Department of Chemical Engineering
Imperial College London

Utilising a Genome Scale Metabolic Model to Design High-Producing CHO Cell Lines through Bi-Level Optimisation

Athanasios Antonakoudis

Supervisors: Dr. Cleo Kontoravdi and Dr. Alexandros Kiparissides

July 2022

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

"*All men naturally desire knowledge*"

Aristotle, 4th century B.C, The Metaphysics, Book I.

Declaration

I hereby certify that unless otherwise stated, the work within this document is my own. Where data or material from other sources has been used the original author has been appropriately credited.

Athanasios Antonakoudis

London, U.K.

26/07/2022

Copyright

The copyright of this thesis rests with the author. Unless otherwise indicated, its contents are licensed under a Creative Commons Attribution-Non-Commercial 4.0 International Licence (CC BY- NC). Under this licence, you may copy and redistribute the material in any medium or format. You may also create and distribute modified versions of the work. This is on the condition that: you credit the author and do not use it, or any derivative works, for a commercial purpose. When reusing or sharing this work, ensure you make the licence terms clear to others by naming the licence and linking to the licence text. Where a work has been adapted, you should indicate that the work has been changed and describe those changes. Please seek permission from the copyright holder for uses of this work that are not included in this licence or permitted under UK Copyright Law.

Abstract

Monoclonal antibodies (mAbs) are a type of biopharmaceutical product, that binds to specific antigens in the body and are administrated mainly for the treatment of cancer or autoimmune disorders. Unfortunately, mAbs suffer from a high market cost, forbidding treatment to many patients. Chinese hamster ovary (CHO) cells are the dominant expression platform of the biopharmaceutical mAb. However, despite increasing global demand and persistent societal needs, mammalian cell-based production systems suffer from low product yields as specific productivity is inversely proportional to cellular growth rates. The low yields can carry a burden on the upstream production phase of mAb production in CHO cell cultures, increasing the cost of the final bio product. The work presented here aims to reduce the upstream production cost of mAbs in CHO cell systems by proposing an increase of their specific productivity. Specifically, the design of a high-producing CHO cell line is attempted herein using a combination of *in silico* and *in vitro* techniques. Initially, the most recently developed genome-scale metabolic model (GEM) iCHO1766, on the time of commencing the research, is manually curated by limiting the cellular uptake and secretion rates based on four different experimental exometabolomic datasets and applying elemental balance algorithms to obtain realistic bounds for the intracellular reactions. Then, bilevel optimization is applied to the GEM, aiming in detecting genetic regulations that would increase specific mAb productivity during the growth phase of the cell culture. The strain optimisation on the CHO-K1 GEM results in a plethora of scenarios, all of which can secrete antibody in a biomass maximising Flux Balance Analysis (FBA) optimisation. These scenarios contain a large number of regulations, making them unsuitable for further investigation. A sensitivity analysis framework is then developed to first reduce the number of regulations to the absolute necessary ones for antibody synthesis, and secondly, to detect the best scenarios based on their FBA-predicted growth rate and specific productivity. An in-depth analysis of the proposed metabolic changes suggests that the regulations create an abundance in one of the amino acids that the cell uptakes, which must be consumed through the mAb synthesis pathway. Finally, two genetic engineering scenarios are screened experimentally. The first contains a knockout of *Aldh4a1* and an overexpression of *Aldh18a1*. However, the protein ALDH18A1 could not be detected, so it was thought that the gene is not being translated in the CHO cell line used for genetic modifications. The second scenario included just a knockdown of *Bcat1*. The cells are then cultured in mild hypothermic fed-batch conditions for 14 days. This modification displayed a 35% increase in the final mAb titer compared to the unmodified CHO cell line. The scenario requires further future testing to determine the underlying reason for the 35% increase in titer; however, the methodology followed shows promise for future use for CHO cell and, more generally, mammalian cell metabolic network engineering.

Acknowledgments

I would like to start by thanking my supervisor, Prof. Cleo Kontoravdi. She provided me with support, opportunities, and guidance from the start of the Ph.D. Always allowing me to pursue my goals, pushing me to achieve better results, and taking risks. During the difficult times of the lockdown, Cleo provided me with support, which I have greatly appreciated. She is a great mentor and teacher, who made the PhD journey something I would gladly take again. Dr. Alex Kiparissides helped me understand the basics of optimisation and metabolic modelling. I am very grateful for the strong foundation that he provided me with. I would like to thank his UCL group, Max, Victoria, Artemis, and Victor, who welcomed me and helped me in the start of the PhD.

I cannot thank Dr. Masue Marbiah enough. She patiently explained molecular biology to someone with absolutely no lab experience and limited biology knowledge. She made the time in the lab such an enjoyable process that I thought of even converting to an experimentalist...

I am also very grateful to be part of an amazing research group. Dr. Pavlos Kotidis has been one of my best friends during the time in imperial. Our talks helped me shape my research and come up with new innovative ideas. I want to thank him for teaching me how to live and play the guitar. I would like to thank Dr. Elli Makrydaki who taught me how to perform a western blot, the sound of which always scared me. She is one my dearest friends, and I'm grateful for her company throughout the PhD journey. Ben Strain and James Morrissey for the company and very insightful talks on metabolic modelling. I could not have moved forward without Roberto. For a year of not being able to knock out our genes, we were supporting each other and crying together. I would also like to thank Rod, Nacho, Sahre, Tom, Mariana, Kosti, Steven and Syazana for making daily life in the office enjoyable and fun.

My friends outside the college also deserve a huge thank you. They have been understanding, helpful, and supportive over the years. Ph.D. is a very long journey, and friends really make the time pass like that. I would also like to thank my family, for their support, guidance, help and love over the years. Last but definitely not least, my brother Sotiris who inspired to pursue a career in biopharma.

Table of contents

Declaration i
Copyright ii
Abstract iii
Acknowledgments iv
Table of contents v
Nomenclature viii
List of figures xii
List of tables xiv
1 Introduction 1
1.1 Monoclonal Antibodies 2
1.1.1 Production of mAbs 4
1.2 Objectives & structure of the thesis 6
2 Literature Review 9
2.1 Metabolic Modelling 9
2.1.1 Flux Balance Analysis 9
2.1.2 Kinetic Modelling 13
2.1.3 Genome Scale Models 14
2.1.4 GEM Curation 17
2.1.5 GEM Applications 18
2.2 Optimisation of CHO cell culture 20
2.2.1 Strain Optimisation 21
2.2.2 Strain Design Algorithms 22
2.3 Aims and Objectives 23
3 Computational Methods 25
3.1 CHO-K1 GEM 25

3.2 FBA Optimisation 26
3.3 Strain Optimisation Algorithm 28
3.3.1 MILP Optimisation 28
3.3.2 Bi-level Optimisation 30
3.3.3 OptReg 31
3.3.4 Conversion to MILP 36
3.4 Sensitivity Analysis 38
4 Strain Design 41
4.1 Dataset definition 42
4.2 Elemental Balance – Dataset Curation 47
4.3 Original CHO cell line 52
4.3.1 Pareto curve 55
4.3.2 Gene and Reaction Essentiality 57
4.4 Strain Design Algorithm 60
4.4.1 Optimisation Settings 61
4.4.2 Optimisation Results 64
4.5 Conclusions 66
5 Sensitivity Analysis 68
5.1 Sensitivity Analysis 69
5.2 Optimisation results 75
5.2.1 Proline Abundance with growth inhibition 77
5.2.2 Leucine abundance with growth inhibition 86
5.2.3 Phenylalanine abundance with growth inhibition 93
5.3 Conclusion 107
6 Experimental Materials and Methods 109
6.1 Cell Line, maintenance, and storage 109
6.2 Fed-Batch culture 109
6.3 Genetic engineering 110
6.3.1 Plasmid transfection 110

6.3.2 FACS Cell Sorting 110
6.3.3 Plasmid design 110
6.3.4 Plasmid DNA purification 112
6.4 Analytical Assays 112
6.4.1 Cell Lysis 112
6.4.2 Western Blot 112
6.4.3 Antibody quantification 114
7 Genetic Engineering 115
7.1 Genetic Engineering 116
7.1.1 Genetic Modification 116
7.1.2 DNA Transfection 118
7.1.3 Stable Cell Line Development 120
7.2 CHO-AA 122
7.2.1 Knock-Out of *Aldh4a1* 122
7.2.2 Overexpression of *Aldh18a1* 130
7.3 CHO-B 134
7.3.1 Knock-out of *Bcat1* 134
7.3.2 CHO-B Analysis 136
7.4 Conclusions 142
8 Conclusions & future work 145
8.1 Summary of results and conclusions 145
8.2 Main Contributions and Conclusions 148
8.3 Future work 150
8.3.1 Strain optimisation 151
8.3.2 Further genetic engineering 152
8.3.3 CHO-B Analysis 152
List of publications 154
Bibliography 155
Appendix 165

Nomenclature

Symbol	Description
y_i^{KO}	Binary variables that chooses the KO of reactions
y_i^{down}	Binary variables that chooses the downregulation of reactions
y_i^{up}	Binary variables that chooses the upregulation of reactions
^{13}C	natural stable isotope of carbon
A	Adenine
Bp	base pairs
C	Cytosine
ccFBA	carbon constrain FBA
CHO	Chinese Hamster Ovary
CoA	Coenzyme A
COBRA	COnstraint-Based Reconstruction and Analysis
CPU	Central Processing Unit
CRISPR	Clustered regularly interspaced short palindromic repeats
CRISPRa	CRISPR activation
D1	Dataset 1
D2	Dataset 2
D3	Dataset 3
D4	Dataset 4
Da	Dalton
DCW	Dry Cell Weight
DFBA	Dynamin FBA
DNA	Deoxyribonucleic acid
DSB	Double Strand Break
E. coli	Escherichia coli
E_a	Activation Energy
e_{down}	Coefficient to adjust the downregulation degree
e_{ep}	Coefficient to adjust the upregulation degree
$Essential^{Genes}$	Number of essential genes
$Essential^{r}$	Essential reactions with no GPR association
$Essential^{r_genes}$	Reactions associated with the essential genes
$Essential^{r_GPR}$	Essential reactions with GPR associations that do not belong in $\boldsymbol{Essential^{Genes}}$
FACS	Fluorescence-activated cell sorting

FBA	Flux Balance Analysis
FDA	Food and Drug Administration
FVA	Flux Variabilty Analysis
G	Guanine
GAMS	General Algebraic Modeling System
GEM	Genome Scale Metabolic Model
GPR	Gene-Protein-Reaction
gRNA	guide RNA
HDR	Homology directed repair
INDEL	insertion or deletion of bases in the genome of an organism
IVCD	Integral of Viable Cell Density
k	Coefficient to adjust the total number of regulations
LB	Lower Bound
LP	Linear Programming
M	Big M variable, takes values of large integers
mAb	Monoclonal antibody
MFA	Metabolic Flux Balance Analysis
MILP	Mixes Integer Linear Programming
mRNA	messenger RNA
MSX	Methionine Sulfoximine
MW	Molecular Weight
NADPH	Nicotinamide adenine dinucleotide phosphate
ncFBA	nitrogen constrain FBA
NHEJ	Non-homologous end joining
ORF	Open Reading Frame
PAM	protospacer adjacent motif
pFBA	Parsimonious FBA
PPP	Pentose phosphate pathway
q_{mAb}	Specific productivity of mAb
qPCR	Real-time polymerase chain reaction
RNA	Ribonucleic acid
S	Stoichiometric matrix
sgRNA	single guide RNA
SSB	Single Strand Break
T	Thymine
TCA	Tricarboxylic acid
TMFA	Thermodynamic Metabolic Flux Balance Analysis

UB	Upper Bound
v	Vector of fluxes
VCD	Viable Cell Density
WB	Western Blot
ΔG	Change in Gibbs free energy
KO	Knockout
μ	Growth rate

Enzymes and Metabolites

1py5c	1-pyrroline-5-carboxylate
1PYR5C	1-pyrroline-5-carboxylate
34hpp	3-(4-hydroxyphenyl) pyruvate
4mop	4-methyl-2-oxopentanoate
AACOAT	Acetoacetyl-CoA transferase
acac	acetoacetate
accoa	acetoacetate CoA ligase
acgam6p	N-Acetyl-D-glucosamine 6-phosphate
ACITL	ATP citrate synthase
ACS	Acetyl-CoA synthetase
ADP	Adenosine diphosphate
ALDH18A1	aldehyde dehydrogenase 18 family, member A1
ALDH4A1	Aldehyde Dehydrogenase 4 Family Member A1
AMP	Adenosine monophosphate
ATP	Adenosine triphosphate
BCAT1	branched chain amino acid transaminase 1
Cas9	CRISPR associated protein 9
DHFR	dihydrofolate reductase
$FADH_2$	Flavin adenine dinucleotide
FEROc	Ferroxidase
G2SADr	L-glutamate 5-semialdehyde dehydratase
G5SDym	glutamate-5-semialdehyde dehydrogenase
GAPDH	Glyceraldehyde 3-phosphate dehydrogenase
glu5sa	L-glutamate 5-semialdehyde
GLUPRT	glutamine phosphoribosyldiphosphate amidotransferase
GMPS	gmp synthase
GS	Glutamine synthetase
HBZOPT10m	hydroxybenzoate decaprenyltransferase

IgG	Imunoglobulin G
IMPD	IMP dehydrogenase
P5CD	1-pyrroline-5-carboxylate dehydrogenase
P5CR	pyrroline-5-carboxylate reductase
paf_cho	1-alkyl 2-acteylglycerol 3-phosphocholine
PPAm	inorganic diphosphatase in mitochondria
PROD2	proline dehydrogenase
pser_L	O-phospho-L-serine
PSP_L	phosphoserine phosphatase
r0074	glutamate-5-semialdehyde oxidoreductase
SpCas9	Streptococcus pyogenes Cas9
TYROX	L-tyrosine oxidase
TYRTA	tyrosine transaminase
xpm	xanthosine 5-phospate

The abbreviations of reactions, metabolites and genes names are the same as in the BiGG databases (47).

List of figures

Figure 1.1: Visual representation of a typical Antibody molecule 2
Figure 2.1: Graphical representation of a metabolic network 11
Figure 2.2: The solution space of an FBA optimisation 12
Figure 2.3: The four types of GPR associations 15
Figure 2.4: Algorithms to improve the network and efficiency of genome-scale metabolic models. 17
Figure 2.5: Algorithms used in strain design for the overproduction of metabolites. 22
Figure 3.1: Generic GEM of the CHO cell organism. 26
Figure 3.2: The four main steps of an FBA optimisation 27
Figure 3.3: Mathematical formulation of the OptReg optimisation 32
Figure 3.4: Mathematical formulation for the optimisation algorithm Bio_Igg 35
Figure 3.5: MILP formulation of the bilevel optimisation algorithm Bio_Igg 37
Figure 3.6: The algorithm used to check in a reaction is vital for IgG production 39
Figure 3.7: Algorithm to check whether any of the combination of the regulations are vital. 40
Figure 4.1: Upper and lower bounds of the ccFBA and ncFBA 51
Figure 4.2: FBA optimisation of the original cell line CHO-K1 for all four datasets 54
Figure 4.3: Trade-off curves between biomass and antibody production 56
Figure 4.4: Formulation of FBA optimisation 63
Figure 4.5: Trade-off curves between biomass and antibody production 64
Figure 5.1: Framework of the SA methodology 70
Figure 5.2: Reduction of the number of regulations 71
Figure 5.3: Reduction of the number of regulations in the high producing scenarios. 75
Figure 5.4: Metabolic network of the original CHO-K1 GEM around proline. 79
Figure 5.5: Metabolic network of the CHO-K1 GEM model around proline. 80
Figure 5.6: Knock-out of G5SADrm_b (A) in the CHO-K1 original GEM. 81
Figure 5.7: The effect that the removal of the proline catabolic pathway has on the metabolism 83
Figure 5.8: Knock-out of the *Aldh4a1* gene and upregulation of *Aldh18a1* gene 84
Figure 5.9: Upregulation degree of *Aldh18a1* and FBA prediction of μ and $q_{mAb.}$ 86
Figure 5.10: Upregulation of the forward direction of RTOTAL2t and RTOTAL3t. 88
Figure 5.11: Metabolic netowork around ACS enzyme in the CHO-K1 GEM 89
Figure 5.12: Catabolism of serine in the CHO-K1 GEM model 91
Figure 5.13: Catabolism of leucine in the CHO-K1 GEM 92
Figure 5.14: Oxidative phosphorylation of the CHO-K1 GEM 97
Figure 5.15: Metabolic network around the metabolism of guanine in the CHO-K1 GEM 99
Figure 5.16: Metabolism of oxalate in the CHO-K1 GEM 100

Figure 5.17: The metabolism of tyrosine and phenylalanine in the original CHO-K1 GEM. 103
Figure 5.18: Removal of GUAt_r in the original CHO-K1 GEM. 105
Figure 5.19: Removal of tyrosine oxidase and tyrosine transaminase enzymes. 106
Figure 7.1: Graphical representation of the different designs for dual sgRNA PAM sequences..... 118
Figure 7.2: Different Methods for DNA delivery in eukaryotic cells. 119
Figure 7.3: The two major ways for selecting cells after DNA transfection 121
Figure 7.4: Western blot analysis of the original CHO and CHO-CoT cell lines..... 124
Figure 7.5: Western blot analysis of the original CHO and CHO-CoT cell lines..... 125
Figure 7.6: Kill curve by the addition of the antibiotic hygromycin..... 128
Figure 7.7: Western blot analysis of the CHO-A cell lysates for each of the six conditions..... 129
Figure 7.8: Western blot analysis of the CHO-AA lysates 132
Figure 7.9: Western blot analysis of the CHO-AA lysates 132
Figure 7.10: Western blot analysis of the CHO-AA lysates 133
Figure 7.11: Non-reducing western blot analysis of the CHO-AA lysates. 133
Figure 7.12: Western blot to identify the degree of BCAT1 expression..... 136
Figure 7.13: Growth curve of the fed-batch culture for all three cell lines used 138
Figure 7.14: Product yield and specific productivity graphs for the fed-batch cultures 139
Figure 7.15: Western blot to identify the degree of BCAT1 expression..... 140

List of tables

Table 2.1: Various omics data that are used to formulate a genome-scale metabolic network (*39*)..... 16
Table 2.2: Advantages and disadvantages of FBA and kinetic modelling..... 24
Table 4.1: The four datasets used in the cell-line design 43
Table 4.2: Constraints used for the metabolites 46
Table 4.3: Results of the element balance algorithm on each of the four datasets 50
Table 4.4: FVA, carbon, and nitrogen balance algorithm results. 52
Table 4.5: Experimental values and FBA predictions for μ q_{mAb} 53
Table 4.6: Control values from the FBA optimisation of the CHO-K1 GEM 55
Table 4.7: Number of essential genes and reactions 59
Table 4.8: Number of reactions that were removed from the possible reaction pool 62
Table 4.9: Parameters of the strain design optimisation process 63
Table 5.1: The 20 high producing scenarios 73
Table 5.2: Predicted μ and q_{mAb} of the proline abundance scenarios 78
Table 5.3: Predicted μ and q_{mAb} of the leucine abundance 49, 52 and 74 scenarios. 87
Table 5.4: FBA predicted μ and q_{mAb} for the high-producing scenarios 94
Table 6.1: Plasmids used for the genetic engineering experiments 111
Table 6.2: Recipe for 2xSDS-PAGE gels 113
Table 7.1: The six different conditions used for developing a selection protocol for the original CHO and CHO-A cell line based on the hygromycin kill-curve..... 127

Table A1: Lower and upper bounds for each of the four experimental datasets.……..……169
Table A2: The 20 high-producing scenarios.……...……170

1

Introduction

Biopharmaceuticals, also known as biological medical products, used for tackling human diseases such as cancer or autoimmune disorders have witnessed a rapid development during the last 20 years. They are pharmaceutical drugs produced or extracted from biological sources, such as living cells. They are complex molecules, with usually large molecular weights in contrast to common pharmaceutical drugs. Gary Walsh has shown in his reviews the increase of the biopharmaceutical sector in the market share of pharmaceuticals (*1-4*). In 2006 there were around 165 biopharmaceutical products approved by the medicine agencies with $33 billion in annual sales. In 2018 the number of total approved biopharmaceutical drugs rose to over 300 with a significant increase in the market share of the field, expecting to reach half a trillion USD in 2027 (4). Half of the approved drugs are monoclonal antibodies (mAbs) and represent eight of the top-ten drugs by sales in 2017 (*5*). The profile of products in late-stage clinical trials suggests that new biopharmaceuticals will continuously get released. Just in 2020 the mAb biopharmaceutical drug Dupixent received approval for use in the European Union, however, it, as well as other similar drugs, suffer from expensive market prices. Another issue biopharmaceuticals face, especially mAbs, is long approval times from medicine administrations (*6*), due to the complexity of their targets and unmet medical needs (condition where there is no satisfactory way of treatment), which, combined with their high market price, makes the administrations sceptical about whether to approve them or not. For example, Dupixent, a monoclonal antibody drug used for the treatment of ectopic dermatitis, is not yet approved in the United Kingdom by the National Institute for Health and Care Excellence (NICE) due to its high market cost (*7*). The rise of biosimilar products can help reduce the cost of mAb production and make them more affordable to patients. The patents of several biopharmaceuticals are expiring, which means that third parties can develop products that exhibit similar therapeutic properties. They require shorter clinical trials because the drug has already been accepted by medicine agencies, which reduces the overall development cost of the product (*8*). A massive increase in the quantity of biosimilar products in the market, as well as the amount of them being researched by third companies gives rise to new innovative ways for optimisation, regarding for example their production cost. The importance of the biopharmaceutical mAb makes it a target for further study and optimisation, and the option of biosimilar production can give rise to novel optimisation techniques.

1.1 Monoclonal Antibodies

Our body needs to defend itself from pathogens or harmful invaders. The defence mechanism is called immune system and saves us every day from death by infection. There are both innate and adaptive immune responses, with their goal being to destroy invading molecules and any toxic products. Innate immune responses are those already present in the organism whereas adaptive are created from the response to a pathogen. (*9*) Generally, any substance capable of stimulating the adaptive immune response system is called an antigen (antibody generator). Adaptive immune responses are carried out by white blood cells called lymphocytes such as B and T cells. They perform two kinds of responses, antibody and cell-mediated responses. In antibody responses, B cells, which are produced in the bone marrow, secrete Immunoglobulins (Ig) also known as antibodies. Antibodies specifically bind to the antigen that stimulated their production and inactivate the invading substance by blocking their ability to bind on host cells or marking them for phagocytic cells to ingest them (*10*). A typical antibody is Y-shaped as seen in Figure 1.1.

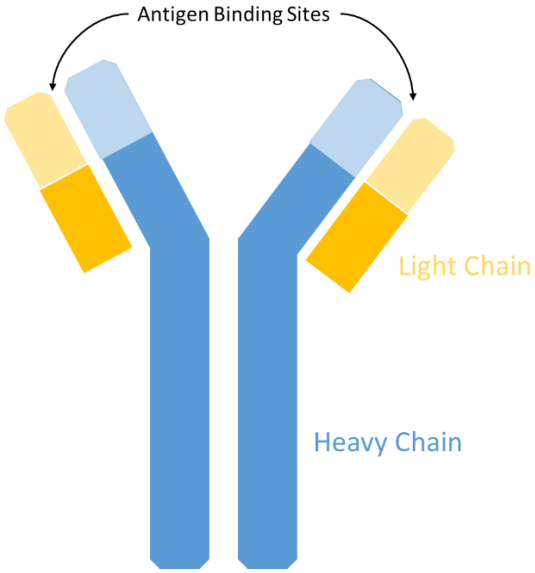


Figure 1.1: Visual representation of a typical Antibody molecule. It is composed by two identical heavy chains and two light chains. The two antigen binding sites are identical and formed in the end of the light and heavy chain. Antigens bind to it as keys to a keychain (*11*).

Antibodies are glycoproteins which are proteins that contain oligosaccharide chains. They consist of four polypeptide chains, two identical heavy and two identical light chains. There are five different classes of Ig, IgA, IgD, IgE, IgG and IgM with α, δ, ε, γ and μ classes of heavy chains respectively. The most common class of immunoglobulin in human blood is IgG, which consists of four chains, two heavy and two light ones. IgG molecules bind to specific receptors called Fc receptors on

macrophages and neutrophils, where these phagocytic cells can bind, ingest, and destroy the infecting microorganisms to which the IgG antibodies have coated (*12*).

One would wonder how the immune system would respond to all those different antigens. The answer lies in clonal selection theory where the host first generates a random diversity of lymphocytes and those that can bind to the antigen are selected and duplicated (*13*). As a result of the clonal selection theory every antibody made from a specific cell has the same antigen specific binding cite, thus, each B cell produces a single species of antibodies, each with a specific binding site. The parts of the antigen that combine with the antibody-binding site are called antigen determinants or epitopes. Most antigens have a variety of different epitopes. A single epitope can activate many different lymphocyte clones. When the response stimulates only one B or T cell, then the antibody is called monoclonal antibody (mAb) (*10*).

The ability of mAbs to target only one single epitope justifies Paul Ehrilich's argument over a magic bullet. Over a century ago he reasoned that if a compound could be made to selectively target a disease-causing organism, then that compound could target the intended cell structure or deliver a toxin to destroy the infectious pathogen (*14*). Hence, a "magic bullet" would be created to destroy only the infections organism it targets. The vision became practical with the development of hybridoma technology by Köhler and Milstein in 1975; they made the development of antibody-drug conjugates feasible, which enables the drug to specifically target the infectious organism. The discovery won a Nobel Prize in Physiology or Medicine 10 years later, showcasing the importance of their work.

Hybridoma technology was used to produce the first monoclonal antibody-secreting cell line. It is the technique that involves the propagation of cell clones from a single antibody secreting B cell, so that a homogeneous solution of antibodies can be obtained in large quantity. Specifically, hybridoma cell lines are formed via fusion between a short-lived antibody-producing B cell and an immortal myeloma cell, to deal with the problem of limited life span of the B cells. Each hybridoma cell constitutively expresses a large amount of one specific mAb, and selected hybridoma cell lines can be cryopreserved for long-lasting mAb production. These hybridoma cell lines are cultured as different clones, each of which produces monoclonal antibodies indefinitely (*15*).

Monoclonal antibodies are important therapeutic agents for cancer, and for over 20 years have been a common component in cancer therapy. They possess several mechanisms with which they can aid our immune system into directly targeting the malignant cells, modifying the host response, delivering cytotoxic moieties or by targeting the malignant cells so the immune system can respond. For an immunotoxin to be effective it needs to bind to the surface antigen on a cancer cell, enter the cell through endocytosis and kill the cell. The density and consistency of the antigen expression on malignant cells and the limited expression on benign cells makes the mAbs desirable for cancer therapy, providing a more targeted treatment (*11, 16*).

1.1.1 Production of mAbs

Monoclonal antibodies have been produced in a variety of different expression platforms including yeast, fungi, bacteria, plant, or mammalian cell lines (*17*). Plant cell lines are great for large-scale bioprocesses especially when high amounts of recombinant protein are required, but they can result in a costly downstream process due to the amounts of material that needs to be processed. They can perform post-translation modifications of proteins like those in animal cells, but complex N-glycans display major differences between species that can result in possible immunogenic reactions. For that reason, chaperone proteins assist the protein folding and the addition of N-glycans is possible due to specific glycosyltransferases, resulting in a very similar to human produced bio-product (*18*). Bacteria, a category dominated by *Escherichia coli*, reaches volumetric yield of 1.5 grams per litre scale (*19*). *E. coli* is also used to produce antibody fragments which are secreted in the periplasm and have to be extracted from there, a not very efficient process. Also, bacteria cannot perform a human-like glycosylation profile which limits their potential (*20*). Yeast combine the properties of eukaryotic cells of short generation time and ease of genetic engineering with the simple culture medium requirements of prokaryotic hosts. They are quick and easy to grow in comparison to mammalian cells and have the simple medium requirements of microbial unicellular hosts. They have been used in food production for many years because they do not produce endotoxins and have gained GRAS (Generally Recognized as Safe) status. Nevertheless, yeast tend to hyper glycosylate heterologous proteins, which affect their activity and is a possible source of unwanted immunogenicity or adverse reactions in humans (*17*). Filamentous fungi can secrete large amounts of proteins into the culture medium making them a cost-effective platform, but they can exhibit protein degradation (*21*). Even though bacterial, yeast, plant and fungi have certain advantages, the advance product quality of mammalian cells makes them the dominant expression platform.

Today, most mAbs are produced in mammalian hosts due to their advanced protein synthesis, creating mAbs indistinguishable from the human ones (*22*). The efficiency in secretion and their ability to produce high-quality large and complex glycosylated bioproducts, results in fewer downstream processing steps compared to the other eukaryotic cells, reducing the effort and cost. Also, the human-like post-translational modifications ensure correct activity of the product and minimize the danger of immunogenicity (*17*). Mammalian cell systems do not usually contain the protein of interest in their genome, so it must be expressed recombinantly by introducing the gene that expresses the protein of interest to the organism, which unfortunately can result in low specific productivities (*16*). The productivity can be increased using genetic and process optimisation techniques. The recombinant expression of proteins faces natural burdens such as the optimal resource allocation resulting in the objectives of cell growth and recombinant protein production to create a competitive for cellular resources relationship. Recent optimisation strategies have increased the titer to significant values. Right now, Chinese hamster ovary (CHO) systems are the workhorse of mAb production, and they can

achieve titers of around 10g/L (*23*). There is a variety of different mammalian cell systems that find application in the biopharmaceutical sector, such as murine myeloma cells (NS0 and Sp2/0), human embryonic kidney cells (HEK293), mouse hybridoma and Chine hamster ovary (CHO) cells.

The most frequently used mammalian cell system is the CHO cell line. Accounting for over 70% of the manufacturing of the currently in the market recombinant proteins (*24*), the expression platform dominates the biopharmaceutical production of drugs and is considered the main expression platform for monoclonal antibody production. CHO cells can grow in serum free suspension cultures with high cell concentrations, which is preferred for large volume bioreactors. Mammalian cells are known for low specific productivity; however, this problem can be overcome with gene amplification. Another important characteristic is the glycosylation of the CHO cells. CHO-derived glycoprotein quality is compatible with humans, having the desired post-translational modifications, due to 99% of CHO genome genes being the same as the human one. Also, CHO cells have been well characterized as their genome is now available and are used for production of biopharmaceuticals for more than two decades, so the FDA approval process is rapid. Decades of research into this host have led to a significant increase of its productivity and optimization of its growth rate (*25, 26*). However, as CHO cells are non-human, initial mAb molecules had the potential to be immunogenic in humans (*27*). Advances in antibody engineering enabled the development of chimeric antibodies – an antibody encoded by genes from more than one species, human and murine for these types of mAbs – with the aim being to prevent immunogenic responses and make the antibodies human-compatible (*16*). These chimeric human-mouse mAbs behave immunologically as natural human IgG and are less likely to be recognised as hostile by the human immune system. They have a half-life similar to human IgG. Also, they are administered in a timeline that is convenient for the patients (2-4 weeks) and are in blood circulation for months at a time, meaning they have great bioavailability (*16*). They have been demonstrated to be safe hosts for high-level recombinant protein production including the production of monoclonal antibodies and have a strong track record with the regulatory authorities. There have been several clonal CHO cell lines developed thought the years. The three most common ones are CHO-S, CHO-K1 and CHO-DG44, each of them with a unique genotype and phenotype (*28*).

Animal cell culture has advanced significantly during the last decades and is considered now a robust, reliable, and mature technology. With the number of bio therapeutics constantly increasing, especially ones based on mAbs (*4*), biopharmaceutical companies trying to figure out new and better ways to produce and deliver them. Maintaining the desired product attributes and reducing the production time and cost are the main challenges today (*29*). The two stages of mAb production are the upstream and downstream process. The upstream bioprocess deals with the production of the drug from the living cells and the downstream is dealing with the purification of the product. The two stages account for similar costs to the production of the bio product and should both be optimised. Another important aspect for production of a biopharmaceutical is the average time length required for the FDA trials. These are clinical trials to ensure safety of the drug, and their average time is 10 years, making

the production time an important factor to optimise (*30*). Optimisation through mathematical modelling can reduce the cost of process development and production or the time required for the product to be produced.

1.2 Objectives & structure of the thesis

Despite their importance as a biopharmaceutical, mAbs suffer from high production cost. Cellular platforms have high running costs which can translate to expensive market prices forbidding the treatment to many patients. The production of mAbs, and most biopharmaceuticals in general, is split between the upstream stage, where the antibody is being produced, and downstream stage, where the product is purified. Both processes are very expensive and increase the cost of the final drug by approximately £1000 per dose (*31*). The research conducted herein is focused on reducing the production cost of mAbs and is centred on optimising the quantity of the produced antibody titer in the upstream phase, specifically in fed-batch production. The cost of the upstream production phase is a function of the size of the bioreactors, the volume of culture media and feed required and the running cost of the fed-batch operation, which are all a function of the required quantity of mAb. An increase in productivity can reduce the cost by either reducing the culture volume or the culture duration, which subsequently, after the application of a techno economic analysis to determine how the increased titer affects the costs of the bioprocess, can be translated to a reduced market price.

Mathematical modelling and model-based optimization are tools that can be used to improve the efficiency of a bioprocess, such as the upstream production of mAbs in CHO cell systems (*32*). The development and production frameworks for the recombinant production of mAbs are somewhat empirical (*33*), usually including laborious clonal selections and/or screening of thousands of strains (*34*). The bioprocess optimisation framework can be sharpened using mathematical modelling. Bioprocess modelling is a crucial tool for understanding and analysing complex behaviour and properties in living organisms (*35*). The vast metabolic network of an organism can be greatly simplified to a metabolic model, which can be optimised to maximise the product of interest using mathematical programming instead of laborious trial and error approaches. Genome Scale Metabolic Models (GEMs) can aid in the optimisation of the upstream mAb production. These are models that contain most known biological information, genes, enzymes, and biochemical reactions about an organism, so they can provide a holistic approach of the under-study organism. Mathematical modelling is a great tool for suggesting optimisation strategies to a bioprocess, but experimental validation is necessary to verify that all the assumptions taken did not negatively impact the quality of the model. Any mathematical model uses assumptions to simplify the mathematical expressions or as an aid because an underlying biological reason or true value of a parameter is unknown. The assumptions can impact on the quality of the result. The experimental result can either verify the output of the optimisation or disprove it. In that case, it might be useful to feed the model with the outputs of the experiment to understand the

reason why it was not verified. A combined modelling and optimisation-based design approach can provide a proof of concept for bioprocess optimisation (*36-38*).

In this work, a model-based approach was used to mathematically represent and analyse the metabolic behaviour of CHO cells, which was later used as a means for the optimization of the expression platform under the aim of increasing the specific mAb productivity. This could increase the titer of mAbs from a CHO cell culture, which could correspond to a potentially reduced market price and a more affordable treatment for patients. Initially, the literature review (Chapter 2) sets the foundation for the rest of the research, presenting the main strategies that will be examined and the relevant literature. The rest of the thesis is separated in three parts, the computational and experimental chapters, that contain the stages for the optimisation of the bioprocess from the *in-silico* modelling to the *in vitro* strain design, and the discussion and conclusions of all the results. In the computational part Chapter 3 explains the *in*-silico methods and techniques. The strain design chapter (Chapter 4) explains the process behind the optimisation. Initially, there is the definition of the experimental data used for to constrain the model and their augmentation to include elementary balances. Then, the model used, which is the GEM of the CHO-K1 cell line, is curated based on the four different experimental datasets. The performance of the metabolic model and of the data are tested using a Flux Balance Analysis (FBA) optimisation by predicting the growth rate and antibody productivity, which are compared to the experimental values of each experimental dataset. Then, the bi-level strain optimisation algorithm is analysed, as well as its mathematical formulation and its mixed-integer linear programming (MILP) reformulation. The optimisation algorithm resulted a plethora of different results, strategies, which consist of modifications for the reaction bounds. These strategies can couple the production of biomass and antibody in the FBA optimisation with biomass the objective function. However, the genetic strategies contained a large number of regulations and some of them had very low FBA predicted growth rate or predicted productivity, which would not be an ideal strategy to perform *in vitro*. A sensitivity analysis framework was developed to analyse the scenarios and determine the most optimal ones, which were then analysed regarding the suggested metabolic changes (Chapter 5). Initially, the number of regulations was reduced to the minimum ones required for the coupling of objective functions to occur. Then, the genetic engineering strategies that predict the highest values of growth rate and antibody specific productivity from an FBA optimisation, were identified as high producing to be studied further regarding the proposed metabolic changes. The study showed that the coupling of the two *in-silico* diametrical objective functions takes place because of an abundance in any of the amino acid that is consumed by the cell and have a limit on their uptake according to the experimental data used to constrain the model. The abundant metabolite is subsequently channelled towards recombinant protein synthesis. The next part of the thesis is the experimental validation of the computational results. Initially, the experimental materials and protocols are described in Chapter 6. Then, the genetic engineering strategies were tested *in vitro* (Chapter 7) using. Two different scenarios resulting from the optimisation were tested, one that focuses on leucine abundance and one on proline abundance. They

both contained gene knockouts which utilised the CRISPR/Cas9 system. The modified cells were tested regarding the produced mAb from a 14-day fed-batch culture. In the final chapter and part of the thesis (Chapter 8), all results are summarised, together with the limitations of the methodology, suggestions for improvements of the whole framework and future perspectives concluded from this thesis.

2

Literature Review

This chapter summarises the important literature related to the aim of the thesis. The chapter initially analyses the field of metabolic modelling, which mainly uses flux balance analysis but can happen using kinetic modelling, then the types of models called genome-scale metabolic models are analysed together with relevant algorithms and applications. Finally, a review of studies for optimisation performed in CHO cell cultures is presented and the proposed way of optimising the titer in the CHO culture under study.

2.1 Metabolic Modelling

Metabolic Modelling is one of the pillars of Metabolic Engineering. It involves the mathematical modelling of the pathways that makeup the metabolism of the cell and is widely used for analysing the fluxes, the allocations of resources and generally the study of the cell's phenotype (*39*). Mammalian metabolism is vast with thousands of enzymes and reactions, making mathematical modelling a convenient way to study it and come up with results and conclusions that can be that can be further used to detect the most important parameters for optimisation. It is a widely used technique with many areas of application, from analysis of different culture media and feeding strategies to the detection of bottlenecks in metabolic pathways in the cell metabolism. Metabolic modelling has frequently been applied to CHO cell culture processes in order to increase the production of monoclonal antibodies or of the growth rate or to achieve a required glycosylation profile or to study the metabolic bottlenecks of the organism (*40, 41*).

2.1.1 Flux Balance Analysis

The most widely used method for the study of a metabolic network is called Flux Balance Analysis (FBA). It is a mathematical technique to calculate the fluxes of the biochemical reactions in a metabolic network using linear programming. The concept of flux calculation in biochemical networks was proposed by Papoutsakis (*42*), who derived a stoichiometric equation from a metabolic map, added constraints to the reactions using both biochemical and thermodynamic constants and calculated the maximal theoretical growth rates and butyric acid yield in bacteria. However, the first mathematical

formulation of a linear programming optimisation, now known as FBA, was proposed later by Savinell and Palsson (*43*). Generally, FBA is used for analysing metabolic networks and making quantitative predictions in the absence of kinetic information (*44*). The simple implementation makes it a common option for metabolic engineers, and it gives rise to many applications with a variety of objectives.

An FBA optimisation has three main sets of parameters that affect the flux distribution and the value(s) of the objective function. First and foremost there is the metabolic network, which is the stoichiometric matrix for the optimisation. The network is usually imported from a biological database such as BiGG (*45*), but it can be created with a bottom-up procedure reaction by reaction. It can also be adjusted to a specific part of the metabolism that is of interest, from a generic large scale model. Generally, the metabolic network of an organism contains two types of reactions, transport and conversion ones. The transport reactions are responsible for moving metabolites from one compartment to another. There are as many cellular compartments in the model as in the cellular environment of the living cell the model represents (Section 2.2.1). The model also contains an extracellular environment which represents the metabolites that are in the culture medium and the outside environment which shows which metabolites are allowed to enter the extracellular environment. The distinction between the two extracellular environments helps to adjust the uptake and secretion rates. When a transport reaction transfers a metabolite from the outside to the extracellular environment it is also known as drain reaction. The reactions of the network can form internal loops, which are also known as type III extreme pathways. Extreme pathways are a set of flux vectors of a metabolic network i.e., possible flux distributions that satisfy the constraints (*46*). The so-called type III extreme pathways, describe the flux distributions that create infeasible loops of fluxes in the metabolic network, as the one of the figure 2.8. A very important reaction of the metabolic networks is the biomass production. All metabolic models represent a living organism, so the must contain a reaction responsible for simulating the conversion of resources to biomass and growth (*47*). The formation of a biomass equation can have three levels, depending on how detailed it is. A biomass equation should define the basic macromolecular content of the cell, and which metabolites are required to form this content. A more detailed equation would also add the energy carriers and the energy conversion from biosynthesis. Finally, by including co-factors, small molecules and other important metabolites the biomass equation represents in detail the growth rate of the modelled organism. The metabolic network of an organism is presented in figure 2.8.

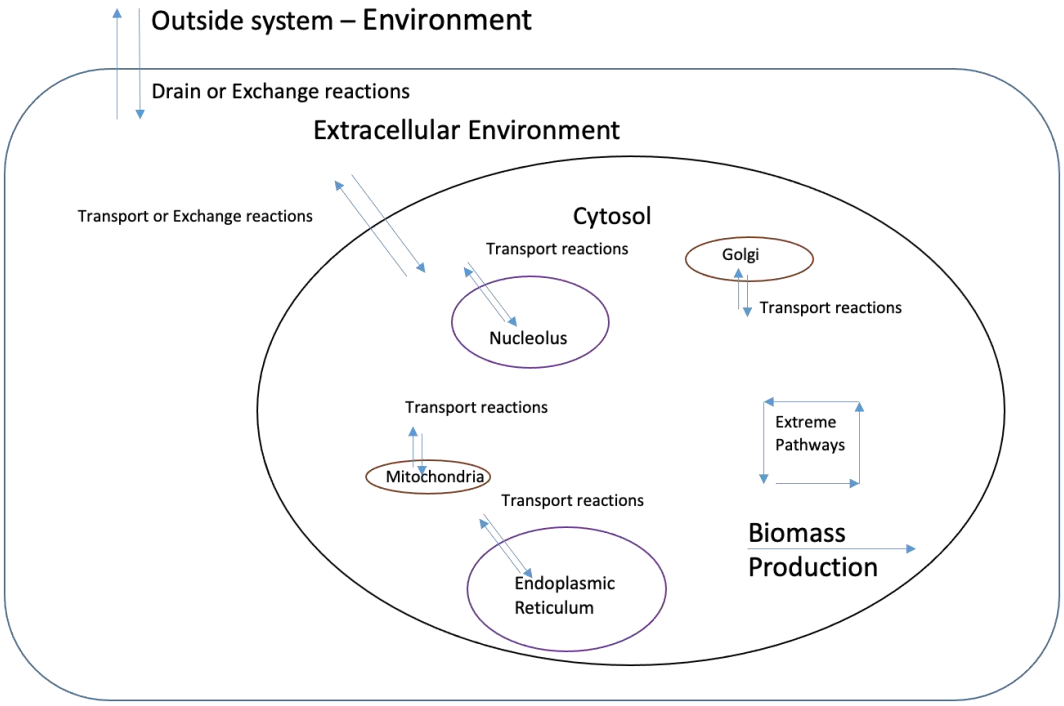


Figure 2.1: Graphical representation of a metabolic network. It can be separated into three main systems. The outside system, environment, which is connected with the extracellular environment through the drain reactions. Drain reactions, enable the uptake or secretion of metabolites. The extracellular environment is connected with the cell using transport reactions, which are used to transfer metabolites and resources from one compartment to another. The cellular system contains several smaller compartments such as the Golgi apparatus or the nucleolus which contain metabolic reactions. A metabolic network also contains a biomass equation, which represent the allocations of resources towards the growth of the cell (*48*).

Each flux of the metabolic network has upper and lower bounds, which are also important for an accurate and realistic flux distribution. These values are either imported from literature or from experimental data or are even extrapolated from computational methods. FBA is a linear optimisation problem, and there are numerous different optimal solutions – flux distributions (Fig 2.9). However, if the fluxes are over constrained then the optimization problem might be infeasible. Together with the choice of bounds, the choice of objective function is also important for the output of the optimisation and should represent the main goal of the cell. Usually, the objective function of an FBA is the maximisation of biomass production, which is the most natural objective of the living cells, and the one better representing the exponential phase of growth. However, there is the option of declaring other reactions, such as the production of metabolite or minimisation of a product. An uptake-rate objective function was used to minimise the sum of amino acid uptake, in order to achieve flux distributions of higher fidelity (*49*). The minimisation of metabolic adjustment (MOMA) between two phenotypes was set as an objective function to better predict flux distributions after applying a modification in a reaction of the metabolic network (*50*). Of course, there is the possibility of declaring more than one reaction as

an objective and forming a multi-objective optimisation problem. The choice of the objective function is responsible for the resulting flux distribution because the solution space will include all reaction fluxes of the system that maximise the objective. If biomass is chosen as objective function, then the resulting fluxes will mainly direct towards the production of biomass (*51*). The balances and bounds define the solution space, thus the more realist constraints a system has, the more realistic the solution will be as is displayed at figure 2.9 (*52*).

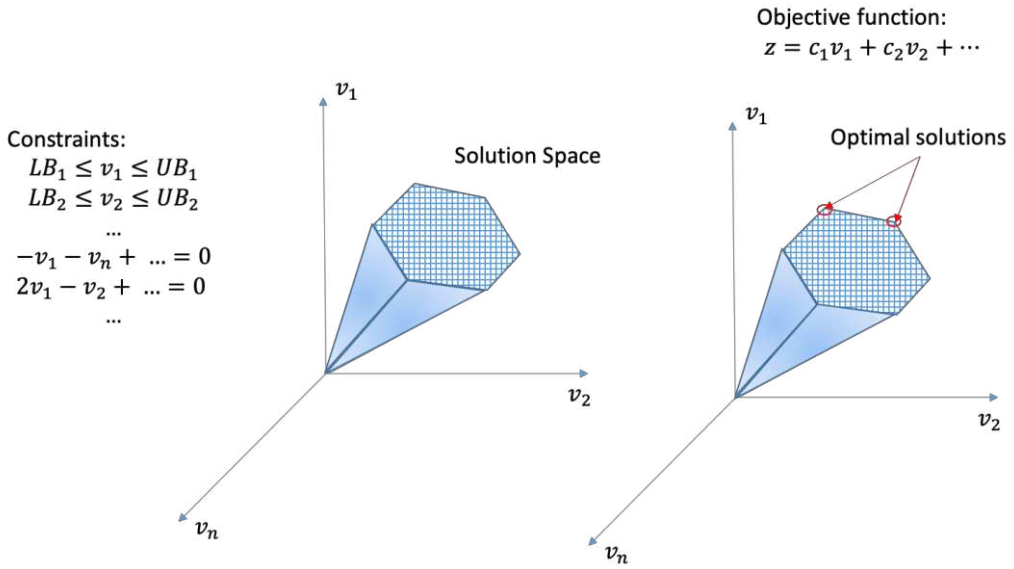


Figure 2.2: The solution space of an FBA optimisation. The choice of bounds defines the solution space, where by solving the optimisation problem with an objective function in mind, the optimal solution fluxes that satisfy the equations of the model can be calculated.

There is one branch of FBA called Metabolic Flux Analysis (MFA), which is an experimental fluxomics technique. It uses data in the form of fluxes to constrain the intracellular reactions, so the total number of degrees of freedom becomes more manageable. Also, MFA usually uses a small-scale metabolic model since the purpose is the prediction of accurate intracellular fluxes, whereas FBA can use larger scale one. The main advantage of MFA, similarly with FBA, is its simplicity and fast implementation. In combination with a supply of a few intracellular fluxes, MFA can result in an accurate phenotype. A common way to gather the flux is the use of ^{13}C-based metabolic flux analysis (^{13}C-MFA). It uses ^{13}C labelled tracers and other labelling measurements in combination with techniques such as NMR or mass spectrometry to allow the reliable estimation of the reaction fluxes. ^{13}C-MFA is a more advanced way for estimating metabolic fluxes in systems that are in a pseudo-steady-state. The cells are cultured with the presence of a ^{13}C-tracer, so that it is incorporated in metabolite intermediates and products. With a proper ^{13}C selection method the fluxes of the metabolites can be measured. A large number of redundant measurements – measurements for a large number of reactions' fluxes – can be easily obtained for flux estimation, so the system of equations will be more constrained with fewer degrees of freedom. The

main drawback of this method, however, is that it is far more resource demanding than simple MFA, both experimentally and computationally (*53*).

2.1.2 Kinetic Modelling

A different approach in metabolic modelling is the use of dynamic equations to represent the rates of the biochemical reactions. Kinetic metabolic models aim in describing dynamic behaviour of the metabolism. They are built on biological principles, and they try to express mathematically the main changes in the metabolism. The formulation of a kinetic model usually includes the model development, the sensitivity analysis of the parameters and their estimation, then finally its validation using independent experimental data (*54*). The main bottleneck in developing kinetic models is the lack of kinetic data, and most kinetic models either simplify the equations or reduce the size of the metabolism they want to model. These models are usually built in a bottom-up manner, meaning that each biochemical reaction's kinetic equation is joined to make up the whole system resulting in a significant number of parameters. Parameters are usually gathered from literature or databases, or they are approximated using parameter estimation (*55*). Parameter estimation is a statistical technique which can be applied by many different ways, like maximum likelihood parameter estimation, or least squares parameter estimation (*56*), where the parameters are calculated from experimentally measured data. It is important that parameter estimation defines parameters within reasonable bounds that obey the physicochemical laws. The main challenge in kinetic modelling is the factor of uncertainty which is part of biological systems (*57*). The underdetermined nature of the mathematical representation of a biological system, results in multiple possible sets of parameters that are able to reproduce a set of experimental data. To combat the uncertainty, frameworks based on Monte Carlo philosophy, which does not follow a deterministic law but uses statistical principles to find the most suitable parameter value (*58*) aid in the reduction of the solution space of the permissible parameters for the mathematical model. The size and content of the network is an issue as well, because the large number of parameters could result in overfitting and poor predictions.

Kinetic modelling has found a plethora of applications in bioprocess optimisation, with the main topic being the study of the dynamic changes in the metabolism. For example, in the study by Chen et al., a kinetic model describing the pathways of glucose in metabolism, coupled with cell population dynamics of CHO cells was developed (*59*). The parameters were gathered from sensitivity analysis and experimental data of batch and fed-batch cultures. Similarly, Robitaille et al. formed a dynamic metabolic model for CHO, describing the central carbon metabolism, where the kinetic parameters were gathered from experimental data (*60*). A similar model was developed for the *E. Coli* metabolism, where the metabolism describing model was parameterised using values from literature adjusting for their uncertainty (*61*). Diab et al. developed a dynamic model to describe the metabolism of CHO cells which was combined with equations describing the fed-batch culture system,

incorporating the variable of pH. That way they were able to predict the dynamic pH and productivity profiles in CHO cell cultures (*62*). Jones and Gerogiorgis developed a dynamic model to describe the behaviour of fed-batch and perfusion CHO culture, enabling them to perform a further analysis concerning the CHO cell as a mAb producing host (*63*). One of the main drawbacks of a kinetic model is the requirement for large experimental datasets, in order to acquire trustworthy kinetic parameters; the combination with FBA tries to deal with the issue.

A purely dynamic model needs to mathematically interpret a large portion of metabolic reactions, which increases the total number of parameters and their uncertainty. A way to deal with the uncertainty of the parameters it to couple the kinetic equations with FBA. FBA as explained above, is an optimisation strategy that calculates the intracellular fluxes from a metabolic network. There are two main schools of thought on dynamic FBA (DFBA), the static and the dynamic optimisation approach. The Static Optimization Approach divides the duration of experiment in discrete intervals. The FBA optimisation in each interval provides a flux distribution, which can give the impression of dynamic changes in fluxes. The Dynamic Optimization Approach involves the optimization of entire time periods of interest to obtain the time profile of fluxes. The dynamic optimization is transformed into a non-linear programming problem and solved using finite elements to parametrises the kinetic equations (*64*). Other possible ways for solving a DFBA system would be the use of B-splines (*65*), use of bi-level optimisation or by applying lexicographic optimisation (*66*). DFMA has been used to predict metabolic shifts, while using parameter estimation to set accurate numerical values for the kinetic model (*67-69*). Generally, kinetic modelling and DFBA have the ability of describing the dynamic behaviour of a biological system, however, when equations with unknown parameters enter the model then parametrisation is very important for the equality of the results.

2.1.3 Genome-Scale Models

Metabolism is highly complex and involves thousands of interconnected reactions, it is therefore necessary to use a technique for an efficient representation of the metabolic model. There are two different approaches for the analysis of metabolism in systems biology: the top-down where data is analysed to develop the model and the bottom-up methodology where detailed models for different compartments are assembled into a holistic model. Genome-scale reconstruction is based on a bottom-up methodology as where genomic and metabolic information is combined to create a holistic model (*39*). Genome scale metabolic models (GEMs) are the most accurate representations of the genotype-phenotype link and a tool for holistic studies of the cell. They are mathematical representations of a cell's metabolic network. They are using matrices to capture most information regarding the genes, enzymes and metabolites take part in the metabolism and their connections, so in that sense they form a database of all known biological information about an organism. A GEM describes the gene-protein-reaction (GPR) associations taking place inside the cell, which are the connections between genes,

enzymes and their biochemical reactions (*70*). These connections are formulated in the GEM as Boolean equations, where the flux of each reaction is dependent on whether the genes that are associated to the reaction are expressed or not. There are four types of GPR associations as seen in figure 2.10, which can be combined creating all kinds of GPR associations for a gene.

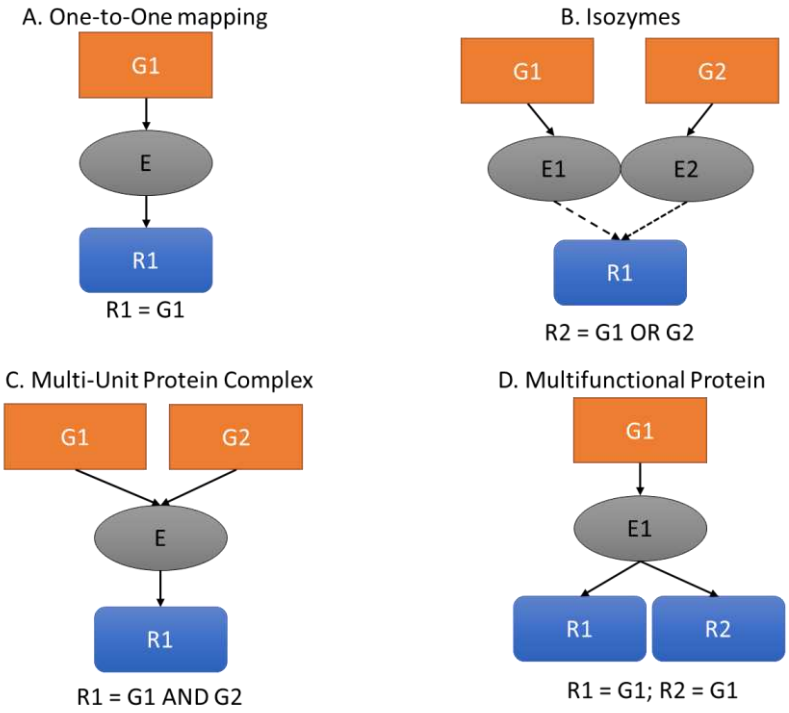


Figure 2.3: The four types of GPR associations. (a) One to One mapping where one gene expresses one single enzyme that catalyses one biochemical reaction (b) Isozymes that a reaction is catalysed by two different enzymes which are expressed by different genes and (c) Multiunit protein complex where a reaction is expressed by an enzyme that is co-expressed by two different genes and (d) Multifunctional protein, where a gene expresses an enzyme that can carry out multiple reactions (*39*).

GEMs can aid in better understanding the complex network of reactions that occur in every cell and to predict the phenotype under specified culture conditions. Due to the extended study of the microbial metabolism more and more GEMs are being reconstructed and published (*39*). Metabolic network reconstruction aims in creating a link between genotype and phenotype of a given organism. Today, with the ability to sequence and annotate genomes, it is possible to create metabolic networks at genome-scale level. However, the process of reconstructing and validating a GEM is a difficult and complex task. There are more than 25,000 completed genome sequences available in public databases (*71*) but only around 100 GEMs published for different organisms (*45*). This gap shows why creating a GEM is such a complex and difficult job. GEMs are built by systematically incorporating various types of omic data, as it is displayed in table 2.1.

Table 2.1: The various omics data that are used to formulate a genome-scale metabolic network (*39*).

Omics Data	*Application*
Metabolomics	Identifications and quantifications of the metabolites in an organism
Fluxomics	Is the study of the fluxome i.e., the total set of fluxes in the metabolic network of the cell
Lipidomics	The study of pathways and networks of lipids inside the cell.
Genomics	Study of the genome
Proteomics	Analysis of the entire protein complement, proteome
Transcriptomics	Degree of expression of each gene

High-throughput biochemical assays gather the different types of omic data, which then allow for a holistic view of the biological system and an in-depth study of the cellular metabolism. They study the transcriptional, translational, and posttranslational levels of the metabolite-enzyme interactions. The analysis of the omics data is performed using statistical methods and yield the required data for the reconstruction of the genome-scale model. Specifically, the reconstruction contains information about substrates, directionalities, stoichiometric coefficients of the reactions and about the compartments in which the reactions are categorized (*72, 73*).

Briefly, the reconstruction of a GEM requires four main steps. Initially, the draft reconstruction takes place, which is based on the genome annotation of the target organism and any other biochemical databases. It is an automated step that gathers information from certain databases, e.g. Kyoto Encyclopaedia of Genes and Genomes (KEGG) (73) or MetaCyc (74), to construct an initial version of the GEM. As the first step is mostly automated, there could be reactions which do not belong in the GEM or lack of co-factor specificity; thus the next step would be to manually curate the model followed by its conversion to a mathematical model. It includes the addition of bounds, definitions of the objective function and additional gap-filling to add missing reactions or enzymes in the model. Finally, the validation of the model and the inclusion of high-throughput data readies the model for publication and general use (*74*).

The main drawback of genome scale models is that they require a lot of experimental data to be constructed. All the omics from the table 2.1, are difficult and costly to obtain, so it is not a simple job to create a GEM for an organism. There are algorithms, as will be analysed in the next section that help in creating a cell line specific GEM, from a generic one for an organism, which uses less omic datasets. However, the number of experiments required for the first genome-scale of an organism, make it a difficult process (*75*). Another drawback the GEMs have, is the very large number of degrees of freedom. Usually, they contain thousands of reactions, and constrain only a very small number using

fluxomic data. Mathematical algorithms have been developed to reduce the degrees of freedom, and further constrain the GEM for mathematical optimisation, but they still cannot be relied upon for the accurate prediction of intracellular fluxes(*76*).

2.1.4 GEM Curation

GEM models, despite their detailed reconstruction and depiction of the metabolism, can be improved using a variety of different algorithms aiming in the curation of different missing pieces of the model. A plethora of GEMs are incomplete due to missing reactions or links, lack of realistic flux bounds and contain erroneous objective functions (*77*). These imperfections are depicted in the model as gaps in the network (dead-end metabolites and blocked reactions), false growth predictions (inaccurate biomass reactions), unrealistic flux predictions (large number of degrees of freedom), thermodynamically infeasible cycles (lack of Gibbs energy change for reactions) and objective function that does not represent the physiology of the organism. The process of fixing a GEM and modifying it from its original state to one that is specifically made for a set of data is called model curation (*78*). There are three main categories of algorithms for improving the performance of GEM model: Size reduction, Error curation and adjusting the intracellular constraints as seen in figure 2.4.

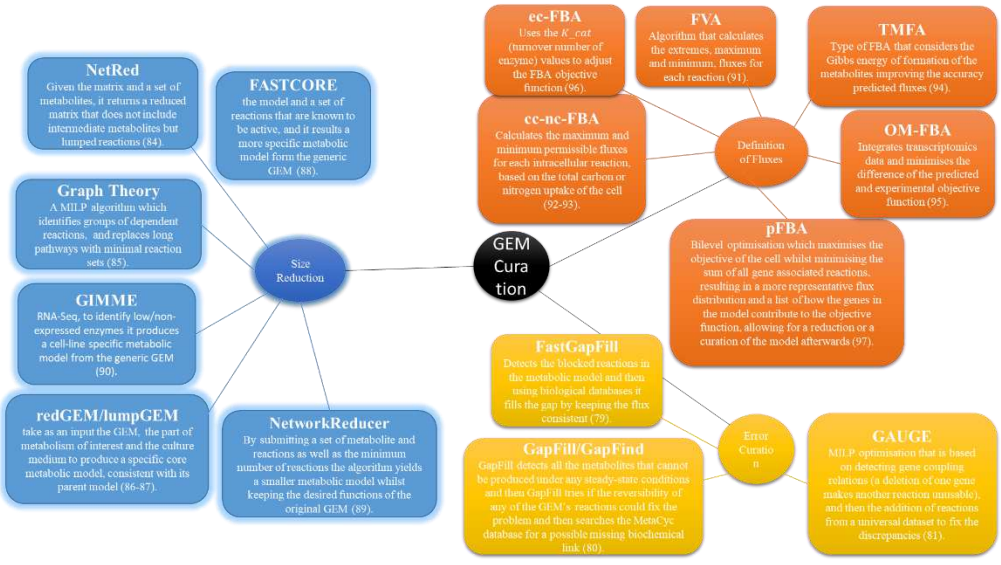


Figure 2.4: Algorithms used to improve the network and efficiency of genome-scale metabolic models.

The curation of the GEMs has three stages, error curation, size reduction and defining of flux distribution. A GEM that is taken from a database or is built with a bottom-up approach might contain mistakes in the form of missing reactions that create gaps in the network. Algorithms like FastGapFill

(*79*), GapFill/GapFind (*80*) or GAUGE (*81*) search the metabolic network of the GEM to detect potential blocked reactions and metabolites that cannot be consumed or produced in the model, and then using an automated search of biological databases the gaps get fixed. Most commonly used databases are the KEGG (*82*) or MetaCyc (*83*).

A common next step is the reduction of the GEM's size. Usually, these kinds of models, contain thousands of reactions which might not be ideal for certain application such as dynamic modelling, which requires a lot of kinetic equations, or C^{13} MFA that might require a lot of exponential data, or even for a simple FBA optimisation. For the size reduction there are two main categories of algorithms that can create a cell-line specific genome-scale model. Those that exploit the stoichiometry to eliminate metabolic intermediates and create lumped reaction such as NetRED (*84*), an algorithm that systematically reduces a metabolic network to a more compact form or a Graph theory method of reduction (*85*). More biologically focused reduction algorithms are redGEM (*86*), lumpGEM (*87*), FASTCORE (*88*) and NetworkReducer (*89*) which are algorithms that reduces a metabolic model focused around a metabolic module of interest or GIMME (*90*) which produces a cell-line specific metabolic model from the generic GEM. The choice of the reduction algorithms depends on the type of model, the objective, or the software of interest.

The size reduction of GEMs also reduces the number of degrees of freedom, but there is still a large solution space with a number of different possible flux distributions belonging in the ranges of each reaction bounds. There have been algorithms developed that algorithmically defined the upper and lower bounds of each reaction from the GEM. Initially, the reactions of the model have large abstract bounds, so the resulting flux destitutions might contain reactions with large unrealistic values as fluxes. Flux Variability Analysis (FVA) (*91*) is an algorithm to calculate the initial constraints for each reaction which can be further constrained using an elementary balance algorithms such as carbon constrain FBA (cc-FBA) (*92*) or nitrogen constrain FBA (nc-FBA) (*93*). Thermodynamics Based Metabolic Flux Analysis (TMFA) (*94*) provides a better understanding whether an enzyme can carry flux and captures the reversibility of the reactions. TMFA is theoretically a great way for the fluxes to always obey biochemical laws, and for the resulting fluxes of the FBA to achieve the minimum produced energy. However, there is a lack of thermodynamic data which decreases the predictive accuracy of the method, especially for large scale models. The resulting flux distribution can be further enhanced i.e., represent a more realistic phenotype by OM-FBA (*95*), ec-FBA (*96*) or Parsimonious FBA (pFBA) (*97*).

2.1.5 GEM Applications

Genome-scale models have been developed for a lot of organisms and have found applications in a plethora of scenarios. Generally, they are extremely useful in the contextualization of high-throughput data. They form a database for known omic data available for an organism, so they can assist in the definition and explanation of the metabolism of an organism. GEMs contain associations between genes

and enzymes, which are very useful for a mathematical optimisation. They can act as a guide for metabolic engineering i.e., the manipulation of genes to affect the yield of a target metabolite. Providing a holistic approach, GEMs can find a variety of applications, limited by the imagination of the researcher (*39*).

GEMs of bacterial organisms have been around the longest (*98*). Dozens of GEMs have been developed describing different strains on *E. coli (99)*, which have then been used in a variety of different applications. Some recent examples worth mentioning are the optimisation of shikimic acid production using the iJO1366 GEM (*100*), a study of the metabolic changes due to the pressure of different antibiotics (*101*) or to model the conversion of phosphosugars to organic acids with the GEM iML1515 (*102*). The GEMs of *E. coli* have also been extensively used for genetic engineering which will be analysed in a following paragraph. Moreover, the GEM of *Streptomyces lividans (103)* was used to increase the efficiency of its secretory pathway, and the GEM of *Acinetobacter baumannii* to study the metabolic responses to antibiotic treatment of the same bacterium. Yeast has also been the centre of many metabolic studies. The GEM of *Saccharomyces cerevisiae* was used to predict genetic engineering strategies that would increase its shikimic and muconic acid production (*104*). The model of *Pichia pastoris* was utilised for a metabolic study of the organism cultures under different oxygenation levels (*105*). The GEM of other types of organisms were also the topic of metabolic study. The essential metabolic pathways and genes of the parasite *Plasmodium berghei* were identified using a GEM as a template (*106*). The Refsum disease was analysed and studied, regarding the specific amino acid metabolic changes, using the human GEM (*107*). In conclusion, there have been numerous applications of GEMs in the study and optimisation of microbial organisms, but the main requirement is the availability of high-fidelity GEMs.

GEMs have recently been developed for CHO cell lines, where they found a variety of applications. The first published GEM for CHO cells is the iCHO1766 GEM (*108*) with 1,766 genes, 6,663 reactions and 4,456 metabolites. Until that time metabolic models for the CHO cell line were focused on a specific part of the metabolism (*109*) and were derived from the mouse GEM. The development of a GEM specific for CHO made strain and bioprocess optimisations in CHO a possibility. The model includes the metabolism of all amino acids, the basic metabolic pathways, lipid, and glycan biosynthesis and contains protein production as well as producing and non-producing biomass equations. From the generic iCHO1766 three cell-line specific models, CHO-K1, CHO-S, CHO-DG44, were derived using RNA-sequencing and proteomic data.

The model of CHO cells was expanded and improved to more complete GEMs. The model was recently update by filling in the metabolic gaps using the GapFill algorithm, an optimisation algorithm with the objective of minimising the gaps by filling the missing reactions from databases, GAUGE and biological databases such as HMDB, KEGG and BiGG. The study resulted in iCHO2101 (*77*) with 300 more genes and 20% less blocked reactions. iCHO1766 was also expanded to iCHO2291 (*96*) using a similar method of removing blocked reactions and adding genes and metabolites from databases. The

iCHO1766 model has also been expanded to include the secretory pathway, enabling the modelling of protein secretion and biosynthesis as well as the creation of product-specific models (*110*). It is important for the GEM to expand and include more enzymes and genes, so the specific cell-line models derived would be as complete and accurate as they can be.

The GEM of CHO has been applied on the field of bioprocess optimisation as well as used for studying of the metabolism (*40*). A few examples are the use of the DG-44 model in an industrial process of mAb production to predict the cell metabolism under different culture conditions. Using experimental data from four independent batch cultures, the model was able to predict the high producing cell lines and intracellular metabolic shifts (*111*). The three models of the different CHO cell lines, K1, DG-44 and S have been used to validate an alternative FBA objective function and minimal amino acid uptake whilst keeping growth above the required minimum and distinguish the metabolic differences between the three cell lineages (*49*). Using the secretome model, Gutierrez et al. were able to estimate the cost of producing various proteins in the CHO cell and pinpoint limiting factors for their secretion (*110*). Using ribosomal profiling data with the secretome model an important conclusion was made; that CHO cells do not secrete energetically unfavourable proteins and the removal of the genes that express these proteins can make the product purification easier (*110*). The theory was experimentally validated by first quantifying the bioenergetic investment of the CHO cells into producing the HCPs, then identifying target HCPs which can be removed from the genome and finally to removing 14 HCP using genome editing (*112*). Another important application of the iCHO1766 GEM is its application in the design of media that increased the antibody production in CHO cells (*113*). Even though it is almost four years since published, the GEM has not yet been utilised as a mean for strain engineering studies.

2.2 Optimisation of CHO cell culture

The bioprocess of mAb production in CHO cell systems has been the subject of many optimisation studies. There are many potential ways to optimise a bioprocess. The production of mAbs in CHO expression systems can be optimised by targeting the increase of antibody productivity, increase of mAb titer, the improvement of mAbs quality regarding its glycosylation profile, the optimisation of the purification stage or with better coupling of the upstream and downstream process. The topic of optimisation herein is the increase of specific productivity of mAbs per CHO cell in a culture. The increase of productivity can increase the total titer of a cell culture, which affects the duration needed by the upstream process to produce a specific quantity of IgG and the cost of the process. An even slight increase in the final mAb titer, can result in a reduced cost of the product of interest. For that reason, there have been many attempts, targeting at different parts of the bioprocess, to try and increase the specific productivity of mAb in a CHO culture.

The improvement of IgG production in a CHO culture has been achieved by optimising different variables of the culture as well as the cell line itself (*26, 114, 115*). Media optimisation is a key factor for the increase of titer. Addition of small molecule compounds in the media has shown significant increases in protein titer (*116, 117*), removal of metabolites from basal media that inhibit the late-phase culture efficiency improved the titer even in a 500L bioreactor (*118*) and *in-silico* media optimisation to define an optimal media composition has also shown positive results regarding the mAb productivity in CHO fed-batch cultures (*119, 120*). Apart from the culture medium, feeding has also been the subject of optimisation. Fed-batch culture can have the feeding composition optimised by analysing with FBA the uptake and secretion rates of metabolites by their exometabolomic concentrations (*121, 122*) and determining the amino acid requirements in each culture phase. The above has been augmented to guide the amino acid feeding as a controller (*123*). Similarly, the optimal time points of feeding have been studied, with the aim of maximising product productivity (*124, 125*). In a CHO culture, the environmental conditions such as the bioreactor conditions (*126*), or the culture temperature (*127*) have shown important effects on the resulting IgG titer. The CHO cell line that produces the product has also been the subject of optimisation. The increase of cell viability through gene engineering (*128*) or the enchantment of the productivity through random gene silencing (*34*). An example of an engineered high-producing CHO cell line, which is widely used in the biopharma sector, is the creation of the GS-CHO cell line. The accumulation of ammonia during the CHO cell culture is known to negatively affecting the growth and product quality (*129*). Glutamine is a major precursor to ammonia production. Thus, CHO cells overexpressing glutamine synthetase (GS), an enzyme that catalyses glutamate to glutamine and reduces the ammonia in the culture medium (*130, 131*).

2.2.1 Strain Optimisation

A way to optimise the production of a biochemical from a bioprocess is by using strain design (*132-135*). Strain design algorithms can be used when a cell has two competitive objective functions and the objective is to change their relationship of the two objectives. An FBA optimisation that maximises one of the two objective functions usually returns zero flux on the other. The coupling of the two objectives would make the FBA optimisation results positive fluxes in both the competitive objective functions, making one function dependent on the other. It encompasses an *in-silico* optimisation of the metabolic network, which results genetic modifications. The result can be tested by engineering the under-study strain *in vitro (136)*. The *in-silico* optimisation predicts several enzymes or genes with a suggested modification strategy for each one, such as knock out, overexpression. Their application *in vitro* could change the flux distribution of the cellular metabolism to one that achieve the objective of the strain optimisation. Most of the strain design algorithms are based on the optimisation technique of bi-level optimisation, where one objective function if maximised of minimised over a set of constraints, and one of the constraints is the maximisation of minimisation of a secondary objective. To solve these bi-level

optimisations, the technique of mixed-integer linear programming (MILP) optimisation is implemented. The key challenges for this type of optimisation technique include the identifications of the genes that can be modified in the genome and the difficulty in performing the genetic modifications *in vitro*.

2.2.2 Strain Design Algorithms

Various strain design algorithms have been developed, each achieving a different goal using a unique methodology. Their main goal is the detection of genes or their associated biochemical reactions whose regulation results in the coupling of two diametrical objective functions in an FBA optimisation. The holistic nature of GEMs makes them ideal candidates for algorithms, as they are able to provide a variety of different combinations of genes and reactions. The genes express enzymes that allow biochemical reactions to take place which form a metabolic network of reactions. A change in the network, which corresponds to a change in the metabolic activity of the cell, can modify the predicted fluxes to achieve the phenotype of choice. Afterwards, the proposed regulations can be tested *in vitro* and verify or reject the optimisation result. A list of strain design algorithms is presented in figure 2.5.

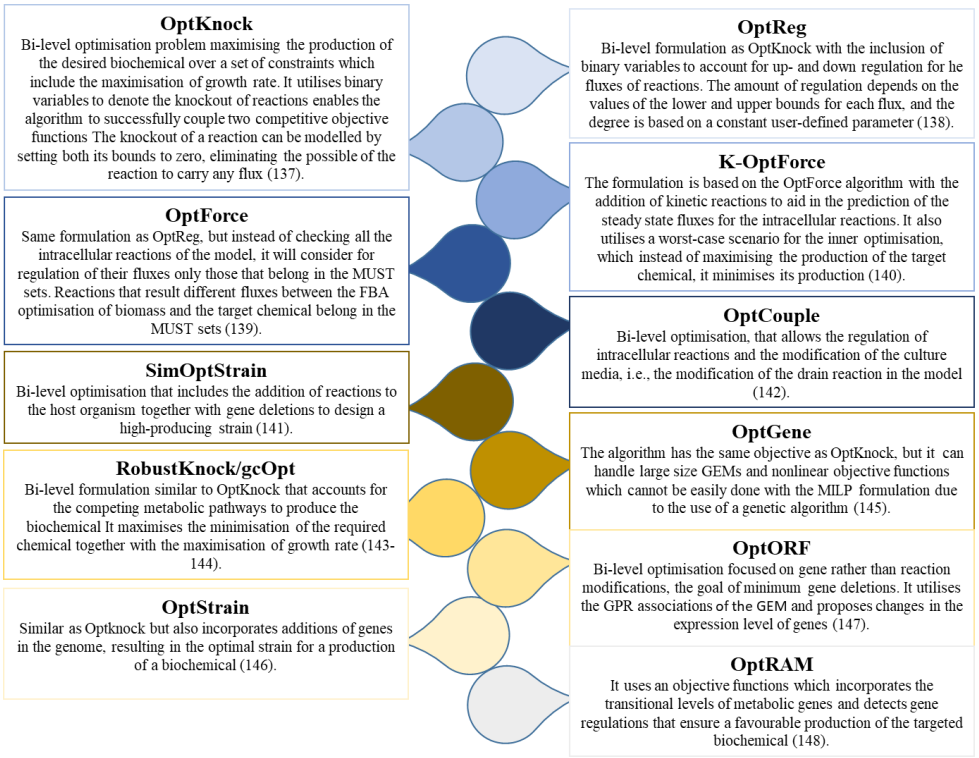


Figure 2.5: Algorithms used in strain design for the overproduction of metabolites.

OptKnock (*137*), from Costas Maranas group, was the first strain design algorithm developed, which detects reaction knockouts, both lower and upper bounds of the reaction flux are set to 0, to increase the production of a biochemical. However, it only detects reaction knockouts from the

metabolic network, and not up- or downregulation of the reaction's flux. OptReg (*138*), OptForce (*139*) and k-OptForce (*140*) include the up-, when the lower bound of the reaction flux increases forcing the reaction to carry more flux in the FBA optimisation, and downregulation of intracellular reactions, when the value of the upper bound reduces. SimOptStrain (*141*) and OptCouple (*142*) are also allowing the addition of reactions and metabolites to the model, in order to couple two competitive objective functions. The two optimisation algorithms of RobustKnock (*143*) and gcOpt (*144*) try to account for the competing pathways that may divert flux from the product of interest, making the suggested genetic strategies more likely to be experimentally validated. The experimental application of the results from one of the above optimisation algorithms would include genetic modification, so the reactions that are proposed should be reversely translated to genes' regulations using the GPR associations of the GEM or by searching in the literature and identifying the genes that express the reactions that are to be knocked out. The algorithms OptGene (*145*), OptStrain (*146*), OptORF (*147*) and OptRAM (*148*) result in genetic modifications. The choice of the algorithm is a function of the purpose of the optimisation and the usability on the metabolic model of the under-study organism.

The strain design algorithms have seen a number of applications on microbial strains, mainly in yeast and bacteria (*99*). *Escherichia coli* has been engineered using OptForce to increase the production of fatty acids (*149*), to increase its octanoic acid production (*150*), its malonyl-CoA production (*151*) and to design glycan production (*152*). Similarly, *Saccharomyces cerevisiae* was engineered by OptForce to increase the aromatic amino acid biosynthesis (*153*) and terpene pathway improvement (*154*). The OptRAM algorithm was applied successfully in yeast to create strains with increased 2,3-butanediol and ethanol (*148*). However, due to the size and complexity of the metabolic network of CHO cells of CHO cells compared to microbial strains, an optimization algorithm in CHO has not yet been applied.

2.3 Aims and Objectives

The thesis is trying to increase the specific productivity of mAbs in a CHO fed-batch culture. In the section **1.1**, the importance of mAbs was analysed as a biopharmaceutical product, and their bottlenecks and main issues in the production were explained. The high production cost can be dealt with an increase of productivity, which can reduce the required running time of a fed-batch to produce the required amount of mAb or reduce the amount of volume needed in a fed-batch culture of mAb producing CHO cells. There have been many attempts to increase the productivity in CHO, as seen in Section **2.2**, but there has not been a successful attempt to apply a strain engineering algorithm on the CHO cell line. The strain design utilises the GEM of the CHO cell, and the FBA optimisation technique to predict the intracellular fluxes. They were chosen against a kinetic model of a small-scale metabolic model, because of the holistic approach that GEMs offer and the lack of kinetic parameters of the FBA optimisation, seen in the table 2.2 below.

Table 2.2: Advantages and disadvantages of FBA and kinetic modelling for the mathematical representation of biological system.

FBA		**Kinetic modelling**	
Advantages	**Drawbacks**	**Advantages**	**Drawbacks**
Use of GEM holistic behaviour	Number of degrees of freedom	Detailed representation	Difficult to model GEM
Representation of GPRs	Lack of dynamic representation	Detailed description of a certain part of the metabolism.	Difficulty in modelling gene-expression
There is no need for parametrisation	Unsuitable for quantitative predictions	Implementation of bioreactor conditions	

A combined computational and experimental methodology were applied to optimise the bioprocess, and the steps are analysed below.

1. Initially, the GEM for the CHO-K1 cell line was optimised using a strain design algorithm. The reactions of the GEM were constrained using exometabolic datasets and elementary balance algorithms. The network of the GEM was then optimised using a bi-level strain algorithm, with the two objectives of growth rate and antibody secretion maximisation.
2. Subsequently, the analysis of the strain optimisation results was performed in two parts. The number of the proposed regulations was reduced to the minimum ones required for antibody secretion in a biomass maximisation FBA, disregarding any unnecessary regulation. Then, the scenarios were analysed, to try and come up with a biological reason of why there positive flux of antibody secreted from the GEM when biomass is maximised. Genetic engineering strategy was proposed for three of the scenarios, which can be implemented *in vitro*, to try and achieve higher IgG specific productivities in CHO cultures.
3. Finally, the scenarios that were calculated and analysed were tested *in vitro*, to try and complete the optimisation process. Using genetic engineering methods, two cell lines were engineered with the goal of increased mAb specific productivity, compared to the unmodified CHO cell line.

3

Computational Methods

The third chapter aims to explain the computational methods used for the modelling of the CHO cellular metabolism and the *in-silico* optimisation part of the thesis. Initially, a description for the genome-scale model of the CHO cell is described, which is then used for the FBA prediction of growth rate and antibody production. The technique of FBA is analysed and the relevant software packages are presented. Concerning the strain optimisation of the fourth chapter, the theory behind bi-level optimisation is explained for a better understand of the optimisation methodology and how the optimal solutions are calculated. Then, the optimisation algorithm OptReg is analysed together with its conversion to MILP from bi-level. Finally, the development of the Sensitivity Analysis framework that is applied on the fifth chapter is explained.

3.1 CHO-K1 GEM

The GEM of the CHO-K1 cell line was used to model the metabolic network of the cell. As explained in section **2.1.3** of the literature review, GEMs are mathematical representation of an organism's metabolism. They contain the genes, proteins and metabolites that take part in the metabolism of the organism. The newest GEM at the time of performing the optimisation was the iCHO1766 (*108*). It is a GEM that contains most known genes and their relevant biochemical reactions for the CHO cell. However, the iCHO1766 is a generic GEM, i.e., it contains information from three different cell line specific GEMs, the CHO-K1, CHO-S and CHO-DG44. The CHO cell line which is under study herein is the CHO-K1, so the use of iCHO1766 would include degrees of freedom in the optimisation that should not be included in the mathematical model, as it can be seen in figure 3.1. For that reason, the GEM used to represent the CHO cellular metabolism is the CHO-K1 GEM.

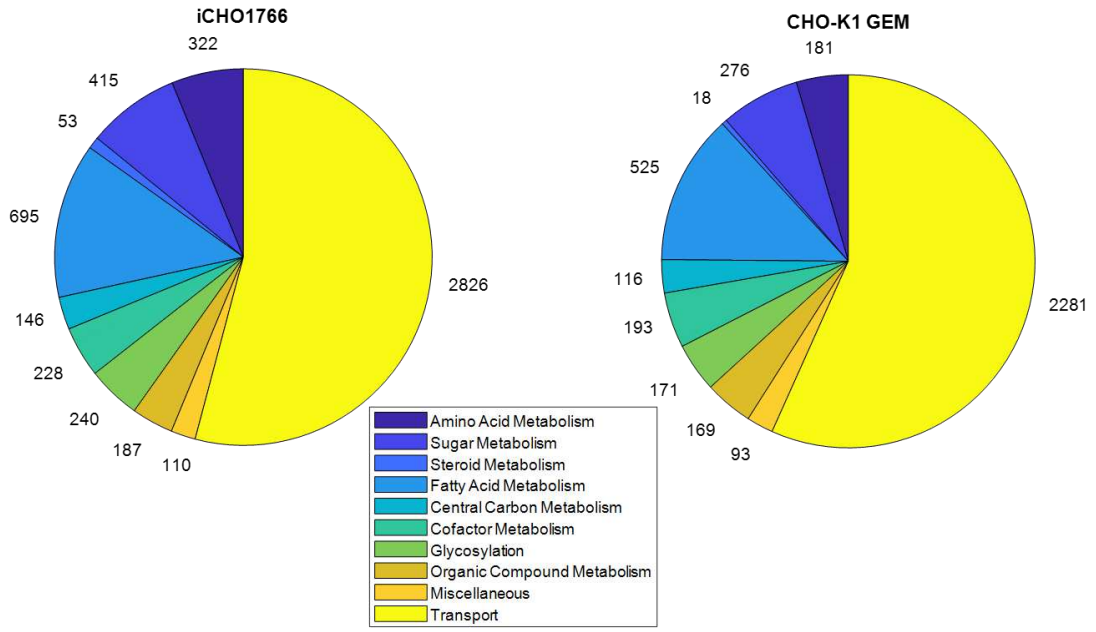


Figure 3.1: Generic GEM of the CHO cell organism (left), and GEM of the CHO-K1 CHO cell line (right). The numbers in the pie charts represent the number of reactions in each of the different parts of metabolism.

3.2 FBA Optimisation

The method of choice for calculating and analysing the fluxes from the metabolic network is FBA. It is a linear programming optimisation method. FBA is based on the assumption of pseudo-steady state and performs an optimisation on the fluxes of the metabolic network. A metabolic network consists of the metabolites and the reactions that take place in an organism (Fig. 3.2A). The goal of the FBA is to calculate in a pseudo steady state the rates of conversion or transport in the biochemical reactions, which are known as fluxes. A pseudo-steady state means there is no accumulation in the internal nodes of the metabolic network, and a linear optimisation can calculate all of the network's fluxes (Fig 3.2B). The assumption of pseudo-steady state is correct under most conditions and also during a batch or a fed-batch culture because of the slow changes in the system's behaviour; only when an abrupt change takes place the assumption does not hold (*155*). It is important to be noted that the pseudo-steady state takes place locally and not throughout the culture period, e.g., only during the exponential or the stationary phase of the culture depending on the purpose of the model. Mathematically, a steady state means that the sum of the fluxes of every metabolite should be equal to zero resulting in a system of linear equations (Fig. 3.2C). Normally, the system of equations would be easy to solve however, in a fully determined stoichiometric matrix, S, there are more reactions, r, than metabolites, m, and the matrix has more columns than rows. The matrix S contains the coefficients of the reactions which are the main constraints of the model, and they impose balance constraints on the system ensuring that the total

amount of flux being produced is the same as the one being consumed, a simple mass balance. The multiplication of the matrix with the fluxes *v*, in pseudo-steady state condition, should be equal to 0

$$S \cdot v = 0 \tag{3.1}$$

This results in an under defined system with *r-m* degrees of freedom, which contains numerous possible solutions and a large solution space. Thus, an optimisation problem is created which requires the definition of upper bounds, *UB,* and lower bounds, *LB*, for the fluxes, *v*, and an objective function. The flux of a reaction, *i*, is then being constrained between the two bounds as:

$$LB_i \leq v_i \leq UB_i \tag{3.2}$$

The resulting flux of the reaction *i* will be between LB_i and UB_i, so these two parameters need to have a realistic value since the resulting flux distribution represents a natural phenomenon and should obey biological principles. Last step is the definition of an objective function, which is a mathematical description of the desired property best representing the biological system. It has the form of a vector with coefficients for every metabolite required to generate biomass content. In metabolic modelling, the objective function is an equation defining the phenotype e.g., the visible character of the individual or more specifically an equation instructing which variable(s) to maximize or minimize (Fig. 3.2D) (*48*). The overall methodology behind an FBA optimisation is presented in figure 3.2.

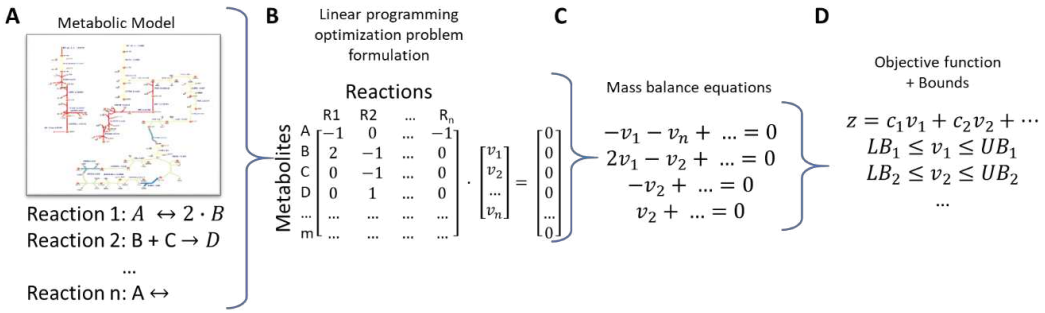


Figure 3.2: The four main steps of an FBA optimisation. Initially, the metabolic model which contains the metabolites and reactions of the organism is formed. The metabolic network is converted to a stoichiometric matrix. The size and its contents are dependent on the objectives of the study. The pseudo-steady state assumption states that all mass balances of the network should be zero e.g., there is no accumulation and the multiplications of the fluxes with each reaction will be equal to 0. The system of linear equations formed is then solved. Any metabolic network contains more reactions than metabolites, so the number of degrees of freedom is greater than 0, so instead of a unique solution there are several possible optimal solution. They are the result of the choice of bound, LB and UB, of each flux and the choice of the objective function, z (*48*).

An FBA optimisation is linear and can be solved in various ways. The most common and user-friendly one is using the Constraint-Based Reconstruction and Analysis (COBRA) toolbox (*156*). It is an open-source MATLAB software suite that provides the algorithms for solving an FBA optimisation problem, creating a genome-scale model from the reaction and metabolites, defining GPRs and lower and upper bounds. It is also useful for the analyses of the resulting flux distribution as well as the ability of implementing custom functions. The majority of FBA optimisation and fluxomic analysis in this study took place in COBRA. The metabolic networks are visualised using the Escher python toolbox (*157*), which is an open-source toolbox that allows the simple visualisation of the metabolic networks together with a representation of the fluxes.

3.3 Strain Optimisation Algorithm

The algorithm used for strain optimisation is bi-level, which was converted to MILP. For that reason, the most important theory behind MILP and bi-level optimisation are presented, before the analysis of the strain optimisation algorithm.

3.3.1 MILP Optimisation

In the FBA section, the optimisation technique of linear programming (LP) was briefly discussed. The canonical form of an LP problem is represented as

$$\max z = c^T x \tag{3.3}$$

$$s.t. Ax \geq b \tag{3.4}$$

$$x \geq 0 \tag{3.5}$$

Where, x is a Nx1 vector of variables, c is a Nx1 vector, also called cost vector for a minimisation problem or profit vector for a maximisation problem. Each element represents the cost or profit for each variable of x, A is an MxN matrix, known as the matrix of coefficients; each of the elements is the coefficient of the corresponding variable x. b is an Mx1 vector, also called the right-hand side vector. It is the simplest and quickest optimisation method, as all terms are linear, and all variables are continuous. The LP problem with *N* number of variables forms a *N-1* dimension hyperplane, and the optimal solution exists in one of vertices, meaning there could be a very large number of optimal solutions (*158*).

An optimisation technique developed from LP is Mixed-Integer Linear Programming (MILP). MILP follows the same formulation as LP but with the inclusion of binary variables. Binary variables

are variables which are restricted to finite set of values, a common set is to take only the values zero or one. MILP has a wide range of applications, one of which is the optimisation of strain design. Strain design algorithms need to detect specific reactions in the metabolic network, so the binary variables act as a decision variable for each reaction, deciding if the reaction should be changed or not from the original metabolic network. An MILP mathematical representation is similar to LP with the addition of a few more terms to represent the binary variables.

$$\max z = c^T x + y \tag{3.6}$$

$$s.t. Ax + By \geq b \tag{3.7}$$

$$x \geq 0 \tag{3.8}$$

$$y \in \{0,1\} \tag{3.9}$$

Where, y is a vector of binary 0 – 1 variable, B is an MxN matrix, called the matrix of coefficients; each of the elements is the coefficient of the corresponding variable y.

The MILP formulation is the same as in LP, but with the addition of binary variables, such as *y* in the example. In these kinds of optimisations, *y*, represents choices and the objective is to determine those that maximise the objective function. The combinatorial nature of y can lead to a huge number of different LPs to solve, for instance only 100 binary variables can lead to 2^{100} different combinations, resulting in the need to solve 2^{100} different LP problems, a computationally very expensive task. There is a variety of different approaches to solve MILP problems and they belong in four main categories. The **logic-based** method which uses disjunctions, if-statements, and logical equations to simplify the model. The **decomposition** method exploits the mathematical structure of the model with the use of variable partitioning, duality and relaxation methods to reduce the large model into smaller ones. The **cutting plane** method, where instead of new subdomains in the feasible region, for each value [0, 1] from the binary variables, new constraints are generated reducing the solution space until an optimal 0-1 solution is found. Finally, the **branch and bound** method, which is the most common method to solve MILP. Branch and bound consists of three main ideas separation, relaxation, and fathoming. The MILP problem (P) can be reduced to a set of sub problems (P_1), (P_2), … (P_n), where a feasible solution of the sub problems is also a solution of P and vice versa, by setting the binary variables to 0 or 1. Figuring out all the possible sub problems requires a huge amount of memory, so, faster ways have been developed such as setting additional constraints just for the binaries. The relaxation of (P) is denoted as (RP); where the set of feasible solutions of (P) is a subset of the feasible solutions of (RP). An important remark is that if the optimal solution of (RP) is feasible in (P) then it is also optimal. Lastly, fathoming in branch and bound is the step where the candidate sub problems of (P) are considered as optimal solutions (*159*).

3.3.2 Bi-level Optimisation

In the canonical formulation of LP and MILP there is only one objective function. However, a lot of real-life problems can be mathematically represented with more than one objective function. Modelling multiple objective functions can be done with either multi-objective or bi-level optimisation. In multi-objective optimisation weighted objective functions can be applied, which are objective functions that incorporate multiple variables multiplied by coefficients representing how important each variable is, or the methodology of ε-constraint can be applied, which maximises one set of variables but applies constraints to the other objectives of the multi objective formulation. In multi-objective optimisation however, the solution is a point in the Pareto front. Strain design algorithms aims in changing the Pareto front, which can be accomplished using bi-level optimisation. A bilevel optimisation, on the other hand, optimises one objective over a set of constraints, which contains a secondary objective function leading to an optimal point. The optimisation results in a complete reformation of the Pareto curve. In bi-level optimisation all variables are partitioned between two optimisations, the inner and outer, and the outer are chosen based on the results from the inner, similar to hierarchical optimisation. The formulation of a bi-level optimisation is

$$\left\{\begin{array}{l}\text{Outer Optimisation}\\ \max z = c1^T x\\ \text{Inner Optimisation}\\ \left\{\begin{array}{l}\max w = c2^T x\\ s.t.\ Ax \geq b\\ x \geq 0\end{array}\right.\end{array}\right.$$

Bi-level problems consist of an inner – upper level - and outer – lower level – optimisation. In the canonical formulation above, inner is the maximisation of *w* and outer the maximisation of *z*. Bi-level optimisation problems are difficult to solve and are not straight-forward as LP or MILP. Extended Mathematical Programming (EMP) is an extension of algebraic modelling languages – programming languages like General Algebraic Modelling System (GAMS) used to describe mathematical optimisation problems – that can be used for hierarchical optimisation i.e., bilevel where using automatic reformulation the nonlinear optimisation problems becomes linear. A common solution method, especially when the lower optimisation problem is convex, is the reformulation to a MILP problem using the Karush–Kuhn–Tucker (KKT) conditions or the strong duality theorem a problem for which robust solution algorithms exist (*136*).

The use of KKT or strong duality is mainly dependant on the presence of nonlinear terms in the objective function. A linear objective function in a bilevel optimisation problem, can use the strong duality theorem, for the reformulation to a MILP problem. The explanation of the strong duality theorem requires first the definition of the dual (D) of an optimization problem (P). Any LP problem [P] with *N*

variables and *M* constraints is associated with another LP [D], defined as dual, with *M* variables and *N* constraints. Its maximisations become minimisations and the inequalities are reversed. The relationship between dual and primal problems holds a very important theorem. The **Weak Duality Theorem** states that if *x* and *y* are the feasible solutions of the primal and the dual problems, respectively, *x* is always greater or equal than *y*. The difference between *x* and *y* is called duality gap. As a result of the weak duality theorem, the **Strong Duality Theorem** states that if the primal or the dual has a finite solution, then so does the other one and their solutions are equal. The strong duality holds if and only if the Slater conditions hold. They state that the problem [P] must be convex, and the problem is strictly feasible i.e., the optimal solution satisfies any nonlinear constraints using inequalities. Then the duality gap is 0, and the two objectives of [P] and [D] are equal. If there is nonlinearity in the objective function, then the use of the three KKT optimality conditions is required, to formulate the bilevel optimisation to a MILP one (*158, 159*).

3.3.3 OptReg

The OptReg algorithm is used to identify different combinations of regulations for the lower and upper bound of the reactions that belong in the CHO-K1 GEM. The objective of the optimisation is to change the natural allocation of resources in the cell. These regulations are classified as upregulations, when the lower bound of a reaction increases forcing the reaction to carry more flux, downregulations, when the upper bound of a reaction decreases or a knockout, when the reaction must carry zero flux. In the original CHO-K1 GEM, when performing an FBA optimisation with objective to maximise the biomass production there is zero flux of secreted antibody as explained in the previous Section **3.3.1**. The changes in the bounds of the reactions suggested for a specific dataset ensures that the FBA optimisation will allocate resources in the synthesis of IgG which will be secreted as a by-product, while the CHO-K1 GEM is being constrained under the same dataset. If the regulations resulted by the strain optimisation on a specific dataset were to be performed on a different dataset, they might not have same result due to the different uptake and secretion rates in each dataset. The optimisation algorithm is the same as in the publication (*138*) with the addition of a few changes that are analysed below.

OptReg is a bilevel optimisation algorithm that maximises both the production of a targeted biochemical and the growth rate, while adjusting the upper and lower bounds of the reactions as seen in the mathematical formulation of figure 3.3.

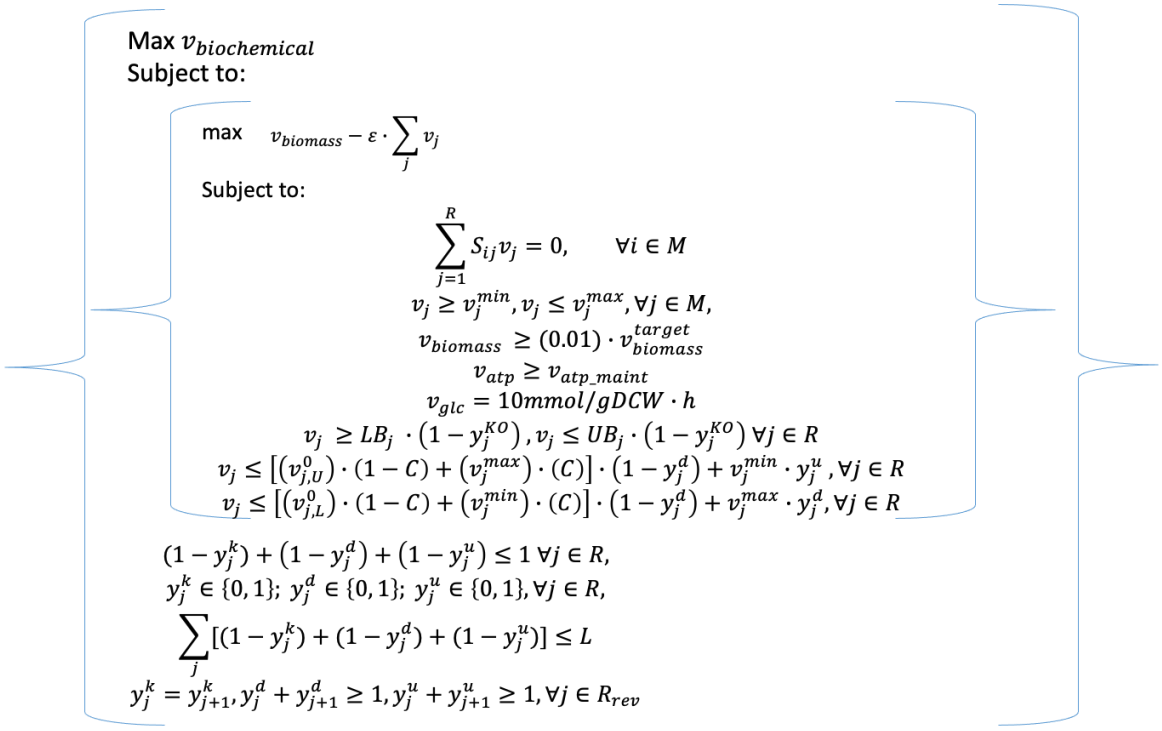


Figure 3.3: Mathematical formulation of the OptReg optimisation, as presented in the original paper of Pharkya and Maranas 2006 (*138*).

The objective function of the outer optimisation has the objective to maximise the flux $v_{biochemical}$, which can represent a biochemical produced from the metabolic network. In the case of the CHO-K1 it represents the flux of IgG secretion. In the outer optimisation there are three additional constraints that take into consideration the binary variables y_i^k which takes the value of 0 when the reaction *j* must have both the lower and the upper bound equal to 0 (knockout), y_i^u which has the value of 0 when the reaction *j* must have the lower bound increased (upregulation) and y_i^d which has the value of 0 when the reaction *j* must have the upper bound decreased (downregulation). A biochemical reaction is not biologically feasible to carry more than one regulation. Mathematically the constrain is represented as

$$(1 - y_j^k) + (1 - y_j^d) + (1 - y_j^u) \leq 1 \; \forall j \in M \tag{3.10}$$

In order to limit the total number of regulations, the user-defined parameter L is used

$$\sum_j [(1 - y_j^k) + (1 - y_j^d) + (1 - y_j^u)] \leq L \tag{3.11}$$

The optimisation uses irreversible reactions. All the reversible reactions, *j*, of the model are converted to their forward, *j*, and backward, *j+1*, reactions, where only one direction can carry a modification

$$y_j^k = y_{j+1}^k, y_j^d + y_{j+1}^d \geq 1, y_j^u + y_{j+1}^u \geq 1, \forall j \in M_{rev} \quad (3.12)$$

In the inner optimisation, the objective function states to maximise $v_{biomass} - \varepsilon \cdot \sum_j v_j$. The first term is the growth rate, the second term tries to minimise flux cycles between the forward and the backward part of a reaction *j* by minimising the total sum of the fluxes, and not allowing the creation of unreasonably large fluxes. The two first constraints formulate the FBA optimisation,

$$\sum_{j=1}^{M} S_{ij} v_j = 0, \ \forall i \in N \quad (3.13)$$

$$v_j \geq v_j^{min}, \ v_j \leq v_j^{max}, \ \forall j \in M \quad (3.14)$$

Where, v_j^{min} and v_j^{max} are calculated from an FVA optimisation.

$$v_{biomass} \geq (0.01) \cdot v_{biomass}^{target} \quad (3.15)$$

$$v_{atp} \geq v_{atp_maint} \quad (3.16)$$

The above two constraints force biomass production to be higher than 1% of the required growth rate and ATP synthesis to be higher than that required for maintenance. The next three constraints adjust the bounds of the reactions, by taking into consideration KO, upregulation, and downregulations.

$$v_j \geq LB_j \cdot \left(1 - y_j^{KO}\right), v_j \leq UB_j \cdot \left(1 - y_j^{KO}\right) \forall j \in J \quad (3.17)$$

$$v_j \geq \left[\left(v_{j,U}^0\right) \cdot (1 - C) + \left(v_j^{max}\right) \cdot (C)\right] \cdot \left(1 - y_j^d\right) + v_j^{min} \cdot y_j^u, \forall j \in M \quad (3.18)$$

$$v_j \leq \left[\left(v_{j,L}^0\right) \cdot (1 - C) + \left(v_j^{min}\right) \cdot (C)\right] \cdot \left(1 - y_j^d\right) + v_j^{max} \cdot y_j^d, \forall j \in M \quad (3.19)$$

Where $v_{j,U}^0$, $v_{j,L}^0$ are the maximum and minimum flux for the reaction j, C is a regulation strength parameter, v_j^{max}, v_j^{min} are the maximum and minimum of each flux subject to the substrate conditions. The bounds of the fluxes change based on the values of the binary variables. For example, if $y_j^{KO} = 0$, then $0 \leq v_j \leq 0$, forcing the flux of reaction j to be equal to 0. If $y_j^u = 0$ then the upper bound of the reaction *j* would transform into $\left(v_{j,U}^0\right) \cdot (1 - C) + \left(v_j^{max}\right) \cdot (C)$ and if $y_j^d = 0$ then the lower bound would transform to $\left(v_{j,U}^0\right) \cdot (1 - C) + \left(v_j^{max}\right) \cdot (C)$. The above mathematical representation of up- and downregulation of reactions shows the important of narrow bounds for the reaction fluxes. If the bounds of the reaction are significantly large e.g., [0, 100], then an upregulation of just 10% would increase the lower bound by 10, a very large value compared to the biomass production which usually is in the order of magnitude of 1E-2.

In the original algorithm, the parameter C has the same value for both upregulation and downregulation which could hinder the detection of optimal solutions. Thus, the first alteration was to adjust the upregulation and downregulation coefficients to not necessarily be the same value. They still were discrete variables but part of a user-defined range of values. Creating continuous variables as coefficients could be a possible solution, however, the non-linearity of the optimisation would impose a difficulty on solving the problem. Another change from the original algorithm is the usage of elemental bounds, instead of the maximum and minimum ones for each reaction. That way, the intracellular fluxes obey the elemental balance and the proposed up- and downregulation strategies are likely to be more accurate. Based on these two changes, the upregulation and downregulation constraints become:

$$v_j \geq \left(LB_j + \left(UB_j - LB_j\right) \cdot e_{up}\right) \cdot y_j^{up} + LB_j \cdot \left(1 - y_j^{up}\right) \quad (3.20)$$

$$v_j \leq \left(UB_j + \left(UB_j - LB_j\right) \cdot e_{down}\right) \cdot y_j^{down} + UB_j \cdot \left(1 - y_j^{down}\right) \quad (3.21)$$

Where LB_j and UB_j are the lower and upper bounds calculated from the elemental balance, respectively, and e_{up} and e_{down} are the up- and downregulation coefficients, respectively. The constraint responsible for reaction knockout remains the same. As seen in the above equations for up- and downregulation, the new bound is a function of the old, which is the reason an FVA was applied beforehand for curating the datasets.

Another change was the addition of two constraints in order to make sure flux is channelled only through either the forward or the backward reaction and eliminate the flux transport between forward and backward reactions. In order to force either only the forward or backward part of a reversible reaction to carry flux a new binary variable, y_j^{rev}, was introduced, which has the value of 0 when a reaction carries flux and 1 when it does not. The reactions derived from the reversible *j* that are forward or backward are known, so the sum y_j^{rev} should be 1 ensuring at no more than one of them is allowed to carry flux.

$$y_{j,forward}^{rev} + y_{j,backward}^{rev} = 1 \quad (3.22)$$

To enforce the flux constraint, the flux of a reaction *j* should be less or equal than the upper bound of *j* multiplied by y_j^{rev} resulting in 0 when y_j^{rev} is 0 and the value of the upper bound when y_j^{rev} is 1.

$$v_j \leq UB_j \cdot \left(1 - y_j^{rev}\right) \quad (3.23)$$

For example, if $y^{rev}_{j,forward} = 1$, then the forward reaction would have flux $v_j \leq 0$ and the backward $y^{rev}_{j,forward} = 0$ could carry flux since $v_{j,backward} \leq UB_j$. To separate the two sets j for $y^{rev}_{j,forward}$ and j for $y^{rev}_{j,backward}$ the command *alias* in GAMS was used to create a second j set of the reactions and match the forward and backward reactions. These two constraints provide a deterministic way to limit the flux cycles between forward and backward parts of the same reaction j, which made the use of the term $-\varepsilon \cdot \sum_j v_j$ in the objective function unnecessary. The strain design algorithm, OptReg, with all the aforementioned changes is named **Bio_Igg,** and its complete mathematical formulation is depicted in Figure 3.4.

$$\text{maximise } z = v_{IgG} \qquad (1)$$

Subject to:

$$\text{maximise } z = v_{biomass} \qquad (2)$$

Subject to:

$$\sum_{j=1}^{R} S_{ij} v_j = 0, \forall i \in M \qquad (3)$$

$$v_{biomass} \geq 10\% \cdot v^{max}_{biomass} \qquad (4)$$

$$v_j \geq \left(LB_j + \left(UB_j - LB_j\right) \cdot e_{up}\right) \cdot y_j^{up} + LB_j \cdot \left(1 - y_j^{up}\right), \forall j \in R \qquad (5)$$

$$v_j \leq \left(UB_j + \left(UB_j - LB_j\right) \cdot e_{down}\right) \cdot y_j^{down} + UB_j \cdot (1 - y_j^{down}), \forall j \in R \qquad (6)$$

$$v_j \geq LB_j \cdot (1 - y_j^{KO}), \forall j \in R \qquad (7)$$

$$v_j \leq UB_j \cdot \left(1 - y_j^{KO}\right), \forall j \in R \qquad (8)$$

$$v_j \leq UB_j \cdot (1 - y_j^{rev}), \forall j \in R \qquad (9)$$

$$y_j^{up} + y_j^{KO} + y_j^{down} \leq 1 \qquad (10)$$

$$\sum_j y_j^{up} + y_j^{KO} + y_j^{down} = k \qquad (11)$$

$$y^{rev}_{j,forward} + y^{rev}_{j,backward} = 1 \qquad (12)$$

$$y_j^{KO} \in \{0,1\}; y_j^{down} \in \{0,1\}; y_j^{up} \in \{0,1\}, \forall j \in R, \qquad (13)$$

Figure 3.4: Mathematical formulation for the optimisation algorithm Bio_Igg. Equations 1,10,11,12 and 13 are part of the outer optimisation with objective of maximising the IgG production and equations 2-9 part of the inner optimisation, dealing with the maximisation of the growth rate.

The Bio_Igg algorithm has both objective functions, inner and outer, as maximisations. A different approach is the implementation of the worst-case scenario set of objective functions, which is also used in the k-OptForce algorithm (*140*). Specifically, instead of maximising both the outer (antibody flux) and the inner (biomass flux) objective functions, the outer optimisation is maximising the secretion of antibody flux and the inner is minimizing the secretion of the antibody flux. Even though it seems that the inner optimisation serves no purpose as it tries to minimise the flux of the antibody secretion, on the contrary it acts as the maximisation of a competitive objective function. The inner minimisation of antibody production is similar to the maximisation of growth, since both try to reduce the allocation of resources in the production of antibody. To sustain a minimal growth, a

constrain for biomass production is implemented ensuring at least 10% of the maximum biomass will be produced

$$v_{bio} \geq v_{bio}^{max} \cdot 0.1 \quad (3.24)$$

Where v_{bio} is the predicted growth rate from the biomass equation of the FBA optimisation and v_{bio}^{max} is the maximum possible growth rate that can be achieved from the CHO-K1 GEM, constrained under each of the four different datasets. The strain design algorithm, called **Igg_Igg**, has a very similar implementation to Bio_Igg but with the changes in objective function and the additional biomass constraint. Overall, two optimisation algorithms were used, the one from OptReg with three additional changes named, Bio_Igg, as seen in figure 3.5, and Igg_Igg, which is the same as Bio_Igg but with a different set of objective functions and an additional ensuring minimum production of biomass.

3.3.4 Conversion to MILP

The two objective functions make the optimization problem nonlinear and difficult to solve with a solver such as gurobi or cplex. Linearising it can result to a simpler version of the strain design algorithm which can be solved using an MILP solver. The idea behind the transformation is based on the strong duality theorem (*158*), which is explained in detail in the section **2.4.2** of the literature review. Briefly, the theorem states that as long as both the primal and the dual of an optimisation problem are convex, then if the primal or dual problem have a finite optimal solution so does the other one and the optimal objective function values are equal to one another, where primal is called an optimization problem with N variables, M constraints, the dual of the primal one is the optimization with M variables, N constraints and the cost vector of the objective function is the right-hand side of the constraints (*136*). In the case of the Bio_Igg algorithm, the inner optimization problem is convex (because the binary variables are chosen by the outer optimization) and the outer optimisation is linear and convex as well, so we can apply the strong duality theorem. First, the dual variables θ_j for y_j^{up}, φ_j for y_j^{down}, μ_j^u and μ_j^l for y_j^{KO}, μ_j^{rev} for y_j^{rev} and λ_i for v_j need to be defined, then to transform the primal constraints into dual ones and set the objective function of the primal equal to the dual one

$$v_{biomass} = \sum UB_j \cdot y_j^{down} \cdot \varphi_j - \sum LB_j \cdot y_j^{up} \cdot \theta_j + \sum LB_j \cdot y_j^{KO} \cdot \mu_j^l - \sum LB_j \cdot y^{up} \cdot \mu_j^u + \sum UB_j \cdot y_j^{rev} \cdot \mu_j^{rev} \quad (3.25)$$

In the end, the constraints of the two inner optimization problems (primal and dual) are combined into one optimisation problem. Also, with the removal of the inner optimization, only the

outer optimisation exists with its own plus the constraints of the primal and dual inner optimization as seen in figure 3.5.

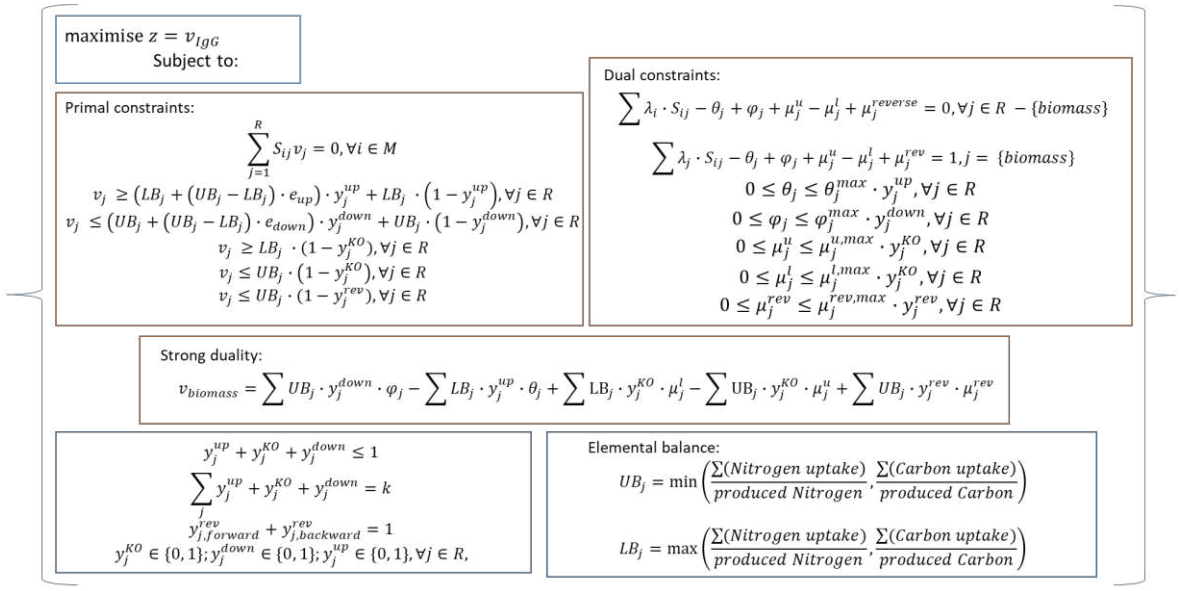


Figure 3.5: MILP formulation of the bilevel optimisation algorithm Bio_Igg. Equations inside the blue squares belong in the outer optimisation, and those in orange squares in the inner optimisation. The primal constraints are the original constraints of the inner optimisation, which are then transformed to dual constraints. The objective function of the primal and the dual optimisation problem are equalised in the strong duality theorem. The whole optimisation problem is MILP.

The non-linear terms $UB_j \cdot y_j^{down} \cdot \varphi_j$ must be linearised, for the optimisation to be MILP. To do that, instead of using the constraint $0 \leq \varphi_j \leq \varphi_j^{max} \cdot y_j^{down}$ one can define a new variable $w\varphi_j$ and a parameter M with a large value (e.g., 100) and use

$$w\varphi_j \leq M \cdot y_j^{down} \quad (3.26)$$

$$w\varphi_j \geq M \cdot y_j^{down} \quad (3.27)$$

$$w\varphi_j \leq \varphi_j + M \cdot (1 - y_j^{down}) \quad (3.28)$$

$$w\varphi_j \geq \varphi_j + M \cdot (1 - y_j^{down}) \quad (3.29)$$

The term $UB_j \cdot y_j^{down} \cdot \varphi_j$ is then changed to $\varphi_j \cdot UB_j - w\varphi_j \cdot UB_j$. For example, if $y_j^{down} = 1$ (meaning the reaction j must be downregulated) then $w\varphi_j = \varphi_j$ and $\varphi_j \cdot UB_j - w\varphi_j \cdot UB_j = 0$. On the other hand, if $y_j^{down} = 0$ (meaning the reaction j does not have to be downregulated) then $w\varphi_j = 0$

and $\varphi_j \cdot UB_j - w\varphi_j \cdot UB_j = \varphi_j \cdot UB_j$. Similar technique can be applied to all other four non-leaner terms.

3.4 Sensitivity Analysis

The Sensitivity Analysis (SA) framework is a 3-stage process removing the reactions initially that have no effect on neither antibody nor biomass production if all the other regulations from the optimisation algorithm take place; then keeping in the pool only the high-producing scenarios; and finally keeping in the regulation pool only the reactions responsible for at least 90% of the antibody and biomass production when all the other regulations take place. It is important to perform the three stages separately and to not initially remove the reactions that are not responsible for the 90% of the high-producing scenarios to evaluate the effect of every SA stage.

The large number of regulations makes it computationally very challenging to try and detect the vital regulations using a combinatorial approach. A way to approach the problem is to initially reduce the total size of regulations by removing the completely unnecessary ones i.e., those that have no visible effect on the production of either biomass or antibody according to an FBA optimisation, for each scenario. To perform that, an iterative approach was followed. Each of the scenarios had one-by-one the regulations removed from the scenario's regulations pool and then an FBA optimisation with biomass as the objective function was performed. If both the flux of antibody secretion and biomass production have a value of at least 99.99% larger than the one with all the regulations active, then the regulation removed from the pool has no effect on the outcome and can be ignored in the genetic engineering strategy. As a result, the regulations in the final pool are only those that are required to produce biomass and antibody. The above check was performed for all the regulations one at a time, thus checking each of the regulation's necessity for the outcome as seen in figure 3.6.

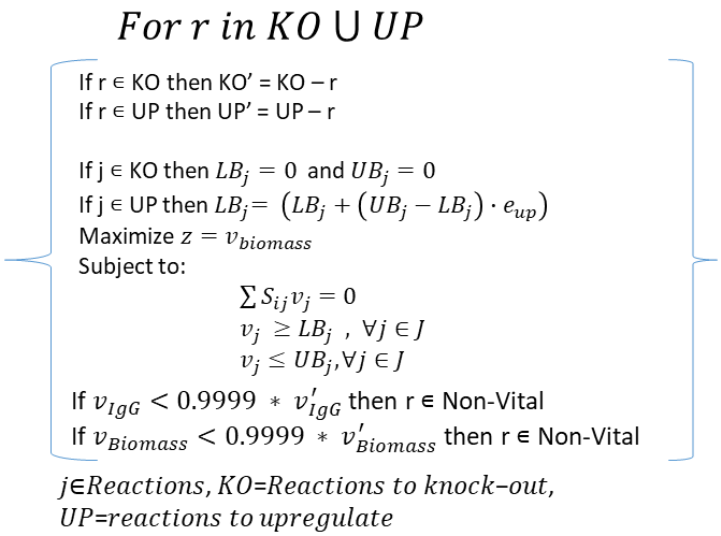


Figure 3.6: The algorithm used to check whether each of the regulations applied on the reactions' bounds has an effect on the required phenotype. The variable *r* iterates over all the regulations, $KO \cup UP$, and in each iteration all the other regulations are applied to the GEM except *r*. Then, the FBA optimisation predicts μ and v_{IgG}, which are compared with the values of the original scenario. If either of them is below the threshold, then *r* is categorized as vital for the outcome regulation. The v'_{IgG} is v_{IgG} when all the initial regulations of a scenario are active and $v_{IgG}^{FBA,exp}$ is the FBA predicted control value of the dataset that the scenario belongs. A similar methodology is applied to the predicted biomass.

After an initial reduction in the number of regulation, the scenarios were selected regarding their FBA predicted biomass and antibody. Their selection was based upon the FBA predicted μ and v_{IgG} in each scenario, and how they perform compared to the FBA predicted values of the original GEM constrained under the experimental datasets. Only the scenarios with both μ and v_{IgG} higher than at least 50% of the predicted FBA control values were considered as high producing ones. It is impossible for a scenario to have both μ and q_{mAb} higher than the control values as they are the result of an FBA optimisation and the control values of μ and q_{mAb} are the maximum possible for each dataset, which is why 50% was considered a cut-off. It was considered a reasonable cut-off because a result that yields low μ but high q_{mAb} could possibly not survive in a fed-batch culture and result in very low VCDs, or similarly high μ and low v_{IgG} would lead to high densities but with very low antibody secretion rates.

In third stage of the SA framework, the same algorithm as in Fig. 4.2 was applied again but allowing μ and $v_{IgG}$to be 90% of v'_{IgG} and μ', which are the antibody flux and growth rate if all the regulations after the first SA take place. That way SA removed reactions that affect slightly the biomass production or the antibody increase; further reducing the number of regulations to make the scenarios easier to study and apply *in vitro*. After the SA, the growth rate or antibody production could be lower than 50% of the control values for the original cell line, making the scenario not as efficient. However,

by changing the upregulation coefficient or by adding back a knockout regulation to the scenario regulations both the predicted growth rate and antibody production can be adjusted. The reasoning behind the 2nd stage SA was to identify the set of reactions that are responsible for the coupling of the two objective functions without keeping a strict constraint on the output, which would make easier to understand why the coupling of the objective takes place. Finally, in order to make sure that the number of regulations are the minimum possible that can achieve a production of IgG, a combinatorial approach was applied. The algorithm is the same as in the 1st stage of SA but instead of iterating over each regulation one by one, the iteration took place for each of all the possible combinations of the remaining regulations after the 1st stage of SA, as seen in figure 3.7.

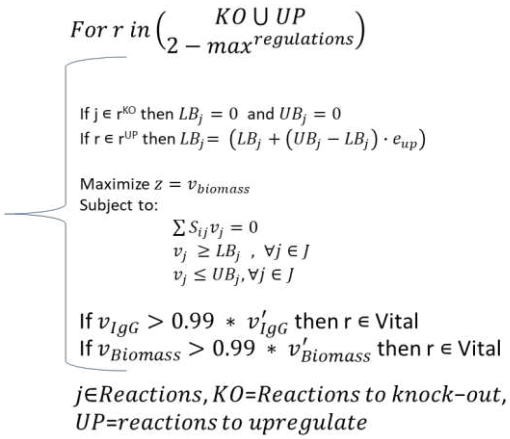


Figure 3.7: Algorithm used to check whether the 2-15 combination of the regulations in enough for the coupling of the two objectives. The variable *r* iterates of all the possible combinations of the set $KO \cup UP$. Then, FBA calculates μ and v_{IgG}, checking if both values are above the threshold, then *r* is categorized as vital for the outcome regulation. v'_{IgG} and $v'_{Biomass}$ are the fluxes of IgG and biomass where all the regulations take place and v_{IgG} and $v_{Biomass}$ are fluxes were all the regulations are active except *r*.

4

Strain Design

Mammalian and generally any type of cell have to face a large diversity of constraints to ensure growth, adaptation, and survival, such as biophysical, thermodynamic, stoichiometric constraints and limitations in its resource allocation (*160*). The allocation of resources in the cell is its ability to dynamically assign any type of internal metabolites and energy-carrying molecules towards the completion of certain objectives (*161*). Even though external resources can be infinitely supplied, there are internal physical limitations such as a limit in DNA translation for protein expression, a limitation in the cytosolic space and density, and general limitations of the metabolic network to produce all metabolic precursors necessary for biomass production or any other objective of the cell (*162, 163*). The objective in the allocation of resources can change from one target to another, such as maximising the growth rate, rapid response to nutritional changes, and survival under stressful conditions. Thus, growth can be seen as a process that allocates cellular resources, limited by environmental and physicochemical constraints (*164*).

A wild-type CHO cell does not express mAbs and the synthesis of recombinant proteins is not in the goal of the resource allocation of its metabolism. However, the expression can be cloned in the genome and then the engineered CHO has an additional objective of recombinantly producing mAbs as analysed in the Section **1.1** of the introduction. Production of mAbs, as well as any recombinant proteins from CHO, is an energetically costly process. The expression, glycosylation, folding and secretion of the recombinant proteins uses a lot of the cell's energy and resources (*108, 165*), as their molecular weight are in the order of magnitude of thousand Daltons. Their production also requires the expression of other supporting proteins involved in post-translation modification and secretion steps, which adds burden on the metabolism of the cell. The very high biosynthetic burden of mAb secretion imposes a burden on the metabolism of CHO cell (*110*), so they tend to suppress its expression as well as other costly to synthesise proteins. Biomass production also uses a large part of the cell's nutrient uptake. The two objectives of the cell form a competitive relationship, as both have a high resource and energy requirement. This argument is also supported by the increase in total titer and growth rate after the removal of unnecessary proteins from the product of interest (*112*). A way to change the diametrical relationship between the two objective functions is by altering the metabolic network, which would redistribute resources in a more optimal manner to the objective of the optimisation.

FBA optimisation can be used to simulate the trade-off between the two objectives. In the optimisation when one of the two objectives is maximised, the other is minimised. A way to change the trade-off and create a point where the maxima of the two objective functions coincide is to alter the metabolic network for the resources to be directed towards antibody production. As presented in the Section **2.2.2** of the literature review, there have been numerous strain design algorithms developed where the goal is the detection of regulations in the network that couple the objective functions in an FBA optimisation. The target of these algorithms is to change the Pareto curve for the maximum points of the two objectives to coincide. Examples like OptKnock, which detects possible reaction knockouts, OptCouple which also detects changes in the media composition or OptForce that allows only specific reactions to be regulated have all had their genetic results been applied to microbial organisms with positive results, as analysed in the Section **2.2.2** of the literature review. However, to the best of my knowledge, none of these genetic engineering algorithms have been successfully applied to the metabolic network of the CHO cell, or generally in any other mammalian models. Using a strain design algorithm on the GEM of the CHO-K1 cell line, antibody and biomass production can be coupled, so in an FBA optimisation with biomass production being the objective function antibody would be synthesised and secreted as a by-product. *In vitro* it would mean that that the engineered cell line would dedicate more resources to the production of antibodies than in the original CHO GEM, resulting in increased mAb specific productivity and potential reduced growth rate as resources would be taken from biomass synthesis and get channelled towards the antibody synthesis pathway. The strain design algorithm used herein is OptReg, as it was deemed the most versatile option and most likely to return a multitude of different genetic engineering scenarios.

Chapter 4 initially defines the four experimental datasets which will be used throughout the thesis to constrain the fluxes of the GEM. Then the four experimental datasets are expanded to also include intracellular reactions using elemental balance algorithms. The OptReg algorithm is explained as well as the three changes applied to the algorithm, including an alternative objective function simulating the worst-case scenario is analysed. The transformation from bi-level optimisation to MILP is presented, which makes the optimisation problem simpler to solve. Lastly, the 93 different genetic strategies that resulted from the optimisation are presented based on the different datasets and objective function used.

4.1 Dataset definition

Four different experimental datasets were used in this study to apply constraints to the GEM which were taken from two different publications (*127, 166*). They are from fed-batch cultures of CHO cells and contain the concentration of the amino acids, glucose, lactate, ammonia, IgG yield and viable cell density (VCD). The definition of the four datasets is presented in table 4.1.

Table 4.1: Each of the four datasets used in the cell-line design, with the corresponding cell line and feeding used for the experimental data.

Dataset	*Cell-Line*	*Feeding*	*Reference*
D1	CHO	FeedC	(*127*)
D2	CHO	FeedC	(*166*)
D3	CHO	FeedAll	(*166*)
D4	CHO	FeedAll+40%	(*166*)

The CHO cell lines come from the same parental cell line, CHO-K1, making the CHO-K1 GEM best suited model to represent the experimental data and conditions. D1 and D2 have the same commercially available feed, FeedC (CD EfficientFeed C AGT, Invitrogen, UK). D3 and D4 both use custom made feeds. The complete explanation and the final amino acid and glucose concentrations can be seen in the original publication (*166*). Briefly, FeedAll assumes that the cell consumption during fed-batch is the same as during a batch culture, up until aspartate and glutamate deplete. The observable consumption rates during the batch culture were scaled up with the goal of tripling the IVCC. The FeedAll+40% follows the same logic behind FeedAll, but it has an increase of 40% in the concentration of the amino acids except tyrosine, which was found to be more than the solubility limits in FeedAll.

The datasets contain experimental values in the form of daily concentrations (mM) for the 20 amino acids, glucose, ammonia, and lactate in the culture medium. The medium that used for the culturing of the cells, in all four datasets, was the commercially available CD-CHO (ThermoFisher Scientific, U.K.). However, D2-4 do not contain any measurements for cysteine, which is an essential amino acid for CHO growth (*108*). The experimental measurement for cysteine was taken from Lularevic et al. (*92*) because of the similarity of experimental conditions. The datasets also contain the daily IgG titer (g/l) and VCD (cells/mL).

The drain reactions of the GEM are constrained using the maximum and minimum uptake/secretion values for each metabolite of the drain reaction. As a first step, the concentrations in each dataset should be converted to fluxes. The concentrations from the datasets can be converted to fluxes by dividing the change in concentration by the change in VCD. Initially, the feeding events, that took place every other day, need to be considered in the calculation of the daily concentrations. The concentrations of each metabolite after each event of feeding were adjusted using:

$$C_{i_t} = \frac{F_{IN} \cdot C_{i_{in}} + C_{i_{t-1}} \cdot V_{t-1}}{V_t} \tag{4.1}$$

Where C_{i_t} is the concentration of component *i* (mM) at the time point *t*, F_{IN} is the feed volume (L), $C_{i_{in}}$ is the concentration of component *i* in F_{IN} and V_{t-1} is the volume of the system (L) before the

event of feeding. The fluxes are calculated from the division of change in the concentration of a metabolite in the culture media by the change in the population of viable cells. However, the reason the change in integral of viable cell density (IVCD) was used instead of change in VCD, is because IVCD incorporates the total time the cells are alive instead of a single point change that the change in VCD would offer. Also, the concentration measured each day represents the cumulative concentration of media, so the change needs to be divided by the cumulative viable cell density which is the IVCD. IVCD is calculated as follows:

$$IVCD_t = IVCD_{t-1} + 0.5 \cdot (VCD_t + VCD_{t-1}) \cdot \Delta t \qquad (4.2)$$

Where $IVCD_t$the integral of viable cell density up to time *t*, VCD_t is the viable cell density at time *t* and Δt is the corresponding time interval. Subsequently, the specific rates of consumption/production are the ratio between the changes in cumulative concentration over the change of IVCD:

$$q_i = \frac{\Delta C_i}{\Delta IVCD} \qquad (4.3)$$

q_i is the specific consumption (if negative) or production (if positive) of component *i* $\left(\frac{pmol}{cell \cdot h}\right)$. The resulting specific rates have the units of pmol/cell/h, however for the biomass to have hr^{-1} units, the units of q_i must be converted to mmol/DCW/h, which requires the multiplication of pmol/cell/h with the dry cell weight (DCW). The DCW of CHO cells during early exponential growth (day 4 of a batch culture) has been measured to be 0.315mg/10^6cells (*125*). All four datasets were gathered for a similar CHO cell line, so the DCW was assumed to be the same between all datasets. The above method results in specific rates of consumption or production for the metabolites measured experimentally for all four datasets. The optimisation is focused on the exponential growth phase of the culture, so the maximum (upper bound) and minimum (lower bound) rates of the exponential growth for each dataset were chosen as bounds for the optimisation.

The four experimental datasets constrain only 23 reactions of the GEM and the rest of the 420 drain reactions remain unbounded, having their flux ranged from [0, 10] mmol/gDCW/hr. However, there are plenty of other metabolites in the culture medium and by setting their lower bound to 0 there is no uptake allowed, disabling a lot of potential metabolic pathways. The most vital and common ones should, therefore, have their lower bound adjusted to a non-zero value. The upper bound can have an arbitrary large value because it accounts for secretion and not uptake. In the culture media, there are metabolites such as oxygen or water that are abundant in the medium and should have the option for unlimited uptake. The specific culture media used for the datasets D1-4 was CD-CHO (ThermoFisher). Using the patent of CD-CHO (*167*) which names the metabolites present in the media, the remaining

metabolites in the culture media can be determined by cross-referencing the drain reactions of the GEM with the CD-CHO metabolites, and define a lower bound for them. Upper bounds for these metabolites were set to an arbitrary high number giving the option for unlimited secretion from the model. Concerning the lower bounds, which are the uptake rates, they were taken either from the metabolomic dataset of Hefzi et al. which also use the CHO-K1 cell line (*108*), or if no information was supplied the lower bound was defined as a fraction of the glucose uptake rate, due to glucose being the most abundant metabolite in the culture media.

$$LB_i = LB_{glucose} \cdot \frac{C_{glucose}}{C_i} \tag{4.4}$$

Where LB_i is the lower bound of the metabolite i, $LB_{glucose}$ is the lower bound of glucose for the specific dataset, $C_{glucose}$ is the number of carbons in glucose and C_i the number of carbons of the metabolite i. In summary the bounds used for the datasets are presented in table 4.2.

Table 4.2: Constraints used for the metabolites without any experimental data in the four datasets. In the first column, the name of the metabolite is associated to the drain reactions that is responsible for the uptake/secretion of the metabolite in the CHO-K1 GEM. The reason why the metabolite is included in the dataset is presented in second column. CD-CHO patent means the metabolite exists in the culture media, according to the patent. GENERIC means the metabolite exists due to its abundance in nature. SINK are the reactions which need to have defined bound, for the sake of the GEM's feasibility. The value of the bound is defined using the method in the third column. Glucose means the bound is calculated as a function of glucose. ICHO1766 means the bound is taken from the iCHO1766 metabolic data (*108*). The generic bounds means that the metabolites are abundant in the culture media so the uptake should be unlimited.

Metabolite	*Source*	*Lower Bound*
Hydroxy-L-proline	CD-CHO Patent	Glucose
Pyridoxine		iCHO1766
Thiamin		Glucose
Oxidized glutathione		Glucose
L-ascorbate		Glucose
Folate		iCHO1766
Riboflavin		Glucose
Biotin		iCHO1766
Choline		iCHO1766
Nicotinamide		Glucose
Myo-inositol		iCHO1766
Spermine		Glucose
Hydrogenphosphate		Generic
Bicarbonate	GENERIC	Generic
Acetate		Generic
Proton		Generic
Water		Generic
Oxygen		Generic
Sulfate		Generic
Protein-linked asparagine residue	SINK	iCHO1766
Protein-linked serine or threonine residue		iCHO1766
Tyr-194 of apo-glycogenin protein		iCHO1766

The intracellular reactions of the CHO-K1 GEM all had all bounds [-10, 10] mmol/gDCW/hr if they were reversible, and [0, 10] mmol/gDCW/hr if they were irreversible, except the ATP maintenance reaction. ATP maintenance is responsible for creating the ATP needed to sustain the metabolic activities of the cell. If the lower bound is 0, then the ATP maintenance would not produce any ATP, because the reaction is simply consuming resources without any positive impact on biomass production. However, with a non-zero lower bound, the model allocates resources for the production of ATP which represents

the cell's ability to sustain its metabolic activities and survive. The value ATP production is a function of the weight of the cell, since the larger the cell the more ATP is required to sustain it. The value is taken from Sheikh et al. who they calculated the ATP maintenance to be 8.6mmol/DCW/hr for a mouse cell line (*168*). A mouse cell line is different from the CHO-K1 that all the experimental data is taken for, however, the two cell lines are similar and it was decided to keep that value. The upper and lower values for each of the four datasets are presented in table A1.

4.2 Elemental Balance – Dataset Curation

As analysed in the previous paragraph, the four different datasets D1, 2, 3, 4 were used to constrain the metabolic model of the CHO cell. They are taken from different experiments, so they contain different uptake and secretion rates for the measured metabolites. However, they only apply constraints to the 46 drain reactions that connect the metabolites to the extracellular environment. The remaining 3904 intracellular reactions of the CHO-K1 GEM are left unbounded. The vast number of unbounded reactions creates a large solution space in which some flux distributions resulting from the FBA optimisation may not obey biological, thermodynamic, or chemical laws. In an attempt to improve the prediction of fluxes from the FBA optimisation, three algorithms were applied to adjust the maximum and minimum bounds for the reactions. Initially, an FVA algorithm, which is explained in Section **2.1.5**, was applied that calculates the maximum and minimum possible fluxes for the 3904 reactions for all four datasets separately. The FVA algorithm does not incorporate any chemical or biological information, and the resulting upper and lower bounds are purely based on the values of the uptake fluxes and the coefficients of the metabolic model's stoichiometric matrix. The resulting bounds are smaller than the abstract high bounds of the model, but they did not reduce the feasible solution space. The main reason why FVA was performed was to reduce the values of the upper and lower bounds, which would aid in simulating an up- or downregulation in the strain optimisation. Furthermore, FVA did not remove the energy-generating futile cycles that result in unrealistically large fluxes. To limit the amount of flux created in these cycles and further constrain the solution space, physicochemical constraints should be applied.

A simple way to further constrain the upper and lower bounds and also apply biological and chemical limitations on the fluxes values is the use of elemental balances. Carbon and nitrogen balance was applied to the model using the algorithm of carbon balance (ccFBA) (*92*) and later augmented to also include nitrogen atoms (ncFBA) (*93*). The carbon balance works by adjusting all of the upper and lower limits of reactions, based on the total carbon uptake. Nitrogen balance works in a similar way, but it adjusts the bounds based on the total nitrogen uptake of the cell. The main reason for the use of nitrogen for elemental balance is because of its abundance in amino acid metabolism. The final bounds of the datasets consist of a combination of the FVA, ccFBA and ncFBA bounds with the objective of

keeping the narrowest value for each flux. As a result, the predicted fluxes of the intracellular reactions by an FBA optimisation would have much smaller values than an unbounded FBA optimisation.

CcFBA and ncFBA limits the amount of flux carried by each reaction based on the total carbon and nitrogen uptake by the cell. Assuming C_{TOT} is the total carbon and N_{TOT} is the total nitrogen flux entering the cell (mmol/gDCW/hr), then a biochemical reaction of the GEM cannot produce more carbon or nitrogen flux than C_{TOT} or N_{TOT} respectively. Thus, the upper and lower bounds of each reaction are adjusted by calculating the total carbon produced in each reaction. For example, the enzyme glutamine synthetase (GS) which catalyses the conversion of glutamate to glutamine in the reaction

$$\text{GS: atp} + \text{glu}_\text{L} + nh_4 \leftrightarrow \text{adp} + \text{gln}_\text{L} + \text{h} + \text{pi}$$

or by replacing the names of the metabolites with their chemical formulas

$$\text{GS: } \text{C}_{10}H_{12}N_5O_{13}P_3 + C_5H_8NO_4 + \text{N}H_4 \leftrightarrow \text{C}_{10}H_{12}N_5O_{10}P_2 + C_5H_{10}N_2O_3 + \text{H} + \text{HP}O_4$$

GS produces/consumes 15 carbon atoms and 7 nitrogen atoms. Based on the chemical equation of GS, the flux of the reaction cannot exceed the value of $\frac{C_{TOT}}{15} mmol/gDCW/hr$ and $\frac{N_{TOT}}{7} mmol/gDCW/hr$, because that way there would be more carbon or nitrogen atoms produced than the total uptake of the cell. However, ATP and ADP act as a cofactor pair, which means that their carbon and nitrogen atoms do not participate in the reactions but are exchanged or cleaved and they should not be considered in the total carbon and nitrogen atoms produced. Therefore, the total number of carbon atoms produced/consumed from the GS reaction would be 5 instead of 15 and the total number of nitrogen atoms 2 instead of 7. The flux of GS should then be lower than $\frac{C_{TOT}}{5} mmol/gDCW/hr$ and $\frac{N_{TOT}}{2} mmol/gDCW/hr$. To enforce both constraints, the flux of GS must be smaller than the smallest out of the two elemental constraints. In the end, the flux of GS will be between the two.

$$-min\left(\frac{C_{TOT}}{5}, \frac{N_{TOT}}{2}\right) \leq v_{GS} \leq min\left(\frac{C_{TOT}}{5}, \frac{N_{TOT}}{2}\right) \quad (4.5)$$

It should be noted that if GS is a one-directional reaction, then the flux would be in the range of

$$0 \leq v_{GS} \leq min\left(\frac{C_{TOT}}{5}, \frac{N_{TOT}}{2}\right) \quad (4.6)$$

The workflow of the elemental balance algorithms is to first calculate C_{TOT} and N_{TOT} by multiplying the maximum uptake of the drain reactions with the corresponding carbon and nitrogen atoms of each metabolite. Co-factors do not take part in the metabolic reactions as explained in the

example above, for that reason next step is the identification of the co-factor pairs of the metabolic model. In the CHO-K1 GEM the co-factor pairs are NAD+/NADH, NADP+/NADPH, FAD/FADH2, ATP/ADP/AMP. Together with cofactor pairs, coenzyme A (CoA) and other CoA molecules, cytochromes, and quinones are also excluded from the elemental balances, due to them acting as currency metabolites and not directly taking part in the reaction. Drain reactions do not take part in the elemental balance, as these are the inputs of the algorithm and provide the total carbon and nitrogen uptake. The calculation of the maximum and minimum fluxes of each reaction based on carbon and nitrogen balance takes place similarly as above. Finally, the bounds are then included in each dataset, with the objective of keeping the narrowest bound and a final FVA is performed to further reduce the values of the bounds.

The quality and importance of elemental constraints is demonstrated by the number of reactions constrained by the two algorithms, ccfBA and ncFBA, and by the constraint applied in growth rate. Initially, the carbon and nitrogen bounds were calculated with the process described above for all four different datasets. All the datasets used were collected from cell lines descendant from CHO-K1, so the CHO-K1 GEM, which is derived from the generic iCHO1766 GEM (*108*), would best represent the metabolic network of the cell lines. The CHO-K1 GEM was then constrained under the D1, 2, 3 and 4 carbon derived and nitrogen derived bounds and under the generic FVA based on the uptake and secretion rates of each dataset. FBA optimisation was performed, with the objective of maximising biomass production, to predict each dataset's growth rate – predicted biomass, which was then compared to the experimental value of the corresponding dataset. The experimental growth rate, μ, is equal to the slope of the logarithm of VCD vs time (hr) during the exponential growth phase. The exponential growth phase in each dataset was selected by plotting the graph of VCD against time. Error propagation was used to determine the standard deviation of μ and q_{mAb}, due to the lack of raw experimental data for the four datasets. The results are presented in table 4.3.

Table 4.3: Results of the element balance algorithm on each of the four datasets. The first two rows show the total number of non-zero bounds adjusted from the carbon and nitrogen balance. The experimental value of μ is then presented as a comparison to the predicted computational value. The value of μ for FVA uses only the uptake and secretion bounds from each of the four different datasets, while the carbon constraint μ and nitrogen constraint μ use bounds calculated by the ccFBA and ncFBA respectively.

		D1	*D2*	*D3*	*D4*
Carbon derived bounds		3321/3420	3322/3440	3319/3377	3331/3369
Nitrogen derived bounds		1983/3420	1996/3440	1985/3377	1996/3369
μ, hr^{-1}	*Experimental*	0.017±0.01	0.017±0.01	0.021±0.01	0.014±0.03
	FVA	0.0123	0.0297	0.0297	0.0297
	ccFBA	0.0453	0.0573	0.0494	0.0634
	ncFBA	0.0206	0.0265	0.0227	0.0293

Two important conclusions can be drawn from the above table. The nitrogen-constrained growth rate is smaller than the carbon one in all four datasets and also smaller than the FVA predicted one in datasets D2-4, whereas the carbon predicted the largest growth rate in all four datasets. The reason lies in the definition of experimental bounds. Carbon-rich metabolites such as glutathione, bicarbonate, and other important vitamins, which do exist in the CD-CHO culture medium (*167*), had their uptake rate defined as a function of glucose in the four experimental datasets due to lack of experimental data. Glucose was chosen because it was the most abundant experimentally measured amino acid in all four cases. This led to an overestimation of carbon uptake, resulting in an over prediction of the intracellular bounds and growth rate, calculated by ccFBA. Therefore, the application of nitrogen-derived bounds acted as a correcting agent for the overestimation of carbon uptake. It was decided to keep the uptakes of the metabolites that were not experimentally measured in the datasets, despite them reducing the effectiveness of ccFBA. The availability of the metabolites can allow flux to go through their catabolic pathway, so they might provide insights for a change in the metabolic network after the application of the strain design optimisation. Antibody production was not constrained further than the FVA calculated maximum value by neither algorithm, which is the reason it is not shown in the above table 4.3.

Another important conclusion is the importance of nitrogen bounds in constraining the intracellular reactions. From the first two rows of Table 3.1 it can be concluded that almost all reactions of the metabolic network contain carbon, and thus the algorithm can calculate possible upper and lower bounds. Although the total number of nitrogen constraint reactions is smaller than the carbon ones, a significant number of metabolic reactions, 75% of the total, contain nitrogen which allows ncFBA calculate upper and lower bounds. Many reactions that participate in amino acid metabolism and

catabolism and contain nitrogen atoms were constrained as seen in Figure 4.1, justifying the importance of nitrogen-derived bounds.

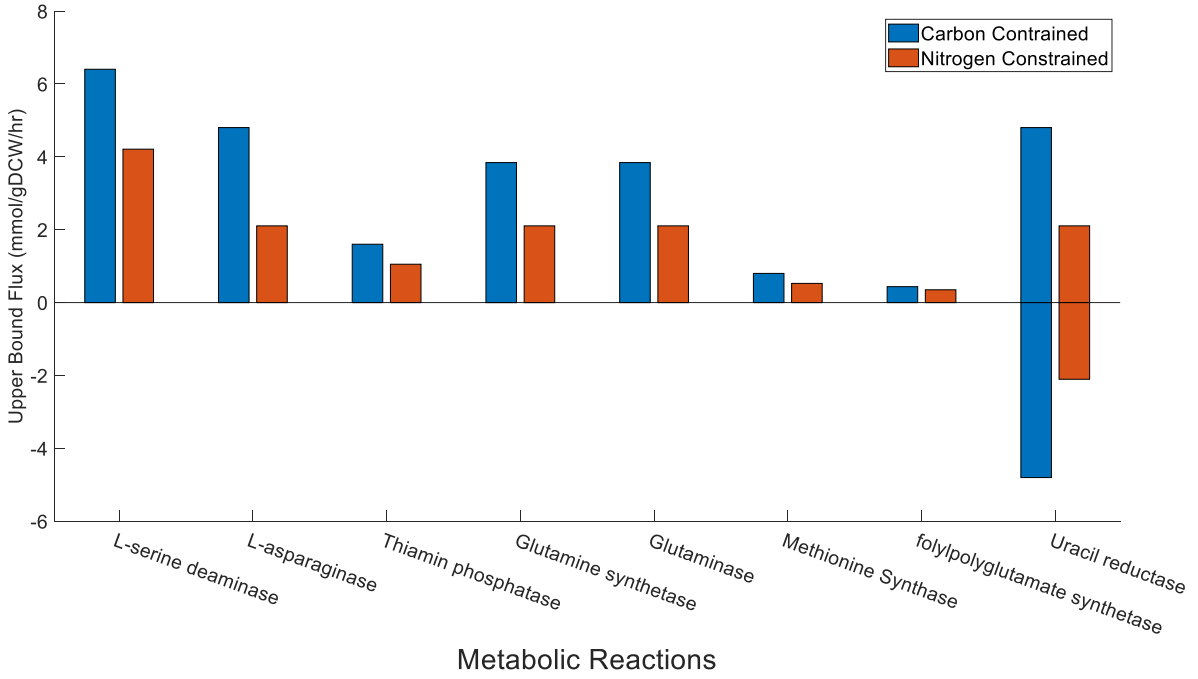


Figure 4.1: Upper and lower bounds of the ccFBA and ncFBA on eight enzymes which are part of the amino acid metabolic/catabolism. The orange column represents the predicted nitrogen bounds, and the blue column are the bounds based on the total carbon uptake of the cell.

In conclusion, nitrogen derived bounds proved to be useful and in a lot of metabolic reactions (40%) providing narrower bounds than the carbon derived ones. The reason other elements, such as oxygen, phosphorus or hydrogen, were not used for elemental balances as well was that they were not exact measurements, and it was impossible to assume an exact uptake rate for oxygen, phosphorus or protons. Carbon and nitrogen can easily be measured though the measurement of extracellular concentrations of amino acids, glucose, lactate, and ammonia. The concentrations of other elements containing them such as vitamins, can either be measured using experimental techniques to provide an accurate calculation of the total carbon and nitrogen atoms entering the cell.

After the calculation of the carbon and nitrogen bounds, the strictest bounds between these two methods were picked since that way the fluxes of the reactions would definitely obey the elemental balances for both carbon and nitrogen. The distribution of reaction bounds that result from each algorithm (FVA, ccFBA, ncFBA) for each of the four datasets is presented in table 4.4. These four datasets will then be used for further optimisation and strain design.

Table 4.4: Number of reactions that had their bounds adjusted by each of the three methods, FVA, carbon, and nitrogen balance.

	Percentage	*D1*	*D2*	*D3*	*D4*
FVA Bounds	38%	1495	1485	1494	1485
Carbon Bounds	39%	1544	1551	1543	1551
Nitrogen Bounds	23%	915	916	915	916
Total reactions	100%	3952	3952	3952	3952

4.3 Original CHO cell line

The datasets D1, 2, 3, 4 were augmented in the previous section to include elementary derived constraints for the intracellular reactions for the CHO-K1 GEM. A comparison between the experimental μ and q_{mAb} of each dataset, and the values predicted by an FBA optimisation of the constraint under each of the four datasets CHO-K1 model can hint the quality of the GEM as well as of the experimental data. Experimental growth rate is calculated as explained in the previous section and has units of hr^{-1}. The experimental q_{mAb} (*169*) is calculated by:

$$q_{mAb}\left(\frac{\frac{pg}{cell}}{day}\right) = \left[\frac{titer_{t+1} - titer_t}{VCD_{t+1} - VCD_t}\right] \cdot daily_growth_rate \qquad (4.7)$$

Where, q_{mAb} is the antibody specific productivity in pg/cell/day, $titer_{t+1} - titer_t$ is the difference in the antibody titer in (g/L) between two time points and $VCD_{t+1} - VCD_t$ the difference in viable cell density (VCD) between two time points. The experimental daily growth rate (hr^{-1}) is given by:

$$daily_growth_{rate}(hr^{-1}) = \frac{\ln(VCD_{t+1}) - \ln(VCD_t)}{\Delta T} \qquad (4.8)$$

Where ΔT is the difference in time between the two points *t+1* and *t* and $\ln(VCD_t)$ is the natural logarithm of the VCD at time point t. The produced biomass by the FBA optimisation has units of hr^{-1}, the same as the experimental one. The units of secreted IgG flux,v_{IgG} are mmol/gDCW/hr which need to be converted to specific productivity, pg/cell/day. Assuming the CHO dry cell weight (DCW) remains constant throughout the exponential growth phase at $0.315mg/10^6$cells, the flux needs to be multiplied by the $0.315mg/10^6$cells (*170*), by 24hr/day and by the molecular weight of the produced antibody, 150 kDa. In the end the FBA predicted q_{mAb} will be

$$q_{mAb}(pg/cell/day) = v_{IgG}(\text{mmol/gDCW/hr}) \cdot 0.315 \cdot 24 \cdot 150{,}000 \quad (4.9)$$

The FBA predicted values were the result of an FBA optimisation in MATLAB using the COBRA toolbox, as explained in section **3.2**. The objective function of the optimisation was biomass production, whilst keeping a lower bound in antibody production (from the experimental dataset) to force antibody production. Similarly, for the prediction of q_{mAb} the objective function was the maximisation of antibody synthesis while imposing a lower bound in biomass production (the experimentally measured growth rate). The rest of the reactions in the CHO-K1 GEM were constrained by the datasets D1, 2, 3, and 4. The values are presented in table 4.5.

Table 4.5: Experimental values and FBA predictions for growth rate, μ, and specific productivity, q_{mAb}, for the four different datasets used. In the two FBA prediction rows, the different between the predicted and the experimental value is displayed in the parenthesis.

	Dataset	*D1*	*D2*	*D3*	*D4*
μ, hr^{-1}	Experimental	0.017±0.01	0.017±0.01	0.021±0.01	0.014±0.03
	FBA Prediction	0.009 (-93%)	0.019 (+15%)	0.022 (+5%)	0.022 (+35%)
q_{mAb},	Experimental	13.27±5.5	40.75±5.5	27.75±14.27	30.42±6.80
pg/cell/day	FBA Prediction	**Infeasible**	47.289 (+14%)	32.401 (+15%)	58.456 (+48%)

FBA optimisation on the CHO-K1 GEM, constraint by three out of four datasets (D2, D3 and D4), over predicted both μ and q_{mAb} compared to the experimental values. FBA is an LP optimisation technique without any constraints focused on substrate inhibition, so given a specific dataset and an objective function, it will always result in the maximum possible value of the objective function. If the objective function is biomass production, as in this case, then FBA would try and channel all the resources with the goal of maximum growth rate, resulting in a possible over prediction of the objective. However, optimisation performed on the constraint by the D1 dataset GEM could not predict the growth rate. As a result, the addition of the experimental growth rate as a lower bound in the antibody maximisation FBA, made the optimisation infeasible. The reason is the much lower value of cysteine uptake than in datasets D2-4. If the FBA was performed with the cysteine flux unbounded i.e., the CHO-K1 GEM could uptake unlimited amounts of cysteine, then growth rate was over predicted by 45% and antibody production by 65%; verifying that the cysteine uptake (lower bound) is the bottleneck in the D1 dataset. Cysteine is essential for growth and cannot be produced naturally in CHO cells (*108*), so a reduced uptake would also lead to reduced biomass production in the FBA optimisation. The bound for cysteine in D1 is measured by Sou et al. (*127*), but due to the lack of a value in D2-4, the cysteine concentration was borrowed from Lularevic et al. (*92*). The reason the value of cysteine from D2-4 was not used in D1 was because of the different cell line used in the experiment designs. Lularevic used the

same exact cell line to the D2-4 experiments, whereas D1 used a slightly different CHO cell line. The metabolic requirements of D2-4 and the experiment by Lularevic were expected to be the same, due to the culture of the same cell line. The main difference between the experiment of Lularevic et al. and of the D2-4 datasets is that the former performed the cell cultures in batch mode whereas D2-4 were performed in fed-batch mode. However, the feeds in all three fed-batch experiments did not contain cysteine, so it was thought that the batch and fed-batch values of cysteine would be similar.

The FBA optimisation is infeasible, when constrained under D1 dataset with the objective to maximise the protein production, as seen above in the table 4.5. The reason is that the value of growth rate used as constraint for the biomass production is the experimental one, $0.017hr^{-1}$, which is higher than the FBA predicted one, $0.009hr^{-1}$ forcing the optimisation to produce more biomass than it can produce with the given resources. A way to make the problem feasible is to reduce the constraint of biomass production. The FBA predicted biomass value when the objective function is biomass production, is a value that is both in the feasible solution space and a reasonable value as it is predicted from the CHO-K1 constraint under the D1 uptake rates and lower bound for IgG secretion. Therefore, after changing the biomass bound on the D1 dataset, FBA is now feasible and there can be both q_{mAb} and μ prediction as seen in figure 4.2.

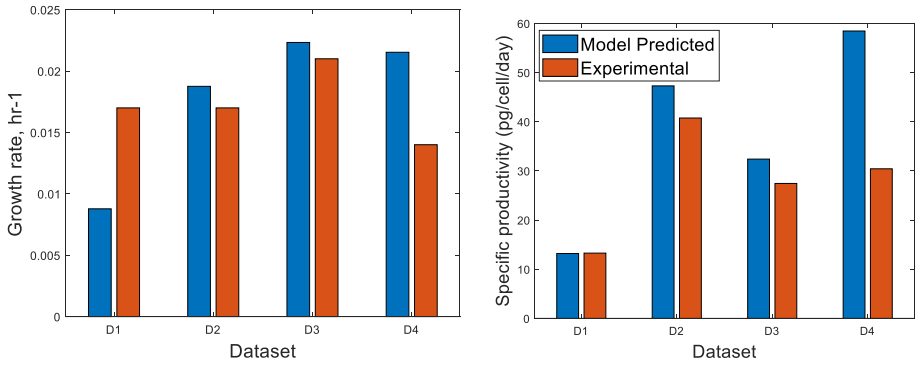


Figure 4.2: FBA optimisation of the original cell line CHO-K1 for all four datasets. The blue bars represent the FBA predicted values, and the orange the experimental values in each of the four datasets. For the optimisation of antibody secretion, D1 uses as lower bound for growth rate the FBA predicted value of 0.009 instead of the experimental 0.017.

The FBA predicted values calculated will be used later as control values for the genetic engineering strategies to rank them based on their quality. The control values are the ones predicted from the FBA optimisation for the CHO-K1 GEM, constrained under the four datasets. The values of μ and IgG flux are predicted straight from the FBA optimisation. The value of q_{mAb} is calculated from the flux of IgG. They are presented in table 4.6.

Table 4.6: Control values from the FBA optimisation of the CHO-K1 GEM, constrain under each of the four datasets.

Control Values/ Dataset	*D1*	*D2*	*D3*	*D4*
FBA Predicted				
Growth Rate, μ (hr^{-1})	0.009	0.019	0.022	0.022
IgG Flux, v_{IgG} (mmol/gDCW/hr)	$1.164 \cdot 10^{-5}$	$4.170 \cdot 10^{-5}$	$2.857 \cdot 10^{-5}$	$5.155 \cdot 10^{-5}$
IgG Specific Productivity, q_{mAb} (pg/cell/day)	13.194	47.289	32.401	58.456
Experimental				
Growth Rate, μ (hr^{-1})	0.017±0.01	0.017±0.01	0.021±0.01	0.014±0.03
IgG Specific Productivity, q_{mAb} (pg/cell/day)	13.27±5.5	40.75±5.5	27.75±14.27	30.42±6.80

4.3.1 Pareto curve

The relationship between antibody production and growth rate in CHO cells has been studied numerous times. Enhanced mAb productivity can take place when there are conditions that inhibit the growth rate of the cell and the specific growth rate is low (*171-173*). The reason is that the cell's metabolism cannot easily support both high demanding objective functions, biomass and antibody production (*174*). Generally, antibodies have an increased production when there is low demand for cellular energy for biomass production or the energy required for cell maintenance. The argument is also supported by mild hypothermia studies, where antibody productivity increases when temperature drops below the optimal amount of 36.5°C, which causes the growth rate to reduce (*127, 175*). The antagonistic effect of the objective functions can be observed *in silico*. The CHO-K1 GEM has two competitive objective functions, biomass and antibody production, when one is maximised, the other is minimised in an *in-silico* optimisation (*176*). The diametrical effect can be simulated using FBA optimisation. For any of the four datasets used to constrain the GEM, FBA optimisation is feasible and results in non-negative flux in both objectives, because of the large flux of amino acids, ATP and other carbon sources required for the synthesis of either. The relationship of the objectives can be seen in the trade-off curve, which is constructed by solving the optimisation problem iteratively while setting a lower bound on the secondary objective, e.g. antibody secretion, and keeping the objective of the optimisation, e.g. biomass production, unbounded, as presented in figure 4.3.

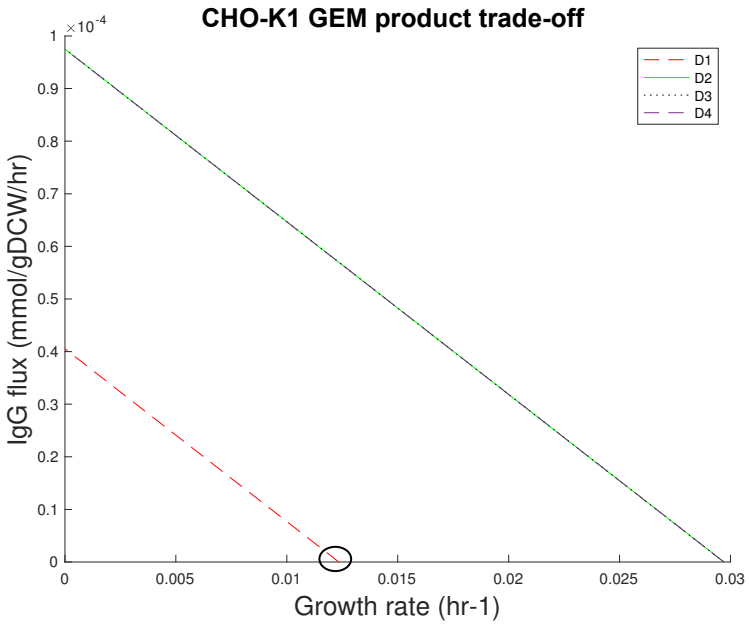


Figure 4.3: Trade-off curves between biomass and antibody production for four of the datasets in the original CHO-K1 GEM. The circled point of D1 dataset is the produced IgG and biomass when biomass is maximised in the FBA optimisation, in the original D1 constrained CHO-K1 GEM.

The creation of the Pareto front requires the objective of the optimisation to be unbounded. Both IgG and biomass production, contained no bounds from ccFBA or ncFBA. When biomass is left unbounded all three datasets predict the same value for biomass (Fig. 3.3), even though there is a difference in the value of the ncFBA biomass predicted between D2, 3 and 4 datasets (Table 3.1). The reason is that the value of the biomass objective is dependent on the values of the metabolite uptakes, which were not constrained so the maximum produced biomass is the same as without the application of the elementary balance algorithms. Nevertheless, figure 3.3 visibly shows the competitive relationship of these two objective functions in an FBA optimisation.

Based on the trade-off curve the objective of a strain design algorithm seems straightforward; the Pareto curve should be restructured so there would be a point where the maxima of biomass production coincides with the maxima of antibody secretion. That way, in an FBA optimisation with biomass being the objective function, there would be a non-zero positive flux in the antibody synthesis pathway. The purpose of strain design algorithms is to identify possible changes in the reactions' bounds that would result in a change in the Pareto front. A change of the Pareto front can be achieved using bilevel optimisation. A multi-objective optimisation with the goal of maximising both biomass and antibody production in the case of CHO-K1 GEM, using for example weighted sum would result a point on top of the figure 3.3 curves. The bi-level optimisation of the CHO-K1 GEM would result in a new optimal point, which would result positive flux of IgG secretion by the GEM when the objective function is biomass maximisation in the FBA optimisation.

4.3.2 Gene and Reaction Essentiality

GPR rules, as explained in the section **2.1.3** of the literature review, are Boolean expressions used to connect genes with their expressed enzymes and with the biochemical reactions the enzymes catalyse. Every biochemical reaction catalysed by enzymes in the GEM has a GPR rule associated to it. The rule decides if the enzyme is expressed, and subsequently if the biochemical reaction in the GEM can carry flux. The GPR rule consists of the genes that express the enzyme in the form of Boolean variables. If the value of the expression for the enzyme *E* is equal to TRUE then *E* is expressed, and the bounds of the associated reaction(s) are the user-defined ones, however if the value is FALSE, then both the lower and upper bound are zero and the reaction(s) cannot carry any flux in the FBA optimisation. Initially, all the genes' Boolean variables have a value of TRUE and thus all GPRs are TRUE, allowing all biochemical reactions of the GEM to carry flux. A gene's Boolean variable of TRUE means that the gene can be translated to a functional enzyme. The knockout or a disruption of the gene *G* result its Boolean variable to change value to FALSE and the GPRs associated with it, should be recalculated again in order to determine whether the result would be FALSE or TRUE. For example, the enzyme ferroxidase (FEROc) is expressed by five different genes and its GPR is GPR^{FERO_c}: *Cp* OR *Heph* OR [*Fth1* AND (*Ftl* OR *LOC100764157*)]. If both *Cp* and *Heph* and *Fthl1* are knocked out, then their Boolean associated variable changes to FALSE from TRUE leading to the expression GPR^{FERO_c} changing to FALSE and therefore the reaction FEROc cannot take place, as the enzyme cannot be expressed. A continuous variable for the gene expression, showcasing the degree of expression, can be implemented into the GEM using methods such as GIMME (*90*) or GX-FBA (*177*), but due to lack of transcriptomic data, the GPR rules of the GEM was the best way to model gene knockouts and enzyme expression from genes.

The first step before detecting any modifications in the GEM is to identify the reactions that must carry non-zero flux to ensure the model's feasibility in biomass or antibody maximising optimisation. The strain design algorithm used herein takes as input and results as output not genes but reactions, so the aim is the detection of reactions essential for biomass or antibody production in the CHO-K1 GEM. The goal is after the implementation of the strain design algorithm to come up with an *in vitro* application of a genetic strategy, so the reactions should not be connected to an essential for growth gene. The metabolic model of CHO-K1 contains several genes that if their Boolean variable changes from TRUE to FALSE, then there would be no flux produced if the objective function of the optimisation is either biomass or antibody production. These genes are defined as essential genes ($\mathbf{essential}^{\mathbf{genes}}$). The essential genes are responsible for the expression, through their GPRs, of essential reactions ($\mathbf{essential}^{\mathbf{r_genes}}$). This set of essential reactions should not be identified by testing whether a reaction is essential, because there can be a reaction of the CHO-K1 GEM which by itself is not essential, but the gene expressing it is. For example, the essential gene, *Rrm1*, co-expresses multiple Ribonucleoside-diphosphate reductases but at least two of them must carry flux to produce biomass in

the FBA optimisation. The detection of essential genes will help when defining a pool of possible reactions for the strain optimisation algorithm.

Apart from reactions that are expressed by essential genes, there are also several other reactions which should not be included in the reactions of the regulation pool in the optimisation algorithm. Reactions that have no GPR association and must carry non-zero flux for biomass or antibody to be produced from the CHO-K1 GEM (**essential^{r}**) should be identified. There are GPR associations for the 1763 out of 3952 reactions of the CHO-K1 GEM, the rest of the reactions are either drain with no gene expressing them, or reactions with unknown gene annotation, or the current GEM iCHO1766 does not have the GPR rule. Latest versions of the CHO GEM contain more genes and GPRs, like the iCHO2101 (*77*), but at the point of performing the optimisations iCHO1766 was currently the latest published GEM for the CHO cell. Therefore, to detect the essential reactions with no associated GPR it is not possible to change the Boolean variables from TRUE to FALSE of the genes in the GEM, but to check individually one by one the reactions whether its non-zero flux is necessary to produce biomass and antibody. Several reactions are expressed by multiple genes (**essential^{r_gpr}**), like the ATP synthase, and should be separately investigated regarding their essentiality. The sum of all essential reactions, those that arise from essential genes, those that do not have GPR associations and those that are expressed by multiple genes, form the set of essential reactions which must carry a non-zero flux for the FBA optimisation to be feasible and produce biomass or antibody.

The identification method for each of the three reaction sets **essential^{r}**, **essential^{r_genes}** and **essential^{r_gpr}** follows a similar iterative approach. The detection of essential for biomass or antibody production reactions, **essential^{r_genes}**, requires initially the identification of essential genes, then using the GPRs they can be converted into essential reactions from the enzymes that the genes express. Following an iterative approach, the genes of the CHO-K1 GEM were knocked out one-by-one and the value of the associated Boolean variables changed from TRUE to FALSE. In each iteration, the associated Boolean expression for each gene was evaluated. If the output of the expression was FALSE then all the reactions connected to the gene and expressed from the GPR would have both bounds equal to 0 (knockout), else both bounds would be taken from the corresponding datasets, D1, 2, 3, 4. Lastly, the value of biomass and antibody produced flux was checked in order to verify the requirement of the gene for the production of both objectives. The detected genes were converted to reactions, **essential^{r_genes}**, using the associated GPRs. Following a similar approach, iterating over the 2189 reactions with no GPR associations, the **essential^{r}** were detected. The remaining 1763 reactions were also checked to identify reactions that require the expression of multiple enzymes and could not be identified by a single iteration of genes, **essential^{r_gpr}**. The three sets of reactions are then joined to form the set of reactions which would not take part in the reactions considered for knockouts and downregulations by the strain design algorithm. The essentiality study was performed for all four datasets and the results can be seen in table 4.7.

Table 4.7: Number of essential genes (***Essential^{Genes}***), reactions associated with the essential genes (***Essential^{r_genes}***), essential reactions with no GPR association (***Essential^{r}***) and essential reactions with GPR associations that do not belong in ***Essential^{Genes}*** (***Essential^{r_GPR}***) for each of the four different datasets. The last row represents the total sum, with no duplicates of essential reactions for each dataset

	D1	*D2*	*D3*	*D4*
Essential^{Genes}	130	126	127	127
Essential^{r_genes}	176	161	163	163
Essential^{r}	44	36	36	36
Essential^{r_GPR}	16	14	14	14
Compendium	236	211	213	213

There are a few important conclusions concerning the bounds of the different datasets that are concluded from the essentiality study. As seen in table 4.7, D2 has one less essential gene than D3 and D4; the remaining 126 other genes are the same. The essential gene that D2 is missing is essential gene *Asrgl1*, responsible for the catabolism of asparagine to aspartate though the enzyme L-asparaginase. The reason is that D3-4 contain more asparagine in the feeding composition which translates to increased uptake of asparagine than D2 in the FBA optimisation. The increased uptake means that there is more asparagine than the cell requires for just biomass production, and it needs to be catabolised thought the enzyme L-asparaginase. The D2-4 datasets have three less essential genes than D1, which are *Bcat1*, responsible for catabolism of isoleucine, valine and leucine, *Aldh4a1*, responsible for catabolism of proline to glutamate and *s0001*, which is not an annotated gene but expresses the spontaneous reactions that converts proline to glutamate through L-glutamate 5-semialdehyde dehydratase. D1 has bounds that uptake much more proline that is needed for biomass production, so it needs to be consumed similarly to asparagine above. *Aldh4a1* converts the remaining proline to glutamate. *S0001* gene expresses similar enzymes with *Aldh4a1,* and it is essential for the same reason as *Aldh4a1*. Lastly, the *Bcat1* expressed enzyme responsible for the catabolism of isoleucine, leucine, and valine is essential for the optimisation not to be infeasible for catabolism of leucine and valine in the D1 dataset. The difference in essential genes leads to a different number of essential reactions in the sets ***Essential^{r_genes}*** of each dataset. Regarding the reactions that are expressed by multiple genes in the set ***Essential^{r_GPR}***, it contains reactions such as ATP synthase or Sphingomyelin synthase which are required for either biomass or antibody production and are expressed by multiple genes. D1 dataset contains two more reactions in the set ***Essential^{r_GPR}*** which are NOS1 and NOS2 and are responsible for catabolising arginine. In D1 dataset, arginine is in abundance, so the whole pathway NOS1 and NOS2 need to be active in order to catabolise the remaining arginine after the biomass or antibody production. The non-GPR essential reactions are the same 36 for all datasets, but D1 contains eight

more reactions which are the secretion of citrulline, drain reactions of arginine and lysine metabolism and PROD2 which catabolises proline to glutamate. Even though, D1 and D2 should result in the same essential genes and reactions, the datasets are taken from different clones of the CHO cell lines and experiments, which corresponds to different measurements of the metabolites resulting in different uptake rates between the two datasets. The four different datasets gave different essential reactions that if were removed from the CHO-K1 GEM, there would be no biomass or antibody flux in an FBA optimisation. In each optimisation, depending on the dataset, the compendium of the different sets of reactions was removed from the knockout pool.

The above method detects the essential genes and metabolic reactions that their knockout has a significant effect on the production of biomass or antibody in an FBA optimisation of the CHO-K1 GEM. Similar work has been done by Chowdhury et al. (*178*), who using MILP optimisation on the CHO 1.2 GEM (*109*) were able to detect essential genes, and propose ways to make the model more robust and consistent concerning its genes and GPRs. CHO cells have a similar genome to mouse cells (*179, 180*) and studies like Tian et al. (*181*) who used machine learning on phenotypic mouse data – data containing changes in the cell's phenotype from gene knockout experiments - to predict gene essentiality; and Kabir et al. who studied gene data for biological information in order to detect the essential genes for murine cells have been useful in the past for getting a hint whether a gene is essential in CHO cells (*182*). However, some genes are classified as essential because of their biological role in the cell survival, and it would be difficult to predict them through computational work. For example, *Gapdh* is responsible for the conversion of g3p to 13dpg in the glycolysis pathway. If *Gapdh* is removed, the optimisation finds different pathways to convert g3p to pyruvate (i.e., thought Methyglyoxylate synthase). However, it is a ubiquitous gene, playing an important part in the energy metabolism and used as a housekeeping gene. The gene *Pi4kb* is also essential for growth. It takes part in cellular signalling and trafficking in the Golgi apparatus. However, the CHO-K1 GEM does not recognise the genes as essential in none of the four datasets, so experimental testing for the essential of genes and enzymes is important. In a recent study, Xiong et al. performed over 75,000 nicks in the genome of CHO-S cells using Cas9, thus resulting in a list of essential genes for growth (*183*). The Xiong et al., database of essential genes was published after the strain design optimisation had been carried out, so it was used in the post-hoc analysis of the proposed genetic engineering strategies to verify which genes can be removed from the genome.

4.4 Strain Design Algorithm

There has been a plethora of strain design algorithms developed, which are described in the section **2.2.2** of the literature review. Three were considered and OptReg (*138*) was deemed to be the most applicable for the specific optimisation problem. Initially, OptForce (*139*) was studied regarding its applicability on the CHO-K1 GEM, as it was the most widely known and cited at the time of research.

The OptForce algorithm performs regulations on the reactions of a metabolic model with the goal of coupling two objective functions and achieving the phenotype of interest. It applies modifications only on reactions that belong in the MUST sets, narrowing down the number of binary variables responsible for regulations. A MUST set contains the metabolic reactions whose flux range in the overproducing strain – objective of the FBA optimisation is maximisation of the product of interest – does not overlap with the range of the fluxes in the wild-type strain – objective of the FBA optimisation is maximisation of biomass production. However, no MUST sets were detected for the iCHO1766 model for any of the datasets. The reason was probably the large size of the model and the large solution space by the optimisation, because a number of MUST sets were detected for the - smaller in size - *E. coli* model iJO1366 (*184*) using just a large uptake of glucose as constraint, together with the required for growth nutrients. Even though the idea behind k-OptForce (*140*) is very interesting, as it couples dynamic equations to calculate the steady state fluxes of the intracellular reactions, the algorithm was not considered either, as it is based on the same concepts of MUST sets. Then, OptKnock (*137*), which is a brute force search version of OptForce was studied since it considers all reactions instead of only those that satisfy a certain criterion. The algorithm did return optimal solutions, but OptReg seemed like a better and more versatile option, since it also considers up- and down- regulations. OptCouple (*142*), a recently developed bilevel algorithm, which considers changes in the media as well as gene engineering strategies, was not considered since the scope of this thesis is the applications of reaction modifications in the metabolic network and not in the media. In conclusion, OptReg is the strain design algorithm of choice and after a few adjustments to the GEM and datasets, it was used to provide genetic strategies that can alter the metabolism of CHO cells in a way that it would increase the specific productivity of IgG during the exponential growth of the culture.

4.4.1 Optimisation Settings

Performing a strain design using the Bio_Igg or the Igg_Igg algorithm on the CHO-K1 GEM requires some preliminary work. The first step is the removal of reactions that cannot be modified by the strain design algorithm due to biological reasons. There are reactions that are associated with genes essential for the growth or antibody production of the cell and reactions which are co-expressed by multiple genes (more than four), and it would be difficult to perform an *in vitro* validation later on. As described in the previous section **4.3.2** the essential genes and reactions were calculated in each dataset, there is a certain number of essential reactions that if they were to carry zero flux it would render the model infeasible in producing either of the two objectives. The essential reactions will not be suggested by the optimisation but they are removed from the knockout pool of reactions to speed up the algorithm and save computational memory. Enzymes co-expressed by more than four genes, which are associated with 337 reactions of the CHO-K1 GEM and are difficult to modify in practice were also removed from the pool of knock-out, up- and downregulation for all four datasets. The optimisation algorithm is focused

on providing changes in the intracellular fluxes, so the 420 drain reactions were removed from the pool, as changing media or feeding composition was not the scope of this optimisation. After the application of the elemental balance algorithm, 693 reactions were categorised as blocked (both lower and upper bounds are zero) for all four datasets and thus they were removed from the optimisation pool. Lastly, important reactions such as the biomass reaction and the reactions associated with antibody production (heavy and light chain synthesis, antibody formation, and its secretion) were also removed from the pool. In conclusion, the total number of reactions removed from the knock-out pool, depending on each dataset, is shown in table 4.8.

Table 4.8: Number of reactions that were removed from the possible reaction pool for modification in during the strain optimisation of the CHO-K1 GEM.

	D1	*D2*	*D3*	*D4*
Essential Reactions	236	211	213	213
Removed Reactions from KO pool	1379	1354	1356	1356
Removed Reactions from Up- and Down- pool	1098	1098	1098	1098

Reactions that are associated with diffusion and those that do not have a GPR association remained in the GEM as a possibility for a reaction regulation. Even though they might be difficult to regulate, they provide insights and hints on other important reactions in the metabolic network that if regulated could result the same phenotype. There is also the possibility of detecting the GPR association later with a search in databases. All the reactions detected with the above method using the reversible CHO-K1 GEM, but the strain design algorithm takes as input irreversible reactions as explained above. Converting the CHO-K1 GEM from reversible to irreversible format is a simple process which requires the substitution of all the reversible reactions with a forward and a backward one, which can be done using the function *convertToIrreversible* from the COBRA toolbox (*156*). As a result, the size of the GEM from 3952 reactions went to 4940.

The final step in performing the strain design optimisation is defining the five user-defined parameters: the number of allowable regulations, the dataset to constrain the GEM, the upregulation and downregulation coefficient and the applied algorithm. The allowable number of regulations ranged from ten, because below ten regulations the optimisation was infeasible, to 225, as it was a large enough number of regulations to contain most optimal solution. Both the up- and downregulation coefficients ranged from 0.4 to 1. The range of the values was chosen to be large enough to contain most possible optimal solutions. However, the initial solutions contained no reaction downregulations as a suggestion by the optimisation leading to the removal of the downregulation option from the remaining

optimisations. The downregulation option corresponds to several equations and thousands of binary variables in the optimisation algorithm, so its removal reduced the complexity and time to reach an optimal solution of the optimisation while still resulting in optimal solutions. Finally, for all four datasets both algorithms Igg_Igg and Bio_Igg were used for the strain design optimisation. To get the maximum possible results, all the variables were varied iteratively to capture all the design space as presented in table 4.9.

Table 4.9: Parameters of the strain design optimisation process. The upregulation coefficient corresponds to the degree of increase in the lower bound of the modified reaction. The allowable number of regulations, *k*, is the number of regulations allowed by the optimisation. The algorithm is the type of algorithm Igg_Igg and Bio_Igg. The final option, dataset, denotes which of the four datasets is used to constrain the CHO-K1 GEM.

Variable	*Upregulation coefficient, e_up*	*Allowable number of regulations, k*	*Algorithm*	*Dataset*
Values	0.4, 0.5, …, 1	10, 20, …, 230	Bio_Igg, Igg_Igg	D1, D2, D3, D4

The total number of optimisations performed was over 1300 since integer cuts were also applied when a successful solution was found to detect other possible combinations with the same number of number of regulations, *k*. Out of all the optimisations, 93 returned a unique optimal solution that would couple the two objective functions in an FBA optimisation with biomass production being the objective function, seen in figure 4.4, and would produce positive flux of IgG.

$$
\begin{aligned}
&\text{Maximize } z = v_{biomass} \\
&\text{Subject to:} \\
&\quad \sum S_{ij} v_j = 0 \\
&\quad v_j \geq \left(LB_j + \left(UB_j - LB_j\right) \cdot e_{up}\right) \cdot y_j^{up} + LB_j \cdot \left(1 - y_j^{up}\right) \\
&\quad v_j \geq LB_j \cdot (1 - y_j^{KO}) \\
&\quad v_j \leq UB_j \cdot (1 - y_j^{KO})
\end{aligned}
$$

Figure 4.4: Formulation of FBA optimisation that implements the changes in the reactions' bound proposed by the strain design algorithm. An optimal solution from the algorithm should provide a positive antibody flux when the regulations are applied to the reactions' bound of the dataset used to acquire the solution.

The optimisations were performed in GAMS (version 28.2.0) with the use of the solver Cplex (version12.9.0.0) using the default settings. They were run on the departmental cluster of chemical engineering in Imperial College London, with 16- and 24- Cores and 48GB of RAM.

4.4.2 Optimisation Results

After the application of the two strain design algorithms Igg_Igg and Bio_Igg on the CHO-K1 GEM constrained by the four different datasets, 93 different sets of regulations on the intracellular reactions bounds of the GEM were suggested. These sets of regulations form a scenario which can change the Pareto curve of the original model as described in section **2.2.1**, to one that the maximisation of biomass in an FBA optimisation results in positive flux of IgG as in the four example curves shown in figure 4.5.

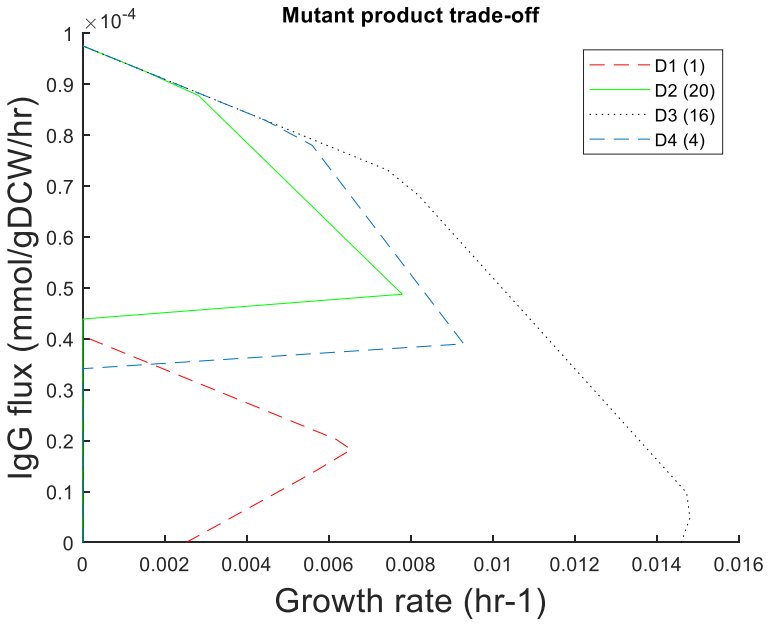


Figure 4.5: Trade-off curves between biomass and antibody production for four scenarios (in parenthesis) that resulted from the strain optimisation.

The Bio_Igg algorithm returned 53 optimal solutions with an average of 131 regulations per solution while Igg_Igg returned 40 solutions with 128 regulations per solution on average. The Bio_Igg algorithm took 3.5 CPU hours on average for the successful optimisations whereas the worst-case scenario algorithm Igg_Igg was much faster in reaching an optimal solution and only took 1.5 hours. Potential reasons include the use of minimisation instead of maximisation in the inner optimisation, the use of a biomass constraint in Igg_Igg in the inner optimisation versus the use of a lower bound in the variable of biomass production in Bio_Igg algorithm, and the use of the same variable in both inner and outer optimisation in Igg_Igg. Concerning the different dataset used, both algorithms Igg_Igg and

Bio_Igg resulted in almost the same number of optimal solutions for each dataset. Overall, out of the 93 solutions, 22 use D1 dataset, 16 D2, 11 D3, and 44 use D4. The fourth dataset has wider bounds than the other three datasets due to the increased feeding quantities used, which is a potential reason it returned more results than the other three datasets. Both algorithms, despite the difference in their objective functions, returned approximately the same number of optimal solutions. However, the average number of regulations for all 93 solutions is 130, a number too high to allow any experimental validation or study of the suggested alterations in the metabolic network.

Regarding the quality of the optimisation results, out of the 40 optimal solutions returned by the use of the Igg_Igg algorithm, the average growth rate is $0.011 \pm 0.005 hr^{-1}$ and the mAb specific productivity is $22.978 \pm 16.912 pg/cell/day$, whereas Bio_Igg out of the 53 solutions it resulted in growth rate $0.009 \pm 0.005 hr^{-1}$ and mAb specific productivity of $30.779 \pm 20.301\ pg/cell/day$. These averages, even with their high standard deviation, are an initial metric for the quality for the optimisation results. They include the solutions from all four datasets, which independently result in different growth rate and antibody productivity values. The difference in both μ and q_{mAb} shows that the replacement of biomass in the objective function with of a constraint in Igg_Igg results in higher μ; but the use of antibody minimisation in the outer optimisation, to simulate a worst-case scenario, instead of the use of maximisation of biomass results in lower q_{mAb} in the Igg_Igg algorithm, compared to Bio_Igg.

An average of growth rate and specific productivity between all optimal solutions for the two optimisation algorithms, might not be subjective criteria to compare the two algorithms. Each individual solution yields both μ and q_{mAb} and they should both be considered. For example, even though the solutions from Igg_Igg, on average, might produce a result with higher growth rate than a solution from Bio_Igg, q_{mAb} could be very low making the suggested genetic strategy unideal for an *in vitro* application. A more subjective way to judge the two algorithms is to add μ and q_{mAb} for each individual solution. Initially all the values of μ and v_{IgG} are normalised, because the difference in order of magnitude between the two values is around 1E2, using the formula

$$x_{i,norm} = \frac{x_i - x_i^{min}}{x_i^{max} - x_i^{min}} \tag{4.10}$$

The new values μ^{norm} and $q_{mAb}{}^{norm}$ are then summed to yield alpha, α, a variable denoting the quality of each individual solution of the optimisation

$$a = \mu^{norm} + v_{IgG}^{norm} \tag{4.11}$$

Finally, the average α for all the optimal solutions produced by the Bio_Igg and by Igg_Igg algorithms was calculated. Bio_Igg has an average of 0.741 ± 0.249 for all 53 solutions and Igg_Igg

has an average of 0.749 ± 0.249 , which argues that both algorithms result in, on average, the same quality of results regarding μ and v_{IgG}. From all the above conclusions, the algorithms Igg_Igg produces an optimal solution much faster than Bio_Igg, and the solution is more likely to have lower q_{mAb} than Bio_Igg. To study the quality of the results further, the analysis of individual scenarios is necessary which takes place in the following chapter.

4.5 Conclusions

This chapter focused on designing CHO-K1 strains with increased q_{mAb} compared to the original CHO-K1 GEM. The idea is based upon coupling μ and v_{IgG} in the GEM, so that in an FBA optimisation with growth rate being the objective function, there would be a positive antibody flux being secreted from the cell as a by-product of the maximisation. Translating this *in vitro*, the cell would direct more carbon sources and energy toward antibody production, resulting in more antibodies being secreted from the cell than in the original derived IgG-producing cell line. The bounds used to constrain the GEM were taken from the exponential phase of growth, so the increase of q_{mAb} would take place during that phase of culture.

Initially, the values of extracellular concentrations of the experimental datasets, D1-4, were converted to fluxes. The datasets were then augmented to include bounds for intracellular reactions with the use of three algorithms, FVA, ccFBA, and ncFBA. The curated datasets contained bounds for two thirds of the intracellular reactions based on the total carbon and nitrogen uptake in each of the datasets, leaving the reaming to be constrained by FVA. As a result, the flux distributions resulting from the FBA optimisation and from the strain design algorithm obeyed chemical balances. Using the four new datasets to constrain the original CHO-K1 GEM, an FBA optimisation was performed on the model to predict the values of μ and q_{mAb}, which were compared with the experimental ones for each dataset, concluding that FBA predicted the values with high precision in all datasets except D1. A main difference between D1 and D2-4 is that D1 contains a bound for cysteine, whereas D2-4 borrow a cysteine bound from a different experiment conducted with similar experiments conditions. This difference in the uptake of the amino acid results in less accurate FBA predictive results for the first dataset. As a preliminary work for the strain design, the essential reactions for biomass and IgG production were detected for each of the four datasets, which were removed from the optimisation pool of the knock-out reactions, together with blocked reactions, reactions that are expressed by more than four genes and the drain reactions which were also removed from the up- and downregulation pool. Finally, the strain design algorithm, OptReg, was modified into two algorithms Bio_Igg and Igg_Igg which were then applied to the CHO-K1 GEM. The strain optimisation resulted 93 optimal solutions for the four datasets, each containing different set of regulations on the reactions' bound to achieve the phenotype of interest. Overall, the speed that Igg_Igg reached an optimal solution is almost half the

time of Bio_Igg, but the genetic engineering strategies proposed by both have similar number of proposed regulation and growth rate and specific productivity. Even though, Igg_Igg seems like a better choice for strain optimisation, the individual analysis of the genetic strategies could provide valuable insights regarding the quality of the results.

The experimental validation would include the application of the suggested by the algorithm reaction regulations. However, the average number of regulations for the 93 genetic regulations is too high to perform *in vitro,* so the next step is reducing the total number of regulations. Furthermore, out of the 93 different scenarios, it is important to choose the scenarios that are more likely to succeed in a genetic engineering experiment *in vitro*, as genetic engineering is expensive and time-consuming as a technique. After the creation of the new high-producing strain, a fed-batch culture will be carried out to simulate the experimental conditions of the datasets used. The hypothesis is that the growth rate would show a slight decrease compared to the original CHO cell line, but the q_{mAb} would increase in the new engineered CHO cell line.

5

Sensitivity Analysis

The optimisation of the CHO-K1 metabolic network resulted in 93 optimal solutions – scenarios – each consisting of a set of regulations on the reactions' upper or lower bounds. The scenarios can couple the biomass and antibody production in an FBA optimisation that maximises biomass production. The optimisation is performed on the CHO-K1 GEM, which is constrained by each of the datasets and obeys the suggested scenario fluxes. However, many of those scenarios have a large number of changes in metabolism, 130 regulations per scenario on average, making both the implementation *in vitro* and the study of the suggested metabolic changes very challenging. Hence, a sensitivity analysis framework was developed to reduce the total number of regulations needed for antibody production increase to take place in an FBA optimisation. The analysis removed unnecessary regulations and kept only the vital ones that are responsible for the required phenotype, reducing the average number of regulations to 9.3 per scenario. Afterwards, the scenarios with the highest predicted monoclonal antibody specific productivity, q_{mAb}, and growth rate, μ, in an FBA optimisation were identified as high producing scenarios. The values of the predicted μ and q_{mAb} were compared to the FBA predicted values for the corresponding datasets - control values – which were calculated in the section **4.3** of the computational strain design, and only the scenarios with both values higher than half of the control values were selected. The 20 high-producing scenarios were then analysed regarding the proposed changes in metabolism. From the analysis a hypothesis is derived which explains the reason behind the coupling of the two diametrical objective functions in the FBA optimisation. According to the results of the optimisation algorithm, antibody synthesis and biomass production can be coupled in an FBA optimisation by creating an abundance in a metabolite inside the cell combined with a reduction of the maximum possible biomass production. FBA assumes a pseudo-steady state, thus there is no accumulation of metabolites. An abundance can be formed by removing possible catabolic pathways of an amino acid. When that happens, FBA finds all the other possible pathways that consume the metabolites. Initially the main pathway would be to direct these metabolites towards the objective function, biomass production in the CHO-K1 for example, however, if an inhibition of the objective function is applied then FBA would direct the abundant metabolites to even pathways that are competitive, such as the IgG synthesis. Lastly, a way to perform *in vitro* validation of three possible scenarios, each changing a different part of the metabolism, is proposed.

5.1 Sensitivity Analysis

The optimisation algorithm applied to the CHO-K1 GEM constrained by the D1-4 datasets resulted in 93 different optimal solutions, each containing a set of regulations applied on the lower or upper bounds of the metabolic reactions, that can couple the production of biomass and antibody in an FBA optimisation. Each solution represents a scenario with a specific set of regulations ranging from 25 to 225, with an average of 130 regulations per scenario, making it extremely difficult and expensive to perform an experimental validation *in vitro* or even study the suggested changes in the metabolic network and to try to understand the reason why the coupling takes place. To tackle the problem, a Sensitivity Analysis (SA) framework was developed, as explained in section **3.4** of the computational methods, with the goal of removing the reactions that have no effect on the resulting phenotype and keeping in the regulations pool of each scenario only the vital ones – those that are necessary to couple growth rate and antibody production. The parameter responsible for the number of regulations allowed by the optimisation algorithm is *k*, is left relaxed with a value in the range of [10-225], as explained in section **4.4.1**. The reason it was not constrained to a smaller number [2-10] to try and instantly get the smallest number of regulations, was that the introduction of such a strict constraint would complicate the optimisation problem due to the large number of binary variables and equations; thus, by relaxing the k-value the algorithm could find optimal solutions faster and subsequently apply meta-analysis to the results to get the minimise the number of regulations. The complete SA framework can be seen in figure 5.1.

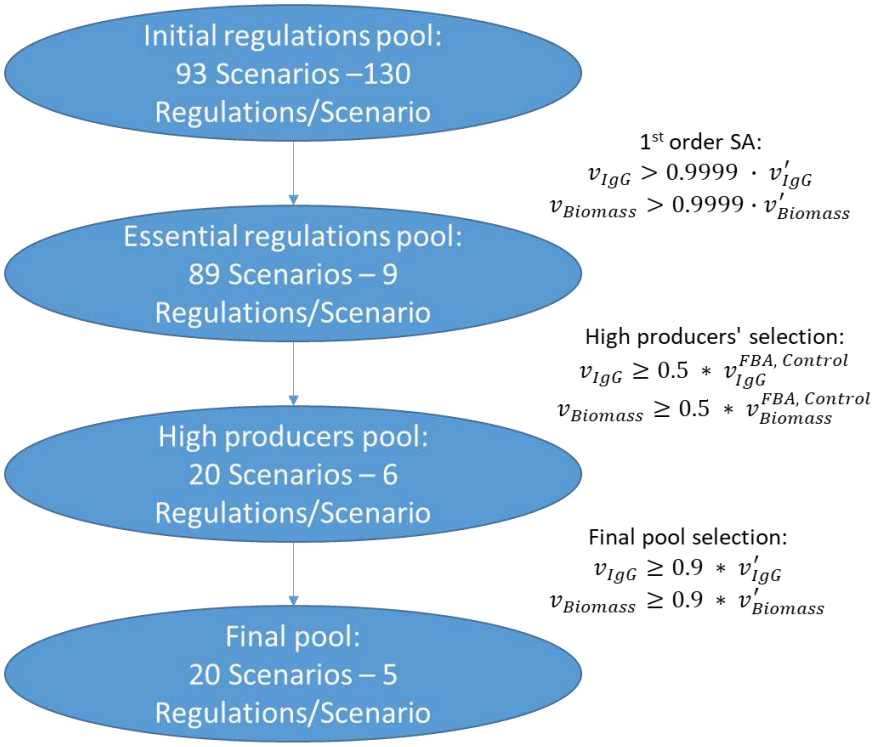


Figure 5.1: Framework of the SA methodology used to reduce the total number of genetic regulations and identify the most optimal genetic strategies. The v_{IgG} is the secreted IgG flux and $v_{biomass}$ is the produced biomass by the FBA optimisation of the CHO-K1 GEM with biomass maximisation. The v'_{IgG} is v_{IgG} when all the initial regulations of a scenario are active and $v_{IgG}^{FBA,exp}$ is the FBA predicted control value of the dataset that the scenario belongs. A similar methodology is applied to the predicted biomass.

The first stage of SA that was applied to the regulations pool of each scenario, removed the regulations that have no impact on the output, as explained in section **3.4** of the computational methods. As a result, the regulations in the final pool are only those that are required for the production of biomass and antibody. SA removed a plethora of regulations for each scenario, reducing the average number from 130 to 9. Two are the most important type of regulations not affecting the coupling of the two objective functions which were removed from the regulation pool after the initial step of SA: regulations on the bounds of transport reactions and downregulations of reactions. An abundance of the regulations proposed by the 93 scenarios consist of transport reactions, 7506/11845 or around 58% of the regulations from all the scenarios. After the reduction algorithm, the number of transport reactions proposed by the optimisation algorithm dropped to 300/800 or about 40% for all the regulations. The majority of transport reactions removed are part of type III thermodynamically infeasible cycles that generate energy in the form of co-factors such as ATP or $FADH_2$. The cycles are not necessary for the coupling of the objectives and removing them does not disrupt the required phenotype. The reason they were proposed by the strain optimisation was due to the large value of the *k* parameter, which allowed

the optimisation to search in a large solution space. Furthermore, SA removed the regulations that downregulate the fluxes of reactions as seen in figure 5.2.

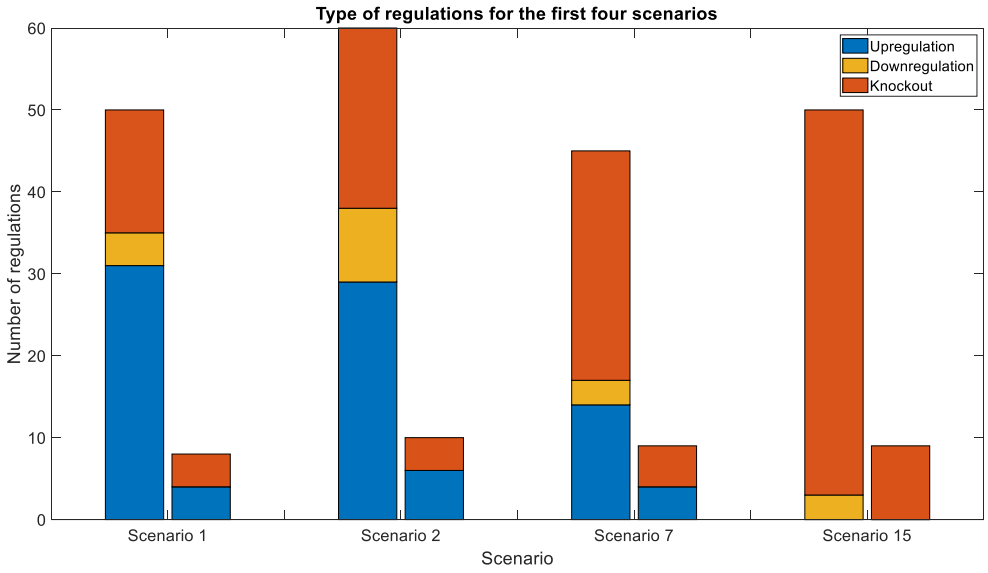


Figure 5.2: Reduction of the number of regulations in the first four scenarios that contained downregulation option using the initial stage of SA. Scenarios 1, 2, 7 and 15 all suggest downregulating the fluxes of reactions, among the knockout and upregulation. After the application of SA, the downregulations in each scenario were removed.

In the first four optimal solutions that contain reaction downregulations, SA removed all downregulations from the regulation pool, as they were unnecessary for the required outcome. Thus, the option for a reaction to be downregulated was removed from the remaining optimisations as mentioned in section **3.4.3** of the computational strain design, since it seemed that the optimisation prefers the option of knock-out. A benefit of removing the downregulation option is the increase in speed of the optimisation, as just the downregulation accounts for around 8,000 binary variables. Downregulation of a reaction is generally very similar to knock-out *in vitro*, and by trying to perform a knock-out a lot of times a knock-down takes place. As is analysed in the later paragraphs, the application of a knock-down instead of a knock-out can yield the same predicted phenotype from the FBA optimisation. Thus, in the post-SA results none of the scenarios contain any downregulations, only upregulations and knock outs.

As a result of the SA, there was a significant decrease in the average number of regulations for all 93 scenarios, resulting in an average of 9 regulations per scenario and reaching as low as two in some scenarios. Moreover, four of the scenarios, after the application of SA, resulted in the same regulations, dataset, and algorithm and upregulation coefficient. The final number of scenarios after the merge of the duplicate ones was 89, with an average number of regulations of 9. Many of the 89 scenarios had either very low μ or v_{IgG}, making them unsuitable candidates for genetic engineering,

thus the next step is picking only the high producing scenarios. Their selection is based upon the FBA predicted μ and v_{IgG} in each scenario, and how they perform compared to the control values calculated from the original CHO-K1 GEM in section **4.3** of the computational strain design. The values were not compared to the experimental ones, because the idea is how well the proposed genetic strategy performs compared to the initial FBA predictions of the original CHO-K1 GEM. As explained in the section **3.4** of the computational methods, out of the 89 scenarios, only those with both μ and v_{IgG} higher than at least 50% of the predicted FBA control values were considered as high producing scenarios. The final 20 high-producing scenarios, have the most chances of creating a high-performing strain, and are presented in table 5.1, and in more detail in the table A2.

Table 5.1: The 20 high producing scenarios with higher μ and v_{IgG} than 50% of the control values calculated for the original CHO-K1. The first column specifies the type of algorithm used, Bio_Igg or Igg_Igg, followed by the scenario number, out of the 89, in column two. The third column specifies the dataset (D1-4) used for constraining the GEM. Then μ and v_{IgG} are compared against the respective control values in columns 5&7. Finally, the number of regulations to achieve the specific outcome, after the application of 1st order SA, is specified in the last column. To calculate the specific productivity from v_{IgG} the same methodology as in paragraph **3.1** was used.

Algorithm	*Scenario*	*Feed*	*Growth Rate (hr^{-1})*	*Change in μ*	*Specific Productivity (pg/cell/day)*	*Change in q_{mAb}*	*Number of Regulations (Post SA)*
Control		*D1*	0.0088	-	16.2000	-	-
Bio_Igg	1		0.0066	↓ 25%	21.2311	↑ 31%	8
Bio_Igg	3		0.0057	↓ 35%	24.7229	↑ 53%	4
Igg_Igg	9		0.0046	↓ 48%	28.9245	↑ 79%	11
Igg_Igg	12		0.0101	↑ 14%	8.4772	↓ 48%	5
Igg_Igg	27		0.0055	↓ 38%	25.4711	↑ 57%	9
Bio_Igg	81		0.0062	↓ 30%	22.9280	↑ 42%	9
Bio_Igg	84		0.0066	↓ 25%	21.4020	↑ 32%	8
Bio_Igg	86		0.0055	↓ 38%	25.6321	↑ 58%	5
Bio_Igg	88		0.0085	↓ 4%	14.3019	↓ 12%	20
Control	76	D2	0.0188		48.0000		
Igg_Igg	30		0.0099	↓ 47%	63.6437	↑ 33%	13
Igg_Igg	31		0.0173	↓ 8%	46.0704	↓ 4%	2
Igg_Igg	74		0.0111	↓ 41%	32.7011	↓ 32%	10
Control		*D3*	0.0223	-	31.7000	-	-
Bio_Igg	32		0.0205	↓ 8%	34.0823	↑ 8%	2
Bio_Igg	33		0.0205	↓ 8%	34.0823	↑ 8%	2
Igg_Igg	47		0.0144	↓ 45%	30.115	↓ 5%	8
Control		*D4*	0.0215	-	51.8000	-	-
Bio_Igg	14		0.0133	↓ 38%	61.0383	↑ 18%	6
Bio_Igg	49		0.0130	↓ 40%	39.3633	↓ 24%	7
Bio_Igg	52		0.0124	↓ 43%	43.6783	↓ 16%	6
Bio_Igg	58		0.0134	↓ 38%	60.7551	↓ 17%	2
Bio_Igg	80		0.1505	↓ 30%	56.98	↑ 10%	8

The majority of the high producing scenarios, 14/20, are optimal solutions that resulted from the Bio_Igg algorithm. In the previous chapter, the conclusion was that both algorithms predict results with almost the same quality, so it would have been expected that they would result the same number of high producing scenarios. The main criterion for judging the quality of each scenario was the sum of the FBA predicted μ and q_{mAb}. Even accounting for the increased number of total solutions by Bio_igg (53) to Igg_Igg (40), the difference in the high-producing scenarios contradicts the conclusion from the last chapter. However, a high-producing scenario is chosen based on the comparison of predicted μ and q_{mAb} to the control values, and not based on the combination of FBA predicted μ and q_{mAb}. The threshold for μ and q_{mAb} is 50% of the corresponding values control values, and 45 scenarios failed to reach the cut-off for growth and 55 for their predicted specific productivity out of the total 89 scenarios. Therefore, it seems that Bio_Igg, which as calculated in Chapter 3, results in scenarios with on average higher specific productivity ($30.779 \pm 20.301\ pg/cell/day$ against 22.978 ± 16.912 of Igg_Igg) and slightly outperforms Igg_Igg when compared to FBA predicted optimal μ and q_{mAb} from experimental datasets. Therefore, it is reasonable that Bio_Igg results in more high-producing scenarios than Igg_Igg.

Finally, the same algorithm as in Fig. 3.6 was applied again but allowing μ and v_{IgG} to be 90% of v'_{IgG} and μ', which are the antibody flux and growth rate if all the regulations after the first SA take place, according to the third stage of the SA framework of section **3.4** of the computational methods. That way SA removed reactions that affect slightly the biomass production or the antibody increase; further reducing the number of regulations to make the scenarios easier to study and apply *in vitro*. After the SA, the growth rate or antibody production could be lower than 50% of the control values for the original cell line, making the scenario not as efficient. However, by changing the upregulation coefficient or by adding back a knockout regulation to the scenario regulations both the predicted growth rate and antibody production can be adjusted. The reasoning behind the 2nd stage SA was to identify the set of reactions that are responsible for the coupling of the two objective functions without keeping a strict constraint on the output, which would make easier to understand why the coupling of the objective takes place. In the end, for the 20 high producing scenarios the average number of regulations is 5 regulations, and the overall reduction in those scenarios can be seen in figure 5.3.

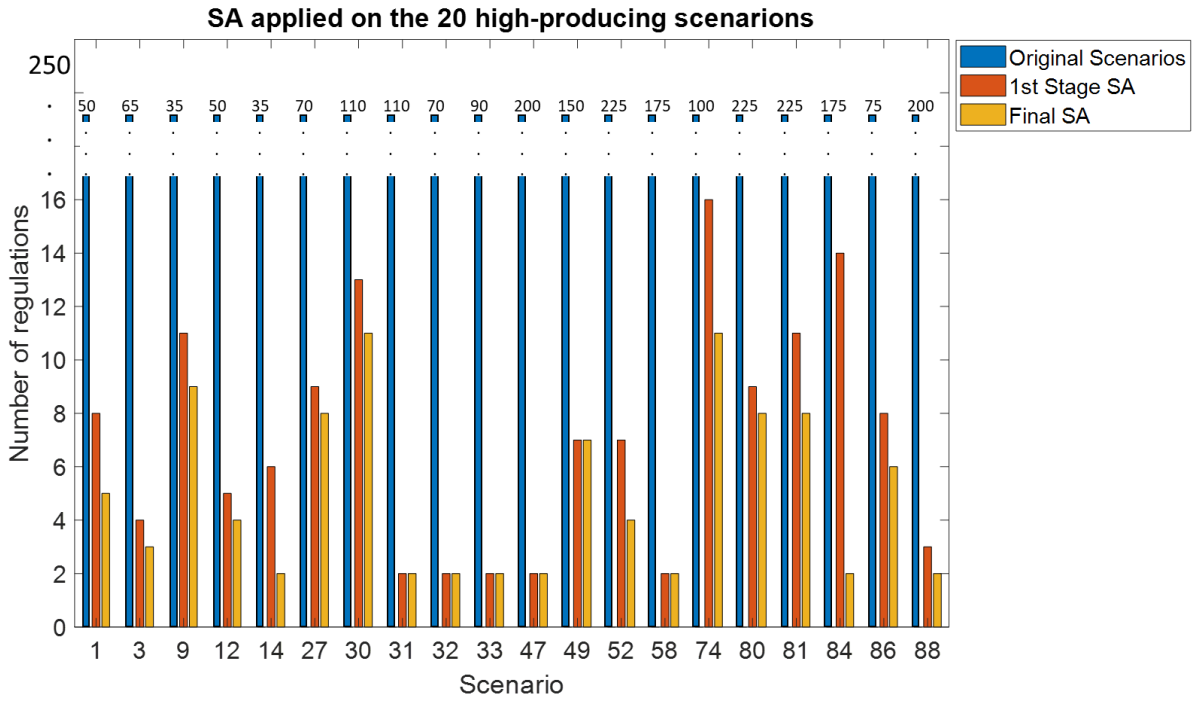


Figure 5.3: Reduction of the number of regulations after the use of SA in the high producing scenarios. The blue columns represent the original regulations for each of the 20 high-producing scenarios. After the application of the 1st stage SA the number of regulation (orange column) reduced significantly. In the yellow column, only the vital for coupling regulations are included in each scenario.

The number of regulations for each scenario was reduced significantly, and it is possible to apply the combinatorial approach to verify that the regulations of each scenario were the absolute necessary to perform the coupling of the two diametrical objective functions. Scenario 74 has the largest number of regulations, 15, and thus the 2nd – 15th order of combinations was checked in the 20 high producing scenarios, in case any of the combinations could be removed without changing either μ or v_{IgG}. The rest 79 scenarios, that predicted the lower growth rate or antibody secretion, were not considered for further SA because their metabolic networks were not analysed nor considered for *in vitro* validation. The combinatorial SA did not further reduce the number of regulations in any of the 20 high-producing scenarios, verifying that the number of regulations is the smallest possible that can achieve the coupling in each scenario. If the threshold in biomass or antibody production reduces, then the number of regulations can be reduced further, but in this stage a manual analysis of each scenario is possible, and it can result in additions or subtractions of regulations.

5.2 Optimisation results

The strain optimisation resulted in 89 different scenarios, where each contained a set of regulations for the bounds of the intracellular reactions that synthesise antibody in a biomass maximisation FBA. SA

reduced the total number of regulations of each scenario by 121 on average, keeping only the vital regulations required for the coupling of the two diametrical objective functions. Furthermore, 20 of the 89 scenarios with the highest combination of q_{mAb} and μ were selected, and their number of regulations was reduced further by allowing the value of the predicted IgG to reduce from v'_{IgG}. The 20 scenarios have only 5 regulations on average, allowing to be studied regarding the suggested changes in the metabolic network.

In the previous chapter the competitive nature of the two objective functions, growth rate and antibody production were analysed; so, it is expected that the proposed regulations also reduce the maximum biomass production form the FBA optimisation. The results of the optimisation algorithm are focused on limiting the maximum possible growth rate of each dataset, which the FBA optimisation can predict. However, it is not enough to reduce the maximum produced biomass as there is no reason for the metabolites to be directed to antibody synthesis. The strain algorithms suggest that, by creating an abundance of an amino acid in the cell, antibody needs to be synthesised in order for the amino acid to be consumed. To create an abundance of an amino acid two things are required, first a limit in the secretion of the amino acid by the cell, which happens because all datasets have upper and lower bounds for the secretion/uptake of amino acids based on experimental data; and secondly the removal of certain catabolic pathways as long as it is not essential for the cell's survival. The amino acid can be any of those that are uptaken by the cell, depending on which ones have upper bound larger than their flux value required for maximum biomass production. That way, after FBA allocates all the resources to produce the maximum possible biomass, there is flux remaining to be channelled towards antibody formation.

In the optimisation of the GEM, out of the 89 scenarios there were six different amino acids that were identified as being in abundance. Initially, three patterns were determined from the analysis of the 20 high producing scenarios – formation of an abundance in **leucine**, **proline**, or **phenylalanine** all coupled with an inhibition in biomass production. Each of the three groups is analysed below, with a specific scenario analysed in depth and a suggestion for an *in vitro* implementation of the scenario. However, the regulations that are based on leucine, proline or phenylalanine account for 70 out of the 89 total optimisation scenarios. The remaining 19 scenarios that were not picked as high producers because they result either lower μ or v_{IgG} in an FBA optimisation, are based on an abundance of different amino acids which was important to be identified for research purposes. Scenarios 5, 13, 36, 60, and 72 are inhibiting the growth rate whilst creating an abundance in **lysine** by not allowing its conversion to peptidyl-L-lysine and the nuclear transport of peptidyl-L-lysine via diffusion which again disables the conversion of lysine to peptidyl-L-lysine. Scenarios 7, 26, 29, 34, 65, and 85 knock-out the enzyme VALTA which catabolises **valine** to glutamate, which, by also imposing a growth inhibition results in antibody production of the CHO-K1 GEM constrained by D1, 3 and D4 datasets. Scenario 89 is the only one that creates an abundance in **asparagine** through the removal of the enzyme L-

asparaginase in cytosol and in mitochondria. Scenario 89 utilised the D2 constraints, which is the only dataset that allows the removal of L-asparaginase due to lower uptake of asparagine compared to the other datasets, as explained in paragraph **4.3.2**. Scenarios 18, 38, 46, 59, 61, 62 and 64 also create a proline inhibition, but with upregulation in the proline metabolism instead of a knockout except scenario 18 that optimisation suggests the removal of the FAD (Flavin Adenine dinucleotide) reductase which creates the required FAD for the enzyme PROD2 to convert proline to glutamate.

The three main groups of proline, leucine or of phenylalanine abundance together with different ways of growth inhibition, are analysed below, together with a proposal on how the genetic engineering can take place *in vitro*. For the analysis, the fluxes from an FBA optimisation with biomass production as the objective function are compared between the original CHO-K1 GEM and the modified one. The original CHO-K1 GEM is the same as in (99), constrained by the datasets D1, 2, 3, 4 which are calculated in section **4.2**. The names of the reactions mentioned below, are the same as in the GEM, which can be found in the BiGG database. A reaction with the symbol _f means forward and _b means backward. In the engineered CHO-K1 GEM, one or more regulations from each scenario analysed are applied on the bounds of the reactions, and the rest stay the same and defined by the dataset used. Even though the regulation is applied on the dataset used to constrain the model, and the metabolic network stays the same, the GEM is defined as engineered.

5.2.1 Proline Abundance with growth inhibition

Proline abundance takes place in six out the 20 high productivity scenarios of the optimisation results and in 35 of all 89 scenarios, making it the most common pattern amongst all strain designs suggested by the optimisation algorithms Igg_Igg and Bio_Igg. The high producing scenarios are presented in table 5.2.

Table 5.2: Predicted μ and q_{mAb} of the proline abundance scenarios. The biomass maximising FBA optimisation took place on the original CHO-K1 GEM, constrained by the dataset of the second column. Each of the eight high producing scenarios make use of proline abundance to couple biomass and antibody production. The scenarios utilised different set of regulations and dataset and result to a different combination of μ and q_{mAb}. The numbers in parenthesis in the 3rd and 4th column, display the change between the FBA predicted values of the original CHO-K1 GEM – control values – and the engineered one, with the red arrow representing decline and green arrow increase.

Scenario	*Dataset*	*μ, hr^{-1}*	*Q_{mAb}, pg/cell/day*	*Number of Regulations*
14	D4	0.0133 (38%) ↓	61.0383 (18%) ↑	6
30	D2	0.0099 (47%) ↓	63.6437 (33%) ↑	13
31	D2	0.0173 (8%) ↓	46.0704 (4%) ↓	2
32	D3	0.0205 (8%) ↓	34.0823 (8%) ↑	2
33	D3	0.0205 (8%) ↓	34.0823 (8%) ↑	2
47	D3	0.0144 (45%) ↓	30.115 (5%) ↓	8
58	D3	0.0134 (38%) ↓	60.7551 (17%) ↓	2
80	D4	0.1505 (30%) ↓	56.98 (10%) ↑	8

Scenarios 14, 30, 47 and 80 suggest a way to inhibit the growth rate by increasing the energy demand of the cell to lower the biomass production, and the removal of proline catabolism which results in the accumulation of proline in the cell. The rest of the scenarios in the above table suggest regulations that result on the same phenotypic changes but with only two regulations, which is the reason these scenarios are analysed in depth below.

Scenarios 31, 32, and 58 of the optimisation results suggest the same two regulations but use different datasets resulting different μ and v_{IgG} by the FBA optimisation. Specifically, the two regulations proposed take part in the glutamate metabolism: upregulation of the conversion of L-glutamate 5-semialdehyde (glu5sa) to 1-pyrroline-5-carboxylate (1py5c) with reaction identifier G5SADr_f by 90% and the knockout of the backward part of the reaction with reaction identifier G5SADr_b.

$$Upregulate\ \text{G5SADr}_\text{f}\ \text{by 90\%}: glu5sa[m] \rightarrow 1pyr5c[m] + h_2O[m] + h[m]$$
$$KO\ \text{G5SADr_b}: 1pyr5c[m] + h_2O[m] + h[m] \rightarrow glu5sa[m]$$

Scenario 33 regulates the transport of glutamate semi-aldehyde from and to mitochondria, which results in the same phenotype. Scenario 31 upregulates and knocks out G5SADr_f and G5SADr_b, which results on the same phenotype as in the scenarios 32, 33 and 58. It utilises the experimental bounds from the 2nd dataset to constrain the GEM. The D2 dataset uses the commercially

available FeedC making it is easier to implement *in vitro* than creating the custom feeds used in D3 and D4, which is the reason the scenario is analysed in depth below.

The genetic engineering strategy of the 31st scenario is focused on proline metabolism. In the metabolic network of the original CHO-K1 GEM, proline is converted to glutamate though proline dehydrogenase (PROD2) and L-glutamate 5-semialdehyde dehydratase (G2SADr) as seen in figure 5.4.

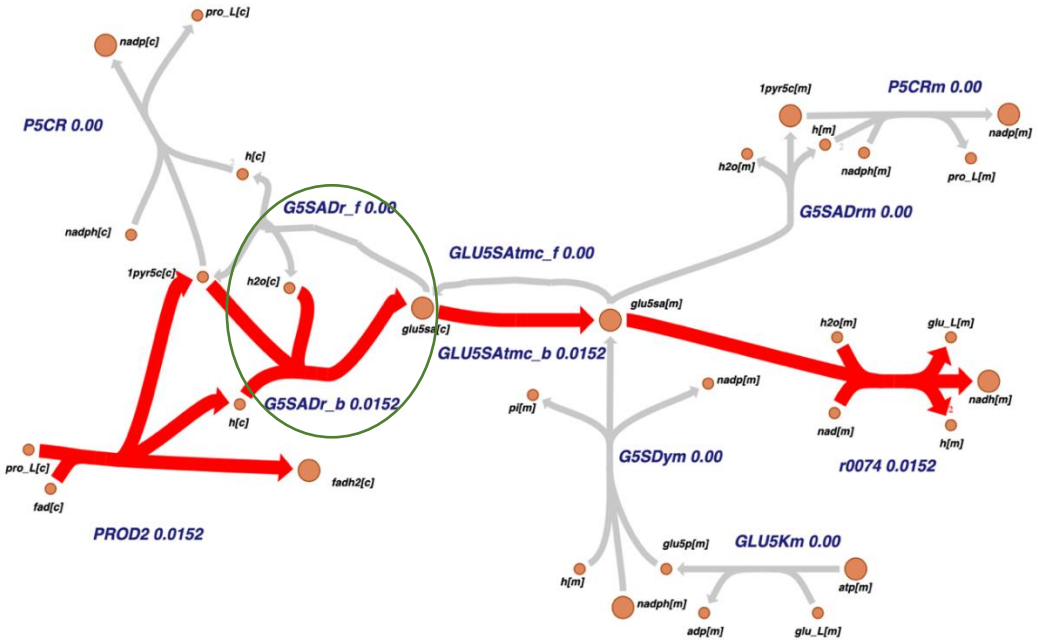


Figure 5.4: Metabolic network of the original CHO-K1 GEM, constrained by the D2 experimental dataset. Proline is directed through the glutamate semi-aldehyde enzyme, to transport from cytosol to mitochondria and convert to glutamate. The genetic strategy of scenario 31 is trying to alter this pathway.

A way to change the catabolism of proline to glutamate is by increasing the flux of G5SADrm_f by 90% of the lower bound in the D2 dataset, that the optimisation proposes, and by blocking the reverse reaction, G5SADrm_b, so proline cannot convert back to glutamate. Or, in the reversible GEM the reaction G5SADrm should always have a positive lower bound. These two changes force glutamate to convert to proline as seen in figure 5.5.

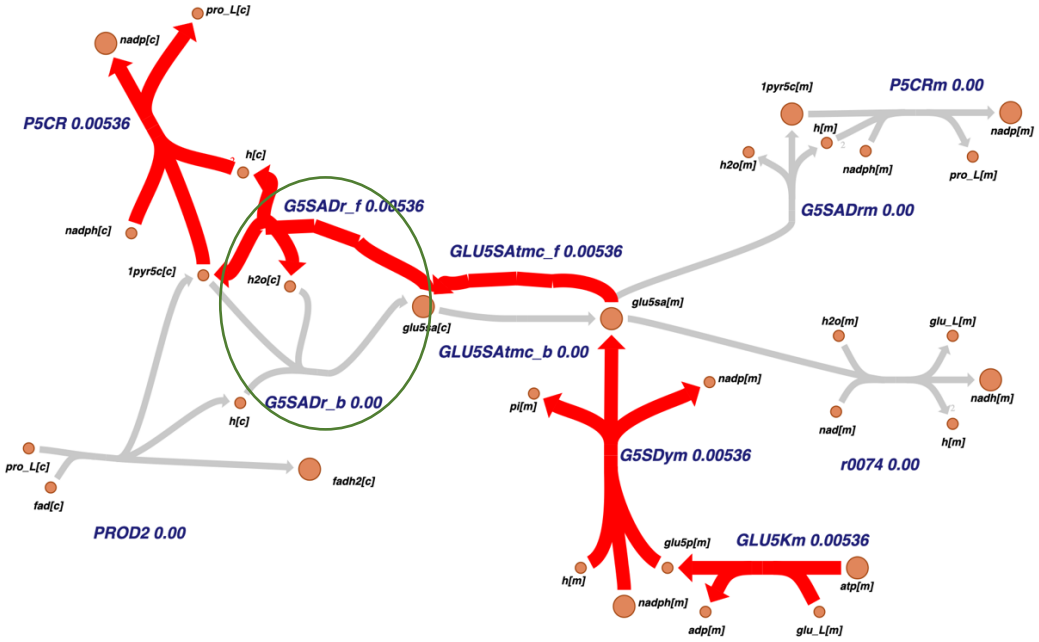


Figure 5.5: Metabolic network of the CHO-K1 GEM model constrained by the D2 dataset with an upregulation of G5SADr_f and knock out of G5SADr_b (green circle). In contrast to the original network, glutamate here is being converted to proline.

The produced 1py5c in cytosol cannot go back to mitochondria and convert to glutamate, because of the removal of the backward reaction, G5SADrm_b. The only catabolic pathways available are through the enzymes glutamate-5-semialdehyde dehydratase (G5SADr) and pyrroline-5-carboxylate reductase (P5CR), both resulting in the production of proline. Both regulations need to take place in order for glutamate to be converted to proline. If only the knock-out takes place, then proline cannot be converted to glutamate and the whole pathway will remain inactive as seen in Fig. 5.6A. If only the upregulation of glu5sa from cytosol to mitochondria takes place then glu5sa will be transferred back to cytosol as seen in Fig. 5.6B, because otherwise there is an abundance of proline that will result to a reduction in growth rate, which would negatively impact the objective function of the optimisation algorithm.

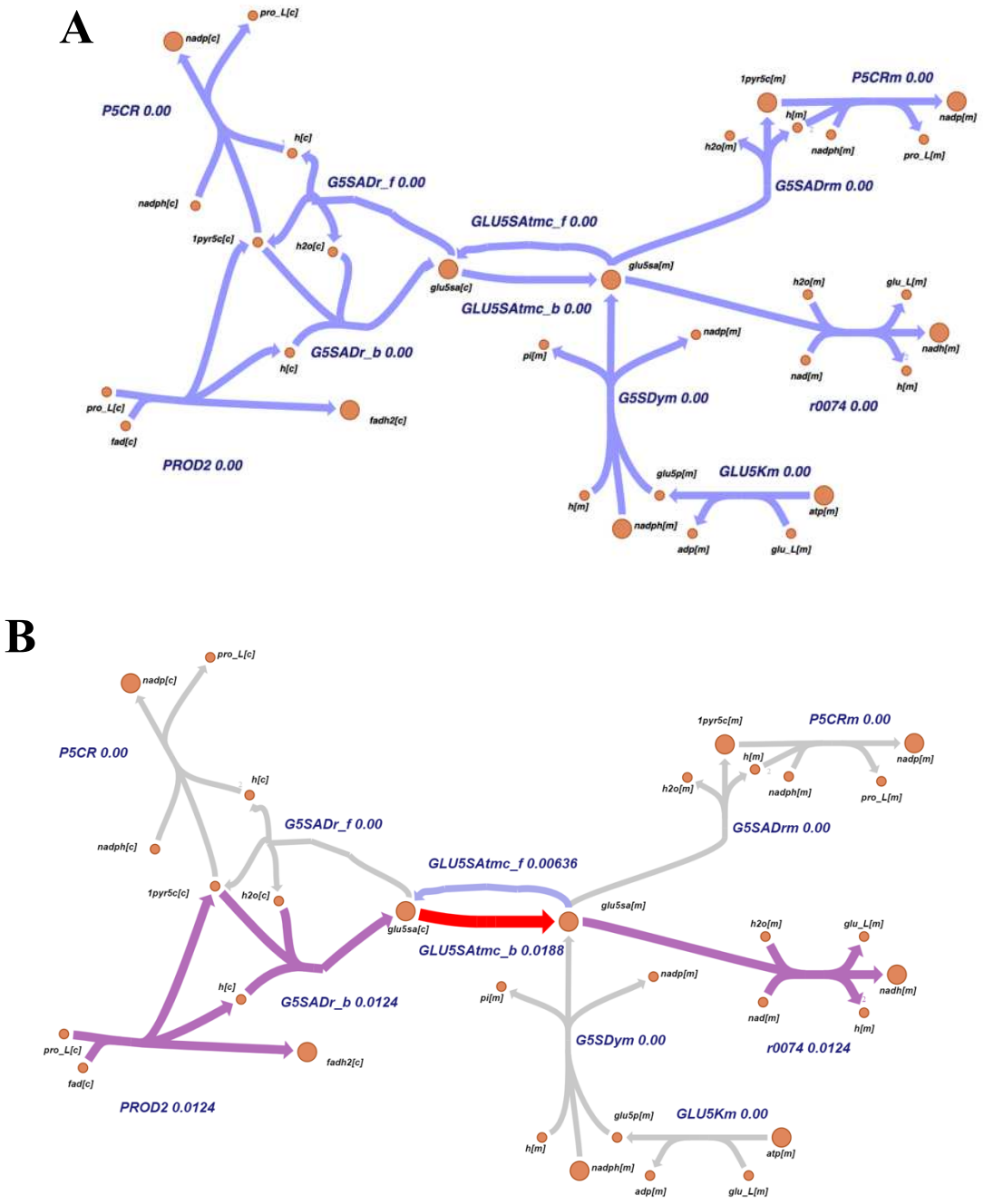


Figure 5.6: Knock-out of G5SADrm_b (A) in the CHO-K1 original GEM, constrained under the D2 dataset, has no effect on the metabolism. The upregulation of G5SADrm _f (B) result in glu5sa to be directed from cytosol to mitochondria. Both scenarios are unable to convert glutamate to proline, which is the target of the optimisation goal to achieve the coupling of antibody and biomass production in an FBA optimisation.

The conversion of glutamate to proline creates an abundance in proline, reducing the growth rate and producing IgG as a result. The main metabolic changes that the strategy is targeting is that, in datasets D2-4 proline cannot be secreted (upper bound is negative), so if there is more than the amount

needed for biomass production, and proline needs to find other pathways in order for it be catabolised. In the D2 dataset, the proline bounds are [-0.024, -0.003], in D3 the bounds are [-0.0208, -0.023] and in D4 they are [-0.019, -0.005]. In all cases there is a negative lower bound in proline not allowing it to be secreted if there is excess proline available in the cell. All three datasets have a maximum possible biomass production of 0.0297 hr^{-1} which corresponds to 0.008mmol of proline/gDCW/hr, based on the biomass stoichiometric coefficient of proline in the biomass producing equation of the CHO-K1 GEM. If a maximum biomass production bound is assumed based on elemental balances as calculated in section **3.2**, then the proline required is 0.007, 0.006 and 0.008 accordingly for D2, D3 and D4 datasets, which is still higher than the minimum proline uptake. Therefore, if there is more than 0.008mmol/gDCW/hr of proline in the cytosol, proline will need to be catabolised since it cannot be secreted. The only two pathways for catabolism of proline, in the CHO-K1 GEM, are either conversion to glutamate or antibody synthesis. The removal of the glutamate metabolic pathway can result in proline not being able to be catabolised in any pathway apart from antibody synthesis. For example, if there is more than 0.006mmol/gDCW/hr of proline being produced in the cell from the conversion of glutamate in D2, then there would be at minimum: 0.003 (minimum uptake of proline based on D2 bounds) + 0.006 (conversion of glutamate to proline by upregulation for the reaction) = 0.009 mmol/gDCW/hr, which is more than the amount of proline, 0.008, required for maximum biomass, so the remaining 0.001mmol/gDCW/hr of proline will be in abundance. In an FBA optimisation with biomass being the objective function, the 0.008mmol/gDCW/hr will be channelled to biomass production and the remaining 0.001mmol of will head towards antibody synthesis. IgG synthesis in the CHO-K1 GEM consumes a large number of amino acids and energy, so growth rate will reduce reaching an optimal trade-off point between the two objectives. The result of an abundance of proline and a biomass growth limitation in the FBA optimisation can be seen at Fig. 5.7.

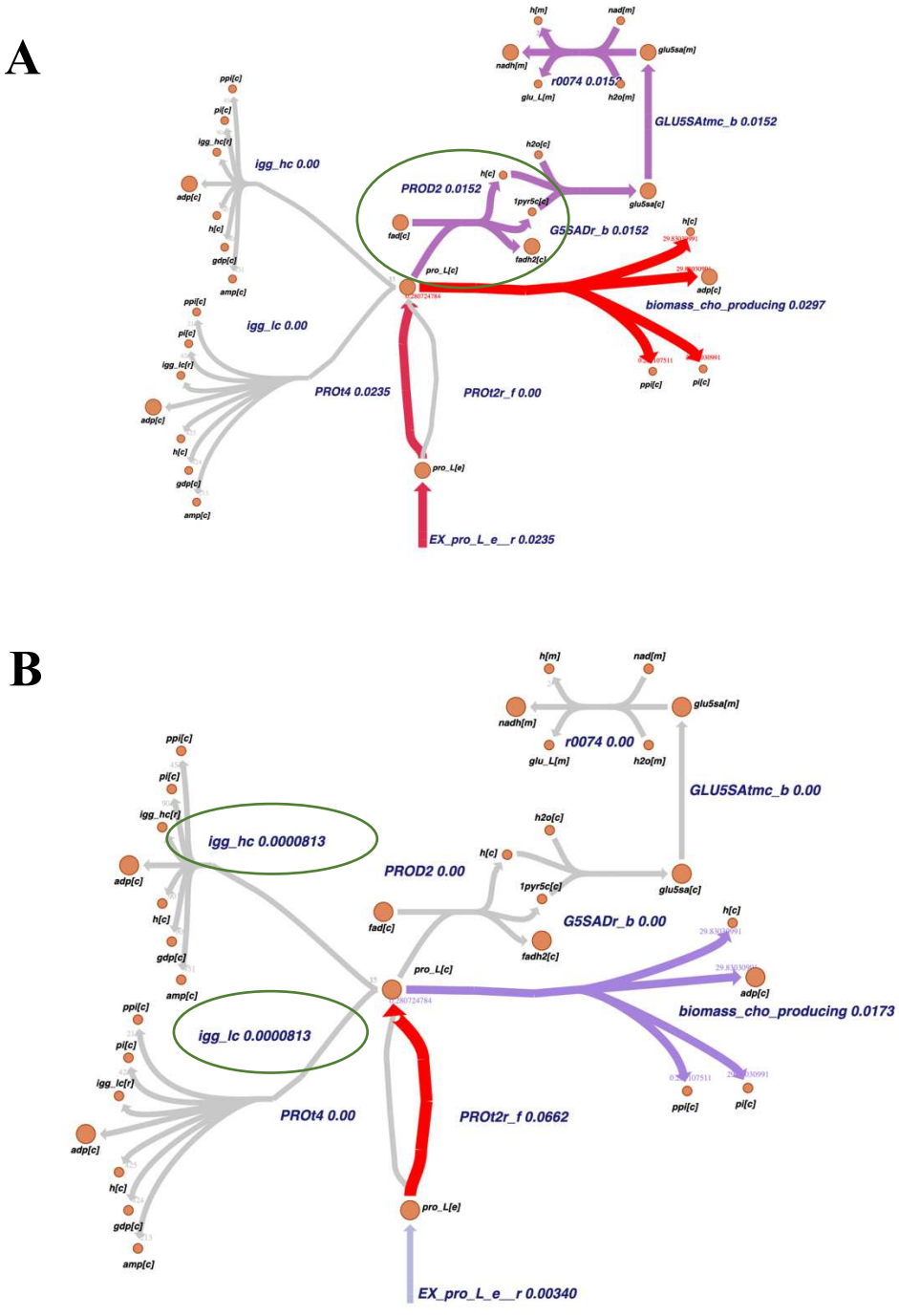


Figure 5.7: The effect that the removal of the proline catabolic pathway has on the metabolism of the CHO-K1 GEM. In the original GEM (A), proline can be converted to biomass or glutamate or take part in the antibody synthesis. The objective function is biomass production, so the majority of proline is directed to biomass and the remaining is converted to glutamate. The KO of the pathway converting proline to glutamate (B), results in proline to be channelled towards antibody synthesis in order for it to be consumed.

The strategy cannot take place in D1 because there is not enough glutamate to sustain the increased demands of cell metabolism, including biomass production and conversion to proline. However, the use of a glutamate feed could potentially fix the problem.

As appealing as the genetic strategy sounds, it requires enzyme engineering to modify G5SADrm to not convert proline to glutamate, which is difficult to perform *in vitro*. The enzyme G5SAD is responsible for the reversible transport of 1pyr5c from cytosol to the mitochondria and backwards. The scenarios 31, 32, 33 and 48 suggest that 1pyr5c can only move from mitochondria to cytosol. By observing the metabolic network, another set of reactions that can perform the same task and expressed by different genes can accomplish the same phenotype. Increasing the flux of the enzyme glutamate-5-semialdehyde dehydrogenase (G5SDym), which converts L-glutamate-5-phosphate (glu5p) to glu5sa in the mitochondria, which is expressed by *Aldh18a1* increases the conversion of glutamate to proline. Also, removing the enzymes 1-pyrroline-5-carboxylate dehydrogenase (P5CD) in mitochondria and glutamate-5-semialdehyde oxidoreductase (r0074), which are both expressed by *Aldh4a1*, does not allow the formation of glutamate from either glu5sa or 1-pyrroline-5-carboxylate (1PYR5C) and generally the formation of glutamate from proline. The resulting phenotype (Fig. 5.8) is the same as in scenario 31 that is analysed above with the main difference in the degree of upregulation of *Aldh18a1*.

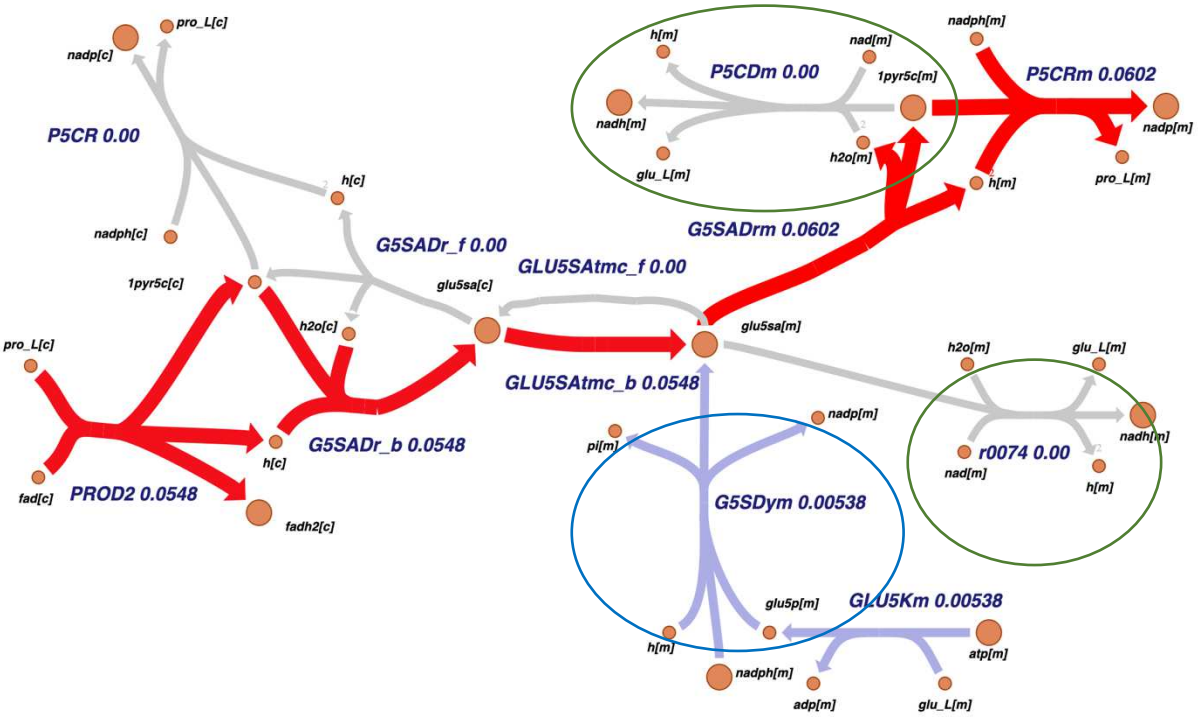


Figure 5.8: Knock-out of the *Aldh4a1* gene (green circle) disables the flux of the reactions P5CD and r0074, and upregulation of *Aldh18a1* gene (blue circle) increases the lower bound of the flux of G5SDym in the CHO-K1 GEM, constrained by the D2 dataset. The resulting phenotype is the same as in the regulations proposed by scenarios 31, 32, 33 and 58.

All three of the enzymes mentioned are expressed by known genes. G5SDym is associated to the gene *Aldh18a1* and P5CD and r0074 are associated to the enzyme ALDH4A1 that is expressed by *Aldh4a1*. Therefore, overexpression of *Aldh18a1* and knock out of *Aldh4a1* forces glutamate to be converted to proline. The amount of proline formation is based on the degree of upregulation, the more flux going through G5SDym the more proline will be formed. The reaction G5SDym has much wider bounds [0, 3.3146] than G5SADr [0, 0.0073] because of the number of carbons and nitrogen participating in the reaction. Thus, the degree of upregulation, which is based on the minimum proline uptake of the D2 dataset (0.003), is 0.2% instead of 90%. The new lower bound of G5SDym will be 3.3146*0.002 = 0.0066mmol/gDCW/hr, which results in total proline in cytosol of 0.0066+0.003 = 0.0086 > 0.008 required for maximum biomass. The abundance of proline results in antibody synthesis as a by-product of the FBA optimisation. Generally, for production of IgG to take place there should more proline in the cell than $v_{proline}^{required\ for\ biomass} + v_{proline}^{minimum\ uptake}$, where $v_{proline}^{required\ for\ biomass}$ is the amount of proline required for the maximum biomass formation and $v_{proline}^{minimum\ uptake}$ the minimum uptake of proline by the cell. If, however, proline is allowed to be secreted then similarly there should be proline in cytosol than $v_{proline}^{maximum\ secretion} - v_{proline}^{required\ for\ biomass}$. For example, if there is a 50% upregulation of the gene *Aldh18a1* then the reaction G5SDym will convert at minimum 3.3146 * 0.5 = 1.6573mmol/gDCW/gr of glutamate to proline in the cell. There is need for 0.008mmol/gDCW/hr of proline for the biomass production, leaving 1.6493mmol/gDCW/hr in the cell. The remaining flux needs to either be secreted or be channelled towards antibody synthesis. If there is a limit in proline secretion, then there would be additional antibody production. Moreover, if instead of a knockout of the *Aldh4a1* gene, there is a knockdown, then there would be proline being converted to glutamate. In order to establish proline abundance, an increased degree of overexpression of *Aldh18a1* should be established then. The effect of the upregulation degree is presented at figure 5.9.

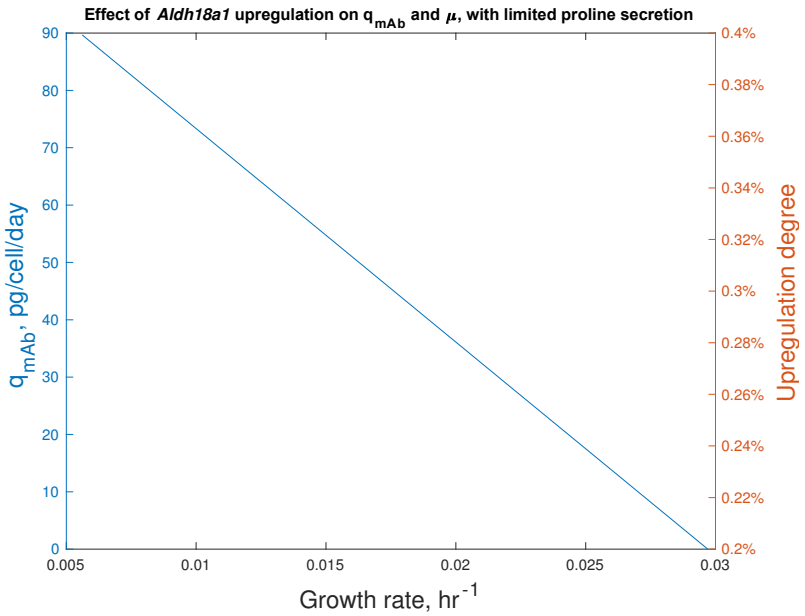


Figure 5.9: The result that the upregulation degree of *Aldh18a1* have on the FBA prediction of μ and $q_{mAb.}$ The optimisation was performed on the CHO-K1 GEM, which was constrained by the modified D2 dataset based on the scenario under study. It is clearly visible that the more *Aldh18a1* is overexpressed the less biomass is produced, and more antibody is secreted from the cell.

The overexpression degree is too small and precise to be achieved *in vitro*, however it can be increased by supplying more glutamate, and thus allowing further overexpression of *Aldh18a1*. Therefore, the best way to achieve this strategy is to try different proline concentrations and degrees of upregulation, as well as possible glutamate feeding due to the upregulation of glutamate catabolism. After the creation of the stable cell line with the two genetic modifications, a fed-batch culture with FeedC can test if there is an increase in q_{mAb}. In summary, proline abundance with growth inhibition can take place by the overexpression of *Aldh18a1* and knock out of *Aldh4a1*, which according to FBA optimisation, could increase q_{mAb} during the exponential growth phase.

5.2.2 Leucine abundance with growth inhibition

Leucine abundance takes place in three out of the 20 high productivity scenarios of the optimisation results and in 15 of the 89 total scenarios. The three high producing ones are presented in table 5.3.

Table 5.3: Predicted μ and q_{mAb} of the leucine abundance 49, 52 and 74 scenarios. The FBA optimisation that utilise leucine abundance and growth inhibition to force antibody production in the CHO-K1 GEM. Each scenario has a different set of regulations (5th column) and datasets (2nd column), which results in a different combination of μ and q_{mAb}. The numbers in parenthesis display the change between the FBA predicted values of the original CHO-K1 GEM – control values – and the engineered one, with the red arrow representing decline and green arrow increase.

Scenario	*Dataset*	*μ, hr^{-1}*	*Q_{mAb}, pg/cell/day*	*Number of Regulations*
49	D4	0.013 (40%) ↓	39.3633 (24%) ↓	7
52	D4	0.0124 (43%) ↓	43.6783 (16%) ↓	6
74	D2	0.0111 (41%) ↓	32.7011 (32%) ↓	10

Scenario 52 has the least number of regulations out of the three, which is the reason it is analysed further below. The scenario suggests two upregulations and four knockouts, using the constraints from the D4 dataset:

$$Upregulate \text{ RTOTAL2t_b } by \text{ } 50\%: Rtotal2[c] \rightarrow Rtotal2[e]$$
$$KO \text{ RTOTAL2t_f: } Rtotal2[e] \rightarrow Rtotal2[c]$$
$$Upregulate \text{ RTOTAL3t_b } by \text{ } 50\%: Rtotal3[c] \rightarrow Rtotal3[e]$$
$$KO \text{ RTOTAL3t_f: } Rtotal3[e] \rightarrow Rtotal3[c]$$
$$KO\ PSP_L: H_2O[c] + pser_{L[c]} \rightarrow pi[c] + ser_L[c]$$
$$KO\ LEUTA: akg[c] + leu_{L[c]} \rightarrow glu_{L[c]} + 4mop[c]$$

Specifically in the scenario 52 with the six changes in the metabolic network of the CHO-K1 GEM constrained by the D4 dataset, FBA predicts μ=0.0124hr^{-1}, 43% decrease, and q_{mAb}=43.6783pg/cell/day, 16% decrease from the control values calculated in section **3.3**. Similarly to the proline abundance scenario that was analysed above, this genetic engineering strategy is focused on creating an abundance of leucine in the cytosol by removing its catabolic pathways and imposing an inhibition in the maximum biomass production of the FBA optimisation. There are five regulations that reduce the maximum FBA predicted growth rate: increase of the transport flux of carboxylate from cytosol to the extracellular environment (RTOTAL2t_b and RTOTAL3t_b), removal of the reverse direction (RTOTAL2t_f, RTOTAL3t_f) and knock-out of phosphoserine phosphatase (PSP_L) combined with the knockout of leucine transaminase (LEUTA) that disables the catabolism of leucine and creates an abundance of leucine inside the cell. Scenarios 49 and 74 propose a different way to create an inhibition in biomass production, but all include the removal of the LEUTA reaction for the creation of leucine abundance in the cell. However, due to the smallest number of regulations needed, scenario 52 is analysed in depth below.

Concerning the biomass production, regulations RTOTAL2t and RTOTAL3t are responsible for the conversion of large amounts of ATP to AMP, consuming energy and reducing the maximum biomass production. With the upregulations of the backward directions of the reactions there is an energy consuming cycle created which uses large amounts of ATP as presented in figure 5.10.

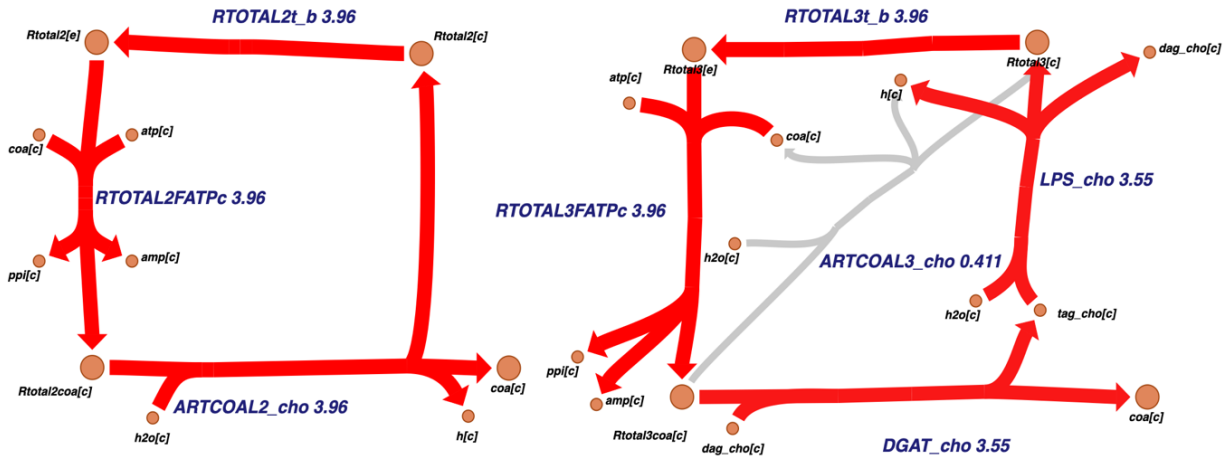


Figure 5.10: Upregulation of the forward direction in combination with knock out of the backward direction of RTOTAL2t and RTOTAL3t, in the CHO-K1 GEM constrained by the D4 dataset. An energy consuming cycle is formed, consuming large amounts of ATP, resulting in a reduced biomass production.

However, the above regulations are very difficult to perform *in vitro* because they both require enzyme engineering, and they are also regulating transport reactions with no known genes expressing them. The changes in the metabolism increase the consumption of energy, which even though can take place in the FBA optimisation they cannot take place *in vitro* and trying to perform a genetic strategy like that would be infeasible. However, in the metabolic network of CHO-K1 there is a reaction whose upregulation can achieve the same result. The upregulation of the enzyme Acetyl-CoA synthetase (ACS) also inhibits growth by consuming ATP, similarly to the upregulation to RTOTAL2t and RTOTAL3t. ACS produces acetyl CoA from acetate and coenzyme A and the complete reaction is:

$$ACS: coa[c] + atp[c] + ac[c] \rightarrow accoa[c] + amp[c] + ppi[c]$$

When ACS is upregulated, there is a significant increase in the conversion of ATP to AMP which negatively affects biomass production since AMP does not carry as much energy as ADP. In the original GEM there is no flux at the ACS reaction (Fig. 4.11A). The upregulation results to increased production of acetate to supply all the required precursors for the increased flux through ACS. Acetate is formed from the conversion of acetyl CoA, conversion of cytosolic N-Acetyl-D-glucosamine 6-

phosphate (acgam6p) and 1-alkyl 2-acteylglycerol 3-phosphocholine (paf_cho) or it is transported from mitochondria where is it formed by the hydrolysis of acetyl CoA. In the figure 4.11B an energy generating cycle there can be seen, which creates all the acetate required for the flux of ACS. If there was no upregulation of ACS, but somehow the cytosolic acetate would still be formed, it would get channelled towards ACS because it is the sole pathway available. However, mitochondrial acetate can form citrate and then move to ATP citrate synthase (ACITL) to convert ATP to ADP and citrate to oxaloacetate, which is more beneficial for biomass production as it can be an intermediate in more reactions than acetyl CoA.

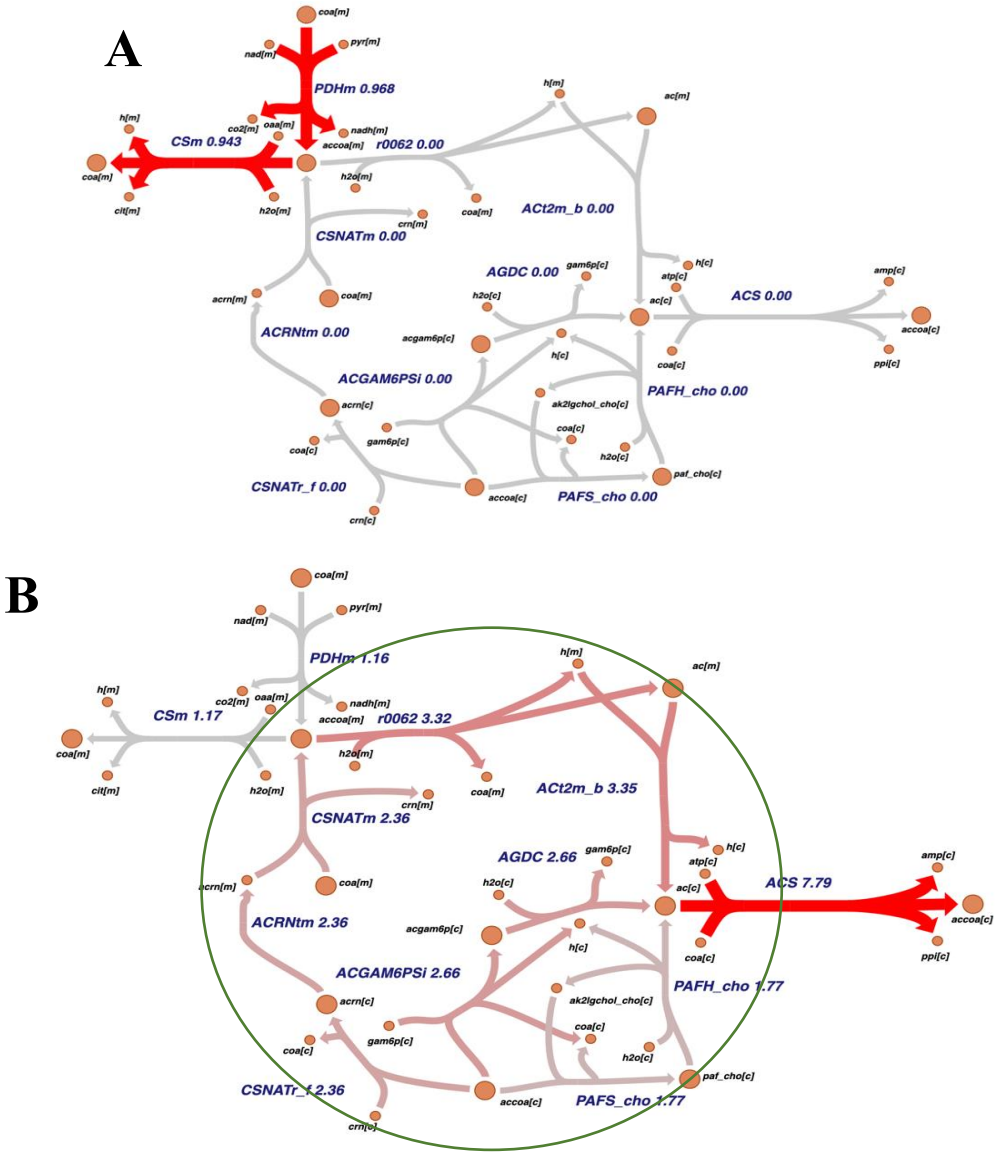
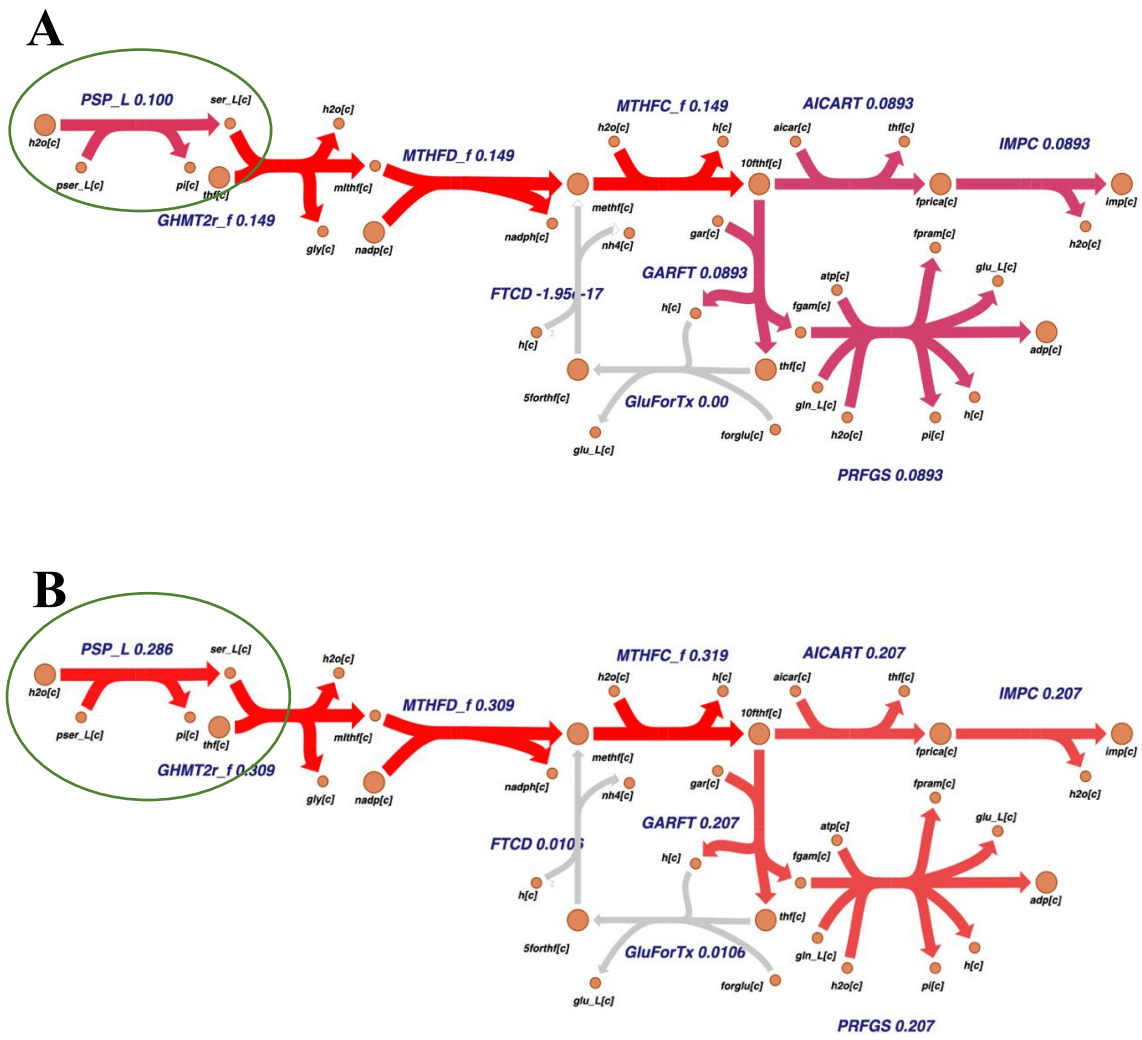


Figure 5.11: Biomass maximisation FBA of the original CHO-K1 GEM constrained under D4 dataset. In the original model (A) there is no flux in the ACS enzyme. However, with its upregulation (B) there is a cycle formed (green circle) that consumes ATP, reducing the maximum predicted growth rate from the FBA optimisation.

The proposed method here to reduce growth is the upregulation of the ACS enzyme together with KO of the enzyme phosphoserine phosphatase (PSP_L). PSP_L is responsible for converting O-phospo-L-serine (pser_L) to serine. Serine is an important for growth amino acid because it is a precursor for phosphatidylserine which is required for biomass production. In the D4 dataset, serine bounds are [-0.0621, -0.0145] and the cell can uptake enough serine for the maximum produced biomass as well as for the formation of other metabolites required for growth. If, however, there is an upregulation in ACS, the reaction PSP_L creates more serine for phosphor-L-serine as seen in figure 5.12A, B. The additional catabolism of serine compensates for the ATP usage of the ACS upregulation. The removal of the reactions PSP_L disables the catabolism of serine (Fig. 5.12C), inhibiting the growth rate.



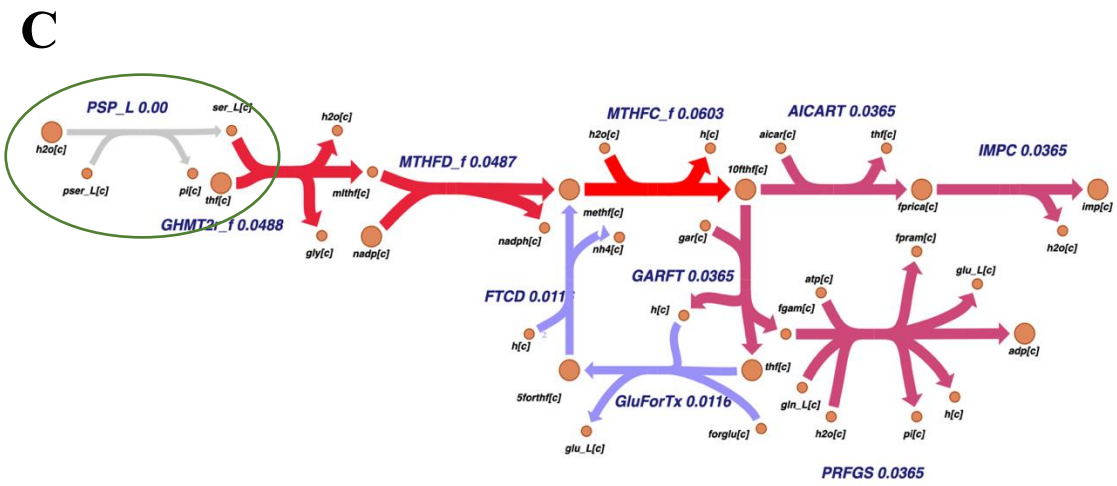


Figure 5.12: Catabolism of serine in the CHO-K1 GEM model constrained by the D4 dataset. In the original model (A), serine is catabolised to glutamate and inosinic acid (IMP). The upregulation of ACS (B) increases the flux of serine catabolism, to counter effect the increased demand of ATP. The KO of PSP (C) reduced the flux towards serine catabolism.

Leucine is catabolised by the LEUTA reaction. It produces both glutamate and 4-methyl-2-oxopentanoate (4mop) from leucine. The latter is used in the mevalonate pathway and for biomass production. Glutamate is an important amino acid which is widely used in the metabolism of mammalian cells. However, the removal of the enzyme does not have an impact in the metabolism due to the decrease of glutamate flux or 4-methyl-2-oxopentanoate, but because of the limitation in catabolising leucine. It results in leucine's only pathways being the antibody synthesis and the production of biomass. The bounds of leucine in the D4 dataset are [-0.0273, -0.0101], and if the biomass growth rate is less than $0.0187hr^{-1}$ (which requires 0.010mmol/gDCW/hr leucine) there would be excess leucine which would be channelled towards antibody synthesis, since no secretion is allowed in the CHO-K1 GEM, constrained by the D4 dataset. Similarly in D2 dataset, which uses the commercially available FeedC, leucine ranges from -0.0466 to -0.0086 and a reduction of growth more than $0.0144hr^{-1}$, leaves excess amounts of leucine inside the cytosol. Therefore, with the removal of LEUTA and a reduction of growth rate the remaining leucine would go to synthesise antibody as seen in figure 5.13A, B.

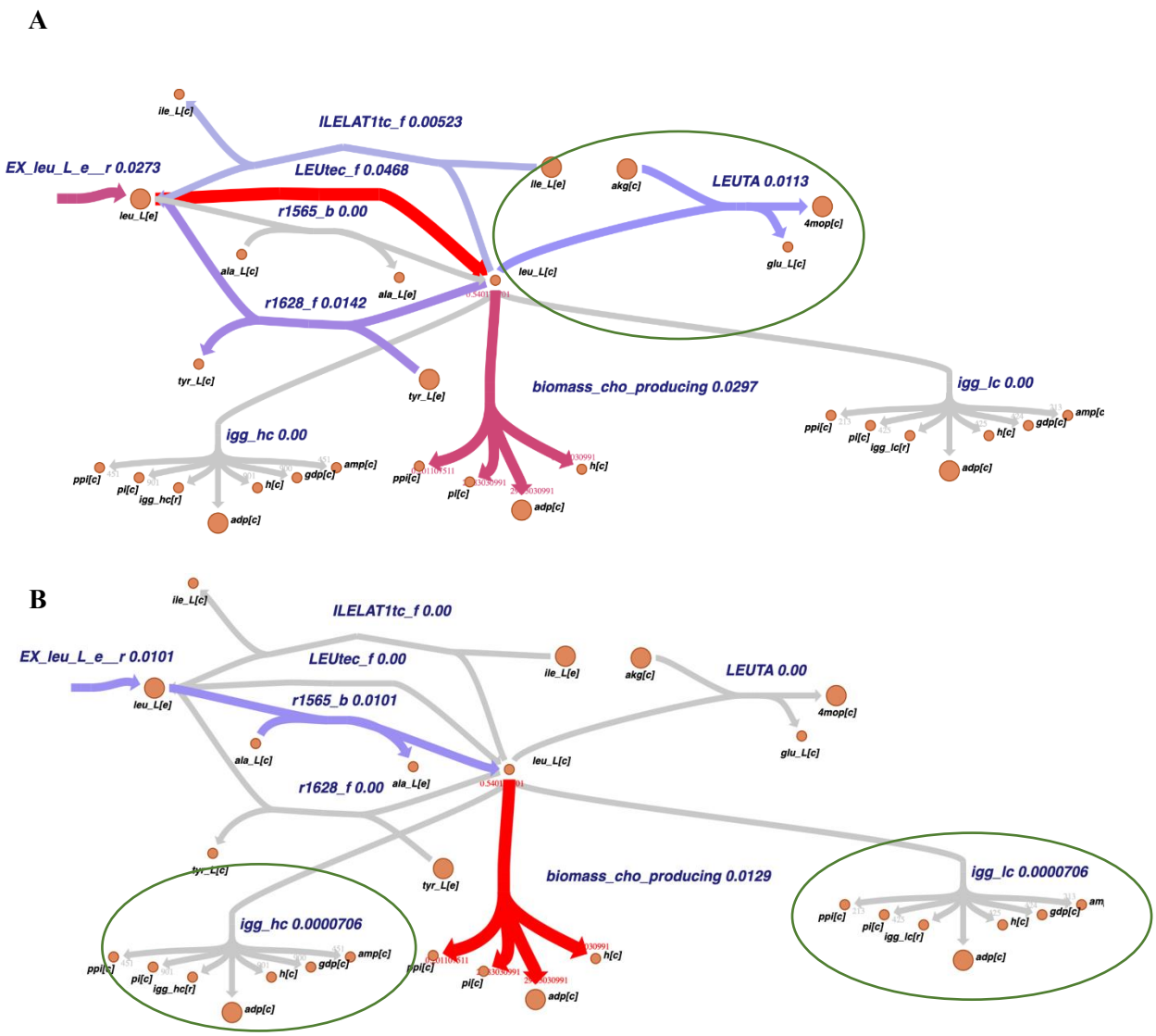


Figure 5.13: Catabolism of leucine in the CHO-K1 GEM constrained under D4 dataset. In the original model (A), leucine enters the cytosol from the extracellular environment, and then it is directed to biomass production, as that is the objective of the FBA optimisation. Any remaining flux of leucine is converted to glutamate thought LEUTA reaction. The removal of LEUTA (B) cause leucine to be in abundance, as it cannot be consumed by any pathway except the antibody synthesis one.

Similarly to the 1st scenario analysed, all three reactions are catalysed by enzymes which are expressed by different genes. The same phenotype can be achieved through genetic modifications of four different genes. Initially, the enzyme ACS is co-expressed by both *Acss2* and *Aacs*, so the upregulation of ACS requires the over-expression of both *Acss2* and *Aacs*. The LEUTA reaction is catalysed by the BCAT1 enzyme, which is expressed by the gene *Bcat1*, and its knock-out would result in the removal of the enzyme from the metabolism of CHO-K1. *Bcat1* is not essential for growth in CHO cells as it was not in the recently published experimentally tested database of essential for growth genes (*183*). The PSP_L reaction is catalysed by the PSPH enzyme which is expressed by *Psph*. It is

not essential for growth in CHO and can be removed from the genome as well. Performing four genetic regulations can be challenging, so instead of overexpressing ACS and knocking out PSP_L the same phenotype can be reached by culturing the cells in conditions of mild hypothermia and simply removing *Bcat1* gene.

Performing a genetic regulation to reduce growth might prove to be difficult, so it was thought to instead culture the cells under mild-hypothermia conditions. A genetic regulation which reduces growth could greatly hinder the cell growth. After transfection, the population of cells is a very small fraction of normal density. The wait time for VCD to reach normal levels could potentially be fatal for the culture. However, the mild-hypothermia culture of cell can solve the problem. The optimal culture temperature for CHO cells is around 36.5^{o}C (*185*). However, there have been studies, where CHO cells were cultured in lower temperatures in order to reduced biomass production and growth rate in order to achieve higher mAb specific productivity (*175, 186, 187*) . A decrease in the biomass production would channel the remaining energy and resources towards the recombinant protein. The reduced biomass production in CHO cells, can be considered as an inhibition in growth, and be used in the genetic strategy instead of performing three genetic modifications. The quality of the product in mild-hypothermia conditions was studied by Sou et al., who found reduced metabolic rates and fewer resources heading towards the glycosylation pathway but a general increase in antibody production (*187*). Overall, a mild-hypothermia approach will be easier to perform than genetic engineering, and according to published literature will inhibit biomass production.

5.2.3 Phenylalanine abundance with growth inhibition

Phenylalanine abundance takes place in most of the high producing scenarios of the optimisation solutions, specifically in 26 out of the 89 total optimal solutions. The high producing scenarios together with their FBA predicted growth rate and q_{mAb} are presented in table 5.4.

Table 5.4: FBA predicted μ and q_{mAb} for the high-producing scenarios, which are based on creating phenylalanine abundance in the cell by altering tyrosine catabolism and resulting in antibody production in the original CHO-K1 GEM. The objective function of the FBA optimisation performed was biomass production and constrained under the datasets of the second column. Each scenario has a different set of regulations yielding a different combination of μ and q_{mAb}. The numbers in parenthesis display the change between the FBA predicted values of the original CHO-K1 GEM – control values – and the engineered one, with the red arrow representing decline and green arrow increase.

Scenario	*Dataset*	*μ, hr^{-1}*	*Q_{mAb}, pg/cell/day*	*Number of Regulations*
1	D1	0. 0066 (25%) ↓	21.2311 (31%) ↑	8
3	D1	0. 0057 (35%) ↓	24.7229 (53%) ↑	4
9	D1	0. 0046 (48%) ↓	28.9245 (79%) ↑	11
12	D1	0.0101 (14%) ↑	32.7011 (48%) ↓	5
27	D1	0.0055 (38%) ↓	25.4711 (57%) ↑	9
81	D1	0.0062 (30%) ↓	22.9280 (42%) ↑	9
84	D1	0.0066 (25%) ↓	21.4020 (32%) ↑	8
86	D1	0.0066 (38%) ↓	25.6321 (58%) ↑	5
88	D1	0.0085 (4%) ↓	14.3019 (12%) ↓	20

There are nine different genetic strategies, suggested by the strain optimisation, which are focused on creating an abundance in phenylalanine by not allowing its conversion to tyrosine. The third scenario is analysed further, because it has the smallest total number of proposed regulations. The rest of the scenarios in table 5.4 are also based on a similar strategy, inhibition of growth rate and removal of phenylalanine's catabolic pathways. The four regulations proposed for the 3rd scenario are one upregulation and three knock-outs:

$$Upregulate \text{ PPAm } by\ 60\%:\ h_2O[m] + ppi[m] \rightarrow h[m] + 2\,pi[m]$$
$$KO \text{ GLXO1}: nad[c] + h_2O[c] + glx[c] \rightarrow 2h[c] + nadh[c] + oxa[c]$$
$$KO \text{ GUAt}: gua[c] \leftrightarrow \text{gua[e]}$$
$$KO \text{ 34HPPte}: 34hpp[c] \rightarrow 34hpp[e]$$

By constraining the CHO-K1 GEM using the D1 dataset and performing an FBA optimisation with growth rate being the objective function, the resulting μ and q_{mAb} are $0.006\,hr^{-1}$ with a 35% decrease and $24.723 pg/cell/day$ with 53% increase from the μ and q_{mAb} control values from the FBA optimisation on the original CHO-K1 GEM, constrained by the D1 dataset. The way this genetic strategy works is by limiting the growth rate and disabling the catabolic pathways of phenylalanine;

forcing it to be channelled towards the antibody synthesis pathway. The three regulations of PPAm, GLXO1 and GUAt inhibit the growth rate and the removal of 3-(4-hydroxyphenyl) pyruvate (34hpp) transport between cytosol and extracellular environment, directs phenylalanine to antibody synthesis. The rest of the scenarios in table 4.4 also contain the removal of the 3-(4-hydroxyphenyl) pyruvate secretion resulting in an abundance of phenylalanine but include a different combination of regulations to inhibit the growth rate. All the scenarios from table 4.4 use the D1 dataset, but the proposed regulations can be also applied on the other three datasets, as in all datasets there is more phenylalanine entering than the amount required for the maximum possible biomass produced, in each dataset.

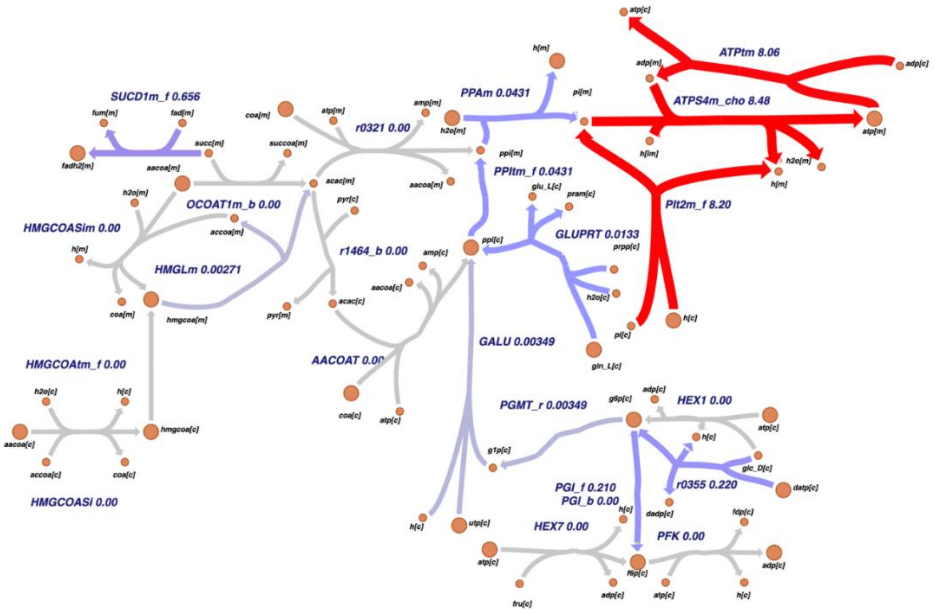
To inhibit the growth rate, scenario three implements a genetic strategy that increases the energy consumption of the cell and therefore reduces the maximum possible growth rate that can be achieved in an FBA optimisation constrained under D1 dataset. The main regulation that inhibits the biomass production of the optimisation is the upregulation of the enzyme inorganic diphosphatase in mitochondria (PPAm). The enzyme is responsible for converting diphosphate (ppi) to phosphate (pi) and is part of the oxidative phosphorylation pathway.

$$PPAm\text{: } h_2O[m] + ppi[m] \rightarrow h[m] + 2\ pi[m]$$

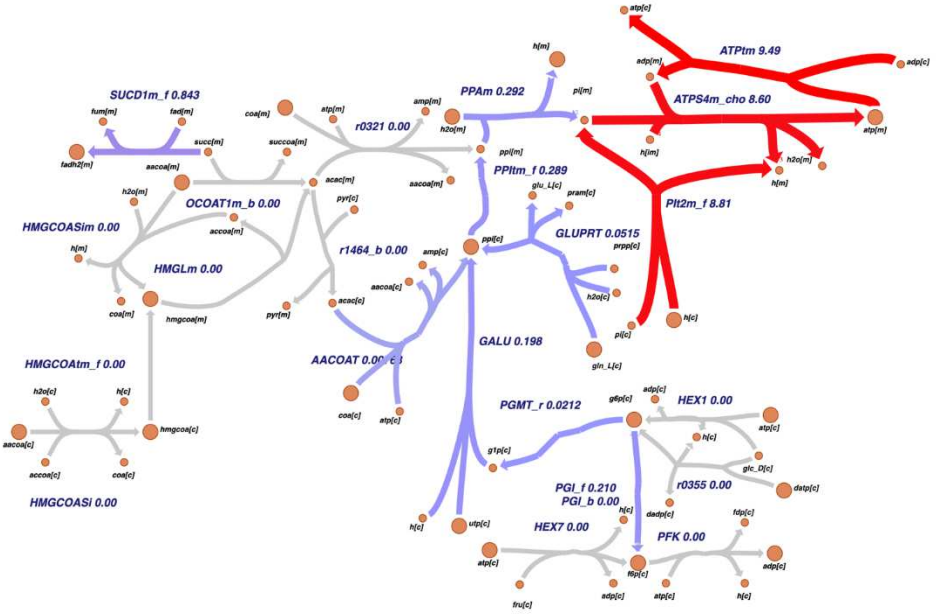
Increasing the flux of PPAm leads to an increase in the demand of diphosphate, which negatively affects biomass production. Mitochondrial phosphate is required for the essential reaction ATP synthase, which converts ADP to ATP, and a minimum flux is required for the survival of the cell. An increased flux, however, misuses the resources of the cell and negatively affects growth. In the original metabolic network constrained by the D1 dataset, PPAm carries a much smaller flux compared to when it is overexpressed (Fig. 5.14A-C), because the majority of phosphate is transferred from cytosol to mitochondria though the enzyme mitochondrial phosphate transport (Pit2m) and takes part in ATP synthase (Fig. 5.14A), without it needed to be converted from diphosphate in mitochondria. An increase in the flux of PPAm increases the demand of mitochondrial diphosphate, which is transferred from cytosol (Fig. 5.14C). The unnecessary conversion of phosphate to diphosphate in cytosol, the transfer of diphosphate to mitochondria and then its conversion back to phosphate in mitochondria is not as energetically favourable as the transfer of phosphate from cytosol to mitochondria; and it imposes a burden on the energy production of the cell. The knockout of the reactions, GLXO1, 34HPPt and GUAt does not have a major impact on the fluxed of PPAm or of ATP synthase (Fig. 4.16A-B) but their flux is affected by the upregulation of PPAm. Diphosphate can be created from a variety of sources; the optimisation decides as the most optimal pathways (smallest number of reactions with maximum energy production) the catabolism of glutamine through the enzyme glutamine phosphoribosyldiphosphate amidotransferase (GLUPRT) and conversion of acetoacetate (acac) to acetoacetate CoA ligase (accoa) through the reaction Acetoacetyl-CoA transferase (AACOAT) in cytosol and in mitochondria (r0321). In the D1 dataset, the supply of glutamine is limited and barely enough for the production of the

maximum biomass predicted from the FBA optimisation. The increase of its consumption takes away glutamine which would have otherwise gone to biomass production, reducing the FBA predicted growth rate. Therefore, increasing the supply of glutamine can counteract the effect of PPAm upregulation, as there would be glutamine for both biomass and its conversion to glutamate, which generates diphosphate. The conversion of acetoacetate to acetoacetate-CoA takes place because it created large amounts of diphosphate, however, it uses up important TCA intermediate metabolites such as fumarate and succinate.

A



B



C

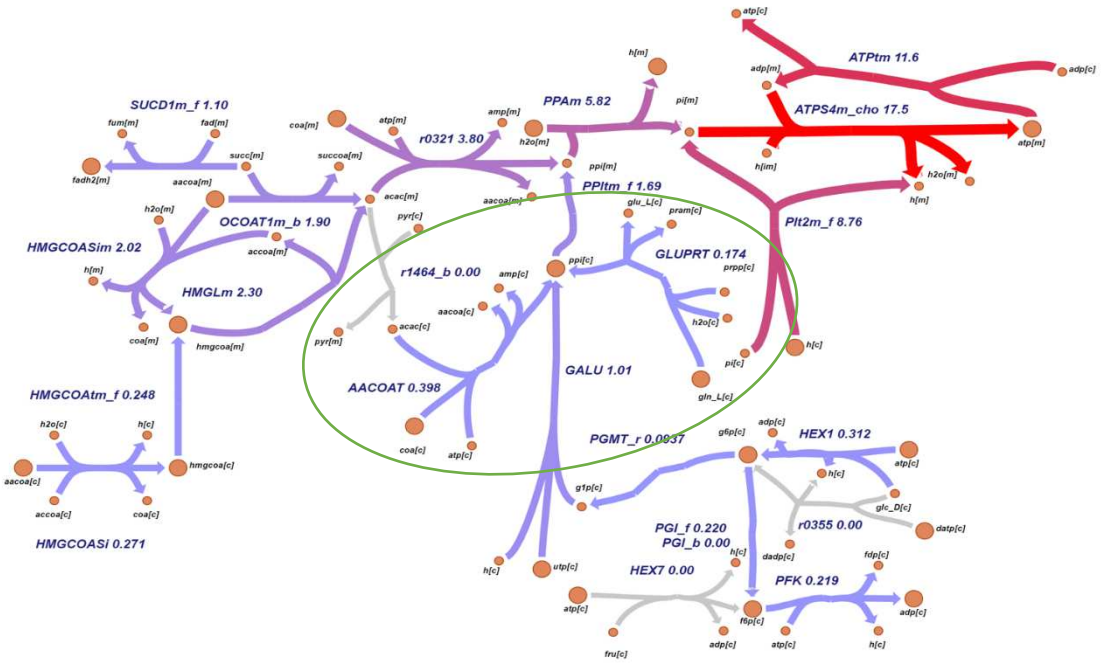
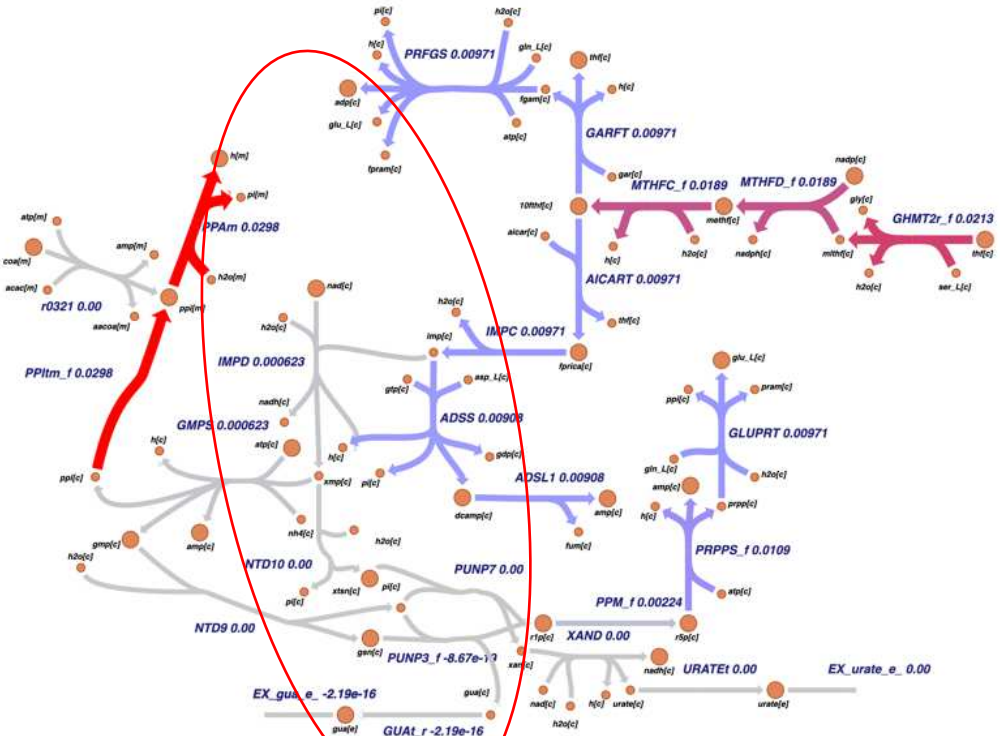


Figure 5.14: Oxidative phosphorylation of the CHO-K1 GEM, constrained by the D1 dataset. The original CHO-K1 GEM network (A), has very similar fluxes compared to (B), which is the network with the removal of GLXO1, GUAt, and 34HPPt. FBA does not increase the flux of neither GLUPRT nor the conversion of acetoacetate to acetoacetate CoA, because both of these enzymes reduce the growth rate and there is no need for extra diphosphate to be produced. However, in (C), which is the network with all the regulations active, including the upregulation of PPAm, both enzymes GLUPRT and AACOAT (green circle) carry flux in order to meet the demand of diphosphate.

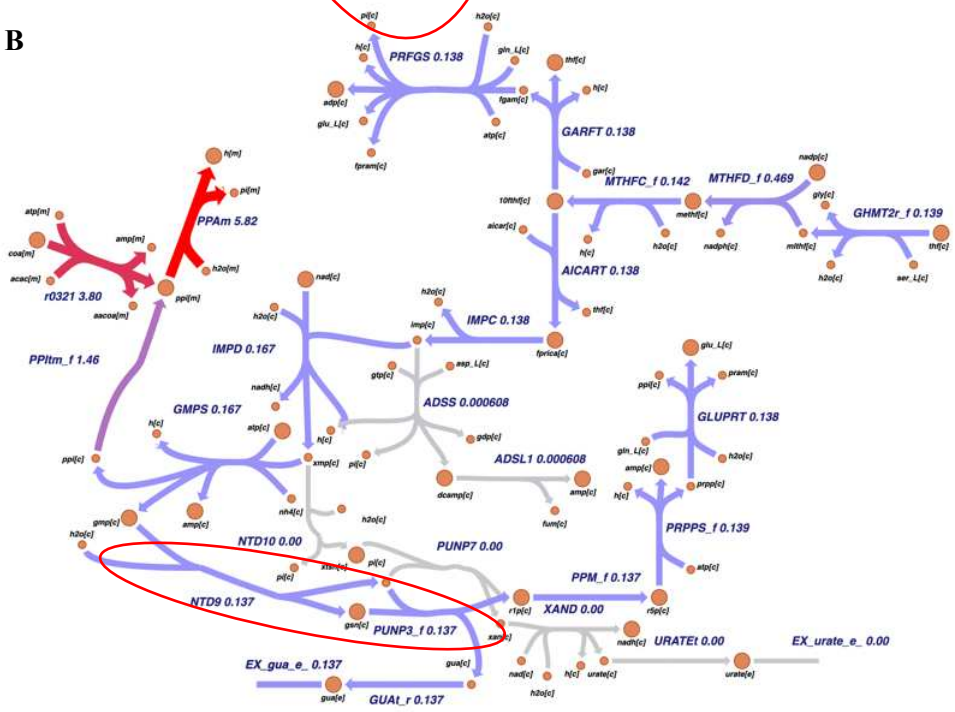
The removal of guanine secretion from the CHO-K1 GEM is energetically unfavourable if PPAm is upregulated. In the original CHO-K1 GEM constrained by the D1 dataset (Fig 5.15A), serine is catabolised to glutamate through the reactions PRFGS and GLUPRT. It synthesises inosinic acid (IMP) which is then converted through the enzyme adenylosuccinate synthase (ADSS) to fumarate and glutamate. There is no need for guanine production or in generally any increased flux through IMP dehydrogenase (IMPD) or gmp synthase (GMPS) as PPAm does not require large amounts of diphosphate. However, with the upregulation of PPAm there is an increase in the requirements of diphosphate, and the pathway metabolising guanine becomes active (Fig 5.15B) to aid the production of diphosphate thought the enzyme GPMS. Removal of the pathway catabolising guanine, results in diphosphate to be synthesised by the catabolism of xanthosine 5-phospate (xpm), a less profitable pathway because of the unnecesery conversion of phosphate to diphosphate (Fig 5.15C). The diphosphate PPAm requires, is transferred from cytosol to mitochondria as seen in the figure (Fig. 5.15A-C) and comes from various sources which are not displayed for clarity. By comparing the fluxes of Fig. 5.15B and Fig. 5.15C, it can be concluded that with the formation of guanine there is an increased

quantity of diphosphate produced in cytosol, which is then transferred to mitochondria for the requirement of PPAm, a scenario that is favourable for biomass production. However, without the metabolism of guanine, diphosphate is created from other energy unfavourable pathways, negatively affecting the growth rate.

A



B



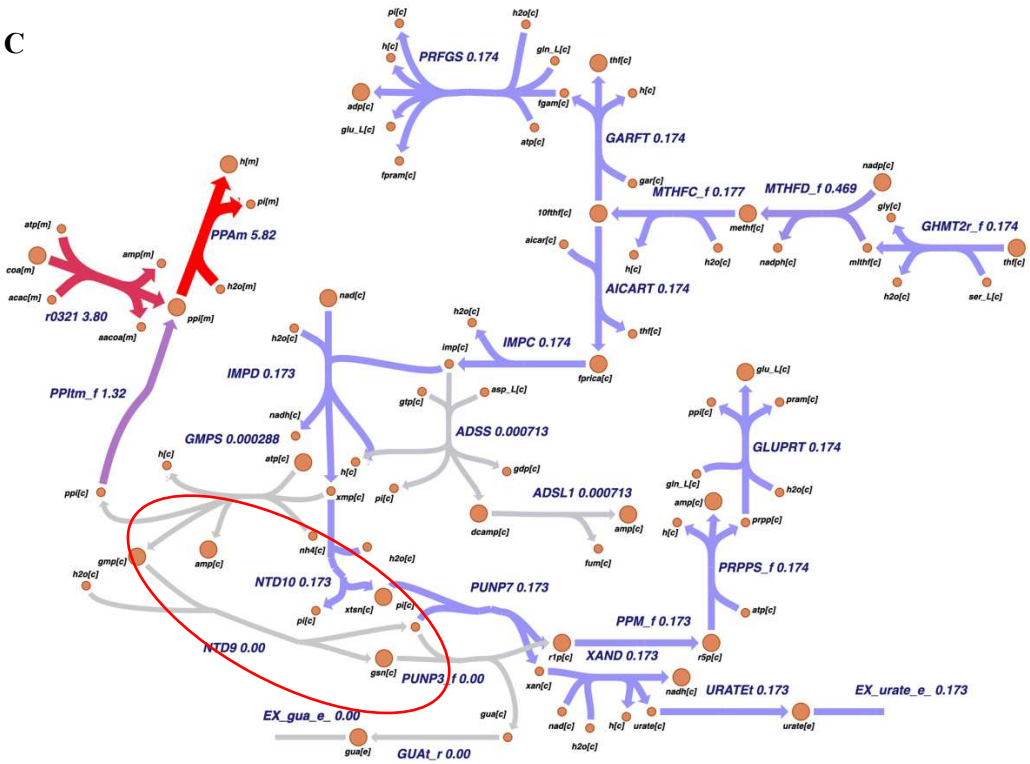


Figure 5.15: Metabolic network around the metabolism of guanine in the CHO-K1 GEM constrained under D1 dataset. After the upregulation of PPAm in (B), there is increased diphosphate demand compared to the original model (A). Synthesis of guanine creates the required diphosphate in cytosol, which can then be transferred to mitochondria for PPAm requirements. However, with the KO of guanine secretion in (C), that option is no longer valid in the FBA optimisation, so the diphosphate is formed in other less energetically favourable pathways.

Catabolism of hydroxy-L-proline does not take place in the original CHO-K1 GEM (Fig 5.16A). However, with the increase demand of diphosphate combined with the removal of guanine secretion the pathway that converts hydroxy-L-proline to oxalane becomes active (Fig 5.16B). During the catabolic pathway, important metabolites such as pyruvate and aspartate are formed, which can then be further catabolised, together with energy carrier molecules of NADH. The KO of GLXO1 disables the conversion of the metabolites and reduces the maximum biomass production from the FBA optimisation.

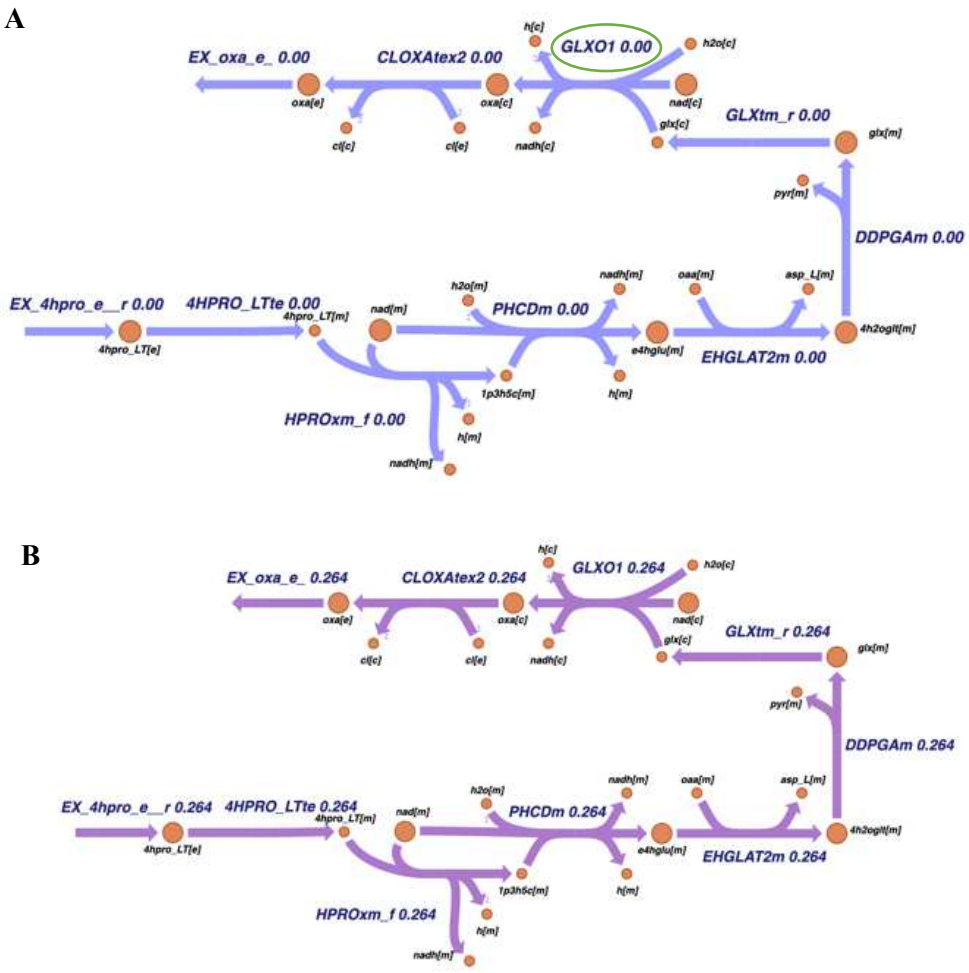


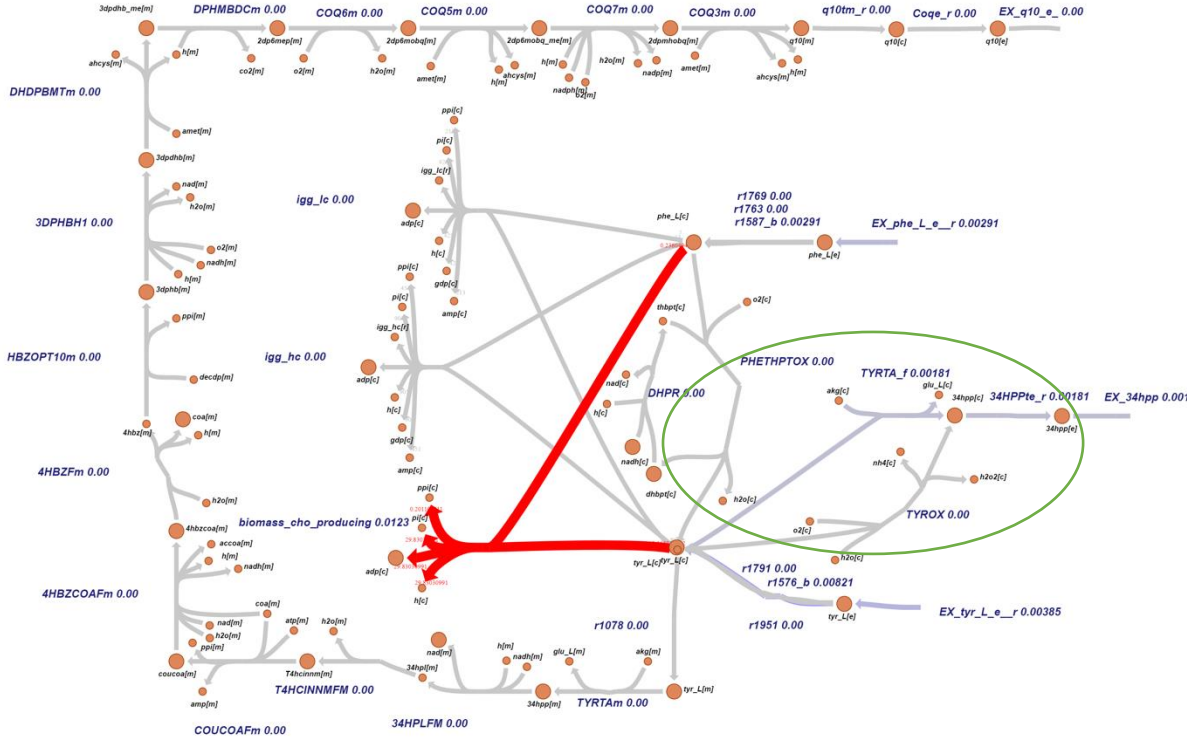
Figure 5.16: Metabolism of oxalate in the CHO-K1 GEM, constrained by the D1 dataset. In the original network (A) the pathway of oxalane catabolism is not active (green circle), as there is no need for the catabolism of hydroxy-L-proline. However, the upregulation of PPAm and the KO of guanine secretion, increased the demand of diphosphate and the oxalane catabolism pathway is an efficient pathway to meet the increase in diphosphate demand.

The next step after reducing the maximum growth rate predicted by the FBA optimisation on the D1 constrained CHO-K1 GEM, is it to create an abundance in phenylalanine inside the cell. The metabolism of phenylalanine in the CHO-K1 GEM is connected to tyrosine's, so both metabolic pathways are analysed below. In the D1 dataset the bounds of tyrosine are [-0.0059, -0.0039] and for phenylalanine are [-0.0074, -0.0029]. The maximum predicted biomass for the D1 dataset is 0.0123hr^{-1} which requires 0.002mmol of tyrosine and 0.0029mmol of phenylalanine. If the biomass produced is less than 0.0123hr^{-1} then there would be phenylalanine remaining in the cell that must be catabolised. In the metabolic network of biomass maximisation CHO-K1 GEM (Fig. 5.17A), remaining tyrosine is exiting the cell through its conversion to 3-(4-hydroxyphenyl) pyruvate. Tyrosine is not to channelled mitochondria through vesicular transport (r1078) because excess tyrosine negatively affects biomass

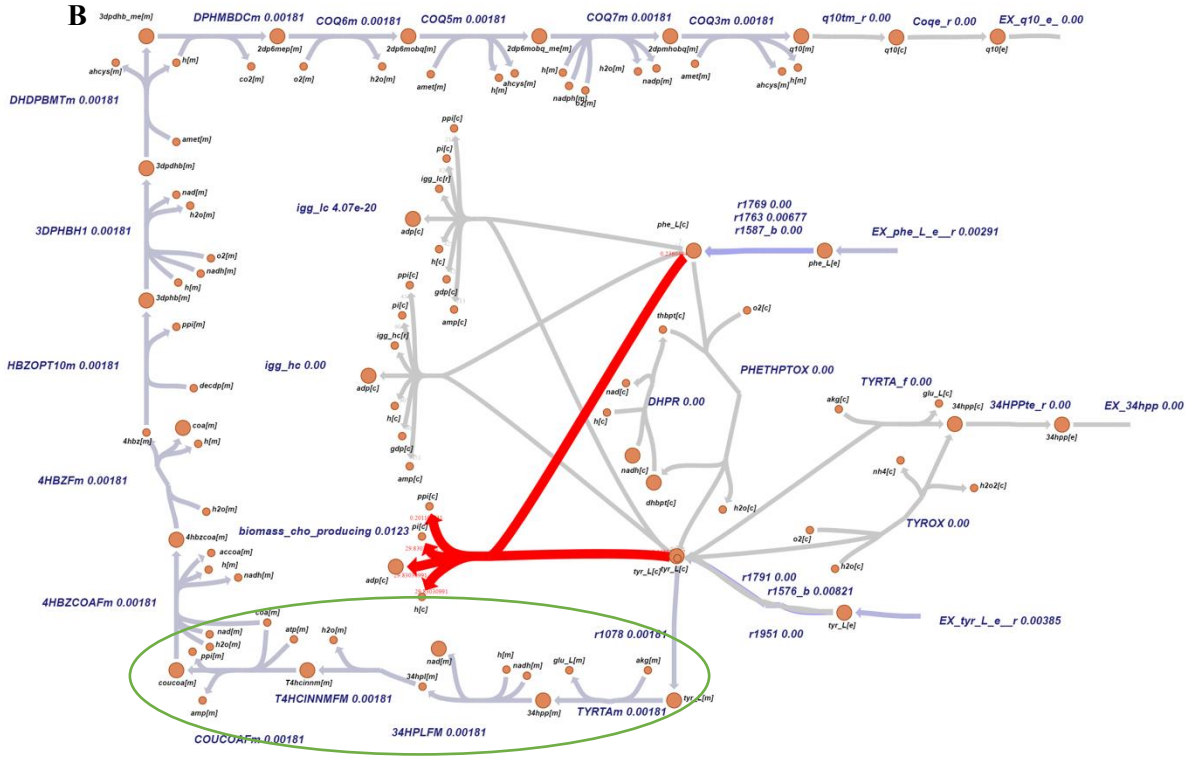
production. When it is transferred to mitochondria it is catabolised in the ubiquinone pathway and converted to ubiquinone-10, which is then secreted from the cell. However, the whole ubiquinone pathway consumes large amounts of energy, and is not beneficial to biomass synthesis, which is the reason the optimisation chooses to allocate the maximum possible tyrosine initially to biomass production and then to convert the remaining to 3-(4-hydroxyphenyl) pyruvate and then secrete it. Phenylalanine, when FBA is performed on the D1 dataset with no regulations, is simply directed to biomass production. The removal of 3-(4-Hydroxyphenyl) pyruvate (34HPP) secretion results in 3-(4-Hydroxyphenyl) pyruvate to not be able to exit the cell (Fig. 5.17B), so it cannot be produced as there is no other pathway it can be catabolised. Thus, the enzymes L-tyrosine oxidase (TYROX) and tyrosine transaminase (TYRTA), which both convert tyrosine to 34HPP, cannot carry any flux and tyrosine is forced to be channelled to biomass or antibody production or be transferred to mitochondria from cytosol.

If there is no biomass inhibition, the remaining tyrosine is channelled to the mitochondria to be converted to ubiquinone-10 and then secreted from the cell, which does not reduce the growth rate because there would still be enough resources to sustain the biomass production. Phenylalanine displays no changes in its catabolism and the biomass production consumes all of phenylalanine entering the cell. An inhibition in biomass growth, which is caused by applying the three regulations of PPAm, GLXO1 and GUAt (Fig. 5.17C) disrupts phenylalanine catabolism. There are no pathways catabolising phenylalanine in the CHO-K1 GEM, therefore it has to convert to tyrosine through the dihydrofolate reductase (DHFR) enzyme. Tyrosine is then converted and secreted from the metabolic network as 3-(4-hydroxyphenyl) pyruvate. The inhibition of growth rate combined with the removal of 3-(4-Hydroxyphenyl) pyruvate secretion (Fig. 5.15D) results in an abundance of phenylalanine, which is converted to tyrosine. In this case however, tyrosine cannot exit the cell as 3-(4-hydroxyphenyl) pyruvate, but only transfer to mitochondria to be catabolised there. The catabolic pathway of tyrosine in mitochondria is energetically unfavourable, leading to a result a trade-off point between the tyrosine that is headed to mitochondria and tyrosine that is headed to antibody production. Phenylalanine is therefore channelled to antibody synthesis as well as in biomass production.

A


DPHMBDCm 0.00
COQ6m 0.00
COQ5m 0.00
COQ7m 0.00
COQ3m 0.00
q10tm_r 0.00
Coqe_r 0.00
EX_q10_e_ 0.00
DHDPBMTm 0.00
3DPHBH1 0.00
igg_lc 0.00
r1769 0.00
r1763 0.00
r1587_b 0.00291
EX_phe_L_e__r 0.00291
HBZOPT10m 0.00
igg_hc 0.00
PHETHPTOX 0.00
TYRTA_f 0.00181
34HPPte_r 0.00181
EX_34hpp 0.001
DHPR 0.00
TYROX 0.00
4HBZFm 0.00
biomass_cho_producing 0.0123
r1791 0.00
r1576_b 0.00821
EX_tyr_L_e__r 0.00385
4HBZCOAFm 0.00
r1078 0.00
r1951 0.00
T4HCINNMFM 0.00
TYRTAm 0.00
COUCOAFm 0.00
34HPLFM 0.00

B


DPHMBDCm 0.00181
COQ6m 0.00181
COQ5m 0.00181
COQ7m 0.00181
COQ3m 0.00181
q10tm_r 0.00
Coqe_r 0.00
EX_q10_e_ 0.00
DHDPBMTm 0.00181
3DPHBH1 0.00181
igg_lc 4.07e-20
r1769 0.00
r1763 0.00677
r1587_b 0.00
EX_phe_L_e__r 0.00291
HBZOPT10m 0.00181
igg_hc 0.00
PHETHPTOX 0.00
TYRTA_f 0.00
34HPPte_r 0.00
EX_34hpp 0.00
DHPR 0.00
TYROX 0.00
4HBZFm 0.00181
biomass_cho_producing 0.0123
r1791 0.00
r1576_b 0.00821
EX_tyr_L_e__r 0.00385
4HBZCOAFm 0.00181
r1078 0.00181
r1951 0.00
T4HCINNMFM 0.00181
TYRTAm 0.00181
COUCOAFm 0.00181
34HPLFM 0.00181

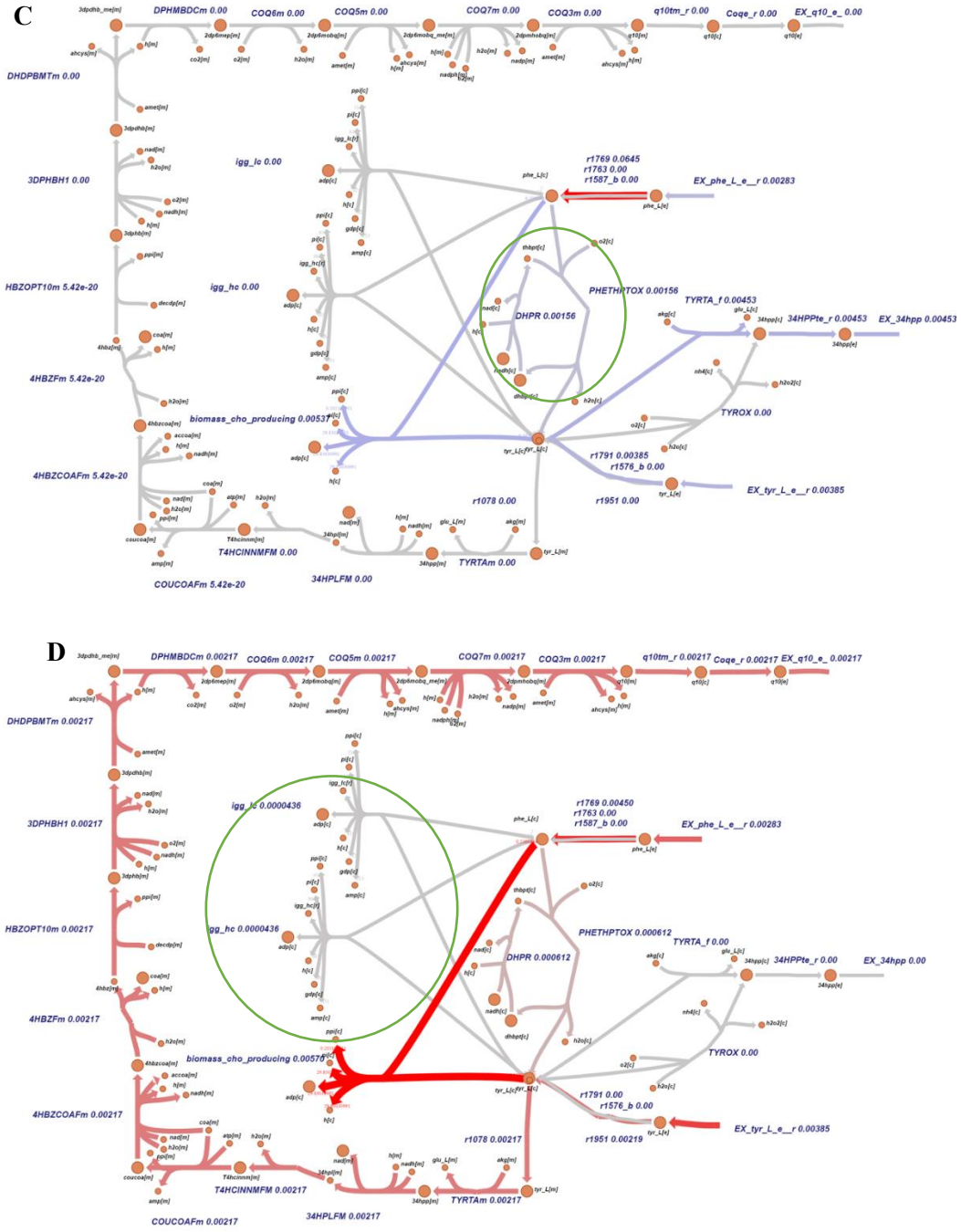
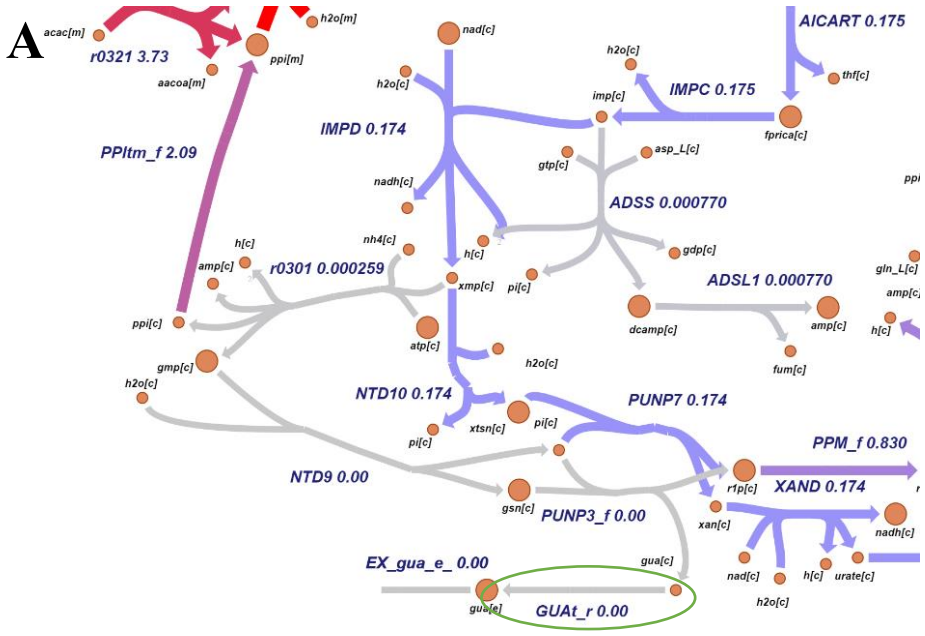


Figure 5.17: In the original CHO-K1 GEM, constrained under D1 dataset, 34hpp can be secreted from the cell (A). In case of an abundance of either tyrosine or phenylalanine, phenylalanine can be converted to tyrosine which can be oxidised to 3-(4-hydroxyphenyl) pyruvate and leave the cellular environment. Removing the option of 3-(4-hydroxyphenyl) pyruvate secretion (B) tyrosine has to transfer to mitochondria and catabolise to q10. Appling an inhibition on biomass production (C) simply directs the remaining phenylalanine and tyrosine to be secreted from the cell in the form of 3-(4-hydroxyphenyl) pyruvate. However, with the combination of both biomass inhibition and removal of the secretory pathway of 3-(4-hydroxyphenyl) pyruvate (D), there is an abundance in phenylalanine leading in the synthesis of monoclonal antibody.

In order to perform the genetic strategy *in vitro*, the regulations of the reactions' bounds need to be converted to genetic modifications that can happen on the CHO-K1 genotype. Initially, both reactions GLXO1 and GUAt must be removed from the metabolic network. GLXO1 is co-expressed by *Ldha* and *Ldhb*, which can be knocked out since they do not belong in the experimentally detected essential genes database (*183*), neither in the essential genes predicted from the FBA optimisation for the four datasets used. PPAm is expressed by *Loc100756805*. The enzyme GUAt is co-expressed by *Loc100754792* and *Slc29a2*, neither of which are essential based on FBA prediction, but *Loc100754792* has been identified as essential for growth in the experimental knockout screen (*183*). However, if guanine cannot be created then it cannot be secreted, so by disrupting its production pathway the result in the metabolic network should be the same. The 5'-nucleotidase enzyme (NTD9), which is co-expressed by *Nt5c* and *Nt5c2*, is responsible for the conversion of gmp to gsn, a vital part of guanine synthesis. The removal of the enzyme can disable the formation of guanine, which can be done since none of the genes expressing it have been identified as essential for growth. Therefore, instead of removing the transport of guanine from cytosol to the extracellular compartment, removing the conversion of gmp to gsn accomplishes the same result as seen in figure 5.18.



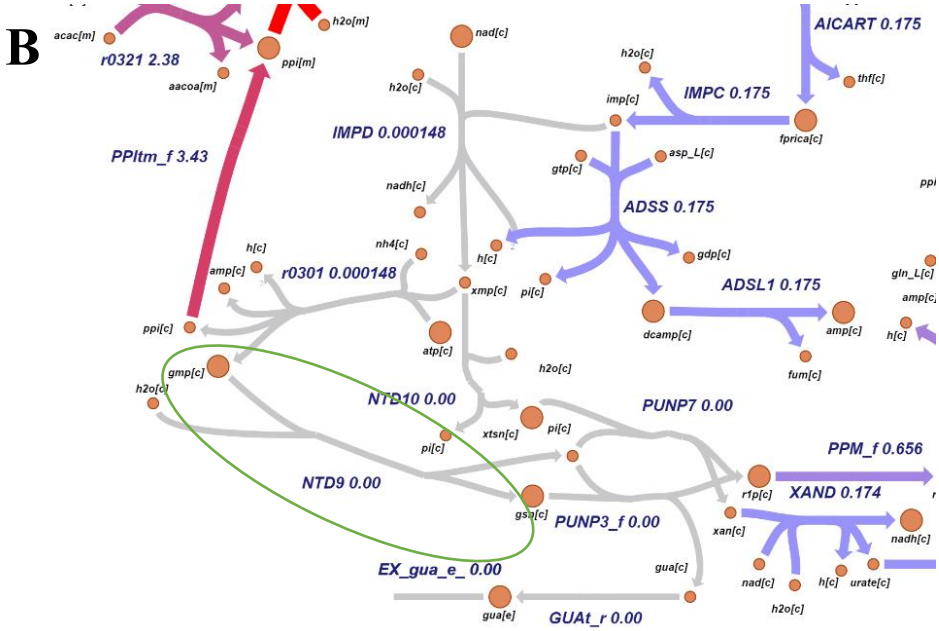


Figure 5.18: The removal of GUAt_r (A), which is co-expressed by *Loc100754792* and *Slc29a2*, results in the same phenotype as with the removal of NTD9, which is co-expressed by the non-essential genes *Nt5c* and *Nt5c2*.

Lastly the removal of the 34hpp secretion forces the antibody production in the CHO-K1 GEM. The secretion is not expressed by any known genes, however if both tyrosine transaminase and tyrosine oxidase are knocked out then 34HPPte cannot be synthesised in the first place, to be secreted afterwards.. Tyrosine transaminase is expressed by the non-essential genes *Got1|1* and *Got1*, but there are no known genes expressing tyrosine oxidase for CHO cells. *Il4il* expresses L-amino acid oxidase, so it could disrupt the expression of tyrosine oxidase. The gene is associated with L-Aspartate, L-Methionine, L-Phenylalanine and L-Tryptophan oxidases in the iCHO 1766 GEM, but it is also associated with L-tyrosine oxidase in the RECON3D GEM (*188*). The CHO-K1 GEM produces both biomass and antibody with all the mentioned oxidases knocked out, so *Il4il* is predicted as non-essential by the FBA optimisation. However, in the gene database categorises it as essential, so it was decided to not alter the pathway (*183*). The upregulation of hydroxybenzoate decaprenyltransferase (HBZOPT10m), conveniently expressed only by the gene *Coq2*, directs tyrosine towards mitochondria to be catabolised in the ubiquinone pathway resulting in a reduced growth rate as explained above. The reason there is no 3-(4-hydroxyphenyl) pyruvate secretion when HBZOPT10 is upregulated is because there is not enough tyrosine to supply both 3-(4-hydroxyphenyl) pyruvate formation and mitochondrial tyrosine catabolism. However, the upregulation of the HBZOPT10m reaction can make the problem infeasible due to lack of tyrosine uptake. There should be an additional feed of tyrosine, in order to supply the increased flux. The similar phenotypes of the two genetic strategies as presented in figure 5.19.

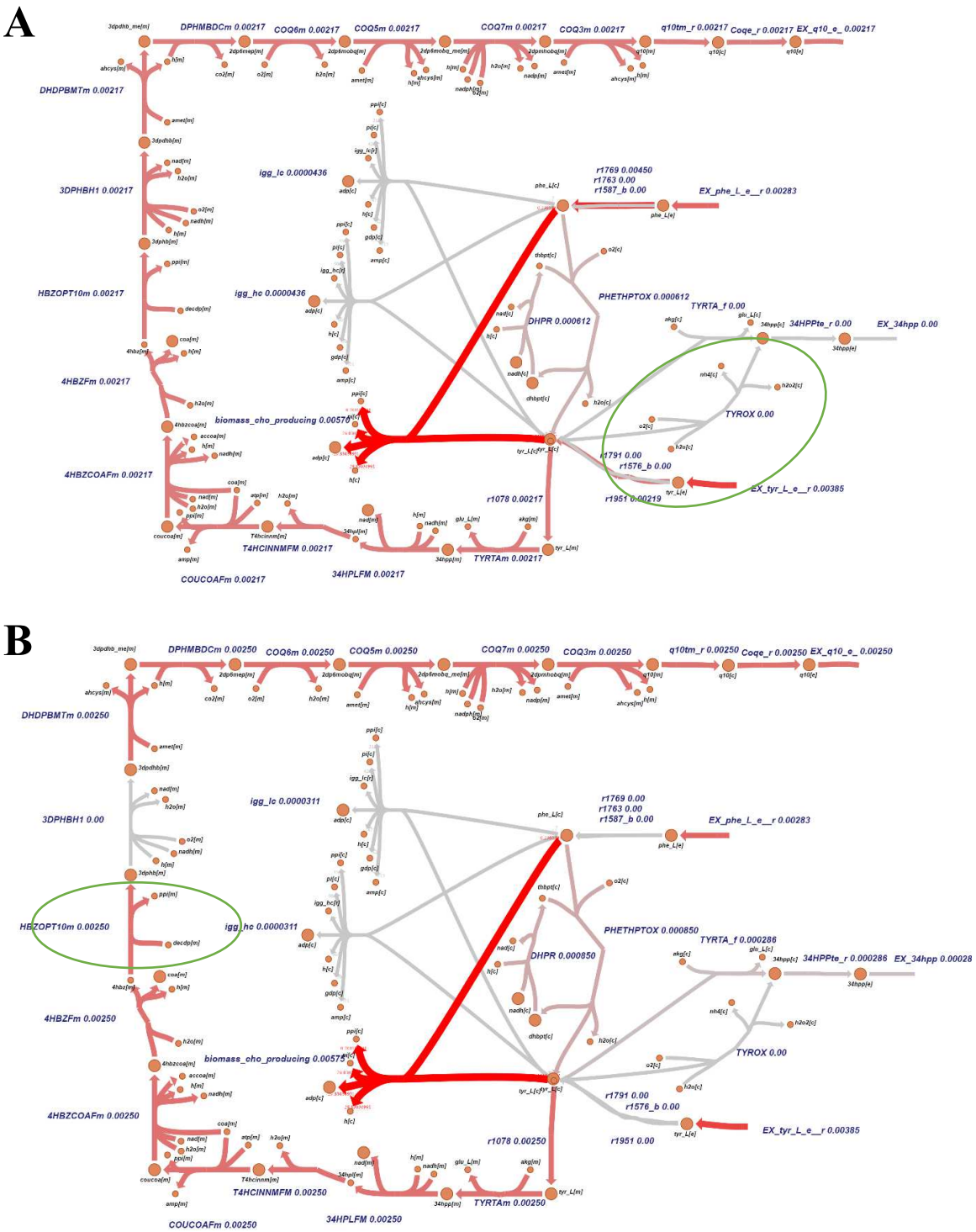


Figure 5.19: The removal of tyrosine oxidase and tyrosine transaminase enzymes (A) has the same effect on the metabolism as with the increase of flux of hydroxybenzoate decaprenyltransferase (B). In both cases there is no secretion of 3-(4-hydroxyphenyl) pyruvate. In the first case because tyrosine cannot get oxidised, and in the second case because there is not enough tyrosine to supply both pathways.

5.3 Conclusion

The strain design algorithm resulted in 93 genetic engineering strategies – scenarios – for the four different experimental datasets used. The scenarios contain regulations for the lower or upper bounds of the intracellular reactions' fluxes, which change the metabolic network of the original CHO-K1 GEM. The resulting network can synthesise antibody in an FBA optimisation, even though the objective function is biomass production. The average number of regulations for the 93 scenarios is 130, which means, on average, each scenario applies 130 changes in the lower or upper bounds of the reactions' fluxes. The very large number of changes makes the study of each scenario, regarding the suggested changes in metabolism very difficult. Moreover, an *in vitro* validation of the scenario is very unlikely and challenging with such a high number of genetic modifications. A sensitivity analysis on the optimisation results can reduce the number of regulation so that the new scenarios would contain only the essential regulations on the reactions bounds, for the production of IgG in the CHO-K1 GEM.

A 3-stage sensitivity analysis framework was developed with the goals of reducing the total number of regulations proposed by each set and identifying the scenarios with the highest μ and q_{mAb} – high productivity scenarios. The reduction of the total number of regulations followed an iterative approach in which one by one the regulations of each scenario were checked to see whether they were beneficial to the output. As a result, the SA reduced the average number of regulations for the 93 scenarios from 130 to 9 with some scenarios having as low a number of regulations as two. After the reduction, four scenarios were identified as duplicates, reducing the total number of scenarios from 93 to 89. 89 scenarios were still a lot to handle, thus it was decided to identify the high producing ones. Each scenario proposed by the strain design algorithm had a value for μ and q_{mAb} associated with it, which was predicted by the FBA optimisation with biomass maximization being the objective function. The comparison between these values and the experimental ones for each of the datasets was a conclusive measure to identify the high performing scenarios. 20 of the 89 scenarios had both μ and q_{mAb} higher than 50% of the control values calculated in the previous chapter. 50% was considered the cut-off point to result in scenarios with high enough growth rate and antibody-specific productivity. The control values are the result from an FBA optimisation making them the highest possible biomass or antibody produced for a specific dataset, meaning that the μ and q_{mAb} from the strain optimisation will be lower. Finally, a combinatorial algorithm verified that the number of regulations in the 20 scenarios are the minimum possible ones for the production of both biomass and antibody from the FBA optimisation.

The 20 final scenarios have a reasonable number of regulations on the reactions' bounds, on average 5 per scenario, and a high enough μ and q_{mAb} to make them ideal candidates for a study concerning the changes of the metabolic network proposed and an *in vitro* application. All the scenarios proposed two important changes in the metabolism, and in general, according to the results of the strain

design optimisation performed, the objectives of biomass and antibody production can be coupled in an FBA optimisation by one way. Both algorithms, Igg_Igg and Bio_Igg, suggested the creation of an amino acid abundance inside the cell by removing its catabolic pathways, combined with an inhibition of growth rate by disabling the most optimal for growth pathways or by increasing the energy demand of the cell or by disabling the most energetically favourable catabolic pathways. This strategy forces the remaining flux of the amino acid to be channelled toward antibody production in order for it to be consumed since there is no accumulation allowed in FBA. The amino acid can be any of the ones that are uptaken by the cell, as long as there is remaining quantity after the production of biomass. Based on that conclusion, the 20 scenarios can be categorised according to the amino acid in abundance. The proline, leucine and phenylalanine scenarios make up the 20 most optimal scenarios, and the rest 69 can be categorised in lysine, valine and asparagine abundance all coupled with growth inhibition. An example scenario from each category of the 20 high-producing scenarios was analysed in detail, showcasing the changes in metabolism that create an amino acid abundance and how the growth rate is being inhibited.

The analysis of the three proposed scenarios suggests a viable way that μ and q_{mAb} can be coupled in an FBA optimisation for the four different experimental datasets used. For the coupling of the two diametrical objectives to take place *in vitro*, genetic engineering is required. Based on the suggested regulations by the optimisation algorithm, a set of genes to be regulated is derived from each of the high-producing scenarios studied in depth. Performing the suggested genetic modifications would potentially result in a reduction of growth and an increase of the monoclonal antibody secretion per cell. All datasets used contain the lower and upper bounds for the exponential phase, so the increase of q_{mAb} is expected to take place during the growth phase of the cell culture.

6

Experimental Materials and Methods

The sixth chapter aims in explaining the materials and protocols used for the cell culturing and the genetic engineering conducted. Initially, the cell-line cultivation protocol is explained and then the process of the fed-batch culture is explained. For the part of the genetic engineering the transfection method is briefly explained together with the design of the plasmids used. After the transfection of the plasmids, it is required to confirm the genetic engineering took place, so, the analytical assays used are explained as well as the method for antibody quantification.

6.1 Cell Line, maintenance, and storage

The original cell line used in this study is a CHO-K1 cell line. It expresses glutamines synthetase and the cB72.3 chimeric IgG4 (immunoglobulin of γ isotype and 4 subclass) antibody. The cell line is revived and cultured at CD-CHO medium (ThermoFisher Scientific, U.K.) with the addition of 1mL/L of medium 25mM MSX, as a selection marker for glutamine synthetase. The seeding density at revival is 3 x 10^5 and for the subsequent subcultures 2 x 10^5 viable cells/mL. Depending on the culture volume, Erlenmeyer flasks (Corning Inc. NY, USA) are used to grow cells, at 140rpm in a humidified 36.5°C incubator with 8% CO_2 supply and are subcultured every 3–4 days. The cell concentration and cell viability are measured using the trypan blue dye exclusion method. For storage, cell pellet containing 1 x 10^7 viable cells is resuspended in 1mL of freezing agent, containing 92.5% CD-CHO medium and 7.5% dimethyl sulfoxide (DMSO; Sigma-Aldrich, Dorset, UK) acting as cryoprotectant. The cell vials are stored at -80°C for 1 day in a Nalgene® Mr Frosty prior to being transferred to liquid nitrogen (-160°C) for long term storage.

6.2 Fed-Batch culture

The feed was prepared according to the manufacturer instructions. Initially 79.6g of CD EfficientFeed^{TM} C AGT^{TM} was added in 0.8L of distilled water and mixed until homogenous. Then the pH was adjusted to 9.2 with the addition of 5N NaOH and then to 7.0 with the addition of 5N HCl. The final volume was adjusted to 1L with the addition of distilled water. FeedC was finally sterilized using membrane filtration and was stored in a dark container at 4^{o}C. The fed-batch culture took place

in 125mL Erlenmeyer flasks in a 30mL working volume. Feeding took place from day 2 of the fed-batch culture and every other day until day 12. In order for the volume to be as close to 30mL as possible, the daily sample volume is 1.5mL, and 3mL (10% of culture volume) was added to the culture in the form of feeding.

6.3 Genetic engineering

The gene modification performed herein requires a plasmid which contains the necessary RNA sequences and a required a method for the plasmid to enter the cell, transfection. The method of choice for transfection is electroporation due to its simplicity and efficiency. Then, each of the six plasmids used in Chapter **6** are described.

6.3.1 Plasmid transfection

For the DNA transfection 40mg of DNA were added to 10^7 cells. Initially, 4 x (1.43 ·10^7) cells were centrifuged at 800rpm for 5min. The cell pellet was then resuspended in prewarmed (37 ºC) CD CHO medium, to a viable cell density of 1.43 ·10^7 cells/mL. 40mg of DNA was diluted in 300mL of PBS (DNA solution). 100µL of the DNA solution were added to an electroporation cuvette of 0.4cm gap width. 700µL of the resuspended cells were transferred to the cuvette and cells were electroporated on the electroporation system Gene Pulser Xcell from BIO-RAD with 300V and 900µF capacitance. The procedure was repeated again two more times to maximise the total number of transfected cells, and the transfected cells were then transferred to a 125mL Erlenmeyer flask with 27mL prewarmed (37 ºC) CD CHO. For the control experiment the 10^7 cells were transfected with 100µL of PBS and the transfected cells were transfected to a 125mL Erlenmeyer flask with 29mL prewarmed (37 ºC) CD CHO.

6.3.2 FACS Cell Sorting

For the cell selection using FACS (Fluorescence-Activated Cell Sorting), initially 10^7 transfected cell were gathered and washed using PBS, pH 7.4 (ThermoFisher, UK). Then, they were centrifuged an isolated to be then added to 1mL of PBS. The 10^7 cells were filtered through a cell strainer-cap FACS tube. Finally, they were transferred to the Department of Life Sciences of Imperial College London for sorting. The collected sample was then transferred back to the lab, and the cells were cultured according to the protocol of **6.1** section.

6.3.3 Plasmid design

Table 6.1 shows all the plasmids used in Chapter 6 to knockout the genes *Aldh4a1*, *Bcat1* and to overexpresess *Aldh18a1*. For the design of most plasmids two different online tools were used: CHOPCHOP (*189*) and CRISPy (*190*). The plasmid used to knockout *Bcat1*, BNu1, was designed based on the paper by Anderson et al. (*191*). All plasmids were designed and ordered in VectorBuilder, except the ORF plasmid A18ORF which was ordered from Origene.

Table 6.1: Plasmids used for the genetic engineering experiments. First column denotes the number of each plasmid. Second column has the name of each plasmid, where Ni means that the Cas9 protein is nickase, and Nu means the Cas9 protein is nuclease. The description contains the basic information about each plasmid. Fourth column contains the sgRNA used for the knockout with the PAM sequence in **bold**, except the ORF plasmid which has no sgRNA.

#	Plasmid	Description	sgRNA Targeting Sequence	Reference
1	A4Ni1	Knockout plasmid, carrying the Cas9 (D10A) nickase protein and the EGFP selection marker. It cleaves the exon 2 of the *Aldh4a1* gene.	5'– GGTGCTTCCCACAGCGGTAA**GGG** – 3'	CHOPCHOP
			5'– CGCACCACCTCTGCGGGAAG**TGG** – 3'	
2	A4Nu1	Knockout plasmid, carrying the Cas9 nuclease protein and the EGFP selection marker. It cleaves the exon 2 of the *Aldh4a1* gene.	5' – GGTGCTTCCCACAGCGGTAA**GGG** – 3'	CRISPy
3	A4Nu2	Knockout plasmid, carrying the Cas9 nuclease protein and the mCherry selection marker. It cleaves the exon 1 of the *Aldh4a1* gene.	5'– GCATACTTCTTAGGTCGGGA**TGG** – 3'	
4	A4Nu3	Knockout plasmid, carrying the Cas9 nuclease protein and the Hygromycin resistance gene as a selection marker. It cleaves the exons 1 and 2 of the *Aldh4a1* gene.	5' – GGTGCTTCCCACAGCGGTAA**GGG** – 3'	CRISPy
			5' – GCATACTTCTTAGGTCGGGA **TGG** – 3'	
5	BNu1	Knockout plasmid, carrying the Cas9 nuclease protein and the Hygromycin resistance gene as a selection marker. It cleaves the	5' – GGCTCTGACATATTTCGGAT**TGG** – 3'	Andersen et al.

		exon 5 or 6 of the *Bcat1* gene in its isoforms X1 and X2.		
6	A18ORF	The NM_019698 (Origene) ORF plasmid, carrying the *Aldh18a1* mRNA sequence and the Neomycin resistance gene as a selection marker.	Mouse *Aldh18a1* mRNA	Origene

6.3.4 Plasmid DNA purification

Plasmid purification was performed using QIAprep Spin Miniprep Kit (Qiagen) and the QIAGEN Plasmid Maxi Kit (Qiagen) as specified by the manufacturer. DNA concentrations were measured using the BioDrop DUO+ (Biodrop, UK) spectrophotometer.

6.4 Analytical Assays

6.4.1 Cell Lysis

1×10^7 viable cell pellets were rinsed with 4°C PBS prior to cell lysis in 200 µL of RIPA Buffer (Sigma-Aldrich) supplemented with 1% (v/v) protease inhibitor cocktail (Sigma–Aldrich). Samples were gently shaken for 10 min before sonication on ice using 3 bursts of 5 s each at 25 s intervals and an amplitude power of 20%. Cell debris was removed by centrifugation at 14,000*g* for 15 min. The supernatant was stored at −80°C prior to Western blot analysis.

6.4.2 Western Blot

Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS –PAGE) was performed with Mini-PROTEAN Tetra Cell (Biorad) using 12% Tris-HCl SDS-PAGE gels (running gel) and 5% Tris-HCl stacking gel. The composition of both is presented in table 6.2.

Table 6.2: Recipe for 2xSDS-PAGE gels

Component	***Volume for running gel***	***Volume for casting gel***
Deionised H_2O	3.191mL	3.400mL
40% Acrylamide/Bis-Acrylamide	4.000mL	0.833mL
1.5M Tris-HCl (pH 8.8)	2.500mL	0.630mL
10% SDS	0.150mL	0.050mL
10% APS (Ammonium Persulfate	0.150mL	0.050mL
TEMED	0.009mL	0.005mL
Final Volume	10mL	5mL

Cell lysates were diluted using a 2x SDS-loading dye (2X SDS loading dye 0.09 M Tris-HCl, pH 6.8; 20% glycerol; 2% SDS; 0.02% bromophenol blue; 0.1 M DTT) to a final concentration of 1X SDS-loading dye and then with PBS to a final protein concentration of 2 – 10mg/mL. All samples were denatured by heating them for 10 minutes at 95 ^{o}C. For loading the gel, 5µl of each protein sample was used as well as 5µl of the Thermo Scientific™ PageRuler™ Prestained Protein Ladder (10–180 kDa) for the identification of the band size. Gels were run at constant amplitude (25mA/gel) for 90 minutes, with 1X 10xTris Glycine SDS buffer (0.25M Tris, 1.92M Glycine, 1% SDS, pH 8.6).

For western blot analysis a Mini Trans-Blot® Cell was used for electroblotting the gels to an Immobilon®-FL PVDF membrane (Millipore Ltd, Herfordshire, UK). The arrangement of the semi-dry western blot involved 1 foam pad (biorad), 2 Blot Filter Paper (Biorad), the gel, the PVDF membrane 2 Filter Papers and 1 foam pad. All components were previously soaked in transfer buffer (per litre: 3.03g tris base, 14.27g glycine, 20% methanol). The PVDF membrane was activated as follows: the membrane was incubated in methanol for 30 seconds, then in deionised H_2O for 2min, and finally in transfer buffer for a minimum of 5 minutes. The gels were washed multiple times with deionised H_2O before placing to transfer buffer. The transfer was run for 60 minutes at 100 Volts/0.35A in ice.

For the blotting following buffers were used: 10X TBS stock, per litre, autoclaved prior to use: 80 g NaCl, 2 g KCl, 30 g Tris Base, pH 8.0 with HCl), 0.05% TBS-T (0.05% Tween-20, 1X TBS); 0.5% TBS-T (Washing Buffer: 0.5% Tween-20, 1X TBS); blocking buffer (5% skimmed milk in 0.05% TBS). All the blotting procedure took place at room temperature on a shaking platform. After the transfer, the PVDF membrane was blocked in blocking buffer for 30 minutes. An hour of primary antibody incubation (diluted in blocking buffer) followed. The primary antibody was diluted in blocking buffered according to the manufacturer's instructions varying between each of the antibodies used. Residual primary antibody was washed off using 3x TBS-T washing buffer for 5 minutes. A secondary antibody (Goat Anti-Rabbit IgG H&L from Abcam) was incubated for 30min (1:2000 dilution in

blocking buffer). Residual secondary antibody was removed with 3x TBS-T washing buffer for 5 minutes. The membrane was finally washed 2 x 2 minutes wash with deionised H_2O. Finally, bands were developed with 5mL of alkaline phosphate substrate BCIP/NBT kit (Thermofisher scientific).

To measure the intensity of the bands from the western blot the software TOTALLAB CLIQS 1D Gel Image Analysis was used. Background subtraction was performed using the rolling ball method.

6.4.3 Antibody quantification

Recombinant protein was quantified using an IgG4 antibody ELISA kit (Bethyl Laboratories, US). Specific rates of monoclonal antibody production were calculated by plotting the integral viable cell concentration (IVCC) vs the cumulative concentration. IgG4 concentration was also verified with a biolayer interferometry device (BLItz® system, ForteBio, Pall, US).

7

Genetic Engineering

In the previous chapter, three different scenarios of modifications for the fluxes of metabolic reactions were selected. When these modifications are applied to the constrained by the D1-4 datasets CHO-K1 GEM, the biomass maximisation FBA results in positive IgG flux. The analysis of the changes in the metabolism suggested that the abundance of an amino acid in the cell leads the FBA optimisation algorithm to synthesise antibody by consuming the excess amino acid after most of it is channelled towards biomass production. It can be any of the amino acids that the cell uptakes, as long as there is remaining quantity after the maximum biomass production. If the catabolic pathways of the amino acid are removed, then it will be in abundance in the cell forcing it to be consumed by the antibody synthesis pathway. In the analysis of the optimisation results, the amino acids were proline, lysine, leucine, asparagine, valine and phenylalanine. The idea behind abundance of amino acids in cells has not been studied, especially connected to increased specific productivity of the recombinant protein product. It is important to test this hypothesis *in vitro* to test its validity.

The best way to test the hypothesis, is to perform the genetic strategy on a CHO cell line, culture it under the same fed-batch conditions as in the dataset used for the optimisation algorithm and observe any potential changes in specific productivity and possibly in the growth rate. Initially, the changes in reaction fluxes suggested by the optimisation were converted to genetic engineering strategies, containing possible knockouts and overexpression of genes that exist in the CHO-K1 genome. These strategies can be tested *in vitro* by applying the genetic modifications on the original mAb producing CHO cell line. The time limitations of the project allowed the testing of two out of the three scenarios that were analysed in depth in the previous chapter, specifically, the abundance of leucine and of proline in the cell and the impact it has on antibody productivity in CHO. The phenylalanine scenario was not tested due to the difficulty of the genetic engineering strategy. The upregulation of the reaction would require an increase of tyrosine in the feed, which could change the feeding composition and then possibly the uptake and secretion rates of the dataset that were used to achieve the optimisation result. The leucine and proline strategies were chosen because the combination of their predicted μ and q_{mAb} and of the small number of required modifications on the genotype. FBA usually over predicts the values of the predicted fluxes as seen in section **3.3**, so μ and q_{mAb} of the engineered GEM were compared to the FBA predicted values of the original unmodified CHO-K1 GEM. The relative increase

of μ and q_{mAb}, of the proline and leucine abundance scenarios was higher than of the other scenarios, which was the reason these two scenarios were picked.

The first strategy is focused on proline metabolism and involves knocking out the gene *Aldh4a1* and overexpressing *Aldh18a1*, creating the CHO-AA engineered strain. The catabolism of leucine is the second strategy that involves disrupting the normal expression of the gene *Bcat1*, thus creating the engineered CHO-B strain. Part of the engineering strategy suggested by the optimisation involved the reduction of specific growth rate. Instead of achieving this through genetic modification by overexpressing the genes *Acss2* and *Aacs*, it was decided to culture the CHO-B strain under mild hypothermic conditions to reduce the growth rate of the cell. CHO-B was cultured under fed-batch conditions and the titer was compared to the original CHO cell line.

The chapter initially provides a theoretical background behind genetic engineering. Then it describes how each of the two cell lines were created and finally it compares the changes in the phenotype of CHO-B with the original CHO cell. In the end, the conclusions are presented, questioning whether the engineering achieved its goal.

7.1 Genetic Engineering

Strain design algorithms suggest interventions in the genome to achieve a specific cell phenotype as described in the section **2.2.2** of the literature review. The interventions can be performed *in vitro* and generate an engineered cell line with the objective of having the same phenotype as the model predicted one. There are many ways a genetic intervention can be achieved using different transfection techniques that are discussed below.

7.1.1 Genetic Modification

The first step in performing a genetic modification is deciding how it can be performed *in vitro*. Genetic modification means a change in the natural expression of a gene that can be modified by either amplifying its expression, overexpression, reducing the level of expression, down-regulating, or disabling the expression with a knockout. Gene overexpression can result in an increased concentration of the translated enzyme in the cell, which would correspond to increased metabolic activity of the specific pathway. Gene knock-out, or also called gene disruption, means the removal or complete silencing of the gene's expression in the genome, so the specific enzyme will not be present in metabolism. A knockout can also be performed by disrupting the normal functionality of an enzyme, so it will not be able to perform its catalytic duties. If the sequence of the gene corresponding to the active site of the enzyme is faulty, it is possible the active site that the substrate binds to the enzyme to not translated properly. The downregulation of a gene results in reduced expression of a gene leading

to reduced levels of the expressed enzyme in the metabolism and a reduction in the metabolic activity. In this chapter, two different genetic modifications are performed, a gene knockout and an overexpression. There are different ways the modifications of the genome can take place *in vitro*; Zinc-fingers, TALENS or CRISPR/Cas9 are all techniques used for the disruption of a gene and rely on engineered site-specific nucleases that introduce double-strand breaks (*192*). Overexpression of a gene can take place using CRISPRa to induce gene expression or by transfecting the ORF sequence of the gene.

Zinc-finger nucleases (ZFNs) are artificial restriction enzymes - enzymes that cleave DNA into fragments - generated by fusing a zinc finger protein, a structural motif with specific coordination of one or more zinc ions, with a restriction enzyme. The zinc finger protein can be engineered to target a specific DNA domain. They induce sequence-specific DNA double-strand breaks (DSBs) (*193*), which can then be repaired using an error-prone nonhomologous end joining (NHEJ) to result in small changes in the targeted genomic loci, using the nuclease domain (*194*). The main limitations of the method are the difficulty in designing the protein with high affinity and the lack of target site selection (*192*).

Transcription activator-like effector nucleases (TALENs) are, similarly to ZFNs, restriction enzymes engineered to cut in specific regions of the genome. They are made using TAL effector, which is a type of protein with a repeated domain and fusing it with a DNA nuclease. The repeated domain of TALEs can bind to any desired DNA sequence, which when it is fused with nuclease it can nick the genome in any place specific region of DNA (*195*). Recently, it was claimed that the TALEN system outperforms CRISPR as a gene editing tool (*196*), however the complicated design and high cost make the method less appealing.

Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) is a family of DNA sequences found in prokaryotic organisms. They are part of the organism's immune system and are responsible for detecting and destroying infected sections of the genome. Cas9 is an enzyme that uses these DNA sequences as a guide to identify and cleave strands of the DNA. The transfection of a Cas9 protein together with an RNA sequence, which is called single guide RNA (sgRNA), will induce a double strand break three bases upstream the protospacer adjacent motif (PAM) sequence, which is usually repaired by NHEJ. The PAM sequence serves as a binding signal for Cas9 and usually has the form 'NGG' (N meaning any of the four bases) (*197*). The sgRNA can be of varying length, usually between 17-24 bases,, obey certain rules, and have the minimum number of exact copies in the genome (*198*). The simplicity of the design, low cost, and high efficiency of the method make it very appealing for genetic engineering.

There are two main types of Cas9 proteins, the widely used SpCas9, which is found in *Streptococcus pyogenes Cas9* and Cas9(D10A), which is a Cas9 nickase protein. SpCas9 induces site-directed double stranded breaks in DNA, which later insertions or deletion (INDELS) mutations cause by NHEJ can negatively affect the expression of the gene. Cas9(D10A) (*199*), is an engineered Cas9 protein which cuts one single strand of DNA that is complementary to the sequence supplied. The use of two different

sequences the Cas9 nickase produces two different single-strand cuts causing a double strand break. SpCas9 can also be used with multiple gRNA sequences, but there is a higher chance of an off-target mutation (*200*). There are two orientations, PAM-out and PAM-in, when two gRNA sequences are supplied as seen in Figure 7.1, where PAM-out is considered much more efficient.

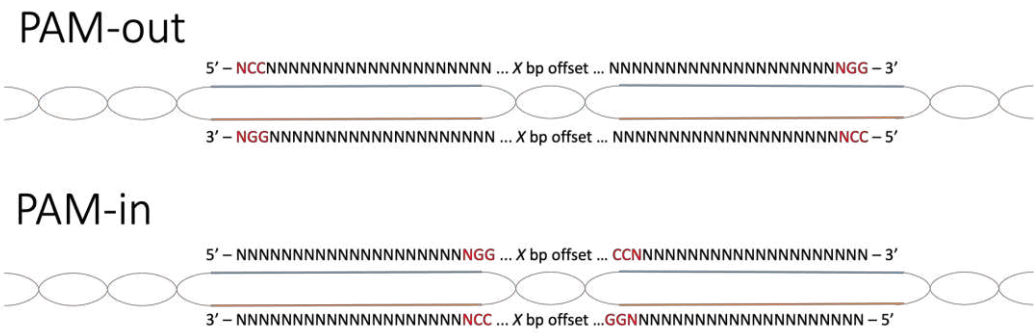


Figure 7.1: Graphical representation of the different designs for dual sgRNA PAM sequences. In the PAM-out design, top, the two PAM sequences have much larger offset between the PAM-in designs, bottom.

Overexpression of a gene can take place due to processes such as gene amplification or by modulating the gene expression. Gene amplification is the addition of extra copies of genes, which can lead to increased levels of mRNA, and, if translated to protein, can then cause an increase in the protein quantity in the cell. A way to add additional copies is to use Open Reading Frames (ORFs). An ORF or protein-coding region carries the required information used for protein synthesis. It carries the same information as in mRNA, which can then be transcribed for the synthesis of the required protein. It is not the addition of additional copies of genes, but of the information they carry. Modulation of expression in genes in the genome can be accomplished using CRISPR activation (CRISPRa). CRISPRa is a modified version of CRISPR/Cas9 that was analysed above. The Cas9 protein is catalytically dead and fused with a transcriptional activator. Transcriptional activators are proteins, or protein domains that assist in the recruitment of important cofactors for the transcription of a gene, resulting in possible increased expression. The use of an ORF plasmid seemed like a better option for overexpressing the gene, due to its higher expression levels and faster design (*201*).

7.1.2 DNA Transfection

Another important step for a genetic modification is the transfer of the DNA material to the cell, known as transfection. There is a variety of different transfection techniques, each of which has a specific set of advantages and disadvantages and is best used in a certain scenario. They can be categorised in three

main categories, chemically, biologically, and physically mediated methods (*202*). Biological methods were not considered, because of the danger to the CHO cells by the use of viruses (*202*). Lipofection, calcium-phosphate precipitation, and polyfection were considered as part of the chemical methods and electroporation as a physical. The optimal method of transfection should have high efficiency, low level of cell toxicity and be easy to use and reproduce. Thus, of all four methods analysed, electroporation was selected as the preferred choice for transfected DNA material to a mammalian cell line because of its simplicity in implementation, low cytotoxicity for the engineering of mammalian cell lines, and low application cost.

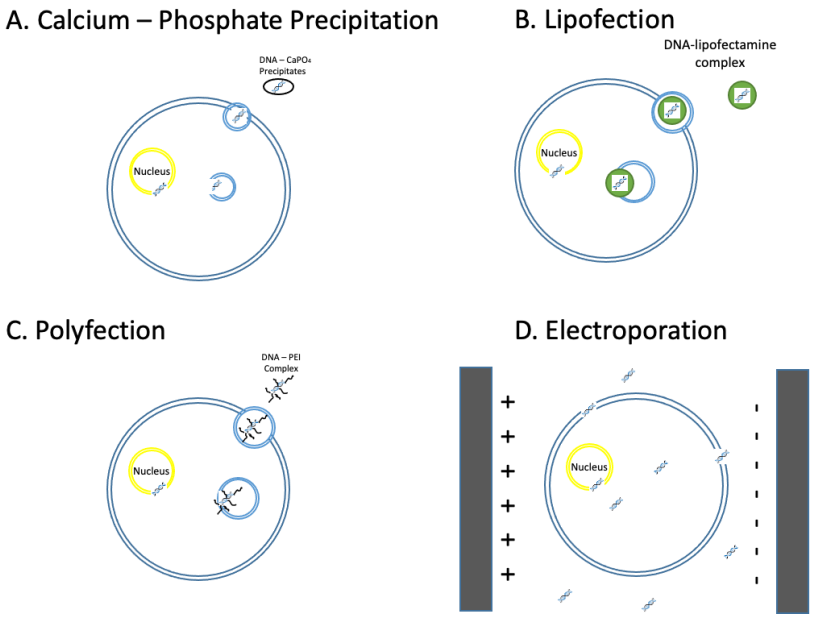


Figure 7.2: Different Methods for DNA delivery in eukaryotic cells. Calcium-Phosphate precipitation (A) takes advantage of a $CaPO_4$ vesicle which carries the DNA in the cell. Similarly, lipofection (B) uses a liposome as a vesicle and polyfection (C) uses a polymer. Electroporation (D), a physical method of transfection makes use of electrical field to break open the cell membrane for the DNA to enter the cell.

Lipofection, or cationic lipid transfection, is one of the most popular chemical ways for introducing foreign DNA into cells. Liposomes trap DNA inside them and form complexes that join with the host cell plasma membrane so that DNA can enter cytosol as seen in figure 7.2A. A variety of different lipids can be the bases for the formation of liposomes such as lipofectin™ or DOTAP. Nowadays lipofectamine is the most commonly used reagent, but the method suffers from being expensive and toxic to certain types of cells (*203*).

Calcium phosphate-mediated transfection (Fig. 7.2B) is a popular technique because the components are readily available and inexpensive. Initially, DNA is mixed with calcium chloride in a

buffered saline/phosphate solution to generate a calcium-phosphate–DNA co-precipitate. The DNA-calcium phosphate precipitates enter the cell though endocytosis. It is effective with many types of cells, but it is prone to variability due to its sensitivity to slight changes in pH, temperature, and buffer salt concentrations. Also, it can be cytotoxic to many types of cell cultures (*202*).

Cationic polymer, one of the oldest transfection methods, is based on the idea of the difference of charge between the DNA vehicle and the cell membrane (Fig. 7.2C). A variety of cationic polymers are used for gene modifications, such as polylysine, polybrene or polyethylenimine. They differ from cationic lipids in that they do not contain a hydrophobic moiety and are completely soluble in water. The nucleic acid-polymer complexes have an overall positive charge, and they interact with the negatively charged host cell membrane, so they adhere, and the foreign DNA is being uptake in the cell though endocytosis. Even though cationic polymers offer stability, reproducible results, and high transfection efficient they also suffer from cytotoxicity, due to them being non-biodegradable because of their high MW (*204*).

Electroporation, on the other hand, is a physical/nonchemical transfection method, utilising electric field pulses to permeabilise the host cell membrane and allow for the DNA to enter the cell as seen in Figure 7.2D. It is a very simple process. Host cells and DNA are suspended in a conductive solution, e.g., culture medium, and an electrical circuit is closed around the mixture. An electrical discharge is discharged through the cell suspension, which allows the DNA to enter the cell through temporary pores. The duration of pulse, usually in milliseconds, and voltage are optimised for each specific cell line. The main advantage of electroporation is its applicability for transient and stable transfection of all types of cells. Furthermore, because electroporation is easy and rapid, it is able to transfect a large number of cells in a short time once optimum electroporation conditions are determined. The advantage of electroporation over chemical transfection reagents is the high transfection efficiency and the low cytotoxicity. The major drawback of electroporation is substantial cell death caused due to high-voltage pulses and only partially successful membrane repair, requiring the use of greater quantities of cells compared to chemical transfection methods. While more modern instrumentation can overcome high cell mortality by distributing the electrical pulse equally among the cells and maintaining a stable pH throughout the electroporation chamber, optimization of pulse and field strength parameters is still required to balance the electroporation efficiency and cell viability (*205*).

7.1.3 Stable Cell Line Development

The next step in creating an engineered strain is to choose whether the DNA modification is transient or not. Stable and transient transfection differ in their long-term effects on the cell. In a stable cell line, the plasmid DNA successfully integrates into the cellular genome and will be passed on to future generations of the cell or the plasmid is forced to stay in the cell because of selection pressure. However,

in transient transfection, the transfected material enters the cell but does not get integrated into the cellular genome, and the cells only express transfected DNA for a short period of time and do not pass it on to their daughter cells. Both stable and transiently transfected cell lines can undergo a 14-day fed-batch culture (*206*), however, it is difficult to perform sequential genetic modification in transiently transfected cells. The first engineered cell line requires a gene knockout and a gene overexpression. Even though the second cell line requires only one knockout, the cell line was engineered using stable selection for potential future regulations on the same cell line. For that reason, the two engineered cell lines undergo a selection process and are considered stable. The main step in creating stable cell lines is the appropriate selection process. After the process of transfection, the resulting pool of cells contains transfected and non-transfected cells. The process of choosing only the transfected cells is known as cell selection and is used for the engineering a stable cell line. There are two main ways for selecting transfected cells, antibiotic resistance and fluorescence, as seen in figure 7.3.

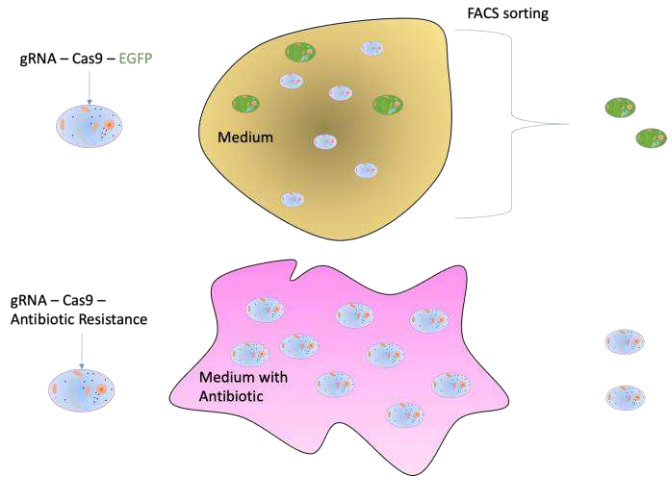


Figure 7.3: The two major ways for selecting cells after DNA transfection. In fluorescence selection (top), the EGFP positive cells are selected by a fluorescence microscope that can detect and then isolate them from the generic cell pool. In antibiotic selection (bottom), the plasmid carries an antibiotic resistance gene. When the cell pool is cultured in the presence of the antibiotic only the transfected cells survive.

For the use of antibiotic selection (Fig. 7.3A), the plasmid DNA needs to also contain an antibiotic resistance gene, that is a gene which prevents the death of the cell from a specific antibiotic. The type of antibiotic used for mammalian selection can be neomycin, hygromycin, kanamycin, etc. To select the transfected cells, the whole cell population is cultured in a pool with antibiotic, which kills all the cells that do not express the resistance marker. Depending on the transfection efficiency, only the cells that have accepted the foreign DNA will produce the antibiotic resistance gene and survive. It is also important to measure the optimal antibiotic concentration in the culture media that kills the cells,

because the resistance genes cannot protect cells from large amounts of the antibiotic. The resulting cell line will be stable and maintain its genetic modifications, as long as there is addition of the antibiotic forcing the expression of the transfected DNA.

For the use of fluoresce (Fig. 7.3B), the plasmids together with the required DNA they carry a protein encoding a fluorescence marker of choice like EGFP or mCherry. The cells emit a light on a specific wavelength when they are exposed to light, which can be later detected using a Fluorescence-Activated Cell Sorting (FACS). After transfection, some cells contain the fluorescence marker and some do not, so FACS choses only the cells with the marker. Later, the cells can be cultured in single colonies or in a pool. For higher accuracy and more homologous pools, the technique of single cell sorting can be applied, where each cell that emits the wavelength of choice is selected and cultured individually. The process is time consuming, but it results on stable cell lines as the single cells will contain the transfected plasmid.

7.2 CHO-AA

The first cell line that was engineered from the original mAb producing CHO cell line was CHO-AA. It modified the expressions of two genes in the CHO genome, the disruption of the *Aldh4a1* expression and the overexpression of the *Aldh18a1* gene. FBA optimisation predicts that when the two modifications are applied on the same cell line and it is then cultured under fed-batch conditions using FeedC, there would be an inhibition on biomass production and an increase in q_{mAb}.

7.2.1 Knock-Out of *Aldh4a1*

The gene *Aldh4a1* expresses the mitochondrial ALDH4A1 protein responsible for the conversion of proline to glutamate. The CHO cell line that was used as parent for the genetic engineering experiments is a descendant from the CHO-K1. The latest version of the CHO-K1 genome (*207*) contains the *Aldh4a1* gene, so the CHO cell line will probably express the gene making the disruption of its expression possible. The technique of CRISPR/Cas9 was used to knock out *Aldh4a1*, due to its simple implementation, design, and efficient results of the method. The goal of the CRISPR gene editing system is to cut a vital piece of the genome, so the DNA repair mechanisms would create an error in the nucleotide sequence resulting to either faulty or no translation at all or for the translated protein to not be functional. *Aldh4a1* is made up of 16 exons with the total sequence exceeding the 3000bp, which raises the question of which parts are the most sensitive and important to perform a mutation on and disable the expression of a working protein. Some exons might be more prone to error from a DNA repair, or some exons known as active sites might be too important for the functionality of the gene by expressing a catalytic domain of the enzyme. An obvious choice is the nicking of an active site that can

disable the function of the protein after it is transcribed. However, if there are no known active sites, the nick in an early exon is also a viable option (*208-211*), to stop the gene transcription in an early point. The gene *Aldh4a1* is annotated by UNIPROT, which predicts two directly involved in catalysis active sites. A mutation in either of these two active sites can rendering the whole protein dysfunctional by not being able to catalyse, in the case of ALDH4A1 the conversion of proline to glutamate. However, the two active sites are predicted by UNIPROT and have not been validated experimentally, so it was not attempted to perform a mutation on either of these two sites. It was decided as a safer choice to try and perform a mutation on the early exons of the gene. Another reason why the target of CRISPR was not chosen to be an active site is that it is more difficult to evaluate the functionality of an expressed protein that its overall existence. The existence of a protein can be determined using a simple western blot analysis, but its functionality requires an activity assay such as metabolomics analysis in order to observe any changes in the metabolome. For example, if there would be an increased uptake of glutamine, it could mean that the ALDH4A1 could not perform its job of converting proline to glutamine. Four different plasmids A4Ni1, A4Nu1, A4Nu2 and A4Nu3, were designed as described in the section **5.3.2** of the experimental materials and methods, with the goal of disrupting the normal function of the ALDH4A1 protein through a knockout of the *Aldh4a1* gene.

Two different CRISPR experimental setups were initially tested to determine the most optimal way for the knockout of the gene of interest. The Cas9 nickase, Cas9 (D10A) and the use of Cas9 nuclease, SpCas9, together with a pair of gRNAs were used to apply a mutation on the early exons of the *Aldh4a1* gene. The Cas9 (D10A) protein creates a single strand break (SSB) on the PAM sequence supplied by the gRNA. Two different gRNA sequences would cause a DSB between the two gRNA sequences. The gRNAs of the plasmid A4Ni1 have 50bp distance from each other and target the second exon of *Aldh4a1*. The two RNA sequences should not be too far or too close to each other to ensure maximum efficiency and the 50bp is considered an efficient length (*200*). The EGFP fluorescence protein was used as a selection marker to isolate the cells that contain the plasmid from the population of un-transfected cells though FACS sorting. For the sorting the FACS machine in the Department of Life Sciences, Imperial College London, was used. The three components were transfected to the CHO cells using the plasmid A4Ni1. A similar strategy made use of the SpCas9 nuclease protein. It creates a DSB just before the PAM sequence of the given gRNA. If multiple gRNAs are supplied, the Cas9 protein would nick the genome at multiple points multiplying the possibility of a positive genetic engineering result. The main advantage that Cas9(D10A) offers against SpCas9 is the limitation of off-target events. Two plasmids, A4Nu1 and A4Nu2, each expressing SpCas9, and different gRNA sequences were used to nick the genome in two different places and create a 350bp gap. The two gRNA sequences target the first and second exon of the gene. The transfection of two different plasmids gives rise to the problem of cell selection. The two plasmids must contain two different selection markers in order to select the cells that carry both plasmids, also known as double positive. The plasmid A4Nu1 contained the EGFP protein and the plasmid A4Nu2 the mCherry fluorescence protein, which both were

co-transfected to the cells. FACS was used to detect the cells that emit both wavelengths and isolate them acquiring the cell pool, CHO-CoT, with cells that contain both plasmids. For the sorting the FACS machine in the Department of Life Sciences, Imperial College London, was used. Similarly, CHO-N is the cell pool which were transfected by the A4Ni1 plasmid. However, due to a malfunction in the shaker's rotation speed, both CHO-N and CHO-CoT cultures formed clumps of cells the day after transfection. CHO-N was contaminated due to contact with the top of the flask and none of the cells survived. CHO-CoT survived the process, but it was not possible to continue a stable culture even after the addition of anti-clumping agent (1:100) for 10 days. After 10 days, the CHO-CoT cells were harvested and lysed to see the effect the co-transfection had on the genome. Specifically, the control CHO cells and the engineered CHO-CoT, were lysed with the protocol presented in the section **6.4.1**, and the proteins were separated and detected using western blot according to the protocol from the section **6.4.2** of the experimental materials and methods. The resulting blots can be seen in figure 7.4.

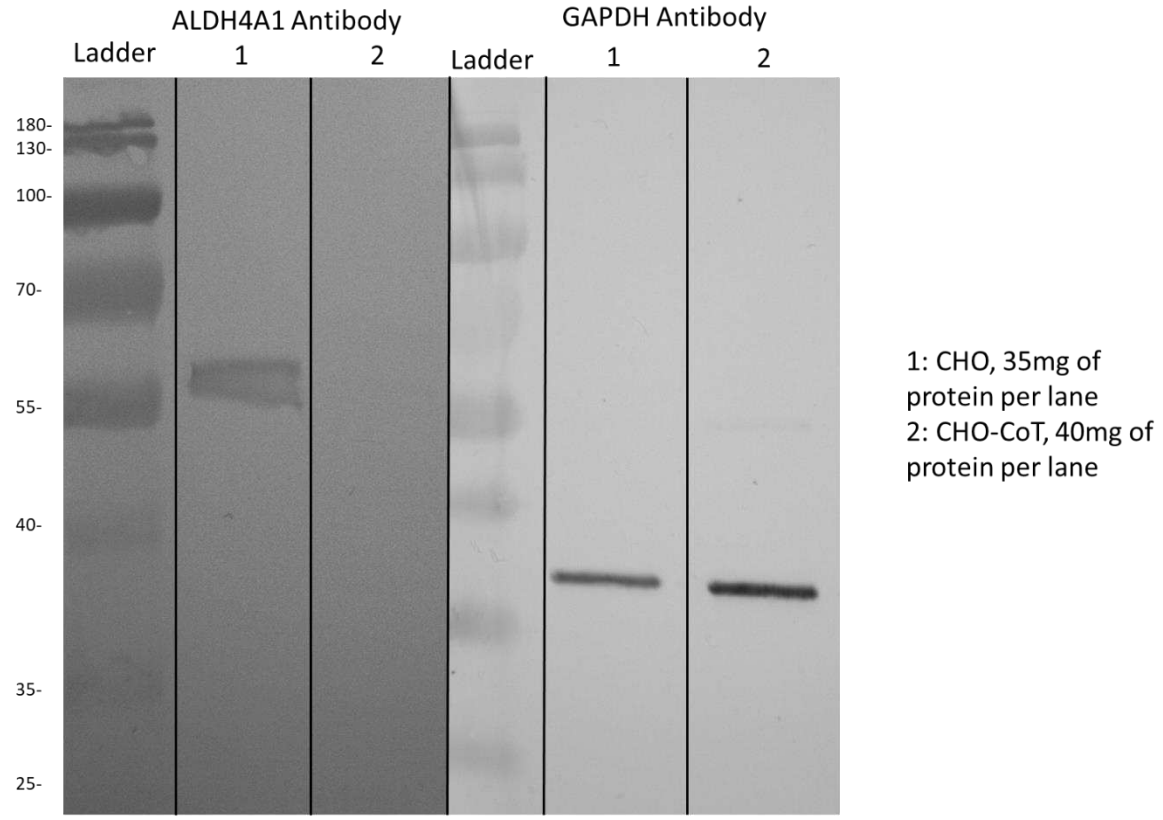


Figure 7.4: Western blot analysis of the original CHO and CHO-CoT cell lines. The right blot shows the quantity of the GAPDH, 36kDa, (ab181602, abcam) protein in the two cell lysates, which acts as a protein loading control for the ALDH4A1 protein, 62kDa, (ab185208, abcam). In the left blot, it can clearly be seen that there is significantly reduced quantity of the ALDH4A1 protein in the CHO-CoT samples compared to that in the CHO cells.

The experiment appeared to be successful as the engineered cell line CHO-CoT resulted in an almost complete knockout of the ALDH4A1 protein. The two gRNA sequences transfected seemed to cause a deleterious INDEL, which is the reason the experimental was replicated and tested again with the same exact conditions. The cell line CHO-N was not replicated due to the positive result of the CHO-CoT. The original CHO cell line was co-transfected again with the two plasmids, and after the FACS selection procedure the CHO-CoT was tested again regarding the expression of the ALDH4A1 protein. For the sorting the FACS machine in the Department of Life Sciences, Imperial College London, was used. Unfortunately, the second time the western blot analysis showed a knockdown of 60% compared to the original CHO-cells, as seen in 7.5.

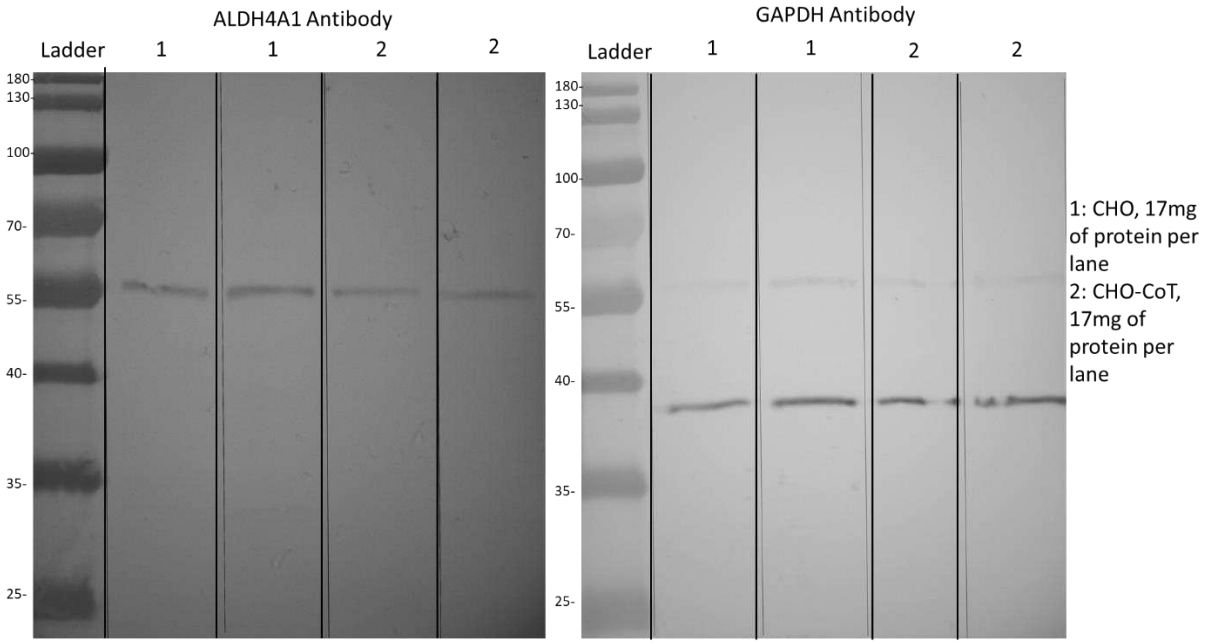


Figure 7.5: Western blot analysis of the original CHO and CHO-CoT cell lines. The right blot shows the quantity of the GAPDH, 36kDa, (ab181602, abcam) protein in the two cell lysates acting as a protein loading control for the ALDH4A1 protein, 62kDa, (ab185208, abcam). In the left blot, the amount of ALDH4A1 protein in the CHO-CoT samples appears to be about 60% less than in the sample of the CHO cells.

The engineered cell line CHO-CoT displayed a downregulation of 60% on the expression of *Aldh4a1* gene and not a complete knockout, which was the initial goal of this genetic engineering experiment. The downregulation of the expression can also cause an abundance of proline in the cell if it is combined with the correct amount of *Aldh18a1* upregulation, as analysed in the previous chapter. It would be necessary to test different degrees of overexpression and keep the ones that display reduced growth rates. However, the downregulation gave rise to a new problem, the heterogeneity of the population. A downregulation in the gene expression indicates that the cell population in the pool consists of original or unmodified CHO cells which express ALDH4A1 at normal levels, and the

knockout cell line CHO-CoT. The two cell lines express ALDH4A1 at in about 60% lower quantity compared to a 100% CHO cell pool. If the culture remains alive as is, the most likely scenario is that the original CHO cells will prevail the engineered CHO-CoT and the cell pool will become 100% untransfected CHO cells due to their advantage in converting proline to glutamate through the ALDH4A1 enzyme. Unfortunately, it is not possible to select the CHO-CoT cells from the cell pool, because they have not kept the A4Nu1 and A4Nu2 plasmids which express the selection markers used for FACS sorting. The plasmids also express Cas9, a protein with a huge resource burden, which negatively affects the growth of the cell and is a resource burden. Therefore, the downregulation degree of *Aldh4a1* expression will probably slowly decrease as the population of the original unmodified CHO cells increase in the cell pool. One way to combat this issue is to transfect the cell pool again with the two plasmids, A4Nu1 and A4Nu2, and hope to achieve better results; but it is not a certain scenario and was not attempted. Selecting single cells, can create cell pools with only the genetically engineered CHO-CoT cells. Ideally, one of these pools will provide a stable culture with a complete knockout of the *Aldh4a1* gene. However, single cell cloning is a very lengthy process, which is the reason it was not considered herein. A safer and easier option is the use of antibiotic as a selection method, which would simulate a stable cell line as long as there is supply of the antibiotic. Antibiotic selection would maintain a population of genetically engineered cells which would express the antibiotic resistance gene. Therefore, there would be a population of genetically modified and unmodified cells, resulting in at least a downregulation of the gene's expression. The plasmid A4Nu3 contains the required components for genetic engineering, the Cas9 protein and the previously tested gRNA sequences, with an antibiotic selection marker. The two different gRNA sequences were cloned in the same plasmid, instead of transfecting them separately which would require the use of two different antibiotics as selection markers, causing a significant antibiotic pressure on the culture.

The A4Nu3 plasmid was transfected to the original CHO cell line to create the CHO-A cell line. It carries the hygromycin resistance gene, acting as a selection marker, the SpCas9 protein and the same two gRNA used by the plasmids A4Nu1 and A4Nu2 to disrupt the normal expression of the *Aldh4a1* gene. The addition of the antibiotic hygromycin naturally kills all original CHO cells, except the ones that have been transfected with the A4Nu3 plasmid, which contains the resistance gene to the antibiotic making them resistant to its effects. An important question it raises is when the optimal time point to add the antibiotic after the transfection of cells. The cells need some time to express the antibiotic resistance gene, but if the wait time is too long then the CHO cells will start to suppress the engineered CHO-A cell population. For this reason, three new conditions of the CHO-A cell line were tested to determine the optimal timing of antibiotic addition to the culture (day 2, day 3, day 4) after the day of transfection (day 0). MSX was also supplied to the cell cultures to maintain increased production of IgG. The combination of the two antibiotics could result in slower recovery rates from the kill curve or possible different levels of *Aldh4a1* knockout. Therefore, six conditions were tested in total as presented in table 7.1.

Table 7.1: The six different conditions used for developing a selection protocol for the original CHO and CHO-A cell line based on the hygromycin kill-curve.

Cell line and condition	*Description*
CHO-A: Day2– Hygromycin (D2AH)	Addition of 400mg/mL of culture hygromycin on day 2 after transfection, for the selection of the CHO-A cell line.
CHO-A: Day2– Hygromycin – MSX (D2AHM)	Addition of 400mg/mL of culture hygromycin and 0.1%v/v 25mM MSX on day 2 after transfection, for the selection of the CHO cell line.
CHO: Day2– Hygromycin (D2CH))	Addition of 400mg/mL of culture hygromycin on day 2 after transfection, for the selection of the CHO-A cell line.
CHO: Day2– Hygromycin – MSX (D2CHM)	Addition of 400mg/mL of culture hygromycin and 0.1%v/v 25mM MSX on day 2 after transfection, for the selection of the CHO cell line.

The conditions for the Day 3 and Day 4 selection follow a similar pattern. In total, twelve conditions were designed and six of them were the CHO control cell line. All conditions were cultured in 6-well plates and their viability after the addition of hygromycin is presented in figure 7.6.

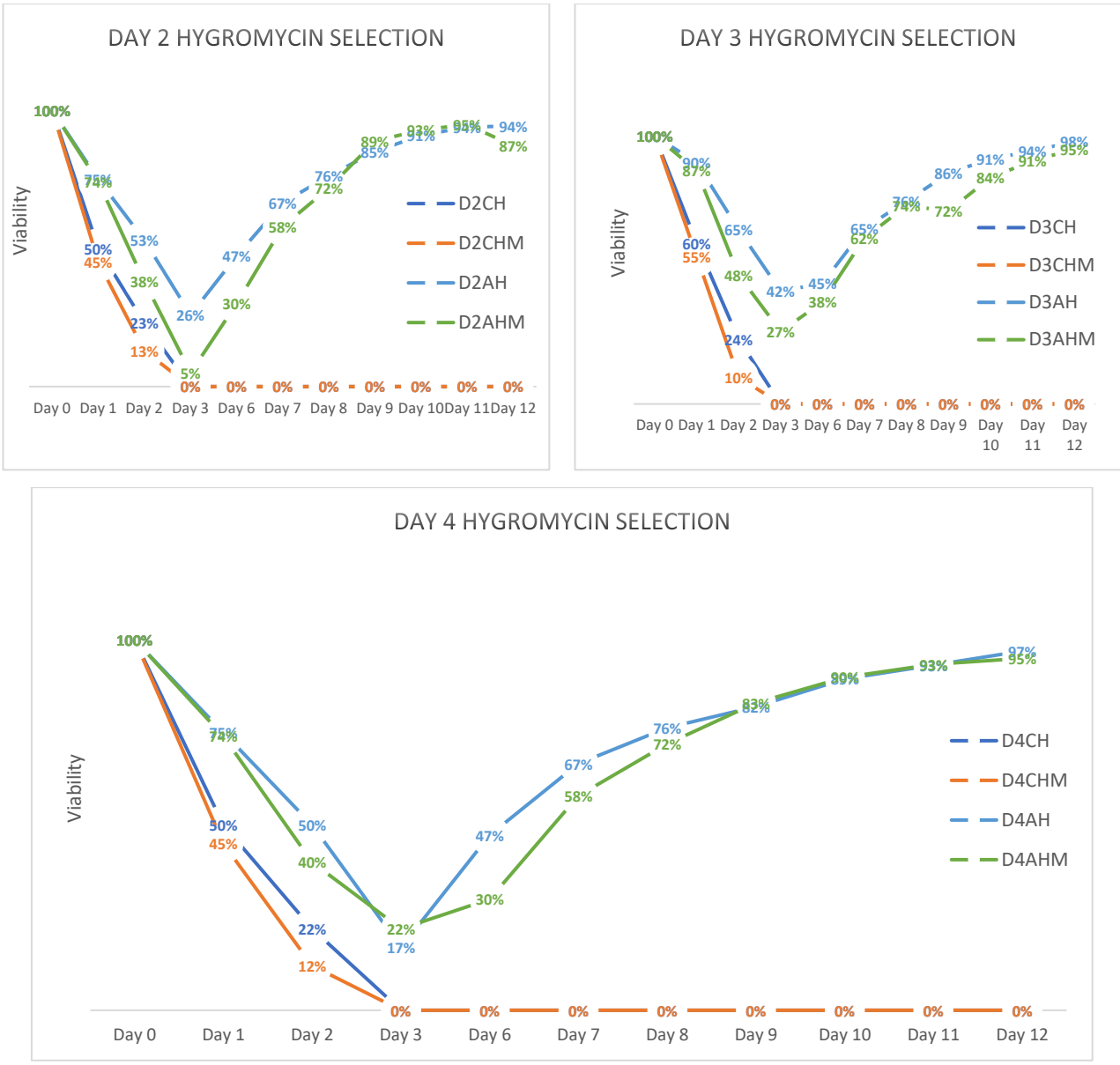


Figure 7.6: Kill curve by the addition of the antibiotic hygromycin. The different days of the addition of antibiotic resulted in different viabilities of the transfected cells. All three cell lines recovered to above 90% viability levels. The control line died in all three cases, meaning that all transfected CHO cells were probably dead in the transfected pools.

All six conditions of the CHO-A engineered cell line survived the kill curve and recovered to reach viabilities greater than 90%. In all cases the control CHO cell lines reached 0% viability. After the antibiotic selection in each of the conditions (the original CHO cells died), the concentration of hygromycin was halved to help the recovery process and achieve high VCD required for further analysis. Due to lack of 125mL Erlenmeyer flasks, the cell lines were maintained in 6-well plates in CD-CHO media with 200mg/mL hygromycin and MSX according to the cell line conditions, according to the subculturing protocol described in section **6.1**. To observe the result of the DNA transfection, a

western blot analysis was performed. All conditions of the CHO-A cell line were lysed, and their protein content was run through an SDS-PAGE, according to the protocol of section **6.4.2** of the experimental materials and methods. The western blot can be seen in figure 7.7.

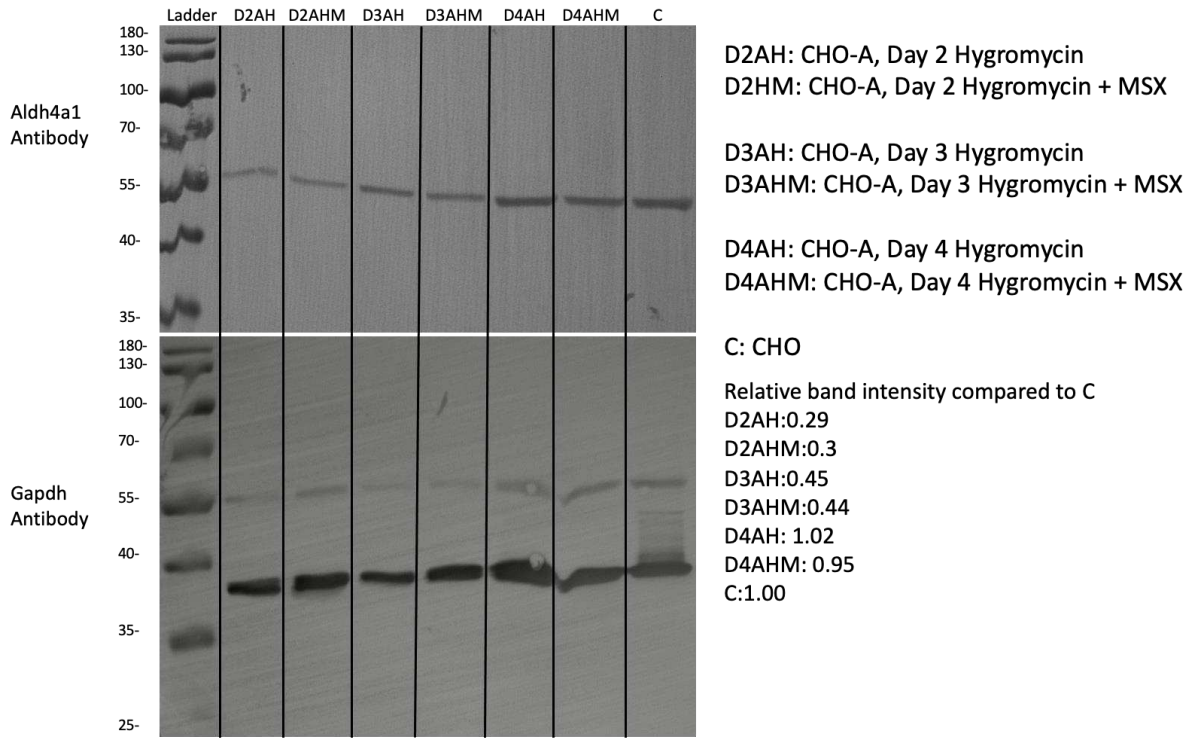


Figure 7.7: Western blot analysis of the CHO-A cell lysates for each of the six conditions, together with the control CHO cell line. The protein load for each lane was 50μg of protein lysate. The molecular weight of GAPDH is around 36kDa and of ALDH4A1 around 62kDa. The top blot shows the quantification of the ALDH4A1 protein (ab185208, abcam) in the lysates, and the bottom blot shows the GAPDH protein (ab181602, abcam) in the samples, which acts as a control for the total protein loading. The first lane is the ladder of the molecular weights. The eighth lane is the CHO cell lysate, acting as a control to observe potential changes in the expression of the ALDH4A1 protein. Lanes 2-7 contain the lysates of the CHO-A cell line. Each of the six conditions exhibit a reduced level of ALDH4A1 compared with the control.

There are two main conclusions from the Western blot analysis of the lysates. First, there is a significant difference in the efficiency of the plasmid in performing a successful genetic modification depending on the day of antibiotic addition. Day 4 exhibited the smallest amount of downregulation, proving that it was too late to add the selection marker. D2AH and D2AHM displayed the largest reduction, 70%, of ALDH4A1 protein content compared to the original CHO cell line. It is possible that the reduced antibiotic concentration permitted the growth of the original CHO cells that have not been modified by CRISPR. A possible way to achieve a complete knockout could have been to not reduce

the antibiotic concentration on day 3 but maintain it in high concentrations. The difference of ALDH4A1 expression from lysates that had and had not been treated with MSX is the same, proving that its addition to the cultures did not affect the efficiency of the CRIRPR editing system. Based on the above Western blot result, the D2AHM was selected for future experiments, while the rest of the cell lines were banked in liquid nitrogen as a backup. Both D2AHM and D2AH reached high enough viabilities (Fig. 7.6) and it was decided to keep the D2AHM CHO-A cell line and disregard the cell line without MSX due to the possibility of lower IgG production. The CHO-A cell line was subcultured under D2AHM conditions in a 125mL Erlenmeyer flask in CD-CHO, with 200mg/mL of media hygromycin and 0.1% v/v MSX, to prepare it for the next genetic engineering of the proline abundance strategy.

7.2.2 Overexpression of *Aldh18a1*

In the previous section, the development of the CHO-A cell line from the original CHO cell line was described. It displayed around 70% reduced ALDH4A1 protein content compared to the original. The heterogeneity of the population in the expression of the protein, means a different degree of expression by the different cells in the CHO-A cell pool. The degree of downregulation should not reduce significantly as long as there is supply of hygromycin. The antibiotic forces the plasmid to express the Cas9 protein and the gRNA required for the disruption of the *Alhd4a1* gene. The only way to stop the supply of hygromycin would be if there was a complete knockout. The cells that do not express a functional ALDH4A1 protein cannot regain its expression, since the DNA sequence cannot return to its original sequence (*212-214*). Therefore, it is possible to continue with the engineering of the cell line by increasing the expression of the gene *Aldh18a1*, while supplying hygromycin.

The gene Aldehyde Dehydrogenase 18 Family, Member A1 (*Aldh18a1* or *P5cs*) expresses the mitochondrial protein ALDH18A1 responsible for converting glutamate to proline. An increase in the expression of the gene can theoretically increase the levels of the enzyme in the mitochondrion and thus the conversion of glutamate to proline, creating an abundance of proline in the cell as explained the section **5.2.1** of the sensitivity analysis of the optimisation results. The gene exists in the CHO-K1 genome, and it has two transcripts in CHO-K1, with protein size around 87kDa (*215*). The overexpression of *Aldh18a1* can happen using the ORF plasmid, A18ORF, which contains the mRNA sequence of the ALDH18A1 protein that can then be transcribed to the protein. A18ORF contains the mRNA sequence for the mouse ALDH18A1 protein which shares 98% similarity with the mRNA sequence with the CHO-K1 protein.

The main difference between an overexpression and a knockout plasmid is that the former needs to remain in the cell while the knockout plasmid can exit the cell if the knockout succeeds. Specifically, a knockout plasmid, like the one used for the knockout of *Aldh4a1*, can exit the cell after it performs the mutation in the genome since the mutation cannot be reversed through DNA repair mechanisms.

Only if there is a knockdown, like in the case of CHO-A, does the plasmid need to remain in the cell and apply pressure on the original cells that have not been successfully mutated by CRISPR. An overexpression ORF plasmid, however, does not change the genome but it supplies the mRNA sequence of the gene of interest, which is then translated to the protein of interest. If there is no selection pressure, the cell has no reason to keep the plasmid. For this reason, the selection marker cannot be a fluorescence marker, but an antibiotic resistance gene to pressure the plasmid in remaining in the cells or some type of metabolic control. The A18ORF plasmid contains the neomycin resistance gene, which similarly to before, enables cells to survive the neomycin antibiotic. The transfection of the CHO-A cell line with the A18ORF plasmid created the cell line CHO-AA. Similarly with the kill-curve of the A4Nu3 plasmid, neomycin was added on days 2, 3 and 4 after the transfection, combined with the addition of MSX. Six different conditions were tested for the CHO-AA cell line, with similar abbreviations as in the knockout of ALDH4A1, as seen in the previous table 7.1. The only difference is the use of 400mg/mL of neomycin, instead of 400mg/mL of hygromycin per mL of culture. Throughout the process of selection, the media was treated with 200mg/mL of hygromycin per mL of culture to maintain the downregulation of the ALDH4A1 protein. All conditions of the engineered CHO-AA underwent through the neomycin kill curve. The overexpressed cell line, similarly to the previous antibiotic selection, recovered to above 90% viability levels. The original CHO cells died in all three cases. The resulting CHO-AA must have contained the A18ORF plasmid, to express the neomycin resistance gene. The mRNA of the gene *Aldh18a1* is expressed by A18ORF, so the remaining cells should overexpress the protein. The concentration of the antibiotic was halved after successful selection, similarly to before, to aid the recovery process. The cell line was maintained in 6-well plates, due to lack of 125mL Erlenmeyer flasks in CD-CHO media, with the addition of 200mg/ml of hygromycin and 200mg/mL of neomycin to maintain the plasmids A4Nu3 and A18ORF in the cell. The next step was to perform a western blot analysis and determine whether an overexpression of *Aldh18a1* took place, and any potential differences between the different conditions of the CHO-AA cell maintenance. The *Aldh4a1* knockout experiment showed no change in the expression of ALDH4A1 between MSX treated and non-treated cells (Fig. 7.7), so it was decided to maintain only the CHO-AA D2NM, D3NM and D4NM. Western blot analysis (Fig. 7.8 and Fig. 7.9) was used to detect the expression levels of the ALDH18A1 and ALDH4A1 proteins and determine any potential changes in the proteins' expression between the engineered cell line CHO-AA and its original. The detection of ALDH4A1 would show whether the knockdown cell line CHO-AA, was still expressing ALDH4A1 in reduced levels compared to the original CHO cell. The conditions D3AN and D4AN were not tested for the expression of ALDH4A1, as all of them were treated with the exact same amount of hygromycin and probably would express the protein in the same quantity. The concentration of hygromycin was halved after the completion of the cell selection, so the degree of knockdown could have decreased between the knockout of *Aldh4a1* (Fig. 7.7) and the overexpression of *Aldh18a1* (Fig. 7.8). Both western blots used the original CHO cell line,

and not the knockdown CHO-A cell line as a control, because it was sought to evaluate by how much ALDH18A1 increased compared to the original CHO cell line.

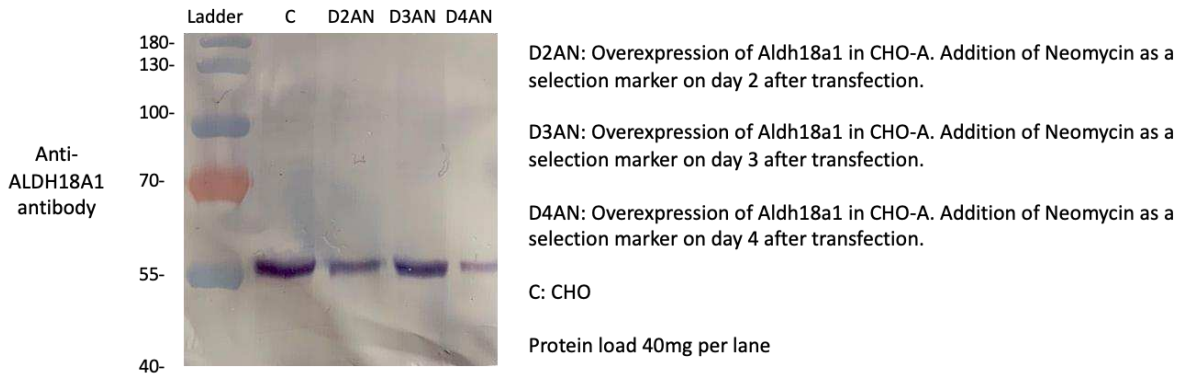


Figure 7.8: Western blot analysis of the CHO-AA lysates, compared to the control, original CHO cells. The antibody used is the ab223713 from abcam. The antibody detects the protein ALDH18A1 and has been verified in human, rat, and mouse tissue cells. However, it was not able to detect the protein in none of the samples, as the molecular weight of ALDH18A1 is around 87 kDa, and the band is at 55 kDa, much smaller than the required weight.

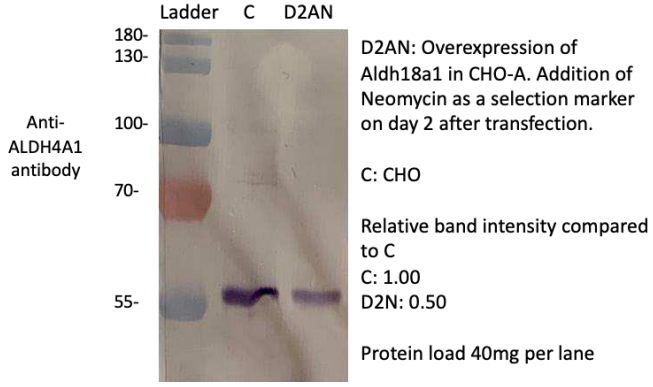


Figure 7.9: Western blot analysis of the CHO-AA lysates, compared to the control, CHO lysate. The antibody used is the ab185208 from abcam, and it detects the ALDH4A1 protein, at 61kDa.

The intensity of D2NM's band is almost half that of the original CHO as shown in figure 7.9. The degree of ALDH4A1 knockdown was reduced, most likely because of the reduction in the concentration of antibiotic, to aid the growth rate. However, even though the knockdown of *Aldh4a1* was still there, the western blot (Fig. 7.8) could not detect the ALDH18A1 protein. It was assumed the problem was the antibody because a band at 85kDA was expected in the control project. A new antibody, ab206682 from Abcam, was ordered which had 100% alignment between the sequences of the antibody and the protein sequence form CHO-K1 and has been verified in human and mouse tissue cells. The second western blot is presented in figure 7.10.

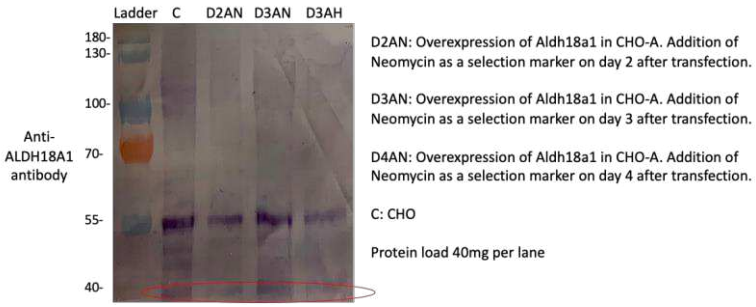


Figure 7.10: Western blot analysis of the CHO-AA lysates, which are compared to the control CHO cell. The antibody used is the ab206682 from abcam.

The western blot was not able to detect the protein in none of the samples, as the molecular weight of ALDH18A1 is around 87kDa, and the band (Fig. 7.10) is again at 55kDa. It was hypothesised that the ALDH18A1 got truncated, as in both cases no bands were appeared at the molecular weight of ALDH18A1 and in the second time (Fig. 7.10) a second band appeared at 40kDa. The next step was to try a non-reducing western blot to try and detect the intact enzyme. A non-reducing western blot follows the same protocol as western blot but does not include DTT in the SDS loading dye. DTT is a reducing agent that breaks down the disulphide bonds, so by not including it in the preparation of the samples for western blot, the proteins will likely not break into its subunits. The non-reducing western blot of the samples is presented in figure 7.11.

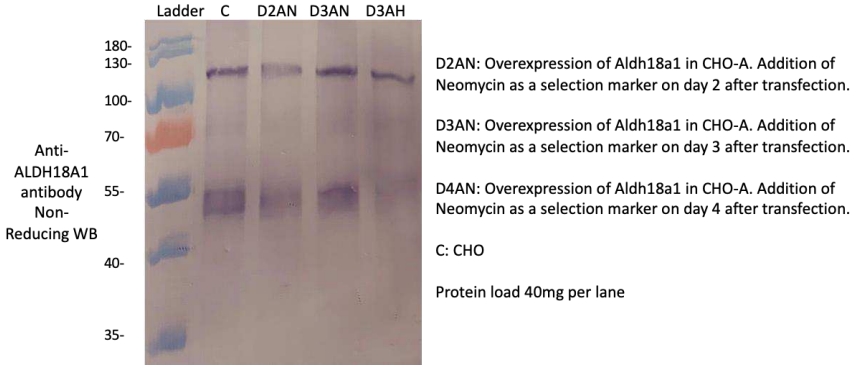


Figure 7.11: Non-reducing western blot analysis of the CHO-AA lysates, compared to the control CHO cell lysate. The antibody used is the ab206682 from abcam.

The band is much higher than before, at 110kDa compared to 55kDa, meaning the non-reducing western blot was successful in not breaking the disulphide bonds of the protein. However, the protein detected in figure 7.11 is not the ALDH18A1 which has a mass of 87 kDa. Even glycosylated the difference of 20 kDa is considered too high. The inability to detect the ALHD18A1 protein means two things. The overexpression experiment either did not work and the protein is not expressed in the original CHO cell line. There have been studies on CHO-S and CHO-K1 cells in which the ALDH18A1 protein was not detected (*216, 217*). A qPCR would detect the mRNA levels, but due to time restrictions

it was not possible to be conducted right now and check the mRNA levels of the ALDH18A1 protein in the original and overexpressed CHO cell line. However, when the iCHO1766 was constructed, there was evidence that the gene is transcribed and there is mRNA of the protein, so it was assumed that the protein would exist in the cell's proteome. The ALDH18A1 enzyme synthesises proline from glutamate and has potentially no reason to be active since there was enough proline in the media for CHO cells to uptake. A way to bypass this issue would be to use a proline deficient culture medium, which would force the ALDH18A1 enzyme to be synthesised. An increase of the antibiotic concentration will probably have no effect since the problem is likely the translation of the mRNA. The use of the ORF sequence of *Aldh18a1* gene of the CHO cell can also be an option. The ORF used for the transfection was from the mouse genome. However, the two genes are 98% similar, so the use of the CHO mRNA would probably not affect the outcome. Another option is the transfection of the ALDH18A1 protein as a whole, known as protein transduction, which would avoid the problem of the untranslated gene. Unfortunately, due to time limitations, the overexpression experiment could not be attempted again.

7.3 CHO-B

The second engineered CHO cell line includes the disruption of the *Bcat1* gene expression in combination with inhibition of growth rate in fed-batch culture. The knockout of the gene utilised the CRISPR system, similar with the knockout of the *Aldh4a1* gene, with the use of hygromycin as a selection marker. The inhibition of the cell's growth rate can happen while performing a culture in mild-hypothermia culture conditions, around 32.5°C, which inhibits the metabolic processes of the cell. After the successful knockdown of the gene, the engineered cell line was cultured under mild hypothermic conditions to inhibit the production of biomass. The strain optimisation results suggested modifications of the natural expression of three genes to inhibit the growth rate. Genetic engineering to inhibit the growth rate might be dangerous because the surviving cells after the transfection and selection stage will be on low VCD and their recovery might be too difficult. Mild hypothermia during the fed-batch culture is a controlled way to reduce the growth rate and simulate the results of the optimisation.

7.3.1 Knock-out of *Bcat1*

The *Bcat1* gene has two transcript variants in the CHO-K1 genome and expresses the cytosolic protein BCAT1 responsible for the conversion of isoleucine, leucine, and valine to glutamate. The FBA optimisation predicts that the knockout of the protein, in combination with reduced growth rate, would create an abundance of leucine in the cell, as explained in the section **4.2.2** of the sensitivity analysis chapter. The original CHO cell line is expected to express the *Bcat1* gene as it is CHO-K1-derived. The knockout of the gene used the same experimental setup as the knockout of *Aldh4a1* with the difference

in the design of the gRNA. Specifically, the CRISPR system of the SpCas9 protein and a gRNA sequence, were transfected into the CHO cell using the plasmid BNu1 as described with the transfection protocol in section **6.3.1** of the experimental materials and methods chapter. The plasmid also carried an antibiotic selection marker to isolate only the transfected cells. The gRNA used was the same as in the publication by Andersen et al. to knock out *Bcat1* in CHO-S (*191*). The sequence targets the fifth exon in the transcript X2 and the sixth exon in the transcript X1 of the gene and has the smallest number of mismatches with only one exact match in the CHO-K1 genome sequence. It is designed to cause a DSB in the *Bcat1* gene of the CHO cell, creating the CHO-B cell line. The antibiotic used to select the transfected cells and keep the plasmid in the cells was hygromycin. For the same reason as explained in the case of the *Aldh4a1* knock out, three different kill curves and subsequently CHO-B conditions were created similarly as in the knockout of the ALDH4A1 protein in section **6.2.1**. The names of the conditions follow a similar pattern as before. In each case, all the control cells completely died at the third day after the addition of the antibiotic, and all conditions of the engineered CHO-B cells recovered in high viabilities.

The corresponding control CHO cell lines reached 0% viability on day 3 of hygromycin addition. The remaining cells, CHO-B, must contain the BNu1 plasmid which expresses the hygromycin resistance gene needed for their survival, as well as the components of SpCas9 and the gRNA sequence used for *Bcat1* knockout. It was decided to reduce the hygromycin concentrations of the culture to 200mg/mL, similarly to the knockout of *Aldh4a1*, so the cell lines could recover to high VCDs. The six different conditions of CHO-B might express the *Bcat1* gene to a different degree. To determine which conditions, produce the lowest amount of the BCAT1 enzyme a western blot analysis is necessary. However, due to the culturing of the cells being done exclusively in 6-well plates, the value for VCD of cultures grown under conditions D3BHM and D4BHM was not sufficient to conduct western blot analysis. Therefore, only four out of the six conditions were analysed. From the previous knockout experiment, there was no visible difference whether cells were treated with MSX, so it was assumed that conditions D3BHM and D4BHM would yield the same results as D3BH and D4BH, respectively. The protocol for cell lysis and western blot analysis is explained in section **6.4.1** and **6.4.2** of the experimental materials and methods, and the results are presented in figure 7.12.

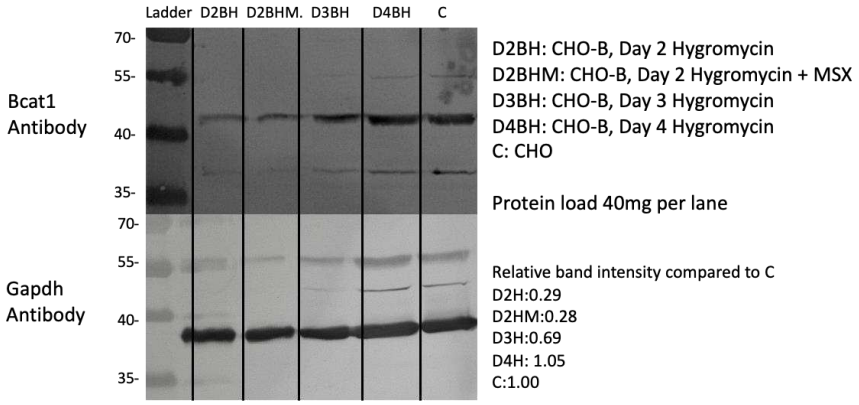


Figure 7.12: Western blot to identify the degree of BCAT1, 43kDa, expression in each of the CHO-B conditions. The first lane is the ladder of the molecular weights. Each of the next four lanes contain the protein lysates of the samples D2H, D2HM, D3H and D4H which are compared to the control, C. The protein load for each lane was 50µg. The molecular weight of BCAT1 (ab232700, abcam) is around 43kDa and of GAPDH 38kDa (ab181602, abcam).

The western blot analysis showed different degrees of downregulation of BCAT1 expression for the four conditions. The addition of antibiotic on day 3 after transfection showed a 30% downregulation of the BCAT1 expression compared to the expression of the original CHO cell line. The addition on day 4 seemed to have no visible effect on the expression of BCAT1. The Day 2 sample had the largest degree of downregulation, 70% lower BCAT1 quantity in the sample compared to the original cell line. The D2BH and D2BHM conditions of the CHO-B cell line resulted in the highest degree of downregulation of the *Bcat1* gene compared to the original CHO cell line. The treatment of MSX seemed to have no effect on the expression of BCAT1 protein in samples D2BH and D2BHM. As mentioned above, the treatment of MSX had no effect in the knockout of ALDH4A1, therefore it was considered as the conditions D3BHM and D4BHM would yield the same results as the D3BH and D4BH respectively. D2BHM was treated with MSX which selects cells with increased copies of the glutamine synthase and IgG antibody, making it a more suitable candidate for a study on the specific productivity of IgG, compared to D2BH. The cell line was transferred from 6-well plates to 125mL Erlenmeyer flask, where it was kept for a week until it reached a suitable viable cell density for the commencing of the fed-batch culture.

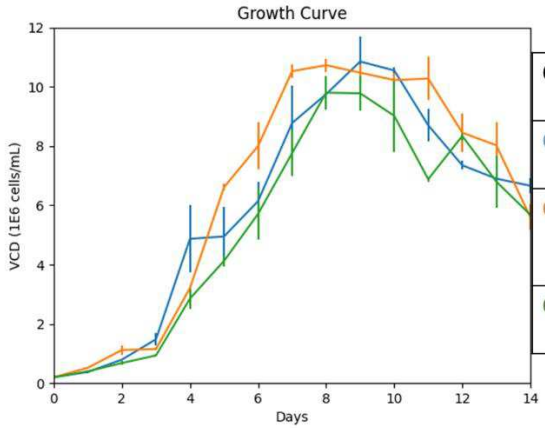
7.3.2 CHO-B Analysis

The final step for the validation of the leucine abundance genetic strategy is the mild hyperthermia fed-batch culture of the engineered CHO strain, CHO-B. In the previous chapter, the analysis of the metabolic engineering results concluded that the removal of the BCAT1 enzyme together with an inhibition of growth rate can result in abundance of leucine in the cell. The pseudo steady-state assumption of FBA assumes no metabolite accumulation in the cell; therefore, leucine would be forced

to be consumed through its conversion to IgG. The BCAT1 enzyme is responsible for the conversion of leucine, isoleucine, and valine to glutamate, however with the dataset provided to the optimisation only leucine can be in abundance. The engineered cell line CHO-B produces less BCAT1 compared to the original CHO cell line, which was the cell line used in the D2 dataset to acquire the genetic engineering result through the optimisation of the D2 constraint CHO-K1 GEM. Specifically, analysis of the resulting blot (Fig. 7.12) showed an almost 70% reduction in the quantity of BCAT1 by the engineered cell line, which means that in the cell there is probably a 70% reduction for the conversion of leucine to glutamate. The engineered cell line does not carry a complete knockout of the gene but a knockdown of its expression. As seen in the section **4.2.2** of the sensitivity analysis chapter, there is a correlation between leucine catabolism and growth rate in the FBA optimisation. A *Bcat1* knocked down cell line would need to have higher inhibition in growth rate than a knocked out one to achieve abundance in leucine. The metabolic dataset, D2, used to constrain the optimisation was taken from a fed-batch culture using the commercially available FeedC. It is important to try and replicate the same culture conditions used in the D2 dataset, so the culture media, feeding strategy and feed was replicated herein. Ideally, the uptake and secretion rates would not deviate a lot from the original dataset. However, the inhibition of growth took place by applying mild hypothermia during cell culture, which changes the parameter of temperature in the experiment. In mild-hypothermic conditions the normal temperature of culture is reduced from 36.5°C to 32.5°C which reduces the metabolic activity of the cell and the total amino acid uptake (*218, 219*). The hypothesis is that the uptake of leucine should not reduce to the degree that there is no remaining leucine in the cell after the inhibition of growth rate. Mild hypothermic conditions inhibit the growth rate, so it was not applied on the start of the culture, but on the end of the early exponential phase, day 4. The hypothesis is that the fed batch culture of the previously engineered cell line CHO-B under mild-hypothermia conditions and using FeedC will increase the specific productivity of IgG compared to the original CHO cell line.

The fed-batch experiment tested whether the modifications in metabolism had an effect in the quantity of the secreted mAb. To examine the effect, the use of appropriate control cell lines was necessary, in order to get the most representative conclusions from the growth rate and resulting mAb titer. The engineered CHO-B was compared to the original CHO cell line with no additional genetic modifications, which was revived a week before the start of the experiment. The comparison showed how the limitation in leucine catabolism can impact antibody production. Another control cell line was the original CHO which has undergone all the procedures that the CHO-B has gone, CHO'. Specifically, it was revived on the same day as the CHO-B cell line and underwent the process of transfection but with no DNA, and it was subcultured in 6-well plates for a month until it was transferred to 125mL Erlenmeyer flasks at the start of the fed-batch experiment. The CHO' cell line showed whether the effect of electroporation and age of the culture had an influence on growth rate or produced mAb. All three cell lines were cultured at 36.5°C until day 4, and then the temperature was reduced to 32.5°C for the rest of the culture. The fed-batch cultures were conducted according to the protocol of section **6.2**

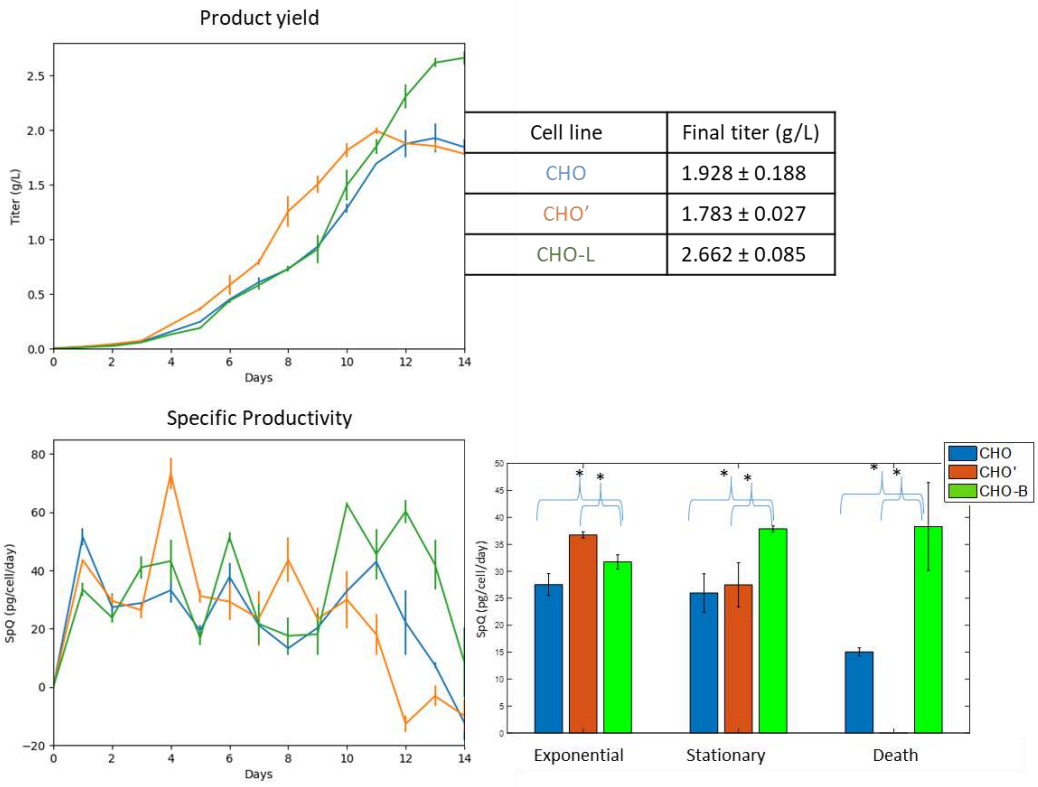
in the experimental materials and methods chapter. Due to lack of Erlenmeyer flasks, the fed-batch cultures were conducted in duplicates, and not triplicates. The growth curve for all three cell lines is presented in figure 7.13.



Cell line	IVCD (1E6 cells/ml/hr)	Growth Rate (hr-1)
CHO	2038.032 ± 174.865	0.015 ± 0.000
CHO'	2212.992 ± 28.307	0.022 ± 0.001
CHO-B	1824.504 ± 95.273	0.019 ± 0.001

Figure 7.13: Growth curve of the fed-batch culture for all three cell lines used. The blue line is the original CHO cell line. CHO' (orange line) is the original cell line but underwent thought all the procedures that CHO-B (green line) went, for its genetic engineering. Next to the graph, the table shows the IVCD of each culture on day 14 of the culture and the growth rate during the exponential phase of growth.

The engineered CHO-B cell line reached the lowest IVCD at the end of the fed-batch compared to the original CHO and CHO' cell lines. The growth rate of all three cultures is similar. The knockdown of BCAT1 could be responsible for additional production of IgG as the total biomass production seems to be lower. However, a student t-test comparing the three cells lines showed the differences were not statitistically significant((p-values f> 0.1). The yield and production graphs are presented in figure 7.14.



Cell line	Final titer (g/L)
CHO	1.928 ± 0.188
CHO'	1.783 ± 0.027
CHO-L	2.662 ± 0.085

Figure 7.14: Product yield (top) and specific productivity (bottom) graphs for the fed-batch cultures of the three cell lines analysed. The blue line is the original CHO cell line. The CHO' cell line (orange line) is the original cell line that underwent all the procedures that the engineered cell line CHO-B (green line) went. Next to the two graphs, the table (top) shows the final IgG titer on day 14 of the culture. The bar chart (bottom) shows the average q_{mAb} in each growth phase of the fed-culture for the three different cell lines. The asterisk indicates significant difference for p-value less than 0.05.

The engineered CHO-B cell line reached higher IgG titer towards the end of the fed-batch culture, with an increase of almost 35% compared to the original CHO cell lines. The IVCD and the growth rate of the CHO-B cell line did not increase significantly compared to the CHO cell lines, which results in a 35% increase of q_{mAb} of CHO-B during stationary phase and an increase of 100% compared to the original CHO cell during death phase. From figure 7.14, it seems that while the original CHO' cells stopped producing IgG after day 10-11, while CHO-B kept producing significant amounts of the protein. During the exponential phase of growth, CHO-B shows a slight increase of q_{mAb} compared to the CHO cell lines, but a decrease when compared to CHO' cell line. The major increase of the CHO' cell line, seems to be cause by the point on day 4, which is the result of a sharp increase in VCD during that time. The significant increase of IgG titer by CHO-B, is a positive result for the use of strain engineering in CHO cell lines for the improvement of their q_{mAb}.

The two original CHO and CHO' cell lines have almost the same average q_{mAb}, during the stationary phase, which is expected since CHO reached higher IgG titer but with a reduced growth rate and final IVCD than the older cell line CHO'. However, during exponential growth phase, CHO'

reaches higher productivities than the CHO cell line, and during death phase, the CHO' cell line reached much smaller specific productivities than the CHO cell. It seems that the older the cell line is, the less productivity it has while reaching higher VCD, which might be the reason they both reached almost the same value for the final IgG titer. However, there is no significant difference between the two cell lines, so the result could be accounted to random events.

Overall, the BCAT1 knockdown cell line has a higher q_{mAb}, however, the increase of titer and q_{mAb} appears to mainly happen after day 8, during the stationary and death phase of culture, while the prediction by the optimisation was for the increase to happen during exponential phase. In order to ensure that the modification was maintained until the end of the fed-batch, I checked that the knockdown was still present at the end of the culture. At the start of the fed-batch culture the cell line CHO-B produced 70% less BCAT1 than the original CHO cell (Fig. 7.12). The reduction in the expression requires antibiotic as pressure to maintain the BNu1 plasmid inside the cells, otherwise it is possible the unmodified CHO cells of the CHO-B cell pool might prevail in the population. However, it is not ideal to treat the cell with hygromycin during fed-batch culture. The synthesis of the hygromycin resistance enzyme and the SpCas9 protein takes away resources from biomass and IgG production, which was the reason there was no supply of antibiotic throughout the duration of the fed-batch culture. The removal of the antibiotic might allow the cells to slowly decrease the degree of BCAT1 downregulation. A way to detect potential changes in the expression of the *Bcat1* gene is a western blot analysis in the end of the fed-batch culture. The analysis of the samples at the end of the culture by western blot is displayed at figure 7.15.

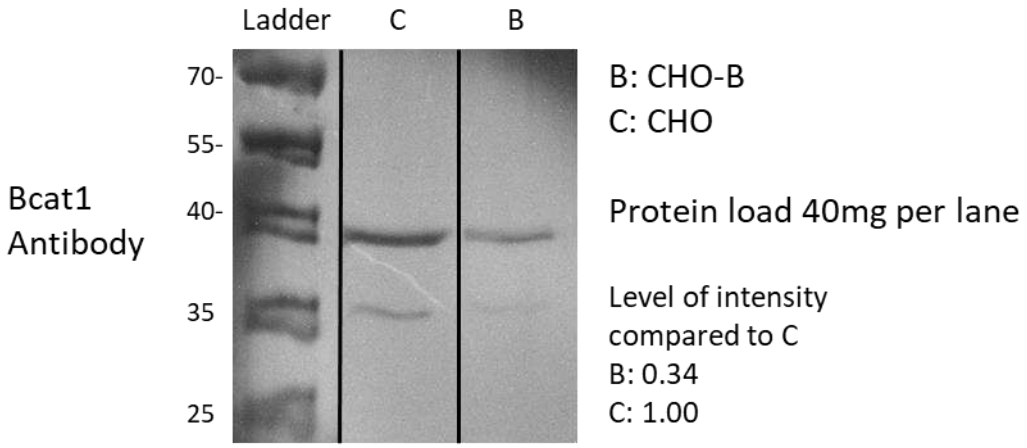


Figure 7.15: The first lane is the ladder of the molecular weights. The protein load for both lanes was 40µg. The molecular weight of BCAT1 (ab232700, abcam) is around 43kDa. The first lane, C, is the original CHO cell line at the end of the 14-day fed batch culture. The second lane, B, is the CHO-B at end of the fed-batch culture.

The knockdown degree of BCAT1 was reduced by almost 5% from the start of the fed-batch to the end. In the start of the fed-batch, CHO-B expressed BCAT1 in 70% reduced levels than the original CHO cell line (Fig. 7.12), whereas in the end of the fed-batch culture (Fig., 7.15) CHO-B displayed a 65% reduced BCAT1 expression. The amount is insignificant because the image analysis technique is not very sensitive and the 5% difference can be considered to be within error. It should be noted that the control CHO cell lines of the two figures (7.12 and 7.16) are different. The initial western blot used the CHO' as control and the final western blot the original CHO cell line. However, they both contain no genetic modification on the *Bcat1* gene, so they should have the exact same genotype. In conclusion, both the initial western blot before the start of the western blot (Fig. 7.14) and the western blot at the end of the fed-batch culture (Fig. 7.17) display a significant reduction in the expression of the BCAT1 enzyme between the engineered CHO-B and the original CHO cell lines. Therefore, it can be concluded that the increase in q_{mAb} in CHO-B takes place during the end of the culture, while the cell line expresses 65% less the BCAT1 protein.

The D2 bounds used for the optimisation were taken from the exponential phase of the fed-batch culture and result in leucine abundance because the minimum uptake of leucine is more than the cell can handle after the removal of the BCAT1 enzyme. Using the same experiment of the D2 dataset, it is possible to get the stationary phase bounds of D2 and constrain the CHO-K1 GEM. Similarly to before, a knockout of *Bcat1* and inhibition of growth rate result in positive IgG secretion flux according to the FBA optimisation. During stationary phase, the value in the D2 dataset for the minimum uptake of leucine is 0.0027mmol/gDCW/hr. Based on the biomass equation of the CHO-K1 GEM, it is the amount of leucine required for growth rate of $0.004hr^{-1}$. The value is very small, but the growth rate during stationary phase is basically zero, and sometimes it can be seen to be negative (Fig. 7.13). If biomass growth rate is reduced to less than $0.004\ hr^{-1}$, then there would be uptake of leucine that must be consumed thought the IgG synthesis pathway. However, the objective of the cell during stationary phase might not be biomass production, as it is during exponential phase (*220, 221*), and FBA optimisation in that period of culture is not a straightforward job. A potential reason there was an abundance of leucine in stationary rather than in exponential phase might be because of the much larger decrease in biomass production rate, due to use of mild hypothermic conditions. As explained above, the temperature shift form 36.5°C to 32.5°C took place on day 4, while the stationary phase, as seen in figure 7.13, started a little after on day 8-9. A control CHO cell line fed-batch culture under normal temperature could provide useful insights, regarding the degree of biomass inhibition due to mild hypothermia.

A similar study in CHO cells also performed knockout of the gene *Bcat1* to remove growth inhibitors which increased the growth rate and q_{mAb}, compared to the cell line that produced the inhibitory by-products (*222*). The enzyme produces by-products such as isovalerate, isobutyrate and 2-methylbutyrate, which act as growth inhibitors for the catabolism of the amino acids, growth rate and recombinant protein production. However, in the case of the CHO-B cell line there was no sign of

growth rate increase (Fig. 7.13), only an increase in the quantity of the produced IgG. It is important to determine the reason behind the increase in titer. It can be either an abundance in leucine, as the model predicted, a reduction of inhibitory by-products, as studied by the similar study, or it can be an irreplicable result.

7.4 Conclusions

The chapter focused on an attempt to validate two of the genetic engineering strategies proposed by the strain optimisation of the CHO-K1 GEM conducted in the previous chapters. The strain design algorithm used experimental datasets to constrain the GEM of CHO-K1, which was then utilised to provide possible reaction regulations that can achieve production of IgG in an FBA optimisation that maximises biomass production. Based on the analysis of the changes in the metabolic network, the conclusion was that IgG is produced if there is an abundance in an amino acid which is then consumed for antibody synthesis, resulting in an increase of q_{mAb}. The changes in metabolic reaction fluxes were translated into gene modifications, which constructed three genetic engineering strategies. Their purpose is the increase of IgG productivity per cell during the exponential phase of growth. Out of the three scenarios that were analysed in depth in the previous chapter, two were tested herein, due to them being the simplest to perform *in vitro*. The first strategy includes the knockout of the gene *Aldh4a1* and the overexpression of *Aldh18a1*, centred around proline metabolism, while the second contains the disruption of the *Bcat1* expression combined with an inhibition of growth rate, which can be done using mild-hypothermia conditions, and is based on the creation of leucine abundance in the cell.

The first cell line, CHO-AA, was not possible to be engineered correctly and was not tested regarding its IgG productivity. Initially, the original CHO cell line had the expression of the gene *Aldh4a1* reduced by 70%. The CRISPR/Cas9 system, using two different gRNA sequences nicked the first and second exon of the *Aldh4a1* gene. The engineered cell line was stable, because of the use of the antibiotic hygromycin acting as a selection marker. Subsequently, the CHO-A cell line was transfected with an ORF plasmid expressing the mRNA sequence of *Aldh18a1* to attempt and create the CHO-AA cell line. However, after multiple western blots there was no proof that the protein ALDH18A1 is expressed in the CHO cell line. The gene is not being translated in CHO-S (*216*) and CHO-K1 (*217*), and according to the results from the western blot analysis the ALDH18A1 protein does not exist in the original mAb producing CHO cell line. The most likely scenario is that the gene *Aldh18a1* is being transcribed to mRNA, since there is evidence that the mRNA exists in other CHO cell lines (*108*), but it is not translated to protein. A way to confirm it is to perform a qPCR to detect the mRNA levels. The reason the mRNA is not translated to protein might be that the cells do not need it for their survival. The CHO cells are proline auxotrophs, which means that they cannot synthesise proline, but they need to uptake it from the culture media. The gene *Aldh18a1* is the only gene capable

of synthesising proline, so if the cells are cultured in L-proline-free culture media it may be possible for the overexpression to work.

The second cell line that was engineered was the *Bcat1* KO cell line, CHO-B. The original CHO cell line had the expression of the gene *Bcat1* disrupted, using the CRISPR/Cas9 system. The engineered CHO-B produced 70% less quantity of the BCAT1 enzyme compared to the original CHO cell line. The selection marker used was the antibiotic hygromycin, which would limit the possibility of the unmodified CHO cell line to prevail in the cell population of CHO-B. Afterwards, the cell line was cultured under mild-hypothermia fed-batch conditions, using the commercially available FeedC, to examine any potential correlations between the disruption of the gene and q_{mAb}. The optimisation suggested the modification of multiple genes to inhibit the growth rate. However, mild hypothermic conditions are easier to control for biomass inhibition, and do not carry the danger of very low growth rate in a very low cell population environment. The productivity of CHO-B was compared to the original CHO and CHO' cell lines. CHO is the original mAb producing cell line, and CHO' is the original cell line but it underwent all the genetic engineering processes of the CHO-B cell line. Three main conclusions can be derived from the fed-batch culture of the cell lines. The two original CHO cell lines displayed differences on their μ and produced titer. CHO reached slightly higher titer than CHO', but lower total IVCD and growth rate. They were cultured under the same conditions, so it seemed like the age of the culture affects the quantity of the produced antibody. Both CHO cell lines reached higher IVCD but lower titers than the knocked down CHO-B cell line, on the final day of the fed-batch culture. The final titer of CHO-B at day 14 was 2.6g/L, about 35% more than the original CHO cell lines. The significant increase in titer was the result of an increase in the average specific productivity of the engineered cells. Towards the end of the culture, the engineered cells display much higher specific productivities. An important conclusion is that the increase in the q_{mAb} of CHO-B compared to the original cell lines takes place mainly during stationary and death phase. The stationary phase of the experimental dataset D2 also predicts an abundance in leucine, however it is important to question why the increase took place for which additional experimental analysis of the metabolome is required. It is possible the increase took place, because of larger growth rate inhibition during that period of culture which can be determined by performing a fed-batch culture at normal temperature of CHO cells an comparing the growth rate reduction due to temperature drop. However, an analysis of the metabolome is also necessary.

The enzyme BCAT1 produces inhibitory by-products isovalerate, isobutyrate and 2-methylbutyrate, which reduce the productivity of the cell. The removal of the enzyme form the CHO cell metabolism has displayed an increase in growth and productivity (*222*). Even though there was no increase in growth in the case of CHO-B, it is still necessary to detect the concentrations of by-products between the original and the engineered cell line. Initially, a fed-batch culture of the original CHO cell line under normal temperature, could show the degree of biomass inhibition due to mild hypothermia. The abundance of leucine does produce additional IgG in an FBA optimisation, but it is necessary to

measure the leucine levels and observe whether the engineered cell line uptakes more or less than the original cell line. It is possible to know if the change happened because of an abundance of leucine or the reduction of inhibitory by-products, but it requires further metabolic analysis. A daily amino acid analysis of the culture media using HPLC could help quantify changes in the amino acid concentrations. Then, it is simple to calculate the different uptake rates of the amino acids. If the uptake rate of leucine is the same as in the original cell lines, then there is strong evidence in favour of the leucine abundances. However, the detection of the BCAT2 protein, which also catabolises leucine would be necessary. The protein does not exist in the CHO-K1 GEM, but a western blot would be better to determine its existence. Furthermore, it is possible to measure the quantities of the by-products secreted by the cell in the culture media, again by HPLC. A reduction of the by-products in the CHO-B cell line will be strong evidence in favour of the theory. A final and important test would include the replication of the experiment with the generation of multiple clones. The 35% increase in titer might not be the same with every clone and it should be tested whether the knockdown of *Bcat1* and inhibition of growth produce a reproducible result.

8

Conclusions & future work

The final chapter of the thesis summarises and discusses all results and suggests relevant future work.

8.1 Summary of results and conclusions

In this thesis, the specific antibody productivity in CHO cells was increased through a combined computational and experimental framework. MAbs are a very important product of the biopharmaceutical industry, being used to treat various types of cancer or autoimmune disorders. Their medical significance is portrayed in the very large market share they own. CHO cells are the workhorse for mAb production in industry. However, CHO cells do not naturally express mAbs and the synthesis of recombinant proteins is not among the cells' main objectives of resource allocation, resulting in low production titre and, by consequence, to increased costs of the upstream production phase of mAbs. A design of a CHO strain that can express mAbs in higher quantities can solve this problem and reduce the cost of their production, which potentially can impact their market price. In the introduction (Chapter 1), the main motivation behind the research is outlined, and a theoretical background behind the mAbs is presented. Then in Chapter 2, the mathematical formulation of the algorithms used for the optimisation is presented, together with the main research that is relevant to the thesis. The following three chapters apply the computational optimisation of the CHO strain, and the subsequent experimental testing of the results, takes place in Chapters 6 and 7.

The fourth chapter of the thesis is focused on the application of the strain design algorithm on CHO cell metabolism. To model the metabolism of CHO cells, the CHO-K1 GEM was used. It is derived from the generic iCHO1766 GEM, which was developed right before the commencement of the present study and was therefore the most up-to-date GEM at that time. GEMs are metabolic models that contain most of the genes and enzymes of an organism. Any metabolic model describes the reaction network of the organism; however, the holistic behaviours of GEMs can provide a large variety of possibilities for strain designs to optimise a cellular objective against small scale networks. Experimental data is required to determine the potential values of metabolic reaction fluxes. Four different datasets, D1, 2, 3, 4, of experimental data were used (*127, 166*), which contained concentrations of extracellular metabolites from fed-batch CHO cells cultures producing IgG. The concentrations were converted to maximum and minimum fluxes for each of the extracellular

metabolites drain reaction of the CHO-K1 GEM. The four datasets constrained only the drain reactions of the CHO-K1, leaving the rest of the intracellular reactions unbounded. The elementary balance algorithms ccFBA and ncFBA were applied to calculate realistic and narrow bounds for the intracellular reactions. The way they work is by applying a constraint on the total atoms of carbon, ccFBA, and nitrogen, ncFBA, which can be consumed or produced by the biochemical reactions that contain carbon or nitrogen atoms, based on the total number of nitrogen and carbon atoms entering the cell in each of the datasets. The reactions of the model that were not reduced by any of ccFBA or ncFBA, were constrained by FVA. Both the elementary constrains and the application of FVA reduced the maximum flux that can be carried by the intracellular reactions, which will be very useful in the upregulation of reaction during the strain optimisation. In order to verify the quality of the bounds of each dataset, an FBA optimisation was performed on the constrained CHO-K1 GEM to predict the q_{mAb} and μ. The comparison of the values to the experimental ones from each dataset showed that D1 contains a very strict bound for the cysteine uptake, which was derived from the extracellular concentration of the amino acid. As a result, the FBA optimisation to maximise IgG synthesis on the D1 constrained CHO-K1 GEM was infeasible, in contrast to the D2, 3, 4 where the constrained GEM was feasible for the maximisation of both biomass and IgG. The D2, 3, 4 datasets did not have a value for cysteine, so they borrowed the value from a different dataset, with similar experimental conditions to them (*92*). To deal with the infeasibility of the D1 dataset, the constraint of biomass production in the model was loosened, and the FBA optimisation was feasible. The predicted IgG and μ, of the CHO-K1 GEM, constrained by the four datasets, are the control values that will determine which strains of the optimisation exhibit the highest increase in these values.

The algorithm used for the q_{mAb} optimisation of the CHO cell line was OptReg. Optreg is a bilevel optimisation that detects reactions to up-, downregulate or knock out. The upregulation of a reaction increases the lower bound of the flux it can carry, while the downregulation reduces the upper bound. The increase or decrease is a function of the range between the initial upper and lower bound, multiplied by a user-defined coefficient. For this reason, it was important to apply strict constraints on the intracellular reactions, so the optimisation would not be forced to pick very small coefficients for the regulation. The knockout of a reaction sets both the upper and lower bounds to zero, forbidding any flux from going through the biochemical reaction. The optimisation was performed on the CHO-K1 GEM, constrained by the ccFBA and ncFBA curated D1, 2, 3, 4 datasets. A mathematical constraint to account for the reversibility of the reactions was also added to the optimisation algorithm, removing the second term of the OptReg outer optimisation objective function which minimised the sum of the fluxes. The strain design algorithm, named Bio_Igg, was also modified to Igg_Igg which instead of maximizing both biomass and antibody production, it maximised antibody production in the outer optimisation while it minimized it in the inner optimisation to simulate a worst-case scenario optimisation. The application of the optimisation on the CHO-K1 GEM, resulted on 93 different scenarios with an average of 130 reactions regulations per scenario. All 93 were able to produce positive flux of IgG in an FBA

optimisation of the CHO-K1 GEM, with biomass the objective function. The scenarios produced by both Bio_Igg and Igg_Igg predicted similar combinations of μ and q_{mAb} but Igg_Igg reached optimal solution much faster than Bio_Igg. Despite the successful implementation of the algorithms, the resulting scenarios had an average of 140 regulations, making them infeasible for *in vitro* testing.

Chapter 5 tried to solve the problem of the very high number of regulations and analyse the changes in the metabolism proposed by the scenarios. The reduction in the number of regulations in each scenario was accomplished using sensitivity analysis. Initially, the regulations were tested one by one to determine whether they are essential to produce IgG on an FBA optimisation with biomass being the objective function. Then, a combinatorial approach was applied to determine the least possible number of modifications in each scenario that can, *in silico*, couple the two diametrical objectives. Also, the SA framework identified the high-producing scenarios. The FBA predicted q_{mAb} and μ were compared to the control values of the original CHO-K1 model, and the ones with higher than at least 50% of the values were considered high-producing. The 20 high-producing scenarios had an average of five regulation. The scenarios were analysed by comparing the flux of an FBA optimisation with biomass being the objective function before and after the application of the regulations. The main conclusion was that all 93 scenarios proposed by the optimisation used a combination of growth rate inhibition and formation of an amino acid abundance in the cell, which would be channelled towards antibody synthesis. The inhibition can take place with a variety of ways, like the upregulation of a pathway that consumes ATP or by blocking optimal pathways with the removal of a reaction. The abundance of an amino acid can happen by the removal of a catabolic reaction of the amino acid. It should be noted that the amino acid can be any of the ones that are the cell uptakes in each of the dataset. The 93 scenarios include strategies for increasing the abundance of proline, valine, arginine, leucine, tryptophan, or lysine. Mathematically, the goal is for the lower bound of the amino acid drain reaction, uptake, to have a value not limiting the biomass production in the GEM. That way, after the maximum biomass is produced, there is remaining amino acid in the cell. Three of the 20 scenarios were analysed in depth, and the suggested regulation on the reactions bounds by the optimisation were converted to gene modifications which can be applied on the CHO-K1 genotype. Two of the scenarios were deemed simple and easy to perform in a real-life CHO-K1 descendant cell line *in vitro*.

The results of Chapter 7 are purely experimental and focused on the validation of the genetic strategies identified in the previous chapter. Two genetic engineering strategies were tested, on the original mAb producing CHO cell line. The cell line is descendant from CHO-K1, which is the cell line the GEM used is based on. The first scenario that was tested included the knockout of the gene *Aldh4a1* in combination with overexpression of the gene *Aldh18a1*. Both genes exist on the CHO-K1 genome and there was also evidence there are transcribed from their mRNA levels in the transcriptomics dataset used to build the CHO-K1 GEM (*108*), so it was thought that the genetic engineering strategy was feasible. The CRISPR system was used to perform the gene knockout with the SpCas9 protein. The SpCas9 protein and two gRNAs were transfected to the cell line, which performed nicks on the first and

second exon of *Aldh4a1*. The early cuts in the gene sequence, did not cause a gene knockout but a downregulation of the gene's expression by 70%. Even though the optimisation suggested a gene knockout the experiment was not tried again due to time limitations. The cell line CHO-A was created, which was stable due to the addition of the antibiotic Hygromycin that maintained the expression of the SpCas9 protein and the two gRNAs. Subsequently, the CHO-A cell line was transfected with an ORF plasmid to overexpress the *Aldh18a1* gene, which contained the mRNA sequence of the ALHD18A1 protein. Ideally, the mRNA should have gotten translated to the ALDH18A1 protein. However, after a series of western blot analysis there was no sign of the protein, neither in the original CHO cell line nor in the transfected with the mRNA CHO-A, and the CHO-AA could not get properly engineered. The second genetic engineered scenario suggested the knockout of the *Bcat1* gene, in combination with two regulations that inhibit the growth rate. However, the application of genetic modifications to reduce the growth rate face the problem of increased difficulty of the engineered cell to reach an optimal VCD after the transfection and selection process. For that reason, it was thought as the safest option to culture the cells under mild hypothermic conditions, which inhibits the growth rate and as a change of the environmental conditions is easier controlled and monitored. The knockout of *Bcat1* used the CRISPR technology similar to before, with the SpCas9 protein and a gRNA targeting the fifth exon of the gene. The stable cell line CHO-B was created. Unfortunately, same as the previous attempt to knockout the *Aldh4a1*, the western blot analysis showed a downregulation in the protein quantity in the cell by 70% instead of a complete knockout that the model suggested. Despite that, the cell line was cultured under fed-batch conditions in 32.5ºC, because the downregulation would still create an abundance in leucine if there was sufficient reduction in growth rate. The 14-day fed-batch culture used the commercially available FeedC, and the temperature was reduced to 32.5C on the day 4 of the culture. The first 4 days the temperature was kept at 36.5 to assist the early exponential phase. At the end of the fed-batch culture, the CHO-B cell line resulted in 35% higher IgG titer and q_{mAb} compared to the two original CHO cell lines.

8.2 Main Contributions and Conclusions

The field of chemical engineering is the study of processes, leading to suggestions for optimisation and improvements regarding the maximisation of an objective function. The research conducted in the thesis is the application of chemical engineering principles in a real-life problem. Initially, the process which is the production of mAbs from CHO cell fed-batch cultures was studied and analysed. Then it was converted to a mathematical model in the form of a GEM, which was optimised to increase the production of the recombinant protein. The results of the optimisation were studied and were experimentally tested, to provide a better understanding of the bioprocess, the mathematical model, and the experiment itself. The main contribution of the thesis is the application of all stages for a bioprocess optimisation.

The production of mAbs from CHO cell fed-batch cultures was optimised with a combined *in silico* and experimental framework. The newest developed GEM of the CHO cell at time of conducting the optimisation was utilised to engineer CHO cell lines with increased q_{mAb}. The novelty of the research lies in the successful application of a strain design algorithm on the metabolic network of CHO cells, to *in-silico* couple the production of mAbs with growth rate. The results were subsequently experimentally tested. The algorithm OptReg was applied on the GEM of the CHO-K1 cell, which was constrained by four experimental datasets. The validation *in vitro* included gene knockouts with CRISPR/Cas9 and overexpression by transfecting ORF plasmids. The overall combined use of computational and experimental methods allowed the practice of many different methods and techniques, and the in depth understanding of the bioprocess. Overall, the thesis provides four novelties on the topic of bioprocess modelling and optimisation.

1. To my knowledge this is the first time, while writing the thesis that a strain design algorithm was successfully applied on a metabolic network of a mammalian organism, and specifically in CHO cells. Mammalian cells have more complicated metabolism that bacteria and yeast. The fact that they also need many more metabolites for survival, against only a few that for example *E. coli* needs, makes the strain optimisation much more complicated due to the possibilities of active pathways. The process of strain optimisation took almost two months of CPU time, showing how difficult of a task it was, due to the number of binary variables and equations of the MILP optimisation. Nevertheless, a plethora of 93 unique scenarios were calculated, that can produce IgG in a biomass maximisation FBA, of the CHO-K1 GEM.
2. The thesis provides a biological explanation for the results of the optimisation, the idea of which has not been found in literature as a possibility for increasing q_{mAb}. The main conclusion, based on the metabolic changes proposed by the results of the strain optimisation is that antibody is synthesised to consume remaining metabolites in cytosol. The optimisation suggested to inhibit the growth rate, which reduces the amount of the amino acids needed to be consumed for biomass production, based on the biomass equation of the CHO-K1 GEM. As a result, the remaining metabolite, which in the results was one of the amino acids that the GEM uptakes, cannot be consumed through the biomass equation and is directed to its catabolic reactions. If the reactions are removed, then the only other available pathway would be the antibody synthesis. For example, if there is leucine remaining in the cell after the maximum possible biomass is formed, leucine could be converted to glutamate by the BCAT1 enzyme. The removal of the BCAT1 protein creates an abundance of leucine which will be forced to be consumed through the other possible pathways, which are the synthesis of heavy and light chain of IgG leading to the IgG secretion. The above reasoning can also be applied for other amino acids, depending on their uptake bounds. Overall, the optimisation suggested that creating an

abundance of a metabolite in the cytosol will channel it towards the mAb synthesis pathway. Despite the novelty of the biological proposal, it is important to test the theory *in vitro*.

3. The strain optimisation of CHO-K1 suggested the knockdown of *Aldh4a1* and the overexpression of *Ald18a1*. However, the engineered cell line could not be created, because the gene *Aldh18a1* is not translated to protein in the CHO cell line used for the strain engineering. Despite the successful transfection of the mRNA to the CHO cell line, western blots showed no evidence of the protein. Transcriptomics data of other studies show that the gene is being transcribed (*108*), but apparently is not translated. CHO cells are proline auxotrophs (*108, 216, 217*) and do not convert glutamate to proline, which might explain the reason there was no evidence of the ALDH18A1 protein. If the culture media was proline deficient, then it is probable the overexpression would have worked, since the expression of the protein would be the only way for proline to be synthesised and take part in metabolism. Nevertheless, the overexpression shows a bottleneck in the CHO-K1 GEM construction. Even if the researchers detect mRNA levels of the gene, the protein should not have been included in the GEM, because it is not detected in the proteome. It is important, for a GEM reconstruction, to consult both transcriptomics and proteomics data of an organism, and to also account for all known autotrophies for an organism.
4. Another scenario that the strain optimisation resulted in, and it was experimentally tested was the knockout of the *Bcat1* gene in combination with two regulations that inhibit the growth rate. The engineered CHO cell line, CHO-B, had 70% reduced BCAT1 expression. The reduction of growth rate took place using mild hypothermia instead of genetic regulations. After a 14-day fed-batch culture, CHO-B reached 35% increased IgG titer than the original mAb producing CHO cell line. CHO-B had slower growth rate and reached smaller IVCD than the original CHO cell line on the end of the fed-batch, but as a result the q_{mAb} of CHO-B was 35% higher than the original CHO cell line. It should be noted that a major increase in titer took place during stationary and death phase, even though the optimisation was performed for the exponential phase of the cell culture. A potential reason could be that the biomass reduction during the late stages of the culture was higher, than during the exponential growth.

8.3 Future work

The following paragraphs present suggested work that can either improve the optimisation framework or give answers to questions born during the research. There are three main distinct part for the future work, the improvement of the strain optimisation framework, additional genetic engineering strategies and the further analysis of the engineered CHO-B cell line.

8.3.1 Strain optimisation

The main tool used for the strain optimisation was the GEM of the CHO-K1 cell. GEMs are mathematical representations of the phenotype and genotype link, and they contain most genes and biochemical reactions belonging in an organism cell. It was used to detect possible biochemical reactions and genetic regulations in the CHO-K1 cell that can increase its q_{mAb} through the coupling of μ and the secreted IgG. However, the GEM used was the one published in 2018, which even though is only four years old at time of writing the thesis, there have been newer more complete GEMs of the CHO cell (*77, 96*). The newer models contain more reactions which translate in additional possible combinations that the diametric objectives can be coupled in the FBA optimisation. Also, newer GEMs contain more genes with their GPR associations, so reactions can be categorised better, and they can easier be associated with genes without having to search an external database. The newer version of CHO cell GEMs includes the protein secretory pathway (*110*), which could potentially result on completely different scenarios of genetic regulations. Another possible optimisation that could take place is one for different cell lines. In the thesis, only the CHO-K1 cell line was tested. The optimisation of a different GEM, such as the GEM of the CHO-S or CHO-DG44, could potentially provide different results. For example, the generic iCHO1766 GEM contains the genes *Bcat1* and *Bcat2*, both of which can catabolise leucine. The CHO-K1 GEM, contains only the first, so a GEM of another CHO cell line, could potentially provide different results to create an abundance in leucine. However, the most complete and reliable GEM for the CHO cell line that can be used in for strain engineering, would be a custom made one for the in-house cell line. Using transcriptomics and proteomics data the generic GEM can be tailored to the cell line that will be used for the experimental validation of the strain optimisation results. That way, the genes and proteins will be in the genome when the optimisation suggests them as an option.

The GEM was designed for an over-production of mAbs, using the strain design algorithm OptReg. The algorithm and the whole optimisation framework could be further modified to search quicker a larger solution space for additional possible genetic modifications. The majority of the strain optimisation results contained a set of reaction regulations that cause an inhibition in growth rate. Therefore, an optimisation that constrain the biomass production would result the reactions that form an abundance in amino acids and possible metabolites. That way, it is possible the optimisation would reach an optimal solution faster and result just the possible combinations to create a metabolite abundance in the cell. Also, the current strain optimisation took place during the exponential phase of growth, while the increase of titer in the CHO-B genetic engineering scenario took place during the late stages of culture. It would be noteworthy to optimise the GEM during the stationary phase of culture and compare the *in silico* results between the two phases. The results of the optimisation could also be optimised by including the essential genes from experimental databases such as the one used herein to analyse post hoc the results (*183*). That way, it would be less likely for a gene suggested by the

optimisation to be essential. The accurate bounds of the reactions are also very important for the strain optimisation results. Herein four different datasets of extracellular bounds were used to constrain the drain reactions, which were then augmented to include the intracellular reaction used elementary balance algorithms. The use of multiple sets of bounds, with the use of different feedings, as well as the usage of thermodynamic data to correctly adjust the reversibility of reactions could help predict realistic data. Concerning the analysis of the metabolic changes, the sampling of the newly created solution space after the application of the changes in the metabolism could potentially give more insights compared to the analysis of fluxes predicted by the FBA optimisation. In Chapter 5, an FBA optimisation to maximise biomass was performed before and after the application of the changes, which showed the difference in the fluxes' values. A sampling of the solution space could show the changes of the new solution space, after the application of the metabolic changes, which is a more generic approach to observe metabolic changes.

8.3.2 Further genetic engineering

Time limitations of the project allowed the testing of only two genetic engineering scenarios, while the amino acid abundance hypothesis could be better supported by the testing of multiple scenarios. There was an effort to create the CHO-AA engineered cell line, which creates an abundance in proline, but the overexpression of *Aldh18a1* was not successful. The use of a proline deficient culture media would force the expression of the ALDH18A1 protein. The protein could also be overexpressed by protein transduction, instead of transfecting the mRNA, since the assumption is the mRNA is not translated to the protein. The above strategy could show if proline abundance would be similar to leucine, could increase the q_{mAb} of the CHO cell line. Other engineering strategies, such as creating abundance in lysine or in asparagine, could also support the amino acid abundance argument. Concerning the knockout of genes, the methodology behind a gene knockout could be optimised by a multiple gRNA screen. The detection of the optimal gRNA sequence than can disrupt the expression of a gene is very important. The quick testing of multiple sequences could provide the most optimal gRNA sequence, which could be transfected to the CHO cell line. Also, future knockouts or overexpression of genes through single cell selection could provide more optimal results. A single cell could provide a complete knockout of a gene, which was not accomplished herein.

8.3.3 CHO-B Analysis

The engineering of CHO-B and its subsequent fed-batch under mild-hypothermia conditions gave rise to more questions than were answered. Even though the final IgG titer in the end of the fed-batch increased by 35% compared to the original mAb producing CHO cell lines, further validation tests are

necessary to establish the reasoning behind the increase. Firstly, more clones of the CHO-B engineered cell line should be tested, regarding their productivity of mAbs. It is possible that a different clone of the CHO-B would not display a 35% increase, and additional mild-hypothermia fed-batch cultures should be conducted to get a complete picture and a better estimate for changes in q_{mAb}. The purpose of the optimisation was the suggestion of an optimisation strategy for biopharmaceutical production of mAbs in CHO fed-batch cultures. The large-scale production takes place in bioreactors or thousands of litres, so it is important to test the effect than the knockdown of *Bcat1* cultured in mild hypothermic conditions have on the production of mAb in a large-scale culture. The conditions of a bioreactor and a shake flask fed batch are very different, so the uptake and secretion fluxes of the cell will probably different as well. In large scale production of mAbs the quality of the product is of great importance. It is important to examine the glycoform of the secreted IgG because the combination of mild-hypothermia and the knockdown of the *Bcat1* gene might change the quality of the product comparted to the original CHO.

Concerning the biological reason for the 35% increase of the q_{mAb}, further tests could be made regarding the extracellular amino acids in culture media. Initially, a control CHO fed-batch culture should be performed, on 36.5°C, to determine the amount of growth rate inhibition due to mild hypothermia. Also, HPLC analysis in the culture media for the 14-day fed-batch culture can show the uptake levels of the CHO-B and the original CHO cell line. If the uptake levels of the amino acids are the same, then it is possible there is abundance of leucine like the optimisation predicted. However, leucine can also be catabolised by the BCAT2 protein which is expressed by *Bcat2*. Even though the BCAT2 enzyme does not exist in the GEM, it does exist in the genome of the CHO-K1. Western blot analysis should be conducted to verify that leucine can only be catabolised by BCAT1 in the CHO cell line used for the engineering of the CHO-B cell line. The amino acid concentrations measure by the HPLC could also be used as feedback for the optimisation. A new batch of strain optimisations on the CHO-K1 GEM could be conducted with the new amino acid data to constrain the model and compare the results. Mulukutla et al. observed that the removal of the *Bcat1* gene, reduces the production of the by-products isovalerate, isobutyrate and 2-methylbutyrate, which resulted in an increase in titer and growth rate. Even though, CHO-B did not show a decrease in growth rate, there was an increase of titer so the theory should be examined. The by-products can be determined using HPLC on the spent media, and compare the concentrations of the original mAb producing CHO and the engineered CHO-B.

List of publications

Articles

1. Antonakoudis, B. Strain, R. Barbosa, I. Jimenez del Val, C. Kontoravdi, Synergising stoichiometric modelling with artificial neural networks to predict antibody glycosylation patterns in Chinese hamster ovary cells. Computers & Chemical Engineering 154, (2021).

Reviews

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

Book Chapters

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

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

Bibliography

1. G. Walsh, Biopharmaceutical benchmarks 2006. *Nature Biotechnology* **32**, (2006).
2. G. Walsh, Biopharmaceutical benchmarks 2010. *Nature Biotechnology* **28**, 917-924 (2010).
3. G. Walsh, Biopharmaceutical benchmarks 2014. *Nature Biotechnology* **32**, 992-1000 (2014).
4. G. Walsh, Biopharmaceutical benchmarks 2018. *Nat Biotechnol* **36**, 1136-1145 (2018).
5. L. Urquhart, Market watch: Top drugs and companies by sales in 2017. *Nat Rev Drug Discov* **17**, 232 (2018).
6. J. Carrol. (https://www.fiercebiotech.com/, 2011), vol. 2022.
7. M. Castro *et al.*, Dupilumab Efficacy and Safety in Moderate-to-Severe Uncontrolled Asthma. *N Engl J Med* **378**, 2486-2496 (2018).
8. . (2017), vol. 2022.
9. T. A. Waldmann, Immunotherapy: past, present and future. *Nature medicine* **9**, 269-277 (2003).
10. B. Alberts *et al.*, *Molecular Biology of the Cell*. (Garland Pub., New York, ed. 4th, 2002).
11. L. M. Weiner, R. Surana, S. Wang, Monoclonal antibodies: versatile platforms for cancer immunotherapy. *Nature Reviews Immunology* **10**, 317-327 (2010).
12. B. P. Gerald, J. B. Lyczak, M. W. Lee, *Immunology, Infection, and Immunity*. (Washington, DC, 2004).
13. F. M. Burnet, A Modification of Jerne's Theoryof Antibody Production using theConcept of Clonal Selection. *CA: A Cancer Journal for Clinicians* **26**, 119-121 (1976).
14. K. S. A. Ullrich, Paul Ehrlich's magic bullet concept: 100 years of progress. *Nature Reviews Cancer* **8**, 473-480 (2008).
15. G. Kohler, C. Milstein, Continuous cultures of fused cells secreting antibody of predefined specificity. *Nature* **256**, 495-497 (1975).
16. G. J. Weiner, Building better monoclonal antibody-based therapeutics. *Nature Reviews Cancer* **15**, 361-370 (2015).
17. A. Frenzel, M. Hust, T. Schirrmann, Expression of recombinant antibodies. *frontiers in Immunology* **4**, (2013).
18. M. Donini, C. Marusic, Current state-of-the-art in plant-based antibody production systems. *Biotechnol Lett* **41**, 335-346 (2019).
19. M. H. Rashid, Full-length recombinant antibodies from Escherichia coli: production, characterization, effector function (Fc) engineering, and clinical evaluation. *mAbs* **14**, 2111748 (2022).
20. P. Martineau, P. Jones, G. Winter, Expression of an antibody fragment at high levels in the bacterial cytoplasm11Edited by J. Karn. *Journal of Molecular Biology* **280**, 117-127 (1998).
21. H. H. Huynh *et al.*, Functional production of human antibody by the filamentous fungus Aspergillus oryzae. *Fungal Biol Biotechnol* **7**, 7 (2020).
22. T. J. C. a. P. S. Dawn M. Ecker, "The Therapeutic Monoclonal Antibody Product Market," (BioProcess International, 2020).
23. C.-C. Kuo *et al.*, The emerging role of systems biology for engineering protein production in CHO cells. *Current Opinion in Biotechnology* **51**, 64-69 (2018).
24. J. Dumont, D. Euwart, B. Mei, S. Estes, R. Kshirsagar, Human cell lines for biopharmaceutical manufacturing: history, status, and future perspectives. *Critical Reviews in Biotechnology* **36**, 1110-1122 (2016).
25. J. Y. Kim, Y. G. Kim, G. M. Lee, CHO cells in biotechnology for production of recombinant proteins: Current state and further potential. *Applied Microbiology and Biotechnology* **93**, 917-930 (2012).
26. S. M. Noh, M. Sathyamurthy, G. M. Lee, Development of recombinant Chinese hamster ovary cell lines for therapeutic protein production. *Current Opinion in Chemical Engineering* **2**, 391-397 (2013).

27. T. T. Puck, S. J. Cieciura, A. Robinson, Genetics of Somatic Mammalian Cells: III. Long-Term Cultivation of Euploic Cells From Human and Animal Subjects. *Journal of Experimental Medicine* **108**, 945-956 (1958).
28. D. Reinhart *et al.*, Bioprocessing of Recombinant CHO-K1, CHO-DG44, and CHO-S: CHO Expression Hosts Favor Either mAb Production or Biomass Synthesis. *Biotechnology Journal* **14**, 1700686 (2019).
29. A. L. Grilo, A. Mantalaris, The Increasingly Human and Profitable Monoclonal Antibody Market. *Trends Biotechnol* **37**, 9-16 (2019).
30. A. F. Shaughnessy, Monoclonal antibodies: magic bullets with a hefty price tag. *BMJ : British Medical Journal* **345**, e8346 (2012).
31. C. R. Fernandez. (www.labiotech.eu, 2017), vol. 2022.
32. J. E. Bailey, Mathematical modeling and analysis in biochemical engineering: Past accomplishments and future opportunities. *Biotechnology Progress* **14**, 8-20 (1998).
33. K. P. Jayapal, K. F. Wlaschin, W.-s. Hu, M. G. S. Yap, Recombinant Protein Therapeutics from CHO Cells — 20 Years and Counting. *Chemical Engineering Progress* **103**, 40-47 (2007).
34. P.-C. Lin *et al.*, Improving Antibody Production in Stably Transfected CHO Cells by CRISPR-Cas9-Mediated Inactivation of Genes Identified in a Large-Scale Screen with Chinese Hamster-Specific siRNAs. *Biotechnology Journal* **16**, 2000267 (2021).
35. M. A. Oberhardt, B. O. Palsson, J. A. Papin, Applications of genome-scale metabolic reconstructions. *Mol Syst Biol* **5**, 320 (2009).
36. M. R. Antoniewicz, Parallel labeling experiments for pathway elucidation and 13C metabolic flux analysis. *Current Opinion in Biotechnology* **36**, 91-97 (2015).
37. K. Kwon *et al.*, Recombinant expression and functional analysis of proteases from Streptococcus pneumoniae, Bacillus anthracis, and Yersinia pestis. *BMC Biochemistry* **12**, 17 (2011).
38. R. H. Ng *et al.*, Constraint-Based Reconstruction and Analyses of Metabolic Models: Open-Source Python Tools and Applications to Cancer. *Frontiers in Oncology* **12**, (2022).
39. J. Nielsen, Systems Biology of Metabolism. *Annual Review of Biochemistry* **86**, 245-275 (2017).
40. A. Antonakoudis, R. Barbosa, P. Kotidis, C. Kontoravdi, The era of big data: Genome-scale modelling meets machine learning. *Comput Struct Biotechnol J* **18**, 3287-3300 (2020).
41. J.-H. Zhang, L.-L. Shan, F. Liang, C.-Y. Du, J.-J. Li, Strategies and Considerations for Improving Recombinant Antibody Production and Quality in Chinese Hamster Ovary Cells. *Frontiers in Bioengineering and Biotechnology* **10**, (2022).
42. E. T. Papoutsakis, Equations and Calculations for Fermentations of Butyric Acid Bacteria. *Biotechnology and Bioengineering* **26**, 174-187 (1984).
43. J. M. Savinell, B. O. Palsson, Network analysis of intermediary metabolism using linear optimization. I. Development of mathematical formalism. *Journal of Theoretical Biology* **154**, 421-454 (1992).
44. J. S. Edwards, M. Covert, B. Palsson, Metabolic modelling of microbes: The flux-balance approach. *Environmental Microbiology* **4**, 133-140 (2002).
45. Z. A. King *et al.*, BiGG Models: A platform for integrating, standardizing and sharing genome-scale models. *Nucleic Acids Res* **44**, D515-522 (2016).
46. J. A. Papin, N. D. Price, B. O. Palsson, Extreme pathway lengths and reaction participation in genome-scale metabolic networks. *Genome Res* **12**, 1889-1900 (2002).
47. A. M. Feist, B. O. Palsson, The biomass objective function. *Current Opinion in Microbiology* **13**, 344-349 (2010).
48. J. D. Orth, I. Thiele, B. O. Palsson, What is flux balance analysis? *Nature Biotechnology* **28**, 245-248 (2010).
49. Y. Chen *et al.*, An unconventional uptake rate objective function approach enhances applicability of genome-scale models for mammalian cells. *NPJ Syst Biol Appl* **5**, 25 (2019).
50. D. Segrè, D. Vitkup, G. M. Church, Analysis of optimality in natural and perturbed metabolic networks. *Proceedings of the National Academy of Sciences* **99**, 15112-15117 (2002).
51. D. Nagrath *et al.*, Integrated energy and flux balance based multiobjective framework for large-scale metabolic networks. *Annals of Biomedical Engineering* **35**, 863-885 (2007).

52. T. Tamura, W. Lu, T. Akutsu, Computational Methods for Modification of Metabolic Networks. *Comput Struct Biotechnol J* **13**, 376-381 (2015).
53. M. R. Antoniewicz, Methods and advances in metabolic flux analysis: a mini-review. *Journal of Industrial Microbiology and Biotechnology* **42**, 317-325 (2015).
54. C. Kontoravdi, E. N. Pistikopoulos, A. Mantalaris, Systematic development of predictive mathematical models for animal cell cultures. *Computers and Chemical Engineering* **34**, 1192-1198 (2010).
55. P. Kotidis *et al.*, Model-based optimization of antibody galactosylation in CHO cell culture. *Biotechnology and Bioengineering* **116**, 1612-1626 (2019).
56. D. G. Kleinbaum, M. Klein, in *Logistic Regression: A Self-Learning Text,* D. G. Kleinbaum, M. Klein, Eds. (Springer New York, New York, NY, 2010), pp. 103-127.
57. S. Choudhury *et al.*, Reconstructing Kinetic Models for Dynamical Studies of Metabolism using Generative Adversarial Networks. *Nature Machine Intelligence* **4**, 710-719 (2022).
58. L. Miskovic, M. Tokic, G. Fengos, V. Hatzimanikatis, Rites of passage: Requirements and standards for building kinetic models of metabolic phenotypes. *Current Opinion in Biotechnology* **36**, 146-153 (2015).
59. N. Chen, M. H. Bennett, C. Kontoravdi, Analysis of Chinese hamster ovary cell metabolism through a combined computational and experimental approach. *Cytotechnology* **66**, 945-966 (2014).
60. J. Robitaille, J. Chen, M. Jolicoeur, A single dynamic metabolic model can describe mAb producing CHO cell batch and fed-batch cultures on different culture media. *PLoS ONE* **10**, (2015).
61. D. Christodoulou *et al.*, Reserve Flux Capacity in the Pentose Phosphate Pathway Enables Escherichia coli's Rapid Response to Oxidative Stress. *Cell Systems* **6**, 569-578 (2018).
62. S. Diab, S. Badr, H. Sugiyama, D. I. Gerogiorgis, in *Computer Aided Chemical Engineering,* S. Pierucci, F. Manenti, G. L. Bozzano, D. Manca, Eds. (Elsevier, 2020), vol. 48, pp. 1657-1662.
63. W. Jones, D. I. Gerogiorgis, in *Computer Aided Chemical Engineering,* L. Montastruc, S. Negny, Eds. (Elsevier, 2022), vol. 51, pp. 1117-1122.
64. R. Mahadevan, J. S. Edwards, F. J. Doyle, Dynamic Flux Balance Analysis of diauxic growth in Escherichia coli. *Biophysical Journal* **83**, 1331-1340 (2002).
65. V. S. Martínez, M. Buchsteiner, P. Gray, L. K. Nielsen, L. E. Quek, Dynamic metabolic flux analysis using B-splines to study the effects of temperature shift on CHO cell metabolism. *Metabolic Engineering Communications* **2**, 46-57 (2015).
66. F. Scott, P. Wilson, R. Conejeros, V. S. Vassiliadis, Simulation and optimization of dynamic flux balance analysis models using an interior point method reformulation. *Computers & Chemical Engineering* **119**, 152-170 (2018).
67. R. P. Nolan, K. Lee, Dynamic model of CHO cell metabolism. *Metabolic Engineering* **13**, 108-124 (2011).
68. A. Provost, G. Bastin, Dynamic metabolic modelling under the balanced growth condition. *Journal of Process Control* **14**, 717-728 (2004).
69. A. Provost, G. Bastin, S. N. Agathos, Y. J. Schneider, Metabolic design of macroscopic bioreaction models: Application to Chinese hamster ovary cells. *Bioprocess and Biosystems Engineering* **29**, 349-366 (2006).
70. D. Machado, M. J. Herrgård, I. Rocha, Stoichiometric Representation of Gene–Protein–Reaction Associations Leverages Constraint-Based Analysis from Reaction to Gene-Level Phenotype Prediction. *PLOS Computational Biology* **12**, e1005140 (2016).
71. . (2022), vol. 2022.
72. Ž. Rejc *et al.*, Computational modelling of genome-scale metabolic networks and its application to CHO cell cultures. *Computers in Biology and Medicine* **88**, 150-160 (2017).
73. A. Conesa, S. Beck, Making multi-omics data accessible to researchers. *Scientific Data* **6**, 251 (2019).
74. I. Thiele, B. Palsson, A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nature Protocols* **5**, 93-121 (2010).

75. A. Walakira, D. Rozman, T. Režen, M. Mraz, M. Moškon, Guided extraction of genome-scale metabolic models for the integration and analysis of omics data. *Computational and Structural Biotechnology Journal* **19**, 3521-3530 (2021).
76. C. Gu, G. B. Kim, W. J. Kim, H. U. Kim, S. Y. Lee, Current status and applications of genome-scale metabolic models. *Genome Biology* **20**, 121 (2019).
77. H. Fouladiha *et al.*, Systematically gap-filling the genome-scale metabolic model of CHO cells. *Biotechnology Letters* **43**, 73-87 (2021).
78. J. Aminian-Dehkordi, S. M. Mousavi, A. Jafari, I. Mijakovic, S.-A. Marashi, Manually curated genome-scale reconstruction of the metabolic network of Bacillus megaterium DSM319. *Scientific Reports* **9**, 18762 (2019).
79. I. Thiele, N. Vlassis, R. M. Fleming, fastGapFill: efficient gap filling in metabolic networks. *Bioinformatics* **30**, 2529-2531 (2014).
80. V. Satish Kumar, M. S. Dasika, C. D. Maranas, Optimization based automated curation of metabolic reconstructions. *BMC Bioinformatics* **8**, 212 (2007).
81. Z. Hosseini, S.-A. Marashi, Discovering missing reactions of metabolic networks by using gene co-expression data. *Scientific Reports* **7**, 41774 (2017).
82. M. Kanehisa, S. Goto, KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research* **28**, 27-30 (2000).
83. R. Caspi *et al.*, MetaCyc: a multiorganism database of metabolic pathways and enzymes. *Nucleic Acids Res* **34**, D511-516 (2006).
84. D. J. Lugar, S. G. Mack, G. Sriram, NetRed, an algorithm to reduce genome-scale metabolic networks and facilitate the analysis of flux predictions. *Metabolic Engineering* **65**, 207-222 (2021).
85. S. Jonnalagadda, R. Srinivasan, An efficient graph theory based method to identify every minimal reaction set in a metabolic network. *BMC Systems Biology* **8**, 28 (2014).
86. M. Ataman, D. F. Hernandez Gardiol, G. Fengos, V. Hatzimanikatis, redGEM: Systematic reduction and analysis of genome-scale metabolic reconstructions for development of consistent core metabolic models. *PLoS Computational Biology* **13**, 1-22 (2017).
87. M. Ataman, V. Hatzimanikatis, lumpGEM: Systematic generation of subnetworks and elementally balanced lumped reactions for the biosynthesis of target metabolites. *PLoS Computational Biology* **13**, 1-21 (2017).
88. N. Vlassis, M. P. Pacheco, T. Sauter, Fast Reconstruction of Compact Context-Specific Metabolic Network Models. *PLOS Computational Biology* **10**, e1003424 (2014).
89. P. Erdrich, R. Steuer, S. Klamt, An algorithm for the reduction of genome-scale metabolic network models to meaningful core models. *BMC Systems Biology* **9**, 48 (2015).
90. S. A. Becker, B. O. Palsson, Context-specific metabolic networks are consistent with experiments. *PLoS Comput Biol* **4**, e1000082 (2008).
91. R. Mahadevan, C. H. Schilling, The effects of alternate optimal solutions in constraint-based genome-scale metabolic models. *Metab Eng* **5**, 264-276 (2003).
92. M. Lularevic, A. J. Racher, C. Jaques, A. Kiparissides, Improving the accuracy of flux balance analysis through the implementation of carbon availability constraints for intracellular reactions. *Biotechnol Bioeng* **116**, 2339-2352 (2019).
93. A. Antonakoudis, B. Strain, R. Barbosa, I. Jimenez del Val, C. Kontoravdi, Synergising stoichiometric modelling with artificial neural networks to predict antibody glycosylation patterns in Chinese hamster ovary cells. *Computers & Chemical Engineering* **154**, (2021).
94. C. S. Henry, L. J. Broadbelt, V. Hatzimanikatis, Thermodynamics-based metabolic flux analysis. *Biophysical Journal* **92**, 1792-1805 (2007).
95. W. Guo, X. Feng, OM-FBA: Integrate Transcriptomics Data with Flux Balance Analysis to Decipher the Cell Metabolism. *PLoS One* **11**, e0154188 (2016).
96. H. C. Yeo, J. Hong, M. Lakshmanan, D.-Y. Lee, Enzyme capacity-based genome scale modelling of CHO cells. *Metabolic Engineering* **60**, 138-147 (2020).
97. N. E. Lewis *et al.*, Omic data from evolved E. coli are consistent with computed optimal growth from genome-scale models. *Molecular Systems Biology* **6**, 390 (2010).
98. J. S. Edwards, B. O. Palsson, Systems Properties of the Haemophilus influenzaeRd Metabolic Genotype*. *Journal of Biological Chemistry* **274**, 17410-17416 (1999).

99. A. M. Feist, B. O. Palsson, The growing scope of applications of genome-scale metabolic reconstructions using Escherichia coli. *Nat Biotechnol* **26**, 659-667 (2008).
100. Y. Kuriya, M. Araki, Dynamic Flux Balance Analysis to Evaluate the Strain Production Performance on Shikimic Acid Production in Escherichia coli. *Metabolites* **10**, (2020).
101. M. Zampieri *et al.*, Metabolic constraints on the evolution of antibiotic resistance. *Molecular Systems Biology* **13**, (2017).
102. C. Iacometti *et al.*, Activating silent glycolysis bypasses in <em>Escherichia coli</em&gt. *bioRxiv*, 2021.2011.2018.468982 (2021).
103. J. R. Valverde, S. Gullón, C. A. García-Herrero, I. Campoy, R. P. Mellado, Dynamic metabolic modelling of overproduced protein secretion in Streptomyces lividans using adaptive DFBA. *BMC Microbiology* **19**, 233 (2019).
104. M. Suastegui *et al.*, Multilevel engineering of the upstream module of aromatic amino acid biosynthesis in Saccharomyces cerevisiae for high production of polymer and drug precursors. *Metab Eng* **42**, 134-144 (2017).
105. P. Torres, P. A. Saa, J. Albiol, P. Ferrer, E. Agosin, Contextualized genome-scale model unveils high-order metabolic effects of the specific growth rate and oxygenation level in recombinant Pichia pastoris. *Metabolic Engineering Communications* **9**, e00103 (2019).
106. R. R. Stanway *et al.*, Genome-Scale Identification of Essential Metabolic Processes for Targeting the Plasmodium Liver Stage. *Cell* **179**, 1112-1128.e1126 (2019).
107. A. B. Wegrzyn *et al.*, Fibroblast-specific genome-scale modelling predicts an imbalance in amino acid metabolism in Refsum disease. *The FEBS Journal* **287**, 5096-5113 (2020).
108. H. Hefzi *et al.*, A Consensus Genome-scale Reconstruction of Chinese Hamster Ovary Cell Metabolism. *Cell Systems* **3**, 434-443.e438 (2016).
109. S. Selvarasu *et al.*, Combined in silico modeling and metabolomics analysis to characterize fed-batch CHO cell culture. *Biotechnology and Bioengineering* **109**, 1415-1429 (2012).
110. J. M. Gutierrez *et al.*, Genome-scale reconstructions of the mammalian secretory pathway predict metabolic costs and limitations of protein secretion. *Nat Commun* **11**, 68 (2020).
111. C. Calmels, A. McCann, L. Malphettes, M. R. Andersen, Application of a curated genome-scale metabolic model of CHO DG44 to an industrial fed-batch process. *Metab Eng* **51**, 9-19 (2019).
112. S. Kol *et al.*, Multiplex secretome engineering enhances recombinant protein production and purity. *Nat Commun* **11**, 1908 (2020).
113. Z. Huang *et al.*, CHO cell productivity improvement by genome-scale modeling and pathway analysis: Application to feed supplements. *Biochemical Engineering Journal* **160**, 107638 (2020).
114. Y. Lim *et al.*, Engineering mammalian cells in bioprocessing – current achievements and future perspectives. *Biotechnology and Applied Biochemistry* **55**, 175-189 (2010).
115. J. Y. Kim, Y.-G. Kim, G. M. Lee, CHO cells in biotechnology for production of recombinant proteins: current state and further potential. *Applied Microbiology and Biotechnology* **93**, 917-930 (2012).
116. M. J. Allen *et al.*, Identification of novel small molecule enhancers of protein production by cultured mammalian cells. *Biotechnology and Bioengineering* **100**, 1193-1204 (2008).
117. J. Zhang, D. Robinson, P. Salmon, 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**, 1188-1197 (2006).
118. J. Xu *et al.*, Improving titer while maintaining quality of final formulated drug substance via optimization of CHO cell culture conditions in low-iron chemically defined media. *mAbs* **10**, 488-499 (2018).
119. Z. Xing *et al.*, Optimizing amino acid composition of CHO cell culture media for a fusion protein production. *Process Biochemistry* **46**, 1423-1429 (2011).
120. B. A. Pérez-Fernández, J. Fernandez-de-Cossio-Diaz, T. Boggiano, K. León, R. Mulet, In-silico media optimization for continuous cultures using genome scale metabolic networks: The case of CHO-K1. *Biotechnology and Bioengineering* **118**, 1884-1897 (2021).
121. Z. Huang *et al.*, CHO cell productivity improvement by genome-scale modeling and pathway analysis: Application to feed supplements. *Biochemical Engineering Journal*, (2020).

122. H. Fouladiha *et al.*, A metabolic network-based approach for developing feeding strategies for CHO cells to increase monoclonal antibody production. *Bioprocess Biosyst Eng* **43**, 1381-1389 (2020).
123. Y. Chen *et al.*, A genome-scale nutrient minimization forecast algorithm for controlling essential amino acid levels in CHO cell cultures. *Biotechnology and Bioengineering* **119**, 435-451 (2022).
124. A. R. Costa, M. E. Rodrigues, M. Henriques, R. Oliveira, J. Azeredo, in *Animal Cell Biotechnology: Methods and Protocols,* R. Pörtner, Ed. (Humana Press, Totowa, NJ, 2014), pp. 105-116.
125. S. Kyriakopoulos, C. Kontoravdi, A framework for the systematic design of fed-batch strategies in mammalian cell culture. *Biotechnology and Bioengineering* **111**, 2466-2476 (2014).
126. S. W. Harcum *et al.*, PID controls: the forgotten bioprocess parameters. *Discover Chemical Engineering* **2**, 1 (2022).
127. S. N. Sou *et al.*, How does mild hypothermia affect monoclonal antibody glycosylation? *Biotechnol Bioeng* **112**, 1165-1176 (2015).
128. M. A. MacDonald *et al.*, Engineering death resistance in CHO cells for improved perfusion culture. *mAbs* **14**, 2083465 (2022).
129. J. Capiaumont *et al.*, Methods for reducing the ammonia in hybridoma cell cultures. *Journal of Biotechnology* **39**, 49-58 (1995).
130. S. L. Bell *et al.*, Genetic engineering of hybridoma glutamine metabolism. *Enzyme and Microbial Technology* **17**, 98-106 (1995).
131. M. I. Cockett, C. R. Bebbington, G. T. Yarranton, High Level Expression of Tissue Inhibitor of Metalloproteinases in Chinese Hamster Ovary Cells Using Glutamine Synthetase Gene Amplification. *Bio/Technology* **8**, 662-667 (1990).
132. S. Y. Lee, H. U. Kim, Systems strategies for developing industrial microbial strains. *Nature Biotechnology* **33**, 1061-1072 (2015).
133. O. D. Kim, M. Rocha, P. Maia, A Review of Dynamic Modeling Approaches and Their Application in Computational Strain Optimization for Metabolic Engineering. *Frontiers in Microbiology* **9**, (2018).
134. A. M. Davy, H. F. Kildegaard, M. R. Andersen, Cell Factory Engineering. *Cell Systems* **4**, 262-275 (2017).
135. S. Landon, J. Rees-Garbutt, L. Marucci, C. Grierson, Genome-driven cell engineering review: in vivo and in silico metabolic and genome engineering. *Essays in Biochemistry* **63**, 267-284 (2019).
136. C. D. Maranas, A. R. Zomorrodi, *Optimization Methods in Metabolic Networks*. (John Wiley & Sons Inc., Hoboken, New Jersey, ed. First, 2016).
137. A. P. Burgard, P. Pharkya, C. D. Maranas, Optknock: a bilevel programming framework for identifying gene knockout strategies for microbial strain optimization. *Biotechnol Bioeng* **84**, 647-657 (2003).
138. P. Pharkya, C. D. Maranas, An optimization framework for identifying reaction activation/inhibition or elimination candidates for overproduction in microbial systems. *Metab Eng* **8**, 1-13 (2006).
139. S. Ranganathan, P. F. Suthers, C. D. Maranas, OptForce: An optimization procedure for identifying all genetic manipulations leading to targeted overproductions. *PLoS Computational Biology* **6**, (2010).
140. A. Chowdhury, A. R. Zomorrodi, C. D. Maranas, k-OptForce: Integrating Kinetics with Flux Balance Analysis for Strain Design. *PLoS Computational Biology* **10**, (2014).
141. J. Kim, J. L. Reed, C. T. Maravelias, Large-Scale Bi-Level Strain Design Approaches and Mixed-Integer Programming Solution Techniques. *PLOS ONE* **6**, e24162 (2011).
142. K. Jensen, V. Broeken, A. S. L. Hansen, N. Sonnenschein, M. J. Herrgard, OptCouple: Joint simulation of gene knockouts, insertions and medium modifications for prediction of growth-coupled strain designs. *Metab Eng Commun* **8**, e00087 (2019).
143. N. Tepper, T. Shlomi, Predicting metabolic engineering knockout strategies for chemical production: accounting for competing pathways. *Bioinformatics* **26**, 536-543 (2010).

144. T. B. Alter, B. E. Ebert, Determination of growth-coupling strategies and their underlying principles. *BMC Bioinformatics* **20**, 447 (2019).
145. K. R. Patil, I. Rocha, J. Förster, J. Nielsen, Evolutionary programming as a platform for in silico metabolic engineering. *BMC Bioinformatics* **6**, 308 (2005).
146. P. Pharkya, A. P. Burgard, C. D. Maranas, OptStrain: a computational framework for redesign of microbial production systems. *Genome Res* **14**, 2367-2376 (2004).
147. J. Kim, J. L. Reed, OptORF: Optimal metabolic and regulatory perturbations for metabolic engineering of microbial strains. *BMC Syst Biol* **4**, (2010,).
148. F. Shen *et al.*, OptRAM: In-silico strain design via integrative regulatory-metabolic network modeling. *PLoS Comput Biol* **15**, e1006835 (2019).
149. S. Ranganathan *et al.*, An integrated computational and experimental study for overproducing fatty acids in Escherichia coli. *Metabolic Engineering* **14**, 687-704 (2012).
150. Z. Tan *et al.*, Engineering of E. coli inherent fatty acid biosynthesis capacity to increase octanoic acid production. *Biotechnol Biofuels* **11**, 87 (2018).
151. P. Xu, S. Ranganathan, Z. L. Fowler, C. D. Maranas, M. A. G. Koffas, Genome-scale metabolic network modeling results in minimal interventions that cooperatively force carbon flux towards malonyl-CoA. *Metabolic Engineering* **13**, 578-587 (2011).
152. J. A. Wayman, C. Glasscock, T. J. Mansell, M. P. DeLisa, J. D. Varner, Improving designer glycan production in Escherichia coli through model-guided metabolic engineering. *Metabolic Engineering Communications* **9**, e00088 (2019).
153. M. Suástegui *et al.*, Multilevel engineering of the upstream module of aromatic amino acid biosynthesis in Saccharomyces cerevisiae for high production of polymer and drug precursors. *Metabolic Engineering* **42**, 134-144 (2017).
154. K. Paramasivan, A. Abdulla, N. Gupta, S. Mutturi, In silico target-based strain engineering of Saccharomyces cerevisiae for terpene precursor improvement. *Integrative Biology* **14**, 25-36 (2022).
155. L. E. Quek, S. Dietmair, J. O. Krömer, L. K. Nielsen, Metabolic flux analysis in mammalian cell culture. *Metabolic Engineering* **12**, 161-171 (2010).
156. J. Schellenberger, Creation and analysis of biochemical constraint-based models: the COBRA Toolbox v3.0. *Language*, (2018).
157. E. Rowe, B. O. Palsson, Z. A. King, Escher-FBA: a web application for interactive flux balance analysis. *BMC Systems Biology* **12**, 84 (2018).
158. J. P. Ignizio, T. M. Cavalier, *Linear programming*. (Prentice-Hall, Inc., 1994), pp. 666.
159. C. A. Floudas, *Nonlinear and Mixed-Integer Optimization: Fundamentals and Applications*. (Oxford Univ. Press, New York, 1995).
160. A. Goelzer, V. Fromion, Resource allocation in living organisms. *Biochemical Society Transactions* **45**, 945-952 (2017).
161. H. Zeng, R. Rohani, W. E. Huang, A. Yang, Understanding and mathematical modelling of cellular resource allocation in microorganisms: a comparative synthesis. *BMC Bioinformatics* **22**, 467 (2021).
162. S. Hui *et al.*, Quantitative proteomic analysis reveals a simple strategy of global resource allocation in bacteria. *Molecular Systems Biology* **11**, 784 (2015).
163. M. Scott, C. W. Gunderson, E. M. Mateescu, Z. Zhang, T. Hwa, Interdependence of Cell Growth and Gene Expression: Origins and Consequences. *Science* **330**, 1099-1102 (2010).
164. S. Müller, D. Széliová, J. Zanghellini, Elementary vectors and autocatalytic sets for resource allocation in next-generation models of cellular growth. *PLOS Computational Biology* **18**, e1009843 (2022).
165. P. T. Man Bock Gu, Dhinakar S. Kompala, Metabolic burden in recombinant CHO cells: effect of dhfr gene amplification and lacZ expression. *Cytotechnology* **18**, 159-166 (1996).
166. S. Kyriakopoulos, (2014).
167. R. M. F. Stephen F. Gorfien, Joyce L. Dzimian, Glenn P. Godwin, Paul J. Price, David A. Epstein, Dale Gruber, Don Mcclure, E. P. SPECIFICATION, Ed. (1997).
168. K. Sheikh, J. Förster, L. K. Nielsen, Modeling Hybridoma Cell Metabolism Using a Generic Genome-Scale Metabolic Model of Mus musculus. *Biotechnology Progress* **21**, 112-121 (2005).

169. C. Clarke *et al.*, Predicting cell-specific productivity from CHO gene expression. *Journal of Biotechnology* **151**, 159-165 (2011).
170. S. Kyriakopoulos, K. M. Polizzi, C. Kontoravdi, Comparative analysis of amino acid metabolism and transport in CHO variants with different levels of productivity. *Journal of Biotechnology* **168**, 543-551 (2013).
171. P. M. Hayter, N. F. Kirkby, R. E. Spier, Relationship between hybridoma growth and monoclonal antibody production. *Enzyme and Microbial Technology* **14**, 454-461 (1992).
172. J. Hagedorn, F. Kargi, Coiled tube membrane bioreactor for cultivation of hybridoma cells producing monoclonal antibodies. *Enzyme and Microbial Technology* **12**, 824-829 (1990).
173. G. M. Lee, A. Varma, B. O. Palsson, Production of monoclonal antibody using free-suspended and immobilized hybridoma cells: Effect of serum. *Biotechnology and Bioengineering* **38**, 821-830 (1991).
174. B. T. Tey, M. Al-Rubeai, Effect of Bcl-2 overexpression on cell cycle and antibody productivity in chemostat cultures of myeloma NS0 cells. *Journal of Bioscience and Bioengineering* **100**, 303-310 (2005).
175. M. Vergara *et al.*, Differential Effect of Culture Temperature and Specific Growth Rate on CHO Cell Behavior in Chemostat Culture. *PLOS ONE* **9**, e93865 (2014).
176. M. Vergara *et al.*, High glucose and low specific cell growth but not mild hypothermia improve specific r-protein productivity in chemostat culture of CHO cells. *PLOS ONE* **13**, e0202098 (2018).
177. A. Navid, E. Almaas, Genome-level transcription data of Yersinia pestis analyzed with a New metabolic constraint-based approach. *BMC Systems Biology* **6**, 150 (2012).
178. R. Chowdhury, A. Chowdhury, C. D. Maranas, Using Gene Essentiality and Synthetic Lethality Information to Correct Yeast and CHO Cell Genome-Scale Models. *Metabolites* **5**, 536-570 (2015).
179. S. Hammond, J. C. Swanberg, M. Kaplarevic, K. H. Lee, Genomic sequencing and analysis of a Chinese hamster ovary cell line using Illumina sequencing technology. *BMC Genomics* **12**, 67 (2011).
180. N. E. Lewis *et al.*, Genomic landscapes of Chinese hamster ovary cell lines as revealed by the Cricetulus griseus draft genome. *Nature Biotechnology* **31**, 759-765 (2013).
181. D. Tian *et al.*, Identifying mouse developmental essential genes using machine learning. *Dis Model Mech* **11**, (2018).
182. M. Kabir, A. Barradas, G. T. Tzotzos, K. E. Hentges, A. J. Doig, Properties of genes essential for mouse development. *PLoS One* **12**, e0178273 (2017).
183. K. Xiong *et al.*, An optimized genome-wide, virus-free CRISPR screen for mammalian cells. *Cell Reports Methods* **1**, (2021).
184. J. D. Orth *et al.*, A comprehensive genome-scale reconstruction of Escherichia coli metabolism—2011. *Molecular Systems Biology* **7**, 535 (2011).
185. C. A. Sellick, in *Genetic Engineering & Biotechnology News*. (2009), vol. 29.
186. C. H. Goey, D. Bell, C. Kontoravdi, CHO cell cultures in shake flasks and bioreactors present different host cell protein profiles in the supernatant. *Biochemical Engineering Journal* **144**, 185-192 (2019).
187. S. Wulhfard *et al.*, Mild Hypothermia Improves Transient Gene Expression Yields Several Fold in Chinese Hamster Ovary Cells. *Biotechnology Progress* **24**, 458-465 (2008).
188. E. Brunk *et al.*, Recon3D enables a three-dimensional view of gene variation in human metabolism. *Nat Biotechnol* **36**, 272-281 (2018).
189. K. Labun *et al.*, CHOPCHOP v3: expanding the CRISPR web toolbox beyond genome editing. *Nucleic Acids Research* **47**, W171-W174 (2019).
190. C. Ronda *et al.*, Accelerating genome editing in CHO cells using CRISPR Cas9 and CRISPy, a web-based target finding tool. *Biotechnology and Bioengineering* **111**, 1604-1616 (2014).
191. S. Pereira *et al.*, BCAT1 and BCAT2 disruption in CHO cells has cell line-dependent effects. *Journal of Biotechnology* **306**, 24-31 (2019).
192. R. M. Gupta, K. Musunuru, Expanding the genetic editing tool kit: ZFNs, TALENs, and CRISPR-Cas9. *The Journal of Clinical Investigation* **124**, 4154-4161 (2014).

193. T. Gaj, J. Guo, Y. Kato, S. J. Sirk, C. F. Barbas, Targeted gene knockout by direct delivery of zinc-finger nuclease proteins. *Nature Methods* **9**, 805-807 (2012).
194. F. D. Urnov, E. J. Rebar, M. C. Holmes, H. S. Zhang, P. D. Gregory, Genome editing with engineered zinc finger nucleases. *Nature Reviews Genetics* **11**, 636-646 (2010).
195. J. C. Miller *et al.*, A TALE nuclease architecture for efficient genome editing. *Nature Biotechnology* **29**, 143-148 (2011).
196. S. Jain *et al.*, TALEN outperforms Cas9 in editing heterochromatin target sites. *Nature Communications* **12**, 606 (2021).
197. M. Jinek *et al.*, A Programmable Dual-RNA–Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science* **337**, 816-821 (2012).
198. Y. Fu, J. D. Sander, D. Reyon, V. M. Cascio, J. K. Joung, Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. *Nature Biotechnology* **32**, 279-284 (2014).
199. M. Gearing. (https://blog.addgene.org/crispr-101-cas9-nickase-design-and-homology-directed-repair, 2018), vol. 2022.
200. F. A. Ran *et al.*, Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. *Cell* **154**, 1380-1389 (2013).
201. S. Liu *et al.*, Neuronal Cell-type Engineering by Transcriptional Activation. *Frontiers in Genome Editing* **3**, (2021).
202. T. K. Kim, J. H. Eberwine, Mammalian cell transfection: the present and the future. *Analytical and Bioanalytical Chemistry* **397**, 3173-3178 (2010).
203. S. L. Holmen, M. W. Vanbrocklin, R. R. Eversole, S. R. Stapleton, L. C. Ginsberg, Efficient lipid-mediated transfection of DNA into primary rat hepatocytes. *In Vitro Cellular & Developmental Biology - Animal* **31**, 347-351 (1995).
204. C. Ehrhardt *et al.*, Polyethylenimine, a cost-effective transfection reagent. *Signal Transduction* **6**, 179-184 (2006).
205. R. Maurisse *et al.*, Comparative transfection of DNA into primary and transformed mammalian cells from different lineages. *BMC Biotechnology* **10**, 9 (2010).
206. S. N. Sou *et al.*, Exploring cellular behavior under transient gene expression and its impact on mAb productivity and Fc-glycosylation. *Biotechnology and Bioengineering* **115**, 512-518 (2018).
207. X. Xu *et al.*, The genomic sequence of the Chinese hamster ovary (CHO)-K1 cell line. *Nat Biotechnol* **29**, 735-741 (2011).
208. M. C. Gundry *et al.*, Technical considerations for the use of CRISPR/Cas9 in hematology research. *Experimental Hematology* **54**, 4-11 (2017).
209. Y. Yamanaka, in *The Tumor Microenvironment: Methods and Protocols,* J. Ursini-Siegel, N. Beauchemin, Eds. (Springer New York, New York, NY, 2016), pp. 261-271.
210. M. Aouida, D. Aljogol, R. Ali, D. Ramotar, A simple protocol to isolate a single human cell PRDX1 knockout generated by CRISPR-Cas9 system. *STAR Protocols* **3**, 101216 (2022).
211. M. W. Popp, L. E. Maquat, Leveraging Rules of Nonsense-Mediated mRNA Decay for Genome Engineering and Personalized Medicine. *Cell* **165**, 1319-1322 (2016).
212. C. V. Campenhout *et al.*, Guidelines for optimized gene knockout using CRISPR/Cas9. *BioTechniques* **66**, 295-302 (2019).
213. V. A. Truong *et al.*, CRISPRai for simultaneous gene activation and inhibition to promote stem cell chondrogenesis and calvarial bone regeneration. *Nucleic Acids Res* **47**, e74 (2019).
214. M. Prabhune. (2021), vol. 2022.
215. B. G. Kremkow, J. Y. Baik, M. L. MacDonald, K. H. Lee, CHOgenome.org 2.0: Genome resources and website updates. *Biotechnology Journal* **10**, 931-938 (2015).
216. T. Sun *et al.*, Development of a Proline-Based Selection System for Reliable Genetic Engineering in Chinese Hamster Ovary Cells. *ACS Synthetic Biology* **9**, 1864-1872 (2020).
217. J. D. Budge *et al.*, A proline metabolism selection system and its application to the engineering of lipid biosynthesis in Chinese hamster ovary cells. *Metabolic Engineering Communications* **13**, e00179 (2021).
218. C. H. Goey, J. M. H. Tsang, D. Bell, C. Kontoravdi, Cascading effect in bioprocessing—The impact of mild hypothermia on CHO cell behavior and host cell protein composition. *Biotechnology and Bioengineering* **114**, 2771-2781 (2017).

219. M. Torres *et al.*, Mild hypothermia upregulates myc and xbp1s expression and improves anti-TNFα production in CHO cells. *PLOS ONE* **13**, e0194510 (2018).
220. M. T. Alam *et al.*, Metabolic modeling and analysis of the metabolic switch in Streptomyces coelicolor. *BMC Genomics* **11**, 202 (2010).
221. N. E. Lewis, H. Nagarajan, B. O. Palsson, Constraining the metabolic genotype-phenotype relationship using a phylogeny of in silico methods. *Nat Rev Microbiol* **10**, 291-305 (2012).
222. B. C. Mulukutla *et al.*, Metabolic engineering of Chinese hamster ovary cells towards reduced biosynthesis and accumulation of novel growth inhibitors in fed-batch cultures. *Metab Eng*, (2019).

Appendix

Table A1: Lower and upper bounds for each of the four experimental datasets. The units of the values are mmol/gDCW/cell. Each of the metabolite is associated to a drain reaction. These drain reactions are then constrained with the values of the table.

No	Metabolite/Reaction	D1		D2		D3		D4	
		LB	UB	LB	UB	LB	UB	LB	UB
1	L-aspartic acid	-0.039	-0.015	-0.061	-0.016	-0.011	-0.002	-0.011	0.001
2	L-asparagine	-0.075	-0.005	-0.098	-0.003	-0.092	-0.034	-0.143	-0.089
3	L-glutamine	-0.003	0.008	0.001	0.006	0.004	0.012	0.002	0.017
4	L-glutamic acid	-0.011	-0.003	-0.016	-0.003	-0.017	-0.012	-0.032	-0.010
5	Ammonium	-0.010	0.027	-0.010	0.041	0.017	0.046	0.019	0.046
6	L-arginine	-0.010	-0.005	-0.016	-0.002	-0.010	-0.004	-0.014	-0.005
7	L-serine	-0.051	-0.016	-0.065	-0.014	-0.038	-0.017	-0.062	-0.014
8	L-histidine	-0.004	0.000	-0.008	-0.001	-0.005	-0.001	-0.014	-0.001
9	L-proline	-0.015	-0.004	-0.024	-0.003	-0.021	-0.002	-0.019	-0.005
10	Glycine	0.003	0.007	0.000	0.022	-0.003	0.012	-0.001	0.014
11	L-alanine	0.007	0.030	0.007	0.019	0.004	0.021	0.009	0.029
12	L-threonine	-0.013	-0.005	-0.020	-0.004	-0.011	-0.005	-0.019	-0.009
13	L-tyrosine	-0.006	-0.004	-0.009	-0.002	-0.008	-0.003	-0.009	-0.003
14	'L-cysteine'	-0.001	0.002	-0.003	0.001	-0.003	0.001	-0.003	0.001
15	L-valine	-0.027	-0.008	-0.034	-0.006	-0.026	-0.005	-0.019	-0.006
16	L-methionine	-0.006	-0.001	-0.007	-0.001	-0.009	-0.001	-0.007	-0.001
17	L-isoleucine	-0.014	-0.007	-0.022	-0.003	-0.021	-0.002	-0.022	-0.003
18	L-leucine	-0.037	-0.013	-0.047	-0.009	-0.039	-0.006	-0.027	-0.010
19	L-lysine	-0.016	-0.007	-0.024	-0.004	-0.034	-0.002	-0.022	-0.008
20	L-phenylananine	-0.007	-0.003	-0.012	-0.002	-0.013	-0.002	-0.010	-0.003
21	L-tryptophan	-0.003	-0.001	-0.005	-0.001	-0.016	0.000	-0.013	-0.002
22	(S)-lactate	-0.036	0.142	-0.029	0.276	-0.035	0.242	-0.065	0.198
23	D-glucose	-0.220	-0.004	-0.279	-0.122	-0.240	-0.138	-0.313	-0.132
24	hydroxy-L-proline	-0.264	10.000	-0.335	10.000	-0.287	10.000	-0.375	10.000
25	pyridoxine-HCl	0.000	10.000	0.000	10.000	0.000	10.000	0.000	10.000
26	Thiamin	-0.110	10.000	-0.139	10.000	-0.120	10.000	-0.156	10.000
27	Oxidized glutathione	-0.132	10.000	-0.167	10.000	-0.144	10.000	-0.188	10.000
28	sulfate	-10.000	10.000	-10.000	10.000	-10.000	10.000	-10.000	10.000
29	L-ascorbate	-0.220	10.000	-0.279	10.000	-0.240	10.000	-0.313	10.000
30	Folate	0.000	10.000	0.000	10.000	0.000	10.000	0.000	10.000
31	Riboflavin	-0.078	10.000	-0.098	10.000	-0.085	10.000	-0.110	10.000

32	Biotin	0.000	10.000	0.000	10.000	0.000	10.000	0.000	10.000
33	Choline	-0.022	10.000	-0.022	10.000	-0.022	10.000	-0.022	10.000
34	Nicotinamide	-0.220	10.000	-0.279	10.000	-0.240	10.000	-0.313	10.000
35	myo-inositol	-0.002	10.000	-0.002	10.000	-0.002	10.000	-0.002	10.000
36	Pyruvate	-0.049	10.000	-0.001	0.003	-0.001	0.004	-0.002	0.008
37	spermine(4+)	-0.132	10.000	-0.167	10.000	-0.144	10.000	-0.188	10.000
38	Bicarbonate	-1.318	10.000	-1.673	10.000	-1.437	10.000	-1.876	10.000
39	acetate	0.000	10.000	0.000	10.000	0.000	10.000	0.000	10.000
40	proton	-10.000	10.000	-10.000	10.000	-10.000	10.000	-10.000	10.000
41	Water	-10.000	10.000	-10.000	10.000	-10.000	10.000	-10.000	10.000
42	Oxygen	-10.000	10.000	-10.000	10.000	-10.000	10.000	-10.000	10.000
43	hydrogenphosphate	-10.000	10.000	-10.000	10.000	-10.000	10.000	-10.000	10.000
44	sulfate	-10.000	10.000	-10.000	10.000	-10.000	10.000	-10.000	10.000
45	protein-linked asparagine residue (N-glycosylation site)	-0.100	10.000	-0.100	10.000	-0.100	10.000	-0.100	10.000
46	protein-linked serine or threonine residue (O-glycosylation site)	-0.100	10.000	-0.100	10.000	-0.100	10.000	-0.100	10.000
47	Tyr-194 of apo-glycogenin protein (primer for glycogen synthesis)	-0.100	10.000	-0.100	10.000	-0.100	10.000	-0.100	10.000
48	ATP Maintenance	8.600	10.000	8.600	10.000	8.600	10.000	8.600	10.000

Table A2: The 20 high-producing scenarios, calculated after the application of SA on the results of the optimisation algorithm. Each table represent a scenario that was resulted by the optimisation of the CHO-K1 GEM. The first row of the table contains the scenario No, the type of algorithm used, the feeding used to constrain the GEM, the total number of regulations, the upregulation coefficient and the predicted q_{mAb} and μ by the FBA optimisation of the modified GEM. The following rows in each of the tables show the reactions that are regulated. The IDs of the metabolites can be found in the BiGG database.

Scenario	**Algorithm**	**Feed**	**Regulations**	**e_up**	**q_{mAb} (pg/cell/day)**	**μ (hr^{-1})**
1	'Bio_Igg'	'D1'	8	0.5	21.231	0.007
UP						
'h[c] -> h[x] '						
'ppi[c] -> ppi[x] '						
'2 h[c] + pi[c] -> 2 h[m] + pi[m] '						

'nad[c] -> nad[x] '
KO
'nad[c] + pi[c] + g3p[c] -> h[c] + nadh[c] + 13dpg[c] '
'h2o[m] + ppi[m] -> h[m] + 2 pi[m] '
'34hpp[c] -> 34hpp[e] '
'h[x] -> h[c] '

Scenario	**Algorithm**	**Feed**	**Regulations**	**e_up**	**q_{mAb} (pg/cell/day)**	**μ (hr^{-1})**
3	'Bio_Igg'	'D1'	4	0.6	24.723	0.006
UP						
'h2o[m] + ppi[m] -> h[m] + 2 pi[m] '						
KO						
'nad[c] + h2o[c] + glx[c] -> 2 h[c] + nadh[c] + oxa[c] '						
'gua[c] -> gua[e] '						
'34hpp[c] -> 34hpp[e] '						

Scenario	**Algorithm**	**Feed**	**Regulations**	**e_up**	**q_{mAb} (pg/cell/day)**	**μ (hr^{-1})**
9	'Igg_Igg'	'D1'	10	0.9	28.925	0.005
UP						
'atp[c] + dump[c] -> adp[c] + dudp[c] '						
'r1p[c] -> r5p[c] '						
'nad[c] + h2o[c] + imp[c] -> h[c] + nadh[c] + xmp[c] '						
'mal_L[m] -> h2o[m] + fum[m] '						
KO						
'2 accoa[m] -> coa[m] + aacoa[m] '						
'34hpp[c] -> 34hpp[e] '						
'nadp[x] -> nadp[c] '						
'ppi[c] -> ppi[m] '						
'pro_L[c] -> pro_L[m] '						
'akg[m] + ala_L[m] -> glu_L[m] + pyr[m] '						
'h[x] + nadh[x] + dhap[x] -> nad[x] + glyc3p[x] '						

Scenario	**Algorithm**	**Feed**	**Regulations**	**e_up**	**q_{mAb} (pg/cell/day)**	**μ (hr^{-1})**
12	'Igg_Igg'	'D1'	5	1	8.4770	0.01

UP
'r1p[c] -> r5p[c] '
KO
'o2[c] + h2o[c] + sprm[c] -> h2o2[c] + spmd[c] + bamppald[c] '
'h[c] + pi[c] -> h[m] + pi[m] '
'nad[c] + pi[c] + g3p[c] -> h[c] + nadh[c] + 13dpg[c] '
'34hpp[c] -> 34hpp[e] '

Scenario	**Algorithm**	**Feed**	**Regulations**	**e_up**	**q_{mAb} (pg/cell/day)**	**μ (hr^{-1})**
14	'Bio_Igg'	'D4'	6	0.9	61.038	0.013000
UP						
'acgam6p[c] -> acgam1p[c] '						
'uri[e] -> uri[c] '						
'glu5sa[m] -> glu5sa[c] '						
KO						
'glu5sa[c] -> glu5sa[m] '						
'ppi[r] -> ppi[c] '						
'pi[c] + ppi[r] -> ppi[c] + pi[r] '						

Scenario	**Algorithm**	**Feed**	**Regulations**	**e_up**	**q_{mAb} (pg/cell/day)**	**μ (hr^{-1})**
27	'Bio_Igg'	'D1'	8	0.7	25.471	0.006
UP						
'2 hco3[c] + cl[e] -> 2 hco3[e] + cl[c] '						
'h[c] + pi[c] -> h[m] + pi[m] '						
'fadh2[m] + q10[m] -> fad[m] + q10h2[m] '						
'k[c] -> k[e] '						
KO						
'coa[c] + atp[c] + acac[c] -> amp[c] + ppi[c] + aacoa[c] '						
'34hpp[c] -> 34hpp[e] '						
'h[c] + k[e] -> h[e] + k[c] '						
'co2[c] + akg[c] -> h[c] + HC01434[c] '						
'cl[c] -> cl[e] '						

Scenario	Algorithm	Feed	Regulations	e_up	q_{mAb} (pg/cell/day)	μ (hr^{-1})
30	'Igg_Igg'	'D2'	13	0.8	63.644	0.01
UP						
'coa[m] + atp[m] + succ[m] -> adp[m] + succoa[m] + pi[m] '						
'glu5sa[m] -> glu5sa[c] '						
'h[c] -> h[x] '						
'R1coa_cho[c] -> Rtotalcoa[c] '						
'h[c] + lac_L[c] -> h[x] + lac_L[x] '						
KO						
'nad[c] + h2o[c] + glx[c] -> 2 h[c] + nadh[c] + oxa[c] '						
'atp[c] + pail_cho[c] -> h[c] + adp[c] + pail4p_cho[c] '						
'fad[m] + glyc3p[c] -> dhap[c] + fadh2[m] '						
'nad[c] + 3pg[c] -> h[c] + nadh[c] + 3php[c] '						
'dhap[m] -> dhap[c] '						
'h[x] + lac_L[x] -> h[c] + lac_L[c] '						
'2 nadp[c] + coa[c] + mev_R[c] -> 2 h[c] + 2 nadph[c] + hmgcoa[c] '						
'glu5sa[c] -> glu5sa[m] '						

Scenario	Algorithm	Feed	Regulations	e_up	q_{mAb} (pg/cell/day)	μ (hr^{-1})
31	'Igg_Igg'	'D2'	2	0.9	46.070	0.017000
UP						
'glu5sa[c] -> h[c] + 1pyr5c[c] + h2o[c] '						
KO						
'h[c] + 1pyr5c[c] + h2o[c] -> glu5sa[c] '						

Scenario	Algorithm	Feed	Regulations	e_up	q_{mAb} (pg/cell/day)	μ (hr^{-1})
32	'Bio_Igg'	'D3'	2	0.9	34.082	0.021000
UP						
'glu5sa[c] -> h[c] + 1pyr5c[c] + h2o[c] '						
KO						
'h[c] + 1pyr5c[c] + h2o[c] -> glu5sa[c] '						

Scenario	Algorithm	Feed	Regulations	e_up	q_{mAb} (pg/cell/day)	μ (hr^{-1})
33	'Bio_Igg'	'D3'	2	0.9	34.082	0.021000
UP						
'glu5sa[m] -> glu5sa[c] '						
KO						
'glu5sa[c] -> glu5sa[m] '						

Scenario	Algorithm	Feed	Regulations	e_up	q_{mAb} (pg/cell/day)	μ (hr^{-1})
47	'Igg_Igg'	'D1'	8	1	13.418	0.009
UP						
'2 h[c] + 1pyr5c[c] + nadph[c] -> pro_L[c] + nadp[c] '						
'atp[c] + cytd[c] -> h[c] + adp[c] + cmp[c] '						
'dtdp[m] + dttp[c] -> dttp[m] + dtdp[c] '						
'2 accoa[m] -> coa[m] + aacoa[m] '						
KO						
'chsterol[c] + R1coa_cho[c] -> coa[c] + xolest_cho[c] '						
'nadp[x] -> nadp[c] '						
'nad[x] -> nad[c] '						
'ppi[c] -> ppi[m] '						

Scenario	Algorithm	Feed	Regulations	e_up	q_{mAb} (pg/cell/day)	μ (hr^{-1})
49	'Bio_Igg'	'D4'	7	0.5	39.363	0.013000
UP						
'h2o[c] + ps_cho[c] -> h[c] + ser_L[c] + pa_cho[c] '						
'h2o[m] + accoa[m] + oaa[m] -> h[m] + coa[m] + cit[m] '						
'coa[c] + atp[c] + acac[c] -> amp[c] + ppi[c] + aacoa[c] '						
'crn[c] + adpcoa[c] -> coa[c] + c6dc[c] '						
'coa[c] + atp[c] + Rtotal3[e] -> amp[c] + ppi[c] + Rtotal3coa[c] '						
KO						
'nad[m] + 3hmbcoa[m] -> h[m] + nadh[m] + 2maacoa[m] '						
'akg[c] + leu_L[c] -> glu_L[c] + 4mop[c] '						

Scenario	Algorithm	Feed	Regulations	e_up	q_{mAb} (pg/cell/day)	μ (hr^{-1})
52	'Bio_Igg'	'D4'	7	0.6	43.678	0.012000
UP						
'Rtotal2[c] -> Rtotal2[e] '						

'Rtotal3[c] -> Rtotal3[e] '
KO
'thr_L[c] -> gly[c] + acald[c] '
'coa[c] + atp[c] + acac[c] -> amp[c] + ppi[c] + aacoa[c] '
'akg[c] + leu_L[c] -> glu_L[c] + 4mop[c] '
'h2o[c] + pser_L[c] -> pi[c] + ser_L[c] '
'ac[m] + oaa[m] -> cit[m] '

Scenario	Algorithm	Feed	Regulations	e_up	q_{mAb} (pg/cell/day)	μ (hr^{-1})
58	'Bio_Igg'	'D4'	2	0.9	60.755	0.013000
UP						
'glu5sa[c] -> h[c] + 1pyr5c[c] + h2o[c] '						
KO						
'h[c] + 1pyr5c[c] + h2o[c] -> glu5sa[c] '						

Scenario	Algorithm	Feed	Regulations	e_up	q_{mAb} (pg/cell/day)	μ (hr^{-1})
74	'Igg_Igg'	'D2'	10	1.1000	32.701	0.011000
KO						
'pro_L[c] + fad[c] -> h[c] + 1pyr5c[c] + fadh2[c] '						
'h[c] + pi[c] -> h[m] + pi[m] '						
'thf[m] + ser_L[m] -> h2o[m] + gly[m] + mlthf[m] '						
'akg[c] + leu_L[c] -> glu_L[c] + 4mop[c] '						
'pro_L[c] -> pro_L[m] '						
'acac[c] + pyr[m] -> pyr[c] + acac[m] '						
'h[m] + lac_L[m] -> h[c] + lac_L[c] '						
'h[m] + orn[c] + citr_L[m] -> h[c] + citr_L[c] + orn[m] '						
'h[m] + glu_L[m] + asp_L[c] -> h[c] + glu_L[c] + asp_L[m] '						
'h[m] + pyr[m] -> h[c] + pyr[c] '						

Scenario	Algorithm	Feed	Regulations	e_up	q_{mAb} (pg/cell/day)	μ (hr^{-1})
80	'Bio_Igg'	'D1'	8	0.5	29.037	0.005
UP						
'h2o[c] + dag_cho[c] -> h[c] + Rtotal[c] + mag_cho[c] '						
'lac_L[x] -> lac_L[c] '						
'pi[x] -> pi[c] '						
'nad[x] + lac_L[x] -> h[x] + nadh[x] + pyr[x] '						
'fadh2[m] + fad[c] -> fad[m] + fadh2[c] '						
KO						
'h2o[c] + dag_cho[c] -> h[c] + Rtotal[c] + mag_cho[c] '						

'lac_L[x] -> lac_L[c] '
'pi[x] -> pi[c] '

Scenario	Algorithm	Feed	Regulations	e_up	q_{mAb} (pg/cell/day)	μ (hr^{-1})
81	'Bio_Igg'	'D1'	9	0.6	22.928	0.006
UP						
'2 accoa[m] -> coa[m] + aacoa[m] '						
'coa[c] + atp[c] + ac[c] -> accoa[c] + amp[c] + ppi[c] '						
'h[c] + pyr[c] -> h[m] + pyr[m] '						
'pi[x] -> pi[c] '						
'ppi[c] -> ppi[m] '						
KO						
'nad[c] + pi[c] + g3p[c] -> h[c] + nadh[c] + 13dpg[c] '						
'34hpp[c] -> 34hpp[e] '						
'nadp[x] -> nadp[c] '						
'fadh2[m] + fad[c] -> fad[m] + fadh2[c] '						

Scenario	Algorithm	Feed	Regulations	e_up	q_{mAb} (pg/cell/day)	μ (hr^{-1})
84	'Bio_Igg'	'D1'	8	0.9	21.402	0.007
UP						
'h[c] + pi[c] -> h[m] + pi[m] '						
'o2[c] -> o2[x] '						
'2 na1[e] + HC02195[e] -> 2 na1[c] + HC02195[c] '						
'h2o[l] -> h2o[c] '						
KO						
'2 h[c] + nadh[c] + L2aadp[c] -> nad[c] + h2o[c] + L2aadp6sa[c] '						
'nad[c] + mal_L[c] -> h[c] + nadh[c] + oaa[c] '						
'34hpp[c] -> 34hpp[e] '						
'hmgcoa[m] -> hmgcoa[c] '						

Scenario	Algorithm	Feed	Regulations	e_up	q_{mAb} (pg/cell/day)	μ (hr^{-1})
86	'Bio_Igg'	'D1'	9	1.1000	25.632	0.005
UP						
'nad[c] + h2o[c] + glx[c] -> 2 h[c] + nadh[c] + oxa[c] '						
'h[c] + pi[c] -> h[m] + pi[m] '						
'34hpp[c] -> 34hpp[e] '						
'h[m] + glu_L[m] -> h[c] + glu_L[c] '						
'adp[m] + succoa[m] + pi[m] -> coa[m] + atp[m] + succ[m] '						
KO						

'nad[c] + pi[c] + g3p[c] -> h[c] + nadh[c] + 13dpg[c] '
'h2o[m] + ppi[m] -> h[m] + 2 pi[m] '
'34hpp[c] -> 34hpp[e] '
'h[x] -> h[c] '

Scenario	**Algorithm**	**Feed**	**Regulations**	**e_up**	**q_{mAb} (pg/cell/day)**	**μ (hr^{-1})**
88	'Bio_Igg'	'D1'	20	1.1000	14.302	0.009
UP						
'thr_L[c] -> gly[c] + acald[c] '						
'h2o[m] + ppi[m] -> h[m] + 2 pi[m] '						
'coa[c] + atp[c] + acac[c] -> amp[c] + ppi[c] + aacoa[c] '						
'h[c] + nadph[c] + acetol[c] -> nadp[c] + 12ppd_S[c] '						
'h2o[m] + accoa[m] + aacoa[m] -> h[m] + coa[m] + hmgcoa[m] '						
'h2o[c] + mi14p[c] -> pi[c] + mi4p_D[c] '						
'atp[m] + hco3[m] + pyr[m] -> h[m] + adp[m] + pi[m] + oaa[m] '						
'34hpp[c] -> 34hpp[e] '						
'Rtotalcoa[c] -> pmtcoa[c] '						
'fadh2[c] -> fadh2[r] '						
'q10h2[c] -> q10h2[e] '						
'sprm[e] -> sprm[c] '						
'glu_L[c] + oaa[c] -> akg[c] + asp_L[c] '						
'h[m] + glu_L[m] -> h[c] + glu_L[c] '						
'h[m] + lac_L[m] -> h[c] + lac_L[c] '						
'h[c] + nadh[c] + oaa[c] -> nad[c] + mal_L[c] '						
'adp[m] + succoa[m] + pi[m] -> coa[m] + atp[m] + succ[m] '						
'h[m] + pyr[m] -> h[c] + pyr[c] '						
'h[c] + nadph[c] + HC01434[c] -> nadp[c] + icit[c] '						
'nad[c] + fadh2[c] -> h[c] + nadh[c] + fad[c] '						