New constraint-based methods to characterize Chinese hamster ovary cell metabolism

Dissertation submitted by

SARAH NOEL GALLEGUILLOS

for obtaining a doctorate degree from the

UNIVERSITY OF NATURAL RESOURCES AND LIFE SCIENCES

Department of Biotechnology

Supervisor

Ao.Univ.Prof. Dipl.-Ing. Dr.nat.techn Nicole Borth

JULY 2019

Abstract

As the host of choice for the production of biopharmaceuticals, Chinese hamster ovary (CHO) cells have been in the spotlight for process optimization in recent years. In order to design rational strategies to further increase protein yields, mechanistic, systems-wide understanding of cellular processes is essential. Genome-scale metabolic network models together with constraint-based analysis methods offer one framework for gaining such an understanding. Moreover, they provide a rational basis for the integration of data from different biological sources and for mathematical modelling of cellular metabolism. Since a community-built metabolic model for CHO was recently published, the full potential of these modelling techniques is now at hand to help us elucidate the complex mechanisms behind protein production in CHO.

The objective of this thesis was to apply constraint-based methods to study of CHO metabolism and to identify potential metabolic engineering targets for efficient energy utilization and increased protein production. Using biased approaches, I identified reactions which correlate with protein production under different experimental conditions. Moreover, based on an unbiased analysis of a medium-scale reconstruction of CHO's central metabolism, I identified various pathways that are prime engineering targets for optimized energy metabolism.

Additionally, I developed a new approach based on concepts of convex analysis for the characterization of the phenotypic capabilities (production envelopes) of an organism. Contrary to currently available methods, this algorithm can be applied simultaneously to several reactions of interest to characterize all feasible steady-state flux distributions within a reduced computation time.

KURZFASSUNG

In den letzten Jahren standen Chinese hamster ovary (CHO) Zellen im Fokus der Prozessoptimierung, da sie die erste Wahl für die Herstellung von Biopharmazeutika waren. Um rationale Strategien zur weiteren Steigerung der Proteinausbeute zu entwickeln, ist ein mechanistisches, systemweites Verständnis der zellulären Prozesse unerlässlich. Metabolische Netzwerkmodelle im Genommaßstab bieten zusammen mit auf Beschränkungen basierenden Analysemethoden einen Rahmen, um ein solches Verständnis zu erlangen. Darüber hinaus bieten sie eine rationale Grundlage für die Integration von Daten aus verschiedenen biologischen Quellen und für die mathematische Modellierung des Zellstoffwechsels. Da kürzlich ein von einer Community erstelltes Stoffwechselmodell für CHO veröffentlicht wurde, können wir jetzt das gesamte Potenzial dieser Modellierungstechniken nutzen, um die komplexen Mechanismen für die Proteinproduktion in CHO aufzuklären.

Ziel dieser Dissertation war es, Constraint-basierte Methoden anzuwenden, um den Stoffwechsel zu untersuchen und potenzielle metabolische Engineering-Ziele für eine effiziente Energienutzung und eine erhöhte Proteinproduktion zu identifizieren. Mit voreingenommenen Ansätzen identifizierte ich Reaktionen, die mit der Proteinproduktion unter verschiedenen experimentellen Bedingungen korrelieren. Auf der Grundlage einer unvoreingenommenen Analyse einer mittelgroßen Rekonstruktion des zentralen Stoffwechsels von CHO identifizierte ich außerdem verschiedene Wege, die die wichtigsten technischen Ziele für einen optimierten Energiestoffwechsel darstellen.

Zusätzlich entwickelte ich einen neuen Ansatz, der auf Konzepten der konvexen Analyse zur Charakterisierung der phänotypischen Fähigkeiten (Produktionshüllen) eines Organismus basiert. Im Gegensatz zu derzeit verfügbaren Methoden kann der Algorithmus gleichzeitig auf mehrere interessierende Reaktionen angewendet werden, um alle möglichen stationären Flussverteilungen innerhalb einer reduzierten Rechenzeit zu charakterisieren.

Dedication and acknowledgements

To my parents, for their unconditional love, support and guidance.

I would like to express my sincere gratitude to my advisor Prof. Nicole Borth, for the support during my PhD, for her enthusiasm and immense knowledge.

To Jungbauer's research group, thank you for the way you made me welcome during my first year at BOKU. Special thanks to Nicole and Flo, for making Austria feel like my second home.

I would like to thank the Austrian Centre of Industrial Biotechnology and the University of Natural Resources and Life Sciences for funding my research.

I am grateful to my colleagues working in the lab, for not letting me forget about the bigger picture and the practical applications of my work.

Thanks to my dear bioinformaticians, Inma and Heena. You have always been an example of hard work and resilience.

Special thanks to Agnes, Verena, Caro, Elisabeth, Sabrina, Mark (and little Laura), Bernie, Daniel, Nico and Magdalena for giving me many moments of joy that kept me motivated throughout my PhD. A special mention to Concetta, for her infinite generosity.

A la familia que se elige y a mis niñas de Bruselas, por su empatía y por los buenos momentos. A Jose Manuel. Por su infinito apoyo, por ser paciente todos estos años y por hacer que yo también lo fuera.

And last but not least, to Jürgen, for...ty-two.

AUTHOR'S DECLARATION

I declare that the work in this dissertation was carried out in accordance with the requirements of the University's Regulations and Code of Practice for Research Degree Programmes and that it has not been submitted for any other academic award. Except where indicated by specific reference in the text, the work is the candidate's own work. Work done in collaboration with, or with the assistance of, others, is indicated as such. Any views expressed in the dissertation are those of the author.

SIGNED: ..... DATE: .....

All discoveries in art and science result from an accumulation of errors.

Marshall McLuhan

Table of Contents

List of Abbreviations

ATP adenosine triphosphate

CH convex hull

CHM convex hull method

CHO Chinese hamster ovary

DCH department of chemistry

EFM elementary flux mode

EFV elementary flux vector

EP extreme point

FBA flux balance analysis

FSEOF flux scanning based on enforced objective flux

FVA flux variability analysis

GSMM genome-scale metabolic model

HP hyperplane

IgG Immunoglobulin G

lac lactate

LP linear programming

PE production envelope

pFBA parsimonious flux balance analysis

PPP pentose phosphate pathway

QbD quality by design

TCA tricarboxylic acid

CHAPTER 1

INTRODUCTION

Mammalian cells have the unique ability to perform complex post-translational modifications on proteins in a way microbial systems are not able to [1]. Therefore, and despite the low product yield and high process cost as compared to prokaryotic platforms [2], they have become the leading platform for the production of biopharmaceuticals. These modifications (e.g. glycosylation) are crucial for serum half-life, therapeutic efficacy and immunogenicity. More specifically, CHO cells are currently the host of choice due to several advantageous features such as their virus resistance and their capability to grow in suspension in serum-free chemically defined cultivation media [3]. Additionally, methods for gene manipulation and clone selection are well characterized in CHO cells and they have a long, well-established history of approvals for clinical applications [4, 5].

In order to meet the increasing demand for recombinant therapeutic proteins in a field worth more than $360 billion in annual revenue [6], major efforts have focused on enhancing protein yield by optimizing both the process and the host cell. However, most engineering strategies to date are the result of time-consuming and costly clone screening processes and oftentimes lack a mechanistic understanding of the underlying biological processes involved.

With the recent development of 'omics' technologies, it has become possible to gain a deeper insight into cell metabolism in a more integrative manner. Towards this end, systems biology relies heavily on mathematical representations of cell biology, such as stoichiometric models. These models contain information about the mass balance relationships of the metabolites in a biochemical network. First metabolic modelling approaches in CHO cells were carried out using models from other animal cells such as baby hamster kidney or hybridoma cells [7, 8]. In 2011 the genome of CHO became available [3, 9, 10], marking a turning point for the development of metabolic modelling approaches for CHO. For a brief summary of metabolic modelling applications

therein, see subsection 1.4.

At the present time, there is a plethora of different models for mammalian cell growth and protein production (empirical, mechanistic and stochastic) which apply different approaches ranging from kinetic models to artificial neural networks [11]. The focus of this thesis lies on metabolic modelling of CHO cells and, more specifically, on constraint-based modelling for the characterization of metabolic capabilities of CHO and potential discovery of metabolic targets.

Constraint-based approaches have been developing for 30 years and have proven to be a valuable tool to predict the metabolic behavior, most successfully in microorganisms, under different experimental conditions [12]. The main advantage of these methods is that they do not rely on kinetic parameters and therefore can be applied to genome-scale metabolic models. The following sections of this thesis outline how I applied methods based on linear algebra and computational geometry to explore the metabolic limitations of CHO cells under different experimental conditions and to establish computational tools to find new potential engineering targets.

1.1 CHO cells

Ever since the first CHO-produced therapeutic protein was approved in 1987 (Alteplase, Genentech, Inc), innovations on the bioprocess have boosted protein yields from 10 to 50 mg/L to more than 10 g/L [13]. Nowadays, top-selling drugs such as Humira, Rituxan or Avastin are produced in CHO [14].

1.1.1 CHO cells as protein factories

The establishment of a new producer cell line encompasses transfection of selection and recombinant genes and screening for the clones of interest (typically high growth and productivity). Genomic sequence analysis of the most commonly used CHO cell lines (CHO-K1, DG44 and CHO-S) highlights the genetic diversity among them [10]. Moreover, CHO cells are known to be genomically unstable [15], mainly due to copy number variations and chromosomal rearrangements [16]. These variations cause clonal heterogeneity [17, 18] and might affect growth and productivity, making the establishment of a producer clone even more challenging.

Genomic instability in CHO limits the applicability of *in silico* methods to different cell lines and experimental conditions [19]. Nevertheless, the rapid increase in the available data sets from different cell lines (e.g. those in CHOgenome.org [20]) enables the characterization of those differences and helps building cell line specific models with the integration of transcriptomics or proteomics data into the metabolic network reconstruction.

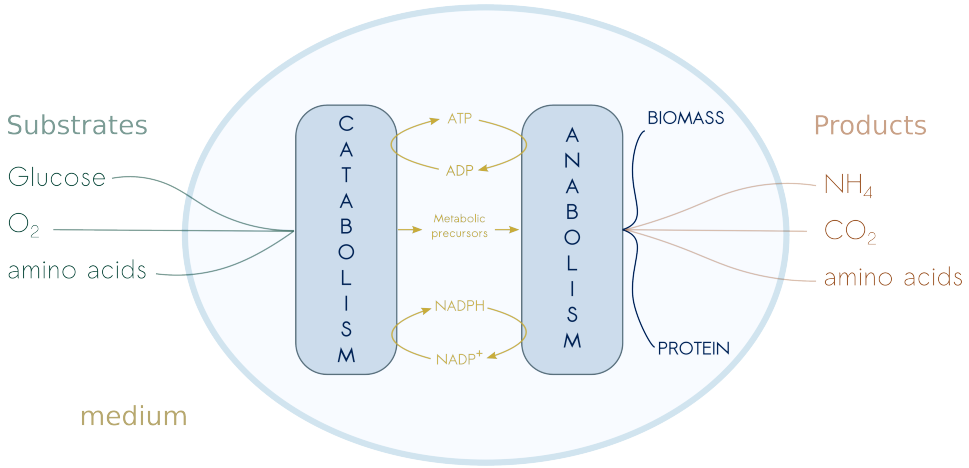


FIGURE 1.1. Schematic representation of cellular metabolism. Source molecules (substrates) are catabolized, generating energy (ATP and NADPH) and precursors for anabolic pathways. Biomass and production of the protein of interest will compete for these resources. Modified from [21]

1.1.2 Metabolism

Protein production depends on the efficiency of the transcription-translation machinery and secreting capabilities of the cell, but it also relies heavily upon cell metabolism, more specifically on the production of the main building blocks (amino acids) of the protein as well as the energy required for the biosynthesis (high-energy phosphates and electrons, primarily in the form of ATP and NADPH, see Figure 1.1).

Like other mammalian cells, CHO cells exhibit high metabolic fluxes through glycolysis and glutaminolysis with the collateral over-production of ammonia, lactate and alanine. Lactate impacts pH and osmolarity [22] and consequently reduces cell growth [23, 24] and protein production [25]. In addition, it is an indication of inefficient use of the main carbon sources, since converting glucose to lactate only generates two molecules of ATP, whereas 36 molecules are produced if glucose is completely oxidized in the tricarboxylic acid (TCA) cycle. Likewise, high concentrations of ammonia severely impact cell growth, productivity and glycosylation [26, 27, 28, 29]. In the past decades, bioprocess optimization strategies have successfully reduced the accumulation of both ammonia and lactate by (i) control of of glucose and glutamine feeds in fed-batch cultures [30, 31], (ii) using alternative substrates [32] and (iii) cell engineering [33, 34].

1.2 Glycosylation of recombinant proteins in CHO

Glycosylation is the most frequent and complex type of post-translational modification. It consists in the addition of carbohydrates at specific protein sites and generates oligosaccharide structures

that differ in length and branching points. This modification highly affects protein folding and secretion, as well as proteins' clinical properties such as clearance or antigenicity [35].

Glycan heterogeneity is known to be influenced by numerous factors. Among them, the metabolic state (glutamine and glucose metabolism, flux through the TCA cycle) [36] and bioprocess conditions such as pH [37], temperature [38] or bioreactor operation [39].

More than twenty enzymes participate in glycosylation along the Endoplasmic Reticulum (ER) and Golgi apparatus (Figure 1.2). Many of them have common substrates and can perform the same enzymatic reaction on different glycans, hence the multiplicity of possible outcome structures. This has challenged the application of glycoengineering techniques to a process once thought to be a consequence of random biochemical effects.

Central carbon metabolism is linked to glycosylation via nucleotide-sugar metabolism; UDP-N-acetylglucosamine (UDP-GlcNAc) and UDP-N-acetylgalactosamine (UDP-GalNAc) are the main precursors of glycosylation. Protein folding and addition/removal of sugars occur along ER and Golgi compartments. Finally, the proteins can stay in the cytoplasm or be transported to the membrane for secretion. The biochemical process of protein glycosylation is summarized in Figure 1.2.

1.3 Systems Biology

During the past decades, cell culture engineering has undergone a period of dramatic increase in product yields [13]. The main trend has been to go from random mutagenesis and clone screening to more targeted (rational) engineering approaches. Genes, transcripts and proteins are not isolated components but show high levels of inter-connectivity in biological systems and are thus better described as networks. In order to predict the effect of any genetic or environmental changes on the phenotype in metabolic engineering, one must consider the system as a whole. In this regard, systems biology aims to integrate different sources of biological information about an organism and use computational approaches to predict and improve cell performance [40]. The significant rise in availability of omics data and analysis [41] has boosted the application of systems biology approaches in CHO cells [42].

Understanding the basis of high productivity and final product quality is crucial in order to establish proper bioprocess control. In recent years, pharmaceutical regulatory agencies such as the Food and Drug Administration (FDA) are promoting the use of quality by design (QbD) approaches for the development of new biopharmaceuticals [43]. These approaches aim to ensure product quality by defining a quality profile based on therapeutic performance and identifying critical parameters affecting product quality in order to be able to control the process and make it more robust [44, 45]. Understanding the factors that influence protein quality implies biochemical characterization of the process. Mathematical modelling facilitates this task and has become an essential tool for rational experimental design and process control.

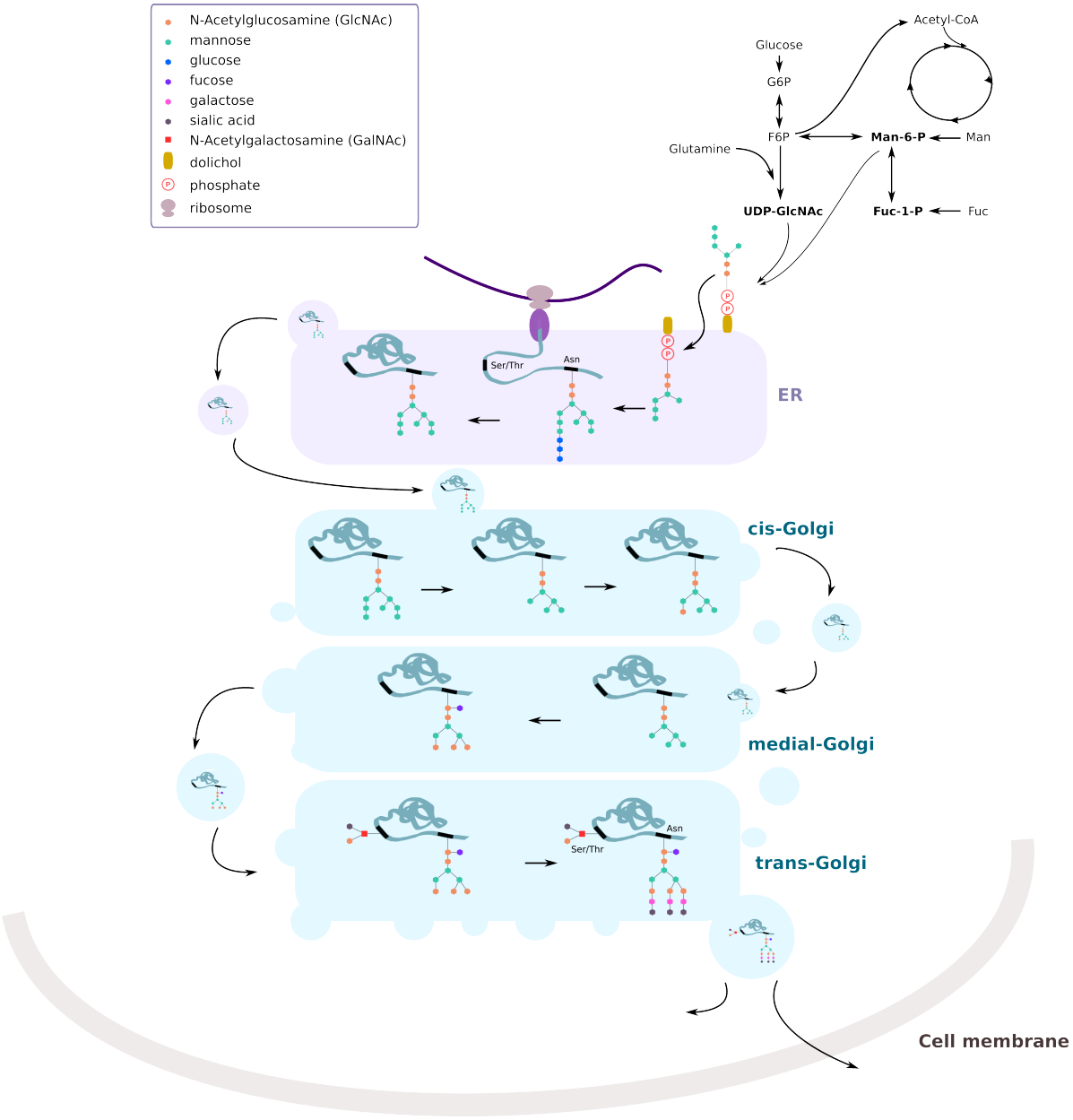


FIGURE 1.2. Glycosylation of proteins through the ER and Golgi. Nucleotide-sugars precursor is transferred from dolichol to a Asn-X-Ser/Thr sequence in the protein. Several glycosyltransferase-mediated steps of trimming, elongation and terminal glycosylation will lead to the final folded, glycosylated structure. Non-completion of the process gives rise to heterogeneity in the final glycan structure.

1.3.1 Biochemical reaction networks

Reaction networks represent the genotype-phenotype relationship (genes to enzymes to substrate and products) and are a key aspect of systems biology. These networks are the scaffold in which omics data can be integrated and they allow for mathematical representation and, consequently, computational modelling. These models allow us to predict phenotypic states based on the genetic content of an organism [21].

The increased availability of high-throughput data in the past decades has enriched the field of systems biology with the concomitant generation of high quality-curated genome-scale metabolic models. These are network-based representations of cellular metabolism; they gather all stoichiometric relationships (reactions or edges) between metabolites (nodes) known for a given organism. They also contain information about reaction kinetics, compartmentalization, reaction directionalities and capacities (flux bounds). When known, gene identifiers for each reaction are also included. Exchange reactions allow uptake/secretion of metabolites from/to the medium and transport between compartments.

1.3.2 Metabolic network reconstruction

Metabolic reconstructions rely on diverse sources of biological data (Figure 1.3) [46] and the steps to follow are well-established in the form of a standard protocol [47]. Metabolic reactions can be included in the model if there exists experimental evidence of the presence of an enzyme; in other cases proteins can be assigned to genes based on homology to other genomes. Genome annotation is essential for model building [21]. Other reactions are "inferred", such as the biomass reaction, which represents the biosynthetic demands for cell growth. These are obtained experimentally from the quantification of biomass components. Metabolic modelling also aids in the process of gap filling, in which reactions that are necessary for a metabolite to be produced (or consumed) are added for model completion. Several toolboxes are currently available for network reconstruction in a semi-automatized way [48, 49]; followed by very labour- and time-intensive manual curation [50].

Despite the avalanche of genomics, proteomics, transcriptomics, metabolomics, fluxomics and lipidomics data which helps us improve our understanding of cell behavior, recent publications highlight the need for standard methods for network reconstruction and quality control [51]. Discrepancies in data obtained from different public databases such as BiGG [52], KEGG [53] or MetaCyc [54] should be removed to alleviate the process of network reconstruction and to generate results which are reproducible and comparable within the scientific community [55, 56].

Metabolic models are generally available in Systems Biology Markup Language (SBML) [57]. There are currently hundreds of software applications that support the SBML format [58]. The use of a standard format for metabolic models facilitates data exchange through the creation of extensive model databases such as BioModels [59] or BiGG [52]. Many of the available toolboxes

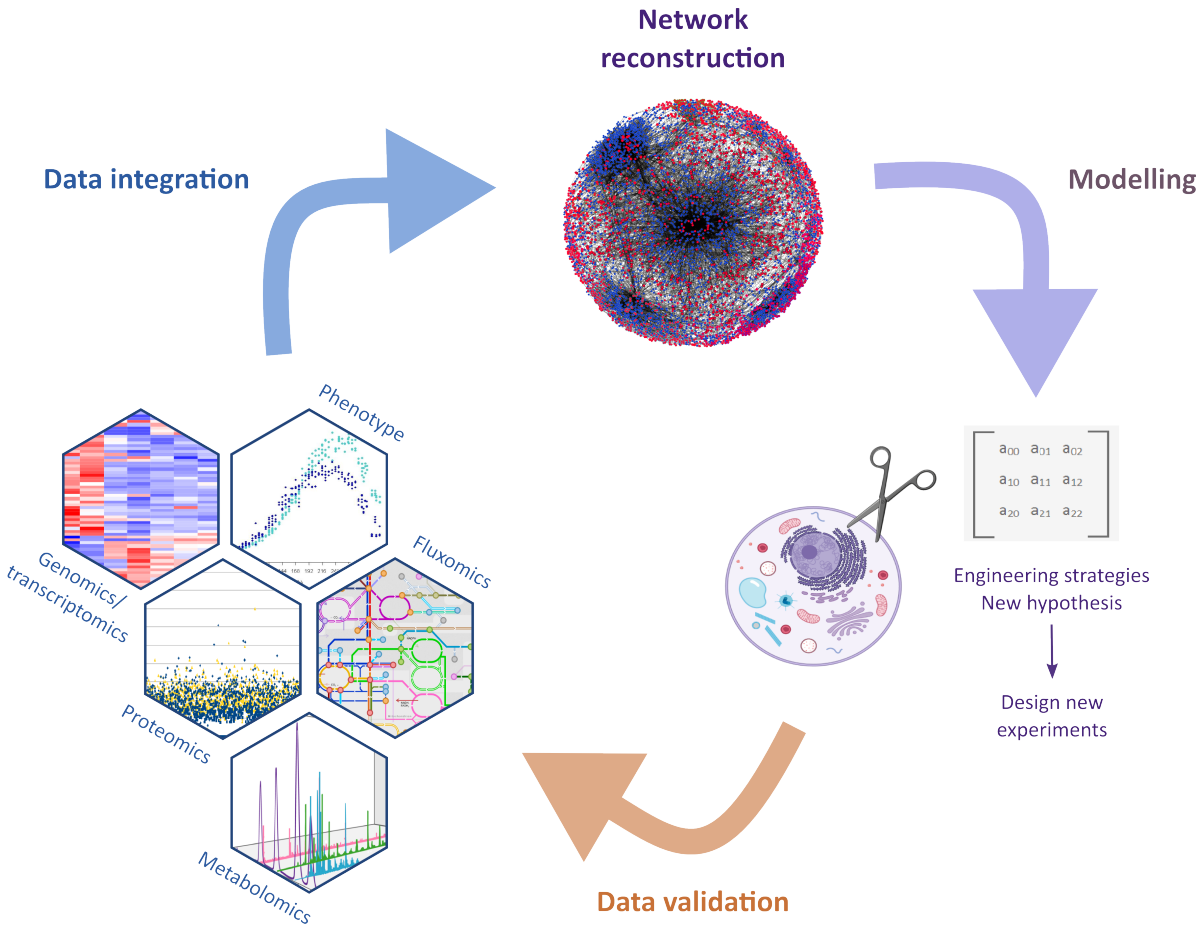


FIGURE 1.3. Genome-scale metabolic networks represent the available knowledge on cell metabolism. These reconstructions derive from the integration of omics data together with biochemical and physiological information. Computational modelling is then applied to develop new strategies for cell line optimization. These strategies are validated with experiments, new data collected, and the reconstruction is refined.

include software for network reconstruction, visualization and analysis (COBRA [60], RAVEN [61]) as well as for *in silico* design of engineering strategies [62].

These reconstructions of all metabolic knowledge of an organism can then be represented as mathematical models, which makes them amenable to computer simulation (Figure 1.3). *In silico* methods have been widely applied to metabolic models of microorganisms [63, 64] and human [65, 66, 67], and are now rapidly developing for CHO cells (see section 1.4).

Metabolic networks can be analyzed and modelled from distinct perspectives: some approaches focus merely on the topology of the network (how nodes are connected, as in protein-protein interaction networks); others aim to mechanistically predict the evolution of metabolic species

in time. However, kinetic information on reaction rates is scarce, limiting the applicability of these methods to small networks. A third major area of network analysis focuses on additional constraints other than topology to characterize the space of feasible metabolic phenotypes [68]. In this thesis, I have applied these so-called constraint-based methods to the recently published metabolic model for CHO, iCHO1766 [69].

1.3.3 CHO metabolic model

Metabolic modelling of CHO cells has been carried out for more than 20 years [70]. Without a metabolic network specific for CHO, first modelling approaches focused mainly on central carbon metabolism [71] and were derived from other mammalian cells (e.g., mouse [72] or hybridoma [8]). Following the sequencing of a CHO cell line in 2011 [3], the publication of a community-built genome-scale metabolic model (GSMM) [69] has ushered in a new era of systems biology for CHO. Transcriptomics and proteomics data were incorporated in the generic model to generate cell-type specific versions for the most widely used cell lines (K1, S, DG44). The generic iCHO1766 model consists of 6663 reactions distributed in 10 compartments (including "external" for exchange reactions). In order to be able to represent transport reactions across compartments, metabolites (2342 in total) are divided into "species" (4455), for example, `M_atp_m` and `M_atp_c` is ATP in the mitochondria and in the cytosol, respectively.

1.4 Mathematical models of CHO metabolism and glycosylation

Section 1.4 is part of the review published by Sarah N.Galleguillos, David Ruckerbauer, Matthias P. Gerstl, Nicole Borth, Michael Hanscho and Jürgen Zanghellini in *Computational and Structural Biotechnology Journal*, January 2017 [73].

> Here we will review recent progress in the computational modelling of CHO cells. Specifically, we will focus on and analyze two main issues associated with recombinant protein production: (i) metabolic burdens affecting growth and thus protein yield and (ii) understanding of the correct glycosylation process of the protein of interest, which is one of the major criteria for product quality.

1.4.1 Modelling CHO metabolism

> To gain mechanistic understanding of these processes, appropriate metabolic models are required that allow one to estimate cellular flux distributions. This can be done in two ways: (i) in a time-dependent or dynamic manner (kinetic analysis) or (ii) in a constraint-based, steady-state analysis. The former approach aims to assess the evolution of the concentrations of metabolites over time and requires a large number of kinetic parameters. Due to the lack of accurate, quantitative data, this approach is currently not feasible on a genome-scale level, but restricted to small-scale models

that consider several tens of selected reactions and interactions. The latter approach, on the other hand, avoids the need for detailed kinetic information by focusing on the steady-state behaviour inside the cell. Disregarding dynamic processes makes this approach, called metabolic flux analysis (MFA), scalable and suitable for genome-wide analysis. For better understanding the modelling approaches are briefly reviewed in Box 1.

Box 1: Common modelling approaches

MFA (Metabolic Flux Analysis): pathway analysis method based on the stoichiometry of metabolic reactions and mass balances under pseudo-steady-state assumption [74]. It can be implemented in several ways. Among them:

- **FBA** (Flux Balance Analysis): an implementation of MFA based on the optimization of a cellular function (such as growth) under specific constraints (experimental metabolic uptake and secretion rates, thermodynamic data, etc) [75, 76].
- **^{13}C MFA**: isotope-labelled substrates are added to the culture media and, once the isotopic steady-state is reached, the distribution of the isotopes is measured via nuclear magnetic resonance or gas chromatography—mass spectrometry [77].

Markov chain Monte Carlo sampling: the glycosylation process is described as a series of states with transition probabilities from one state to the other. In the references reviewed herein, it is used to overcome the lack of kinetic parameters (metabolic and glycosylation enzymes) [78].
Artificial Neural Network models aim to predict the behaviour of complex, non-linear systems by detecting and "learning" patterns and relationships within a training set which can be applied then to the input data [79].

In the following section we review current advances in metabolic modelling of CHO cells (listed chronologically in Figure 1.4), focusing on those that investigate the accumulation of the two main metabolic by-products that are detrimental to cell growth, i.e. lactate and ammonia.

1.4.1.1 The metabolic fate of lactate

Altamirano *et al.* [8] investigated the metabolic fate of lactate on a metabolic network of CHO core metabolism. They argued that, when re-metabolized, lactate is not used as an energy source, as their experimentally measured low oxygen uptake rate was inconsistent with a full oxidation of lactate via the TCA cycle. Consequently, they

proposed alternative pathways for the non-oxidative decarboxylation of pyruvate, which are known to exist in cancer cells [80], to be present in CHO cells too. Nevertheless, the accumulation of the end product of these pathways, i.e. acetoin, was not experimentally proven. In a more recent work, Martinez *et al.* [81] were able to refute this hypothesis. In their study, they analyzed the metabolic switch from lactate production to lactate uptake by means of FBA in a reduced mouse-derived metabolic model. Contrary to Altamirano *et al.*, Martinez *et al.* showed that their oxygen uptake rate measurements were consistent with lactate oxidation in the TCA cycle. This suggests that the metabolic network of Altamirano *et al.* might have been too simplistic to capture the metabolic changes between the phases. Compared to Martinez, Altamirano's model lacked fatty acid, steroid and glycogen metabolism. In addition, the prediction of the ATP yield per mol carbon identified lactate consumption to be energetically more efficient than glucose consumption. Furthermore, they showed that the estimation of ranges for the metabolic fluxes (due to the insufficient amount of experimentally measured data in an underdetermined network) provides a valuable, semi-quantitative description of the changes between the two metabolic states. This concept was also supported by Zamorano *et al.* [82], who performed MFA in an under-determined network containing 100 reactions of the core metabolism and obtained narrow intervals for the fluxes with a relatively low amount of extracellular measurements.

FBA can be combined with isotopomer analysis to improve the accuracy of the predicted fluxes. Sengupta *et al.* [83] studied the main metabolic fluxes in a simplified network during the stationary phase of cell culture by ^{13}C MFA. This phase is typically characterized by reduced production of lactate and high protein yields. Likewise, Templeton *et al.* [84] performed ^{13}C MFA to understand the metabolic changes between growth and stationary phases in a producer CHO cell line. They found that, during the antibody production peak (stationary phase), fluxes through the TCA cycle were maximal while lactate was not produced. Moreover, this increased activity of the TCA cycle correlated with increased fluxes through the oxidative pentose phosphate pathway (PPP) when compared to the exponential phase, where high glycolytic fluxes predominate. They provide several explanations for the activation of the oxidative PPP: to regenerate NADPH/NADP$^+$, to compensate reduction during exponential growth, to suppress oxidative stress or to cover NADPH requirements during protein folding and secretion. Irrespective of the ultimate reason, these findings point towards metabolic engineering to increase oxidative TCA cycle (CO_2-producing reactions) and PPP fluxes which would help achieve higher protein yields.

1.4.1.2 Lactate as a beneficial medium component?

More recently, Chen *et al.* [85] even suggested that adding small amounts of lactate at the beginning of the culture process increases the metabolic efficiency. They used a kinetic model of the central carbon metabolism (i.e. glycolysis, PPP and TCA cycle) coupled with a model of the population dynamics and computed the time-dependent yield of lactate with respect to glucose. They found this yield decreased with increasing (yet not toxic) initial extracellular concentrations of lactate, meaning more efficient use of glucose. These findings were supported by Li *et al.* [86], who found that lactate can be fed as a major carbon source when glucose concentrations are kept low in culture.

Lactate uptake in the presence of galactose was also studied by flux balance analysis (FBA) in tissue plasminogen activator producing CHO cells in batch cultures [87]. Main changes were observed to occur in the pyruvate metabolism; the slow utilization of galactose as compared to glucose does not provide enough pyruvate to fulfill the energy requirements. This causes lactate dehydrogenase to reverse its mode of operation, transforming lactate into pyruvate, which then enters the TCA cycle. Consequently, intracellular pyruvate and lactate concentrations are reduced, which activates the monocarboxylate transporter towards lactate uptake.

The importance of taking compartments into consideration when modelling metabolism has been demonstrated by analyzing enzyme localized activity together with non-stationary ^{13}C techniques. These allow a more accurate assessment of metabolic fluxes [88], mostly for those pathways that cannot be resolved using steady state approaches, such as cyclic or parallel pathways (e.g. glycolysis and PPP). In this study, Nicolae *et al.* also discussed the sources of lactate production in both cytosol and mitochondria. Taking into account not only the time-evolution of the metabolites, but also their spatial localization, proved that there is an additional control factor of precursor availability for both glycolysis and TCA cycle [89].

Likewise, Ahn *et al.* [90, 91] performed high precision ^{13}C MFA on a network containing 79 reactions and resolved metabolic fluxes accurately. During the exponential phase, characterized predominantly by high fluxes through glycolysis, 70% of the glucose was converted to lactate. They also observed a decrease in glycolytic fluxes and an increase in the oxidative PPP in the stationary phase, as reported previously [84].

1.4.1.3 What makes a "good" growth medium?

As already mentioned, the addition of alternative energy feedstocks can reduce the accumulation of undesired by-products. The effects of these alternative carbon sources on metabolism and protein production were studied with MFA on a reduced

metabolic network by Altamirano *et al.* [92]. They showed that replacing glutamine by glutamate indeed resulted in reduced accumulation of ammonia, although at the price of a lower glucose uptake rate. This lowered metabolism has a negative impact on the specific protein production rate, as carbon is predominantly captured to sustain growth, leaving little for protein production.

In a follow-up work, Altamirano *et al.* [8] considered co-feeding strategies with galactose added to the medium, as galactose-glutamate media are known to significantly reduce by-product formation, but unfortunately, also cell growth. However, they showed that after glucose depletion, cells were able to maintain growth on galactose by simultaneously utilizing previously produced lactate. Interestingly, CHO cells do not metabolize lactate when it is offered as the sole carbon source.

MFA has also been applied for media optimization. Xing *et al.* performed MFA in continuous culture to assess the metabolic demands (in terms of amino acids) of antibody producing CHO cells [93], which resulted in a modified medium where final concentrations of ammonia and lactate were reduced and higher viable cell densities and higher productivities were achieved.

The steady state assumption might be problematic when modelling the inherently time-dependent fed-batch processes [94]. Hence, several efforts have been made to perform kinetic metabolic analysis while keeping a reduced, tractable set of reactions to avoid dealing with too many kinetic parameters. One of the first attempts in this direction was made by Nolan *et al.* [95], who included kinetic expressions in a reduced, lumped model containing 34 reactions. They studied the metabolic lactate switch by linking glucose concentration in the medium to cytosolic levels of NADH and lactate metabolic rate (lower levels of cytosolic NADH leading to net lactate consumption). This study also analyzed the intracellular concentrations of 24 metabolites in different cell lines and found that 20 of them either remained constant during the process or that their concentration changes were negligible compared to the fluxes, supporting the validity of the pseudo-steady state assumption [74] also for fed-batch processes.

Goudar *et al.* [96] made remarkable progress towards quasi real-time estimation of the metabolic rates in perfusion culture of CHO cells for optimal process control based on metabolite balancing. They observed that reducing the initial concentrations of glucose and glutamine resulted in an increased flux towards the TCA cycle and decreased production of waste metabolites, mainly lactate.

Xing *et al.* [97] applied a Markov chain Monte Carlo method to develop a kinetic model of fed-batch cultures and predicted optimal initial concentrations of glucose and glutamine that minimized the production of ammonia and lactate.

The effects of decreasing concentrations of glutamine in the media, namely the

increased uptake of other carbon sources and the reduction of secreted ammonia and other products, was studied by dynamic MFA on fed-batch CHO cultures with different glutamine concentrations [98]. They show how controlled feeding prevents glutamine metabolism to be coupled to waste producing pathways and, moreover, stabilizes the flux through the TCA cycle.

Similarly, Sheikholeslami *et al.* [99] used ^{13}C MFA to compare two semicontinuous cultures grown on chemically defined media with 1mM and 5mM glutamine, respectively, and found that low glutamine uptake (in the 1mM culture) was more metabolically efficient in terms of the proportion of pyruvate that enters the TCA cycle (and therefore is not converted to lactate). Furthermore, the CHO cell line used in this study was found to be particularly efficient, mostly under hypothermic conditions, as confirmed on their previous work [100]. In this case, the use of ^{13}C MFA was simplified by analyzing only extracellular ^{13}C -labelled metabolites and then performing MFA to predict the intracellular fluxes.

Another interesting feeding strategy was suggested by Naderi *et al.* [101]. In their work, they used MFA to reduce the metabolic network to a set of significant reactions and coupled them to a dynamic cell growth model to asses the differences between growing and apoptotic cells. They highlighted the differences on the metabolic rates for the different cell subpopulations (growing, resting and apoptotic cells) and suggested a feeding strategy based on the "aging" of the cell culture: when glutamine is in excess in late phases of the process (where the non-growing cells become predominant), there is a switch from glycolytic reactions towards deamination of glutamine (and concomitant ammonia accumulation), which could be prevented by gradually lowering the concentrations of glutamine in the feed as the culture ages.

Some other compounds, such as sodium butyrate, have shown to improve productivity in CHO cells [102]; Ghorbaniaghdam *et al.* [103] used a kinetic model to assess the effects of this compound on metabolism in a non-compartmentalized model assuming Michaelis-Menten kinetics. They found cells to become more energetically efficient (in terms of the lactate to glucose ratio) when sodium butyrate was added at the mid-exponential phase. Moreover, they made noteworthy improvements in describing energy metabolism (in terms of ATP) and redox potential (in terms of NADH, NAD^+, NADPH and $NADP^+$). Adding sodium butyrate to the media generates an increased flux through the TCA cycle and a high cell redox potential, while not significantly changing the ATP production rates.

MFA has also been combined with statistical analysis methods (such as principal component analysis) to determine key metabolites linked to the accumulation of ammonia and lactate. In their study, Selvarasu *et al.* [72] analyzed profiles of extracellular and intracellular species and integrated this information in a mouse-derived

GSMR with the goal of finding pathways related to growth limitation. In addition to glucose and glutamine, they identified asparagine to be correlated with the accumulation of ammonia in the medium, most probably via its conversion to aspartate, then glutamate and finally α-ketoglutarate.

1.4.1.4 The future starts now: iCHO1766, a comprehensive, genome-scale metabolic reconstruction of CHO

As outlined above, the results derived from a model-based analysis have significantly improved our understanding of the underlying metabolic processes. This is all the more remarkable as, so far, a truly CHO-specific GSMR was missing. All the applications summarized above used either small-scale metabolic models or adapted reconstructions developed for related organisms like mouse or humans. However, after the complete genomic sequence of CHO-K1 was published in 2011 [3], several research groups around the world joined forces in creating the first community-curated GSMR of CHO, which just now became available [69]. This model consists of 4,455 metabolites participating in 6,663 reactions and contains 1,766 annotated genes. In a first demonstration of possible applications of this CHO GSMR, typical process engineering strategies were analyzed for their effects on the predicted maximum product yield. In all tested cases, the model suggested that these processes are not even close to tapping the full potential of CHO cells.

Furthermore, the transcriptome [104] and proteome [105] of CHO cells can be now used to obtain cell-line specific models that provide a more precise characterization of metabolic capabilities [106]. Metabolomics data can further refine these models to make better predictions under the given culture conditions. Thus, given the advances in high-throughput technology, we expect that the model based-analysis of systems-level data like the transcriptome and proteome will help to further unravel the complexity of CHO metabolism.

Regardless of these promising results, model performance has to be further evaluated. Ever since the first modelling approaches appeared, the accuracy of experimental measurements has been shown to be an important factor to obtain meaningful results [107]. Moreover, it has been shown that biomass composition varies among different cell lines [72]. It is also known that the biomass composition has a great effect on model predictions [108]. Therefore, factors influencing the robustness of CHO metabolic models is a question that still remains to be addressed.

1.4.2 Glycosylation

Modelling metabolism aims at reducing the metabolic burden on the cells induced by the recombinant production of the protein of interest. It aims to increase the

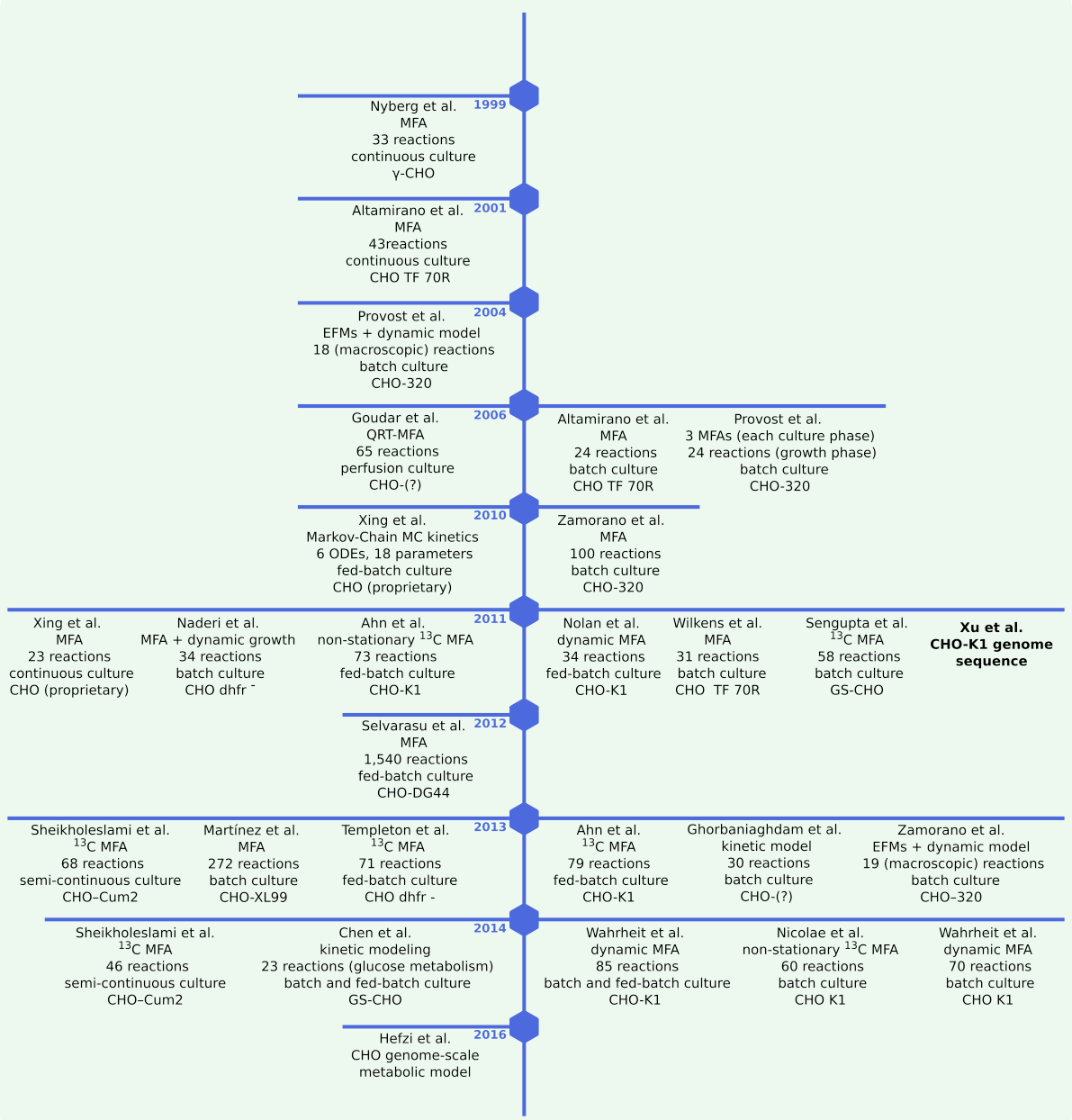


Figure 1.4: Metabolic modelling efforts in CHO listed in chronological order. Abbreviations: QRT, quasi-real-time; dhfr, dihydrofolate reductase

protein yield. However, the biopharmaceutical industry is not only faced with the problem of producing therapeutic proteins efficiently, but also to produce them at high quality. A major quality attribute of many biopharmaceuticals is correct glycosylation, as the correct function of most therapeutic proteins depends on it [109]. Glycosylation consists of the addition of an oligosaccharide chain to an amino acid residue, predominantly asparagine (N-linked) or serine/threonine (O-linked

glycosylation) and takes place in the Endoplasmic reticulum and Golgi apparatus along the protein secretory pathway. These sugar modifications play a fundamental role in protein conformation, stability, solubility, receptor recognition and antigenicity as well as cytotoxicity [110, 111, 112, 113]. Thus glycosylation essentially modifies the pharmacological properties of a protein.

Glycosylation patterns are naturally and in general heterogeneous. There are two main sources of variability in glycosylation: macroheterogeneity, which refers to the fact that a particular site in the protein might or might not be glycosylated; and microheterogeneity, when different glycan structures can be found on the same site. However, this natural variability presents a particular challenge for the production of biosimilars, were the glycosylation patterns of the primary drugs have to be reproduced within tight tolerance regions defined by regulatory authorities.

1.4.2.1 Modelling glycosylation in CHO

Many factors are known to influence glycosylation in cell culture: concentration of metabolites in the medium (both substrate and waste products), pH, temperature and cell viability [114, 115]. The mechanisms by which these factors affect micro- and macroheterogeneity remain, however, unclear. Thus a systematic analysis is called for. Computational modelling provides a powerful framework for such an analysis. In fact, there have been remarkable advances in the development of mathematical models of glycosylation (listed chronologically in Figure 1.5), supported by the detailed knowledge of the glycosylation pathways [116]. Generally, these models aim to reduce the combinatorial explosion in the number of possible glycan distributions. To this end, models make some general assumptions, while keeping compartmentalization (each compartment is modelled differently since they contain different sets of enzymes) and finally linking glycosylation to metabolism. The complexity of the process, together with the many intervening factors, makes modelling glycosylation quite a challenging task.

One of the first attempts to deterministically describe protein glycosylation focused on macroheterogeneity. In 1996, Shelikoff *et al.* [117] proposed a mathematical model to predict how site-occupancy is affected by different factors such as the expression levels of glycotransferases, the protein production rate, the concentrations of nucleotide sugars and the mRNA elongation rate. They used a plug-flow reactor-based model and included protein folding as a competing event that occurs concurrently with glycosylation.

Shortly after, Monica *et al.* [118] modeled sialylation of N-linked oligosaccharides in a single, isotropic compartment (trans-Golgi). The predictions were in agreement with experimental data of CD4 glycoprotein produced in CHO cells.

Umaña and Bailey (1997) [119] presented the first attempt to model glycoform microheterogeneity based on expression and spatial localization of the enzymes involved in N-linked glycosylation. Parameters such as the half-life of the protein in the Golgi, the protein productivity and the volume of the Golgi compartments were also included in this model. Furthermore, they modified the model to take the competition for the glycosylation machinery between endogenous and recombinant proteins into account. Kontoravdi *et al.* used this model of glycosylation and included it in a simple dynamic mathematical model of cell growth, death and metabolism. With this reduced model they predicted the evolution of oligosaccharide molar fractions over time. However, these results could not be validated due to the lack of experimental data [120].

Several years later, in 2005, Krambeck and Betenbaugh [121] extended Umaña's model (which contained 33 glycan structures and 33 reactions), by adding around 7,500 oligosaccharide structures and more than 22,000 reactions. Among these, reactions for fucosylation and sialylation were included in the model, which are of special relevance for recombinant proteins [122, 123]. In contrast to the model of Umaña and Bailey, this model adjusts enzyme concentrations to fit an experimentally observed glycopattern, thereby calibrating it to a specific protein. They argue that the reason for having a case-specific, adjusted model is the inherent variability of glycosylation: the glycan structures do not only depend on the specific protein, but also on the glycosylation site. Their results were validated with N-glycan structures observed in recombinant human thrombopoietin expressed in CHO cells [124]. This model was then used as a prototype for further development by other research groups.

In 2009, Krambeck *et al.* applied the previously developed model to predict enzyme expression that resulted in an observed mass spectrometry spectrum. Reciprocally, the model was used to automatically annotate spectra to the corresponding glycan structures [125].

Both models (Umaña and Bailey, Krambeck and Betenbaugh) were combined in two different studies to predict the sensitivity of N-Glycan branching with respect to the hexosamine flux [126] and key enzymes involved in glycan branching [127].

Senger and Karim [128] used a plug-flow reactor model to describe the differences in glycosylation of recombinant tissue plasminogen activator in CHO under shear stress conditions. They found decreased site occupancy to be related to low residence times of the protein in the Endoplasmic reticulum due to high protein production rates, caused by increasing levels of shear stress.

In a follow-up study, Senger and Karim used artificial neural network models to predict glycosylation from primary sequence information around the glycosylation site (glycosylation window). The model was used to classify macroheterogeneity as ei-

ther robust (invariant with culture conditions) or variable, according to this sequence information [129]. They improved this approach further by using information about the secondary structure and solvent accessibility, resulting in the prediction of two main types of glycan branching: high mannose type and complex-type [130]. Artificial Neural Networks had already been applied to predict glycosylation sites [131, 132]. The complexity of the impact of protein conformation in the surroundings of the glycosylation site on glycotransferase activity hinders the creation of a mathematical model that could describe the process deterministically. Therefore, they presented the Neural-Network approach as a valuable workaround to construct prediction tools. The main advantage of this approach with respect to the previous models is that it does not require a large number of parameters, but only the protein sequence (from which they predict the secondary structure). In addition, it highlights the influence of protein secondary and tertiary structure on the accessibility of the enzymes. In another instance, Gerken *et al.* [133] considered the inhibitory effect of the presence of glycan structures on neighboring sites of glycosylation.

Built on the premise that glycan biosynthesis is controlled by the expression of glycotransferases, Kawano *et al.* [134] predicted a set of glycan structures from DNA microarray data. This set was further expanded by Suga *et al.* [135] with the prediction of new structures (Kawano's set of predicted glycans was limited to those included in the database of known structures). This approach was refined several years later with high-throughput RNA microarray data [136].

Hossler *et al.* [137] compared the prediction performance of two main models for protein maturation in the Golgi: four continuous mixing-tanks (4CSTR) for vesicular transport and four plug-flow reactors (4PFR) in series for the maturation model. They claimed that the latter describes the process more accurately and they emphasised the importance of the residence time in the Golgi and enzyme localization as key parameters to be considered when modelling glycosylation.

The plug-flow reactor model was then used to describe monoclonal antibody (mAb) glycosylation [138]. The major improvement over the previous model was to include the transport of nucleotide sugar donors. This was the first step towards coupling cellular metabolism (and therefore measurable variables like glucose uptake) to glycosylation. Kaveh *et al.* [139] pursued this goal and performed a dynamic analysis of extracellular metabolite concentrations via MFA and linked those of glutamine and glucose to nucleotide sugar biosynthesis and glycolysis using the previous models (del Val 2011 [138] and Hossler 2007 [137]). The model successfully predicted dynamic trends of the glycopatterns of mAb produced in CHO batch culture. In another study [140], they combined dynamic MFA with the GLYCOVIS software developed by Hossler *et al.* [141] to predict, based on experimentally observed glycopatterns, how

different concentrations of glutamine, glucose, ammonia and different pH values affect the glycosylation process. Yet more progress was made by Jedrzejewski *et al.* [142], who used a dynamic model for cell death and growth together with the dynamic model from del Val [138] to predict glycosylation patterns. In this case, experimental data from mAb producing mouse hybridoma cells was used for the calculations. A similar study was applied to mAb producing CHO fed-batch cultures [143]. As a result, recent models have succeeded in linking cell growth, metabolism, protein production rate and glycosylation [144].

The majority of these models describe N-glycosylation. Liu *et al.* [145] presented a reaction network for the formation of the O-glycosylation of the sialyl Lewis-X epitope. In their work, they introduce the concept of "subset-modelling", where the whole set of reactions in the network is divided into "sub-networks" and then a search is performed for the one that fits the experimental data best. Furthermore, they use genetic algorithm-based optimization, hierarchical clustering and principal component analysis to fit subsets of reaction networks to the observed glycan structure distribution, thereby reducing the parameterisation of the model. Recently, the same group developed a software for the automated creation, analysis and visualization of glycosylation reaction networks, called GNAT (Glycosylation Network Analysis Toolbox) [146, 147]. GNAT was further expanded to include a higher number of enzymes [148].

Kim *et al.* [149] also exploited the modularity of the glycosylation pathways to propose new engineering strategies based on targeting modules instead of specific enzymes.

In a simpler approach, FBA was applied to assess the effect of low temperature conditions on metabolism and nucleotide sugar availability for glycosylation in mAb producing CHO cells [150]. A similar MFA-based method was applied to analyze the effects of different concentrations of glutamine in the media on nucleotide sugar intracellular concentrations and N-glycan content of recombinant human chorionic gonadotrophin in CHO cells [151].

In the past year, a simple stoichiometric model was also used to compute the nucleotide sugar demands for glycosylation of recombinant proteins in CHO for rational feeding strategies [152].

In order to avoid the requirement of a high number of kinetic parameters, Spahn *et al.* [153] used a Markov chain model to describe glycosylation as a stochastic process in which each glycan state has a transition probability to reach the next glycan state. These probabilities are linked to the steady state solution given by FBA for a reduced network of the reactions contributing to the observed glycoprofile. By using this protein-specific model, they successfully predicted the effect of an enzyme

knock-down on an antibody producer CHO cell line [154].

1.4.3 Parameters and general assumptions

The parameters involved in glycosylation include reaction kinetic parameters, compartment residence times, enzyme distributions between compartments, compartment volumes, total glycan concentration and donor cosubstrate concentrations. These parameters are either obtained via optimization or taken from literature [155]. Imaging techniques for green fluorescent protein-labelled proteins can be used to measure residence time and protein flux through the secretory machinery [156]. Kinetic parameters are commonly derived from independent enzymology experiments [157], which are arduous and should be carried out for each enzyme. However, there have been remarkable advances on high-throughput technologies that allow more accurate assessment of kinetic parameters of glycosyltransferases [158].

Due to the sequential nature of glycosylation, models have to incorporate time-dependent equations. The majority of the kinetic models reviewed herein assume Michaelis-Menten Kinetics. Over time, more terms were included in these models' equations, with increasing complexity, e.g. competitive inhibition terms in their enzyme-kinetic expressions.

The main limitation of glycosylation models is the high grade of parameterisation required to describe the process. Moreover, most of the parameters are derived from in vitro experiments, even though they might be different in an intracellular environment. As previously mentioned, various factors influence glycosylation at different points of the process [115] and the effects are cell line [159], glycoprotein [160] and even glycosylation site specific [119], which reduces the general applicability of the models. Thus, despite the tremendous advances achieved over the last years in this field, the ultimate goal of predicting the effect of cell line specific behaviour of different protein sequences or structures, or of process related changes, on glycosylation still requires further work and optimisation to be fully achieved.

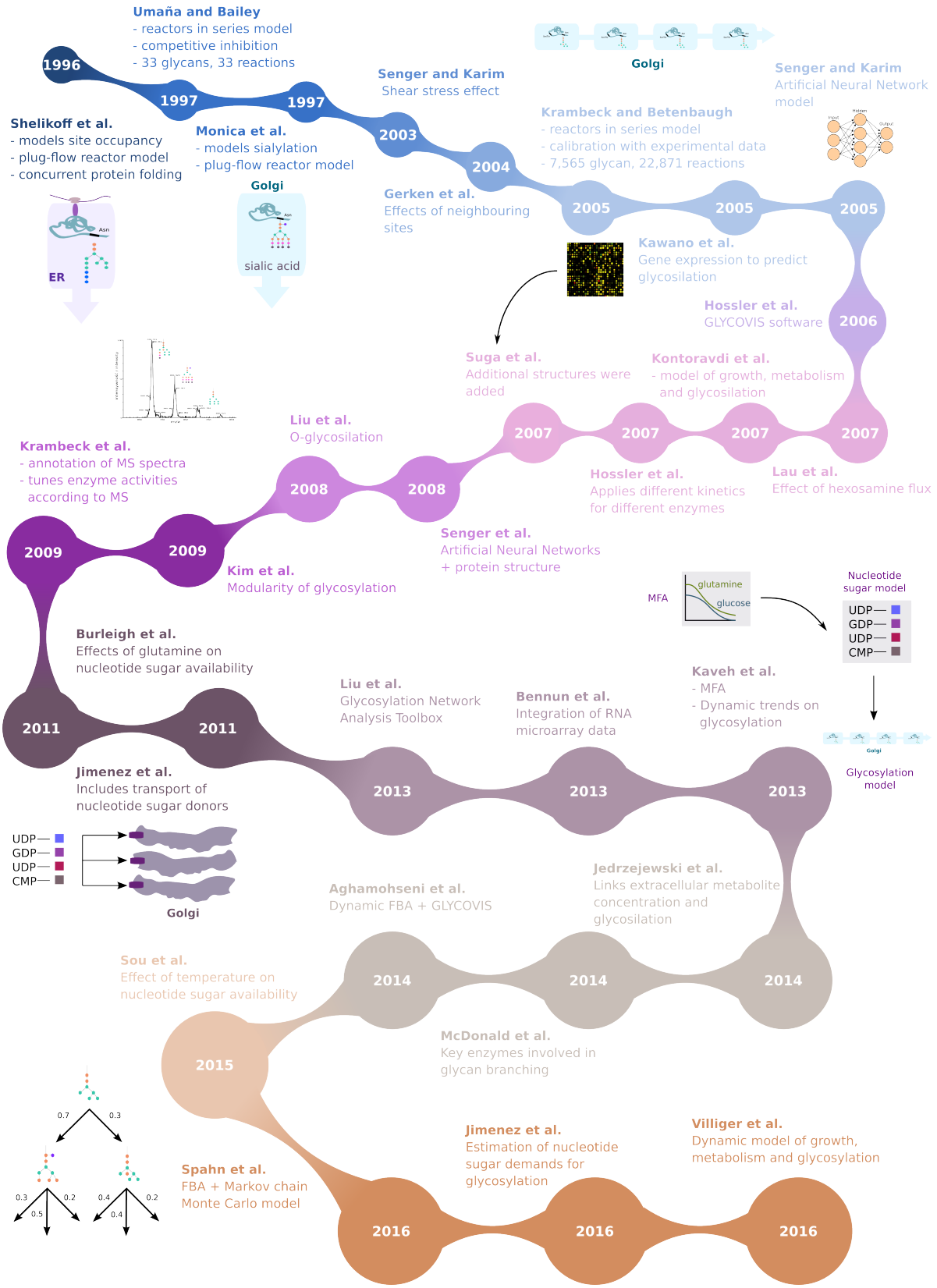


Figure 1.5: Models for protein glycosylation in CHO listed in chronological order. Abbreviations: MS, Mass-spectrometry;

CHAPTER 2

MATERIALS AND METHODS

2.1 Constraint-based analysis of metabolic networks

Cell metabolic behavior is limited by a multitude of constraints which reduce the set of potential phenotypic states of an organism [161]. These constraints are determined by genetic, physico-chemical laws (thermodynamics, maximum diffusion rates, mass balances) and environmental factors (nutrient and oxygen availability) and can be represented mathematically as (in)equalities for the reaction fluxes. In the following sections, we summarize the basic mathematical concepts to represent and analyze metabolic networks.

2.1.1 Mathematical principles of constraint-based modelling

Stoichiometric matrix

The accumulation of any given metabolite in the network depends on its flow into or out of the compartment (exchange reactions) and reactions generating and consuming the metabolite. This can be described as:

$$\frac{d\boldsymbol{c}}{dt} = \boldsymbol{N}\boldsymbol{r} - \mu\boldsymbol{c} \tag{2.1}$$

where $\boldsymbol{c}$ is the vector of intracellular concentrations, $\boldsymbol{r} \in \mathbb{R}^r$ the vector of reaction fluxes and μ is the growth rate. $\boldsymbol{N}$ is a matrix containing the net stoichiometric coefficients for each metabolite and has dimensions $m \times n$ (see Figure 2.1 for an example of the stoichiometric matrix for a small network). The dilution due to growth is generally small, so the term $\mu\boldsymbol{c}$ can be neglected.

In this way, the stoichiometric matrix translates the biochemical reactions into a mathematical term in which computational methods can be developed.

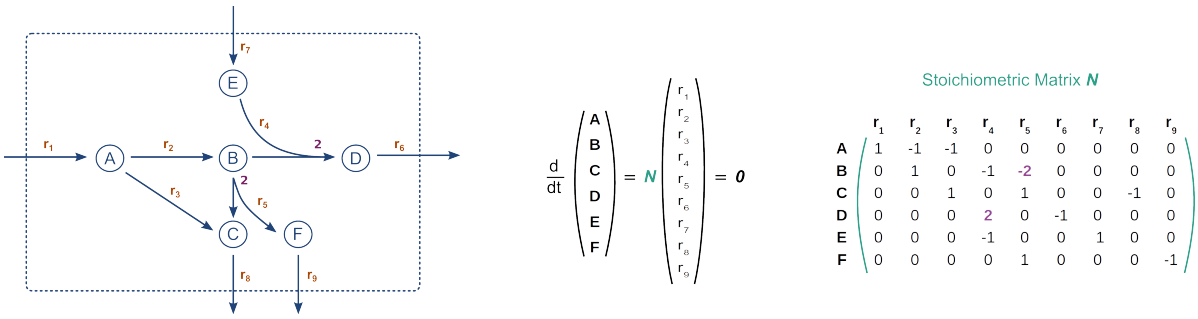


FIGURE 2.1. Small metabolic network and the corresponding stoichiometric matrix. Dashed line represents the system's boundary. All metabolites inside this boundary are in steady state.

The flux polyhedron

Since the number of metabolites is typically less than the number of reactions (i.e., the number of mass-balance equations is less than the number of unknowns), the system is under-determined, that is, there exists an infinite number of solutions. In order to reduce this solution space to a meaningful set of feasible flux distributions, one has to impose additional constraints.

If we consider that fluxes through metabolites in the cell are much faster than external environmental changes (quasi-steady state assumption), all generating and consuming fluxes (multiplied by the stoichiometric coefficient) for each internal metabolite are assumed to be in equilibrium. This is represented as:

$$\boldsymbol{N}\boldsymbol{r} = \boldsymbol{0} \tag{2.2}$$

The set of reaction fluxes that satisfy the linear system (2.2) is the nullspace of $\boldsymbol{N}$. If additional information about reversibilities or enzyme allocation is considered, we obtain a subset of this nullspace. The resulting solution subspace is given by intersecting the non-negative half-spaces given by reversibility constraints ($\boldsymbol{r}_{irr} \geq 0$) with the nullspace given by (2.2):

$$FC = \{\boldsymbol{r} \in \mathbb{R}^n \mid \boldsymbol{N}\boldsymbol{r} = \boldsymbol{0}, r_i \geq 0 \text{ for } i \in irr\} \tag{2.3}$$

In geometrical terms, this corresponds to a convex polyhedral cone. In the specific context of metabolic networks, it is generally referred to as the flux cone (FC).

In some cases, additional constraints on upper and lower flux bounds are also known (e.g. experimental uptake and secretion rates). They can be written as:

$$r_i^{\mathrm{l}} \leq r_i \leq r_i^{\mathrm{u}}$$

Which gives:

$$FP = \{\boldsymbol{r} \in \mathbb{R}^n \mid \boldsymbol{N}\boldsymbol{r} = \boldsymbol{0}, \boldsymbol{r}_{irr} \geq 0, \boldsymbol{G}\boldsymbol{r} \geq \boldsymbol{h}\} \tag{2.4}$$

Where $\boldsymbol{G}$ is a matrix with coefficients for the reactions that are constrained by lower (coefficient +1, corresponding to $\boldsymbol{r}_i \geq \boldsymbol{r}_i^{lb}$) and/or upper bounds (coefficient -1, corresponding to $-\boldsymbol{r}_i \geq -\boldsymbol{r}_i^{ub}$). Vector h contains the upper and lower bounds on the rates. Equivalently, equation (2.4) can be written in matrix form as:

$$\begin{pmatrix} \boldsymbol{N} \\ -\boldsymbol{N} \\ \boldsymbol{I}_{irr} \\ \boldsymbol{G} \end{pmatrix} \boldsymbol{x} \geq \begin{pmatrix} 0 \\ 0 \\ 0 \\ \boldsymbol{h} \end{pmatrix}$$

These inhomogeneous constraints reduce the set of admissible fluxes from a cone to a polyhedron. In the specific case where a polyhedron is bounded, it is said to be a polytope (Figure 2.2).

A polyhedron can be described by the intersection of half spaces [as in (2.4)] or by linear combination of generators. If the flux polyhedron is pointed, which is often the case for metabolic models [162], it can be shown that the set of extreme points (EPs) and extreme rays of the polyhedron are a unique and minimal set of generators [163]. In this case,

$$FP = \{\boldsymbol{r} \in \mathbb{R}^r \mid \boldsymbol{r} = \boldsymbol{P}\boldsymbol{\lambda} + \boldsymbol{R}\boldsymbol{\mu},\ \boldsymbol{\lambda}, \boldsymbol{\mu} \geq \boldsymbol{0},\ \textstyle\sum \lambda_i = 1\}, \tag{2.5}$$

can be represented as a convex combination of EPs (represented by the columns of $\boldsymbol{P}$) plus a conical (i.e. nonnegative) linear combination of extreme rays (contained in the columns of $\boldsymbol{R}$).

Geometrically, EPs and extreme rays are the vertices of the bounded part of the polytope and edges of the unbounded part of the polytope, respectively. Thus, any element of a pointed polytope can be represented by a non-negative linear combination of its boundary elements. For instance, the point $(2,1,1)$ represents a feasible flux distribution in the toy network in Figure 2.3a. It sits in the flux polyhedron and can be represented as $\boldsymbol{e}_0/3 + \boldsymbol{e}_4/2 + \boldsymbol{e}_5/6 = (2,1,1)$, i.e., as a convex combination of the EPs listed in Figure 2.3c. However, this representation is in general not unique. Note that the flux polyhedron in 2.3b is bounded and therefore does not contain extreme rays.

An alternative set of generators, which are often used in biological applications, are known as elementary flux vectors (EFVs) [162]. EFVs are the proper generalization of the classical concept of elementary flux modes [164], which cannot handle linear, inhomogeneous constraints in metabolic models. Here we only note that the set of EFVs is a super-set of the set of EPs and extreme rays [163]. For instance, the point $\boldsymbol{e}_3 = (2,0,2)$ sits at the edge of the polytope in Figure 2.3b and is an EFV of the metabolic network in Figure 2.3a, but not an EP as it can be represented as a combination of two vertices, i.e. $(\boldsymbol{e}_3 = \boldsymbol{e}_2 + \boldsymbol{e}_4)/2$.

With a set of generators at hand, projecting the flux polyhedron onto the reactions of interest becomes straightforward, as the production envelope (PE) is simply the convex hull of the projected generators. However, enumeration of generators, more specifically the enumeration of EFVs in genome-scale metabolic models, is a long standing problem in systems biology, and currently computationally intractable as the total number of generators grows combinatorially

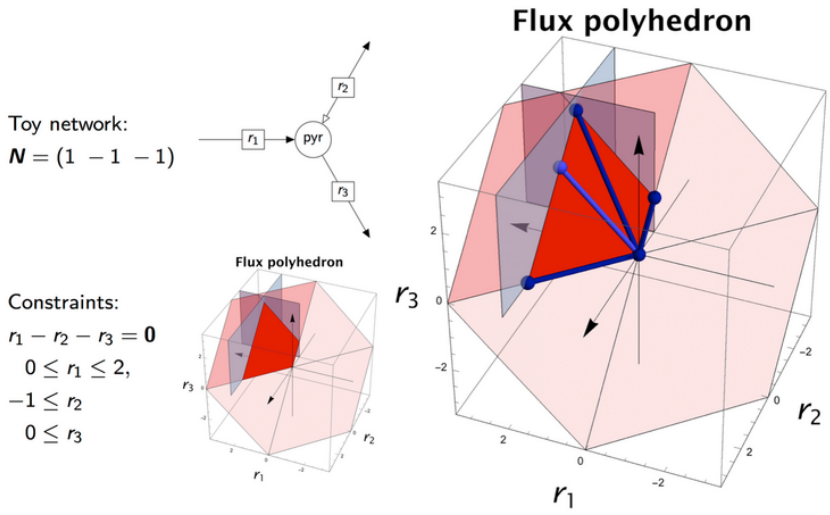


FIGURE 2.2. Reduction of the solution space to a flux polyhedron by linear constraints in a small example network. In red, the constraint space of feasible solutions. Pyr: pyruvate

with the size of the network [165]. To circumvent this issue, we adopt the convex hull method (CHM) (section 3.2.2) to compute the PE spanned by multiple reactions of interest in a genome-scale metabolic model.

The flux polyhedron represents all the feasible phenotypic states of the organism. The potential optimal cell behavior lies within this polyhedron, and so does the actual cellular state of the organism.

Constraints reduce the amount of possible flux distributions or phenotypic capabilities of the organism. Methods to analyze this set of biochemically "allowed" fluxes are divided in two main groups: the *biased* approaches, in which linear programming is applied to obtain a set of optimal fluxes and *unbiased* approaches, which aim to analyze the whole set of steady-state flux distributions [106]. In the following, we include a list of constraint-based methods (biased and unbiased) that we have applied for the analysis of CHO metabolic model.

2.2 Biased approaches

2.2.1 Flux Balance Analysis

Flux balance analysis (FBA) uses linear optimization to find a steady-state flux distribution within the flux polytope that gives an optimum for a given objective function [76]. If the objective

function is also linear, this corresponds to a linear programming (LP) problem, expressed as:

$$\begin{aligned} &\text{Maximize} && \textstyle\sum c_i r_i \\ &\text{s.t} && \boldsymbol{N}\boldsymbol{r} = \mathbf{0} \\ &&& r_{irr} \geq 0 \\ &&& r_i^{\mathrm{l}} \leq r_i \leq r_i^{\mathrm{u}} \end{aligned} \tag{2.6}$$

The most common objective function used in metabolic modelling is cell growth, which has proven to be a compelling assumption mostly for bacteria and yeast [166, 167], the argument being that, evolutionary speaking, organisms tend to maximize biomass. Other valid objective functions are minimization of ATP production, maximization of metabolite production (such as ethanol for *E. coli* or protein for CHO), minimization of nutrient uptake or minimization of redox potential [168].

Despite its growing popularity, FBA has its limitations: there are multiple possible flux distributions that result in the same objective optimum; FBA is not able to predict flux distributions in the case of parallel metabolic pathways. Other quantitative approaches such as ^{13}C Metabolic Flux Analysis (MFA) [74] can be employed to estimate intracellular fluxes. Yet the complexity of these methods in terms of experimental design and data analysis exclude genome-scale metabolic models from the scope of application [169]. Other approaches solve this issue by minimizing the overall flux through the network (see section 2.2.3 below). Omics data can also be used to obtain flux distributions consistent with experimental observations: this is the case of the Gene Inactivity Moderated by Metabolism and Expression (GIMME) algorithm, in which reactions are considered to be inactive if gene expression values lie below a given threshold [170].

FBA is the oldest and most commonly used method for constraint-based analysis of metabolic networks for its simplicity and utility [106], and provides the basis of the succeeding computational tools.

2.2.2 Flux Variability Analysis

In flux variability analysis (FVA), a particular metabolic state is fixed (by constraints on other reactions) and fluxes are maximized and minimized [171]. Given a set of uptake and secretion rates, FVA gives the feasible flux range for a reaction under those conditions. For this reason it can be used to assess the metabolic capabilities of the constraint network, by sampling the edges of phenotypic space (Phenotypic Phase Plane Analysis - see section 2.3.2). For instance, the maximum and minimum protein production can be computed for fixed combinations of biomass and ATP fluxes. Applications of FVA range from exploring alternative optima [172] to analyzing network robustness [173] and redundancy [174].

2.2.3 Parsimonious Flux Balance Analysis

As already mentioned, there are multiple flux distributions that satisfy maximal objective function in (2.6). FBA solution is one (arbitrary) flux distribution within this set. To address this degeneracy, parsimonious flux balance analysis (pFBA) performs a step-wise linear programming optimization: first, maximization of the objective function (as in 2.6), then this maximum, e.g. biomass, is set as constraint and a second LP is performed in which all the fluxes in the network are minimized. In the general implementation of the algorithm, the Manhattan Norm is minimized. Since in our case some fluxes can be negative, we minimized the euclidean norm instead and solved the following quadratic problem:

$$\begin{aligned} \text{Minimize} \quad & \sum r_i^2 \\ \text{s.t} \quad & \boldsymbol{N}\boldsymbol{r} = \mathbf{0} \\ & r_{irr} \geq 0 \\ & r_i^{\text{l}} \leq r_i \leq r_i^{\text{u}} \\ & r_{bm} = max(r_{bm}) \end{aligned}$$

pFBA is based on the concept of parsimonious enzyme usage [175], which assumes that the cell will attempt to retrieve its optimal state (maximum growth in our case) while minimizing the enzymatic cost (activity), therefore minimizing the metabolic flux through the network.

2.2.4 Flux scanning based on enforced objective flux

Flux scanning based on enforced objective flux (FSEOF) is one of the constraint-based methods specifically developed for strain design [106]. Other approaches of this group are OptKnock, which aims to find deletion strategies to couple metabolite production with cell growth [176]; RobustKnock, which improves the former by considering alternate optima [177]; or Optforce [178], which compares experimentally measured fluxes of wild type Vs engineered strains for overproduction of a given metabolite in order to identify reactions consistent with the high producer phenotype.

In FSEOF, the objective function is maximized in successive LPs in which the flux towards the product of interest (POI) is gradually (ϵ) increased from the experimental value to 90% of its theoretical maximum (since theoretical maximum from FBA would be unrealistic and would only occur at the expense of zero growth). FSEOF has been primarily applied for *in silico* identification of metabolic targets [179].

$$
\begin{aligned}
\text{Minimize} \quad & \sum r_{bm} \\
\text{s.t} \quad & \boldsymbol{N}\boldsymbol{r} = \mathbf{0} \\
& r_{irr} \geq 0 \\
& r_i^{\text{l}} \leq r_i \leq r_i^{\text{u}} \\
& r_{POI} = r_{POI} + \epsilon
\end{aligned}
$$

Those fluxes which decrease with increasing protein production will be suitable candidates for knockout or knock down, whereas those increasing will be considered for overexpression.

2.3 Unbiased approaches

As already mentioned, the set of inhomogeneous constraints form the flux polytope, which is a convex set of points (for this and other definitions related to convex analysis, see Box 2.3.1). Two main methods for the analysis of this convex polytope have been developed in the past decades: elementary modes and extreme pathways. Both methods aim to explore the full capabilities of the metabolic network by applying concepts of computational geometry.

2.3.1 Elementary Flux Modes

Elementary flux modes (EFMs) are topologically feasible, minimal sets of reactions that support a steady-state [180]. In other words, each EFM is defined by a set of active reaction that satisfy (2.3). The set of all EFMs is a generating set of the flux cone, i.e., every flux distribution can be expressed as a non-negative linear combination of EFMs. This approach analyzes the metabolic capabilities of the network, both in terms of rates or yields [181] without introducing any biased optimization process. Other applications are the quantification of the network robustness [182] and the optimization of microbial strains [183]. Nevertheless, combinatorial explosion when enumerating EFMs has limited the application to small to medium scale metabolic networks [164, 184].

Elementary Flux Vectors (EFVs) were introduced by Urbanczick [185] and aim to extend the concept of EFMs from flux cones to general flux polyhedra. Analogously to the cone and EFMs, thorough examination of the flux polyhedra can be done by enumerating all EFVs or by random sampling [186]. As for EFMs, this computation is restricted to small scale networks.

Box 1: basic concepts

- **Convex set**. Given points $x_1, x_2, \ldots, x_m \in \mathbb{R}, \sum_{i=1}^{m} \lambda_i x_i$ is said to be a convex set if
  - $\sum_i \lambda_i = 1$ (it is affine)
  - $\lambda_i \geq 0$ (it is conical)
- **Convex hull** of a set of points S is the smallest convex set containing all the points.
- **Hyperplane**: set of all points $\boldsymbol{x} \in \mathbb{R}^n$ that satisfy $\boldsymbol{a}^T \boldsymbol{x} = b$ for some $\boldsymbol{a} \in \mathbb{R}^n$ and $b \in \mathbb{R}$.
- **Half-space**: either of the two parts into which a hyperplane divides an affine space, i.e., the set of all points $\boldsymbol{x}$ such that $\boldsymbol{a}^T \boldsymbol{x} \leq b$ for some $\boldsymbol{a} \in \mathbb{R}^n$ and $b \in \mathbb{R}$.
- **Polyhedron**: is the intersection of finitely many half-spaces, i.e., the set $\{\boldsymbol{x} \in \mathbb{R}^n | \boldsymbol{A}\boldsymbol{x} \leq \boldsymbol{b}\}$ for a matrix $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ and a vector $\boldsymbol{b} \in \mathbb{R}^{m \times 1}$. It can also be written as a sum of a polytope Q and a cone C, i.e., $P = Q + C = \{\boldsymbol{x} + \boldsymbol{y} \mid \boldsymbol{x} \in Q, \boldsymbol{y} \in C\}$.
- **Polyhedral cone**: the intersection of finitely many half-spaces that contain the origin, i.e., $\{\boldsymbol{x} \in \mathbb{R}^n | \boldsymbol{A}\boldsymbol{x} \leq \boldsymbol{0}\}$ for a matrix $\boldsymbol{A} \in \mathbb{R}^{m \times n}$.
- **Polytope**: is a bounded polyhedron. Alternatively, it can be defined as the convex hull of a finite set of points.

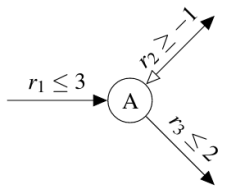
2.3.2 Phenotypic Phase Planes

Section 2.3.2 is part of *Fast computation of multi-dimensional production envelopes in genome-scale metabolic models* by Sarah N.Galleguillos, Matthias P. Gerstl, Norbert Auer, Nicole Borth and Jürgen Zanghellini to Scientific Reports

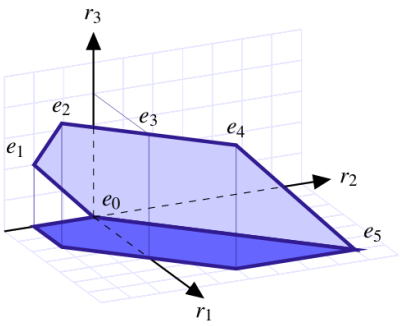
Phenotypic phase planes, sometimes called production envelopes, are an important tool in the constraint-based analysis of metabolic networks. Geometrically, they correspond to the projection of the flux polyhedron onto the reactions of interest. They allow characterization of the full metabolic capabilities of an organism as a function of a subset of reaction fluxes. Phase planes are most often evaluated for only two reactions of interest (typically a product of interest as function of growth), as the computational work load scales exponentially with the number of selected fluxes.

Consider the toy metabolic network in Figure 2.3a. One reversible and two irreversible reactions produce or consume a metabolite A. The network is in steady state ($r_1 - r_2 - r_3 = 0$) and the enzyme capacities of the reactions are limited (flux values are bounded). These constraints form the flux polyhedron, depicted as the light blue area in Figure 2.3b. All feasible flux distributions sit in this area only. If we want to know all feasible combinations of r_1 and r_2, then we are interested in the

(a)



(b)

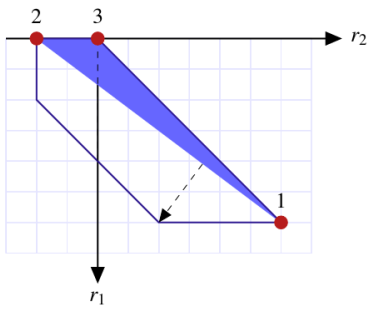


(c)

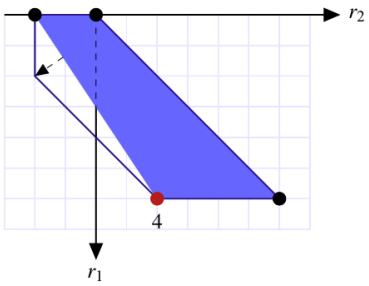
EFV	r_1	r_2	r_3	EP
e_0	0	0	0	yes
e_1	0	-1	1	yes
e_2	1	-1	2	yes
e_3	2	0	2	no
e_4	3	1	2	yes
e_5	3	3	0	yes

FIGURE 2.3. Constraint toy network (panel a), its flux polytope along with its projection onto the plane r_1, r_2 (panel b), and a list of EFVs and EPs (panel c) of the flux polytope.

(a)



(b)



(c)

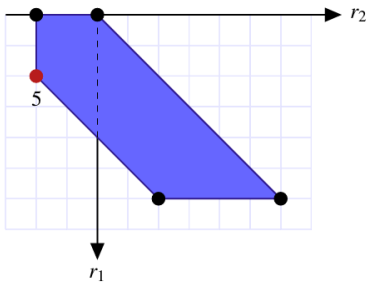


FIGURE 2.4. Iterations in the convex hull algorithm. The minimal set of extreme points (red circles) for the initial hull are computed for the first two dimensions r_1 and r_2 (A). In the following iterations (B and C), hyperplanes (lines) containing these vertices are mimimized/maximized (indicated with dashed black arrows) until all facets of the projection are terminal, i.e., optimization does not lead to a new extreme point.

projection of the flux polyhedron onto the plane spanned by those reactions, that is, the PE of r_1 and r_2 (see dark blue area in Figure 2.3b). Note that (i) the projection of a polyhedron is also a polyhedron and (ii) the components of an element in the PE do not necessarily add up to zero (here $r_1 + r_2 = r_3 \geq 0$) although r_1 and r_2 are components of a feasible steady-state flux distribution of the full network.

Typically, PEs are computed by sampling their boundaries [187] using FVA [188, 189]. This requires fixing all reactions of interest except for one, whose feasible

range is then determined by minimizing and maximizing that reaction by FBA. Although two- and three-dimensional PEs can be readily evaluated with standard software packages [190], the calculation of higher dimensional PEs becomes quickly intractable as the computational effort grows with p^{n_R}, where n_R and p denote the number of reactions of interest and the number of sampling points per reaction, respectively. Furthermore, sampling approaches do not facilitate a compact representation of a multi-dimensional PE.

In section 3.2.2, we present a new approach for the fast calculation of PEs. This approach exploits geometric features of the bounded flux polyhedron and focuses exclusively on the projection onto a particular set of dimensions. We have successfully applied this method for the computation of phenotypic spaces for three metabolic networks: *Escherichia coli* core (519 metabolites and 499 reactions) and genome-scale (1805 metabolites and 2583 reactions) and CHO cells (4456 metabolites and 6663 reactions). Compared to the current approach based on FVA, this method provides a simple and fast way to compute phenotypic spaces, broadening the scope of application to genome-scale metabolic models.

CHAPTER 3

RESULTS

3.1 Biased approaches

3.1.1 FBA predictions on cell growth

CHO-K1 cells previously adapted to grow in glutamine-free media were cultivated in 0mM and 8mM Glutamine CD-CHO medium in shaking flasks. Batch fermentations were performed in-house and metabolite concentrations were measured in two independent facilities (Biocrates and department of chemistry (DCH) of the University of Natural Resources and Life Sciences) in one and three technical replicates respectively.

The resulting metabolic uptake and secretion rates were subsequently used as constraints on the metabolic model to predict cell growth with FBA.

3.1.1.1 Estimation of metabolic rates

First, I established at which time points the cell is at metabolic steady steady state, choosing cell growth as a reference. Figure 3.1 shows the Viable Cell Density (VCD) in a semi-log scale for both 0mM and 8mM glutamine batch experiments. Time points were added or removed from the exponential phase when the standard error of the regression did not increase significantly. The resulting time intervals (data points 3-5 in 0mM and 1-5 in 8mM) were used for the computation of metabolic rates.

Since fluxes during exponential phase are in a near steady-state (balanced growth), the uptake and secretion rates are constant and can be derived by fitting an exponential function. First, I obtain the growth rate μ and the initial biomass BM_0 from the VCD by exponential

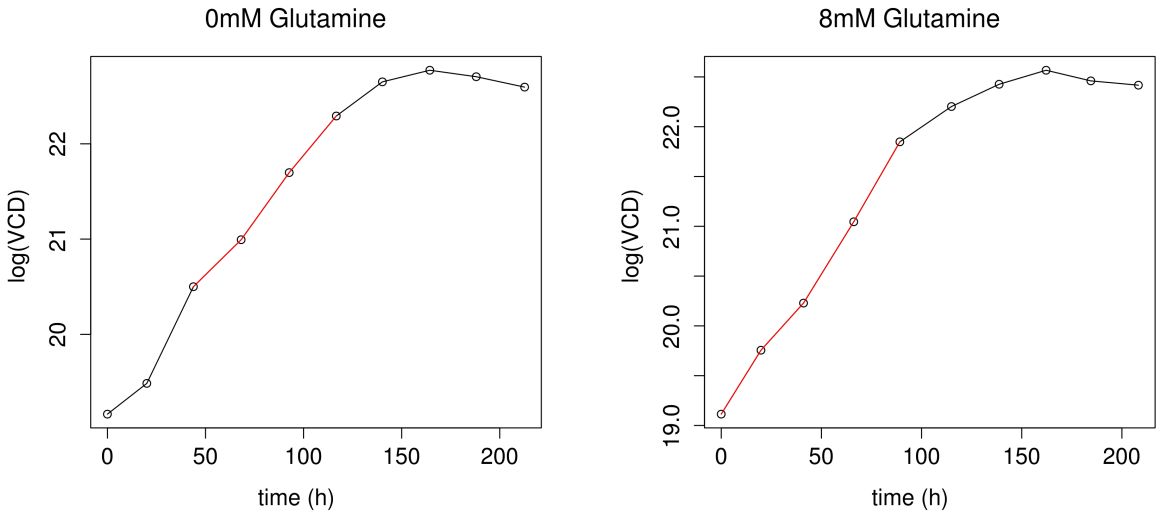


FIGURE 3.1. Viable cell density (VCD) of CHO-K1 cells in 0mM (left) and 8mM (right) glutamine media. The exponential growth phase is shown in red.

fitting:

$$BM = BM_0 \exp(\mu t) \tag{3.1}$$

These parameters are used to compute the metabolic rates after exponential fitting of the concentration profiles:

$$S = \frac{q_S X_0}{\mu}\left(\exp(\mu t) - 1\right) + S_0 \tag{3.2}$$

In order to test the metabolic model of CHO with in-house experimental data, I run FBAs for each set of uptake and secretion rates (Biocrates and DCH). Cysteine (not measured) is an essential amino acid for CHO and uptake was chosen to be the minimum rate to support experimental growth. Oxygen uptake rate is 1.41 mmol/gDW/h for both data sets (value taken from similar batch experiments performed at our collaborators' facilities).

The metabolic model iCHO1766 was downloaded from the BiGG database [52]. Linear optimization problems with biomass production as the objective function were solved using GLPK [191].

Ammonia and lactate production decrease in culture when cells are grown in 0mM glutamine medium. In the presence of glutamine, aspartate is secreted since it is one of the byproducts of glutaminolysis. On the other hand, when glutamine is not present, aspartate and glutamate are consumed from the medium (see appendix A for the uptake and secretion rates).

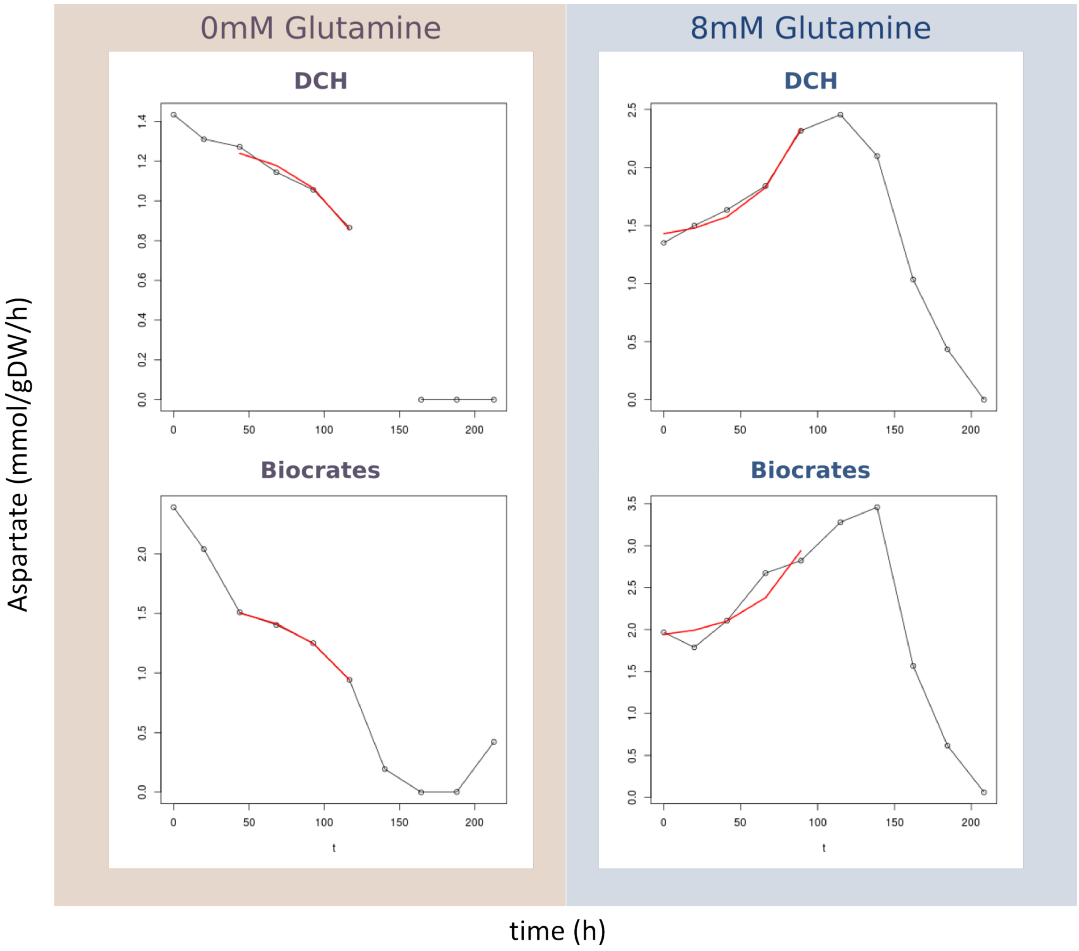


FIGURE 3.2. Concentration profiles for aspartate in 0mM (left) and 8mM glutamine (right) experiments for DCH (upper plots) and Biocrates (bottom). Exponential fit for the days selected as exponential phase is shown in red. Despite the discrepancies between the data sets, the overall trend is preserved.

I realized some concentration profiles from DCH and Biocrates differed on the values although the overall trend was similar (see Figure 3.2), so I tried to minimize the experimental error by merging the two data sets into a third one, given the fact that they correspond to the same experiments. All rates for the three data sets are listed in appendix A.

Figure 3.3 shows FBA-predicted Vs experimental growth rates in iCHO1766 metabolic model for the three data sets of constraints. Merging the data sets improved predictions on cell growth. Nevertheless, we wanted to further analyze the effect of accurate measurements of supernatant metabolite concentrations on model predictions.

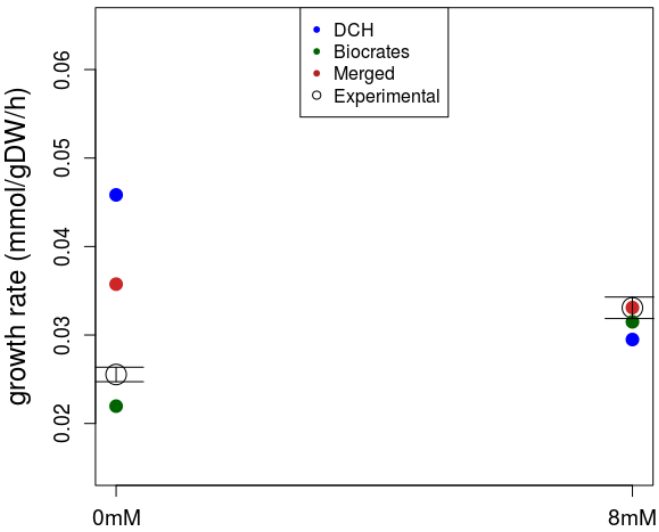


FIGURE 3.3. FBA-predictions on cell biomass with iCHO1766 metabolic model for 0mM and 8 mM glutamine experiments using three data sets (DCH, Biocrates, or both merged). Horizontal bars indicate experimental error.

3.1.1.2 Effects of the biomass composition and uptake rates on model predictions

I realized that small variations in metabolic rates had great effects on FBA results for predictions on cell growth. I computed shadow prices for the uptake and secretion rates to determine how those changes could affect the predictions (see Figure 3.5). The effects of uptake rates on biomass are more visible for essential substrates with low uptake rates, for which catabolic reactions have almost zero flux (most of the flux goes directly towards biomass generation). Experiments with 0mM and 8mM glutamine (one biological sample, three and one technical replicates for Biocrates and DCH) show that fluxes for exchange reactions cannot be computed accurately with such low number of samples and replicates. Therefore we decided to repeat these experiments with CHO-K1 and add other cell lines. In this case, sampling time was reduced and all were done in three biological replicates.

Figure 3.4 shows results for FBA predicted Vs experimental growth for 3-4 samples compared to one sample per day in 13 batch processes. These results show that experimental error on the rates can be significantly reduced by increasing the sample points per day, thereby improving model predictions. Vertical error bars on predicted values result from sampling experimental rates (1000 FBA runs) within the experimental standard error of the measurements.

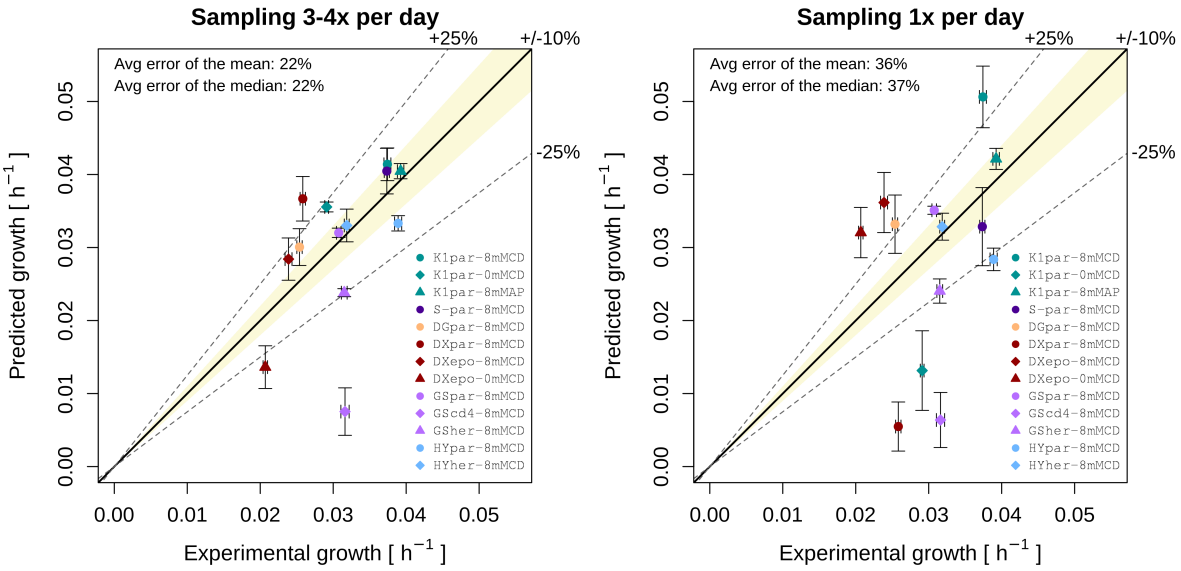


FIGURE 3.4. FBA prediction Vs measured growth rates for different cell lines and different conditions. Vertical error bars on predicted values result from sampling experimental rates (1000 FBA runs) within the experimental standard error of the measurements.

Shadow prices

Shadow prices are the derivatives of the objective function with respect to an exchange flux, and they can be used to assess the sensitivity of the objective function to the availability of a certain substrate.

$$\pi_i = \frac{\delta Z}{\delta b_i} \tag{3.3}$$

Figure 3.5 shows how phenylalanine and tryptophan uptake impacts biomass production in 0mM and 8mM experiments, respectively. This is another example of how obtaining accurate rates is crucial for model predictions, mostly in the case of essential amino acids with lower uptake rates.

3.1.2 FSEOF with literature data

I performed FSEOF with pFBA for 8 different sets of uptake and secretion rates with which the model was originally tested in literature [69]. In Carinhas et al. [192], 4 experiments were run with high and low producer CHO-K1 derived cell lines with and without sodium butyrate treatment. Rates from early and late exponential phases in Selvarasu et al. [72] from CHO-DG44 were used separatedly for FBA predictions. In Martinez et al. [193], CHO-K1 producer cell lines

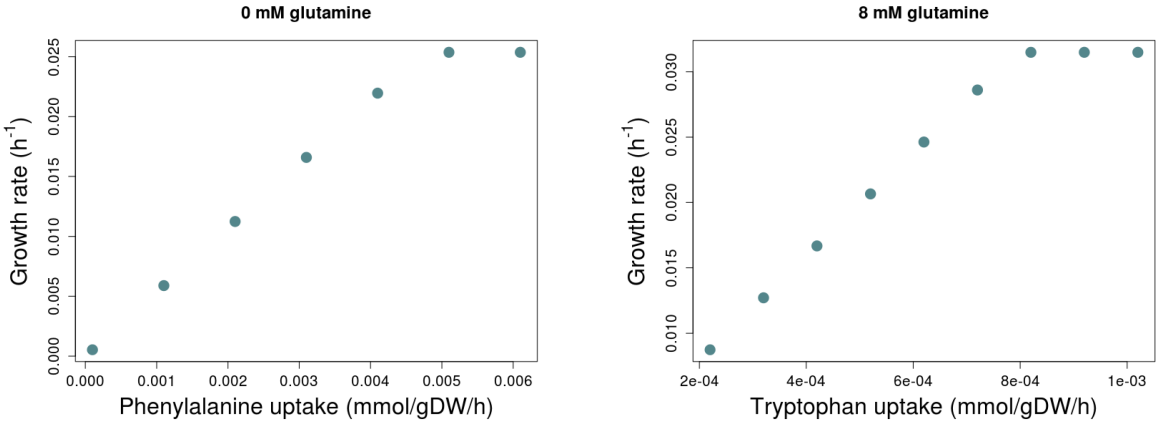


FIGURE 3.5. Shadow prices: biomass production with respect to exchange fluxes of phenylalanine in 0mM glutamine (left) and tryptophan in 8mM glutamine (right) using Biocrates set of experimental constraints

Data set	**experiment**	**IgG_{exp}** [μM/gDW/h]	**90% of IgG_{max}** [μM/gDW/h]
Carinhas (CHO-K1)	HP	0.02	0.1
	HP + NaBu	0.03	0.11
	LP	0.007	0.045
	LP + NaBu	0.012	0.114
Selvarasu (CHO-DG44)	early	0.013	0.038
	late	0.024	0.044
Martinez (CHO-K1)	Cold1	0.005	0.068
	Cold2	0.005	0.142

TABLE 3.1. Data sets with which FSEOF was performed. HP: high producer; LP: low producer; NaBu: sodium butyrate; IgG_{exp}: experimental value for protein production. IgG_{max}: maximum theoretical IgG production computed with FBA.

were exposed to a temperature shift in 2 independent experiments (Cold1 and Cold2). All cell lines were Immunoglobulin G (IgG) producers.

The cell lines, metabolic model and experimental conditions used are summarized in table 3.1.

I found some fluxes which vary with increasing protein production to be specific to the given experimental constraints. These might reflect only limitations on substrate uptake that affect protein production under those specific conditions. For instance, I found many reactions from the TCA cycle to be negatively correlated to protein production in the case of Selvarasu et al. (early exponential phase). This might be due to the fact that glucose uptake is the lowest of all data

sets and it is not enough to cover pentose phosphate pathway (PPP) and TCA cycle demands. In order to search for more general metabolic targets, I compared FSEOF results for all 8 data sets to find reactions which were decreasing or increasing with protein production in all cases.

For those reactions which decrease with increasing protein production, I performed *in silico* knockouts (by setting the flux to zero) to check if they are essential to sustain growth.

I found only few reaction candidates for overexpression. Among them, reactions associated to valine, leucine and isoleucine metabolism and fructose and mannose metabolism. An interesting reaction present in all data sets to be positively correlated with protein production is that catalyzed by the phosphoglucomutase:

$$\text{glucose-1-phosphate} \rightleftharpoons \text{glucose-6-phosphate}$$

Glucose-6-phosphate is the main substrate for TCA and PPP. As already mentioned, several reactions from the TCA were found to correlate inversely with flux towards protein production in Selvarasu et al. Therefore, our assumption is that increasing the glucose-6-phosphate towards the PPP could have positive effects in protein production, consistent to experiments performed in yeast [194].

In agreement with these results for overexpression targets, I found the production of fructose 1,6-bisphosphate to inversely correlate with IgG production for all data sets. As one of the compounds in early steps of glycolysis, downregulation of this pathway could imply that flux is diverted towards the PPP.

I found the following pathways to decrease with higer flux towards protein:

$$\text{D-Mannose 1-phosphate} \rightleftharpoons \text{D-Mannose 6-phosphate}$$
$$\text{D-Mannose 6-phosphate} \rightleftharpoons \text{D-Fructose 6-phosphate}$$
$$\text{D-Fructose 6-phosphate + UTP} \rightleftharpoons \text{UDP + } H^+ \text{ + D-Fructose 1,6-bisphosphate}$$

And:

$$\text{D-Fructose + dATP} \rightleftharpoons H^+ \text{ + D-Fructose 6-phosphate + dADP}$$

The flux of the latter, identified as `R_r0358` is plotted as an example of flux decreasing with increasing protein production in Figure 3.6.

I did not find experimental validation for these reactions in literature; 3-phosphoglycerate dehydrogenase, however, is a glycolytic enzyme and it was found to be down-regulated in two different high producer CHO cell lines when compared to low producers [195, 196].

I found more targets to appear in all data sets for knockdown and knockout. Most of the reactions found belong to cholesterol and lipid metabolism pathways. In terms of metabolic precursors, these pathways could be "wasting" resources that could otherwise be used for IgG production. However, several studies show that cholesterol and lipid metabolic pathways are overexpressed in high producer CHO cell lines [195, 197], which agrees with the fact that organelles and cell membrane play an important role in the synthesis and the transport of recombinant proteins. These

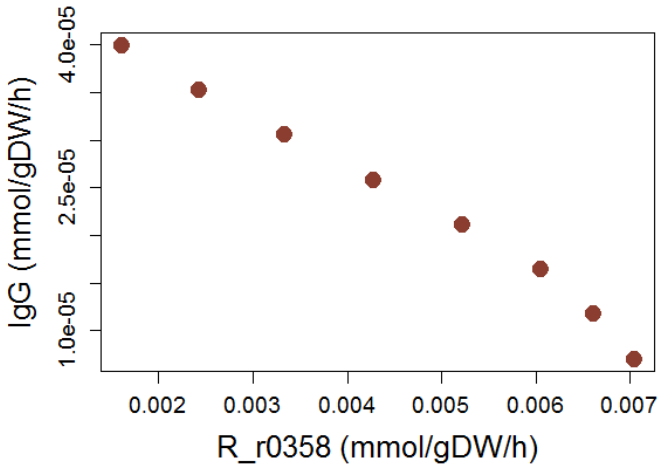


FIGURE 3.6. Flux predicted by FSEOF to decrease with increasing IgG production

results could point at possible limitations regarding the interpretation of metabolic modelling simulations.

A complete list of the subsystems for the pathways found to be correlated with protein production can be found in appendix B.

3.2 Unbiased approaches

3.2.1 EFM analysis for energetically efficient pathways

In order to be able to compute the EFMs, I reduced the metabolic model of CHO under a set of experimental constraints on minimal substrate uptake. I analyzed EFMs regarding NADH/NADPH and ATP production, as fundamental determinants of metabolic performance. Our goal is to find potential engineering targets which utilize energetically efficient pathways. In fact, NADPH has been hypothesized to be limiting for proliferation and lipid biosynthesis [198].

3.2.1.1 Model reduction

EFM analysis on the whole metabolic network (6663 reaction) is currently not computationally feasible due to the combinatorial explosion in the enumeration of the complete set of steady state pathways in the network. Thus, I generated a reduced version of CHO metabolic model. Besides the central carbon metabolism, which encompasses all the major fluxes [74], the model contains the fluxes to sustain growth and protein production with only the minimal substrate requirements. Since our goal was to investigate energetically efficient pathways, I included

glutamine (which is not essential for growth), since it is a major energy source in mammalian cells [199, 200, 201].

I used *NetworkReducer* for the reduction of the metabolic network. The algorithm is included in the *CellNetAnalyzer* toolbox for MATLAB [202].

Model pre-processing

In order to decrease computation time in the main reduction algorithm, I performed several pre-processing steps to remove blocked reactions and generate a consistent metabolic model.

The starting network is the generic model iCHO1766, downloaded from the BiGG database [52]. This generic model includes the biosynthetic pathway for arginine. This pathway was removed from the model due to the lack of omics data on cell-line specific experiments (see [69]).

Following the assumption that both biomass and protein production are ultimately delimited by essential metabolites, I simplified the constraints on the uptake to glucose, glutamine, oxygen and essential amino acids (arginine, histidine, isoleucine, leucine, lysine, methionine, phenylalanine, threonine, tryptophan and valine). The values for the uptake rates were taken from experimental data in exponential phase of CHO-DG44 batch experiments [72]. Since the experimental rate of protein production is no longer feasible under minimal media conditions, I set a value for IgG production so that the flux ratio protein/biomass is the same as in the original set of constraints ($r_{IgG} = 0.0075$ mmol/gDW/h).

In the first steps, transport reactions that are inactive due to missing metabolites in the media are removed from the model. Transport reactions are often coupled to cofactors such as Na^+ or Cl^-. These cofactors are considered to be in excess in the cell environment, so I artificially set exchange reactions to be cofactor independent.

The following step requires running FVA with the above-mentioned minimal experimental constraints. This method computes the feasible flux range of each reaction under the given conditions. "Blocked" reactions with $r_{max} = r_{min} = 0$ (inactive under these conditions) are removed from the model. Reactions that always occur in one direction are set to be irreversible. The resulting model is said to be "consistent" and will be the input for the *NetworkReducer* algorithm [203].

NetworkReducer

These are the main inputs required by the algorithm:

- The metabolic model (pre-processed)
- Protected phenotypes: constraints on uptake rates; minimal biomass and protein production required
- List of protected reactions

- Other parameters: which solver to use (CPLEX); minimal degree of freedom (1); minimal number of reactions remaining in the reduced model (1) and whether or not the protected reactions should be feasible (yes).

Protected reactions are those that are considered to be part of the core metabolism of CHO and must not be removed. These were taken from [7] and include glycolysis, TCA cycle, glutaminolysis, amino acid metabolism, PPP, biomass and protein production. Reactions in literature were in most cases lumped and had to be mapped into CHO metabolic model to obtain the complete pathway. The final list of protected reactions can be found in appendix C.

Protected phenotypes: additionally to uptake and secretion rates from [72], I included minimum cell growth (at least 99% of maximal biomass as of FBA) and adenosine triphosphate (ATP) maintenance energy. The latter (catabolism of ATP in ADP and P_i) accounts for energy consumption not associated with cell growth. I set a lower bound of 3mmol/gDW/h. Although I can assume this value is higher in mammalian cells, I could not find a valid experimental value for CHO cells in literature. Our working group is currently running continuous fermentations to determine maintenance energy and experimental values will soon be available.

The *NetworkReducer* algorithm iteratively removes reactions based on FVA. It is assumed that if a reaction has a small flux range under the given conditions, it can be removed from the network without highly affecting the overall flux variability of the system. These candidates will be removed only if the protected phenotypes and the protected reactions remain feasible. Figure 3.7 shows a summary of the main steps taken for reducing the genome-scale metabolic network.

The resulting reduced model consists of 291 metabolites (415 species) and 405 reactions. It reproduces the FBA results for biomass and protein production in the generic model (6,663 reactions).

3.2.1.2 Computation of EFMs on the reduced model

CellNetAnalyzer also incorporates efmtool [204] for the enumeration of EFMs. Enumeration of the whole set of EFMs took ≈ 6 minutes, and returned 529,329 EFMs. Distribution of NAD(P)H and ATP yields in the whole set of modes is very similar (see Figure 3.8 for the correlation of ATP Vs NAD(P)H yields), therefore only EFMs with ATP yield distributions are shown (Figure 3.9). Only those modes in which protein production and biomass were higher than 40% of theoretical maximum were selected for further analysis of energetically efficient pathways (modes in which energy but neither biomass nor protein is produced are of no biological interest). This subset of modes is shown in gray in Figures 3.10 and 3.11.

This subset of EFMs with biomass and protein production higher than 40% was then sorted with respect to moles of 1) ATP and 2) NADPH/NADH carbon yield. I then compared the union of the supports (active reactions in each EFM) of the two most efficient EFMs against the intersection of the sets of reactions in the two least efficient modes; idem for the three most and least efficient modes, and so forth (up to 20). The top 20 most and least efficient modes are shown

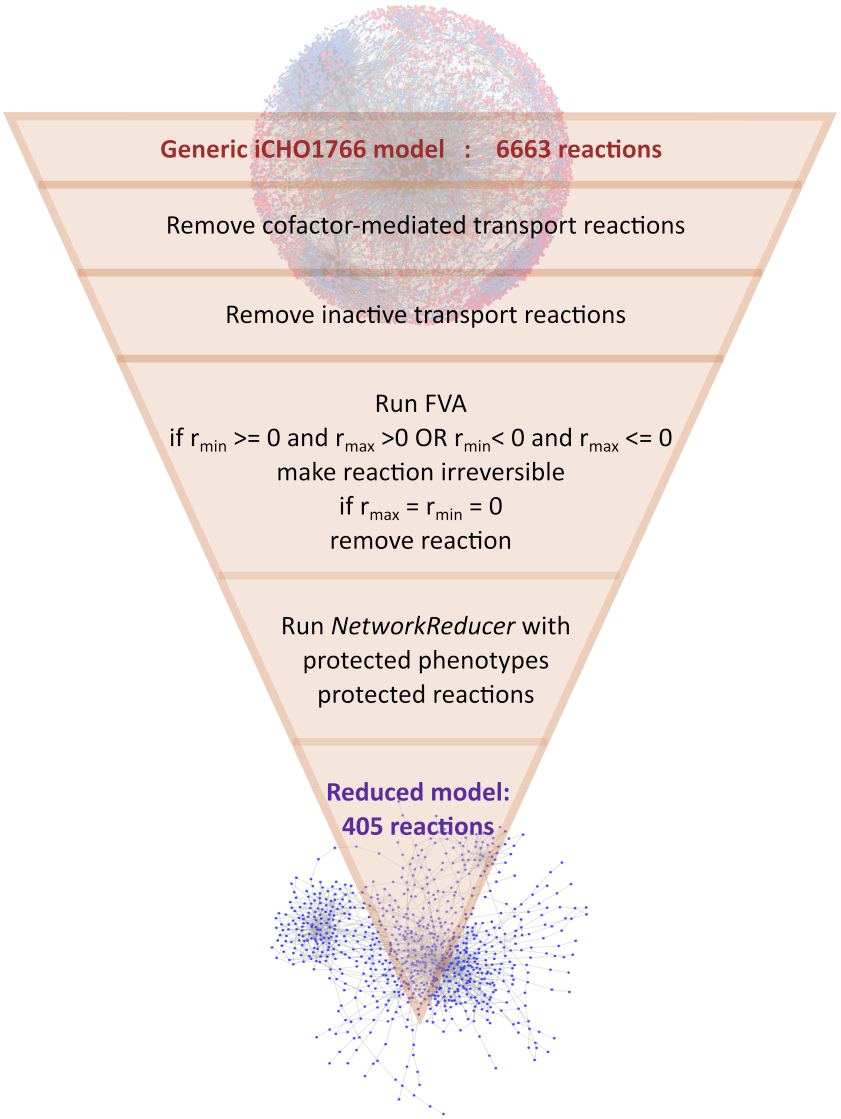


FIGURE 3.7. Main steps of the process for network reduction. "Hairball" representations of the generic (top) and reduced (bottom) models were done with Cytoscape.

in green and red in Figures 3.10 and 3.11. Our goal was to find reactions which are linked to an efficient use of the metabolic substrates and which could be potentially used as metabolic engineering targets.

I found no reactions in the most and not in the least efficient modes for ATP-producing EFMs. On the other hand, when ranked according to NAD(P)H carbon yield, I found 3 pathways that appear only in the most efficient modes:

- Glycolysis/gluconeogenesis

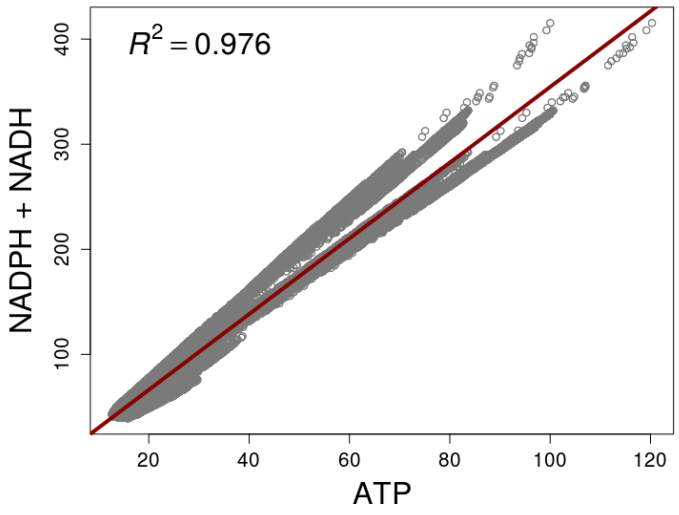


FIGURE 3.8. ATP and NADPH/NADH yields for the total set of EFMs

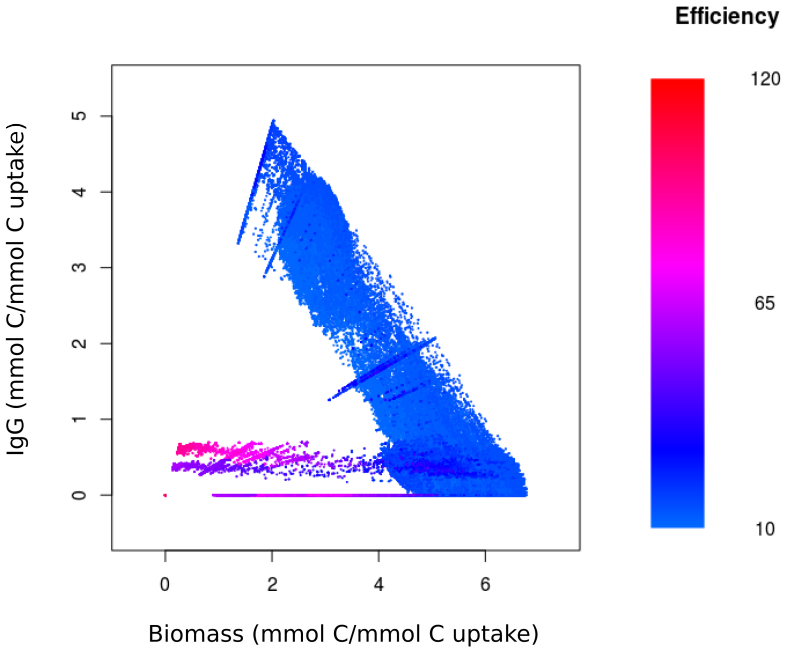


FIGURE 3.9. Total set of EFMs of the reduced metabolic model. Values correspond to biomass and protein production (normalized by uptake). They are colored according to the sum of the fluxes over ATP consuming reactions (normalized).

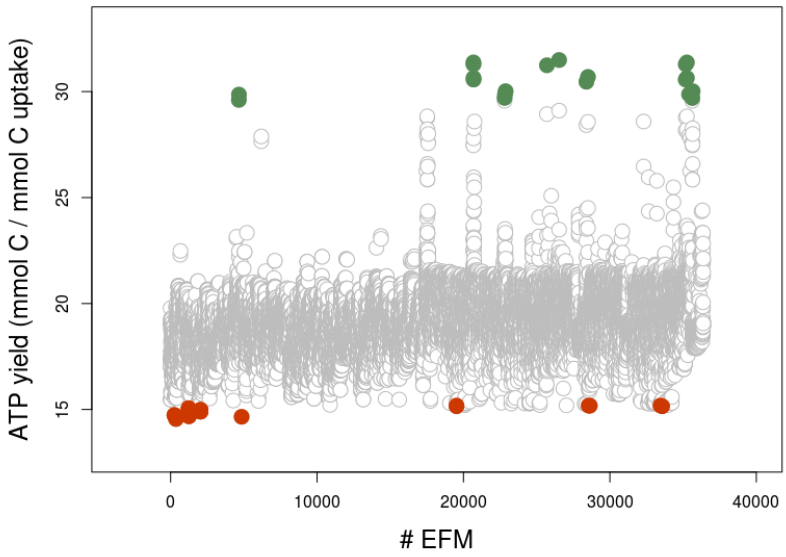


FIGURE 3.10. EFMs in which protein and biomass production are higher than 40% of the maximum. 20 most and least efficient EFMs in terms of ATP are shown in green and red, respectively

3-Phospho-D-glycerate → D-Glycerate 2-phosphate

D-Glycerate 2-phosphate → H_2O + phosphoenolpyruvate

phosphoenolpyruvate + ADP + H^+ → ATP + pyruvate

- Tyrosine metabolism

α-ketoglutarate + tyrosine → glutamate + 3-(4-hydroxyphenyl)pyruvate

3-(4-hydroxyphenyl)pyruvate + O_2 → CO_2 + homogentisate

homogentisate + O_2 → 4-maleylacetoacetate + H^+

4-maleylacetoacetate → 4-fumarylacetoacetate(2-)

4-fumarylacetoacetate(2-) + H_2O → H^+ acetoacetate + fumarate

- Urea cycle

arginine + H_2O → ornithine + urea

ornithine + α-ketoglutarate → glutamate + L-Glutamate 5-semialdehyde

Two of these pathways contain α-ketoglutarate, which is key metabolite of the TCA and it is known to have a crucial role in cellular energy metabolism. However, the pathways that I found to be energetically efficient are not part of the TCA but of phenylalanine and tyrosine

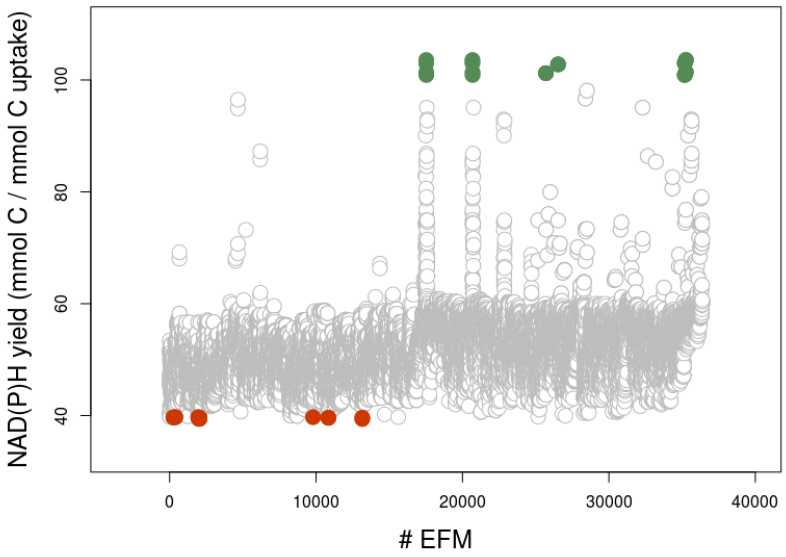


FIGURE 3.11. EFMs in which protein and biomass production are higher than 40% of the maximum. 20 most and least efficient EFMs in terms of NAD(P)H are shown in green and red, respectively

catabolism; fumarate is not produced via succinate but via homogentisate. This alternative pathway could generate fumarate to be incorporated to the TCA cycle (with the concomitant increase in production of NAD(P)H).

The second pathway from α-ketoglutarate belongs to the urea cycle. This pathway is also coupled with the TCA via the glutamate synthase which catalizes the following reaction:

$$\text{L-glutamine + 2-oxoglutarate + NADPH + } H^+ \rightleftharpoons \text{2 L-glutamate + } NADP^+$$

3.2.2 The convex hull method

Section 3.2.2 is part of *Fast computation of multi-dimensional production envelopes in genome-scale metabolic models* by Sarah N.Galleguillos, Matthias P. Gerstl, Norbert Auer, Nicole Borth and Jürgen Zanghellini to Scientific Reports.

The CHM is an established algorithm used in computational geometry for projecting multi-dimensional polytopes onto lower-dimensional spaces. As a formal description of the algorithm is available elsewhere [205], I will introduce the CHM by way of example.

Suppose we want to calculate the PE spanned by r_1 and r_2 for the toy network in Figure 2.3a, i.e., I want to compute the projection of the three-dimensional flux

polyhedron in Figure 2.3b onto r_1 and r_2. We assume that the matrices and vectors are sorted such that the first $n_R = 2$ columns and elements correspond to the reactions of interest that span the PE and that the PE is full-dimensional. Then the algorithm proceeds as follows:

Initialization phase. The coordinates of the first EP (point 1 in Figure 2.4a) are obtained by maximizing the first flux of interest [$\max(r_1) = 3$] in a standard FBA problem; then the maximum of the second reaction of interest is computed by FBA with $r_1 = 3$ fixed to its maximum [$\max(r_2) = 3$ if $r_1 = 3$]. Similarly, the coordinates of the second EP are obtained by minimization. Next, the line containing the initial points is drawn outwards by solving the associated FBA problem. This defines a new EP, point 3 in Figure 2.4a. To make sure that the point is a vertex and does not sit on a line between two adjacent vertices, its coordinates are successively optimized as described above (this step is not necessary in the toy example as point 3 is not degenerated). The three points now define a plane, which is again moved to obtain a new vertex. This process is repeated until $n_R + 1$ initial vertices are computed (in the toy example, $n_R + 1 = 3$).

Iteration phase. The initial EPs represent an approximation of the convex hull (CH). They define a smaller polytope that sits completely inside the projection of the flux polytope (see blue area in Figure 2.4a). The projection is confined by (hyper-)planes (or lines as in 2.4a). Some of these hyperplanes (HPs) might be "terminal", which means that they cannot be pushed further outward (see the line segments between the points 2-3 and 1-3 in 2.4a). The idea of the iteration phase is to shift all confining, non-terminal HPs (line segment 2-4 in 2.4b) outward as far as possible (dashed line in 2.4a), which then defines a new vertex (point 4 in 2.4b). As above, the coordinates of this point are obtained by successive optimization. With the new EP, the CH is updated, i.e., new confining HPs are added for all possible combinations of existing points. Previously confining HPs that now sit inside the CH are removed (compare Figures 2.4a and 2.4b). This incremental refinement of the CH continues until all HPs are terminal, i.e., until they correspond to the confining HPs of the wanted projection; see 2.4c. The flowchart in 3.12 summarizes the procedure.

Implementation

An implementation of the CHM based on exact, rational arithmetic is available [206]. However, I found that the program terminated with a segmentation fault if run with the genome-scale metabolic model *i*JO1366 [207]. I implemented two different versions of the CH algorithm: double-precision MATLAB version (64-bit, $\approx$ 16 digits)

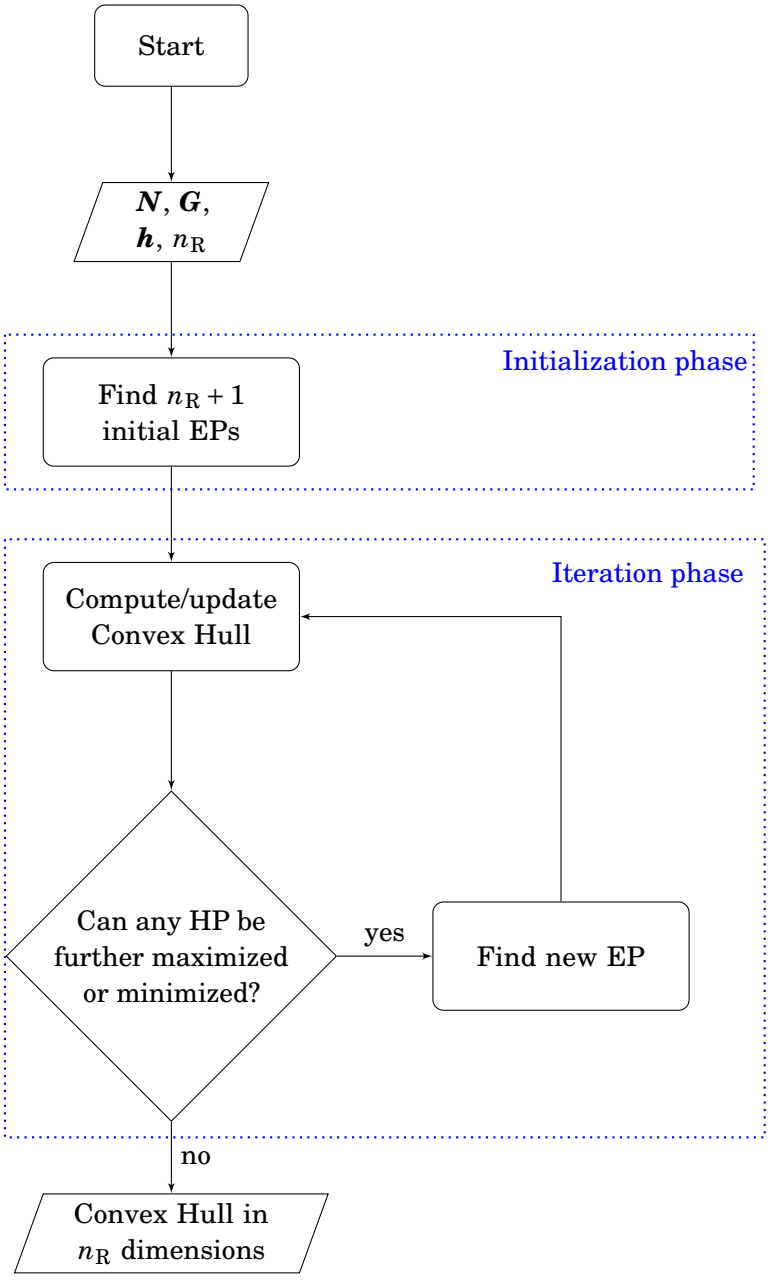


Figure 3.12: CHM flowchart. The inputs for the algorithm are the stoichiometric matrix $\boldsymbol{N}$, linear constraints $\boldsymbol{Gr} \leq \boldsymbol{h}$, and n_{R} dimensions onto which the flux polytope is to be projected. The initial CH is given by the EPs in the first dimensions. This initial hull is refined by maximizing/minimizing the HPs containing the EPs until all the facets of the projection are terminal.

and exact, rational arithmetic version in Python3. For the linear optimization, I used CPLEX and *Qsopt_ex* [208] respectively. The source codes of both implementations are available under the GNU General Public License at `https://bokubox.boku.ac.at/index.php/#200f56a1ab47d20a3f4e3718d0872981` (A GitHub link will be

Parameter	**Use**	**Value**
Precision	general	12 decimals
Precision	$\leq$, $\geq$, =	6 decimals
ϵ	if $\mid x \mid < \epsilon$, then $x = 0$	1×10^{-5}
τ	$\boldsymbol{a}^T \boldsymbol{r} \geq b - \tau$, $\boldsymbol{a}^T \boldsymbol{r} \leq b + \tau$	1×10^{-5}

Table 3.2: List of parameters used in the double-precision implementation of the CHM.

provided upon acceptance.)

Specifics of the double precision implementation. The general precision of the double precision CHM was set to 12 decimal places. This setting affect matrix multiplications, generating HPs or EPs.

During the iterative phase of the algorithm, I encountered some infeasible linear programs caused by accumulation of rounding-off errors. For instance, once a HP $\boldsymbol{a}^T \boldsymbol{r} = b$ is computed, it should be added as an equality constraint to the problem in order to find a new EP. However, equality constraints in linear programs sometimes cause numerical issues for solvers. For this reason, I allow for an explicit tolerance τ in the double-precision implementation. Rather than adding an equality constraint $\boldsymbol{a}^T \boldsymbol{r} = b$, I add $\boldsymbol{a}^T \boldsymbol{r} \leq b + \tau$, and $\boldsymbol{a}^T \boldsymbol{r} \geq b - \tau$.

On the other hand, I reduced the number of decimal digits when comparing numerical values in order to increase the robustness of the method against round-off errors. This applies, for example, to removing hyperplanes if extreme points lie on the right and left. This is, a hyperplane $h = \boldsymbol{a}^T \boldsymbol{r}$ is removed from the convex hull if there are extreme points e such that $h(e) < b$ and $h(e) > b$, where b is the distance from the origin to the plane (see Hesse normal form equation of the plane in table 3.3). The final values of the parameters used in the double-precision implementation of CHM are listed in Table 3.2 and are the result of several rounds of parameter optimization, with the main goals of removing infeasibilities and keeping a consistent set of EPs when the reactions in n_{R} were sorted differently.

Escherichia coli

I used the core metabolic model EColiCore2 [209] and the genome-scale metabolic model of *i*JO1366 [207] – downloaded from the BIGG database [52] and simulated growth on glucose minimal medium under oxygen-limited conditions. Maximum oxygen and glucose uptake were constrained to 5 and 10 mmol/gDW/h, respectively. The lower bound on the rate for ATP (accounting for maintenance energy) was 3.15 mmol/gDW/h.

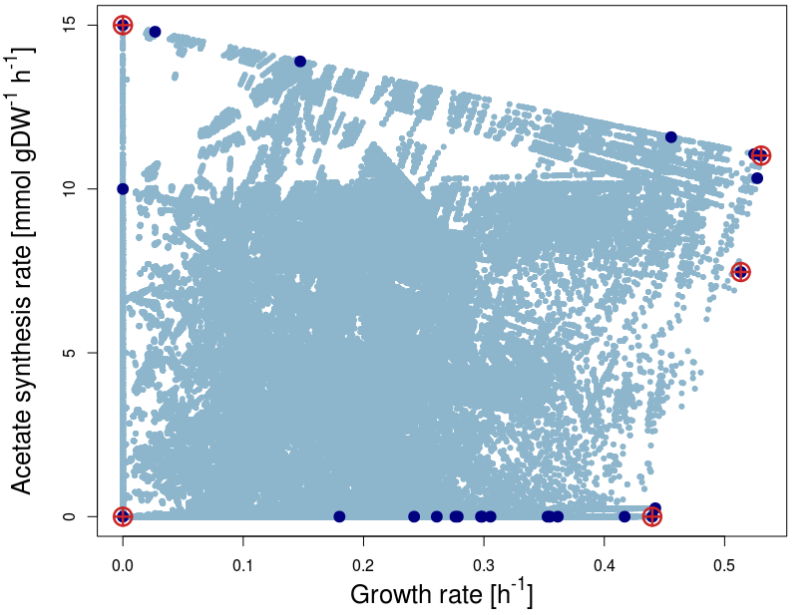


Figure 3.13: PE for acetate as function of growth in EColiCore2. Blue-gray dots represent the complete set of EFVs computed with efmtool; vertices of the convex hull encompassing the EFVs as computed by MATLAB `convhull` function are shown in blue filled circles. EPs computed with CHM (double and exact implementations) are depicted as crossed red circles.

Validation on a core metabolic model

I computed the PE for acetate with respect to growth in EColiCore2 [209] with our implementations of the CHM. Both computed an identical set of EPs, see the red points in Figure 3.13. Next, I computed the complete set of EFVs via the CellNetAnalyzer toolbox, version 2019.1 [190], projected them onto acetate and cell growth (light blue dots in Figure 3.13) and calculated the EPs of the resulting convex hull with the MATLAB function `convhull` (dark blue filled circles). All vertices that were found by our CHM were also found by MATLAB. However, `convhull` found additional vertices that visually appear to sit on line segments between the red points in Figure 3.13. In fact, these additional points not true EPs but caused by inaccuracies of `convhull` as all of them sit on line segments spanned by the red vertices in Figure 3.13.

Application to a genome-scale metabolic model

Next, I computed the PE for the five main fermentative products (succinate, lactate, formate, acetate and ethanol [210]) in *i*JO1366 [207]. In this case, I find that the double precision implementation misses some EPs compared to the exact imple-

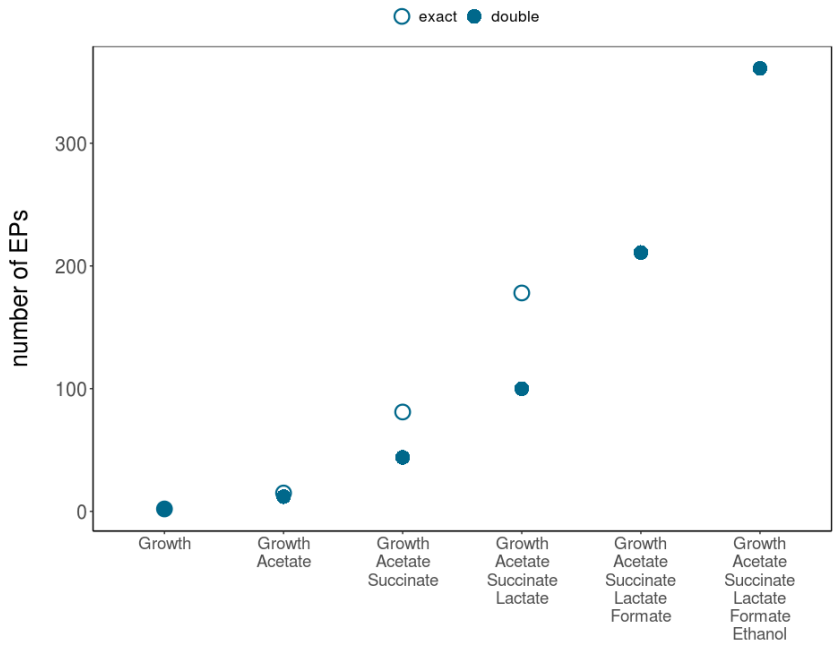


Figure 3.14: Number of extreme points found in the exact implementation (open circles) compared to double (filled circles) for the CHM for increasing number of n_{R}.

mentation, see Figure 3.14. This discrepancy becomes more prominent for higher dimensions. However, I verified that the EPs found by the exact but not by the double precision CHM are within the algorithm specific error margin of ε and sit either close to a terminal HP or another EP. Thus I conclude that the missing points are due to the limited accuracy of the double precision method.

I compared run times of our implementations of the CHM with current alternatives such as sampling the CH with FVA [187]. Results are shown in Figure 3.15. FVA was performed at equally spaced flux values along each dimension i of interest with 40 or 60 supporting points between $\max(r_i)$ and $\min(r_i)$. CPLEX was used to solve the associated FBA problems. CHM in double-precision was found to be significantly faster and scaled better (with respect to the number of reactions of interest) than all other methods. The performance of CHM however, is dependent on the order of the reactions of interest, see Figure 3.15. For instance, consider again the network in Figure 2.3. Suppose you were to start the iteration phase with the EPs 1, 3, and 5 instead of 1, 2, 3 (see Figure 2.4), then all EPs of the projection would be retrieved already during the first iteration, which results in run times differences, see Figure 3.15.

Despite the run time differences, I verified that our implements always get the same set of EPs independent of the order of the reactions of interest.

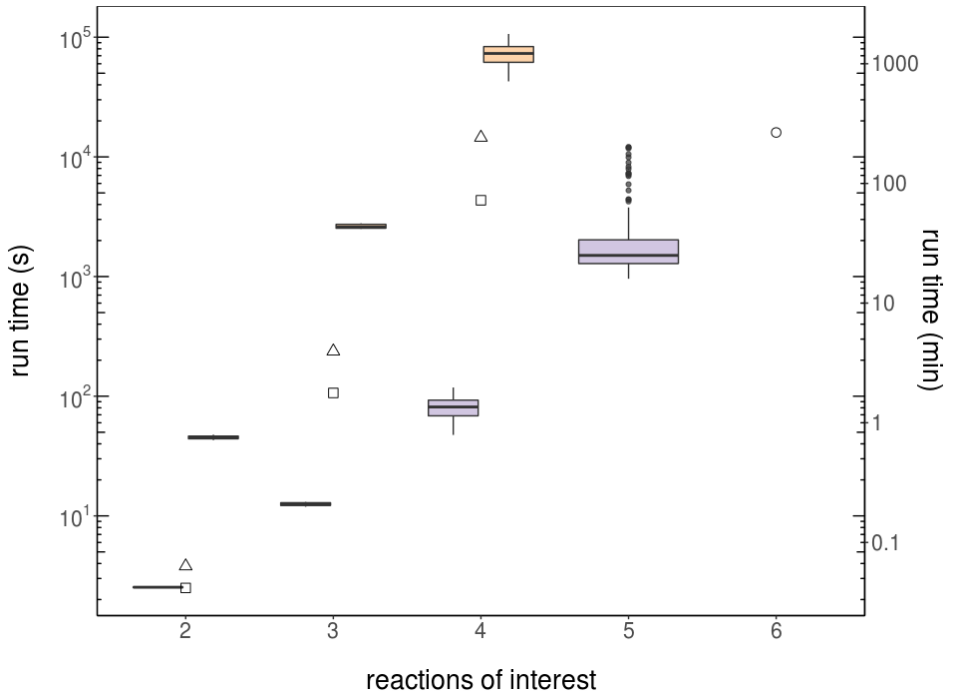


Figure 3.15: Run time distribution for all possible permutations of n_{R} in rational arithmetic (upper boxplot, in orange) and double-precision (purple). Run times for FVA with 40 and 60 sampling values per dimension are plotted as squares and triangles respectively. For 6 dimensions, one particular set of n_{R} is shown (circle). Note that all computations timed out after 10^5 s.

CHO cells

Next to the product, CHO cells secret lactate and ammonia during the fermentation process, both of which have a detrimental effects on growth and product quality [29, 211]. Thus I computed the PE for cell growth, production of IgG, lactate and ammonia in CHO.

CHO was represented by the genome-scale metabolic model iCHO1766 [69] and downloaded from the BiGG database [52]. This model includes bounds on all uptake reactions used.

Two 3D projections of the four dimensional PE are shown in Figure 3.16. The terminal (hyper-)planes of the CH are referenced according to the lists in Tables 3.3 and 3.4. I observe that for secretion rates below 0.5 mmol/gDW lactate and 0.23 mol/gDW ammonia neither of these is a metabolic burden for growth and IgG production. Thus, reducing typical secretion rates (0.1208 mmol/gDW for lactate and 0.083 mmol/gDW for ammonia [72]) frees no additional metabolic resources, but only reduces the toxic side effects of these products.

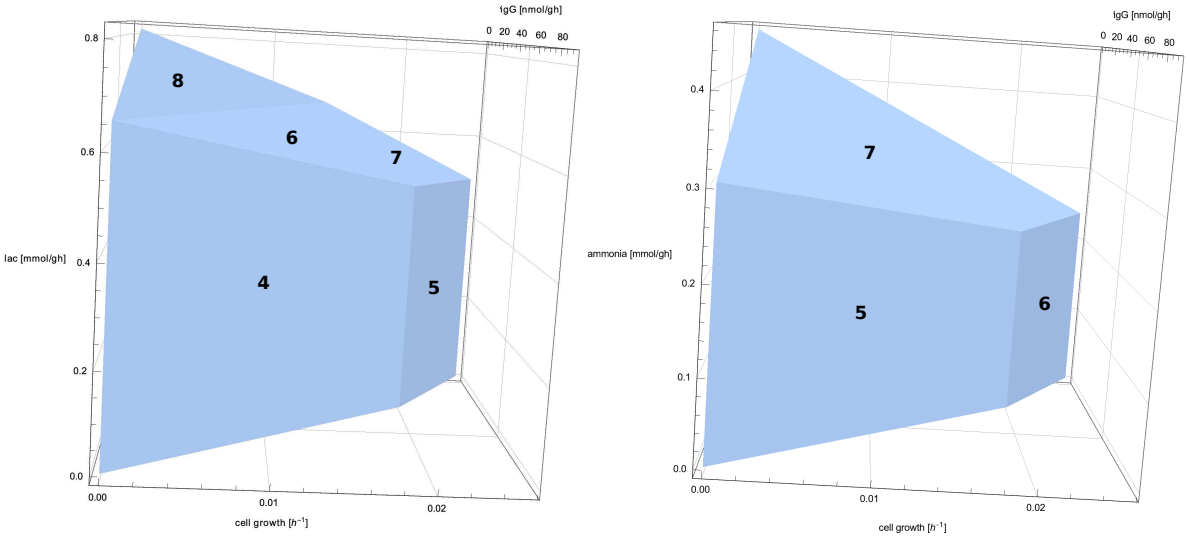


Figure 3.16: Projection of the phenotypic space for iCHO1766 onto three reactions of interest: IgG and lac (left) and ammonia (right) production with respect to growth. In different shades of blue, the HPs (facets) of the projection, listed in Tables 3.3 and 3.4

HP	a	b	c	d
1	-1000	0	0	0
2	0	-1000	0	0
3	0	0	-1000	0
4	0	999.9956	2.9523	0.0924
5	0	999.9826	5.9011	0.1511
6	0.4476	999.9796	6.3516	0.3863
7	0.4616	999.9772	6.7308	0.4009
8	0.5942	999.9799	6.2863	0.4825

Hesse Normal form ($\times 10^3$)

Table 3.3: Coefficients for the Hesse Normal form equation of the hyperplane ($a\,r_{lactate} + b\,r_{IgG} + c\,r_{growth} = d$) obtained with the CHM applied in iCHO1766 for the production of IgG and lactate with respect to cell growth.

HP	a	b	c	d
1	-1000	0	0	0
2	0	-1000	0	0
3	0	0	-1000	0
4	-999.9999	0.2696	0.0009	0
5	0	999.9956	2.9523	0.0924
6	0	999.9826	5.9011	0.1511
7	0.6053	999.9861	5.2377	0.2776

Hesse Normal form ($\times 10^3$)

Table 3.4: Coefficients for the Hesse Normal form equation of the hyperplane ($a\, r_{ammonia} + b\, r_{IgG} + c\, r_{growth} = d$) obtained with the CHM applied in iCHO1766 for the production of IgG and ammonia with respect to cell growth.

CHAPTER 4

DISCUSSION

The complexity of mammalian cells is a double-edged sword: on the one hand, it allows production of recombinant proteins with complex post-translational modifications, unlike prokariotic expression systems [212]. On the other hand, the intricate interactions among all components hinder our control over the production process. In order to convert CHO cells into efficient biopharmaceutical factories, we need to fight against natural genetic evolution; we need to rewire the components of the network to make cells divert resources from cell growth to production of the recombinant protein of interest. Strategies followed in the past years for cell line engineering were based on modification of single targets identified in basic research [213] and labour and cost intensive clone screening [214], with limited success rate [215]. This long process can be drastically reduced if we understand cell behaviour in order to modify it to meet our demands. Significant strides have been made in developing computational and analytical tools to gather information from every level of the cell systems, from genes to phenotypic traits. Systems biology aims to integrate this information and build mathematical models to be able to predict cell behavior.

In the present work, I have focused in constraint-based analysis of metabolic networks. These methods have long been applied in microbial systems [216, 64] and are rapidly developing for CHO since the full genome sequence became available [3, 42]. One example of application of these methods is the identification of potential candidates for gene overexpression and downregulation using FSEOF, which I applied in this thesis for 8 literature data sets on experimental constraints using CHO metabolic model. Results for downregulation suggest that decreasing the flux through the TCA cycle would improve protein production, consistent with proteomics analysis in high producer CHO cell lines [195, 196]. I foresee these methods will be successfully applied to CHO as they have been to bacteria [217] and yeast [218]. These reaction candidates will be further evaluated experimentally by our research group. On the other hand, downregulation of lipid synthesis which contradicts biological evidence, and reflects limitations of the model in

representing additional processes other than metabolism such as protein secretion.

I have reviewed the numerous applications of mathematical modelling to describe CHO metabolism. However, I am aware that this field is still in its infancy for CHO and there are still important limitations of *in silico* methods to be overcome in the near future. In the following, I mention some of those which are relevant to my work.

- I already mentioned that correct annotation of the genome is crucial for metabolic network reconstruction [21]. Compared to other metabolic models, iCHO1766 still lacks gene annotation for many of the metabolic reactions (only 1,766 out of 6,663 have a gene ID associated). Some potential targets predicted with metabolic modelling were not associated to any gene ID which poses a challenge for experimental validation of those targets.

- Proper assignation of reactions reversibilities, which is essential for performing constraint-based metabolic modelling [216] or metabolic flux analysis [77].

- Also problematic are transport reactions between compartments, which need to be described more accurately, even more if we take glycosylation, a highly compartimentalized process, into account.

- Improved cell-type specific models should be generated and validated for better predictions [219]. In our research group, we are currently working on assessing the impact of the biomass composition on modelling output. Currently, all cell line specific metabolic models for CHO (K1, DG44 and S) share the same biomass equation. Preliminary results on biomass composition show that protein and lipid content vary significantly among CHO cell lines (manuscript in preparation).

- In addition to the above-mentioned improvements on the metabolic model, we emphasize the importance of biological replicates and sampling intervals for the accuracy of the experimental data to be integrated in the model. Our results show that uncertainties on uptake and secretion rates impact model predictive capabilities and consequently any potential biological interpretation of the modelling output. In the era of high-throughput technology, coverage of different experimental setups often prevails at the expense of the number of replicates for a given experimental study. In order for systems biology to achieve the goal of reducing the number of experiments required, data-driven hypotheses should rest upon statistically significant results, more specifically in the case of CHO cells, which are known to be genetically unstable and batch-to-batch variations are observed.

- Major improvements are still required in the incorporation of recombinant protein glycosylation and secretion in iCHO1766. In our results, we saw an example in which metabolic modelling (FSEOF in this case) yielded results that were inconsistent with membrane-associated changes that come along with protein production and secretion. Recent developments in metabolism with gene expression (ME) models include protein translocation and

compartmentalization [220], in addition to enzyme expense to operate a metabolic pathway [221]. CHO metabolic model should also aim to integrate this kind of information for better predictions.

Constraint-based methods such as FBA overcome the drawbacks of kinetic models, but rely heavily on the definition of an objective function. One can find reasonable arguments to assume maximization of cell growth in microbial systems growing exponentially [168], but it may not be a suitable (or the only) objective function for engineered cells. FBA variants have been developed to address this issue, such as MOMA (minimization of metabolic adjustment) [222], in which the differences between the engineered and wild type strain flux distributions are minimized. Alternatively, one can perform an unbiased analysis of the solution space. The latter is a valuable alternative that can enlighten us on the phenotypic capabilities of CHO cells without the assumption of optimality.

Pathway analysis aims to obtain a greater picture of the metabolic landscapes of the network, contrary to methods such as FBA, in which only one optimal flux distribution is obtained. Despite their potential applications and versatility [223, 224], the applicability of these methods has been hindered by the complexity of the results and the computational limitations in large scale networks [225].

In order to perform EFM analysis in CHO, I have systematically reduced the genome-scale metabolic network to a biologically meaningful set of reactions that fulfills functions such as growth and protein production as the generic model with the minimal substrate requirements (oxygen and essential amino acids) and major carbon and energy sources (glucose and glutamine) [201]. The purpose of EFM analysis in this reduced model was to find genetic engineering targets to improve metabolism of CHO towards a more efficient use of the available resources. This is of special interest given the tendency to perform fed-batch cultures at high cell densities [226], where protein production and cell growth are competing for the limited energy capacity of the cell. I found 3 pathways that appear to be correlated with higher production of NAD(P)H per mol of substrate. Experimental validation of these targets is still ongoing.

EFMs have emerged as a valuable mathematical tool for network analysis that allows one to have a broader understanding of the system as a whole. Nevertheless, we still bear in mind that the enumeration of metabolic pathways was done in a reduced model and the extensive metabolic capabilities and flexibility of whole-cell metabolism are not fully represented.

In this regard, we have introduced an approach based in computational geometry to analyze the flux space for the CHO GSMM in an unbiased manner. Until now, this phenotypic space was computed by finding the maximum and minimum of the fluxes of interest with consecutive LPs by fixing other fluxes in small intervals, limiting the computation to only a few reactions of interest [227]. With the CHM, we were able to extend the computation to up to six dimensions in a genome-scale metabolic network within a reasonable computation time frame.

Additionally, recent advances in metabolic modelling applied to microbial communities have exposed a demand for new mathematical tools which include multiple organisms and metabolic targets [228, 229]. In some cases, reduced, core-models for each organism are assembled in a community model [230], with the potential risk of removing the metabolic flexibility for each network when exposed to cohabitation.

The applications of the CHM are of special interest in the case of microbial communities; the effects of the environment (uptake and secretion rates), gene modifications, or cohabitation with other species can be analyzed now in terms of particular changes to the phenotypic capabilities on multiple products of interest in a fast, unbiased manner.

Vertex enumeration methos such as the CHM rely on accurate numerics. Round-off errors are inherent to floating-point computation and are known to yield erroneous results in computational geometry [231]. Nevertheless, solution time of current exact solvers hamper the applicability of the CHM in systems of more than 4 dimensions. Instead, we argue that the double-precision implementation of the CHM is a valid approximation to compute PEs, all the more so if we consider the intrinsic inaccuracies of the stoichiometric coefficients of the reactions. These are obtained experimentally as biochemical compositions and the precision is limited to that of the measuring methods.

CHAPTER 5

CONCLUSION

Given the burgeoning demand for more efficient bioprocesses and more effective and safe therapeutic proteins, drug manufacturing in CHO has undergone major progress in bioprocess control, media optimization and genetic engineering of the host. This progress took place regardless of the dearth of mechanistic knowledge of the cellular processes and was primarily based on trial and error. The whole process of cell line development for a given therapeutic product takes between 16 and 24 months and includes screening of hundreds of cell lines to find stable high producers and formulation of specific optimal media conditions [214]. It is now for the first time that we can make use of years of data acquisition, more computational power and better modelling tools to aid rational design of experiments to obtain controlled product quality and higher yields. It is only with detailed characterization of cellular processes such as gene regulation, metabolism, post-translational modifications and protein secretion that we can aim to control the production process. The complexity of the genotype-phenotype relationship, mostly in higher organisms such as mammalian cells, goes beyond intuitive comprehension and requires mathematical modelling [232].

Metabolic modelling of mammalian cells has been hampered by the inherent complexity of the cell structure (compartmentalization) and the large variability of media compositions and process perturbations under which the culture processes are carried out. To these challenges, we have to add clonal variation and instability, intrinsic to CHO cells. New techniques such as single cell analysis would add another layer of information to the rampant field of omics technologies. In this regard, it is important to establish standardized methods for data integration and validation within the modelling community. All together, these layers of information about genes, transcripts, proteins and metabolites are slowly but steadily helping us to understand cell metabolism in a systems-level manner.

To date, the vast majority of modelling approaches in CHO have been applied in a reduced set of reactions. These usually include glycolysis, the TCA cycle, the PPP and amino acid metabolism. However, in 2016 a full genome-scale metabolic model of CHO has become available [69], unleashing the capabilities of genome-scale metabolic modelling. In the present work, I have applied constraint-based modelling methods to characterize CHO metabolism and to search for potential engineering targets related to the efficient use of energy resources and for increased protein production. Our results suggest that decreasing flux through the TCA cycle could boost protein production. Pathways from tyrosine metabolism, urea cycle and central carbon metabolism were reported to make an efficient use of metabolic resources in terms of NAD(P)H. These potential engineering targets will be experimentally tested for changes in protein production or cell growth.

Pathway analysis tools such as EFMs or EFVs are computationally demanding, mostly in the case of genome-scale metabolic models. Even though computing power increases exponentially, there are still limitations when it comes to solve mathematical problems in a reasonable amount of time. We have developed a new, fast method for the characterization of the full feasible flux space of the metabolic network based in computational geometry that bypasses the need of sampling the phenotypic space with FVA. The CHM runs LPs to efficiently find only the vertices of the PE for the reactions of interest. The algorithm can be applied for a higher number of reactions than state-of-the-art algorithms for phenotypic phase plane analysis, with low time and memory requirements.

Despite being a promising tool for bioprocess optimization, metabolic modelling in CHO is still challenging. I already mentioned the need for better annotation. On the other hand, predictions can only be as close to reality as the experimental constraints are. Parallel improvement on analysis techniques will undoubtedly boost the prediction capabilities of metabolic models. These still have to be further developed to generate more accurate results, e.g., metabolite concentrations for uptake and secretion rates or biomass constituents [72, 233]. In addition, the iterative process of model building can benefit from predictions which fail to describe cell behaviour.

Besides metabolic burdens for production, modelling glycosylation still remains a challenge given the combinatorial nature of the process. Many factors seem to affect the final glycan distribution, and controlling product quality (and therapeutic efficacy) depends heavily on knowing which factors are the key players and what biochemical process they trigger. Glycan patterns vary from batch to batch and from cell line to cell line, making it difficult to model the process deterministically. Statistical models of protein glycosylation perform well in predicting final glycan structures from key bioprocess parameters [234] but still lack mechanistic understanding of the underlying causes. On the bright side, it has been recently shown that only a limited amount of CHO proteins account for the majority of glycosylation, which could ease the approaches dealing with the dynamic evolution of glycosylation by focusing solely on these highly contributing proteins [152]. Even though the mechanisms by which the culture conditions and enzyme expression affect glycosylation are still unknown, the modelling efforts discussed in

section 1.4 have taken a significant step forward in media optimization by linking glycosylation to metabolism. A future step in this direction would be including glycan compounds in the biomass stoichiometric equation, since it has been shown that the metabolic demands towards glycosylation of both recombinant and host proteins are significant [152].

In summary, the ability to accurately predict cell behaviour is of major relevance to the development of new biopharmaceuticals. Models for recombinant protein production in CHO cells are rapidly evolving and data acquisition must increase in both quantity and quality; the main goal is to build a merged model which integrates all levels of information from the cell systems, from genes to enzymes to metabolites; and processes such as post-translational modifications, protein folding and secretion.

APPENDIX A

APPENDIX A

Metabolic rates for 0mM and 8mM glutamine experiments

All rates are expressed in mmol/gDW/h.

Metabolite	0mM	8mM
Alanine	0.03817	0.07632
Arginine	-0.00246	-0.00549
Asparagine	-0.05911	-0.04403
Aspartate	-0.01529	0.03622
Citrate	0	0
Glutamine	0.00242	-0.19898
Glutamate	-0.02209	0.02288
Glycine	0.01797	0.01823
Histidine	-0.00292	-0.00887
Isoleucine	-0.02457	-0.04169
Leucine	-0.05797	-0.1
Lysine	-0.01488	-0.03095
Methionine	-0.00739	-0.01113
Ornitine	0.00005	0.00003
Phenylalanine	-0.0041	-0.00588
Proline	-0.0036	-0.00969
Serine	-0.05607	-0.0633
Threonine	-0.00648	-0.02114
Tryptophan	-0.00416	-0.00082
Tyrosine	-0.00597	-0.01297
Valine	-0.01536	-0.02385

Metabolite	**0mM**	**8mM**
Citrate	0.00356	0.00232
Isoleucine	-0.01468	-0.01224
Leucine	-0.0274	-0.0183
Malate	0.0012	0.00184
Succinate	0.00084	0.00093
Alanine	0.03611	0.0996
Arginine	-0.01136	0.00425
Aspargine	-0.08733	-0.07006
Aspartate	-0.01046	0.0328
Glutamine	0.00412	-0.21143
Glutamate	-0.03983	-0.02133
Histidine	-0.00464	-0.00563
Lysine	-0.01863	-0.02167
Methionine	-0.00509	-0.00407
Phenylalanine	-0.00926	-0.01247
Proline	0	-0.00054
Serine	-0.04655	-0.06776
Threonine	-0.01523	-0.00753
Tryptophan	-0.00301	-0.00565
Tyrosine	-0.00551	-0.00708
Valine	-0.01479	-0.01841

Table A.1: Metabolic rates for Biocrates (left) and DCH (right).

Metabolite	0mM	8mM
Glucose	-0.51149	-0.54198
Glutamine	0	-0.12719
Glutamate	-0.02066	0.02782
Lactate	0.52359	0.83569
Ammonia	0.00831	0.21855

Table A.2: Metabolic rates from in-house bioprocess data

APPENDIX B

APPENDIX B

Subsystems for the pathways found with FSEOF

GLYCOLYSIS/GLUCONEOGENESIS	1
VALINE, LEUCINE, AND ISOLEUCINE METABOLISM	4
NUCLEOTIDE INTERCONVERSION	11
TRANSPORT, GOLGI APPARATUS	1
PYRIMIDINE SYNTHESIS	1
TRANSPORT, ENDOPLASMIC RETICULAR	1
N-GLYCAN SYNTHESIS	1
EXCHANGE/DEMAND REACTION	1
GLYCEROPHOSPHOLIPID METABOLISM	1
FRUCTOSE AND MANNOSE METABOLISM	1
TRANSPORT, MITOCHONDRIAL	1

TABLE B.1. Subsystems for the reactions found with FSEOF to be correlated positively with protein production, and the number of occurrences

GLYCEROPHOSPHOLIPID METABOLISM	10
BILE ACID SYNTHESIS	1
TRANSPORT, ENDOPLASMIC RETICULAR	3
EXCHANGE/DEMAND REACTION	1
GALACTOSE METABOLISM	1
PROPANOATE METABOLISM	1
BIOMASS	1
NUCLEOTIDE INTERCONVERSION	4
EICOSANOID METABOLISM	2
CHOLESTEROL METABOLISM	12
STARCH AND SUCROSE METABOLISM	3
TRANSPORT, PEROXISOMAL	2
PYRIMIDINE SYNTHESIS	2
TRIACYLGLYCEROL SYNTHESIS	2
UNASSIGNED	1
TRANSPORT, MITOCHONDRIAL	2
FRUCTOSE AND MANNOSE METABOLISM	4
TRANSPORT, EXTRACELLULAR	2

TABLE B.2. Subsystems for the reactions found with FSEOF to be correlated negatively with protein production, and the number of occurrences

APPENDIX C

APPENDIX C

Protected reactions for model reduction

The following list contains the identifiers from iCHO1766 to be protected during the model reduction. Modified from the lumped version in [7].

R_HEX1	R_PGI	R_PFK	R_FBA	R_TPI
R_GAPD	R_PGK	R_PGM	R_ENO	R_PYK
R_LDH_L	R_PDHm	R_CSm	R_ACONTm	R_ICDHxm
R_AKGDm	R_SUCOAS1m	R_SUCD1m	R_FUMm	R_MDHm
R_ME2m	R_GLUNm	R_GLNS	R_GLUDxm	R_r0081
R_r0193	R_GHMT2rm	R_GLYCLm	R_r0060	R_ASNN
R_ASNS1	R_ASPTA	R_r0399	R_TYRTA	R_34HPPOR
R_HGNTOR	R_MACACIr	R_FUMAC	R_AACOAT	R_ACACT1r
R_VALTAm	R_OIVD2m	R_ACOAD9m	R_ECOAH12m	R_3HBCOAHLm
R_HIBDm	R_MMTSADm	R_MMMm	R_METAT	R_GNMT
R_AHC	R_CYSTS	R_CYSTGL	R_OBDHc	R_PPCOACm
R_MMEm	R_HISD	R_URCN	R_IZPN	R_GluForTx
R_ARGNm	R_ORNTArm	R_r0074	R_PRO1xm	R_G5SADrm
R_GLU5Km	R_G5SDym	R_P5CRm	R_LEUTAm	R_OIVD1m
R_ACOAD8m	R_MCCCrm	R_MGCHrm	R_HMGLm	R_SACCD3m
R_SACCD4m	R_AASAD3m	R_r0450	R_2OXOADOXm	R_GLUTCOADHm
R_ECOAH1m	R_HACD1m	R_ILETAm	R_OIVD3m	R_ACOAD10m
R_HACD9m	R_ACACT10m	R_THRA	R_ALDD2x	R_ACS
R_G6PDH2r	R_PGL	R_GND	R_RPI	R_biomass_cho_producing
R_IgG1_production		R_DM_atp_LSQBKT_c_RSQBKT_		

TABLE C.1. Protected reactions for model reduction

APPENDIX D

CURRICULUM VITAE

2015 -	PhD student, University of Natural Resources and Life Sciences, Vienna
2013 - 2014	Master thesis, University of Natural Resources and Life Sciences, Vienna
2013	Master's internship Developmental Biology Institute, Seville
2012 - 2014	Master's degree in Bioinformatics and Modelling, Université Libre de Bruxelles
2005 - 2011	Five-year degree in Biotechnology, Universidad Pablo de Olavide, Seville

List of publications

Galleguillos SN, Ruckerbauer D, Gerstl MP, 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. 2017.

Galleguillos SN, Gerstl MP, Auer N, Borth N, Zanghellini J. Fast computation of multi-dimensional production envelopes in genome-scale metabolic models. Scientific Reports (submitted).

Galleguillos SN, Borth N, Zanghellini J. Finding energetically efficient pathways by Elementary Flux Mode analysis in a reduced CHO metabolic model. Manuscript in preparation.

Széliová D, Ruckerbauer DE, Galleguillos SN, Petersen LB, Natter K, Hanscho M, Troyer C, Panholzer S, Schoeny H, Christensen HB, Lee DY, Lewis NE, Koellensperger G, Hann S, Nielsen LK, Zanghellini J, Borth B. Quantitative analysis of biomass composition on various Chinese hamster ovary cell lines under different culture conditions and their impact on metabolic model predictions. Manuscript in preparation.

BIBLIOGRAPHY

[1] Butler M. Animal cell cultures: recent achievements and perspectives in the production of biopharmaceuticals. Applied microbiology and biotechnology. 2005;68(3):283–291.

[2] Kaufman RJ. Overview of vector design for mammalian gene expression. In: Recombinant Gene Expression Protocols. Springer; 1997. p. 287–300.

[3] Xu X, Nagarajan H, Lewis NE, Pan S, Cai Z, Liu X, et al. The genomic sequence of the Chinese hamster ovary (CHO)-K1 cell line. Nature biotechnology. 2011;29(8):735–741.

[4] Kyriakopoulos S, Kontoravdi C. Analysis of the landscape of biologically-derived pharmaceuticals in Europe: Dominant production systems, molecule types on the rise and approval trends. European Journal of Pharmaceutical Sciences. 2013;48(3):428–441.

[5] Walsh G. Biopharmaceutical benchmarks 2014. Nature biotechnology. 2014;32(10):992–1000.

[6] Ade N, Koirala Y, Mannan MS. Towards an inherently safer bioprocessing industry: A review. Journal of Loss Prevention in the Process Industries. 2019;.

[7] Ferreira AR, Dias JM, Teixeira AP, Carinhas N, Portela RM, Isidro IA, et al. Projection to latent pathways (PLP): a constrained projection to latent variables (PLS) method for elementary flux modes discrimination. BMC systems biology. 2011;5(1):181.

[8] Altamirano C, Illanes A, Becerra S, Cairó JJ, Gòdia F. Considerations on the lactate consumption by CHO cells in the presence of galactose. Journal of Biotechnology. 2006;125(4):547–556.

[9] Hammond S, Swanberg JC, Kaplarevic M, Lee KH. Genomic sequencing and analysis of a Chinese hamster ovary cell line using Illumina sequencing technology. BMC Genomics. 2011 Jan;12(1):67. Available from: https://doi.org/10.1186/1471-2164-12-67.

[10] Lewis NE, Liu X, Li Y, Nagarajan H, Yerganian G, O'brien E, et al. Genomic landscapes of Chinese hamster ovary cell lines as revealed by the Cricetulus griseus draft genome. Nature biotechnology. 2013;31(8):759.

[11] Shirsat NP, English NJ, Glennon B, Al-Rubeai M. Modelling of mammalian cell cultures. In: Animal cell culture. Springer; 2015. p. 259–326.

[12] Bordbar A, Monk JM, King ZA, Palsson BØ. Constraint-based models predict metabolic and associated cellular functions. Nature Reviews Genetics. 2014;15(2):107–120.

[13] Wurm FM. Production of recombinant protein therapeutics in cultivated mammalian cells. Nature biotechnology. 2004;22(11):1393.

[14] Ecker DM, Jones SD, Levine HL. The therapeutic monoclonal antibody market. In: MAbs. vol. 7. Taylor & Francis; 2015. p. 9–14.

[15] Derouazi M, Martinet D, Besuchet N, Flaction R, Wicht M, Bertschinger M, et al. Stability and Cytogenetic Characterization of Recombinant CHO Cell Lines Established by Microinjection and Calcium Phosphate Transfection. In: Smith R, editor. Cell Technology for Cell Products. Dordrecht: Springer Netherlands; 2007. p. 443–446.

[16] Deaven LL, Petersen DF. The chromosomes of CHO, an aneuploid Chinese hamster cell line: G-band, C-band, and autoradiographic analyses. Chromosoma. 1973 Jun;41(2):129–144. Available from: https://doi.org/10.1007/BF00319690.

[17] Kim NS, Kim SJ, Lee GM. Clonal variability within dihydrofolate reductase-mediated gene amplified Chinese hamster ovary cells: Stability in the absence of selective pressure. Biotechnology and bioengineering. 1998;60(6):679–688.

[18] Pilbrough W, Munro TP, Gray P. Intraclonal protein expression heterogeneity in recombinant CHO cells. PloS one. 2009;4(12):e8432.

[19] Chen C, Le H, Goudar CT. Integration of systems biology in cell line and process development for biopharmaceutical manufacturing. Biochemical Engineering Journal. 2016;107:11–17.

[20] Hammond S, Kaplarevic M, Borth N, Betenbaugh MJ, Lee KH. Chinese hamster genome database: an online resource for the CHO community at www. CHOgenome. org. Biotechnology and bioengineering. 2012;109(6):1353–1356.

[21] Palsson BØ. Systems biology: properties of reconstructed networks. Cambridge university press; 2006.

[22] Li F, Vijayasankaran N, Shen AY, Kiss R, Amanullah A. Cell culture processes for monoclonal antibody production. MAbs. 2010;2(5):466–479.

[23] Hu W, Dodge T, Frame K, Himes V. Effect of glucose on the cultivation of mammalian cells. Developments in biological standardization. 1986;66:279–290.

[24] 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(1-2):113–128.

[25] Glacken M, Fleischaker R, Sinskey A. 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):1376–1389.

[26] Schneider M, Marison IW, von Stockar U. The importance of ammonia in mammalian cell culture. Journal of Biotechnology. 1996;46(3):161–185.

[27] Hansen HA, Emborg C. Influence of ammonium on growth, metabolism, and productivity of a continuous suspension Chinese hamster ovary cell culture. Biotechnology progress. 1994;10(1):121–124.

[28] Andersen D, Goochee CF. 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.

[29] Yang M, Butler M. Effects of ammonia on CHO cell growth, erythropoietin production, and glycosylation. Biotechnology and bioengineering. 2000;68(4):370–380.

[30] Ljunggren J, Häggström L. Catabolic control of hybridoma cells by glucose and glutamine limited fed batch cultures. Biotechnology and bioengineering. 1994;44(7):808–818.

[31] Xie L, Wang DI. Fed-batch cultivation of animal cells using different medium design concepts and feeding strategies. Biotechnology and bioengineering. 1994;43(11):1175–1189.

[32] Altamirano C, Paredes C, Cairo J, Godia F. Improvement of CHO cell culture medium formulation: simultaneous substitution of glucose and glutamine. Biotechnology progress. 2000;16(1):69–75.

[33] Zhou M, Crawford Y, Ng D, Tung J, Pynn AF, Meier A, et al. 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.

[34] Kim SH, Lee GM. Functional expression of human pyruvate carboxylase for reduced lactic acid formation of Chinese hamster ovary cells (DG44). Applied microbiology and biotechnology. 2007;76(3):659–665.

[35] Sinclair AM, Elliott S. Glycoengineering: the effect of glycosylation on the properties of therapeutic proteins. Journal of pharmaceutical sciences. 2005;94(8):1626–1635.

[36] Borys MC, Linzer DI, Papoutsakis ET. 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.

[37] Trummer E, Fauland K, Seidinger S, Schriebl K, Lattenmayer C, Kunert R, et al. Process parameter shifting: Part I. Effect of DOT, pH, and temperature on the performance of Epo-Fc expressing CHO cells cultivated in controlled batch bioreactors. Biotechnology and bioengineering. 2006;94(6):1033–1044.

[38] Yoon SK, Song JY, Lee GM. Effect of low culture temperature on specific productivity, transcription level, and heterogeneity of erythropoietin in Chinese hamster ovary cells. Biotechnology and bioengineering. 2003;82(3):289–298.

[39] Jenkins N, Parekh RB, James DC. Getting the glycosylation right: implications for the biotechnology industry. Nature biotechnology. 1996;14(8):975.

[40] Stephanopoulos G, Alper H, Moxley J. Exploiting biological complexity for strain improvement through systems biology. Nature biotechnology. 2004;22(10):1261.

[41] Clarke C, Barron N, Meleady P, Clynes M. Statistical methods for mining Chinese hamster ovary cell 'omics data: from differential expression to integrated multilevel analysis of the biological system. Pharmaceutical Bioprocessing. 2014;2(5):469–481.

[42] Kildegaard HF, Baycin-Hizal D, Lewis NE, Betenbaugh MJ. The emerging CHO systems biology era: harnessing the 'omics revolution for biotechnology. Current opinion in biotechnology. 2013;24(6):1102–1107.

[43] Yu LX. Pharmaceutical Quality by Design: Product and Process Development, Understanding, and Control. Pharmaceutical Research. 2008 Apr;25(4):781–791. Available from: https://doi.org/10.1007/s11095-007-9511-1.

[44] Lawrence XY, Amidon G, Khan MA, Hoag SW, Polli J, Raju G, et al. Understanding pharmaceutical quality by design. The AAPS journal. 2014;16(4):771–783.

[45] Lawrence XY. Pharmaceutical quality by design: product and process development, understanding, and control. Pharmaceutical research. 2008;25(4):781–791.

[46] Rocha I, Förster J, Nielsen J. Design and application of genome-scale reconstructed metabolic models. In: Microbial Gene Essentiality: Protocols and Bioinformatics. Springer; 2008. p. 409–431.

[47] Thiele I, Palsson BØ. A protocol for generating a high-quality genome-scale metabolic reconstruction. Nature protocols. 2010;5(1):93.

[48] Dias O, Rocha M, Ferreira EC, Rocha I. Reconstructing genome-scale metabolic models with merlin. Nucleic acids research. 2015;43(8):3899–3910.

[49] Machado D, Andrejev S, Tramontano M, Patil KR. Fast automated reconstruction of genome-scale metabolic models for microbial species and communities. Nucleic acids research. 2018;46(15):7542–7553.

[50] Feist AM, Herrgård MJ, Thiele I, Reed JL, Palsson BØ. Reconstruction of biochemical networks in microorganisms. Nature Reviews Microbiology. 2009;7(2):129.

[51] Lieven C, Beber ME, Olivier BG, Bergmann FT, Ataman M, Babaei P, et al. Memote: A community-driven effort towards a standardized genome-scale metabolic model test suite. BioRxiv. 2018;p. 350991.

[52] King ZA, Lu J, Dräger A, Miller P, Federowicz S, Lerman JA, et al. BiGG Models: A platform for integrating, standardizing and sharing genome-scale models. Nucleic acids research. 2015;44(D1):D515–D522.

[53] Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. Nucleic acids research. 2000;28(1):27–30.

[54] Karp PD, Riley M, Paley SM, Pellegrini-Toole A. The metacyc database. Nucleic acids research. 2002;30(1):59–61.

[55] Heavner BD, Price ND. Transparency in metabolic network reconstruction enables scalable biological discovery. Current opinion in biotechnology. 2015;34:105–109.

[56] Bernard T, Bridge A, Morgat A, Moretti S, Xenarios I, Pagni M. Reconciliation of metabolites and biochemical reactions for metabolic networks. Briefings in bioinformatics. 2012;15(1):123–135.

[57] Hucka M, Finney A, Sauro HM, Bolouri H, Doyle JC, Kitano H, et al. The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. Bioinformatics. 2003;19(4):524–531.

[58] Voll LM, Nikoloski Z. Engineering Synthetic Metabolons: From Metabolic Modelling to Rational Design of Biosynthetic Devices. Frontiers Research Topics. Frontiers Media SA; 2016. Available from: `https://books.google.at/books?id=YhUvDwAAQBAJ`.

[59] Büchel F, Rodriguez N, Swainston N, Wrzodek C, Czauderna T, Keller R, et al. Path2Models: large-scale generation of computational models from biochemical pathway maps. BMC Systems Biology. 2013 Nov;7:116.

[60] Schellenberger J, Que R, Fleming RM, Thiele I, Orth JD, Feist AM, et al. Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2. 0. Nature protocols. 2011;6(9):1290.

[61] Agren R, Liu L, Shoaie S, Vongsangnak W, Nookaew I, Nielsen J. The RAVEN toolbox and its use for generating a genome-scale metabolic model for Penicillium chrysogenum. PLoS computational biology. 2013;9(3):e1002980.

[62] Rocha I, Maia P, Evangelista P, Vilaça P, Soares S, Pinto JP, et al. OptFlux: an open-source software platform for in silico metabolic engineering. BMC systems biology. 2010;4(1):45.

[63] Gombert AK, Nielsen J. Mathematical modelling of metabolism. Current opinion in biotechnology. 2000;11(2):180–186.

[64] Edwards JS, Covert M, Palsson B. Metabolic modelling of microbes: the flux-balance approach. Environmental microbiology. 2002;4(3):133–140.

[65] Thiele I, Price ND, Vo TD, Palsson BØ. Candidate metabolic network states in human mitochondria impact of diabetes, ischemia, and diet. Journal of Biological Chemistry. 2005;280(12):11683–11695.

[66] Kell DB. Systems biology, metabolic modelling and metabolomics in drug discovery and development. Drug discovery today. 2006;11(23-24):1085–1092.

[67] Folger O, Jerby L, Frezza C, Gottlieb E, Ruppin E, Shlomi T. Predicting selective drug targets in cancer through metabolic networks. Molecular systems biology. 2011;7(1):501.

[68] Stelling J. Mathematical models in microbial systems biology. Current opinion in microbiology. 2004;7(5):513–518.

[69] Hefzi H, Ang KS, Hanscho M, Bordbar A, Ruckerbauer D, Lakshmanan M, et al. A consensus genome-scale reconstruction of Chinese hamster ovary cell metabolism. Cell systems. 2016;3(5):434–443.

[70] Wu P, Ray N, Shuler M. A Single-Cell Model for CHO Cellsa. Annals of the New York Academy of Sciences. 1992;665(1):152–187.

[71] Provost A, Bastin G, Agathos SN, Schneider YJ. Metabolic design of macroscopic bioreaction models: application to Chinese hamster ovary cells. Bioprocess and Biosystems Engineering. 2006 Dec;29(5):349–366. Available from: https://doi.org/10.1007/s00449-006-0083-y.

[72] Selvarasu S, Ho YS, Chong WP, Wong NS, Yusufi FN, Lee YY, et al. Combined in silico modeling and metabolomics analysis to characterize fed-batch CHO cell culture. Biotechnology and bioengineering. 2012;109(6):1415–1429.

[73] Galleguillos SN, Ruckerbauer D, Gerstl MP, 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.

[74] Stephanopoulos G, Aristidou AA, Nielsen J. Metabolic engineering: principles and methodologies. Elsevier; 1998.

[75] Varma A, Palsson BO. Metabolic Flux Balancing: Basic Concepts, Scientific and Practical Use. Bio/technology. 1994;12:.

[76] Orth JD, Thiele I, Palsson BØ. What is flux balance analysis? Nature biotechnology. 2010;28(3):245.

[77] Wiechert W. 13 C metabolic flux analysis. Metabolic engineering. 2001;3(3):195–206.

[78] Hastings WK. Monte Carlo sampling methods using Markov chains and their applications. Biometrika. 1970;57(1):97–109.

[79] Shanmuganathan S, Samarasinghe S. Artificial Neural Network Modelling. Studies in Computational Intelligence. Springer International Publishing; 2016.

[80] Baggetto L. Deviant energetic metabolism of glycolytic cancer cells. Biochimie. 1992;74(11):959–974.

[81] Martínez VS, Dietmair S, Quek LE, Hodson MP, Gray P, Nielsen LK. 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.

[82] Zamorano F, Wouwer AV, Bastin G. A detailed metabolic flux analysis of an underdetermined network of CHO cells. Journal of biotechnology. 2010;150(4):497–508.

[83] Sengupta N, Rose ST, Morgan JA. Metabolic flux analysis of CHO cell metabolism in the late non-growth phase. Biotechnology and bioengineering. 2011;108(1):82–92.

[84] Templeton N, Dean J, Reddy P, Young JD. Peak antibody production is associated with increased oxidative metabolism in an industrially relevant fed-batch CHO cell culture. Biotechnology and bioengineering. 2013;110(7):.

[85] Chen N, Bennett MH, Kontoravdi C. Analysis of Chinese hamster ovary cell metabolism through a combined computational and experimental approach. Cytotechnology. 2014;66(6):945–966.

[86] Li J, Wong CL, 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.

[87] Wilkens CA, Altamirano C, Gerdtzen ZP. Comparative metabolic analysis of lactate for CHO cells in glucose and galactose. Biotechnology and Bioprocess Engineering. 2011;16(4):714–724.

[88] Nicolae A, Wahrheit J, Bahnemann J, Zeng AP, Heinzle E. Non-stationary 13 C metabolic flux analysis of Chinese hamster ovary cells in batch culture using extracellular labeling highlights metabolic reversibility and compartmentation. BMC systems biology. 2014;8(1):1.

[89] Wahrheit J, Niklas J, Heinzle E. Metabolic control at the cytosol–mitochondria interface in different growth phases of CHO cells. Metabolic engineering. 2014;23:9–21.

[90] Ahn WS, Antoniewicz MR. 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.

[91] Ahn WS, Antoniewicz MR. Parallel labeling experiments with [1, 2-13C] glucose and [U-13C] glutamine provide new insights into CHO cell metabolism. Metabolic engineering. 2013;15:34–47.

[92] Altamirano C, Illanes A, Casablancas A, Gámez X, Cairó J, Gòdia C. Analysis of CHO Cells Metabolic Redistribution in a Glutamate-Based Defined Medium in Continuous Culture. Biotechnology Progress. 2001;17(6):1032–1041.

[93] Xing Z, Kenty B, Koyrakh I, Borys M, Pan SH, Li ZJ. Optimizing amino acid composition of CHO cell culture media for a fusion protein production. Process Biochemistry. 2011;46(7):1423–1429.

[94] Deshpande R, Yang TH, Heinzle E. Towards a metabolic and isotopic steady state in CHO batch cultures for reliable isotope-based metabolic profiling. Biotechnology journal. 2009;4(2):247–263.

[95] Nolan RP, Lee K. Dynamic model of CHO cell metabolism. Metabolic engineering. 2011;13(1):108–124.

[96] Goudar C, Biener R, Zhang C, Michaels J, Piret J, Konstantinov K. Towards industrial application of quasi real-time metabolic flux analysis for mammalian cell culture. In: Cell culture engineering. Springer; 2006. p. 99–118.

[97] Xing Z, Bishop N, Leister K, Li ZJ. Modeling kinetics of a large-scale fed-batch CHO cell culture by Markov chain Monte Carlo method. Biotechnology progress. 2010;26(1):208–219.

[98] Wahrheit J, Nicolae A, Heinzle E. Dynamics of growth and metabolism controlled by glutamine availability in Chinese hamster ovary cells. Applied microbiology and biotechnology. 2014;98(4):1771–1783.

[99] Sheikholeslami Z, Jolicoeur M, Henry O. Elucidating the effects of postinduction glutamine feeding on the growth and productivity of CHO cells. Biotechnology progress. 2014;30(3):535–546.

[100] Sheikholeslami Z, Jolicoeur M, Henry O. Probing the metabolism of an inducible mammalian expression system using extracellular isotopomer analysis. Journal of biotechnology. 2013;164(4):469–478.

[101] Naderi S, Meshram M, Wei C, McConkey B, Ingalls B, Budman H, et al. Development of a mathematical model for evaluating the dynamics of normal and apoptotic Chinese hamster ovary cells. Biotechnology progress. 2011;27(5):1197–1205.

[102] McMurray-Beaulieu V, Hisiger S, Durand C, Perrier M, Jolicoeur M. Na-butyrate sustains energetic states of metabolism and t-PA productivity of CHO cells. Journal of bioscience and bioengineering. 2009;108(2):160–167.

[103] Ghorbaniaghdam A, Henry O, Jolicoeur M. A kinetic-metabolic model based on cell energetic state: study of CHO cell behavior under Na-butyrate stimulation. Bioprocess and biosystems engineering. 2013;36(4):469–487.

[104] Becker J, Hackl M, Rupp O, Jakobi T, Schneider J, Szczepanowski R, et al. Unraveling the Chinese hamster ovary cell line transcriptome by next-generation sequencing. Journal of biotechnology. 2011;156(3):227–235.

[105] Baycin-Hizal D, Tabb DL, Chaerkady R, Chen L, Lewis NE, Nagarajan H, et al. Proteomic analysis of Chinese hamster ovary cells. Journal of proteome research. 2012;11(11):5265–5276.

[106] Lewis NE, Nagarajan H, Palsson BO. Constraining the metabolic genotype–phenotype relationship using a phylogeny of in silico methods. Nature Reviews Microbiology. 2012;10(4):291.

[107] Nyberg GB, Balcarcel RR, Follstad BD, Stephanopoulos G, Wang DI. Metabolism of peptide amino acids by Chinese hamster ovary cells grown in a complex medium. Biotechnology and bioengineering. 1999;62(3):324–335.

[108] Dikicioglu D, Kırdar B, Oliver SG. Biomass composition: the “elephant in the room” of metabolic modelling. Metabolomics. 2015;11(6):1690–1701.

[109] Durocher Y, Butler M. Expression systems for therapeutic glycoprotein production. Current Opinion in Biotechnology. 2009;20(6):700–707.

[110] Lu D, Yang C, Liu Z. How hydrophobicity and the glycosylation site of glycans affect protein folding and stability: a molecular dynamics simulation. The Journal of Physical Chemistry B. 2012 Jan;116(1):390–400.

[111] Jefferis R. Glycosylation as a strategy to improve antibody-based therapeutics. Nature Reviews Drug Discovery. 2009;8(3):226–234.

[112] Cumming DA. Glycosylation of recombinant protein therapeutics: control and functional implications. Glycobiology. 1991;1(2):115–130.

[113] Varki A. Biological roles of oligosaccharides: all of the theories are correct. Glycobiology. 1993;3(2):97–130.

[114] Andersen DC, Goochee CF. The effect of cell-culture conditions on the oligosaccharide structures of secreted glycoproteins. Current Opinion in Biotechnology. 1994;5(5):546–549.

[115] Hossler P, Khattak SF, Li ZJ. Optimal and consistent protein glycosylation in mammalian cell culture. Glycobiology. 2009;19(9):936–949.

[116] Kornfeld R, Kornfeld S. Assembly of asparagine-linked oligosaccharides. Annual review of biochemistry. 1985;54(1):631–664.

[117] Shelikoff M, Sinskey A, Stephanopoulos G. A modeling framework for the study of protein glycosylation. Biotechnology and bioengineering. 1996;50(1):73–90.

[118] Monica TJ, Andersen DC, Goochee CF. A mathematical model of sialylation of N-linked oligosaccharides in the trans-Golgi network. Glycobiology. 1997;7(4):515–521.

[119] Umaña P, Bailey JE. A mathematical model of N-linked glycoform biosynthesis. Biotechnology and bioengineering. 1997;55(6):890–908.

[120] Kontoravdi C, Asprey SP, Pistikopoulos EN, 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.

[121] Krambeck FJ, Betenbaugh MJ. A mathematical model of N-linked glycosylation. Biotechnology and Bioengineering. 2005;92(6):711–728.

[122] Jenkins N, Curling EM. Glycosylation of recombinant proteins: problems and prospects. Enzyme and microbial technology. 1994;16(5):354–364.

[123] Weikert S, Papac D, Briggs J, Cowfer D, Tom S, Gawlitzek M, et al. Engineering Chinese hamster ovary cells to maximize sialic acid content of recombinant glycoproteins. Nature biotechnology. 1999;17(11):1116–1121.

[124] Inoue N, Watanabe T, Kutsukake T, Saitoh H, Tsumura H, Arai H, et al. Asn-linked sugar chain structures of recombinant human thrombopoietin produced in Chinese hamster ovary cells. Glycoconjugate journal. 1999;16(11):707–718.

[125] Krambeck FJ, Bennun SV, Narang S, Choi S, Yarema KJ, Betenbaugh MJ. A mathematical model to derive N-glycan structures and cellular enzyme activities from mass spectrometric data. Glycobiology. 2009;19(11):1163–1175.

[126] Lau KS, Partridge EA, Grigorian A, Silvescu CI, Reinhold VN, Demetriou M, et al. Complex N-glycan number and degree of branching cooperate to regulate cell proliferation and differentiation. Cell. 2007;129(1):123–134.

[127] McDonald AG, Hayes JM, Bezak T, Głuchowska SA, Cosgrave EF, Struwe WB, et al. Galactosyltransferase 4 is a major control point for glycan branching in N-linked glycosylation. J Cell Sci. 2014;127(23):5014–5026.

[128] Senger RS, Karim MN. 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.

[129] Senger RS, Karim MN. Variable Site-Occupancy Classification of N-Linked Glycosylation Using Artificial Neural Networks. Biotechnology progress. 2005;21(6):1653–1662.

[130] Senger RS, Karim MN. Prediction of N-linked glycan branching patterns using artificial neural networks. Mathematical biosciences. 2008;211(1):89–104.

[131] Hansen JE, Lund O, Tolstrup N, Gooley AA, Williams KL, Brunak S. NetOglyc: prediction of mucin type O-glycosylation sites based on sequence context and surface accessibility. Glycoconjugate journal. 1998;15(2):115–130.

[132] Julenius K, Mølgaard A, Gupta R, Brunak S. Prediction, conservation analysis, and structural characterization of mammalian mucin-type O-glycosylation sites. Glycobiology. 2005;15(2):153–164.

[133] Gerken TA. Kinetic modeling confirms the biosynthesis of mucin core 1 (β-Gal (1-3) α-GalNAc-O-Ser/Thr) O-glycan structures are modulated by neighboring glycosylation effects. Biochemistry. 2004;43(14):4137–4142.

[134] 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.

[135] Suga A, Yamanishi Y, Hashimoto K, Goto S, Kanehisa M. An improved scoring scheme for predicting glycan structures from gene expression data. Genome Informatics. 2007;18:237–246.

[136] Bennun SV, Yarema KJ, Betenbaugh MJ, Krambeck FJ. Integration of the transcriptome and glycome for identification of glycan cell signatures. PLoS Comput Biol. 2013;9(1):e1002813.

[137] Hossler P, Mulukutla BC, Hu WS. Systems analysis of N-glycan processing in mammalian cells. PloS one. 2007;2(8):e713.

[138] Jimenez del Val I, Nagy JM, 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.

[139] Kaveh O, Hengameh A, Johannes G, Murray MY, Raymond LL, Jeno S, et al. Novel dynamic model to predict the glycosylation pattern of monoclonal antibodies from extracellular cell culture conditions. IFAC Proceedings Volumes. 2013;46(31):30–35.

[140] Aghamohseni H, Ohadi K, Spearman M, Krahn N, Moo-Young M, Scharer JM, et al. Effects of nutrient levels and average culture pH on the glycosylation pattern of camelid-humanized monoclonal antibody. Journal of biotechnology. 2014;186:98–109.

[141] Hossler P, Goh LT, Lee MM, Hu WS. GlycoVis: Visualizing glycan distribution in the protein N-glycosylation pathway in mammalian cells. Biotechnology and bioengineering. 2006;95(5):946–960.

[142] Jedrzejewski PM, del Val IJ, Constantinou A, Dell A, Haslam SM, Polizzi KM, et al. Towards controlling the glycoform: a model framework linking extracellular metabolites to antibody glycosylation. International journal of molecular sciences. 2014;15(3):4492–4522.

[143] Villiger TK, 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.

[144] 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.

[145] Liu G, Marathe DD, Matta KL, Neelamegham S. Systems-level modeling of cellular glycosylation reaction networks: O-linked glycan formation on natural selectin ligands. Bioinformatics. 2008;24(23):2740–2747.

[146] Liu G, Puri A, Neelamegham S. Glycosylation Network Analysis Toolbox: a MATLAB-based environment for systems glycobiology. Bioinformatics. 2013;29(3):404–406.

[147] Liu G, Neelamegham S. A computational framework for the automated construction of glycosylation reaction networks. PloS one. 2014;9(6):e100939.

[148] Hou W, Qiu Y, Hashimoto N, Ching WK, Aoki-Kinoshita KF. A systematic framework to derive N-glycan biosynthesis process and the automated construction of glycosylation networks. BMC bioinformatics. 2016;17(7):240.

[149] Kim PJ, Lee DY, Jeong H. Centralized modularity of N-linked glycosylation pathways in mammalian cells. PloS one. 2009;4(10):e7317.

[150] Sou SN, Sellick C, Lee K, Mason A, Kyriakopoulos S, Polizzi KM, et al. How does mild hypothermia affect monoclonal antibody glycosylation? Biotechnology and bioengineering. 2015;112(6):1165–1176.

[151] Burleigh SC, van de Laar T, Stroop CJ, van Grunsven WM, O'Donoghue N, Rudd PM, et al. Synergizing metabolic flux analysis and nucleotide sugar metabolism to understand the control of glycosylation of recombinant protein in CHO cells. BMC biotechnology. 2011;11(1):1.

[152] Del Val IJ, Polizzi KM, Kontoravdi C. A theoretical estimate for nucleotide sugar demand towards Chinese Hamster Ovary cellular glycosylation. Scientific Reports. 2016;6(28547):610–623.

[153] Spahn PN, Hansen AH, Hansen HG, Arnsdorf J, Kildegaard HF, Lewis NE. A Markov chain model for N-linked protein glycosylation–towards a low-parameter tool for model-driven glycoengineering. Metabolic engineering. 2016;33:52–66.

[154] Imai-Nishiya H, Mori K, Inoue M, Wakitani M, Iida S, Shitara K, et al. 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):1.

[155] Taniguchi N, Honke K, Fukuda M. Handbook of glycosyltransferases and related genes. Springer Science & Business Media; 2011.

[156] Hirschberg K, Lippincott-Schwartz J. Secretory pathway kinetics and in vivo analysis of protein traffic from the Golgi complex to the cell surface. FASEB journal: official publication of the Federation of American Societies for Experimental Biology. 1999;13:S251.

[157] Marathe DD, Chandrasekaran E, Lau JT, Matta KL, Neelamegham S. Systems-level studies of glycosyltransferase gene expression and enzyme activity that are associated with the selectin binding function of human leukocytes. The FASEB Journal. 2008;22(12):4154–4167.

[158] Yang M, Brazier M, Edwards R, Davis BG. High-Throughput Mass-Spectrometry Monitoring for Multisubstrate Enzymes: Determining the Kinetic Parameters and Catalytic Activities of Glycosyltransferases. ChemBioChem. 2005;6(2):346–357.

[159] Zhang P, Chan KF, Haryadi R, Bardor M, Song Z. CHO glycosylation mutants as potential host cells to produce therapeutic proteins with enhanced efficacy. In: Future Trends in Biotechnology. Springer; 2012. p. 63–87.

[160] Thaysen-Andersen M, Packer NH. Site-specific glycoproteomics confirms that protein structure dictates formation of N-glycan type, core fucosylation and branching. Glycobiology. 2012;22(11):1440–1452.

[161] Covert MW, Famili I, Palsson BO. Identifying constraints that govern cell behavior: a key to converting conceptual to computational models in biology? Biotechnology and bioengineering. 2003;84(7):763–772.

[162] Klamt S, Regensburger G, Gerstl MP, Jungreuthmayer C, Schuster S, Mahadevan R, et al. From elementary flux modes to elementary flux vectors: Metabolic pathway analysis with arbitrary linear flux constraints. PLOS Computational Biology. 2017;13(4):e1005409.

[163] Müller S, Regensburger G. Elementary Vectors and Conformal Sums in Polyhedral Geometry and their Relevance for Metabolic Pathway Analysis. Frontiers in Genetics. 2016;7:90.

[164] Zanghellini J, Ruckerbauer DE, Hanscho M, Jungreuthmayer C. Elementary flux modes in a nutshell: properties, calculation and applications. Biotechnology journal. 2013;8(9):1009–1016.

[165] Klamt S, Stelling J. Combinatorial Complexity of Pathway Analysis in Metabolic Networks. Molecular Biology Reports. 2002;29(1):233–236.

[166] Edwards JS, Ibarra RU, Palsson BO. In silico predictions of Escherichia coli metabolic capabilities are consistent with experimental data. Nature biotechnology. 2001;19(2):125.

[167] Ibarra RU, Edwards JS, Palsson BO. Escherichia coli K-12 undergoes adaptive evolution to achieve in silico predicted optimal growth. Nature. 2002;420(6912):186.

[168] Feist AM, Palsson BO. The biomass objective function. Current opinion in microbiology. 2010;13(3):344–349.

[169] Antoniewicz MR. Methods and advances in metabolic flux analysis: a mini-review. Journal of Industrial Microbiology & Biotechnology. 2015 Mar;42(3):317–325. Available from: https://doi.org/10.1007/s10295-015-1585-x.

[170] Becker SA, Palsson BO. Context-specific metabolic networks are consistent with experiments. PLoS computational biology. 2008;4(5):e1000082.

[171] Mahadevan R, Schilling C. The effects of alternate optimal solutions in constraint-based genome-scale metabolic models. Metabolic engineering. 2003;5(4):264–276.

[172] Thiele I, Fleming RM, Bordbar A, Schellenberger J, Palsson BØ. Functional characterization of alternate optimal solutions of Escherichia coli's transcriptional and translational machinery. Biophysical journal. 2010;98(10):2072–2081.

[173] Puchałka J, Oberhardt MA, Godinho M, Bielecka A, Regenhardt D, Timmis KN, et al. Genome-scale reconstruction and analysis of the Pseudomonas putida KT2440 metabolic network facilitates applications in biotechnology. PLoS computational biology. 2008;4(10):e1000210.

[174] Segura D, Mahadevan R, Juárez K, Lovley DR. Computational and experimental analysis of redundancy in the central metabolism of Geobacter sulfurreducens. PLoS computational biology. 2008;4(2):e36.

[175] Lewis NE, Hixson KK, Conrad TM, Lerman JA, Charusanti P, Polpitiya AD, et al. Omic data from evolved E. coli are consistent with computed optimal growth from genome-scale models. Molecular systems biology. 2010;6(1):390.

[176] Burgard AP, Pharkya P, Maranas CD. Optknock: a bilevel programming framework for identifying gene knockout strategies for microbial strain optimization. Biotechnology and bioengineering. 2003;84(6):647–657.

[177] Tepper N, Shlomi T. Predicting metabolic engineering knockout strategies for chemical production: accounting for competing pathways. Bioinformatics. 2009;26(4):536–543.

[178] Ranganathan S, Suthers PF, Maranas CD. OptForce: an optimization procedure for identifying all genetic manipulations leading to targeted overproductions. PLoS computational biology. 2010;6(4):e1000744.

[179] Choi HS, Lee SY, Kim TY, Woo HM. In Silico Identification of Gene Amplification Targets for Improvement of Lycopene Production. Applied and Environmental Microbiology. 2010;76(10):3097–3105. Available from: https://aem.asm.org/content/76/10/3097.

[180] Schuster S, Hilgetag C. On elementary flux modes in biochemical reaction systems at steady state. Journal of Biological Systems. 1994;2(02):165–182.

[181] Klamt S, Müller S, Regensburger G, Zanghellini J. A mathematical framework for yield (vs. rate) optimization in constraint-based modeling and applications in metabolic engineering. Metabolic Engineering. 2018;47:153 – 169. Available from: http://www.sciencedirect.com/science/article/pii/S1096717617303853.

[182] Stelling J, Klamt S, Bettenbrock K, Schuster S, Gilles ED. Metabolic network structure determines key aspects of functionality and regulation. Nature. 2002;420(6912):190.

[183] Trinh CT, Carlson R, Wlaschin A, Srienc F. Design, construction and performance of the most efficient biomass producing E. coli bacterium. Metabolic engineering. 2006;8(6):628–638.

[184] Trinh CT, Unrean P, Srienc F. Minimal Escherichia coli cell for the most efficient production of ethanol from hexoses and pentoses. Applied and environmental microbiology. 2008;74(12):3634–3643.

[185] Urbanczik R. Enumerating constrained elementary flux vectors of metabolic networks. IET systems biology. 2007;1(5):274–279.

[186] Wiback SJ, Famili I, Greenberg HJ, Palsson BØ. Monte Carlo sampling can be used to determine the size and shape of the steady-state flux space. Journal of theoretical biology. 2004;228(4):437–447.

[187] Klamt S, Müller S, Regensburger G, Zanghellini J. A mathematical framework for yield (vs. rate) optimization in constraint-based modeling and applications in metabolic engineering. Metabolic Engineering. 2018;47:153–169.

[188] Gudmundsson S, Thiele I. Computationally efficient flux variability analysis. BMC Bioinformatics. 2010;11:489.

[189] Mahadevan R, Schilling CH. The effects of alternate optimal solutions in constraint-based genome-scale metabolic models. Metabolic Engineering. 2003;5:264–276.

[190] von Kamp A, Thiele S, Hädicke O, Klamt S. Use of CellNetAnalyzer in biotechnology and metabolic engineering. Journal of Biotechnology. 2017;261:221–228.

[191] Makhorin A. The GNU Linear Programming Kit (GLPK); 2012. `https://www.gnu.org/software/glpk/glpk.html`.

[192] Carinhas N, Duarte TM, Barreiro LC, Carrondo MJ, Alves PM, Teixeira AP. Metabolic signatures of GS-CHO cell clones associated with butyrate treatment and culture phase transition. Biotechnology and bioengineering. 2013;110(12):3244–3257.

[193] Martínez VS, Buchsteiner M, Gray P, Nielsen LK, Quek LE. 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.

[194] Nocon J, Steiger M, Mairinger T, Hohlweg J, Rußmayer H, Hann S, et al. Increasing pentose phosphate pathway flux enhances recombinant protein production in Pichia pastoris. Applied microbiology and biotechnology. 2016;100(13):5955–5963.

[195] Meleady P, Henry M, Gammell P, Doolan P, Sinacore M, Melville M, et al. Proteomic profiling of CHO cells with enhanced rhBMP-2 productivity following co-expression of PACEsol. Proteomics. 2008;8(13):2611–2624.

[196] Carlage T, Hincapie M, Zang L, Lyubarskaya Y, Madden H, Mhatre R, et al. Proteomic profiling of a high-producing Chinese hamster ovary cell culture. Analytical chemistry. 2009;81(17):7357–7362.

[197] Schaub J, Clemens C, Schorn P, Hildebrandt T, Rust W, Mennerich D, et al. CHO gene expression profiling in biopharmaceutical process analysis and design. Biotechnology and bioengineering. 2010;105(2):431–438.

[198] Jeon SM, Chandel NS, Hay N. AMPK regulates NADPH homeostasis to promote tumour cell survival during energy stress. Nature. 2012;485(7400):661.

[199] Reitzer LJ, Wice BM, Kennell D. Evidence that glutamine, not sugar, is the major energy source for cultured HeLa cells. Journal of Biological Chemistry. 1979;254(8):2669–2676.

[200] Zielke HR, Zielke CL, Ozand PT. Glutamine: a major energy source for cultured mammalian cells. In: Federation proceedings. vol. 43; 1984. p. 121–125.

[201] Glacken MW. Catabolic control of mammalian cell culture. Bio/technology. 1988;6(9):1041.

[202] Klamt S, Saez-Rodriguez J, Gilles ED. Structural and functional analysis of cellular networks with CellNetAnalyzer. BMC systems biology. 2007;1(1):2.

[203] Erdrich P, Steuer R, Klamt S. An algorithm for the reduction of genome-scale metabolic network models to meaningful core models. BMC systems biology. 2015;9(1):48.

[204] Terzer M, Stelling J. Large-scale computation of elementary flux modes with bit pattern trees. Bioinformatics. 2008;24(19):2229–2235.

[205] Lassez C. Quantifier elimination for conjunctions of linear constraints via a convex hull algorithm. Symbolic and Numerical Computation for Artificial Intelligence. 1992;p. 103–122.

[206] Apte J. Convex Hull Method (chmlib); 2014. Available from: `https://sites.google.com/site/jayantapteshomepage/home/software`.

[207] Orth JD, Conrad TM, Na J, Lerman JA, Nam H, Feist AM, et al. A comprehensive genome-scale reconstruction of Escherichia coli metabolism – 2011. Molecular Systems Biology. 2011;7:535.

[208] Applegate DL, Cook W, Dash S, Espinoza DG. Exact solutions to linear programming problems. Operations Research Letters. 2007;35(6):693–699.

[209] Hädicke O, Klamt S. EColiCore2: a reference network model of the central metabolism of Escherichia coli and relationships to its genome-scale parent model. Scientific Reports. 2017;7:39647.

[210] Chang DE, Jung HC, Rhee JS, Pan JG. Homofermentative production of d-orl-lactate in metabolically engineered Escherichia coli RR1. Appl Environ Microbiol. 1999;65(4):1384–1389.

[211] Lao MS, 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.

[212] Walsh G, Jefferis R. Post-translational modifications in the context of therapeutic proteins. Nature biotechnology. 2006;24(10):1241.

[213] Kuystermans D, Krampe B, Swiderek H, Al-Rubeai M. Using cell engineering and omic tools for the improvement of cell culture processes. Cytotechnology. 2007;53(1-3):3–22.

[214] Chartrain M, Chu L. Development and production of commercial therapeutic monoclonal antibodies in Mammalian cell expression systems: an overview of the current upstream technologies. Current pharmaceutical biotechnology. 2008;9(6):447–467.

[215] Bailey JE, Sburlati A, Hatzimanikatis V, Lee K, Renner WA, Tsai PS. Inverse metabolic engineering: a strategy for directed genetic engineering of useful phenotypes. Biotechnology and Bioengineering. 1996;52(1):109–121.

[216] Price ND, Reed JL, Palsson BØ. Genome-scale models of microbial cells: evaluating the consequences of constraints. Nature Reviews Microbiology. 2004;2(11):886.

[217] Park JM, Park HM, Kim WJ, Kim HU, Kim TY, Lee SY. Flux variability scanning based on enforced objective flux for identifying gene amplification targets. BMC systems biology. 2012;6(1):106.

[218] Nocon J, Steiger MG, Pfeffer M, Sohn SB, Kim TY, Maurer M, et al. Model based engineering of Pichia pastoris central metabolism enhances recombinant protein production. Metabolic engineering. 2014;24:129–138.

[219] Hyduke DR, Lewis NE, Palsson BØ. Analysis of omics data with genome-scale models of metabolism. Molecular BioSystems. 2013;9(2):167–174.

[220] Liu JK, O'Brien EJ, Lerman JA, Zengler K, Palsson BO, Feist AM. Reconstruction and modeling protein translocation and compartmentalization in Escherichia coli at the genome-scale. BMC systems biology. 2014;8(1):110.

[221] Lerman JA, Hyduke DR, Latif H, Portnoy VA, Lewis NE, Orth JD, et al. In silico method for modelling metabolism and gene product expression at genome scale. Nature communications. 2012;3:929.

[222] Segre D, Vitkup D, Church GM. Analysis of optimality in natural and perturbed metabolic networks. Proceedings of the National Academy of Sciences. 2002;99(23):15112–15117.

[223] Trinh CT, Wlaschin A, Srienc F. Elementary mode analysis: a useful metabolic pathway analysis tool for characterizing cellular metabolism. Applied microbiology and biotechnology. 2009;81(5):813.

[224] Llaneras F, Picó J. Stoichiometric modelling of cell metabolism. Journal of bioscience and bioengineering. 2008;105(1):1–11.

[225] Papin JA, Price ND, Wiback SJ, Fell DA, Palsson BO. Metabolic pathways in the post-genome era. Trends in biochemical sciences. 2003;28(5):250–258.

[226] Kundu PK, Prasad N, Datta D. Monoclonal antibody: high density culture of hybridoma cells and downstream processing for IgG recovery. Indian journal of experimental biology. 1998;36(2):125–135.

[227] Edwards JS, Ramakrishna R, Palsson BO. Characterizing the metabolic phenotype: a phenotype phase plane analysis. Biotechnology and bioengineering. 2002;77(1):27–36.

[228] Biggs MB, Medlock GL, Kolling GL, Papin JA. Metabolic network modeling of microbial communities. Wiley Interdisciplinary Reviews: Systems Biology and Medicine. 2015;7(5):317–334.

[229] Mahadevan R, Henson MA. Genome-based modeling and design of metabolic interactions in microbial communities. Computational and structural biotechnology journal. 2012;3(4):e201210008.

[230] Koch S, Benndorf D, Fronk K, Reichl U, Klamt S. Predicting compositions of microbial communities from stoichiometric models with applications for the biogas process. Biotechnology for biofuels. 2016;9(1):17.

[231] Schirra S. Robustness and precision issues in geometric computation. 1998;.

[232] Kitano H. Systems biology: a brief overview. science. 2002;295(5560):1662–1664.

[233] Quek LE, Dietmair S, Krömer JO, Nielsen LK. Metabolic flux analysis in mammalian cell culture. Metabolic Engineering. 2010;12(2):161–171.

[234] Strnad J, Brinc M, Spudić V, Jelnikar N, Mirnik L, Čarman B, et al. Optimization of cultivation conditions in spin tubes for Chinese hamster ovary cells producing erythropoietin and the comparison of glycosylation patterns in different cultivation vessels. Biotechnology progress. 2010;26(3):653–663.