnology* 1995, *39* (1), 49-58.

275. Lim, Y.; Wong, N. S. C.; Lee, Y. Y.; Ku, S. C. Y.; Wong, D. C. F.; Yap, M. G. S., Engineering mammalian cells in bioprocessing – current achievements and future perspectives. *Biotechnology and Applied Biochemistry* 2010, *55* (4), 175-189.

276. Chen, K.; Liu, Q.; Xie, L.; Sharp, P. A.; Wang, D. I. C., Engineering of a mammalian cell line for reduction of lactate formation and high monoclonal antibody production. *Biotechnology and bioengineering* 2001, *72* (1), 55-61.

277. Kim, S. H.; Lee, G. M., Down-regulation of lactate dehydrogenase-A by siRNAs for reduced lactic acid formation of Chinese hamster ovary cells producing thrombopoietin. *Applied Microbiology and Biotechnology* 2007, *74* (1), 152-159.

278. Noh, S. M.; Park, J. H.; Lim, M. S.; Kim, J. W.; Lee, G. M., Reduction of ammonia and lactate through the coupling of glutamine synthetase selection and downregulation of lactate dehydrogenase-A in CHO cells. *Applied Microbiology and Biotechnology* 2017, *101* (3), 1035-1045.

279. Jeon, M. K.; Yu, D. Y.; Lee, G. M., Combinatorial engineering of ldh-a and bcl-2 for reducing lactate production and improving cell growth in dihydrofolate reductase-deficient Chinese hamster ovary cells. *Applied Microbiology and Biotechnology* 2011, *92* (4), 779-790.

280. Dorai, H.; Kyung, Y. S.; Ellis, D.; Kinney, C.; Lin, C.; Jan, D.; Moore, G.; Betenbaugh, M. J., Expression of anti-apoptosis genes alters lactate metabolism of Chinese Hamster Ovary cells in culture. *Biotechnology and bioengineering* 2009, *103* (3), 592-608.

281. Zhou, M.; Crawford, Y.; Ng, D.; Tung, J.; Pynn, A. F. J.; Meier, A.; Yuk, I. H.; Vijayasankaran, N.; Leach, K.; Joly, J.; Snedecor, B.; Shen, A., Decreasing lactate level and increasing antibody production in Chinese Hamster Ovary cells (CHO) by reducing the expression of lactate dehydrogenase and pyruvate dehydrogenase kinases. *Journal of biotechnology* 2011, *153* (1), 27-34.

282. Gupta, S. K.; Srivastava, S. K.; Sharma, A.; Nalage, V. H. H.; Salvi, D.; Kushwaha, H.; Chitnis, N. B.; Shukla, P., Metabolic engineering of CHO cells for the development of a robust protein production platform. *PLOS ONE* 2017, *12* (8), e0181455.

283. Hu, W. S.; Dodge, T. C.; Frame, K. K.; Himes, V. B., Effect of glucose on the cultivation of mammalian cells. *Dev Biol Stand* 1987, *66*, 279-290.

284. Li, J.; Wong, C. L.; Vijayasankaran, N.; Hudson, T.; Amanullah, A., Feeding lactate for CHO cell culture processes: Impact on culture metabolism and performance. *Biotechnology and bioengineering* 2012, *109* (5), 1173-1186.

285. Altamirano, C.; Paredes, C.; Cairó, J. J.; Gòdia, F., Improvement of CHO Cell Culture Medium Formulation: Simultaneous Substitution of Glucose and Glutamine. *Biotechnology Progress* 2000, *16* (1), 69-75.

286. Omasa, T.; Higashiyama, K.-I.; Shioya, S.; Suga, K.-i., Effects of lactate concentration on hybridoma culture in lactate-controlled fed-batch operation. *Biotechnology and bioengineering* 1992, *39* (5), 556-564.

287. Freund, N. W.; Croughan, M. S., A Simple Method to Reduce both Lactic Acid and Ammonium Production in Industrial Animal Cell Culture. *Int J Mol Sci* 2018, *19* (2), 385.

288. Liu, B.; Spearman, M.; Doering, J.; Lattová, E.; Perreault, H.; Butler, M., The availability of glucose to CHO cells affects the intracellular lipid-linked oligosaccharide distribution, site occupancy and the N-glycosylation profile of a monoclonal antibody. *Journal of biotechnology* 2014, *170*, 17-27.

289. Fan, Y.; Jimenez Del Val, I.; Müller, C.; Lund, A. M.; Sen, J. W.; Rasmussen, S. K.; Kontoravdi, C.; Baycin-Hizal, D.; Betenbaugh, M. J.; Weilguny, D.; Andersen, M. R., A multi-pronged investigation into the effect of glucose starvation and culture duration on fed-batch CHO cell culture. *Biotechnology and bioengineering* 2015, *112* (10), 2172-2184.

290. Naik, H. M.; Majewska, N. I.; Betenbaugh, M. J., Impact of nucleotide sugar metabolism on protein N-glycosylation in Chinese Hamster Ovary (CHO) cell culture. *Current Opinion in Chemical Engineering* 2018, *22*, 167-176.

291. Sha, S.; Agarabi, C.; Brorson, K.; Lee, D.-Y.; Yoon, S., N-Glycosylation Design and Control of Therapeutic Monoclonal Antibodies. *Trends in Biotechnology* 2016, *34* (10), 835-846.

292. Gramer, M. J.; Eckblad, J. J.; Donahue, R.; Brown, J.; Shultz, C.; Vickerman, K.; Priem, P.; van den Bremer, E. T. J.; Gerritsen, J.; van Berkel, P. H. C., Modulation of antibody galactosylation through feeding of uridine, manganese chloride, and galactose. *Biotechnology and bioengineering* 2011, *108* (7), 1591-1602.

293. Grainger, R. K.; James, D. C., CHO cell line specific prediction and control of recombinant monoclonal antibody N-glycosylation. *Biotechnology and bioengineering* 2013, *110* (11), 2970-2983.

294. St. Amand, M. M.; Tran, K.; Radhakrishnan, D.; Robinson, A. S.; Ogunnaike, B. A., Controllability Analysis of Protein Glycosylation in Cho Cells. *PLOS ONE* 2014, *9* (2), e87973.

295. Rouiller, Y.; Périlleux, A.; Vesin, M.-N.; Stettler, M.; Jordan, M.; Broly, H., Modulation of mAb quality attributes using microliter scale fed-batch cultures. *Biotechnology Progress* 2014, *30* (3), 571-583.

296. Crowell, C. K.; Grampp, G. E.; Rogers, G. N.; Miller, J.; Scheinman, R. I., Amino acid and manganese supplementation modulates the glycosylation state of erythropoietin in a CHO culture system. *Biotechnology and bioengineering* 2007, *96* (3), 538-549.

297. Wong, N. S. C.; Wati, L.; Nissom, P. M.; Feng, H. T.; Lee, M. M.; Yap, M. G. S., An investigation of intracellular glycosylation activities in CHO cells: Effects of nucleotide sugar precursor feeding. *Biotechnology and bioengineering* 2010, *107* (2), 321-336.

298. Gawlitzek, M.; Valley, U.; Wagner, R., Ammonium ion and glucosamine dependent increases of oligosaccharide complexity in recombinant glycoproteins secreted from cultivated BHK-21 cells. *Biotechnology and bioengineering* 1998, *57* (5), 518-28.

299. Grammatikos, S. I.; Valley, U.; Nimtz, M.; Conradt, H. S.; Wagner, R., Intracellular UDP−N-Acetylhexosamine Pool Affects N-Glycan Complexity: A Mechanism of Ammonium Action on Protein Glycosylation. *Biotechnology Progress* 1998, *14* (3), 410-419.

300. Gu, X.; Wang, D. I. C., Improvement of interferon-γ sialylation in Chinese hamster ovary cell culture by feeding of N-acetylmannosamine. *Biotechnology and bioengineering* 1998, *58* (6), 642-648.

301. Zanghi, J. A.; Mendoza, T. P.; Schmelzer, A. E.; Knop, R. H.; Miller, W. M., Role of Nucleotide Sugar Pools in the Inhibition of NCAM Polysialylation by Ammonia. *Biotechnology Progress* 1998, *14* (6), 834-844.

302. Baker, K. N.; Rendall, M. H.; Hills, A. E.; Hoare, M.; Freedman, R. B.; James, D. C., Metabolic control of recombinant protein N-glycan processing in NS0 and CHO cells. *Biotechnology and bioengineering* 2001, *73* (3), 188-202.

303. Hills, A. E.; Patel, A.; Boyd, P.; James, D. C., Metabolic control of recombinant monoclonal antibody N-glycosylation in GS-NS0 cells. *Biotechnology and bioengineering* 2001, *75* (2), 239-251.

304. Clark, K. J. R.; Griffiths, J.; Bailey, K. M.; Harcum, S. W., Gene-expression profiles for five key glycosylation genes for galactose-fed CHO cells expressing recombinant IL-4/13 cytokine trap. *Biotechnology and bioengineering* 2005, *90* (5), 568-577.

305. Davidson, M. B.; Hunt, K.; Fernandez-Mejia, C., The hexosamine biosynthetic pathway and glucose-induced down regulation of glucose transport in L6 myotubes. *Biochimica et Biophysica Acta (BBA) - General Subjects* 1994, *1201* (1), 113-117.

306. Sha, S.; Yoon, S., An investigation of nucleotide sugar dynamics under the galactose supplementation in CHO cell culture. *Process Biochemistry* 2019, *81*, 165-174.

307. Cha, H.-M.; Lim, J.-H.; Lee, K.-S.; Kim, D.-I., Nucleotide sugar precursor feeding strategy to enhance sialylation of albumin-erythropoietin in CHO cell cultures. *Process Biochemistry* 2018, *66*, 197-204.

308. Huang, C., Jr.; Lin, H.; Yang, J., A robust method for increasing Fc glycan high mannose level of recombinant antibodies. *Biotechnology and bioengineering* 2015, *112* (6), 1200-1209.

309. Slade, P. G.; Caspary, R. G.; Nargund, S.; Huang, C., Jr., Mannose metabolism in recombinant CHO cells and its effect on IgG glycosylation. *Biotechnology and bioengineering* 2016, *113* (7), 1468-1480.

310. Berrios, J.; Altamirano, C.; Osses, N.; Gonzalez, R., Continuous CHO cell cultures with improved recombinant protein productivity by using mannose as carbon source: Metabolic analysis and scale-up simulation. *Chemical Engineering Science* 2011, *66* (11), 2431-2439.

311. Jedrzejewski, P. A platform for the optimisation of metabolic pathways for glycosylation to achieve a narrow and targeted glycoform distribution. Imperial Collge London, 2015.

312. Kildegaard, H. F.; Fan, Y.; Sen, J. W.; Larsen, B.; Andersen, M. R., Glycoprofiling effects of media additives on IgG produced by CHO cells in fed-batch bioreactors. *Biotechnology and bioengineering* 2016, *113* (2), 359-366.

313. Rodriguez, J.; Spearman, M.; Huzel, N.; Butler, M., Enhanced Production of Monomeric Interferon-β by CHO Cells through the Control of Culture Conditions. *Biotechnology Progress* 2005, *21* (1), 22-30.

314. Gawlitzek, M.; Estacio, M.; Fürch, T.; Kiss, R., Identification of cell culture conditions to control N-glycosylation site-occupancy of recombinant glycoproteins expressed in CHO cells. *Biotechnology and bioengineering* 2009, *103* (6), 1164-1175.

315. Yin, B.; Wang, Q.; Chung, C.-Y.; Bhattacharya, R.; Ren, X.; Tang, J.; Yarema, K. J.; Betenbaugh, M. J., A novel sugar analog enhances sialic acid production and biotherapeutic sialylation in CHO cells. *Biotechnology and bioengineering* 2017, *114* (8), 1899-1902.

316. Yin, B.; Wang, Q.; Chung, C.-Y.; Ren, X.; Bhattacharya, R.; Yarema, K. J.; Betenbaugh, M. J., Butyrated ManNAc analog improves protein expression in Chinese hamster ovary cells. *Biotechnology and bioengineering* 2018, *115* (6), 1531-1541.

317. Senger, R. S.; Karim, M. N., Optimization of fed-batch parameters and harvest time of CHO cell cultures for a glycosylated product with multiple mechanisms of inactivation. *Biotechnology and bioengineering* 2007, *98* (2), 378-390.

318. Walther, J.; Lu, J.; Hollenbach, M.; Yu, M.; Hwang, C.; McLarty, J.; Brower, K., Perfusion Cell Culture Decreases Process and Product Heterogeneity in a Head-to-Head Comparison With Fed-Batch. *Biotechnology journal* 2019, *14* (2), 1700733.

319. Butler, M., Optimisation of the Cellular Metabolism of Glycosylation for Recombinant Proteins Produced by Mammalian Cell Systems. *Cytotechnology* 2005, *50* (1), 57.

320. Axelsson, M. A. B.; Karlsson, N. G.; Steel, D. M.; Ouwendijk, J.; Nilsson, T.; Hansson, G. C., Neutralization of pH in the Golgi apparatus causes redistribution of glycosyltransferases and changes in the O-glycosylation of mucins. *Glycobiology* 2001, *11* (8), 633-644.

321. Borys, M. C.; Linzer, D. I. H.; Papoutsakis, E. T., Culture pH Affects Expression Rates and Glycosylation of Recombinant Mouse Placental Lactogen Proteins by Chinese Hamster Ovary (CHO) Cells. *Bio/Technology* 1993, *11* (6), 720-724.

322. Ivarsson, M.; Villiger, T. K.; Morbidelli, M.; Soos, M., Evaluating the impact of cell culture process parameters on monoclonal antibody N-glycosylation. *Journal of biotechnology* 2014, *188*, 88-96.

323. Müthing, J.; Kemminer, S. E.; Conradt, H. S.; Šagi, D.; Nimtz, M.; Kärst, U.; Peter-Katalinić, J., Effects of buffering conditions and culture pH on production rates and glycosylation of clinical phase I anti-melanoma mouse IgG3 monoclonal antibody R24. *Biotechnology and bioengineering* 2003, *83* (3), 321-334.

324. Yoon, S. K.; Choi, S. L.; Song, J. Y.; Lee, G. M., Effect of culture pH on erythropoietin production by Chinese hamster ovary cells grown in suspension at 32.5 and 37.0°C. *Biotechnology and bioengineering* 2005, *89* (3), 345-356.

325. Michl, J.; Park, K. C.; Swietach, P., Evidence-based guidelines for controlling pH in mammalian live-cell culture systems. *Communications Biology* 2019, *2* (1), 144.

326. Zhu, M. M.; Goyal, A.; Rank, D. L.; Gupta, S. K.; Boom, T. V.; Lee, S. S., Effects of Elevated pCO2 and Osmolality on Growth of CHO Cells and Production of Antibody-Fusion Protein B1: A Case Study. *Biotechnology Progress* 2005, *21* (1), 70-77.

327. Ozturk, S. S.; Palsson, B. O., Growth, Metabolic, and Antibody Production Kinetics of Hybridoma Cell Culture: 2. Effects of Serum Concentration, Dissolved Oxygen Concentration, and Medium pH in a Batch Reactor. *Biotechnology Progress* 1991, *7* (6), 481-494.

328. deZengotita, V. M.; Kimura, R.; Miller, W. M., Effects of CO2 and osmolality on hybridoma cells: growth, metabolism and monoclonal antibody production. *Cytotechnology* 1998, *28* (1), 213-227.

329. Kim, M. S.; Kim, N. S.; Sung, Y. H.; Lee, G. M., Biphasic culture strategy based on hyperosmotic pressure for improved humanized antibody production in Chinese hamster ovary cell culture. *In vitro cellular & developmental biology. Animal* 2002, *38* (6), 314-9.

330. Nasseri, S. S.; Ghaffari, N.; Braasch, K.; Jardon, M. A.; Butler, M.; Kennard, M.; Gopaluni, B.; Piret, J. M., Increased CHO cell fed-batch monoclonal antibody production using the autophagy inhibitor 3-MA or gradually increasing osmolality. *Biochemical Engineering Journal* 2014, *91*, 37-45.

331. Qin, J.; Wu, X.; Xia, Z.; Huang, Z.; Zhang, Y.; Wang, Y.; Fu, Q.; Zheng, C., The effect of hyperosmolality application time on production, quality, and biopotency of monoclonal antibodies produced in CHO cell fed-batch and perfusion cultures. *Applied Microbiology and Biotechnology* 2019, *103* (3), 1217-1229.

332. Konno, Y.; Kobayashi, Y.; Takahashi, K.; Takahashi, E.; Sakae, S.; Wakitani, M.; Yamano, K.; Suzawa, T.; Yano, K.; Ohta, T.; Koike, M.; Wakamatsu, K.; Hosoi, S., Fucose content of monoclonal

antibodies can be controlled by culture medium osmolality for high antibody-dependent cellular cytotoxicity. *Cytotechnology* 2012, *64* (3), 249-265.

333. Hippach, M. B.; Schwartz, I.; Pei, J.; Huynh, J.; Kawai, Y.; Zhu, M. M., Fluctuations in dissolved oxygen concentration during a CHO cell culture process affects monoclonal antibody productivity and the sulfhydryl-drug conjugation process. *Biotechnology Progress* 2018, *34* (6), 1427-1437.

334. Goey, C. H.; Tsang, J. M. H.; Bell, D.; Kontoravdi, C., Cascading effect in bioprocessing—The impact of mild hypothermia on CHO cell behavior and host cell protein composition. *Biotechnology and bioengineering* 2017, *114* (12), 2771-2781.

335. Sou, S. N.; Sellick, C.; Lee, K.; Mason, A.; Kyriakopoulos, S.; Polizzi, K. M.; Kontoravdi, C., How does mild hypothermia affect monoclonal antibody glycosylation? *Biotechnology and bioengineering* 2015, *112* (6), 1165-1176.

336. Trummer, E.; Fauland, K.; Seidinger, S.; Schriebl, K.; Lattenmayer, C.; Kunert, R.; Vorauer-Uhl, K.; Weik, R.; Borth, N.; Katinger, H.; Müller, D., Process parameter shifting: Part II. Biphasic cultivation—A tool for enhancing the volumetric productivity of batch processes using Epo-Fc expressing CHO cells. *Biotechnology and bioengineering* 2006, *94* (6), 1045-1052.

337. Borys, M. C.; Dalal, N. G.; Abu-Absi, N. R.; Khattak, S. F.; Jing, Y.; Xing, Z.; Li, Z. J., Effects of culture conditions on N-glycolylneuraminic acid (Neu5Gc) content of a recombinant fusion protein produced in CHO cells. *Biotechnology and bioengineering* 2010, *105* (6), 1048-1057.

338. Serrato, J. A.; Palomares, L. A.; Meneses-Acosta, A.; Ramírez, O. T., Heterogeneous conditions in dissolved oxygen affect N-glycosylation but not productivity of a monoclonal antibody in hybridoma cultures. *Biotechnology and bioengineering* 2004, *88* (2), 176-188.

339. Lewis, A. M.; Croughan, W. D.; Aranibar, N.; Lee, A. G.; Warrack, B.; Abu-Absi, N. R.; Patel, R.; Drew, B.; Borys, M. C.; Reily, M. D.; Li, Z. J., Understanding and Controlling Sialylation in a CHO Fc-Fusion Process. *PLOS ONE* 2016, *11* (6), e0157111.

340. Chotigeat, W.; Watanapokasin, Y.; Mahler, S.; Gray, P. P., Role of environmental conditions on the expression levels, glycoform pattern and levels of sialyltransferase for hFSH produced by recombinant CHO cells. *Cytotechnology* 1994, *15* (1), 217-221.

341. Kunkel, J. P.; Jan, D. C. H.; Butler, M.; Jamieson, J. C., Comparisons of the Glycosylation of a Monoclonal Antibody Produced under Nominally Identical Cell Culture Conditions in Two Different Bioreactors. *Biotechnology Progress* 2000, *16* (3), 462-470.

342. Kunkel, J. P.; Jan, D. C. H.; Jamieson, J. C.; Butler, M., Dissolved oxygen concentration in serum-free continuous culture affects N-linked glycosylation of a monoclonal antibody. *Journal of biotechnology* 1998, *62* (1), 55-71.

343. Restelli, V.; Wang, M.-D.; Huzel, N.; Ethier, M.; Perreault, H.; Butler, M., The effect of dissolved oxygen on the production and the glycosylation profile of recombinant human erythropoietin produced from CHO cells. *Biotechnology and bioengineering* 2006, *94* (3), 481-494.

344. Yamane-Ohnuki, N.; Kinoshita, S.; Inoue-Urakubo, M.; Kusunoki, M.; Iida, S.; Nakano, R.; Wakitani, M.; Niwa, R.; Sakurada, M.; Uchida, K.; Shitara, K.; Satoh, M., Establishment of FUT8 knockout Chinese hamster ovary cells: An ideal host cell line for producing completely defucosylated antibodies with enhanced antibody-dependent cellular cytotoxicity. *Biotechnology and bioengineering* 2004, *87* (5), 614-622.

345. Ripka, J.; Adamany, A.; Stanley, P., Two chinese hamster ovary glycosylation mutants affected in the conversion of GDP-mannose to GDP-fucose. *Archives of Biochemistry and Biophysics* 1986, *249* (2), 533-545.

346. Imai-Nishiya, H.; Mori, K.; Inoue, M.; Wakitani, M.; Iida, S.; Shitara, K.; Satoh, M., Double knockdown of α1,6-fucosyltransferase (FUT8) and GDP-mannose 4,6-dehydratase (GMD) in antibody-producing cells: a new strategy for generating fully non-fucosylated therapeutic antibodies with enhanced ADCC. *BMC Biotechnology* 2007, *7* (1), 84.

347. Louie, S.; Haley, B.; Marshall, B.; Heidersbach, A.; Yim, M.; Brozynski, M.; Tang, D.; Lam, C.; Petryniak, B.; Shaw, D.; Shim, J.; Miller, A.; Lowe, J. B.; Snedecor, B.; Misaghi, S., FX knockout CHO hosts can express desired ratios of fucosylated or afucosylated antibodies with high titers and comparable product quality. *Biotechnology and bioengineering* 2017, *114* (3), 632-644.

348. Zhang, P.; Haryadi, R.; Chan, K. F.; Teo, G.; Goh, J.; Pereira, N. A.; Feng, H.; Song, Z., Identification of functional elements of the GDP-fucose transporter SLC35C1 using a novel Chinese hamster ovary mutant. *Glycobiology* 2012, *22* (7), 897-911.

349. Yang, Z.; Wang, S.; Halim, A.; Schulz, M. A.; Frodin, M.; Rahman, S. H.; Vester-Christensen, M. B.; Behrens, C.; Kristensen, C.; Vakhrushev, S. Y.; Bennett, E. P.; Wandall, H. H.; Clausen, H., Engineered CHO cells for production of diverse, homogeneous glycoproteins. *Nature Biotechnology* 2015, *33* (8), 842-844.

350. Bydlinski, N.; Maresch, D.; Schmieder, V.; Klanert, G.; Strasser, R.; Borth, N., The contributions of individual galactosyltransferases to protein specific N-glycan processing in Chinese Hamster Ovary cells. *Journal of biotechnology* 2018, *282*, 101-110.

351. Amann, T.; Hansen, A. H.; Kol, S.; Lee, G. M.; Andersen, M. R.; Kildegaard, H. F., CRISPR/Cas9-Multiplexed Editing of Chinese Hamster Ovary B4Gal-T1, 2, 3, and 4 Tailors N-Glycan Profiles of Therapeutics and Secreted Host Cell Proteins. *Biotechnology journal* 2018, *13* (10), e1800111.

352. Schulz, M. A.; Tian, W.; Mao, Y.; Van Coillie, J.; Sun, L.; Larsen, J. S.; Chen, Y.-H.; Kristensen, C.; Vakhrushev, S. Y.; Clausen, H.; Yang, Z., Glycoengineering design options for IgG1 in CHO cells using precise gene editing. *Glycobiology* 2018, *28* (7), 542-549.

353. Voss, M.; Künzel, U.; Higel, F.; Kuhn, P.-H.; Colombo, A.; Fukumori, A.; Haug-Kröper, M.; Klier, B.; Grammer, G.; Seidl, A.; Schröder, B.; Obst, R.; Steiner, H.; Lichtenthaler, S. F.; Haass, C.; Fluhrer, R., Shedding of glycan-modifying enzymes by signal peptide peptidase-like 3 (SPPL3) regulates cellular N-glycosylation. *The EMBO Journal* 2014, *33* (24), 2890-2905.

354. Chung, C.-y.; Wang, Q.; Yang, S.; Ponce, S. A.; Kirsch, B. J.; Zhang, H.; Betenbaugh, M. J., Combinatorial genome and protein engineering yields monoclonal antibodies with hypergalactosylation from CHO cells. *Biotechnology and bioengineering* 2017, *114* (12), 2848-2856.

355. Deutscher, S. L.; Hirschberg, C. B., Mechanism of galactosylation in the Golgi apparatus. A Chinese hamster ovary cell mutant deficient in translocation of UDP-galactose across Golgi vesicle membranes. *Journal of Biological Chemistry* 1986, *261* (1), 96-100.

356. Maszczak-Seneczko, D.; Olczak, T.; Jakimowicz, P.; Olczak, M., Overexpression of UDP-GlcNAc transporter partially corrects galactosylation defect caused by UDP-Gal transporter mutation. *FEBS Letters* 2011, *585* (19), 3090-3094.

357. Varki, A., Sialic acids in human health and disease. *Trends Mol Med* 2008, *14* (8), 351-360.

358. Lin, C.-W.; Tsai, M.-H.; Li, S.-T.; Tsai, T.-I.; Chu, K.-C.; Liu, Y.-C.; Lai, M.-Y.; Wu, C.-Y.; Tseng, Y.-C.; Shivatare, S. S.; Wang, C.-H.; Chao, P.; Wang, S.-Y.; Shih, H.-W.; Zeng, Y.-F.; You, T.-H.; Liao, J.-Y.; Tu, Y.-C.; Lin, Y.-S.; Chuang, H.-Y.; Chen, C.-L.; Tsai, C.-S.; Huang, C.-C.; Lin, N.-H.; Ma, C.; Wu, C.-Y.; Wong, C.-H., A common glycan structure on immunoglobulin G for enhancement of effector functions. *Proceedings of the National Academy of Sciences* 2015, *112* (34), 10611.

359. Raymond, C.; Robotham, A.; Spearman, M.; Butler, M.; Kelly, J.; Durocher, Y., Production of α2,6-sialylated IgG1 in CHO cells. *MAbs* 2015, *7* (3), 571-583.

360. Rose, R. J.; van Berkel, P. H. C.; van den Bremer, E. T. J.; Labrijn, A. F.; Vink, T.; Schuurman, J.; Heck, A. J. R.; Parren, P. W. H. I., Mutation of Y407 in the CH3 domain dramatically alters glycosylation and structure of human IgG. *MAbs* 2013, *5* (2), 219-228.

361. Jassal, R.; Jenkins, N.; Charlwood, J.; Camilleri, P.; Jefferis, R.; Lund, J., Sialylation of Human IgG-Fc Carbohydrate by Transfected Rat α2,6-Sialyltransferase. *Biochemical and Biophysical Research Communications* 2001, *286* (2), 243-249.

362. Chung, C.-y.; Wang, Q.; Yang, S.; Yin, B.; Zhang, H.; Betenbaugh, M., Integrated Genome and Protein Editing Swaps α-2,6 Sialylation for α-2,3 Sialic Acid on Recombinant Antibodies from CHO. *Biotechnology journal* 2017, *12* (2), 1600502.

363. Chung, C.-Y.; Yin, B.; Wang, Q.; Chuang, K.-Y.; Chu, J. H.; Betenbaugh, M. J., Assessment of the coordinated role of ST3GAL3, ST3GAL4 and ST3GAL6 on the α2,3 sialylation linkage of mammalian glycoproteins. *Biochemical and Biophysical Research Communications* 2015, *463* (3), 211-215.

364. Eckhardt, M.; Gotza, B.; Gerardy-Schahn, R., Mutants of the CMP-sialic Acid Transporter Causing the Lec2 Phenotype. *Journal of Biological Chemistry* 1998, *273* (32), 20189-20195.

365. Lim, S. F.; Lee, M. M.; Zhang, P.; Song, Z., The Golgi CMP-sialic acid transporter: A new CHO mutant provides functional insights. *Glycobiology* 2008, *18* (11), 851-860.

366. Wong, N. S.; Yap, M. G.; Wang, D. I., Enhancing recombinant glycoprotein sialylation through CMP-sialic acid transporter over expression in Chinese hamster ovary cells. *Biotechnology and bioengineering* 2006, *93* (5), 1005-16.

367. Ha, T. K.; Hansen, A. H.; Kildegaard, H. F.; Lee, G. M., Knockout of sialidase and pro-apoptotic genes in Chinese hamster ovary cells enables the production of recombinant human erythropoietin in fed-batch cultures. *Metabolic Engineering* 2020, *57*, 182-192.

368. Kwak, C.-Y.; Park, S.-Y.; Lee, C.-G.; Okino, N.; Ito, M.; Kim, J. H., Enhancing the sialylation of recombinant EPO produced in CHO cells via the inhibition of glycosphingolipid biosynthesis. *Scientific Reports* 2017, *7* (1), 13059.

369. Son, Y.-D.; Jeong, Y. T.; Park, S.-Y.; Kim, J. H., Enhanced sialylation of recombinant human erythropoietin in Chinese hamster ovary cells by combinatorial engineering of selected genes. *Glycobiology* 2011, *21* (8), 1019-1028.

370. Jeong, Yeon T.; Choi, O.; Son, Young D.; Park, Seung Y.; Kim, Jung H., Enhanced sialylation of recombinant erythropoietin in genetically engineered Chinese-hamster ovary cells. *Biotechnology and Applied Biochemistry* 2009, *52* (4), 283-291.

371. Yin, B.; Gao, Y.; Chung, C.-y.; Yang, S.; Blake, E.; Stuczynski, M. C.; Tang, J.; Kildegaard, H. F.; Andersen, M. R.; Zhang, H.; Betenbaugh, M. J., Glycoengineering of Chinese hamster ovary cells for enhanced erythropoietin N-glycan branching and sialylation. *Biotechnology and bioengineering* 2015, *112* (11), 2343-2351.

372. Wang, Q.; Wang, T.; Yang, S.; Sha, S.; Wu, W. W.; Chen, Y.; Paul, J. T.; Shen, R.-F.; Cipollo, J. F.; Betenbaugh, M. J., Metabolic engineering challenges of extending N-glycan pathways in Chinese hamster ovary cells. *Metabolic Engineering* 2020, *61*, 301-314.

373. Galleguillos, S. N.; Ruckerbauer, D.; Gerstl, M. P.; Borth, N.; Hanscho, M.; Zanghellini, J., What can mathematical modelling say about CHO metabolism and protein glycosylation? *Computational and Structural Biotechnology Journal* 2017, *15*, 212-221.

374. Bordbar, A.; Monk, J. M.; King, Z. A.; Palsson, B. O., Constraint-based models predict metabolic and associated cellular functions. *Nature Reviews Genetics* 2014, *15* (2), 107-120.

375. Guo, W.; Feng, X., OM-FBA: Integrate Transcriptomics Data with Flux Balance Analysis to Decipher the Cell Metabolism. *PLOS ONE* 2016, *11* (4), e0154188.

376. Kim, M. K.; Lun, D. S., Methods for integration of transcriptomic data in genome-scale metabolic models. *Computational and Structural Biotechnology Journal* 2014, *11* (18), 59-65.

377. Selvarasu, S.; Ho, Y. S.; Chong, W. P.; Wong, N. S.; Yusufi, F. N.; Lee, Y. Y.; Yap, M. G.; Lee, D. Y., Combined in silico modeling and metabolomics analysis to characterize fed-batch CHO cell culture. *Biotechnology and bioengineering* 2012, *109* (6), 1415-29.

378. Volkova, S.; Matos, M. R. A.; Mattanovich, M.; Marín de Mas, I., Metabolic Modelling as a Framework for Metabolomics Data Integration and Analysis. *Metabolites* 2020, *10* (8).

379. Henry, C. S.; Broadbelt, L. J.; Hatzimanikatis, V., Thermodynamics-Based Metabolic Flux Analysis. *Biophysical Journal* 2007, *92* (5), 1792-1805.

380. Pandey, V.; Gardiol, D. H.; Chiappino-Pepe, A.; Hatzimanikatis, V., TEX-FBA: A constraint-based method for integrating gene expression, thermodynamics, and metabolomics data into genome-scale metabolic models. *bioRxiv* 2019, 536235.

381. Pandey, V.; Hadadi, N.; Hatzimanikatis, V., Enhanced flux prediction by integrating relative expression and relative metabolite abundance into thermodynamically consistent metabolic models. *PLOS Computational Biology* 2019, *15* (5), e1007036.

382. Mahadevan, R.; Schilling, C. H., The effects of alternate optimal solutions in constraint-based genome-scale metabolic models. *Metabolic Engineering* 2003, *5* (4), 264-276.

383. Lularevic, M.; Racher, A. J.; Jaques, C.; Kiparissides, A., Improving the accuracy of flux balance analysis through the implementation of carbon availability constraints for intracellular reactions. *Biotechnology and bioengineering* 2019, *116* (9), 2339-2352.

384. Yeo, H. C.; Hong, J.; Lakshmanan, M.; Lee, D.-Y., Enzyme capacity-based genome scale modelling of CHO cells. *Metabolic Engineering* 2020, *60*, 138-147.

385. Orth, J. D.; Thiele, I.; Palsson, B. Ø., What is flux balance analysis? *Nature Biotechnology* 2010, *28* (3), 245-248.

386. Fouladiha, H.; Marashi, S.-A.; Li, S.; Li, Z.; Masson, H. O.; Vaziri, B.; Lewis, N. E., Systematically gap-filling the genome-scale metabolic model of CHO cells. *Biotechnology Letters* 2020.

387. Burgard, A. P.; Maranas, C. D., Optimization-based framework for inferring and testing hypothesized metabolic objective functions. *Biotechnology and bioengineering* 2003, *82* (6), 670-677.

388. Knorr, A. L.; Jain, R.; Srivastava, R., Bayesian-based selection of metabolic objective functions. *Bioinformatics* 2007, *23* (3), 351-357.

389. Chen, Y.; McConnell, B. O.; Gayatri Dhara, V.; Mukesh Naik, H.; Li, C.-T.; Antoniewicz, M. R.; Betenbaugh, M. J., An unconventional uptake rate objective function approach enhances applicability of genome-scale models for mammalian cells. *npj Systems Biology and Applications* 2019, *5* (1), 25.

390. Edwards, J. S.; Ibarra, R. U.; Palsson, B. O., In silico predictions of Escherichia coli metabolic capabilities are consistent with experimental data. *Nature Biotechnology* 2001, *19* (2), 125-130.

391. Pan, X.; Dalm, C.; Wijffels, R. H.; Martens, D. E., Metabolic characterization of a CHO cell size increase phase in fed-batch cultures. *Applied Microbiology and Biotechnology* 2017, *101* (22), 8101-8113.

392. Gudmundsson, S.; Thiele, I., Computationally efficient flux variability analysis. *BMC Bioinformatics* 2010, *11* (1), 489.

393. Segrè, D.; Vitkup, D.; Church, G. M., Analysis of optimality in natural and perturbed metabolic networks. *Proceedings of the National Academy of Sciences* 2002, *99* (23), 15112.

394. Varma, A.; Palsson, B. O., Metabolic Flux Balancing: Basic Concepts, Scientific and Practical Use. *Bio/Technology* 1994, *12* (10), 994-998.

395. Wiechert, W., 13C Metabolic Flux Analysis. *Metabolic Engineering* 2001, *3* (3), 195-206.

396. Gopalakrishnan, S.; Maranas, C. D., 13C metabolic flux analysis at a genome-scale. *Metabolic Engineering* 2015, *32*, 12-22.

397. Wiechert, W.; Möllney, M.; Petersen, S.; de Graaf, A. A., A Universal Framework for 13C Metabolic Flux Analysis. *Metabolic Engineering* 2001, *3* (3), 265-283.

398. Bonarius, H. P. J.; Hatzimanikatis, V.; Meesters, K. P. H.; de Gooijer, C. D.; Schmid, G.; Tramper, J., Metabolic flux analysis of hybridoma cells in different culture media using mass balances. *Biotechnology and bioengineering* 1996, *50* (3), 299-318.

399. Martínez, V. S.; Buchsteiner, M.; Gray, P.; Nielsen, L. K.; Quek, L.-E., Dynamic metabolic flux analysis using B-splines to study the effects of temperature shift on CHO cell metabolism. *Metabolic Engineering Communications* 2015, *2*, 46-57.

400. Nargund, S.; Qiu, J.; Goudar, C. T., Elucidating the role of copper in CHO cell energy metabolism using 13C metabolic flux analysis. *Biotechnology Progress* 2015, *31* (5), 1179-1186.

401. Ahn, W. S.; Antoniewicz, M. R., Metabolic flux analysis of CHO cells at growth and non-growth phases using isotopic tracers and mass spectrometry. *Metabolic Engineering* 2011, *13* (5), 598-609.

402. Mahadevan, R.; Edwards, J. S.; Doyle, F. J., Dynamic Flux Balance Analysis of Diauxic Growth in Escherichia coli. *Biophysical Journal* 2002, *83* (3), 1331-1340.

403. Leighty, R. W.; Antoniewicz, M. R., Dynamic metabolic flux analysis (DMFA): A framework for determining fluxes at metabolic non-steady state. *Metabolic Engineering* 2011, *13* (6), 745-755.

404. Schuster, S.; Dandekar, T.; Fell, D. A., Detection of elementary flux modes in biochemical networks: a promising tool for pathway analysis and metabolic engineering. *Trends in Biotechnology* 1999, *17* (2), 53-60.

405. Zanghellini, J.; Ruckerbauer, D. E.; Hanscho, M.; Jungreuthmayer, C., Elementary flux modes in a nutshell: Properties, calculation and applications. *Biotechnology journal* 2013, *8* (9), 1009-1016.

406. Klamt, S.; Stelling, J., Combinatorial Complexity of Pathway Analysis in Metabolic Networks. *Molecular Biology Reports* 2002, *29* (1), 233-236.

407. Burgard, A. P.; Vaidyaraman, S.; Maranas, C. D., Minimal Reaction Sets for Escherichia coli Metabolism under Different Growth Requirements and Uptake Environments. *Biotechnology Progress* 2001, *17* (5), 791-797.

408. Gerstl, M. P.; Jungreuthmayer, C.; Zanghellini, J., tEFMA: computing thermodynamically feasible elementary flux modes in metabolic networks. *Bioinformatics* 2015, *31* (13), 2232-2234.

409. Sambamoorthy, G.; Raman, K., MinReact: a systematic approach for identifying minimal metabolic networks. *Bioinformatics* 2020, *36* (15), 4309-4315.

410. Llaneras, F.; Picó, J., Stoichiometric modelling of cell metabolism. *Journal of Bioscience and Bioengineering* 2008, *105* (1), 1-11.

411. Zomorrodi, A. R.; Suthers, P. F.; Ranganathan, S.; Maranas, C. D., Mathematical optimization applications in metabolic networks. *Metabolic Engineering* 2012, *14* (6), 672-686.

412. Dai, Z.; Locasale, J. W., Understanding metabolism with flux analysis: From theory to application. *Metabolic Engineering* 2017, *43*, 94-102.

413. Gagneur, J.; Klamt, S., Computation of elementary modes: a unifying framework and the new binary approach. *BMC Bioinformatics* 2004, *5* (1), 175.

414. Kyriakopoulos, S.; Ang, K. S.; Lakshmanan, M.; Huang, Z.; Yoon, S.; Gunawan, R.; Lee, D.-Y., Kinetic Modeling of Mammalian Cell Culture Bioprocessing: The Quest to Advance Biomanufacturing. *Biotechnology journal* 2018, *13* (3), 1700229.

415. Goudar, C. T.; Joeris, K.; Konstantinov, K. B.; Piret, J. M., Logistic Equations Effectively Model Mammalian Cell Batch and Fed-Batch Kinetics by Logically Constraining the Fit. *Biotechnology Progress* 2005, *21* (4), 1109-1118.

416. Ben Yahia, B.; Malphettes, L.; Heinzle, E., Macroscopic modeling of mammalian cell growth and metabolism. *Applied Microbiology and Biotechnology* 2015, *99* (17), 7009-7024.

417. Sidoli, F. R.; Asprey, S. P.; Mantalaris, A., A Coupled Single Cell-Population-Balance Model for Mammalian Cell Cultures. *Industrial & Engineering Chemistry Research* 2006, *45* (16), 5801-5811.

418. Alhuthali, S.; Fadda, S.; Goey, C. H.; Kontoravdi, C., Multi-stage population balance model to understand the dynamics of fed-batch CHO cell culture. In *Computer Aided Chemical Engineering*, Espuña, A.; Graells, M.; Puigjaner, L., Eds. Elsevier: 2017; Vol. 40, pp 2821-2826.

419. Kontoravdi, C.; Asprey, S. P.; Pistikopoulos, E. N.; Mantalaris, A., Development of a dynamic model of monoclonal antibody production and glycosylation for product quality monitoring. *Computers & Chemical Engineering* 2007, *31* (5), 392-400.

420. Rügen, M.; Bockmayr, A.; Legrand, J.; Cogne, G., Network reduction in metabolic pathway analysis: Elucidation of the key pathways involved in the photoautotrophic growth of the green alga Chlamydomonas reinhardtii. *Metabolic Engineering* 2012, *14* (4), 458-467.

421. Stanford, N. J.; Lubitz, T.; Smallbone, K.; Klipp, E.; Mendes, P.; Liebermeister, W., Systematic Construction of Kinetic Models from Genome-Scale Metabolic Networks. *PLOS ONE* 2013, *8* (11), e79195.

422. Xing, Z.; Li, Z.; Chow, V.; Lee, S. S., Identifying Inhibitory Threshold Values of Repressing Metabolites in CHO Cell Culture Using Multivariate Analysis Methods. *Biotechnology Progress* 2008, *24* (3), 675-683.

423. Nielsen, J.; Keasling, Jay D., Engineering Cellular Metabolism. *Cell* 2016, *164* (6), 1185-1197.

424. Chakrabarti, A.; Miskovic, L.; Soh, K. C.; Hatzimanikatis, V., Towards kinetic modeling of genome-scale metabolic networks without sacrificing stoichiometric, thermodynamic and physiological constraints. *Biotechnology journal* 2013, *8* (9), 1043-1057.

425. Smallbone, K.; Simeonidis, E.; Swainston, N.; Mendes, P., Towards a genome-scale kinetic model of cellular metabolism. *BMC Systems Biology* 2010, *4* (1), 6.

426. Khodayari, A.; Maranas, C. D., A genome-scale Escherichia coli kinetic metabolic model k-ecoli457 satisfying flux data for multiple mutant strains. *Nature Communications* 2016, *7* (1), 13806.

427. Srinivasan, S.; Cluett, W. R.; Mahadevan, R., Constructing kinetic models of metabolism at genome-scales: A review. *Biotechnology journal* 2015, *10* (9), 1345-1359.

428. Jang, J. D.; Barford, J. P., An unstructured kinetic model of macromolecular metabolism in batch and fed-batch cultures of hybridoma cells producing monoclonal antibody. *Biochemical Engineering Journal* 2000, *4* (2), 153-168.

429. Robitaille, J.; Chen, J.; Jolicoeur, M., A Single Dynamic Metabolic Model Can Describe mAb Producing CHO Cell Batch and Fed-Batch Cultures on Different Culture Media. *PLOS ONE* 2015, *10* (9), e0136815.

430. Machado, D.; Costa, R. S.; Ferreira, E. C.; Rocha, I.; Tidor, B., Exploring the gap between dynamic and constraint-based models of metabolism. *Metabolic Engineering* 2012, *14* (2), 112-119.

431. Strutz, J.; Martin, J.; Greene, J.; Broadbelt, L.; Tyo, K., Metabolic kinetic modeling provides insight into complex biological questions, but hurdles remain. *Current Opinion in Biotechnology* 2019, *59*, 24-30.

432. Kovárová-Kovar, K.; Egli, T., Growth Kinetics of Suspended Microbial Cells: From Single-Substrate-Controlled Growth to Mixed-Substrate Kinetics. *Microbiology and Molecular Biology Reviews* 1998, *62* (3), 646.

433. Gutenkunst, R. N.; Waterfall, J. J.; Casey, F. P.; Brown, K. S.; Myers, C. R.; Sethna, J. P., Universally Sloppy Parameter Sensitivities in Systems Biology Models. *PLOS Computational Biology* 2007, *3* (10), e189.

434. Kontoravdi, C.; Asprey, S. P.; Pistikopoulos, E. N.; Mantalaris, A., Application of Global Sensitivity Analysis to Determine Goals for Design of Experiments: An Example Study on Antibody-Producing Cell Cultures. *Biotechnology Progress* 2005, *21* (4), 1128-1135.

435. Kiparissides, A.; Koutinas, M.; Kontoravdi, C.; Mantalaris, A.; Pistikopoulos, E. N., 'Closing the loop' in biological systems modeling — From the in silico to the in vitro. *Automatica* 2011, *47* (6), 1147-1155.

436. Kiparissides, A.; Kucherenko, S. S.; Mantalaris, A.; Pistikopoulos, E. N., Global Sensitivity Analysis Challenges in Biological Systems Modeling. *Industrial & Engineering Chemistry Research* 2009, *48* (15), 7168-7180.

437. Heijnen, J. J.; Verheijen, P. J. T., Parameter identification of in vivo kinetic models: Limitations and challenges. *Biotechnology journal* 2013, *8* (7), 768-775.

438. Chowdhury, A.; Khodayari, A.; Maranas, C. D., Improving prediction fidelity of cellular metabolism with kinetic descriptions. *Current Opinion in Biotechnology* 2015, *36*, 57-64.

439. Zeng, H.; Yang, A., Bridging substrate intake kinetics and bacterial growth phenotypes with flux balance analysis incorporating proteome allocation. *Scientific Reports* 2020, *10* (1), 4283.

440. von Stosch, M.; Hamelink, J.-M.; Oliveira, R., Hybrid modeling as a QbD/PAT tool in process development: an industrial E. coli case study. *Bioprocess and Biosystems Engineering* 2016, *39* (5), 773-784.

441. Rana, P.; Berry, C.; Ghosh, P.; Fong, S. S., Recent advances on constraint-based models by integrating machine learning. *Current Opinion in Biotechnology* 2020, *64*, 85-91.

442. Zelezniak, A.; Vowinckel, J.; Capuano, F.; Messner, C. B.; Demichev, V.; Polowsky, N.; Mülleder, M.; Kamrad, S.; Klaus, B.; Keller, M. A.; Ralser, M., Machine Learning Predicts the Yeast Metabolome from the Quantitative Proteome of Kinase Knockouts. *Cell systems* 2018, *7* (3), 269-283.e6.

443. Angione, C.; Lió, P., Predictive analytics of environmental adaptability in multi-omic network models. *Scientific Reports* 2015, *5* (1), 15147.

444. von Stosch, M.; Rodrigues de Azevedo, C.; Luis, M.; Feyo de Azevedo, S.; Oliveira, R., A principal components method constrained by elementary flux modes: analysis of flux data sets. *BMC Bioinformatics* 2016, *17* (1), 200.

445. Folch-Fortuny, A.; Teusink, B.; Hoefsloot, H. C. J.; Smilde, A. K.; Ferrer, A., Dynamic elementary mode modelling of non-steady state flux data. *BMC Systems Biology* 2018, *12* (1), 71.

446. Folch-Fortuny, A.; Marques, R.; Isidro, I. A.; Oliveira, R.; Ferrer, A., Principal elementary mode analysis (PEMA). *Molecular BioSystems* 2016, *12* (3), 737-746.

447. Bhadra, S.; Blomberg, P.; Castillo, S.; Rousu, J., Principal metabolic flux mode analysis. *Bioinformatics* 2018, *34* (14), 2409-2417.

448. González-Martínez, J. M.; Folch-Fortuny, A.; Llaneras, F.; Tortajada, M.; Picó, J.; Ferrer, A., Metabolic flux understanding of Pichia pastoris grown on heterogenous culture media. *Chemometrics and Intelligent Laboratory Systems* 2014, *134*, 89-99.

449. Bordbar, A.; Yurkovich, J. T.; Paglia, G.; Rolfsson, O.; Sigurjónsson, Ó. E.; Palsson, B. O., Elucidating dynamic metabolic physiology through network integration of quantitative time-course metabolomics. *Scientific Reports* 2017, *7* (1), 46249.

450. Sridhara, V.; Meyer, A. G.; Rai, P.; Barrick, J. E.; Ravikumar, P.; Segrè, D.; Wilke, C. O., Predicting Growth Conditions from Internal Metabolic Fluxes in an In-Silico Model of E. coli. *PLOS ONE* 2014, *9* (12), e114608.

451. Wu, S. G.; Wang, Y.; Jiang, W.; Oyetunde, T.; Yao, R.; Zhang, X.; Shimizu, K.; Tang, Y. J.; Bao, F. S., Rapid Prediction of Bacterial Heterotrophic Fluxomics Using Machine Learning and Constraint Programming. *PLOS Computational Biology* 2016, *12* (4), e1004838.

452. Antonakoudis, A.; Barbosa, R.; Kotidis, P.; Kontoravdi, C., The era of big data: Genome-scale modelling meets machine learning. *Computational and Structural Biotechnology Journal* 2020, *18*, 3287-3300.

453. Petsagkourakis, P.; Sandoval, I. O.; Bradford, E.; Zhang, D.; del Rio-Chanona, E. A., Reinforcement learning for batch bioprocess optimization. *Computers & Chemical Engineering* 2020, *133*, 106649.

454. Bradford, E.; Imsland, L.; Zhang, D.; del Rio Chanona, E. A., Stochastic data-driven model predictive control using gaussian processes. *Computers & Chemical Engineering* 2020, *139*, 106844.

455. Katz, J.; Pappas, I.; Avraamidou, S.; Pistikopoulos, E. N., Integrating deep learning models and multiparametric programming. *Computers & Chemical Engineering* 2020, *136*, 106801.

456. Del Rio-Chanona, E. A.; Ahmed, N. R.; Wagner, J.; Lu, Y.; Zhang, D.; Jing, K., Comparison of physics-based and data-driven modelling techniques for dynamic optimisation of fed-batch bioprocesses. *Biotechnology and bioengineering* 2019, *116* (11), 2971-2982.

457. Zhang, D.; Del Rio-Chanona, E. A.; Petsagkourakis, P.; Wagner, J., Hybrid physics-based and data-driven modeling for bioprocess online simulation and optimization. *Biotechnology and bioengineering* 2019, *116* (11), 2919-2930.

458. del Rio-Chanona, E. A.; Wagner, J. L.; Ali, H.; Fiorelli, F.; Zhang, D.; Hellgardt, K., Deep learning-based surrogate modeling and optimization for microalgal biofuel production and photobioreactor design. *AIChE Journal* 2019, *65* (3), 915-923.

459. Gunther, J. C.; Baclaski, J.; Seborg, D. E.; Conner, J. S., Pattern matching in batch bioprocesses—Comparisons across multiple products and operating conditions. *Computers & Chemical Engineering* 2009, *33* (1), 88-96.

460. Sokolov, M.; Ritscher, J.; MacKinnon, N.; Bielser, J.-M.; Brühlmann, D.; Rothenhäusler, D.; Thanei, G.; Soos, M.; Stettler, M.; Souquet, J.; Broly, H.; Morbidelli, M.; Butté, A., Robust factor selection in early cell culture process development for the production of a biosimilar monoclonal antibody. *Biotechnology Progress* 2017, *33* (1), 181-191.

461. Sokolov, M.; Soos, M.; Neunstoecklin, B.; Morbidelli, M.; Butté, A.; Leardi, R.; Solacroup, T.; Stettler, M.; Broly, H., Fingerprint detection and process prediction by multivariate analysis of fed-batch monoclonal antibody cell culture data. *Biotechnology Progress* 2015, *31* (6), 1633-1644.

462. Thomassen, Y. E.; van Sprang, E. N. M.; van der Pol, L. A.; Bakker, W. A. M., Multivariate data analysis on historical IPV production data for better process understanding and future improvements. *Biotechnology and bioengineering* 2010, *107* (1), 96-104.

463. Mercier, S. M.; Diepenbroek, B.; Dalm, M. C. F.; Wijffels, R. H.; Streefland, M., Multivariate data analysis as a PAT tool for early bioprocess development data. *Journal of biotechnology* 2013, *167* (3), 262-270.

464. Kirdar, A. O.; Conner, J. S.; Baclaski, J.; Rathore, A. S., Application of Multivariate Analysis toward Biotech Processes: Case Study of a Cell-Culture Unit Operation. *Biotechnology Progress* 2007, *23* (1), 61-67.

465. Teixeira, A. P.; Portugal, C. A. M.; Carinhas, N.; Dias, J. M. L.; Crespo, J. P.; Alves, P. M.; Carrondo, M. J. T.; Oliveira, R., In situ 2D fluorometry and chemometric monitoring of mammalian cell cultures. *Biotechnology and bioengineering* 2009, *102* (4), 1098-1106.

466. Dors, M.; Simutis, R.; Lübbert, A., Advanced Supervision of Mammalian Cell Cultures Using Hybrid Process Models. *IFAC Proceedings Volumes* 1995, *28* (3), 72-77.

467. Narayanan, H.; Sokolov, M.; Butté, A.; Morbidelli, M., Decision Tree-PLS (DT-PLS) algorithm for the development of process: Specific local prediction models. *Biotechnology Progress* 2019, *35* (4), e2818.

468. Narayanan, H.; Sokolov, M.; Morbidelli, M.; Butté, A., A new generation of predictive models: The added value of hybrid models for manufacturing processes of therapeutic proteins. *Biotechnology and bioengineering* 2019, *116* (10), 2540-2549.

469. Noe, D. A.; Delenick, J. C., Quantitative analysis of membrane and secretory protein processing and intracellular transport. *Journal of Cell Science* 1989, *92* (3), 449.

470. Shelikoff, M.; Sinskey, A. J.; Stephanopoulos, G., A modeling framework for the study of protein glycosylation. *Biotechnology and bioengineering* 1996, *50* (1), 73-90.

471. Monica, T. J.; Andersen, D. C.; Goochee, C. F., A mathematical model of sialylation of N-linked oligosaccharides in the trans-Golgi network. *Glycobiology* 1997, *7* (4), 515-521.

472. Umaña, P.; Bailey, J. E., A mathematical model of N-linked glycoform biosynthesis. *Biotechnology and bioengineering* 1997, *55* (6), 890-908.

473. Krambeck, F. J.; Betenbaugh, M. J., A mathematical model of N-linked glycosylation. *Biotechnology and bioengineering* 2005, *92* (6), 711-728.

474. Krambeck, F. J.; Bennun, S. V.; Narang, S.; Choi, S.; Yarema, K. J.; Betenbaugh, M. J., A mathematical model to derive N-glycan structures and cellular enzyme activities from mass spectrometric data. *Glycobiology* 2009, *19* (11), 1163-1175.

475. Krambeck, F. J.; Bennun, S. V.; Andersen, M. R.; Betenbaugh, M. J., Model-based analysis of N-glycosylation in Chinese hamster ovary cells. *PLOS ONE* 2017, *12* (5), e0175376.

476. Hossler, P.; Mulukutla, B. C.; Hu, W.-S., Systems Analysis of N-Glycan Processing in Mammalian Cells. *PLOS ONE* 2007, *2* (8), e713.

477. Bibila, T. A.; Flickinger, M. C., Use of a structured kinetic model of antibody synthesis and secretion for optimization of antibody production systems: I. Steady-state analysis. *Biotechnology and bioengineering* 1992, *39* (3), 251-61.

478. 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. *Biotechnology Progress* 2011, *27* (6), 1730-1743.

479. Kaveh, O.; Hengameh, A.; Johannes, G.; Murray, M.-Y.; Raymond, L. L.; Jeno, S.; Hector, B. M., Novel Dynamic Model to Predict the Glycosylation Pattern of Monoclonal Antibodies from Extracellular Cell Culture Conditions. *IFAC Proceedings Volumes* 2013, *46* (31), 30-35.

480. Sou, S. N.; Jedrzejewski, P. M.; Lee, K.; Sellick, C.; Polizzi, K. M.; Kontoravdi, C., Model-based investigation of intracellular processes determining antibody Fc-glycosylation under mild hypothermia. *Biotechnology and bioengineering* 2017, *114* (7), 1570-1582.

481. Villiger, T. K.; Scibona, E.; Stettler, M.; Broly, H.; Morbidelli, M.; Soos, M., Controlling the time evolution of mAb N-linked glycosylation - Part II: Model-based predictions. *Biotechnology Progress* 2016, *32* (5), 1135-1148.

482. Kotidis, P.; Jedrzejewski, P.; Sou, S. N.; Sellick, C.; Polizzi, K.; del Val, I. J.; Kontoravdi, C., Model-based optimization of antibody galactosylation in CHO cell culture. *Biotechnology and bioengineering* 2019, *116* (7), 1612-1626.

483. Aghamohseni, H.; Spearman, M.; Ohadi, K.; Braasch, K.; Moo-Young, M.; Butler, M.; Budman, H. M., A semi-empirical glycosylation model of a camelid monoclonal antibody under hypothermia cell culture conditions. *Journal of Industrial Microbiology & Biotechnology* 2017, *44* (7), 1005-1020.

484. Sha, S.; Huang, Z.; Agarabi, C. D.; Lute, S. C.; Brorson, K. A.; Yoon, S., Prediction of N-linked Glycoform Profiles of Monoclonal Antibody with Extracellular Metabolites and Two-Step Intracellular Models. *Processes* 2019, *7* (4).

485. Zhang, L.; Wang, M.; Castan, A.; Stevenson, J.; Chatzissavidou, N.; Hjalmarsson, H.; Vilaplana, F.; Chotteau, V., Glycan Residues Balance Analysis - GReBA: A novel model for the N-linked glycosylation of IgG produced by CHO cells. *Metabolic Engineering* 2020, *57*, 118-128.

486. Jimenez del Val, I.; Fan, Y.; Weilguny, D., Dynamics of immature mAb glycoform secretion during CHO cell culture: An integrated modelling framework. *Biotechnology journal* 2016, *11* (5), 610-623.

487. McDonald, A. G.; Hayes, J. M.; Bezak, T.; Głuchowska, S. A.; Cosgrave, E. F. J.; Struwe, W. B.; Stroop, C. J. M.; Kok, H.; van de Laar, T.; Rudd, P. M.; Tipton, K. F.; Davey, G. P., Galactosyltransferase 4 is a major control point for glycan branching in N-linked glycosylation. *Journal of Cell Science* 2014, *127* (23), 5014.

488. Bennun, S. V.; Yarema, K. J.; Betenbaugh, M. J.; Krambeck, F. J., Integration of the Transcriptome and Glycome for Identification of Glycan Cell Signatures. *PLOS Computational Biology* 2013, *9* (1), e1002813.

489. Kremkow, B. G.; Lee, K. H., Glyco-Mapper: A Chinese hamster ovary (CHO) genome-specific glycosylation prediction tool. *Metabolic Engineering* 2018, *47*, 134-142.

490. Kawano, S.; Hashimoto, K.; Miyama, T.; Goto, S.; Kanehisa, M., Prediction of glycan structures from gene expression data based on glycosyltransferase reactions. *Bioinformatics* 2005, *21* (21), 3976-3982.

491. Suga, A.; Yamanishi, Y.; Hashimoto, K.; Goto, S.; Kanehisa, M., An improved scoring scheme for predicting glycan structures from gene expression data. *Genome informatics. International Conference on Genome Informatics* 2007, *18*, 237-46.

492. Arigoni-Affolter, I.; Scibona, E.; Lin, C.-W.; Brühlmann, D.; Souquet, J.; Broly, H.; Aebi, M., Mechanistic reconstruction of glycoprotein secretion through monitoring of intracellular N-glycan processing. *Science Advances* 2019, *5* (11), eaax8930.

493. 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. *Metabolic Engineering* 2017, *43*, 9-20.

494. Hutter, S.; Wolf, M.; Papili Gao, N.; Lepori, D.; Schweigler, T.; Morbidelli, M.; Gunawan, R., Glycosylation Flux Analysis of Immunoglobulin G in Chinese Hamster Ovary Perfusion Cell Culture. *Processes* 2018, *6* (10).

495. Losfeld, M.-E.; Scibona, E.; Lin, C.-W.; Villiger, T. K.; Gauss, R.; Morbidelli, M.; Aebi, M., Influence of protein/glycan interaction on site-specific glycan heterogeneity. *The FASEB Journal* 2017, *31* (10), 4623-4635.

496. Spahn, P. N.; Hansen, A. H.; Hansen, H. G.; Arnsdorf, J.; Kildegaard, H. F.; Lewis, N. E., A Markov chain model for N-linked protein glycosylation – towards a low-parameter tool for model-driven glycoengineering. *Metabolic Engineering* 2016, *33*, 52-66.

497. Spahn, P. N.; Hansen, A. H.; Kol, S.; Voldborg, B. G.; Lewis, N. E., Predictive glycoengineering of biosimilars using a Markov chain glycosylation model. *Biotechnology journal* 2017, *12* (2), 1600489.

498. Liang, C.; Chiang, A. W. T.; Hansen, A. H.; Arnsdorf, J.; Schoffelen, S.; Sorrentino, J. T.; Kellman, B. P.; Bao, B.; Voldborg, B. G.; Lewis, N. E., A Markov model of glycosylation elucidates isozyme specificity and glycosyltransferase interactions for glycoengineering. *Current Research in Biotechnology* 2020, *2*, 22-36.

499. Fisher, P.; Spencer, H.; Thomas-Oates, J.; Wood, A. J.; Ungar, D., Modeling Glycan Processing Reveals Golgi-Enzyme Homeostasis upon Trafficking Defects and Cellular Differentiation. *Cell Reports* 2019, *27* (4), 1231-1243.e6.

500. Fisher, P.; Thomas-Oates, J.; Wood, A. J.; Ungar, D., The N-Glycosylation Processing Potential of the Mammalian Golgi Apparatus. *Frontiers in Cell and Developmental Biology* 2019, *7* (157).

501. Zhang, L.; Wang, M.; Castan, A.; Hjalmarsson, H.; Chotteau, V., Probabilistic model by Bayesian network for the prediction of antibody glycosylation in perfusion and fed-batch cell cultures. *Biotechnology and bioengineering* 2021.

502. Sokolov, M.; Ritscher, J.; MacKinnon, N.; Souquet, J.; Broly, H.; Morbidelli, M.; Butté, A., Enhanced process understanding and multivariate prediction of the relationship between cell culture process and monoclonal antibody quality. *Biotechnology Progress* 2017, *33* (5), 1368-1380.

503. Zürcher, P.; Sokolov, M.; Brühlmann, D.; Ducommun, R.; Stettler, M.; Souquet, J.; Jordan, M.; Broly, H.; Morbidelli, M.; Butté, A., Cell culture process metabolomics together with multivariate data

analysis tools opens new routes for bioprocess development and glycosylation prediction. *Biotechnology Progress* 2020, *36* (5), e3012.

504. Sokolov, M.; Morbidelli, M.; Butté, A.; Souquet, J.; Broly, H., Sequential Multivariate Cell Culture Modeling at Multiple Scales Supports Systematic Shaping of a Monoclonal Antibody Toward a Quality Target. *Biotechnology journal* 2018, *13* (4), 1700461.

505. Brühlmann, D.; Sokolov, M.; Butté, A.; Sauer, M.; Hemberger, J.; Souquet, J.; Broly, H.; Jordan, M., Parallel experimental design and multivariate analysis provides efficient screening of cell culture media supplements to improve biosimilar product quality. *Biotechnology and bioengineering* 2017, *114* (7), 1448-1458.

506. Kontoravdi, C.; N. Pistikopoulos, E.; Mantalaris, A., *Systematic development of predictive mathematical models for animal cell cultures*. 2010; Vol. 34, p 1192-1198.

507. Nolan, R. P.; Lee, K., Dynamic model of CHO cell metabolism. *Metab Eng* 2011, *13* (1), 108-24.

508. Qasba, P. K.; Ramakrishnan, B.; Boeggeman, E., Structure and function of beta -1,4-galactosyltransferase. *Curr Drug Targets* 2008, *9* (4), 292-309.

509. Wieczorke, R.; Dlugai, S.; Krampe, S.; Boles, E., Characterisation of Mammalian GLUT Glucose Transporters in a Heterologous Yeast Expression System. *Cellular Physiology and Biochemistry* 2003, *13* (3), 123-134.

510. Zhao, F.-Q.; Keating, A. F., Functional properties and genomics of glucose transporters. *Curr Genomics* 2007, *8* (2), 113-128.

511. Kanehisa, M.; Goto, S., KEGG: kyoto encyclopedia of genes and genomes. *Nucleic acids research* 2000, *28* (1), 27-30.

512. Schomburg, I.; Chang, A.; Schomburg, D., BRENDA, enzyme data and metabolic information. *Nucleic acids research* 2002, *30* (1), 47-49.

513. del Val, I. J.; Polizzi, K. M.; Kontoravdi, C., A theoretical estimate for nucleotide sugar demand towards Chinese Hamster Ovary cellular glycosylation. *Scientific Reports* 2016, *6*, 28547.

514. Waldman, B. C.; Rudnick, G., UDP-GlcNAc transport across the Golgi membrane: electroneutral exchange for dianionic UMP. *Biochemistry* 1990, *29* (1), 44-52.

515. Presley, J. F.; Cole, N. B.; Schroer, T. A.; Hirschberg, K.; Zaal, K. J. M.; Lippincott-Schwartz, J., ER-to-Golgi transport visualized in living cells. *Nature* 1997, *389* (6646), 81-85.

516. Planinc, A.; Dejaegher, B.; Vander Heyden, Y.; Viaene, J.; Van Praet, S.; Rappez, F.; Van Antwerpen, P.; Delporte, C., Batch-to-batch N-glycosylation study of infliximab, trastuzumab and bevacizumab, and stability study of bevacizumab. *Eur J Hosp Pharm* 2017, *24* (5), 286-292.

517. Kotidis, P.; Demis, P.; Goey, C. H.; Correa, E.; McIntosh, C.; Trepekli, S.; Shah, N.; Klymenko, O. V.; Kontoravdi, C., Constrained global sensitivity analysis for bioprocess design space identification. *Computers & Chemical Engineering* 2019, *125*, 558-568.

518. Kotidis, P.; Kontoravdi, C., Strategic Framework for Parameterization of Cell Culture Models. *Processes* 2019, *7* (3), 174.

519. Kyriakopoulos, S.; Kontoravdi, C., A framework for the systematic design of fed-batch strategies in mammalian cell culture. *Biotechnology and bioengineering* 2014, *111* (12), 2466-2476.

520. Kucherenko, S.; Zaccheus, O. SobolGSA Software. https://www.imperial.ac.uk/process-systems-engineering/research/free-software/sobolgsa-software/ (accessed March 12, 2021).

521. Li, G.; Wang, S.-W.; Rabitz, H., Practical Approaches To Construct RS-HDMR Component Functions. *The Journal of Physical Chemistry A* 2002, *106* (37), 8721-8733.

522. Klymenko, O. V.; Kucherenko, S.; Shah, N., Constrained Global Sensitivity Analysis: Sobol' indices for problems in non-rectangular domains. In *Computer Aided Chemical Engineering*, Espuña, A.; Graells, M.; Puigjaner, L., Eds. Elsevier: 2017; Vol. 40, pp 151-156.

523. Kucherenko, S.; Klymenko, O. V.; Shah, N., Sobol' indices for problems defined in non-rectangular domains. *Reliability Engineering & System Safety* 2017, *167*, 218-231.

524. Li, J.-H.; Huang, W.; Lin, P.; Wu, B.; Fu, Z.-G.; Shen, H.-M.; Jing, L.; Liu, Z.-Y.; Zhou, Y.; Meng, Y.; Xu, B.-Q.; Chen, Z.-N.; Jiang, J.-L., N-linked glycosylation at Asn152 on CD147 affects protein folding and stability: promoting tumour metastasis in hepatocellular carcinoma. *Scientific Reports* 2016, *6* (1), 35210.

525. Houde, D.; Peng, Y.; Berkowitz, S. A.; Engen, J. R., Post-translational Modifications Differentially Affect IgG1 Conformation and Receptor Binding. *Molecular & Cellular Proteomics* 2010, *9* (8), 1716.

526. Khoder-Agha, F.; Sosicka, P.; Escriva Conde, M.; Hassinen, A.; Glumoff, T.; Olczak, M.; Kellokumpu, S., N-acetylglucosaminyltransferases and nucleotide sugar transporters form multi-enzyme–multi-transporter assemblies in golgi membranes in vivo. *Cellular and Molecular Life Sciences* 2019, *76* (9), 1821-1832.

527. Liang, C.; Chiang, A. W. T.; Hansen, A. H.; Arnsdorf, J.; Schoffelen, S.; Sorrentino, J. T.; Kellman, B. P.; Bao, B.; Voldborg, B. G.; Lewis, N. E., A Markov model of glycosylation elucidates isozyme specificity and glycosyltransferase interactions for glycoengineering. *Current Research in Biotechnology* 2020.

528. Zhang, L.; Luo, S.; Zhang, B., Glycan analysis of therapeutic glycoproteins. *MAbs* 2016, *8* (2), 205-215.

529. Everest-Dass, A. V.; Moh, E. S. X.; Ashwood, C.; Shathili, A. M. M.; Packer, N. H., Human disease glycomics: technology advances enabling protein glycosylation analysis – part 1. *Expert Review of Proteomics* 2018, *15* (2), 165-182.

530. Gaunitz, S.; Nagy, G.; Pohl, N. L. B.; Novotny, M. V., Recent Advances in the Analysis of Complex Glycoproteins. *Anal Chem* 2017, *89* (1), 389-413.

531. Medlock, G. L.; Papin, J. A., Guiding the Refinement of Biochemical Knowledgebases with Ensembles of Metabolic Networks and Machine Learning. *Cell systems* 2020, *10* (1), 109-119.e3.

532. García-Contreras, R.; Vos, P.; Westerhoff, H. V.; Boogerd, F. C., Why in vivo may not equal in vitro – new effectors revealed by measurement of enzymatic activities under the same in vivo-like assay conditions. *The FEBS Journal* 2012, *279* (22), 4145-4159.

533. Sumit, M.; Dolatshahi, S.; Chu, A.-H. A.; Cote, K.; Scarcelli, J. J.; Marshall, J. K.; Cornell, R. J.; Weiss, R.; Lauffenburger, D. A.; Mulukutla, B. C.; Figueroa, B., Dissecting N-Glycosylation Dynamics in Chinese Hamster Ovary Cells Fed-batch Cultures using Time Course Omics Analyses. *iScience* 2019, *12*, 102-120.

534. Lancashire, L. J.; Lemetre, C.; Ball, G. R., An introduction to artificial neural networks in bioinformatics—application to complex microarray and mass spectrometry datasets in cancer studies. *Briefings in Bioinformatics* 2009, *10* (3), 315-329.

535. Darsey, J. A.; Griffin, W. O.; Joginipelli, S.; Melapu, V. K., Architecture and Biological Applications of Artificial Neural Networks: A Tuberculosis Perspective. In *Artificial Neural Networks*, Cartwright, H., Ed. Springer New York: New York, NY, 2015; pp 269-283.

536. Shahid, N.; Rappon, T.; Berta, W., Applications of artificial neural networks in health care organizational decision-making: A scoping review. *PLOS ONE* 2019, *14* (2), e0212356.

537. Julenius, K.; Mølgaard, A.; Gupta, R.; Brunak, S., Prediction, conservation analysis, and structural characterization of mammalian mucin-type O-glycosylation sites. *Glycobiology* 2004, *15* (2), 153-164.

538. Senger, R. S.; Karim, M. N., Variable Site-Occupancy Classification of N-Linked Glycosylation Using Artificial Neural Networks. *Biotechnology Progress* 2005, *21* (6), 1653-1662.

539. Senger, R. S.; Karim, M. N., Prediction of N-linked glycan branching patterns using artificial neural networks. *Mathematical Biosciences* 2008, *211* (1), 89-104.

540. Senger, R. S.; Karim, M. N., Effect of Shear Stress on Intrinsic CHO Culture State and Glycosylation of Recombinant Tissue-Type Plasminogen Activator Protein. *Biotechnology Progress* 2003, *19* (4), 1199-1209.

541. Kotidis, P.; Kontoravdi, C., Harnessing the potential of artificial neural networks for predicting protein glycosylation. *Metabolic Engineering Communications* 2020, *10*, e00131.

542. McCulloch, W. S.; Pitts, W., A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics* 1943, *5* (4), 115-133.

543. del Rio-Chanona, E. A.; Manirafasha, E.; Zhang, D.; Yue, Q.; Jing, K., Dynamic modeling and optimization of cyanobacterial C-phycocyanin production process by artificial neural network. *Algal Research* 2016, *13*, 7-15.

544. Lairson, L. L.; Henrissat, B.; Davies, G. J.; Withers, S. G., Glycosyltransferases: Structures, Functions, and Mechanisms. *Annual Review of Biochemistry* 2008, *77* (1), 521-555.

545. Wang, Q.; Chung, C.-Y.; Chough, S.; Betenbaugh, M. J., Antibody glycoengineering strategies in mammalian cells. *Biotechnology and bioengineering* 2018, *115* (6), 1378-1393.

546. Tulsyan, A.; Garvin, C.; Ündey, C., Advances in industrial biopharmaceutical batch process monitoring: Machine-learning methods for small data problems. *Biotechnology and bioengineering* 2018, *115* (8), 1915-1924.

547. North, S. J.; Huang, H.-H.; Sundaram, S.; Jang-Lee, J.; Etienne, A. T.; Trollope, A.; Chalabi, S.; Dell, A.; Stanley, P.; Haslam, S. M., Glycomics Profiling of Chinese Hamster Ovary Cell Glycosylation Mutants Reveals N-Glycans of a Novel Size and Complexity. *Journal of Biological Chemistry* 2010, *285* (8), 5759-5775.

548. Bojar, D.; Powers, R. K.; Camacho, D. M.; Collins, J. J., Deep-Learning Resources for Studying Glycan-Mediated Host-Microbe Interactions. *Cell Host & Microbe* 2020.

549. Li, F.; Zhang, Y.; Purcell, A. W.; Webb, G. I.; Chou, K.-C.; Lithgow, T.; Li, C.; Song, J., Positive-unlabelled learning of glycosylation sites in the human proteome. *BMC Bioinformatics* 2019, *20* (1), 112.

550. del Val, I. J.; Kyriakopoulos, S.; Polizzi, K. M.; Kontoravdi, C., An optimized method for extraction and quantification of nucleotides and nucleotide sugars from mammalian cells. *Analytical Biochemistry* 2013, *443* (2), 172-180.

551. Ceroni, A.; Maass, K.; Geyer, H.; Geyer, R.; Dell, A.; Haslam, S. M., GlycoWorkbench: a tool for the computer-assisted annotation of mass spectra of glycans. *J Proteome Res* 2008, *7* (4), 1650-9.

552. King, Z. A.; Lu, J.; Dräger, A.; Miller, P.; Federowicz, S.; Lerman, J. A.; Ebrahim, A.; Palsson, B. O.; Lewis, N. E., BiGG Models: A platform for integrating, standardizing and sharing genome-scale models. *Nucleic Acids Research* 2016, *44* (D1), D515-D522.

553. Rowe, E.; Palsson, B. O.; King, Z. A., Escher-FBA: a web application for interactive flux balance analysis. *BMC Systems Biology* 2018, *12* (1), 84.

554. Benyair, R.; Ron, E.; Lederkremer, G. Z., Chapter Five - Protein Quality Control, Retention, and Degradation at the Endoplasmic Reticulum. In *International Review of Cell and Molecular Biology*, Jeon, K. W., Ed. Academic Press: 2011; Vol. 292, pp 197-280.

555. Alberts, B.; Johnson, A.; Lewis, J.; Raff, M.; Roberts, K.; Walter, P., *Molecular Biology of the Cell*. 4th ed.; Garland Science: New York, 2002.

556. Spang, A., Retrograde Traffic from the Golgi to the Endoplasmic Reticulum. *Cold Spring Harbor Perspectives in Biology* 2013, *5* (6).

557. Brandizzi, F.; Barlowe, C., Organization of the ER–Golgi interface for membrane traffic control. *Nature Reviews Molecular Cell Biology* 2013, *14* (6), 382-392.

558. Klumperman, J., Architecture of the mammalian Golgi. *Cold Spring Harbor perspectives in biology* 2011, *3* (7), a005181.

559. Klumperman, J.; Kuliawat, R.; Griffith, J. M.; Geuze, H. J.; Arvan, P., Mannose 6–Phosphate Receptors Are Sorted from Immature Secretory Granules via Adaptor Protein AP-1, Clathrin, and Syntaxin 6–positive Vesicles. *Journal of Cell Biology* 1998, *141* (2), 359-371.

560. Posakony, J. W.; England, J. M.; Attardi, G., Mitochondrial growth and division during the cell cycle in HeLa cells. *Journal of Cell Biology* 1977, *74* (2), 468-491.

561. Zhang, L.; Castan, A.; Stevenson, J.; Chatzissavidou, N.; Vilaplana, F.; Chotteau, V., Combined effects of glycosylation precursors and lactate on the glycoprofile of IgG produced by CHO cells. *Journal of biotechnology* 2019, *289*, 71-79.

562. Owen, O. E.; Kalhan, S. C.; Hanson, R. W., The Key Role of Anaplerosis and Cataplerosis for Citric Acid Cycle Function*. *Journal of Biological Chemistry* 2002, *277* (34), 30409-30412.

563. Jakkamsetti, V.; Marin-Valencia, I.; Ma, Q.; Pascual, J. M., Chapter 29 - Pyruvate dehydrogenase, pyruvate carboxylase, Krebs cycle, and mitochondrial transport disorders. In *Rosenberg's Molecular and Genetic Basis of Neurological and Psychiatric Disease (Sixth Edition)*, Rosenberg, R. N.; Pascual, J. M., Eds. Academic Press: 2020; pp 427-436.

564. Curtis, M.; Kenny, H. A.; Ashcroft, B.; Mukherjee, A.; Johnson, A.; Zhang, Y.; Helou, Y.; Batlle, R.; Liu, X.; Gutierrez, N.; Gao, X.; Yamada, S. D.; Lastra, R.; Montag, A.; Ahsan, N.; Locasale, J. W.; Salomon, A. R.; Nebreda, A. R.; Lengyel, E., Fibroblasts Mobilize Tumor Cell Glycogen to Promote Proliferation and Metastasis. *Cell Metabolism* 2019, *29* (1), 141-155.e9.

565. TeSlaa, T.; Bartman, C. R.; Jankowski, C. S. R.; Zhang, Z.; Xu, X.; Xing, X.; Wang, L.; Lu, W.; Hui, S.; Rabinowitz, J. D., The Source of Glycolytic Intermediates in Mammalian Tissues. *Cell Metabolism*.

566. McCorvie, T. J.; Gleason, T. J.; Fridovich-Keil, J. L.; Timson, D. J., Misfolding of galactose 1-phosphate uridylyltransferase can result in type I galactosemia. *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease* 2013, *1832* (8), 1279-1293.

567. Crews, C.; Wilkinson, K. D.; Wells, L.; Perkins, C.; Fridovich-Keil, J. L., Functional Consequence of Substitutions at Residue 171 in Human Galactose-1-phosphate Uridylyltransferase *. *Journal of Biological Chemistry* 2000, *275* (30), 22847-22853.

568. Milla, M. E.; Clairmont, C. A.; Hirschberg, C. B., Reconstitution into proteoliposomes and partial purification of the Golgi apparatus membrane UDP-galactose, UDP-xylose, and UDP-glucuronic acid transport activities. *The Journal of biological chemistry* 1992, *267* (1), 103-7.

569. Lu, G.; Crihfield, C. L.; Gattu, S.; Veltri, L. M.; Holland, L. A., Capillary Electrophoresis Separations of Glycans. *Chem Rev* 2018, *118* (17), 7867-7885.

570. Grinna, L. S.; Robbins, P. W., Glycoprotein biosynthesis. Rat liver microsomal glucosidases which process oligosaccharides. *Journal of Biological Chemistry* 1979, *254* (18), 8814-8818.

571. Burns, D. M.; Touster, O., Purification and characterization of glucosidase II, an endoplasmic reticulum hydrolase involved in glycoprotein biosynthesis. *Journal of Biological Chemistry* 1982, *257* (17), 9991-10000.

572. Sakono, M.; Seko, A.; Takeda, Y.; Hachisu, M.; Ito, Y., Biophysical properties of UDP-glucose:glycoprotein glucosyltransferase, a folding sensor enzyme in the ER, delineated by synthetic probes. *Biochemical and Biophysical Research Communications* 2012, *426* (4), 504-510.

573. Karaveg, K.; Moremen, K. W., Energetics of substrate binding and catalysis by class 1 (glycosylhydrolase family 47) alpha-mannosidases involved in N-glycan processing and endoplasmic reticulum quality control. *The Journal of biological chemistry* 2005, *280* (33), 29837-48.

574. Miyazaki, T.; Matsumoto, Y.; Matsuda, K.; Kurakata, Y.; Matsuo, I.; Ito, Y.; Nishikawa, A.; Tonozuka, T., Heterologous expression and characterization of processing α-glucosidase I from Aspergillus brasiliensis ATCC 9642. *Glycoconjugate Journal* 2011, *28* (8), 563-571.

Appendix

Appendix I - Figures & Tables

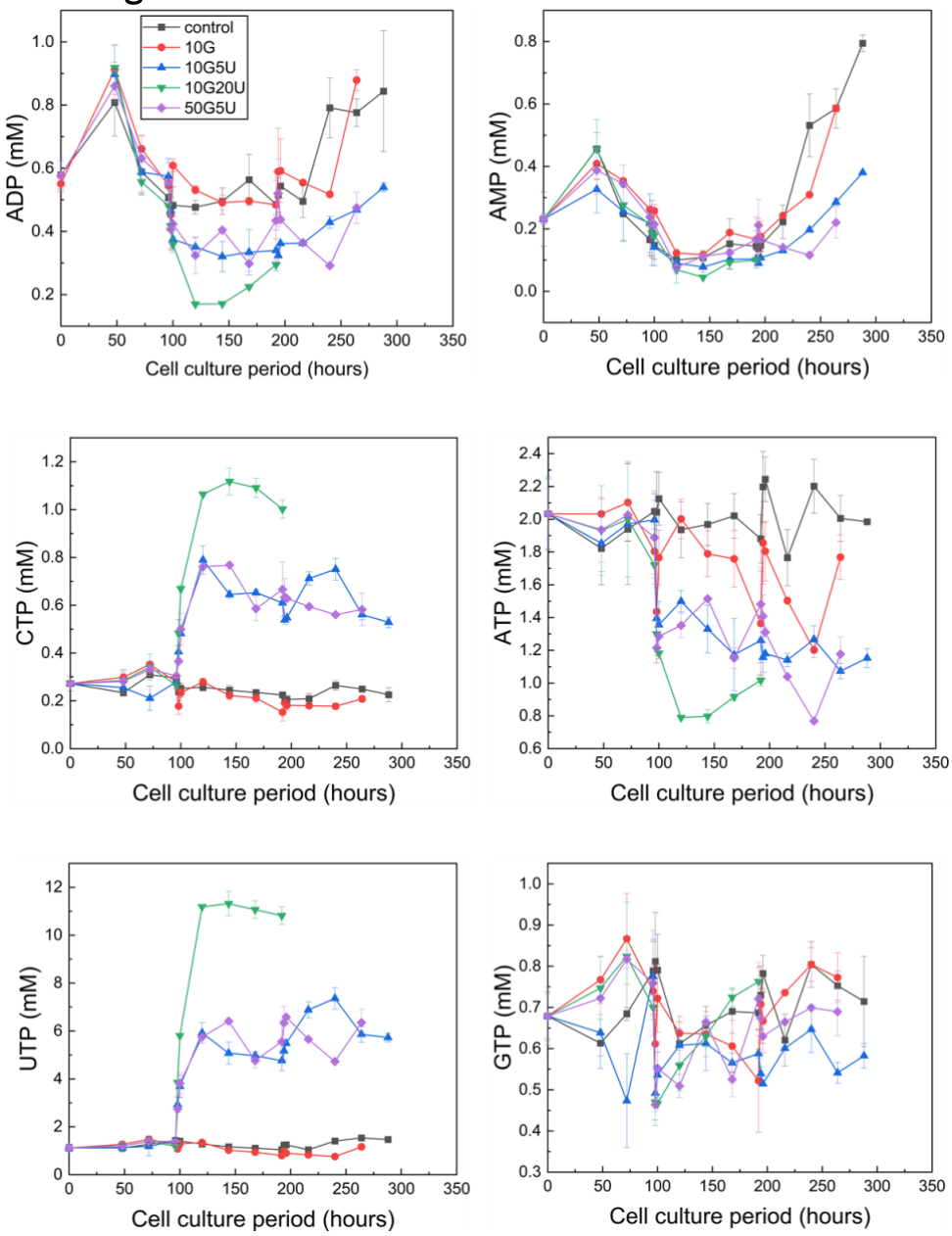

Figure A1: Nucleotides concentration for the experiments used in Chapter 2 of this thesis and throughout the cell culturing period. Adapted from Jedrzejewski[311].

Table A1: Estimated parameters for the *Metabolism* and *NSD submodels*. Taken with permission from Kotidis et al.[482].

Parameter	Value	Unit	95% Confidence Interval
		Metabolism	
$Y_{X_{Glc}}$	1.01×10^{9}	cell·mmol^{-1}	1.50×10^{8}
$Y_{X_{Lac}}$	5.46×10^{7}	cell·mmol^{-1}	7.60×10^{6}
$Y_{X_{Gln}}$	4.64×10^{9}	cell·mmol^{-1}	2.05×10^{8}
$Y_{X_{Glu}}$	1.46×10^{10}	cell·mmol^{-1}	3.29×10^{9}
$Y_{X_{Asn}}$	7.68×10^{8}	cell·mmol^{-1}	7.23×10^{6}
$Y_{X_{Amm}}$	2.36×10^{9}	cell·mmol^{-1}	2.15×10^{7}
$Y_{X_{Gal}}$	1.38×10^{8}	cell·mmol^{-1}	4.09×10^{6}
$Y_{X_{Urd}}$	1.61×10^{9}	cell·mmol^{-1}	2.39×10^{7}
$Y_{X_{Asp}}$	3.59×10^{9}	cell·mmol^{-1}	fixed[480]
$Y_{Gln/Amm}$	0.10	mmol·mmol^{-1}	2.24×10^{-2}
$Y_{Lac/Glc}$	1.56	mmol·mmol^{-1}	fixed[480]
$Y_{Asn/Asp}$	0.10	mmol·mmol^{-1}	0.03
$Y_{Asp/Asn}$	0.13	mmol·mmol^{-1}	fixed[480]
$Y_{Amm/Urd}$	2	mmol·mmol^{-1}	0.03
m_{Glc}	3.43×10^{-11}	mmol·cell^{-1}·h^{-1}	1.50×10^{-12}
m_{lac}	1.87×10^{-10}	mmol·cell^{-1}·h^{-1}	4.65×10^{-11}
Kc_{Gal}	5.27	mM	0.31
f_{Gal}	0.35	N/A	0.03
Lac_{max1}	21.20	mM	0.73
Lac_{max2}	16	mM	0.50
K_{Gal}	18.23	mM	0.74
K_{Urd}	7	mM	0.26
		NSD	
f_{Gln}	2.00×10^{-2}	N/A	3.30×10^{-4}
$K_{M1_{Gln}}$	0.42	mM	6.23×10^{-3}
$K_{M1_{sink}}$	4.00×10^{-2}	mM	1.98×10^{-3}
$KI_{1_{sink}}$	1.21×10^{-4}	mM	2.18×10^{-6}
$K_{M2_{Glc}}$	78.12	mM	6.719
$K_{M2b_{UDPGal}}$	2.48×10^{-2}	mM	3.22×10^{-3}
KI_{2A}	1.05×10^{-6}	mM	1.54×10^{-7}
KI_{2B}	92.11	mM	9.82×10^{7}
KI_{2C}	1.33×10^{-2}	mM	1.28
KI_{2D}	2.66×10^{-6}	mM	1.23×10^{-7}
$K_{M3_{Glc}}$	50	mM	4.05
$K_{M4_{UDPGlcNAc}}$	2.31	mM	0.30
$K_{M5_{UDPGlcNAc}}$	2.70×10^{-2}	mM	8.21×10^{-3}
KI_5	1000	mM	8.92×10^{6}
$K_{M6_{UDPGlc}}$	1.6×10^{-2}	mM	3.22×10^{-4}
KI_{6A}	1.10×10^{-7}	mM	6.73×10^{-9}
KI_{6B}	4.57	mM	2.81×10^{5}
KI_{6C}	4.84×10^{-6}	mM	5.08×10^{-7}
$K_{6_{sink}}$	0.13	mM	1.88×10^{-3}
$KI_{6_{sink}}$	10	mM	3.83×10^{-3}

$K_{regulator}$	1.00×10^{-5}	mM	fixed
$K_{M7_{GDPMan}}$	0.99	mM	0.10
KI_7	1.64×10^{-2}	mM	3.57×10^{-3}
$K_{M7_{sink}}$	8.88	mM	0.65
$V_{max,1_{Urd}}$	0.15	$mmol_{NSD}\cdot L_{cell}^{-1}\cdot h^{-1}$	6.02×10^{-3}
$V_{max,2_{Urd}}$	4.52×10^{-2}	$mmol_{NSD}\cdot L_{cell}^{-1}\cdot h^{-1}$	4.23×10^{-4}
$V_{max,4_{Urd}}$	1.27×10^{-2}	$mmol_{NSD}\cdot L_{cell}^{-1}\cdot h^{-1}$	2.70×10^{-3}
$V_{max,6_{Urd}}$	5.34	$mmol_{NSD}\cdot L_{cell}^{-1}\cdot h^{-1}$	1.10×10^{-2}
$V_{max7_{sink}}$	10.94	$mmol_{NSD}\cdot L_{cell}^{-1}\cdot h^{-1}$	0.81
$K_{M,1_{Urd}}$	6.08	mM	0.51
$K_{M,2_{Urd}}$	13.63	mM	0.69
$K_{M,4_{Urd}}$	6.25	mM	2.62
$K_{M,6_{Urd}}$	0.44	mM	1.91×10^{-3}
$V_{max1_{sink}}$	25.49	$mmol_{NSD}\cdot L_{cell}^{-1}\cdot h^{-1}$	0.46
$V_{max6_{sink}}$	7.30	$mmol_{NSD}\cdot L_{cell}^{-1}\cdot h^{-1}$	0.04
$V_{max6_{Gal}}$	40.90	$mmol_{NSD}\cdot L_{cell}^{-1}\cdot h^{-1}$	9.86
$K_{M6_{Gal}}$	0.60	mM	0.14
KI_{6D}	1.00×10^{-2}	mM	3.75×10^{-4}
KI_{6E}	99.63	mM	525.2
KI_{6F}	9.11×10^{-4}	mM	8.74×10^{-6}
$K_{TP_{UDPGlc}}$	0.90	mM	4.47×10^{-2}
$K_{TP_{UDPGlcNAc}}$	5.05	mM	2.31
$K_{TP_{UDPGal}}$	7.15	mM	0.33
$K_{TP_{UDPGalNAc}}$	11.06	mM	9.12
$K_{TP_{GDPMan}}$	0.13	mM	3.42×10^{-2}
$K_{TP_{GDPFuc}}$	0.10	mM	0.75
$K_{TP_{CMPNeu5Ac}}$	503.21	mM	824
V_{cell}	1.12×10^{-12}	L	fixed
V_{Golgi}	2.50×10^{-14}	L	fixed
MW_{mAb}	165.17×10^{3}	$g_{mAb}\cdot mol_{mAb}^{-1}$	fixed
$N_{HCP/Lipids_{UDPGlc}}$	1.56×10^{-12}	$mmol_{NSD}\cdot cell^{-1}$	fixed[513]
$N_{HCP/Lipids_{UDPGlcNAc}}$	1.25×10^{-12}	$mmol_{NSD}\cdot cell^{-1}$	fixed[513]
$N_{HCP/Lipids_{UDPGal}}$	2.29×10^{-12}	$mmol_{NSD}\cdot cell^{-1}$	fixed[513]
$N_{HCP/Lipids_{UDPGalNAc}}$	1.25×10^{-12}	$mmol_{NSD}\cdot cell^{-1}$	fixed[513]
$N_{HCP/Lipids_{GDPMan}}$	3.54×10^{-12}	$mmol_{NSD}\cdot cell^{-1}$	fixed[513]
$N_{HCP/Lipids_{GDPFuc}}$	1.40×10^{-13}	$mmol_{NSD}\cdot cell^{-1}$	fixed[513]
$N_{HCP/Lipids_{CMPNeu5Ac}}$	1.85×10^{-12}	$mmol_{NSD}\cdot cell^{-1}$	fixed[513]
$N_{precursor_{UDPGlc}}$	40.39×10^{-6}	$mmol_{NSD}\cdot mg_{mAb}^{-1}$	fixed[513]
$N_{precursor_{GDPMan}}$	121.2×10^{-6}	$mmol_{NSD}\cdot mg_{mAb}^{-1}$	fixed[513]
$N_{precursor_{UDPGlcNAc}}$	26.67×10^{-6}	$mmol_{NSD}\cdot mg_{mAb}^{-1}$	fixed[513]

Table A2: Reactions considered in the FBA. The FBA ant the reactions' table was kindly provided by Dr. Ioscani Jimenez del Val (University College Dublin). The shown coefficients of amino acids contribution towards biomass and IgG synthesis are derived from the control experiment (P5). The coefficients were adapted according to the experiment.

Flux	Rev	Reactions
		Glycolylsis
R_1	0	Glc + ATP --> G6P + ADP
R_2	-1	G6P <--> F6P
R_3	0	F6P + ATP --> DHAP + GAP + ADP
R_4	-1	DHAP <--> GAP
R_5	-1	GAP + NAD + ADP <--> 3PG + $NADH_{cyt}$ + ATP
R_6	0	3PG + ADP --> Pyr_{cyt} + ATP
		TCA cycle & oxidative phosporylation (mitochondrion)
R_7	-1	Pyr_{cyt} <--> Pyr_{mit}
R_8	0	Pyr_{mit} + NAD^+ + CoASH --> AcCoA + CO_2 + $NADH_{mit}$
R_9	0	AcCoA + $Oxal_{mit}$ --> Cit + CoASH
R_{10}	0	Cit + NAD^+ --> αKG_{mit} + CO_2 + $NADH_{mit}$
R_{11}	0	αKG_{mit} + CoASH + NAD --> SucCoA + CO2 + $NADH_{mit}$
R_{12}	-1	SucCoA + GDP <--> Succ + GTP + CoASH
R_{13}	-1	Succ + FAD^+ <--> Fum + $FADH_2$
R_{14}	-1	Fum <--> Mal_{mit}
R_{15}	-1	Mal_{mit} + NAD^+ <--> $Oxal_{mit}$ + $NADH_{mit}$
R_{16}	0	Pyr_{mit} + ATP --> $Oxal_{mit}$ + ADP
R_{17}	0	Mal_{mit} + NAD --> Pyr_{mit} + $NADH_{mit}$
R_{18}	0	$FADH_2$ --> $1.5ATP_{energy}$ + FAD
R_{19}	0	$NADH_{mit}$ --> $2.5ATP_{energy}$ + NAD^+
		Aspartate-Malate Shuttle & NAD^+ and NADPH synthesis
R_{20}	0	Glu_{cyt} + Asp_{mit} --> Glu_{mit} + Asp_{cyt}
R_{21}	-1	Mal_{cyt} + αKG_{mit} <--> Mal_{mit} + αKG_{cyt}
R_{22}	-1	$Oxal_{mit}$ + Glu_{mit} <--> αKG_{mit} + Asp_{mit}
R_{23}	-1	αKG_{mit} + $NADH_{mit}$ + NH_4^+ <--> Glu_{mit} + NAD^+
R_{24}	-1	$Oxal_{cyt}$ + $NADH_{cyt}$ <--> Mal_{cyt} + NAD^+
R_{25}	0	$NADH_{cyt}$ + ATP --> $NADPH_{cyt}$ + ADP
R_{26}	0	R5P + Trp + Gln + NADPH + 2ATP --> NAD^+ + Ala + $NADP^+$ + Glu_{cyt}
		Pyruvate fates
R_{27}	-1	Pyr_{cyt} + $NADH_{cyt}$ <--> Lac + NAD^+
R_{28}	-1	Pyr_{cyt} + Glu_{cyt} <--> Ala + αKG_{cyt}
		Pentose Phosphate pathway
R_{30}	0	3G6P + $6NADP^+$ --> $3CO_2$ + 3R5P + 6NADPH
		Anaplerotic reaction
R_{31}	-1	Mal_{cyt} + NAD^+ <--> Pyr_{cyt} + $NADH_{cyt}$
		Amino acid metabolism
R_{32}	-1	Glu_{cyt} + NAD^+ <--> αKG_{cyt} + NH_4^+ + $NADH_{cyt}$
R_{33}	-1	Asp_{cyt} + αKG_{cyt} <--> $Oxal_{cyt}$ + Glu_{cyt}
R_{34}	-1	Gln + ADP <--> Glu_{cyt} + ATP + NH_4^+
R_{35}	0	Thr + NAD^+ + CoASH --> Gly + $NADH_{cyt}$ + AcCoA

R_{36}	-1	Ser + THF + NAD^+ <--> Gly + $NADH_{cyt}$ + N10FTHF
R_{37}	-1	N10FTHF + ADP <--> ATP + Formate + THF
R_{38}	0	Ser --> Pyr_{cyt} + NH_4^+
R_{39}	0	Thr --> αKb + NH_4^+
R_{40}	0	αKb + CoASH + NAD^+ + HCO_3 + ATP --> SucCoA + ADP + $NADH_{cyt}$ + CO_2
R_{41}	0	Trp --> Ala + $2CO_2$ + αKa
R_{42}	0	Lys + $2\alpha KG_{cyt}$ + $3NAD^+$ + FAD^{2+} --> αKa + $2Glu_{cyt}$ + $3NADH_{cyt}$ + $FADH_2$
R_{43}	0	αKa + 2CoASH + $2NAD^+$ --> 2AcCoA + $2NADH_{cyt}$ + $2CO_2$
R_{44}	0	Val + αKG_{cyt} + CoASH + NAD^+ --> IsobutCoA + Glu_{cyt} + CO_2 + $NADH_{cyt}$
R_{45}	0	IsobutCoA + FAD^{2+} + $2NAD^+$ + HCO_3 + ATP --> SucCoA + ADP + $FADH_2$ + $2NADH_{cyt}$ + CO_2
R_{46}	0	IsobutCoA --> Isobut
R_{47}	0	Ile + αKG_{cyt} + 2CoASH + $2NAD^+$ + FAD^{2+} + HCO_3 + ATP --> AcCoA + SucCoA + ADP + Glu_{cyt} + CO_2 + $2NADH_{cyt}$ + $FADH_2$
R_{48}	0	Leu + αKG_{cyt} + CoASH + NAD^+ --> IsovalCoA + Glu_{cyt} + CO_2 + $NADH_{cyt}$
R_{49}	0	IsovalCoA + FAD^{2+} + ATP + CO_2 + SucCoA + CoASH --> 3AcCoA + Succ + $FADH_2$ + ADP
R_{50}	0	IsovalCoA --> Isoval
R_{51}	0	Phe + $NADH_{cyt}$ --> Tyr + NAD^+
R_{52}	0	Tyr + αKG_{cyt} + SucCoA + CoASH --> Fum + 2AcCoA + Succ + Glu_{cyt} + CO_2
R_{53}	0	Met + Ser + ATP --> αKb + NH_4^+ + AMP
R_{54}	-1	Asn <--> Asp_{cyt} + NH_4^+
R_{55}	-1	Pro + NAD^+ <--> Glu_{cyt} + $NADH_{cyt}$
R_{56}	0	2Arg + 3NADPH + $4O_2$ + $3H^+$ --> 2Citruline + 2NO + $3NADP^+$ + $4H_2O$
R_{57}	0	His --> Glu_{cyt} + NH_4^+
R_{58}	0	3PG + NAD^+ + Glu_{cyt} + H_2O --> Ser + αKG_{cyt} + $NADH_{cyt}$ + H^+ + Pi
R_{59}	-1	Gly + acetaldehyde <--> Thr
		Urea cycle
R_{60}	0	Arg + H_2O --> Ornithine + Urea
R_{61}	0	NH_4^+ + 2ATP + CO_2 + H_2O + Ornithine --> Citruline + 2Pi + 2ADP
R_{62}	0	Citruline + ATP + Asp_{mit} --> Fum + Arg + AMP + 2Pi
R_{63}	-1	Ornithine + αKG_{cyt} + $NADH_{cyt}$ <--> Glu_{cyt} + Pro + NAD^+
R_{64}	0	Ornithine --> Putrescine + CO_2
R_{65}	0	Putrescine --> Spermidine
		Glycogen synthesis
R_{66}	0	G6P --> G1P
R_{67}	0	G1P + UMPRN + 2ATP --> UDP-Glc + 2ADP
R_{68}	0	UDP-Glc --> Glycogen + UDP
		Nucleotide synthesis
R_{69}	0	R5P + ATP --> PRPP + AMP
R_{70}	0	PRPP + 2Gln + Gly + Asp_{cyt} + 5ATP + CO_2 + 2N10FTHF --> IMP + $2Glu_{cyt}$ + Fum + 5ADP + 2THF
R_{71}	0	IMP + Asp_{cyt} + GTP --> AMPRN + Fum + GDP
R_{72}	0	IMP + Gln + ATP + NAD^+ --> GMPRN + Glu_{cyt} + AMP + $NADH_{cyt}$
R_{73}	0	HCO_3 + NH_4^+ + Asp_{cyt} + 2ATP + NAD^+ --> Orotate + 2ADP + $NADH_{cyt}$
R_{74}	0	Orotate + PRPP --> UMPRN + CO_2
R_{75}	0	UMPRN + Gln + ATP --> CMPRN + Glu_{cyt} + ADP

R_{76}	0	AMPRN --> dAMP
R_{77}	0	GMPRN --> dGMP
R_{78}	0	CMPRN --> dCMP
R_{79}	0	UMPRN --> dTMP
		Lipid synthesis
R_{80}	0	Choline + ATP --> Pcholine + ADP
R_{81}	0	Pcholine + 18AcCoA + Glyc3P + 22ATP + 33$NADH_{cyt}$ --> PC + 16ADP + 6AMP + 33NAD^+
R_{82}	-1	PC + Ser --> PS + Choline
R_{83}	0	PS --> PE + CO_2
R_{84}	-1	Choline + Glyc3P --> Glyc3PC
R_{85}	0	G6P --> Inositol
R_{86}	0	Inositol + 18AcCoA + Glyc3P + 22ATP + 33$NADH_{cyt}$ --> PI + 16ADP + 6AMP + 33NAD^+
R_{87}	0	18AcCoA + 2Glyc3P + 22ATP + 33$NADH_{cyt}$ --> PG + 16ADP + 6AMP + 33NAD^+
R_{88}	0	2PG --> DPG + Glyc
R_{89}	0	16AcCoA + Ser + Choline + 16ATP + 29NADPH --> SM + 2CO_2 + 14ADP + 2AMP + 29$NADP^+$
R_{90}	0	18AcCoA + 18ATP + 14NADPH --> Cholesterol + 9CO_2 + 18ADP + 14$NADP^+$
		Biomass formation
R_{91}	0	110.599Ala + 80.3954Arg + 54.8137Asn + 82.7637Asp + 64.1741Gln + 116.2137Glu + 103.1927Gly + 30.9908His + 72.7055Ile + 127.6291Leu + 105.0358Lys + 36.044Met + 51.7991Phe + 72.8871Pro + 96.1535Ser + 76.0632Thr + 13.5816Trp + 41.1743Tyr + 95.4303Val + 3.2393dAMP + 2.1669dCMP + 2.1669dGMP + 3.2393dTMP + 7.2871AMPRN + 12.1673CMPRN + 13.7793GMPRN + 7.2871UMPRN + 61.101Glycogen + 3.942Cholesterol + 0.657DPG + 15.111PC + 5.694PE + 0.219PG + 2.19PI + 0.657PS + 1.752SM + 2.0712ATP + 0.4131CTP + 0.5219GTP + 2.0338UTP + 0.4709NAD + 0.0285NADP + 0.0022FAD + 0.011NADH + 0.0876NADPH + 7.665putrescine + 1.533Spermidine + 0.73762UDP-Gal + 1.79789GDP-Man + 0.513899UDP-GlcNAc + 0.833033CMP-Neu5Ac + 0CMP-Neu5Gc + 0.0543580GDP-Fuc + 0.6581091UDP-GalNAc --> 1Biomass + 11,694,463ADP
		Other by-products
R_{92}	-1	AcCoA + AMP --> Acetate + CoASH + ATP
R_{93}	-1	DHAP + $NADH_{cyt}$ --> Glyc3P + NAD^+
R_{94}	-1	Glyc3P + FAD --> DHAP + $FADH_2$
R_{95}	-1	Glyc3P --> Glyc
		IgG glycosylation
R_{96}	-1	UDP-Glc --> UDP-Gal
R_{97}	0	Glc + ATP + GTP --> GDP-Man + ADP
R_{98}	0	F6P + Gln + AcCoA + UTP --> UDP-GlcNAc + Glu_{cyt} + CoASH
R_{99}	0	UDP-GlcNAc + ATP + 3PG + CTP --> CMP-Neu5Ac + UDP + ADP
R_{100}	0	GDP-Man + NADPH --> GDP-Fuc + $NADP^+$
R_{101}	-1	UDP-GlcNAc <--> UDP-GalNAc
R_{102}	-1	CMP-Neu5Ac <--> CMP-NeuGc

		IgG formation
R_{103}	0	423.8Ala + 266.4Arg + 314.8Asn + 302.7Asp + 363.3Glu + 363.3Gln + 581.2Gly + 133.2His + 605.4Leu + 520.7Lys + 72.7Met + 266.4Phe + 532.8Pro + 1138.2Ser + 593.3Thr + 121.1Trp + 387.5Tyr + 690.2Val + 10.44GDP-Fuc + 53.01UDP-GlcNAc + 120GDP-Man + 7.97UDP-Gal + 0CMP-Neu5Ac --> 127.8566GDP + 54.8008UDP + 2.01845CMP + 1IgG
		Transport reactions
R_{104}	-1	--> Asp_{cyt}
R_{105}	-1	Gly -->
R_{106}	-1	--> Ser
R_{107}	-1	--> Glu_{cyt}
R_{108}	-1	--> Tyr
R_{109}	-1	Ala -->
R_{110}	-1	--> Arg
R_{111}	-1	--> Asn
R_{112}	-1	--> Gln
R_{113}	0	--> His
R_{114}	0	--> Ile
R_{115}	0	--> Leu
R_{116}	0	--> Lys
R_{117}	0	--> Met
R_{118}	0	--> Phe
R_{119}	-1	--> Pro
R_{120}	0	--> Thr
R_{121}	0	--> Trp
R_{122}	0	--> Val
R_{123}	-1	--> Choline
R_{124}	-1	NH_4^+ -->
R_{125}	0	CO_2 -->
R_{126}	-1	Cit -->
R_{127}	-1	Fum -->
R_{128}	-1	Pyr -->
R_{129}	-1	Succ -->
R_{130}	-1	Mal -->
R_{131}	-1	Glyc -->
R_{132}	-1	Pcholine -->
R_{133}	-1	Glyc3PC -->
R_{134}	-1	Formate -->
R_{135}	-1	Acetate -->
R_{136}	0	Isobut -->
R_{137}	0	Isoval -->
R_{138}	0	Biomass -->
R_{139}	0	--> Glc
R_{140}	-1	Lac -->
R_{141}	0	IgG -->
R_{142}	-1	Urea -->

Table A3: Kinetic constants for HCPs *N*-linked glycosylation. The (-) symbol indicates that the value was assigned to the designated parameter due to trace levels of the glycans generated from the respective reaction in which the parameter is involved. N/A indicates that no confidence interval was given by the parameter estimation.

Parameter	Value (µM)	95% CI or Reference
$K_{di,ManIA,HCPs}$	2,565	± 1,562
$K_{di,ManIB,HCPs}$	2,836	± 2,306
$K_{di,ManIC,HCPs}$	100	Krambeck et al.[474]
$K_{di,ManID,HCPs}$	1,882	± 1,052
$K_{di,ManIE,HCPs}$	284	± 14,285
$K_{di,ManIF,HCPs}$	1,276	Krambeck et al.[474]
$K_{di,ManIG,HCPs}$	100	Krambeck et al.[474]
$K_{di,ManIIA,HCPs}$	353	± 239
$K_{di,ManIIB,HCPs}$	219	± 148
$K_{di,a6FucTA,HCPs}$	7,000	-
$K_{di,a6FucTB,HCPs}$	3,943	Krambeck et al.[474]
$K_{di,a6FucTC,HCPs}$	1,996	Krambeck et al.[474]
$K_{di,a6FucTD,HCPs}$	30	Krambeck et al.[474]
$K_{di,GnTIA,HCPs}$	397	± 253
$K_{di,GnTIIA,HCPs}$	2,643	± 1,814
$K_{di,GnTIIIA,HCPs}$	688	N/A
$K_{di,GnTIVA,HCPs}$	1,879	N/A
$K_{di,GnTIVB,HCPs}$	17,000	Krambeck et al.[474]
$K_{di,GnTIVC,HCPs}$	5,100	Krambeck et al.[474]
$K_{di,GnTIVD,HCPs}$	5,100	Krambeck et al.[474]
$K_{di,GnTVA,HCPs}$	9,072	N/A
$K_{di,GnTVB,HCPs}$	10,000	Krambeck et al.[474]
$K_{di,iGnTA,HCPs}$	76	± 78
$K_{di,b4GalTA,HCPs}$	811	± 59
$K_{di,b4GalTB,HCPs}$	2,622	± 161
$K_{di,b4GalTC,HCPs}$	16,197	Krambeck et al.[474]
$K_{di,b4GalTD,HCPs}$	4,000	± 11,265
$K_{di,b4GalTE,HCPs}$	10,000	-
$K_{di,b4GalTF,HCPs}$	10,000	-
$K_{di,b4GalTG,HCPs}$	6,280	Del Val et al.[478]
$K_{di,b4GalTH,HCPs}$	6,280	Del Val et al.[478]
$K_{di,b4GalTK,HCPs}$	44	± 127
$K_{di,b4GalTL,HCPs}$	6,280	Del Val et al.[478]
$K_{di,b4GalTM,HCPs}$	6,280	Del Val et al.[478]
$K_{di,b4GalTN,HCPs}$	6,280	Del Val et al.[478]
$K_{di,b4GalTO,HCPs}$	6,280	Del Val et al.[478]
$K_{di,b4GalTP,HCPs}$	79	± 73
$K_{di,b4GalTR,HCPs}$	51	± 406
$K_{di,b4GalTS,HCPs}$	54	± 150
$K_{di,a3SiaTA,HCPs}$	2,711	± 209
$K_{di,a3SiaTB,HCPs}$	1,001	± 258
$K_{dk,ManIA,HCPs}$	0	Del Val et al.[478]
$K_{dk,ManIB,HCPs}$	0	Del Val et al.[478]
$K_{dk,ManIC,HCPs}$	0	Del Val et al.[478]
$K_{dk,ManID,HCPs}$	0	Del Val et al.[478]
$K_{dk,ManIE,HCPs}$	0	Del Val et al.[478]

$K_{dk,ManIF,HCPs}$	0	Del Val et al.[478]
$K_{dk,ManIG,HCPs}$	0	Del Val et al.[478]
$K_{dk,ManIIA,HCPs}$	0	Del Val et al.[478]
$K_{dk,ManIIB,HCPs}$	0	Del Val et al.[478]
$K_{dk,a6FucTA,HCPs}$	46	Del Val et al.[478]
$K_{dk,a6FucTB,HCPs}$	4,600	-
$K_{dk,a6FucTC,HCPs}$	46	Del Val et al.[478]
$K_{dk,a6FucTD,HCPs}$	46	Del Val et al.[478]
$K_{dk,GnTIA,HCPs}$	170	Del Val et al.[478]
$K_{dk,GnTIIA,HCPs}$	960	Del Val et al.[478]
$K_{dk,GnTIIIA,HCPs}$	3,100	Krambeck et al.[474]
$K_{dk,GnTIVA,HCPs}$	8,300	Krambeck et al.[474]
$K_{dk,GnTIVB,HCPs}$	8,300	Krambeck et al.[474]
$K_{dk,GnTIVC,HCPs}$	8,300	Krambeck et al.[474]
$K_{dk,GnTIVD,HCPs}$	8,300	Krambeck et al.[474]
$K_{dk,GnTVA,HCPs}$	8,300	Krambeck et al.[474]
$K_{dk,GnTVB,HCPs}$	3,500	Krambeck et al.[474]
$K_{dk,iGnTA,HCPs}$	3,500	Krambeck et al.[474]
$K_{dk,b4GalTA,HCPs}$	55	Krambeck et al.[474]
$K_{dk,b4GalTB,HCPs}$	65	Del Val et al.[478]
$K_{dk,b4GalTC,HCPs}$	65	Del Val et al.[478]
$K_{dk,b4GalTD,HCPs}$	65	Del Val et al.[478]
$K_{dk,b4GalTE,HCPs}$	65	Del Val et al.[478]
$K_{dk,b4GalTF,HCPs}$	65	Del Val et al.[478]
$K_{dk,b4GalTG,HCPs}$	65	Del Val et al.[478]
$K_{dk,b4GalTH,HCPs}$	65	Del Val et al.[478]
$K_{dk,b4GalTK,HCPs}$	65	Del Val et al.[478]
$K_{dk,b4GalTL,HCPs}$	65	Del Val et al.[478]
$K_{dk,b4GalTM,HCPs}$	65	Del Val et al.[478]
$K_{dk,b4GalTN,HCPs}$	65	Del Val et al.[478]
$K_{dk,b4GalTO,HCPs}$	65	Del Val et al.[478]
$K_{dk,b4GalTP,HCPs}$	65	Del Val et al.[478]
$K_{dk,b4GalTR,HCPs}$	65	Del Val et al.[478]
$K_{dk,b4GalTS,HCPs}$	65	Del Val et al.[478]
$K_{dk,a3SiaTA,HCPs}$	57	Krambeck et al.[474]
$K_{dk,a3SiaTB,HCPs}$	57	Krambeck et al.[474]

Table A4: Kinetic constants for IgG *N*-linked glycosylation. The high 95% CIs are a result of the parameter unidentifiability that characterizes this model.

Parameter	Value (µM)	95% CI or Reference
$\mathbf{K_{di,ManIA,IgG}}$	61	Del Val et al.[478]
$\mathbf{K_{di,ManIB,IgG}}$	110	Del Val et al.[478]
$\mathbf{K_{di,ManIC,IgG}}$	31	Del Val et al.[478]
$\mathbf{K_{di,ManID,IgG}}$	74	Del Val et al.[478]
$\mathbf{K_{di,ManIIA,IgG}}$	221	N/A
$\mathbf{K_{di,ManIIB,IgG}}$	100	Del Val et al.[478]
$\mathbf{K_{di,GnTIA,IgG}}$	74	N/A
$\mathbf{K_{di,GnTIIA,IgG}}$	104	N/A
$\mathbf{K_{di,GnTIIIA,IgG}}$	1×10^{12}	-

$K_{di,a6FucTA,IgG}$	25	Del Val et al.[478]
$K_{di,a6FucTB,IgG}$	1,432	± 279
$K_{di,b4GalTA,IgG}$	483	± 879
$K_{di,b4GalTB,IgG}$	6,280	Del Val et al.[478]
$K_{di,b4GalTC,IgG}$	703	± 630
$K_{di,b4GalTD,IgG}$	6,280	Del Val et al.[478]
$K_{di,a3SiaTA,IgG}$	1 x 10^{12}	Del Val et al.[478]
$K_{di,a3SiaTB,IgG}$	1 x 10^{12}	Del Val et al.[478]
$K_{di,a3SiaTC,IgG}$	1 x 10^{12}	Del Val et al.[478]
$K_{di,a3SiaTD,IgG}$	1 x 10^{12}	Del Val et al.[478]
$K_{dk,ManIA,IgG}$	0	Del Val et al.[478]
$K_{dk,ManIB,IgG}$	0	Del Val et al.[478]
$K_{dk,ManIC,IgG}$	0	Del Val et al.[478]
$K_{dk,ManID,IgG}$	0	Del Val et al.[478]
$K_{dk,ManIIA,IgG}$	0	Del Val et al.[478]
$K_{dk,ManIIB,IgG}$	0	Del Val et al.[478]
$K_{dk,GnTIA,IgG}$	170	Del Val et al.[478]
$K_{dk,GnTIIA,IgG}$	960	Del Val et al.[478]
$K_{dk,GnTIIIA,IgG}$	1 x 10^{12}	Del Val et al.[478]
$K_{dk,a6FucTA,IgG}$	46	Del Val et al.[478]
$K_{dk,a6FucTB,IgG}$	46	Del Val et al.[478]
$K_{dk,b4GalTA,IgG}$	3668	± 9177
$K_{dk,b4GalTB,IgG}$	65	Del Val et al.[478]
$K_{dk,b4GalTC,IgG}$	257	± 541
$K_{dk,b4GalTD,IgG}$	65	Del Val et al.[478]
$K_{dk,a3SiaTA,IgG}$	1 x 10^{12}	Del Val et al.[478]
$K_{dk,a3SiaTB,IgG}$	1 x 10^{12}	Del Val et al.[478]
$K_{dk,a3SiaTC,IgG}$	1 x 10^{12}	Del Val et al.[478]
$K_{dk,a3SiaTD,G}$	1 x 10^{12}	Del Val et al.[478]

Table A5: Parameters for NST concentration and distribution. 95% CIs are not available as the estimated parameters ($[NST]$ and $distr_{NST}{}^{UDPGal-transporter}(c)$) were calculated using the Optimization gPROMS entity.

Enzyme	$[NST]$ (µM)	$distr_{NST}(1)$	$distr_{NST}(2)$	$distr_{NST}(3)$	$distr_{NST}(4)$
UDPGal-transporter	7.8 x 10^{-9}	0	0.085	0.324	0.591
GDPFuc-transporter	2.4 x 10^{-7}	0.200	0.450	0.200	0.150
UDPGlcNac-transporter	4.2 x 10^{-8}	0.050	0.700	0.200	0.050
CMPNeu5Ac-transporter	8.9 x 10^{-8}	0	0.050	0.200	0.750

Table A6: Parameters for NSD transport kinetics. A 95% CI is not available for $K_m^{UDPGal-transporter}$ as the parameter was estimated using the Optimization gPROMS entity.

Enzyme	k_f^{NST} (min^{-1})	K_m^{NST} (µM)	Reference
UDPGal-transporter	948	7.2 (estimated)	Del Val et al.[478]
GDPFuc-transporter	130	7.5	Del Val et al.[478]
UDPGlcNac-transporter	1422	7.1	Del Val et al.[478]
CMPNeu5Ac-transporter	592	1.3	Del Val et al.[478]

Table A7: ER glycosylation model parameters. If a designation of 'HCPs' or 'mAb' is not mentioned, then the same value was used for both proteins' glycosylation. For the estimated parameters, a 95% CI is not provided due to the unidentifiability of the system. The resulting parameter values are by no means a realistic representation of the variables, as they are a single solution of the many that can satisfy the problem.

Parameter	Value	Units	Reference
$k_{f,\alpha GluI}$	24.2	min^{-1}	Grinna & Robbins[570]
$k_{f,\alpha GluI,A}$	432	min^{-1}	Burns & Touster[571]
$k_{f,\alpha GluI,B}$	432	min^{-1}	Burns & Touster[571]
$k_{f,UGGT}$	25	min^{-1}	Sakono et al. 2012[572]
$k_{f,ERManI}$	222	min^{-1}	Karaveg & Moremen[573]
$K_{di,\alpha GluI}$	6.1	µM	Miyazaki et al.[574]
$K_{di,\alpha GluII,A}$	850	µM	Burns & Touster[571]
$K_{di,\alpha GluII,B}$	850	µM	Burns & Touster[571]
$K_{di,UGGT}$	69	µM	Sakono et al. 2012[572]
$K_{dk,UGGT}$	4	µM	Sakono et al. 2012[572]
$K_{di,ERManI,HCPs}$	9988.5	µM	-
$K_{di,ERManI,mAb}$	1362.7	µM	-
$[\alpha GluI]$	0.32	µM	-
$[\alpha GluII]$	9.9	µM	-
$[UGGT]$	0.01	µM	-
$[ERManI]$	0.17	µM	-

Table A8: Glycosidases and glycosylatransferases concentration as estimated against the HCPs glycoprofile. 95% CIs are not available as all parameters were estimated using the Optimization gPROMS entity.

Enzyme	Concentration (µM)
ManI	1.11
ManII	2.10
GnTI	0.19
GnTII	3.80
GnTIII	2.35×10^{-3}
GnTIV	3.45
GnTV	1.70
iGnT	1.89
a6FucT	2.31
b4GalT	0.29
a3SiaT	0.77
a6SiaT	0
a3FucT	0
a3GalT	0

Appendix II - Model reduction

Model reduction was initially performed manually, prior to its integration in the GLYMMER software through the "*Lumping*" function. The manual model reduction was performed for a previously published dataset of CHO HCPs[547]. The utilized HCPs glyosylation profile presents 21 different glycans[513], with the majority of them being high mannose (≥ 5 molecules of mannose in the oligosaccharide) glycans that collectively account for the 48.8% of the total distribution. The remaining 51.2% is mainly distributed between bi-, tri- and tetra-antennary glycans and glycans that incorporate poly-LacNAc branches. The experimentally measured glycosylation profile of CHO-HCPs is presented in Fig. A2.

The glycosylation network from Krambeck et al. 2017 ("Large" network) was generated with the GLYMMER software and included 18,276 structures (oligosaccharide conformations – glycans) and 52,600 reactions. The enzymes included in the reactions network are listed in Table A9. The "Large" network was then reduced by defining the "essential glycans", meaning the glycans that are experimentally observed in the glycosylation profile of interest. The distribution of the essential glycans is not used for network reduction (thus making the reduction based on solely qualitatively data). The network was reduced both manually and automatically ("*Lumping"*) with the methods as described below. For highly branched glycans, the essential glycans list included only a small number of the possible isoforms, in an effort to maximise the reduction of the network's complexity. The same values for the kinetic constants were used for all three networks, apart from the enzyme concentrations that were estimated for each case and are presented in Fig. A3. The GLYMMER software was used for the estimation of the enzyme concentrations for all networks.

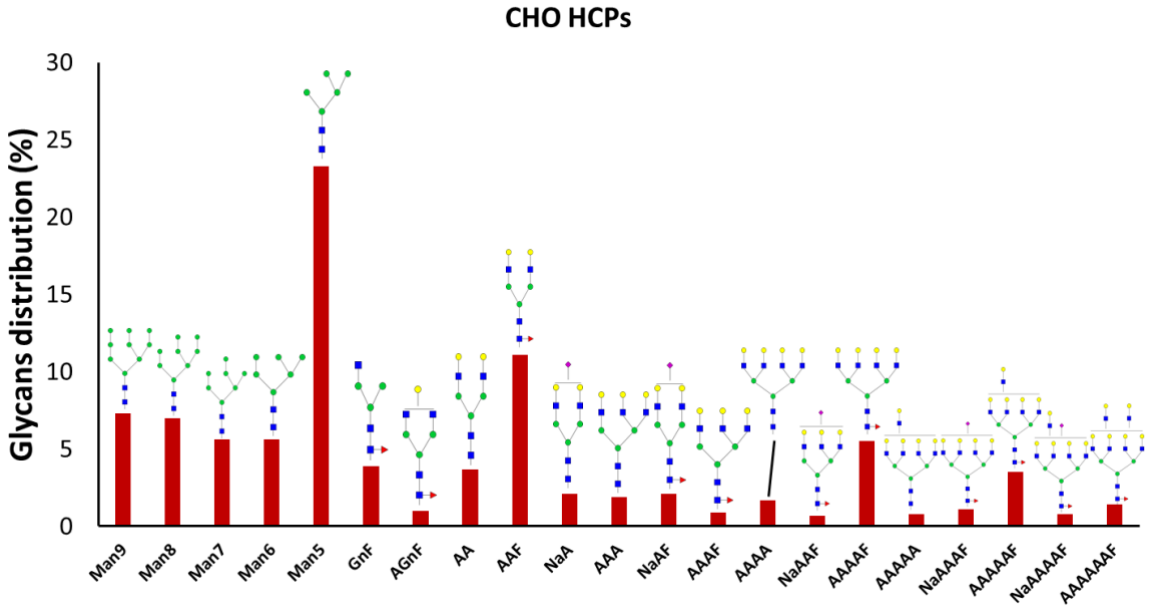

Figure A2: Glycosylation profile of intracellular host cell proteins of CHO cells as reported in North et al.[547]

The manual reduction method has the following workflow: Starting from each essential glycan and going backwards (backpropagation), all the reactions necessary for each essential glycan production are included. If during the backpropagation another essential glycan is encountered, the backpropagation for the examined essential glycans is terminated. No reactions related to the consumption of non-essential glycans that are not leading either directly or indirectly to essential glycans are included. Also, reactions describing essential glycans consumption are included during reduction, but no subsequent reaction was included unless it led to another essential glycan. The manually reduced network includes 346 structures and 596 reactions, leading to a significant decrease of the number of both the structures and the reactions of the network. The "Manual" network includes all the experimentally observed glycans and as shown in Fig. A3 is able to closely describe the vast majority of glycans and even present a slightly lower total error, when compared to the large network. Notably, the "Manual" network is unable to detect the tri-antennary AAAF and NaAAF and the highly branched AAAAA and NaAAAAF that are found in the "Large" network.

Table A9: Enzymes used in the glycosylation network of CHO-HCPs. ManI and ManII are referred twice as they present different kinetic constants for different substrates.

Enzyme	Cosubstrate	Coproduct	Substrate	Product
ManI	water	mannose	(Ma2Ma	(Ma
ManI	water	mannose	(Ma3(Ma2Ma3(Ma6)Ma6)	(Ma3(Ma3(Ma6)Ma6)
ManII	water	mannose	(Ma3(Ma6)Ma6	(Ma6Ma6
ManII	water	mannose	(Ma6Ma6	(Ma6
a6FucT	GDP-Fuc	GDP	GNb4GN	GNb4(Fa6)GN
GnTI	UDP-GlcNAc	UDP	(Ma3(Ma3(Ma6)Ma6)Mb4	(GNb2Ma3(Ma3(Ma6)Ma6)Mb4
GnTII	UDP-GlcNAc	UDP	(GNb2|Ma3(Ma6)Mb4	(GNb2|Ma3(GNb2Ma6)Mb4
GnTIII	UDP-GlcNAc	UDP	GNb2|Ma3	GNb2|Ma3(GNb4)
GnTIV	UDP-GlcNAc	UDP	(GNb2Ma3	(GNb2(GNb4)Ma3
GnTV	UDP-GlcNAc	UDP	(GNb2Ma6	(GNb2(GNb6)Ma6
iGnT	UDP-GlcNAc	UDP	(Ab4GN	(GNb3Ab4GN
b4GalT	UDP-Gal	UDP	(GN	(Ab4GN
a3SiaT	CMP-NeuAc	CMP	(Ab4GN	(NNa3Ab4GN
a3FucT	GDP-Fuc	GDP	(...Ab4GNb	(Fa3(...Ab4)GNb
a3GalT	UDP-Gal	UDP	(Ab4GN	(Aa3Ab4GN

The automatic reduction method follows a similar procedure as in the first step only the reactions and structures required to generate the essential glycans are included. The automatic method is considered as an automatic implementation of the "Manual" network reduction, which is critical due to the extensively tedious, time-consuming and error prone process followed for manual reduction. The *Lumping* method was incorporated in the GLYMMER software by Dr. Fred Krambeck, who is also the developer of the GLYMMER software. Subsequently the network was enriched with reactions of which substrates and products were essential glycans. The automatically reduced "*Lumped*" network includes only 119 structures and 145 reactions, reducing both the number of structures and reactions

more than three (3x) times when compared to the "Manual" network and more than 100x when compared to the "Large" network. As presented in Table A10, the "Lumped" network presents the smallest total error when compared to the other configurations. Unlike the "Manual" network, the "Lumped" network is able to qualitatively detect the AAAF and NaAAF glycan but fails to find both the AAAAA and NaAAAAF (one poly-LacNAc molecule) glycans. The low levels of the AAAAA and AAAAAF glycans could be attributed to the increased concentration of the iGnT (the enzyme responsible for the poly-LacNAc extension) that leads to the formation of mainly AAAAAAF (two poly-LacNAc molecules). Moreover, all three networks were found to closely describe the high mannose glycans. Thus, the "Lumped" network that significantly improves the parameter estimation times when compared to the "Large" network, can be utilized to both qualitatively and quantitatively describe complex glycoprofiles using a very limited number of reactions and structures.

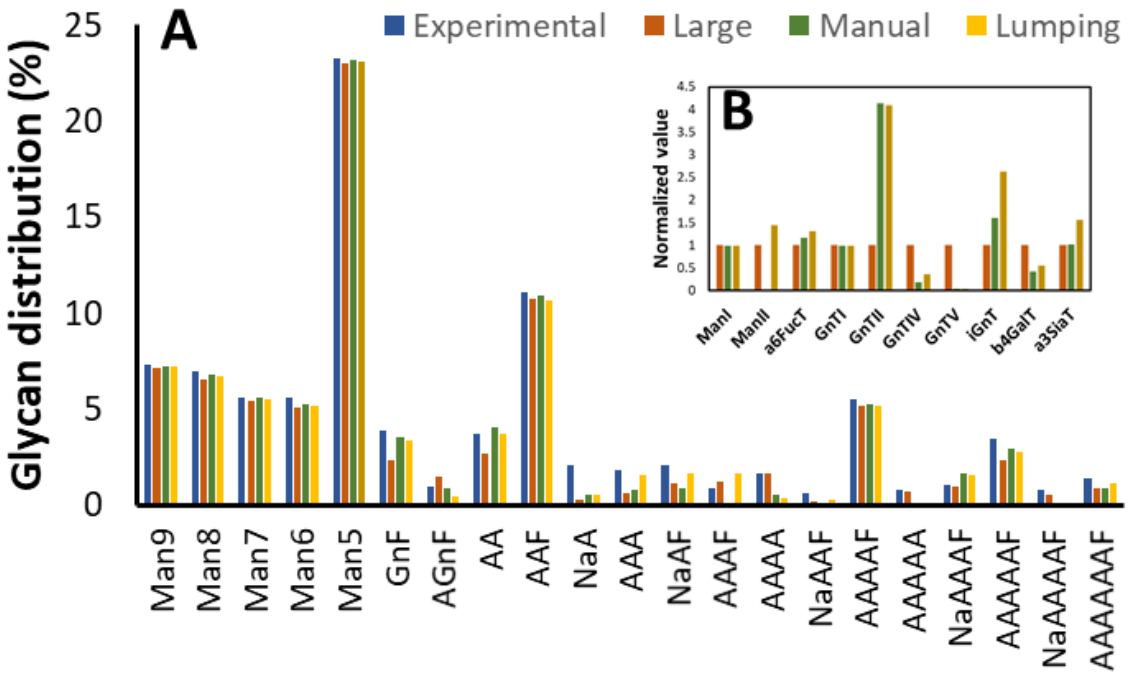

Figure A3: (A) Comparison of the glycans distribution for each model simulation and the experimental data. (B) Estimated enzyme concentrations of the enzymes involved in all the networks (normalized by the value in the "Large" network). The following enzymes: GnTIII, a3FucT and a3GalT are not included in the graph as their concentration was estimated to be zero.

As shown in Fig. A3, the concentration of ManI, a6FucT and a3SiaT is relatively the same between the different networks. Interestingly, ManII shows a very low concentration in the "Manual" network despite the fact that the estimations for the Man5 and GnF glycans are accurate. However, the actual concentration of ManII in the "Manual" experiment is ~1.44μM, which is approximatelly 35x lower compared to the concentration in the "Large" network (~51μM). Interestingly, the GnTII enzyme shows higher concentration in the reduced networks, increasing that way the flux towards the more

complex glycans that are then regulated by the lower concentration of the GnTIV and GnTV enzymes (enzymes responsible for the formation of the tri- and tetra- antennary structures).

Table A10: Total errors for the fitting of each network.

	Large	Manual	Lumping
Total error (%)	11.97	11.60	10.41

Overall, the networks generated by the model reduction methods (either “Manual” or “Lumping”) were found to closely describe the experimental data, without reducing model performance on data fitting. Therefore, the Lumping method was considered reliable in creating a reduced reaction network for the experimentally measured HCPs of the CHO-T127 cell line, as presented in Chapter 5 of this thesis.

Appendix III - Copyrights

JOHN WILEY AND SONS LICENSE
TERMS AND CONDITIONS

Mar 11, 2021

This Agreement between Mr. Pavlos Kotidis ("You") and John Wiley and Sons ("John Wiley and Sons") consists of your license details and the terms and conditions provided by John Wiley and Sons and Copyright Clearance Center.

License Number	5026021275747
License date	Mar 11, 2021
Licensed Content Publisher	John Wiley and Sons
Licensed Content Publication	Biotechnology & Bioengineering
Licensed Content Title	Model-based optimization of antibody galactosylation in CHO cell culture
Licensed Content Author	Cleo Kontoravdi, Ioscani Jimenez del Val, Karen Polizzi, et al
Licensed Content Date	Mar 21, 2019
Licensed Content Volume	116
Licensed Content Issue	7

Licensed Content Pages	15
Type of use	Dissertation/Thesis
Requestor type	Author of this Wiley article
Format	Print and electronic
Portion	Figure/table
Number of figures/tables	14
Will you be translating?	No
Title	Investigating CHO cell metabolism and protein glycosylation based on integrated computational and experimental frameworks
Institution name	Imperial College London
Expected presentation date	Jul 2021
Portions	Figure 1, Figure 2, Figure 3, Figure 4, Figure 5, Figure 6, Table 1, Table 2, Figure S1, Figure S2, Figure S3, Table S1, Table S2, Table S3
Requestor Location	Mr. Pavlos Kotidis Kensington London, England SW7 2AZ United Kingdom Attn: Imperial College London
Publisher Tax ID	EU826007151

Total 0.00 GBP

Terms and Conditions

TERMS AND CONDITIONS

This copyrighted material is owned by or exclusively licensed to John Wiley & Sons, Inc. or one of its group companies (each a"Wiley Company") or handled on behalf of a society with which a Wiley Company has exclusive publishing rights in relation to a particular work (collectively "WILEY"). By clicking "accept" in connection with completing this licensing transaction, you agree that the following terms and conditions apply to this transaction (along with the billing and payment terms and conditions established by the Copyright Clearance Center Inc., ("CCC's Billing and Payment terms and conditions"), at the time that you opened your RightsLink account (these are available at any time at http://myaccount.copyright.com).

Terms and Conditions

- The materials you have requested permission to reproduce or reuse (the "Wiley Materials") are protected by copyright.
- You are hereby granted a personal, non-exclusive, non-sub licensable (on a stand-alone basis), non-transferable, worldwide, limited license to reproduce the Wiley Materials for the purpose specified in the licensing process. This license, **and any CONTENT (PDF or image file) purchased as part of your order,** is for a one-time use only and limited to any maximum distribution number specified in the license. The first instance of republication or reuse granted by this license must be completed within two years of the date of the grant of this license (although copies prepared before the end date may be distributed thereafter). The Wiley Materials shall not be used in any other manner or for any other purpose, beyond what is granted in the license. Permission is granted subject to an appropriate acknowledgement given to the author, title of the material/book/journal and the publisher. You shall also duplicate the copyright notice that appears in the Wiley publication in your use of the Wiley Material. Permission is also granted on the understanding that nowhere in the text is a previously published source acknowledged for all or part of this Wiley Material. Any third party content is expressly excluded from this permission.
- With respect to the Wiley Materials, all rights are reserved. Except as expressly granted by the terms of the license, no part of the Wiley Materials may be copied, modified, adapted (except for minor reformatting required by the new Publication), translated, reproduced, transferred or distributed, in any form or by any means, and no derivative works may be made based on the Wiley Materials without the prior permission of the respective copyright owner.**For STM Signatory Publishers clearing permission under the terms of the STM Permissions Guidelines only, the terms of the license are extended to include subsequent editions and for editions in other languages, provided such editions are for the work as a whole in situ and does not involve the separate exploitation of the permitted figures or extracts,** You may not alter, remove or suppress in any manner any copyright, trademark or other notices displayed by the Wiley Materials. You may not license, rent, sell, loan, lease, pledge, offer as security, transfer or assign the Wiley Materials on a stand-alone

basis, or any of the rights granted to you hereunder to any other person.

- The Wiley Materials and all of the intellectual property rights therein shall at all times remain the exclusive property of John Wiley & Sons Inc, the Wiley Companies, or their respective licensors, and your interest therein is only that of having possession of and the right to reproduce the Wiley Materials pursuant to Section 2 herein during the continuance of this Agreement. You agree that you own no right, title or interest in or to the Wiley Materials or any of the intellectual property rights therein. You shall have no rights hereunder other than the license as provided for above in Section 2. No right, license or interest to any trademark, trade name, service mark or other branding ("Marks") of WILEY or its licensors is granted hereunder, and you agree that you shall not assert any such right, license or interest with respect thereto

- NEITHER WILEY NOR ITS LICENSORS MAKES ANY WARRANTY OR REPRESENTATION OF ANY KIND TO YOU OR ANY THIRD PARTY, EXPRESS, IMPLIED OR STATUTORY, WITH RESPECT TO THE MATERIALS OR THE ACCURACY OF ANY INFORMATION CONTAINED IN THE MATERIALS, INCLUDING, WITHOUT LIMITATION, ANY IMPLIED WARRANTY OF MERCHANTABILITY, ACCURACY, SATISFACTORY QUALITY, FITNESS FOR A PARTICULAR PURPOSE, USABILITY, INTEGRATION OR NON-INFRINGEMENT AND ALL SUCH WARRANTIES ARE HEREBY EXCLUDED BY WILEY AND ITS LICENSORS AND WAIVED BY YOU.

- WILEY shall have the right to terminate this Agreement immediately upon breach of this Agreement by you.

- You shall indemnify, defend and hold harmless WILEY, its Licensors and their respective directors, officers, agents and employees, from and against any actual or threatened claims, demands, causes of action or proceedings arising from any breach of this Agreement by you.

- IN NO EVENT SHALL WILEY OR ITS LICENSORS BE LIABLE TO YOU OR ANY OTHER PARTY OR ANY OTHER PERSON OR ENTITY FOR ANY SPECIAL, CONSEQUENTIAL, INCIDENTAL, INDIRECT, EXEMPLARY OR PUNITIVE DAMAGES, HOWEVER CAUSED, ARISING OUT OF OR IN CONNECTION WITH THE DOWNLOADING, PROVISIONING, VIEWING OR USE OF THE MATERIALS REGARDLESS OF THE FORM OF ACTION, WHETHER FOR BREACH OF CONTRACT, BREACH OF WARRANTY, TORT, NEGLIGENCE, INFRINGEMENT OR OTHERWISE (INCLUDING, WITHOUT LIMITATION, DAMAGES BASED ON LOSS OF PROFITS, DATA, FILES, USE, BUSINESS OPPORTUNITY OR CLAIMS OF THIRD PARTIES), AND WHETHER OR NOT THE PARTY HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES. THIS LIMITATION SHALL APPLY NOTWITHSTANDING ANY FAILURE OF ESSENTIAL PURPOSE OF ANY LIMITED REMEDY PROVIDED HEREIN.

- Should any provision of this Agreement be held by a court of competent jurisdiction to be illegal, invalid, or unenforceable, that provision shall be deemed amended to achieve as nearly as possible the same economic effect as the original provision, and the legality, validity and enforceability of the remaining provisions of this Agreement shall not be affected or impaired thereby.

- The failure of either party to enforce any term or condition of this Agreement shall not constitute a waiver of either party's right to enforce each and every term and condition of this Agreement. No breach under this agreement shall be deemed waived or excused by either party unless such waiver or consent is in writing signed by the party granting such waiver or consent. The waiver by or consent of a party to a breach of any provision of this Agreement shall not operate or be construed as a waiver of or consent to any other or subsequent breach by such other party.

- This Agreement may not be assigned (including by operation of law or otherwise) by you without WILEY's prior written consent.

- Any fee required for this permission shall be non-refundable after thirty (30) days from receipt by the CCC.

- These terms and conditions together with CCC's Billing and Payment terms and conditions (which are incorporated herein) form the entire agreement between you and WILEY concerning this licensing transaction and (in the absence of fraud) supersedes all prior agreements and representations of the parties, oral or written. This Agreement may not be amended except in writing signed by both parties. This Agreement shall be binding upon and inure to the benefit of the parties' successors, legal representatives, and authorized assigns.

- In the event of any conflict between your obligations established by these terms and conditions and those established by CCC's Billing and Payment terms and conditions, these terms and conditions shall prevail.

- WILEY expressly reserves all rights not specifically granted in the combination of (i) the license details provided by you and accepted in the course of this licensing transaction, (ii) these terms and conditions and (iii) CCC's Billing and Payment terms and conditions.

- This Agreement will be void if the Type of Use, Format, Circulation, or Requestor Type was misrepresented during the licensing process.

- This Agreement shall be governed by and construed in accordance with the laws of the State of New York, USA, without regards to such state's conflict of law rules. Any legal action, suit or proceeding arising out of or relating to these Terms and Conditions or the breach thereof shall be instituted in a court of competent jurisdiction in New York County in the State of New York in the United States of America and each party hereby consents and submits to the personal jurisdiction of such court, waives any objection to venue in such court and consents to service of process by registered or certified mail, return receipt requested, at the last known address of such party.

WILEY OPEN ACCESS TERMS AND CONDITIONS

Wiley Publishes Open Access Articles in fully Open Access Journals and in Subscription journals offering Online Open. Although most of the fully Open Access journals publish open access articles under the terms of the Creative Commons Attribution (CC BY) License only, the subscription journals and a few of the Open Access Journals offer a choice of Creative Commons Licenses. The license type is clearly identified on the article.

The Creative Commons Attribution License

The Creative Commons Attribution License (CC-BY) allows users to copy, distribute and transmit an article, adapt the article and make commercial use of the article. The CC-BY license permits commercial and non-

Creative Commons Attribution Non-Commercial License

The Creative Commons Attribution Non-Commercial (CC-BY-NC)License permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.(see below)

Creative Commons Attribution-Non-Commercial-NoDerivs License

The Creative Commons Attribution Non-Commercial-NoDerivs License (CC-BY-NC-ND) permits use, distribution and reproduction in any medium, provided the original work is properly cited, is not used for commercial purposes and no modifications or adaptations are made. (see below)

Use by commercial "for-profit" organizations

Use of Wiley Open Access articles for commercial, promotional, or marketing purposes requires further explicit permission from Wiley and will be subject to a fee.

Further details can be found on Wiley Online Library
http://olabout.wiley.com/WileyCDA/Section/id-410895.html

Other Terms and Conditions:

v1.10 Last updated September 2015

Questions? customercare@copyright.com or +1-855-239-3415 (toll free in the US) or +1-978-646-2777.