Titre: Title:	A Dynamic Constraint-Based Modelling Approach of Cell Metabolism
Auteur: Author:	Mohammadreza Yasemi
Date:	2022
Type:	Mémoire ou thèse / Dissertation or Thesis
Référence: Citation:	Yasemi, M. (2022). A Dynamic Constraint-Based Modelling Approach of Cell Metabolism [Ph.D. thesis, Polytechnique Montréal]. PolyPublie. https://publications.polymtl.ca/10552/

POLYTECHNIQUE MONTRÉAL
affiliée à l'Université de Montréal

A Dynamic Constraint-based Modelling Approach Of Cell Metabolism

MOHAMMADREZA YASEMI
Département de génie chimique

Thèse présentée en vue de l'obtention du diplôme de *Philosophiæ Doctor*
Génie chimique

Août 2022

POLYTECHNIQUE MONTRÉAL
affiliée à l'Université de Montréal

Cette thèse intitulée :

A Dynamic Constraint-based Modelling Approach Of Cell Metabolism

présentée par **Mohammadreza YASEMI**
en vue de l'obtention du diplôme de *Philosophiæ Doctor*
a été dûment acceptée par le jury d'examen constitué de :

Gregory DE CRESCENZO, président
Mario JOLICOEUR, membre et directeur de recherche
Olivier HENRY, membre
Maria I. KLAPA, membre externe

DEDICATION

Dedicated to the lasting memory of my parents...

ACKNOWLEDGEMENTS

It was an exciting and wonderful journey since January 1, 2019. Many people had a direct or indirect role in the achievement of this dissertation.

The first person on this list is Dr. Mario Jolicoeur. A mentor with a unique personality and teaching method that made this experience possible for me in the first place. Thank you Mario, I will not forget your support and trust in me throughout the way. I like to thank Dr. Bala Srinivasan from whom I learnt immensely despite his short presence as my co-supervisor. I would like to thank the other members of my committee, Dr. Gregory De Crescenzo, Dr. Olivier Henry, and Dr. Maria I. Klapa for agreeing to be part of my PhD thesis committee.

I would also like to thank Dr. Yaman Arkun, my master's supervisor, for introducing me to the systems biology field and optimal control theory in my master's studies. I was a lucky person to be the child of a teacher, and from elementary school until now I have enjoyed the blessings of inspiring teachers, all of whom I appreciate.

More than half of this thesis was written during the COVID pandemic and I was away from my family, but their support for me never diminished, it would not have been possible without their support and encouragement. Also, I benefited from the support of my lab mates and previous students of the research group, Léa, Pablo and Jorge, whose help and guidance I appreciate. Last but not least, I would like to thank Wendy for being by my side in Montreal helping me overcome the difficulties and make memories.

RÉSUMÉ

La modélisation métabolique mathématique est une approche systématique pour déterminer les principales causes d'un changement métabolique observé dans un biosystème, et estimer les implications d'une perturbation métabolique induite. En fournissant des informations mécanistiquement pertinentes au niveau des systèmes sur un réseau de bioréactions, la modélisation dynamique basée sur les contraintes (DCBM) s'est révélée prometteuse en ingénierie métabolique et pour la conception de bioprocédés. Aussi, nous avons d'abord dressé l'état de l'art en approches de modélisation du métabolisme de cellules eucaryotes, sous la forme d'une revue. Nous avons, par la suite, développé une approche DCBM qui utilise une boîte à outils mathématique d'optimisation convexe et de régression non linéaire pour estimer les distributions dynamiques des flux métaboliques intracellulaires dans deux biosystèmes : dans un premier temps les globules rouges stockés comme élément critique pour la médecine transfusionnelle, et dans un deuxième temps, des cellules de l'ovaire de hamster chinois (CHO), lignée cellulaire en tant que principal organisme hôte pour produire des produits biopharmaceutiques recombinant par culture cellulaire. Nous avons créé un réseau métabolique ad hoc comportant 77 réactions et 74 métabolites pour les globules rouges. Nous avons acquis une dynamique de flux à grain fin des processus intracellulaires. Ensuite, pour une analyse dynamique de l'équilibre des flux (DFBA), nous avons créé quatre fonctions objectives liées à l'accumulation de stress oxydatif dans les globules rouges stockés. Des prédictions de flux résolues dans le temps ont été obtenues dans les quatre situations tout en respectant les exigences d'égalité et d'inégalité requises. Enfin, pour calculer la distance euclidienne entre les vecteurs de flux optimaux dynamiques, nous avons utilisé une approche de programmation quadratique (QP). L'approche DCBM que nous avons créée ici, couplée au réseau métabolique que nous avons développé, s'est avérée adaptée à l'analyse informatique du comportement métabolique des globules rouges, et on s'attend à ce qu'elle soit bénéfique pour d'autres biosystèmes. De plus, en raison de la complexité inhérente des cellules eucaryotes, l'optimisation de la dynamique de croissance cellulaire et de la bio-production à partir de cultures de cellules de mammifères est une tâche complexe au niveau cellulaire. En conséquence, les approches expérimentales heuristiques en l'ingénierie métabolique des organismes hôtes sont fréquemment complétées par des modèles mathématiques de culture cellulaire afin d'améliorer l'efficacité des bioprocédés et d'identifier les causes des améliorations connues. En utilisant la stœchiométrie des bilans de masse du réseau, les modèles métaboliques structurés sont capables de représenter avec précision la complexité d'un réseau métabolique. Les modèles métaboliques basés sur les contraintes ont progressé au cours des

deux dernières décennies, passant de leur utilisation pour fournir une description linéaire des systèmes métaboliques à l'état d'équilibre, à la modélisation de la dynamique des systèmes non linéaires dans une formulation de problème d'optimisation résolue dans le temps ou dynamique. Cependant, bon nombre de ces approches échouent lorsqu'elles sont testées dans un contexte autre que celui pour lequel elles ont été développées, en raison d'un grand nombre d'hypothèses formulées au cours du processus de développement du modèle, ainsi que de diverses normes de modélisation utilisées par différents groupes de recherche, tous visant à révéler la complexité inhérente des réseaux métaboliques. Nous avons suivi les normes établies par la communauté telles que dans la reconstruction de réseaux et leur analyse basées sur les contraintes (COBRA). La plateforme COBRA in silico a été utilisée pour proposer une technique de modélisation dynamique basée sur les contraintes (gDCBM) à l'échelle du génome, et a permis de fournir la dynamique résolue dans le temps d'un réseau métabolique structuré pour le métabolisme des cellules CHO. Nous avons amélioré le réseau métabolique à l'aide d'un modèle métabolique de référence à l'échelle du génome (GSMM) de CHO, à savoir iCHO DG44 v1, puis imposé des contraintes dynamiques sur ses flux de transport à l'aide de données métabolomiques générées en interne. Pour anticiper les changements physiologiques dans les variants clonaux de CHO, nous avons utilisé cette technique gDCBM. Le modèle peut prédire les flux intracellulaires de manière continue et résolue dans le temps (par heure de temps de culture) pendant la croissance et vers le changement métabolique de non-croissance, a été confirmé en prédisant les concentrations en métabolites extracellulaires, y compris les acides aminés, ainsi que leur dynamique dans le temps. En conséquence, nous pouvons générer des hypothèses d'intervention et étudier les effets d'altération in silico avant ou en plus des expériences de culture cellulaire, qui sont chronophages et coûteuses. Le modèle est également utilisé pour décrire les altérations métaboliques globales entre les lignées cellulaires parentales et productrices élevées dans une autre application. Nous avons démontré que l'approche de modélisation établie peut être utilisée pour étendre ou réduire le réseau métabolique de manière systématique.

ABSTRACT

Mathematical metabolic modelling is a systematic approach to determining the major causes of a metabolic change seen in a biosystem and estimating the implications of an induced metabolic perturbation. By providing mechanistically relevant systems-level information about a network of bioreactions, dynamic constraint-based modelling (DCBM) has shown promise in metabolic engineering and bioprocess design. Also, we first drew up the state of the art in approaches to modelling the metabolism of eukaryotic cells, in the form of a review article. Thus, we have developed a DCBM approach that uses a mathematical toolkit of convex optimization and nonlinear regression to estimate dynamic intracellular metabolic flux distributions in two biosystems: Firstly, stored red blood cells (RBCs) as a critical element for transfusion medicine, and secondly, Chinese hamster ovary (CHO) cell line as the main host organism producing recombinant biopharmaceuticals in cell culture technology. We created an *ad hoc* metabolic network including 77 reactions and 74 metabolites for RBCs. We acquired fine-grained flow dynamics of intracellular processes. Then, for a dynamic Flux Balance Analysis (DFBA), we created four objective functions related to the accumulation of oxidative stress in stored RBCs. Time-resolved flux predictions were obtained in all four situations while adhering to the required equality and inequality requirements. Finally, to calculate the Euclidean distance between the dynamic optimum flux vectors, we used a quadratic programming (QP) approach. The DCBM approach we created here, coupled with the metabolic network we developed, proved to be suitable for the computational analysis of RBC metabolic behaviour, and it is expected to be beneficial for other biosystems. In addition, because of the inherent complexity of eukaryote cells, optimising cell growth dynamics and bioproduction from mammalian cell cultures is a complex task at the cellular level. As a result, heuristic experimental approaches in metabolic engineering of host organisms are frequently complemented with mathematical models of cell culture in order to improve the odds of enhancing bioprocess efficiency and identifying causes for known improvements. By utilising the network's mass balances' stoichiometry, structured metabolic models are able to accurately represent the complexity of a metabolic network. Constraint-based metabolic models have advanced over the last two decades from being utilised to provide a linear description of metabolic systems at steady states to modelling nonlinear system dynamics in a time-resolved or dynamic optimization problem formulation. Many of these approaches, however, fail when tested in a setting other than the one for which they were developed, owing to a large number of assumptions made during the model development process, as well as diverse modelling standards used by different research groups, all aimed at revealing the

inherent complexity of metabolic networks. We have followed the community's established standards as in Constraint-based reconstruction and analysis (COBRA). The COBRA in silico platform has been used to propose a genome-scale dynamic constraint-based modelling (gDCBM) technique that allows delivering time-resolved dynamics of a structured metabolic network for CHO cell metabolism. We improved the metabolic network using a reference genome-scale metabolic model (GSMM) of CHO, i.e., iCHO DG44 v1, and then imposed dynamic constraints on its transport fluxes using metabolomics data generated in-house. To anticipate physiological changes in CHO clonal variants, we used this gDCBM technique. The model can predict intracellular fluxes in a continuous time-resolved (per hour of culture time) manner during the growth to non-growth metabolic switch, and it has been confirmed by predicting concentrations of extracellular metabolites, including amino acids, dynamics of change in time. As a result, we may generate intervention hypotheses and study the alteration effects in silico before or in addition to the time-consuming and expensive cell culture experiments. The model is also used to describe global metabolic alterations between parental and high producer cell lines in another application. We demonstrated that the established modelling approach may be used to extend or reduce the metabolic network in a systematic way.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS AND ACRONYMS

bio-CRN	Bio-Chemical Reactions Network
CHO	Chinese hamster ovary
COBRA	Constraint-based reconstruction and analysis
DCBM	dynamic constraint-based modelling
DFBA	dynamic flux balance analysis
FBA	Flux Balance Analysis
gDCBM	genome-scale dynamic constraint-based modelling
GSMM	Genome-scale Metabolic Model
mAb	monoclonal Antibody
MFA	Metabolic Flux Analysis
ORFs	open reading frames
QP	quadratic programming
RBC	Red blood cell

LIST OF APPENDICES

CHAPTER 1 INTRODUCTION

1.1 Research motivation

Numerous bioprocess improvements have been achieved through molecular biology, such as DNA engineering and genome sequences. However, yet these alone are insufficient for understanding a biosystem behaviour. A change of perspective from a reductionist to a system-level understanding of cells is needed. Although improvements in precise quantitative experimental methods are undoubtedly still on-going, acquiring insights into the operation of biosystems demand a mix of reductionist and systems-level methods. This is a result of biological systems' inherent complexity [5]. It is not surprising, since cells have evolved over millions of years building many interconnected pathways causing complex phenotypes.

Systems Biology relies on models of biological systems as tools to elaborate the systemic relationship within an organism and its environment [6]. When it comes to the cell as a biosystem, mathematical models are developed to describe cellular functions. Model simulations may concern deriving the behaviour of either overall cellular functions or the specific function of an individual cellular process related to signaling, gene regulation or metabolism. Metabolic models are particularly useful to describe the flow of mass and energy in metabolic networks, in addition to describing the regulatory actions taking place inside the cell. Two main modelling methods can be taken into consideration depending on how the metabolic network is formalised: (i) constraint-based methods, that incorporate reaction stoichiometry and thermodynamic knowledge to infer metabolic flux distributions and predict complex metabolic phenotypes; and (ii) kinetic models that simulate changes in metabolite concentrations over time by including biochemical network stoichiometry, mechanistic reaction rate laws, and the associated kinetic parameters [7]. Together, these complimentary strategies can be utilised to reveal hidden patterns, explain emergent features, and generate new hypotheses.

In the field of biotechnology, mathematical models of cellular metabolism are of particular interest. The cell factory produces a wide range of valuable compounds, and mathematical modelling can be used to both develop new production methods and enhance already existing ones. The systematic investigation of the metabolism becomes possible using stoichiometric and kinetic models. These models were used to identify ideal bioprocess conditions as well as to guide genetic alterations to develop high-producer strains or cell lines. In addition to being a tool for understanding the cellular metabolism and physiology, mathematical models are also becoming increasingly useful due to the growing availability of genomic data and modern analytical approaches. The analytical power in laboratory and the adequacy of developed

models are two main factors that determine success of this practice. However, the current power of available molecular biological techniques points to a considerable disparity with the ability to rationally simulate and design biochemical networks [1, 8].

The well-conserved approach, for the analysis and the design of bioprocess engineering, still relies on having a mathematical model of the process under study. Metabolic models were primarily intended to describe a biosystem's static state, making the models less effective for describing manufacturing processes such as in batch and fed-batch modes that are inherently dynamic [9, 10]. There are unstructured dynamic models available that only vaguely consider the cell metabolism and instead include lumped mass balances to describe dynamics with Monod equations, for instance. These models, if extended to also describe intracellular metabolism, typically include a large number of parameters, making them extremely susceptible to model error and measurement noise in the data used for model calibration [11]. As a result, there is a strong incentive to develop structured metabolic models that accurately account for the interactions between extracellular and intracellular environments while also generating fewer calibration parameters [12]. These new models are expected to be more robust and more suitable for model-based prediction and optimization of bioprocesses under study across a wide range of operating parameters.

The best route across a stoichiometric network, within specific physicochemical constraints, is highlighted by constraint-based modelling (CBM) approach. A minimal amount of biological information is needed to draw quantitative conclusions regarding network behaviour using this method [13]. CBM techniques for metabolism organise biochemical, genetic, and genomic knowledge into a mathematical framework. It enables the prediction of cellular activity from a genotype. In order to enable a mechanistic description of metabolic physiology, over the past 30 years, the use of constraint-based techniques has evolved, and recently, more studies have integrated models with high-throughput data sets for prospective experimentation [14]. As a result of these research, a greater and more significant capacity for pertinent biological predictions have been confirmed.

Therefore, to the best of our knowledge and taking all the above, the current challenge relies on simulating overall intracellular and extracellular properties based on the extensively linking these. Simulation attempts to predict a system dynamics, and thus relevancy of underlying assumptions, can then be tested. The model is validated, when experimental observation and detailed behaviour of computer-executable simulations are contrasted. On the one hand, occurrence of inconsistency at this point indicates that the assumptions we are using to reflect our understanding of the system are insufficient, irrelevant or erroneous. On the other hand, if our models pass first validation, then it can be applied to answer

research questions as well as generate predictions that could be confronted and validated from experiments. Therefore, the hypothesis that the establishment of a dynamic constraint-based *in-silico* platform contributes to predict and analyze non-steady state metabolic fluxes, and to quantify medium dynamics, to enable understanding and optimizing a living organism metabolism have motivated this research.

1.2 Project objectives and methodology

The primary objective was to develop a framework with CBM properties and then to extend it to simulate dynamic metabolic behaviours. The goal of this study was thus to create appropriate algorithms for developing a dynamic constraint-based modelling (DCBM) approach for both non-growing Red Blood Cells (RBCs) and growing mammalian cells (CHO). Unlike dynamic kinetic modelling, which relies on a huge number of kinetic expressions and tuning parameters, the current approach is based on the parameter-free Dynamic Flux Balance Analysis (DFBA) algorithm [15]. The method developed in this study, is based on the formulation of an optimization problem in which a range of flux distribution in a metabolic network is desired, as well as the optimization of a set of specific objective functions under certain flux constraints. According to the premise underpinning DFBA, evolution had conditioned the cell to operate as an agent that optimally distributes resources to achieve a set of biological objectives. However, due to the higher complexity of mammalian cells compared to bacteria, no optimization objectives have yet been specified for mammalian cells. As a result, one of the objectives of this research was to identify objective functions especially for the non-growth phase of mammalian cells, for which we studied a non-growing biosystem (RBC) to analyze impact of alternate objective functions before designing new objective functions for mammalian cell culture (CHO). Therefore, objective functions and certain mathematical methods are designed coupled to produce an appropriate optimization problem whose solution describes the experimental behaviour of mammalian cells. The resulting DFBA model can then be used in support to guide process optimization approaches such as genetic modification, media design, or bioprocess management.

Despite the fact that our ultimate goal was to develop a dynamic metabolic flux model for CHO, we initially devised an *ad hoc* model for RBCs biosystem. The following were the motivations for designing the initial method for RBCs: (i) a basic metabolic network of RBCs was available, which allowed testing our framework on a rather simpler metabolic network. (ii) a collaborating scientist identified the importance of studying storage lesions in stored RBCs. (iii) the assumption of growth rate was irrelevant for this non-growing biosystem as the objective function, therefore, other reliable objective functions were identified. We established

systematic ways to treat raw experimental data and to develop a metabolic network coupled with the flux constraints that are the basic elements of the dynamic metabolic flux models. Algorithms and mathematical techniques that make use of raw data from literature were created in order to accomplish the project's objectives. In the work on RBC, we estimated the course of changes in metabolites concentration using non-linear regression before switching to cubic spline estimation in the work on CHO. In order to provide a workable optimization problem, we consolidated a metabolic network for RBC as well as for CHO, validated with experimental data sets, in support of formulating dynamic optimization problems.

Genome-scale metabolic models (GSMM) provide the most accurate *in silico* mapping of the genotype-phenotype link and thus have become increasingly important for the optimization of bioproduction processes. In the absence of a genome-wide annotation of open reading frames (ORFs) or because of modelling purposes, one may choose to develop an *ad-hoc* metabolic model in which a predefined subset of reactions (and metabolites) are explicitly considered. In the work on CHO, we created a genome-scale dynamic constraint-based modelling (gD-CBM) framework employing methodologies for systematic identification of constraints and reliable objective functions. As a result, in this thesis, we worked on the metabolic network of a non-growing biosystem (RBC) and a growing one (CHO). In parallel, we developed an *ad-hoc* metabolic model to study stored RBCs oxidative lesions and a GSMM to study monoclonal antibody production in Chinese hamster ovary (CHO) cell culture. The models were rectified based on the constraint-based reconstruction and analysis (COBRA) guidelines [16], to eventually they provide robust *in-silico* computational platforms to estimate input-output relations between extracellular and intracellular environments.

1.3 Organization of the thesis

This thesis includes 9 chapters and is organized as follows. In Chapter one we presented the research motivation and the project objectives. Chapter two presents a literature review on modelling cell metabolism. We critically analyzed a wide array of approaches in metabolic modelling including Dynamic Kinetic Modelling (DKM) and Constraint-based Modelling (CBM) and their variants. This review concluded with a discussion of knowledge gaps and future perspectives for model development. This review was published in *Processes* with the doi:10.3390/pr9020322. Chapter three is a modelling study of a non-growing biological system, i.e., Red blood cells (RBCs). In that study, we analyzed impacts of alternative objective functions on the dynamic modelling of the RBCs storage process. Hypothetical objective functions regarding oxidative defense mechanisms were considered developing a dynamic constraint-based modelling (DCBM) approach. This chapter is in press for publication

in IFAC-PapersOnLine. Chapter four describes a genome-scale dynamic constraint-based modelling (gDCBM) framework to model CHO clonal variations. Interactive components of this framework are explained and the dynamics of extracellular metabolites is simulated followed by the prediction of the range of the intracellular flux distributions. This chapter has been submitted to *Metabolic Engineering*, and is under review for publication. Chapter five contains a general discussion and the original contributions of this thesis. Chapter six presents recommendations for future work and Chapter seven proposes the Conclusion part.

CHAPTER 2 LITERATURE REVIEW

2.1 ARTICLE 1 : MODELLING CELL METABOLISM: A REVIEW ON CONSTRAINT-BASED STEADY-STATE AND KINETIC APPROACHES

Mohammadreza Yasemi[1] and Mario Jolicoeur[1,*]

[1] Research Laboratory in Applied Metabolic Engineering, Department of Chemical Engineering, École Polytechnique de Montréal, P.O. Box 6079, Centre-ville Station, Montréal, QC H3C 3A7, Canada

*Corresponding author: mario.jolicoeur@polymtl.ca

(Published 9 February 2021, in Processes 2021, 9(2), 322; https://doi.org/10.3390/pr9020322)

Abstract

Studying cell metabolism serves a plethora of objectives such as the enhancement of bioprocess performance, advancement in the understanding of cell biology, of drug target discovery, and in metabolic therapy. Remarkable successes in these fields emerged from heuristics approaches, for instance, with the introduction of effective strategies for genetic modifications, drug developments and optimization of bioprocess management. However, heuristics approaches have showed significant shortcomings, such as to describe regulation of metabolic pathways and to extrapolate experimental conditions. In the specific case of bioprocess management, such shortcomings limit their capacity to increase product quality, while maintaining desirable productivity and reproducibility levels. For instance, since heuristics approaches are not capable of prediction of the cellular functions under varying experimental conditions, they may lead to sub-optimal processes. Also, such approaches used for bioprocess control often fail in regulating a process under unexpected variations of external conditions. Therefore, methodologies inspired by the systematic mathematical formulation of cell metabolism have been used to address such drawbacks and achieve robust reproducible results. Mathematical modelling approaches are effective for both the characterization of the cell physiology, and the estimation of metabolic pathways utilization, thus allowing to characterize a cell population metabolic behavior. In this article, we present a review on methodology used and promising mathematical modelling approaches, focusing primarily to investigate metabolic events and regulation. Proceeding from a topological representation of the metabolic networks, we

first present the metabolic modelling approaches that investigate cell metabolism at steady state, complying to the constraints imposed by mass conservation law and thermodynamics of reactions reversibility. Constraint-based models (CBMs) are reviewed highlighting the set of assumed optimality functions for reaction pathways. We explore models simulating cell growth dynamics, by expanding flux balance models developed at steady state. Then, discussing a change of metabolic modelling paradigm, we describe dynamic kinetic models that are based on the mathematical representation of the mechanistic description of nonlinear enzyme activities. In such approaches metabolic pathway regulations are considered explicitly as a function of the activity of other components of metabolic networks and possibly far from the metabolic steady state. We have also assessed the significance of metabolic model parameterization in kinetic models, summarizing a standard parameter estimation procedure frequently employed in kinetic metabolic modelling literature. Finally, some optimization practices used for the parameter estimation are reviewed.

keyword constraint-based modelling approach, kinetic modelling, metabolic network, dynamic metabolic flux analysis, metabolic flux regulation, metabolic network structure, metabolic model parameterization, Gibbs free energy, thermodynamic constraints, metabolic control analysis

2.1.1 Introduction

Expanding the modelling paradigm, from a microscopic description to systems biology perspective, has offered a deeper capacity to describe cell behaviour and its interaction with extracellular environment. The improvement was necessary to address complex problems related to the characterization of intracellular events [6, 17, 18]. Indeed, one-reaction models have then rapidly evolved to reaction network-level models, and recently reached a genome-wide level when applicable. Various mathematical approaches have been proposed to first integrate a reaction network and stoichiometry, for then integrating reaction kinetics and regulation mechanisms. While evolving, models have gained the capacity to question cell behaviour by both interpolating within and extrapolating from experimental data, for then allow extracting data that are tedious to impossible to be directly acquired, such as intracellular metabolic flux rates [19]. The predictive power that network-level models offer is an important advantage of such systemic mathematical representation compared to heuristics approaches. In fact, metabolic models have evolved to encompass higher levels of the knowledge regarding a biosystem, i.e. from gene to protein to reaction, to thus enhance the predictive power. Cell metabolism is a complex biosystem of study, even for unicellular organisms. Indeed, metabolic modelling is particularly challenging because many cellular

functions are the result of a combination of factors. Therefore, unravelling causal relations and correlations between an observed phenomenon and its associated factors necessitates a model capable of describing the underlying biosystem, the network of biochemical interactions, and their connectivity and regulation. A model is still a modest approximation [20]. The merits and drawbacks of a model are determined by adequacy to describe a biological reality, further, a compromise between model correctness and the extent to which it is mathematically tractable is unavoidable [21]. Thus, simplifying assumptions must be considered making a metabolic model definite in its scope and solvable regarding available experimental data. These assumptions are a consolidation of heuristics experiences for a model, the current knowledge of the biosystem and of its interaction with its surroundings. Typical assumptions concern the cell compartments homogeneity level, a diversity in biological time scales, the level of distributing versus lumping entities of interest, among others. These assumptions are particularly of interest because they are the main factors determining the relation between an observation in reality and the result of the simulation of an abstract mathematical model. The required metabolic modelling approach is shaped to accommodate the assumptions and provide the sought-after applications [18, 22–24]. The metabolic network models have been developed to understand the dominant qualitative features of cell metabolic behaviour in a bioprocess [25], or in a metabolic disease [26]. A comprehensive metabolic model can also improve conventional wisdom regarding the cell metabolism by making important corrections, such as gap-filling reactions suggested by genome-wide metabolic models [27, 28]. When required to interfere with cell functioning, a metabolic model can contribute to discovering new strategies [26, 29, 30], and to organize the disparate accumulated information into a coherent body of practical knowledge [18, 31]. In a complex system such as the metabolic network of the cell, a model helps to think (and calculate) logically about what components and interactions are important and what other components can be neglected [1]. However, the extent of progress in the model application to accomplish each of these goals depends on several factors, including the industrial importance and the scientific significance of sought-after results. Based on this premise, metabolic modelling has gained substantial importance to enhance bioprocess optimization [32, 33], where the increased efficiency leads to cost reduction in million-dollar order of magnitude, and to develop metabolic therapies [26, 29], i.e. to push forward cancer treatment. In a broad classification, metabolic models are divided into four categories based on the capability of the model to distinguish between sub-populations of biomass (unsegregated/segregated) and the extent of recognizing the intracellular biochemical components (unstructured/structured) [1, 34], (Figure 2.1).

In this review, we have clustered modelling approaches which are mostly understood as *unsegregated and structured*. Thus, such a mathematical model describes *multi-compartment*

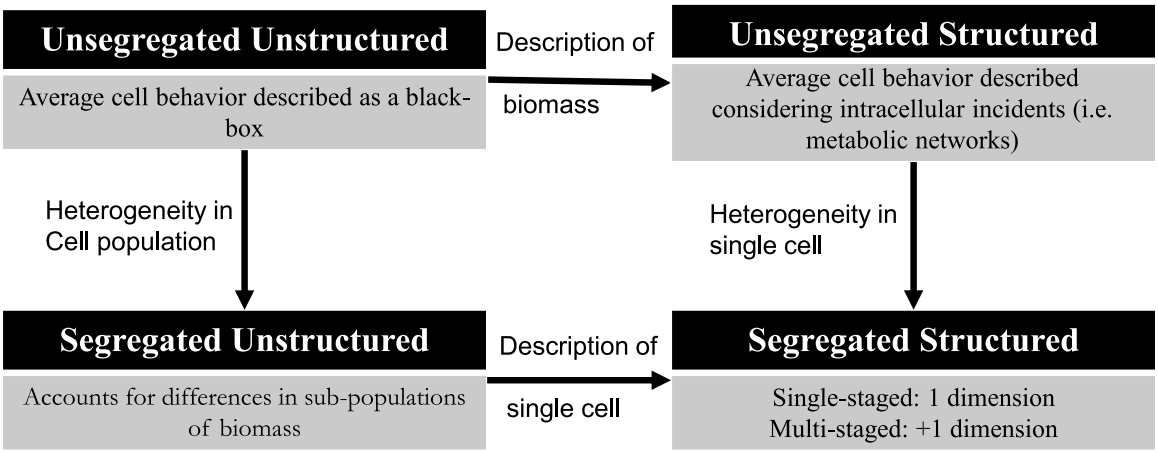

Figure 2.1: **A broad classification of the mathematical metabolic models.** The difference in the models that are shown originates from an average cell approximation and a balanced growth approximation in regard with (1) cell population, (2) a single cell. Reproduced with permission from [1].

average cell and when it is desired the model extends the description to the whole population by taking into account the quantity (size) of the population, i.e. biomass concentration. Present approaches proposed in literature were compared based on their ability to elucidate the structure of a network, to then characterize allowable and empirical flux distribution, and to finally take into account flux dynamics regulation. This review is not by any means a comprehensive tabulation of the exhaustively diverse range of the published metabolic models, but a methodology motivated review of the keystones of metabolic modelling. Thus, significant trade-offs between any two categories were acknowledged. Also, from another crucial perspective, the capability of the models under steady state and dynamic state of the metabolic network is assessed. The revised modelling approaches have been used for severely different scales of metabolic systems and typically to serve distinct strategies, therefore, it deems essential to notice the context where models are primarily used to avoid confusion. When possible, a simple running example is used to illustrate the approach being discussed. The running example is first introduced in Figure 2.2.

2.1.2 Topological representation of metabolic networks

Representation based on the graph-theory

In this approach, cell metabolic networks are denominated Bio-Chemical Reactions Network (bio-CRN), including the reactions that govern the cell metabolism. Owing to genome sequencing technologies, annotated genomes for several organisms made possible to reconstruct metabolic network accounting for the major active constituents of the cell, i.e. proteins, DNA, RNA and other metabolites [35, 36]. Modelling approaches based on the graph theory have been used to unravel network complexity by quantitative methods. In (the major variants of) this approach, the metabolic network substrates are modelled as nodes of the graph and reactions as the (directed) edges [35, 37]. It has been shown, in comparative studies for different species that the graph of a metabolic network can be modelled by a scale-free connectivity structure [38]. In a scale-free structure, the probability distribution of finding a node with k connections follows a power-law function, i.e. $P(k) \approx k^{-\lambda}$, with λ being a species-specific positive constant. In a metabolic network context, it means that the probability of finding a substrate involved in k reactions decreases with a power-law relationship to the increasing number of reactions [37, 39, 40]. (refer to Figure 1 of [35] for illustration of the connectivity structures.)

Reconstruction of the metabolic network for 80 fully sequenced organisms taken from eukaryotes, archaea and bacteria conclude that most of the metabolites participate in a few reactions and a few ones drive many reactions, with these latter defined as hub-metabolites [35, 38, 40, 41]. Interestingly, this statistically heterogeneous distribution of metabolites connectivity resembles an error-tolerant system structure, which is found in human-made networks such as the Internet [42]. It was shown that random elimination of reactions due to mutations of enzyme-coding genes are tolerated by metabolic networks and the structural redundancy allows the modified metabolic networks to result in viable reorganizations of the metabolic routes, i.e. the *in-silico* model maintained the potential to support growth [43]. In fact, it is argued that "the small world architecture" of metabolic networks, which refers to the scale-free structure, "may serve to minimize transition times between metabolic states [. . .]" [40] as well. On the other hand, network behaviour, i.e. fluxes rates and distribution, is sensitive to changes of so-called hub-metabolites [38, 44, 45], which motivated further structural investigation of these special nodes. A sub-network mostly made of hub-metabolites is named *the giant strong component* of a metabolic network, and it represents the core metabolites typically consisting of less than one third of the total metabolites included in the metabolic network [38]. In [40], the authors identify a node in the substrate graph as a hub-metabolite if its degree exceeds the mean metabolite degree of the network by three times

the standard deviation. Moreover, considering the most linked metabolite is ranked first in a network, and it is a shared metabolite between tens of organisms with an average rank of R, it has been shown that metabolites in higher average ranks see less deviation in their ranks among the different organisms. Conversely, the metabolites of lower average ranks, i.e. the ones participating in a few reactions, show species-specific changes in the number of reactions they are participating in. Topological analysis of metabolic networks structure also provides the average reaction path between any two substrates. As expected, this number increases with the addition of more metabolites to the network. However, some individual metabolites then become more and more connected, thus the average length of reaction paths remains in a much narrower range compared to the range of added metabolites. As an example, if the number of metabolites integrated in the metabolic network increases from 200 to 1000, the average reaction path increases from 4 to 12 (refer to Figure 5 in [37]). Worth mentioning, the calculated average reaction path increases when the connections through the energetic metabolites, such as ATP/ADP, $NADH/NAD+$ and $NADPH/NADP+$, are eliminated, highlighting their significant role as co-metabolites of many reactions. Excluding these metabolites, the ten most connected metabolites are the intermediates in the glycolysis pathway (Glycerate-3-phosphate, D-Fructose 6-phosphate, D-glyceraldehyde-3-phosphate), the intermediates in the pentose-phosphate pathway (D-ribose-5-phosphate, D-xylulose-5-phosphate), pyruvate and acetyl-CoA, the precursor for purine and histidine synthesis (5-phospho-d-ribose 1-diphosphate), L-glutamate and L-aspartate [37].

Representation based on the Petri net theory

Petri net theory supports another topological oriented modelling approach of metabolic networks [46–48]. A Petri net is a directed graph, whose nodes have two different types, namely, *places* and *transitions*. Thus, substrates (and products) are considered as the places, and reactions are the transitions. In this approach, one end of an edge is connected to a metabolite and the other end is connected to a reaction. Therefore, as opposed to the graphs reviewed in 2.1.2, in Petri net graphs edges are not representative of reactions. Instead, edges have *weights* resembling the stoichiometric coefficients [49]. Petri net models provide a reliable analysis of the consumption/production relations but not of regulatory interactions. However, several extensions of the primary formalism have been introduced to increase the quantitative suitability of Petri net theory for biological systems analysis. For example, coloured Petri nets, stochastic Petri nets [50, 51], self-modified Petri nets [52, 53], hybrid and hybrid functional Petri nets [47] . Nevertheless, the Petri net representation of metabolic networks still deserves more methodological developments.

The stoichiometric matrix of a metabolic network

As for a chemical reaction, the stoichiometry of a biochemical reaction tells the relative number of moles on either side of a reaction involving an enzyme in a balanced reaction [18]. In a bio-CRN, the matrix of stacked stoichiometric vectors of all the network reactions forms the stoichiometric matrix of the reaction network. The stoichiometric models of metabolic networks for prokaryotic and eukaryotic cells are available in literature [54]. The stoichiometric matrix for a metabolic network with m metabolites and n reactions is as follows:

$$S_{m\times n} = \left(s_{i,j}\right)_{m\times n} \{i = 1, ..., m | j = 1, ..., n\} \tag{2.1}$$

where the i^{th} row consists of reaction coefficients for metabolite i, with regard to (w.r.t) all reactions of the network, s_{ij} is the coefficient of metabolite i in reaction j. A negative s_{ij} implies that the metabolite i is on the left side of a reaction j and thus assumed to be consumed, if positive it is receiving flux of matter in the reaction j (v_j), and zero if it does not participate in the reaction j:

$$\frac{dx_i}{dt} = \sum_{j=1}^{n} s_{ij}^{+} v_j - \sum_{j=1}^{n} s_{ij}^{-} v_j, \qquad i = 1, \ldots, m \tag{2.2}$$

Therefore, we have one column vector for each reaction and Eq. A.2 can be written in matrix-vector notation for the whole metabolic network as follows:

$$\frac{d\vec{x}}{dt} = S_{m\times n}\vec{v} \tag{2.3}$$

The stoichiometric matrix can be reshaped into four partitions, each of which with specific characteristics (Figure 2.2). Structural analysis of metabolic networks can be more elegantly performed by using the matrix formalism instead of topological theories, as it is transferable to a graph representation while such transfer from the topological representations is not always possible [18, 21]. Either way, the analysis is not restricted only to genome-scale metabolic networks, and it can be efficiently used in an early step of modelling to draw a non-intuitive sense of network reactions connectivity. However, relying solely on structural information to extract connectivity interactions suffers from the flawed assumption that every biochemical reaction is worthing same weight, regardless of its level of activity. Therefore, the structural analysis results can hardly be directly useful for investigating the cellular mechanisms for flux regulation [55, 56]. Contribution of the stoichiometric matrix to the formulation of the further approaches is fundamental. The algebraic properties of stoichiometric matrix and

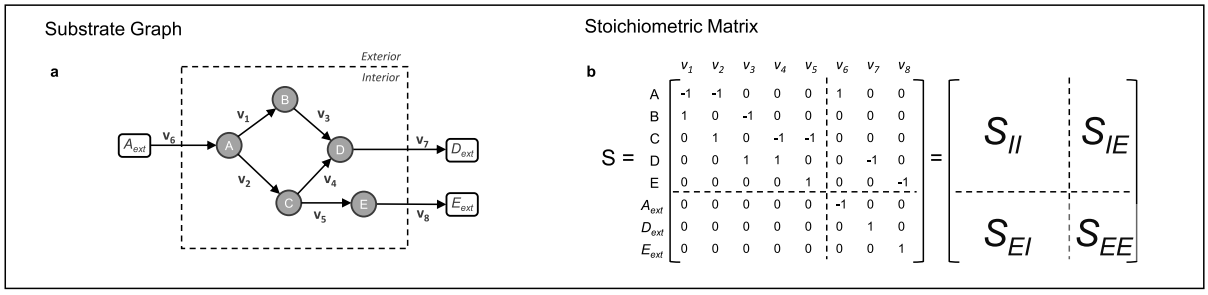

Figure 2.2: **Partitioning of the stoichiometric matrix of the running example.** (**a**) The substrate graph representation of the running example metabolic network. (**b**) The corresponding stoichiometric matrix divided in the four characteristic sub-matrices. S_{II}: Intracellular metabolites reaction coefficients w.r.t intracellular reactions. S_{IE}: Intracellular metabolites reaction coefficients w.r.t exchange reactions. S_{EI}: Extracellular metabolites reaction coefficients w.r.t intracellular reactions, always zero. S_{EE}: Extracellular metabolites reaction coefficients w.r.t exchange reactions, always diagonal.

nowadays matrix computation capacity has fuelled further use of the stoichiometric matrix as an important part of any metabolic mathematical models. The significance of its role in the representation of metabolic networks will become evident through the following sections.

2.1.3 Metabolism is a constrained system

Approaches based on the metabolic steady state hypothesis have been first proposed for estimating theoretically active enzymatic reactions [57–60], and to estimate metabolic flux rates from the reaction network stoichiometry and experimental time-data of extracellular metabolite concentrations [26, 61–66]. The mass conservation constraint at the steady state is formulated through the stoichiometric matrix for a network of biochemical reactions at any size, and this constraint has a central role in the determination of the solution space in many approaches (following sections). The observation that the changes of intracellular fluxes occur at a faster dynamic compared to the changes of the extracellular concentrations supports the validity of the steady state assumption [67, 68]. A detailed mathematical justification of the (pseudo) steady state assumption in the mathematical modelling of biochemical reaction networks is proposed in 1983 in [67], however, in the context of this review article, this assumption is valid at constant exponential growth phase in batch cultures, and in continuous cultures operating at steady state [69]. The advantage of the constraint-based modelling is that these approaches do not need to deal with enzyme kinetics and post-translational regulatory mechanisms [12]. In the following, we review the most relevant approaches and the associated constraints that are imposed on metabolic networks.

The null space of stoichiometric matrix

From a mathematical point of view, the stoichiometric matrix S, denotes a linear mapping of the reaction rate vector into the space of accumulated concentration time derivatives of metabolites [70]. To illustrate, consider N to be a linear mapping function, let:

$$N : R^n \rightarrow R^m$$

be defined by:

$$N([v_1, v_2, \ldots, v_n]) = [\dot{x}_1, \dot{x}_2, \ldots, \dot{x}_m] \tag{2.4}$$

since N is linear, we can write:

$$N(\vec{v}_{(n\times 1)}) = v_1 N(\vec{e}_1) + v_2 N(\vec{e}_2) + \cdots + v_n N(\vec{e}_n) \tag{2.5}$$

$$= \left[N(\vec{e}_1) N(\vec{e}_2) \ldots N(\vec{e}_n) \right]_{(m\times n)} \vec{v}_{(n\times 1)} \tag{2.6}$$

$$= S_{(m\times n)} \vec{v}_{(n\times 1)} = \vec{\dot{x}}_{(m\times 1)} \tag{2.7}$$

For the particular case of metabolic network operating at steady state, we have $S_{(m\times n)}\vec{v}^*_{(n\times 1)} = \vec{\dot{x}}_{(m\times 1)} = \vec{0}$. The set of vectors $\vec{v}^*_{(n\times 1)}$ are linearly dependent to the members of *the basis for the null space* (Eq. A.4) and thus, a subset of the null space of the stoichiometric matrix S (Eq. A.5). Where r is the rank of stoichiometric matrix S, the basis for the null space is a set with $q = (n - r)$ linearly independent column vectors of dimension n, which generates the null space for stoichiometric matrix S when spanned [70].

$$K(S) = \{\vec{b}_i \in R^n \quad (i = 1, \ldots, q) \quad | \quad S.\vec{b}_i = \vec{0} \quad \text{and} \quad \vec{b}_i.\vec{b}_j = \vec{0} \quad (i \neq j)\} \tag{2.8}$$

$$\textit{Null space}(S) = \textit{span}\{\vec{b}_1, \ldots, \vec{b}_q\}, \qquad q = (n - r) \tag{2.9}$$

where $K(S)$ denotes the basis for the null space of stoichiometric matrix S.

Metabolic Pathway Analysis (MPA)

Metabolic networks are an interconnected set of the metabolic pathways which interconnect at branch points and through shared metabolites. It is common, inside a metabolic network, to have several redundant routes for the transformation of one given metabolite to

another. It is particularly true for linking external substrates to external metabolic products [71, 72]. These multiple routes, which obey to mass conservation principle at steady state, are considered a main cause of biological robustness [73, 74]. Thus, model development by the decomposition and characterization of a metabolic network into definitive pathways has attracted attention to allow identifying the dominant routes [24, 75, 76]. Particularly, MPA modelling approaches adapt a convex basis analysis methodology to estimate a theoretically feasible relative flux distribution. In MPA, *component pathways* of the network and genetically independent reaction pathways are identified [77–79], such insight is particularly of interest in gene-knockout and synthetic pathway construction studies.

MPA is usually used as an umbrella term for two closely related approaches, namely elementary flux mode analysis (EFMA) and extreme pathway analysis (EPA). A *flux mode* is a flux distribution vector that satisfies the mass conservation constraint for the intermediate metabolites at steady state, and does not violate direction restriction of irreversible reactions. The term flux modes was coined by Schuster et al. [21, 75] in 1994, to address the independent enzyme sets responsible for a component pathway. If it is not possible to further decompose a flux mode to a linear combination of simpler flux modes, it is considered as an *elementary flux mode*. In biochemical terms, it is a minimal set of enzymes "that can operate at steady state, with all irreversible reactions used in the appropriate direction" [77]. In mathematical definition, the convex basis vectors for the null space of the stoichiometric matrix of a metabolic network represent elementary flux modes (EFMs) [80] (see Table S1, Supplementary Materials, for properties of the different MPA methods).

EPA benefits from the algebraic tools developed for metabolic pathway analysis. EPA is closely related to EFMA, and the important condition in their formulation is the net direction of each reaction. Reactions that can be reversible may flow in two different directions at different states of cell metabolism. Knowing that, EPA is assumed to extract the minimal subset of EFMs [81]. Considering the n-dimensional vector of intracellular fluxes, elementary flux modes form a convex cone in the solution space, while extreme pathways characterize the edges of this solution cone [82–84]. Therefore, EPA is computationally less demanding than EFMA [85, 86]. EFMs reproduce all the admissible reaction pathways at steady state in an unbiased manner, therefore, enumerated EFMs of a metabolic model can quickly exceed computational capacity of a typical computing system of the date. However, various studies show that usually a small percentage of these calculated EFMs are biologically and thermodynamically feasible [84, 87], thus, the integration of constraints from biology and thermodynamics is utilized to ameliorate computational cost of enumerating EFMs for genome-scale models (the following sections). Use of these approaches complemented promising methods to identify co-regulated reactions and co-expressed genes [88], and composition of the minimal required

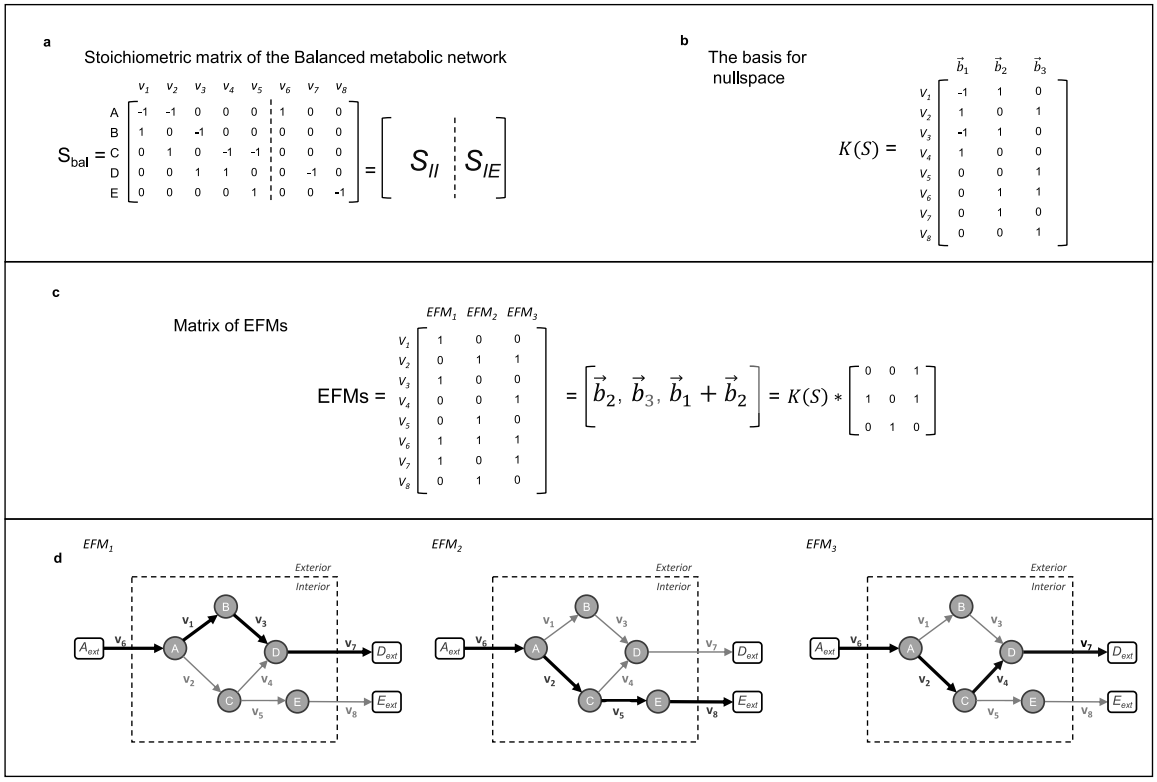

Figure 2.3: **Metabolic pathway analysis** (**a**) The balanced metabolic network includes the metabolites for which the quasi-steady state assumption holds. (**b**) The basis for the null space is the kernel matrix of the stoichiometric matrix. (**c**) EFMs are enumerated assuming that all the reactions are irreversible. Each EFM is a linear combination of the basis vectors. (**d**) The flux maps for the EFMs are shown on the substrate graph of the network (see Figure S1, Supplementary Material, for the network of reversible reactions).

substrates in order to produce valuable metabolic products [88, 89] (see Table 2.1 in [90] for a list of software packages for enumerating EFMs).

Metabolic Flux Analysis (MFA): admissible metabolic fluxes

Flux quantification in a metabolic system reveals significant information about the cell physiological state. Such information, particularly characterizes the cell response to genetic manipulation or changes in its environment, thus, may be used to improve a bioprocess toward predefined optimization objectives [1, 91]. Moreover, it was shown in several studies that the rates of exchange reactions, i.e. intermembrane transport reactions, can be used not only to quantify the rates of the intracellular reactions, but also to describe global processes taking place in the corresponding metabolic state of the cell [1, 91–93]. In a comprehensive metabolic network, the number of metabolites is typically smaller than the number of intracellular and exchange reactions combined. Thus, the associated stoichiometric matrix renders an under-determined system of equations, where the metabolites form balance equations, and the reaction rates are unknowns. When the system of equations is solved, MFA provides an empirical flux map [94–96], instead of the relative distribution of the fluxes provided by MPA. The mathematical formulation is as follows [68]:

$$\begin{aligned} S.\vec{v} &= 0 \\ \begin{bmatrix} S_u & S_k \end{bmatrix} \begin{bmatrix} \vec{v}_u \\ \vec{v}_k \end{bmatrix} &= 0 \\ \vec{v}_u &= -[(S_u^T S_u)^{-1} S_u^T] S_k \vec{v}_k \end{aligned} \tag{2.10}$$

where the vector v_u represents unknown fluxes, the vector v_k the measured (known) fluxes, and S_u , S_k the stoichiometric matrix subsets for unknown and measured fluxes, respectively.

To have a determined system, the degree of freedom of a system must equal zero. Generally, some reaction pathways are excluded from a comprehensive metabolic network to qualify it for MFA [8], and experimental data regarding concentration of medium components are used to further decrease the degree of freedom, i.e. identifying more exchange reactions. Measurements provide the estimation of cell concentration and uptake/secretion rates, which are the characteristic parameters of cell activity. Conversely, the degree of freedom increases as a result of the branches of pathways diverging and then rejoining later in the network, reversible reactions and metabolic cycles. These structures add reaction columns to the stoichiometric matrix, which are linearly dependent to the existing ones [97, 98]. For instance, the authors in [99] show that in a mammalian cell, it is essential to determine the uptake/secretion rates

of ammonium and the secretion rate of either CO_2 or O_2. Because these are the cometabolites of the TCA cycle reactions, thus, determination of them enhances the observability of the fluxes in the TCA cycle. However, generally the balance equations alone are not capable of addressing such issues, therefore, additional constraining balance equations must be introduced to determine the system and consequently solve for the unknown fluxes [97, 98].

13C-Metabolic Flux Analysis More sophisticated substrate labelling experiments can be performed to overcome limitations of MFA approach such as from reversible and cyclic reactions [92, 93, 100–103]. Metabolite and isotopomer balancing have been developed since 1995 [104]. In this method, a labelled substrate (e.g. 13C-glucose, 13C-glutamate, 14N-glutamine) is administered to the cell, leading to the labelled atoms in different metabolites and in the macromolecular biomass constituents and thus as well as in the secreted products [102, 105]. Henry et. al [92, 93], used 13C-MFA to optimize the timing of induction of a biphasic Chinese hamster ovary (CHO) cell culture, by assessing the intracellular flux maps before and after production induction. Parallel labelling experiments as an alternative to traditional single tracer experiments [106] was used by Antoniewicz et. al to characterize flux distribution in growth and non-growth phases of CHO cell culture [66], among numerous other applications of this method [91].

Chromatography methods are used to separate labelled and non-labelled metabolites from each other (i.e. metabolite identification), then, mass spectrometry is used to measure their abundance by mass determination (i.e. metabolite quantification). Thus, use of a coupled either liquid or gas-chromatography mass spectrometry apparatus (LC-MS or GC-MS) is a prevalent strategy for metabolite profiling [100, 107]. MS is particularly more popular in comparison with nuclear magnetic resonance (NMR) and liquid-chromatography mass-spectrometry (LC-MS), because of its good performance, extensive databases, relative cost of operation and the ease of maintenance [108,109], but exact quantification is more straightforward with LC-MS because of the use of known calibration curve with standards. The acquired data has to undergo correction and normalization to provide the measurements in a form that has biological information content. Kanani et. al [110], discussed the main sources of biases which arise in GC-MS and troubleshooted some of these biases, following the efforts to standardize this methodology as an integral part of the metabolomics data generation [111, 112].

The non-experimental stage of metabolite labelling flux analysis consists of solving the isotopomer balance equations by computational algorithms. However, the capability of these algorithms to provide interpretation of experimental data and to support strategic experimental design is limited. In fact, increasing the power of frameworks to model isotope labelling sys-

tem, with the least (or no) loss of information, is an active topic of research [105,108,113,114]. Recently modified versions of 13C-MFA were used to overcome the challenges of measuring metabolic fluxes in the distinct compartments of the eukaryotic cells (e.g. cytosol, mitochondrion). In a recent article, the authors provide an updated detailed protocol to perform 13C-MFA for the quantification of intracellular fluxes [108], and in a comprehensive review the latest developments in the MFA has been described [8].

2.1.4 Constraints augmented with a hypothesized objective function for the cell

The information available about carbon flow distribution in a metabolic system mostly suggests that the cell primary task is biomass synthesis, then homeostasis and maintenance of cell activities, and to a lesser extent, the production of an important product such as monoclonal antibodies. The priority and significance of these tasks change depending on the organism origin, time and mode of cell growth. Thus, the assumption that the cell metabolism has a particular *predefined* objective, is used to determine metabolic fluxes distribution [54, 102, 115]. The commonly assumed objective function for flux balance models is the maximization of biomass growth (See Table S2, Supplementary Materials, for a list of other objectives). Flux Balance Analysis (FBA) is the framework used in this approach since it allows determining ranges for *the allowable flux distributions* based on the hypothesized relevant biological objectives of the cell [94]. The mathematical formulation of a cell objective, accounting for defined fluxes restrictions with bounding limits for the reaction fluxes, is developed as a linear programming optimization problem (Eq. A.7).

$$\begin{aligned} &\max \vec{C^T}.\vec{v} \\ \text{subject to:} \\ &S.\vec{v} = 0 \\ &lb < \vec{v} < ub \end{aligned} \tag{2.11}$$

Considering that the stoichiometrically defined domain of flux vectors is dictated by the cell *metabolic genotype*, the solutions of an optimized objective function will be pertained within the *metabolic phenotype* of a specific cell-line or strain under study [116]. The arrays of the row vector $C^T_{(n)}$ are the linear coefficients representing weights of the fluxes in the objective. The solution for one particular objective function, such as the maximal growth, may be different and even contrasting from another objective, such as for maximal production of a secondary metabolite [54, 102, 117]. It is exemplified in Figure 2.4 that how FBA is used to estimate the metabolic network flux maps for the running example.

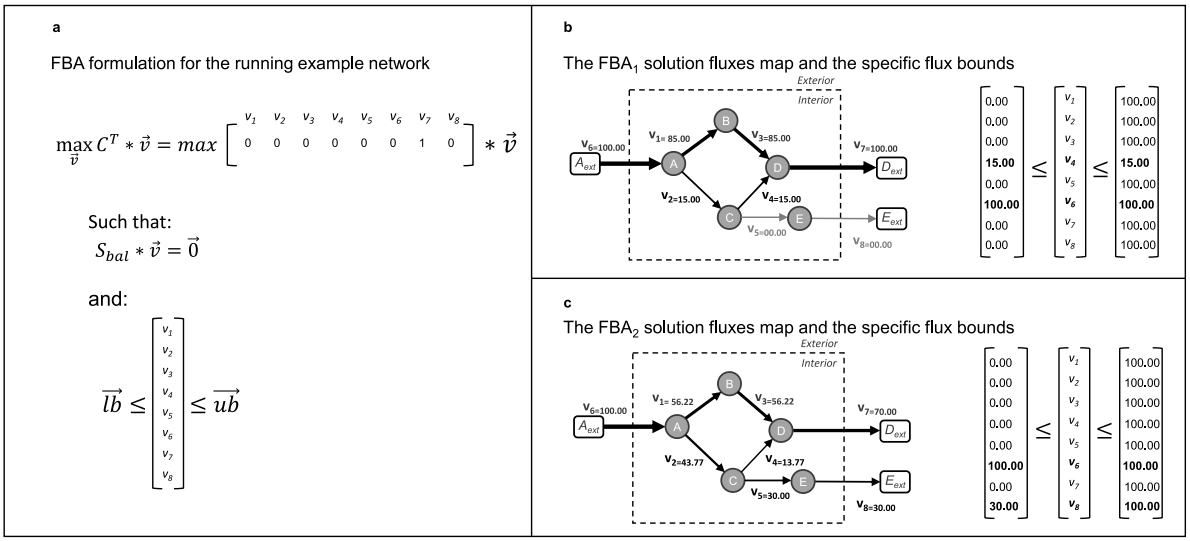

Figure 2.4: **Flux Balance Analysis (FBA)** (**a**) The formulation of FBA as a linear programming optimization problem implemented on the running example. v_6 denotes the measured uptake rate and v_7 is the objective flux to be maximized. (**b**) The solution fluxes map (left) for the imposed flux bonds (right). It is assumed that v_4 is measured/known in addition to v_6 (FBA_1). (**c**) The solution fluxes map (left) for the imposed flux bonds (right). It is assumed that v_8 cannot be less than 30 units (FBA_2). **Units:** Flux rate units such as $\frac{(nmol)}{1e6cell-hr}$ or $\frac{mmol}{gDW-hr}$.

The particular solution for the relative distribution of the fluxes can be used to interpret, describe and predict experimental results. The workflow to perform FBA is as follows:

1. To curate metabolic reactions from the annotated genome data.
2. To identify the topology and the structure of a network.
3. To identify the flux vector solution space at steady state.
4. To impose constraints and bounds on fluxes.
5. To define hypothesized objective function of interest for a biosystem.
6. To realize the vector space of the solution and identify the optimum solution (i.e. the flux map).

The accuracy of FBA solutions is partially determined by the accuracy of constraints imposed on the fluxes, and also, by the relevancy of the biological objective function that is assumed. Thus, it is not rare that the FBA solution provides unfeasible predictions of the flux distribution. To reconcile the associated issues, the annotated genome information and the results of experimental research must be combined [118, 119]. Different advancements such as the inclusion of regulatory and thermodynamics constraints in the FBA have been introduced to address the challenges associated with the application of this approach (see Table S3, Supplementary Material, for the list of FBA enhancements and Table S4 for the comparison of FBA and MFA).

The constraint-based modelling approaches are particularly useful when cellular activities involved in a particular cell physiology of interest are not well understood, hence hindering the application of mechanistic modelling approaches. The lack of knowledge of the mechanistic basis of cell activities results in the increasing uncertainty of corresponding model parameters. In such a situation, a hypothesized objective function based on the established knowledge on the global cell behaviour (i.e. growth rate, ...) justifies the rationale behind modelling the metabolic network using constraint-based approaches. However, one of the adverse consequences of this is that the constraint-based models are generally poor at extrapolating outside the imposed experimental conditions. These models are also weak in accommodating sequential changes in the conditions, i.e. in the fed-batch regime, which are frequently occurring in bioprocesses, as well as describing medical cases.

2.1.5 Thermodynamics-based constraints

Thermodynamics laws dictate that the direction of every spontaneous reaction is from lower entropy to higher. Thus, a living organism encompassing the vital nonspontaneous reactions must have a continual energy input and then dissipate some of this energy in the form of unusable heat, in order to continue existing. Boltzman had already understood this principle when he hypothesized that:

> The general struggle for existence of animate beings is therefore not a struggle for raw materials, nor for energy, but a struggle for entropy which becomes available through the transition of heat from the hot sun to the cold earth (Boltzman, 1886).

Hence, the biological systems cannot attain true equilibrium state, and their thermodynamics is studied under the non-equilibrium steady state (NESS). Classical thermodynamics can describe the direction of change for systems nearing equilibrium, however, when the distance from equilibrium state exceeds a critical value, the system may exhibit *non-equilibrium structured states* that are maintained only with a continual input of energy [120, 121]. This is one of the reasons that makes circulation of energy in the cell possible and necessary. Thus, the direction of reactions in the metabolic pathways is determined as the result of the consistent manifestation of reactions taking place far from equilibrium. The true equilibrium is never reached unless the cell is practically dead, i.e. insignificant regulated metabolic activity is observed [122–125]. Imposing the constraints that emerge from this underlying theory narrows down the solution space of the metabolic model and thus ensures complying to the thermodynamics laws [2].

Estimation of the Gibbs free energy change

A positive net change of the Gibbs free energy of a reaction ($\Delta_r G$) suggests that the reaction cannot occur spontaneously, i.e. it is endergonic, unfavourable or nonspontaneous. Conversely, an exergonic reaction is one that releases *work energy* and can be assumed spontaneous or favourable. By the means of Gibbs free energy, the degree of thermodynamic favorability of the reactions in a metabolic network can be quantified [126]. The spontaneity described by the Gibbs free energy change is concerned with whether or not the reaction needs *continual input of energy* to take place. However, the change in Gibbs free energy between substrates and products reveals no information regarding the rate of the reaction.

This extensive variable has the following expression (Eq. 2.12) [127]:

$$\begin{aligned} \Delta_r G^{\circ'} &= -RT \ln K \\ \Delta_r G' &= \Delta_r G^{\circ'} + RT \ln Q \end{aligned} \tag{2.12}$$

where K denotes the equilibrium constant of the reaction, Q is the reaction quotient. The standard state which is denoted by ($^\circ$) is defined as $T = 298(K)$, $pH = 7$ and concentrations of molecular compounds, except for H^+, OH^- and H_2O, equal to $1mol/L(M)$.

Theoretically, the experimental values of a reaction equilibrium constant can be measured by making a solution of the enzyme and the molecular compounds participating in the reaction, to then allow them to reach equilibrium state; where the substrate and product concentrations are fixed and then measured to calculate the equilibrium constant [128]. A large number of chemical reactions have been investigated in this manner, and the acquired data have been collected in thermodynamics databases (see Table S3 of the Supplementary Materials). In addition, tables of thermodynamic data are available in literature such as in the extensive works of Alberty [129,130]; Thauer [131,132], and Dolfing [133,134]. However, for the majority of the biochemical reactions experimental derivation of $\Delta_r G^{\circ'}$ is troublesome [127], mainly, because of the lack of reliable models for metabolic chemistry, the difficulty of conducting *in vivo* measurements, and the loss of accuracy of *in vitro* values when transformed to the *in vivo* conditions [135]. Thus, experimental data of the equilibrium constants of metabolic reactions are scarce and mostly unreliable. To address this shortcoming, the Gibbs free energy change of the metabolic reactions is estimated as the sum of the Gibbs free energies of formation of the participating reactants and products (Eq. 2.13).

$$\Delta_r G_j^{\circ'} = \sum_{i=1}^{m} s_{i,j} \Delta_f G_i'^0, \tag{2.13}$$

where $s_{i,j}$ is the stoichiometric coefficient of compound i in the reaction j, and $\Delta_f G_i'^\circ$ denotes the standard Gibbs energy change of formation for the compound i. As a result, $\Delta_r G_j'^\circ$ is calculated for the reaction j. Consequently, the problem then becomes how to estimate the Gibbs energy change of formation for the participating molecular compounds. The Gibbs energy change of formation for a compound can be calculated from its standard Gibbs energy of formation and its thermodynamic activity [136]. Group Contribution Methods (GCMs) GCMs are statistical estimation methods for the estimation of Gibbs free energy of formation [137]. In the GCMs, it is assumed that the standard Gibbs formation energy of a metabolite is a linear summation of the formation energies of its constituent molecular substructures (or groups, denoted as P_i) [126, 127, 138]. Moreover, a common reference of estimation for

all functional groups that are involved is defined. If a specific sub-structure appears more than once in the main molecular compound, number of the occurrences must be taken into account as a coefficient applied to the contribution of that specific substructure (N_i). The general formulation of the property calculation is as follows (Eq. 2.14).

$$P = P_0 + \sum_{i=1}^{q} N_i P_i, \tag{2.14}$$

Developed from the classic work of Mavrovouniotis [127], Jankowski et. al [126] provided a version of GCM tailored for biochemical networks. By this method one can estimate the standard Gibbs free energy of formation $\Delta_f G'^0$, and consequently the standard Gibbs free energy of reaction $\Delta_r G'^0$, based on Eq. 2.13. In the recent years, the accuracy and the scope of application of GCM have been continually improved, but, however, there are still some issues that are limiting the confidence interval of this method estimation. As categorized by Du et. al (Figure 1 in [128]), the issues are associated with estimation accuracy, data quality and convergence, and inherent difficulties of the GCM methodology. Some promising directions to address the current shortcomings consist in gathering more curated thermodynamic data for fitting of Gibbs formation energies, and to extend current methods so they can calculate equilibrium constants as a function of temperature [139]. In addition, uncertainty at different degrees arises from several factors that affect GCM calculation, such as ionic strength, ion concentration, pH and temperature [139] and certain group interactions cause large errors in the total value of the property estimated [127, 140].

Modelling approaches complying to the thermodynamics-based constraints

In order to rule out closed reaction cycles from FBA, *Energy Balance Analysis (EBA)* enforces nonlinear thermodynamics-based constraints on chemical potential [141, 142]. When metabolomics data is available, Network-embedded Thermodynamic analysis (NET analysis) developed by Kummel et al. [143] can be used as a computational thermodynamics-based framework for the estimation of the feasible range of the Gibbs free energy change of reactions of the network under physiological conditions. In addition to the standard Gibbs free energy change of formation of the metabolites, NET analysis requires metabolomics data, predefined directions for the reactions and a stoichiometric metabolic model. However, the lack of the direction assignment does not lead to wrong results, but less insight from the metabolomics data. Interestingly, NET analysis does not require a predefined metabolic objective of the cell and thus the identification of the underlying thermodynamic infeasibilities is not biased. The types of results that can be obtained from using NET analysis comprise ascertaining

the thermodynamic consistency of measured metabolite data, prediction of concentrations of unmeasured metabolites and identification of the potential enzymes sites for regulatory actions [117, 143].

In the method proposed by Hatzimanikatis and co-workers [144], named Thermodynamics-based Metabolic Flux Analysis (TMFA), the CBMs is augmented with the Gibbs free energy change of the reactions to form a mixed integer linear programming (MILP) problem. The solution of this MILP optimization problem eliminates any thermodynamically infeasible flux distributions and, moreover, estimates feasible metabolite activity ranges. The level of insight provided by TMFA is largely determined by the number of reactions for which the Gibbs free energy change is known. In a genome-scale model of *E. coli (iJR904)* [145], the experimental values of Gibbs free energy change of the reactions and formations are only available for 5.6% of the reactions and 11% of the compounds, respectively. However, employing GCMs resulted in estimating the Gibbs free energy change for more than 90% of the reactions and metabolites [126], thus, allowing the implementation of TMFA on the *iJR904* metabolic model. The suggested ways to overcome the superficial infeasibility of the essential reactions includes correcting for uncertainty involved in the estimation of Gibbs free energy changes and adjusting the metabolites activity ranges. The impact of such remedies is studied for the reaction dihydroorotase in the *E. coli* genome-scale metabolic model [144]. (For the equations and inequalities that describe the mathematical formulation of TMFA refer to Text S1 in the Supplementary Materials.)

Generally, the implementation of CBMs with the thermodynamic constraints divides the reactions of a network into three categories. First, bottlenecks on flux direction, which are the reactions with $\Delta_r G'$ close to zero. The second is the reactions that have exclusively highly negative values of $\Delta_r G'$, meaning they are independent of the concentration of metabolites. The reactions of this category receive a special attention as regulatory points of the cell metabolism. The third category is the reactions in which $\Delta_r G'$ is highly dependent on the metabolites concentrations. These reactions can be used to determine the feasible range of concentration or concentration ratios. In many biochemical reactions, we observe interconversion of some known pairs of molecular compounds such as ATP/ADP, $NAD^+/NADH$ and $NADP^+/NADPH$. These metabolites are assumed to be tightly regulated and have a relatively complicated structure, but minimal structural difference between one another, therefore, a more accurate contribution share for their interconversion is estimated by GCMs [127]. Henry et al. [144] found the feasible ranges for $NADP+/NADPH$ and $NAD+/NADH$ in a genome-scale metabolic model of *E.coli* and showed that the ratio ranges comply to the assumption that these energetic pairs are tightly regulated. In fact, the cell maintains the ratio of the former energetic pair close to the maximum allowable value, while the ratio of

the latter is kept close to its minimum value. It is in contrary to the majority of metabolites where concentrations are maintained close to the logarithmic means of minimum and maximum allowable activities, suggesting the wider range of fluctuations allowed for their activities. The advantages of using TMFA can be described into three aspects [138]:

- Assignment of thermodynamically feasible directions to all reactions in a network.
- Elimination of futile cycles (infeasible closed cycles).
- Consideration of thermodynamically coupled reactions.

In addition, by coupling thermodynamics of a metabolic network with the EFMA, one can reduce the number of relevant EFMs [2, 146]. In fact, it was demonstrated that to have a thermodynamically feasible intracellular flux distribution, all the founding EFMs must be thermodynamically feasible too. Based on this, the authors of [117] discarded 46% of a total of 71.3 million EFMs generated for a compartmentalized model of central metabolism in *S. cerevisiae*. Peres et. al [146] compared the traditional MPA approach, with an approach where they incorporate the external metabolite concentrations in addition to the standard Gibbs free energy change of reactions. Consequently, the incorporation of external metabolites allowed for the enumeration of different EFMs in the different culture conditions and cell growth phases. However, to overcome the need of discarding the already generated EFMs and thus to decrease computational cost, *tEFMA* was developed that generates only thermodynamically feasible EFMs in the first place [147]. Figure 2.5 gives a quick overview of the context of applications of thermodynamics-based theories.

Extensions of constraint-based flux balance models simulate growth dynamics

Even though enzyme activities in a cell are constantly changing as a result of metabolic regulation, the validity of the steady state assumption is justified in many cases, such as in bioprocesses in continuous cell cultures and in early exponential growth phase of batch fermentation. However, there are many practical situations like in batch and fed-batch industrial bioreactors and in medical cases, where cells exhibit a dynamic behaviour [148]. In recent efforts, further developments have been done on the constraint-based approaches to form dynamic flux balance analysis (dFBA) and dynamic metabolic flux analysis (dMFA) [26,61,63,64,96]. Studying a biosystem under transient conditions using dFBA and dMFA approaches uses steady state models sequentially, evaluating the state of the biosystem at each sampling time interval [15]. In this convenient way, the model can be used to estimate the dynamic changes

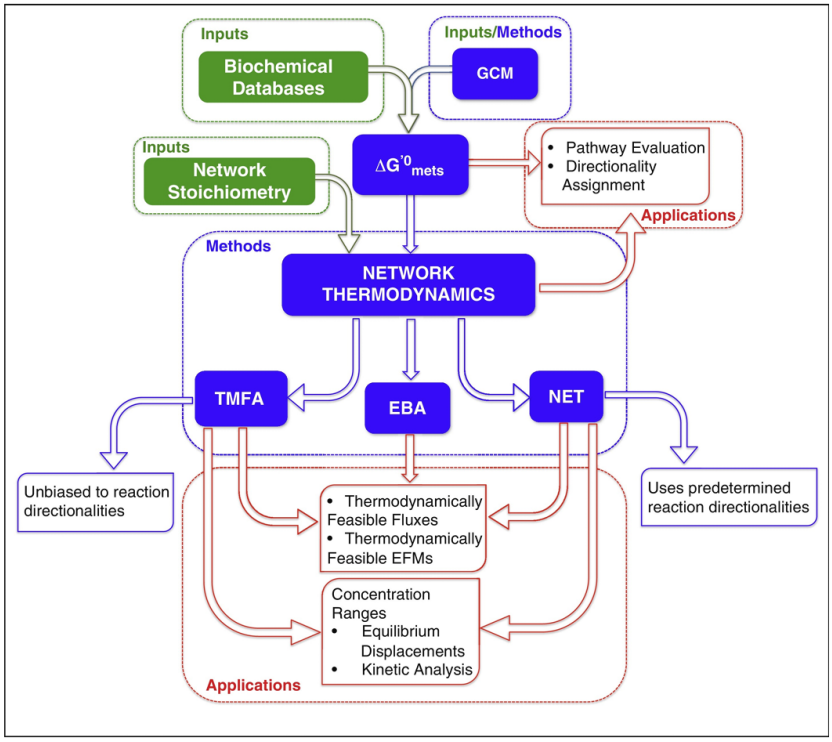

Figure 2.5: **Thermodynamics-based modelling of metabolic networks.** Thermodynamics integration to metabolic networks modelling allows computation of fluxes and EFMs that are thermodynamically feasible. The feasible concentration ranges of metabolites can be estimated and then used for estimating the distance from the equilibrium. Reused with permission from [2].

.

Table 2.1: Thermodynamically constrained models along with their specifications and reference to their origin.

Thermodynamics-based Constrained Model		
Theory Name	**Specification**	**Reference**
Energy balance analysis (EBA)	Identifies reactions' direction and flux limits based on the Gibbs free energy value of reactions.	[141]
Network-embedded thermodynamic analysis (NET)	Identifies thermodynamically feasible flux range or metabolite concentration range, based on the predefined standard Gibbs free energy and predetermined flux directions.	[143]
Thermodynamics-based metabolic flux analysis (TMFA)	Identifies flux direction, allowable flux range and also concentration ranges. Incorporates a MILP optimization. Based on the predefined standard Gibbs free energy.	[135]

of intracellular fluxes and to decipher the major activated metabolic pathways in the reaction network without the need for a large set of predefined kinetic parameters [149]. The particular advantage of this enhancement is its capability to describe the growth dynamics of biomass [150]. This approach uses the constraint-based models at each assumed steady state to estimate the (momentarily) growth constant, to then find a new set of the exchange rates through integrating the macroscopic mass balance equations of substrate consumption and product formation. The calculated exchange rates are then used to constrain the next iteration of the constraint-based model simulation updating the growth constant, and the iterations go forward. To have a smooth continuous simulation of the bioprocess dynamics, the time between two consecutive operation points, i.e. two consecutive cell count measurements, is modelled with a continuously differentiable function, such as Monod model. The application of this line of metabolic model development has been shown in describing dynamic growth on multiple carbon sources for a small metabolic network [151, 152]. In essence such model obeys a discrete-continuous framework, where the discrete states are decided based on logical transcriptional regulatory rules and then continuous states, i.e. concentrations and fluxes, are computed according to the specific parameters of each discrete state.

2.1.6 Kinetic modelling

Dynamic kinetic models attempt to provide a mechanistic description of the biosystem in terms of the cell enzymatic activity and mass balances over intracellular metabolites. Thus, the mass conservation equations for the components form a system of ordinary differential equations (ODEs). The values of the intracellular fluxes and concentrations are derived as a numerical solution of the conservation equations. Kinetic models bring about the possibility to interpolate and extrapolate dynamic behaviour of a system, in a consistent fashion, in situations other than the one on which the model was originally validated [13, 153, 154]. Moreover, this type of mathematical models can simulate changes in the relative activation of enzymatic reactions as a function of parameters of the system and the initial concentrations [155–158]. The general mathematical formulation is given in Eq. 2.15 as follows:

$$\frac{d\vec{x}(t)}{dt} = S.\vec{v}(t), \qquad v_j(t) = f_j(\vec{x}(t), u, \theta), \qquad \text{with:} \qquad x(t_0) = x_0(\theta) \tag{2.15}$$

$$x_i(t) = \int_{t_0}^{t} \left(\sum_{j=1}^{n} s_{ij}.v_j(t) \right) dt + x_0(\theta_i) = \int_{t_0}^{t} \left(\sum_{j=1}^{n} s_{ij}.f_j(\vec{x}(t), u, \theta) \right) dt + x_0(\theta_i) \tag{2.16}$$

where the vector $x(t)$ is a vector of dimension m of time-dependent state variables such as extracellular and intracellular metabolites concentration, the vector v is here a nonlinear time-dependent vector of the reaction fluxes, which depends on $x(t)$, a vector of regulatory inputs $u(t)$, and θ which stands for the collective set of kinetic parameters and initial states [73, 159, 160]. In some cases, algebraic equations or additional ODEs are added to this general representation to reflect, among others, conserved moieties total concentrations, volume changes or supplementary descriptive variables [159]. The flux vector is equivalent of the vector of biochemical reaction rates and is reported in the units of mole of reacting matter per mass of biomass per time. The inclusion of reaction kinetics in Eq. 2.15 introduces nonlinearity to the formerly linear mapping of a reaction rate vector into a space of accumulated concentration time derivatives of metabolites ($\dot{x}(t)$) [70], and in return, provides the capability to calculate concentrations change and intracellular fluxes at transient, i.e. not only at steady state. The function f_j is to provide a mechanistically valid description of reaction rates of the reactions taking place in an underlying metabolic network. The mathematical formulation of the reaction rate is driven by kinetics theories (i.e. transition state theory), thermodynamics theories (i.e. theory of equilibrium, linear and nonlinear non-equilibrium theory) and first principles (i.e. mass balances of metabolites). The impact from all the reaction rates (metabolic fluxes) where the metabolite i participated is calculated based on the corresponding row of the stoichiometric matrix S, (first term on the right-hand side of the

Eq. 2.16), integrating over this term for a period ending at t provides the concentration of the metabolite i at t. However, the numerical solution of the underlying ODEs integration is quite troublesome due to the large number of parameters, the nonlinear nature of enzymatic reaction rates and the stiffness caused by the difference between the time scale of the underlying bioprocesses. Addressing the issues that are obstructing the use of kinetic modelling on large-scale metabolic models, i.e. genome-scale models, has been focus of the research community. In the following, the most relevant mathematical formulations are presented and their merits and drawbacks are discussed.

Approximate kinetic formats and the quantification of metabolic regulation

Alternative approaches to mechanistic kinetic modelling that uses non-mechanistic kinetic models have primarily emerged based on a reasoning that suggests, because of the homeostasis constraint, redesign of metabolic networks does not require detailed mechanistic models [23, 161]. Thus, the development of approximate kinetic formats provided mathematical equations to approximate reactions kinetics in the neighbourhood of a reference metabolic state, with a decreased number of kinetic parameters. The approximate kinetics that are formed generally have four characteristics in common: proportionality between reaction rate and enzyme concentration, capability to reach the plateau phase for the reaction rate against substrate concentration, involving the least possible kinetic parameters and providing analytical solution at steady state [162]. Approximate kinetic formats explicitly incorporate metabolite concentrations and provide a framework for the quantification of metabolic regulation [73, 162–166].

Metabolic Control Analysis (MCA) Sensitivity analysis (SA) of metabolic model parameters for determining their influence on the model simulation is crucial [167]. The initial values of metabolite concentrations and of kinetic parameters, as well as the constraints on flux rates may be considered as sensitive factors influencing model behaviour [168]. The SA divides into two major groups, namely, local and global SA. Local SA methods consider that only one input varies at a time and the perturbation results are observed in outputs, while the rest of parameters is maintained unchanged. Conversely, in global SA, all inputs vary simultaneously and the resulting effect on model output can be investigated by intense computational cost. In fact, global SA uses repeated sampling from the probability distribution of each input parameter to obtain numerical results covering the entire range of variation of the metabolic model parameters [169]. Local sensitivity analysis is widely used for characterizing the effect of parameter changes on the solution of dynamic models in the neighbourhood of a certain reference state. For a metabolic model, the linearization about the reference state is generally

expressed as a first order Taylor series approximation [11, 169, 170]:

$$\frac{\partial c^i(t,p^\circ)}{\partial p_g} \approx \frac{c^i(t,p_g^\circ+\Delta p_g,p^\circ_{k=1,\ldots,m,k\neq g})-c^i(t,p_g^\circ-\Delta p_g,p^\circ_{k=1,\ldots,m,k\neq g})}{2\Delta p_g} \tag{2.17}$$

where c^i is a property of the metabolic system, i.e. the metabolite concentration, p° is the initial set of the parameters, and p_g is the parameter subject to change. Explicit differentiation can be used when the model explicitly states the relation between the desired output and the input, in this case sensitivity function represents an analytical solution which can be used to calculate the sensitivity. When the output is provided as an implicit function of the input parameter(s), implicit differentiation might be used. However, in practice local sensitivity coefficients are calculated with numerical approximation [11]. Simulation is used to determine a variable of interest (i.e. steady state metabolic flux) at two different parameter values (i.e. enzyme total concentration), then various numerical forms of the Eq. 2.17 formula are used to calculate the absolute and relative local sensitivity [17].

One of the well-known SA frameworks in the metabolic modelling context is metabolic control analysis (MCA) that is widely used to study metabolic pathways interactions [171–174]. By the use of MCA, the effect of a parameter perturbation on metabolic fluxes or metabolite concentrations can be estimated or reported in an experimental scenario [174]. The three main coefficients in MCA are defined based on the notation in [161] in Table 2.2.

On theoretical grounds, dynamical stability of a stable metabolic system dictates that upon the perturbation of one state of the system at steady state, the states of the system, i.e. metabolite concentrations and intracellular fluxes, eventually returns to the same steady state. When one parameter is perturbed, structural stability ensures that the system eventually returns to the same steady state or *to one close by*. If the system is unstable, such perturbations result in the diverge of some or all of the states. Hence, a subset of the system states is chosen that relative to the period of the observations, the concept of stability and steady state for them is associated with the same time period, i.e. the phenomena have close enough relaxation times [174]. Then, the study of the metabolic system is confined in the neighbourhood of such stable steady state of the biosystem. Thus, in practice, MCA has a limited capacity in quantification of the control distribution within the metabolic network as the theory stays only valid for small variations in parameters, where the linearization in Eq. 2.17 is a good approximation of the nonlinear kinetics behaviours. The rigorous math-

Table 2.2: **The main MCA coefficients.**

MCA coefficients table		
Name	**Mathematical formulation**	**Description**
2*Control coefficients	Flux: $C_{ij}^{J} = \frac{e_j^\circ}{J_i^\circ}\frac{dJ_i}{de_j}$ Intracellular metabolite: $C_{ij}^{x} = \frac{e_j^\circ}{x_i^\circ}\frac{dx_i}{de_j}$	The coefficients are a measure of the relative change in a flux or concentration upon a relative change of an enzyme activity level
2*Response coefficients	Flux: $R_{ij}^{J} = \frac{c_j^\circ}{J_i^\circ}\frac{dJ_i}{dc_j}$ Intracellular metabolite: $R_{ij}^{x} = \frac{c_j^\circ}{x_i^\circ}\frac{dx_i}{dc_j}$	The coefficients are a measure of the effect of a change of an external parameter, on intracellular fluxes and concentrations
Elasticity coefficient	Intracellular reaction rate: $\varepsilon_{ij}^{x} = \frac{x_j^\circ}{J_i^\circ}\frac{\partial v_i}{\partial x_j}$	The elasticity is a local measure quantifying the relative change in reaction rate upon a relative change in metabolite concentration, while maintaining other concentrations and parameters constant

ematical formulation of MCA was exploited to derive several theorems to relate the MCA coefficients in global relationships, i.e. the summation and connectivity theorems, among others (for a mathematically rich elaboration see [173] and [174]). Attempts to extend the application of MCA lead to the expansion of enzyme concentration changes for which the theory holds valid [161,175]. These *approximate kinetic formats* are log-lin and lin-log based on the approximation procedure employed to derive each one. However, they are typically formulated directly from the stoichiometry of the metabolic network as given in Table 2.3. Biochemical System Theory (BST) Power-law representation of the ODEs (Eq. 2.15) is the key ingredient of more general Biochemical System Theory (BST) developed primarily by Michael Savageau and Eberhard Voit in the 1960s [176, 177] prior to MCA development. In the two global formats of BST, namely, Generalized Mass Action (GMA) and canonical S-system representations the sums and differences of multivariate products of power-law functions are used to represent dynamics of changes in biochemical processes and metabolite pools, respectively [163–166]. Sharing much of the modelling philosophy with MCA, BST gives the same steady states solution, and the same local stability properties (at the same reference state), which implies that the different parameters of BST can be derived from the coefficients used in MCA, i.e. kinetic orders in BST from elasticities in MCA [178–180]. This theory has been extensively studied and discussed in several books and journal articles, one can refer to [178] for a comprehensive review.

where m denotes all concentrations including the dependent (intracellular concentrations) and independent (extracellular concentrations) variables; k and j count reactions and variables respectively. γ represents rate constants (either measured or estimated); f denotes kinetic orders in BST representation. When reformulated to show S-system representation, the first term on the right-hand side is an aggregated term for all the reactions flowing in the pool and the second term denotes all of the fluxes leaving the metabolite pool [181]. In BST kinetic orders play a major role in the purpose of modelling, unlike mass action kinetics they can acquire negative or non-integer values. For example, an inhibitory variable can be included with a negative kinetic order. The direct biological meaning of parameters in BST representation is an advantage [178, 179]. Du et al. [23] listed two general situations where there is a higher chance of rate law approximations success:

- In the domain where the underlying assumptions of a specific rate law approximation remain valid and not substantially violated throughout simulations;
- The rate laws are not the most important single factor in determination of dynamic behaviour of the network.

Table 2.3: **Various approximate kinetic formats based on MCA and BST using reference parameters and reference elasticities (See the following paragraph for the notation definition).**

Approximate kinetic formats based on MCA and BST		
Type of approximate rate law	**Mathematical formulation**	**Eq.**
Log-lin	$\frac{v_j}{J_j^\circ} = 1 + \ln\left(\frac{e_j}{e_j^\circ}\right) + \sum_{i=1}^{m} \varepsilon_{ji}^\circ \ln\left(\frac{x_i}{x_i^\circ}\right)$	(1)
Lin-log	$\frac{v_j}{J_j^\circ} = \frac{e_j}{e_j^\circ}\left[1 + \sum_{i=1}^{m} \varepsilon_{ji}^\circ \ln\left(\frac{x_i}{x_i^\circ}\right)\right]$	(2)
S-system	$\frac{dx_i}{dt} = V_i^+ - V_i^- = \alpha_i \prod_j^m x_j^{g_{ij}} - \beta_i \prod_j^m x_j^{h_{ij}}$	(3)
General mass action (GMA)	$v_j = \left(\gamma_{+j} \prod_i^m x_i^{f_{ij}^-} - \gamma_{-j} \prod_i^m x_i^{f_{ij}^+}\right)$	(4)

In other studies by Heijnen [162] and Visser and Heijnen [161], biochemical system theory (BST), generalized mass action (GMA) and lin-log approaches were compared, and authors concluded that the lin-log approach has the best approximation capacity and its solutions are valid for large changes in enzyme activities. However, Wang et al. [179] argue Heijnen's theoretical conclusion claiming that lin-log and log-lin representations might misbehave when the accompanying substrate concentrations approach zero, it is the opposite of the problem with GMA highlighted by Heijnen, where the governing equations' outcome approaches infinity for unbounded concentrations, i.e. inability to simulate saturation. In another review study, Voit proposed that for metabolic design and theoretical study of the principles of operation, S-system format can provide a better default point to start because it allows the user to perform algebraic calculations at steady states and more straightforward stability and sensitivity analysis [181].

Regulated kinetic metabolic models

Flux regulation enabling continuous cellular activities resemble applying certain control strategies to maintain, increase or decrease the rate of production or consumption of biomolecular compounds (i.e. metabolites, enzymes or signalling proteins). *Feedback repression* and

induction are two examples of these regulation scenarios in epigenetic level, which is the study of interactions in genetic level affecting gene expression, i.e. transcriptional regulation. Their counter-scenarios at metabolic levels are *feedback inhibition* and *enzyme activation* [182]. Competitive inhibition and allosteric regulation are two prevalent ways among others, which have been formulated mathematically to model enzyme activity regulation, i.e. post-translational modifications [183]. The regulated kinetic models of metabolic network reactions with the condition-specific parameters of affinity constants and maximum reaction velocities can simulate the nonlinear metabolic behaviour of biosystems in an extended time frame, i.e. cellular growth dynamics. The rationale justifying this practice is in two folds, first, generality of the governing assumptions for first-principle modelling covers biological systems such as the viable cell. Second, culture dynamics is a phenotypic behaviour of the viable cell during the process period. Such dynamic models either do not consider transcriptional regulation or consider it in a non-kinetic way, for example, qualitatively through Boolean logic or multi-valued logical rules [21, 184]. When modelling a biological regulatory system of interest, special care must be devoted to the time frames of reactions and processes involved, as it appears commonly in biology that a system in its whole may span over various time scales of nanoseconds [185] to days [17]. As a rule of thumb, reactions that happen with a time constant one order of magnitude larger than time constant of the system are taken into account as frozen, and concentration of metabolites associated with them is considered as fixed variables. However, it is trickier to handle the faster reactions, in this case two assumptions are used to approximate underlying concentrations. First, the rapid equilibrium assumption and second the quasi-steady state (QSS) assumption [17, 68]. Briefly, QSS assumption states that where the enzyme is available in catalytic amount, ES complex reaches steady state fast enough that we can assume its concentration remains constant. Rapid equilibrium assumption implies that the substrate rapidly reaches equilibrium with the enzyme-substrate complex, with the rates of both directions of the binding reaction being faster than the conversion step of the enzyme-substrate complex to enzyme-product complex (see the Convenience Kinetics (CK) section). It is worth mentioning that the time frames of regulation at the metabolic level are usually notably shorter compared to the regulations imposed at the epigenetic level. It is believed that the former takes place in the order of minutes, while the latter happens in hours [186]. Accordingly, the metabolic level regulation of enzyme activity can be viewed as fine-tuning, as opposed to the coarse control associated with regulatory mechanisms at the epigenetic level [3]. The comparative time scale of cellular operations along with a schematic representation of genetic and metabolic level regulatory mechanisms are given (Figure (2.6)).

Metabolic regulation is increasingly considered in metabolic modelling studies, where its

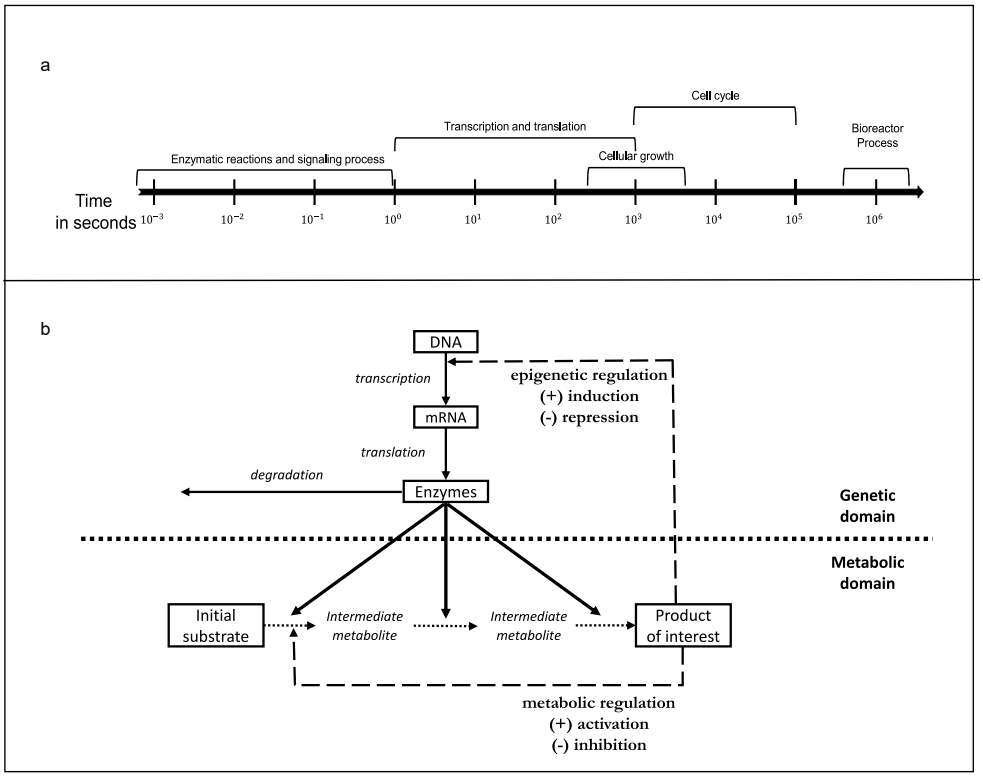

Figure 2.6: **Bioprocesses time frames and possible regulation scheme.** (**a**) Time scale of cellular operations. (**b**) Interacting regulatory mechanism in genetic and metabolic level (inspired by p. 16 [3]).

importance is clearly demonstrated [158, 187–190]. However, particularly the hierarchical nature of relaxation times in regulatory mechanisms along with the inherent interactions at signalling pathways makes the time-dependent quantification of the cellular regulations challenging [190–192]. The kinetic metabolic model formulations that can accommodate dynamics of the regulatory interactions between modifiers and enzymes have the advantage to be used to generate hypotheses for a wider set of experiment designs. Following the enhancements of GMA in [21,67], Drager et al. [166] reformulated GMA to describe a reaction including a modification term as follows:

$$v_j(x,\theta) = F_j(x,\theta)\left(\gamma_{+j}\prod_i^m x_i^{f_{ij}^-} - \gamma_{-j}\prod_i^m x_i^{f_{ij}^+}\right) \qquad \text{GMA representation} \tag{2.18}$$

The F_j function is used to introduce activation and inhibition effects [160]:

$$F_j(x,\theta) = \frac{[A]}{k_j^A + [A]} \qquad \text{with} k_j^A \geq 0 \tag{2.19}$$

$$F_j(x,\theta) = \frac{1}{1 + k_j^I[I]} \qquad \text{with} k_j^I \geq 0 \tag{2.20}$$

In the following kinetic rate laws, the metabolic regulation can be accommodated explicitly and with a mechanistic justification.

Michaelis-Menten kinetic expression for enzymatic reactions

L. Michaelis and M.L. Menten [193] developed a general theory in 1913 to explain enzyme kinetics, following V. Henri [194] who had already taken important steps toward describing saturable enzyme kinetics. The mathematical model developed based on this theory is commonly used to describe the kinetic expression of the enzymatic reactions. It simulates experimental results by assigning two parameters, v_m^+ and K_S^M. This model has the least error when applied on a reaction with one substrate being converted to one product after forming the substrate-enzyme complex as the intermediate [193, 195]. Michaelis-Menten model is the default choice to model an enzymatic reaction [181, 190], numerous studies incorporated Michaelis-Menten kinetics in their kinetic models [25, 158, 166, 196, 197]. The schematic of Michaelis-Menten mechanism and the corresponding mathematical formulations are as shown in Figure 2.7.

The reversible Michaelis-Menten expression is given in Figure 2.7-a for a uni-uni reaction, by equating $[P]$ to zero it reduces to the well-known classical Michaelis-Menten equation, the formula for the case where inhibition is introduced is given (Figure 2.7-b) [21]. In a

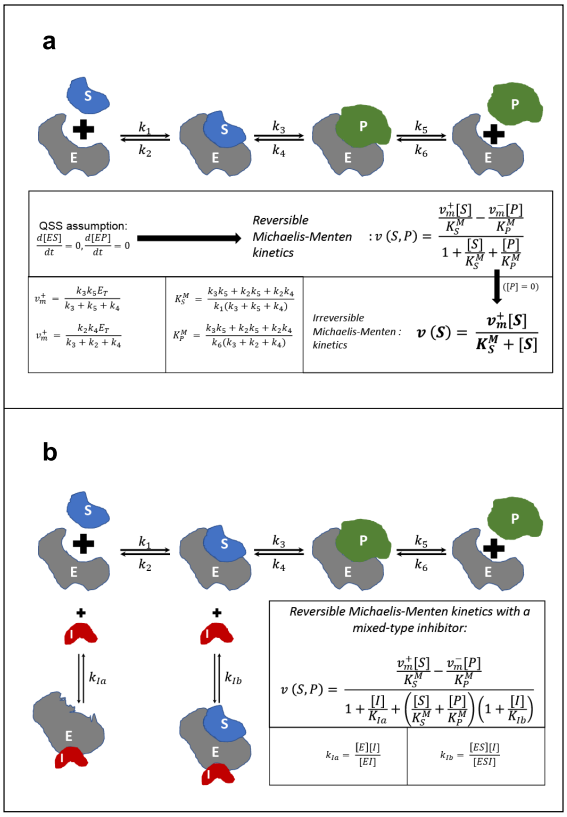

Figure 2.7: **Michaelis-Menten enzymatic reaction mechanism.** (**a**) Substrate S is converted to product P through an enzymatic reaction catalyzed by the enzyme E. (**b**) The inhibitor I is introduced and inhibits both the enzyme and enzyme-substrate complex.

review study Tummler et. al [160] discuss the assumptions accompanied with enzyme kinetics and how new types of experimental data can be incorporated to enhance estimation and calculation of Michaelis-Menten parameters. The experimental measurements are categorized for flux v, enzyme E_T, kinetic parameters and substrate concentration S as follows:

1. Metabolic flux quantification: Carbon labelling experiments, uptake rate of nutrients, secretion rate of products.
2. Enzyme concentrations measurement: MS/MS technique, absolute quantification by standard peptides or label-free methods—in general, quantitative and qualitative proteomics.
3. Kinetic parameters: database values for *in vitro* assays, existing models in literature with similar settings, estimation techniques.
4. Substrate concentrations: Metabolomics.

Indeed, Ghorbaniaghdam et al. [25] considered regulatory functions from energy shuttles ATP/ADP and cofactors $NADH/NADPH$ in their dynamic model based on the multiplicative Michaelis-Menten kinetics, to simulate the behaviour of CHO cells. They modified flux kinetic equations to consider the effects of inhibitors and activators, and obtained satisfactory results simulating experimental data of the extracellular (and intracellular) analytics with low errors.

Convenience Kinetics (CK)

This framework is the most similar to the conventional Michaelis-Menten representation of enzymatic reactions in terms of formulating kinetics of a reversible enzymatic reaction. The authors in [22] point out thermodynamic dependency of the kinetic parameters in the rate law representation of the reactions in a metabolic network as an undesired phenomenon, because it obstructs the scan of the kinetic parameter space for parameter estimation methods, i.e. optimization algorithms. They argue that this thermodynamic dependency arises from the theory that the Gibbs free energies of formation of the metabolites are determinants of the equilibrium constants of the network reactions, and the equilibrium constants are kinetic parameters affecting the model behaviour. Thus, it is likely for a mechanistic model of metabolism to produce thermodynamically wrong outputs if such dependencies are not addressed. To resolve this issue, they derive CK from a random-order enzyme mechanism with simplifying assumptions with respect to the order and the binding energies of enzyme binding and dissociation. CK can be extended to model kinetic laws in an entire metabolic

network, then a set of independent parameters are estimated and used to determine the rest of kinetic parameters. Also, the formulation justifies the regulatory terms mechanistically, in this case, the assumption is that the enzyme has sites for the activation and inhibition terms where binding mechanisms is comparable to the binding mechanisms of the formation of enzyme-substrate complexes. Parameters k^A, k^I in Eq. 2.23 are defined similarly to the same parameters in Eq. 2.20 and denote the dissociation constants for the activation and inhibition enzyme binding. The mathematical formulation for a reversible reaction in a metabolic network of any size is derived as follows (Eq. 2.21):

$$v_j = E_j \frac{k^{cat}_{+j} \prod_i \left(\frac{x_i}{K^M_{ji}}\right)^{s^-_{ij}} - k^{cat}_{-j} \prod_i \left(\frac{x_i}{K^M_{ji}}\right)^{s^+_{ij}}}{\prod_i \sum_{m=0}^{s^-_{ij}} \left(\frac{x_i}{K^M_{ji}}\right)^m + \prod_i \sum_{m=0}^{s^+_{ij}} \left(\frac{x_i}{K^M_{ji}}\right)^m - 1} \tag{2.21}$$

$$\tag{2.22}$$

where $s^{\pm}_{ij}$ are absolute values of all positive and negative elements of stoichiometric matrix S. E_j, denotes enzyme concentration in the reaction j in (mM). In the original formulation, $\tilde{x}_i$ equals x_i/K^M_{ji}, where K^M_{ji} is the counterpart of Michaelis-Menten constants denoting the dissociation constant of enzyme j for the metabolite i (in mM). $k^{cat}_{\pm j}$ denote turnover rates in (s^{-1}) for the reversible reaction of substrate-enzyme and product-enzyme complexes. Similar to Eq. 2.20, h_A and h_I functions are introduced for the activation and inhibition effects from a metabolite (m) on the reaction j. k^A_{jm} and k^I_{jm} denote the dissociation constants for activation and inhibition modifiers [22, 166].

$$v^r_j = \prod_m h_A(x_m, k^A_{jm})^{w^+_{jm}} h_I(x_m, k^I_{jm})^{w^-_{jm}} \times v_j \tag{2.23}$$

$W^{\pm}$ being regulation matrix, with $w_{jm} = 1$, -1 or 0 denoting activation, inhibition or no regulatory action from the metabolite m on the enzyme j, respectively [22, 166].

The kinetics formulation of reaction fluxes is the single most important factor that determines success of dynamic kinetic modelling. It is different from constraint-based models (CBMs), where the formulation of a hypothesized objective function determines the model performance [1, 175]. Enforcing the Wegsheider condition constrains the equilibrium constant values of the reactions of the network and consequently provides relationships between the kinetic parameters, such relationship are employed to determine dependent kinetic parameters from the independent ones.

2.1.7 Parameter estimation formulation

Possibility of constructing large-scale kinetic models significantly depends on the availability of physiologically valid kinetic parameters. The kinetic parameters for mechanistic models include enzyme concentration, turnover rate of the enzyme for both the reaction directions, dissociation constants for reactants bound to the enzyme, inhibition and activation constants and equilibrium constants, i.e. the standard Gibbs free energy of the enzymatic reaction. The resulting large and non-convex parametric space with undefined dimensions renders the estimation of parameters particularly challenging and intractable when the network size is enlarged [22, 198, 199]. For the isolated enzymatic reactions, the kinetic parameters are estimated directly through *in vitro* assays specifically designed to characterize particular enzymes, and the values are collected in the enzymes properties databases [198,200]. However, it is quite cost and labour intensive to characterize the majority of enzymes in the metabolic network of an organism, in comparable experiment conditions. In fact, it has been shown that enzyme characteristics have considerable dependencies to the thermodynamic state of the reaction solution. This means that the possible differences in pH, ionic strength, temperature and abundance of cofactors between the observation situation and the original measurement setup may cause the measured kinetic parameters to be incompatible or inappropriate for a certain model of the observation. In the parameter estimation of a metabolic network, thus, it is inevitable in most studies to implement an indirect estimation of parameters through minimization of the discrepancy between experimental data and model simulations. Yet provided data in databases can be used as initial guesses and/or to impose bounds on the parameters to be estimated [163, 201]. Parameter estimation in the kinetic metabolic modelling can be formulated as an optimization problem subject to the context specific types of constraints [175,202,203]. When the problem is formulated, the estimation procedure starts from a set of initial guesses of the parameter values, which then are evaluated according to a quality measure such as fitness value of the comparison between model results and available data. Then, the parameter values are refined by scanning the parameter space in various directions until a satisfactory outcome is achieved, i.e. the parameters minimized an objective function which is usually the weighted sum of squared distances between model simulations and the associated experimental values. An optimization algorithm must be used to find *the best direction*, along which the parameter values are changed to calculate a reduced objective function value within the imposed constraints, consistent with the non-equilibrium thermodynamics of biosystem [166,204–208]. Of importance, a mechanistic kinetic metabolic model includes a set of nonlinear functions of the kinetic parameters and consequently forms a nonlinear optimization problem for which linear (or mixed integer linear) programming optimization methods perform poorly. However, regardless of the implemented optimization

algorithm, the workflow of parameter estimation can be summarized in the following global steps:

1. Objective function formulation: enhancements can be introduced by collection of the data in several duplicates to then adjust the relative weight of the error in states experimental values.

2. Constraints definition: bounds on parameter values should be introduced to keep them in the feasible biological ranges.

3. Algorithm and solver selection: nature of the optimization problem and available data play a significant role in this step.

4. Optimization option assignment: a set of options must be assigned for the solver. These options are descriptive of optimization algorithm's details and accuracy.

Given the following generic objective function formulation (Eq. 2.24), countless methods have been employed to solve for a vector of parameters θ minimizing the function $S(\theta)$ for a metabolic network under study.

$$\begin{aligned} y &= f(x, \theta) \\ S(\theta) &= \sum_{j=1}^{n} [\hat{y}_j - f_j(x_i, \theta)]^T Q_j [\hat{y}_j - f_j(x_i, \theta)], \end{aligned} \tag{2.24}$$

Evidently, the number of parameters increases with the size of the network, the extent of regulation details and the extent of interactions between the variables formulated by kinetic rate laws of the biochemical reactions. The wide array of optimization techniques search to find the minimum of $S(k)$ in a constrained subspace of R^p, where n is the number of free parameters, i.e. length of vector k if all parameters are independent. Mathematical formulation of the constraints is not trivial and demands an understanding of the biosystem. While global optimization algorithms are used to find the global minimum within the range of all inputs, local optimization methods are used to search for the local minima in the neighbourhood of a vector of initial guesses, i.e. a reference steady state of the biosystem. Global optimization methods are highly computationally demanding as opposed to the local optimization [159, 163]. Optimization methods can be categorized into deterministic and stochastic algorithms. In a deterministic optimization, no random element appears and algorithms are based on mathematical scheduling, i.e. derivative-based algorithms. Conversely, in stochastic optimization, random searching procedures are applied to find a next step in the direction of extrema, in this manner it is more likely that the algorithm will not be trapped

in local optimum [204, 205, 209]. Various studies confessed that there is not a single recipe to tackle the parameter estimation problem of metabolic modelling approaches [210–212], selection of an appropriate method depends on the type of formalism used to represent kinetic rate laws, the experimental data that is available to the modeller as well as the capacity of accessible computational frameworks [213]. In an ideal situation, a comprehensive set of collected data can be utilized to examine more than one complementary method of kinetic parameter estimation [211, 214]. To reduce the computational cost of parameter estimation, several approaches have been proposed in literature [153], here we list the most popular ones:

- To employ meta-heuristics optimization algorithms such as particle swarm [215] and genetic algorithm [216].
- To impose a reference state and estimate the biosystem parameters around this state [217].
- To introduce thermodynamic limitations and therefore limit the parameter solution space [218].
- To introduce local stability constraints [219].
- To reduce the model directly through the model reduction techniques (listed in [153]).

Drager et al. compared different evolutionary algorithms for the optimization of mixed models comprising CK and Michaelis-Menten kinetics in Corynebacterium glutamicum. They found that differential evolution (DE) and particle swarm optimization (PSO) resulted in the best parameters approximation, moreover, Tribes algorithm is useful for the first optimization attempts because of its performance and its user-friendly traits [166, 220]. In a related study of Spieth et al. [221], evolution strategy (ES) and DE led to the best parameters estimation. Taken together these studies, it can be concluded that DE is an adequate estimation method for kinetic models [220]. DE is an accurate, reliable, robust and fast estimation algorithm which has few control parameters, and an easy to use and powerful search capability [222,223]. However, due to its fast convergence speed and low risk of divergence, it may get trapped within local minima. In addition, it is sensitive to its control parameters and if the population size increases, it becomes more computationally demanding [201, 204, 223, 224]. On the other hand, authors in a recent study [207], developed a (deterministic) gradient-based kinetic parameter estimation algorithm that provides the best fit to multiple sets of metabolic data, i.e. fluxomics, metabolomics. A large metabolic network of *E. coli* (k-ecoli307) with 307 reactions and 258 metabolites is used to develop a kinetic model with 2367 kinetic parameters. The applicability of *K-FIT* is demonstrated in the estimation of the kinetic parameters

by fitting the model to 13C-labelling data of multiple genetic perturbation mutants. It is reported that the algorithm works three orders of magnitude faster in comparison with meta-heuristics for estimating kinetic parameters of the test model [**?**]. A general workflow for kinetic parameter estimation is given in Figure 2.8.

As pointed out several times in this review, the modelling of metabolic networks is intrinsically an iterative effort. In Figure 2.9, we show schematically where the reviewed modelling approaches fit in modelling of cell metabolism.

2.1.8 Perspectives and challenges ahead

In this review, we assessed the mostly used modelling approaches of metabolic networks. The metabolic network was understood as a highly interconnected system of bioreactions, where various carbon sources are metabolized to produce energy and product molecular compounds. Then, the approaches to characterize the metabolic networks of distinct organisms was discussed. We showed that the *in silico* description of metabolic fluxes in large networks at steady state is performed successfully, by the modelling approaches based on the understanding that resulting outcome of the cell operation is optimized with regard to certain objectives. In fact, the mathematical formulation of the metabolic networks relying on the stoichiometric matrix of the network motivated several approaches collectively referred to as constrained-based models (CBMs). This modelling paradigm has been further established in recent years by the development of theoretical and computational extensions based on CBMs principles. Following the contributions to develop dFBA based on both static and dynamic optimization methods [15, 54], efforts also focused on the integration of transcriptional regulation [225, 226], and the consideration of resource allocations in terms of enzyme production cost [27, 227]. The challenges associated with the expansion of the CBMs can be related to the biological justification of the underlying assumptions and the efficiency of the computational methods employed to solve the core optimization problems. The CBMs collectively proved successful in complementing some particular types of interests including recombinant DNA technology and optimal growth medium formulation by finding answers to the questions regarding estimation of maximum theoretical yield, estimation of the growth rate of a certain strain, or identification of candidate genes for knockouts or gene overexpression. Typically, CBMs perform well as far as the objective of modelling practice is not to quantify how the intracellular fluxes are evoked and regulated by the activities of enzymes and what their response to perturbations are. Also, in more complex metabolic networks such as mammalian cells that are simultaneously growing on multiple carbon substrates, the applicability of the various constraint-based methods are still rather limited. The established principles

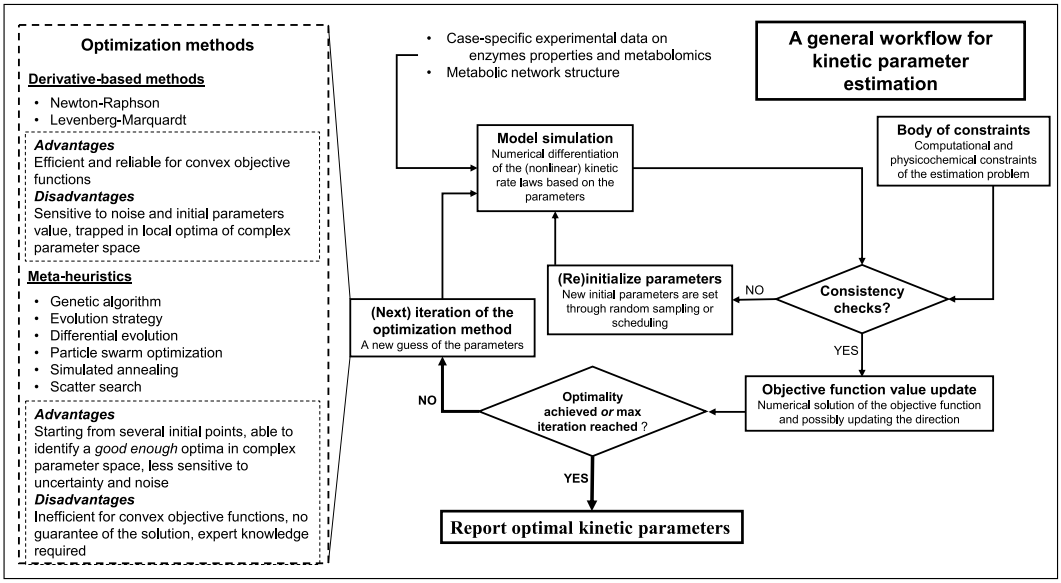

Figure 2.8: **A general workflow of kinetic parameter estimation.**

and mathematical justification of constraint-based modelling and its successful application for metabolic engineering purposes will support its continuous enhancement to define the hypothetical *best* the cell can do with decreasing uncertainty and thus identifying what the cell is not capable of doing.

In a conceptually different modelling paradigm, we assessed the established modelling approaches to model flux kinetics of the regulated enzymatic reactions based on the hypothesis that changes in enzyme activities during the cell operation is mechanistically describable in terms of chemical activity of substances which regulate or determine enzyme activities. We reviewed MCA and BST as two founding theories for the systemic kinetics analysis of the biochemical reaction networks at (or close to) steady state, where the network responses to perturbations can be quantified and the distribution of flux control between several regulatory sites (i.e. enzymes) is appreciated. In general, kinetic modelling adds accuracy to the prediction of perturbation outcomes and gives flexibility to the simulation of different scenarios, mainly because it is rooted in the inclusion of the detailed mechanistic description of the regulatory and compensatory mechanisms of the cell by use of a wide array of kinetic formats. In addition to the metabolic fluxes, dynamic models can reflect intracellular metabolites concentration not only at steady state but also in the transient time [25, 26, 156]. It consequently makes dynamic models a better candidate in comparison with their constraint-based counterparts for the implementation of monitoring and control strategies in bioprocess

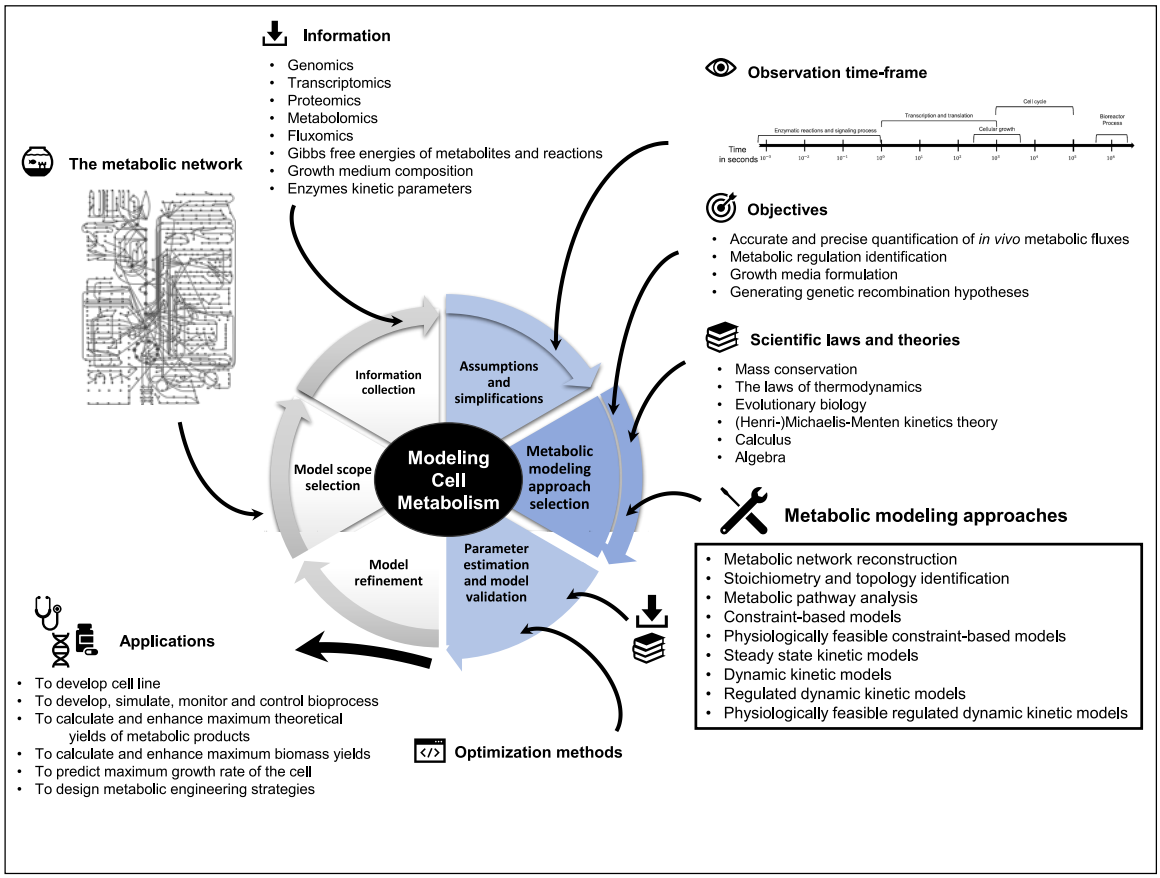

Figure 2.9: **An overview of modelling Cell Metabolism.**

management [32, 33, 92, 93, 155]. However, the hindering obstacles for all kinetic models to reach genome-scale are limited available data for intracellular concentrations and the complex procedure of selecting kinetic rate laws with identifiable mechanistic parameters. To circumvent, hybrid models have been proposed where for the part of network for which mechanistic data is available kinetic rate laws are used and the rest of the network is retained in its purely stoichiometric format. Observance of the emergent properties such as ultra-sensitivity (switch-like behaviour), bistability and oscillations which cannot be attributed to any single reaction or constituent of the network, but is only explainable with a systems understanding, further necessitates a dynamic nonlinear analysis of the underlying metabolic system. The importance of control and regulatory functional units inherent in a metabolic network can be hardly overestimated. Furthermore, in industrial applications where cells are used as small factories to produce invaluable therapeutics and commercial products, the shortcomings of steady state models are on display as they are incapable of reflecting cell response to the imposed control actions and fluctuating surrounding conditions. In fact, after comparing Michaelis-Menten with approximate rate laws in deriving a kinetic model for Red Blood Cells (RBCs), Du et al. [23] concluded that it is best to construct "mechanistically detailed enzyme modules" whenever the available data on enzyme properties allows that. It is reported that mixed models defined with Michaelis-Menten equations for one substrate-one product reactions combined with CK has the best prediction capacity [166]. It is justified based on the drawbacks of Michaelis-Menten formalism as a sole representation for the reactions in a network of interconnected reactions compared with the system-level definition of CK. Also, CK formulates complex regulatory mechanisms independently and thus reduces the complexity of regulated Michaelis-Menten equation on the way of interpreting the underlying biological principles. In addition, the convenience kinetics rate law is easy to use for parameter estimation and optimization and its models have comparatively accurate approximations and results [22].

For the parameter estimation of dynamic kinetic models incorporating mechanistic parameters, several *good enough* practices have been reported in literature, i.e. evolutionary meta-heuristics algorithms. Particularly, when the gradient-based optimization methods appear to get trapped in suboptimal extrema, meta-heuristics optimization approaches are employed to circumvent the inherent nonlinearity of metabolic networks and the expansion of parameters in an exceeding number of dimensions. We envision that the development of kinetic models that can encompass as many as possible different data types for its fit, particularly enhances the parameter estimation quality. With nowadays expansion of quantitative omics data and increasing biochemical thermodynamics data, we envision that the successful metabolic modelling approaches of the future will help integrate more insight from these in-

puts in a consistent computational framework. The power of mathematical modelling would be appreciated greatly, when it becomes more involved in the inception and design of new practical technologies. The fast-paced emerges of genome sequencing, poses the question that "what does the product of these genes do? How do various gene products interact to determine healthy or pathological states of the whole organism?" [1] (or high-producing and low-producing strains). Finding the answer, demands vigorous complementary analytical and computational technologies to be developed, in addition to the pile of information from (reductionist) biological sciences. This cannot be done without a significant contribution, from the mathematical modelling approaches, in order to understand and simulate the relationships among genotype, phenotype, and environment of the cell.

2.1.9 Author Contributions

Writing—original draft preparation, M.Y.; writing—review and editing, M.Y., M.J.; visualization, M.Y.; supervision, M.J.; funding acquisition, M.J. All authors have read and agreed to the published version of the manuscript.

2.1.10 Funding

M.J. received the NSERC (https://www.nserc-crsng.gc.ca/) Discovery Grant RGPIN-2019-05050. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

2.1.11 Acknowledgments

Authors wish to thank Bala Srinivasan for helpful discussions.

CHAPTER 3 ARTICLE 2 : A DYNAMIC CONSTRAINT-BASED MODELLING (DCBM) APPROACH WITH ALTERNATIVE METABOLIC OBJECTIVE FUNCTIONS PREDICTS THE IMPACT OF OXIDATIVE STRESS ON STORED RED BLOOD CELLS (RBCS)

Mohammadreza Yasemi[1], Michel Prudent[2,3,4], and Mario Jolicoeur[1,*]

[1] Research Laboratory in Applied Metabolic Engineering, Department of Chemical Engineering, École Polytechnique de Montréal, P.O. Box 6079, Centre-ville Station, Montréal, QC H3C 3A7, Canada.

[2] Laboratoire de Recherche sur les Produits Sanguins, Transfusion Interrégionale CRS, Epalinges, Switzerland.

[3] Centre de transfusion sanguine, Faculté de Biologie et Médecine, Université de Lausanne, Lausanne, Switzerland.

[4] Center for Research and Innovation in Clinical Pharmaceutical Sciences, Lausanne University Hospital and University of Lausanne, Lausanne, Switzerland.

[*]Corresponding author: mario.jolicoeur@polymtl.ca

(Published 27-29 July 2022 in IFAC-PapersOnLine, 55(20), 385-390.
doi:https://doi.org/10.1016/j.ifacol.2022.09.125)

3.1 Abstract

Mathematical metabolic modelling is a systematic endeavour to allow identifying the main causes of an observed metabolic change and to estimate the consequences of an imposed metabolic perturbation regarding a biosystem. Dynamic Constraint-based modelling (DCBM) has delivered promising results in metabolic engineering and in bioprocess design by providing mechanistically relevant systems-level knowledge of a network of bioreactions. Here, we seek to establish a DCBM approach that leverages convex optimization and nonlinear regression mathematical toolkit to estimate dynamic intracellular metabolic flux distributions in stored Red Blood Cells (RBCs) for transfusion purposes. First, we developed an *ad-hoc* metabolic

network including 77 reactions and 74 metabolites, second, we adapted Flux Variability Analysis (FVA) technique to quantify the connection between exometabolomic dynamics and the dynamics of feasible intracellular reaction flux ranges. We have obtained fine-grained flux range dynamics of the intracellular reactions for the benchmark data published in [4]. Then, we defined four objective functions regarding the accumulation of oxidative stress in stored RBCs for performing a dynamic Flux Balance Analysis (DFBA). In all four cases, time-resolved flux predictions were obtained respecting the imposed equality and inequality constraints. Last, we adapted a quadratic programming (QP) approach to calculate the Euclidean distance between the dynamic optimum flux vectors. The DCBM approach we have developed herein along with the developed metabolic network showed being suitable for the computational analysis of RBCs metabolic behaviour, and it is thought to be useful for other biosystems.

keyword

Mathematical Programming, Biological Systems Modelling, Metabolic Network Modelling, Constraint-based Modelling, Linear Programming Problem, Oxidative Stress, Red Blood Cells

3.2 Introduction

The integration of metabolomics data into mechanistic models can provide a practical understanding of a metabolic network at the system level ([18, 228, 229]). Guiding model construction and validation, the availability of experimental data allows elucidating mechanisms behind an observed metabolic change. Thus, mathematical modelling enables researchers to estimate consequences of an imposed metabolic perturbation in several research interests such as looking at metabolic energy states, bioproduction optimization, computational strain design and drug development ([230]). The current state of metabolic modelling approaches and computational systems biology methods have been reviewed recently ([7, 231]). In this work, we have selected quiescent non-growing stored Red Blood Cells (RBCs) as the biosystem under study, integrating metabolomics data to study the accumulation of storage lesions in stored RBCs ([232–236]). In transfusion medicine, the storage lesions may have detrimental health consequences for the patients transfused with old stored RBC concentrates ([237, 238]). Here, we developed an *in-silico* workflow for investigating the dynamics of the enzymatic and non-enzymatic oxidative stress defence mechanisms in RBC storage conditions shown in Fig. 3.1.

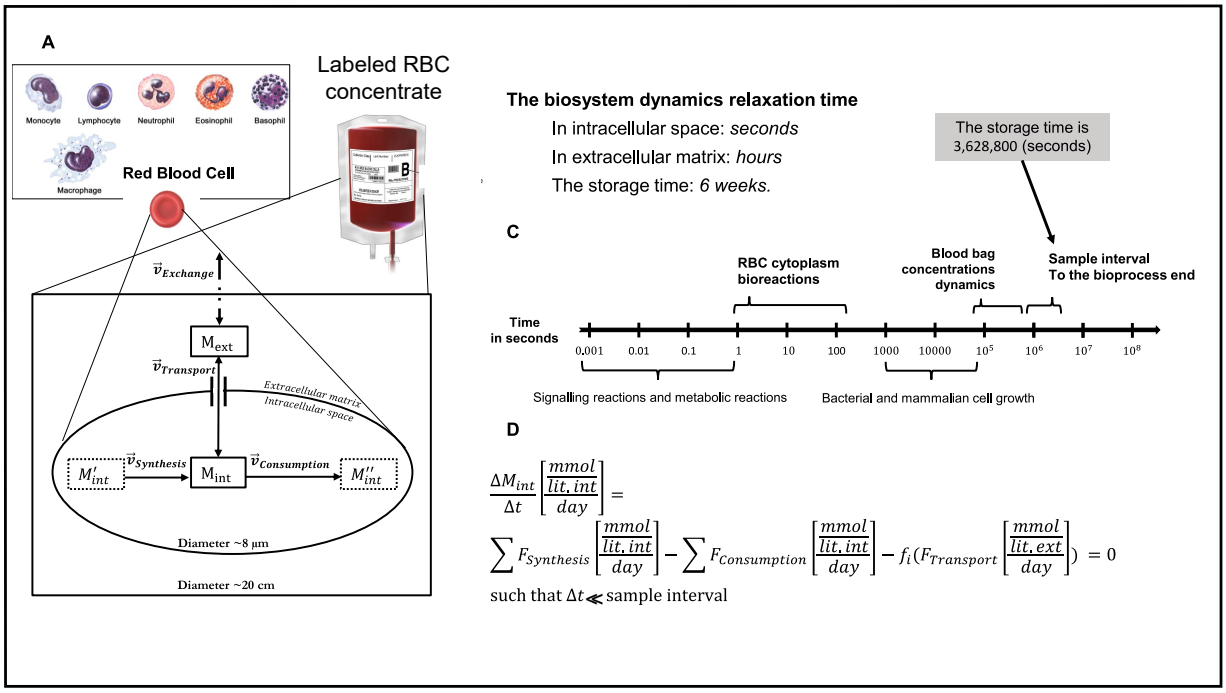

Figure 3.1: **The biosystem layout.** **(A)** The physical borders of the nested biosystems are presented and a global illustration of the different blood cell types. **(B)** The relaxation time of the nested biosystems dynamics. **(C)** A logarithmic scale of the relaxation time of the involved bioreactions. **(D)** This equation is based on the quasi-steady-state assumption. The third term on the right-hand side of the equation ($f_i(.)$) represents transport reactions impact on intracellular metabolites. Also, *lit.int* and *lit.ext* refer to the unit volume of cytoplasm and extracellular matrix, respectively.

In this study, we present a dynamic constraint-based modelling (CBM) workflow that integrates the extracellular time series data of stored RBC for an *ad-hoc* metabolic network of RBC. The goal is to predict the unknown intracellular fluxes, from a combination of measured transport fluxes, without any *a priori* assumptions regarding potential metabolic shifts in the metabolic network magnitude and direction during the RBC storage time.

3.3 Stoichiometric model reconstruction and partitioning

The stoichiometric model of the RBC metabolic network was reconstructed iteratively from the physiological knowledge on RBCs and the previously published genome-wide metabolic model for this biological system ([236]). The stoichiometric matrix was partitioned based on a distinction between the intra- and extracellular spaces as given in equation (3.2),

$$\frac{\Delta \vec{M}_{int}}{\Delta t} = S.\vec{v}$$

$$= \begin{bmatrix} S_{II} & S_{IT} \end{bmatrix} \begin{bmatrix} \vec{v_{Intra}} \\ \vec{v_{Transport}} \end{bmatrix} = \vec{b_m}, \tag{3.1}$$

$$\vec{lb} \leq \vec{v} \leq \vec{ub} \tag{3.2}$$

where the vector ΔM_{int} is a vector of dimension m of intracellular metabolites change, the vector v represents reaction fluxes with dimension n, which stands for the collective set of v_{Intra} and $v_{Transport}$. The stoichiometric matrix S of dimension $m \times n$ is divided into S_{II} representing the stoichiometric relation of intracellular metabolites to the intracellular reactions and S_{IT}, which represents the relationship between the metabolites passing the system boundary and transport reactions. The vector b_m is non-zero right-hand side of the equation, lb_n and ub_n are the lower and upper bounds of the flux vector, respectively. We adapted the pseudo steady state assumption if $\Delta t << $ sample intervals with t being a continuous time vector. Thus, equation (3.2) simplifies to the following equation.

$$S.\vec{v} = \begin{bmatrix} S_{II} & S_{IT} \end{bmatrix} \begin{bmatrix} \vec{v_{Intra}} \\ \vec{v_{Transport}} \end{bmatrix} = 0 \tag{3.3}$$

In the iterative procedure of the metabolic network reconstruction, we modified the metabolites by lumping and distributing, and we modified the reactions by adding (or removing) transport and sink reactions followed by explanatory simulations ([9,239]). The final version of our stoichiometric model describing the metabolic network was used throughout the next

sections (table 3.1).

3.4 Nonlinear fitting of the exometabolomics

To avoid model infeasibility due to the inaccuracies emerging from noisy experimental measurements, we run a nonlinear regression data-fit for each of the twelve measured extracellular metabolites. A set of independent exponential continuous functions is fitted to the extracellular metabolites concentration curves as described in equation (3.4).

$$M_{ext}^{sim}(p,t) = p_1 * \exp\left(p_2 * t\right) \tag{3.4}$$

where p is the vector of parameters to be estimated, and M_{ext}^{sim} represents the simulation values of extracellular metabolites as a function of the continuous time vector t. The parameter estimation formulation and the nonlinear objective function are given in equation (3.5).

$$\begin{aligned} \min_{\vec{p}\in R^2} f(\vec{p}) &:= \sum_{k=1}^{N} \left(M_{ext}(t_k) - M_{ext}^{sim}(p,t_k) \right)^2 \\ \text{Such that } t_1 = 0,\ t_N = 42,\ k &\in \{1,2,\ldots,N\} \\ N &= 14 \end{aligned} \tag{3.5}$$

$M_{ext}(t_k)$ represents the experimental value of an extracellular metabolite at time point t_k (in days), where k is the number of experimental data acquisition ending at day 42 after a total of $N = 14$ observations (see [4] for detailed explanation of the experimental setup). Then, the 95% confidence interval of the estimated parameters is determined by *nlparci* function in MATLAB using the Jacobian of the nonlinear fit function, and the instantaneous transport rates are calculated based on the analytical derivative of the parameterized continuous function in equation (3.4).

Table 3.1: The RBC metabolic model configuration

Model component	Total Number	# Intracellular	# Transport or Extracellular (measured)	# Sink	# Blocked or Dead end
Reactions	77	53	16 (12)	8	0
Metabolites	74	57	17 (12)	N/A	0

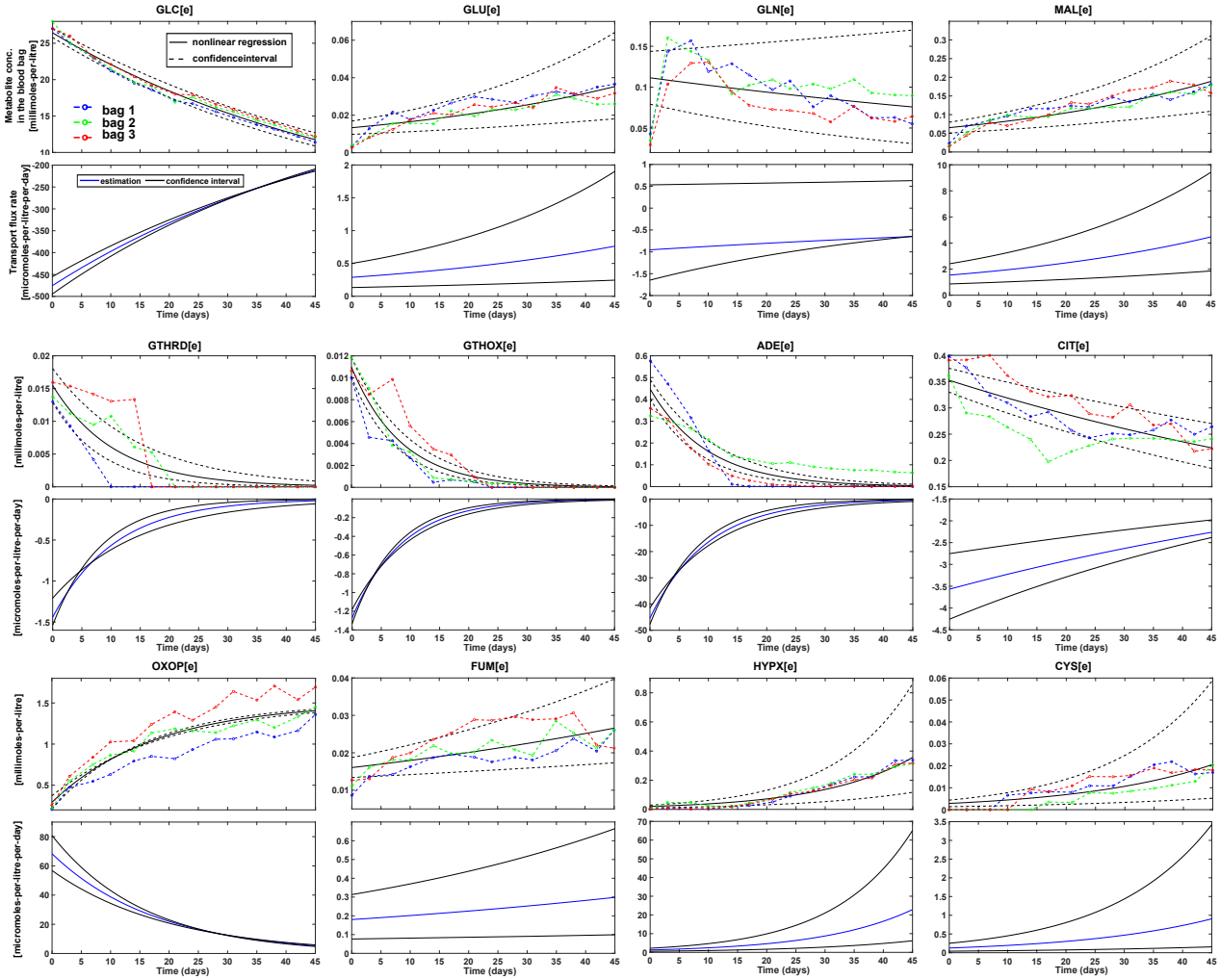

Figure 3.2: **Nonlinear regression fit of the blood bag ingredients concentration values and the associated transport flux rates** The exponential regression fit curves (solid black lines) with 95% confidence interval of the estimated parameters (dashed black lines). The lower and upper bounds of the twelve transport flux rates are assigned based on this panel. The bag numbers and experimental values (circles) for $N = 14$ experimental time points are taken from the second supplementary document of ([4]).

3.5 Defining simulation time

We define a separate time vector $tsim$ to use in the following sections, this is the time vector on which our constraint-based modelling method goes forward in time until ending at $tsim_L = t_N = 42\ (days)$.

$$\begin{gathered} t_1 \leq tsim_q \leq t_N \\ \text{Such that } q \in \{1, 2, \ldots, L\} \\ L = 45 \end{gathered} \tag{3.6}$$

where L represents the number of simulation points. In this study, we defined this number equal to 45 in a trade-off between model feasibility and the predefined computational cost limit. Moreover, the simulation time points are linearly distributed and each interval is roughly equal to one day in the experimental context.

3.6 Dynamic Flux Variability Analysis (DFVA) to identify the allowable intracellular flux ranges

We ran FVA at each of the simulation time points to identify the intracellular flux bounds based on the solution of $2n$ number of Linear Programming (LP) problems optimizing for min/max of each reaction flux. The advantage of calculating intracellular flux ranges by this method is that the estimated ranges are unbiased with respect to any assumed objective function for the cell functioning during the storage time. However, for the calculated ranges to be surrounded by default flux bounds, i.e., not to be redundant, iterative modification of the metabolic network structure was required. The ranges identified by DFVA were narrower than the default -1000 to 1000 $(mmol.L^{-1}.day^{-1})$ range for $\frac{51}{61} = 83\%$ of the intracellular fluxes.

$$\begin{array}{lll} \min / \max\limits_{\vec{v} \in R^n} f(\vec{v}) & := v_i & v_i \in \begin{bmatrix} \vec{v_{Intra}} \\ \vec{v_{Transport}} \end{bmatrix} \\ \text{Subject to} & S.\vec{v} = 0; & \\ lb_j \leq & v_j \leq ub_j & , v_j \in \vec{v_{Transport}} \end{array} \tag{3.7}$$

where lb_j and ub_j values were calculated in section 3.4. Moreover, to determine reaction directions, we tabulated all the reactions that could acquire a negative minimum flux value and a positive maximum one after the inspection of DFVA results. Then, we modified the list based on the biological knowledge for the removal of unrealistic reversibility predictions.

3.7 Dynamic Flux balance analysis (DFBA) based on several theoretical objectives

First, we constrained a sub-set of the transport fluxes $v_{\vec{Transport}}$ including twelve fluxes as shown in Fig. 3.2. The selected measured consumption/secretion rates constituted $12/16 = 75\%$ and $12/57 = 21\%$ of transport and intracellular fluxes, respectively (table 3.1). Then, we defined four different objective functions and studied the solutions (see Results and Discussion). The objective functions were set to the following $f_i(\vec{v})$'s (reactions are given in Fig. 3.4).

1. Set 1: Limited glucose ($-v1$).
2. Set 2: Maximal ROS tolerance ($v53$).
3. Set 3: Minimal ROS tolerance ($-v53$).
4. Set 4: Maximal Oxidative PPP (Ox-PPP) activity ($v14$).

We considered Set 1 the nominal scenario, Set 2 the ideal antioxidant scenario within the experimental constraints, Set 3 the worst case scenario, and Set 4 as a relevant metabolic scenario of interest. Thus, the LP problem in equation (3.8) was solved regarding each set at consecutive simulation time points.

$$\begin{array}{lll} \min\limits_{\vec{v}\in R^n} f_i(\vec{v}) & := c_i^T \vec{v} & i = 1,2,3,4 \\ \text{Subject to} & S.\vec{v} = 0; & \\ & lb_i \leq v_i \leq ub_i \quad , v_i \in v_{\vec{Intra}} & \\ & lb_j \leq v_j \leq ub_j \quad , v_j \in v_{\vec{Transport}} & \end{array} \tag{3.8}$$

where (lb_i,ub_i) and (lb_j,ub_j) values were calculated in sections 3.6 and 3.4, respectively.

3.8 Results and discussion

3.8.1 Development of an *ad-hoc* metabolic network

The metabolic network was reconstructed for explaining RBCs oxidative metabolism, the model scope involves cofactor-dependent enzymes participating in Reactive Oxygen Species (ROS) termination bioreactions, and glutathione metabolism as our central modelling objective. Moreover, the major pathways dominating intracellular metabolism of RBCs were accounted for as shown in Fig. 3.4. To validate the developed metabolic network, we optimized the nominal objective function (Set 1) at the N consecutive experimental time points

($N = 14$) reported in ([4]) for RBCs that were suspended in 100 mL of SAGM (Saline, Adenine, Glucose, Mannitol) additive solution. At first, the reconstruction needed several rounds of modifications to find dynamic feasible solutions. However, when the feasible solutions appeared for the first time, i.e., the metabolic network was validated, the extracted stoichiometric model supported DFBA simulations for $L = 45$ simulation time points regarding the four objectives with only minor modifications (Fig. 3.3).

3.8.2 Linking the transport flux rates to unbiased intracellular flux ranges

The developed extension of CBM for describing time-resolved dynamics of the RBC internal metabolic network starts off from the nonlinear exometabolomics fitting (section 3.4), which converts the discrete measurements into differentiable analytical functions. However, it is noteworthy that neither the exponential analytical functions nor the estimated independent parameters in equation (3.4) have any biological meaning. The aim of this step is to define a set of sufficiently accurate and smooth exchange flux constraints that also address the solution infeasibility issue, hindering the estimation of continuous intracellular flux predictions in dynamic constraint-based models. Then, the 95% confidence interval of transport fluxes were imposed as inequality constraints on $v_{Transport}$ vector in equation (3.2). The generated ranges of intracellular fluxes estimated by DFVA were treated as unbiased bounds with regard to any possibly assumed objective function for the LP problems solved in DFBA (section 3.7). Of importance, it is a methodologically distinct use of FVA technique in CBM approaches than what the authors followed in ([240]). In their work, the LP problem was first optimized, for example, to maximize the growth rate and then the FVA technique was used to determine the range of the possible alternate optima. The *ad-hoc* metabolic network developed here supports generation of dynamic unbiased intracellular flux ranges as described in section 3.6 on a personal computer with Intel(r) Core(TM) i5-8250U CPU @ 1.60 GHz and 8 GB RAM memory. However, a similar analysis on the genome-wide metabolic model of erythrocytes demands significantly higher computation power emphasizing the importance of metabolic modelling integration with the state-of-the-art computing techniques (see [5]).

3.8.3 Multiple alternative objective functions and analysis of the Euclidean distance between dynamic flux distribution optima

In section 3.7, we solved the DFBA problem for multiple alternative objective functions, but within the same equality and inequality constraints. Then, we used Quadratic Programming (QP) to calculate the Euclidean distance between the four sets of dynamic flux distribution

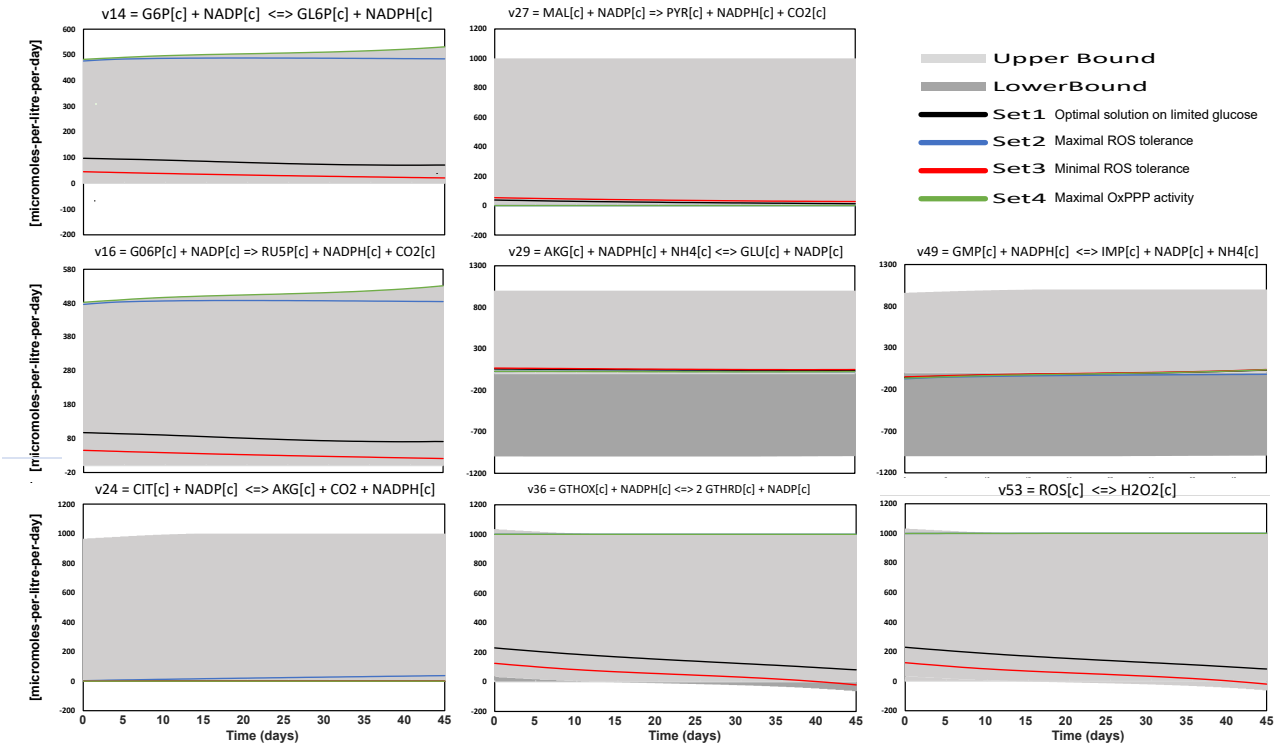

Figure 3.3: **The dynamics of ROS detoxification and NADPH-dependent reaction rates constrained by dynamic intracellular flux bounds** The colour-coded curves represent optimal flux dynamics associated with the solution set of the four objectives (section 4.3.4) and the grey-shaded background represents the dynamic intracellular flux ranges. Set 2 (maximal ROS tolerance in solid blue) and Set 4 (maximal PPP activity in solid green) overlap in v27, v29, v36, and v53.

optima as shown in equation (3.9),

$$dist(set_i, set_j) = \left(\sum_{k=1}^{L} \sum_{r=1}^{n} \left(v_{i,k,r} - v_{j,k,r} \right)^2 \right)^{\frac{1}{2}} \quad i, j = 1, 2, 3, 4 \tag{3.9}$$

In fact, we intend to show that systemic generation of multiple relevant biological objective functions followed by model-driven analysis of the generated flux distribution optima provides a robust computational approach to systems-level scrutiny of a metabolic network of interest. Thus, in table 3.2, we reported the symmetrical matrix of dynamic flux distributions Euclidean distance. The results revealed that Set 2 and Set 4 are the closest dynamic flux distribution optima.

3.8.4 Mitigating the solution space infeasibility issue

We know that the biosystem acquires a mass balanced carbon flux distribution at any point within the experimental scope, therefore, we expect the model to find feasible solutions within the same *in-silico* conditions. Thus, we argue that if the LP optimization solver cannot find a feasible solution in the enzymatic flux vector space R^n, then, there exist some limitations in the model (e.g., unfilled gaps, inaccurate mechanistic assumptions, inconsistent model scope, etc). We showed that the connectedness of the developed metabolic network and the smoothness of the simulated exometabolomics play important roles in tackling the infeasibility issue regularly associated with stoichiometric-based models ([241, 242]). In general, when finding a flux distribution solution is infeasible, an undesired CBM practice would be to include more sink reactions for a larger set of the intracellular metabolites, because such that the mass balance constraints on more metabolites are loosened and the optimization solver finds feasible solutions in an enzymatic flux vector space of $R^{n+\xi}$, with

Table 3.2: The Euclidean distance between the DFBA solution sets.

	Set 1	Set 2	Set 3	Set 4
Set 1	0	13737	1591.8	13874
Set 2	13737	0	15308	787.52
Set 3	1591.8	15308	0	15446
Set 4	13874	787.52	15446	0

ξ being the number of added sink reactions. However, the newly predicted intracellular flux distributions via this CBM practice will suffer from the inaccurate prediction of blocked intracellular reactions. Herein, we managed to generate feasible solutions by including sink reactions only for a small fraction of the intracellular metabolites, i.e., $8/57 = 14\%$, and thus avoiding artifact blocked reactions prediction as shown in the last column of table 3.1.

3.8.5 Using the validated model for generating systems-level biological hypotheses

The results reported in table 3.2 is in agreement with previous findings supporting the positive correlation between PPP activity and antioxidant defence mechanisms in other biological systems ([243, 244]). In fact, the oxidative PPP to hexokinase turnover ratio could increase from 106 % to 197 % in the PPP hyperactivity scenario (set 4). This range was from $\frac{97.77}{455.04} = 21\%$ at day zero to $\frac{71.24}{221.02} = 32\%$ at day 42 of the storage time in the glucose limited nominal scenario (set 1). The carbon flux through the oxidative branch of PPP with 2 moles NADPH turnover for each mole of glucose intake represents a metabolic route to generate reducing cofactors (NADPH), i.e., $v14$, $v15$, and $v16$. We show that the termination of higher fluxes of ROS modelled via $v53$ to preserve redox homeostasis is possible through a sustained activation of this pathway. We also found that hyper activation of the PPP pathway can pull $v2$ in its reverse direction and also suppress citrate to alpha-ketoglutarate and malate to pyruvate reactions, possibly because the latter ones are competing for NAD substrates. Such *in-silico* analysis results may suggest a rationale for increasing citrate concentrations in blood bags to compete for the storage lesions consumption through an alternative metabolic route. We also envisage that the same *in-silico* approach could be helpful to characterize the influence of urate on RBC metabolism during storage. Indeed, we showed that the RBC preparation triggers a progressive loss of urate during the first week of storage ([235]). The compensation of this leak by adding urate and ascorbic acid was suspected to reroute the metabolism (switch between oxPPP and glycolysis) ([234]). Collectively, these simulation results emphasize the inherent variability in dynamics of cell metabolism and the possible implications of this heterogeneity for the regulation of antioxidant defence machinery. Similar model biological systems particularly emerged as an illuminating case in systems analysis of energy metabolism ([244–246]).

In the maximal ROS tolerance scenario, the predicted reaction rates of $v24 : CIT[c] + NADP[c] <=> AKG[c] + CO2[c] + NADPH[c]$ and $v25 : ATP[c] + CIT[c] + COA[c] <=> ACCOA[c] + ADP[c] + OAA[c]$ suggest that citrate is consumed in the former to provide reducing power in form of $NADPH$ cofactor and it is produced in the latter allowing the

ATP regeneration. Of importance, dynamics of these fluxes (Fig. 3.3) in the alternative sets 2 and 4 predict opposite directions for $v25$. In Set 2, ATP and CIT are produced whereas in Set 4 $v25$ proceeds in the forward direction thus consuming citrate and energetic cofactor ATP.

3.9 Conclusion

We have set up a constraint-based model of RBC metabolic network for the mechanistic estimation of the biochemical reaction network fluxes under twelve narrow transport reaction constraints evolving with storage time. The dynamic CBM approach presented above incorporates several data and model processing steps as discussed in the manuscript. The model complemented with different hypotheses can predict flux rates of antioxidant defence systems. Indeed, we demonstrated that the model captures time-dependent switches in reversible intracellular reactions, and also, it predicts time-resolved activity patterns of enzymatic reaction rates under four distinct metabolic objectives (Fig. 3.3). Finally, we used the model to calculate the distance between the optimal dynamic flux distribution solutions. Two considerable shortcomings of the presented dynamic metabolic flux methodology are, first, not providing intracellular metabolite levels at this stage, and also, the lack of explicit integration of the metabolic regulation events. Addressing these limitations will be covered in a further modelling study.

3.10 Software

All the calculations and computations were done using MATLAB R2020b (The Mathworks; Natick, MA, USA) and glpk, gurobi, or CPLEX optimization solvers (academic licenses) (gnu.org/software/glpk/, gurobi.com, ibm.com/analytics/cplex-optimizer) with functions from COBRA Toolbox library, if needed ([16]). The metabolic map was drawn in CellDesigner ([247]), and PowerPoint, Excel and InkScape were used for the figures preparation.

3.11 Acknowledgement

The authors wish to acknowledge Natural Sciences and Engineering Research Council (NSERC) funding for this project (M.J.).

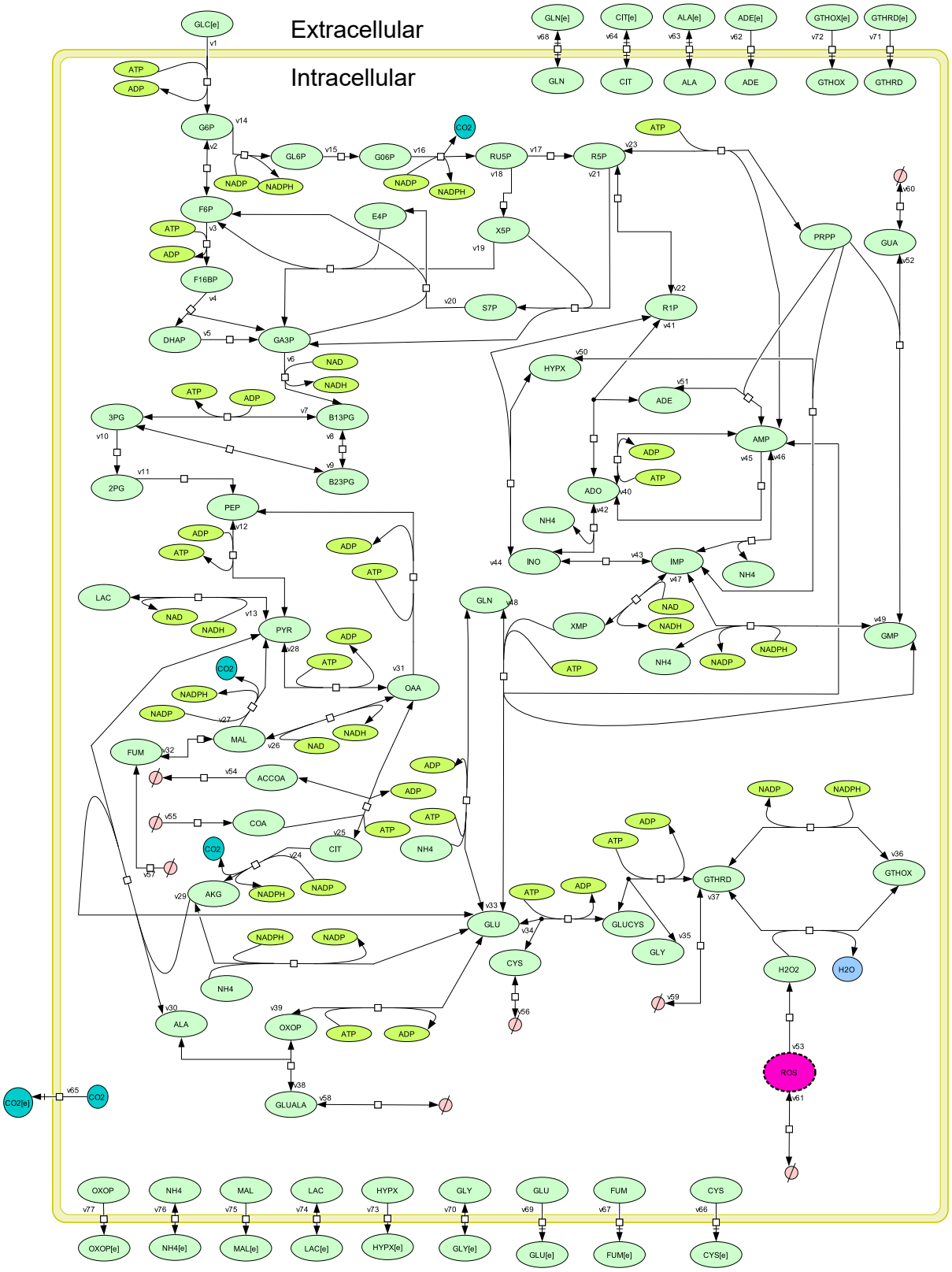

Figure 3.4: **The *ad-hoc* RBC metabolic map** The metabolic network can be divided into ten bioreaction subsystems including, glycolysis (v1-v13), pentose phosphate pathway (v14-v23), TCA cycle (v24-v32), glutamate and glutathione metabolism (v33-v39), purine catabolism (v40-v45), Nucleotides (v46-v49), Salvage pathway (v50-v52), ROS detoxification (v53), sink and demand reactions (v54-v61), and transport reactions (v62-v77).

CHAPTER 4 ARTICLE 3 : A GENOME-SCALE DYNAMIC CONSTRAINT-BASED MODELLING (GDCBM) FRAMEWORK PREDICTS GROWTH DYNAMICS, MEDIUM COMPOSITION AND INTRACELLULAR FLUX DISTRIBUTIONS IN CHO CLONAL VARIATIONS

Mohammadreza Yasemi[1,*], and Mario Jolicoeur[1]

[1] Research Laboratory in Applied Metabolic Engineering, Department of Chemical Engineering, École Polytechnique de Montréal, P.O. Box 6079, Centre-ville Station, Montréal, QC H3C 3A7, Canada.

[*]Corresponding author: mohammadreza.yasemi@polymtl.ca

(In revision with Metabolic Engineering, Submitted July 8, 2022)

4.1 Abstract

Optimizing mammalian cell growth and bioproduction is a tedious task. However, due to the inherent complexity of eukaryotic cells, heuristic experimental approaches such as, metabolic engineering and bioprocess design, are frequently integrated with mathematical models of cell culture to improve biological process efficiency and find paths for improvement. Constraint-based metabolic models have evolved over the last two decades to be used for dynamic modelling in addition to providing a linear description of steady-state metabolic systems. Formulation and implementation of the underlying optimization problems require special attention to the model's performance and feasibility, lack of defects in the definition of system components, and consideration of optimal alternate solutions, in addition to processing power limitations. Here, the time-resolved dynamics of a genome-scale metabolic network of Chinese hamster ovary (CHO) cell metabolism are shown using a genome-scale dynamic constraint-based modelling framework (gDCBM). The metabolic network was adapted from a reference model of CHO genome-scale metabolic model (GSMM), iCHO_DG44_v1, and dynamic restrictions were imposed to its exchange fluxes based on experimental results. We used this framework for predicting physiological changes in CHO clonal variants [248]. Because of the methodical creation of the components for the flux balance analysis optimization problem and the integration of a switch time, this model can generate sequential predictions

of intracellular fluxes during growth and non-growth phases (per hour of culture time) and transparently reveal the shortcomings in such practice. As a result of the differences exploited by various clones, we can understand the relevance of changes in intracellular flux distribution and exometabolomics. The integration of various omics data into the given gDCBM framework, as well as the reductionist analysis of the model, can further help bioprocess optimization.

4.2 Introduction

Mammalian cells have the ability to produce *in vitro* complex biologics of high commercial interest. Successful genetic alterations, growth media optimization, and an effective understanding and management of cellular regulatory mechanisms are expected to yield the next generation of cell factories producing biotherapeutics. Because of their stable growth, high cell density peak, high innate protein expression level, and functional post-translatory regulatory mechanisms, Chinese hamster ovary (CHO) cell lines have been widely used in industry for biotherapeutics manufacturing [249]. Recombinant CHO cell lines, the gold standard in biotherapeutics production, are responsible for stable and transient protein expressions in bioreactors of various capacities, up to fully monitored production vessels with a working volume of over 20,000 L [250]. Biotechnological strategies, on the other hand, necessitate extensive quantitative information on cellular metabolism, i.e., establishing a descriptive link between culture conditions and cell productivity. As a result, approaches that provide insight into the flow of carbon, energy, and electrons are extremely desirable because they can quantify metabolic alterations caused by intentional or unintentional perturbations in genetic or environmental factors [1, 8]. Hence, metabolic flux analysis (MFA) is a powerful tool for discovering new bioprocess design solutions that improve the target metabolic product.

Mathematical metabolic models, in combination with high-performance computing and state-of-the-art biomolecules measurement capacity, provides a systems biology approach to describe cell factories, complementing wet-lab methods to improve bioprocess efficiency [9, 10]. On the one hand there is Dynamic Kinetic Modelling (DKM) [155, 184, 251], which requires cell culture experimental data for model kinetic parameters identification, and on the other hand, there is Constraint-based Modelling (CBM) [15, 102, 252] (reviewed in [253]), which is based on the stoichiometry of metabolic networks. Metabolic Flux Analysis (MFA) [254] and Flux Balance Analysis (FBA) [94] are two variants of constraint-based models, each representing determined (or overdetermined) and under-determined metabolic networks, respectively. Without the use of kinetic parameters, these methods can provide credible cell

culture models behaviour. The theories that support the validity of FBA provide a fertile ground for the development of more complex approaches for exploring the behaviour of metabolic networks. Flux Variability Analysis (FVA) is one such methodology, in which minimum (and maximum) achievable enzyme activity of each reaction is defined as the objective function, and the flux rates bounds are established by performing linear programming (LP) optimization within the initial problem constraints [240]. A quadratic minimization of the estimated fluxes is integrated as a bi-level optimization problem in *parsimonious FBA (pFBA)*, on top of the primary objective of FBA, i.e., maximum growth rate [27]. In order to switch from descriptive to predictive models, various crucial innovations have come out in the past years. Apart from the more traditional options, such as optimising growth, researchers have looked into various objective functions for cell functionalities. Indeed, the premise of maximal growth loses its usefulness, especially during the non-growing period or slow growth fed-batch of mammalian cell culture, and other goals such as related to physiological factors and complex proteins processing steps are mostly considered.

The use of CBM extensions to simulate growth and by-products dynamics in non-steady state phases proved being particularly compelling [15, 102]. Dynamic Constraint-based Modelling (DCBM) technique is the name given to algorithms with predictive power algorithms developed in this area. In DCBM, the well-known quasi-steady state (QSS) assumption is employed in conjunction with experimental limitations on substrates, oxygen, and energy requirements of growth and production to account for faster intracellular dynamics relative to extracellular matrix dynamics. In this category, there are two ways to follow: sequential and simultaneous methods. The sequential technique, which is used in this study, divides the entire process time into small steps, and the FBA optimization algorithm assumes QSS at each time step, but not at the transition point. The solution to the optimization problem generates new growth and uptake rates, which are then employed in a system of ordinary differential equations (ODEs), whose integration yields in component concentration change over time. If numerous measurement time points are averaged as a single metabolic phase, this method loses information about the real-time evolution of the fluxes. On the other hand, if the time steps are too small, noise in the measured concentrations leads to erroneous uptake rate estimations, and the computing cost becomes a bottleneck. The answer to the optimization problem is obtained in an iterative *static* optimization as long as the unknown fluxes and the objective function are linearly connected. The states dynamics are achieved in this case from the subsequent integral solutions.

Comprehensive experimental research have been conducted to find active metabolic pathways in hybridoma [99, 104, 255–257] and then in CHO cell lines [258–260], for developing pre-genomic structured metabolic networks. Subsequent investigations of CHO [62, 261] en-

larged these bioreaction networks from a set of less than fifty biological reactions to include a more thorough metabolism of amino acids and nucleotides, including roughly 100 reactions. After the publication of the first genome-wide metabolic model of CHO in 2011 [262], and a consensus community-based CHO genome-scale model in 2016 [28], new opportunities for metabolic modelling of CHO cell cultures have emerged. Some recent studies on CHO metabolism [27, 227, 263] addressed revising the consensus GSMM for constructing context-specific metabolic models. In fact, multiple approaches for extracting consistent metabolic models from the reduction of GSMMs have been derived, each based on distinct assumptions and screening criteria for including the content of a reference GSMM [264–266]. These metabolic model extraction methods provide a crucial computational toolkit for extracting context-specific metabolic models from an organism's most comprehensive metabolic model. Other omics data, such as exometabolomics or transcriptomics data, may be used in computational algorithms for designing GSMMs [267, 268]. Other frequent requirements include establishing a list of preserved metabolites and reactions, as well as requiring the reduced model to perform specific metabolic tasks [269]. Furthermore, the complexity of eukaryotic cells necessitates considerations of compartmentalization, metabolic process energetic costs, and biomass composition.

Here, we revised a recent GSMM of CHO [27], as deposited on http://bigg.ucsd.edu/models/iCHOv1_DG44 (mat file), as the metabolic network for developing a gDCBM model of a bioprocess. The underlying metabolic network was built and then the corresponding stoichiometric matrix was determined to be used in the main optimization problem that is subsequently formulated. A smoothing spline fit was performed on exometabolomics to convert the raw experimental data from three strains of CHO to dynamic exchange flux bounds. The data corresponds to parental, low-producing and high-producing CHO strains that have been described in a previous study [248]. The gDCBM framework is consequently used for modelling and the results are discussed.

4.3 Methods

4.3.1 Metabolic network development

We selected the model contributed in [27] because the authors updated the consensus GSMM [28] by considering the gene expression data in CHO_DG44 by applying Gene Inactivity Moderated by Metabolism and Expression (GIMME) algorithm [267]. In a preliminary analysis using swiftcore algorithm [270], we realized that 470 metabolites out of 2750 are dead-ends and 1810 out of 3942 reactions are inconsistent, i.e., unable to carry metabolic flux under any

conditions. Thus, the network was reduced by removal of such nonfunctional components. The configurations of our stoichiometric model used throughout the next sections is given in Table 4.1 and the complete list of the reactions is given in the supplementary material file (Reactions List).

Table 4.1: Comparison of the original and curated genome-scale metabolic models

	Initial model	Curated model
Number of reactions	3942	2123
Number of blocked reactions	1810	0
Number of balanced reactions	3211	1931
Number of metabolites	2751	1285
Number of dead-end metabolites	470	0
Number of extracellular metabolites	580	226

4.3.2 Phenomenological reactions

Here, the phenomenological reactions are particularly referring to the net biochemical transformations represented as a single (pseudo-) reaction. The scope of the metabolic network requires assuming the following (pseudo-) reactions.

Biomass synthesis

The biomass synthesis reaction directly affects predictions for *in silico* cell growth and maintenance. The weight fractions for the cell building blocks, i.e., lipids, proteins, DNA, RNA, and carbohydrates, with respect to the cell dry mass, molar fractions of the macromolecular building blocks, macromolecules average molar weights and CHO_DG44 dry mass weight were all taken from [263], ID/DGpar-8mMCD. In [263] flux units are different than this study ($mmol\ gDW^{-1}\ h^{-1}$ vs. $nmol\ 10^{-6} cell\ h^{-1}$), therefore, we converted the stoichiometric coefficients accordingly. Then the cell components synthesis reactions were lumped to form the biomass synthesis reaction. Finally, the original BIOMASS_cho_producing_1 reaction in iCHOv1_DG_44 was replaced by the new reaction. The calculation details are given in the supplementary material files (Biomass_eqn_1 and Biomass_eqn_2). We named this reaction MY_BIOMASS_cho_producing_1 for distinction.

Antibody production

In the experimental study, Ghorbaniaghdam et al. [248] proposed a model mAb against CD20 protein, we assumed this mAb has the same amino acid sequence as Rituximab (*Mabthera*) [271] the most similar commercial human mAb. Then, the amino acid sequence for light- and heavy- chains and an average molecular mass of 143.86 ($g/mmol$) were retrieved from [272]. The antibody production reaction was developed based on the procedure used in [273]. The polymerisation energy demand was coupled to the reaction by inclusion of 2 GTP and 1.306 ATP hydrolysis to 2 GDP, 1 AMP and 0.306 ADP per peptide bound formed [274]. Thus, the balanced reactions specific to Rituximab heavy-chain and light-chain replaced reactions *igg_hc_1* and *igg_lc_1* in iCHO_v1_DG44, respectively, and the *igg_formation* reaction remained intact. The detailed calculations are given in the supplementary material file (mAb_eqn).

Energetic requirements

We considered the growth-associated and non-growth associated energy consumption for the cell maintenance (mATP). The growth-associated maintenance ATP consumption was included in the biomass synthesis reaction after lumping the macromolecular biosynthesis reactions. As such, -1360 ($nmol$) ATP and -2000 ($nmol$) GTP is hydrolyzed for synthesis of 10^6 cells. To account for the non-growth associated maintenance ATP, for example due to membrane leak or protein turnover costs, we used the estimated values in a recent study on the impacts of mATP inclusion on the intracellular flux predictions [273]. Particularly, we used the estimated values for [92, 93], which have similar conditions with the Ghorbaniaghdam et al. [248]. Thus, we imposed a lower bound of 718.2 ($nmol\ 10^{-6} cell\ h^{-1}$) (equal to 3.6 ($mmol\ gDW^{-1}\ h^{-1}$)) on the mATP reaction.

Oxidative stress termination

Reactive oxygen species (ROS) are produced when there is (even partly) aerobic metabolism involved. We used the lumped reaction 2.0 gthrd_c + 3.0 h2o2_c $\longrightarrow$ gthox_c + 4.0 h2o_c + o2_c [27] to represent H2O2 termination.

4.3.3 Defining metabolic network stoichiometric matrix

The tailored CHO metabolic network transforms into a stoichiometric matrix that maps intracellular and transport reactions (between the cytosol and the external environment) to intracellular and extracellular metabolites dynamic of change, i.e. mass balances. As a

result, the stoichiometric matrix (and metabolic fluxes vector) were partitioned as shown in the equation below.

$$\frac{\Delta \vec{Z}_{bal}}{\Delta t} = S_{bal}\vec{v} = \begin{bmatrix} S_{II} & S_{IT} \end{bmatrix} \begin{bmatrix} \vec{v_{Int}} \\ \vec{v_{Tr}} \end{bmatrix} = \vec{b_m} \tag{4.1}$$

$$\frac{\Delta \vec{Z}_{nbal}}{\Delta t} = S_{nbal}\vec{v}X = \begin{bmatrix} S_{EI} & S_{ET} \end{bmatrix} \begin{bmatrix} \vec{v_{Int}} \\ \vec{v_{Tr}} \end{bmatrix} X = \vec{r_l}X, \tag{4.2}$$

where $Z_{bal} \in R^m$ and $Z_{nbal} \in R^l$ are the concentrations of intracellular and extracellular metabolites respectively in $(nmol/10^6 cell)$ and $(\mu mol/litre)$. The vector v represents specific fluxes in $(nmol/10^6 cell/hr)$ with dimension n. The stoichiometric matrix S of dimension $(m+l) \times n$ is divided into S_{II}, S_{IT}, $S_{EI} = 0$ and S_{ET}. The row decomposition is used in the calculations to distinguish between the metabolites for which the steady state assumption is imposed (S_{bal}), or not (S_{nbal}). Thus, the vectors b_m and r_l present the associated accumulation rates. Based on the QSS, vector b_m equals zero [68].

4.3.4 Formulating constrained optimization problem of gDCBM

We defined the dynamic optimization problem in the form of a differential-algebraic equation (DAE). Thus, the main problem is formulated in Equations (4.3-4.12).

$$\max_{v(\vec{t}),Z(\vec{t}),X(t)} \quad \alpha J_1 + \beta J_2 \tag{4.3}$$

Subject to

$$\frac{\Delta \vec{Z}_{bal}}{\Delta t} = 0 \tag{4.4}$$

$$r_l^{\vec{lb}} \leq \frac{\Delta \vec{Z}_{nbal}}{\Delta t} \leq r_l^{\vec{ub}} \tag{4.5}$$

$$\frac{\Delta X}{\Delta t} = \mu . X(t) \tag{4.6}$$

$$\mu = MY\_CHO\_Biomass \tag{4.7}$$

$$t_j = t_0 + j\frac{t_f - t_0}{N_{ts}}, \qquad j = 0, .., N_{ts}, \qquad \frac{t_f - t_0}{N_{ts}} = 1(hr) \tag{4.8}$$

$$\alpha = \begin{cases} 1, & \text{if} \quad t \leq t_{switch} \\ 0, & \text{otherwise.} \end{cases}, \qquad \beta = 1 \tag{4.9}$$

$$Z \geq 0 \quad X \geq 0 \qquad \forall t \in [t_0, t_f] \tag{4.10}$$

$$Z(t_0) = Z_0 \quad X(t_0) = X_0 \tag{4.11}$$

$$lb \leq v \leq ub, \qquad \forall t \in [t_0, t_f] \tag{4.12}$$

Where J_1 and J_2 in Equation (4.3) are the objective functions, α and β are the corresponding weightings. Equation (4.4) represent the QSS assumption, and r_l^{lb} and r_l^{ub} denote the bounds on the measured cell specific uptake rates (explained in the next section). μ is the growth rate equal to the rate of MY_CHO_Biomass reaction. t_j represents the descretized time points, moreover, N_{ts} is chosen as such the time interval between simulation points is equal to one hour. In Equation (4.9), t_{switch} was determined to minimize the fit error measurement. The non-negativity constraint is imposed on the metabolite and cell concentrations. in Equation (4.11), the initial concentrations are assigned based on the experimental data. Finally, each flux value v is bounded by a lower bound (lb) and an upper bound (ub) [239].

The objective functions for the above optimization problem is detailed in the Equations 4.13 and 4.14.

$$J_1(\vec{v}, \vec{Z_{nbal}}, X) = \mu(t)|_{t<t_{switch}}, \qquad \text{growth dependent,} \tag{4.13}$$

$$J_2(\vec{v}, \vec{Z_{nbal}}, X) = \sum_{j=1}^{N_{sf}} -v_j, \qquad \text{growth independent} \tag{4.14}$$

J_1 is the assumption of the growth rate maximization at any given time, which is valid before the switch to the non-growth phase. J_2 is the uptake rate objective functions which were defined for a subset N_{sf} of amino acids. We hypothesize that since it is a nutrient-rich environment, the goal of the cell is to maximize utilization of resources, and not to minimize nutrient consumption [275]. We included EX_gln___L_e, EX_val___L_e, EX_ala___L_e, EX_tyr___L_e, EX_phe___L_e, EX_met___L_e, EX_leu___L_e, EX_his___L_e, EX_arg___L_e, EX_cys___L_e, which are the exchange reactions for the associated metabolites. We chose the metabolites that would decrease the prediction error measurement, other metabolites would either worsen the fit or do not change it. Therefore, the optimization has multiple objectives when it comes to the non-growth phase.

4.3.5 Metabolomics data integration as exchange flux bounds

The data for this study was collected from six out of ten batch studies, including two parental, two low-producing, and two high-producing batch cell cultures, and eliminating two induced low-producing, two induced high-producing batch cell cultures [248]. We excluded the induced data sets because, first, the analysis in the original paper showed that induction does not cause significantly different behaviour (in the defined experimental conditions), and second, the pseudo-steady state assumption is likely to fail where a dramatic transient is expected, such as in transcription factor induction [254], resulting in gross model error. Measurements for glucose, lactate, ammonium (3), amino acids (14), CD-20 (1), oxygen uptake rate (1), and cell concentration (1) are included in the three data sets (in duplicates) of non-induced cultures. The data spans growth and non-growth stages, from seed culture inoculation until the end of the sixth day (144 hours), when viability decreases below 70%. To determine exchange fluxes for constraining the model, the following method was adapted.

Spline fitting of the exometabolomics and framing uncertainty in the concentration measurements

Because metabolomics data is often noisy and sparse, two challenges arise when such data is differentiated to compute exchange fluxes. First, the computed results are non-smooth,

and second, the calculated exchange fluxes are erroneous because they do not account for the inherent uncertainty of experimental measurements. Thus, imposing exact values for individual uptake fluxes is prone to inaccuracy and considerably increases infeasibility issues. We used zero mean standard noise with a standard deviation twice the experimental value to populate the experimental data, and then piece-wise cubic spline functions to fit the measured concentrations. The numerical derivatives of the evaluated fit functions were then calculated to produce a range of specific uptake rate values (r_l^{lb} and r_l^{ub}) as imposed by Equation 4.5. The exchange fluxes were constrained at each time point using the maximum fitting uptake or secretion rates.

(4.15).

$$PP_i = f(t, Z_{nbal,i}, p) \tag{4.15}$$

$$r_i = \frac{dPP_i}{dt} \tag{4.16}$$

where PP_i is the piece-wise polynomial fit on the $Z_{nbal,i}$ at time t, and p is the estimated spline parameters. $\frac{dPP_i}{dt}$ is the numeral derivative representing specific uptake rates r_i.

4.3.6 Sequential dynamic optimization

The optimization problem in Equation (4.3-4.12) contains both algebraic and differential equation models that makes its simultaneous solution impossible for LP solvers. It can be solved either as a nonlinear programming (NLP) problem and by methods such as orthogonal collocations on finite elements, or as LP problems solved and integrated sequentially [15,276]. The sequential solution of the problem is used here, the procedure involves discretizing a scalar dimension (time), into a set of intervals, i.e., $t_{j+1} - t_j$ that suffice error minimization, and converting the differential equations to a set of approximating algebraic equations, which are then integrated to yield the states trajectories.

4.3.7 Identifying the range of alternate optima for intracellular flux predictions

We did FVA at each of the simulation time points and after the main optimization problem was solved to identify the intracellular flux ranges based on the solution of $2n$ number of LP problems (n being the number of reactions) optimizing for min/max of each reaction flux [240]. The default information on reversibility/irreversibility was based on the iCHOv1_DG44_v1 flux bounds.

$$\begin{aligned} &\min_{\vec{v}\in R^n} / \max f(\vec{v}) && := v_i && v_i \in \begin{bmatrix} \vec{v_{Intra}} \\ \vec{v_{Transport}} \end{bmatrix} \\ &\text{Subject to} && \text{Equation}(4.3-4.12); \end{aligned} \tag{4.17}$$

4.3.8 Prediction error measurement

To measure the prediction error, we used the following normalized root mean squared error (NRMSE) formula [277],

$$NRMSE = \sqrt{\frac{1}{N_D}\sum_{k=1}^{N_e}\sum_{j=1}^{N_{y,k}}\frac{\sum_{i=1}^{N_{t,k,j}}(y_{ijk}-\tilde{y}_{ijk})^2}{(\max_i \tilde{y}_{ijk}-\min_i \tilde{y}_{ijk})^2}} \tag{4.18}$$

Here N_D is the total number of data points for 18 observables, N_e is the number of experiments, $N_{y,k}$ is the number of observables in the k-th experiment, $N_{t,k,j}$ is the number of time points in the k-th experiments for the j-th observable. y_{ijk} is the model prediction for the data $\tilde{y}_{ijk}$. This formula computes the root of the sum of squared error between model prediction and data for each observable, and normalizes it by the squared range of the data corresponding to that observable. Thus, the observables are properly scaled.

4.4 Results and discussions

4.4.1 Development of a genome-scale dynamic constraint-based modelling (gD-CBM) framework

DCBM techniques were used to provide a modular framework for evaluating CHO dynamic metabolism in batch cell cultures with CHO clonal variants in this study. As shown in Figure 4.8, the framework starts with a set of conventional cell culture data and a reference GSMM. The output includes time-resolved dynamics of medium components (mostly amino acids) and intracellular flux distributions with the alternate optima range of the fluxes. We demonstrated that this approach may be used to predict cell culture dynamics without the necessity for biosystem kinetic data.

We upgraded the CHO GSMM by removing the blocked reactions and changing the phenomenological reactions to fit the demands of the investigation. The traditional application of the molar mass conservation rule is challenging and expensive due to the unprecedented complexity and uncertainty of biomass composition. Based on a recent detailed assessment of CHO biomass composition in various strains as explained in the Methods section (4.3.2),

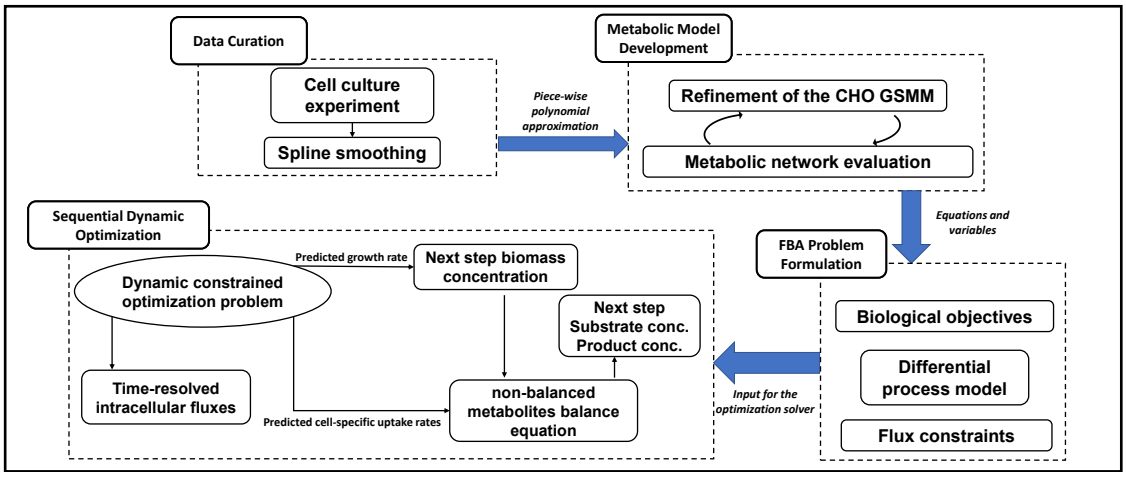

Figure 4.1: **The gDCBM framework** Four interactive components make up the framework. Smooth splines are used to approximate the data in the first module. The metabolic model is developed from the reference GSMM in the second module, while dynamic limitations imposed by the piece-wise polynomials formed in the data curation module are taken into account. The body of the FBA problem is formulated in the third module, and we employ optimization solvers to derive the final answers in the fourth module.

we elected to create a standard version of the model with a biomass synthesis reaction. To investigate nutrient metabolism in CHO cells, we looked at three different types of growth medium components. For each category, the gDCBM predictions of extracellular metabolites are presented alongside the experimental results in the accompanying figures.

1. The main metabolites: glucose (GLC), lactate (LAC), glutamine (GLN) and ammonia (NH4).

2. The nonessential amino acids: alanine (ALA), arginine (ARG), cysteine (CYS), glutamate (GLU), glycine (GLY), serine (SER), tyrosine (TYR).

3. The essential amino acids: histidine (HIS), isoleucine (ILE), leucine (LEU), methionine (MET), phenylalanine (PHE), valine (VAL).

4.4.2 The gDCBM framework accommodates *a posteriori* determination of the switch time

The switch time linked with each clone was calculated by looking for the time point with the least NRMSE. For the different clones, the exercise resulted in three different switch time

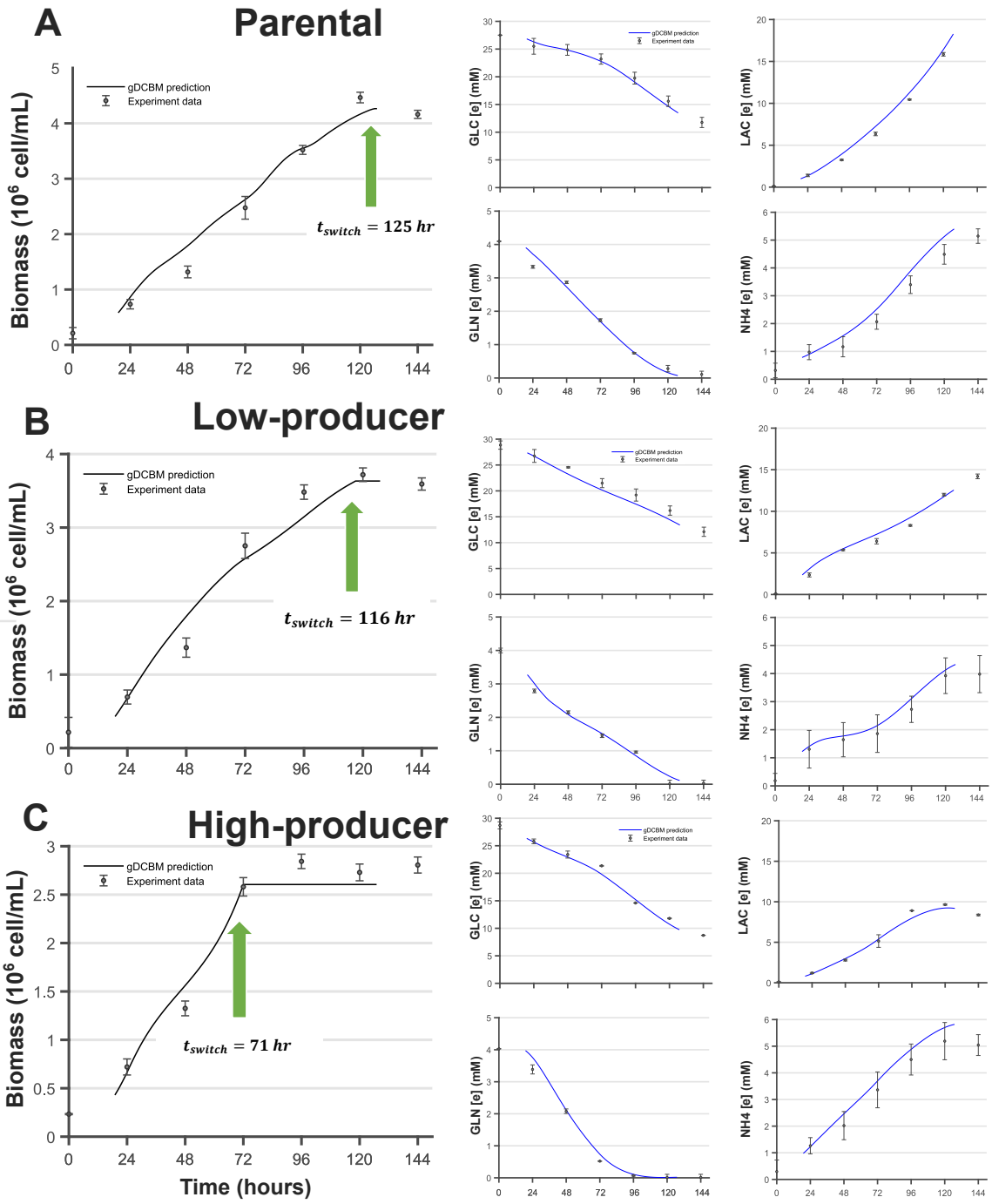

Figure 4.2: **Cell growth and concentrations of the extracellular metabolites: Main metabolites** The growth dynamics is on the left and the concentrations dynamics are on the right for GLC, GLN, LAC and NH4. The green arrow indicates the switch time from the growth to non-growth phase.

points. The parental clone demonstrated a late switch, followed by a fall in cell viability, as is common in mammalian cell culture, whereas the high-producer clone converted to non-growth phase as early as 71 hours after inoculation. In every example, the gDCBM predictions are very close to the experimental results. Indeed, we demonstrated that the framework allows for *a posteriori* determination of the transition from growth to non-growth phase based on the reduction of the NRMSE, which accounts for the prediction versus experimental data mismatch among all observable states.

4.4.3 The assumption of varying protein content of the cell

Because the initial cell composition was assessed in a producing cell line, the derived biomass synthesis reaction predicted the high-producing clone with the least error compared to the other two clones. We discovered that the biomass synthesis reaction results in a considerable underestimating of the growth rate for low-producer and parental clones in our preliminary simulations. We hypothesised that the disagreement stemmed from differences in clone amino acid compositions, thus we calculated a multiplier for the amino acid coefficients in the cell biosynthesis reaction. Thus, the reaction coefficients of the 18 amino acids involved in biomass synthesis (MET, ASN, CYS, GLN, SER, THR, ARG, GLY, PHE, GLU, ASP, VAL, TRP, TYR, HIS, LEU, ILE, and LYS) were multiplied by a factor of 0.7 and 0.9, resulting in the protein content for parental and low-producer, respectively. The modification is based on the notion that the cell makeup of the clones varies, which is supported by literature [278]. This adjustment is required to accommodate changes in cell lines, and it only needs to be done once; the coefficients will remain fixed after that. Another consequence of the modification is that, in the absence of kinetic parameters, amino acid biomass synthesis coefficients emerged as a significant source of sensitivity.

4.4.4 The gDCBM approach with uptake rates as objective functions predicts medium composition dynamics

The growth optimization assumption is based on the output of cellular activity, and while it leads to successful *in silico* results (especially for organisms simpler than mammals), it loses relevance in the non-growth phase and must be replaced by other objective functions to meet the necessary conditions of an optimization problem formulation. Moreover, further proof that a cell can design optimal behaviour for more than one metabolic objective [279] begs the question of what those objectives could be and how important they are in determining cell behaviour. In contrast to the limited nutrient consumption assumption [275], we assume that in a rich cell culture media, CHO cells attempt to ingest as many nutrients as their internal

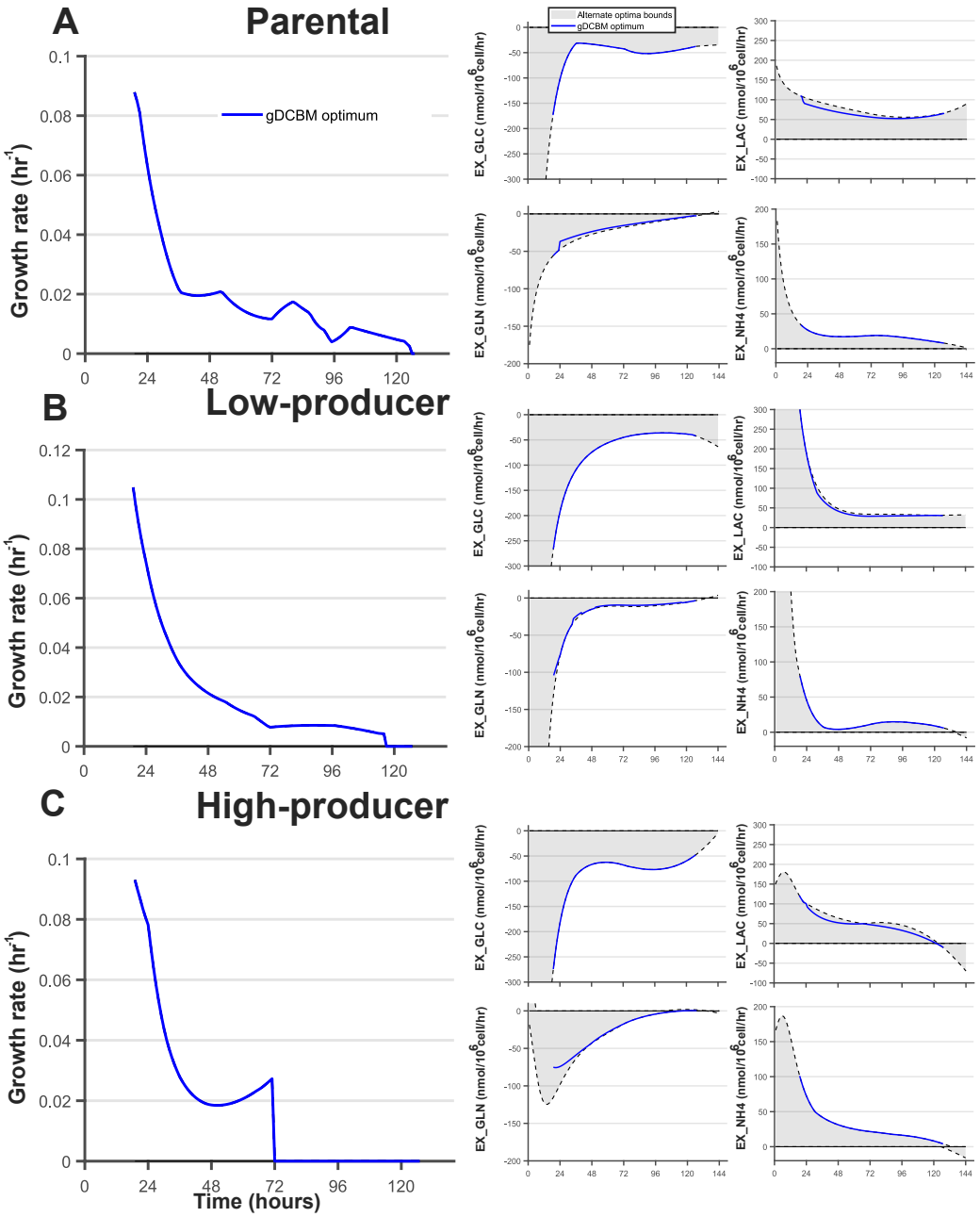

Figure 4.3: **Growth rate and specific uptake or secretion rates: Main metabolites** The instantaneous growth rates are on the left and the specific rates are on the right for EX_GLC, EX_GLN, EX_LAC and EX_NH4 for the parental **(A)**, low-producer **(B)**, and high-producer **(C)** clones. The shaded area shows allowed bounds for the specific rates and the blue line represents the gDCBM predictions.

metabolism allows and secrete as much product as their internal metabolism determines. As a result, we focus on the cell's input and presume that intracellular activity is the limiting factor

in the exchange of medium constituents across the cell membrane. These substances' input flows can be used as objectives in the primary FBA optimization problem. We were able to acquire the simulations in Figures 4.2,4.4 and 4.6 owing to this approach. We hypothesised a maximum intake of 10 amino acids, as outlined in the Method section. This systemic synthesis of numerous important biological objective functions, followed by model-driven analysis of the resulting flux distribution optima, could provide a viable computational technique for systems-level inspection of context-specific metabolic behaviour. If the individual fluxes had not been restricted previously by the derivative of the fitted cubic splines, the procedure would fail, highlighting the importance of following order of the modules in the framework application.

The results are analysed to establish the limitations of the culture media that may have contributed to the cell decline or switch from growth to non-growth. Overfeeding glucose slows lactate uptake, and this combination of glucose and glutamine feed is not optimal, resulting in a non-optimal metabolism in which a considerable percentage of glucose is converted to lactate, while glutamine is rapidly consumed, especially in the producing cell line. However, both ammonia and lactate remain below lethal levels, leading us to believe that the switch to non-growth phase is caused by a glutamine shortage. The clonal diversity had no effect on cell survival or concentration until the metabolic switch to the non-growth phase, although the peak cell density in the high-producer clone is significantly lower.

In the production phase of the high-producer, HIS, ILE, and MET depleted from the essential amino acids which may have caused a reduction in the monoclonal antibody. The depletion does not happen in the other two clones, however, the mentioned amino acids remain available in small amounts. For the non-essential amino acids, we observe depletion of SER and TYR especially in the high-producer clone. Most probably either higher initial concentrations or intermediate feeding of the mentioned five amino acids could fuel cell metabolism to produce more antibody. Interestingly, the model captures the metabolic shift from alanine production to alanine consumption accurately with the most pronounced shift for the high-producer clone.

4.4.5 Solution space underdeterminancy and infeasibility

We know that the biosystem acquires a mass balanced carbon flux distribution at the experimental conditions, thus, we expect the model to find feasible solutions within the same *in-silico* conditions. We argue that if the LP optimization solver cannot find a feasible solution in the flux vector space R^n, then, there exist some limitations in the model such as unfilled gaps, inaccurate mechanistic assumptions, inconsistent model scope, etc. Relaxations

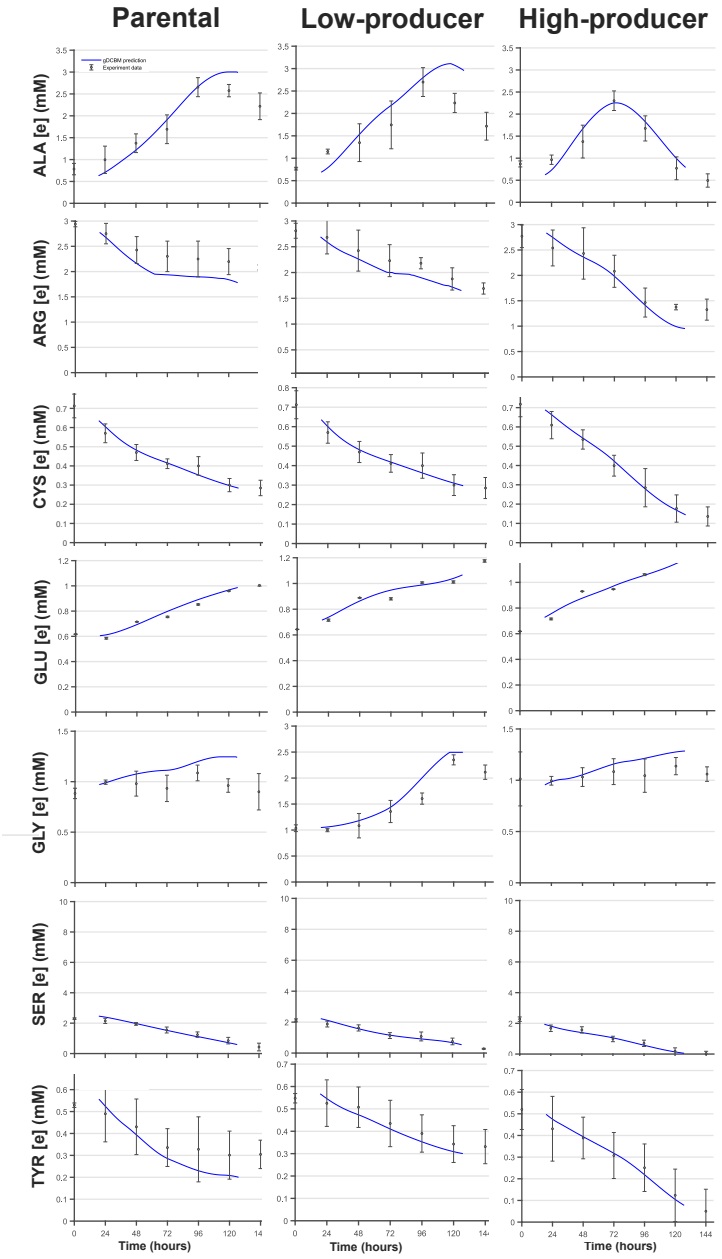

Figure 4.4: **Concentrations of the extracellular metabolites: Non-Essential amino acids** The rows reflect non-essential amino acid concentrations for which measurements were obtained, and the columns represent parental, low-producer, and high-producer clones. The blue lines reflect the gDCBM prediction of the concentrations, while the error bars reveal the experimental data.

in the imposed exchange fluxes to account for the exometabolomics uncertainty and resolving conflicting equations by modifying metabolic network connectedness both reduced the infeasibility. After we have had adequate experience with the preliminary model simulations,

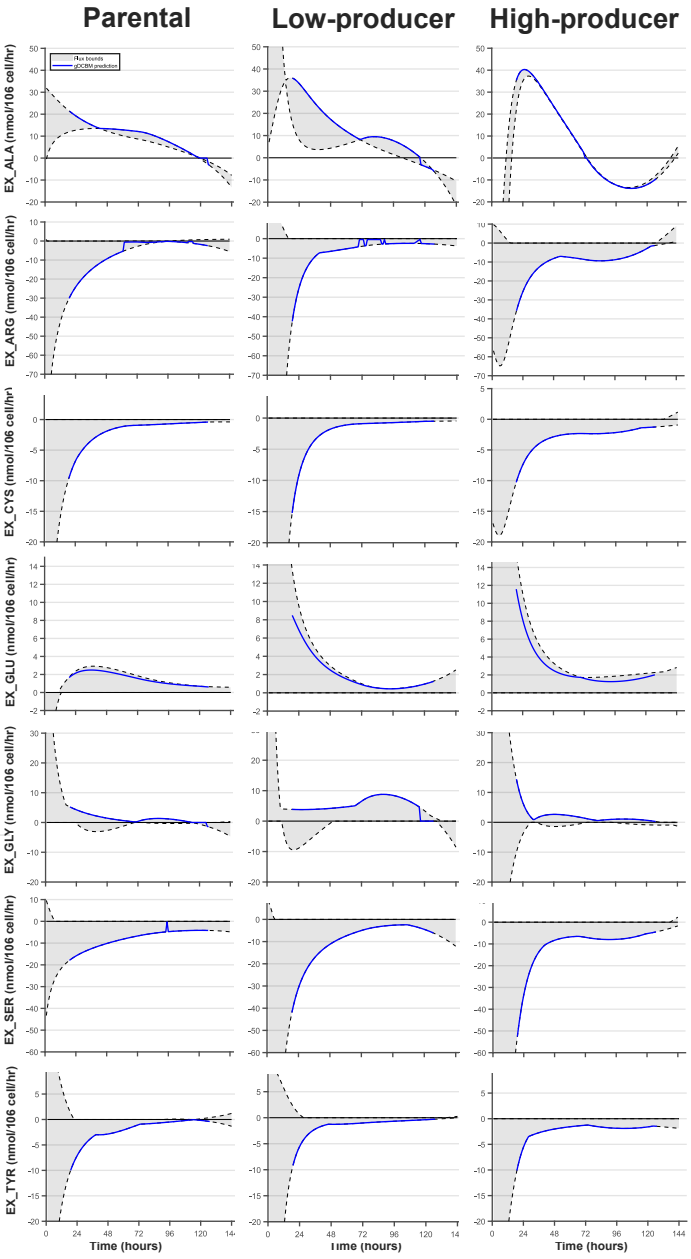

Figure 4.5: **Specific uptake or secretion rates: Non-Essential amino acids** The columns correspond to the parental, low-producer, and high-producer clones, respectively, and the rows to non-essential amino acids for which measurements were available. The blue lines reflect the gDCBM prediction of the specific rates, while the shaded areas show allowed bounds for the specific rates.

we differentiated between relaxations and conflicting equations. We addressed the former by adopting spline smoothing method followed by constraints from one way for the exchange fluxes and the latter by tailoring the network to remove blocked reactions and modifying

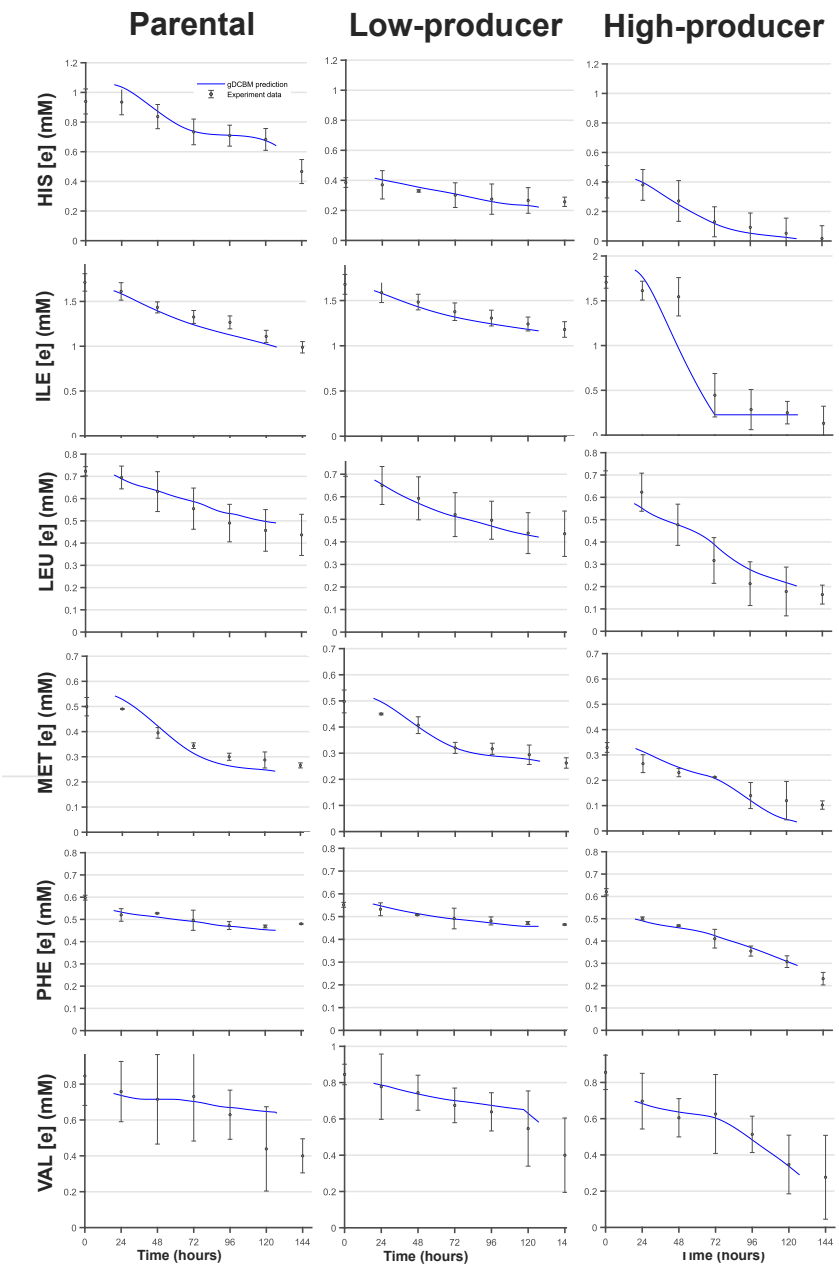

Figure 4.6: **Concentrations of the extracellular metabolites: Essential amino acids** The rows reflect essential amino acid concentrations for which measurements were obtained, and the columns represent parental, low-producer, and high-producer clones. The blue lines reflect the gDCBM prediction of the concentrations, while the error bars reveal the experimental data.

bounds on the remaining reactions. Hence, both these items play important roles in tackling the infeasibility issue regularly associated with stoichiometric-based models ([241, 242]). We used COBRA Toolbox library functions to check for the consistency of reactions, carbon

balance, removal of blocked reactions and frameworking computations.

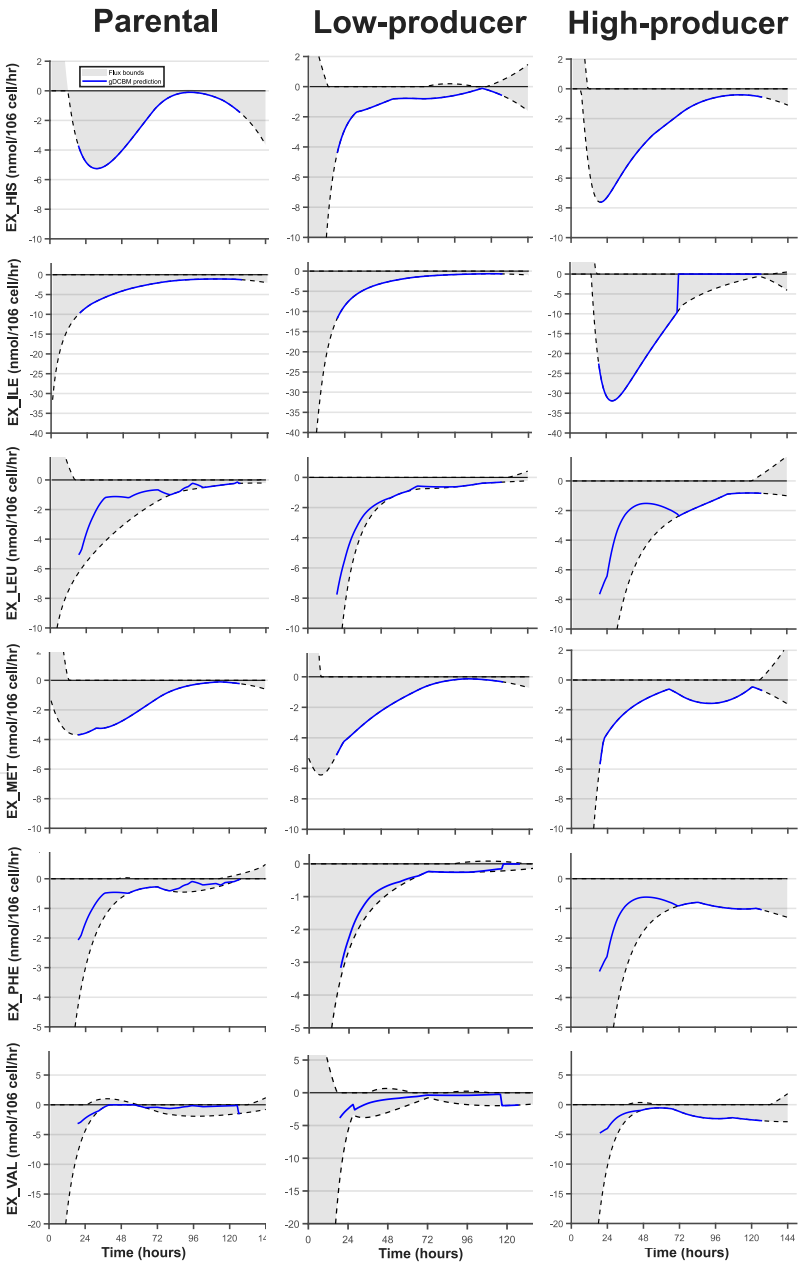

Figure 4.7: **Specific uptake or secretion rates: Essential amino acids** The columns correspond to the parental, low-producer, and high-producer clones, respectively, and the rows to essential amino acids for which measurements were available. The blue lines reflect the gDCBM prediction of the specific rates, while the shaded areas show allowed bounds for the specific rates.

It is shown in Figure 4.7 that HIS and ILE are consumed more rapidly during the growth-phase of the high-producer clone. This could suggest that production begins prior to the

transition to the non-growth phase. In the case of ILE, the simulation matches experimental data when practically all consumption in the non-growth interval ceases. Unlike the parental and low-producer clones, the high-producer clone's intake of LEU, MET, and PHE remains stable until the end of the culture time. Notable is also the fact that, at least for LEU, PHE, and VAL, the consumption rate in the majority of culture time points is far from the estimated limit value, demonstrating the efficacy of our method in viewing the specific rates as a range rather than a single value.

4.4.6 Indicators of increased efficiency

It is known that high-producing cell lines have a higher efficiency in carbon utilization than non-producing cell lines. In our simulations, the model can capture the differences between the global metabolism of clones. For example, in Figures 4.2 comparing panel A and panel C shows that lactate is produced much less in high-producer clone compared to the parental and low-producing cell lines and even the high-producer starts consuming lactate at the end of the cell culture. Similar trend with a more pronounced uptake is observed for alanine, which is continuously produced even after the glutamine consumption in the parental cell culture but it is consumed in the high-producer when glutamine is depleted. We hypothesize that the cellular control machinery will expend energy to create (or activate) a pathway for utilization of the less favorable nitrogen source and in return the cell productivity is sustained.

4.4.7 Comparison between clones with different mAb productivity levels

We argue that understanding the cellular physiology dynamics during the cell culture is the key to creating high-producing cell lines. The flux magnitude of model antibody production is relatively small when compared to the flux value through biomass synthesis, thus, the productivity level does not have a considerable impact on flux prediction results. However, the fact that a clone is engineered to be a high-producer rewires its metabolism. There are puzzling areas of CHO metabolism that obstacle harnessing an efficient energy metabolism for achieving yields closer to the theoretical production yields. The results show that glucose and glutamine are the primary sources for energy and carbon skeleton, depletion of each of which resulting in the growth limitation. Moreover, the time-resolved dynamics of asparagine, alanine, glycine and glutamate shows distributed nitrogen source utilization throughout the culture, especially after glutamine is depleted. Based on the results provided here, it seems that glucose must be fed in minimal amounts as this has advantages both for the limitation of undesired Warburg effect and possible commence of substrate utilization for product formation rather than growth maximization. Interestingly, even though nitrogen intake in-

creased in the high-producer, the waste $NH4$ did not exceed its levels in the parental and low-producer clone. It shows that the increase in nitrogen consumption encountered with an increase in its demand from the intracellular environment.

4.4.8 Intracellular flux predictions

For the sake of brevity, we show in Figure 9 five reactions from upper glycolysis and pentose phosphate pathway to show the ability of the framework to model intracellular reactions. Noteworthy is the demonstration of the optimal points along with the alternate optima range, which both addresses the well-known problem of FBA, namely, appearance of more than one optimum solution and shows that although some reactions work in a certain direction, but they can also work in other directions. For reaction HEX1 we observe that there are fractures in the optimal path passing zero activity, these fractures indicate the existence of different ways to convert glucose to g6p and are also a function of the sequential optimization method discussed in the Method section indicating a sudden change in the solution compared to the neighbouring points. This refers to how the answer is calculated step by step and that the answer of the equation at each point is independent of the previous or next point. The results show that at the beginning of the culture time the range of optimal alternative points is very narrow but it widens from the middle of the culture time. While for the high-producer clone this expansion coincides with the switch to the non-growth phase, it is not clear to the authors what is the reason for this change in the other two clones. We observed that in most of the intracellular flux predictions large bands are obtained for the reactions. This shows that the restrictions were not enough, especially when a reaction is further from the exchange reactions or there are more than one way of exchanging material in and out of the cell. This pattern also points to the inherent robustness of the cell by having several routes and it shows why inactivation of single enzymes is not enough to navigate metabolic flux.

Several important aspects of metabolism such as enzyme regulation cannot be covered by this framework. In fact, the ability to model genome-scale phenotype of reactions was achieved in trade-off with losing the ability to study kinetic regulations. Thus, this study is in a sense complementary to the previous study conducted in our group which was focused on the development of a dynamic kinetic model for the same biosystem [248,280]. Some intracellular flux predictions for example for the fluxes stemming from pyruvate indicates either inactivity or activity at the lower (upper) bound of the fluxes, it is in contrary to our findings with a reduced model of the same system [248]. We justify that this happens because of the inclusion of pyruvate in four compartments (extracellular, cytoplasm, mitochondria and peroxisome) participating in a total of 30 reactions. This many degree of freedom for the mass balance on

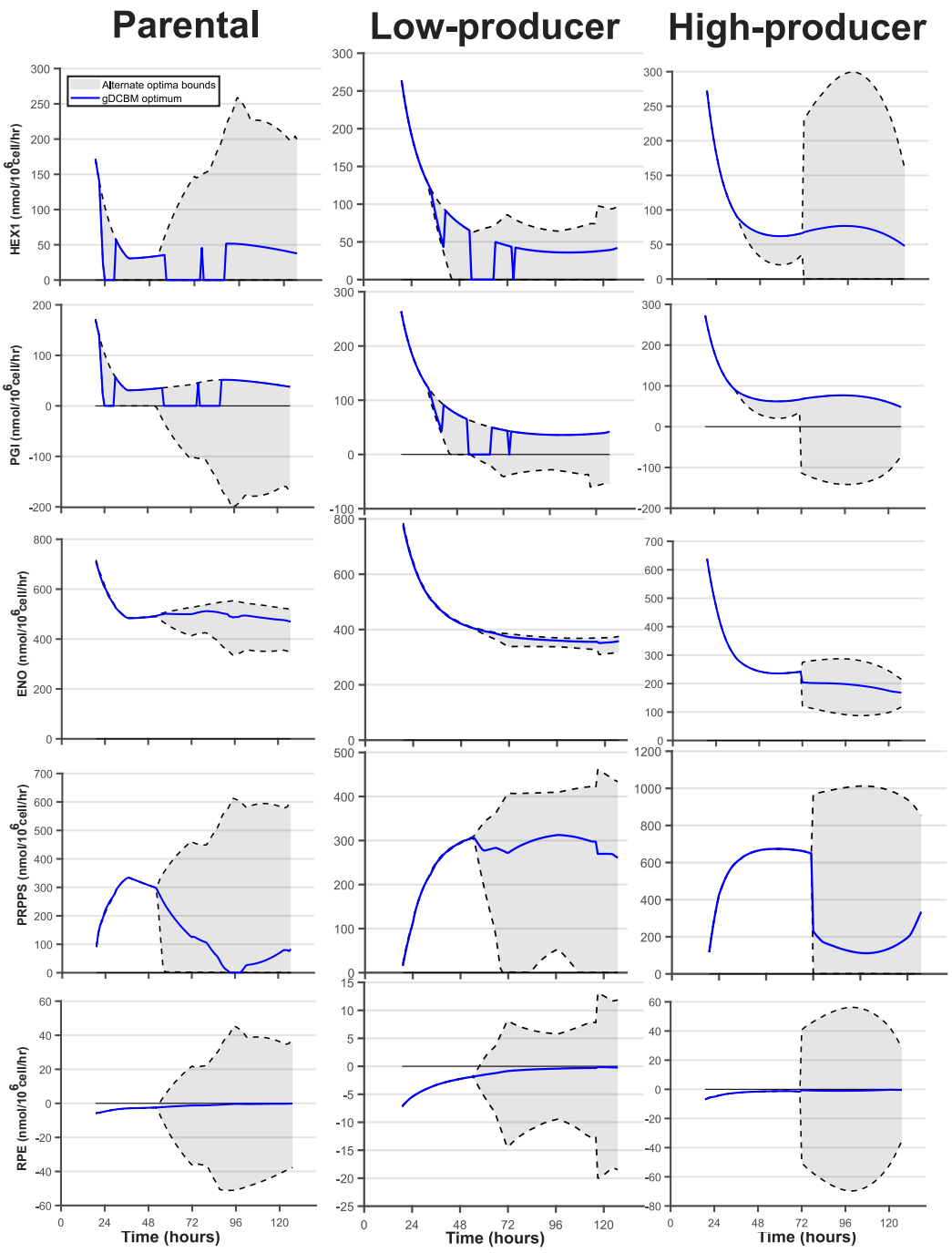

Figure 4.8: **Selected intracellular fluxes dynamics** The blue lines indicate the optimum flux estimation and the shaded area shows the alternate optima bounds calculated for each flux based on FVA. HEX1: Hexokinase, PGI: Glucose-6-phosphate isomerase, ENO: Enolase, PRPPS: Phosphoribosylpyrophosphate synthetase, RPE: Ribulose 5-phosphate 3-epimerase.

pyruvate may cause some false inactive fluxes and some unrealistically large fluxes. It points to the necessity of collecting more data such as carbon labeling data for further constraining the model and therefore navigating the flux into plausible reactions.

4.5 Conclusion

In this work we built a modelling approach named gDCBM upon two main principles: (i) metabolic network dynamics is key to describe cell physiology, and (ii) optimality hypotheses support generating predictions in cell culture. We proposed a gDCBM framework in which a metabolic network for the CHO cell line was selected, bound by cell culture specific uptake rates, and then tuned to predict growth dynamics with a one-hour temporal resolution. In fact, cell internal metabolism was linked to the dynamic of change for external moieties. This global genome-wide understanding of mammalian cell metabolism was then used to investigate differences in clonal variations of CHO cell lines. The model showed being able to successfully predict the time trajectory of all available experimental measurements except for the model antibody, which was produced in small quantities and on flux levels that were significantly lower than other fluxes. We believe that the proposed mechanistic gDCBM metabolic network modelling technique can help to overcome some of the current challenges in *in-vitro* and *in-silico* bioprocess optimization. This technique is likely to benefit complex organisms like mammalian cells, which can absorb a variety of diverse nutritional sources. With such a metabolic modelling framework, better applications for investigating cell metabolism and physiology, as well as media optimization and biomanufacturing control, may be achievable.

4.6 Software

All the calculations and computations were done using MATLAB R2020b (The Mathworks; Natick, MA, USA) as compiler. The computational algorithms are cited in the text and were implemented through developed scripts using glpk, gurobi, or CPLEX optimization solvers (academic licenses) (gnu.org/software/glpk/, gurobi.com, ibm.com/analytics/cplex-optimizer) and with borrowed functions from COBRA Toolbox library ([16]) and CellNet-Analyzer library of MATLAB functions. The computational demand of gDCBM for a system of 2132 reactions at 100 time points exceeded capability of a personal computer with Intel(r) Core(TM) i5-8250U CPU @ 1.60 GHz and 8 GB RAM memory. Thus, the analysis were performed on remote high-performance computing (HPC) clusters of Compute Canada. It took two hours to run the algorithm once on a single computation node with 12 CPU-workers

and 20 GB RAM.

4.7 Supporting information

S1 File. Reactions List. List of all the considered reactions.

S2 File. Biomass_eqn_1. Calculations performed to change the dimension of reaction coefficients of biomass synthesis from per gram to per million cell.

S3 File. Biomass_eqn_2. Calculations performed to determine the reaction coefficients of individual macromolecular synthesis and biomass synthesis.

S4 File. mAb_eqn. Calculations performed to determine the reaction coefficients of amino acids and energetic cofactors in Rituximab production reaction.

Data availability

The authors declare that all data supporting the findings of this study are available within the paper and its supplementary material files. Additional data are available from the corresponding author upon reasonable request.

Authors' contributions

MY, Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing - original draft, Writing - review editing; MJ, Conceptualization, Funding acquisition, Methodology, Project administration, Resources, Software, Supervision, Writing - original draft, Writing - review editing.

Funding

MJ received the NSERC (https://www.nserc-crsng.gc.ca/) Discovery Grant RGPIN- 2019-05050. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Declaration of competing interest

The authors declare that they have no competing interests.

CHAPTER 5 GENERAL DISCUSSION

The metabolic modelling of two distinct biosystems, namely RBC and CHO, was at the heart of the fundamental challenge that we attempted to solve. The first was focused on reducing oxidative lesions, while the second was focused on further optimising mAb production. We were particularly interested in simulating and predicting transient metabolic behaviour of stored RBCs and cultured CHO cell lines in response to changing medium composition, and then devising ways for interpreting the underlying biological mechanisms using models.

When considering a solution to these problems in the context of metabolic modelling, we have faced issues like: What is the importance of a dynamic metabolic model for describing cell culture dynamics? What criteria affect a dynamic metabolic model's success? What are some unique ways for getting the most out of combining diverse modelling approaches? Which experiments must be carried out in order to validate the model? What conclusions can be drawn from the interaction between model simulations and experiment findings? In response to these inquiries, we devised additional hypotheses to be investigated along with this thesis. We postulated that establishing a modular *in silico* platform to evaluate non-steady state metabolic fluxes and quantify medium dynamics could help to better understanding and optimising different biosystems. In other words, while the various modelling approaches discussed in Chapter 2 each have their own set of benefits and drawbacks, we believe that the most useful information can be derived when they are used in a modular interactive fashion, beginning with those that retrieve the structure of the metabolic network and ending with those that quantify flux values within the boundaries of imposed physicochemical constraints.

In the case of RBCs, the modelling approach starts with processing experimental data so that noise in the measurements does not lead to infeasibility or erroneous estimation bounds. We employed nonlinear regression with exponential fit functions to smooth out the measured experimental data in Chapter 3. The metabolic network was subsequently rebuilt to describe RBC oxidative metabolism. The model scope includes cofactor-dependent enzymes involved in ROS termination bioreactions, as well as glutathione metabolism, which was our major modelling target. The major RBC intracellular metabolism pathways were also taken into account. The reconstruction required multiple rounds of adjustments at the start in order to identify dynamically plausible solutions. When the viable solutions appeared for the first time, i.e., the metabolic network was confirmed, the recovered stoichiometric model supported DFBA simulations considering the four objectives with only minor changes. Our method in assessing distinct objective functions showed proximity of intracellular flux estimations for the

coupling biological phenomena such as oxidative stress termination and NADH production.

In Chapter 4 we upgraded the method to use cubic splines to give a lower error in the primary simulation of the experimental data. Furthermore, instead of an *ad hoc* metabolic network, as was the case with RBCs, we tailored a GSMM. Four interactive components made up the framework for gDCBM. In the first module, smooth splines were employed to approximate the data. The metabolic model was built from the reference GSMM in the second module, taking into consideration the dynamic constraints imposed by the piece-wise polynomials produced in the data curation module. In the third module, we formulated the body of the FBA problem, and in the fourth module, we used optimization solvers to obtain the integrated concentration trajectories.

The CHO GSMM was updated by deleting the blocked reactions and altering the phenomenological reactions to meet the investigation's needs. Due to the enormous complexity and uncertainty of biomass composition, traditional implementation of the molar mass conservation rule is difficult and costly. We chose to develop a standard version of the model with a biomass synthesis reaction based on a recent extensive study of CHO biomass composition in various strains, as discussed in Chapter 4.

Modeling the non-growth phase of mammalian cell culture proved difficult, thus we used the findings from the quiescent RBCs publication to build ideas for capturing metabolic behaviour in the non-growth phase in this thesis. In this phase, we were able to successfully describe the cell dynamics by considering multiple metabolic goal functions related to amino acid exchange reactions. The time point with the least error was used to compute the switch time associated to each clone. The exercise yielded three different switch time points for the various clones. The gDCBM predictions were quite close to the experimental outcomes in every case. Indeed, based on the decrease of the NRMSE, which accounts for the prediction versus experimental data mismatch among all observable states, we proved that the framework enables for *a posteriori* detection of the transition from growth to non-growth phase.

In a rich cell culture media, we hypothesised that CHO cells would try to ingest as many nutrients as their internal metabolism would allow and produce as much product as their internal metabolism would also allow. As a result, we concentrate on the cell's input, assuming that intracellular activity is the limiting factor in medium constituent exchange across the cell membrane. In the fundamental FBA optimization problem, the input flows of these chemicals can be employed as objectives. Because of this method, we were able to obtain the simulations in Chapter 4. This systemic synthesis of multiple essential biological objective functions, followed by model-driven analysis of the resulting flux distribution optima, should provide a viable computational technique for context-specific metabolic behaviour assessment

at the systems level. The approach would fail if the individual fluxes had not been previously constrained by the derivative of the fitted cubic splines, emphasising the significance of following the order of the modules in the framework application.

The scientific research finally performed in this thesis appeared as a consolidation of theory, specialized methodology and proposed ideas to domesticate cross-referenced topics.

Finally, an important contribution of this work was the generation of a Matlab–based framework devoted to incorporate COBRA toolbox with several novel methods used in the course of the project conduction. The prospect for the developed script is to be implementable for calibration of the model to other datasets with the least adjustment required, in addition to providing the user with all the necessary information to reproduce or implement the methodologies applied in this project.

5.1 Original contributions and implications

The original contributions of this thesis to the scientific community are briefed as follows,

- In Chapter 2, we provided a review analysis of the literature on cell modeling and provide goals to address existing barriers. It is believed that the literature review structure in this chapter can benefit novice modelers with biology background and mathematicians with interests in biology equally. Such structure can also highlight a new perspective regarding fusion of algebra techniques and nonlinear dynamics analysis techniques. The review article is particularly focused on the classification and comparison of dynamic kinetic models and constraint-based models of metabolism.

- In Chapter 3, the development of an *ad-hoc* metabolic network for RBCs is an original contribution followed by its study through dynamic stoichiometry based methods. The metabolic network was meticulously developed by considering the genome-scale metabolic model of RBCs and accumulated physiological data on the RBCs metabolic behaviour. In this chapter, data was related to a six-week blood cell maintenance process to test the capabilities of the proposed model. The model accounts for the effects of different metabolic goals on intracellular flow distributions, particularly fluxes linked to oxidative stress reduction. Through this methodology, the capabilities of DCBM are extended. For example, in this study FVA was adapted to derive unbiased ranges for the intracellular fluxes prior to the implementation of any metabolic objective function.

- In Chapter 4, the first contribution was to tailor a previously developed GSMM, then it was followed by the development of a dynamic modeling framework for modeling

data on CHO clonal variations. This study involves novel assumptions regarding cell composition, switch time from non-growth to growth phase of the culture, and formulation of the biomass synthesis reaction. In the context of CHO profiting from the findings of the study on non-growing RBCs, the model proposed a set of objectives for prediction of growth dynamics and intracellular flux patterns in both growth and non-growth phases, which are original and novel. Therefore, the modelling framework makes it possible to model the metabolic behaviour of three different strains in their production phase as well as the growth phase.

CHAPTER 6 RECOMMENDATIONS

Similar investigations can make advantage of the framework entire structure with replacement of the experimental data. To improve the capacity to identify fluxes based on scarce experimental data, alternative exchange paths for extracellular metabolites can be blocked in a follow-up investigation. One way to continue this research could be to study specific subsets of this genome-scale model by implementing model extraction methods on the available metabolic network. Such models could be kinetically determined and thus subject to dynamic kinetic modelling. The advantage of using the model developed in this study as the basis, is that it has already been proven to be able to result in feasible solutions. Moreover, it was not possible to solve the gDCBM optimization problem in simultaneous format because of computational limitation. A reduced model extracted based on the available data can be integrated simultaneously over the culture time. For such system, the solution can be interpreted as an optimal dynamic behaviour. A dynamic nonlinear analysis of the underlying metabolic system is also required to observe emergent properties such as ultra-sensitivity (switch-like behaviour), bistability, and oscillations, which cannot be attributed to any single reaction or constituent of the network and can only be explained with an enhanced systems understanding.

In general, kinetic modelling improves the accuracy of perturbation outcome predictions and allows for more flexibility in simulation of different scenarios, owing to the inclusion of detailed mechanistic descriptions of the cell's regulatory and compensatory mechanisms through the use of a variety of kinetic formats. In terms of future study, developing a reduced version of the metabolic network by limiting the network to the links between the detected extracellular metabolites is a step toward developing a hybrid model with kinetic properties. Such network can be characterized based on the recognised kinetic modelling methodologies. The flux kinetics of the regulated enzymatic processes can be modelled as nonlinear constraints for a differential algebraic optimization problem similar to what we defined in this thesis. This modelling technique is based on the notion that changes in enzyme activity during cell operation may be described mechanistically in terms of chemical activity of substances that govern or determine enzyme activity. Unfortunately, such mechanistic information is not implementable at genome-scale yet. At this scale we observed that many intracellular reactions cannot be determined, for example, had huge bounds, such parts of the genome-scale model are almost of no use for metabolic flux analysis. After all, there are more than one thousand entity in this model of which we only have measurements for less than fifty ones. In a functional hybrid model, kinetic rate laws are applied for the parts of

the network for which mechanistic data is available, but the rest of the network is kept in its purely stoichiometric shape. This stoichiometric shape must be a subset of the metabolic network developed in this thesis.

For parameter estimation of dynamic kinetic models containing mechanistic parameters, several adequate techniques, i.e. those employing meta-heuristics algorithms, have been reported in the literature. When gradient-based optimization approaches seem to become trapped in unsatisfactory extrema, meta-heuristic optimization procedures are employed to prevent the intrinsic nonlinearity of metabolic networks and the growth of parameters in an excessive number of dimensions. We forecast that parameter estimate quality will be considerably enhanced by the creation of kinetic models that can match as many different data sources as possible. Given the current overabundance of quantitative omics data and growing biochemical thermodynamics data, we believe that successful metabolic modelling approaches in the future will help incorporate further knowledge from both inputs in a coherent computational framework. Its strength will be clearly understood when mathematical modelling plays a larger role in the development and design of new useful technologies. The question of "what does the result of these genes do?" has become more important with the rapid improvement of genome sequencing. What interactions between various gene products dictate which strains produce more and which produce less? To identify the solution, strong complementary analytical and computational approaches will need to be developed in addition to the vast amount of data from (reductionist) biological sciences. Without a significant contribution from mathematical modelling techniques, it will be impossible to explain and recreate the interplay between genetics, phenotype, and the environment of the cell.

CHAPTER 7 CONCLUSION

The development of efficient bioprocess management strategies necessitates the use of mathematical models. Most mathematical models published in the literature that describe the dynamics of nutrients and by-products during cell cultures either require a high number of parameters or are concerned with the organisms simpler than mammalian cells. Recognizing that these models are susceptible to parameters change or experimental data noise and inaccuracy, the use of an interactive modular DCBM approach was investigated. This type of model is known to be more compact in terms of the number of parameters. Our working premise was that, based on a thorough preliminary investigation in the literature, dynamic metabolic flux models will be able to represent the transient behaviour of biosystems during a bioprocess with fewer parameters. To objectively assess this theory, we developed mathematical tools for describing a non-growing biosystem, i.e., RBCs during their storage time, and a growing biosystem, i.e., CHO cells during batch cell culture.

Initially, metabolic modelling algorithms were employed to describe RBC storage lesions with emphasis on examining different biological objective functions. In this part of the thesis, we used nonlinear regression to provide a smooth approximation of the experimental measurements of the blood bags nutrients, then in a DCBM approach, we derived dynamic intracellular flux distributions with regard to different hypothetical objective functions. Next, our attention was turned to CHO cell metabolism as it is the workhorse for biotherapeutics production. In this part, in a shift from the metabolic modelling approach previously developed in our research group, we formulated the problem of simulating dynamic of medium components as a constraint-based differential algebraic optimization problem. The objective function for this optimization problem has two parts that depend on growth or non-growth phase of the cell culture. We derived bounds of exchange fluxes based on the derivative of a cubic smooth spline approximation of the measured experimental data, which enabled us to find feasible genome-scale intracellular flux estimation ranges of the revised CHO GSMM.

Theoretically, developing dynamic constraint-based models of metabolism necessitates the identification of one or a set of meaningful biological objective functions to be maximized/minimized as well as active constraints on the fluxes. This thesis proposed key strategies to address these issues. A gDCBM framework was developed in which a metabolic network for CHO cell line was tailored and then constrained by cell culture specific uptake rates, and finally optimized to predict the growth dynamics in one-hour time resolution. We also developed an *ad-hoc* model of the stored RBCs, which was used to model the impact of

alternate objective functions in a non-growing organism. We leveraged the systems-based understanding of the mammalian cell metabolism and the RBCs metabolism for investigating the differences in clonal variations of CHO cell lines and for studying RBCs oxidative lesions, respectively.

We envisage that the provided mechanistic metabolic network modelling approach may address some of the current obstacles on the way of *in-vitro* and *in-silico* bioproduction optimization. The approach is likely to be particularly suitable for complex organisms such as mammalian cells, which can metabolize from multiple distinct nutrient inputs. Such metabolic modelling framework may allow an improved capacity for analyzing cell metabolism and physiology as well as media optimization and biomanufacturing control.

REFERENCES

[1] J. E. Bailey, "Mathematical modeling and analysis in biochemical engineering: past accomplishments and future opportunities," *Biotechnol Prog*, vol. 14, no. 1, pp. 8–20, 1998.

[2] M. Ataman and V. Hatzimanikatis, "Heading in the right direction: thermodynamics-based network analysis and pathway engineering," *Curr Opin Biotechnol*, vol. 36, pp. 176–82, 2015.

[3] B. O. PALSSON, "Mathematical modelling of dynamics and control in metabolic networks," Ph.D. dissertation, University of Wisconsin-Madison, 1984.

[4] A. Bordbar *et al.*, "Identified metabolic signature for assessing red blood cell unit quality is associated with endothelial damage markers and clinical outcomes," *Transfusion*, vol. 56, no. 4, pp. 852–62, 2016.

[5] H. Kitano, "Computational systems biology," *Nature*, vol. 420, no. 6912, pp. 206–10, 2002.

[6] J. D. III, *Dynamic Systems Biology Modeling and Simulation*, 1st ed. Academic Press, 2013.

[7] S. Volkova *et al.*, "Metabolic modelling as a framework for metabolomics data integration and analysis," *Metabolites*, vol. 10, no. 8, 2020.

[8] M. R. Antoniewicz, "A guide to metabolic flux analysis in metabolic engineering: Methods, tools and applications," *Metabolic Engineering*, 2020.

[9] V. L. Porubsky *et al.*, "Best practices for making reproducible biochemical models," *Cell Syst*, vol. 11, no. 2, pp. 109–120, 2020.

[10] P. Kotidis *et al.*, "Model-based optimization of antibody galactosylation in cho cell culture," *Biotechnol Bioeng*, vol. 116, no. 7, pp. 1612–1626, 2019.

[11] D. Degenring *et al.*, "Sensitivity analysis for the reduction of complex metabolism models," *Journal of Process Control*, vol. 14, no. 7, pp. 729–745, 2004.

[12] K. Smallbone *et al.*, "Something from nothing: bridging the gap between constraint-based and kinetic modelling," *FEBS J*, vol. 274, no. 21, pp. 5576–85, 2007.

[13] ——, "Something from nothing: bridging the gap between constraint-based and kinetic modelling," *FEBS J*, vol. 274, no. 21, pp. 5576–85, 2007.

[14] A. Bordbar *et al.*, "Constraint-based models predict metabolic and associated cellular functions," *Nat Rev Genet*, vol. 15, no. 2, pp. 107–20, 2014.

[15] R. Mahadevan, J. S. Edwards, and F. J. Doyle, "Dynamic flux balance analysis of diauxic growth in escherichia coli," *Biophysical journal*, vol. 83, no. 3, pp. 1331–1340, 2002.

[16] L. Heirendt *et al.*, "Creation and analysis of biochemical constraint-based models using the cobra toolbox v.3.0," *Nat Protoc*, vol. 14, no. 3, pp. 639–702, 2019.

[17] B. P. Ingalls, *Mathematical Modeling in Systems Biology: An Introduction*. The MIT Press, 2013.

[18] B. O. Palsson, *Systems Biology: Properties of Reconstructed Networks*. Cambridge, UK: Cambridge University Press, 2006.

[19] M. Jolicoeur, "Modeling cell behavior: moving beyond intuition," *AIMS Bioengineering*, vol. 1, no. 1, pp. 1–12, 2014.

[20] W. Zhou *et al.*, "Alteration of mammalian cell metabolism by dynamic nutrient feeding," *Cytotechnology*, vol. 24, no. 2, pp. 99–108, 1997.

[21] S. S. Reinhart Heinrich, *The Regulation of Cellular Systems*. Springer, 1996.

[22] W. Liebermeister and E. Klipp, "Bringing metabolic networks to life: convenience rate law and thermodynamic constraints," *Theoretical Biology and Medical Modelling*, vol. 3, no. 1, p. 41, 2006. [Online]. Available: https://tbiomed.biomedcentral.com/track/pdf/10.1186/1742-4682-3-41?site=tbiomed.biomedcentral.com

[23] B. Du *et al.*, "Evaluation of rate law approximations in bottom-up kinetic models of metabolism," *BMC Syst Biol*, vol. 10, no. 1, p. 40, 2016.

[24] C. H. Schilling *et al.*, "Metabolic pathway analysis: basic concepts and scientific applications in the post-genomic era," *Biotechnology progress*, vol. 15, no. 3, pp. 296–303, 1999.

[25] A. Ghorbaniaghdam, O. Henry, and M. Jolicoeur, "A kinetic-metabolic model based on cell energetic state: study of cho cell behavior under na-butyrate stimulation," *Bioprocess and Biosystems Engineering*, vol. 36, no. 4, pp. 469–487, 2013.

[26] G. Goffaux, I. Hammami, and M. Jolicoeur, "A dynamic metabolic flux analysis of myeloid-derived suppressor cells confirms immunosuppression-related metabolic plasticity," *SCIENTIFIC REPORTS*, vol. 7, 2017.

[27] C. Calmels *et al.*, "Application of a curated genome-scale metabolic model of cho dg44 to an industrial fed-batch process," *Metabolic Engineering*, vol. 51, pp. 9–19, 2019.

[28] H. Hefzi *et al.*, "A consensus genome-scale reconstruction of chinese hamster ovary cell metabolism," *Cell Syst*, vol. 3, no. 5, 2016.

[29] J. da Veiga Moreira *et al.*, "Metabolic therapies inhibit tumor growth in vivo and in silico," *Scientific Reports*, vol. 9, no. 1, p. 3153, 2019.

[30] R. Katzir *et al.*, "The landscape of tiered regulation of breast cancer cell metabolism," *Scientific Reports*, vol. 9, no. 1, p. 17760, 2019.

[31] I. Thiele *et al.*, "A community effort towards a knowledge-base and mathematical model of the human pathogen salmonella typhimurium lt2," *BMC Syst Biol*, vol. 5, p. 8, 2011.

[32] M. Cloutier *et al.*, "A systems approach to plant bioprocess optimization," *Plant biotechnology journal*, vol. 7, no. 9, pp. 939–951, 2009.

[33] L. Montegut *et al.*, "Combining lipoic acid to methylene blue reduces the warburg effect in cho cells: From tca cycle activation to enhancing monoclonal antibody production," *PLoS One*, vol. 15, no. 4, p. e0231770, 2020.

[34] A. G. Fredrickson, R. D. Megee, and H. M. Tsuchiya, *Mathematical Models for Fermentation Processes**During preparation of this review, the authors were supported in part by USDA Grant No. 12-14-100-9178 and USDPH Grant No. GM 16692.* Academic Press, 1970, vol. 13, pp. 419–465.

[35] H. Jeong *et al.*, "The large-scale organization of metabolic networks," *Nature*, vol. 407, no. 6804, pp. 651–654, 2000. [Online]. Available: https://doi.org/10.1038/35036627https://www.nature.com/articles/35036627.pdf

[36] P. D. Karp *et al.*, "Integrated pathway-genome databases and their role in drug discovery," *Trends Biotechnol*, vol. 17, no. 7, pp. 275–81, 1999.

[37] H. Ma and A.-P. Zeng, "Reconstruction of metabolic networks from genome data and analysis of their global structure for various organisms," *Bioinformatics*, vol. 19, no. 2, pp. 270–277, 2003.

[38] H. W. Ma and A. P. Zeng, "The connectivity structure, giant strong component and centrality of metabolic networks," *BIOINFORMATICS*, vol. 19, no. 11, pp. 1423–1430, 2003.

[39] D. A. Fell and A. Wagner, "The small world of metabolism," *Nature Biotechnology*, vol. 18, no. 11, pp. 1121–1122, 2000.

[40] A. Wagner and D. A. Fell, "The small world inside large metabolic networks," *Proceedings of the Royal Society of London. Series B: Biological Sciences*, vol. 268, no. 1478, pp. 1803–1810, 2001.

[41] D. Koschützki, H. Schwöbbermeyer, and F. Schreiber, "Ranking of network elements based on functional substructures," *Journal of Theoretical Biology*, vol. 248, no. 3, pp. 471–479, 2007.

[42] M. R. Meiss, F. Menczer, and A. Vespignani, "Structural analysis of behavioral networks from the internet," *Journal of Physics A: Mathematical and Theoretical*, vol. 41, no. 22, p. 224022, 2008.

[43] J. S. Edwards and B. O. Palsson, "The escherichia coli mg1655 in silico metabolic genotype: its definition, characteristics, and capabilities," *Proc Natl Acad Sci U S A*, vol. 97, no. 10, pp. 5528–33, 2000.

[44] D. Koschützki and F. Schreiber, "Centrality analysis methods for biological networks and their application to gene regulatory networks," *Gene regulation and systems biology*, vol. 2, p. 193, 2008.

[45] S. Wuchty and P. F. Stadler, "Centers of complex networks," *J Theor Biol*, vol. 223, no. 1, pp. 45–53, 2003.

[46] V. N. Reddy, M. L. Mavrovouniotis, and M. N. Liebman, "Petri net representations in metabolic pathways," in *ISMB*, vol. 93, 1993, Conference Proceedings, pp. 328–336.

[47] M. Chen, M. Hu, and R. Hofestadt, "A systematic petri net approach for multiple-scale modeling and simulation of biochemical processes," *Appl Biochem Biotechnol*, vol. 164, no. 3, pp. 338–52, 2011.

[48] I. Zevedei-Oancea and S. Schuster, "Topological analysis of metabolic networks based on petri net theory," *Stud Health Technol Inform*, vol. 162, pp. 17–37, 2011.

[49] C. Chaouiya, "Petri net modelling of biological networks," *BRIEFINGS IN BIOINFORMATICS*, vol. 8, no. 4, pp. 210–219, 2007.

[50] M. A. Marsan *et al.*, "Modelling with generalized stochastic petri nets," *ACM SIGMETRICS Performance Evaluation Review*, vol. 26, no. 2, 1998.

[51] P. J. E. Goss and J. Peccoud, "Quantitative modeling of stochastic systems in molecular biology by using stochastic petri nets," *Proceedings of the National Academy of Sciences*, vol. 95, no. 12, p. 6750, 1998.

[52] R. Hofestadt and S. Thelen, "Quantitative modeling of biochemical networks," *In Silico Biol*, vol. 1, no. 1, pp. 39–53, 1998.

[53] R. Valk, "Self-modifying nets, a natural extension of petri nets," in *Automata, Languages and Programming*, G. Ausiello and C. Böhm, Eds. Springer Berlin Heidelberg, 1978, Conference Proceedings, pp. 464–476.

[54] A. Varma and B. O. Palsson, "Metabolic flux balancing: basic concepts, scientific and practical use," *Nature Biotechnology*, vol. 12, no. 10, pp. 994–998, 1994.

[55] E. Ravasz *et al.*, "Hierarchical organization of modularity in metabolic networks," *science*, vol. 297, no. 5586, pp. 1551–1555, 2002.

[56] S. Bilke and C. Peterson, "Topological properties of citation and metabolic networks," *Physical Review E*, vol. 64, no. 3, p. 036106, 2001.

[57] D. A. Fell and J. R. Small, "Fat synthesis in adipose tissue. an examination of stoichiometric constraints," *Biochemical Journal*, vol. 238, no. 3, pp. 781–786, 1986.

[58] R. Majewski and M. Domach, "Simple constrained-optimization view of acetate overflow in e. coli," *Biotechnology and bioengineering*, vol. 35, no. 7, pp. 732–738, 1990.

[59] E. T. Papoutsakis, "Equations and calculations for fermentations of butyric acid bacteria," *Biotechnology and bioengineering*, vol. 26, no. 2, pp. 174–187, 1984.

[60] J. M. Savinell and B. O. Palsson, "Network analysis of intermediary metabolism using linear optimization. i. development of mathematical formalism," *Journal of theoretical biology*, vol. 154, no. 4, pp. 421–454, 1992.

[61] S. Fernandes *et al.*, "Dynamic metabolic flux analysis of underdetermined and overdetermined metabolic networks," *IFAC-PapersOnLine*, vol. 49, no. 26, pp. 318–323, 2016.

[62] F. Zamorano, A. V. Wouwer, and G. Bastin, "A detailed metabolic flux analysis of an underdetermined network of cho cells," *J Biotechnol*, vol. 150, no. 4, pp. 497–508, 2010.

[63] W. L. Robert and R. A. Maciek, "Dynamic metabolic flux analysis (dmfa): A framework for determining fluxes at metabolic non-steady state," *Metabolic Engineering*, vol. 13, no. 6, pp. 745–755, 2011.

[64] W. S. Ahn and M. R. Antoniewicz, "Towards dynamic metabolic flux analysis in cho cell cultures," *Biotechnology Journal*, vol. 7, no. 1, pp. 61–74, 2012.

[65] J. Niklas and E. Heinzle, "Metabolic flux analysis in systems biology of mammalian cells," *Advances in biochemical engineering/biotechnology*, vol. 127, pp. 109–132, 2012. [Online]. Available: http://europepmc.org/abstract/MED/21432052https://doi.org/10.1007/10_2011_99

[66] W. S. Ahn and M. R. Antoniewicz, "Metabolic flux analysis of cho cells at growth and non-growth phases using isotopic tracers and mass spectrometry," *Metab Eng*, vol. 13, no. 5, pp. 598–609, 2011.

[67] M. Schauer and R. Heinrich, "Quasi-steady-state approximation in the mathematical modeling of biochemical reaction networks," *Mathematical Biosciences*, vol. 65, no. 2, pp. 155–170, 1983.

[68] G. Stephanopoulos, A. A. Aristidou, and J. Nielsen, *Metabolic engineering: principles and methodologies*. Academic press, 1998.

[69] J. Niklas *et al.*, "Quantitative characterization of metabolism and metabolic shifts during growth of the new human cell line age1.hn using time resolved metabolic flux analysis," *Bioprocess and biosystems engineering*, vol. 34, no. 5, pp. 533–545, 2011.

[70] I. Famili and B. O. Palsson, "The convex basis of the left null space of the stoichiometric matrix leads to the definition of metabolically meaningful pools," *Biophysical Journal*, vol. 85, no. 1, pp. 16–26, 2003.

[71] F. Llaneras and J. Pico, "Which metabolic pathways generate and characterize the flux space? a comparison among elementary modes, extreme pathways and minimal generators," *JOURNAL OF BIOMEDICINE AND BIOTECHNOLOGY*, 2010.

[72] S. Klamt *et al.*, "From elementary flux modes to elementary flux vectors: Metabolic pathway analysis with arbitrary linear flux constraints," *PLoS computational biology*, vol. 13, no. 4, pp. e1 005 409–e1 005 409, 2017.

[73] S. Srinivasan, W. R. Cluett, and R. Mahadevan, "Constructing kinetic models of metabolism at genome-scales: A review," *Biotechnology Journal*, vol. 10, no. 9, pp. 1345–1359, 2015.

[74] H. Kitano, "Biological robustness," *Nature Reviews Genetics*, vol. 5, no. 11, pp. 826–837, 2004.

[75] S. Schuster and C. Hilgetag, "On elementary flux modes in biochemical reaction systems at steady state," *Journal of Biological Systems*, vol. 02, no. 02, pp. 165–182, 1994. [Online]. Available: https://doi.org/10.1142/S0218339094000131

[76] S. Schuster, D. A. Fell, and T. Dandekar, "A general definition of metabolic pathways useful for systematic organization and analysis of complex metabolic networks," *Nature biotechnology*, vol. 18, no. 3, pp. 326–332, 2000.

[77] S. Schuster *et al.*, "Reaction routes in biochemical reaction systems: Algebraic properties, validated calculation procedure and example from nucleotide metabolism," *Journal of Mathematical Biology*, vol. 45, no. 2, pp. 153–181, 2002.

[78] C. T. Trinh, A. Wlaschin, and F. Srienc, "Elementary mode analysis: a useful metabolic pathway analysis tool for characterizing cellular metabolism," *Applied Microbiology and Biotechnology*, vol. 81, no. 5, pp. 813–826, 2009.

[79] C. H. Schilling *et al.*, "Metabolic pathway analysis: basic concepts and scientific applications in the post-genomic era," *Biotechnology progress*, vol. 15, no. 3, pp. 296–303, 1999.

[80] Pfeiffer *et al.*, "Metatool: for studying metabolic networks," *Bioinformatics*, vol. 15, no. 3, pp. 251–257, 1999.

[81] J. Papin *et al.*, "Comparison of network-based pathway analysis methods," *Trends in biotechnology*, vol. 22, no. 8, pp. 400–405, 2004.

[82] J. A. Papin *et al.*, "Metabolic pathways in the post-genome era," *Trends in biochemical sciences*, vol. 28, no. 5, pp. 250–258, 2003.

[83] S. Klamt and J. Stelling, "Two approaches for metabolic pathway analysis?" *Trends in Biotechnology*, vol. 21, no. 2, pp. 64–69, 2003.

[84] S. Peres, S. Schuster, and P. Dague, "Thermodynamic constraints for identifying elementary flux modes," *Biochem Soc Trans*, vol. 46, no. 3, pp. 641–647, 2018.

[85] J. Gagneur and S. Klamt, "Computation of elementary modes: a unifying framework and the new binary approach," *BMC bioinformatics*, vol. 5, no. 1, p. 175, 2004.

[86] S. Klamt and J. Stelling, "Combinatorial complexity of pathway analysis in metabolic networks," *Molecular biology reports*, vol. 29, no. 1, pp. 233–236, 2002.

[87] J. Zanghellini *et al.*, "Elementary flux modes in a nutshell: Properties, calculation and applications," *Biotechnology Journal*, vol. 8, no. 9, pp. 1009–1016, 2013.

[88] C. H. Schilling and B. Ø. Palsson, "Assessment of the metabolic capabilities of haemophilus influenzae rd through a genome-scale pathway analysis," *Journal of Theoretical Biology*, vol. 203, no. 3, pp. 249–283, 2000.

[89] R. Carlson and F. Srienc, "Fundamental escherichia coli biochemical pathways for biomass and energy production: Identification of reactions," *Biotechnology and Bioengineering*, vol. 85, no. 1, pp. 1–19, 2004. [Online]. Available: https://doi.org/10.1002/bit.10812

[90] C. T. Trinh and R. A. Thompson, *Elementary mode analysis: A useful metabolic pathway analysis tool for reprograming microbial metabolic pathways.* Springer, 2012, pp. 21–42.

[91] S. I. Vernardis, C. T. Goudar, and M. I. Klapa, "Metabolic profiling reveals that time related physiological changes in mammalian cell perfusion cultures are bioreactor scale independent," *Metabolic Engineering*, vol. 19, pp. 1–9, 2013.

[92] Z. Sheikholeslami, M. Jolicoeur, and O. Henry, "Elucidating the effects of postinduction glutamine feeding on the growth and productivity of cho cells," *Biotechnology Progress*, vol. 30, no. 3, pp. 535–546, 2014.

[93] ——, "The impact of the timing of induction on the metabolism and productivity of cho cells in culture," *Biochemical Engineering Journal*, vol. 79, pp. 162–171, 2013.

[94] J. D. Orth, I. Thiele, and B. Ø. Palsson, "What is flux balance analysis?" *Nature biotechnology*, vol. 28, no. 3, pp. 245–248, 2010.

[95] N. J. Kruger and R. G. Ratcliffe, "Fluxes through plant metabolic networks: measurements, predictions, insights and challenges," *Biochemical Journal*, vol. 465, pp. 27–38, 2015.

[96] K. Raman and N. Chandra, "Flux balance analysis of biological systems: applications and challenges," *Briefings in Bioinformatics*, vol. 10, no. 4, pp. 435–449, 2009.

[97] K. Schügerl and K.-H. Bellgardt, *Bioreaction Engineering - Modeling and Control.* Springer, 2000.

[98] J. V. l. J. N. l. G. Liden, *Bioreaction Engineering Principles*, third edition ed. Springer, 2011.

[99] H. P. J. Bonarius *et al.*, "Metabolic flux analysis of hybridoma cells in different culture media using mass balances," *Biotechnology and Bioengineering*, vol. 50, no. 3, pp. 299–318, 1996.

[100] M. I. Klapa *et al.*, "Metabolite and isotopomer balancing in the analysis of metabolic cycles: I. theory," *Biotechnology and Bioengineering*, vol. 62, no. 4, pp. 375–391, 1999.

[101] S. M. Park *et al.*, "Metabolite and isotopomer balancing in the analysis of metabolic cycles: Ii. applications," *Biotechnology and Bioengineering*, vol. 62, no. 4, pp. 392–401, 1999.

[102] A. Varma and B. O. Palsson, "Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type escherichia coli w3110," *Applied and environmental microbiology*, vol. 60, no. 10, pp. 3724–3731, 1994.

[103] M. R. Antoniewicz, J. K. Kelleher, and G. Stephanopoulos, "Determination of confidence intervals of metabolic fluxes estimated from stable isotope measurements," *Metabolic Engineering*, vol. 8, no. 4, pp. 324–337, 2006.

[104] C. Zupke and G. Stephanopoulos, "Intracellular flux analysis in hybridomas using mass balances and in vitro 13c nmr," *Biotechnology and Bioengineering*, vol. 45, no. 4, pp. 292–303, 1995.

[105] M. R. Antoniewicz, J. K. Kelleher, and G. Stephanopoulos, "Elementary metabolite units (emu): a novel framework for modeling isotopic distributions," *Metab Eng*, vol. 9, 2007. [Online]. Available: https://doi.org/10.1016/j.ymben.2006.09.001

[106] M. R. Antoniewicz, "Parallel labeling experiments for pathway elucidation and 13c metabolic flux analysis," *Current Opinion in Biotechnology*, vol. 36, pp. 91–97, 2015.

[107] T. Szyperski, "Biosynthetically directed fractional 13c-labeling of proteinogenic amino acids," *European Journal of Biochemistry*, vol. 232, no. 2, pp. 433–448, 1995.

[108] C. P. Long and M. R. Antoniewicz, "High-resolution (13)c metabolic flux analysis," *Nat Protoc*, vol. 14, no. 10, pp. 2856–2877, 2019. [Online]. Available: https://www.ncbi.nlm.nih.gov/pubmed/31471597

[109] K. Spagou *et al.*, "A gc-ms metabolic profiling study of plasma samples from mice on low- and high-fat diets," *Journal of Chromatography B*, vol. 879, no. 17, pp. 1467–1475, 2011.

[110] H. Kanani, P. K. Chrysanthopoulos, and M. I. Klapa, "Standardizing gc-ms metabolomics," *Journal of Chromatography B*, vol. 871, no. 2, pp. 191–201, 2008.

[111] H. H. Kanani and M. I. Klapa, "Data correction strategy for metabolomics analysis using gas chromatography-mass spectrometry," *Metabolic Engineering*, vol. 9, no. 1, pp. 39–51, 2007.

[112] M. P. Papadimitropoulos *et al.*, "Untargeted gc-ms metabolomics," *Methods Mol Biol*, vol. 1738, pp. 133–147, 2018.

[113] J. Choi and M. R. Antoniewicz, "Tandem mass spectrometry: A novel approach for metabolic flux analysis," *Metabolic Engineering*, vol. 13, no. 2, pp. 225–233, 2011.

[114] P. Millard *et al.*, "Scalaflux: A scalable approach to quantify fluxes in metabolic subnetworks," *PLOS Computational Biology*, vol. 16, no. 4, p. e1007799, 2020.

[115] J. S. Edwards and B. O. Palsson, "How will bioinformatics influence metabolic engineering?" *Biotechnology and Bioengineering*, vol. 58, no. 2-3, pp. 162–169, 1998.

[116] C. Wagner and R. Urbanczik, "The geometry of the flux cone of a metabolic network," *Biophysical journal*, vol. 89, no. 6, pp. 3837–3845, 2005.

[117] S. J. Jol *et al.*, "System-level insights into yeast metabolism by thermodynamic analysis of elementary flux modes," *PLOS Computational Biology*, vol. 8, no. 3, p. e1002415, 2012.

[118] M. W. Covert *et al.*, "Metabolic modeling of microbial strains in silico," *Trends Biochem Sci*, vol. 26, no. 3, pp. 179–86, 2001. [Online]. Available: https://www.ncbi.nlm.nih.gov/pubmed/11246024

[119] N. E. Lewis *et al.*, "Omic data from evolved e. coli are consistent with computed optimal growth from genome-scale models," *Mol Syst Biol*, vol. 6, no. 1, p. 390, 2010.

[120] A. Kibele *et al.*, "Stable, unstable and metastable states of equilibrium: Definitions and applications to human movement," *Journal of sports science & medicine*, vol. 14, no. 4, pp. 885–887, 2015.

[121] H. Qian, D. A. Beard, and S.-d. Liang, "Stoichiometric network theory for nonequilibrium biochemical systems," *European Journal of Biochemistry*, vol. 270, no. 3, pp. 415–421, 2003.

[122] Y. Demirel and S. I. Sandler, "Thermodynamics and bioenergetics," *Biophysical Chemistry*, vol. 97, no. 2, pp. 87–111, 2002.

[123] B. Schoepp-Cothenet *et al.*, "On the universal core of bioenergetics," *Biochim Biophys Acta*, vol. 1827, no. 2, pp. 79–93, 2013.

[124] E. Branscomb and M. J. Russell, "Turnstiles and bifurcators: the disequilibrium converting engines that put metabolism on the road," *Biochim Biophys Acta*, vol. 1827, no. 2, pp. 62–78, 2013.

[125] H. Qian, "Entropy production and excess entropy in a nonequilibrium steady-state of single macromolecules," *Phys Rev E Stat Nonlin Soft Matter Phys*, vol. 65, no. 2 Pt 1, p. 021111, 2002.

[126] M. D. Jankowski *et al.*, "Group contribution method for thermodynamic analysis of complex metabolic networks," *Biophys J*, vol. 95, no. 3, pp. 1487–99, 2008.

[127] M. L. Mavrovouniotis, "Group contributions for estimating standard gibbs energies of formation of biochemical compounds in aqueous solution," *Biotechnol Bioeng*, vol. 36, no. 10, pp. 1070–82, 1990.

[128] B. Du, D. C. Zielinski, and B. O. Palsson, "Estimating metabolic equilibrium constants: Progress and future challenges," *Trends Biochem Sci*, vol. 43, no. 12, pp. 960–969, 2018.

[129] R. A. Alberty, *Thermodynamics of Biochemical Reactions*. John Wiley & Sons, 2003.

[130] ——, "Calculation of standard transformed formation properties of biochemical reactants and standard apparent reduction potentials of half reactions," *Archives of Biochemistry and Biophysics*, vol. 358, no. 1, pp. 25–39, 1998.

[131] R. K. Thauer, "Biochemistry of methanogenesis: a tribute to marjory stephenson. 1998 marjory stephenson prize lecture," *Microbiology*, vol. 144 (Pt 9), pp. 2377–406, 1998.

[132] R. K. Thauer, K. Jungermann, and K. Decker, "Energy conservation in chemotrophic anaerobic bacteria," *Bacteriological reviews*, vol. 41, no. 1, pp. 100–180, 1977.

[133] J. Dolfing and B. K. Harrison, "Gibbs free energy of formation of halogenated aromatic compounds and their potential role as electron acceptors in anaerobic environments," *Environmental Science & Technology*, vol. 26, no. 11, pp. 2213–2218, 1992.

[134] J. Dolfing and D. B. Janssen, "Estimates of gibbs free energies of formation of chlorinated aliphatic compounds," *Biodegradation*, vol. 5, no. 1, pp. 21–28, 1994.

[135] C. S. Henry, L. J. Broadbelt, and V. Hatzimanikatis, "Thermodynamics-based metabolic flux analysis," *Biophys J*, vol. 92, no. 5, pp. 1792–805, 2007.

[136] R. A. Alberty, "Equilibrium compositions of solutions of biochemical species and heats of biochemical reactions," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 88, no. 8, pp. 3268–3271, 1991.

[137] C. Zerfass, M. Asally, and O. S. Soyer, "Interrogating metabolism as an electron flow system," *Curr Opin Syst Biol*, vol. 13, pp. 59–67, 2019.

[138] J. J. Hamilton, V. Dwivedi, and J. L. Reed, "Quantitative assessment of thermodynamic constraints on the solution space of genome-scale metabolic models," *Biophys J*, vol. 105, no. 2, pp. 512–22, 2013.

[139] B. Du *et al.*, "Temperature-dependent estimation of gibbs energies using an updated group-contribution method," *Biophys J*, vol. 114, no. 11, pp. 2691–2702, 2018.

[140] E. Noor *et al.*, "An integrated open framework for thermodynamics of reactions that combines accuracy and coverage," *Bioinformatics*, vol. 28, no. 15, pp. 2037–44, 2012.

[141] D. A. Beard, S.-d. Liang, and H. Qian, "Energy balance for analysis of complex metabolic networks," *Biophysical Journal*, vol. 83, no. 1, pp. 79–86, 2002.

[142] H. Qian, D. A. Beard, and S.-d. Liang, "Stoichiometric network theory for nonequilibrium biochemical systems," *European Journal of Biochemistry*, vol. 270, no. 3, pp. 415–421, 2003.

[143] A. Kummel, S. Panke, and M. Heinemann, "Putative regulatory sites unraveled by network-embedded thermodynamic analysis of metabolome data," *Mol Syst Biol*, vol. 2, p. 2006 0034, 2006.

[144] C. S. Henry, L. J. Broadbelt, and V. Hatzimanikatis, "Thermodynamics-based metabolic flux analysis," *Biophys J*, vol. 92, no. 5, pp. 1792–805, 2007.

[145] J. L. Reed *et al.*, "An expanded genome-scale model of escherichia coli k-12 (ijr904 gsm/gpr)," *Genome Biology*, vol. 4, no. 9, p. R54, 2003.

[146] S. Peres *et al.*, "How important is thermodynamics for identifying elementary flux modes?" *PLoS One*, vol. 12, no. 2, p. e0171440, 2017.

[147] M. P. Gerstl, C. Jungreuthmayer, and J. Zanghellini, "tefma: computing thermodynamically feasible elementary flux modes in metabolic networks," *Bioinformatics*, vol. 31, no. 13, pp. 2232–2234, 2015.

[148] Y. Fan, D. Ley, and M. R. Andersen, "Fed-batch cho cell culture for lab-scale antibody production," *Methods Mol Biol*, vol. 1674, pp. 147–161, 2018.

[149] S. Kyriakopoulos *et al.*, "Kinetic modeling of mammalian cell culture bioprocessing: The quest to advance biomanufacturing," *Biotechnology Journal*, vol. 13, no. 3, p. 1700229, 2018.

[150] H. M. Sauro, "Enzyme kinetics for systems biology," *Ambrosius Publishing*, 2011.

[151] L. Liu and A. Bockmayr, "Formalizing metabolic-regulatory networks by hybrid automata," *Acta Biotheoretica*, vol. 68, no. 1, pp. 73–85, 2020.

[152] ——, "Regulatory dynamic enzyme-cost flux balance analysis: A unifying framework for constraint-based modeling," *Journal of Theoretical Biology*, vol. 501, p. 110317, 2020.

[153] J. Strutz *et al.*, "Metabolic kinetic modeling provides insight into complex biological questions, but hurdles remain," *Current Opinion in Biotechnology*, vol. 59, pp. 24–30, 2019.

[154] M. Ederer and E. D. Gilles, "Thermodynamically feasible kinetic models of reaction networks," *Biophysical Journal*, vol. 92, no. 6, pp. 1846–1857, 2007.

[155] M. Cloutier *et al.*, "Kinetic metabolic modelling for the control of plant cells cytoplasmic phosphate," *Journal of theoretical biology*, vol. 259, no. 1, pp. 118–131, 2009.

[156] R. Julien *et al.*, "A single dynamic metabolic model can describe mab producing cho cell batch and fed-batch cultures on different culture media," *PloS One*, vol. 10, no. 9, 2015.

[157] O. Henry, A. Kamen, and M. Perrier, "Monitoring the physiological state of mammalian cell perfusion processes by on-line estimation of intracellular fluxes," *Journal of Process Control*, vol. 17, no. 3, pp. 241–251, 2007.

[158] G. Atefeh, H. Olivier, and J. Mario, "An in-silico study of the regulation of cho cells glycolysis," *Journal of Theoretical Biology*, vol. 357, no. C, pp. 112–122, 2014.

[159] J. Almquist *et al.*, "Kinetic models in industrial biotechnology - improving cell factory performance," *Metabolic Engineering*, vol. 24, pp. 38–60, 2014.

[160] K. Tummler *et al.*, "New types of experimental data shape the use of enzyme kinetics for dynamic network modeling," *FEBS Journal*, vol. 281, no. 2, pp. 549–571, 2014.

[161] D. Visser and J. J. Heijnen, "Dynamic simulation and metabolic re-design of a branched pathway using linlog kinetics," *Metabolic Engineering*, vol. 5, no. 3, pp. 164–176, 2003.

[162] J. J. Heijnen, "Approximative kinetic formats used in metabolic network modeling," *BIOTECHNOLOGY AND BIOENGINEERING*, vol. 91, no. 5, pp. 534–545, 2005.

[163] K. Sriyudthsak, F. Shiraishi, and M. Y. Hirai, "Mathematical modeling and dynamic simulation of metabolic reaction systems using metabolome time series data," *Frontiers in Molecular Biosciences U6*, vol. 3, 2016.

[164] F. Horn and R. Jackson, "General mass action kinetics," *Archive for rational mechanics and analysis*, vol. 47, no. 2, pp. 81–116, 1972.

[165] V. Chellaboina *et al.*, "Modeling and analysis of mass-action kinetics," *IEEE Control Systems*, vol. 29, no. 4, 2009.

[166] A. Dräger *et al.*, "Modeling metabolic networks in c. glutamicum: a comparison of rate laws in combination with various parameter optimization strategies," *BMC Systems Biology*, vol. 3, 2009.

[167] D. Hamby, "A review of techniques for parameter sensitivity analysis of environmental models," *Environmental monitoring and assessment*, vol. 32, no. 2, pp. 135–154, 1994.

[168] P. Cazzaniga *et al.*, "Computational strategies for a system-level understanding of metabolism," *Metabolites*, vol. 4, no. 4, 2014.

[169] J. Di Maggio, J. C. Diaz Ricci, and M. S. Diaz, "Global sensitivity analysis in dynamic metabolic networks," *Computers & Chemical Engineering*, vol. 34, no. 5, pp. 770–781, 2010.

[170] K. Chan, E. M. Scott, and A. Saltelli, *Sensitivity analysis: edited by Andrea Saltelli, Karen Chan, E. Marian Scott*. Chichester, England;Toronto;: Wiley, 2000.

[171] H. Kacser and J. A. Burns, "The control of flux," *Symp Soc Exp Biol*, vol. 27, pp. 65–104, 1973.

[172] R. Heinrich and T. A. Rapoport, "A linear steady-state treatment of enzymatic chains," *European Journal of Biochemistry*, vol. 42, no. 1, pp. 97–105, 1974.

[173] D. Visser and J. J. Heijnen, "The mathematics of metabolic control analysis revisited," *Metabolic Engineering*, vol. 4, no. 2, pp. 114–123, 2002.

[174] D. A. Fell, "Metabolic control analysis: a survey of its theoretical and experimental development," *Biochemical Journal*, vol. 286, no. Pt 2, p. 313, 1992.

[175] V. Hatzimanikatis, C. A. Floudas, and J. E. Bailey, "Analysis and design of metabolic reaction networks via mixed-integer linear optimization," *AIChE Journal*, vol. 42, no. 5, pp. 1277–1292, 1996.

[176] M. A. Savageau, "Biochemical systems analysis. ii. the steady-state solutions for an n-pool system using a power-law approximation," *Journal of theoretical biology*, vol. 25, no. 3, p. 370, 1969.

[177] ——, "Biochemical systems analysis: I. some mathematical properties of the rate law for the component enzymatic reactions," *Journal of theoretical biology*, vol. 25, no. 3, pp. 365–369, 1969.

[178] E. O. Voit, "Biochemical systems theory: A review," *ISRN Biomathematics*, vol. 2013, pp. 1–53, 2013.

[179] F.-S. Wang, C.-L. Ko, and E. O. Voit, "Kinetic modeling using s-systems and lin-log approaches," *Biochemical Engineering Journal*, vol. 33, no. 3, pp. 238–247, 2007.

[180] V. Hatzimanikatis and J. E. Bailey, "Mca has more to say," *J Theor Biol*, vol. 182, no. 3, pp. 233–42, 1996.

[181] E. O. Voit, "The best models of metabolism," *Wiley interdisciplinary reviews. Systems biology and medicine*, vol. 9, no. 6, p. 10.1002/wsbm.1391, 2017.

[182] M. Shuler and F. Kargı, *Bioprocess engineering: basic concepts*, ser. Prentice-Hall international series in the physical and chemical engineering sciences. Prentice Hall, 1992.

[183] M. Audagnotto and M. Dal Peraro, "Protein post-translational modifications: In silico prediction tools and molecular modeling," *Computational and structural biotechnology journal*, vol. 15, pp. 307–319, 2017.

[184] R. P. Nolan and K. Lee, "Dynamic model of cho cell metabolism," vol. 13, no. 1, pp. 108–124, 2011. [Online]. Available: https://dx.doi.org/10.1016/j.ymben.2010.09.003

[185] O. Bozovic *et al.*, "Real-time observation of ligand-induced allosteric transitions in a pdz domain," *Proc Natl Acad Sci U S A*, vol. 117, no. 42, pp. 26 031–26 039, 2020.

[186] A. Goldbeter and S. R. Caplan, "Oscillatory enzymes," *Annual Review of Biophysics and Bioengineering*, vol. 5, no. 1, pp. 449–476, 1976.

[187] M. Alaka *et al.*, "The cumate gene-switch: a system for regulated expression in mammalian cells," *BMC Biotechnology*, vol. 6, 2006.

[188] B. Korzeniewski and P. Liguzinski, "Theoretical studies on the regulation of anaerobic glycolysis and its influence on oxidative phosphorylation in skeletal muscle," *Biophysical chemistry*, vol. 110, no. 1, pp. 147–169, 2004.

[189] M. W. Covert *et al.*, "Integrating metabolic, transcriptional regulatory and signal transduction models in escherichia coli," *Bioinformatics*, vol. 24, no. 18, pp. 2044–2050, 2008.

[190] S. M. Fendt *et al.*, "Tradeoff between enzyme and metabolite efficiency maintains metabolic homeostasis upon perturbations in enzyme capacity," *Mol Syst Biol*, vol. 6, p. 356, 2010.

[191] H. V. Westerhoff *et al.*, "Thermodynamics of complexity - the live cell," *Thermochimica Acta*, vol. 309, no. 1-2, pp. 111–120, 1998.

[192] Y. Arkun and M. Yasemi, "Dynamics and control of the erk signaling pathway: Sensitivity, bistability, and oscillations," *PLoS One*, vol. 13, no. 4, p. e0195513, 2018.

[193] K. A. Johnson and R. S. Goody, "The original michaelis constant: Translation of the 1913 michaelis-menten paper," *Biochemistry*, vol. 50, no. 39, pp. 8264–8269, 2011.

[194] V. Henri, "Théorie générale de l'action de quelques diastases par victor henri [c. r. acad. sci. paris 135 (1902) 916-919]," *Comptes rendus biologies*, vol. 329, no. 1, pp. 47–50, 2006.

[195] A. Cornish-Bowden, "One hundred years of michaelis-menten kinetics," *Perspectives in Science*, vol. 4, pp. 3–9, 2015.

[196] Bajzer and E. E. Strehler, "About and beyond the henri-michaelis-menten rate equation for single-substrate enzyme kinetics," *Biochemical and Biophysical Research Communications*, vol. 417, no. 3, pp. 982–985, 2012. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0006291X11022601

[197] X. Ren *et al.*, "A kinetic metabolic study of lipid production in chlorella protothecoides under heterotrophic condition," *Microbial Cell Factories*, vol. 18, no. 1, p. 113, 2019.

[198] I. Schomburg *et al.*, "Brenda, the enzyme database: updates and major new developments," *Nucleic Acids Res*, vol. 32, no. Database issue, pp. D431–3, 2004.

[199] R. N. Goldberg, Y. B. Tewari, and T. N. Bhat, "Thermodynamics of enzyme-catalyzed reactions- a database for quantitative biochemistry," *Bioinformatics*, vol. 20, no. 16, pp. 2874–2877, 2004. [Online]. Available: https://doi.org/10.1093/bioinformatics/bth314

[200] R. Caspi *et al.*, "Metacyc: a multiorganism database of metabolic pathways and enzymes," *Nucleic Acids Research*, vol. 34, no. $\text{suppl}_1, pp.D511 - -D516, 2006$.

[201] C. Khim Chong *et al.*, "A review on modelling methods, pathway simulation software and recent development on differential evolution algorithms for metabolic pathways in systems biology," *Current Bioinformatics*, vol. 9, no. 5, pp. 509–521, 2014.

[202] P. K. Polisetty, E. O. Voit, and E. P. Gatzke, "Identification of metabolic system parameters using global optimization methods," *Theoretical Biology and Medical Modelling*, vol. 3, no. 1, p. 4, 2006.

[203] F. Reali, C. Priami, and L. Marchetti, "Optimization algorithms for computational systems biology," *Frontiers in Applied Mathematics and Statistics*, vol. 3, 2017.

[204] M. Cavazzuti, *Optimization methods: from theory to design scientific and technological aspects in mechanics*. Springer Science & Business Media, 2012.

[205] C. A. Floudas, *Deterministic global optimization: theory, methods and applications*. Springer Science & Business Media, 2013, vol. 37.

[206] V. Dominique, L. Filip, and I. Jan Van, "Dynamic estimation of specific fluxes in metabolic networks using non-linear dynamic optimization," *BMC Systems Biology*, vol. 8, 2014. [Online]. Available: https://bmcsystbiol.biomedcentral.com/track/pdf/10.1186/s12918-014-0132-0?site=bmcsystbiol.biomedcentral.com

[207] S. Gopalakrishnan, S. Dash, and C. Maranas, "K-fit: An accelerated kinetic parameterization algorithm using steady-state fluxomic data," *Metab Eng*, vol. 61, pp. 197–205, 2020.

[208] T. Lubitz and W. Liebermeister, "Parameter balancing: consistent parameter sets for kinetic metabolic models," *Bioinformatics*, vol. 35, no. 19, pp. 3857–3858, 2019.

[209] R. Horst and H. Tuy, *Global optimization: Deterministic approaches*. Springer Science & Business Media, 2013.

[210] R. L. Dutton, J. M. Scharer, and M. Moo-Young, "Descriptive parameter evaluation in mammalian cell culture," *Cytotechnology*, vol. 26, no. 2, pp. 139–52, 1998.

[211] I. C. Chou and E. O. Voit, "Recent developments in parameter estimation and structure identification of biochemical and genomic systems," *Mathematical Biosciences*, vol. 219, no. 2, pp. 57–83, 2009.

[212] Z. Kutalik, V. Moulton, and W. Tucker, "S-system parameter estimation for noisy metabolic profiles using newton-flow analysis," *IET Systems Biology*, vol. 1, no. 3, pp. 174–180, 2007.

[213] C. A. Floudas, *Deterministic global optimization: theory, methods and applications*. Springer Science & Business Media, 2013, vol. 37.

[214] E. O. Voit *et al.*, "Estimation of metabolic pathway systems from different data sources," *IET Systems Biology*, vol. 3, no. 6, pp. 513–522, 2009.

[215] D. R. Penas *et al.*, "Parameter estimation in large-scale systems biology models: a parallel and self-adaptive cooperative strategy," *BMC Bioinformatics*, vol. 18, no. 1, p. 52, 2017. [Online]. Available: https://doi.org/10.1186/s12859-016-1452-4

[216] A. Khodayari and C. D. Maranas, "A genome-scale escherichia coli kinetic metabolic model k-ecoli457 satisfying flux data for multiple mutant strains," *Nature Communications*, vol. 7, no. 1, p. 13806, 2016. [Online]. Available: https://doi.org/10.1038/ncomms13806https://www.nature.com/articles/ncomms13806.pdf

[217] Y. Tan *et al.*, "Reducing the allowable kinetic space by constructing ensemble of dynamic models with the same steady-state flux," *Metabolic Engineering*, vol. 13, no. 1, pp. 60–75, 2011.

[218] C. S. Henry *et al.*, "Genome-scale thermodynamic analysis of escherichia coli metabolism," *Biophys J*, vol. 90, no. 4, pp. 1453–61, 2006.

[219] J. L. Greene *et al.*, "Acceleration strategies to enhance metabolic ensemble modeling performance," *Biophysical Journal*, vol. 113, no. 5, pp. 1150–1162, 2017.

[220] A. Dräger *et al.*, "Benchmarking evolutionary algorithms on convenience kinetics models of the valine and leucine biosynthesis in c. glutamicum," in *Evolutionary Computation, 2007. CEC 2007. IEEE Congress on*. IEEE, 2007, Conference Proceedings, pp. 896–903.

[221] C. Spieth, N. Hassis, and F. Streichert, "Comparing mathematical models on the problem of network inference," in *Proceedings of the 8th annual conference on Genetic and evolutionary computation*. ACM, 2006, Conference Proceedings, pp. 279–286.

[222] A. Salman, A. P. Engelbrecht, and M. G. Omran, "Empirical analysis of self-adaptive differential evolution," *European Journal of operational research*, vol. 183, no. 2, pp. 785–804, 2007.

[223] R. Thangaraj, M. Pant, and A. Abraham, "A simple adaptive differential evolution algorithm," in *Nature & Biologically Inspired Computing, 2009. NaBIC 2009. World Congress on*. IEEE, 2009, Conference Proceedings, pp. 457–462.

[224] L. Feng, Y.-F. Yang, and Y.-X. Wang, "A new approach to adapting control parameters in differential evolution algorithm," *Simulated Evolution and Learning*, pp. 21–30, 2008.

[225] M. W. Covert *et al.*, "Integrating high-throughput and computational data elucidates bacterial networks," *Nature*, vol. 429, no. 6987, pp. 92–96, 2004.

[226] M. J. Herrgård *et al.*, "Integrated analysis of regulatory and metabolic networks reveals novel regulatory mechanisms in saccharomyces cerevisiae," *Genome research*, vol. 16, no. 5, pp. 627–635, 2006.

[227] H. C. Yeo *et al.*, "Enzyme capacity-based genome scale modelling of cho cells," *Metabolic Engineering*, vol. 60, pp. 138–147, 2020.

[228] B. Laflaquiere *et al.*, "Identifying biomarkers of wharton's jelly mesenchymal stromal cells using a dynamic metabolic model: The cell passage effect," *Metabolites*, vol. 8, no. 1, 2018.

[229] M. Zampieri *et al.*, "Frontiers of high-throughput metabolomics," *Curr Opin Chem Biol*, vol. 36, pp. 15–23, 2017.

[230] P. Lakrisenko and D. Weindl, "Dynamic models for metabolomics data integration," *Current Opinion in Systems Biology*, vol. 28, p. 100358, 2021.

[231] M. Yasemi and M. Jolicoeur, "Modelling cell metabolism: A review on constraint-based steady-state and kinetic approaches," *Processes*, vol. 9, no. 2, 2021.

[232] G. Paglia *et al.*, "Biomarkers defining the metabolic age of red blood cells during cold storage," *Blood*, vol. 128, no. 13, pp. e43–50, 2016.

[233] J. D. Roback *et al.*, "Metabolomics of adsol (as-1) red blood cell storage," *Transfus Med Rev*, vol. 28, no. 2, pp. 41–55, 2014.

[234] M. Bardyn *et al.*, "Restoration of physiological levels of uric acid and ascorbic acid reroutes the metabolism of stored red blood cells," *Metabolites*, vol. 10, no. 6, 2020.

[235] ——, "The antioxidant capacity of erythrocyte concentrates is increased during the first week of storage and correlated with the uric acid level," *Vox Sang*, vol. 112, no. 7, pp. 638–647, 2017.

[236] A. Bordbar, N. Jamshidi, and B. O. Palsson, "iab-rbc-283: A proteomically derived knowledge-base of erythrocyte metabolism that can be used to simulate its physiological and patho-physiological states," *BMC Syst Biol*, vol. 5, no. 1, p. 110, 2011.

[237] M. Bardyn *et al.*, "Red blood cells ageing markers: a multi-parametric analysis," *Blood Transfus*, vol. 15, no. 3, pp. 239–248, 2017.

[238] T. Yoshida, M. Prudent, and A. D'Alessandro, "Red blood cell storage lesion: causes and potential clinical consequences," *Blood transfusion = Trasfusione del sangue*, vol. 17, no. 1, pp. 27–52, 2019.

[239] I. Thiele and B. O. Palsson, "A protocol for generating a high-quality genome-scale metabolic reconstruction," *NATURE PROTOCOLS*, vol. 5, no. 1, pp. 93–121, 2010.

[240] R. Mahadevan and C. H. Schilling, "The effects of alternate optimal solutions in constraint-based genome-scale metabolic models," *Metab Eng*, vol. 5, no. 4, pp. 264–76, 2003.

[241] A. L. Meadows *et al.*, "Application of dynamic flux balance analysis to an industrial escherichia coli fermentation," *Metabolic engineering*, vol. 12, no. 2, pp. 150–160, 2010.

[242] R. P. Nolan and K. Lee, "Dynamic model for cho cell engineering," *J Biotechnol*, vol. 158, no. 1-2, pp. 24–33, 2012.

[243] S. J. Moon *et al.*, "Oxidative pentose phosphate pathway and glucose anaplerosis support maintenance of mitochondrial nadph pool under mitochondrial oxidative stress," *Bioeng Transl Med*, vol. 5, no. 3, p. e10184, 2020.

[244] D. Christodoulou *et al.*, "Reserve flux capacity in the pentose phosphate pathway enables escherichia coli's rapid response to oxidative stress," *Cell Syst*, vol. 6, no. 5, pp. 569–578 e7, 2018.

[245] C. J. Edwards and J. Fuller, "Oxidative stress in erythrocytes," *Comparative Haematology International*, vol. 6, no. 1, pp. 24–31, 1996.

[246] M. Bardyn, J. D. Tissot, and M. Prudent, "Oxidative stress and antioxidant defenses during blood processing and storage of erythrocyte concentrates," *Transfus Clin Biol*, vol. 25, no. 1, pp. 96–100, 2018.

[247] H. Kitano *et al.*, "Using process diagrams for the graphical representation of biological networks," *Nat Biotechnol*, vol. 23, no. 8, pp. 961–6, 2005.

[248] A. Ghorbaniaghdam *et al.*, "Analyzing clonal variation of monoclonal antibody-producing cho cell lines using an in silico metabolomic platform," *PLOS ONE*, vol. 9, no. 3, p. e90832, 2014.

[249] J. Y. Kim, Y. G. Kim, and G. M. Lee, "Cho cells in biotechnology for production of recombinant proteins: current state and further potential," *Appl Microbiol Biotechnol*, vol. 93, no. 3, pp. 917–30, 2012.

[250] J. Gašperšič *et al.*, "Metabolic network modelling of chinese hamster ovary (cho) culture bioreactors operated as microbial cell factories," *Acta Chimica Slovenica*, vol. 65, no. 4, pp. 769–786, 2018.

[251] S. Sha *et al.*, "Mechanistic modeling and applications for cho cell culture development and production," *Current Opinion in Chemical Engineering*, vol. 22, pp. 54–61, 2018.

[252] J. Schellenberger *et al.*, "Quantitative prediction of cellular metabolism with constraint-based models: the cobra toolbox v2. 0," *Nature protocols*, vol. 6, no. 9, p. 1290, 2011.

[253] M. Yasemi and M. Jolicoeur, "Modelling cell metabolism: A review on constraint-based steady-state and kinetic approaches," *Processes*, vol. 9, no. 2, 2021.

[254] L. E. Quek *et al.*, "Metabolic flux analysis in mammalian cell culture," *Metab Eng*, vol. 12, no. 2, pp. 161–71, 2010.

[255] J. M. Savinell and B. O. Palsson, "Optimal selection of metabolic fluxes for in vivo measurement. ii. application to escherichia coli and hybridoma cell metabolism," *Journal of Theoretical Biology*, vol. 155, no. 2, pp. 215–242, 1992.

[256] C. Zupke and G. Stephanopoulos, "Modeling of isotope distributions and intracellular fluxes in metabolic networks using atom mapping matrixes," *Biotechnology Progress*, vol. 10, no. 5, pp. 489–498, 1994.

[257] L. Z. Xie and D. I. C. Wang, "Material balance studies on animal cell metabolism using a stoichiometrically based reaction network," *Biotechnology and Bioengineering*, vol. 52, no. 5, pp. 579–590, 1996.

[258] G. B. Nyberg *et al.*, "Metabolism of peptide amino acids by chinese hamster ovary cells grown in a complex medium," *Biotechnol Bioeng*, vol. 62, no. 3, pp. 324–35, 1999.

[259] C. Altamirano *et al.*, "Analysis of cho cells metabolic redistribution in a glutamate-based defined medium in continuous culture," *Biotechnology Progress*, vol. 17, no. 6, pp. 1032–1041, 2001.

[260] ——, "Considerations on the lactate consumption by cho cells in the presence of galactose," *Journal of Biotechnology*, vol. 125, no. 4, pp. 547–556, 2006.

[261] S. Fernandes de Sousa *et al.*, "Dynamic metabolic flux analysis using a convex analysis approach: Application to hybridoma cell cultures in perfusion," *Biotechnol Bioeng*, vol. 113, no. 5, pp. 1102–12, 2016.

[262] S. Hammond *et al.*, "Genomic sequencing and analysis of a chinese hamster ovary cell line using illumina sequencing technology," *BMC Genomics*, vol. 12, p. 67, 2011.

[263] D. Széliová *et al.*, "What cho is made of: Variations in the biomass composition of chinese hamster ovary cell lines," *Metabolic Engineering*, vol. 61, pp. 288–300, 2020.

[264] M. Ataman *et al.*, "redgem: Systematic reduction and analysis of genome-scale metabolic reconstructions for development of consistent core metabolic models," *PLoS Comput Biol*, vol. 13, no. 7, p. e1005444, 2017.

[265] P. Erdrich, R. Steuer, and S. Klamt, "An algorithm for the reduction of genome-scale metabolic network models to meaningful core models," *BMC Syst Biol*, vol. 9, p. 48, 2015.

[266] D. J. Lugar, S. G. Mack, and G. Sriram, "Netred, an algorithm to reduce genome-scale metabolic networks and facilitate the analysis of flux predictions," *Metab Eng*, vol. 65, pp. 207–222, 2021.

[267] S. A. Becker and B. O. Palsson, "Context-specific metabolic networks are consistent with experiments," *PLoS Comput Biol*, vol. 4, no. 5, p. e1000082, 2008.

[268] M. P. Pacheco and T. Sauter, "The fastcore family: For the fast reconstruction of compact context-specific metabolic networks models," *Methods Mol Biol*, vol. 1716, pp. 101–110, 2018.

[269] A. Richelle *et al.*, "Increasing consensus of context-specific metabolic models by integrating data-inferred cell functions," *PLoS Comput Biol*, vol. 15, no. 4, p. e1006867, 2019.

[270] M. Tefagh and S. P. Boyd, "Swiftcore: a tool for the context-specific reconstruction of genome-scale metabolic networks," *BMC Bioinformatics*, vol. 21, no. 1, p. 140, 2020.

[271] M. Aoyama *et al.*, "Effects of amino acid substitutions on the biological activity of anti-cd20 monoclonal antibody produced by transgenic silkworms (bombyx mori)," *Biochem Biophys Res Commun*, vol. 503, no. 4, pp. 2633–2638, 2018.

[272] [Online]. Available: https://go.drugbank.com/drugs/DB00073

[273] D. Szeliova *et al.*, "Inclusion of maintenance energy improves the intracellular flux predictions of cho," *PLoS Comput Biol*, vol. 17, no. 6, p. e1009022, 2021.

[274] K. Sheikh, J. Forster, and L. K. Nielsen, "Modeling hybridoma cell metabolism using a generic genome-scale metabolic model of mus musculus," *Biotechnol Prog*, vol. 21, no. 1, pp. 112–21, 2005.

[275] Y. Chen *et al.*, "An unconventional uptake rate objective function approach enhances applicability of genome-scale models for mammalian cells," *npj Systems Biology and Applications*, vol. 5, no. 1, p. 25, 2019.

[276] J. E. Cuthrell and L. T. Biegler, "On the optimization of differential-algebraic process systems," *AIChE Journal*, vol. 33, no. 8, pp. 1257–1270, 1987.

[277] A. Gábor and J. R. Banga, "Robust and efficient parameter estimation in dynamic models of biological systems," *BMC systems biology*, vol. 9, pp. 74–74, 2015.

[278] A. M. Feist and B. O. Palsson, "The biomass objective function," *Current Opinion in Microbiology*, vol. 13, no. 3, pp. 344–349, 2010.

[279] R. Schuetz *et al.*, "Multidimensional optimality of microbial metabolism," *Science*, vol. 336, no. 6081, pp. 601–4, 2012.

[280] J. Robitaille, J. Chen, and M. Jolicoeur, "A single dynamic metabolic model can describe mab producing cho cell batch and fed-batch cultures on different culture media," *PLoS One*, vol. 10, no. 9, p. e0136815, 2015.

[281] V. Acuña *et al.*, "A note on the complexity of finding and enumerating elementary modes," *BioSystems*, vol. 99, no. 3, pp. 210–214, 2010.

[282] C. Jungreuthmayer and J. Zanghellini, "Designing optimal cell factories: Integer programing couples elementary mode analysis with regulation," *BMC Systems Biology*, vol. 6, no. 1, p. 103, 2012.

[283] J. Stelling *et al.*, "Metabolic network structure determines key aspects of functionality and regulation," *Nature*, vol. 420, no. 6912, pp. 190–193, 2002.

[284] J. Behre *et al.*, "Structural robustness of metabolic networks with respect to multiple knockouts," *Journal of Theoretical Biology*, vol. 252, no. 3, pp. 433–441, 2008.

[285] S. Klamt, J. Saez-Rodriguez, and E. D. Gilles, "Structural and functional analysis of cellular networks with cellnetanalyzer," *BMC SYSTEMS BIOLOGY*, vol. 1, no. 1, pp. 2–2, 2007.

[286] S. Hoops *et al.*, "Copasi—a complex pathway simulator," *Bioinformatics*, vol. 22, no. 24, pp. 3067–3074, 2006.

[287] M. Terzer and J. Stelling, "Large-scale computation of elementary flux modes with bit pattern trees," *Bioinformatics*, vol. 24, no. 19, pp. 2229–2235, 2008.

[288] R. Schwarz *et al.*, "Integrated network reconstruction, visualization and analysis using yanasquare," *BMC bioinformatics*, vol. 8, no. 1, p. 313, 2007.

[289] C. H. Schilling, D. Letscher, and B. Ø. Palsson, "Theory for the systemic definition of metabolic pathways and their use in interpreting metabolic function from a pathway-oriented perspective," *Journal of Theoretical Biology*, vol. 203, no. 3, pp. 229–248, 2000.

[290] S. L. Bell and B. Ø. Palsson, "Expa: a program for calculating extreme pathways in biochemical reaction networks," *Bioinformatics*, vol. 21, no. 8, pp. 1739–1740, 2004.

[291] J. M. Savinell and B. O. Palsson, "Network analysis of intermediary metabolism using linear optimization. i. development of mathematical formalism," *Journal of theoretical biology*, vol. 154, no. 4, pp. 421–454, 1992.

[292] D. A. Fell and J. R. Small, "Fat synthesis in adipose tissue. an examination of stoichiometric constraints," *Biochemical Journal*, vol. 238, no. 3, pp. 781–786, 1986.

[293] H. G. Holzhutter, "The principle of flux minimization and its application to estimate stationary fluxes in metabolic networks," *Eur J Biochem*, vol. 271, no. 14, pp. 2905–22, 2004.

[294] J. S. Edwards, R. U. Ibarra, and B. O. Palsson, "In silico predictions of escherichia coli metabolic capabilities are consistent with experimental data," *Nature biotechnology*, vol. 19, no. 2, 2001.

[295] A. Ghorbaniagham, "Development of a dynamic model to describe cho cells metabolic network and regulation," 2013.

[296] S. Waldherr, D. A. Oyarzún, and A. Bockmayr, "Dynamic optimization of metabolic networks coupled with gene expression," *Journal of Theoretical Biology*, vol. 365, pp. 469–485, 2015.

[297] A. Goelzer, V. Fromion, and G. Scorletti, "Cell design in bacteria as a convex optimization problem," *Automatica*, vol. 47, no. 6, pp. 1210–1218, 2011. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0005109811001397

[298] G. Jeanne *et al.*, "Dynamical resource allocation models for bioreactor optimization," *IFAC-PapersOnLine*, vol. 51, no. 19, pp. 20–23, 2018.

[299] M. W. Covert, C. H. Schilling, and B. Palsson, "Regulation of gene expression in flux balance models of metabolism," *Journal of Theoretical Biology*, vol. 213, no. 1, pp. 73–88, 2001.

[300] M. W. Covert and B. O. Palsson, "Transcriptional regulation in constraints-based metabolic models of escherichia coli," *J Biol Chem*, vol. 277, no. 31, pp. 28 058–64, 2002.

[301] T. Shlomi *et al.*, "A genome-scale computational study of the interplay between transcriptional regulation and metabolism," *Molecular systems biology*, vol. 3, pp. 101–101, 2007.

[302] S. Chandrasekaran and N. D. Price, "Probabilistic integrative modeling of genome-scale metabolic and regulatory networks in escherichia coli and mycobacterium tuberculosis," *Proceedings of the National Academy of Sciences*, vol. 107, no. 41, p. 17845, 2010.

[303] M. W. Covert *et al.*, "Integrating metabolic, transcriptional regulatory and signal transduction models in escherichia coli," *Bioinformatics*, vol. 24, no. 18, pp. 2044–2050, 2008. [Online]. Available: https://doi.org/10.1093/bioinformatics/btn352

[304] J. Min Lee *et al.*, "Dynamic analysis of integrated signaling, metabolic, and regulatory networks," *PLOS Computational Biology*, vol. 4, no. 5, p. e1000086, 2008.

[305] D.-Y. Lee *et al.*, "Metafluxnet: the management of metabolic reaction information and quantitative metabolic flux analysis," *Bioinformatics*, vol. 19, no. 16, pp. 2144–2146, 2003.

[306] N. Zamboni, E. Fischer, and U. Sauer, "Fiatflux – a software for metabolic flux analysis from 13c-glucose experiments," *BMC Bioinformatics*, vol. 6, no. 1, p. 209, 2005.

[307] L.-E. Quek *et al.*, "Openflux: efficient modelling software for 13c-based metabolic flux analysis," *Microbial Cell Factories*, vol. 8, no. 1, p. 25, 2009.

[308] J. Schellenberger *et al.*, "Quantitative prediction of cellular metabolism with constraint-based models: the cobra toolbox v2. 0," *Nature protocols*, vol. 6, no. 9, p. 1290, 2011.

[309] A. Hoppe *et al.*, "Fasimu: flexible software for flux-balance computation series in large metabolic networks," *BMC bioinformatics*, vol. 12, no. 1, p. 28, 2011.

[310] A. Flamholz *et al.*, "equilibrator—the biochemical thermodynamics calculator," *Nucleic Acids Research*, vol. 40, no. D1, pp. D770–D775, 2011.

[311] C. W. Bale *et al.*, "Factsage thermochemical software and databases, 2010–2016," *Calphad*, vol. 54, pp. 35–53, 2016.

APPENDIX A TEXT S1 : MATHEMATICAL FORMULATIONS OF CONSTRAINT-BASED MODELS

Text S1: Mathematical formulations of constraint-based models

Stoichiometric matrix formulation

The stoichiometric matrix for a metabolic network with m metabolites and n reactions is as follows:

$$S_{m\times n} = \left(s_{i,j}\right)_{m\times n} \{i = 1, ..., m | j = 1, ..., n\} \tag{A.1}$$

$$\frac{dx_i}{dt} = \sum_{j=1}^{n} s_{ij}^{+} v_j - \sum_{j=1}^{n} s_{ij}^{-} v_j, \qquad i = 1, \ldots, m \tag{A.2}$$

Therefore, we have one column vector for each reaction and Eq. A.2 can be written in matrix-vector notation for the whole metabolic network as follows:

$$\frac{d\vec{x}}{dt} = S_{m\times n}\vec{v} \tag{A.3}$$

The null space of stoichiometric matrix formulation

Where r is the rank for stoichiometric matrix S, the basis for the null space is a set with $q = (n - r)$ linearly independent column vectors of dimension n, which generates the null space for stoichiometric matrix S when spanned [70].

$$K(S) = \{\vec{b}_i \in R^n \quad (i = 1, \ldots, q) \quad | \quad S.\vec{b}_i = \vec{0} \quad \text{and} \quad \vec{b}_i.\vec{b}_j = \vec{0} \quad (i \neq j)\} \tag{A.4}$$

$$Null\ space(S) = span\{\vec{b}_1, \ldots, \vec{b}_q\}, \qquad q = (n - r) \tag{A.5}$$

Metabolic Flux Analysis (MFA)

MFA provides an empirical flux map:

$$\begin{aligned} S.\vec{v} &= 0 \\ \begin{bmatrix} S_u & S_k \end{bmatrix} \begin{bmatrix} \vec{v}_u \\ \vec{v}_k \end{bmatrix} &= 0 \\ \vec{v}_u &= -[(S_u^T S_u)^{-1} S_u^T] S_k \vec{v}_k \end{aligned} \tag{A.6}$$

Flux Balance Analysis (FBA)

The mathematical formulation of a cell objective, accounting for defined fluxes restrictions with bounding limits for the reaction fluxes, is developed as a linear programming optimization problem for FBA:

$$\begin{aligned} &\max \vec{C^T}.\vec{v} \\ &\text{subject to:} \\ &S.\vec{v} = 0 \\ &lb < \vec{v} < ub \end{aligned} \tag{A.7}$$

Modeling approaches complying to the thermodynamics-based constraints formulation

The formulation of TMFA combines constraints from FBA, directionality and constraints on metabolites concentrations as follows:

$$S.\vec{v} = 0, \tag{A.8}$$

$$0 \leq v_i \leq z_i v_{Max}, \{i = 1, ..., r\}, \tag{A.9}$$

$$\Delta_r G_i^{'} - K + K z_i < 0, \{i = 1, ..., r | \Delta_r G_i^{'\circ} \text{is known}\}, \tag{A.10}$$

$$\Delta_r G_i^{'\circ} + RT \sum_{j=1}^{m} n_{i,j} \ln x_j = \Delta_r G_i^{'}, \{i = 1, ..., r + L | \Delta_r G_i^{'\circ} \text{is known}\}, \tag{A.11}$$

$$\Delta_r G_i^{'} - K y_i < 0, \{i = 1, ..., r + L\}, \tag{A.12}$$

$$y_i + \sum_{j=1}^{r} \alpha_{i,j} z_j \leq \sum_{j=1}^{r} \alpha_{i,j}, \{i = 1, ..., r + L\} \tag{A.13}$$

In this method, each reversible reaction is decomposed to two backward and forward reactions. Therefore, there will be no negative v_i value. z_i is a binary number equal to one when there is non-zero flux or otherwise equals to zero. In Eq A.10, K is fixed as a value to make sure the inequality is satisfied when z_i and v_i are zeros. Of interest, the whole inequality checks that the solution for flux distribution obeys the second law of thermodynamics. The Gibbs free energy calculation for each reaction i is carried out in Eq A.11 with considering the activity x_j of involved metabolites.

Figure S1: EFMs enumeration for the running example with reversible reactions

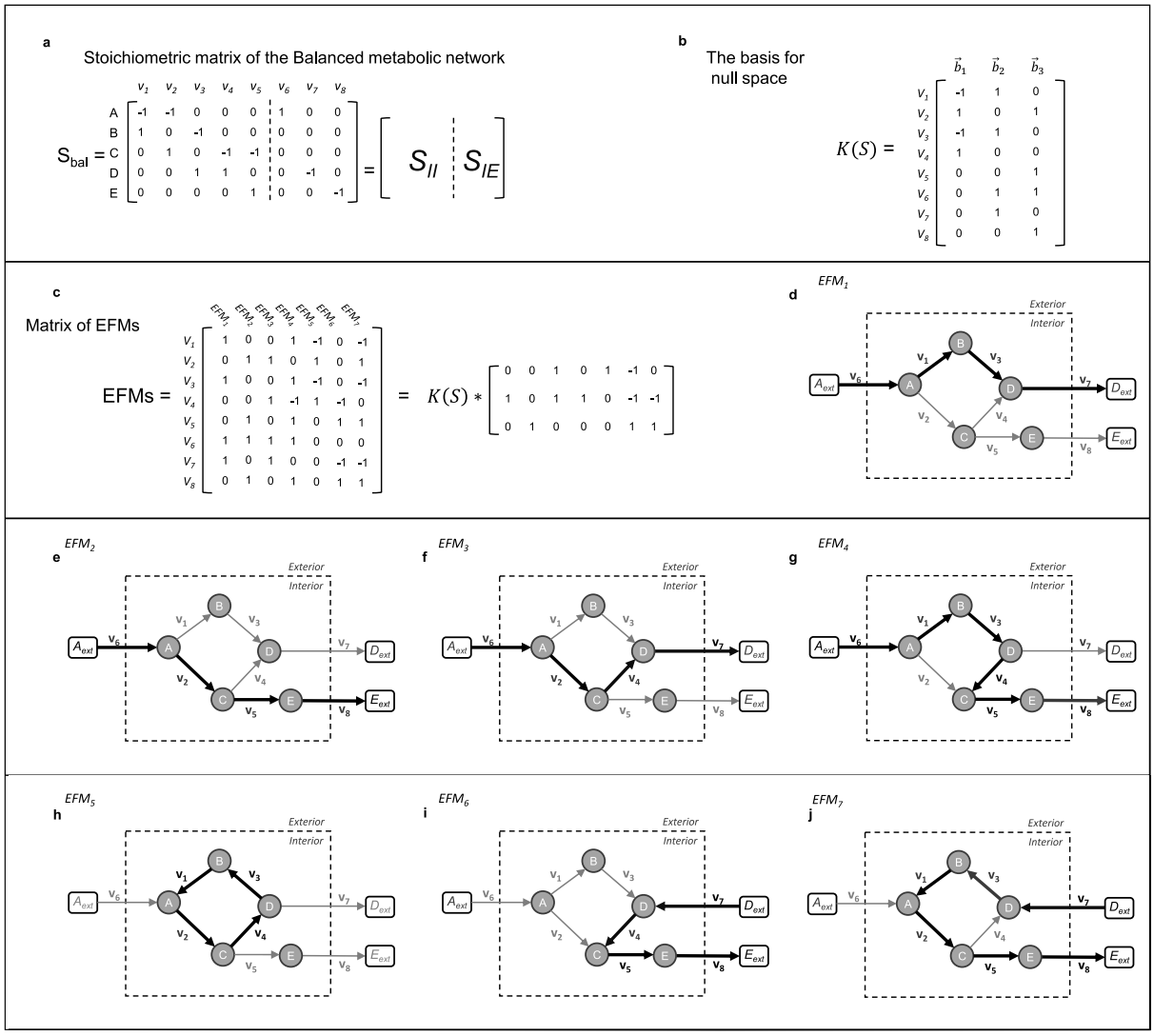

Figure A.1: **EFMs enumeration for the running example with reversible reactions** (**a**) The balanced metabolic network is same as in the main text. (**b**) The basis for the null space is the kernel matrix of the stoichiometric matrix. (**c**) 7 EFMs are enumerated assuming that all the reactions are reversible, except for v_6 which is inward. Each EFM is a linear combination of the basis vectors and for this system with reversible reactions, the negative coefficients in the columns 6 and 7 appear. (**d-j**) The flux maps for the EFMs are shown on the substrate graph of the network. (**g,i,j**) The EFMs that include some reactions in the reverse direction. (**h**) The EFM that includes a futile cycle.

Table A.1: **Properties of the different MPA methods.**

MPA Methods Properties				
Method Name	**Advantages**	**Disadvantages**	**Applications**	**Toolbox/Software**
Elementary Flux Mode (EFM)	No need for flux measurement and optimality objective function [78]; Every flux distribution can be decomposed into basic functional units without cancellation [77, 87]	Calculation of all EFMs is computationally demanding [87, 281]; Small percentage of EFMs are biologically relevant [87]	Suitable for small to medium scale models [87,282]; Identification of coregulated and coexpressed reactions and genes [88]; Composition of the minimal substrate required [88–90]; Quantification of cellular robustness [87,283,284]; Study of gene deletions and knockouts [83,87]; Finding operational modes and optimal routes [83]; Finding pathway length [83]	METATOOL [80]; CellNetAnalyzer [285]; COPASI [286]; Efmtool [287]; YANAsquare [288]
Extreme Pathway Analysis (EPA)	No need for flux measurement and optimality objective function [78]; Reducing number of pathways and applicability for assessing network properties [83]	Fluxes can cancel out [81, 87, 289]; Do not decompose reversible exchange reactions [199]; Not suitable to study mutations and the effect of reaction removal [83]	Suitable for small to medium scale models [87,282]; To caluclate the edges of solution space cone [81]	Expa [290]; METATOOL [80]; CellNetAnalyzer [285]; YANAsquare [288]

Table A.2: **The hypothesized objective functions of the cell and their biological rationales.**

Particular Objectives of the Cell		
Objective	**Biological rationale**	**Reference**
Maximum growth	In the exponential growth phase, cells utilize their resources to maximize proliferation, i.e. to form new biomass. This assumption is biologically relevant for both prokaryotes and eukaryotes, although a large part of cell resources are directed to maintenance cell in eukaryotes.	[]
Maximum bioenergetic production	Cells aim for the maximal ATP production, or maximum reducing power, which dictates a distinct distribution of carbon flux.	[291, 292]
Minimum overall intracellular flux	Cells aim for the minimum sum of the squares of fluxes, based on the assumed maximum enzymatic efficiency for cellular growth.	[99, 119, 293]
Maximum product formation	Recombinant DNA (rDNA) facilitates the design of modified cell lines with desired traits beyond predetermined cell objectives.	[]
Minimum reaction steps	Assuming that the cell chooses the shortest metabolic path from a substrate to a product.	[]
Minimum redox potential	Focus cell capability to generate energy through ubiquitous redox reactions.	[]

Table A.3: List of the flux balance analysis enhancements.

Enhancements of Flux Balance Analysis				
Method Name	**Approach**	**Regulation**	**Enzyme-cost**	**Reference application**
Flux Balance Analysis (FBA)	Steady-state constraint-based	NO	NO	Studying gene knockouts [94, 294], Studying cell growth on different media [94], Finding metabolic gaps [96]
Dynamic FBA (dFBA)	Iterative or dynamic	NO	NO	Description of unsteady-state growth and by-product secretion in aerobic batch, fed-batch, and anaerobic batch cultures [15], analysis of diauxic growth in Escherichia coli [15], study of carbon storage metabolism in microalgae [295]
Dynamic enzyme-cost FBA (deFBA)	Dynamic	NO	YES	Prediction of dynamic changes in metabolic fluxes and biomass composition during metabolic adaptations [296]
Resource balance analysis (RBA)	Steady-state	NO	YES	Prediction of the cell composition of bacteria with respect to their medium [297]
Dynamic resource balance analysis (dRBA)	Dynamic	NO	YES	Optimization of production of added-value compounds by bacteria [298]
Regulatory FBA (rFBA)	Iterative, Boolean logic	YES	NO	Regulation of gene expression [299, 300], predict of high-throughput experiments outcome [225], indication of knowledge gaps and unknown components and interactions [225], identification of potential targets for transcription factors [226]
Steady-state regulatory FBA (SR-FBA)	Steady-state	YES	NO	Prediction of gene expression and metabolic fluxes [301]
Probabilistic regulation of metabolism (PROM)	Steady-state, probabilistic	YES	NO	Statistical inference of regulatory network, automated quantification of regulatory Interactions [302]
Integrated FBA (iFBA)	Iterative, Boolean logic, ODE	YES	NO	Encapsulation of dynamics of internal metabolites, consideration of signaling molecules [303]
Integrated dynamic FBA (idFBA)	Iterative, Boolean logic, ODE	YES	YES	Integration of metabolic, signaling and regulatory networks [304]
Parsimonious FBA (pFBA)	Steady-state	NO	YES	describing metabolic states of the cell by minimizing the overall enzymatic fluxes to eliminate the alternative flux solutions [27]
Enzyme-cost FBA (ecFBA)	Steady-state	YES	YES	incorporating enzyme kinetic information in FBA framework, reduces the flux variability [227]
Regulatory dynamic enzyme-cost FBA (r-deFBA)	Dynamic	YES	YES	Consideration of enzyme production cost and regulatory events, integration of metabolic, signaling and regulatory networks [152]

Table A.4: Comparison of MFA and FBA.

Metabolic Flux Analysis (MFA)	**Flux Balance Analysis (FBA)**
$\vec{v}_u = -S_u^{-1} S_m \vec{v}_m$ $\vec{v}_u$: unknown flux vector; $\vec{v}_m$: measured flux vector; S_u , S_m: stoichiometric matrix subsets for unknown and measured fluxes respectively [68].	$\max \vec{c}^T.\vec{v}$; subject to: $S.\vec{v} = 0$ and lb $< \vec{v} <$ ub $\vec{v}$: flux vector; $\vec{c}$: weights associated to fluxes based on the objective of optimization; lb , ub: allowed lower and upper bound respectively for the flux values [94, 239].
Applicable for small to medium scale metabolic networks at steady-state. For calculation of the empirical value of unknown intracellular metabolic fluxes from measured extracellular metabolites' concentration changes.	Applicable for small to genome-scale metabolic networks at steady-state. For calculation of the feasible flux distribution from objective function linear optimization.
Accurate and simple approach; No need for kinetic parameters; Giving empirical results. *but*, Does not consider kinetics; Valid at steady-state conditions; There is a trade-off between the number of measured fluxes and laborious experimental work.	Simple approach; No need for kinetic parameters; Giving feasible results set. *but*, Does not consider kinetics; Valid at steady-state conditions; Optimization objective may not be realistic; Highly likely to have more than one feasible solution; Affected by incomplete genome annotation.
MetaFluxNet [305]; FiatFlux [306]; OpenFlux [307].	MetaFluxNet [305]; COBRA Toolbox [308]; FASIMU [309].

Table A.5: **Databases for retrieving the equilibrium constants.**

Databases for retrieving the equilibrium constants		
Name	**Specification**	**Reference**
eQuilibrator	Biochemical thermodynamics data calculator.	[310]
Factsage	Thermodynamics properties database and calculator.	[311]
NIST	National Institute of Standards and Technology.	[199]

APPENDIX B MATLAB CODE FOR ARTICLE 3

Data integration

```
%% C:\Users\ryase\OneDrive - ...
    polymtl.ca\Documents\MyMatlabDir\RBC_Code\RY_MATLAB scripts
%% The data integration module
%to find approximate place of datas for a given state, e.g. X
%         [i,j] = find(contains(statename,'X'))

%% Culture: Two- Control and Treated
    [stateval,statename,RAW]=xlsread('DATA_CHO_clonal.xlsx','AG-2014','A7:V28');
    [statevalSD,¬,¬]=xlsread('DATA_CHO_clonal.xlsx','AG-2014_std','A7:V28');

    No_obser = 7;
    No_metabphase = 1;
    No_conditions = 3; %Parental, LP, HP
    No_replicate = 1;   %AG-2014 reported mean and STD
    No_clmns = No_replicate*length(statename(1,3:end));%without time ...
        and Cnd id
    t_meas = stateval(1:No_obser,1); %Time vector of measurements
    Name = statename(1,3:end);%in case the descriptive names are entered
    NameID = statename(1,3:end);
    stateval = stateval(:,2:end);
    statevalSD = statevalSD(:,2:end);
    Data_struct = ...
        struct('TimeVector',t_meas,'Name',{cell(1,No_clmns)},'NameID',{cell(1,No_clmns)},
         'Type',{cell(1,No_clmns)},...
         'Value',nan(No_obser,No_clmns),...
         'Rate',nan(No_obser,No_clmns),...
         'ValueSD',zeros(No_obser,No_clmns/No_replicate),...
         'RateSD',zeros(No_obser,No_clmns/No_replicate));

    for i = 1:No_clmns/No_replicate
        for j = 1:No_conditions
            Data_struct.Name{No_conditions*(i-1)+j} = ...
                append('Cnd-',num2str(j),'-',Name{i});
            Data_struct.NameID{No_conditions*(i-1)+j} = ...
                append('Cnd-',num2str(j),'-',NameID{i});
```

```
31              Data_struct.Value(:,No_conditions*(i-1)+j) = ...
                    stateval((1+(j-1)*No_obser):(j*No_obser),i);
32              Data_struct.ValueSD(:,No_conditions*(i-1)+j) = ...
                    statevalSD((1+(j-1)*No_obser):(j*No_obser),i);
33
34              if contains(Data_struct.NameID{No_conditions*(i-1)+j},'MAB[e]')
35              %correct MAB unit/value based on the molar mass, AG-2014
36              %reported in g/L, we want mmol/L, the MW of Rituximab is
37              %143859.7 (Da) or 143859.7 (g/mol) = 143.8597 g/mmol
38              Data_struct.Value(:,No_conditions*(i-1)+j) = (1/143.8597)* ...
                    Data_struct.Value(:,No_conditions*(i-1)+j);
39              Data_struct.ValueSD(:,No_conditions*(i-1)+j) = ...
                    (1/143.8597)* statevalSD((1+(j-1)*No_obser):(j*No_obser),i);
40              end
41
42              if endsWith(Data_struct.NameID{No_conditions*(i-1)+j},'[e]')
43                  if ...
                        (sum(isnan(Data_struct.Value(:,No_conditions*(i-1)+j))) ...
                        == No_obser)
44                      Data_struct.Type{No_conditions*(i-1)+j} = ...
                            'ext_nonmeas';
45                  else
46                      Data_struct.Type{No_conditions*(i-1)+j} = 'ext_meas';
47                      if ...
                            contains(Data_struct.NameID{No_conditions*(i-1)+j},'BIOM[e]')
48                          %specific growth rate: 1/hr
49                          Data_struct.Rate(:,No_conditions*(i-1)+j) = ...
                                [0;log(Data_struct.Value(2:end,No_conditions*(i-1)+j)./Data_s
50                      elseif ...
                            contains(Data_struct.NameID{No_conditions*(i-1)+j},'O22[e]')
51                          %O2 data is given for the O2 consumption rate:
52                          %mmol/10^6cell/hr (see AG-2014)
53                          Data_struct.Rate(:,No_conditions*(i-1)+j) = ...
                                -10^6*([0;Data_struct.Value(2:end,No_conditions*(i-1)+j)]);
54                      else
55                          %specific transport rates: nanomol/10^6cell/hr
56                          Data_struct.Rate(:,No_conditions*(i-1)+j) = ...
                                1000*([0;Data_struct.Value(2:end,No_conditions*(i-1)+j)-Data_
57                      end
58                  end
59
60              elseif ...
                    endsWith(Data_struct.NameID{No_conditions*(i-1)+j},'[c]')
```

```
61                  if ...
                        (sum(isnan(Data_struct.Value(:,No_conditions*(i-1)+j))) ...
                        == No_obser)
62                      Data_struct.Type{No_conditions*(i-1)+j} = ...
                            'int_nonmeas';
63                  else
64                      Data_struct.Type{No_conditions*(i-1)+j} = 'int_meas';
65                      Data_struct.Rate(:,No_conditions*(i-1)+j) = ...
                            [0;Data_struct.Value(2:end,No_conditions*(i-1)+j)-Data_struct.Val
66                  end
67              else
68                  fprintf('\nwrong name entry at column %d of data ...
                        sheet\n',i);
69              end
70          end
71
72      end
73
74  save DATA_CHO_clonal_GeM.mat Data_struct
75  clear
```

Smoothing splines

```
1   clear
2   close all
3   load DATA_CHO_clonal_processed_ver3.mat Data_struct
4
5   % global sim_time_vec
6   %% The simulation time step and time vector
7   h = 1; %time step
8   sim_time_vec = 0:h:Data_struct.TimeVector(end);
9
10  %% The select transport metabolites
11  % 1:BIOM[e] 2:ALA[e] 3:ARG[e] 4:CYS[e] 5:GLC[e] 6:GLN[e] 7:GLU[e] ...
        8:GLY[e]
12  % 9:HIS[e] 10:ILE[e] 11:LAC[e] 12:LEU[e] 13:MET[e] 14:NH4[e] 15:PHE[e]
13  % 16:SER[e] 17:TYR[e] 18:VAL[e] 19:MAB[e]
14  select_ext_mets = {'BIOM[e]','GLC[e]', 'GLN[e]', 'GLU[e]',...
15          'LAC[e]', 'NH4[e]','O22[e]',...
16          'ALA[e]','ARG[e]','CYS[e]',...
17          'GLY[e]','HIS[e]','ILE[e]',...
```

```
18          'LEU[e]','MET[e]','PHE[e]','SER[e]','TYR[e]','VAL[e]','MAB[e]'};
19  % select_ext_mets = {'BIOM[e]';'ALA[e]';'ARG[e]';'CYS[e]';'GLC[e]';...
20  %     'GLN[e]';'GLU[e]';'GLY[e]';'HIS[e]';'ILE[e]';'LAC[e]';'LEU[e]';...
21  %     'MET[e]';'NH4[e]';'PHE[e]';'SER[e]';'TYR[e]';'VAL[e]';};
22  %% To initialize the concentration and flux rate simulation vector
23  % global sim_conc sim_conc_lb sim_conc_ub
24      sim_conc = nan(length(sim_time_vec),length(select_ext_mets));
25      sim_conc_lb = nan(length(sim_time_vec),length(select_ext_mets));
26      sim_conc_ub = nan(length(sim_time_vec),length(select_ext_mets));
27          sim_conc_cnd_1 = sim_conc; sim_conc_cnd_2 = sim_conc; ...
                sim_conc_cnd_3 = sim_conc;
28          sim_conc_lb_cnd_1 = sim_conc_lb; sim_conc_lb_cnd_2 = ...
                sim_conc_lb; sim_conc_lb_cnd_3 = sim_conc_lb;
29          sim_conc_ub_cnd_1 = sim_conc_ub; sim_conc_ub_cnd_2 = ...
                sim_conc_ub; sim_conc_ub_cnd_3 = sim_conc_ub;
30  % global sim_rate sim_rate_lb sim_rate_ub
31      sim_rate = nan(length(sim_time_vec)-1,length(select_ext_mets));
32      sim_rate_lb = nan(length(sim_time_vec)-1,length(select_ext_mets));
33      sim_rate_ub = nan(length(sim_time_vec)-1,length(select_ext_mets));
34          sim_rate_cnd_1 = sim_rate; sim_rate_cnd_2 = sim_rate; ...
                sim_rate_cnd_3 = sim_rate;
35          sim_rate_lb_cnd_1 = sim_rate_lb; sim_rate_lb_cnd_2 = ...
                sim_rate_lb; sim_rate_lb_cnd_3 = sim_rate_lb;
36          sim_rate_ub_cnd_1 = sim_rate_ub; sim_rate_ub_cnd_2 = ...
                sim_rate_ub; sim_rate_ub_cnd_3 = sim_rate_ub;
37  % x_stored = nan(length(select_ext_mets),2); %how many paramteres does ...
      your fit-function have?
38  % tiledlayout(3,2)
39  spline_func_stored = cell(length(select_ext_mets),1);
40  % fig_panel_let = ...
      {'A','B','C','D','E','F','G','H','I','J','K','L','M','N','O'};
41
42  %% To fit biomass concentratin and calculate the specific growth rate
43
44          dumindex = find(contains(Data_struct.NameID,select_ext_mets{1}));
45          cell2fit_cnd1 = ...
                [Data_struct.Value(:,dumindex(1)),Data_struct.Value(:,dumindex(1))+Data_struc
46          cell2fit_cnd2 = ...
                [Data_struct.Value(:,dumindex(2)),Data_struct.Value(:,dumindex(2))+Data_struc
47          cell2fit_cnd3 = ...
                [Data_struct.Value(:,dumindex(3)),Data_struct.Value(:,dumindex(3))+Data_struc
48          ppp_cnd1 = RY_splinefit(Data_struct.TimeVector,cell2fit_cnd1',2);
49          ppp_cnd2 = RY_splinefit(Data_struct.TimeVector,cell2fit_cnd2',2);
50          ppp_cnd3 = RY_splinefit(Data_struct.TimeVector,cell2fit_cnd3',2);
```

```
51          cell_splinefitted_cnd1 = ...
               ppval(ppp_cnd1,sim_time_vec);%triplicate in rows
52          cell_splinefitted_cnd2 = ppval(ppp_cnd2,sim_time_vec);%triplicate
53          cell_splinefitted_cnd3 = ppval(ppp_cnd3,sim_time_vec);%triplicate
54              %fill up the sim conc matrix
55              sim_conc_cnd_1(:,1) = cell_splinefitted_cnd1(1,:);
56              sim_conc_cnd_2(:,1) = cell_splinefitted_cnd2(1,:);
57              sim_conc_cnd_3(:,1) = cell_splinefitted_cnd3(1,:);
58
59              sim_conc_lb_cnd_1(:,1) = cell_splinefitted_cnd1(2,:);
60              sim_conc_lb_cnd_2(:,1) = cell_splinefitted_cnd2(2,:);
61              sim_conc_lb_cnd_3(:,1) = cell_splinefitted_cnd3(2,:);
62
63              sim_conc_ub_cnd_1(:,1) = cell_splinefitted_cnd1(3,:);
64              sim_conc_ub_cnd_2(:,1) = cell_splinefitted_cnd2(3,:);
65              sim_conc_ub_cnd_3(:,1) = cell_splinefitted_cnd3(3,:);
66              %fill up the sim rate matrix
67              sim_rate_cnd_1(:,1) = ...
                   (diff(sim_conc_cnd_1(:,1))/h)./sim_conc_cnd_1(2:end,1);
68              sim_rate_cnd_2(:,1) = ...
                   (diff(sim_conc_cnd_2(:,1))/h)./sim_conc_cnd_2(2:end,1);
69              sim_rate_cnd_3(:,1) = ...
                   (diff(sim_conc_cnd_3(:,1))/h)./sim_conc_cnd_3(2:end,1);
70
71              sim_rate_lb_cnd_1(:,1) = ...
                   (diff(sim_conc_lb_cnd_1(:,1))/h)./sim_conc_lb_cnd_1(2:end,1);
72              sim_rate_lb_cnd_2(:,1) = ...
                   (diff(sim_conc_lb_cnd_2(:,1))/h)./sim_conc_lb_cnd_2(2:end,1);
73              sim_rate_lb_cnd_3(:,1) = ...
                   (diff(sim_conc_lb_cnd_3(:,1))/h)./sim_conc_lb_cnd_3(2:end,1);
74
75              sim_rate_ub_cnd_1(:,1) = ...
                   (diff(sim_conc_ub_cnd_1(:,1))/h)./sim_conc_ub_cnd_1(2:end,1);
76              sim_rate_ub_cnd_2(:,1) = ...
                   (diff(sim_conc_ub_cnd_2(:,1))/h)./sim_conc_ub_cnd_2(2:end,1);
77              sim_rate_ub_cnd_3(:,1) = ...
                   (diff(sim_conc_ub_cnd_3(:,1))/h)./sim_conc_ub_cnd_3(2:end,1);
78
79
80  for k = 2:length(select_ext_mets)
81      %% spline fit
82          dumindex = find(contains(Data_struct.NameID,select_ext_mets{k}));
83          y2fit_cnd1 = ...
                [Data_struct.Value(:,dumindex(1)),Data_struct.Value(:,dumindex(1))+Data_struc
```

```
84          y2fit_cnd2 = ...
               [Data_struct.Value(:,dumindex(2)),Data_struct.Value(:,dumindex(2))+Data_struc
85          y2fit_cnd3 = ...
               [Data_struct.Value(:,dumindex(3)),Data_struct.Value(:,dumindex(3))+Data_struc
86          ppp_cnd1 = RY_splinefit(Data_struct.TimeVector,y2fit_cnd1',2);
87          ppp_cnd2 = RY_splinefit(Data_struct.TimeVector,y2fit_cnd2',2);
88          ppp_cnd3 = RY_splinefit(Data_struct.TimeVector,y2fit_cnd3',2);
89          y_splinefitted_cnd1 = ppval(ppp_cnd1,sim_time_vec);%triplicate
90          y_splinefitted_cnd2 = ppval(ppp_cnd2,sim_time_vec);%triplicate
91          y_splinefitted_cnd3 = ppval(ppp_cnd3,sim_time_vec);%triplicate
92              %concentrations in mmol/L
93              sim_conc_cnd_1(:,k) = y_splinefitted_cnd1(1,:);
94              sim_conc_cnd_2(:,k) = y_splinefitted_cnd2(1,:);
95              sim_conc_cnd_3(:,k) = y_splinefitted_cnd3(1,:);
96
97              sim_conc_lb_cnd_1(:,k) = y_splinefitted_cnd1(2,:);
98              sim_conc_lb_cnd_2(:,k) = y_splinefitted_cnd2(2,:);
99              sim_conc_lb_cnd_3(:,k) = y_splinefitted_cnd3(2,:);
100
101             sim_conc_ub_cnd_1(:,k) = y_splinefitted_cnd1(3,:);
102             sim_conc_ub_cnd_2(:,k) = y_splinefitted_cnd2(3,:);
103             sim_conc_ub_cnd_3(:,k) = y_splinefitted_cnd3(3,:);
104             %fill up the sim "specific" rate matrix:*1000 to have it in
105             %nanomol/10^6cell/hr
106             sim_rate_cnd_1(:,k) = ...
                   1000*(diff(sim_conc_cnd_1(:,k))/h)./sim_conc_cnd_1(2:end,1);
107             sim_rate_cnd_2(:,k) = ...
                   1000*(diff(sim_conc_cnd_2(:,k))/h)./sim_conc_cnd_2(2:end,1);
108             sim_rate_cnd_3(:,k) = ...
                   1000*(diff(sim_conc_cnd_3(:,k))/h)./sim_conc_cnd_3(2:end,1);
109
110             sim_rate_lb_cnd_1(:,k) = ...
                   1000*(diff(sim_conc_lb_cnd_1(:,k))/h)./sim_conc_lb_cnd_1(2:end,1);
111             sim_rate_lb_cnd_2(:,k) = ...
                   1000*(diff(sim_conc_lb_cnd_2(:,k))/h)./sim_conc_lb_cnd_2(2:end,1);
112             sim_rate_lb_cnd_3(:,k) = ...
                   1000*(diff(sim_conc_lb_cnd_3(:,k))/h)./sim_conc_lb_cnd_3(2:end,1);
113
114             sim_rate_ub_cnd_1(:,k) = ...
                   1000*(diff(sim_conc_ub_cnd_1(:,k))/h)./sim_conc_ub_cnd_1(2:end,1);
115             sim_rate_ub_cnd_2(:,k) = ...
                   1000*(diff(sim_conc_ub_cnd_2(:,k))/h)./sim_conc_ub_cnd_2(2:end,1);
116             sim_rate_ub_cnd_3(:,k) = ...
                   1000*(diff(sim_conc_ub_cnd_3(:,k))/h)./sim_conc_ub_cnd_3(2:end,1);
```

```
117         if k == 7 %oxygen
118                 sim_rate_cnd_1(:,k) = -10^6*sim_conc_cnd_1(2:end,k);
119                 sim_rate_cnd_2(:,k) = -10^6*sim_conc_cnd_2(2:end,k);
120                 sim_rate_cnd_3(:,k) = -10^6*sim_conc_cnd_3(2:end,k);
121
122                 sim_rate_lb_cnd_1(:,k) = -10^6*sim_conc_lb_cnd_1(2:end,k);
123                 sim_rate_lb_cnd_2(:,k) = -10^6*sim_conc_lb_cnd_2(2:end,k);
124                 sim_rate_lb_cnd_3(:,k) = -10^6*sim_conc_lb_cnd_3(2:end,k);
125
126                 sim_rate_ub_cnd_1(:,k) = -10^6*sim_conc_ub_cnd_1(2:end,k);
127                 sim_rate_ub_cnd_2(:,k) = -10^6*sim_conc_lb_cnd_2(2:end,k);
128                 sim_rate_ub_cnd_3(:,k) = -10^6*sim_conc_lb_cnd_3(2:end,k);
129         end
130     %% concentration simulation values
131     %Cnd_1: nominal
132 figure
133     plot(sim_time_vec,sim_conc_cnd_3(:,k),'k-','LineWidth',1)
134     hold on
135     plot(sim_time_vec,sim_conc_lb_cnd_3(:,k),'b-')
136     plot(sim_time_vec,sim_conc_ub_cnd_3(:,k),'b-')
137     % The extracellular metabolite measurements plot with error bars
138     errorbar(Data_struct.TimeVector,...
139     Data_struct.Value(:,dumindex(3)),...
140     Data_struct.ValueSD(:,dumindex(3)),'r+')
141     str = ['Cnd_3','     ', select_ext_mets{k},'_{conc}']; %MUST ...
            comment-out for single plotting
142     title(str)
143     xlabel('Time (hours)')
144     str = {'Extracellular metabolite ...
            concentration';'[millimol-per-litter]'};
145     ylabel(str)
146     hold off
147
148     %save the figure
149 %     figdum = gcf;
150 %     figname = ['Cnd 3 high-producing','     ', ...
        select_ext_mets{k},'_{conc}'];
151 %     saveas(figdum,figname,'png')
152 %% The rate simulation COMMON between tiled plot and single high ...
        quality plot
153 figure
154     plot(sim_time_vec(1:end-1),sim_rate_cnd_2(:,k),'k-','LineWidth',1)
155     hold on
156     plot(sim_time_vec(1:end-1),sim_rate_lb_cnd_2(:,k),'b-')
```

```
157      plot(sim_time_vec(1:end-1),sim_rate_ub_cnd_2(:,k),'b-')
158      % incremental experimental rate
159  %     errorbar(Data_struct.TimeVector(1:end-1),...
160  %         ...
         Data_struct.Ratemean(2:end,find(strcmp(Data_struct.NameIDmean,select_ext_mets{k}))),
161  %         ...
         Data_struct.RateSD(2:end,find(strcmp(Data_struct.NameIDmean,select_ext_mets{k}))),'r-
162      str = [select_ext_mets{k},'  ','specific flux rate'];
163      title(str)
164      xlabel('Time (hours)')
165      str = {'[nanomol-per-10^6 cell-per-hour]'};
166      ylabel(str)
167      hold off
168      %save figure
169  %     figdum = gcf;
170  %     figname = ['Cnd 2 low producing','     ', ...
         select_ext_mets{k},'_{specific transport flux}'];
171  %     saveas(figdum,figname,'png')
172  %     if mod(k,4) == 0  %For tiled plotting
173  %         fig_handle = gcf;
174  %         fig_name = ...
         ['sim_and_exp_concen_and_rate_values_Sept',select_ext_mets{(k-3):k},'.PDF'];
175  % %         ...
         exportgraphics(fig_handle,fig_name,'Resolution',300,'ContentType','vector');
176      end
177
178
179
180  save CHO_splinefitresults_AG-2014_3conditions_ver0 ...
181      sim_time_vec ...
182      sim_conc_cnd_1 sim_conc_cnd_2 sim_conc_cnd_3 ...
183      sim_conc_lb_cnd_1 sim_conc_lb_cnd_2 sim_conc_lb_cnd_3 ...
184      sim_conc_ub_cnd_1 sim_conc_ub_cnd_2 sim_conc_ub_cnd_3 ...
185      sim_rate_cnd_1 sim_rate_cnd_2 sim_rate_cnd_3 ...
186      sim_rate_lb_cnd_1 sim_rate_lb_cnd_2 sim_rate_lb_cnd_3 ...
187      sim_rate_ub_cnd_1 sim_rate_ub_cnd_2 sim_rate_ub_cnd_3
```

Simulation

```
1  %% Implementation of G-DCBM on Updated iCHOv1_DG44/After:04/April/2022/
2  %% Mario Jolicoeur Lab
```

```
%% Author: r.yasemi@gmail.com
% function my_cost = G_DCBM_MY_MJ_cndns_April_01_simulation_spare3(inn)
close all
clear all
% initCobraToolbox(false)
[solverOK, solverInstalled] = changeCobraSolver('gurobi','LP');
%% Preamble: Parallel computing settings
delete(gcp('nocreate'))
% local_cluster = parcluster()
% mypool = parpool();
%% Preamble: The tasks
do_DFBA = true;
do_PLOT = true;
do_INT_PRED = false;
do_CALC_res = true;
global COND_INDEX DFBA_init_time_index DFBA_fin_time_index sorted_pred_mets
    COND_INDEX = 3;
    switch COND_INDEX
        case 1
            DFBA_init_time_index = 20;
            DFBA_fin_time_index = 128;
            DFBA_shift_time = 125;
        case 2
            DFBA_init_time_index = 20;
            DFBA_fin_time_index = 128;
%             DFBA_fin_time_index = 144;
            DFBA_shift_time = 116;
        case 3
            DFBA_init_time_index = 20;
            DFBA_fin_time_index = 128;
%             DFBA_fin_time_index = 130;
            DFBA_shift_time = 71;
    end
%% Preamble: Loading the inputs
% Metabolic network in COBRA format
% load MODEL_ver_8.mat MODEL
load MODEL_ver_10.mat MODEL
% Spline smoothed concentrations and exchange rates
load CHO_splinefitresults_AG-2014_3conditions_ver0 ...
    sim_time_vec ...
    sim_conc_cnd_1 sim_conc_cnd_2 sim_conc_cnd_3 ...
    sim_conc_lb_cnd_1 sim_conc_lb_cnd_2 sim_conc_lb_cnd_3 ...
    sim_conc_ub_cnd_1 sim_conc_ub_cnd_2 sim_conc_ub_cnd_3 ...
    sim_rate_cnd_1 sim_rate_cnd_2 sim_rate_cnd_3 ...
```

```
47      sim_rate_lb_cnd_1 sim_rate_lb_cnd_2 sim_rate_lb_cnd_3 ...
48      sim_rate_ub_cnd_1 sim_rate_ub_cnd_2 sim_rate_ub_cnd_3
49  % Unbiased dynamic flux bounds on all reactions based on the imposed
50  % measured exchange reactions
51  % load series_DFVA_GeM_cnd_1_Feb17_11constraints.mat ...
        cell_array_DFVA_output DCBM_time_vec common_trnspt_mets ...
        main_trnspt_mets dum_cons
52  % FVA_FminFmax_vectors = cell_array_DFVA_output{20};
53
54  %% Step 0: Tailoring the model
55  GeM_CHO = MODEL;
56
57  switch COND_INDEX
58      case 1
59          MODEL.ub(startsWith(MODEL.rxns,'SK_ser')) = 0;
60          MODEL.lb(startsWith(MODEL.rxns,'SK_ser')) = -860;
61          MODEL.ub(startsWith(MODEL.rxns,'SK_val')) = 0;
62          MODEL.lb(startsWith(MODEL.rxns,'SK_val')) = -1;
63  %         MODEL.ub(startsWith(MODEL.rxns,'SK_gly')) = 2000;
64  %         MODEL.lb(startsWith(MODEL.rxns,'SK_gly')) = 1500;
65      case 2
66          MODEL.ub(startsWith(MODEL.rxns,'SK_ser')) = 0;
67          MODEL.lb(startsWith(MODEL.rxns,'SK_ser')) = -570;
68          MODEL.ub(startsWith(MODEL.rxns,'SK_val')) = 0;
69          MODEL.lb(startsWith(MODEL.rxns,'SK_val')) = -1;
70  %         MODEL.ub(startsWith(MODEL.rxns,'SK_gly')) = 600;
71  %         MODEL.lb(startsWith(MODEL.rxns,'SK_gly')) = 440;
72      case 3
73          MODEL.ub(startsWith(MODEL.rxns,'SK_ser')) = 0;
74          MODEL.lb(startsWith(MODEL.rxns,'SK_ser')) = -230;
75          MODEL.ub(startsWith(MODEL.rxns,'SK_val')) = 0;
76          MODEL.lb(startsWith(MODEL.rxns,'SK_val')) = -1;
77  %         MODEL.ub(startsWith(MODEL.rxns,'SK_gly')) = 500;
78          MODEL.lb(startsWith(MODEL.rxns,'SK_gly')) = 10;
79  end
80  %% Step 1: Determining the secretion/uptake rates of measured ...
        extracellular species [EXCHANGE fluxes]
81  ModelFluxUnitLongFormat ='nanomolebiomolecule-per-millioncell-per-hour';
82  TransportFluxUnit = 'nanomolepermillioncellperhour';
83  % The complete set of species for which time-series measurements are
84  % available
85  %DO NOT CHANGE THIS LIST
86      common_trnspt_mets = {'BIOM_e','glc__D_e', 'gln__L_e', 'glu__L_e',...
87          'lac__L_e', 'nh4_e','o2_e',...
```

```
88          'ala__L_e','arg__L_e','cys__L_e',...
89          'gly_e','his__L_e','ile__L_e',...
90          'leu__L_e','met__L_e','phe__L_e','ser__L_e','tyr__L_e','val__L_e','igg_e'};
91  % List of the main sources of carbon and nitrogen, electron acceptor,
92      main_trnspt_mets = {'glc__D_e', 'gln__L_e', 'glu__L_e',...
93          'lac__L_e', 'nh4_e','o2_e','ala__L_e','tyr__L_e','met__L_e',...
94          'his__L_e','cys__L_e','arg__L_e','ile__L_e'};
95  % dum_cons = {'arg__L_e','cys__L_e','gly_e','his__L_e','ile__L_e',...
96  %          ...
      'leu__L_e','met__L_e','phe__L_e','ser__L_e','tyr__L_e'};%,'val__L_e'};
97      dum_cons = {'leu__L_e','phe__L_e','val__L_e','gly_e','ser__L_e'};
98
99
100 imposed_trnspt_mets = [main_trnspt_mets, dum_cons];
101 sim_index = ismember(common_trnspt_mets,imposed_trnspt_mets);
102 %% The involved transport rates/ pay attention to the cell culture ...
      condition
103 %(Condition 1:Parental- Condition 2:LP - Condition 3:HP)
104     sim_rate_lb_ub_3Dmat = ...
          repmat(zeros(length(sim_rate_lb_cnd_1),1),[1,2,sum(sim_index)]);
105     j = 0;
106     for i=1:length(sim_index)
107         switch COND_INDEX
108             case 1
109                 if sim_index(i)
110                     j = j+1;
111                     sim_rate_lb_ub_3Dmat(:,1,j) = ...
                            min([sim_rate_cnd_1(:,i),...
112                     sim_rate_lb_cnd_1(:,i),sim_rate_ub_cnd_1(:,i)]');
113
114                     sim_rate_lb_ub_3Dmat(:,2,j) = ...
                            max([sim_rate_cnd_1(:,i),...
115                     sim_rate_lb_cnd_1(:,i),sim_rate_ub_cnd_1(:,i)]');
116                     %This is required to make sure only the "upper ...
                            bounds" are
117                     %imposed
118                     for jj = 1:length(sim_rate_lb_cnd_1)
119                         if ...
                                sim_rate_lb_ub_3Dmat(jj,1,j)*sim_rate_lb_ub_3Dmat(jj,2,j) ..
                                > 0 ...
120                              &&  sim_rate_lb_ub_3Dmat(jj,1,j) > 0 && j ≠ ...
                                    2 && j ≠ 3 && j ≠ 4 && j ≠5
121                              sim_rate_lb_ub_3Dmat(jj,1,j) = 0;
122                         end
```

```
123                         if ...
                               sim_rate_lb_ub_3Dmat(jj,1,j)*sim_rate_lb_ub_3Dmat(jj,2,j) ..
                               > 0 ...
124                             &&  sim_rate_lb_ub_3Dmat(jj,1,j) < 0 && j ≠ 2
125                             sim_rate_lb_ub_3Dmat(jj,2,j) = 0;
126                         end
127                     end
128                 end
129             case 2
130                 if sim_index(i)
131                     j = j+1;
132                     sim_rate_lb_ub_3Dmat(:,1,j) = ...
                           min([sim_rate_cnd_2(:,i),...
133                     sim_rate_lb_cnd_2(:,i),sim_rate_ub_cnd_2(:,i)]');
134
135                     sim_rate_lb_ub_3Dmat(:,2,j) = ...
                           max([sim_rate_cnd_2(:,i),...
136                     sim_rate_lb_cnd_2(:,i),sim_rate_ub_cnd_2(:,i)]');
137                 for jj = 1:length(sim_rate_lb_cnd_1)
138                         if ...
                               sim_rate_lb_ub_3Dmat(jj,1,j)*sim_rate_lb_ub_3Dmat(jj,2,j) ..
                               > 0 ...
139                             &&  sim_rate_lb_ub_3Dmat(jj,1,j) > 0 && j ≠ ...
                                   2 && j ≠ 3 && j ≠ 4 && j ≠ 5
140                             sim_rate_lb_ub_3Dmat(jj,1,j) = 0;
141                         end
142                         if ...
                               sim_rate_lb_ub_3Dmat(jj,1,j)*sim_rate_lb_ub_3Dmat(jj,2,j) ..
                               > 0 ...
143                             &&  sim_rate_lb_ub_3Dmat(jj,1,j) < 0 && j ≠ 2
144                             sim_rate_lb_ub_3Dmat(jj,2,j) = 0;
145                         end
146                     end
147                 end
148             case 3
149                 if sim_index(i)
150                     j = j+1;
151                     sim_rate_lb_ub_3Dmat(:,1,j) = ...
                           min([sim_rate_cnd_3(:,i),...
152                     sim_rate_lb_cnd_3(:,i),sim_rate_ub_cnd_3(:,i)]');
153
154                 sim_rate_lb_ub_3Dmat(:,2,j) = max([sim_rate_cnd_3(:,i),...
155                     sim_rate_lb_cnd_3(:,i),sim_rate_ub_cnd_3(:,i)]');
156                 for jj = 1:length(sim_rate_lb_cnd_1)
```

```
157                         if ...
                                sim_rate_lb_ub_3Dmat(jj,1,j)*sim_rate_lb_ub_3Dmat(jj,2,j) ..
                                > 0 ...
158                              &&  sim_rate_lb_ub_3Dmat(jj,1,j) > 0 && j ≠ ...
                                     2 && j ≠ 3 && j ≠ 4 && j ≠5
159                              sim_rate_lb_ub_3Dmat(jj,1,j) = 0;
160                         end
161                         if ...
                                sim_rate_lb_ub_3Dmat(jj,1,j)*sim_rate_lb_ub_3Dmat(jj,2,j) ..
                                > 0 ...
162                              &&  sim_rate_lb_ub_3Dmat(jj,1,j) < 0 && j ≠ 2
163                              sim_rate_lb_ub_3Dmat(jj,2,j) = 0;
164                         end
165                     end
166                 end
167         end
168     end
169 display('**$$$** To assign lower and upper bounds of the transport ...
        reaction(s) from the extracellular data:')
170
171 %% APM
172 addpath('../apm')
173
174 %% Dynamic FBA (all time points)
175 solverOK = changeCobraSolver('gurobi', 'LP');
176 my_cost = 0;
177 flag_infeasible = 0;
178 if solverOK == 1 && do_DFBA
179     DCBM_time_vec = sim_time_vec(DFBA_init_time_index:DFBA_fin_time_index);
180
181     FVA_FminFmax_vectors = ...
            repmat([MODEL.lb,MODEL.ub],[1,1,length(DCBM_time_vec)]);
182
183     post_FVA_FminFmax_vectors = FVA_FminFmax_vectors;
184
185     imposed_exch = append('EX_',common_trnspt_mets(sim_index));
186
187     index_EX_rxn = ismember(MODEL.rxns,imposed_exch);
188
189     G_DCBM_flux_vectors = ...
            repmat(nan(length(MODEL.rxns),1),[1,3,length(DCBM_time_vec)]);
190
191     G_DCBM_flux_vectors(:,1:2,:) = FVA_FminFmax_vectors;
192
```

```
193     %storing optimization information
194     G_DCBM_opt_solution_infos = cell(length(DCBM_time_vec),3);
195     G_DCBM_opt_solution_infos(:,1) = num2cell(DCBM_time_vec,1);
196     G_DCBM_opt_solution_infos(:,2) = cell(length(DCBM_time_vec),1);
197
198     %FBA conditions:%the model is sensitive to the existence of loops
199     allowLoops = 0;%0:loopless solutions
200     printLevel = 0;%2;%prints results
201     optPercentage = 100; osenseStr = 'max'; method = '2-norm'; minNorm ...
            = 1e-6;
202     solverParams = struct();
203         solverParams.zeroNormApprox = 'all';
204         solverParams.verify = 'true';
205     format shortG
206
207     if sum(index_EX_rxn,'all') ≠ length(imposed_exch)
208         formatSpec = '\n\n*%8.4f imposed exchange rates WERE found in ...
                the GeM\n %8.4f imposed exchange rates were not found in the ...
                GeM.\n';
209         warning(formatSpec,sum(index_EX_rxn),length(common_trnspt_mets(sim_index))-sum(i
210     end
211
212 %The reaction ID for the biomass reaction included in the genome-scale
213     %model: 'MY_BIOMASS_cho_producing_1'
214     dum_index = find(contains(MODEL.rxns,'MY_BIOMASS_cho_producing_1'));
215     %Extracting the participating species list
216     metaboliteList_biom = MODEL.mets(sum(full(MODEL.S(:,dum_index)),2)≠0);
217     coeffsList_biom = ...
            sum(full(MODEL.S(ismember(MODEL.mets,metaboliteList_biom),dum_index)),2)
218     InitGuess = coeffsList_biom;
219     aa_indices = [8;11;12;13;14;15;16;18;19;20;21;23;24;25;27;28;29;30];
220 switch COND_INDEX
221     case 1
222         InitGuess(aa_indices) = InitGuess(aa_indices)*0.7;
223     case 2
224         InitGuess(aa_indices) = InitGuess(aa_indices)*0.9;
225     case 3
226         InitGuess(aa_indices) = InitGuess(aa_indices)*1;
227 end
228    %remove the reaction to avoid duplicate reactions
229     MODEL = removeRxns(MODEL,'MY_BIOMASS_cho_producing_1');
230    %Add the new reaction considering a new THE_LIST value
231     MODEL = ...
            addReaction(MODEL,'MY_BIOMASS_cho_producing_1','reactionName',...
```

```
232     'Biomass for a producing DG44 cell ...
           line','metaboliteList',metaboliteList_biom,...
233     'stoichCoeffList',InitGuess,'reversible',false);
234   %% PREDICTIONS: Concentrations back calculation, Cell Specific Fluxes
235     %metabolites that must be predicted and for which we have an initial
236     %concentration value
237     pred_mets = common_trnspt_mets;
238     pred_EXflux = append('EX_',pred_mets);
239     pred_index = ismember(common_trnspt_mets,pred_mets);
240     [sorted_pred_EXflux, sorted_pred_index] = sort(pred_EXflux);
241     sorted_pred_mets = pred_mets(sorted_pred_index);
242
243     sim_rate_for_graph_lb_ub_3Dmat = ...
           repmat(zeros(length(sim_rate_lb_cnd_1),1),[1,2,sum(pred_index)]);
244     j = 0;
245
246 for i=1:length(pred_index)
247     switch COND_INDEX
248         case 1
249             if pred_index(i)
250                 j = j+1;
251                 conc_sim_data = sim_conc_cnd_1(:,sorted_pred_index);
252                 conc_sim_datalb = sim_conc_lb_cnd_1(:,sorted_pred_index);
253                 conc_sim_dataub = sim_conc_ub_cnd_1(:,sorted_pred_index);
254             end
255
256         case 2
257             if pred_index(i)
258                 j = j+1;
259                 conc_sim_data = sim_conc_cnd_2(:,sorted_pred_index);
260                 conc_sim_datalb = sim_conc_lb_cnd_2(:,sorted_pred_index);
261                 conc_sim_dataub = sim_conc_ub_cnd_2(:,sorted_pred_index);
262
263             end
264
265         case 3
266             if pred_index(i)
267                 j = j+1;
268                 conc_sim_data = sim_conc_cnd_3(:,sorted_pred_index);
269                 conc_sim_datalb = sim_conc_lb_cnd_3(:,sorted_pred_index);
270                 conc_sim_dataub = sim_conc_ub_cnd_3(:,sorted_pred_index);
271
272             end
273         end
```

```
274 end
275
276 %Initializing the matrices for storing dynamic concentration results
277     pred_conc_trajectory = zeros(length(DCBM_time_vec),sum(pred_index));
278 %Assigning the initial values
279     pred_conc_trajectory(1,:) = conc_sim_data(DFBA_init_time_index,:);
280 %To exclude BIOM_e vector and assign its initial value (initial biomass
281 %concentration)
282     biom_pred_data(1) = pred_conc_trajectory(1,ismember( ...
            sorted_pred_mets,'BIOM_e'));
283 %     pred_conc_trajectory(:,ismember(sorted_pred_mets,'BIOM_e')) = [];
284
285 %Ensuring that the model IDs for the exchange fluxes are synced
286     [sorted_pred_uptake, sorted_pred_uptake_index] = ...
287         sort(MODEL.rxns(ismember(MODEL.rxns,pred_EXflux)));
288
289 %% Reading out the predicted growth and optimum fluxes from DFBA
290     pred_uptake = ...
            G_DCBM_flux_vectors(ismember(MODEL.rxns,pred_EXflux),3:3:end);
291     sorted_pred_uptake = pred_uptake(sorted_pred_uptake_index,:);
292     mu = sorted_pred_uptake(ismember(sorted_pred_EXflux,'EX_BIOM_e'),:);
293     sorted_pred_uptake(ismember(sorted_pred_EXflux,'EX_BIOM_e'),:) = [];
294     %Step size (1 hour)
295     Δ_t = DCBM_time_vec(3) - DCBM_time_vec(2);
296
297     %% Dynamic Flux balance analysis Linear Programming optimization
298     solverOK = changeCobraSolver('gurobi', 'LP');
299     allowLoops = true;%0:loopless solutions
300     printLevel = 0;%2;%prints results
301     optPercentage = 100; osenseStr = 'max'; method = '2-norm'; minNorm ...
            = 1e-6;
302     solverParams = struct();
303     solverParams.zeroNormApprox = 'all';
304     solverParams.verify = 'true';
305     format shortG
306
307 %% DFBA-dFVA iterating over time
308 for j = 1:length(DCBM_time_vec)
309     %Imposing constraints on the model
310     MODEL = changeRxnBounds(MODEL, imposed_exch, ...
            sim_rate_lb_ub_3Dmat(DCBM_time_vec(j),1,:), 'l');
311     MODEL = changeRxnBounds(MODEL, imposed_exch, ...
            sim_rate_lb_ub_3Dmat(DCBM_time_vec(j),2,:), 'u');
312     G_DCBM_flux_vectors(:,1,j) = MODEL.lb;
```

```
313     G_DCBM_flux_vectors(:,2,j) = MODEL.ub;
314     formatSpec = '\n## \n**The exchange rate bounds in time index %d ...
           are set.\n t_initial = %.2f hour\n';
315     fprintf(formatSpec,j,DCBM_time_vec(j))
316
317     %Conditional flux bounds constraints
318     oxid_stress_rxns = {'SPODM';'TYROX';'ILEOX';'SMOX';'SPMDOX';...
319     '5HOXINDACTO2OX';'XAO';'XAO2';'ASCBOX';'SPODMx';'GLYOp';...
320     'XAO2x';'XAOx';'SPODMm';'SPODMn'};    % COBRA function: optimizeCbModel
321     C_ineq = ismember(MODEL.rxns,oxid_stress_rxns);
322     d_ineq = 1; %min H2O2 generation from literature
323 %     MODEL.C = C_ineq';
324 %     MODEL.d = d_ineq;
325 %     MODEL.dsense = 'G';
326 %
327     % To introduce the objective function
328     fprintf('Testing flux balance analysis using %s ... ', 'gurobi');
329     fprintf('\n *Optimal solution growth\n');
330     DCBM_time_vec(j)
331
332     %Initializing the objective function
333     MODEL.c = zeros(length(MODEL.c),1);
334
335 if DCBM_time_vec(j) ≤ DFBA_shift_time
336     MODEL.c(ismember(MODEL.rxns,'MY_BIOMASS_cho_producing_1')) = ...
           1;%biomass synthesis
337     switch COND_INDEX
338         case 1
339             MODEL.c(contains(MODEL.rxns,'EX_ile__L_e')) = -1;
340             MODEL.c(contains(MODEL.rxns,'EX_his__L_e')) = -1;
341             MODEL.c(contains(MODEL.rxns,'EX_cys__L_e')) = -1;
342             MODEL.c(contains(MODEL.rxns,'EX_met__L_e')) = -1;
343 %             MODEL.c(contains(MODEL.rxns,'EX_gly_e')) = -1;
344             MODEL.c(contains(MODEL.rxns,'EX_tyr__L_e')) = -1;
345         case 2
346             MODEL.c(contains(MODEL.rxns,'EX_ile__L_e')) = -1;
347             MODEL.c(contains(MODEL.rxns,'EX_his__L_e')) = -1;
348             MODEL.c(contains(MODEL.rxns,'EX_cys__L_e')) = -1;
349             MODEL.c(contains(MODEL.rxns,'EX_met__L_e')) = -1;
350 %             MODEL.c(contains(MODEL.rxns,'EX_gly_e')) = -1;
351             MODEL.c(contains(MODEL.rxns,'EX_tyr__L_e')) = -1;
352         case 3
353             MODEL.c(contains(MODEL.rxns,'EX_ile__L_e')) = -1;
354             MODEL.c(contains(MODEL.rxns,'EX_his__L_e')) = -1;
```

```
355             MODEL.c(contains(MODEL.rxns,'EX_cys__L_e')) = -1;
356             MODEL.c(contains(MODEL.rxns,'EX_met__L_e')) = -1;
357 %             MODEL.c(contains(MODEL.rxns,'EX_gly_e')) = -1;
358             MODEL.c(contains(MODEL.rxns,'EX_tyr__L_e')) = -1;
359     end
360 else
361       switch COND_INDEX
362        case 1
363             MODEL.c(ismember(MODEL.rxns,'MY_BIOMASS_cho_producing_1')) ...
                 = 0;%biomass synthesis
364             MODEL.c(contains(MODEL.rxns,'EX_glc__D_e')) = -1;
365             MODEL.c(contains(MODEL.rxns,'EX_ser__L_e')) = -1;
366             MODEL.c(contains(MODEL.rxns,'EX_gln__L_e')) = -1;
367             MODEL.c(contains(MODEL.rxns,'EX_val__L_e')) = -1;
368             MODEL.c(contains(MODEL.rxns,'EX_ala__L_e')) = -1;
369             MODEL.c(contains(MODEL.rxns,'EX_tyr__L_e')) = -1;
370             MODEL.c(contains(MODEL.rxns,'EX_phe__L_e')) = -1;
371             MODEL.c(contains(MODEL.rxns,'EX_met__L_e')) = -1;
372             MODEL.c(contains(MODEL.rxns,'EX_leu__L_e')) = -1;
373             MODEL.c(contains(MODEL.rxns,'EX_ile__L_e')) = -1;
374             MODEL.c(contains(MODEL.rxns,'EX_his__L_e')) = -1;
375             MODEL.c(contains(MODEL.rxns,'EX_gly_e')) = -1;
376             MODEL.c(contains(MODEL.rxns,'EX_arg__L_e')) = -1;
377             MODEL.c(contains(MODEL.rxns,'EX_cys__L_e')) = -1;
378
379        case 2
380             MODEL.c(ismember(MODEL.rxns,'MY_BIOMASS_cho_producing_1')) ...
                 = 0;%biomass synthesis
381             MODEL.c(contains(MODEL.rxns,'EX_glc__D_e')) = 0;
382             MODEL.c(contains(MODEL.rxns,'EX_ser__L_e')) = 0;
383             MODEL.c(contains(MODEL.rxns,'EX_gln__L_e')) = -1;
384             MODEL.c(contains(MODEL.rxns,'EX_val__L_e')) = -1;
385             MODEL.c(contains(MODEL.rxns,'EX_ala__L_e')) = -1;
386             MODEL.c(contains(MODEL.rxns,'EX_tyr__L_e')) = -1;
387 %             MODEL.c(contains(MODEL.rxns,'EX_phe__L_e')) = -1;
388             MODEL.c(contains(MODEL.rxns,'EX_met__L_e')) = -1;
389             MODEL.c(contains(MODEL.rxns,'EX_leu__L_e')) = -1;
390             MODEL.c(contains(MODEL.rxns,'EX_ile__L_e')) = -1;
391             MODEL.c(contains(MODEL.rxns,'EX_his__L_e')) = -1;
392             MODEL.c(contains(MODEL.rxns,'EX_gly_e')) = -1;
393             MODEL.c(contains(MODEL.rxns,'EX_arg__L_e')) = -1;
394             MODEL.c(contains(MODEL.rxns,'EX_cys__L_e')) = -1;
395
396        case 3
```

```
397             MODEL.c(ismember(MODEL.rxns,'MY_BIOMASS_cho_producing_1')) ...
                    = 0;%biomass synthesis
398             MODEL.c(contains(MODEL.rxns,'EX_glc__D_e')) = 0;
399             MODEL.c(contains(MODEL.rxns,'EX_ser__L_e')) = 0;
400             MODEL.c(contains(MODEL.rxns,'EX_gln__L_e')) = -1;
401             MODEL.c(contains(MODEL.rxns,'EX_val__L_e')) = -1;
402             MODEL.c(contains(MODEL.rxns,'EX_ala__L_e')) = -1;
403             MODEL.c(contains(MODEL.rxns,'EX_tyr__L_e')) = -1;
404             MODEL.c(contains(MODEL.rxns,'EX_phe__L_e')) = -1;
405             MODEL.c(contains(MODEL.rxns,'EX_met__L_e')) = -1;
406             MODEL.c(contains(MODEL.rxns,'EX_leu__L_e')) = -1;
407 %             MODEL.c(contains(MODEL.rxns,'EX_ile__L_e')) = -1;
408             MODEL.c(contains(MODEL.rxns,'EX_his__L_e')) = -1;
409 %             MODEL.c(contains(MODEL.rxns,'EX_gly_e')) = -1;
410             MODEL.c(contains(MODEL.rxns,'EX_arg__L_e')) = -1;
411             MODEL.c(contains(MODEL.rxns,'EX_cys__L_e')) = -1;
412
413         otherwise
414             break
415         end
416     end
417 %% SOLVE Linear Programming problem
418
419      solution_1 = optimizeCbModel(MODEL);
420
421
422      if strcmpi(solution_1.origStat,'INFEASIBLE') || ...
            strcmpi(solution_1.origStat,'NUMERIC')
423          display('Infeasible solution')
424          my_cost = inf;
425          return
426      else
427 %         f = MODEL.c;
428 %         A = [];
429 %         b = [];
430 %         Aeq = MODEL.S;
431 %         beq = MODEL.b;
432 %         lb = MODEL.lb;
433 %         ub = MODEL.ub;
434 %       options = optimoptions(@linprog);
435 %       [solution_1,fval,exitflag,output,lambda] = ...
        linprog(f,A,b,Aeq,beq,lb,ub,options)
436 %         G_DCBM_flux_vectors(:,3,j) = solution_1;
437          G_DCBM_flux_vectors(:,3,j) = solution_1.x;
```

```
438         G_DCBM_opt_solution_infos{j,2} = solution_1;
439         L2_norm = solution_1.x'*solution_1.x;
440         G_DCBM_opt_solution_infos{j,3} = L2_norm;
441         %One can add necessary information to be stored from the
442         %optimization protocol
443     end
444 %% DFVA
445 % %     MODEL = changeRxnBounds(MODEL, {'MY_BIOMASS_cho_producing_1'}, ...
        solution_1.x(ismember(MODEL.rxns,'MY_BIOMASS_cho_producing_1')), 'b');
446 % %     G_DCBM_flux_vectors(:,1,j) = MODEL.lb;
447 % %     G_DCBM_flux_vectors(:,2,j) = MODEL.ub;
448 %     formatSpec = '\n## \n**The FVA for the predicted growth rate is ...
        started at time\n t_initial = %.2f hour\n';
449 %     fprintf(formatSpec,DCBM_time_vec(j))
450 %     optPercentage =100; osenseStr= 'max'; printLevel =1; allowLoops = ...
        false;
451 % %     [G_DCBM_flux_vectors(:,1,j), G_DCBM_flux_vectors(:,2,j)] = ...
        fluxVariability(MODEL, optPercentage, osenseStr, MODEL.rxns, ...
        printLevel, allowLoops, [], [], [])
452 %     [FVA_FminFmax_vector(:,1,j),FVA_FminFmax_vector(:,2,j)] = ...
        fluxVariability(MODEL);
453 %
454 %% Reading out the predicted growth and optimum fluxes from DFBA
455     pred_uptake = ...
            G_DCBM_flux_vectors(ismember(MODEL.rxns,pred_EXflux),3,j);
456     sorted_pred_uptake = pred_uptake(sorted_pred_uptake_index);
457     mu = sorted_pred_uptake(ismember(sorted_pred_EXflux,'EX_BIOM_e'));
458 %     sorted_pred_uptake(ismember(sorted_pred_EXflux,'EX_BIOM_e')) = [];
459     %Step size (1 hour)
460
461         if j < length(DCBM_time_vec)
462         %The place where the formula for integration over biomass can ...
              be modified
463         biom_pred_data(j+1) = biom_pred_data(j)*exp(mu*(Δ_t));%biomass ...
              trajectory
464
465         %Update concentrations
466         pred_conc_trajectory(j+1,:) = pred_conc_trajectory(j,:) + ...
              0.001*sorted_pred_uptake'*biom_pred_data(j+1)*(Δ_t);
467
468         %update the cost
469         my_cost = my_cost + ...
470             (biom_pred_data(j+1) - ...
                 conc_sim_data(j+1,ismember(sorted_pred_mets,'BIOM_e'))).^2 ...
```

```
                +...
471             (pred_conc_trajectory(j+1,ismember(sorted_pred_mets,'glc__D_e'))-conc_sim_da
                /(max(conc_sim_data(:,ismember(sorted_pred_mets,'glc__D_e')))-min(conc_si
                + ...
472             (pred_conc_trajectory(j+1,ismember(sorted_pred_mets,'gln__L_e'))-conc_sim_da
                /(max(conc_sim_data(:,ismember(sorted_pred_mets,'gln__L_e')))-min(conc_si
                ...
473             (pred_conc_trajectory(j+1,ismember(sorted_pred_mets,'glu__L_e'))-conc_sim_da
                /(max(conc_sim_data(:,ismember(sorted_pred_mets,'glu__L_e')))-min(conc_si
                ...
474             (pred_conc_trajectory(j+1,ismember(sorted_pred_mets,'lac__L_e'))-conc_sim_da
                /(max(conc_sim_data(:,ismember(sorted_pred_mets,'lac__L_e')))-min(conc_si
                ...
475             (pred_conc_trajectory(j+1,ismember(sorted_pred_mets,'nh4_e'))-conc_sim_data(
                /(max(conc_sim_data(:,ismember(sorted_pred_mets,'nh4_e')))-min(conc_sim_c
                ...
476             (pred_conc_trajectory(j+1,ismember(sorted_pred_mets,'ala__L_e'))-conc_sim_da
                + ...
477             (pred_conc_trajectory(j+1,ismember(sorted_pred_mets,'arg__L_e'))-conc_sim_da
                + ...
478             (pred_conc_trajectory(j+1,ismember(sorted_pred_mets,'cys__L_e'))-conc_sim_da
                + ...
479             (pred_conc_trajectory(j+1,ismember(sorted_pred_mets,'gly_e'))-conc_sim_data(
                + ...
480             (pred_conc_trajectory(j+1,ismember(sorted_pred_mets,'his__L_e'))-conc_sim_da
                + ...
481             (pred_conc_trajectory(j+1,ismember(sorted_pred_mets,'met__L_e'))-conc_sim_da
                + ...
482             (pred_conc_trajectory(j+1,ismember(sorted_pred_mets,'phe__L_e'))-conc_sim_da
                + ...
483             (pred_conc_trajectory(j+1,ismember(sorted_pred_mets,'ser__L_e'))-conc_sim_da
                + ...
484             (pred_conc_trajectory(j+1,ismember(sorted_pred_mets,'tyr__L_e'))-conc_sim_da
                + ...
485             (pred_conc_trajectory(j+1,ismember(sorted_pred_mets,'val__L_e'))-conc_sim_da
                + ...
486             (pred_conc_trajectory(j+1,ismember(sorted_pred_mets,'ile__L_e'))-conc_sim_da
                /(max(conc_sim_data(:,ismember(sorted_pred_mets,'ile__L_e')))-min(conc_si
                ...
487             (pred_conc_trajectory(j+1,ismember(sorted_pred_mets,'leu__L_e'))-conc_sim_da
488
489         end
490
491 end
```

```
%to calculate Normalized Root Mean Square Error (NRMSE)
my_cost = sqrt(my_cost/(18*length(DCBM_time_vec)));
    formatSpec = 'my_cost value is %.2f \n';
    fprintf(formatSpec,my_cost)


end
%% Save output
save G_DCBM_dFBA_cnd_3_April_18 G_DCBM_flux_vectors ...
    pred_conc_trajectory biom_pred_data DCBM_time_vec
```

Visualization

```
%% Implementation of G-DCBM on Updated iCHOv1_DG44/After:04/April/2022/
%% Mario Jolicoeur Lab
%% Author: r.yasemi@gmail.com
% function my_cost = G_DCBM_MY_MJ_cndns_April_01_simulation_spare3(inn)
close all
clear all
do_PLOT = false;
do_INT_FLUX = true;
global COND_INDEX DFBA_init_time_index DFBA_fin_time_index sorted_pred_mets
    COND_INDEX = 1;
    switch COND_INDEX
        case 1
            load G_DCBM_dFVA_cnd_1_April_12 G_DCBM_flux_vectors ...
                FVA_FminFmax_vector pred_conc_trajectory biom_pred_data ...
                DCBM_time_vec
            DFBA_init_time_index = 20;
            DFBA_fin_time_index = 128;
            DFBA_shift_time = 125;
        case 2
            load G_DCBM_dFVA_cnd_2_April_14 G_DCBM_flux_vectors ...
                FVA_FminFmax_vector pred_conc_trajectory biom_pred_data ...
                DCBM_time_vec
            DFBA_init_time_index = 20;
            DFBA_fin_time_index = 128;
%             DFBA_fin_time_index = 144;
            DFBA_shift_time = 116;
        case 3
```

```
24              load G_DCBM_dFVA_cnd_3_April_14 G_DCBM_flux_vectors ...
                    FVA_FminFmax_vector pred_conc_trajectory biom_pred_data ...
                    DCBM_time_vec
25              DFBA_init_time_index = 20;
26              DFBA_fin_time_index = 128;
27  %             DFBA_fin_time_index = 130;
28              DFBA_shift_time = 71;
29      end
30  %% Preamble: Loading the inputs
31  % Metabolic network in COBRA format
32  %% Plotting growth dynamics
33  load DATA_CHO_clonal_GeM.mat Data_struct
34  % load MODEL_ver_9.mat MODEL
35  load MODEL_ver_10.mat MODEL
36  % Spline smoothed concentrations and exchange rates
37  load CHO_splinefitresults_AG-2014_3conditions_ver0 ...
38      sim_time_vec ...
39      sim_conc_cnd_1 sim_conc_cnd_2 sim_conc_cnd_3 ...
40      sim_conc_lb_cnd_1 sim_conc_lb_cnd_2 sim_conc_lb_cnd_3 ...
41      sim_conc_ub_cnd_1 sim_conc_ub_cnd_2 sim_conc_ub_cnd_3 ...
42      sim_rate_cnd_1 sim_rate_cnd_2 sim_rate_cnd_3 ...
43      sim_rate_lb_cnd_1 sim_rate_lb_cnd_2 sim_rate_lb_cnd_3 ...
44      sim_rate_ub_cnd_1 sim_rate_ub_cnd_2 sim_rate_ub_cnd_3
45
46  load considered_rxns.mat considered_rxns
47  %% Step 0: Tailoring the model
48  GeM_CHO = MODEL;
49  %% Step 1: Determining the secretion/uptake rates of measured ...
        extracellular species [EXCHANGE fluxes]
50  ModelFluxUnitLongFormat ='nanomolebiomolecule-per-millioncell-per-hour';
51  TransportFluxUnit = 'nanomolepermillioncellperhour';
52  %DO NOT CHANGE THIS LIST
53      common_trnspt_mets = {'BIOM_e','glc__D_e', 'gln__L_e', 'glu__L_e',...
54          'lac__L_e', 'nh4_e','o2_e',...
55          'ala__L_e','arg__L_e','cys__L_e',...
56          'gly_e','his__L_e','ile__L_e',...
57          'leu__L_e','met__L_e','phe__L_e','ser__L_e','tyr__L_e','val__L_e','igg_e'};
58
59      [sorted_mets, sorted_mets_index] = sort(common_trnspt_mets);
60      EXflux = append('EX_',sorted_mets);
61      switch COND_INDEX
62          case 1
63              conc_sim_data = sim_conc_cnd_1(:,sorted_mets_index);
64              conc_sim_datalb = sim_conc_lb_cnd_1(:,sorted_mets_index);
```

```
65              conc_sim_dataub = sim_conc_ub_cnd_1(:,sorted_mets_index);
66          case 2
67              conc_sim_data = sim_conc_cnd_2(:,sorted_mets_index);
68              conc_sim_datalb = sim_conc_lb_cnd_2(:,sorted_mets_index);
69              conc_sim_dataub = sim_conc_ub_cnd_2(:,sorted_mets_index);
70          case 3
71              conc_sim_data = sim_conc_cnd_3(:,sorted_mets_index);
72              conc_sim_datalb = sim_conc_lb_cnd_3(:,sorted_mets_index);
73              conc_sim_dataub = sim_conc_ub_cnd_3(:,sorted_mets_index);
74      end
75
76
77  %% The involved transport rates/ pay attention to the cell culture ...
        condition
78  %(Condition 1:Parental- Condition 2:LP - Condition 3:HP)
79      sim_rate_lb_ub_3Dmat = ...
            repmat(zeros(length(sim_rate_lb_cnd_1),1),[1,2,length(sorted_mets_index)]);
80      j = 0;
81      for i=1:length(sorted_mets_index)
82          switch COND_INDEX
83              case 1
84                      j = j+1;
85                      sim_rate_lb_ub_3Dmat(:,1,j) = ...
                            min([sim_rate_cnd_1(:,sorted_mets_index(i)),...
86                      sim_rate_lb_cnd_1(:,sorted_mets_index(i)),sim_rate_ub_cnd_1(:,sorted
87
88                      sim_rate_lb_ub_3Dmat(:,2,j) = ...
                            max([sim_rate_cnd_1(:,sorted_mets_index(i)),...
89                      sim_rate_lb_cnd_1(:,sorted_mets_index(i)),sim_rate_ub_cnd_1(:,sorted
90                      %This is required to make sure only the "upper ...
                            bounds" are
91                      %imposed
92                      for jj = 1:length(sim_rate_lb_cnd_1)
93                          if ...
                                sim_rate_lb_ub_3Dmat(jj,1,j)*sim_rate_lb_ub_3Dmat(jj,2,j) ..
                                > 0 ...
94                               &&  sim_rate_lb_ub_3Dmat(jj,1,j) > 0 && j ≠ ...
                                     2 && j ≠ 3 && j ≠ 4 && j ≠5
95                               sim_rate_lb_ub_3Dmat(jj,1,j) = 0;
96                          end
97                          if ...
                                sim_rate_lb_ub_3Dmat(jj,1,j)*sim_rate_lb_ub_3Dmat(jj,2,j) ..
                                > 0 ...
98                               &&  sim_rate_lb_ub_3Dmat(jj,1,j) < 0 && j ≠ 2
```

```
99                          sim_rate_lb_ub_3Dmat(jj,2,j) = 0;
100                     end
101                 end
102
103             case 2
104                 j = j+1;
105                 sim_rate_lb_ub_3Dmat(:,1,j) = ...
                        min([sim_rate_cnd_2(:,sorted_mets_index(i)),...
106                 sim_rate_lb_cnd_2(:,sorted_mets_index(i)),sim_rate_ub_cnd_2(:,sorted
107
108                 sim_rate_lb_ub_3Dmat(:,2,j) = ...
                        max([sim_rate_cnd_2(:,sorted_mets_index(i)),...
109                 sim_rate_lb_cnd_2(:,sorted_mets_index(i)),sim_rate_ub_cnd_2(:,sorted
110             for jj = 1:length(sim_rate_lb_cnd_1)
111                     if ...
                            sim_rate_lb_ub_3Dmat(jj,1,j)*sim_rate_lb_ub_3Dmat(jj,2,j) ..
                            > 0 ...
112                          &&  sim_rate_lb_ub_3Dmat(jj,1,j) > 0 && j ≠ ...
                                 2 && j ≠ 3 && j ≠ 4 && j ≠ 5
113                          sim_rate_lb_ub_3Dmat(jj,1,j) = 0;
114                     end
115                     if ...
                            sim_rate_lb_ub_3Dmat(jj,1,j)*sim_rate_lb_ub_3Dmat(jj,2,j) ..
                            > 0 ...
116                          &&  sim_rate_lb_ub_3Dmat(jj,1,j) < 0 && j ≠ 2
117                          sim_rate_lb_ub_3Dmat(jj,2,j) = 0;
118                     end
119                 end
120
121             case 3
122                 j = j+1;
123                 sim_rate_lb_ub_3Dmat(:,1,j) = ...
                        min([sim_rate_cnd_3(:,sorted_mets_index(i)),...
124                 sim_rate_lb_cnd_3(:,sorted_mets_index(i)),sim_rate_ub_cnd_3(:,sorted
125
126                 sim_rate_lb_ub_3Dmat(:,2,j) = ...
                        max([sim_rate_cnd_3(:,sorted_mets_index(i)),...
127                 sim_rate_lb_cnd_3(:,sorted_mets_index(i)),sim_rate_ub_cnd_3(:,sorted
128             for jj = 1:length(sim_rate_lb_cnd_1)
129                     if ...
                            sim_rate_lb_ub_3Dmat(jj,1,j)*sim_rate_lb_ub_3Dmat(jj,2,j) ..
                            > 0 ...
130                          &&  sim_rate_lb_ub_3Dmat(jj,1,j) > 0 && j ≠ ...
                                 2 && j ≠ 3 && j ≠ 4 && j ≠5
```

```
131                                 sim_rate_lb_ub_3Dmat(jj,1,j) = 0;
132                             end
133                             if ...
                                    sim_rate_lb_ub_3Dmat(jj,1,j)*sim_rate_lb_ub_3Dmat(jj,2,j) ..
                                    > 0 ...
134                                 &&  sim_rate_lb_ub_3Dmat(jj,1,j) < 0 && j ≠ 2
135                                 sim_rate_lb_ub_3Dmat(jj,2,j) = 0;
136                             end
137                         end
138                     end
139             end
140
141 display('**$$$** To assign lower and upper bounds of the transport ...
        reaction(s) from the extracellular data:')
142
143     %% Reading out the predicted growth and optimum fluxes from DFBA
144     pred_uptake = zeros(length(EXflux),floor(size(G_DCBM_flux_vectors,3)));
145     for i = 1:length(EXflux)
146         pred_uptake(i,:) = ...
                G_DCBM_flux_vectors(ismember(MODEL.rxns,EXflux(i)),3:3:end);
147     end
148     mu = pred_uptake(ismember(EXflux,'EX_BIOM_e'),:);
149     %Step size (1 hour)
150     Δ_t = DCBM_time_vec(3) - DCBM_time_vec(2);
151
152 % biomass
153 dumindex = find(contains(Data_struct.NameID,'BIOM[e]'));
154 dumindex_cnd_ = dumindex(COND_INDEX);
155 figure('Units', 'pixels', ...
156     'Position', [100 100 500 375]);
157 hold on;
158 figPRED     = plot(DCBM_time_vec,biom_pred_data,'k-','LineWidth',1);
159 % figSPLINE   = ...
        plot(sim_time_vec,conc_sim_data(:,ismember(pred_mets(sorted_pred_index),'BIOM_e')),'k
160 figDATA = errorbar(Data_struct.TimeVector,...
161 Data_struct.Value(:,dumindex_cnd_),...
162 Data_struct.ValueSD(:,dumindex_cnd_));
163 set(figDATA, 'LineStyle','none','Marker','.','Color',[.3 .3 .3]);
164 set(figDATA, 'LineWidth',1,'Marker','o','MarkerSize',3,...
165     'MarkerEdgeColor',[.2 .2 .2],'MarkerFaceColor',[.5 .5 .5]);
166 % Add Legend and Labels
167 switch COND_INDEX
168     case 1
169         hTitle  = title (['Case 1 - ','Growth dynamics']);
```

```
170     case 2
171         hTitle  = title (['Case 2 - ','Growth dynamics']);
172     case 3
173         hTitle  = title (['Case 3 - ','Growth dynamics']);
174 end
175 hXLabel = xlabel('Time (hr)','FontSize', 12);
176 hYLabel = ylabel({'Cell concentration';'(10^6 cell/mL)'});
177 hLegend = legend([figPRED, figDATA],...
178     'gDCBM prediction' , 'Experiment data','location', 'NorthWest' );
179 % Adjust Error Bar Width
180 figDATA_c = get(figDATA,'Children');
181     errorbarXData = get(figDATA_c,'XData');
182     errorbarXData(4:9:end) = errorbarXData(1:9:end)-0.2;
183     errorbarXData(7:9:end) = errorbarXData(1:9:end)-0.2;
184     errorbarXData(5:9:end) = errorbarXData(1:9:end)+0.2;
185     errorbarXData(8:9:end) = errorbarXData(1:9:end)+0.2;
186 set(figDATA_c, 'XData', errorbarXData);
187 % Adjust Font and Axes Properties
188 set(gca,'FontName','Helvetica');
189 set([hTitle, hXLabel, hYLabel],'FontName','AvantGarde');
190 set([hLegend, gca],'FontSize',8);
191 set(hXLabel,'FontSize',12);
192 set(hYLabel,'FontSize',12);
193 set( hTitle,'FontSize',12,'FontWeight','bold');
194 set(gca,'Box','off','FontSize',12,'TickDir','out','TickLength',[.02 ...
        .02],...
195     'XMinorTick','off','YMinorTick','off','YGrid','on',...
196     'XColor',[.3 .3 .3],'YColor',[.3 .3 .3]);
197 set(gca,'XTick',0:24:144,'LineWidth',2);
198 %Set PaperPositionMode to auto so that the exported figure looks like ...
        it does on the screen.
199 set(gcf, 'PaperPositionMode', 'auto');
200 figdum = gcf;
201 figname = ['Case',num2str(COND_INDEX),' - Cell growth'];
202 % saveas(figdum,figname,'png')
203 % saveas(figdum,figname,'pdf')
204 % print -depsc2 figname.eps
205 hold off
206
207 %% Exometabolomics dynamics (in mM)
208 if do_PLOT,
209 conc_sim_data(:,ismember(sorted_mets,'BIOM_e')) = [];
210 sorted_mets(ismember(sorted_mets,'BIOM_e')) = [];
211
```

```
212 for k = 1:length(sorted_mets)
213     if startsWith(sorted_mets{k},'igg_')
214         dummet = 'MAB';
215     elseif startsWith(sorted_mets{k},'o2_')
216         dummet = 'O22';
217     else
218         dummet = upper(extractBefore(sorted_mets{k},'_'));
219     end
220     dummet = append(dummet,'[e]');
221     dumindex = find(contains(Data_struct.NameID,dummet));
222     dumindex_cnd_ = dumindex(COND_INDEX);
223 %adjustments in the initial conidtions within the experimental error
224     switch COND_INDEX
225         case 1
226             if startsWith(dummet,'glc','IgnoreCase',true)
227                 pred_conc_trajectory(:,k) = pred_conc_trajectory(:,k) + 1;
228             end
229             if startsWith(dummet,'ala','IgnoreCase',true)
230                 pred_conc_trajectory(:,k) = pred_conc_trajectory(:,k) - ...
                        0.3;
231             end
232             if startsWith(dummet,'cys','IgnoreCase',true)
233                 pred_conc_trajectory(:,k) = pred_conc_trajectory(:,k) + ...
                        0.05;
234             end
235             if startsWith(dummet,'gln','IgnoreCase',true)
236                 pred_conc_trajectory(:,k) = pred_conc_trajectory(:,k) + ...
                        0.3;
237             end
238             if startsWith(dummet,'his','IgnoreCase',true)
239                 pred_conc_trajectory(:,k) = pred_conc_trajectory(:,k) + ...
                        0.1;
240             end
241             if startsWith(dummet,'met','IgnoreCase',true)
242                 pred_conc_trajectory(:,k) = pred_conc_trajectory(:,k) + ...
                        0.05;
243             end
244             if startsWith(dummet,'ser','IgnoreCase',true)
245                 pred_conc_trajectory(:,k) = pred_conc_trajectory(:,k) + ...
                        0.25;
246             end
247             if startsWith(dummet,'tyr','IgnoreCase',true)
248                 pred_conc_trajectory(:,k) = pred_conc_trajectory(:,k) + ...
                        0.05;
```

```
249             end
250
251         case 2
252             if startsWith(dummet,'ala','IgnoreCase',true)
253                 pred_conc_trajectory(:,k) = pred_conc_trajectory(:,k) - ...
                        0.3;
254             end
255             if startsWith(dummet,'cys','IgnoreCase',true)
256                 pred_conc_trajectory(:,k) = pred_conc_trajectory(:,k) + ...
                        0.05;
257             end
258             if startsWith(dummet,'gln','IgnoreCase',true)
259                 pred_conc_trajectory(:,k) = pred_conc_trajectory(:,k) + ...
                        0.3;
260             end
261             if startsWith(dummet,'his','IgnoreCase',true)
262                 pred_conc_trajectory(:,k) = pred_conc_trajectory(:,k) + ...
                        0.04;
263             end
264             if startsWith(dummet,'met','IgnoreCase',true)
265                 pred_conc_trajectory(:,k) = pred_conc_trajectory(:,k) + ...
                        0.04;
266             end
267             if startsWith(dummet,'phe','IgnoreCase',true)
268                 pred_conc_trajectory(:,k) = pred_conc_trajectory(:,k) + ...
                        0.02;
269             end
270             if startsWith(dummet,'ser','IgnoreCase',true)
271                 pred_conc_trajectory(:,k) = pred_conc_trajectory(:,k) + ...
                        0.25;
272             end
273             if startsWith(dummet,'tyr','IgnoreCase',true)
274                 pred_conc_trajectory(:,k) = pred_conc_trajectory(:,k) + ...
                        0.03;
275             end
276
277         case 3
278             if startsWith(dummet,'ala','IgnoreCase',true)
279                 pred_conc_trajectory(:,k) = pred_conc_trajectory(:,k) - ...
                        0.1;
280             end
281             if startsWith(dummet,'arg','IgnoreCase',true)
282                 pred_conc_trajectory(:,k) = pred_conc_trajectory(:,k) + ...
                        0.2;
```

```
283             end
284             if startsWith(dummet,'cys','IgnoreCase',true)
285                 pred_conc_trajectory(:,k) = pred_conc_trajectory(:,k) + ...
                        0.05;
286             end
287             if startsWith(dummet,'gln','IgnoreCase',true)
288                 pred_conc_trajectory(:,k) = pred_conc_trajectory(:,k) + ...
                        0.3;
289             end
290             if startsWith(dummet,'gly','IgnoreCase',true)
291                 pred_conc_trajectory(:,k) = pred_conc_trajectory(:,k) - ...
                        0.05;
292             end
293             if startsWith(dummet,'his','IgnoreCase',true)
294                 pred_conc_trajectory(:,k) = pred_conc_trajectory(:,k) + ...
                        0.02;
295             end
296             if startsWith(dummet,'leu','IgnoreCase',true)
297                 pred_conc_trajectory(:,k) = pred_conc_trajectory(:,k) - ...
                        0.1;
298             end
299             if startsWith(dummet,'met','IgnoreCase',true)
300                 pred_conc_trajectory(:,k) = pred_conc_trajectory(:,k) + ...
                        0.05;
301             end
302             if startsWith(dummet,'phe','IgnoreCase',true)
303                 pred_conc_trajectory(:,k) = pred_conc_trajectory(:,k) - ...
                        0.03;
304             end
305             if startsWith(dummet,'ser','IgnoreCase',true)
306                 pred_conc_trajectory(:,k) = pred_conc_trajectory(:,k) + ...
                        0.1;
307             end
308             if startsWith(dummet,'tyr','IgnoreCase',true)
309                 pred_conc_trajectory(:,k) = pred_conc_trajectory(:,k) + ...
                        0.05;
310             end
311
312     end
313     % concentration simulation values
314     figure('Units', 'pixels', ...
315     'Position', [100 100 500 375]);
316     hold on;
```

```
317     figPRED = ...
            plot(DCBM_time_vec,pred_conc_trajectory(:,k),'b-','LineWidth',1);
318 %     figSPLINE = plot(sim_time_vec,conc_sim_data(:,k),'b');
319     figDATA = errorbar(Data_struct.TimeVector,...
320     Data_struct.Value(:,dumindex_cnd_),...
321     Data_struct.ValueSD(:,dumindex_cnd_));
322     set(figDATA, 'LineStyle','none','Marker','.','Color',[.3 .3 .3]);
323     set(figDATA, 'LineWidth',1,'Marker','o','MarkerSize',3,...
324     'MarkerEdgeColor',[.2 .2 .2],'MarkerFaceColor',[.5 .5 .5]);
325     % Add Legend and Labels
326     switch COND_INDEX
327         case 1
328             hTitle  = title (['Case 1 - ',dummet]);
329         case 2
330             hTitle  = title (['Case 2 - ',dummet]);
331         case 3
332             hTitle  = title (['Case 3 - ',dummet]);
333     end
334 %     hXLabel = xlabel('Time (hr)');
335 %     hYLabel = ylabel({'Concentration';'(mM)'});
336     hLegend = legend([figPRED, figDATA],...
337     'gDCBM prediction' , 'Experiment data','location', 'NorthWest' );
338     % Adjust Error Bar Width
339     figDATA_c = get(figDATA,'Children');
340         errorbarXData = get(figDATA_c,'XData');
341         errorbarXData(4:9:end) = errorbarXData(1:9:end)-0.2;
342         errorbarXData(7:9:end) = errorbarXData(1:9:end)-0.2;
343         errorbarXData(5:9:end) = errorbarXData(1:9:end)+0.2;
344         errorbarXData(8:9:end) = errorbarXData(1:9:end)+0.2;
345     set(figDATA_c, 'XData', errorbarXData);
346     % Adjust Font and Axes Properties
347     set(gca,'FontName','Helvetica');
348     set([hTitle, hXLabel, hYLabel],'FontName','AvantGarde');
349     set([hLegend, gca],'FontSize',8);
350     set([hXLabel, hYLabel],'FontSize',12);
351     set( hTitle,'FontSize',12,'FontWeight','bold');
352     set(gca,'Box','off','FontSize',12,'TickDir','out','TickLength',[.02 ...
            .02],...
353     'XMinorTick','off','YMinorTick','off','YGrid','on',...
354     'XColor',[.3 .3 .3],'YColor',[.3 .3 .3]);
355     set(gca,'XTick',0:24:144,'LineWidth',2);
356     %Set PaperPositionMode to auto so that the exported figure looks ...
            like it does on the screen.
357     set(gcf, 'PaperPositionMode', 'auto');
```

```
358
359     if startsWith(sorted_mets{k},'val')
360         ylim([0 1])
361     elseif startsWith(sorted_mets{k},'tyr')
362         ylim([0 0.7])
363     elseif startsWith(sorted_mets{k},'ser')
364         ylim([0 10])
365     elseif startsWith(sorted_mets{k},'glc')
366         ylim([0 30])
367     elseif startsWith(sorted_mets{k},'gln')
368         ylim([0 5])
369     elseif startsWith(sorted_mets{k},'lac')
370         ylim([0 20])
371     elseif startsWith(sorted_mets{k},'nh4')
372         ylim([0 6])
373     elseif startsWith(sorted_mets{k},'glu')%C5H9NO4
374         ylim([0 1.2])
375     elseif startsWith(sorted_mets{k},'phe')%C9H11NO2
376         ylim([0 0.8])
377     elseif startsWith(sorted_mets{k},'met')
378         ylim([0 0.7])
379     elseif startsWith(sorted_mets{k},'leu')
380         ylim([0 0.8])
381     elseif startsWith(sorted_mets{k},'ile')
382         ylim([0 2])
383     elseif startsWith(sorted_mets{k},'o2')
384         ylim([0 20])
385     elseif startsWith(sorted_mets{k},'arg')
386         ylim([0 3])
387     elseif startsWith(sorted_mets{k},'cys')
388         ylim([0 0.8])
389     elseif startsWith(sorted_mets{k},'gly')
390         ylim([0 1.5])
391     elseif startsWith(sorted_mets{k},'his')
392         ylim([0 1.2])
393     elseif startsWith(sorted_mets{k},'ala')
394         ylim([0 3.5])
395     end
396 %
397     figdum = gcf;
398     figname = ['Case ',num2str(COND_INDEX),'-CONC-',sorted_mets{k}];
399 %     saveas(figdum,figname,'png')
400 %     saveas(figdum,figname,'pdf')
401
```

```
402     hold off
403 end
404
405 %% Dynamic Genome-scale flux distributions
406 %for measured fluxes
407 for k = 1:length(EXflux)
408     if strcmp(EXflux{k},'EX_BIOM_e')
409         continue
410     end
411     figure('Units', 'pixels', ...
412     'Position', [100 100 500 375]);
413     hold on;
414     figPRED = plot(DCBM_time_vec,pred_uptake(k,:)','b-','LineWidth',2);
415     figSPLINE = ...
            plot_ci(sim_time_vec(2:end),[zeros(length(sim_time_vec(2:end)),1),...
416          sim_rate_lb_ub_3Dmat(:,1,k),sim_rate_lb_ub_3Dmat(:,2,k)],'PatchColor', ...
                'k',...
417          'PatchAlpha', 0.1, 'MainLineWidth', 0.1, 'MainLineStyle', '-',...
418          'MainLineColor', 'k','LineWidth', 1.5, 'LineStyle','--', ...
                'LineColor', 'k');
419 %     set(figSPLINE, 'LineStyle','none','Marker','.','Color',[0.6 0.6 ...
        0.6]);
420 %     set(figSPLINE, 'LineWidth',1,'Marker','o','MarkerSize',0.1,...
421 %     'MarkerEdgeColor',[.2 .2 .2],'MarkerFaceColor',[.5 .5 .5]);
422     % Add Legend and Labels
423     switch COND_INDEX
424         case 1
425             hTitle  = title (['Case 1 - ',EXflux{k}]);
426         case 2
427             hTitle  = title (['Case 2 - ',EXflux{k}]);
428         case 3
429             hTitle  = title (['Case 3 - ',EXflux{k}]);
430     end
431 %     hXLabel = xlabel('Time (hr)');
432 %     hYLabel = ylabel({'Exchange Flux';'(nanomol/10^6cell/hr)'});
433     hLegend = legend('gDCBM prediction' , 'Flux bounds' , 'location', ...
            'NorthWest' );
434     % Adjust Font and Axes Properties
435     set(gca,'FontName','Helvetica');
436     set([hTitle, hXLabel, hYLabel],'FontName','AvantGarde');
437     set([hLegend, gca],'FontSize',8);
438     set([hXLabel, hYLabel],'FontSize',12);
439     set( hTitle,'FontSize',12,'FontWeight','bold');
```

```
440     set(gca,'Box','off','FontSize',12,'TickDir','out','TickLength',[.02 ...
            .02],...
441     'XMinorTick','off','YMinorTick','off','YGrid','on',...
442     'XColor',[.3 .3 .3],'YColor',[.3 .3 .3]);
443     set(gca,'XTick',0:24:144,'LineWidth',2);
444     uistack(figPRED,'top')
445     %Set PaperPositionMode to auto so that the exported figure looks ...
            like it does on the screen.
446     set(gcf, 'PaperPositionMode', 'auto');
447     %Y-axis limits for better visualization
448     if startsWith(EXflux{k},'EX_val')
449         ylim([-20 10])
450     elseif startsWith(EXflux{k},'EX_tyr')
451         ylim([-20 10])
452     elseif startsWith(EXflux{k},'EX_ser')
453         ylim([-60 10])
454     elseif startsWith(EXflux{k},'EX_glc')
455         ylim([-300 10])
456     elseif startsWith(EXflux{k},'EX_gln')
457         ylim([-200 10])
458     elseif startsWith(EXflux{k},'EX_lac')
459         ylim([-100 300])
460     elseif startsWith(EXflux{k},'EX_nh4')
461         ylim([-20 200])
462     elseif startsWith(EXflux{k},'EX_glu')%C5H9NO4
463         ylim([-2 15])
464     elseif startsWith(EXflux{k},'EX_phe')%C9H11NO2
465         ylim([-5 1])
466     elseif startsWith(EXflux{k},'EX_met')
467         ylim([-10 2])
468     elseif startsWith(EXflux{k},'EX_leu')
469         ylim([-10 2])
470     elseif startsWith(EXflux{k},'EX_ile')
471         ylim([-40 5])
472     elseif startsWith(EXflux{k},'EX_o2')
473         ylim([-300 20])
474     elseif startsWith(EXflux{k},'EX_arg')
475         ylim([-70 10])
476     elseif startsWith(EXflux{k},'EX_cys')
477         ylim([-20 5])
478     elseif startsWith(EXflux{k},'EX_gly')
479         ylim([-20 30])
480     elseif startsWith(EXflux{k},'EX_his')
481         ylim([-10 2])
```

```
482     elseif startsWith(EXflux{k},'EX_ala')
483         ylim([-20 50])
484     end
485     figdum = gcf;
486     figname = ['Case ',num2str(COND_INDEX),'-FLUX-',EXflux{k}];
487 %     saveas(figdum,figname,'png')
488 %     saveas(figdum,figname,'pdf')
489     hold off
490 end
491 end
492 
493 %% For plotting a subset of reactions
494 if do_INT_FLUX
495 %     dumreactions = considered_rxns(301:end);
496 % dumreactions = ...
       {'GLCt1','HEX1','PGI','PFK','FBA','TPI','GAPD','PGK','PGM','PYK','PYK6','PC','PDHm',
       'RPE',...
497 %     ...
       'PPM','PRPPS','TKT1','TKT2','CSm','ICDHym','FUMm','MDH','MDHm','CSm','PDHm','LDH_L',
       'FUMm','ENO', ...
498 %     ...
       'RPE','GLUDym','ASPTA','SPTc','SUCD1m_1','GLYCLm','MTHFC','MTHFCm','MTHFDm','MTHFR2',
499 %     'P5CDm','ILETA','GTMLTe','ME1m','MDHm','MDH',...
500 %     'SPODM','GTHS','GTHOr','GTHPi','GTHDH','GTHP_CAT',...
501 %     'ATPM','ATPS4m_cho','NADH2_u10m_cho','NADHtru', ...
       'CYOOm3_cho','CYOR_u10m_cho'};
502 
503 dumreactions = {'PFK'}
504 % dumreactions = findRxnsFromMets(MODEL, 'gln__L_c');
505 % dumreactions = ...
       {'GLCt1','GBA_cho','SBTR','GK_adp','HEX1','GLCter','r0354','r0355','r1392','RE1342C',
506 % dumreactions = {'GTHS','GTHOr','GTHPi','GTHDH','GTHP_CAT'}
507 % dumreactions = {'SPODM';'TYROX';'ILEOX';'SMOX';'SPMDOX';...
508 %     '5HOXINDACTO2OX';'XAO';'XAO2';'ASCBOX';'SPODMx';'GLYOp';...
509 %     'XAO2x';'XAOx';'SPODMm';'SPODMn'};%oxid_stress_rxns
510 selected_rxns_to_analyze_index      = ismember(MODEL.rxns,dumreactions);
511 selected_rxns_to_analyze            = ...
       MODEL.rxns(selected_rxns_to_analyze_index);
512 selected_rxns_to_analyze_formula    = ...
       printRxnFormula(MODEL,selected_rxns_to_analyze)
513 
514 for k = 1:length(selected_rxns_to_analyze)
515     %Assigning the dFBA solutions and dFVA solutions to the fluxes which
516     %are about to be plotted.
```

```
517     y_hat = ...
            reshape(G_DCBM_flux_vectors(ismember(MODEL.rxns,selected_rxns_to_analyze(k)),3,:)
518         [length(DCBM_time_vec) 1]);
519     y_hat_lb = ...
            reshape(FVA_FminFmax_vector(ismember(MODEL.rxns,selected_rxns_to_analyze(k)),1,:)
520         [length(DCBM_time_vec) 1]);
521     y_hat_ub = ...
            reshape(FVA_FminFmax_vector(ismember(MODEL.rxns,selected_rxns_to_analyze(k)),2,:)
522         [length(DCBM_time_vec) 1]);
523 %       y_hat_lb = ...
        reshape(G_DCBM_flux_vectors(ismember(MODEL.rxns,selected_rxns_to_analyze(k)),3,:),..
524 %         [length(DCBM_time_vec) 1]);
525 %     y_hat_ub = ...
        reshape(G_DCBM_flux_vectors(ismember(MODEL.rxns,selected_rxns_to_analyze(k)),3,:),..
526 %         [length(DCBM_time_vec) 1]);
527
528 figure('Units', 'pixels', ...
529     'Position', [100 100 500 375]);
530     hold on;
531     figPRED = plot(DCBM_time_vec,y_hat,'b-','LineWidth',2);
532     figSPLINE = plot_ci(DCBM_time_vec,[zeros(length(DCBM_time_vec),1),...
533         y_hat_lb,y_hat_ub   ],'PatchColor', 'k',...
534         'PatchAlpha', 0.1, 'MainLineWidth', 0.1, 'MainLineStyle', '-',...
535         'MainLineColor', 'k','LineWidth', 1.5, 'LineStyle','--', ...
                'LineColor', 'k');
536 %     set(figSPLINE, 'LineStyle','none','Marker','.','Color',[0.6 0.6 ...
        0.6]);
537 %     set(figSPLINE, 'LineWidth',1,'Marker','o','MarkerSize',0.1,...
538 %     'MarkerEdgeColor',[.2 .2 .2],'MarkerFaceColor',[.5 .5 .5]);
539     % Add Legend and Labels
540     switch COND_INDEX
541         case 1
542             hTitle  = title (['Case 1 - ...
                    ',selected_rxns_to_analyze{k},'-',selected_rxns_to_analyze_formula{k}]);
543         case 2
544             hTitle  = title (['Case 2 - ...
                    ',selected_rxns_to_analyze{k},'-',selected_rxns_to_analyze_formula{k}]);
545         case 3
546             hTitle  = title (['Case 3 - ...
                    ',selected_rxns_to_analyze{k},'-',selected_rxns_to_analyze_formula{k}]);
547     end
548     hXLabel = xlabel('Time (hr)');
549     hYLabel = ylabel({'Intracellular Flux';'(nanomol/10^6cell/hr)'});
```

```
550     legend('gDCBM optimum' , 'Alternate optima bounds' , 'location', ...
            'NorthWest' );
551     % Adjust Font and Axes Properties
552     set(gca,'FontName','Helvetica');
553     set([hTitle, hXLabel, hYLabel],'FontName','AvantGarde');
554     set([hLegend, gca],'FontSize',8);
555     set([hXLabel, hYLabel],'FontSize',12);
556     set( hTitle,'FontSize',12,'FontWeight','bold');
557     set(gca,'Box','off','FontSize',12,'TickDir','out','TickLength',[.02 ...
            .02],...
558     'XMinorTick','off','YMinorTick','off','YGrid','on',...
559     'XColor',[.3 .3 .3],'YColor',[.3 .3 .3]);
560     set(gca,'XTick',0:24:144,'LineWidth',2);
561     uistack(figPRED,'top')
562     %Set PaperPositionMode to auto so that the exported figure looks ...
            like it does on the screen.
563     set(gcf, 'PaperPositionMode', 'auto');
564
565 %     if startsWith(selected_rxns_to_analyze{k}, PAT)
566 %         DO somtheitng on the plot
567 %     end
568
569     figdum = gcf;
570     figname = ['Case ...
            ',num2str(COND_INDEX),'-DCBM-IntracellularFlux-',selected_rxns_to_analyze{k}];
571     saveas(figdum,figname,'png')
572     saveas(figdum,figname,'pdf')
573     hold off
574 end
575
576 end
577
578 return
```