UNIVERSITY OF CALIFORNIA SAN DIEGO

Systems Biotechnology of the Mammalian Secretory Pathway: Analysis of Energetic Trade-offs in Protein Production

A dissertation submitted in partial satisfaction of the
requirements for the degree
Doctor of Philosophy

in

Bioengineering

by

Jahir Mauricio Gutierrez Bugarin

Committee in charge:

Professor Nathan E. Lewis, Chair
Professor Bernhard Ø. Palsson, Co-Chair
Professor Jeffrey D. Esko
Professor Prashant Mali
Professor Christian M. Metallo

2018

The dissertation of Jahir Mauricio Gutierrez Bugarin is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

_____

_____

Co-Chair

_____

Chair

University of California San Diego

2018

DEDICATION

To Má and Pá, for all their love and support throughout my whole life.

To Mónica, for patiently waiting for me all these years as my best friend, my girlfriend, and finally now as my wife. ¡Te quiero!

EPIGRAPH

I got into science because I thought that,
with inspiration and hard work,
I could figure out how life works.
— Randy Scheckman

In order to discover something truly new,
at least one of your basic assumptions has to change
— Uri Alon

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

There are many individuals that I would like to thank for their guidance, mentorship, generosity, and support during my doctoral studies at UC San Diego. First and foremost, I would like to thank Nathan Lewis, my friend and thesis advisor, for his leadership and continuous interest in my work. He provided me with countless opportunities and taught me the value of maintaining a critical mind when approaching scientific questions. His life is an example for me and I will always be in debt to him for all his support.

I also wish to thank my thesis Co-advisor, Bernhard Palsson, for his mentorship and advice throughout my masters and doctoral studies. He taught me the value of hard work and the importance of making concise presentations. I really enjoyed being his Teaching Assistant for the BENG 212 class and the experience has inspired me to wanting to teach systems biology to young bioengineers.

I owe a great deal of gratitude to Amir Feizi. I deeply thank Amir for being my mentor and for introducing me to his techniques for modeling the secretory pathway. This dissertation and the manuscripts it comprises would not exist without his pioneering work. I owe him all my respect and I am truly fortunate to be his friend. I would especially like to thank Hooman Hefzi, whose friendship, leadership and advice were key to accomplish many of the goals in my dissertation.

It has been an honor to work alongside present and past members of the Lewis Lab: Austin Chiang, Shangzhong Li, Chintan Joshi, Philipp Spahn, Curtis Kuo, Alex Thomas, Joanne Liu, Anne Richelle, Matt Schinn, and Benjamin Kellman. I will treasure and cherish each one of the memories from our scientific discussions, lab outings, conferences, holiday parties, and food adventures. I would like to thank the researchers and staff of the Center for Biosustainability at Technical University of Denmark: Bjorn Voldborg, Helene Kildegaard, Manuel Camacho, Kim

Marquart, and Lise Grav. I have also had the pleasure to meet and learn from present and past researchers and students of the Bioengineering Department at UC San Diego, in particular: Daniel Zielinski, Cristal Zuniga, Karsten Zengler, Ali Ebrahim, Zak King, Aarash Bordbar, Edward O'Brien, and Jose Utrilla. In my second year of graduate school, I was honored to serve as President of the Bioengineering Graduate Society. I am deeply grateful to the graduate students in my cohort who helped me become a better person and leader during that time, including: Andrew Richards, Christopher Ahn, Sindhu Raghunandan, Kara Johnson, Kimberly McCabe, Nick Forsch, Tony Yen, and Shawn Hsu. Special thanks to John Watson, Jan Lenington, Helder Balelo, and Irene Jacobo for their support and for teaching me the values of leadership, kindness, and service. I am also fortunate for having Marlon Bouzy as my close friend and colleague. We successfully won the Game Changing Innovation world competition organized by Roche in 2014 and the experience forever changed my life. I thank Marlon for his ideas, his stories, and for the endless weekend hackathons.

During the summer of my fourth year as a graduate student, I moved to Boston to work for Indigo Agriculture as a Data Analysis intern. I would like to thank Gerardo Toledo and David Weisman for being great mentors and friends during that time and for teaching me about translating ideas into industry.

I am extremely grateful to CONACYT and UC Mexus for the Ph.D. scholarship they awarded me so that I could complete my studies in the United States. I have no words (in Spanish or English) to thank them enough for their financial support which helped make this dream a reality. Graduate school would have been an impossible task without the emotional support I owe to my dear friends, in particular: Ana Moreno, Sara Hariri, Harsha Kabra, Shawn Mailo, and Ali Athar. Thank you for being my extended family in San Diego.

I would also like to thank my loving parents: Juan Manuel and Tereza, as well as my brothers and best friends: Juan Manuel, Jose Miguel, and Jorge Mario. I owe you so much and I will forever love you all.

Finally, I would like to thank my loving and beautiful wife, Monica. Your undying love, patience, and perseverance have made me the man I am today. This thesis is dedicated to you.

Chapter 1 in part is a reprint of the material *Gutierrez JM, Lewis NE. Optimizing eukaryotic cell hosts for protein production through systems biotechnology and genome-scale modeling. Biotechnol. J. (2015) 10:939-949*. The dissertation author was the primary author.

Chapter 2 in part is a reprint of the material *Golabgir A*, Gutierrez JM*, Hefzi H, Li S, Palsson BO, Herwig C, Lewis NE. Quantitative feature extraction from the Chinese hamster ovary bioprocess bibliome using a novel meta-analysis workflow. Biotechnol. Advances (2016) 34:621-633*. The dissertation author was one of the two primary authors.

Chapter 3 in part is from a manuscript that is under review for publication *Gutierrez JM*, Feizi A*, Li S, Kallehauge TB, Grav LM, Hefzi H, Ley D, Hizal DB, Betenbaugh MJ, Voldborg B, Kildegaard HF, Lee GM, Palsson BO, Nielsen J, Lewis NE. Genome scale reconstruction of the mammalian secretory pathway predict metabolic costs and limitations of protein secretion*. The dissertation author was one of the two primary authors.

Chapter 4 in part is from a manuscript that is being prepared for publication *Gutierrez JM, Lewis NE. Mapping protein turnover rates onto the human secretory pathway reconstruction reveals tissue-specific modular control of pathway usage.* The dissertation author is the primary author.

Chapter 5 in part is a reprint of the material *Gutierrez JM, Lewis NE. Optimizing eukaryotic cell hosts for protein production through systems biotechnology and genome-scale modeling. Biotechnol. J. (2015) 10:939-949*. The dissertation author was the primary author.

VITA

2011	B.S. in Biomedical Engineering, University of Guadalajara, Mexico
2014	M.S. in Bioengineering, University of California San Diego
2018	Ph.D. in Bioengineering, University of California San Diego

PUBLICATIONS

1. **Gutierrez JM**, Lewis NE. Mapping protein turnover rates onto secretory pathway reconstructions reveals modular control of pathway usage. *In preparation*

2. **Gutierrez JM**, Chiaro TR, Round JL, Lewis NE. A metabolic network of human and commensal yeast elucidates the effect of diet on the mycobiome during onset and progression of ulcerative colitis. *In preparation*

3. **Gutierrez JM,** Feizi A, Li S, Kallehauge TB, Grav LM, Hefzi H, Ley D, Hizal DB, Betenbaugh MJ, Voldborg B, Kildegaard HF, Lee GM, Palsson BO, Nielsen J, Lewis NE. Genome scale reconstruction of the mammalian secretory pathway predicts metabolic costs and limitations of protein secretion. *Under review*

4. **Gutierrez JM,** Lewis NE. Optimizing eukaryotic cell hosts for protein production through systems biotechnology and genome-scale modeling. Biotechnol J (2015) 10:939-949.

5. Golabgir A, **Gutierrez JM**, Hefzi H, Li S, Palsson BO, Herwig C, Lewis NE. Quantitative feature extraction from the Chinese hamster ovary bioprocess bibliome using a novel meta-analysis workflow. Biotechnol. Advances (2016) 34:621-633.

6. Swainston N, Smallbone K, Hefzi H, Dobson PD, Brewer J, Hanscho M, Zielinski DC, Ang KS, Gardiner NJ, **Gutierrez JM**, Kyriakopoulos S, Lakshmanan M, Li S, Liu J, Martinez VS, Orellana CA, Quek LE, Thomas A, Zanghellini J, Borth N, Lee DY, Nielsen LK, Kell DB, Lewis NE, Mendes P. Recon 2.2: from reconstruction to model of human metabolism. Metabolomics (2016) 12:109.

7. Hefzi H, Ang KS, Hanscho M, Bordbar A, Ruckerbauer D, Lakshmanan M, Orellana CA, Hizal DB, Huang Y, Ley D, Martinez VS, Kyriakopoulos S, Jimenez NE, Zielinski DC, Quek LE, Wulff T, Arnsdorf J, Li S, Lee JS, Paglia G, Loira N, Spahn PN, Pedersen LE, **Gutierrez JM**, Nagarajan H, Thomas A, Abdel-Haleem AM, Zanghellini J, Kildegaard HF, Gerdtzen ZP, Betenbaugh MJ, Palsson BO,

Andersen MR, Nielsen LK, Borth N, Lee DY, Lewis NE. A community built, genome-scale reconstruction of CHO cell metabolism recapitulates cell physiology and responses to bioprocess treatments. Cell Systems (2016) 3:434-443.

8. Kuo C, Chiang A, Shamie I, Samoudi M, **Gutierrez JM**, Lewis NE. The emerging role of systems biology for engineering protein production in CHO cells. Curr. Opin. Biotechnol. (2017) 51:64-69.

ABSTRACT OF THE DISSERTATION

Systems Biotechnology of the Mammalian Secretory Pathway: Analysis of Energetic Trade-offs in Protein Production

by

Jahir Mauricio Gutierrez Bugarin

Doctor of Philosophy in Bioengineering

University of California San Diego, 2018

Professor Nathan E. Lewis, Chair
Professor Bernhard Ø. Palsson, Co-Chair

The secretory pathway is a fundamental process of eukaryotic cells and it is responsible for synthesizing, folding, and packaging thousands of membrane and secreted proteins. These proteins play important roles in cells as signaling molecules, hormones, receptors, and structural components. Today, many of the most important biotherapeutics and monoclonal antibodies are produced via the secretory pathway of animal cells in culture. Thus, it has become clear that a mechanistic understanding of the function and regulation of the secretory pathway is of prime importance for the advancement of biotechnology and bioprocessing. In this doctoral dissertation, computational methods are developed and applied to quantify the energetic burden that the secretory pathway imposes on animal cell metabolism at the systems level. First, a meta-analysis workflow to extract quantitative features from the cell bioprocessing literature is presented. These

quantitative features are consistent across studies and culture conditions and thus provide insight into fundamental properties of cell bioprocessing. Second, genomic and proteomic data are utilized to construct genome-scale computational reconstructions of the human, mouse, and Chinese hamster secretory pathways. These reconstructions are used to expand the scope of existing genome-scale metabolic networks and to investigate the energetic trade-off between cell growth and protein secretion during bioprocessing. Model simulations recapitulate bioprocess measurements and enable the quantification of energetic costs associated to cellular productivity in a product-specific manner. Finally, a mathematical expression for computing the energetic cost of protein synthesis is formulated and used to map the energetic cost landscape of a cell secretome. The energetic cost of proteins negatively correlates with protein expression levels and this negative correlation is stronger in highly secretory animal cell lines and human tissues. Finally, protein turnover rates are used to investigate the robustness and usage levels of the secretory pathway across human tissues. The degree of secretory pathway usage is linked to the degradation rate of the half-lives of secretory pathway components in a modular manner that is shaped by the product-specific demands of the secreted proteins in each cell type. The results from this analysis may help design strategies for engineering the secretory pathway in CHO cells.

Chapter 1 - Optimizing eukaryotic cell hosts for protein production through systems biotechnology and genome-scale modeling

1.1 Introduction

Eukaryotic cell lines, including Chinese hamster ovary cells, yeast, and insect cells, are invaluable hosts to produce many recombinant proteins. With the advent of genomic resources, it is now possible to leverage genome-scale computational modeling of cellular pathways to rationally engineer eukaryotic host cells. Genome-scale models of metabolism include all known biochemical reactions occurring in a specific cell. By describing these mathematically and using tools such as flux balance analysis, the models can simulate cell physiology and provide targets for cell engineering that could lead to enhanced cell viability, titer, and productivity. This introductory chapter highlights examples in which metabolic models in eukaryotic cell cultures have been used to rationally select targets for genetic modification, improve cellular metabolic capabilities, design media supplementation, and interpret high-throughput omics data. As more comprehensive models of metabolism and other cellular processes are developed for eukaryotic cell culture, these will enable further exciting developments in cell line engineering, thus accelerating recombinant protein production and biotechnology in the years to come.

Eukaryotic cells are the dominant production hosts in the therapeutic protein industry and contribute substantially to the $140 billion dollars in annual sales [1]. Common hosts, such as Chinese Hamster Ovary (CHO) cells, are particularly desirable for their capacity to fold and make human-compatible post-translational modifications on recombinant proteins [2]. As the demand for improved quantity, purity, and quality in biotherapeutic products continues to increase, novel strategies for engineering efficient eukaryotic cells become more necessary. Traditional strategies for increasing protein titers and improving cellular performance during culture relied primarily on mutant screens and bioprocess optimizations. For example, culture temperature can be lowered, or culture media can be varied to identify conditions resulting in high titers [3, 4]. However, with the advent of high throughput omic technologies and the application of computational methods in systems biology, it is now possible to elucidate the molecular basis of eukaryotic cell physiology and elucidate the mechanisms influencing their production capabilities at the genome-scale [5]. Some initial attempts to utilize metabolic networks on eukaryotic cells for metabolic engineering used dynamic modeling for estimating flux distributions [6, 7]. Such efforts involve reconstructed and refined genome-scale metabolic network models [8, 9]. These models enable the quantitative analysis of intracellular metabolic fluxes *in silico* (i.e. in a computer simulation) and the prediction of phenotype from genotype [10, 11]. Such predictions are possible since all precursors needed for synthesizing cell biomass and maintaining cell viability are produced through metabolic pathways. Thus, the metabolic fluxes directly influence cell physiology and their quantification is of great importance to bioprocess engineering [12]. Among the different methodologies, the constraint-based reconstruction and analysis (COBRA) approach has proven quite useful for studying cell metabolism at the genome scale, using algorithms such as flux balance analysis (FBA). Detailed methodologies for COBRA and its implementation for scientific computing have been thoroughly

developed over the last couple decades [13, 14]. This introductory chapter begins with a brief description of the fundamental goals of systems biotechnology as an emerging field and of COBRA as a modeling framework. Then, several research efforts that applied these models to characterize and engineer eukaryotic cell metabolism for bioprocessing are highlighted to show the recent state of the field. Finally, a roadmap for the next chapters is presented, where computational methodologies and constraint-based modeling are utilized to further the predictive capabilities of these models in bioprocessing.

1.2 Systems biotechnology and metabolic models

1.2.1 Introduction to genome-scale reconstructions

Systems biotechnology combines computational and experimental approaches to comprehensively describe the biomolecular mechanisms relevant to bioprocessing [15]. This approach frequently utilizes high-throughput omics data to study and quantify the function of specific pathways (e.g., using pathway maps [16–18], metabolic networks [19], or other interaction databases). In this context, genome-scale metabolic networks contain a comprehensive collection of all known biochemical (i.e., metabolic) information of a specific organism [20, 21]. These networks represent a structured database of the totality of known metabolic processes that take place in the cell, including the metabolites involved, the enzymes catalyzing each of the reactions and the genes that code for the necessary machinery for these processes (Figure 1.1). With the proliferation of genome sequencing efforts, many metabolic network reconstructions have been built including eukaryotic genome-scale models that are relevant to industry and medicine [22–24]. These include the filamentous fungi Saccharomyces cerevisiae [25] and Pichia pastoris [26]

for industrial applications, as well as Homo sapiens [27] and Mus musculus [28] which are important models for medicine and drug design.

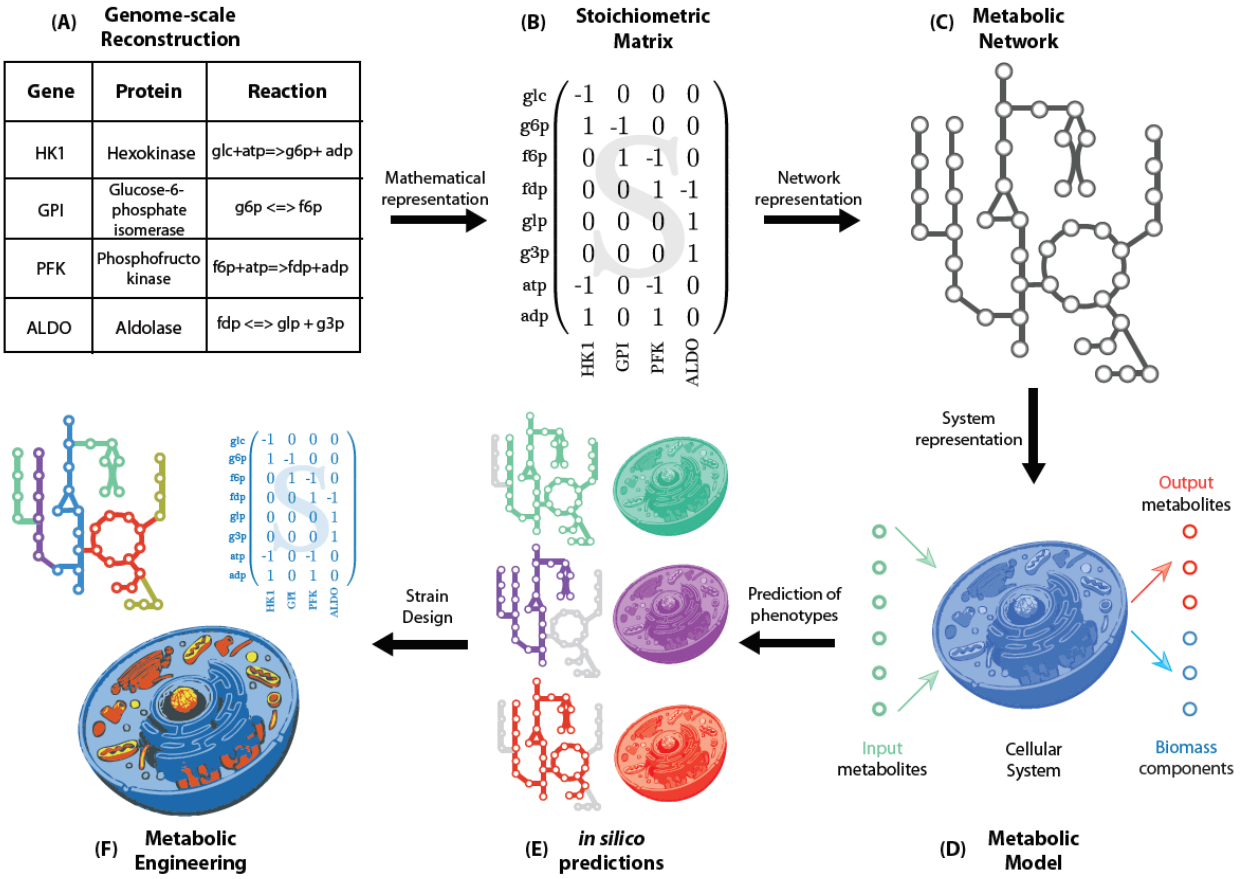

Figure 1.1: General framework for using genome-scale reconstructions in systems biotechnology. (**A**) First, the reconstruction is assembled from the organism-specific parts list (e.g. genes, proteins, metabolites, and reactions). (**B**) The metabolic reactions in the cell are described mathematically in a stoichiometric matrix, which contains the stoichiometric coefficients for each metabolite (row) in each reaction (column). (**C**) The stoichiometric matrix can be represented graphically as a metabolic network. (**D**) From the metabolic network, a system representation (i.e. metabolic model) of the cell can be obtained by identifying which metabolites are consumed or secreted, as well as the biomass components the cell needs to produce for growing (e.g. ribosomes, proteins, lipids, nucleic acids, etc.). (**E**) By using computational methods such as constraint-based analysis, different phenotypes of interest can be computed by simulating gene knock-outs or nutritional limitations in the media (represented by the different coloring patterns in the networks). (**F**) Finally, the results from these predictions serve as the basis for engineering the metabolism of the host cell towards a desired phenotype.

Biotechnological applications of genome-scale models include metabolic engineering [15], phenotype prediction and characterization [29], identification of genetic targets for cellular

engineering [30], and interpretation of high-throughput omics data [31]. Metabolic engineering of production strains has also been facilitated by in silico predictions of gene deletions, alternative metabolic pathways, metabolic coupling of growth rate with secretion of target molecule, and estimations of minimum nutrients in culture media for optimizing growth [32, 33]. Among all the different types of predictions done with metabolic models, one of particular interest to industrial biotechnology is the computation of maximum yield of a target molecule from a given substrate [24].

1.2.2 Constraint-based reconstruction and analysis of metabolic networks

To capture the biologically meaningful pathway usage, or flux distributions, of a metabolic network under a given condition in silico, it is valuable to use approaches that apply known physicochemical constraints, such as mass balance and thermodynamics of each reaction. The Constraint-Based Reconstruction and Analysis (COBRA) approach uses such constraints to narrow down the range of feasible flux distributions to recapitulate real pathway usage. COBRA further provides a diverse range of analytical tools for constructing and analyzing genome-scale metabolic networks. The networks are reconstructed by enumerating all biochemical reactions in the organism of interest. Each reaction can be described mathematically using a stoichiometric matrix, which contains the stoichiometric coefficients for each metabolite (rows in the matrix) in each reaction (columns in the matrix, see Figure 1.1B). To analyze stoichiometric networks and quantify the metabolic flux distribution of a particular phenotype, COBRA models often assume a steady-state flux and apply fundamental constraints derived from mass conservation and thermodynamics [34]. These constraints can allow for identification of steady-state flux distributions that are thermodynamically feasible and biologically meaningful. Such feasible flux distributions form a solution space, which is a mathematical space containing all possible

combinations of steady state reaction fluxes in the metabolic network (Figure 1.2, Solution Space). Once the solution space is defined, the next step is to choose an objective function, which is a reaction whose flux is sought to be maximized or minimized (see Figure 1.2, Objective and constraints). Finally, by applying linear programming algorithms, a solution that satisfies both the constraints and the objective function is computed, which provides a prediction of the flux level through each reaction. This optimization technique is commonly called Flux Balance Analysis (FBA) and it is a fundamental COBRA method [35].

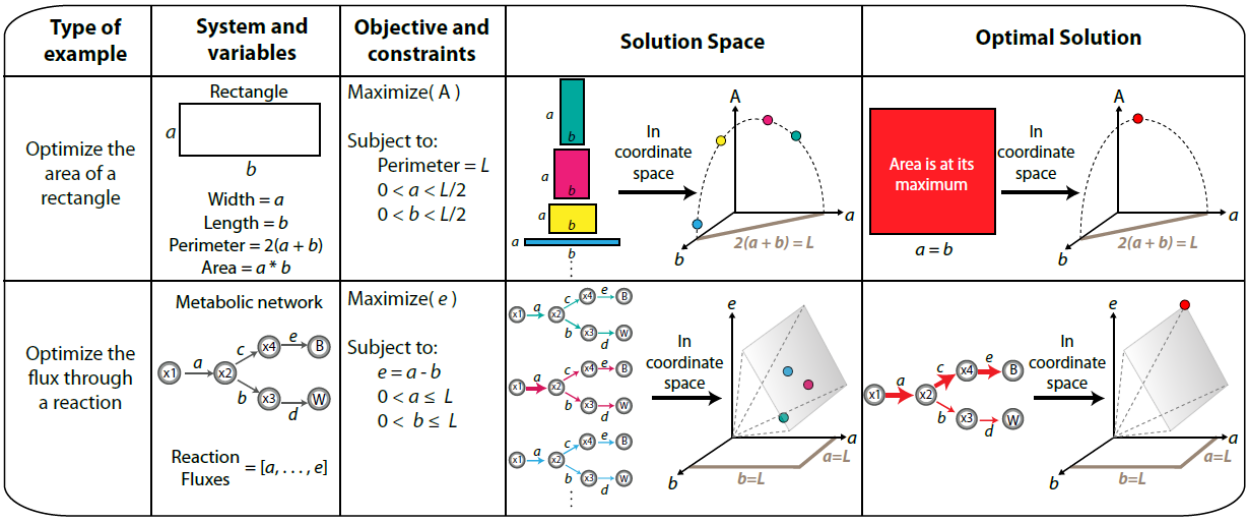

Figure 1.2: Exemplification of two optimization problems. Two examples of optimization problems are shown here to illustrate flux balance analysis. The first example appears in the context of Euclidean geometry (top row) and the second in the context of metabolic networks (bottom row). In the small metabolic network shown, x1–x4 represent intermediate metabolites, B represents produced biomass and W the secreted waste products. The arrows represent the reactions that connect the metabolites in the network and their width is proportional to the flux.

In short, FBA consists of a linear programming problem that requires: 1) the set of all biochemical reactions in the system (in the form of a stoichiometric matrix), 2) an objective function, and 3) a set of constraints that define the conditions under which the system can operate. This method will be now described conceptually with a simple optimization problem of maximizing the area of a rectangle, and then relate this to modeling metabolism (Figure 1.2).

When optimizing the area of a rectangle given a constrained perimeter, the rectangle is the system in question and this system can be described with two independent variables: the width a and the length b. The area of the rectangle in this case is the objective function, which is computed by taking the product of a and b. It is possible to construct an infinite number of rectangles by varying the values of a and b. However, the constraint requiring the perimeter of the rectangle to be the value L shortens the range of possible values that both the length and the width can take. Therefore, a solution space arises, and it is sought to identify the values of a and b that maximize the area of the rectangle. This optimal solution is obtained only when a and b are equal (i.e. when they form a square; Figure 1.2, Optimal Solution). For metabolic models, a and b are reaction fluxes of the metabolic network. The perimeter and the area of the rectangle are also fluxes of the system since their values depend on a and b. However, there exists a constraint upon the perimeter as it can only take a constant value (L). The set of all possible rectangles with perimeter L defines the solution space. Finally, the area of the rectangle represents the objective function that is sought to maximize while satisfying the given constraint. In metabolic models, a common objective function is growth which is represented by the biomass function, a pseudo reaction in which all metabolites required for the synthesis of cell parts consumed [36]. The constraints in metabolism include the directionality of the biochemical reactions or the allowed rates of substrate uptake (see bottom panel in Figure 1.2). COBRA methods have been used and implemented to study metabolism for over 30 years now and the universe of possible applications in quite vast. Many applications [11], including strategies for interpreting high-throughput omics data in the context of metabolic networks [37], have been developed and some will be highlighted in the following sections.

1.3 Applications of metabolic models in systems biotechnology for bioprocessing

Metabolic models and stoichiometric equations have been used to gain a systemic understanding of how metabolism dictates the phenotype of various eukaryotic cells in four major applications. These include media optimization, characterization of phenotypes under different culture conditions, improvement of cell density, and maximization of protein yield. A summary of these examples can be found in Table 1.1.

Table 1.1: Overview of Systems Biotechnology applications of stoichiometric equations and metabolic models presented in this introductory chapter

Reference	Organism/Cell line	Aim(s) of study	Summary of key results
[32]	CRL-1606	To construct a simplified stoichiometric network that allows for determination of material balances in animal cell metabolism and potential nutrient supplementations in culture media.	Good agreement between model predictions and experimental data covered in literature. Predictions on media supplementations turned out correct in experiments.
[44]	GS-CHO and CHO-320	To elucidate the effects of chemical composition from plant-derived supplements on the metabolic flux distribution.	The amino acid and trace element content of wheat hydrolysates induce important variations in central and amino acid metabolism of mammalian cells. Flux distributions with higher cell growth rates were found to have highly active glycine and serine metabolism.
[50]	CRL-1606	To gain a mechanistic insight into the effect of pH on mammalian cell metabolism.	Significant physiological differences between metabolic flux distributions under two pH conditions were identified.

Table 1.1: Overview of Systems Biotechnology applications (continued)

Reference	Organism/Cell line	Aim(s) of study	Summary of key results
[53]	CHO-XL99	To use a metabolic network to understand the metabolic fluxes that trigger a metabolic switch in lactate uptake and secretion.	The main differences before and after the metabolic switch were described in terms of ATP usage and redistribution through the core metabolic pathways.
[41]	in-house IgG-producing CHO cell line and CHO M250-9	To develop a framework for integrating metabolomic data into metabolic networks to gain a mechanistic insight of CHO cell physiology during fed-batch culture and identify the metabolite profile in different growth phases.	Cell-specific biomass composition may lead to erroneous *in silico* predictions if not properly calculated. Flux distributions of pentose phosphate, amino acid and fatty acid biosynthetic pathways are higher during initial exponential growth phase compared to late exponential growth phase.
[59]	GS-CHO	To contextualize the effects of sodium butyrate on cellular metabolism in a stoichiometric network in the context of low- and high-producing cell lines.	Computational predictions agree very well with experimental data and GS-CHO cell lines' metabolism was found to be characterized by high asparagine uptake and higher metabolic efficiency than other CHO cell lines. Butyrate treatment has a marked effect on increasing biosynthetic activity during stationary phase.

Table 1.1: Overview of Systems Biotechnology applications (continued)

Reference	Organism/Cell line	Aim(s) of study	Summary of key results
[61]	*Spodoptera frugiperda* Sf9 cells	To understand the cell density drop effect observed in high concentration cultures of insect cells infected with a baculovirus expression vector for recombinant protein production.	Redox homeostasis and ATP synthesis, but not byproduct accumulation nor nutrient depletion, have a drastic change after infection, which translates into cell growth arrest and higher conversion of pyruvate to acetyl-CoA.
[64]	*Spodoptera frugiperda* Sf9 cells	To optimize protein production of insect cells and bypass the cell density drop effect by identifying nutrient supplementations from a metabolic network.	It is demonstrated that supplementation of pyruvate and α-ketoglutarate has a 6 to 7-fold increase in yield.
[65]	*Saccharomyces cerevisiae*	To study the metabolic burden that heterologous protein production imposes on cell growth.	Protein secretion causes a redistribution of the carbon source in the metabolic network of yeast and thus limits growth.
[69]	*Pichia pastoris X-33-hSOD*	To engineer central metabolism of *P. pastoris* to enhance protein production by identifying beneficial mutations (i.e. gene knockouts, gene overexpression) via *in silico* predictions.	The genome scale model used in this study ([67]) accurately predicts flux changes caused by recombinant protein secretion. About 50% of the single gene mutations significantly improved recombinant protein production.

1.3.1 Identifying effective cell culture media supplementations

The metabolic phenotype of mammalian cell systems often involves high levels of glucose and glutamine uptake and excessive lactate secretion [38]. Thus, mammalian cell culture media includes nutrients that promote both cell growth and the synthesis of the target recombinant protein. Some nutrients, like essential amino acids, vitamins and inorganic salts, cannot be synthesized from the basic carbon (e.g. glucose) and nitrogen (e.g. glutamine) sources. Other nutrients can be synthesized from basic nutrient sources and their supplementation prevents the excessive accumulation of harmful metabolic byproducts (e.g., ammonia in the case of nonessential amino acids). Based on this idea, Xie and Wang formulated a stoichiometric metabolic model to study the nutritional demands for cell growth and protein production in mammalian cell cultures [39]. Using measured cell composition data, the model allowed them to determine the coefficients of a stoichiometric equation governing cell growth. The stoichiometric equation accounts for energy production and synthesis of carbohydrates, lipids, nucleotides and proteins. The researchers subsequently used these results to develop a new medium that allowed for a dramatic improvement in product titers when used in fed-batch cultures of a CRL-1606 hybridoma cell line [33]. Years later, researchers brought this in silico approach for the determination of medium supplementation to a higher level of complexity by incorporating multivariate statistical analysis and data preprocessing [40]. This allowed for the inference of optimal amino acid concentrations that could be incorporated into the nutrient medium. Furthermore, some negative correlations between non-essential amino acids and cell growth were found, suggesting a way to increase cell viability by reducing the concentrations of some media components [4, 41].

Another common media supplementation in CHO cell cultures includes plant-derived protein hydrolysates from soy, rice or wheat [42]. These supplements support cellular growth and productivity as they serve as raw materials for protein biosynthesis. However, plant-derived protein hydrolysates suffer from high compositional variability which translates into unpredictable culture performance and final product quality. Researchers investigated this issue from a systems biology perspective to elucidate the effects of wheat hydrolysates' composition on the metabolic flux distribution of CHO cells [43]. Based on a CHO-320 metabolic network [44], the researchers constructed a constraint-based metabolic model and applied FBA to estimate the metabolic fluxes in cultures with different wheat hydrolysate supplementations. Then, by using principal component analysis (PCA) and partial least squares (PLS), they interpreted the results obtained from FBA and found important characteristics in the central and amino acid metabolic pathways that varied according to the amino acid composition of wheat hydrolysates. These results confirmed the usefulness of constraint-based analysis in determining the metabolic regulation in cell cultures under different media supplementations, which have the potential to guide rational design of culture media composition and appropriate supplementations.

1.3.2 Characterizing cell physiology under different culture conditions

When cultured mammalian cells grow with excess glucose, lactate dehydrogenase activity increases, leading to a high turnover of intracellular pyruvate and subsequent secretion of lactate into the extracellular medium. As lactate accumulates, both cell growth and cell productivity decrease [45] and certain enzymes in the glycolytic pathway are downregulated [46]. Therefore, an important objective in bioprocess control is to reduce lactate secretion in mammalian cell culture. To achieve this, techniques have been proposed to modulate metabolic pathways via

genetic mutations [47] or media optimization [48]. In a recent study [49], however, researchers managed to limit lactate formation and consumption by controlling media pH in CRL-1606 hybridoma cell cultures. The researchers applied FBA to a metabolic network [50] in order to see the effect that pH had on lactate metabolism. A reaction for ATP production was chosen as the objective function and thus it was maximized in their constraint-based simulations. The results of this study led to the conclusion that hybridoma cells become more energy-efficient and synthesize more monoclonal antibody at low (6.8) pH levels. The authors were able to identify the consequences of pH on intracellular fluxes, particularly the activation of gluconeogenic enzymes at an unfavorable pH level of 7.8 that regulate the TCA cycle. Importantly, these consequences could not be captured in gene expression analysis under both pH conditions, which highlights the relevance of looking at metabolic fluxes through computational models.

One limitation of mammalian cell cultures is that cells sometimes experience a metabolic switch, leading to an inefficient phenotypic state, e.g. when lactate is secreted while glucose is highly consumed [50]. To understand the mechanism of this phenomenon in the context of metabolic fluxes, researchers derived a CHO XL99 cell metabolic model from a previous mouse genome scale model [51] and performed flux balance analysis (FBA) to yield a detailed analysis of the differences in flux distributions between two phenotypic states: lactate secretion, known to be metabolically inefficient, and lactate consumption, which was surprisingly found to be more energy efficient [52]. For example, by comparing the fluxes in key metabolic pathways (TCA, glycolysis), researchers found that the lactate-consuming phenotype of CHO cells represents a more efficient state, producing about six times more ATP (80% destined to cell maintenance and 20% to biomass production) compared to the high-lactate-secretion phenotype. The results of this

study highlight the power of metabolic models to interpret the consequences of phenotypic changes on cellular metabolism.

In another study, researchers presented an integrated framework to characterize the physiology of CHO cells in fed-batch cultures [53]. Their framework consists of combining fed-batch culture data, metabolomics, and in silico metabolic network modeling. This led to an in-depth study of three metabolic pathways associated with limitation of CHO cell growth. One surprising finding was the significant differences in biomass composition (i.e., fraction of lipids, amino acids, nucleic acids, etc. that make up cell biomass) across five different CHO cell lines that the researchers were able to analyze. This emphasizes the need for careful quantification of a cell line being studied, since accurate cell biomass composition is important for many modeling uses, such as media optimization. Otherwise, models may lead to spurious conclusions if biomass examination is not properly realized [54–56].

A common strategy used in CHO cell cultures to stimulate over-expression of the target protein involves treating the cells with sodium butyrate, a histone deacetylase inhibitor that arrests cell growth but sustains recombinant protein productivity [57]. Although this technique increases the specific productivity of CHO cells, it also increases the risk of apoptosis dramatically and can compromise the entire bioprocess. Metabolic models can be used to address pertinent questions on how to optimally culture CHO cells under sodium butyrate treatment. In the year 2013, researchers realized precisely this in the context of a metabolic network with 117 reactions and 24 metabolites in a glutamine synthetase (GS)-CHO cell line [58]. By integrating exometabolomic data from different clones at specific growth phases with a metabolic network, the researchers characterized important metabolic trends of GS-CHO cells that influence metabolic transitions in high- and low-producing CHO cell cultures under control and butyrate treatment conditions. Specifically, the

study reveals the metabolic efficiency of GS-CHO cells during the transition from exponential to stationary growth, and it also demonstrates a differentiated nitrogen metabolism of GS-CHO cells that is characterized by an increased uptake of asparagine for energy generation.

1.3.3 Analyzing the energetic basis of cell density to improve cell productivity

Insect cells represent a safe and effective way to produce heterologous proteins and vaccines with protein yields above 500 mg of protein per liter. In this context, baculovirus expression vectors (BEVs) are transfected into insect cell hosts and form a production platform of high volumetric productivity [59]. However, a common problem with this system is called the cell density drop effect [60]. This phenomenon refers to a significant reduction of specific productivity (i.e. mass of product produced per cell per unit time) of the insect cells when they have been infected with the BEV at high cell densities [61]. The cell density drop effect thus forces one to perform the BEV transfection at low insect cell concentrations to obtain acceptable titers. To understand what happens to insect cells' metabolism before and after BEV infection, researchers embarked on the mission of constructing a core metabolic model of the Spodoptera frugiperda Sf9 cell line and performed metabolic flux analysis on the basis of material balances under both conditions [59]. Their core model consisted of 52 internally balanced metabolites and 73 reactions, including reactions from 1) central metabolic pathways such as glycolysis, the pentose phosphate pathway and TCA cycle and 2) reactions that account for the energetic costs of biomass formation and membrane transport. Interestingly, the results of this study suggest that neither byproduct accumulation nor depletion of nutrients in the culture media are responsible for the cell density drop effect observed in insect cell cultures with high density. Nevertheless, this work sheds light on metabolic regulation occurring in insect cells after infection with BEVs. These include changes

in redox homeostasis, augmented ATP synthesis, and enhanced consumption of disaccharides after infection, thus resulting in a higher flux through the conversion of pyruvate into acetyl-CoA. Based upon these results, the same research team subsequently altered Sf9 energy metabolism combining experimental and computational methods, and successfully enhanced protein production [62]. Their strategy involved supplementing the culture media with alpha-ketoglurate and pyruvate at the time of infection, which resulted in a 6-fold increase in yield. These two studies highlight the potential of metabolic models in identifying key culture manipulations for enhancing productivity in a bioprocess, even when the information required to build a genome-scale network is not available.

1.3.4 Characterizing the energetic trends that favor protein production

Recombinant protein production in yeast is commonly increased using different strategies that range from codon usage to manipulating protein folding processes. However, increasing protein secretion has a draining effect on central metabolic fluxes. In one study of Saccharomyces cerevisiae metabolism, researchers presented a core stoichiometric model (81 metabolites, 78 reactions) of a human superoxide dismutase (SOD)-producing cell line, and used the model to calculate the metabolic flux distributions in wild type and protein-producing yeast strains [63]. The fundamental differences between both strains were captured in this study even when glucose consumption and ethanol production remained the same, the key contrasting features lie in the distribution of the carbon source to produce biomass (i.e. growth rate). The synthesis of the recombinant SOD protein was linked to higher fermentation and lower ATP synthesis compared to the wild type strain. This study successfully pin-pointed the energetic trade-off between cell growth and protein synthesis by means of a metabolic model, and thus set the foundations for

subsequent research efforts aimed to characterize yeast metabolism via comprehensive stoichiometric networks [64].

Pichia pastoris is a methylotrophic yeast that has drawn the attention of many systems biologists, since it is an effective host for heterologous protein production. Several fully compartmentalized genome-scale metabolic reconstructions of this organism are now available [26, 65, 66]. Using a previous genome-scale reconstruction [65], researchers [67] demonstrated significant changes in flux distributions of a Pichia pastoris strain when forced to produce recombinant protein. They utilized the algorithms Minimization of Metabolic Adjustment (MOMA, [68]) and Flux Scanning based on Enforced Objective Function (FSEOF, [69]) to predict appropriate genetic modifications (i.e. knockout or overexpression) that would translate into increased recombinant protein production. From there, the researchers were able to highlight the most important features of the regulatory flexibility of P. pastoris metabolic network to redirect resources for protein production thanks to the predicted genetic manipulations (see Table 1.1). This study goes to show that metabolic models not only provide powerful descriptions of yeast metabolism to enhance secretion of small molecules (e.g. succinate, sesquiterpenes [70]) but also secretion of macromolecules and polymers.

1.3 Current challenges and new modeling methodologies

The use of genome-scale metabolic models for enhancing recombinant protein production is still in its infancy. As can be inferred from the studies included in this introductory chapter, the discovery of more sophisticated and novel biotechnological strategies for enhancing recombinant protein production will rely on the refinement and analysis of these models. Some immediate areas

of research that will have the greatest impact on model-based improvements of protein secretion are as follow. First, advances that address the higher complexity and compartmentalization of metabolic processes in eukaryotes will be invaluable. Second, the physiology of only a few eukaryotes (e.g. yeast [64]) have been studied, and so continued efforts in characterizing the more complex metabolism in higher order organisms (e.g. mouse, hamster) will enable more detailed and accurate predictions with genome-scale metabolic models for these protein secretion hosts. Third, technological advances regarding the generation of complex high-throughput datasets (beyond the genome or the proteome) will further benefit work with eukaryotes. The areas of glycobiology and phosphoproteomics, when mapped to metabolic and genetic networks, will help understand how to control post-translational modification of products and better account for key regulatory events in the cell. Fortunately for CHO cells, there have been several efforts to generate these types of datasets for the N-glycoproteome [71], O-glycoproteome [72] and the transcriptome [73].

Major successes in the use of genome-scale models for metabolic engineering have been achieved in the development of production hosts for small molecules [74–76]. Recent expansions of these models have given place to the next generation of genome-scale models of bacteria, also known as ME-models (Metabolic and gene Expression models). These models incorporate non-enzymatic events such as transcription, translation [77, 78] as well as translocation [79], and allow for the estimation of the optimal functional proteome required by the prokaryotic cell under particular conditions [80, 81]. Although the task would be enormous, the ME model framework could be used to expand and refine eukaryotic cell models. Beyond transcription and translation and signaling, the coupling of additional process such as protein secretion and associated post-translational modifications can also benefit the development of eukaryotic protein production

hosts. For example, protein folding in the endoplasmic reticulum via chaperone activity imposes an additional energetic cost (e.g., consuming ATP, sugar nucleotides, etc.) that is not explicitly accounted in metabolic models simply because this process cannot be stoichiometrically described. The same applies to redox balancing when creating disulfide bonds in proteins or to the impact of amino acid composition on metabolic flux distributions [82, 83]. A recent study addressed these issues where the first genome-scale model of the yeast protein secretory pathway was reconstructed [84]. Furthermore, significant progress in modeling the eukaryotic glycosylation pathways has been made. These research efforts aim to gain a systemic insight of the glycosylation capabilities of cell hosts [85, 86]. The computational tools derived from these efforts could be easily incorporated into the systems biotechnology toolbox for practical applications soon. As these models continue to be deployed, it is anticipated that they will prove exceptionally valuable for engineering the next generation of protein-producing eukaryotic cell factories. Specifically, they will help identify targets for genetic modification, improve cellular metabolic capabilities, optimize media, and interpret high-throughput omics data to elucidate the biomolecular mechanisms controlling recombinant protein production yield and quality.

1.4 Toward data analysis and computational models of protein secretion

This introductory chapter has discussed case studies of successful applications of data processing and genome-scale modeling in systems biotechnology. The remaining chapters of this dissertation utilize data analysis and computational models of protein secretion for gaining insight into the energetic costs and trade-offs associated to protein production inside animal cells. In Chapter 2, a catalog of the CHO cell bioprocessing literature spanning 20 years of research is

constructed. Then, a novel meta-analysis workflow for analyzing the legacy bioprocessing data from the literature is developed. This workflow can extract quantitative features that explain the differences in bioprocess performance across culture conditions. In Chapter 3, I focus on the construction of the first genome-scale model of the mammal secretory pathway. This model enables three new extended metabolic reconstructions that account for the cost of protein secretion in Human, mouse, and the Chinese hamster. Then, applications of the reconstruction are presented in the context of studying CHO cell bioprocesses. In Chapter 4, I focus on utilizing the secretory pathway reconstructions of human and yeast to contextualize proteomic data. The analysis presented in this chapter reveals interesting properties of the secretory pathway and provides insight into the way that cells regulate the half-lives of its secretory components and the turnover rates of intracellular proteins.

1.5 Acknowledgements

This work was completed with generous support from the Novo Nordisk Foundation provided to the Center for Biosustainability at the Technical University of Denmark and a graduate fellowship from CONACYT and the University of California Institute for Mexico and the United States.

Chapter 1 in part is a reprint of the material *Gutierrez JM, Lewis NE. Optimizing eukaryotic cell hosts for protein production through systems biotechnology and genome-scale modeling. Biotechnol. J. (2015) 10:939-949*. The dissertation author was the primary author.

Chapter 2 - Quantitative feature extraction from the Chinese hamster ovary bioprocess bibliome using a novel meta-analysis workflow

2.1 Abstract

The scientific literature concerning Chinese hamster ovary (CHO) cells grows annually due to the importance of CHO cells in industrial bioprocessing of therapeutics. In an effort to start to catalogue the breadth of CHO phenotypes, or phenome, we present the CHO bibliome. This bibliographic compilation covers all published CHO cell studies from 1995-2015, and each study is classified by the types of phenotypic and bioprocess data contained therein. Using data from selected studies, we also present a quantitative meta-analysis of bioprocess characteristics across diverse culture conditions, yielding novel insights and addressing the validity of long held assumptions. Specifically, we show that bioprocess titers can be predicted using indicator variables derived from viable cell density, viability, and culture duration. We further identified a positive correlation between the cumulative viable cell density (VCD) and final titer, irrespective of cell line, media, and other bioprocess parameters. In addition, growth rate was negatively correlated with performance attributes, such as VCD and titer. In summary, despite assumptions that technical diversity among studies and opaque publication practices can limit research re-use in this

field, we show that the statistical analysis of diverse legacy bioprocess data can provide insight into bioprocessing capabilities of CHO cell lines used in industry.

2.2 Introduction

The scientific Chinese hamster ovary (CHO) cells have been utilized for academic and industrial purposes since the 1950s [87]. Today CHO cells represent the preferred cellular factory for the production of important recombinant proteins and biotherapeutics [1], including six of the top ten selling biotherapeutics in 2014. In the early days of recombinant protein production, the complexity of desired products required a mammalian host. Thus, new CHO cell lines were developed, and an entire new field of CHO bioprocessing was born. Throughout the history of CHO cell culturing, major technological advances have continued to expand its use in industry.

Over the past two decades, there has been a steady increase in the number of published studies on CHO cell culturing and bioprocessing. However, we still do not fully understand the factors that determine the optimal performance of CHO cells during culture. A deeper understanding of these factors may possibly be obtained from the retrospective analysis of large amounts of carefully collated and curated legacy data on CHO cells. Examples of questions that could be explored with the use of an organized repository of CHO bioprocessing data include the following. What are the main phenotypic differences across CHO cell lines (e.g., CHO-K1, DG44, DUKXB11, CHO-S, etc.)? How do the different culture conditions affect the performance of cells? Is it possible to predict titer, viable cell density (VCD) or viability over time given appropriate information on the cell line and culture conditions? Are there any significant differences between parental cell lines that ultimately translate into a sustained effect in culture performance? Answers

to these and many other questions can have important implications on CHO cell bioprocessing and help improve recombinant protein quality.

In an initial step to explore such questions, we compiled and curated the literature between January 1995 and June 2015 and identified studies containing biotech-relevant data on CHO cells. In addition, we classified each article based on the type of data it contains. Next, we extracted the detailed experimental data from a sample of 74 articles [88–161] that contain relevant data on CHO cell phenotype, culture performance and production characteristics. Through several statistical analyses, we identified significant trends across bioprocesses corresponding to specific attributes, such as parental cell lines, culture conditions, growth rates, production capabilities, and other research parameters. While it is often assumed that technical variation and opaque publication practices limit research re-use in this field, here we successfully integrate data from diverse studies to quantitatively validate long-held assumptions in bioprocessing. Thus, the collation and analysis of the ever-increasing data on CHO bioprocessing can provide valuable insights for future bioprocessing efforts.

2.3 Methods

The methodology for realizing the presented quantitative review involved two main phases: 1) bibliographic compilation of scientific literature on CHO, or "bibliome", along with the extraction and digitization of the metadata to be used for 2) statistical analysis. Figure 2.1 illustrates the step-by-step workflow and detailed descriptions about each step are provided in the following sections.

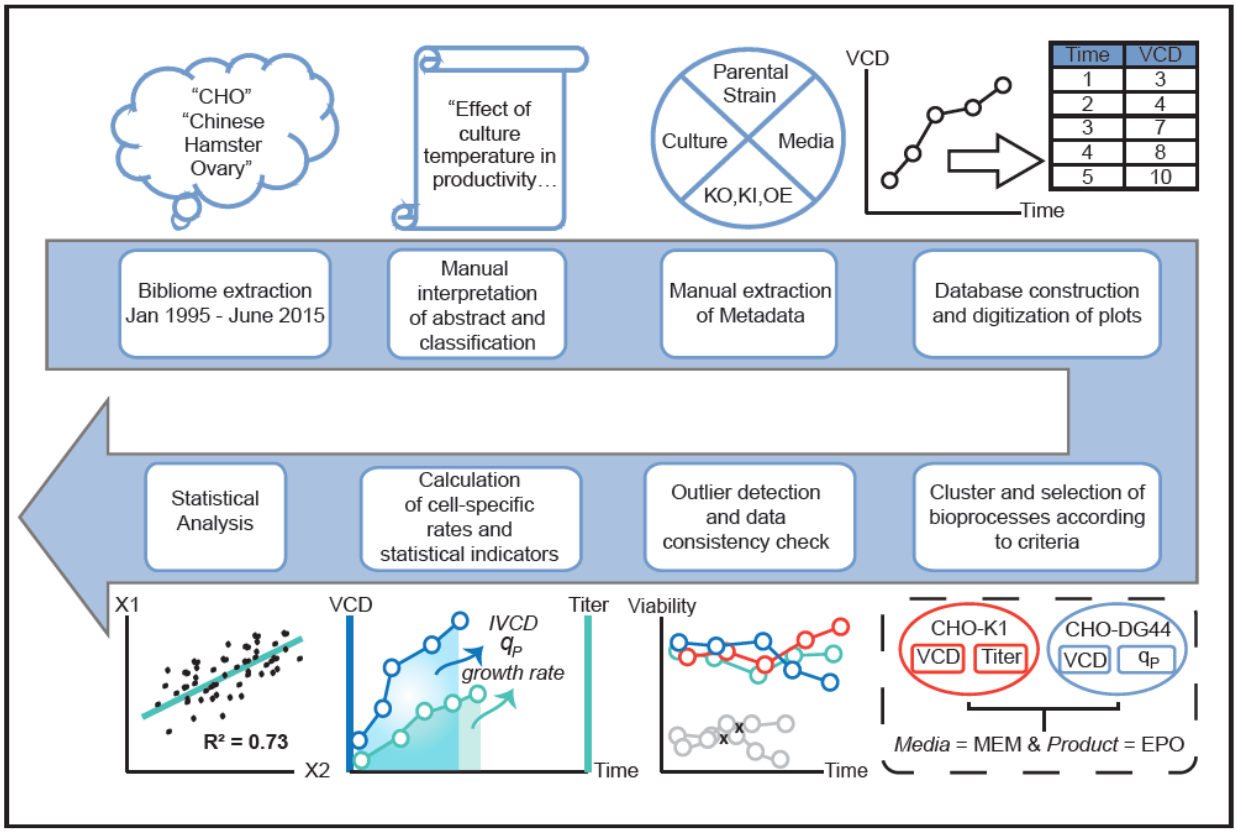

Figure 2.1: Overview of the meta-analysis workflow. The abstracts of all studies published between 1995 and July 2015 containing the keywords "CHO cells" or "Chinese Hamster Ovary" were downloaded from Thomson Reuters Web of Science. Each abstract was reviewed and classified according to its relevance and data type provided. Once the appropriate information was extracted and the corresponding figures digitized, the time series data were organized into bioprocesses, appropriate filters and outlier detection methods were applied, and calculations of cell-specific reaction rates were performed. Finally, statistical indicators were applied, and the data were analyzed using a variety of univariate and multivariate statistical methods. KO = gene knockout, KI = gene knock-in, OE = gene overexpression, VCD = viable cell density, qP = specific productivity, MEM = minimal essential media, EPO = erythropoietin, IVCD = integral of viable cell density.

2.3.1 Identification and selection of publications

Thomson Reuters Web of Science was queried to search for all research articles published between January 1995 and June 2015 that contained the keywords "CHO cells" and/or "Chinese hamster ovary" in the title or abstract. Although the first mention of CHO cells in the scientific

literature dates back to 1958 [87] we focused here on studies published within the last 20 years to focus more on CHO cell bioprocesses that employ current technologies. This initial set of articles was then manually filtered by removing any study involving characterization of a recombinant protein expressed in CHO for basic science purposes (e.g., localization, interaction within the cell, effects of mutations or consequences of exposure to UV light or radiation).

2.3.2 Extraction of metadata

Most articles in our bibliome utilize graphs and time course plots to present the results. Thus, WebPlotDigitizer [162] was used to digitize the data contained in the corresponding articles of our sample. From here on, the data extracted from these figures and the associated meta-features will be referred as the **Phenotype and Production Characteristics dataset**. To make the proposed analysis comprehensive, we manually annotated each article and figure with experimental details that may influence cell phenotypes of interest (see Table 2.1).

Data series were grouped based on their associated metadata to facilitate subsequent analyses. To do this, we assigned a bioprocess identification number (bioprocess ID) to each data series corresponding to the same experiment. That is, a bioprocess ID was assigned to each set of data series with the same values in each of the metadata features such as **cell line**, **culture media** and **culture conditions**. Many articles contain multiple bioprocess IDs since there can be more than one bioprocess in a single study (e.g. when a study tests the performance of two cell lines under same culture conditions).

Table 2.1 Annotated features used for each study included in meta-analysis (n = 74)

FEATURE	QUESTIONS ADDRESSED	EXAMPLES
Author affiliation	Was the study conducted by an academic group or a company?	
Parental Cell Line	What is the origin of the CHO cell line used in the study?	CHO-K1
		CHO-DG44
		CHO-DUKXB11
Recombinant Protein Produced	Which protein is secreted by the CHO cells in this study?	tPA
		EPO
		Proprietary mAb
Culture Media	What type of media was used to cultivate the cells?	DMEM
		F-12
		Iscoves Medium
Media Supplementations	Were there any supplements added to the media?	Fetal calf serum
		Fetal bovine serum
		Non-essential amino acids
Selection Method	What was the method for selecting high-producing CHO cells?	Methotrexate
		Zeocin
		Geneticin
Genetic manipulations	Were any genes silenced, inserted or overexpressed in CHO cells for the study?	Knock-in
		Knock-out
	Were any genes switched to be constitutively expressed?	Overexpressed
		Inducible expression
Culture type	How were the CHO cells cultured?	Batch
		Fed-batch
		Continuous
Culture Conditions	Were there any deviations from standard culture conditions?	Temperature shifts
		Low/high pH
		Osmolality
	Were there any chemical treatments?	Addition of NaBu
		Addition of sorbitol

Table 2.1 Annotated features used for each study included in meta-analysis (continued)

FEATURE	QUESTIONS ADDRESSED	EXAMPLES
Data in digitized figure or plot	What are the measured variables and what were their units?	Viable cell density [millions of cells per unit volume] versus time [days]

2.3.3 Data consistency, outlier detection, and calculations

Computations on data extracted from publications were performed using MATLAB (2014b) (Mathworks Inc., USA). Calculations for analyses on raw-variables were done using bioprocesses that contained the necessary variables and required meta-information. For instance, calculation of cellular growth rate in a bioprocess requires that VCD is known. Similarly, calculation of the specific production rate (q_P) requires knowledge of both the VCD and the titer. Whenever a bioprocess did not include important variables or when data were reported in non-standard units (e.g. % for titer), the bioprocess was eliminated from the analysis dataset. The selection criteria were adjusted according to the requirements of the analysis. For example, both VCD and titer of a bioprocess should be reported in appropriate physical units to analyze the growth and production characteristics. Once a set of bioprocesses satisfying these conditions was identified, calculations were performed. The growth rate μ (in units of 1/h) was calculated from the VCD [cells/mL] variable according to Eq. 2.1.

$$\mu(t) = \frac{1}{VCD(t)} \frac{d}{dt} VCD(t) \qquad \text{(Equation 2.1)}$$

Similarly, the specific production rate, q_P [mg/cell/h], was calculated from the reported titer (or product concentration), c_P [g/mL] and VCD (Eq. 2.2).

$$q_P(t) = \frac{1}{VCD(t)} \frac{d}{dt} c_P(t) \qquad \text{(Equation 2.2)}$$

The derivatives of VCD and titer were computed using a Savitzky-Golay filter [163] with a window size of 3 data points and polynomial degree of 1. The derivative of the growth rate was used for estimating the duration of the adaptation and growth phases and the time point at which this derivative becomes negative was selected as the ending point of the exponential growth phase.

To obtain a reliable dataset, outlier bioprocesses were detected and removed. Visual inspection of the dataset revealed that some reported values were unlikely (e.g., the end value of some VCD curves was 20 times higher than the median of all other bioprocesses). For detecting such extreme cases, an outlier removal test based on median absolute deviation [164] was implemented with lenient cutoffs of greater than 20. The outlier removal method was applied on the median of VCD and growth rate on the level of bioprocesses. Outliers detected with the cutoff of 20 were attributed to errors during digitization or incorrect axis labels and could be corrected (only 2.1% of VCD data points were detected as outliers). To remove outliers in individual variables, the Hampel filter [103] method was applied using a moving-window approach with a window size of 3 and a cutoff value of 3. Each signal value at a time-point was compared to the median of itself and its adjacent neighbors. Values with absolute distances higher than three times the median absolute deviation were replaced by the median. Replacing outliers with the median was chosen to avoid potential reduction of the number of data points. The Hampel filter has been shown to be an effective and statistically robust measure against potential outliers [103]. Thus, this

approach allowed the elimination of outlier data points that would otherwise distort the quality of extracted information.

2.3.4 Data consistency, outlier detection, and calculations

The dataset consists of multiple experiments, each containing several variables reported over time; thus, the dataset has three modes: experiment, variable, and time. To apply existing 2D statistical methods, such as pairwise correlation analysis and Partial Least Squares (PLS) regression, the time mode was eliminated by either taking the value of a variable at a given time point or calculating indicator variables over a specified time range using statistical operators, such as the mean or the maximum value. Table 2.2 contains the list of indicator variables that are related to process variables (X) in addition to variables that are related to quality attributes or production aspects (Y). Prior to the application of multivariate analysis methods, variables with high skewness were transformed approximately into a normal distribution according to the functions specified in Table 2. The criterion for deciding which variables needed transformation was based on a combination of skewness statistics and minimum and maximum values [110]. For calculating some of the indicator variables, the duration of the growth phase in each bioprocess (t_{growth}) was determined by calculating the derivative of VCD and detecting the time point corresponding to the value of the derivative becoming zero. Thus, the duration of the growth phase (t_{growth}) is the time it takes to reach maximum VCD. The total duration of each culture is indicated by t_{end}, which can include stationary phase and culture die-off. Culture duration (t_{end}) is greater or equal to the duration of growth phase (t_{growth}).

2.3.5 Statistical analysis

To determine whether different classes of bioprocesses (e.g. different parental cell line or process mode), are different with respect to indicator variables, the nonparametric Kruskal–Wallis test [165] was used under the assumption that the groups have identically shaped distributions with the null hypothesis being that the medians of all groups are equal. This test is suitable when the measurement variable does not meet the normality assumption of one-way ANOVA. Since we transformed skewed variables prior to testing, the assumption of identical distributions can be made, and the test can be used to assess the differences in medians or means. When significant differences were detected, the Dunn's test [166] was employed for determining which of the sample pairs were significantly different. Determination of correlations between pairs of variables was performed by calculation of the Spearman rank correlation coefficient (r_S).

PLS regression was used for relating process variables with quality attributes [167]. PLS models were calculated using the software SIMCA ver. 13.0.3.0 (Umetrics, Sweden). The X and Y variables for all PLS models, were transformed into approximate normal distributions by applying transformation functions as listed in Table 2.2. Individual models were generated for each of the three response variables listed in Table 2.2. All variables were mean centered and scaled to unit variance. For estimating the predictive power of the model, the Q^2 value was calculated by cross-validation, dividing the data into 7 parts. To determine the importance of each variable for both X and Y, Variable Importance for Projection (VIP), which is a sum of squares of the PLS weights with the weights calculated from the amount of Y-variance of each PLS component, were calculated [168].

Table 2.2: Indicator variables for summarizing bioprocess characteristics. Variables were either calculated over the specified time span or at a point. Data transformation was performed to achieve approximate normal distributions. For $Viab_{end}$ and dViab, the applied transformation changes the direction of the effect

ID	Abbreviation	Description	Units	Data transformation	Time span
X_1	μ_{ave}	Average growth rate	[1/h]	$\log(X_1)$	$[t_0, t_{growth}]$
X_2	μ_{max}	Maximum growth rate	[1/h]	$\log(X_2)$	$[t_0, t_{growth}]$
X_3	t_{end}	Duration of culture	[h]	$\log(X_3)$	t_{end}
X_4	t_{growth}	Duration of growth phase	[h]	X_4	t_{growth}
X_5	VCD_{max}	Maximum value of VCD	[cells/ml]	$\log(X_5)$	$[t_0, t_{end}]$
X_6	VCD_{end}	Final value of VCD	[cells/ml]	$\log(X_6)$	t_{end}
X_7	VCD_{int}	Integral of VCD over culture duration.	[cells*h/ml]	$\log(X_7)$	$[t_0, t_{end}]$
X_8	$Viab_{end}$	Final value of viability	[%]	$\log(100- X_8)$	t_{end}
X_9	dViab	Minimum rate of change of viability. $dViab/dt \leq 0$	[%/h]	$\log(-X_9)$	$[t_0, t_{end}]$
Y_1	P	End value of titer (product concentration)	[mg/ml]	$\log(Y_1)$	t_{end}
Y_2	q_P	Average of specific productivity	[mg/cell/h]	$\log(Y_2)$	$[t_0, t_{end}]$
Y_3	$Y_{P/X}$	Yield of product / viable cells	[mg/cell]	$\log(Y_3)$	t_{end}

2.4 Results

2.4.1 The CHO bioprocessing bibliome

The initial raw list of publication abstracts downloaded from Thomson Reuters Web of Science included 10,279. After filtering out non-bioprocessing CHO studies, a final list of 1157 biotech-relevant articles were included in the final CHO bibliome and each of these studies was classified into one or more of the 16 categories listed in Table 2.3 according to the type of information contained. Figure 2.2A reveals the evolution of the trends in publication types over time and the predominance of the "Phenotype and Production characteristics" category. The stacked bars of Figure 2A are grouped in three sets to capture interesting features across time periods and to ease visualization. The bottom set (1995-2001) is characterized by having a large proportion of enzyme-related analysis and review articles. The middle set (2002-2008) shows a "dip" in the number of articles relevant to CHO bioprocessing published annually. The latter part of this period is characterized by a slow increase in the number of studies using high-throughput data (e.g. gene expression, proteomics). Finally, as expected, the omics studies on CHO cells are more abundant in the top set (2009-2015) due to the appearance of novel technologies for DNA sequencing, mass spectrometry and metabolomics. Note that the number of articles in the year 2015 does not include studies published after the month of June of that year.

A total of 618 studies were classified under the **Phenotype and Production Characteristics** category. With respect to the authorship of the corresponding authors, 23% of these 618 articles were led by a corresponding author affiliated to industry whereas the rest were led by a corresponding author affiliated to an academic institution (see Figure 2.3A). From this category, 74 articles were included in the meta-analysis. The distribution of cell lines reported in

these 74 articles is like that of the bibliome. Furthermore, these articles provided ample diversity of bioprocess information such as cell viability, VCD, titer, specific productivity, growth rate, and specific consumption/secretion rates of important metabolites (e.g. glucose, glutamine, lactate and ammonia). From here, 529 relevant figures were identified, and these yielded a total of 1489 data series after digitization.

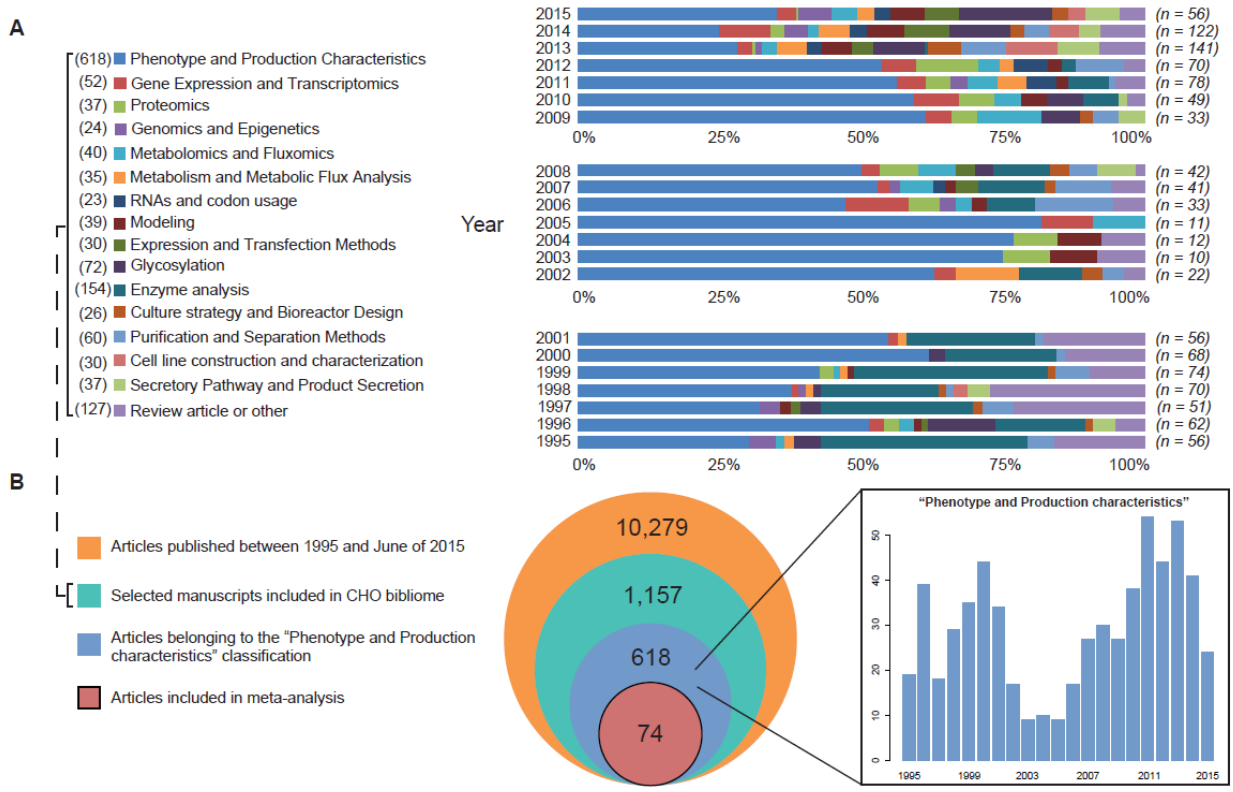

Figure 2.2: Graphical representation of the CHO bibliome. (A) Distribution of the number of articles in each category of the CHO bibliome over time as a percentage of the annual total. **(B)** Schematic of revised literature used to construct the bibliome. The orange set represents the total number of articles downloaded from Web of Science. The green set contains the studies included in the CHO bibliome. The blue set contains all studies under the "Phenotype and Production Characteristics" category and an additional bar chart shows the distribution of the number of articles in this category over time. Finally, the red set contains the articles included in our meta-analysis. The numbers inside each circle indicate the size of the corresponding set. The insert shows a bar plot of the number of articles in the "Phenotype and Production characteristics" category over time.

Table 2.3: Categories used to classify relevant research articles in the CHO bibliome according to the data contained in each study. Studies were assigned to more than one category when relevant.

Category	Examples of data contained within category	Total number of studies in category
Phenotype and Production Characteristics	➔ Yield (mass of target protein per mass of substrate) over culture time ➔ Titer (mass of target protein per unit of culture volume) over culture time ➔ Cellular percentage viability over time ➔ Effects of culture conditions on protein secretion ➔ Dependence of growth rate on culture temperature, pH, osmolarity ➔ Changes in viable cell density according to culture media supplementations ➔ Cellular adaptation to different concentrations of methotrexate during selection	618
Enzyme Analysis	➔ Functional characterization of specific enzymes that are important in CHO cells	154
Purification and Separation Methods	➔ Recombinant mAb purification methods in CHO cultures ➔ Analysis of recombinant protein composition after separation	60

Table 2.3: Categories used to classify relevant research articles in the CHO bibliome according to the data contained in each study (continued)

Category	Examples of data contained within category	Total number of studies in category
Proteomics	➔ Prediction of presence of a protein in CHO cells from genetic sequences ➔ Effects of culture conditions on proteome composition	37
Gene expression and Transcriptomics	➔ Differential gene expression under different culture conditions (e.g. treatment with NaBu) ➔ Analysis of CHO transcriptome	52
Glycosylation	➔ Glycomics profiling for different CHO cell lectin mutants ➔ Effects of culture vessel volume and revolutions per minute on glycosylation profile of target protein ➔ Comparison of glycoprofiles between humans and CHO cells	72
Metabolism and Metabolic Flux Analysis	➔ Estimation of enzyme kinetic parameters ➔ Metabolic flux analysis in culture at steady state ➔ CHO cell doubling times ➔ Effects of media supplementations on metabolic pathways	35

Table 2.3: Categories used to classify relevant research articles in the CHO bibliome according to the data contained in each study (continued)

Category	Examples of data contained within category	Total number of studies in category
Modeling	➔ Kinetic modeling of batch / fed batch bioprocesses with CHO ➔ In silico metabolic analysis (e.g. Flux balance analysis) ➔ Models for predicting optimal culture conditions	39
Secretory Pathway and Product Secretion	➔ Measurements of protein secretion rates ➔ Characterization of proteins/enzymes involved in the secretory pathway ➔ Dynamics of the unfolded protein response during cell culturing	37
RNAs and codon usage	➔ Identification of miRNA during specific culture conditions ➔ Cell engineering with the use of non-coding RNAs ➔ Role of RNAs on apoptosis	23
Genomics and Epigenetics	➔ CHO cell line-specific genome sequences ➔ Effects of genome methylation on production stability	24

Table 2.3: Categories used to classify relevant research articles in the CHO bibliome according to the data contained in each study (continued)

Category	Examples of data contained within category	Total number of studies in category
Culture Strategy and Bioreactor Design	➔ Automation of bioprocesses ➔ Comparison of CHO cells performance in different types of bioreactors ➔ Novel strategies for culturing CHO cells	26
Expression and Transfection Methods	➔ Promoter-based control of expression vectors ➔ Transient transfection methods in cultivation systems	30
Cell line construction and characterization	➔ Characterization of mutant cell lines and their associated phenotype	30
Metabolomics and Fluxomics	➔ Systems-level measurement of intracellular metabolite concentrations ➔ Large-scale measurement of uptake/secretion rates	40
Review article or other	➔ Review articles on CHO cells, high-throughput data for CHO culturing, and bioprocessing control ➔ Hydrodynamic stress analysis of CHO cells in culture	127

The compiled database of Phenotype and Production Characteristics provides a valuable dataset through which insights into general trends in CHO bioprocessing can be gained and key

attributes can be identified. By classifying each study and extracting bioprocess data, it is possible to address several questions related to cell phenotypes, the interactions between process variables, and to test long-held assumptions in the field of CHO bioprocessing.

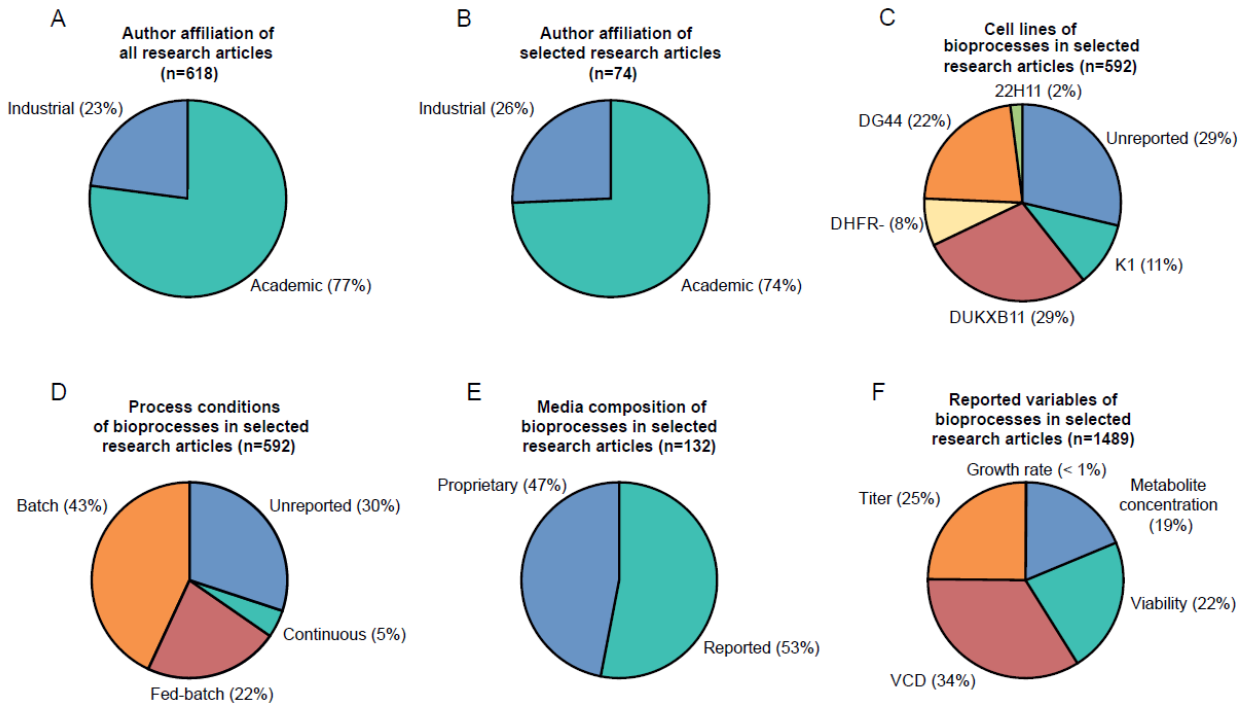

Figure 2.3: Overview of the CHO bibliome and the subset of articles used for meta-analysis. **(A)** Author affiliation of all articles. **(B)** Author affiliation of the analyzed articles. **(C)** Distribution of cell lines in all bioprocesses. DHFR- corresponds to bioprocesses where the authors did not divulge the exact parental cell line but did report that the cell line used is dihydrofolate reductase-deficient. **(D)** Distribution of bioprocess operation modes. **(E)** Availability of media composition. **(F)** Type of reported variables.

2.4.2 Overview of the Phenotype and Production Characteristics Dataset

The 74 articles, from which the Phenotype and Production Characteristics dataset was constructed, consist of a range of cell lines, process conditions, and experimental setups (Figure 2.3). The 1489 time-series were grouped together based on similar process parameters, resulting in the identification of 592 individual bioprocesses and the distribution of cell lines within these is

shown in Figure 2.3C. Around 29% of bioprocesses did not specify a parental cell line (referred as "Unreported" parental cell line in this study) meaning that the authors utilized proprietary CHO cell lines or did not disclose the information for other reasons. In other bioprocesses, the exact parental cell line is not specified but at least it is indicated that the parental cell line is dihydrofolate reductase deficient (DHFR-). The parental cell line for these bioprocesses was thus simply named DHFR-. The DHFR- group in the classification could include DUKXB11 or DG44, but the authors did not specify. However, among the other bioprocesses in which the authors disclosed the parental cell line, the DUKXB11 and DG44 lines were found to be the most prevalent in our sample representing 29% and 22% percent of the bioprocesses, respectively. For an in-depth review of the origins and properties of the different CHO cell lines, the reader is referred to [105]. Finally, the subset of articles included in the meta-analysis showed a distribution of industrial vs. academic authorship that was very similar to that of the 618 articles in the Phenotype and Production Characteristics category (Figures 2.3A-B).

Bioprocesses that utilize the major types of culturing modes were included in our study as shown in Figure 2.3D. Most of the bioprocesses (43%) were found to be batch cultures, while fed-batch and continuous processes accounted for 22% and 5% of the total, respectively. Although it is widely known that fed-batch and continuous (e.g. perfusion) cultures commonly outperform batch processes with respect to production amount, the high prevalence of batch processing is likely due to the less challenging implementation requirements and shorter culture durations, making batch cultures suitable for rapid analysis of various experimental conditions. Media compositions (Figure 2.3E) were obtained from each article (n=74) and a total of 135 unique media compositions were identified as some studies described more than one formulation. We observed

that almost half of the reported media compositions were proprietary, suggesting that the use of proprietary media formulations is prevalent in both academic as well as industrial settings.

Across the studies, several types of data were reported in the extracted time-series variables (n=1489), hinting at process variables commonly deemed as important or critical by most authors (Figure 2.3F). Viable cell density was the most commonly reported data type (34%), followed by titer (25%). Specific rates, such as the growth rate, which are often important variables due to their suitability for providing physiological insights, were reported rarely (<1%).

In summary, the Phenotype and Production Characteristics dataset provides a diverse collection of phenotype-related variables that can lead to generation of insights into various aspects of CHO bioprocessing. In the following sections, a combination of statistical methods to extract information will be presented.

2.4.3 Fed-batch bioprocesses exhibit longer growth phases and higher titer

Bioprocess setup and cultivation strategies significantly influence product yields and quality [169]. By comparing the performance of batch and fed-batch processes within the subset of 592 bioprocesses, it was possible to recapitulate known differences between these two processing strategies, thereby increasing the general confidence in the dataset and the soundness of the statistical methods employed here.

Fed-batch processes were found to be superior with respect to production variables [104]. The final titer was significantly higher in fed-batch processes compared to batch processes ($p<0.01$). However, the differences in specific productivity (q_P) could not be statistically established ($p=0.057$). Similarly, the product yield (or product to biomass coefficient) $Y_{P/X}$ was also not significantly different between the two process modes ($p=0.053$). Therefore, the difference

in final product yield could be attributed to higher maximum viable cell densities in fed-batch cultures ($p<0.01$), resulting from a longer growth phase duration as well as longer total duration of culture ($p<0.01$). Comparison of the growth rate of the two culture modes revealed that the average growth rate during the growth phase was not significantly different. Thus, due to the longer growth phase in fed-batch cultures, fed-batch bioprocessing allows for higher viable cell densities via regular supplementation of nutrients. While these results recapitulate known differences between batch and fed-batch cultures, this validation demonstrates that relationships between process variables and outcomes can be identified despite the heterogeneity among the studies.

2.4.4 Common production CHO cell lines show significant physiological differences

Recombinant protein production is influenced by many factors, including growth rate, viability, culture and growth phase duration, and media conditions [170]. To shed light on some of the differences among CHO cell lines, three main indicators, namely average growth rate, average cell-specific productivity, and the maximum viable cell density were compared among different cell lines found in the meta-analysis subset.

As shown in Figure 2.4A, different cell lines show significant differences with respect to the average growth rate during the exponential growth (log) phase. Interestingly, despite some reports that the growth rate of the K1 cell line is higher than DHFR-deficient lines, such as DG44 and DUKXB11 [106], our dataset shows the K1 cell line ($n=27$) grew significantly slower than all other cell line groups (see Supplementary Table 4). Also, growth rate in bioprocesses involving the DUKXB11 ($n=77$) and Unreported ($n=61$) parental cell lines were found to be significantly higher than the other cell lines ($p<0.01$). The high growth rate of the Unreported group could be

since publications without cell line specification often utilized proprietary lines that have been optimized for higher growth rates. The difference between the high-growth group (DUKXB11 and Unreported) and another group consisting of DG44 (n=79) and DHFR- (n=32) was also statistically significant. To our knowledge, comparison of growth rate between the main DHFR-deficient CHO cell lines, that is DUKXB11 and DG44, has not been addressed in the scientific literature to-date.

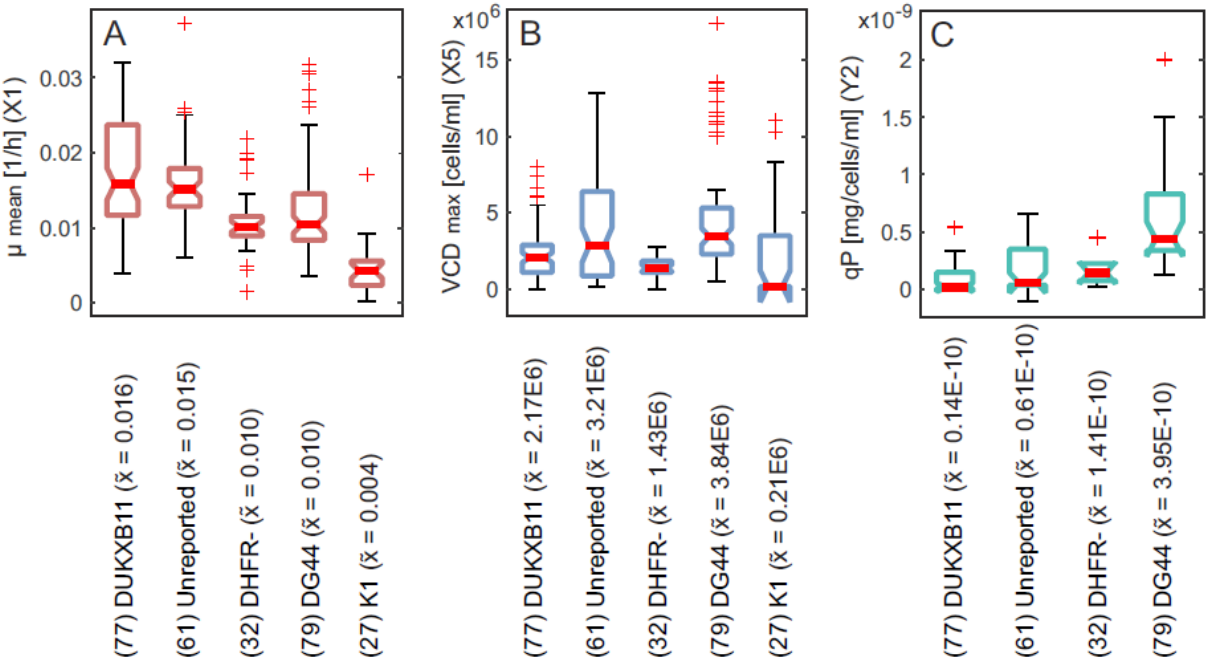

Figure 2.4: Phenotypic differences among main CHO cell lines. (A) Comparison of the mean growth rates of different cell lines over the growth phase duration of batch and fed-batch bioprocesses. **(B)** Comparison of the maximum viable cell densities of different cell lines over the growth phase duration. **(C)** Comparison of the specific production rate of different cell lines. The q_P for the DG44 cell line is significantly ($p<0.01$) larger than the DUKXB11 and Unreported cell lines (CHO-K1 was excluded for this experiment due to lack of titer data). The difference between the median of DHFR- and other cell lines is not statistically significant due to the small number of bioprocesses in the DHFR- group. Numbers in brackets correspond to the number of bioprocesses in each group. The medians of each box plot are indicated with $\tilde{x}$.

VCD is considered an important quality attribute of CHO bioprocesses since higher culture viability and delayed loss of viability often result in higher product titers. Comparison of the

maximum value of the VCD signal between different cell lines (Figure 2.4B) and its integral over the duration of the growth-phase (IVC) revealed that DG44 and Unreported cell line groups show the highest viable cell densities. The difference between the maximum VCD of the DG44 cell line and DUKXB11, K1, and DHFR- groups is significant ($p < 0.05$). These results suggest that although the growth rate of DG44 is lower than DUKXB11, higher VCDs are reached most probably due to a more sustained growth phase by DG44.

The DG44 cell line demonstrated significantly higher specific production rates than the DUKXB11 and Unreported cell lines ($p < 0.05$) (Figure 2.4C). To compare the production characteristics between different cell lines, bioprocesses were selected based on the availability of VCD and titer signals. Since many articles did not report these signals simultaneously, only a subset of the dataset could be used (n=100 bioprocesses). In addition, many publications reported titer in relative units, which resulted in their elimination from further analysis. The specific production rate, q_P, was calculated over process duration for each remaining bioprocess, and the average value of q_P over time was calculated as an indicator variable. Comparison of the median of this indicator variable among different cell lines was performed using Kruskal–Wallis and Dunn's tests. The tests showed that DG44 cells exhibit significantly higher specific productivity than DUKXB11 and Unreported cell lines ($p<0.01$). As the selected subset of bioprocesses did not contain any bioprocess belonging to the K1 group, comparisons with CHO-K1 could not be performed. Viability was not found to be significantly different among the remaining cell lines. In addition to the mean and median of viability over process duration, the rate of decrease of viability was calculated and a significant difference was not detected.

In summary, significant differences among cell lines were observed with respect to growth rate, viable cell density, and specific production rate. Overall, the Unreported and DUKXB11 cell

lines were found to exhibit similar growth and production characteristics, leading to the hypotheses that industrially-derived cell lines could be based on DUKXB11. A comparison of DG44 and DUKXB11 cell lines substantiates the long-held assumption that growth rate and specific productivity are inversely correlated, and bioprocesses with a low growth rates and more sustained growth phases exhibit higher specific productivities. The large number of bioprocesses in the DG44 (n=79) and DUKXB11 (n=77) cell lines made a direct comparison of these two groups possible, leading to the conclusion that the DG44 line could be more suitable for production of recombinant proteins that would benefit from its higher maximum viable cell densities and specific productivities.

2.4.5 Industrial bioprocesses show higher VCD but not titer

Industrial bioprocessing has been particularly successful at process optimization, resulting in cell lines with much higher yields. Therefore, we wondered if the published work from industry demonstrated a clear improvement over research coming from academic groups. Thus, we compared the performance of bioprocesses with respect to corresponding author affiliation (industrial vs. academic) and found no significant difference between the mean and maximum growth rates. However, industrial publications reported significantly higher maximum VCDs compared to academic publications. Similarly, the integral of VCD was also higher in industrial bioprocesses. Despite higher VCD values in industrial publications, we did not detect a significant difference in the specific productivity between industrial and academic publications. Furthermore, the difference between maximum titer was not significant between the two groups. In short, while the VCD of industrial publications appears to be higher than the other group, the higher VCD values do not translate into higher production rates in published work from industry.

2.4.6 High growth rate is anti-correlated with many process quality attributes

Using the previously presented comparison of bioprocess characteristics among different classes, it was possible to reveal some of the relationships between individual bioprocess variables. To arrive at a more comprehensive overview of the relationships between variables in the dataset, pairwise correlation analysis using the Spearman rank correlation coefficient was conducted (Figure 2.5). The objective of this analysis was to find relationships that are present across a wider range of bioprocesses irrespective of their class assignments, e.g. cell line. Data used for the analysis originated from a subset of 100 bioprocesses that contained the required time-series variables for the calculations (including VCD, viability, and titer). The bioprocesses were not selected according to any other criteria; therefore, they contained variations with respect to cell line and other class conditions. The significance of the correlation coefficients was determined by testing the null hypothesis of no correlation against the alternative of a nonzero correlation. Although some of the observed correlations were expected, such as between VCD_{max} and VCD_{int}, the correlation matrix also highlights some more interesting relationships which hint at the underlying physiology of CHO cells.

Higher growth rates during exponential phase are preferable since they enable the rapid expansion of seed trains and are expected to achieve higher cell densities and titers more quickly. However, we found that several process quality attributes anti-correlated with the average growth rate of bioprocesses during the growth phase (μ_{ave}). Specifically, growth rate was negatively correlated with the final titer (P), suggesting that processes with lower growth rates have higher amounts of target product at the end of the process and vice versa. The inverse relationship between productivity and growth rate has been previously highlighted [171, 172] and our results also

recapitulate a positive correlation between titer and VCD as discussed in Clarke et al. [90] (see Figure 2.5). Average growth rate was also negatively correlated with specific productivity (q_P), the yield of product per viable cells ($Y_{P/X}$), maximum viable cell density (VCD_{max}), time integral of viable cell density (VCD_{int}), total duration of culture (t_{end}), and duration of the growth phase (t_{growth}). The surprising result of negative correlations between average growth rate and production-related variables could be attributed to higher flux through glycolysis and therefore overflow metabolism, which results in the production of toxic byproducts such as lactate [173] and a loss of the energetic efficiency of CHO cells. These effects would adversely affect protein titers [107], reduce viable cell density [173], and deplete sugars, leading to earlier culture termination. In the present dataset, analysis of lactate production and consumption revealed a link between the glucose concentration in the medium and the ability of CHO cells to switch to low glycolytic flux, partially confirming results from a previous study [108].

While growth rates were anti-correlated with various process quality attributes, we found that the time integral of viable cell density over the entire culture duration (VCD_{int}) positively correlates with final titer, consistent with previous reports [174]. In summary, our results suggest that while high growth rates and high VCDs might be good for seed train expansion and scale-up, a more sustained growth phase at lower rates is associated with higher production of recombinant proteins.

2.4.7 Viable cell density significantly predicts product titer

The diverse range of parameters (i.e. different cell lines, media, process modes, etc.) that were varied among the studies opens opportunities to explore correlations between process variables and quality attributes using multivariate methods such as PLS. Such relationships can provide hints about possible underlying mechanisms and relationships between important

phenotypic variables. Using the input matrix consisting of variables (X), three different PLS models were generated, each for prediction of a different target (Y) variable (Table 2.2). To assess model quality, the Q^2 value was used to quantify a model's predictive power using cross-validation. As summarized in Table 2.4, a model was developed that could predict final titer (P) satisfactorily (Q^2>0.5).

Table 2.4: Figures of merit for PLS models. LVs: number of latent variables of the PLS model. R2X: X-block variance, R2Y: Y-block variance, Q2: model's predictive power assessed via cross-validation (acceptable models should be >0.5).

Target variable	LVs	R2X (cum)	R2Y (cum)	Q^2
P: final titer	3	0.847	0.664	0.635
q_P: specific productivity	2	0.736	0.364	0.324
$Y_{P/X}$: yield P/X	2	0.741	0.405	0.362

According to the Variable Importance for Projection plot (Figure 2.5A), the integral of viable cell density (VCD_{int}) is the most important variable in the model for prediction of product titer, and higher values of this variable imply higher final titers. The final value of VCD (VCD_{end}) is the second most important variable in the model and has a negative coefficient. Culture duration (t_{end}) is the third most important variable in the model and has a positive correlation with the end titer (Figure 2.5B). Using pairwise correlation analysis (Figure 2.6), the final value of VCD (VCD_{end}) was found to show a small positive correlation with the final value of titer (P). However, in the multivariate PLS model, VCD_{end} was found to have a large negative overall coefficient for predicting titer (Figure 2.6B). The negative coefficient for VCD_{end} is caused by considering several compounded relationships in multivariate models. The PLS model consists of three latent variables

(LVs), each representing a selected portion of the decomposed variance in X variables. The first LV showed positive loadings for VCD_{int}, duration of growth (t_{growth}), duration of culture (t_{end}), VCD_{max}, and VCD_{end}. This LV can be interpreted as representing the expected association between higher VCD and higher final titer. The second LV showed significant negative loadings for average growth rate (μ_{ave}) and maximum growth rate (μ_{max}), hinting generally at the inverse association between growth rate and final titer and confirming the results found previously using pairwise correlation analysis. The third LV had a positive loading for culture duration (t_{end}) and a large negative loading for final VCD_{end}. The relationship represented by the third LV can be attributed to cell lysis, leading to the release of intracellular products (higher titer) and loss of viability (lower VCD). By combining these effects, the multivariate model can make a significant prediction of final titer, revealing relationships in the dataset that are not readily accessible using univariate methods.

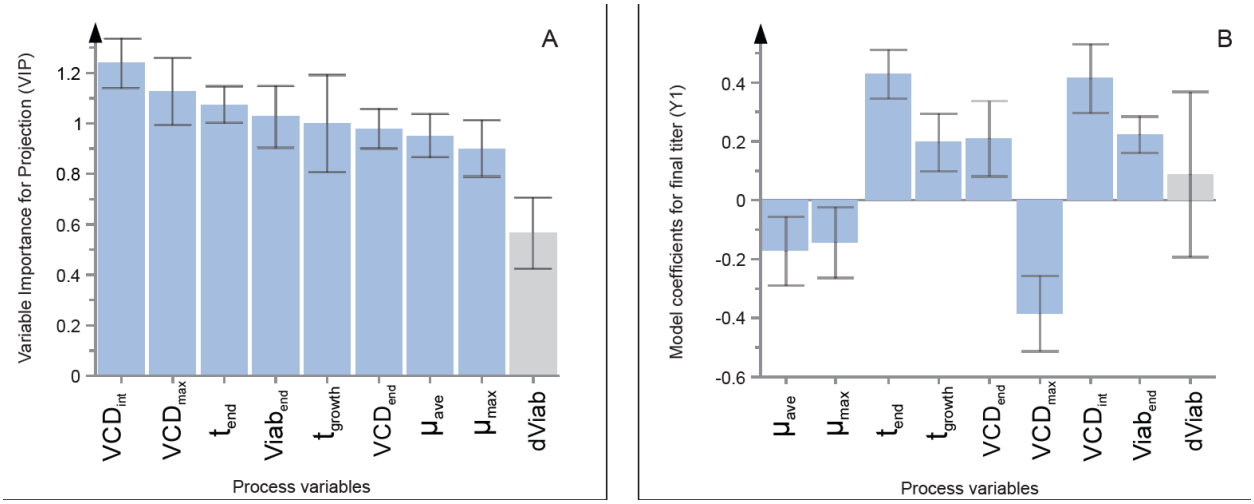

Figure 2.5: Prediction of titer from process variables. (A) Variable Importance for Projection (VIP) plot of a PLS model for prediction of final titer. A variable with a VIP Score close to or greater than 1 can be considered important in given model (blue). Variables with VIP scores significantly less than 1 are less important and might be good candidates for exclusion from the model (gray). **(B)** Coefficients of the PLS model for prediction of final titer (significant coefficients are blue).

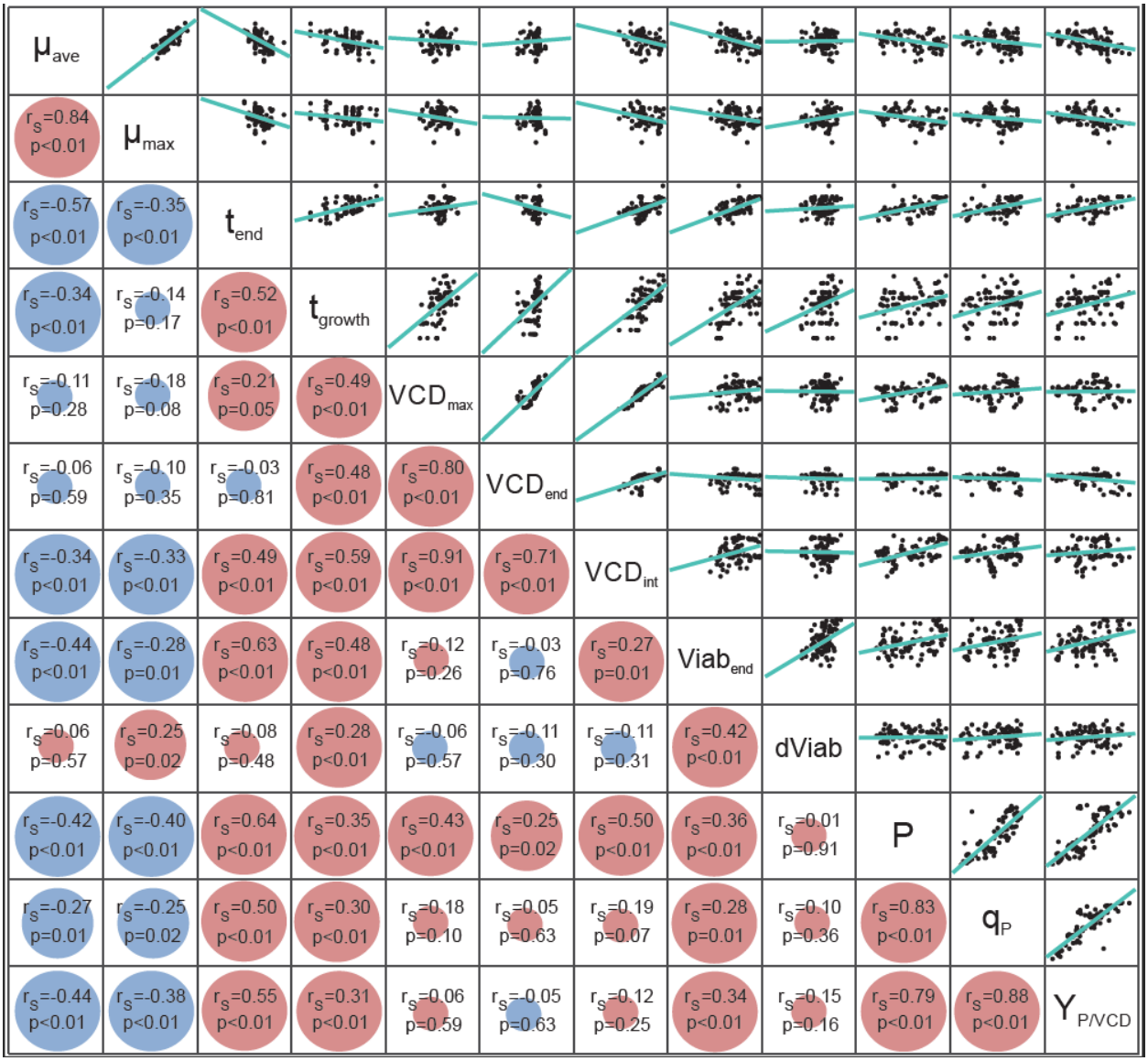

Figure 2.6: Spearman's rank correlation matrix. Significant correlation coefficients ($p<0.01$) are marked by full size circles. Red circles: positive correlation, blue circles: negative correlation. For more information about the variables, see Table 2.2.

2.5 Discussion

In this chapter, we have presented a large bibliographic collection of all published CHO cell research from the past 20 years. This bibliome includes 1,157 articles classified into 16 categories according to the type of data they contain. Using this resource, we conducted a meta-analysis including data from 592 individual bioprocesses published in 74 articles from the

"Phenotype and Production Characteristics" category in our bibliome; thus, we present here the largest published meta-analysis of CHO bioprocessing to date.

In our meta-analysis, we elucidated physiological differences across several CHO cell lines. These differences generally appear in the form of mean growth rate, specific productivity, and maximum viable cell density. Since our analysis used data from a sample of all studies, we anticipate that higher resolution insights can be obtained as data are extracted from additional papers in a targeted manner to address specific questions. For example, for questions related to protein production rates, there are many bioprocess and cell-specific factors that influence this critical bioprocess attribute, such as the type of production process, protein size, its chemical characteristics, and the host cell line. Future studies including larger datasets could possibly consider these diverse factors by specifically looking at one or more classes of product and controlling for the various bioprocess parameters that may influence protein production.

We were also able to investigate correlations between culture variables (e.g. growth rate, titer, percentage viability, viable cell density). Many correlations agreed with knowledge in the field. For instance, we saw that the average growth rate during the growth phase exhibited an inverse correlation with production-related characteristics, such as specific productivity and final titer. This inverse relationship has previously been attributed to factors, such as energetic burden of protein expression [175], overflow metabolism at higher growth rates accompanied by secretion of toxic by-products [173], and a less efficient energetic metabolic state during overflow metabolism [173]. Our ability to recapitulate such effects here demonstrates the validity of the methodologies used for analysis and that the physiological characteristics of CHO cells seen here are largely independent of variations of cell line, selection conditions, media composition and other bioprocess parameters.

Further application of the presented meta-analysis methodologies could help identify control and optimization strategies, especially in industrial settings when huge amounts of historical bioprocess data are available for analysis. As shown here, general trends and product-independent characteristics of a production platform could be identified, facilitating the design of new strategies for improving cell line selection and process conditions. For instance, in the presented meta-analysis, the time integral of VCD was identified as one of the most important factors for predicting maximum titer (Figure 2.5A), which is consistent with previous reports [174]. Thus, performance improvement strategies, such as media design or cell line selection, could be devised accordingly. In other words, despite the variance across such a heterogeneous set of studies, we identified significant correlations and factors that are common in all studies and thus could be transferable to any CHO culture as the main variables in a bioprocess (i.e. titer, VCD, viability) conserve inherent relationships that seem to be independent of the variations in culture conditions. From this, one can imagine the potential of combining the knowledge of bioprocessing control with systems-level understanding of the fundamental metabolic, glycosylation and secretory capabilities of CHO cells [100, 101, 109].

Going forward, this chapter demonstrates how bioprocess data can be collected and analyzed to deepen our understanding of parameters that influence quality attributes. In our analysis, we used heterogeneous data, collected from numerous different researchers over two decades. With better standardization and reporting of research findings, similar meta-analysis methodologies can be applied for answering diverse questions in academic and industrial bioprocess optimization. To fully enable this, several improvements in data reporting are necessary. First, organization of key metadata would facilitate more rapid re-use of information. Specifically, more complete datasets would be available if a standard format for reporting relevant

information of bioprocessing experiments were adopted by the bioprocessing community. For instance, authors should completely report all aspects of a bioprocess instead of citing other studies when it comes to important conditions, such as the parental CHO cell line from which the cells under study were derived. Similarly, a standard set of experimental parameters could be reported, such as temperature, pH or the osmolality of the experiments. Such standards have been established for many other fields, requiring "Minimum Information" for a given data type upon publication [176–178]. Adherence to such a standard in our community would enable improved data re-use and increase the knowledge in the field.

Standardized reporting and formatting of data can be difficult, given the proprietary nature of products and processes (e.g., proprietary media and cell lines) used in the biopharmaceutical industry. However, one recommendation derived from this study would be to include as many details as possible when reporting culture conditions and to not underestimate the value of reporting variables such as osmolality, pH and temperature since it has been observed that these parameters can affect the metabolism and physiology of CHO cells, and thereby affect the overall outcome of a bioprocessing experiment [92, 179]. We also recommend the inclusion of numerical values of all measurements, when reporting relevant measurements such as VCD, viability and product concentration (titer). Some journals are aiding in this by allowing authors to include supplementary data tables that provide the exact data used for the article figures [180]. This would facilitate the comparison of the results and would prevent any potential biases that result from digitization of plots (especially in log scale plots).

Finally, as we enter the so-called "Big Data" driven era of research, there is an ever-increasing need to standardize how data are reported in each field. In the case of CHO bioprocesses, this could be accomplished by proposing a unifying format for summarizing the

methods and results of a study as to structure them in the abstract of the article. Furthermore, an important driver that motivated the realization of this study is the fact that the CHO community lacks a public database of bioprocessing data. Such a database would be invaluable for the advancement of the field and for fostering collaboration around the globe with experts in different areas of the CHO bioprocessing field. Thus, standardizing the language and the semantics used for reporting CHO data will be indeed a necessary preliminary step for the construction and maintenance of such knowledge repository.

2.6 Conclusions

Here we present a bibliome and meta-analysis for CHO, consisting of a combination of data mining tools, bioprocess-specific data treatment methodologies, and tailored statistical analysis methods. Together, these tools lead to the identification of important trends and relationships among bioprocess parameters. The extracted features are indeed of high importance for process developers using CHO. Therefore, our methodology could be expanded and applied to generate valuable hypotheses from a more diverse set of experimental data. Such hypotheses could be tested experimentally and refined by means of systems biology modeling [181] and other computational tools that could then help in the development of mechanistic models for CHO cell culturing [9]. Furthermore, retrospective meta-analyses of literature data, such as presented here, provide correlations that can be invaluable for subsequent experimental design in order to arrive at causal relationships, and ultimately the rational design of novel cell factories [10] through genome engineering [182]. Finally, we hope that the present analysis triggers further discussions on the establishment of data reporting standards for bioprocessing experiments and motivates the future development of more studies on the topic.

2.7 Acknowledgements

This work was completed with generous support from CONACYT and a graduate fellowship from the University of California Institute for Mexico and the United States. Financial support was also provided by the Austrian research funding association (FFG) under the scope of the COMET program within the research project “Industrial Methods for Process Analytical Chemistry (imPACts)” (contract # 843546). This work was completed with support from the Novo Nordisk Foundation that had been provided to the Center for Biosustainability at the Technical University of Denmark.

Chapter 2 in part is a reprint of the material *Golabgir A*, Gutierrez JM*, Hefzi H, Li S, Palsson BO, Herwig C, Lewis NE. Quantitative feature extraction from the Chinese hamster ovary bioprocess bibliome using a novel meta-analysis workflow. Biotechnol. Advances (2016) 34:621-633*. The dissertation author was one of the two primary authors.

Chapter 3 - A genome-scale reconstruction of the mammalian secretory pathway for analysis of energetic trade-offs in bioprocesses

3.1 Abstract

In mammalian cell lines, >25% of proteins are synthesized and exported through the secretory pathway. The pathway complexity, however, obfuscates its impacts on the secretion of different biopharmaceuticals. Here we delineate the core secretory pathway functions and integrate them with the genome-scale metabolic model of Chinese hamster ovary (CHO) cells. The resulting reconstruction, called iCHO2048s, enables the computation of the energetic costs and machinery demanded by secreted proteins. We predicted metabolic costs and maximum productivities of biotherapeutics and identified protein features that most significantly impact protein secretion. By integrating additional metabolomic, glycoproteomic and ribosomal profiling data, we further found that CHO cells have adapted to reduce expression and secretion of expensive host cell proteins. Our work represents a knowledge-base of the mammalian secretory pathway that serves as a novel tool for systems biotechnology.

3.2 Introduction

To interact with their environment, cells produce numerous signaling proteins, hormones, receptors, and structural proteins. In mammals, these include >3000 secreted proteins (e.g.,

enzymes, hormones, antibodies, extracellular matrix proteins, etc.) and >5500 membrane proteins [183], most of which are synthesized and processed in the secretory pathway. The secretory pathway is a complex series of processes predominantly in the endoplasmic reticulum (ER), Golgi apparatus, and other components of the endomembrane system. In these compartments, the synthesis of the thousands of membrane and secreted proteins is facilitated by hundreds of other proteins that make up the secretory pathway machinery. These are used to translate, fold, post-translationally modify, test for quality, sort and translocate the secreted proteins.

The secretory pathway is particularly important in biotechnology and the biopharmaceutical industry, since most therapeutic proteins are secreted. Mammalian cell lines such as HEK293 or PerC6 are used extensively to ensure that a secreted biotherapeutic is properly folded and contains the necessary post-translational modifications (PTMs). In this regard, Chinese hamster ovary (CHO) cells stand out as the most widely used cell line and account for most top-selling biotherapeutics [184]. For any given biotherapeutic, different machinery in the secretory pathway may be needed, and each step can exert a non-negligible metabolic demand on the cells. The complexity of this pathway, however, makes it unclear how the biosynthetic cost and cellular needs vary for different secreted proteins, each of which exerts different demands for cellular resources. Therefore, a detailed understanding of the biosynthetic costs of the secretory pathway could guide efforts to engineer host cells and bioprocesses for any desired product. The energetic and material demands of the mammalian secretory pathway can be accounted for by substantially extending the scope of metabolic models. Indeed, recent studies have incorporated portions of the secretory pathway in metabolic models of yeast [84, 185, 186].

We present the first genome-scale reconstruction of metabolism and protein secretion in CHO cells, called iCHO2048s. We first demonstrate that product-specific secretory pathway

models can be built to estimate CHO cell growth rates given the specific productivity rate of the specific product as a constraint. Second, we identify the features of secreted proteins that have the highest impact on protein cost and productivity rates. Third, we use our model to identify proteins that compete for cell resources, thereby presenting targets for cell engineering. Finally, we derive an expression for computing the energetic cost of synthesizing and secreting a product in terms of molecules of ATP equivalents per protein molecule. We use this expression and analyze how the energetic burden of protein secretion has led to an overall suppression of more expensive proteins in CHO cells. Through this study we demonstrate that a systems-view of the secretory pathway now enables the analysis of many biomolecular mechanisms controlling the efficacy and cost of protein expression in mammalian cells. We envision our models as valuable tools for the study of normal physiological processes and engineering cell bioprocesses in biotechnology.

3.2.1 The secretory pathway of animal cells

Historically, most of the knowledge on the secretory pathway was obtained by studying protein transport processes and secretion in *Saccharomyces cerevisiae* [187]. Albeit quite similar in core functions, the secretory pathways of mammalian cells and fungi differ significantly in some of the steps which have been evolved based on species-specific secretion phenotypes [188]. The following paragraphs briefly overview the mammalian secretory pathway and highlights pathways exclusive to animals not present in fungi. The last section provides an in-depth comparison of the yeast and animal secretory pathways while highlighting the most important differences between both.

3.2.2 Translocation and processing in endoplasmic reticulum

Proteins destined to the secretory pathway generally bear a signal peptide at the amino-terminus which targets the proteins to the endoplasmic reticulum (ER) where the initial post-translational modifications (PTMs) take place. This transport requires translocating the target protein across the ER membrane through two general pathways: co-translational translocation (GTP dependent) and post-translational translocation (ATP dependent) [189]. An additional pathway for tail-anchored (TA) proteins into the ER membrane has also been discussed in the literature and included in our iCHO1921s reconstruction [190, 191]. Once inside the ER lumen, the target proteins are folded by the action of several transmembrane ER proteins, including calnexin, calreticulin, and other luminal chaperones [192–194]. In the event of protein misfolding, a target protein may go through a "quality control" system (exclusive in the mammalian secretory pathway) that attempts to correct for folding errors [195, 196]. However, if the misfolded state of the protein is sustained for too long, the protein then enters the ER associated degradation pathway, or ERAD, which involves retrotranslocation of the misfolded protein back to the cytosol, ubiquitination and proteasomal degradation [197–199].

Besides folding, a target protein may acquire additional PTMs while inside the ER such as attachment of a glycosylphosphatidylinositol (GPI) anchor [200, 201], formation of disulfide bonds [202], and N-linked glycosylation [203–206]. After these PTMs are successfully completed , the target proteins are transported to the Golgi apparatus via COPII-coated vesicles that bud from the ER [207, 208] whereas misfolded proteins are retro-translocated to the cytoplasm [209, 210] for proteasomal degradation via the ER-associated degradation pathway (ERAD) [211, 212]. In the Golgi apparatus, N-glycans are processed into branched and complex glycoforms and proteins

are further glycosylated with O-linked glycans [213–215] and then sorted to their final destination (e.g. lysosome, extracellular medium) via clathrin-coated secretory vesicles [216–219].

In co-translational translocation, proteins destined to the secretory pathway bear a hydrophobic signal sequence at the amino-terminus that promotes the targeting of ribosome-nascent chain (RNC) complexes to the ER via binding to the signal recognition particle (SRP). The SRP recognizes the signal peptide as soon as it emerges from the ribosome during translation. Then, the newly formed SRP-RNC complex is recognized by the SRP receptor on the ER membrane where translocation is initiated by interaction with the Sec61 complex (Sec61C) and assisted by the chaperone BiP to increase the efficiency and ensure the unidirectionality of this process [216].

Post-translational translocation, in contrast to co-translational translocation, occurs independently of SRP and its receptor [220]. Furthermore, this process does not rely too heavily on the Sec61C to translocate the target protein and instead utilizes the protein Sec62 as a safe route that guarantees the efficient translocation of small proteins (<160 amino acids in length) [221].

Finally, the pathway for inserting TA proteins into the ER membrane also occurs post-translationally due to the fact that the ER targeting signal of TA proteins is located very close to the carboxy-terminus, which allows the ribosome to release the protein before it is recognized and localized to the ER [222]. This pathway depends on ATP and one of the main players in the process is a transmembrane recognition complex known as TRC40 or Asna1 [223].

3.2.2 Important differences between the yeast and animal secretory pathways

Core functions of the secretory pathway are conserved between mammalian and yeast cells. These core functions include: Translocation through endoplasmic reticulum; primary glycosylation in ER (N-linked glycans) and Golgi (N-linked and O-linked glycans); protein folding and quality control in ER; anterograde and retrograde vesicular transport between ER and Golgi via COPII and COPI vesicles, respectively; dolichol pathway for N-linked core glycan translocation through the ER membrane; endoplasmic reticulum associated degradation (ERAD); GPI biosynthesis; and the unfolded protein response (UPR). Nevertheless, minor and major differences exist between the yeast and mammalian secretory pathways. Some of these differences [188] are summarized in Table 3.1 below. Here, we highlight the major differences between both secretory pathways that are relevant for modeling purposes.

Table 3.1: Summary of differences between mammalian and yeast secretory pathways

Description of difference	Mammalian secretory pathway	Yeast secretory pathway	Importance for modeling purposes
Chaperones involved in translocation	The main chaperone is BiP	The main chaperone is Kar2	Minor
Presence of heat-shock proteins (HSPs) in ER	Mainly presence of proteins in the Hsp90 family	Not found in yeast	Minor
Enzymes for detoxification of reactive oxygen species in ER	Contains several enzymes such as Ero1 and glutathione peroxidases	Not found in yeast	Major
Oxidation state of Protein disulfide isomerase (PDI)	PDI is mainly reduced	PDI is mainly oxidized	Minor

Table 3.1: Summary of differences between mammalian and yeast secretory pathways (continued)

Description of difference	Mammalian secretory pathway	Yeast secretory pathway	Importance for modeling purposes
Components of calnexin-calreticulin cycle	Includes an enzyme coded by the UGGT gene to transfer glucose residues to core N-linked glycans in misfolded proteins	Lacks UGGT and instead directs misfolded proteins to ER exit	Major
ERAD pathway branches for degrading misfolded proteins	Capable of directing misfolded proteins towards the ERAD pathway by trimming N-linked glycan residues in the A, B and C branches	Capable of directing misfolded proteins towards the ERAD pathway by trimming N-linked glycan residues only in B and C branches	Major
Components of COPII vesicles	Contains four isoforms of Sec24	Expresses Sec24 with three cargo binding sites as well as Sec24 homologs Sfb2-3	Minor

3.3 Methods

3.3.1 Reconstruction of the mammalian secretory pathway

A list of proteins and enzymes in the mammalian secretory pathway was compiled from literature curation, UniProt, NCBI Gene, NCBI Protein and CHOgenome.org. To facilitate the reconstruction process, the secretory pathway was divided into twelve subsystems or functional modules (Figure 3.1) to sort the components according to their function. These subsystems correspond to the major steps required to process and secrete a protein. The components from a prior yeast secretory pathway reconstruction [84] were used as a starting reference. To build

species-specific models, orthologs for human, mouse and the Chinese hamster were identified and used, while yeast components and subsystems that are not present in the mammalian secretory pathway were removed. Additional subsystems were added when unique to higher eukaryotes, such as the calnexin-calreticulin cycle in the ER [224]. These were constructed *de novo* and added to the reconstruction. The databases and literature were then consulted to identify the remaining components involved in each subsystem of the mammalian secretory pathway. Since most components in the mammalian secretory pathway have been identified in mouse and human, BLAST was utilized to identify the corresponding Chinese hamster orthologs by setting human as the reference organism and a cutoff of 60% of sequence identity.

3.3.2 Protein Specific Information Matrix (PSIM)

The PSIM contains the necessary information to construct a protein-specific secretory model from the template reactions in our reconstruction. The columns in the PSIM are: presence of a signal peptide (SP), number of disulfide bonds (DSB), presence of Glycosylphosphatidylinositol (GPI) anchorage, number of N-linked (NG) and O-linked (OG) glycans, number of transmembrane domains (TMD), subcellular location, protein length, and molecular weight. For most proteins, the information in the PSIM was obtained from the Uniprot database. When necessary, computational tools were used to predict signal peptides (PrediSi [225]) and GPI anchors (GPI-SOM [226]). Finally, additional information on the number of O-linked glycosylation sites of certain proteins were obtained from experimental data in previous studies [227],[228]. The PSIMs of the CHO and human secretomes are a subset of the full PSIM and contains only the proteins with a signal peptide (predicted or confirmed in Uniprot).

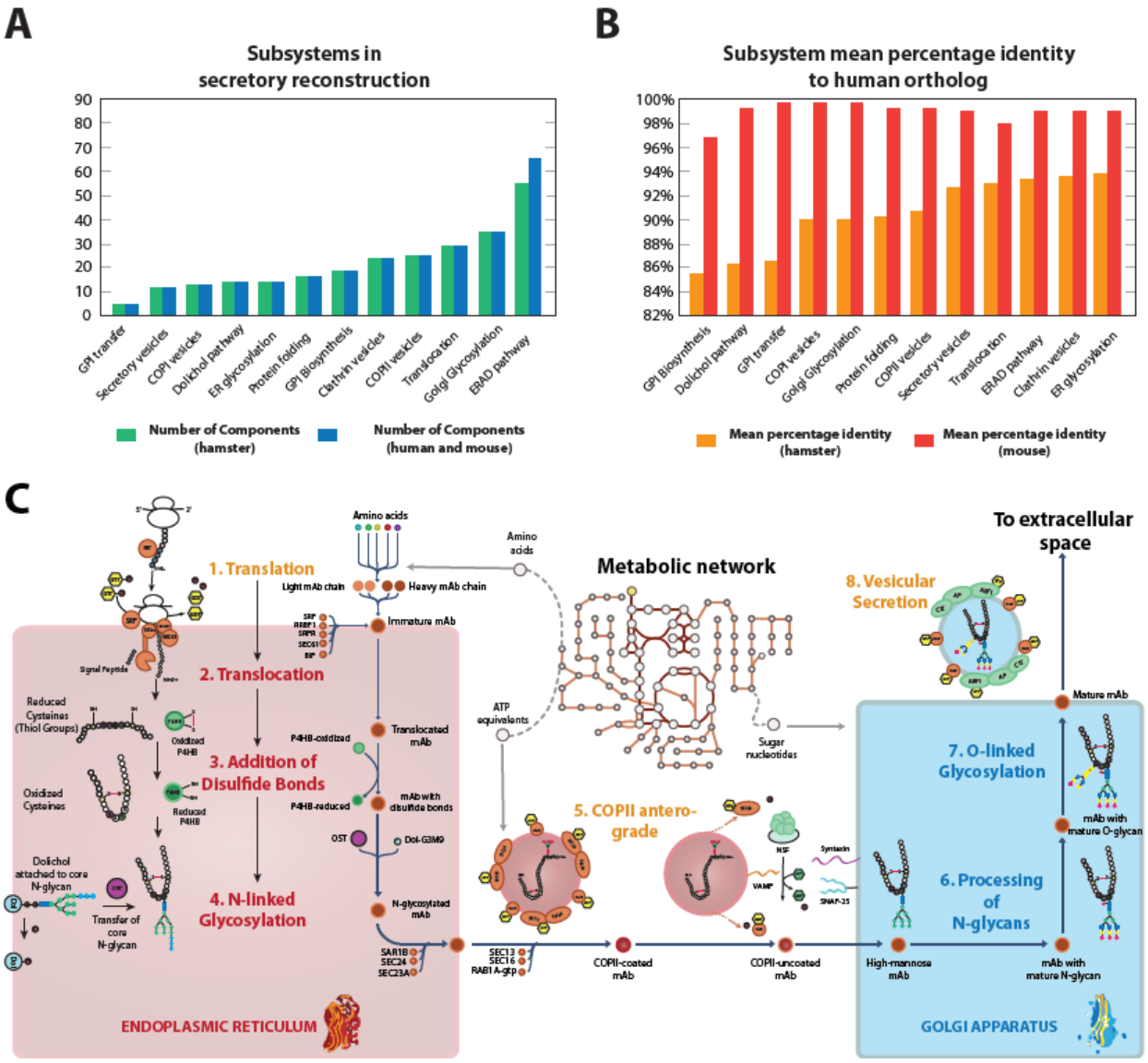

Figure 3.1: Components in the reconstruction of secretory pathway in mammalian cells. (A) The reconstruction comprises 261 proteins in CHO cells and 271 proteins in human and mouse that are distributed across 12 subsystems or functional modules. The different component numbers arise from the fact that the Chinese hamster proteome annotation only contains one alpha and one beta proteasome subunits, whereas the human and mouse contain 12 subunits of different subtypes. The detailed description of all components can be found in Supp. File 1. **(B)** High similarities were seen for proteins in CHO and human, with a high mean percentage identity in each subsystem (calculated with the sequence alignment tool BLAST). **(C)** Simplified schematic of reactions and subsystems involved in the secretion of a monoclonal antibody (mAb). A total of eight subsystems are necessary to translate, fold, transport, glycosylate, and secrete a mAb. The color of the subsystem names indicates if the reactions occur in the cytoplasm (yellow), the ER lumen (red) or the Golgi apparatus (blue). GPI = Glycosylphosphatidylinositol, ER = Endoplasmic Reticulum, ERAD = ER associated degradation.

3.3.3 Detection of N-linked glycosylation sites via mass spectrometry

The number of N-linked glycosylation sites in the PSIM was determined computationally and experimentally as follows. CHO-K1 cells (ATCC) were lysed, denatured, reduced, alkylated and digested by trypsin enzyme. Desalted peptides were incubated with 10 mM sodium periodate in dark for 1 hour before coupling to 50 μL of (50% slurry) hydrazide resins. After overnight reaction, non-glycosylated peptides were washed with 1.5 M NaCl and water. The N-glycosylated peptides were released with PNGaseF at 37 °C and desalted by using C18 SepPak column. Strong cation exchange (SCX) column was used to separate the sample into 8 fractions. Each fraction was analyzed by LTQ-Orbitrap Velos (Thermo Electron, Bremen, Germany) mass spectrometer. During the mass spectrometry data search, carbamidomethylation was set as a fixed modification while oxidation, pyroglutamine and deamidation were variable modifications.

3.3.4 Construction of secretory models and constraint-based analysis

A Python script that takes a row from the PSIM as input to produce an expanded iCHO2048s metabolic model with the product-specific secretory pathway of the corresponding protein was written. Flux balance analysis (FBA [229]) and all other constraint-based analyses were done using the COBRA toolbox [230] in MATLAB R2014a and the Gurobi solver version 6.0.0. For the iCHO2048s models secreting human proteins, we set the same constraints in all models and computed the theoretical maximum productivity (max_{qp}) while maintaining a growth rate (in units of 1/h) of 0.01. Finally, since the exact glycoprofiles of most proteins in CHO are unknown and some even change over time in culture [231], we simplified our models by only adding the core N-linked and O-linked glycans to the secreted proteins.

3.3.5 Batch cultivation

Two isogenic CHO-S cell lines adapted to grow in suspension, one producing Enbrel (Etanercept) and the other producing human plasma protease C1 inhibitor (SERPING1), were seeded at 3 x 10^5 cells/mL in 60 mL CD-CHO medium (Thermo Fisher Scientific, USA) supplemented with 8 mM L-Glutamine and 1 µL/mL anti-clumping agent, in 250 mL Erlenmeyer shake flasks. Cells were incubated in a humidified incubator at 37°C, 5% CO_2 at 120 rpm. Viable cell density and viability were monitored every 24 hours for 7 days using the NucleoCounter NC-200 Cell Counter (ChemoMetec). Daily samples of spent media were taken for extracellular metabolite concentration and titer measurements by drawing 0.8 mL from each culture, centrifuging it at 1000 g for 10 minutes and collecting the supernatant and discarding the cell pellet.

3.3.6 Titer determination

To quantify Enbrel/SERPING1, biolayer interferometry was performed using an Octet RED96 (Pall Corporation, Menlo Park, CA). ProA biosensors (Fortebio 18-5013) were hydrated in PBS and preconditioned in 10 mM glycine pH 1.7. A calibration curve was prepared using Enbrel (Pfizer) or SERPING1 at 200, 100, 50, 25, 12.5, 6.25, 3.13, 1.56, 0.78 µg/ml. Culture spent media samples were collected after centrifugation and association was performed for 120 s with a shaking speed of 200 rpm at 30 °C. Octet System Data Analysis 7.1 software was used to calculate binding rates and absolute protein concentrations.

3.3.7 Extracellular metabolite concentration measurements

The concentrations of glucose, lactate, ammonium (NH4), and glutamine in spent media were measured using the BioProfile 400 (Nova Biomedical). Amino acid concentrations were determined via High Performance Liquid Chromatography using the Dionex Ultimate 3000 autosampler at a flow rate of 1mL/min. Briefly, samples were diluted 10 times using 20 µL of sample, 80 µL MiliQ water, and 100 µL of an internal amino acid standard. Derivatised amino acids were monitored using a fluorescence detector. OPA-derivatised amino acids were detected at 340ex and 450em nm and FMOC-derivatised amino acids at 266ex and 305em nm. Quantifications were based on standard curves derived from dilutions of a mixed amino acid standard (250 ug/mL). The upper and lower limits of quantification were 100 and 0.5 ug/mL, respectively.

3.3.8 Estimation of the energetic cost secreting a protein as the number of ATP equivalent molecules

We estimated the energetic cost of synthesizing and secreting all 5,641 endogenous CHO cell proteins. These proteins were chosen for containing a signal peptide in their sequence and/or for being localized in the cell membrane (according to the UniProt database). The energetic cost (in units of number of ATP equivalents) of secreting each protein (length L) was computed using the following formulas and assumptions:

1. **Energy cost of translation**. For each protein molecule produced, ***2L*** ATP molecules are cleaved to AMP during charging of the tRNA with a specific amino acid; ***1*** GTP molecule is consumed during initiation and ***1*** GTP molecule for termination; L **-** ***1*** GTP molecules

are required for the formation of L-1 peptide bonds; ***L - 1*** GTP molecules are necessary for L-1 ribosomal translocation steps. Thus, the total cost of translation (assuming no proofreading) is ***4L***.

2. **Average cost of signal peptide degradation**. On average, signal peptides have a length of 22 amino acids. Thus, the average cost of degrading all peptide bonds in the signal peptide is ***22***. This average cost was assigned to all proteins analyzed.
3. **Energetic cost of translocation across the ER membrane**. During activation of the translocon, ***2*** cytosolic GTP molecules are hydrolyzed. From there, a GTP molecule bound to the folding-assisting chaperone BiP is hydrolyzed to GDP for every 40 amino acids that pass through the translocon pore [232]. Thus, the cost of translocation is **L/40 + 2**.
4. **Energetic cost of vesicular transport and secretion**. We used published data[233–235] (see Supp. File 1) to compute stoichiometric coefficients for reactions involving vesicular transport. That is, the number of GTP molecules bound to RAB and coat proteins in each type of vesicle (COPII and secretory vesicles). We found that a total of 192 and 44 GTPs must be hydrolyzed to transport one COPII or secretory (i.e. clathrin coated) vesicle from the origin membrane to the target membrane, respectively. Since vesicles do not transport single protein molecule at a time, we estimated the number of secreted protein molecules that would fit inside a spherical vesicle (see estimated and assumed diameters in Supp. File 1). For that, we assumed that the secreted protein is globular and has a volume V_P (nm^3) that is directly proportional to its molecular weight MW [236]:

$$V_P = MW \times 0.00121$$

Finally, we assumed that only 70 percent of the vesicular volume can be occupied by the target protein. Thus, the cost of vesicular transport via COPII vesicles with Volume V_{COPII} is:

$$192\ GTPs\ \div (V_{COPII}\ \times 0.7\ \div V)$$

Similarly, the cost of vesicular secretion is:

$$44\ GTPs\ \div (V_{Secretory}\ \times 0.7\ \div V)$$

3.3.9 Constraints used in models and Pareto optimality frontiers

All models were constrained using different sets of experimental uptake rates which can be found in Supp. File 3. To construct Pareto optimality frontiers, we used the robustAnalysis function from the COBRA Toolbox in Matlab 2015b using biomass as the control and secretion of the recombinant protein as the objective reactions, respectively.

3.3.10 Analysis of gene expression versus protein cost

Ribosome-profiling data [237] were used to quantify the ribosomal occupancy of each transcript in CHO cells. A cutoff of 1 RPKM was used to remove genes with low expression (10,045 genes removed from day 3 analysis and 10,411 from day 6 analysis). We used Spearman correlation to assess the variation of expression levels with respect to protein ATP cost.

3.3.11 CHO-DG44 model and prediction of neoR knock-out effect on specific productivity

Ribosome-profiling data, specific productivity, product sequence, and growth rates of an IgG-producing CHO-DG44 cell line were obtained previously [237]. From the same cultures, we obtained further cell dry weight and metabolomic data from spent culture medium for this study. The mCADRE algorithm [238],[239] was used to construct a DG44 cell line-specific iCHO2048s model. The specific productivity and the RPKM values of the secreted IgG were used to estimate the translation rate for the neoR selection marker gene. We assumed that the flux (in units of mmol/gDW/h) through the neoR translation reaction (v_{neoR}) should be proportional to that of the IgG translation rate (v_{IgG}, calculated from the measured specific productivity) and related to their expression ratios (i.e. the RPKM values of their genes in the ribosome-profiling data).

$$v_{neoR} = \frac{RPKM_{neoR}}{2(RPKM_{light} + RPKM_{heavy})} v_{IgG}$$

Finally, a reaction of neoR peptide translation (which is expressed in cytosol and is not processed in the secretory pathway) was added to construct a neoR-specific iCHO2048s model. Uptake and secretion rates of relevant metabolites on days 3 and 6 of cell culture were used to constrain our model. Because recombinant proteins represent 20% of total cell protein [240], we scaled the coefficients of all 20 amino acids in the model's biomass reaction accordingly (i.e. each coefficient was multiplied by 0.8). We then used FBA to predict the specific productivity of IgG with or without neoR. Finally, we used LIMAA to compute the residual fluxes of every amino acid in the medium to identify the limiting nutrients in each phase of the μ/q_p curves (i.e. the metabolites causing the hinges in the curve).

3.3.12 Cell dry weight measurements

For cell dry weight measurements, 6 tubes containing 2 mL of culture samples of known viable cell density and viability were freeze dried, weighed, washed in PBS, and weighed again. The difference in weight was used to calculate the mass per cell. The procedure resulted in an average cell dry weight of 456 pg per cell. As a simplification, we assumed that cell dry weight does not significantly differ from this average measured value during culture and thus was used when computing flux distributions in all simulations.

3.3.13 Calculation of amino acid uptake, growth rates and specific productivity from experimental data

Experimental uptake and secretion rates from different studies were used to constrain the iCHO2048s models [237, 240, 241]. When rates were not explicitly stated in the studies we consulted, we used a method we developed in Chapter 2 [242]. Briefly, appropriate viable cell density, titer, and metabolite concentration plots were digitized using WebPlot Digitizer software and we computed the corresponding rates as follows:

- Growth rate (in units of inverse hours):

$$\mu = \frac{1}{VCD}\frac{d}{dt}VCD$$

  Where VCD is the viable cell density (in units of cells per milliliter)

- Specific productivity (in units of picograms per cell per hour):

$$q_p = \frac{1}{VCD}\frac{d}{dt}Titer$$

- Consumption or production rate v_x of metabolite x (in units of millimoles per gram dry weight per hour):

$$v_x = \frac{1}{VCD}\frac{d}{dt}[x]$$

3.4 Results

3.4.1 *In silico* reconstruction of the mammalian protein secretion pathway

We mapped out the core processes involved in the synthesis of secreted and membrane proteins in mammalian cells (i.e. human, mouse, and Chinese hamster). This included 261 components (gene products) in CHO cells and 271 components in both human and mouse. The components are involved in secretory reactions across 12 subsystems (i.e., functional modules of the secretory pathway; Figure 3.1A). These components represent the core secretory machinery needed in the transition of a target protein from its immature state in the cytosol (i.e., right after translation) to its final form (i.e., when it contains all post-translational modifications and is secreted to the extracellular space). Each component in the reconstruction either catalyzes a chemical modification on the target protein (e.g., N-linked glycosylation inside ER lumen/Golgi) or participates in a multi-protein complex that promotes protein folding and/or transport. This distinction between catalytic enzymes and complex-forming components is important for modeling purposes as a catalytic component consumes or produces metabolites that are directly connected to the metabolic network (e.g., ATP, sugar nucleotides). Because all components of the core secretory pathway were conserved across human, mouse and hamster (Figure 3.1B), we generated species-specific reconstructions and used them to expand the respective genome-scale metabolic network reconstructions (Recon 2.2 [243], iMM1415 [244], iCHO1766 [241]) and

called these metabolic-secretory reconstructions Recon 2.2s, iMM1685s, and iCHO2048s, respectively.

3.4.2 Validation of iCHO2048s model predictions

We first validated the accuracy of iCHO2048s predictions using published growth and specific productivity rates of IgG-producing CHO cell lines from two independent studies [237, 245]. For this, we built an IgG-secreting iCHO2048s model using the information in the PSIM matrix for the therapeutic mAb Rituximab. We then constrained the model's Rituximab-specific secretory pathway with the reported productivity value in each study and used FBA to predict growth (Figure 3.2A). Later, to assess iCHO2048s ability to predict growth rates in cases where CHO cells produce proteins different from antibodies, we collected data from two batch culture experiments using Enbrel- and SERPING1-producing isogenic CHO cell lines. We constructed two iCHO2048s models for each case and predicted growth rates on days 1 (early exponential growth phase) through 5 (late exponential growth phase) of culture while constraining the protein secretion rate to the measured specific productivity value (Figure 3.2B-C). The model's predictions were in very good agreement with the reported/measured values. There were cases where iCHO2048s predicted a much higher growth rate than what was measured in the first days of batch culture (Figure 3.2B-C). Since FBA computes theoretical maximum growth rates given a set of constraints, these over-prediction cases point at situations where CHO cells do not direct resources towards biomass production (during very early stages of culture), a discrepancy that is attenuated in later stages of culture (days 4-5 in Figure 3.2B-C). In conclusion, these results confirm the ability of protein-specific reconstructions to capture the specific energetic requirements that each recombinant product imposes on CHO cell metabolism.

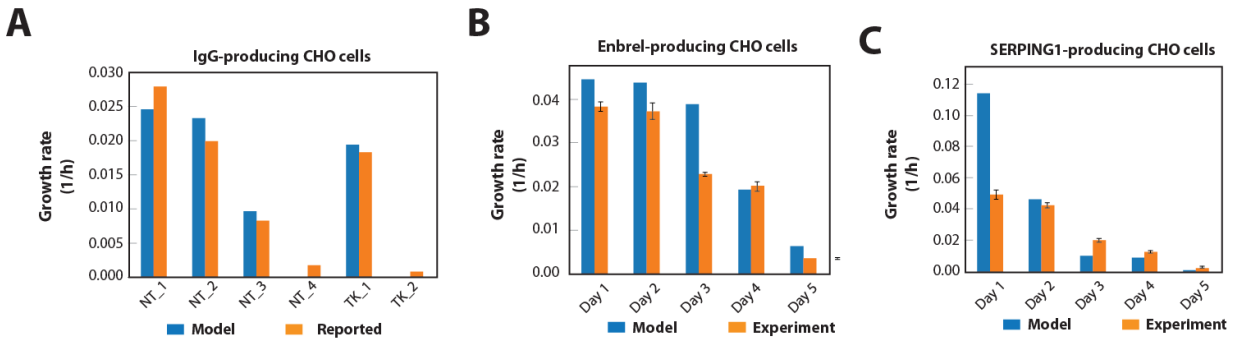

Figure 3.2: Validation of iCHO2048s predictions. (A) Comparison of growth rates reported in six datasets from two previous studies using IgG-producing cell lines and rates predicted by an IgG-specific iCHO2048s model. NT and TK stand for the main author initials in both studies (Neil Templeton, Thomas Kallehauge). **(B)** Comparison between experimental and predicted growth rates of Enbrel-producing CHO cells. **(C)** Comparison between experimental and predicted growth rates of SERPING1-producing CHO cells. Error bars in B and C represent the standard deviation of three biological replicates. In all cases, the iCHO2048s models were constrained to produce the recombinant protein at the reported/measured specific productivity rate.

3.4.3 Protein composition and complexity significantly impact model-predicted productivity

To produce a specific product, CHO cells must utilize different modules of the secretory pathway according to the protein attributes and post-translational modifications (PTMs). For example, the synthesis of a monoclonal antibody (mAb) requires the use of multiple processes and consumes several different metabolites, such as amino acids for protein translation, ATP equivalents for vesicular transport, and sugar nucleotides for protein glycosylation (Figure 3.1C). Having validated our iCHO2048s model predictions, we set out to generate eight product-specific secretory pathway models for biotherapeutics commonly produced in CHO cells (Figure 3.3A): bone morphogenetic proteins 2 and 7 (BMP2, BMP7), erythropoietin (EPO), Etanercept, factor VIII (F8), interferon beta 1a (IFNB1), Rituximab, and tissue plasminogen activator (tPA). The resulting iCHO2048s models were used to compute Pareto optimality frontiers between maximum cell growth (μ) and specific productivity (q_P) assuming all eight CHO cells grow under the same

conditions. That is, all eight models were given the same measured glucose and amino acid uptake rates [241] as model constraints.

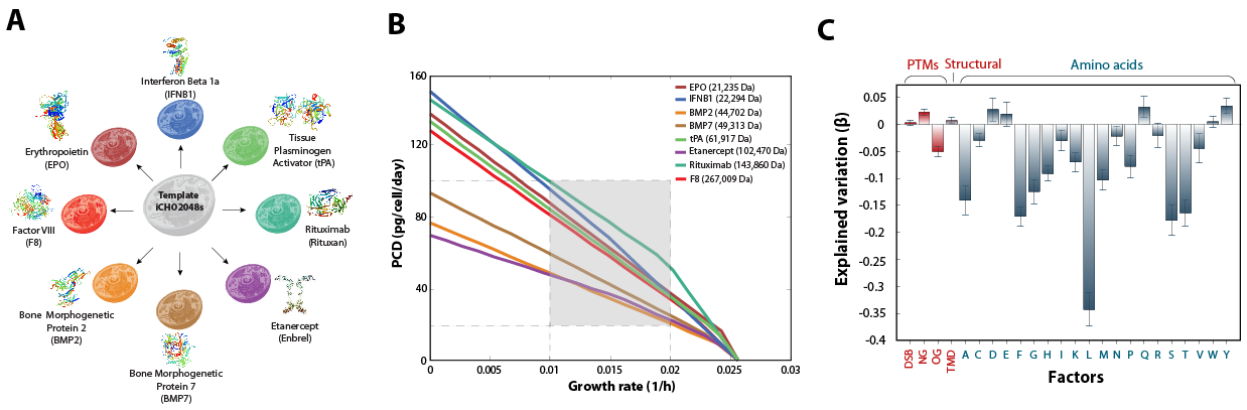

Figure 3.3: Construction of product-specific iCHO2048s models. (A) Eight product-specific iCHO2048s models were constructed for biotherapeutics commonly produced in CHO cells. **(B)** Pareto optimality frontiers of growth/productivity (μ/q_p) trade-off curves were computed for the eight iCHO2048s models using the same constraints and experimental data from Supp. File 3. The shaded region corresponds to commonly observed cell growth rates in CHO cell cultures. The molecular weight (in Daltons) of each biotherapeutic is shown in the legend. **(C)** All protein features (PTMs, transmembrane domains, and amino acid compositions) were analyzed to quantify their contribution to the explained variation of specific productivity.

We computed the tradeoff between growth (μ; inverse hours or 1/h) and specific productivity (q_p; pg protein produced per cell per day or PCD) as a Pareto optimal "μ/q_p curve" for each protein (Figure 3.2B). This curve defines the frontier of maximum specific productivity and maximum growth rates under the assumption that CHO cells can utilize all available resources towards production of biomass and recombinant protein only. The hinges in some of the μ/q_p curves are indicative of a transition between regions in the μ/q_p that are limited by distinct amino acids.

An analysis of the μ/q_p curves for the eight biotherapeutics demonstrate that under the measured growth conditions, maximum productivities vary from 20-100 PCD at common growth rates (Figure 3.2B, shaded region) to 70-150 PCD for senescent cells. Neither the molecular weight

(MW) nor product length can explain the 2-fold range differences in maximum productivity for different proteins. For example, the μ/q_p curves show tPA (MW = 61,917 Da) can express at higher PCD than BMP2 (MW = 44,702 Da) despite being larger, because the N-glycans in BMP2 reduce productivity due to the higher cost of synthesizing core N-glycans (see Table 3.2), consistent with previous observations in yeast [186]. Furthermore, the degree and directionality of these effects will depend on the nutrient uptake rates used to constrain the model, highlighting the need in CHO bioprocessing to tailor culture media in a host cell and product-specific manner. Thus, while intuitively larger proteins would be expected to exert more bioenergetic cost on protein secretion, we find that specific compositional attributes of both the recombinant protein and the culture media significantly impact biosynthetic capacity.

Table 3.2: Protein specific information matrix of biotherapeutics secreted in eight iCHO2048s models

Protein Name	Total number of amino acids in biotherapeutic	Molecular Weight [Da]	Total number of disulfide bonds in mature protein	Total number of N-glycans in mature protein	Total number of O-glycans in mature protein	Estimated secretory cost [ATP equivalents]
IFNB1	187	22294	1	1	0	777
EPO	193	21037	2	3	1	801
BMP2	396	44702	4	5	0	1618
BMP7	431	49313	4	4	0	1759
tPA	562	61917	17	3	1	2286
Etanercept**	934	102470	7	6	26	3784
Rituximab*	1328	143860	17	2	0	5370
F8	2351	267009	8	22	0	9488

*** Rituximab is a tetramer (2 light and 2 heavy chains)**
**** Etanercept is a dimer**

3.4.4 iCHO2048s accurately predicts protein productivity increase following gene knock-down

A recent study by Kallehauge [237] demonstrated that a CHO-DG44 cell line producing an antiviral mAb [246] also expressed high levels of the neoR selection-marker gene (Figure 3.4A-B). Upon neoR knockdown, the titer and maximum viable cell densities of the CHO-DG44 cell line were increased. To test if iCHO2048s could replicate these results, we constructed a model for the Kallehauge DG44 cell line and measured exometabolomics, and dry cell weight to parameterize the model. Since expression of neoR uses resources that could be used for antibody production, we predicted how much additional antibody could be synthesized with the elimination of the neoR gene. We simulated antibody production following a complete knockout of neoR (Table 3.3 and Figure 3.4B) and predicted that the deletion of neoR could increase specific productivity by 4% and 29% on days 3 (early exponential phase) and 6 (late phase) of culture, respectively (Fig. 4C). This was consistent with the experimentally observed values of 2% and 14%. We then computed the μ/q_p curves for both the control and the neoR *in silico* knockout conditions on day 6. We found that the length of the μ/q_p curve (i.e. the size of the set of Pareto optimal flux distributions, here denoted by delta (Δ)) increased by 18% percent when neoR production is eliminated (Figure 3.4D). Thus, iCHO2048s can quantify how much non-essential gene knockouts can boost growth and productivity in CHO cells by freeing energetic and secretory resources. In fact, the ribosome-profiling data from Kallehauge revealed that only 30 secretory proteins in CHO cells account for more than 50% of the ribosomal load directed towards translation of protein bearing a signal peptide (Figure 3.4E).

Table 3.3: Experimental data used for validation of Kallehauge iCHO2048s model

Experimental value description	Day 3 (early exponential growth phase)	Day 6 (late growth phase)
Growth Rate [1/day]	0.44	0.02
Specific Productivity [Picograms of IgG/cell/day]	16	5.5
Total IgG ribosomal footprint [RPKM]	40258	13356
Total neoR ribosomal footprint [RPKM]	36952	25679

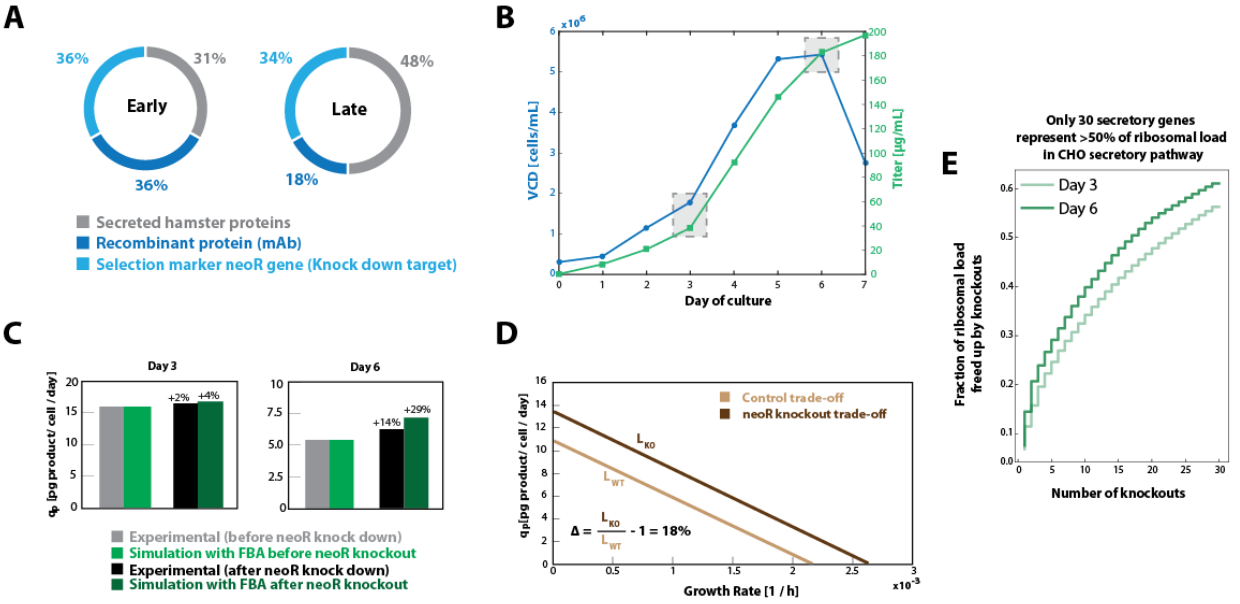

Figure 3.4: iCHO2048s recapitulates experimental results of neoR knock-down in silico. (A) Ribosome occupancy was measured with ribosomal profiling during early (left) and late (right) exponential growth phases in the study by Kallehauge [237]. **(B)** Time profiles of viable cell density (VCD) and titer in experimental culture. Shaded boxes indicate the time points corresponding to early (day 3) and late (day 6) growth phases. **(C)** Flux balance analysis predictions of specific productivity (q_p) with iCHO2048s model before and after *in silico* knockout of neoR gene. **(D)** Growth/productivity (μ/q_p) trade-offs predicted by iCHO2048s demonstrated a potential 18% increase after the neoR *in silico* knockout. The formula for calculating the trade-off improvement (Δ) is shown in the plot. L_{WT} = length of μ/q_p curve before knockout, L_{KO} = length of μ/q_p curve after knockout. **(E)** Cumulative plot of the ribosomal load demanded by the top 30 secreted proteins in CHO cells based on Kallehauge ribosome profiling data on days 3 and 6 of culture.

3.4.5 CHO cells have suppressed expression of expensive proteins in their secretome

In any cell, the secretory machinery is concurrently processing thousands of secreted and membrane proteins, which all compete for secretory pathway resources and pose a metabolic burden. To quantify this burden, we estimated the energetic cost of synthesizing and secreting all 5,641 endogenous proteins in the CHO secretome and membrane proteome in terms of total number of ATP equivalent molecules consumes. These protein costs were compared to the cost of the eight recombinant proteins previously analyzed. To refine estimates, we experimentally measured the number of N-linked glycans in the CHO proteome and integrated published numbers of O-linked glycans in CHO proteomic data [227]. Across the CHO secretome, protein synthesis cost varies substantially, and recombinant products are on average more expensive (Fig. 5A). For example, F8 is a "difficult-to-express" protein in CHO cells due to its propensity to aggregate in the ER, which promotes its premature degradation [247, 248]. Our analysis further highlights that each molecule of F8 requires an excessive amount of ATP to produce (9488 ATP molecules). This imposes a significant burden to the secretory machinery of CHO cells, which typically expresses much less expensive proteins. With the broad range of biosynthetic costs for different proteins, we wondered if gene expression in CHO cells has been influenced by the ATP cost of secreted proteins, by suppressing host cell protein expression to more efficiently allocate nutrients. That is, unless specific proteins are essential, CHO cells would preferentially suppress energetically expensive proteins. To test this, we analyzed ribosomal profiling (Ribo-seq) data from a mAb-producing CHO cell line [237] and compared translation of each transcript against the ATP cost of the associated secreted protein (see Methods). Indeed, there was a significant negative correlation (Spearman) of -0.43 and -0.36 (p value < $1x10^{-20}$) between ribosomal occupancy and

ATP cost during early and late phases of culture, respectively (Figure 3.5B). Wondering if the reduced translation was regulated transcriptionally, we further analyzed RNA-seq data from the same mAb-producing cell line and from another, non-producing CHO-K1 cell line [249]. The RNA expression also negatively correlated with ATP cost (see Figure 3.6).

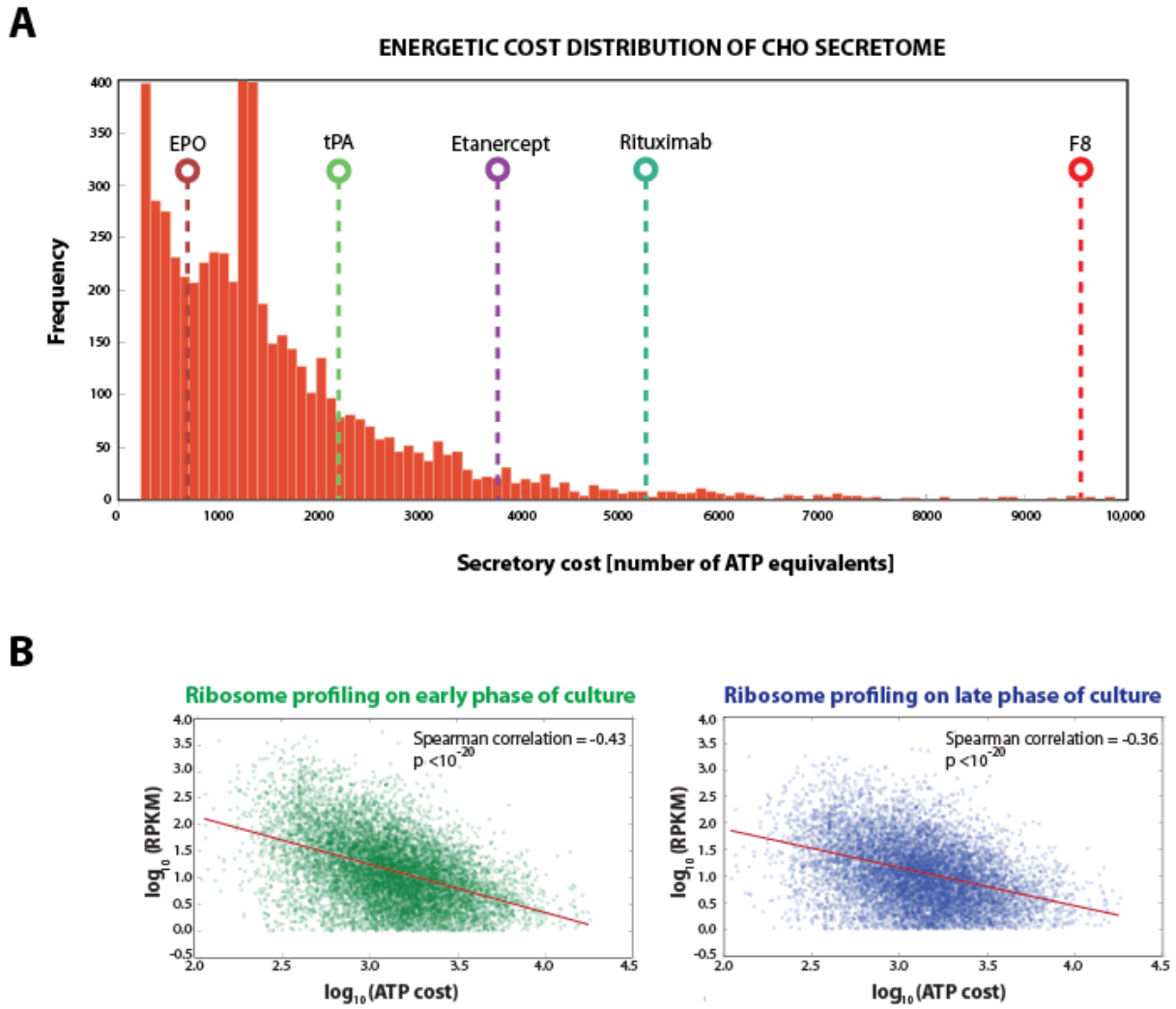

Figure 3.5: Energetic cost of CHO secretome and gene expression as a function of protein cost. (A) The bioenergetic cost of each secreted CHO protein was computed, 28 of the 5641 proteins in CHO secretome had a cost > 10,000 ATP equivalents and were therefore not included in the histogram for the sake of ease of visualization. The biosynthetic costs of 5 representative biotherapeutics are shown for comparison purposes (see Table 3.2) **(B)** Scatter plots and Spearman correlation of gene expression and protein cost (in number of ATP equivalent molecules) from Kallehauge [237] during early (left) and late (right) phases of culture. RPKM = reads per kilobase of transcript per million.

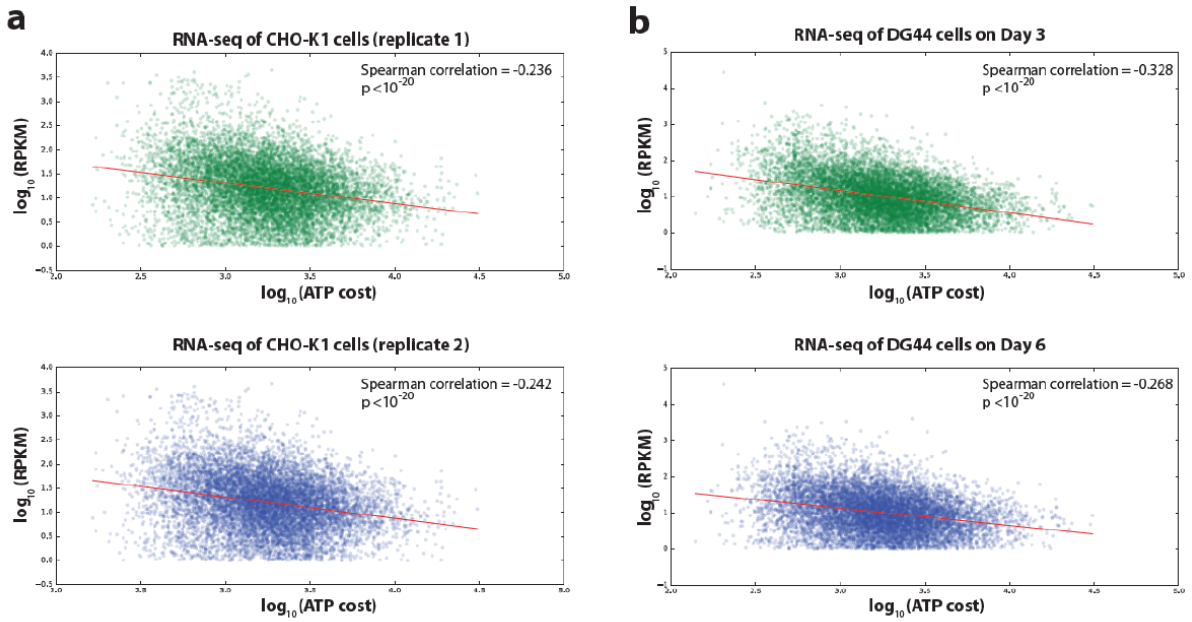

Figure 3.6: Negative correlation between ATP cost and gene expression is consistent even for RNA-seq data. (**A**) Analysis of RNA-seq data from a different CHO cell line (non-producing CHO-K1) for two biological replicates. (**B**) Analysis of RNA-seq from the same cell line as in Figure 3.5. The correlation tends to be stronger for CHO cell lines producing a recombinant protein.

Finally, we analyzed RNA-seq data from human tissues and immortalized cell lines in the Human Protein Atlas (HPA) [183]. All RNA-seq datasets in the HPA samples also negatively correlated with ATP cost. Interestingly, we found that highly secretory tissues such as liver, pancreas and salivary gland had the strongest correlations, although none as strong as that of the mAb-producing CHO cells (Figure 3.7). A recent study by Feizi and colleagues recently found that these tissues fine-tune the expression of protein disulfide isomerase genes [250], suggesting that a similar regulatory process may take place in the ER of CHO cells as the secreted mAb contains a relatively high number (17) of disulfide bonds. In conclusion, there is a clear preference in CHO cells to suppress the expression and translation of proteins that are costly to synthesize, fold, and secrete.

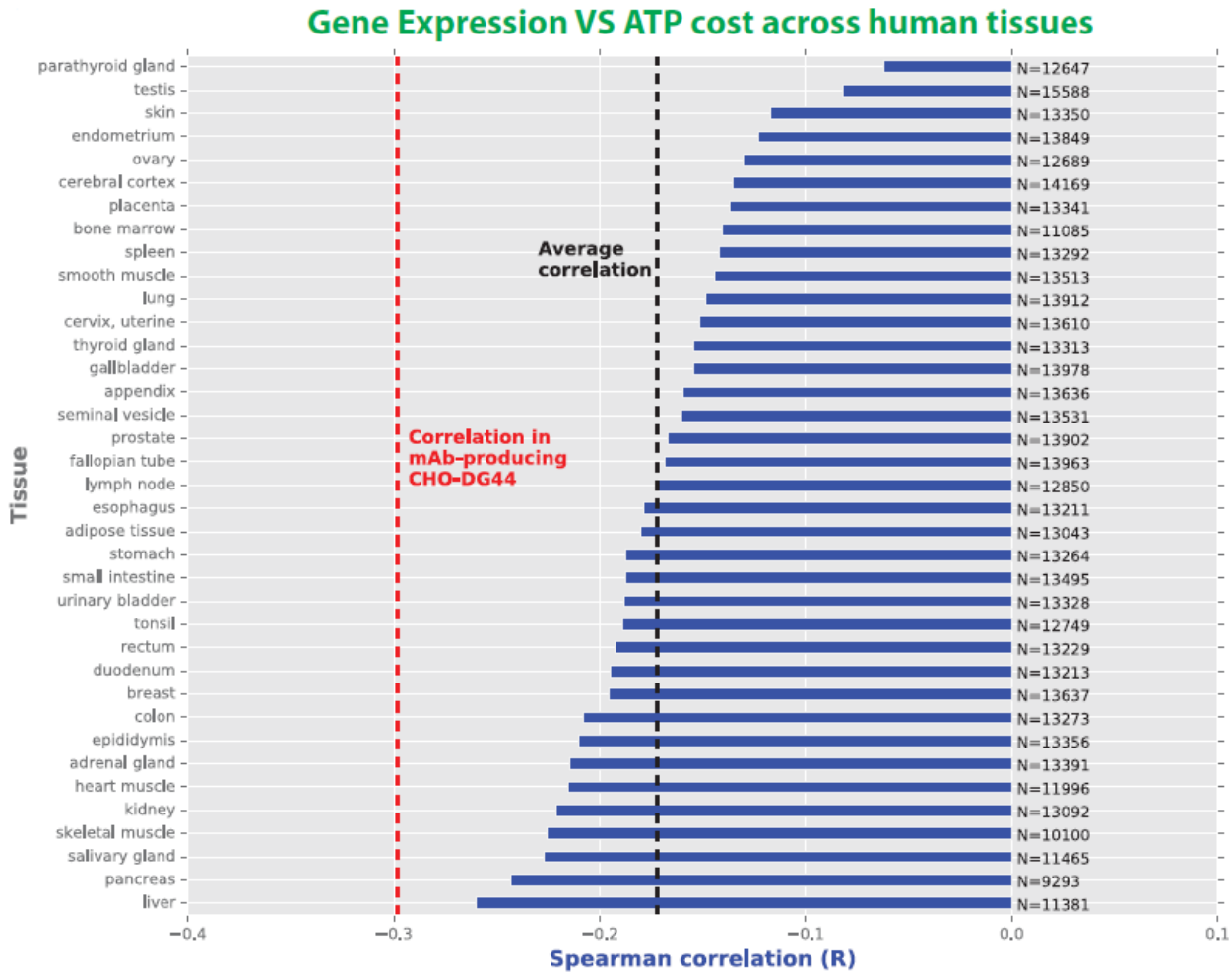

Figure 3.7: Correlations of ATP cost and gene expression across human tissues. Spearman correlations are stronger for highly secretory tissues such as liver and pancreas and lower for non-secretory tissues like the endometrium and cerebral cortex. The Spearman correlation of the CHO-DG44 cell line (red dashed line) is shows for the sake of comparison

3.5 Discussion

Mammalian cells synthesize and process thousands of proteins through their secretory pathway. Many of these proteins, including hormones, enzymes, and receptors, are essential for mediating mammalian cell interactions with their environment. Therefore, many have therapeutic importance either as drugs or as targets. CHO cells in particular, have been excellent hosts for large-scale production of therapeutic proteins, and also effective at processing post-translational

modifications (PTMs) on products passing through their secretory pathway, rendering these products suitable and safe for human prescription [251, 252]. The expression and secretion of recombinant proteins represents a significant anabolic demand that drains several substrates from cellular metabolism (e.g., amino acids, sugar nucleotides, ATP) [253, 254]. There has been an increasing interest in engineering the CHO secretory pathway [255–257]. Despite important advances in the field [242], current strategies to engineer the secretory pathway have remained predominantly empirical [258, 259]. Recent modeling approaches, however, have enabled the analysis of the metabolic capabilities of important eukaryotic cells (including CHO) under different genetic and environmental conditions [181, 241, 260, 261]. With the development of the first genome-scale model of CHO cell metabolism [241] it is now possible to gain a systems-level understanding of the CHO phenotype [262].

For instance, Lund and colleagues [263] have recently reconstructed the most comprehensive network of the mouse secretory pathway to date. By comparing the mouse and CHO-K1 genomes mapping CHO gene expression data onto this network, the authors identified potential targets for CHO cell engineering, demonstrating the great potential of systems biology to interrogate and understand protein secretion in animal cells. This network reconstruction, although useful for contextualizing omics data (e.g., RNA-seq), is not set up for simulations of protein production, nor integrated with additional cellular processes such as metabolism. Therefore, to quantify the cost and cellular capacity for protein production, it is important to delineate the mechanisms of all biosynthetic steps and bioenergetic processes in the cell.

3.6 Conclusion

Here we have presented the first genome-scale reconstruction of the secretory pathway in mammalian cells. We connected this to current metabolic networks, yielding models of protein secretion and metabolism for human, mouse and CHO cells. These models compile decades of research in biochemistry and cell biology of higher eukaryotes and present it in a mathematical model. Using our model, we quantitatively estimated the energetic cost of producing several therapeutic proteins and all proteins in the CHO cell and human secretomes. We also identified factors limiting the secretion of individual products and observed that these depend on both the complexity of the product and the composition of the culture media. Furthermore, by integrating ribosomal profiling data with our model we found that CHO cells have selectively suppressed the expression of energetically expensive secreted proteins. Expanding upon this observation, we demonstrated that specific productivities can be predictably increased following the knock-down of an energetically expensive, non-essential protein.

It is important to note that while our models capture major features of secreted proteins, there are additional PTMs (e.g., phosphorylation, gamma carboxylation), pathway machinery (e.g., chaperones), and cell processes that could possibly be captured in further expansions of the modeling framework [263] (e.g, the unfolded protein response). These could be included as energetic costs associated to building and maintaining the secretory machinery (chaperones [84], glycosyltransferases [264]); protein stability and turnover rates [265]; solubility constraints [266] and molecular crowding effects [267]. As these are captured by the models in a protein product-specific manner, predictions of protein production capacity will improve, and the models could provide further insights for cell engineering for biotechnology or to obtain a deeper understanding of mechanisms underlying amyloid diseases. Finally, a simplification of our secretory model is

that it only computes the bioenergetic cost of synthesizing and attaching core N- and O-linked glycans to secreted proteins. Thus, an immediate potential expansion of our secretory model would involve coupling it to existing computational models of protein glycosylation [85]. For example, given an N-glycan reaction network that captures the glycoform complexity of a target protein [268], one could build secretory reactions for the specific glycoforms of interest and compute the metabolic demands associated with each of them as to identify potential targets and nutrient supplementations for glycoengineering.

In conclusion, the results of our study have important implications regarding the ability to predict protein expression based on protein specific attributes and energetic requirements. The secretory pathway models here stand as novel tools to study mammalian cells and the energetic trade-off between growth and protein secretion in a product- and cell-specific manner. We presented algorithms that provide novel insights with our models, and expect that many other methods can be developed to answer a wide array of questions surrounding the secretory pathway, as seen for metabolism [269].

3.7 Acknowledgements

The authors would like to thank Philipp Spahn, Austin Chiang, and Chih-Chung Kuo for their insightful comments on this manuscript. This work was supported by generous funding from the Novo Nordisk Foundation provided to the Center for Biosustainability at the Technical University of Denmark (NNF10CC1016517 and NNF16CC0021858), and from NIGMS (R35 GM119850), and a fellowship from the Government of Mexico (CONACYT) and the University of California Institute for Mexico and the United States (UC-MEXUS).

Chapter 3 in part is from a manuscript that is under review for publication *Gutierrez JM*, Feizi A*, Li S, Kallehauge TB, Grav LM, Hefzi H, Ley D, Hizal DB, Betenbaugh MJ, Voldborg B, Kildegaard HF, Lee GM, Palsson BO, Nielsen J, Lewis NE. Genome scale reconstruction of the mammalian secretory pathway predict metabolic costs and limitations of protein secretion.* The dissertation author was one of the two primary authors.

Chapter 4 - Mapping protein turnover rates onto the human secretory pathway reconstruction reveals tissue-specific modular control of pathway usage

4.1 Abstract

Gene expression is a key determinant of a cell's phenotype and is determined by both the synthesis and degradation rates of proteins and transcripts. Across human tissues, 20% of the genes transcribed translates into the secretome (the entire collection of secreted and membrane-bound proteins). Because the secretome holds the largest fraction of the tissue-specific proteome, it is important to understand how the expression of secretory pathway genes is modulated to satisfy the secretory demands of each cell. Here we apply a systems biology approach to characterize the expression of secretory pathway genes across three highly secretory human cell types. We found that the usage levels of secretory pathway components significantly differ across three human cell types even though the energetic expenditure in all cell types is similar. Furthermore, the usage levels seem to be determined by how fast cells renew their secretory pathway components (that is, the component turnover rates). The results from this analysis extend our current understanding of gene expression in the secretory pathway and may have important applications in cell engineering for biotherapeutic production.

4.2 Introduction

Mammalian cells are continuously synthesizing numerous signaling proteins, hormones, receptors, and structural proteins that determine their function in the context of an organism. According to a recent analysis of the human proteome [183], the collection of proteins synthesized across tissues includes >3000 secreted proteins (e.g., enzymes, hormones, antibodies, extracellular matrix proteins, etc.) and >5500 membrane proteins, which are processed in the secretory pathway. In fact, the secretome explains the largest fraction of tissue-specific proteomic differences [250]. Because secreted proteins are continuously synthesized and exported out of the cell, the secretory pathway must cope with the energetic and usage demands imposed by the secreted proteins' turnover rates. Due to its importance, many studies have explored the properties of protein expression and protein turnover rates in human cells [270, 271]. However, these studies have only provided a general quantitative assessment of protein levels while paying very little attention to the secretory pathway and the demands that the secretome might impose upon it. In a recent study, however, researchers used mRNA levels from human tissues to investigate the differences in expression levels of secretory pathway genes [250]. Interestingly, they found tissue-specific fine tuning of secretory pathway gene expression that seems to satisfy the specific secretory demands of the tissue's secretome. However, it has been recognized that protein synthesis and degradation rates, and not transcription rates alone, heavily contribute to protein concentrations [272]. Therefore, there is still room for exploring the properties of secretory gene expression in the context of translation and transcription rates.

In this chapter, we present an analysis of secretory gene expression and secretory pathway usage across three human cell lines, which are characterized by high levels of protein secretion: hepatocytes, monocytes, and B cells, which are capable of secreting large quantities of albumin,

cytokines, and antibodies, respectively. First, we use protein abundances and protein turnover rates of hepatocytes, monocytes and B cells to derive a quantitative estimate of their secretory pathway demands. We found that the three cells utilize about the same energy to produce their secretomes, but the way in which they utilize different secretory pathway genes differs significantly. In fact, how much a cell type utilizes a secretory pathway subsystem corresponds very well with the mean half-life of the proteins that comprise such subsystem.

The results from this chapter have important implications in cellular physiology and biotechnology and expand our current understanding of the fundamental properties of protein secretion. Understanding such properties will aid the development of novel expression systems to produce recombinant biotherapeutics in animal cells (e.g. CHO or HEK293 cells).

4.3 Methods

4.3.1 Datasets

We obtained transcription and translation rates as well as protein abundances in human cell lines from [273]. Protein turnover rates of hepatocytes, monocytes and B cells were obtained from the supplementary materials in [274] whereas protein abundances were downloaded from the PaxDB (version 4.0) website [275].

4.3.2 Secretory pathway genes and subsystems

Human secretory pathway genes were taken from Recon 2.2s, which we developed in Chapter 3 of this dissertation. Each component in Recon 2.2s is associated to one of twelve subsystems (e.g. protein folding, protein translocation) which we used to make more specific comparisons across cells.

4.3.3 Computation of ATP costs and turnover in secretome production

We used the ATP cost formulation from Chapter 3 (section 3.3.8) to compute the ATP cost of all secreted proteins in the hepatocyte, B cell, and monocyte datasets. We then multiplied each ATP cost by the protein's degradation rate (α_p, in proteins per hour) which we calculated from the reported half-life ($t_{1/2}$) of that protein:

$$\alpha_p = \frac{\ln(2)}{t_{1/2}}$$

Note that here we assume first-order exponential decay. Finally, we computed the overall ATP turnover due to secretome cost by adding up all individual ATP turnovers from all secreted proteins:

$$ATP_{turnover} = \sum_{i=secreted\ protein} atp_{cost,i} * \alpha_{p,i}$$

4.3.4 Computation of secretory pathway usage

The secretory pathway usage for a specific module is defined as the number of catalytic cycles per hour at steady-state. To compute this, we first wrote a Python script to generate protein-specific stoichiometric reactions for each of the secreted proteins in the protein turnover dataset. The stoichiometric coefficients and the protein turnover rates from each reaction were used to estimate the flux through each secretory reaction. The mean flux through the reactions comprising a secretory pathway module was used to estimate the pathway usage demanded by each secreted protein. In total, the protein turnover rates data from monocytes, hepatocytes and B cells included 219 out of the 271 (80%) secretory pathway components in Recon 2.2s and 1075 out of the 4269

(25.2%) proteins in the human secretome. From these components, secreted proteins, and turnover rates, the global pathway usage of all proteins was computed by adding up the individual protein's pathway usages:

$$PU_i = \frac{\sum_j r_j c_i}{N}$$

Where PU_i is the pathway usage demanded by the ith secreted protein, whose concentration is c_i and the secretory flux it demands from the jth secretory reaction is r_j, normalized by the number of reactions (N) in the secretory module.

4.4 Results

4.3.1 Distribution of secretory pathway gene expression

Gene expression is fundamentally determined by the four rates of the central dogma of molecular biology: mRNA transcription, protein translation, mRNA degradation, and protein degradation (Figure 4.1A). These four quantities determine the steady-state concentration of each protein and can be mapped onto a four-dimensional space called the Crick space [273]. Because mRNA and protein degradation are not measured as often as transcription and translation rates, researchers often reduce the dimensionality of Crick space to two dimensions only (Figure 4.1B). Nevertheless, protein levels are determined by both synthesis and degradation rates so that translation rates alone are only able to capture one half of the protein dynamics equation [276]. Fortunately, high-throughput technologies that measure protein degradation rates have recently been developed [277, 278].

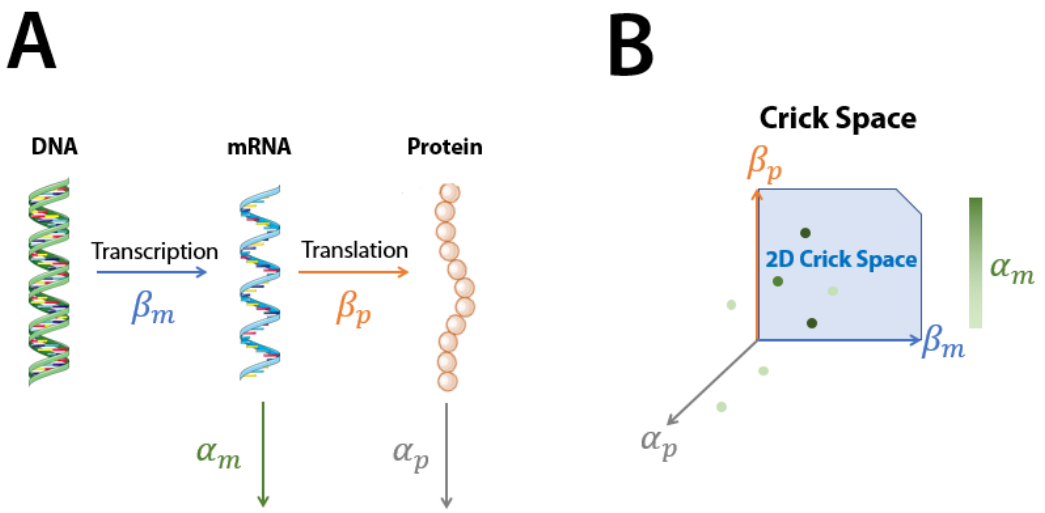

Figure 4.1: The four fundamental rates of gene expression. (**A**) The steady-state concentration of any protein is determined by four rates: transcription rate (β_m), mRNA degradation rate (α_m), translation rate (β_p), and protein degradation rate (α_p). (**B**) Crick space is defined as the four-dimensional space containing these four rates for each gene. Two-dimensional Crick space comprises transcription and translation rates only.

Here we asked whether the genes in the secretory pathway of human cells have a significantly different distribution of protein degradation rates (in terms of protein half-lives). First, we looked at the differences in global distribution of protein half-lives across three human cell types. We found that the global distribution (N=8623) of protein half-lives does not differ significantly across the three cells (p=0.47, Mann-Whitney U test), as shown in Figure 4.2A. However, the distribution of secretory pathway genes (N=219) is significantly different (p<0.001, Mann-Whitney U test, Figure 4.2B). Furthermore, the distributions significantly differ for some of the secretory pathway subsystems, but not for all (Figure 4.3A).

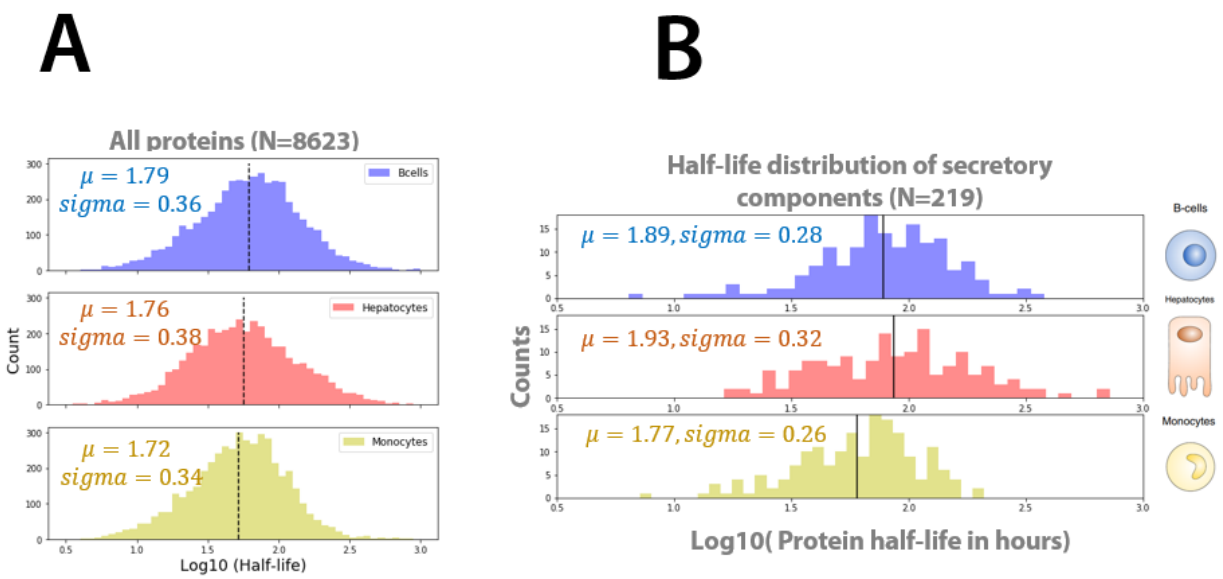

Figure 4.2: Distribution of secretory pathway gene half-lives. (**A**) Distribution of global protein half-lives across B cells, hepatocytes, and monocytes. The three distributions do not differ significantly (p=0.47, Mann-Whitney U test). (**B**) On the other side, the distribution of secretory pathway components does differ across cells ($p<0.001$ for all three pairwise comparisons, Mann-Whitney U test).

The above results suggest that the degradation of secretory pathway genes is regulated at the subsystem level. Furthermore, we wondered if the observed differences in secretory pathway gene half-lives (that is, along the 3rd dimension in Crick space) could be associated to the usage levels of each secretory subsystem. In other words, we wanted to investigate if half-lives of secretory pathway components are linked to the degree at which the component is utilized in the cell. To do so, we set out to formulate an expression for computing usage levels for each of the secretory components and subsystems and compare those against their measured half-lives [274].

4.3.2 The mean half-life of secretory machinery components differs across cell types and across subsystems

We observed significant differences are found when looking at the mean half-lives of the components in each secretory pathway module (Figure 4.3B). The proteins associated to protein

folding and glycosylation in the endoplasmic reticulum (ER) have a significant (2-fold) longer half-life in hepatocytes compared to the monocytes and B cells. This is interesting since the main secreted protein in hepatocytes is albumin, which demands a total of 17 disulfide bonds and 24 N-linked glycans upon entering the ER lumen. In fact, it has been estimated that a single hepatocyte secretes 1.3 pg of albumin every hour [279]. Therefore, this result suggests that hepatocytes regulate the half-life of the secretory pathway components in the ER to cope with the very high demand of albumin in the body (Figure 4.4). In contrast, monocytes are known for secreting cytokines such as TNF-α, IL-8, IL-1β, and IL-6 at levels ranging between 5 and 15 fg per hour [280]. Cytokines lack post-translational modifications and thus further supports the observation that monocytes have the fastest turnover in their ER secretory pathway components among the three cell types (Figure 4.3B).

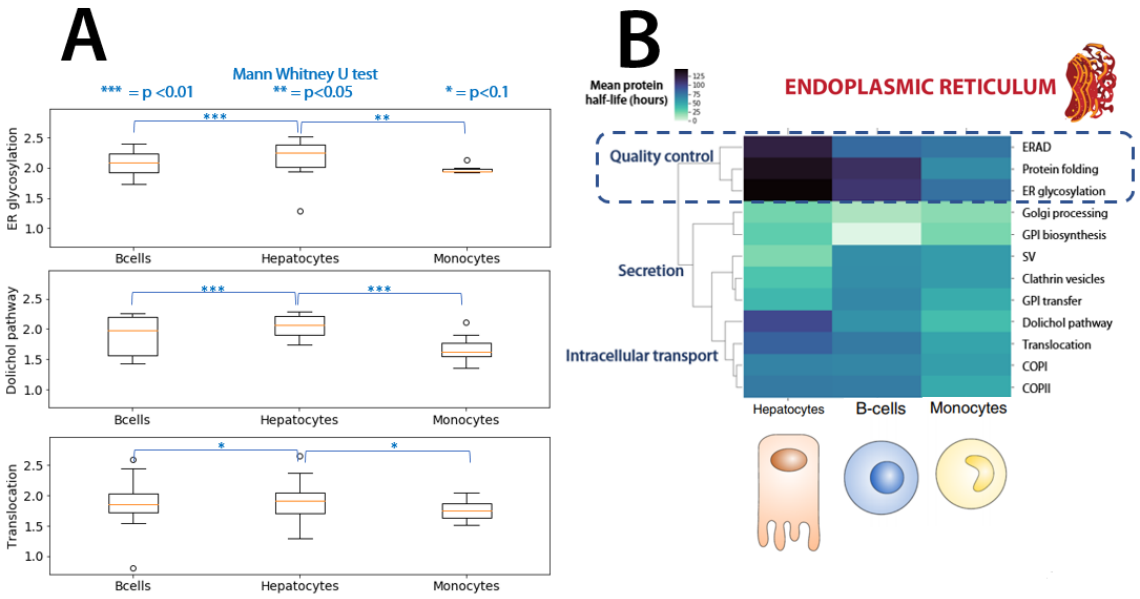

Figure 4.3: Mean half-lives of secretory pathway components across subsystems. (**A**) The distribution of secretory pathway components half-lives (in units of hours) across monocytes, hepatocytes, and B cells varies from one secretory subsystem to the other. In the vertical axis, half-lives were log10 transformed. (**B**) Heatmap of mean half-life of secretory pathway subsystems. Hepatocytes show the longest mean half-lives in their ER processing modules, probably due to the high requirements for folding and glycosylation that albumin demands from these cells.

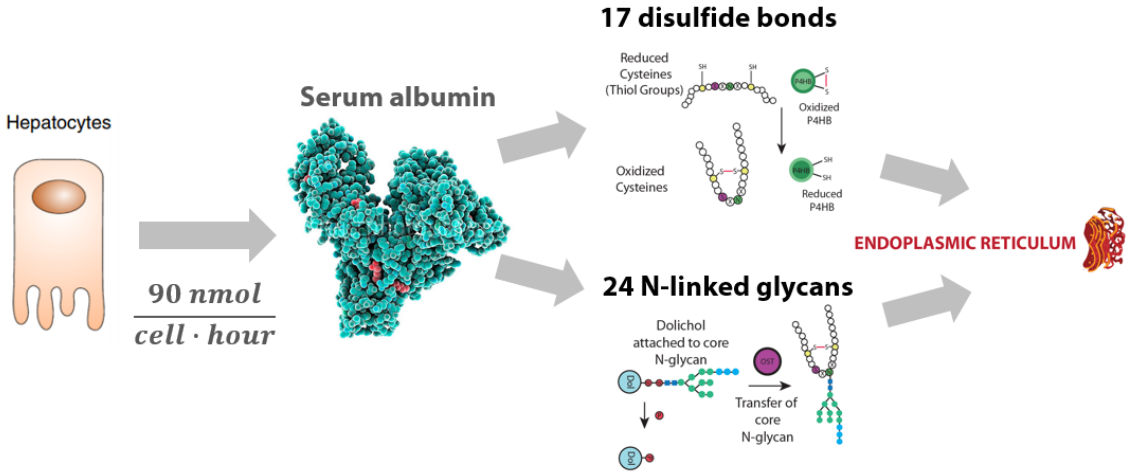

Figure 4.4: Secretory demands in hepatocytes due to albumin production. Albumin, the most secreted protein in hepatocytes, contains 17 disulfide bonds and 24 N-linked glycans in its structure. These post-translational modifications impose a significant demand on the secretory pathway components in the ER of hepatocytes. This may explain the longer mean half-lives of the ER-associated subsystems observed (Figure 4.3).

4.3.3 The ATP turnover for synthesizing cell-specific secretomes does not explain the differences in secretory pathway expression

Given the above results, we wondered if the secretory pathway usage across the three cell types also imposed different energetic demands, which could contribute to explaining the differences in mean module half-lives we observed. That is, we pondered the possibility of energy constraints as a selective pressure for modulating secretory pathway gene expression. We set out to investigate this in the context of ATP cost (as defined in Chapter 3) and ATP turnover for all the secretory proteins produced in the three cell types and that were included in the protein turnover dataset. We found that the total ATP turnover necessary for synthesizing the three cell secretomes is almost the same since the largest difference between the minimum and maximum ATP global turnovers was only 15% (Figure 4.5A). This is interesting considering how, as we showed in Chapter 3, different tissues regulate the gene expression of secreted proteins based on their ATP

cost (Figure 4.5B). Thus, ATP turnover and energetic costs of producing the three different secretomes do not capture the differences observed in secretory pathway module expression levels.

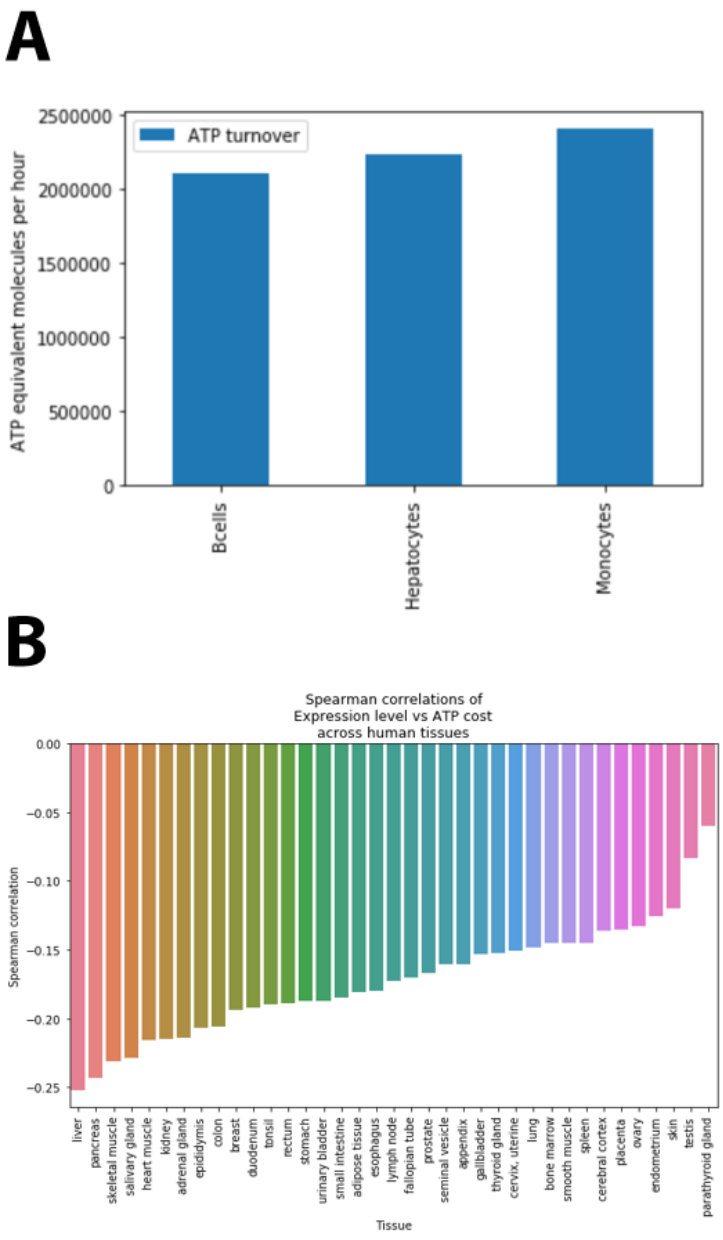

Figure 4.5: ATP turnover in three cell types. (**A**) ATP turnover (in units of number of ATP molecules per hour) that is required to produce the secretome in hepatocytes, monocytes and B cells. (**B**) Spearman correlations between ATP cost and gene expression levels across human tissues. The correlation is stronger in highly secretory tissues such as liver and pancreas.

4.3.3 Secretory pathway usage varies across cell types in a modular manner, resembling mean half-lives of secretory pathway components

Finally, to complete our analysis, we computed the specific secretory pathway usage in a modular way across the three human cell lines. To do so, we used Recon 2.2s to compute the secretory pathway usage required for secreting each one of the secreted proteins and multiplied this usage by the degradation rate of each protein and scaled it by the protein abundance. Surprisingly, we found that modular usage resembles the pattern we observed when looking only at the mean half-lives of secretory pathway components (Figure 4.6A). Therefore, our results suggest that cells regulate the half-lives of secretory pathway components based on how much each module is needed to satisfy the demands imposed by the secretome (Figure 4.6B).

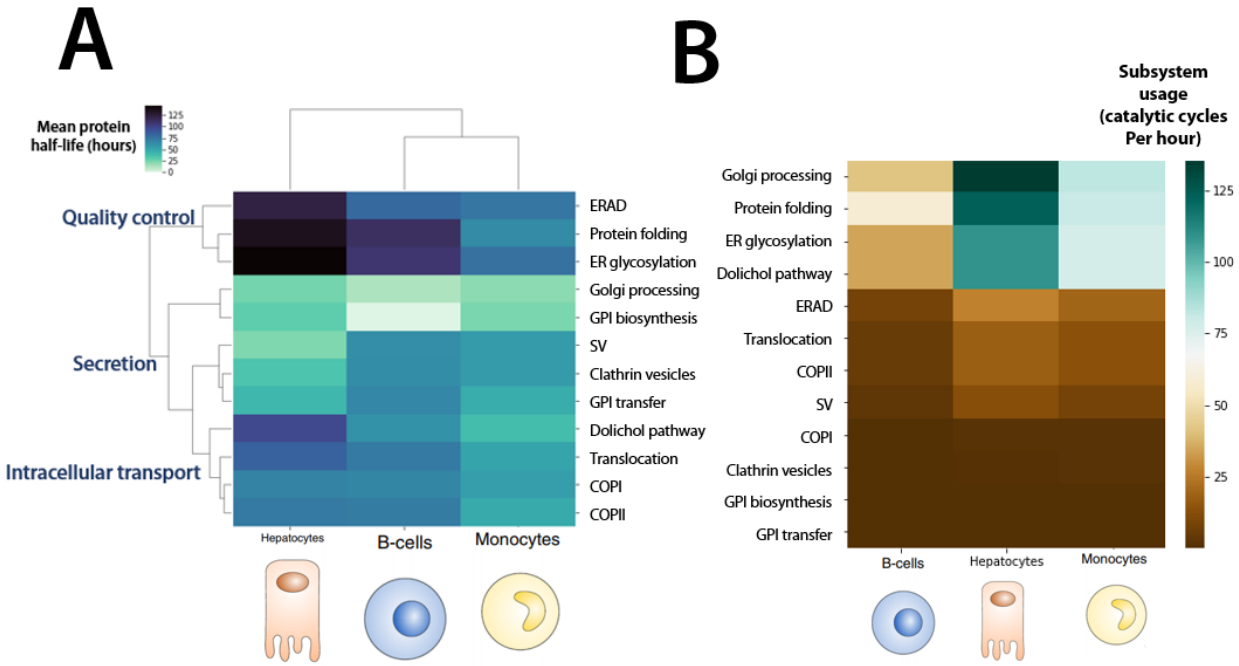

Figure 4.6: Secretory pathway usage correlates with mean half-lives in a modular fashion. (**A**) A clustergram of the mean half-lives computed for each of the secretory pathway modules as in Figure 4.3B. The major function of the three main clusters are indicated to the left of the clustergram. (**B**) Heatmap of the global secretory pathway usage computed in this study for each module. The secretory pathway usage resembles the patterns observed in Figure 4.3B, indicative of protein half-lives being regulated according to its usage.

4.5 Discussion

In this chapter, we have analyzed gene expression levels and protein turnovers in the context of the human secretory pathway reconstruction from Chapter 3. We first demonstrated that significant expression differences in secretory pathway genes are difficult to recover with transcription and translation rates only (i.e. the first two dimensions of Crick space). However, the third dimension in Crick space (i.e. protein degradation rates) enables the differentiation of secretory pathway usage across three highly secretory human cell types. We showed that ATP turnover is also unable to capture significant differences across secretory pathway usage in cell types. However, for the first time we have established a correlation between two fundamental properties of the gene expression of secretory pathway genes at the modular level. The expression for computing secretory pathway module usage resembles the pattern in mean half-lives of across the same modules. That is, cells tend to protect those secretory pathway components from degradation based on how much they use them to process their secretome. This extends upon an observation made recently by Feizi and colleagues [250] on the fine-tuning of secretory pathway components in a tissue-dependent way.

Our results demonstrate that a complete picture of secretory pathway regulation and usage is incomplete without considering protein half-lives. In other words, at least three dimensions of Crick space (transcription, translation, and protein degradation rates) are necessary to capture the differences in secretory pathway usage across cell types. Nevertheless, it is not clear how cells could regulate the degradation rates of secretory pathway components since proteins control protein abundance mainly via transcription and translation [272, 273]. To our knowledge, all previous studies comparing gene expression and protein turnovers across human cells ignored the secretory pathway, highlighting the novelty and the relevance of our results in the context of a

computational reconstruction of human secretion. Future experiments should aim to address the question of what factors regulate protein half-lives across cell types.

In conclusion, highly secretory cells might not be defined by the levels of transcription and or translation of their secretory pathway genes. Instead, as shown in this chapter, the differences seem to span from the ability of a cell to preserve secretory pathway components for longer to make its usage more robust according to the secretome demands. These results could have could have implications in the biotechnology industry for cellular engineering as it challenges the current paradigm for engineering the secretory pathway where only one or two secretory pathway components are overexpressed. Perhaps a better strategy would be to learn from cells like hepatocytes to make secretory pathway components more robust against degradation. That would require a collective effort to experimentally measure the protein turnover rates of important animal cell lines for bioproduction such as HEK293 and CHO cells at different time points in culture.

4.6 Acknowledgements

Chapter 4 in part is from a manuscript that is being prepared for publication *Gutierrez JM, Lewis NE. Mapping protein turnover rates onto the human secretory pathway reconstruction reveals tissue-specific modular control of pathway usage.* The dissertation author is the primary author.

Chapter 5 - Conclusions

5.1 Recapitulation

In this thesis, I have developed computational and systems biotechnology methods to advance our understanding of the animal secretory pathway. A novel workflow for extracting knowledge from legacy data in the CHO cell literature was presented. I demonstrate how the use of such workflow elucidates fundamental features of CHO cell bioprocessing which determine cell growth and protein productivity. Next, a genome-scale computational model of the mammalian secretory pathway was developed. I have validated the predictive power of the model in the context of CHO cell bioprocessing with reported and experimental data. The model is versatile in that it can be constructed for specific recombinant proteins. Thus, it aids in the understanding of cellular physiology at different stages of culture and for different productivity demands. Finally, I demonstrated how an often neglected omic dataset, namely protein turnover rates, reveals differences in secretory pathway usage across human cells that are missed by gene expression data alone. The results point at potential new ways to engineer the secretory pathway of CHO cells for improving cell productivity and overall performance. The work presented here has implications in biotechnology and bioengineering and highlights the importance of using a systems-level view for solving problems in biotherapeutic production.

5.2 The emergence of Systems Biotechnology

Today, 8 out of the 10 top selling drugs are recombinant proteins produced in mammalian cells such as CHO and HEK293 cells [1]. However, the industry still faces three main challenges that relate to quality attributes of the secreted biotherapeutic (Figure 5.1). To effectively tackle these challenges, a paradigm shift towards a systematic view of the producing cell was necessary and Systems Biotechnology has emerged as a new discipline that incorporates models and omics data to study bioprocessing phenomena. Rather than focusing on culture conditions to improve culture performance, Systems Biotechnology sees the producing cell as a system of interconnected modules that give rise to the cell phenotype, which in consequence affect the productivity and growth rate of the cell during culture (Figure 5.2). Overall, the improvements in product yield achieved over the last two decades are the result of a deeper understanding of the effects of both extracellular (temperature, culture media, bioreactor design) and intracellular (gene expression, metabolic fitness) features on culture performance.

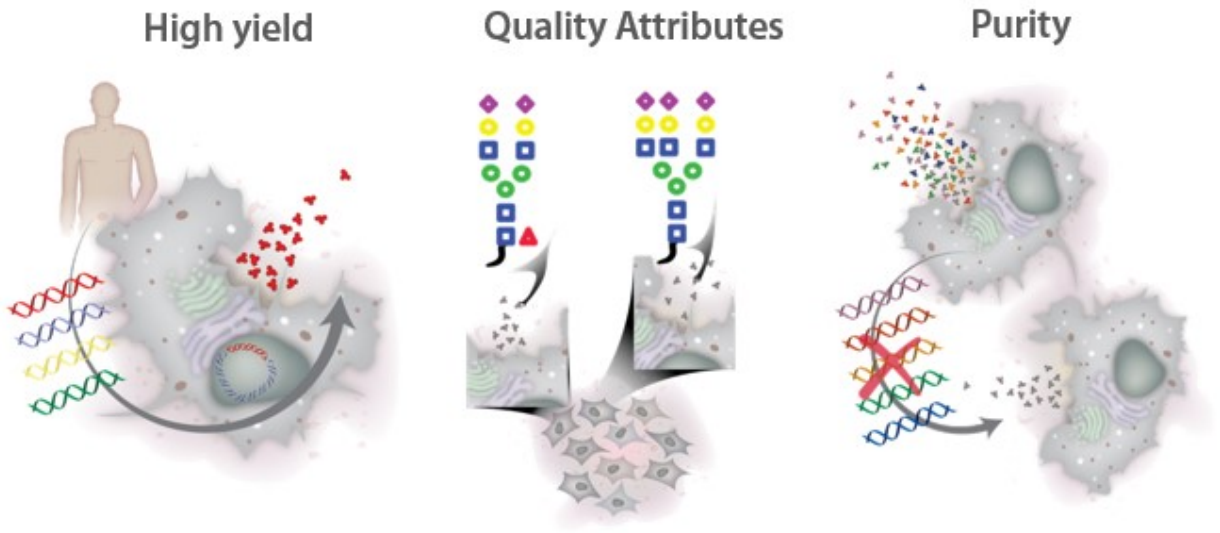

Figure 5.1: Current challenges in the biopharmaceutical industry. To ensure proper quality of secreted therapeutic proteins, the producing cell line should have high yield, confer quality attributes (glycosylation, folding), and avoid secreted too much of the native cells (purity).

In this context, Systems Biotechnology has helped researchers uncover non-obvious relationships between the cellular genotype and the culture conditions via the utilization of high throughput omics data. For example, the identification of gene targets for overexpression were identified using transcriptomic data whereas the identification of metabolites that inhibit cellular growth and protein production was enabled via metabolomics [2]. Therefore, beyond traditional strategies of clonal selection, expression vectors or bioreactor design, Systems Biotechnology aims to utilize elegant modeling and data analysis techniques to gain a mechanistic understanding of cellular bioprocessing. This dissertation has contributed to the field of Systems Biotechnology with three novel approaches: a meta-analysis workflow for analyzing legacy CHO bioprocessing data, a genome-scale reconstruction of the mammalian secretory pathway, and a method for mapping omic data onto secretory pathway models.

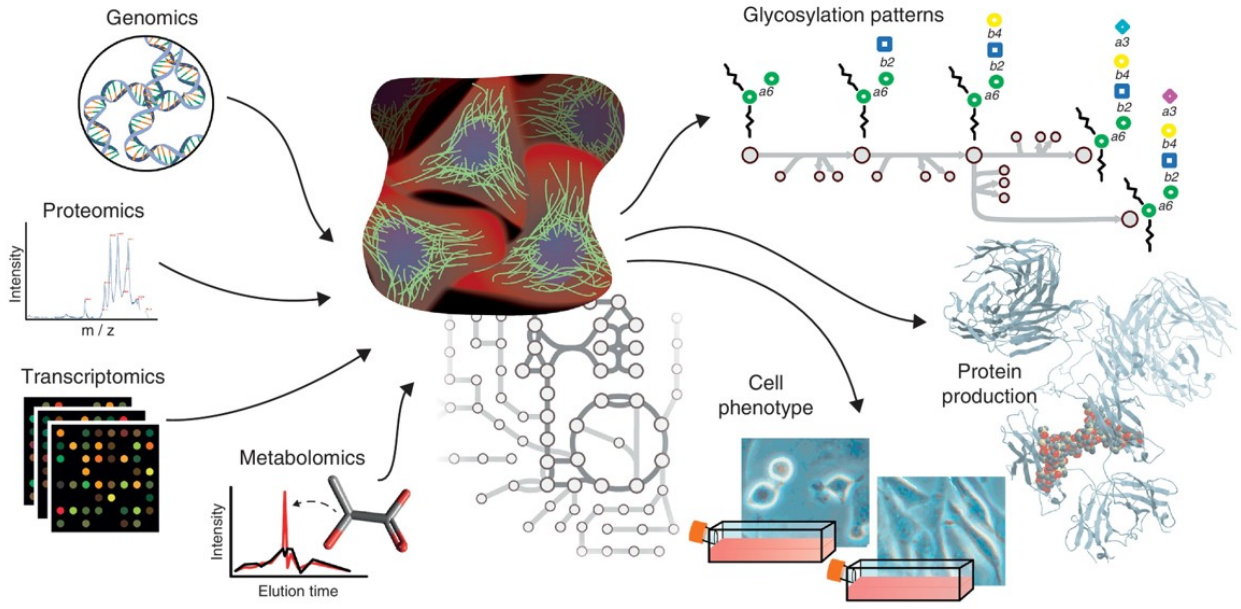

Figure 5.2: The new paradigm of Systems Biotechnology. The producing cell defines the quality of the product, and not just the culture conditions. The producing cell contains intricated networks of interacting components that give raise to the protein production phenotype. Adapted from [5]

5.3 Perspectives for the future

The continued success of Systems Biotechnology will depend on the development and refinement of both modeling and experimental methodologies for analyzing the different parameters that are relevant to bioprocesses. As discussed in Chapter 1, our compilation of the CHO bibliome reveals a growing trend in the number of publications that utilize omic data and modeling techniques for studying CHO bioprocesses [281]. This forecasts a bright future for Systems Biotechnology of animal cell cultures. Furthermore, recent studies have provided large-scale collections of omic data measurements relevant to CHO cells such as glycoproteomics [72], ribosomal profiling [237], and metabolomics [58, 282]. We anticipate that these datasets will motivate the progress in statistical and modeling techniques that will identify novel targets for cellular engineering. Furthermore, the knowledge gained through the application of Systems Biotechnology will make a positive impact in medicine, molecular biology, and synthetic biology.

5.4 Acknowledgements

Chapter 5 in part is a reprint of the material *Gutierrez JM, Lewis NE. Optimizing eukaryotic cell hosts for protein production through systems biotechnology and genome-scale modeling. Biotechnol. J. (2015) 10:939-949*. The dissertation author was the primary author.

Bibliography

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

[2] Kuo CC, Chiang AW, Shamie I, Samoudi M, Gutierrez JM, Lewis NE (2018) The emerging role of systems biology for engineering protein production in CHO cells. Current Opinion in Biotechnology 51:64–69.

[3] Kaufmann H, Mazur X, Fussenegger M, Bailey JE, Zu C (1999) Influence of Low Temperature on Productivity , Proteome and Protein Phosphorylation of CHO Cells.

[4] Chen P, Harcum SW (2005) Effects of amino acid additions on ammonium stressed CHO cells. Journal of Biotechnology 117:277–286.

[5] Kildegaard HF, Baycin-Hizal D, Lewis NE, Betenbaugh MJ (2013) The emerging CHO systems biology era: harnessing the 'omics revolution for biotechnology. Current opinion in biotechnology 24:1102–7.

[6] Provost A, Bastin G (2004) Dynamic metabolic modelling under the balanced growth condition. Journal of Process Control 14:717–728.

[7] Llaneras F, Picó J (2007) A procedure for the estimation over time of metabolic fluxes in scenarios where measurements are uncertain and/or insufficient. BMC bioinformatics 8:421.

[8] Smallbone K, Mendes P (2013) Large-Scale Metabolic Models: From Reconstruction to Differential Equations. Industrial Biotechnology 9:179–184.

[9] Kaas CS, Fan Y, Weilguny D, Kristensen C, Kildegaard HF, Andersen MR (2014) Toward genome-scale models of the Chinese hamster ovary cells: incentives, status and perspectives. Pharmaceutical Bioprocessing 2:437–448.

[10] Hefzi H, Lewis NE (2014) From random mutagenesis to systems biology in metabolic engineering of mammalian cells. Pharmaceutical Bioprocessing 2:355–358.

[11] Bordbar A, Monk JM, King ZA, Palsson BO (2014) Constraint-based models predict metabolic and associated cellular functions. Nature reviews Genetics 15:107–20.

[12] Lewis NE, Nagarajan H, Palsson BO (2012) Constraining the metabolic genotype-phenotype relationship using a phylogeny of in silico methods. Nature reviews Microbiology 10:291–305.

[13] Schellenberger J, Que R, Fleming RMT, Thiele I, Orth JD, Feist AM, Zielinski DC, Bordbar A, Lewis NE, Rahmanian S, Kang J, Hyduke DR, Palsson BØ (2011) Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0.

Nature protocols 6:1290–307.

[14] Orth JD, Thiele I, Palsson BØ (2010) What is flux balance analysis? Nature biotechnology 28:245–8.

[15] Lee SY, Lee D-Y, Kim TY (2005) Systems biotechnology for strain improvement. Trends in Biotechnology 23:349–358.

[16] Kanehisa M, Goto S, Sato Y, Kawashima M, Furumichi M, Tanabe M (2014) Data, information, knowledge and principle: back to metabolism in KEGG. Nucleic acids research 42:D199-205.

[17] Cerami EG, Gross BE, Demir E, Rodchenkov I, Babur O, Anwar N, Schultz N, Bader GD, Sander C (2011) Pathway Commons, a web resource for biological pathway data. Nucleic acids research 39:D685-90.

[18] Croft D, Mundo AF, Haw R, Milacic M, Weiser J, Wu G, Caudy M, Garapati P, Gillespie M, Kamdar MR, Jassal B, Jupe S, Matthews L, May B, Palatnik S, Rothfels K, Shamovsky V, Song H, Williams M, Birney E, Hermjakob H, Stein L, D'Eustachio P (2014) The Reactome pathway knowledgebase. Nucleic acids research 42:D472-7.

[19] Kim TY, Sohn SB, Kim HU, Lee SY (2008) Strategies for systems-level metabolic engineering. Biotechnology journal 3:612–23.

[20] Carinhas N, Oliveira R, Alves PM, Carrondo MJT, Teixeira AP (2012) Systems biotechnology of animal cells: the road to prediction. Trends in Biotechnology 30:377–385.

[21] Price ND, Reed JL, Palsson BØ (2004) Genome-scale models of microbial cells: evaluating the consequences of constraints. Nature reviews Microbiology 2:886–97.

[22] Kim TY, Sohn SB, Kim Y Bin, Kim WJ, Lee SY (2012) Recent advances in reconstruction and applications of genome-scale metabolic models. Current Opinion in Biotechnology 23:617–623.

[23] Mardinoglu A, Nielsen J (2012) Systems medicine and metabolic modelling. Journal of internal medicine 271:142–54.

[24] King ZA, Lloyd CJ, Feist AM, Palsson BO (2015) Next-generation genome-scale models for metabolic engineering. Current Opinion in Biotechnology 35:23–29.

[25] Herrgard MJ, Swainston N, Dobson P, Dunn WB, Arga KY, Arvas M, Blüthgen N, Borger S, Costenoble R, Heinemann M, Hucka M, Le Novère N, Li P, Liebermeister W, Mo ML, Oliveira AP, Petranovic D, Pettifer S, Simeonidis E, Smallbone K, Spasić I, Weichart D, Brent R, Broomhead DS, Westerhoff H V, Kirdar B, Penttilä M, Klipp E, Palsson BØ, Sauer U, Oliver SG, Mendes P, Nielsen J, Kell DB (2008) A consensus yeast metabolic network reconstruction obtained from a community approach to systems biology. Nature biotechnology 26:1155–60.

[26] Chung BK, Selvarasu S, Andrea C, Ryu J, Lee H, Ahn J, Lee H, Lee D-Y (2010) Genome-scale metabolic reconstruction and in silico analysis of methylotrophic yeast Pichia pastoris for strain improvement. Microbial cell factories 9:50.

[27] Thiele I, Swainston N, Fleming RMT, Hoppe A, Sahoo S, Aurich MK, Haraldsdottir H, Mo ML, Rolfsson O, Stobbe MD, Thorleifsson SG, Agren R, Bölling C, Bordel S, Chavali AK, Dobson P, Dunn WB, Endler L, Hala D, Hucka M, Hull D, Jameson D, Jamshidi N, Jonsson JJ, Juty N, Keating S, Nookaew I, Le Novère N, Malys N, Mazein A, Papin JA, Price ND, Selkov E, Sigurdsson MI, Simeonidis E, Sonnenschein N, Smallbone K, Sorokin A, van Beek JHGM, Weichart D, Goryanin I, Nielsen J, Westerhoff H V, Kell DB, Mendes P, Palsson BØ (2013) A community-driven global reconstruction of human metabolism. Nature biotechnology 31:419–25.

[28] Selvarasu S, Karimi IA, Ghim G-H, Lee D-Y (2010) Genome-scale modeling and in silico analysis of mouse cell metabolic network. Molecular bioSystems 6:152–61.

[29] Templeton N, Lewis A, Dorai H, Qian E a., Campbell MP, Smith KD, Lang SE, Betenbaugh MJ, Young JD (2014) The impact of anti-apoptotic gene Bcl-2Δ expression on CHO central metabolism. Metabolic Engineering 1–11.

[30] Naderi S, Meshram M, Wei C, McConkey B, Ingalls B, Budman H, Scharer J (2011) Development of a mathematical model for evaluating the dynamics of normal and apoptotic Chinese hamster ovary cells. Biotechnology progress 27:1197–205.

[31] Oberhardt M a, Palsson BØ, Papin J a (2009) Applications of genome-scale metabolic reconstructions. Molecular systems biology 5:320.

[32] Zhou M, Crawford Y, Ng D, Tung J, Pynn AFJ, Meier A, Yuk IH, Vijayasankaran N, Leach K, Joly J, Snedecor B, Shen A (2011) Decreasing lactate level and increasing antibody production in Chinese Hamster Ovary cells (CHO) by reducing the expression of lactate dehydrogenase and pyruvate dehydrogenase kinases. Journal of biotechnology 153:27–34.

[33] Xie L, Wang DI (1996) Material balance studies on animal cell metabolism using a stoichiometrically based reaction network. Biotechnology and bioengineering 52:579–90.

[34] Palsson BØ (2015) Systems Biology : Constraint-based Reconstruction and Analysis. Cambridge University Press

[35] Schilling CH, Letscher D, Palsson BO (2000) Theory for the systemic definition of metabolic pathways and their use in interpreting metabolic function from a pathway-oriented perspective. Journal of theoretical biology 203:229–48.

[36] Feist AM, Palsson BO (2010) The biomass objective function. Current opinion in microbiology 13:344–9.

[37] Hyduke DR, Lewis NE, Palsson BØ (2013) Analysis of omics data with genome-scale models of metabolism. Molecular bioSystems 9:167–74.

[38] Quek L-E, Dietmair S, Krömer JO, Nielsen LK (2010) Metabolic flux analysis in mammalian cell culture. Metabolic engineering 12:161–71.

[39] Xie L, Wang DI (1994) Stoichiometric analysis of animal cell growth and its application in medium design. Biotechnology and bioengineering 43:1164–74.

[40] Selvarasu S, Kim DY, Karimi IA, Lee D-Y (2010) Combined data preprocessing and multivariate statistical analysis characterizes fed-batch culture of mouse hybridoma cells for rational medium design. Journal of biotechnology 150:94–100.

[41] Broadhurst ER, Butler M (2000) The inhibitory effect of glutamate on the growth of a murine hybridoma is caused by competitive inhibition of the x(-) (C) transport system required for cystine utilization. Cytotechnology 32:31–43.

[42] Ballez JS, Mols J, Burteau C, Agathos SN, Schneider YJ (2004) Plant protein hydrolysates support CHO-320 cells proliferation and recombinant IFN-gamma production in suspension and inside microcarriers in protein-free media. Cytotechnology 44:103–114.

[43] Lee HW, Christie A, Starkey J a., Read EK, Yoon S (2015) Intracellular metabolic flux analysis of CHO cells supplemented with wheat hydrolysates for improved mAb production and cell-growth. Journal of Chemical Technology & Biotechnology 90:291–302.

[44] Zamorano F, Vande Wouwer A, Jungers RM, Bastin G (2013) Dynamic metabolic models of CHO cell cultures through minimal sets of elementary flux modes. Journal of biotechnology 164:409–22.

[45] Ahn WS, Antoniewicz MR (2011) Metabolic flux analysis of CHO cells at growth and non-growth phases using isotopic tracers and mass spectrometry. Metabolic engineering 13:598–609.

[46] Leite TC, Coelho RG, Da Silva D, Coelho WS, Marinho-Carvalho MM, Sola-Penna M (2011) Lactate downregulates the glycolytic enzymes hexokinase and phosphofructokinase in diverse tissues from mice. FEBS letters 585:92–8.

[47] Paredes C, Prats E, Cairó JJ, Azorín F, Cornudella L, Gòdia F (1999) Modification of glucose and glutamine metabolism in hybridoma cells through metabolic engineering. Cytotechnology 30:85–93.

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

[49] Ivarsson M, Noh H, Morbidelli M, Soos M (2015) Insights into pH-induced metabolic switch by flux balance analysis. Biotechnology Progress n/a-n/a.

[50] Mulukutla BC, Gramer M, Hu W-S (2012) On metabolic shift to lactate consumption in fed-batch culture of mammalian cells. Metabolic engineering 14:138–49.

[51] Quek L-E, Nielsen LK (2008) On the reconstruction of the Mus musculus genome-scale metabolic network model. Genome informatics International Conference on Genome Informatics 21:89–100.

[52] Martínez VS, Dietmair S, Quek L-E, Hodson MP, Gray P, Nielsen LK (2013) Flux balance analysis of CHO cells before and after a metabolic switch from lactate production to consumption. Biotechnology and bioengineering 110:660–6.

[53] Selvarasu S, Ho YS, Chong WPK, Wong NSC, Yusufi FNK, Lee YY, Yap MGS, Lee D-Y (2012) Combined in silico modeling and metabolomics analysis to characterize fed-batch CHO cell culture. Biotechnology and Bioengineering 109:1415–1429.

[54] Szappanos B, Kovács K, Szamecz B, Honti F, Costanzo M, Baryshnikova A, Gelius-Dietrich G, Lercher MJ, Jelasity M, Myers CL, Andrews BJ, Boone C, Oliver SG, Pál C, Papp B (2011) An integrated approach to characterize genetic interaction networks in yeast metabolism. Nature genetics 43:656–62.

[55] Carnicer M, Baumann K, Töplitz I, Sánchez-Ferrando F, Mattanovich D, Ferrer P, Albiol J (2009) Macromolecular and elemental composition analysis and extracellular metabolite balances of Pichia pastoris growing at different oxygen levels. Microbial cell factories 8:65.

[56] Raghunathan AU, Pérez-Correa JR, Bieger LT (2003) Data reconciliation and parameter estimation in flux-balance analysis. Biotechnology and bioengineering 84:700–9.

[57] Jiang Z, Sharfstein ST (2008) Sodium butyrate stimulates monoclonal antibody over-expression in CHO cells by improving gene accessibility. Biotechnology and bioengineering 100:189–94.

[58] Carinhas N, Duarte TM, Barreiro LC, Carrondo MJT, Alves PM, Teixeira AP (2013) Metabolic signatures of GS-CHO cell clones associated with butyrate treatment and culture phase transition. Biotechnology and bioengineering 110:3244–57.

[59] Bernal V, Carinhas N, Yokomizo AY, Carrondo MJT, Alves PM (2009) Cell density effect in the baculovirus-insect cells system: a quantitative analysis of energetic metabolism. Biotechnology and bioengineering 104:162–80.

[60] Doverskog M, Ljunggren J, Ohman L, Häggström L (1997) Physiology of cultured animal cells. Journal of biotechnology 59:103–15.

[61] Taticek RA, Shuler ML (1997) Effect of elevated oxygen and glutamine levels on foreign protein production at high cell densities using the insect cell-baculovirus expression system. Biotechnology and bioengineering 54:142–52.

[62] Carinhas N, Bernal V, Monteiro F, Carrondo MJT, Oliveira R, Alves PM (2010) Improving baculovirus production at high cell density through manipulation of energy metabolism. Metabolic engineering 12:39–52.

[63] Gonzalez R, Andrews B a., Molitor J, Asenjo J a. (2003) Metabolic analysis of the synthesis

of high levels of intracellular human SOD in Saccharomyces cerevisiae rhSOD 2060 411 SGA122. Biotechnology and Bioengineering 82:152–169.

[64] Graf A, Dragosits M, Gasser B, Mattanovich D (2009) Yeast systems biotechnology for the production of heterologous proteins. FEMS Yeast Research 9:335–348.

[65] Sohn SB, Graf AB, Kim TY, Gasser B, Maurer M, Ferrer P, Mattanovich D, Lee SY (2010) Genome-scale metabolic model of methylotrophic yeast Pichia pastoris and its use for in silico analysis of heterologous protein production. Biotechnology journal 5:705–15.

[66] Caspeta L, Shoaie S, Agren R, Nookaew I, Nielsen J (2012) Genome-scale metabolic reconstructions of Pichia stipitis and Pichia pastoris and in-silico evaluation of their potentials. BMC Systems Biology 6:24.

[67] Nocon J, Steiger MG, Pfeffer M, Sohn SB, Kim TY, Maurer M, Rußmayer H, Pflügl S, Ask M, Haberhauer-Troyer C, Ortmayr K, Hann S, Koellensperger G, Gasser B, Lee SY, Mattanovich D (2014) Model based engineering of Pichia pastoris central metabolism enhances recombinant protein production. Metabolic Engineering 24:129–138.

[68] Segrè D, Vitkup D, Church GM (2002) Analysis of optimality in natural and perturbed metabolic networks. Proceedings of the National Academy of Sciences of the United States of America 99:15112–7.

[69] Choi HS, Lee SY, Kim TY, Woo HM (2010) In Silico Identification of Gene Amplification Targets for Improvement of Lycopene Production. Applied and Environmental Microbiology 76:3097.

[70] Asadollahi MA, Maury J, Patil KR, Schalk M, Clark A, Nielsen J (2009) Enhancing sesquiterpene production in Saccharomyces cerevisiae through in silico driven metabolic engineering. Metabolic engineering 11:328–34.

[71] Baycin-Hizal D, Tabb DL, Chaerkady R, Chen L, Lewis NE, Nagarajan H, Sarkaria V, Kumar A, Wolozny D, Colao J, Jacobson E, Tian Y, O'Meally RN, Krag SS, Cole RN, Palsson BO, Zhang H, Betenbaugh M (2012) Proteomic analysis of Chinese hamster ovary cells. Journal of proteome research 11:5265–76.

[72] Yang Z, Halim A, Narimatsu Y, Joshi HJ, Steentoft C, Bennett P, Levery SB, Vakhrushev SY, Clausen H (2014) The GalNAc-type O-Glycoproteome of CHO cells.

[73] Vishwanathan N, Yongky A, Johnson KC, Fu H-Y, Jacob NM, Le H, Yusufi FNK, Lee DY, Hu W-S (2014) Global insights into the Chinese hamster and CHO cell transcriptomes. Biotechnol Bioeng. doi: 10.1002/bit.25513

[74] Yim H, Haselbeck R, Niu W, Pujol-Baxley C, Burgard A, Boldt J, Khandurina J, Trawick JD, Osterhout RE, Stephen R, Estadilla J, Teisan S, Schreyer HB, Andrae S, Yang TH, Lee SY, Burk MJ, Van Dien S (2011) Metabolic engineering of Escherichia coli for direct production of 1,4-butanediol. Nature chemical biology 7:445–52.

[75] Lee SK, Chou H, Ham TS, Lee TS, Keasling JD (2008) Metabolic engineering of microorganisms for biofuels production: from bugs to synthetic biology to fuels. Current opinion in biotechnology 19:556–63.

[76] Lee JW, Na D, Park JM, Lee J, Choi S, Lee SY (2012) Systems metabolic engineering of microorganisms for natural and non-natural chemicals. Nature chemical biology 8:536–46.

[77] Thiele I, Jamshidi N, Fleming RMT, Palsson BØ (2009) Genome-scale reconstruction of Escherichia coli's transcriptional and translational machinery: a knowledge base, its mathematical formulation, and its functional characterization. PLoS computational biology 5:e1000312.

[78] Lerman JA, Hyduke DR, Latif H, Portnoy VA, Lewis NE, Orth JD, Schrimpe-rutledge AC, Smith RD, Adkins JN, Zengler K, Palsson BO, Jolla L (2012) In silico method for modeling metabolism and gene product expression at genome scale - Supplementary Information. Nat. Commun.

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

[80] O'Brien EJ, Lerman J a, Chang RL, Hyduke DR, Palsson BØ (2013) Genome-scale models of metabolism and gene expression extend and refine growth phenotype prediction. Molecular systems biology 9:693.

[81] O'Brien EJ, Palsson BO (2015) Computing the functional proteome: recent progress and future prospects for genome-scale models. Current Opinion in Biotechnology 34:125–134.

[82] Oddone GM, Mills DA, Block DE (2009) A dynamic, genome-scale flux model of Lactococcus lactis to increase specific recombinant protein expression. Metabolic engineering 11:367–81.

[83] Margittai E, Sitia R (2011) Oxidative protein folding in the secretory pathway and redox signaling across compartments and cells. Traffic (Copenhagen, Denmark) 12:1–8.

[84] Feizi A, Österlund T, Petranovic D, Bordel S, Nielsen J (2013) Genome-Scale Modeling of the Protein Secretory Machinery in Yeast. PLoS ONE 8:e63284.

[85] Spahn PN, Lewis NE (2014) Systems glycobiology for glycoengineering. Current opinion in biotechnology 30C:218–224.

[86] Neelamegham S, Liu G (2011) Systems glycobiology: biochemical reaction networks regulating glycan structure and function. Glycobiology 21:1541–53.

[87] Tjio JH, Puck TT (1958) Genetics of somatic mammalian cells. II. Chromosomal constitution of cells in tissue culture. The Journal of experimental medicine 108:259–68.

[88] Keen MJ, Rapson NT (1995) Development of a serum-free culture medium for the large

scale production of recombinant protein from a Chinese hamster ovary cell line. Cytotechnology 17:153–63.

[89] Schumpp-Vonach B, Kresbach G, Schlaeger EJ, Steiner B (1995) Stable expression in Chinese hamster ovary cells of a homogeneous recombinant active fragment of human platelet glycoprotein Ib alpha. Cytotechnology 17:133–41.

[90] Kim NS, Lee GM (2002) Inhibition of sodium butyrate-induced apoptosis in recombinant Chinese hamster ovary cells by constitutively expressing antisense RNA of caspase-3. Biotechnology and bioengineering 78:217–28.

[91] Oettl K, Höfler G, Ness GC, Sattler W, Malle E (2003) An apparent decrease in cholesterol biosynthesis in peroxisomal-defective Chinese hamster ovary cells is related to impaired mitochondrial oxidation. Biochemical and biophysical research communications 305:957–63.

[92] Yoon SK, Song JY, Lee GM (2003) Effect of low culture temperature on specific productivity, transcription level, and heterogeneity of erythropoietin in Chinese hamster ovary cells. Biotechnology and bioengineering 82:289–98.

[93] Fox SR, Patel U a, Yap MGS, Wang DIC (2004) Maximizing interferon-gamma production by Chinese hamster ovary cells through temperature shift optimization: experimental and modeling. Biotechnology and bioengineering 85:177–84.

[94] Fox SR, Tan HK, Tan MC, Wong SCNC, Yap MGS, Wang DIC (2005) A detailed understanding of the enhanced hypothermic productivity of interferon-gamma by Chinese-hamster ovary cells. Biotechnology and applied biochemistry 41:255–64.

[95] Kim NY, Kim JH, Kim HJ (2005) Effect of low adapted temperature and medium composition on growth and erythropoietin (EPO) production by Chinese hamster ovary cells. Archives of pharmacal research 28:220–6.

[96] Kim DY, Lee JC, Chang HN, Oh DJ (2006) Development of serum-free media for a recombinant CHO cell line producing recombinant antibody. Enzyme and Microbial Technology 39:426–433.

[97] Yoon SK, Hong JK, Choo SH, Song JY, Park HW, Lee GM (2006) Adaptation of Chinese hamster ovary cells to low culture temperature: cell growth and recombinant protein production. Journal of biotechnology 122:463–72.

[98] Zhang J, Robinson D, Salmon P (2006) A novel function for selenium in biological system: selenite as a highly effective iron carrier for Chinese hamster ovary cell growth and monoclonal antibody production. Biotechnology and bioengineering 95:1188–97.

[99] Choi YS, Lee DY, Kim IY, Kim HJ, Park HW, Choe TB, Kim I-H (2007) Enhancement of erythropoietin production in recombinant Chinese hamster ovary cells by sodium lactate addition. Biotechnology and Bioprocess Engineering 12:60–72.

[100] Zang M, Trautmann H, Gandor C, Messi F, Asselbergs F, Leist C, Fiechter A, Reiser J (1995) Production of recombinant proteins in Chinese hamster ovary cells using a protein-free cell culture medium. Bio/technology (Nature Publishing Company) 13:389–92.

[101] Chun B-H, Kim J-H, Lee H-J, Chung N (2007) Usability of size-excluded fractions of soy protein hydrolysates for growth and viability of Chinese hamster ovary cells in protein-free suspension culture. Bioresource technology 98:1000–5.

[102] Figueroa B, Ailor E, Osborne D, Hardwick JM, Reff M, Betenbaugh MJ (2007) Enhanced cell culture performance using inducible anti-apoptotic genes E1B-19K and Aven in the production of a monoclonal antibody with Chinese hamster ovary cells. Biotechnology and bioengineering 97:877–92.

[103] Fujiwara M, Tsukada R, Tsujinaga Y, Takagi M (2007) Fetal calf serum-free culture of Chinese hamster ovary cells employing fish serum. Applied microbiology and biotechnology 75:983–7.

[104] Hong JK, Kim Y-G, Yoon SK, Lee GM (2007) Down-regulation of cold-inducible RNA-binding protein does not improve hypothermic growth of Chinese hamster ovary cells producing erythropoietin. Metabolic engineering 9:208–16.

[105] Kim SH, Lee GM (2007) Down-regulation of lactate dehydrogenase-A by siRNAs for reduced lactic acid formation of Chinese hamster ovary cells producing thrombopoietin. Applied microbiology and biotechnology 74:152–9.

[106] Liu C-H, Chen L-H (2007) Enhanced recombinant M-CSF production in CHO cells by glycerol addition: model and validation. Cytotechnology 54:89–96.

[107] Liu C-H, Chen L-H (2007) Promotion of recombinant macrophage colony stimulating factor production by dimethyl sulfoxide addition in Chinese hamster ovary cells. Journal of bioscience and bioengineering 103:45–9.

[108] Muller N, Derouazi M, Van Tilborgh F, Wulhfard S, Hacker DL, Jordan M, Wurm FM (2007) Scalable transient gene expression in Chinese hamster ovary cells in instrumented and non-instrumented cultivation systems. Biotechnology letters 29:703–11.

[109] Takuma S, Hirashima C, Piret JM (2007) Dependence on glucose limitation of the pCO2 influences on CHO cell growth, metabolism and IgG production. Biotechnology and bioengineering 97:1479–88.

[110] Yoon SK, Ahn Y-H (2007) Application of sodium propionate to the suspension culture of Chinese hamster ovary cells for enhanced production of follicle-stimulating hormone. Biotechnology and Bioprocess Engineering 12:497–501.

[111] Matsuzawa A, Hattori K, Aoki J, Arai H, Inoue K (1997) Protection against oxidative stress-induced cell death by intracellular platelet-activating factor-acetylhydrolase II. The Journal of biological chemistry 272:32315–20.

[112] Yoon SK, Ahn Y-H, Jeong MH (2007) Effect of culture temperature on follicle-stimulating hormone production by Chinese hamster ovary cells in a perfusion bioreactor. Applied microbiology and biotechnology 76:83–9.

[113] Yoon SK, Jeong MH (2007) Adaptation of Chinese hamster ovary cells to high potassium ion-containing medium for enhancement of follicle-stimulating hormone production. Biotechnology and Bioprocess Engineering 12:399–403.

[114] Ahn WS, Jeon J-J, Jeong Y-R, Lee SJ, Yoon SK (2008) Effect of culture temperature on erythropoietin production and glycosylation in a perfusion culture of recombinant CHO cells. Biotechnology and bioengineering 101:1234–44.

[115] Allen MJ, Boyce JP, Trentalange MT, Treiber DL, Rasmussen B, Tillotson B, Davis R, Reddy P (2008) Identification of novel small molecule enhancers of protein production by cultured mammalian cells. Biotechnology and bioengineering 100:1193–204.

[116] Astley K, Al-Rubeai M (2008) The role of Bcl-2 and its combined effect with p21CIP1 in adaptation of CHO cells to suspension and protein-free culture. Applied microbiology and biotechnology 78:391–9.

[117] Baik JY, Joo EJ, Kim YH, Lee GM (2008) Limitations to the comparative proteomic analysis of thrombopoietin producing Chinese hamster ovary cells treated with sodium butyrate. Journal of biotechnology 133:461–8.

[118] Becker E, Florin L, Pfizenmaier K, Kaufmann H (2008) An XBP-1 dependent bottle-neck in production of IgG subtype antibodies in chemically defined serum-free Chinese hamster ovary (CHO) fed-batch processes. Journal of biotechnology 135:217–23.

[119] Hwang S-J, Kim SH, Kim H-Z, Steinmetz MO, Koh GY, Lee GM (2008) High-level expression and purification of a designed angiopoietin-1 chimeric protein, COMP-Ang1, produced in Chinese hamster ovary cells. The protein journal 27:319–26.

[120] Lee JC, Kim DY, Oh DJ, Chang HN (2008) Two-stage depth filter perfusion culture for recombinant antibody production by recombinant Chinese hamster ovary cell. Biotechnology and Bioprocess Engineering 13:560–565.

[121] Lee JC, Kim DY, Oh DJ, Chang HN (2008) Long-term operation of depth filter perfusion systems (DFPS) for monoclonal antibody production using recombinant CHO cells: Effect of temperature, pH, and dissolved oxygen. Biotechnology and Bioprocess Engineering 13:401–409.

[122] Mastrangelo a J, Hardwick JM, Zou S, Betenbaugh MJ (2000) Part II. Overexpression of bcl-2 family members enhances survival of mammalian cells in response to various culture insults. Biotechnology and bioengineering 67:555–64.

[123] Lee S-B, Park JS, Lee S, Park J, Yu S, Kim H, Kim D, Byun TH, Baek K, Ahn Y-J, Yoon J (2008) Overproduction of recombinant human VEGF (vascular endothelial growth factor) in Chinese hamster ovary cells. Journal of microbiology and biotechnology 18:183–7.

[124] Majors BS, Arden N, Oyler GA, Chiang GG, Pederson NE, Betenbaugh MJ (2008) E2F-1 overexpression increases viable cell density in batch cultures of Chinese hamster ovary cells. Journal of biotechnology 138:103–6.

[125] Majors BS, Betenbaugh MJ, Pederson NE, Chiang GG (2008) Enhancement of transient gene expression and culture viability using Chinese hamster ovary cells overexpressing Bcl-x(L). Biotechnology and bioengineering 101:567–78.

[126] Ohya T, Hayashi T, Kiyama E, Nishii H, Miki H, Kobayashi K, Honda K, Omasa T, Ohtake H (2008) Improved production of recombinant human antithrombin III in Chinese hamster ovary cells by ATF4 overexpression. Biotechnology and bioengineering 100:317–24.

[127] Omasa T, Takami T, Ohya T, Kiyama E, Hayashi T, Nishii H, Miki H, Kobayashi K, Honda K, Ohtake H (2008) Overexpression of GADD34 enhances production of recombinant human antithrombin III in Chinese hamster ovary cells. Journal of bioscience and bioengineering 106:568–73.

[128] Tan HK, Lee MM, Yap MGS, Wang DIC (2008) Overexpression of cold-inducible RNA-binding protein increases interferon-gamma production in Chinese-hamster ovary cells. Biotechnology and applied biochemistry 49:247–57.

[129] Zhao DX, Fike R, Naumovich B, Stramaglia M (2008) Improving Protein Production in CHO Cells.

[130] Florin L, Pegel A, Becker E, Hausser A, Olayioye MA, Kaufmann H (2009) Heterologous expression of the lipid transfer protein CERT increases therapeutic protein productivity of mammalian cells. Journal of biotechnology 141:84–90.

[131] Gigout A, Levasseur S, Girard-Lauriault P-L, Buschmann MD, Wertheimer MR, Jolicoeur M (2009) CHO cells adhering to nitrogen-rich plasma-polymerised ethylene exhibit high production of a specific recombinant protein. Macromolecular bioscience 9:979–88.

[132] Hwang SO, Lee GM (2009) Effect of Akt overexpression on programmed cell death in antibody-producing Chinese hamster ovary cells. Journal of biotechnology 139:89–94.

[133] Nishijima K, Fujiki T, Kojima H, Iijima S (2000) The effects of cell adhesion on the growth and protein productivity of animal cells. Cytotechnology 33:147–55.

[134] Ju HK, Hwang S-J, Jeon C-J, Lee GM, Yoon SK (2009) Use of NaCl prevents aggregation of recombinant COMP-angiopoietin-1 in Chinese hamster ovary cells. Journal of biotechnology 143:145–50.

[135] Kim Y-G, Kim JY, Mohan C, Lee GM (2009) Effect of Bcl-xL overexpression on apoptosis and autophagy in recombinant Chinese hamster ovary cells under nutrient-deprived condition. Biotechnology and bioengineering 103:757–66.

[136] Mohan C, Lee GM (2009) Calnexin overexpression sensitizes recombinant CHO cells to

apoptosis induced by sodium butyrate treatment. Cell stress & chaperones 14:49–60.

[137] Nam JH, Ermonval M, Sharfstein ST (2009) The effects of microcarrier culture on recombinant CHO cells under biphasic hypothermic culture conditions. Cytotechnology 59:81–91.

[138] Peng R-W, Fussenegger M (2009) Molecular engineering of exocytic vesicle traffic enhances the productivity of Chinese hamster ovary cells. Biotechnology and bioengineering 102:1170–81.

[139] Yee JC, Gerdtzen ZP, Hu W-S (2009) Comparative transcriptome analysis to unveil genes affecting recombinant protein productivity in mammalian cells. Biotechnology and bioengineering 102:246–63.

[140] Chong WPK, Reddy SG, Yusufi FNK, Lee D-Y, Wong NSC, Heng CK, Yap MGS, Ho YS (2010) Metabolomics-driven approach for the improvement of Chinese hamster ovary cell growth: overexpression of malate dehydrogenase II. Journal of biotechnology 147:116–21.

[141] Cost GJ, Freyvert Y, Vafiadis A, Santiago Y, Miller JC, Rebar E, Collingwood TN, Snowden A, Gregory PD (2010) BAK and BAX deletion using zinc-finger nucleases yields apoptosis-resistant CHO cells. Biotechnology and bioengineering 105:330–40.

[142] Fan L, Zhao L, Ye Z, Sun Y, Kou T, Zhou Y, Tan W-S (2010) Effect of culture temperature on TNFR-Fc productivity in recombinant glutamine synthetase-chinese hamster ovary cells. Biotechnology letters 32:1239–44.

[143] Han YK, Kim Y-G, Kim JY, Lee GM (2010) Hyperosmotic stress induces autophagy and apoptosis in recombinant Chinese hamster ovary cell culture. Biotechnology and bioengineering 105:1187–92.

[144] Takagi M, Hayashi H, Yoshida T (2000) The effect of osmolarity on metabolism and morphology in adhesion and suspension chinese hamster ovary cells producing tissue plasminogen activator. Cytotechnology 32:171–9.

[145] Kaneko Y, Sato R, Aoyagi H (2010) Evaluation of Chinese hamster ovary cell stability during repeated batch culture for large-scale antibody production. Journal of bioscience and bioengineering 109:274–80.

[146] Kantardjieff A, Jacob NM, Yee JC, Epstein E, Kok Y-J, Philp R, Betenbaugh M, Hu W-S (2010) Transcriptome and proteome analysis of Chinese hamster ovary cells under low temperature and butyrate treatment. Journal of biotechnology 145:143–59.

[147] Malphettes L, Freyvert Y, Chang J, Liu P-Q, Chan E, Miller JC, Zhou Z, Nguyen T, Tsai C, Snowden AW, Collingwood TN, Gregory PD, Cost GJ (2010) Highly efficient deletion of FUT8 in CHO cell lines using zinc-finger nucleases yields cells that produce completely nonfucosylated antibodies. Biotechnology and bioengineering 106:774–83.

[148] Rodrigues Goulart H, Arthuso F dos S, Capone MVN, de Oliveira TL, Bartolini P, Soares

CRJ (2010) Enhancement of human prolactin synthesis by sodium butyrate addition to serum-free CHO cell culture. Journal of biomedicine & biotechnology 2010:405872.

[149] Altamirano C, Paredes C, Cairó JJ, Gòdia F (2000) Improvement of CHO cell culture medium formulation: simultaneous substitution of glucose and glutamine. Biotechnology progress 16:69–75.

[150] Kim BJ, Chang HN, Oh DJ (2007) Application of a cell-once-through perfusion strategy for production of recombinant antibody from rCHO cells in a Centritech Lab II centrifuge system. Biotechnology progress 23:1186–97.

[151] Kim Y-G, Lee GM (2009) Bcl-xL overexpression does not enhance specific erythropoietin productivity of recombinant CHO cells grown at 33 degrees C and 37 degrees C. Biotechnology progress 25:252–6.

[152] Majors BS, Betenbaugh MJ, Pederson NE, Chiang GG (2009) Mcl-1 overexpression leads to higher viabilities and increased production of humanized monoclonal antibody in Chinese hamster ovary cells. Biotechnology progress 25:1161–8.

[153] Prentice HL, Ehrenfels BN, Sisk WP (2007) Improving performance of mammalian cells in fed-batch processes through “bioreactor evolution”. Biotechnology progress 23:458–64.

[154] Shen D, Kiehl TR, Khattak SF, Li ZJ, He A, Kayne PS, Patel V, Neuhaus IM, Sharfstein ST (2010) Transcriptomic responses to sodium chloride-induced osmotic stress: a study of industrial fed-batch CHO cell cultures. Biotechnology progress 26:1104–15.

[155] Yamamoto S, Matsuda H, Takahashi T, Xing XH, Tanji Y, Unno H (2000) Aggregate formation of rCHO cells and its maintenance in repeated batch culture in the absence of cell adhesion materials. Journal of bioscience and bioengineering 89:534–8.

[156] Sung YH, Lee JS, Park SH, Koo J, Lee GM (2007) Influence of co-down-regulation of caspase-3 and caspase-7 by siRNAs on sodium butyrate-induced apoptotic cell death of Chinese hamster ovary cells producing thrombopoietin. Metabolic engineering 9:452–64.

[157] Sunley K, Tharmalingam T, Butler M (2008) CHO cells adapted to hypothermic growth produce high yields of recombinant beta-interferon. Biotechnology progress 24:898–906.

[158] Tharmalingam T, Sunley K, Butler M (2008) High yields of monomeric recombinant beta-interferon from macroporous microcarrier cultures under hypothermic conditions. Biotechnology progress 24:832–8.

[159] Wulhfard S, Tissot S, Bouchet S, Cevey J, De Jesus M, Hacker DL, Wurm FM (2008) Mild hypothermia improves transient gene expression yields several fold in Chinese hamster ovary cells. Biotechnology progress 24:458–65.

[160] Kim MS, Kim NS, Sung YH, Lee GM (2002) Biphasic culture strategy based on hyperosmotic pressure for improved humanized antibody production in Chinese hamster ovary cell culture. In vitro cellular & developmental biology Animal 38:314–9.

[161] Kim NS, Lee GM (2002) Response of recombinant Chinese hamster ovary cells to hyperosmotic pressure: effect of Bcl-2 overexpression. Journal of biotechnology 95:237–48.

[162] Rohatgi A (2014) WebPlotDigitizer.

[163] Savitzky A, Golay MJE (1964) Smoothing and Differentiation of Data by Simplified Least Squares Procedures. ANALYTICAL CHEMISTRY 36:1627–1639.

[164] Rousseeuw PJ, Hubert M (2011) Robust statistics for outlier detection. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery 1:73–79.

[165] Kruskal WH, Wallis WA (1952) Use of Ranks in One-Criterion Variance Analysis. Journal of the American Statistical Association 47:583–621.

[166] Dunn OJ (1964) Multiple Comparisons Using Rank Sums. Technometrics 6:241–252.

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

[168] Wold S, Sjöström M, Eriksson L (2001) PLS-regression: a basic tool of chemometrics. Chemometrics and Intelligent Laboratory Systems 58:109–130.

[169] Kim DY, Chaudhry MA, Kennard ML, Jardon MA, Braasch K, Dionne B, Butler M, Piret JM (2013) Fed-batch CHO cell t-PA production and feed glutamine replacement to reduce ammonia production. Biotechnology Progress 29:165–175.

[170] Zalai D, Golabgir A, Wechselberger P, Putics A, Herwig C (2015) Advanced Development Strategies for Biopharmaceutical Cell Culture Processes. Curr Pharm Biotechnol. doi: 10.2174/1389201016666150724100450

[171] Edros R, McDonnell S, Al-Rubeai M (2014) The relationship between mTOR signalling pathway and recombinant antibody productivity in CHO cell lines. BMC biotechnology 14:15.

[172] Chusainow J, Yang YS, Yeo JH, Toh PC, Asvadi P, Wong NS, Yap MG (2009) A study of monoclonal antibody-producing CHO cell lines: What makes a stable high producer? Biotechnology and bioengineering 102:1182–1196.

[173] Martínez VS, Dietmair S, Quek L-E, Hodson MP, Gray P, Nielsen LK (2013) Flux balance analysis of CHO cells before and after a metabolic switch from lactate production to consumption. Biotechnology and bioengineering 110:660–6.

[174] Savinell JM, Lee GM, Palsson BO (1989) On the orders of magnitude of epigenic dynamics and monoclonal antibody production. Bioprocess Engineering 4:231–234.

[175] Gu MB, Todd P, Kompala DS (1995) Metabolic burden in recombinant CHO cells: effect ofdhfr gene amplification andlacZ expression. Cytotechnology 18:159–166.

[176] Brazma A, Hingamp P, Quackenbush J, Sherlock G, Spellman P, Stoeckert C, Aach J, Ansorge W, Ball CA, Causton HC, Gaasterland T, Glenisson P, Holstege FC, Kim IF, Markowitz V, Matese JC, Parkinson H, Robinson A, Sarkans U, Schulze-Kremer S, Stewart J, Taylor R, Vilo J, Vingron M (2001) Minimum information about a microarray experiment (MIAME)-toward standards for microarray data. Nature genetics 29:365–71.

[177] Waltemath D, Adams R, Beard DA, Bergmann FT, Bhalla US, Britten R, Chelliah V, Cooling MT, Cooper J, Crampin EJ, Garny A, Hoops S, Hucka M, Hunter P, Klipp E, Laibe C, Miller AK, Moraru I, Nickerson D, Nielsen P, Nikolski M, Sahle S, Sauro HM, Schmidt H, Snoep JL, Tolle D, Wolkenhauer O, Le Novère N (2011) Minimum Information About a Simulation Experiment (MIASE). PLoS computational biology 7:e1001122.

[178] Le Novère N, Finney A, Hucka M, Bhalla US, Campagne F, Collado-Vides J, Crampin EJ, Halstead M, Klipp E, Mendes P, Nielsen P, Sauro H, Shapiro B, Snoep JL, Spence HD, Wanner BL (2005) Minimum information requested in the annotation of biochemical models (MIRIAM). Nature biotechnology 23:1509–15.

[179] Zhu MM, Goyal A, Rank DL, Gupta SK, Vanden Boom T, Lee SS (2005) Effects of elevated pCO2 and osmolality on growth of CHO cells and production of antibody-fusion protein B1: a case study. Biotechnology progress 21:70–7.

[180] (2013) Raising standards. Nature Biotechnology 31:366–366.

[181] Gutierrez JM, Lewis NE (2015) Optimizing eukaryotic cell hosts for protein production through systems biotechnology and genome-scale modeling. Biotechnology journal 10:939–49.

[182] Lee JS, Grav LM, Lewis NE, Kildegaard HF (2015) CRISPR/Cas9-mediated genome engineering of CHO cell factories: Application and perspectives. Biotechnology journal 10:979–94.

[183] Uhlén M, Fagerberg L, Hallström BM, Lindskog C, Oksvold P, Mardinoglu A, Sivertsson Å, Kampf C, Sjöstedt E, Asplund A, Olsson I, Edlund K, Lundberg E, Navani S, Szigyarto CA-K, Odeberg J, Djureinovic D, Takanen JO, Hober S, Alm T, Edqvist P-H, Berling H, Tegel H, Mulder J, Rockberg J, Nilsson P, Schwenk JM, Hamsten M, von Feilitzen K, Forsberg M, Persson L, Johansson F, Zwahlen M, von Heijne G, Nielsen J, Pontén F (2015) Tissue-based map of the human proteome. Science (80-.). 347:

[184] Walsh G (2014) Biopharmaceutical benchmarks 2014. Nature Biotechnology 32:992–1000.

[185] Liu L, Feizi A, Österlund T, Hjort C, Nielsen J (2014) Genome-scale analysis of the high-efficient protein secretion system of Aspergillus oryzae. BMC Systems Biology 8:73.

[186] Irani ZA, Kerkhoven EJ, Shojaosadati SA, Nielsen J (2016) Genome-scale metabolic model of Pichia pastoris with native and humanized glycosylation of recombinant proteins. Biotechnology and Bioengineering 113:961–969.

[187] Schekman R, Novick P (2004) 23 genes, 23 years later. Cell 116:S13–5, 1 p following S19.

[188] Delic M, Valli M, Graf AB, Pfeffer M, Mattanovich D, Gasser B (2013) The secretory pathway: exploring yeast diversity. FEMS Microbiology Reviews 37:872–914.

[189] Mandon EC, Trueman SF, Gilmore R (2013) Protein translocation across the rough endoplasmic reticulum. Cold Spring Harb Perspect Biol. doi: 10.1101/cshperspect.a013342

[190] Shao S, Hegde RS (2011) Membrane Protein Insertion at the Endoplasmic Reticulum. Annual Review of Cell and Developmental Biology 27:25–56.

[191] Borgese N, Fasana E (2011) Targeting pathways of C-tail-anchored proteins. Biochimica et Biophysica Acta (BBA) - Biomembranes 1808:937–946.

[192] Gidalevitz T, Stevens F, Argon Y (2013) Orchestration of secretory protein folding by ER chaperones. Biochimica et biophysica acta 1833:2410–24.

[193] Araki K, Nagata K (2011) Protein folding and quality control in the ER. Cold Spring Harbor perspectives in biology 3:a007526.

[194] Braakman I, Hebert DN (2013) Protein folding in the endoplasmic reticulum. Cold Spring Harbor perspectives in biology 5:a013201.

[195] Sousa M, Parodi AJ (1995) The molecular basis for the recognition of misfolded glycoproteins by the UDP-Glc:glycoprotein glucosyltransferase. The EMBO journal 14:4196–203.

[196] Caramelo JJ, Parodi AJ (2008) Getting In and Out from Calnexin/Calreticulin Cycles. Journal of Biological Chemistry 283:10221–10225.

[197] Xu C, Ng DTW (2015) Glycosylation-directed quality control of protein folding. Nature Reviews Molecular Cell Biology 16:742–752.

[198] Xu C, Wang S, Thibault G, Ng DTW (2013) Futile Protein Folding Cycles in the ER Are Terminated by the Unfolded Protein O-Mannosylation Pathway. Science 340:978–981.

[199] Olzmann JA, Kopito RR, Christianson JC (2013) The mammalian endoplasmic reticulum-associated degradation system. Cold Spring Harb Perspect Biol. doi: 10.1101/cshperspect.a013185

[200] Ikezawa H (2002) Glycosylphosphatidylinositol (GPI)-anchored proteins. Biological & pharmaceutical bulletin 25:409–17.

[201] Fujita M, Kinoshita T (2010) Structural remodeling of GPI anchors during biosynthesis and after attachment to proteins. FEBS Letters 584:1670–1677.

[202] Ron D, Harding HP (2012) Protein-folding homeostasis in the endoplasmic reticulum and

nutritional regulation. Cold Spring Harb Perspect Biol. doi: 10.1101/cshperspect.a013177

[203] Ron E, Shenkman M, Groisman B, Izenshtein Y, Leitman J, Lederkremer GZ (2011) Bypass of glycan-dependent glycoprotein delivery to ERAD by up-regulated EDEM1. Molecular Biology of the Cell 22:3945–3954.

[204] Mohorko E, Glockshuber R, Aebi M (2011) Oligosaccharyltransferase: the central enzyme of N-linked protein glycosylation. Journal of Inherited Metabolic Disease 34:869–878.

[205] Aebi M, Bernasconi R, Clerc S, Molinari M (2010) N-glycan structures: recognition and processing in the ER. Trends in biochemical sciences 35:74–82.

[206] Aebi M (2013) N-linked protein glycosylation in the ER. Biochimica et Biophysica Acta (BBA) - Molecular Cell Research 1833:2430–2437.

[207] Lord C, Ferro-Novick S, Miller EA (2013) The Highly Conserved COPII Coat Complex Sorts Cargo from the Endoplasmic Reticulum and Targets It to the Golgi. Cold Spring Harbor Perspectives in Biology 5:a013367–a013367.

[208] Dancourt J, Barlowe C (2010) Protein Sorting Receptors in the Early Secretory Pathway. Annual Review of Biochemistry 79:777–802.

[209] Bagola K, Mehnert M, Jarosch E, Sommer T (2011) Protein dislocation from the ER. Biochimica et Biophysica Acta (BBA) - Biomembranes 1808:925–936.

[210] Nakatsukasa K, Brodsky JL, Kamura T (2013) A stalled retrotranslocation complex reveals physical linkage between substrate recognition and proteasomal degradation during ER-associated degradation. Molecular biology of the cell 24:1765–75, S1-8.

[211] Nakatsukasa K, Kamura T, Brodsky JL (2014) Recent technical developments in the study of ER-associated degradation. Current opinion in cell biology 29:82–91.

[212] Hebert DN, Bernasconi R, Molinari M (2010) ERAD substrates: Which way out? Seminars in Cell & Developmental Biology 21:526–532.

[213] Croset A, Delafosse L, Gaudry J-P, Arod C, Glez L, Losberger C, Begue D, Krstanovic A, Robert F, Vilbois F, Chevalet L, Antonsson B (2012) Differences in the glycosylation of recombinant proteins expressed in HEK and CHO cells. Journal of Biotechnology 161:336–348.

[214] Moremen KW, Tiemeyer M, Nairn A V. (2012) Vertebrate protein glycosylation: diversity, synthesis and function. Nature Reviews Molecular Cell Biology 13:448–462.

[215] Stanley P (2011) Golgi glycosylation. Cold Spring Harb Perspect Biol. doi: 10.1101/cshperspect.a005199

[216] Bitsikas V, Corrêa IR, Nichols BJ (2014) Clathrin-independent pathways do not contribute significantly to endocytic flux. eLife 3:e03970.

[217] Banfield DK (2011) Mechanisms of Protein Retention in the Golgi. Cold Spring Harbor Perspectives in Biology 3:a005264–a005264.

[218] Gannon J, Bergeron JJM, Nilsson T (2011) Golgi and Related Vesicle Proteomics: Simplify to Identify. Cold Spring Harbor Perspectives in Biology 3:a005421–a005421.

[219] Ispolatov I, Müsch A (2013) A model for the self-organization of vesicular flux and protein distributions in the Golgi apparatus. PLoS computational biology 9:e1003125.

[220] Johnson N, Powis K, High S (2013) Post-translational translocation into the endoplasmic reticulum. Biochimica et Biophysica Acta (BBA) - Molecular Cell Research 1833:2403–2409.

[221] Lakkaraju AKK, Thankappan R, Mary C, Garrison JL, Taunton J, Strub K (2012) Efficient secretion of small proteins in mammalian cells relies on Sec62-dependent posttranslational translocation. Molecular Biology of the Cell 23:2712–2722.

[222] Vilardi F, Lorenz H, Dobberstein B (2011) WRB is the receptor for TRC40/Asna1-mediated insertion of tail-anchored proteins into the ER membrane. Journal of Cell Science 124:1301–1307.

[223] Stefanovic S, Hegde RS (2007) Identification of a Targeting Factor for Posttranslational Membrane Protein Insertion into the ER. Cell 128:1147–1159.

[224] Araki K, Inaba K (2012) Structure, mechanism, and evolution of Ero1 family enzymes. Antioxidants & redox signaling 16:790–9.

[225] Hiller K, Grote A, Scheer M, Munch R, Jahn D (2004) PrediSi: prediction of signal peptides and their cleavage positions. Nucleic Acids Research 32:W375–W379.

[226] Fankhauser N, Maser P (2005) Identification of GPI anchor attachment signals by a Kohonen self-organizing map. Bioinformatics 21:1846–1852.

[227] Yang Z, Halim A, Narimatsu Y, Jitendra Joshi H, Steentoft C, Schjoldager KT-BG, Alder Schulz M, Sealover NR, Kayser KJ, Paul Bennett E, Levery SB, Vakhrushev SY, Clausen H (2014) The GalNAc-type O-Glycoproteome of CHO cells characterized by the SimpleCell strategy. Molecular & cellular proteomics : MCP 13:3224–35.

[228] Baycin-Hizal D, Tabb DL, Chaerkady R, Chen L, Lewis NE, Nagarajan H, Sarkaria V, Kumar A, Wolozny D, Colao J, Jacobson E, Tian Y, O'Meally RN, Krag SS, Cole RN, Palsson BO, Zhang H, Betenbaugh M (2012) Proteomic Analysis of Chinese Hamster Ovary Cells. Journal of Proteome Research 11:5265–5276.

[229] Orth JD, Thiele I, Palsson BØ (2010) What is flux balance analysis? Nature biotechnology 28:245–8.

[230] Schellenberger J, Que R, Fleming RMT, Thiele I, Orth JD, Feist AM, Zielinski DC, Bordbar A, Lewis NE, Rahmanian S, Kang J, Hyduke DR, Palsson BØ (2011) Quantitative

prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0. Nature Protocols 6:1290–1307.

[231] Grainger RK, James DC (2013) CHO cell line specific prediction and control of recombinant monoclonal antibody *N* -glycosylation. Biotechnology and Bioengineering 110:2970–2983.

[232] Araki K, Nagata K (2011) Protein Folding and Quality Control in the ER. Cold Spring Harbor Perspectives in Biology 3:a007526–a007526.

[233] Borner GHH, Antrobus R, Hirst J, Bhumbra GS, Kozik P, Jackson LP, Sahlender DA, Robinson MS (2012) Multivariate proteomic profiling identifies novel accessory proteins of coated vesicles. The Journal of Cell Biology 197:141–160.

[234] Cheng Y, Boll W, Kirchhausen T, Harrison SC, Walz T (2007) Cryo-electron Tomography of Clathrin-coated Vesicles: Structural Implications for Coat Assembly. Journal of Molecular Biology 365:892–899.

[235] Takamori S, Holt M, Stenius K, Lemke EA, Grønborg M, Riedel D, Urlaub H, Schenck S, Brügger B, Ringler P, Müller SA, Rammner B, Gräter F, Hub JS, De Groot BL, Mieskes G, Moriyama Y, Klingauf J, Grubmüller H, Heuser J, Wieland F, Jahn R (2006) Molecular Anatomy of a Trafficking Organelle. Cell 127:831–846.

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

[237] Kallehauge TB, Li S, Pedersen LE, Ha TK, Ley D, Andersen MR, Kildegaard HF, Lee GM, Lewis NE (2017) Ribosome profiling-guided depletion of an mRNA increases cell growth rate and protein secretion. Scientific Reports 7:40388.

[238] Wang Y, Eddy JA, Price ND (2012) Reconstruction of genome-scale metabolic models for 126 human tissues using mCADRE. BMC Systems Biology 6:153.

[239] Opdam S, Richelle A, Kellman B, Li S, Zielinski DC, Lewis NE, Raza A, Mukherjee S, Manley JL, Asplund A, al. et (2017) A Systematic Evaluation of Methods for Tailoring Genome-Scale Metabolic Models. Cell Systems 4:318–329.e6.

[240] González-Leal IJ, Carrillo-Cocom LM, Ramírez-Medrano A, López-Pacheco F, Bulnes-Abundis D, Webb-Vargas Y, Alvarez MM (2011) Use of a Plackett-Burman statistical design to determine the effect of selected amino acids on monoclonal antibody production in CHO cells. Biotechnology Progress 27:1709–1717.

[241] Hefzi H, Ang KS, Hanscho M, Bordbar A, Ruckerbauer D, Lakshmanan M, Orellana CA, Baycin-Hizal D, Huang Y, Ley D, Martinez VS, Kyriakopoulos S, Jiménez NE, Zielinski DC, Quek L-E, Wulff T, Arnsdorf J, Li S, Lee JS, Paglia G, Loira N, Spahn PN, Pedersen LE, Gutierrez JM, King ZA, Lund AM, Nagarajan H, Thomas A, Abdel-Haleem AM, Zanghellini J, Kildegaard HF, Voldborg BG, Gerdtzen ZP, Betenbaugh MJ, Palsson BO,

Andersen MR, Nielsen LK, Borth N, Lee D-Y, Lewis NE (2016) A Consensus Genome-scale Reconstruction of Chinese Hamster Ovary Cell Metabolism. Cell Systems 3:434–443.e8.

[242] Golabgir A, Gutierrez JM, Hefzi H, Li S, Palsson BO, Herwig C, Lewis NE (2016) Quantitative feature extraction from the Chinese hamster ovary bioprocess bibliome using a novel meta-analysis workflow. Biotechnology Advances 34:621–633.

[243] Swainston N, Smallbone K, Hefzi H, Dobson PD, Brewer J, Hanscho M, Zielinski DC, Ang KS, Gardiner NJ, Gutierrez JM, Kyriakopoulos S, Lakshmanan M, Li S, Liu JK, Martínez VS, Orellana CA, Quek L-E, Thomas A, Zanghellini J, Borth N, Lee D-Y, Nielsen LK, Kell DB, Lewis NE, Mendes P (2016) Recon 2.2: from reconstruction to model of human metabolism. Metabolomics 12:109.

[244] Sigurdsson MI, Jamshidi N, Steingrimsson E, Thiele I, Palsson BØ (2010) A detailed genome-wide reconstruction of mouse metabolism based on human Recon 1. BMC Systems Biology 4:140.

[245] Templeton N, Dean J, Reddy P, Young JD (2013) Peak Antibody Production is Associated With Increased Oxidative Metabolism in an Industrially Relevant Fed-Batch CHO Cell Culture. Biotechnol Bioeng 110:2013–2024.

[246] Kim SJ, Kim NS, Ryu CJ, Hong HJ, Lee GM (1998) Characterization of chimeric antibody producing CHO cells in the course of dihydrofolate reductase-mediated gene amplification and their stability in the absence of selective pressure. Biotechnology and bioengineering 58:73–84.

[247] Kaufman RJ, Wasley LC, Davies M V, Wise RJ, Israel DI, Dorner AJ (1989) Effect of von Willebrand Factor Coexpression on the Synthesis and Secretion of Factor VIII in Chinese Hamster Ovary Cells. MOLECULAR AND CELLULAR BIOLOGY 9:1233–1242.

[248] Pipe SW, Morris JA, Shah J, Kaufman RJ (1998) Differential interaction of coagulation factor VIII and factor V with protein chaperones calnexin and calreticulin. The Journal of biological chemistry 273:8537–44.

[249] van Wijk XM, Döhrmann S, Hallström BM, Li S, Voldborg BG, Meng BX, McKee KK, van Kuppevelt TH, Yurchenco PD, Palsson BO, Lewis NE, Nizet V, Esko JD (2017) Whole-Genome Sequencing of Invasion-Resistant Cells Identifies Laminin α2 as a Host Factor for Bacterial Invasion. mBio 8:e02128-16.

[250] Feizi A, Gatto F, Uhlen M, Nielsen J (2017) Human protein secretory pathway genes are expressed in a tissue-specific pattern to match processing demands of the secretome. npj Systems Biology and Applications 3:22.

[251] Kim JY, Kim Y-G, Lee GM (2012) CHO cells in biotechnology for production of recombinant proteins: current state and further potential. Applied Microbiology and Biotechnology 93:917–930.

[252] Wurm FM (2004) Production of recombinant protein therapeutics in cultivated mammalian cells. Nature Biotechnology 22:1393–1398.

[253] Gu MB, Todd P, Kompala DS (1996) Metabolic burden in recombinant CHO cells: effect of dhfr gene amplification and lacZ expression. Cytotechnology 18:159–166.

[254] Gu MB, Todd P, Kompala DS (1994) Analysis of foreign protein overproduction in recombinant CHO cells. Effect of growth kinetics and cell cycle traverse. Annals of the New York Academy of Sciences 721:194–207.

[255] Hansen HG, Pristovšek N, Kildegaard HF, Lee GM (2017) Improving the secretory capacity of Chinese hamster ovary cells by ectopic expression of effector genes: Lessons learned and future directions. Biotechnology Advances 35:64–76.

[256] Delic M, Göngrich R, Mattanovich D, Gasser B (2014) Engineering of Protein Folding and Secretion—Strategies to Overcome Bottlenecks for Efficient Production of Recombinant Proteins. Antioxidants & Redox Signaling 21:414–437.

[257] Le Fourn V, Girod P-A, Buceta M, Regamey A, Mermod N (2014) CHO cell engineering to prevent polypeptide aggregation and improve therapeutic protein secretion. Metabolic Engineering 21:91–102.

[258] Borth N, Mattanovich D, Kunert R, Katinger H (2008) Effect of Increased Expression of Protein Disulfide Isomerase and Heavy Chain Binding Protein on Antibody Secretion in a Recombinant CHO Cell Line. Biotechnology Progress 21:106–111.

[259] Ku SCY, Ng DTW, Yap MGS, Chao S-H (2008) Effects of overexpression of X-box binding protein 1 on recombinant protein production in Chinese hamster ovary and NS0 myeloma cells. Biotechnology and Bioengineering 99:155–164.

[260] Yusufi FNK, Lakshmanan M, Ho YS, Loo BLW, Ariyaratne P, Yang Y, Ng SK, Tan TRM, Yeo HC, Lim HL, Ng SW, Hiu AP, Chow CP, Wan C, Chen S, Teo G, Song G, Chin JX, Ruan X, Sung KWK, Hu W-S, Yap MGS, Bardor M, Nagarajan N, Lee D-Y (2017) Mammalian Systems Biotechnology Reveals Global Cellular Adaptations in a Recombinant CHO Cell Line. Cell systems 4:530–542.e6.

[261] Selvarasu S, Ho YS, Chong WPK, Wong NSC, Yusufi FNK, Lee YY, Yap MGS, Lee D-Y (2012) Combined in silico modeling and metabolomics analysis to characterize fed-batch CHO cell culture. Biotechnology and Bioengineering 109:1415–1429.

[262] Galleguillos SN, Ruckerbauer D, Gerstl MP, Borth N, Hanscho M, Zanghellini J (2017) What can mathematical modelling say about CHO metabolism and protein glycosylation? Computational and structural biotechnology journal 15:212–221.

[263] Lund AM, Kaas CS, Brandl J, Pedersen LE, Kildegaard HF, Kristensen C, Andersen MR (2017) Network reconstruction of the mouse secretory pathway applied on CHO cell transcriptome data. BMC Systems Biology 11:37.

[264] Jimenez Del Val I, Polizzi KM, Kontoravdi C (2016) A theoretical estimate for nucleotide sugar demand towards Chinese Hamster Ovary cellular glycosylation. Sci Rep. doi: 10.1038/srep28547

[265] O'Brien EJ, Lerman JA, Chang RL, Hyduke DR, Palsson BO (2014) Genome-scale models of metabolism and gene expression extend and refine growth phenotype prediction. Molecular Systems Biology 9:693–693.

[266] Vazquez A, Oltvai ZN (2016) Macromolecular crowding explains overflow metabolism in cells. Sci Rep. doi: 10.1038/srep31007

[267] Beg QK, Vazquez A, Ernst J, de Menezes MA, Bar-Joseph Z, Barabási A-L, Oltvai ZN (2007) Intracellular crowding defines the mode and sequence of substrate uptake by Escherichia coli and constrains its metabolic activity. Proceedings of the National Academy of Sciences of the United States of America 104:12663–8.

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

[269] Lewis NE, Nagarajan H, Palsson BO (2012) Constraining the metabolic genotype-phenotype relationship using a phylogeny of in silico methods. Nature reviews Microbiology 10:291–305.

[270] Cambridge SB, Gnad F, Nguyen C, Bermejo JL, Krüger M, Mann M (2011) Systems-wide Proteomic Analysis in Mammalian Cells Reveals Conserved, Functional Protein Turnover. Journal of Proteome Research 10:5275–5284.

[271] Boisvert F-M, Ahmad Y, Gierliński M, Charrière F, Lamont D, Scott M, Barton G, Lamond AI (2012) A Quantitative Spatial Proteomics Analysis of Proteome Turnover in Human Cells. Molecular & Cellular Proteomics 11:M111.011429.

[272] Kristensen AR, Gsponer J, Foster LJ (2014) Protein synthesis rate is the predominant regulator of protein expression during differentiation. Molecular Systems Biology 9:689–689.

[273] Hausser J, Mayo A, Alon U (2018) Central dogma rates and the trade-off between precision and economy. bioRxiv 276139.

[274] Mathieson T, Franken H, Kosinski J, Kurzawa N, Zinn N, Sweetman G, Poeckel D, Ratnu VS, Schramm M, Becher I, Steidel M, Noh K-M, Bergamini G, Beck M, Bantscheff M, Savitski MM (2018) Systematic analysis of protein turnover in primary cells. Nature Communications 9:689.

[275] Wang M, Herrmann CJ, Simonovic M, Szklarczyk D, von Mering C (2015) Version 4.0 of PaxDb: Protein abundance data, integrated across model organisms, tissues, and cell-lines. PROTEOMICS 15:3163–3168.

[276] Hinkson I V., Elias JE (2011) The dynamic state of protein turnover: It's about time. Trends in Cell Biology 21:293–303.

[277] Eden E, Geva-Zatorsky N, Issaeva I, Cohen A, Dekel E, Danon T, Cohen L, Mayo A, Alon U (2011) Proteome Half-Life Dynamics in Living Human Cells. Science 331:764–768.

[278] Yen H-CS, Xu Q, Chou DM, Zhao Z, Elledge SJ (2008) Global Protein Stability Profiling in Mammalian Cells. Science 322:918–923.

[279] Feldhoff RC, Taylor JM, Jefferson LS (1977) Synthesis and secretion of rat albumin in vivo, in perfused liver, and in isolated hepatocytes. Effects of hypophysectomy and growth hormone treatment. The Journal of biological chemistry 252:3611–6.

[280] Schutte RJ, Parisi-Amon A, Reichert WM (2009) Cytokine profiling using monocytes/macrophages cultured on common biomaterials with a range of surface chemistries. Journal of biomedical materials research Part A 88:128–39.

[281] Golabgir A, Gutierrez JM, Hefzi H, Li S, Palsson BO, Herwig C, Lewis NE (2016) Quantitative feature extraction from the Chinese hamster ovary bioprocess bibliome using a novel meta-analysis work fl ow. Biotechnology Advances 34:621–633.

[282] Selvarasu S, Ho YS, Chong WPK, Wong NSC, Yusufi FNK, Lee YY, Yap MGS, Lee D-Y (2012) Combined in silico modeling and metabolomics analysis to characterize fed-batch CHO cell culture. Biotechnology and bioengineering 109:1415–29.