



Dinis Granadeiro Pires Madeira

Licenciatura em Engenharia Química

**Development of a toolbox for *in silico* experiment free
culture media design based on *prior* metabolic knowledge**

Dissertação para obtenção do Grau de Mestre em
Engenharia Química e Bioquímica

Orientador: Rui Oliveira, Professor Associado , FCT-UNL

Co-orientador: Rafael S. Costa, Doutor, FCT-UNL

Júri:

Presidente: Professor Doutor Mário Fernando José Eusébio

Arguentes: Professor Doutor Ludwig Krippahl

Vogais: Professor Doutor Rui Manuel Freitas Oliveira

Março, 2020



FACULDADE DE
CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE NOVA DE LISBOA

Development of a toolbox for *In silico* experiment free culture media design based on *prior* metabolic knowledge

Copyright © Dinis Granadeiro Pires Madeira, Faculdade de Ciências e Tecnologia,
Universidade Nova de Lisboa.

A Faculdade de Ciências e Tecnologia e a Universidade Nova de Lisboa têm o direito, perpétuo e sem limites geográficos, de arquivar e publicar esta dissertação através de exemplares impressos reproduzidos em papel ou de forma digital, ou por qualquer outro meio conhecido ou que venha a ser inventado, e de a divulgar através de repositórios científicos e de admitir a sua cópia e distribuição com objetivos educacionais ou de investigação, não comerciais, desde que seja dado crédito ao autor e editor.

Resumo

O desenho de meios de cultura desempenha um papel vital na performance dos meios de cultura nomeadamente na concentração celular, viabilidade das células, rendimento e qualidade de produto produzido. O principal objetivo da presente dissertação é desenvolver um método computacional em MATLAB para o design de meios de cultura *in silico* com base em conhecimento metabólico prévio, sem a necessidade de experiências laboratoriais. Foi usado como caso de estudo uma rede metabólica para as células de ovário do *hamster* chinês (CHO), as células CHO são atualmente as principais células de mamífero utilizadas na produção de proteínas recombinantes para uso na indústria farmacêutica e aplicações terapêuticas. A técnica da análise dos modos elementares foi usada para calcular o número total de *elementary modes* (EM) da rede metabólica CHO, estes são usados para calcular a *footprint* (pegada) dos metabolitos extracelulares no meio de cultura para os diferentes metabolismos celulares. O conjunto de todas as *footprints* da rede metabólica foi nomeado de *footprintome*. O método “Pattern clustering” foi utilizado para reduzir o *footprintome*. Duas *footprints* foram calculadas para representar a totalidade do *footprintome*, usando dois métodos de cálculo, a média aritmética de todas as *footprints* presentes no *footprintome* e a média ponderada pela energia livre de Gibbs. Comparou-se os resultados obtidos com dados da literatura e observou-se que 21 dos 26 metabolitos extracelulares presentes nos dados da literatura estão dentro do intervalo de variância dos resultados obtidos. Também foi desenvolvido um método para calcular meios de cultura específicos baseados no fenótipo de cada *footprint*, duas fórmulas de meios de cultura foram calculadas uma para crescimento celular e outra para produção de imunoglobulina G (IgG), a fórmula para crescimento celular foi comparada com uma fórmula de meio de cultura CHO testada em laboratório, concluindo-se que esta ferramenta computacional é capaz de calcular fórmulas de meios de cultura, que posteriormente podem ser testadas em laboratório.

Palavras Chave: desenho de meios de cultura, *análise dos modos elementares*, redes metabólicas, Chinese Hamster Ovary (CHO)

Abstract

Culture media design plays a vital role in culture performance, namely cell density, cell viability, product yield and product quality. The main goal of the present *M.Sc.* thesis is to develop a method for *in silico* culture media design based on prior metabolic knowledge of the targeted cell line, reducing the burden of laboratory experiments. A Chinese Hamster Ovary (CHO) cell line was used as case study. CHO cells are the main mammalian host used today in the biopharmaceutical industry for the production of recombinant proteins for therapeutics and pharmaceutical applications. Elementary mode analysis (EMA) was applied to compute the total number of elementary modes (EM) of a representative CHO metabolic network. From the full set of EMs, the extracellular metabolite footprints for different metabolic states were computed. The full set of metabolic footprints was named *footprintome*. Footprintome reduction was achieved by pattern clustering. Two footprints were computed representing all of the Footprintome by averaging based on number and averaging based on Gibbs free energy. These footprints were compared with literature data resulting in significant similarities. More specifically, 21 out of the 26 extracellular metabolites present in the literature data are within the variance intervals of the computed footprintome. Afterwards a phenotype-targeted design was applied to compute custom culture media formulas for cell growth and for Immunoglobulin G (IgG) production. The cell growth formula was compared with a lab tested CHO culture medium formula, showing that this tool is capable of computing culture media formulas that can be further tested in the lab.

Keywords: culture media design methods, elementary mode analysis, metabolic networks, Chinese Hamster Ovary (CHO)

Index

Abstract	V
Figure Index.....	VIII
1. Introduction.....	1
1.1 Culture Media Design Methods	3
1.1.1 Plackett and Burman design	5
1.1.2 Taguchi design.....	6
1.1.3 Response Surface Methodology	7
1.1.4 Artificial neural network.....	8
1.1.5 Genetic Algorithm	10
1.2 Metabolic modelling methods	12
1.2.1 Metabolic flux analysis (MFA)	13
1.2.2 Flux balance analysis (FBA).....	14
1.2.3 Elementary mode analysis	15
1.4 M.Sc. thesis objectives	17
2.4.1 Calculation of elementary mode Gibbs free energy of reaction (ΔG°_r).....	21
2.4.2 Clustering based on phenotype similarity.....	22
3. Results and discussion.....	25
3.1 Chinese Hamster Ovary Cells (CHO) metabolic network	25
3.2 Computation of CHO Elementary Modes.....	26
3.3 Computation and reduction of CHO <i>footprintome</i>	29
3.4. Comparison with experimental data.....	33
3.4. Phenotype-targeted culture media design	36
3.4.1 Minimization of lactate and ammonium buildup	37
3.4.2 Minimization of osmolarity buildup.....	39
3.4.3 Final culture media concentrations.....	41
4. Conclusions.....	44
5. References.....	46
Appendix	53
Appendix A – CHO Metabolic Networks	53
Appendix C - Standard Gibbs free energy of formation of metabolites values	66
Appendix D - Literature data cell growth rate and metabolic rates ⁶²	67
Appendix E - Arithmetic footprint and computed formulas for Biomass and IgG producing networks.	68

Figure Index

Figure 1 - Culture media design methodologies.....	3
Figure 2 - Response surface for an exemplary model.....	7
Figure 3 - Multi layered perceptron neural network	9
Figure 4 - GA basic flow scheme.....	10
Figure 5 - Example of culture media optimization using GA.....	11
Figure 6 - Example of a simple metabolic Network.....	12
Figure 7 - FBA example, restrictions and objective function.....	14
Figure 8 - Elementary modes of the Fig. 6 network.....	15
Figure 9 - Admissible Flux Space of the Fig. 6 network elementary modes.....	15
Figure 10 - Example of a Chinese Hamster Ovary Cell elementary mode.....	18
Figure 11 - Reduced <i>footprintome</i> for biomass production after 2 steps of reduction: Step 1- biomass production, Step 2-Thermodynamic reduction.....	31
Figure 12 - Reduced <i>footprintome</i> for biomass production after 3 steps of reduction: Step 1- biomass production, Step 2-Thermodynamic reduction, Step 3- Pattern clustering with arithmetic averagings.....	32
Figure 13 - Comparison between theoretical footprints computed by the arithmetic average method and the weighted average method with experimental footprint	34
Figure 14 - Histogram ΔG°_r of each EM for the biomass producing network.....	36
Figure 15 - Reduced <i>footprintome</i> for biomass production after 4 steps of reduction: Step 1- biomass production, Step 2-Thermodynamic reduction, Step 3- Pattern clustering with arithmetic averaging, Step 4- Elimination of all footprints that either produce lactate or ammonium.	38
Figure 16 - Comparison between the optimal EM computed for the biomass producing network and the respective arithmetic footprint.....	40
Figure 17 - Comparison between the optimal EM computed for the IgG producing network and the respective arithmetic footprint.....	41
Figure 18 - Comparison between the computed formula for biomass synthesis with CHO culture media.....	42

Table Index

Table 1 Culture media design studies over the years.....	13
Table 2 Plackett and Burman design of experiments.....	14
Table 3 Taguchi Signal -to -Noise Ratios.....	15
Table 4 Elementary mode count of the biomass and product producing networks.....	37
Table 5 Footprint of the EM represented in Appendix B, Table B.1.....	39
Table 6. Automatic <i>footprintome</i> reduction results.....	40
Table 7 MSE between measured and theoretical footprint.....	46
Table 8 Computed formulas for IgG and biomass syntheses and Lab tested CHO culture media.....	54

1. Introduction

The optimisation of the composition of culture media has been historically a very important factor for the performance of cell culture. Historically the main drivers have been the maximization of cell density, cell viability, product titer and product yield. In many cases the positive impact of culture media surpasses that of genetic engineering. The creation and optimization of culture media that supports the growth of microorganisms or cells comes with a set of challenges, particularly, the creation of methods that can achieve the best results, while being time and cost effective. There are several methodologies in use today, in this introduction we will be giving a brief history and definition about the different methods for culture media design. A more detail explanation about this methods, their advantages and disadvantages will be discussed in further chapters of this work.

Culture media design started with classical methods using one-factor-at-time experiments which consists in changing one factor for each experiment until we have a desired result. This method was primarily used in culture media design until the mid-twenties century, when it was replaced by design of experiments methods (DoE). One of the early forms of this was the Plackett and Burman design (PBD) created by Plackett R. L. and Burman J. P.¹ in 1946, it is a two level design method used to detect the top contributing factor, assuming that the interaction between two or more factors are negligible, PBD is still in use today, primarily for the elimination of non-contributing factors (screening) in the early stages of culture media design.

A few years later the Central Composite Design (CCD) DoE was created by Box G. E. P. and Wilson K. B.² in 1951 to overcome some of the limitations of PBD. The CCD is composed by 3 sets of points: a factorial set, each having two levels, a center set, whose values are the median values of each factor, and an axial set that are identical to the center points except for one factor. Today this design is mainly used in the Response Surface Methodology (RSM) also developed by Box G. E. P. and Wilson K. B.², a DoE is ran and then RSM is used to create a mathematical model that uses statistical design of experiments and regression analysis to obtain the best formulations This method is still to date, the most used method in the industry to optimise culture media.

The Taguchi method was created by Genichi Taguchi³ in 1992, this method is based on noise analysis (uncontrollable variables in the experiments), making it a useful method for measuring the characteristics deviation of the target value. Taguchi method uses an orthogonal array of experiments and unlike PBD, it analyses the main effect and two factor iterations.

The artificial neural networks (ANN) are a computing system inspired by the neural connections in the animal brains, the first computational model was created by Warren M. and Walter P.⁴ in 1943, the technology continued to advance through the century and the first application of ANN in culture media design was the work done by Glassey J et. al.⁵ in a *Escherichia coli batch* fermentation in 1994. ANN can process a large amount of information, which is useful in culture media design, that often contains hidden patterns and large amounts of variables.

A genetic algorithm (GA) mimics the process of natural selection by relying on biologically inspired operators such as mutation, crossover and selection⁶. These types of algorithms started to be applied in culture media design in the 1990s. One of the first examples of this approach used in culture media design was the work proposed by Weuster-Botz D and Wandrey⁷, with the optimization of a culture media for a *Candida boidinii* fermentation.

In this work will also be discussed different analysis tools to determine metabolic fluxes in metabolic networks. These types of analysis are the basis for metabolic engineering which is defined as the “direct improvement of product formation or cellular properties through the modification of specific biochemical reactions or introduction of new ones with the use of recombinant DNA technology”, as such metabolic engineering uses metabolic networks to determine metabolic fluxes that characterize cell physiology and control cell metabolism⁸. The analysis of a metabolic networks is based on the principle of mass conservation of internal metabolites within a system⁹. Depending on what type of analysis we want, there are three main techniques: metabolic flux analysis(MFA)¹⁰, flux balance analysis (FBA)¹¹ and elementary mode analysis (EMA)¹². These techniques and how they can be applied in culture media design will be explain in more detail in the next chapters of this work.

1.1 Culture Media Design Methods

Culture media design methods can be classified as (Fig. 1):

- Traditional one-factor-at-a-time
- Advanced statistical/mathematical methods (the most frequently applied today)
- Advanced system biology methods

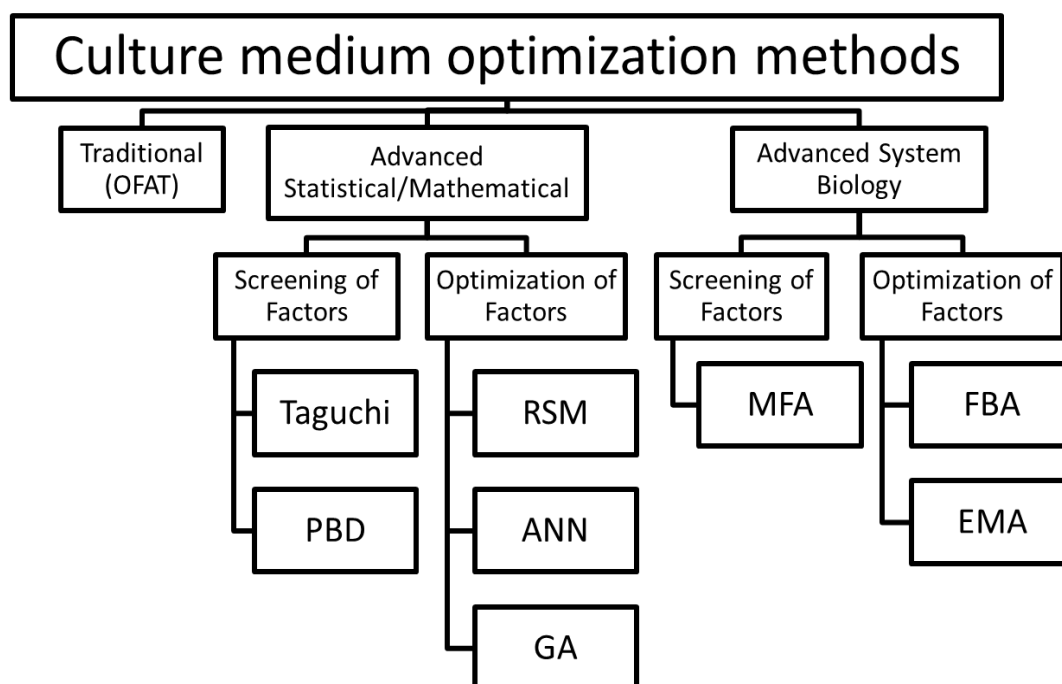


Figure 1 - Culture media design methodologies . PBD: Plackett and Burman Design; RSM: Response Surface Method;(Adapted from: Singh V, Haque S, Niwas R, Srivastava A, Pasupuleti M and Tripathi C K M 2017 Strategies for fermentation media optimization: An in-depth review Front. Microbiol. 7⁸⁵)

In this context, it is important to distinguish two different problems: screening and optimization of factors. Screening of factors is typically the first step in a culture media design project. It aims to identify the principal contributing factors that should be optimized in the proceeding steps. Optimization of factors aims at optimizing the media components quantities (typically concentrations) to enforce maximal product synthesis and/or maximal cell growth while minimizing unwanted by-products, such as lactate and ammonia. Table 1 overviews some of the published studies for culture media design methods.

Table 1 - Culture media design studies over the years

Screening method / DOE	Optimization method	Experiments	Number of Factors	Metabolite	Cell line	Optimization	Year	Reference
Taguchi	-	16	4	hyaluronic acid	<i>Streptococcus sp.</i>	112%	2009	13
Taguchi	-	9	4	N-acetylchitooligosaccharide	<i>Streptomyces chilikensis</i>	126,86%	2019	14
Taguchi	-	16	4	Bioethanol	<i>Saccharomyces cerevisiae</i>	37,50%	2019	15
Taguchi	-	8	6	Mixed alcohols	<i>Clostridium pasteurianum</i>	78,48%	2013	16
Taguchi	-	18	8	Amidase	<i>Geobacillus subterraneus</i>	113,79%	2016	17
PBD	-	16	12	Lipase	<i>Candida rugosa</i>	-	2008	18
Taguchi - CCD	RSM	12 - 27	10-4	Vanillin	<i>Psychrobacter sp.</i>	-	2012	19
Taguchi - CCD	RSM	18 -27	8-4	Lipase	<i>Rhizopus chinensis</i>	120%	2008	20
CCD	RSM	20	3	b-carotene	<i>Daucus carota</i>	141,33%	2008	21
CCD	RSM	50	5	Lipase	<i>Aryabhattai SE3-PB</i>	618,45%	2018	22
CCD	RSM	17	3	Y-aminobutyric acid	<i>Streptococcus thermophilus</i>	20%	2018	23
PBD - CCD	RSM	20- 32	12-5	Glycolipopeptide Biosurfactant	<i>Pseudomonas aeruginosa</i>	-	2017	24
PBD - CCD	RSM	8 - 13	5-2	Exopolysaccharides	<i>Cordyceps militaris</i>	96,00%	2010	25
-	GA	80	11	-	<i>Helicoverpa zea (insect cell)</i>	550,00%	2002	26
-	GA	98	14	2-Phenylethanol	<i>Kluyveromyces marxianus</i>	32,00%	2004	27
-	GA	270	12	Eicosapentaenoic acid	<i>Nannochloropsis gaditana</i>	23,00%	20015	28
-	GA	544	26	yessotoxins	<i>Dinoflagellates</i>	60,00%	2011	29
CCD	ANN	20	3	avermectin B1b	<i>Streptomyces avermitilis</i>	50,78%	2014	30
CCD	ANN-GA	36	5	Cholesterol Oxidase	<i>Streptomyces Sp.</i>	132,14%	2015	31
	RSM-GA					49,60%		
CCD	ANN-GA	20	3	Lipase	<i>Soil Microorganism (Not Specified)</i>	19,41%	2007	32
	RSM-GA					10,56%		

(Experiments and number of factors are represented for screen method and Doe method respectively in cases were both take part in the study).

1.1.1 Plackett and Burman design

The PBD method focuses on the main effect of each factor, assuming that, interactions of two or more factors are negligible, it is used for factor screening in the beginning of a culture media design process. PBD is a two-level factor design, a high level (+) and a low level (-), there are two types of variables: “real variables”, whose values change during experiments, and “dummy variables”, whose values remain the same. Classical experiments before the execution of the PBD help in the selection of independent and dummy variables. Table 2 represents a Plackett and Burman design of experiments for 12 runs and 11 two-level factors, usually in culture media design, factor values represent concentration values of different compounds in the media.

Table 2 - Plackett and Burman design of experiments

RUN	FACTORS										
	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11
1	+	+	+	+	+	+	+	+	+	+	+
2	-	+	-	+	+	+	-	-	-	+	-
3	-	-	+	-	+	+	+	-	-	-	+
4	+	-	-	+	-	+	+	+	-	-	-
5	-	+	-	-	+	-	+	+	+	-	-
6	-	-	+	-	-	+	-	+	+	+	-
7	-	-	-	+	-	-	+	-	+	+	+
8	+	-	-	-	+	-	-	+	-	+	+
9	+	+	-	-	-	+	-	-	+	-	+
10	+	+	+	-	-	-	+	-	-	+	-
11	-	+	+	+	-	-	-	+	-	-	+
12	+	-	+	+	+	-	-	-	+	-	-

The effect of each factor is given by Eq. 1:

$$Eff = 2 * \left(\sum Y_h - \sum Y_l \right) / N \quad (1)$$

Where Eff is the effect of a factor (objective function), Y_h is the value given for the high level experiments, Y_l is the value given for the low level experiments and N is the total number of runs.

The experimental error is calculated by using the effects of the “dummy variables” (Eq. 2).

$$V_{eff} = \sum \frac{Ed^2}{n} \quad (2)$$

Where V_{eff} is the variance of the effect Ed , Ed is the effect of a “dummy variable” and n is the number of “dummy variables”.

The standard error is the square root of the variance V_{eff} and the significant level of the effect of each variable is determined by a t-test distribution and the variables with a confidence level greater than 90% or 95% are chosen.

Examples of culture media design utilizing PBD are the statistical evaluation of a culture media components for lipase production by *Pseudomonas fluorescens* made by Rajendran A. et. al.¹⁸ and the screening of components for protease production by *Bacillus safensis* in a submerged fermentation³³.

Its main advantages are the screening of high number of factors using a low number of experiments, capable of screening n factors in $n + 1$ experiments, saving both time and money at the cost of neglecting two factor interaction, remaining a simple and practical screening method to use in the initial stages of culture media design.

1.1.2 Taguchi design

Taguchi method is based in an orthogonal array of experiments. Taking focal point in noise factor analysis (uncontrollable factors that affect the result of experiments and generally cause loss of quality), through the use of a signal-to-noise ratio (S/N), Taguchi methodology can be used in the initial stages of culture media design to identify which factors are more influential.

First we identify the control factors (controllable factors), their levels and the appropriate orthogonal array (based on the degrees of freedom), once we got the measured experiment values, we calculate the (S/N) ration for each experiment. There are three signal-to-noise ratios: *smaller the better*, *larger the better* and *nominal the best*. Table 3 describes these three ratios.

Table 3 - Taguchi Signal -to -Noise Ratios

<i>Smaller the better</i>	$\left(\frac{S}{N}\right) = -10 * \log_{10} \left[\left(\frac{1}{n}\right) * \sum_{i=1}^n Y_i^2 \right]$	(3)
<i>Larger the better</i>	$\left(\frac{S}{N}\right) = -10 * \log_{10} \left[\left(\frac{1}{n}\right) * \sum_{i=1}^n \frac{1}{Y_i^2} \right]$	(4)
<i>Nominal the best</i>	$\left(\frac{S}{N}\right) = 10 * \log_{10} \left(\frac{Mean^2}{Variance} \right)$	(5)

Where Y is the experiment result, n is the total number of experiments and for the nominal the best S/N ratio, the Mean and Variance are for a chosen set value.

Next, we calculate the S/N value for each level of the control factors. Eq. 6 shows the S/N values for the control factor $\{i\}$ level $\{j\}$.

$$\frac{S}{N}_{ij} = \sum_k^n \frac{S}{N}_k \quad (6)$$

With n being the total number of experiments with the factor $\{i\}$ on level $\{j\}$.

The S/N values for the different levels of a factor are compared between each other and those with the highest value are chosen, being the levels that minimize the effect of noise. With Taguchi method we can choose the set of control factors levels that minimizes the effect of noise for the objective function in study.

An example of culture media design study using Taguchi method is the work done by Makowski K. et. al.³⁴. It comprises the optimization of a culture media, for the production of microorganisms active in odorous compound removal. Another example is the optimization of media component for the production of N-acetylchitooligosaccharide from chitin by *Streptomyces chilikensis*¹⁴.

Contrary to PBD method, the Taguchi is a more complex and complete approach, that can detect 2 factors interaction but is still far away of describing a whole microorganism metabolism.

1.1.3 Response Surface Methodology

Response surface methodology (RSM) it's a complex mathematical method for the optimization of culture media design, which includes statistical experimental design and regression analysis². It is used after the implementation of a PBD or CCD design of experiments. RSM optimization includes three main steps, experimental design (screening of factors), the path of steepest ascent/descent and the quadratic regression mode.

After identifying the main effects, the next step is to explore the region of the operation conditions. This region is called the *response surface* (Fig. 2) and the goal is to conduct a series of experiments to find the path of the steepest ascend or descend given by the initial set of experiences³⁵. A first order model can be used (Eq. 7), containing only the main effects and its interactions.

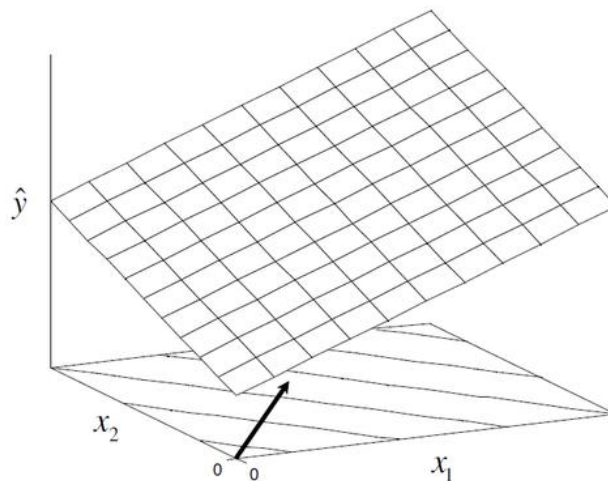


Figure 2 - Response surface for an exemplary model

$$Y = a_0 + a_1x_1 + a_2x_2 + a_{12}x_1x_2 + e \quad (7)$$

Where Y is the predicted response, e represents the effects of uncontrollable variables (noise) and a_i the regression coefficients of a factor i .

In RSM, when close to the optimum region, in which a first order model is no longer adequate, it is also applied a second order model. The second order regression model is presented in Eq. 8:

$$Y = a_0 + \sum_{i=1}^N a_i X_i + \sum_{i < j}^N a_{ij} X_i X_j + \sum_{i=1}^N a_{ii} X_i^2 \quad (8)$$

Where $a_i X_i$ are the linear terms, $a_{ij} X_i X_j$ are the interaction terms and $a_{ii} X_i^2$ are the square terms.

The main advantage of using RSM is the ability to study the effects of the factors in the response throughout the entire surface region, capable of predicting an optimum response. Examples of culture media design using RSM are the culture media optimization for β -carotene and biomass production in *Dunaliella salina* in mixotrophic culture³⁶, and the optimization of lipase production in *Bacillus aryabhattai* (²²).

However, the RSM also has its limitations, the complexity of microorganism metabolism, its nonlinear nature and the low availability of quality kinetic data³⁷, makes the modeling of biological reaction systems a challenge. Also it is difficult to study interaction of five or more factors, this is a problem because culture media can have up to 100 different components, this coupled with the fact that the experiment numbers needed to optimize 100 factors would be unfeasible to do explains the reason why RSM optimization in culture media design is limited to 3-5 factors (Table 1).

1.1.4 Artificial neural network

Artificial neural networks (ANN) are based on the structural aspect of the network of neurons in the brain. It's a mathematical system that adapts to the information flowing through the network, during the learning "stage". Given a set of training data the network learns to output certain data based on the input given, ANN are useful in culture media design because they are able to compute and learn every type of function, acting as a black box model for solving complex functions that describe microorganism metabolisms³⁸.

The structure of the neural network is comprised in layers of "neurons", a "neuron" is a mathematical function that model the functioning of a biological neuron, that computes the so called activation value. The first layer of the structure it's called the input layer where it's given the initial information, this layer of neurons is then connected to other layers until it reaches the output layer.

There are many types of neural networks Fig. 3 shows a simplistic structure for a multi layered perceptron neural network³⁹ which comprises multi layers of connected neurons.

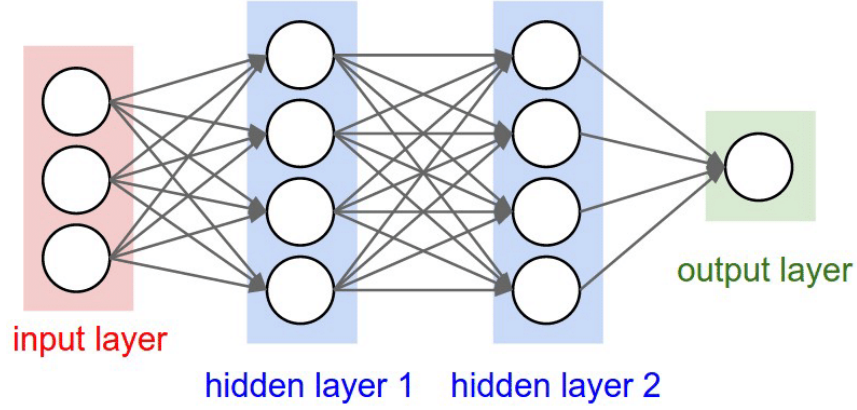


Figure 3 - Multi layered perceptron neural network

The activation value of a neuron is comprised by the sum of all connections values from the previous layer, a connection value is given by the activation value from the originating neuron and a weight value associated to that connection as presented by Eq. 9.

$$A = \sum_i^n w_{ia} a_i \quad (9)$$

Where A is the activation value of a neuron, n is the total number of neurons in the previous layer, w is the specific weight associated to a connection $\{i\}$ and a the is the activation value from the neuron were connection $\{i\}$ originated in the previous layer. The activation value is also subjected to an activation function that is mainly there to convert an input signal to an output one that can be used by the network, the activation function adds a layer of non-linearity, without this the neural network would be nothing more than a one-degree polynomial function, being just able to solve linear functions. The most used activation function is the sigmoid function (Eq. 10) that comprises the activation value between 0 and 1.

$$\sigma(A) = \frac{1}{1 + e^{-A}} \quad (10)$$

The neural network then learns to give a desired output value by running a set of learning data, utilizing a technique called backpropagation, the weights of each connection are modified so the network outputs the respective desired value. The objective is then, like RSM, to predict new data. An example of culture media design using ANN is the enhance of production of Avermectin B1b by *Streptomyces avermitilis*³⁰. Also, various studies were conducted comparing RSM to ANN, showing that ANN as an overall better performance that RSM, with higher optimization and precision in predicted values^{40 41 42}.

ANN is able to process large amounts of data, which is suited to conduct culture media design, although studies using ANN with more than 3- 5 factors (Table 1) are, to our knowledge, inexistent in the literature. This is due to the fact that an ANN needs an initial set of training data, usually CCD is used to create this training set, needing a higher number of experiments the more factors we have in study, which can be costly and time consuming. The access to the training data is the main limitation factor in culture media design using ANN.

1.1.5 Genetic Algorithm

Genetic algorithm (GA) mimics the process of natural selection and it's based in the principle "survival of the fittest"⁴³. Different factors (genes) are encoded in a string (chromosome), the best performing individual is unchanged and the rest can "mate" in performance order.

The GA optimize for each generation the values of a fitness functions using mainly three types of rules to create the next generation:

- *Selection rule*, selects the individuals known as parents that contribute to the population of the next generation.
- *Crossover rule* combines two parents to form children for the next generation.
- *Mutation rule* applies random changes to individual parents to form children.

Fig. 4 represents the basic workflow of a genetic algorithm⁴⁴.

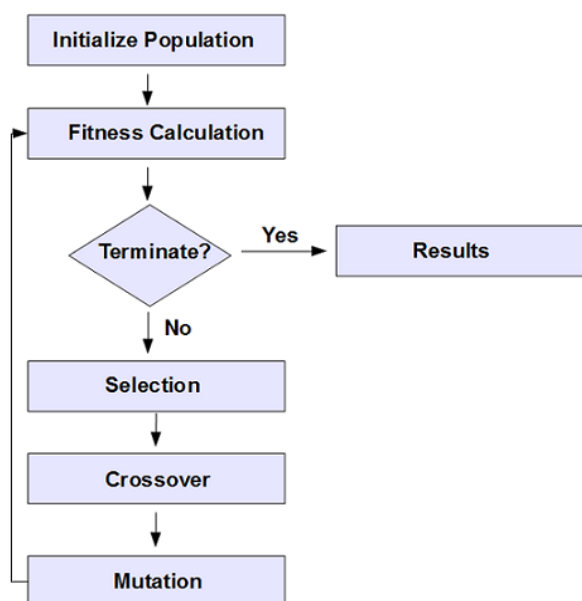


Figure 4 - GA basic flow scheme

Having the previous scheme (figure 4) as example, a generic algorithm can be implemented into culture media design (Figure 5).

Fig. 5 shows how a genetic algorithm can be implemented into culture media design²⁶

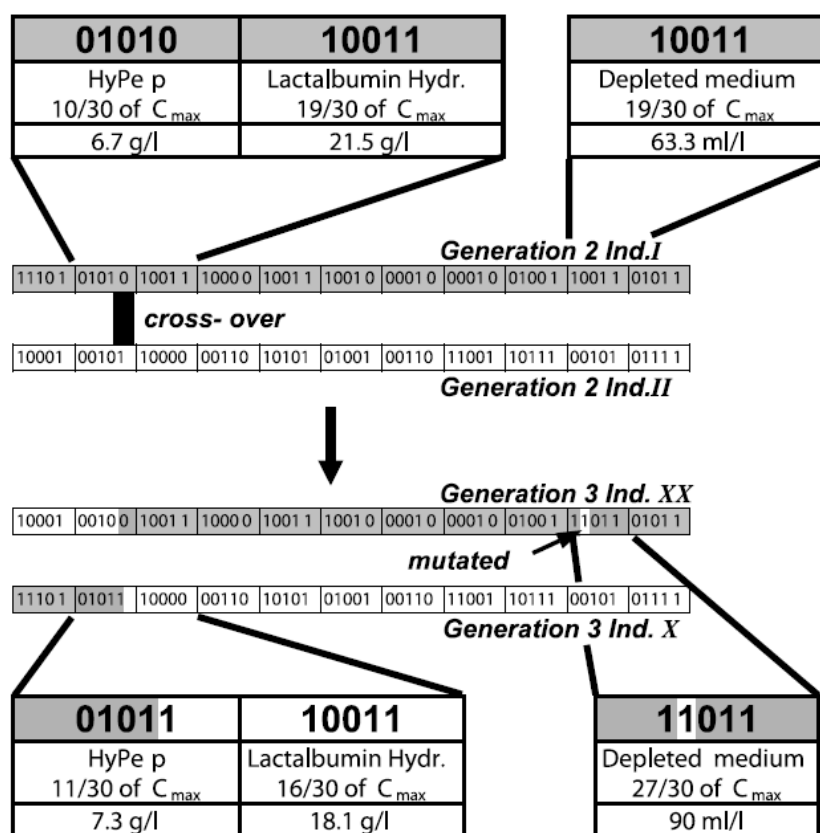


Figure 5 Example of culture media optimization using GA

In this example, the factors encoded were media components concentrations and the parents were selected for maximum cell density and growth rate (fitness function). For every generation a new set of media components was created and tested.

The main advantage of the GA is that, there is no guessing between each experiment as the direction of experiments is set automatically by experimental data, like so, the error between predicted and experimented values doesn't exist, unlike RSM that relies on a second order model to predict the response of complex cell metabolisms, with an increasing number of factors this can often result in poor estimation of optimal formulations⁴⁵, this is not a limitation in the case of GA. Various studies were carried out for culture media design using GA^{26 29 28 7}, and although these studies shows good optimization values and higher number of factor when compared with other methods the main downside is the high number of experiments needed (Table 1).

Various studies were conducted comparing RSM coupled with GA (RSM-GA) and ANN coupled with GA (ANN-GA), where the genetic algorithm is used to optimize the mathematical models created by RSM and ANN. In these studies, ANN-GA performed better than RSM-GA³² (Table 1), making ANN-GA one of the best methods to use in culture media design.

1.2 Metabolic modelling methods

All the culture media design methods described above are empirical in nature, i.e. they disregard knowledge on the biological mechanisms underlying cell growth and product synthesis. They rely instead on intensive experimentation to acquire cause-effect data for a very high number of media modifications. A step further could be the use of more rational design methods based on metabolic networks which are currently widespread in the literature^{46 47 48}. The reconstruction of these networks allows to understand the interconnectivity and functional relationships between all biochemical reactions of a biological system. It is currently relatively easy to synthesize a detailed metabolic network of most of the cell lines/strains used for industrial production. In some organisms, a genome scale reconstructed metabolic network can be found in specific databases like BIGG (Genetic and Genomic knowledgebase of large scale metabolic reconstructions)⁴⁹.

A metabolic network is a system where metabolites (nodes) are linked to each other by enzyme catalysed reactions (edges) where directionality of connection means mass conversion (arrow) (see Fig. 6 for illustration). Reactions that transform metabolites within the system are considered internal reactions, while reactions involving the transport of metabolites in and out of the system are considered exchange reactions. Figure 6 illustrates this concept with 5 internal metabolites (A, B, C, D, P) and 4 external metabolites (A_{ext}, B_{ext}, D_{ext}, P_{ext})⁵⁰.

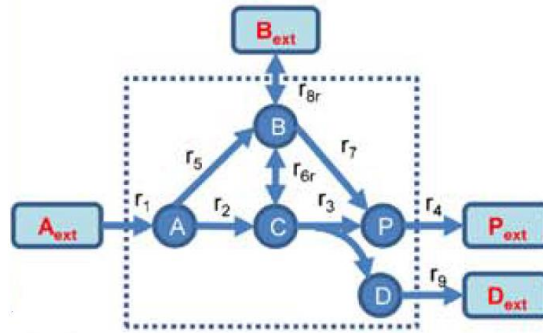


Figure 6 - Example of a simple metabolic Network

In this example reaction r_1 , r_4 , r_8 , and r_9 are considered exchange/transport reactions while the other are the internal reaction. The analysis of a metabolic networks is based on the principle of mass conservation of internal metabolites within a system⁹. Under the hypothesis of a well-mixed system, the general material balance equation applies (Eq. 11).

$$\frac{d}{dt}C = S \times r - \mu \times C \quad (11)$$

Where C is the metabolite concentration vector, r is the reaction rate vector, μ is the specific growth rate and S is the stoichiometry matrix. The stoichiometry matrix S contains the stoichiometry of the network metabolites in each reaction, where the rows are the network metabolites and the columns are the network reactions.

Eq.12 illustrates the stoichiometry matrix for the metabolic network of Fig. 6.

$$\underline{S} = \begin{array}{c|cccccccc} & r_1 & r_2 & r_3 & r_4 & r_5 & r_{6r} & r_7 & r_{8r} & r_9 \\ \hline A & 1 & -1 & 0 & 0 & -1 & 0 & 0 & 0 & 0 \\ B & 0 & 0 & 0 & 0 & 1 & -1 & -1 & -1 & 0 \\ C & 0 & 1 & -1 & 0 & 0 & 1 & 0 & 0 & 0 \\ D & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & -1 \\ P & 0 & 0 & 1 & -1 & 0 & 0 & 2 & 0 & 0 \end{array} \quad (12)$$

At steady state, there is no accumulation of internal metabolites in the system and Eq. 11 can be simplified into Eq. 13.

$$S \times r = 0 \quad (13)$$

An additional restriction is needed (Eq. 14), due to thermodynamic constraints stating that irreversible reactions $\{j\}$ have to proceed in the appropriate direction and require to have positive flux values

$$r_j \geq 0 \quad (14)$$

There are 3 main techniques to solve Eq. 13 with de thermodynamic restriction (Eq. 14): metabolic flux analysis (MFA), flux balance analysis (FBA) and elementary mode analysis (EMA). These are briefly reviewed in the next sub-sections.

1.2.1 Metabolic flux analysis (MFA)

In MFA the flux vector is divided into two vectors, a vector containing the measured metabolic fluxes r_m , and a vector contain all the unmeasured fluxes r_u , in this technique we solve Eq. 13 as:

$$S_u \times r_u = S_m \times r_m \quad (15)$$

With the S_u and S_m being the stoichiometric matrix of the unmeasurable and measurable fluxes respectively. We want to measure enough fluxes so that, the matrix S_u becomes invertible⁵¹ and we can solve Eq. 15 like Eq.16:

$$r_u = -S_u^{-1} \times S_m \times r_m \quad (16)$$

Then we can obtain the unmeasurable flux vector r_u by solving Eq 14 15 and 16 in a system. MFA relies on measuring enough metabolic fluxes until we can calculate the unmeasured ones. Note that this technique only obtains a single metabolic flux vector for a specific growth condition, changing this growth condition will result in a different r_m vector leading to a different metabolic flux vector, an example of this is the study by Wilkens et al.⁵², lactate production was compared in a CHO cell line grown on two different substrates, namely glucose and lactose.

MFA has been applied in the context of culture media design as a “screening of factors” approach. MFA is typically applied for in-depth analysis of the effect of media factors on carbon flux distribution^{53 54}. It is however not a technique that can be applied for quantitative design of media composition.

1.2.2 Flux balance analysis (FBA)

FBA is a technique that can be used for undetermined systems, i.e. when measured fluxes are not enough to invert matrix S_u . FBA implies an optimization according to some objective function alongside several flux constraints, such as, substrate consumption, product secretion, thermodynamic constraints, etc. Like MFA, FBA calculates one flux vector for a given growth condition. Fig. 7 illustrate a typical FBA problem for the small network shown of Fig. 6⁵⁰:

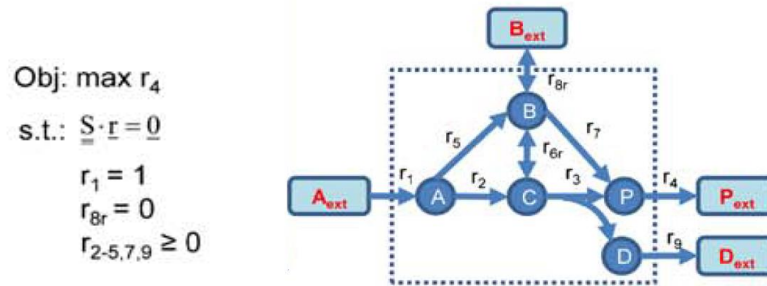


Figure 6 - FBA example, restrictions and objective function

In this example our objective function is to maximize the production of the metabolite $P(r_4)$ while applying substrate consumption (r_1), product excretion (r_{8r}) and thermodynamic ($r_{2-5,7,9}$) constraints.

The main challenge of this approach is to define an objective function that best describes the metabolism of a given growth condition^{55 56}. FBA was successfully implemented to develop a culture media supporting high cell density growth of *Bacillus coagulans*⁵⁷.

1.2.3 Elementary mode analysis

Elementary modes of a cell are defined as the unique set of reactions to support steady state operation of a metabolic network. Fig. 8⁵⁰ shows all elementary modes present in the simple network of Fig. 6.

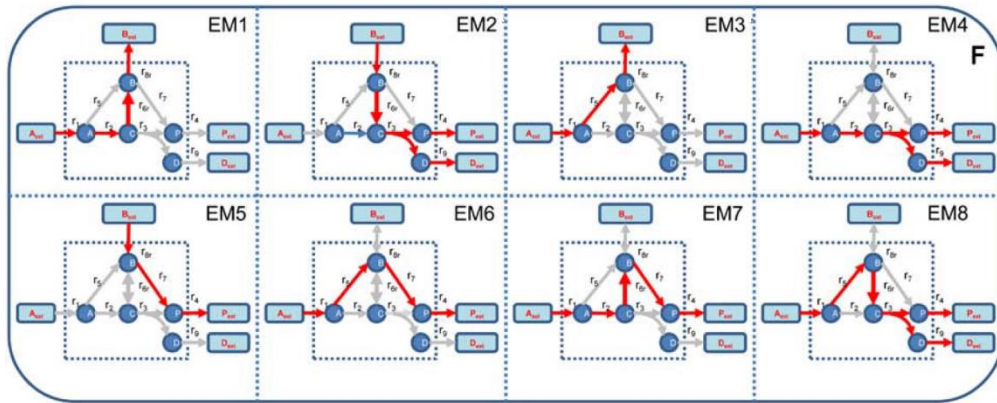


Figure 7 - Elementary modes of the Fig. 6 network

The universe of solutions of Eq. 13 together with the inequality constraint (Eq. 14) takes the form of a convex polyhedral cone⁵⁸, containing an infinite number of solutions, i.e. the admissible flux space. Fig. 9 represents the admissible flux space for the network in Fig. 6.

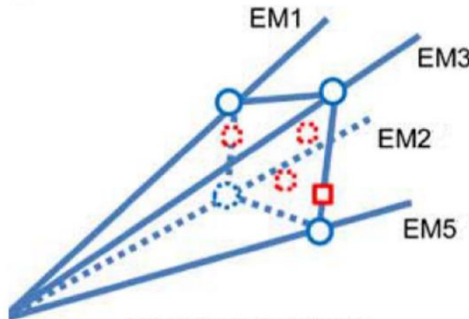


Figure 8.- Admissible Flux Space of the Fig. 6 network elementary modes

Elementary modes must also obey to the non-decomposability constraint. This non-decomposability constrain means that one EM vector cannot be described by other two or more vectors, meaning that each elementary mode its unique in the network.

After EMs calculation we obtain the EMs matrix that has all the EMs of the metabolic network Eq.17.

$$\underline{\underline{EM}} = \begin{matrix} & \begin{matrix} EM_1 & EM_2 & EM_3 & EM_4 & EM_5 & EM_6 & EM_7 & EM_8 \end{matrix} \\ \begin{matrix} r_1 \\ r_2 \\ r_3 \\ r_4 \\ r_5 \\ r_{6f} \\ r_7 \\ r_{8f} \\ r_9 \end{matrix} & \begin{bmatrix} 1 & 0 & 1 & 1 & 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 & 2 & 2 & 2 & 1 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 \\ -1 & 1 & 0 & 0 & 0 & 0 & -1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 \\ 1 & -1 & 1 & 0 & -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 \end{bmatrix} \end{matrix} \quad (17)$$

The rows of the EM matrix represent metabolic reactions while the columns represent the elementary modes of the network. The values in the matrix represent the “weights” (i.e. participation strength) of a given reaction in a given EM. Note that the EM1 vector (1,1,0,0,0,-1,0,1,0)^T may be translated into a flux vector (2,2,0,0,-2,0,2,0)^T by multiplying a positive scalar factor, for example 2. Negative reaction “weights” apply only for reversible reactions, meaning that the particular reaction takes place in the opposite direction of the positive flux direction in that particular EM. Flux calculations are thus obtained by applying a linear combination of EMs according to Eq.18

$$r = \sum_i \lambda_i EM_i \quad (18)$$

With:

$$\lambda \geq 0 \quad (19)$$

There is a parallelism between Eqs. 18-19 and steady state Eqs 13-14. There is however a fundamental difference when doing flux calculations by applying EM analysis. EMs embody knowledge on regulation of metabolic processes. This is so because one EM is interpreted as a metabolic state, i.e. it is not only the metabolic fluxes but also the genes, mRNAs, proteins, that support that particular steady state flux distribution. For example, Tunahan Çakır et al.(2007)⁵⁹ correlated transcriptomic data to active EMs in *Saccharomyces cerevisiae*.

1.4 M.Sc. thesis objectives

The general objective of this thesis is the development of a rational design method for culture media composition customized to the cell line and/or product. This method should act as a proof of concept for the future development of a toolbox. Instead of an empirical design approach where culture media composition is optimized resorting to intensive lab experimentation, the goal here is to develop an *In silico*, experiment-free culture media design method. Given that the knowledge of the metabolic network is easily available in public domain, this novel methodology should be based on prior knowledge of the metabolic network of the target cells. Moreover, since the elementary modes embody additional information on regulation, the elementary modes framework will be privileged over the MFA or FBA techniques. As such, the specific thesis objectives are the following:

Objective 1: Development of an *In silico* culture media design method based on prior knowledge of the metabolic network that drastically reduces the experimental activity for culture media optimization. In limit it should be experiment-free.

Objective 2: The *In silico* culture media design method should allow to compute culture media formulations customized to the cell line and/or product expressed by the cell line.

Objective 3. The method to be developed should be based on elementary modes framework. The advantage of the EM framework is the inclusion of high level regulatory information that is easily accessible.

Objective 4: Implement the EM culture media design method in MATLAB. The input information is the metabolic network of the target cell line including detailed information of its biochemical composition and the synthesis of the target product. The output of the method is theoretical culture media formulations for the particular cell line and product.

Objective 5: Apply the method/tool to design *In silico* culture media formulas for Chinese Hamster Ovary (CHO) cells as illustrative case study and compare computed culture media formulations with literature data.

2. SBEMedia: a toolbox for *In silico* culture media design

In this thesis the method for the SBEMedia toolbox was developed for *In silico* experiment-free culture media design. The methodology applied in this thesis is based on the concept of elementary modes and the respective footprint. Fig. 10 illustrates a cell growth elementary flux mode for Chinese Hamster Ovary Cells (CHO). The nodes in green represent the exchange of materials with the extracellular media for the cell to be able to grow according to the particular elementary mode.

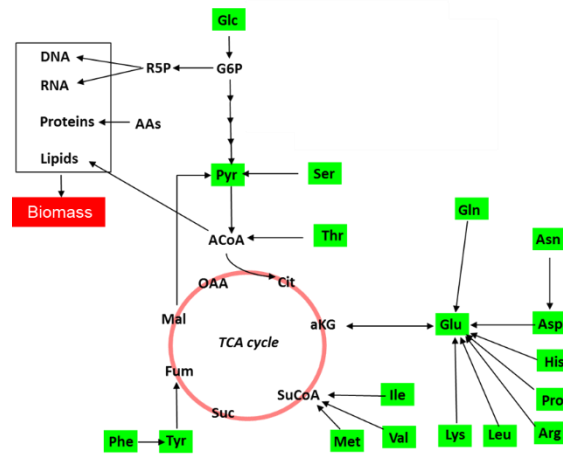


Figure 9 - Example of a Chinese Hamster Ovary Cell elementary mode

The SBEMedia toolbox implements the following sequence of steps to compute *In silico* culture media formulations based on prior knowledge of the metabolic network:

Step 1. Input metabolic network: The metabolic network comprising a list of representative metabolic reactions including i) the synthesis of the target product and ii) the exchange reactions of media components for the synthesis of the target product, are inputted to the system;

Step 2. Computation of elementary modes: The full set of elementary modes of inputted metabolic network are calculated using elementary mode analysis algorithms; the resulting full set of elementary modes displays the full range of possible metabolic states for the cells to realize their biological function.

Step 3. Computation of the footprintome: The metabolic footprint of each elementary mode is computed and merged together in a full matrix comprising all possible modes to interact with the extracellular environment. The full set of footprints is denominated “footprintome”.

Step 4. Automatic *footprintome* reduction: The *footprintome* is reduced based on three steps: eliminating EMs that don't produce the target product (this will later be explain), eliminating EMs that are thermodynamically unfeasible and clustering EMs that have the same phenotype (same list of external metabolites being produced and consumed).

Step 5. Case-dependent *footprintome* reduction based on optimal metabolic criteria. In this step, a subset of elementary modes and respective footprints are selected such as to enforce a desired metabolic state. This step is case dependent and implies the definition of optimal metabolic criteria for the target cell

Step 6. *In silico* culture media formulations: Computation *In silico* of culture media composition based on the final reduced *footprintome*.

All methods were implemented in MATLAB version R2019a⁶⁰. What follows are the details of each step provided.

2.1. Input metabolic network

The metabolic network is inputted as a MATLAB data structure with a list of reactions. To illustrate this, we specify below the metabolic reactions of the simple network of Fig. 6

Network = {

Internal reaction:

'A -->B'

'A -->C'

'B -->P'

'C -->P+D'

'C<-->B'

External/Transport reactions

'-->A'

'<-->B'

'P-->'

'D-->'

}

The exchange reactions, which must be placed in the end of the reactions list, implies a decision on the identity of the culture media components whose concentrations need to be optimised.

A more realistic metabolic network, considering the production of biomass and product is described below in the CHO case study.

2.2 Computation of Elementary modes

The metabolic network is parsed and processed into a MATLAB data structure using the *efmtool* package version 4.7.1. This MATLAB package is freely available for academia and may be accessed at <https://csb.ethz.ch/tools/software/efmtool.html>. The *efmtool* package builds: i) a list of metabolites, separated as intracellular and extracellular, ii) a list of metabolic reactions, separated as intracellular and extracellular, iii) creates the stoichiometric matrix S (Eq. 12) and internal and external stoichiometric submatrices. It also computes the elementary modes and delivers the result in the form of elementary modes matrix EM (Eq. 17). The method used to calculate the EMs is described in detail in Terzer et al.⁶¹. The determination of elementary modes by this method suffers from the computational explosion problem⁶². It can be thus applied only to small and medium scale networks, it cannot be applied to genome scale networks.

2.3 Computation of the *footprintome*

Once the network elementary modes are computed, the next step is the computation of the *footprintome*. The footprint of an elementary mode may be seen as the extracellular phase modification pattern characteristic of the particular elementary mode. This footprint is computed from the subset of exchange reactions, in and out, and their respective weights in the EM matrix. Thus the first procedure is to reorganize the EM matrix by considering the exchange reactions only, and eliminating all others, resulting in Eq. 20.

$$EM^{exch} = \begin{matrix} & \text{EM's} \\ \text{Exchange Reaction} & \begin{bmatrix} X_{11} & & & X_{1j} \\ \cdot & \cdot & & \cdot \\ \cdot & & \cdot & \cdot \\ X_{n1} & & & X_{nj} \end{bmatrix} \end{matrix} \quad (20)$$

The stoichiometric matrix S is also reorganized by considering only the stoichiometry of the exchange reaction (Eq.21).

$$S^{exch} = \begin{matrix} & \text{Exchange Reaction} \\ \text{Metabolites} & \begin{bmatrix} M_{11} & & & M_{1j} \\ \cdot & \cdot & & \cdot \\ \cdot & & \cdot & \cdot \\ M_{n1} & & & M_{nj} \end{bmatrix} \end{matrix} \quad (21)$$

The *footprintome* of the cell results from the multiplication of these two matrices (Eq.22):

$$footprintome = -S^{exch} \times EM^{exch} \quad (22)$$

This operation eliminates all the intracellular metabolites, resulting in macroscopic reactions defining the stoichiometric conversion of extracellular substrates into extracellular products. Note that the minus sign is required because the metabolic reactions are defined in the perspective of the interior of the cell, i.e. the positive direction of an exchange flux is from the extracellular to the intracellular phase. With the minus signal the perspective is flipped, such that all culture media substrates take a negative stoichiometric coefficient, whereas all products of the network take a positive coefficient. The final result is thus a matrix with rows representing culture media components, columns representing elementary mode, and the matrix values are stoichiometric conversion factors.

2.4 Automatic *footprintome* reduction

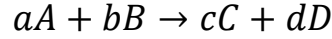
The *footprintome* is typically a large matrix that comprises all possible interaction modes with the culture media. It may be reduced automatically by applying the following criteria:

- ***Elimination based on selected product.*** Footprints that do not produce the target product are not useful for design purpose thus eliminated. The target products are typically biomass and/or recombinant product. As example, if the objective is to design a cell growth culture media, then only elementary modes with biomass production are selected for further analysis while all other eliminated. After this reduction step, the *footprintome* is normalized column wise by dividing the footprint vector by the stoichiometric coefficient of the target product. In this way the footprint of each EM is interpreted as the quantities of substrates consumed and sub-products produced per unit of target product synthesized.
- ***Thermodynamic reduction.*** In this step a thermodynamic elimination of elementary modes based on the Gibbs free energy of reaction of the metabolites in each EM is performed. All elementary modes with positive Gibbs free energy of reaction are considered unfeasible thus eliminated. More details on the calculation of ΔG°_r are provided next.
- ***Clustering based on footprint similarity (Pattern Clustering).*** Many elementary modes footprints have the same phenotype (same list of external metabolites being produced and consumed). These are clustered together.

2.4.1 Calculation of elementary mode Gibbs free energy of reaction (ΔG°_r)

The change in the Gibbs free energy of formation of a substance is the difference between the free energy of a substance and the free energies of its elements in their most thermodynamically stable states at standard-state conditions. The Gibbs free energy of reaction, ΔG°_r , is a measure of reaction spontaneity ($\Delta G^{\circ}_r < 0$) for constant pressure and temperature processes which describes the majority of the biological systems⁶³.

Reaction 1 – Example of a reaction



For reaction 1, the standard Gibbs free energy of reaction is given by Eq. 23:

$$\Delta G^0 r = c * \Delta G^0 f(C) + d * \Delta G^0 f(D) - (a * \Delta G^0 f(A) + b * \Delta G^0 f(B)) \quad (23)$$

Where ΔG^0_f is the metabolite standard Gibbs free energy of formation. ΔG depends only on the difference in free energy of products and reactants (or final state and initial state). ΔG is independent of the path of the transformation⁶⁴. With this in mind, the EMs are considered to be macroscopic reactions that transform a number of extracellular substrates into a number of extracellular products. The ΔG^0_r was calculated for each EMs and whenever $\Delta G^0_r > 0$ (non-spontaneous reactions) the respective EM is eliminated from the list. The metabolite standard energy of formation data were taken from the *eQuilibrator*⁶⁵ database, that uses the group contribution methodologies⁶⁶ to estimate the Gibbs free energy of formation of metabolites.

For IgG the ΔG^0_f value was not found in the literature alongside with that of biomass of CHO cells. For these complex products the ΔG^0_f was computed from the respective reaction of formation in the network of IgG (reaction 78) and biomass (reaction 69) respectively, as shown in Eq. 24.

$$\Delta G^0 f = \sum_i^n a_i \times \Delta G^0 f(M_i) \quad (24)$$

With n being the number of metabolite reactants in the reaction, a_i are stoichiometric coefficients and $\Delta G^0_f(M_i)$ the standard Gibbs free energy of formation of the metabolite i .

2.4.2 Clustering based on phenotype similarity.

In this step, after grouping the footprints together based on their phenotype the centroid position of each cluster is computed based on the arithmetic average of the metabolite values (Eq. 25).

$$F = \frac{\sum_i^n footprint_i}{n} \quad (25)$$

With F being the centroid footprint and n the number of footprints under consideration.

Note that Eq.21 will also be applied to do the arithmetic average of all EM in a *footprintome* in chapter 3.4, “Comparison with experimental data”, alongside a second method, the weighted average of the *footprintome* by the ΔG^0_r of each EM, shown in Eq. 26.

$$F = \frac{\sum_i^n \Delta G^0 r_i \times EM_i}{\sum_i^n \Delta G^0 r_i} \quad (26)$$

2.5 Case-dependent *footprintome* reduction based on optimal metabolic criteria

A much smaller set of elementary modes and respective footprints are selected based on optimal metabolic criteria. The optimal metabolic criteria are case dependent. A typical objective might be the elimination of undesired metabolic byproducts secreted to the culture media. This is achieved by eliminating all elementary modes that produce the particular byproduct. In the CHO case study, hypothetical culture media that eliminate the production of lactate and ammonium is targeted. Another important criteria, especially for animal cell lines, is the media osmolarity. Culture media formulations that lead to a large increase in osmolarity should be eliminated (more to this will be discussed in the case study).

2.6 *In silico* culture media formulations

The final step is the computation of culture media formulas from the remaining EMs. At this stage only a small set of elementary modes survives the elimination/selection process.

It is possible to choose an optimum EM from the remaining set of EMs. For each criteria implemented, elementary modes were given a score value between 0 and 1, dependent on the criteria value of the respective elementary mode. The lower the score value, the better the criteria value of an elementary mode is, compared with other EMs. If the objective is to minimize the criteria value, then Eq. 27 is applied:

$$Score_i = \frac{Criteria_i}{\max(Criteria)} \quad (27)$$

With Score and Criteria, being vectors containing the score and the criteria values of the EMs for the criteria{i} and max(Criteria) being the highest values in the Criteria vector.

In cases where we want to maximize a criteria value, Eq.28 is applied:

$$Score_i = 1 - \left(\frac{Criteria_i}{\max(Criteria)} \right) \quad (28)$$

The Total Score vector, comprises the final score for each EM, it is the sum of the score vectors of all criteria implemented (Eq. 29).

$$Total\ Score = \sum_i^n Score_i \quad (29)$$

With n being the total number of criteria implemented.

The EM with the lowest total score value is chosen as the optimum one. Note that the final culture media formula is a vector whose elements represent the stoichiometric quantities of culture media components per unit of target product produced. To transform it in concentrations we need to multiply the culture media vector by the concentration of the target product, ($c_{product}$), as shown in Eq. 30.

$$C = Op \times c_{product} \quad (30)$$

With C being the concentration vector of metabolites and Op the optimal EM chosen.

3. Results and discussion

3.1 Chinese Hamster Ovary Cells (CHO) metabolic network

To illustrate the application of this method, a Chinese Hamster Ovary (CHO) cell line expressing an antibody (IgG) was selected as a case study. CHO cells are currently the workhorse in the biopharmaceutical industry to produce monoclonal antibodies. It is probably the most important cell line in the biopharmaceutical industry today⁶⁷.

The metabolic network used in this thesis was adapted from the work published by Duarte et al.⁶⁸. In this study the authors adapted and validated the CHO metabolic network from previous published works (Quek et al.⁶⁹, Sengupta et al.⁷⁰, Zamorano et al.⁷¹) and complemented the model using ¹H-NMR (Proton nuclear magnetic resonance) exometabolomic analysis to quantify supernatant metabolites along culture time, under butyrate-treated conditions.

In this thesis the network was simplified by eliminating ATP, NADH, NADPH and FADH₂ and respective oxidative phosphorylation reactions. Thus the simplified metabolic network closes the carbon and nitrogen balances, but does not close the oxygen and hydrogen balances. The simplified CHO metabolic network has 114 reactions (79 intracellular and 35 exchange reactions) and 81 metabolites (46 intracellular and 35 extracellular). The full list of reactions is provided in Appendix A.1. The target network “products” are biomass (X) and antibody (IgG). The synthesis reaction of CHO biomass sets the amounts of compounds in nmol required to synthesize 1 unit of CHO biomass. One unit of CHO biomass corresponds to 10⁶ cells or a CDW of 271 µg.

Reaction 2 - Biomass synthesis reaction

160.1015 Ala + 235.2056 Glu + 70.3787 Gln + 174.6799 Gly + 114.9787 Ser + 147.4132 Lys + 157.4070 Leu + 82.6648 Ile + 91.8543 Arg + 169.492 Asp + 95.7754 Thr + 118.3569 Val + 40.0354 Met + 67.4027 Phe + 47.4956 Tyr + 36.2551 His + 55.559 Pro + 70.3787 Asn + 8.943 AMPRN + 4.878 Cholesterol + 14.9321 CMPRN + 4.0108 dAMP + 2.6829 dCMP + 2.6829 dGMP + 4.0108 dTMP + 0.813 DPG + 75.609 Glycogen + 16.9104 GMPRN + 18.699 PC + 7.046 PE + 0.271 PG + 2.71 PI + 0.813 PS + 2.168 SM + 8.943 UMPRN + 9.2297 Trp --> 1 X

The IgG synthesis reaction reflects the amount of compounds in nmol needed to synthesize 1 mg of IgG.

Reaction 3 – Antibody IgG synthesis reaction

428.7 Ala + 362.75 Glu + 351.76 Gln + 516.64 Gly + 934.36 Ser + 472.67 Lys + 516.64 Leu + 175.88 Ile + 307.79 Arg + 296.8 Asp + 626.57 Thr + 714.51 Val + 65.954 Met + 285.8 Phe + 285.8 Tyr + 164.89 His + 505.65 Pro + 263.82 Asn + 142.9 Trp + 10.992 GDPFuc + 54.962 UDPNAG + 32.977 GDPMann + 21.985 UDPGal + 21.985 CMPSialic --> 1 P

The culture media components that will be subject to optimization in the following section are those present in the exchange reactions with positive influx to the intracellular phase. These are:

Glucose (Glc), Glutamate(Glu), Serine(Ser), Lysine(Lys), Leucine(Leu), Isoleucine(Ile), Arginine(Arg), Aspartate(Asp), Threonine(Thr), Valine(Val), Methionine(Met), Phenylalanine(Phe), Tyrosine(Tyr), Histidine(His), Proline(Pro), Choline, Asparagine(Asn), Tryptophan(Trp) and Pyruvate(Pyr).

In the following sections, custom culture media will be developed for the specific cell line growth specified by reaction (2) and for the specific IgG specified by reaction (3) by determining the optimal quantities of the exchange compounds listed above. Obviously, these exchange compounds do not cover the complete set of media components in a typical CHO culture media, which has typically more than 100 components. In future studies, this network could be enlarged to include additional media exchange components.

3.2 Computation of CHO Elementary Modes

The complete set of elementary modes was computed for the CHO metabolic network using the *efmtool*⁷². The elementary modes count of medium and large metabolic networks can be very high, in the order of millions⁷³. To facilitate the computation of elementary modes, it was considered two different scenarios:

- i) The metabolic network includes the biomass synthesis reaction only (without the IgG synthesis) resulting in the biomass elementary modes (biomass-EMs).
- ii) The metabolic network includes the IgG synthesis reaction only (without the biomass synthesis reaction) resulting in the product elementary modes (product-EMs).

By removing the biomass synthesis and IgG synthesis reactions of the network, further reactions needed to be removed in order to not create metabolic dead ends for specific intracellular metabolites. As a result, for the biomass producing network the IgG glycosylation reactions were also removed, resulting in a 106 reaction network (Appendix A, Table A.2), for the IgG producing network the lipid synthesis reactions were removed, as well as some reactions in the nucleotide synthesis and glycogen synthesis metabolisms, this also led to the removal of the Choline and Glyc3PC exchange reactions. The final number of reactions in the IgG producing network was 92 (Appendix A, Table A.3)

The final count of Elementary Modes obtained is shown in Table 4

Table 4 Elementary mode count of the biomass and product producing networks.

	Without ATP/NADH/NADPH/FADH2	With ATP/NADH/NADPH/FADH2
Biomass-EMs	218 538	1 987 460
Product-EMs	313 523	1 768 927

To note that the inclusion of oxidative phosphorylation reactions increases very expressively the EMs count. The overall computation time for the biomass producing network increases from 30,420s to 267,63s, and in the IgG producing network from 27,72s to 206,80s, roughly a 10-fold increase in both networks. Although the computation time of EM for the more complex metabolic network wouldn't take a deterrent amount of time for the making of this work the computation of the respective *footprintome* visualizations would. For this reason, we have adopted in the proceeding studies the simplified version of the metabolic network.

Note also, that the number of final EMs is not dependent on the stoichiometry of the biomass or IgG synthesis reactions.

As an illustrative example, Appendix B, Table B.1, represents one EM from the biomass producing network that uses the minimum number of metabolic reactions required to synthesize 1 unit of biomass. This EM is potentially very efficient for cell growth using the minimum regulatory resources to synthesize one unit of biomass. Even so, only 42 from the 106 reactions in the network do not participate in this EM. By multiplying the elementary modes coefficients by the exchange reactions stoichiometry (reactions 73 to 106), one gets the EM footprint shown in Table 5. The metabolites in green are end-products secreted to the culture media, whereas the red metabolites are substrates needed to synthesize biomass. The footprint may be interpreted as the resources needed to synthesize one unit of biomass and respective by-products if the cells grows according to the EM shown in Appendix B.

Table 5 – Footprint of the EM represented in Appendix B

<i>Metabolite</i>	<i>Footprint</i> (nmol/10 ⁶ cell)
<i>X</i>	1,00
<i>Glc</i>	-269,67
<i>His</i>	-37,17
<i>Isobut</i>	0,00
<i>Ile</i>	-84,75
<i>Isoval</i>	0,00
<i>Leu</i>	-161,33
<i>Lys</i>	-151,08
<i>Met</i>	-41,00
<i>Phe</i>	-254,08
<i>Thr</i>	-323,83
<i>Trp</i>	-9,50
<i>Val</i>	-121,33
<i>CO2</i>	972,83
<i>NH4</i>	0,00
<i>Acetate</i>	0,00
<i>Ala</i>	0,00
<i>Arg</i>	-311,33
<i>Asn</i>	-72,17
<i>Asp</i>	-342,50
<i>Gln</i>	0,00
<i>Cit</i>	0,00
<i>Choline</i>	21,42
<i>Formate</i>	0,00
<i>Glu</i>	-232,33
<i>Glyc</i>	0,83
<i>Gly</i>	0,00
<i>Lac</i>	0,00
<i>Pro</i>	-56,92
<i>Pyr</i>	0,00
<i>Ser</i>	-128,17
<i>Tyr</i>	0,00
<i>Mal</i>	0,00
<i>Glyc3PC</i>	0,00

3.3 Computation and reduction of CHO *footprintome*

The *footprintome* may be defined as, the complete set of footprints a cell may resort to, realize its physiological objectives. The *footprintome* is computed by repeating the process described in the previous chapter for the complete set of EMs. The footprint of each EM (a column vector) are stacked together in a large matrix in the final form of the *footprintome* matrix. Each footprint is a potential culture media “candidate”, thus it is imperative to reduce as much as possible the size of the *footprintome* by eliminating nonessential footprints. The full set of EMs undergoes three steps of automatic reduction as described below for the biomass-EMs (A similar reduction procedure was applied to the IgG-EMs with final results presented in Appendix F Fig. F.1-F.2) as follows:

- The biomass-EMs (count = 218538) contain several EMs without biomass production. These were eliminated reducing the total EMs count to 211580. All columns of the *footprintome* are afterwards normalized by dividing by the biomass coefficient respectively. As a result, the biomass rows are always one, whereas the other rows read as nutrient consumption/production per unit of biomass production.
- The next step is the thermodynamic reduction. The $\Delta G^{\circ r}$ was computed for each EM and those with nonnegative $\Delta G^{\circ r}$ were eliminated. This step further reduced the EMs count to 203137. The result of this reduction is shown in Fig. 11. The $\Delta G^{\circ f}$ values of each metabolite can be found in Appendix C.
- The next step is *footprintome* reduction by clustering based on phenotype similarity. Footprints that have the exact same list of substrates and end-products are clustered together. These clusters are averaged into a representative footprint by the arithmetic average (Eq. 25). This procedure ensures that the reduced set of footprints centroids (for simplicity sake during this work we will still be calling the footprints centroids as only footprints, although they are no longer representative of a singular EM) are linearly independent and characteristic of a unique cellular phenotype. This reduces the footprints count to 3488. This step was the one that achieved a higher degree of reduction. The result of this reduction is shown in Fig. 12

These results are summarized in Table 6.

Table 6. Automatic *footprintome* reduction results

	BIOMASS-EMS	IGG-EMS
BEFORE REDUCTION	218538	313523
ELIMINATION BASED ON SELECTED PRODUCT	211580	307404
THERMODYNAMIC REDUCTION	203137	307402
CLUSTERING	3488	4704

After the automatic reduction steps, 3488 potential cell growth footprints remain, which still is very high to test in the lab. Nevertheless, in a scenario of abundant resources it would be feasible to test them in a high throughput cell culture equipment such as the Sartorius AMBR® system.

Other studies in the literature have attempted EM reduction using different criteria. Hyun-Seob Song et. al.⁷⁴ proposed a reduction method based on yield analysis, in which the authors also compute EMs ratios of specific metabolites and use it to characterize different EMs phenotypes. The “yield analysis” method reduces EMs count to a small set representative of 99% of all phenotypic states. This reduction is however based on experimental data. Folch-Fortuny et al.⁷⁵ proposed a method for discrimination of active EMs based on fluxomic datasets and tested it in *Escherichia coli* and *Pichia pastoris* cultures. It reduces the full set of EMs into a small number of representative EMs based on independent measurements of metabolic fluxes. This method explores the analogy of EM and principal component analysis of a measured fluxomics dataset. Ferreira et al. (2011)⁷⁶ proposed a method to discriminate active EMs also based on experimental measurements of metabolic fluxes. A reduced set of EMs is discriminated by maximizing the correlation of the EM weighting factor and measured media composition.

All the above methods are good options for EMs reduction when sufficient measured data of metabolic fluxes and/or metabolic footprints are available. Therefore, they could not be applied in this thesis as the objective in this thesis is to design *In silico* (100% experiment-free) the composition of culture media.

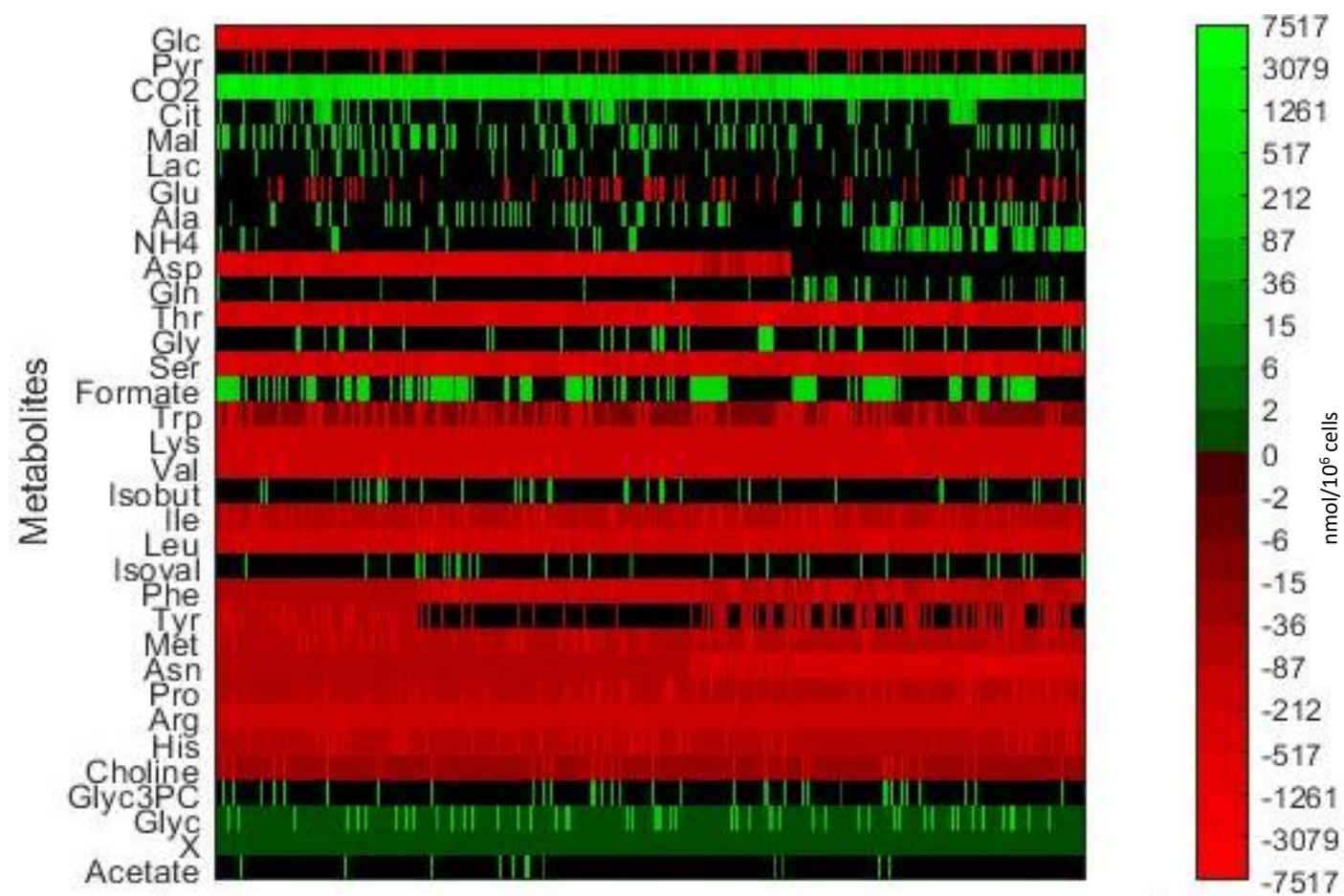


Figure 10 - Reduced footprint for biomass production (normalized to unit) after 2 steps of reduction: Step 1- biomass production, Step 2-thermodynamic reduction. The columns represent the reduced set of EMs (203137). The colour green means that the compound is being produced by the EM. The colour red means it's being consumed by the EM. The black means it is neither consumed nor produced by the EM.

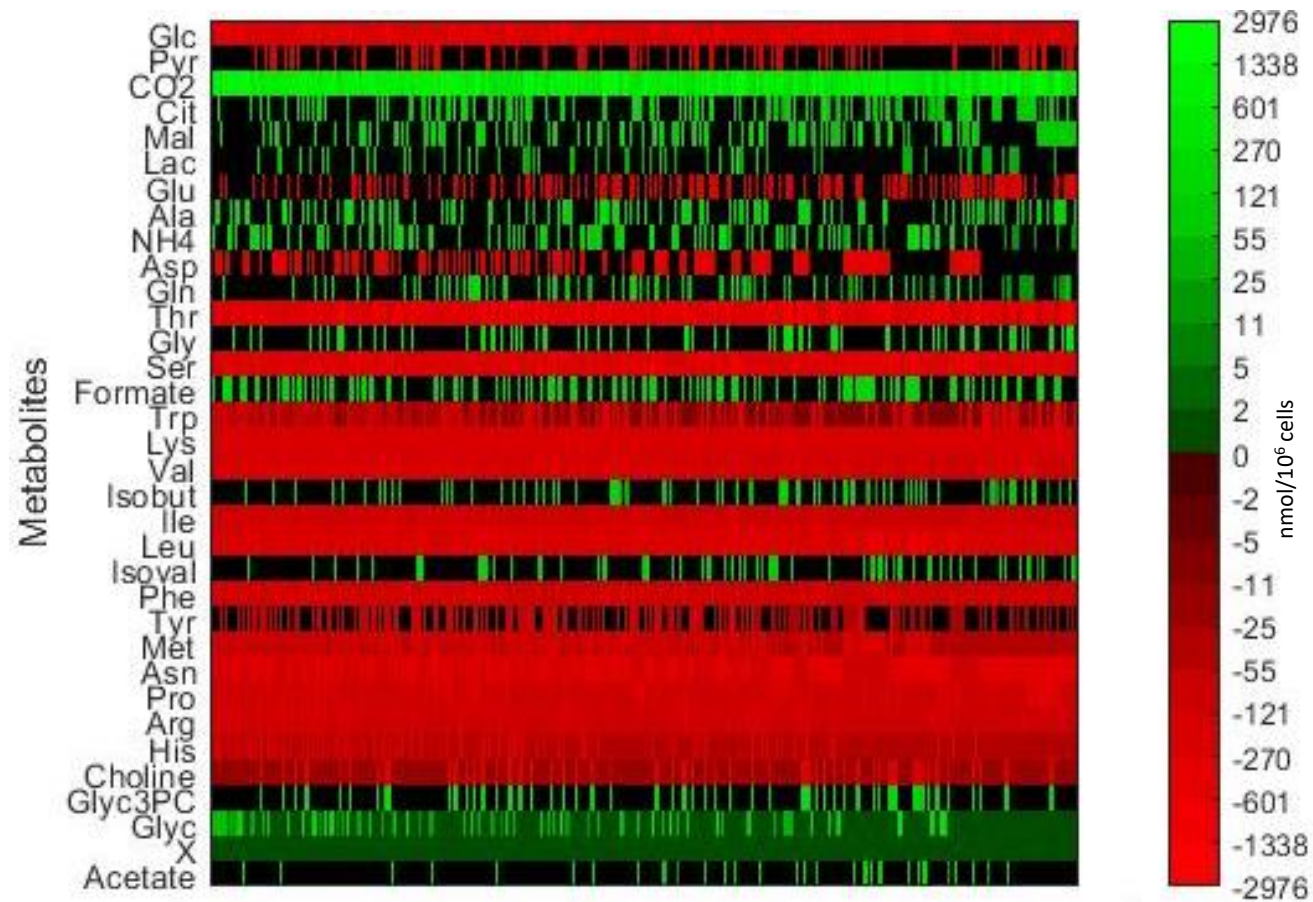


Figure 11 - Reduced footprint for biomass production (normalized to unit) after 3 steps of reduction: Step 1- biomass production, Step 2- Thermodynamic reduction, Step 3- Pattern clustering with arithmetic averaging. The rows represent extracellular media components. The columns represent average EM clusters (3488). The green gradient represents compound production associated with the production of 1 biomass unit. The red gradient represents compound consumption to generate 1 biomass unit. The black colour means that the compound is neither consumed nor produced by the EM cluster.

3.4. Comparison with experimental data

The objective here is to compare the theoretical *footprintome* (computed from the metabolic network) with measured data of uptake/production rates of media compounds in CHO cultures (measured footprint). More specifically, the data published by Duarte et. al. ⁷⁷, who studied the metabolic response of CHO cells metabolism in experiments with varying concentration of asparagine and serine in the culture media, this data was adopted to validate the theoretical biomass growth *footprintome* after thermodynamic selection (Fig. 11). In order to compare the theoretical *footprintome* to a measured time point, it is required to average the full set of footprints into a representative footprint. Two different averaging methods were applied:

- i) Arithmetic average of all EM footprints (Eq. 25) in the *footprintome* of Fig. 11. The idea behind this method is that the number distribution of nutrient requirements in the *footprintome* is a measure of probability, thus the culture media formulation of a particular metabolic model corresponds to the mean concentration based on number.
- ii) Weighted average by the ΔG° of each EM footprint (Eq. 26) in the *footprintome* of Fig. 11. The idea behind this method is that thermodynamically more favourable EMs have a higher contribution to the observed footprint.

Fig. 13A shows the overall results for the arithmetic average and the weighted average, compared with the measured footprint data (Fig. 13B), that was obtained from the raw data published by Duarte et. al. (2014) (Appendix D), normalized by the cell growth rate.

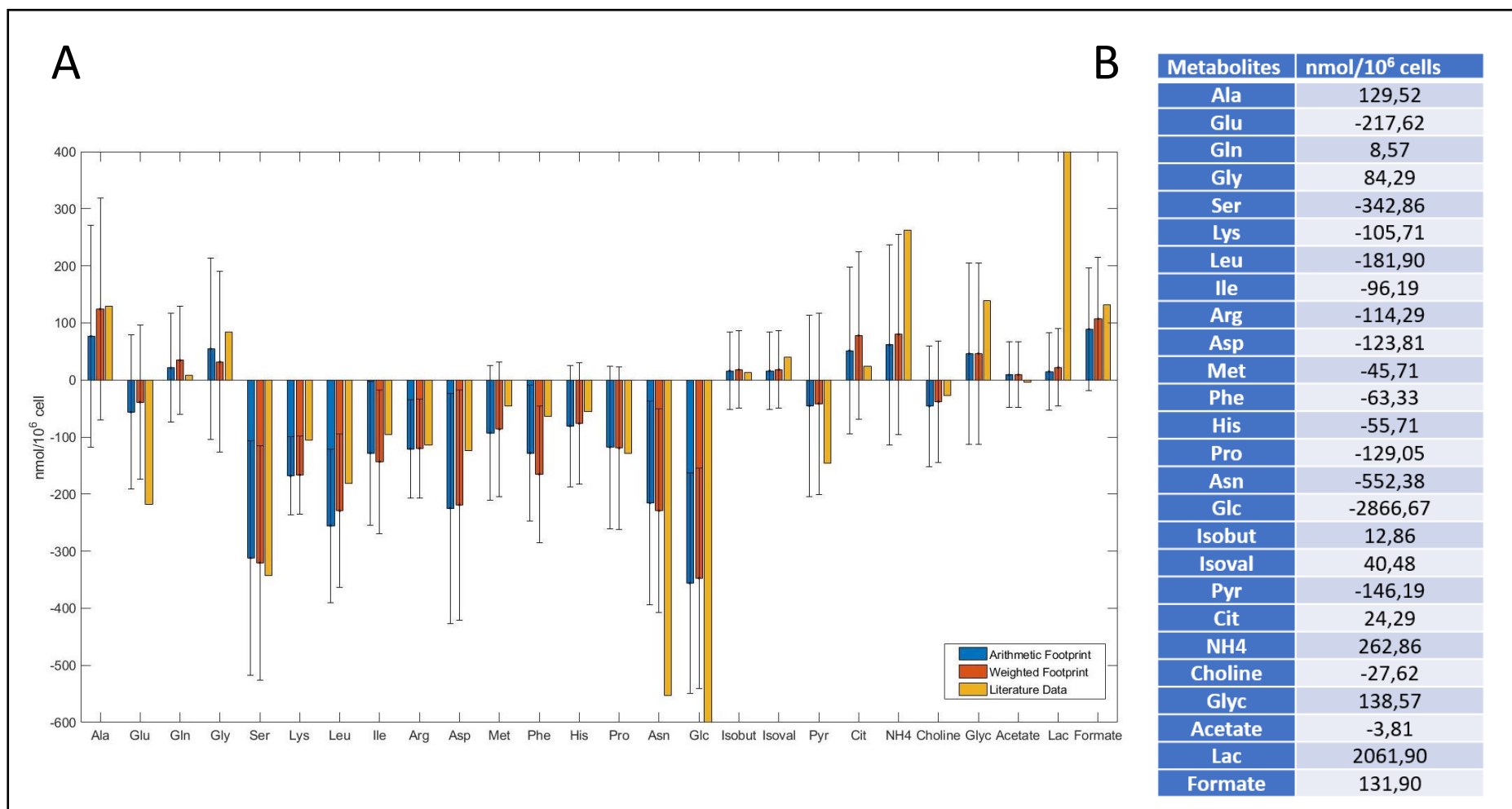


Figure 12 - (A) Comparison between theoretical footprints computed by the arithmetic average method (blue bar) and the weighted average method (orange bar) with experimental footprint (yellow bar). (B) Experimental footprint taken from Duarte et. al. (2014)⁵⁸. The error bars correspond to the standard deviation of footprints in relation to the average

All in all these results show a remarkable concordance between the theoretical footprint and the measured footprint. Moreover, the measured footprint is within the variance interval of the theoretical footprints computed for most nutrients (exceptions discussed below) as shown by the error bars in Figs. 16A. It should be highlighted that the theoretical footprint was computed resorting to the metabolic network information only. It is completely *In silico* and experiment free.

The broad error bars displayed in Figs. (13A) reflect the variability in theoretical footprints. Even if it is not possible to know *a priori* without any experimental evidence which are the “active EMs” used by the cells to grow, it is possible to determine the average theoretical footprint and theoretical variability around the mean.

The only large theory-measurement mismatch observed in Fig. 13A relates to the glucose and lactate nutrients. The measured footprint is characterized by a very high glucose consumption yield ($-2866,67 \text{ nmol}/10^6 \text{ cells}$) which is linked with a very high lactate production yield ($2061,90 \text{ nmol}/10^6 \text{ cells}$). This type of glucose overflow metabolism is typical of high glucose concentration in the media, which is indeed the case in the experiments described by Duarte et al.⁷⁷, where the cell culture was initially fed with 50mM of glucose. The high lactate production can be prevented with a low and steady feed of glucose as shown by Fan Y et al.⁷⁸ Very likely, if these experiments were repeated with low glucose concentrations, the glucose and lactate data would likely be concordant with the theoretical footprint.

To compare the arithmetic and weighted averaging methods, the Mean-Square-Error (MSE) between theoretical and experimental footprints were computed. The glucose and lactate data were considered outliers in the computation of the MSE. The results are shown in Table 7.

Table 7 MSE between measured and theoretical footprint

METHOD	MSE
ARITHMETIC FOOTPRINT	9.0320e+03
WEIGHTED FOOTPRINT	8.7426e+03

Although the weighted average method presented a lower MSE then the arithmetic method, the difference is not very significative. Comparing the arithmetic footprint with the weighted footprint we can see that the differences between the footprints obtained by both methods are not significative. This could suggest a low variance in the ΔG_r^0 values of EMs. Fig 14 displays the computed ΔG_r^0 distribution, which is clearly a non-uniform distribution. This distribution has a mean value of $-4.0756e-04 \text{ kJ/nmol}$ and a variance coefficient of 73,29% which is very significant. It is not possible with these results to make a final conclusion on which method is better to compute the footprint given the small differences obtained. It is also important to mention that in simplifying the metabolic network we also influenced the ΔG_r^0 distribution of the EM. More studies are required in the future before a final conclusion can be taken.

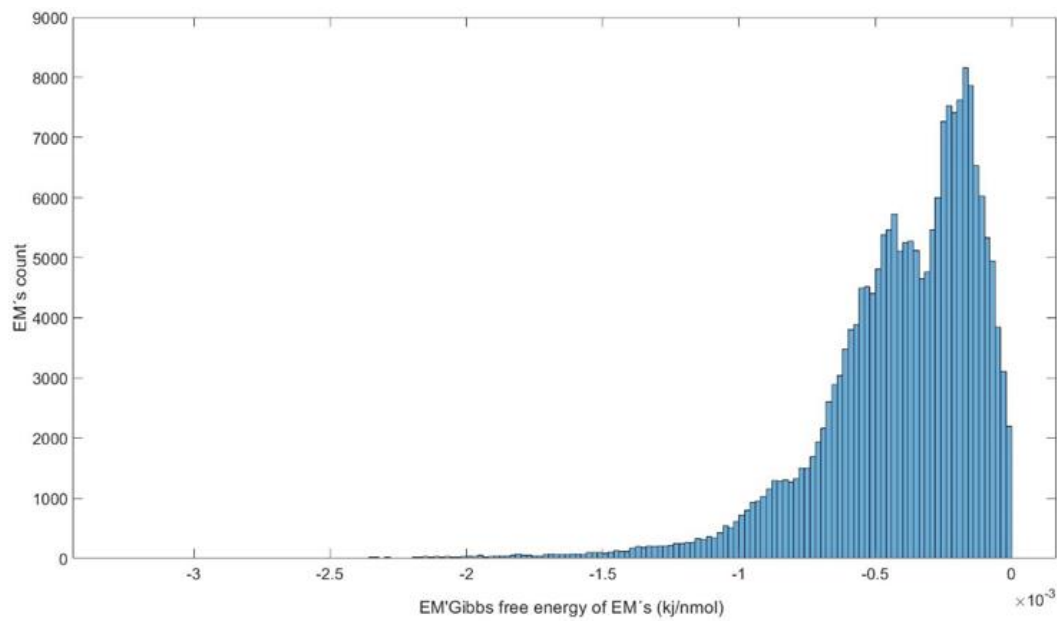


Figure 13 - Histogram ΔG°_r of each EM for the biomass producing network after the thermodynamic reduction step (Fig. 11)

3.4. Phenotype-targeted culture media design

The objective in this section is to design *In silico* culture media formulations that are targeted to a given desired phenotype. In here the following phenotypic optimal criteria were chosen

- Minimization of lactate accumulation
- Minimization of ammonium accumulation
- Minimization of osmolarity build-up
- Minimization of CO₂ production

The methodology adopted is to further reduce the *footprintome* by eliminating undesired footprints that do not comply with the above enumerated criteria.

3.4.1 Minimization of lactate and ammonium buildup

Lactate and ammonium are by products that when accumulated to high concentrations in the media are toxic to mammalian cells. Lactate production is the by-product of an inefficient catabolism, as it only produces 2 ATP molecules compared to the 36 ATP molecules that result from the full oxidation of glucose in the TCA cycle. Lactate also acidifies the media and causes high osmolarity⁷⁹, which reduces specific growth rate⁸⁰ and protein yield⁸¹. High ammonium concentration in the media has also a detrimental effect on the cell culture, also reducing specific growth⁸⁰ and protein yield⁸². It is therefore of high interest to design culture media formulations that minimize the accumulation of lactate and ammonium in the media.

Taking a closer look to the reduced *footprintome* for biomass production (Fig. 12) and for IgG production (Appendix F, Fig. F.2), we observe that they contain a high number of footprints that produce either lactate, ammonium or both metabolites simultaneously. In order to design a culture media that theoretically eliminates the accumulation of lactate and ammonium in the media, a phenotype targeted *footprintome* reduction is applied by removing all footprints that produce lactate and/or ammonium, as follows:

- For the biomass producing network, of the 3488 footprints present in the in the *footprintome* of Fig. 12, 1549 footprints either produce lactate and/or ammonium, and were therefore removed. The footprint count after this step is 1939.
- For the IgG producing network, of the 4704 footprints present in the *footprintome* of Appendix F (Fig. F.2), 1804 footprints produce lactate and/or ammonium and were therefore removed. The footprints count after this step is 2900.

Fig. 15 represents the biomass producing *footprintome* after the elimination of the lactate and/or ammonia producing EMs.

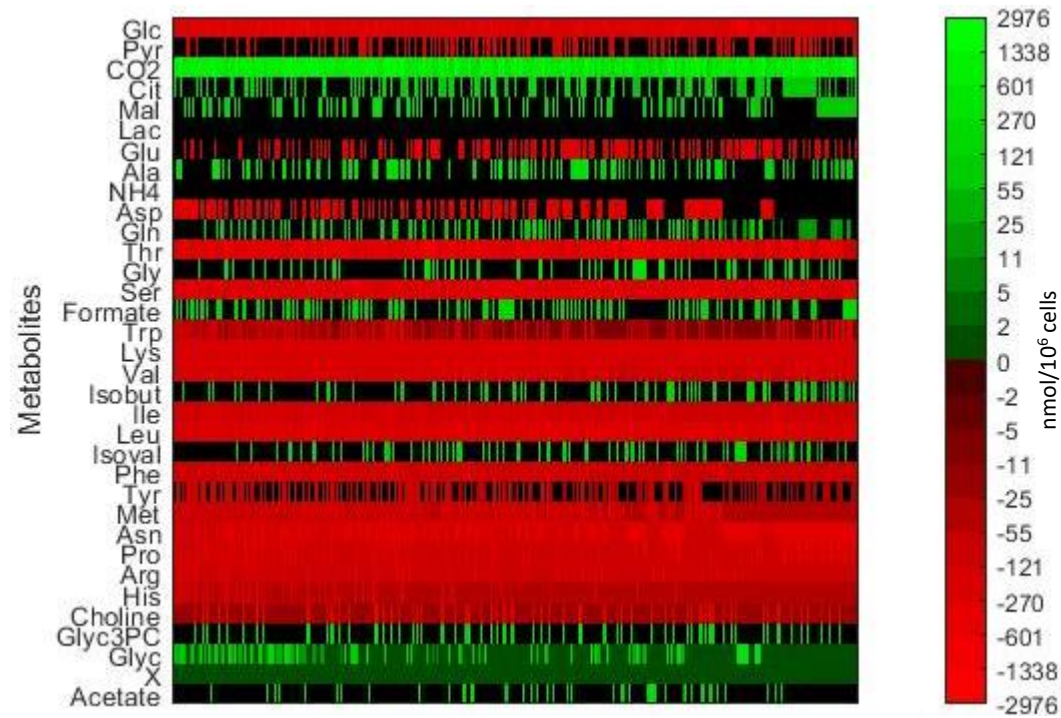


Figure 14 - Reduced footprintome for biomass production (normalized to unit) after 4 steps of reduction: Step 1- biomass production, Step 2- Thermodynamic reduction, Step 3- Pattern clustering with arithmetic averaging, Step 4- Elimination of all footprints that either produce lactate or ammonium. The rows represent extracellular media components. The columns represent average EM clusters (1939). The green gradient represents compound production associated with the production of 1 biomass unit. The red gradient represents compound consumption to generate 1 biomass unit. The black colour means that the compound is neither consumed nor produced by the EM cluster.

3.4.2 Minimization of osmolarity buildup

The next criteria considered are the osmolarity and CO₂ of each footprint. In the study by Marie M. Zhu et.al⁸³ it is shown that osmolarity and partial pressure of CO₂ have a very significant impact in a large scale CHO cell culture. High osmolarity in the culture media is shown to have a negative effect on specific growth rate and viable cell density. Moreover, it causes a shift in CHO cells metabolism leading to an increased production of lactate and ammonia. Furthermore, the combined effects of high partial pressure of CO₂ (pCO₂) and high osmolarity caused a more prominent effect on viable cell density than just high osmolarity alone.

The osmolarity build-up associated with each footprint was computed. The osmolarity value of a footprint was approximated by the absolute sum of metabolite values (Eq.31). As we are dealing with metabolite yields and we lack any sense of volume, these osmolarity values don't describe the real culture formula osmolarity. They can be however objectively compared between each other.

$$\sum_i^n abs(footprint_i) \quad (31)$$

Where n is the total number of metabolites in a footprint and $abs(footprint_i)$ is the absolute value of a metabolite i .

For the minimization of the osmolarity and CO₂ criteria footprints associated with low osmolarity build-up and low CO₂ production were selected. For this purpose, a score value ($Score_i$) is computed for each footprint $\{i\}$ according to the Eq. 32:

$$Score_i = \frac{Osm_i}{\max(Osm)} + \frac{CO2_i}{\max(CO2)} \quad (32)$$

Where, Osm_i is the osmolarity build-up associated with footprint $\{i\}$, Osm is a vector containing the osmolarity values of all footprints, $CO2_i$ is the CO₂ production by footprint $\{i\}$ and $CO2$ is a vector containing the CO₂ values of all footprints. The EM with the lowest score is the one chosen for culture media design. Figs. 16-17 show the final optimal footprint with the lowest score, for the biomass producing network (Fig. 16) and IgG producing network (Fig. 17) in comparison with their respective arithmetic footprints representatives of the *footprintome* (Fig. 13).

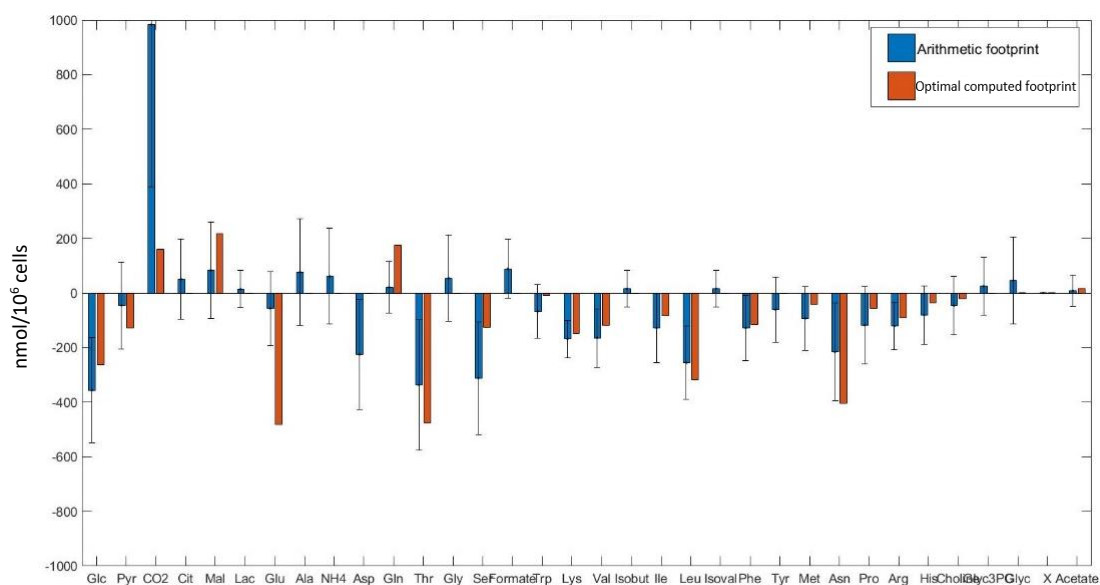


Figure 15 - Comparison between the optimal EM (orange bars) computed for the biomass producing network and the respective arithmetic footprint (blue bars).

By comparing the optimal EM with the respective arithmetic footprint (Fig. 16, for biomass production), it becomes evident the absence of lactate and ammonium production, as expected. The same can be observed for the IgG producing network (Fig.17).

In term of substrate consumption, Fig. 16 shows that by minimizing osmolarity some substrates have a lower yield (lysine, valine, proline, histidine etc.) compared with the arithmetic footprint while others aren't consumed at all. The most prominent case is the metabolite aspartate, where in the optimal EM no consumption is predicted whereas its consumption yield is 224,78 nmol/10⁶ cells for the arithmetic footprint. The main differences in substrate consumption are the lower glucose consumption of 263,06 nmol/10⁶ cells compared to 356,20 nmol/10⁶ cells, a higher glutamate consumption, 482,09 nmol/10⁶ cells in comparison with 56,52 nmol/10⁶ cells, an also higher consumption of asparagine 404,50 nmol/10⁶ cells compared with 215,80 nmol/10⁶ cells and a lower consumption of serine ,125,01 nmol/10⁶ cells while the arithmetic footprint being 312,07 nmol/10⁶ cells.

It may also be observed a lack of by-products formation in the optimal EM. Examples of this are the, isobut, isoval, alanine, glycine and formate. On the contrary, malate and glutamine are increased. Also the CO₂ production is much lower in the formula chosen as a result of minimizing CO₂ production, being 161,46 nmol/10⁶ cells compared to the arithmetic footprint of 983,73 nmol/10⁶ cells. The full list of results can be found in Appendix E, Table E.1

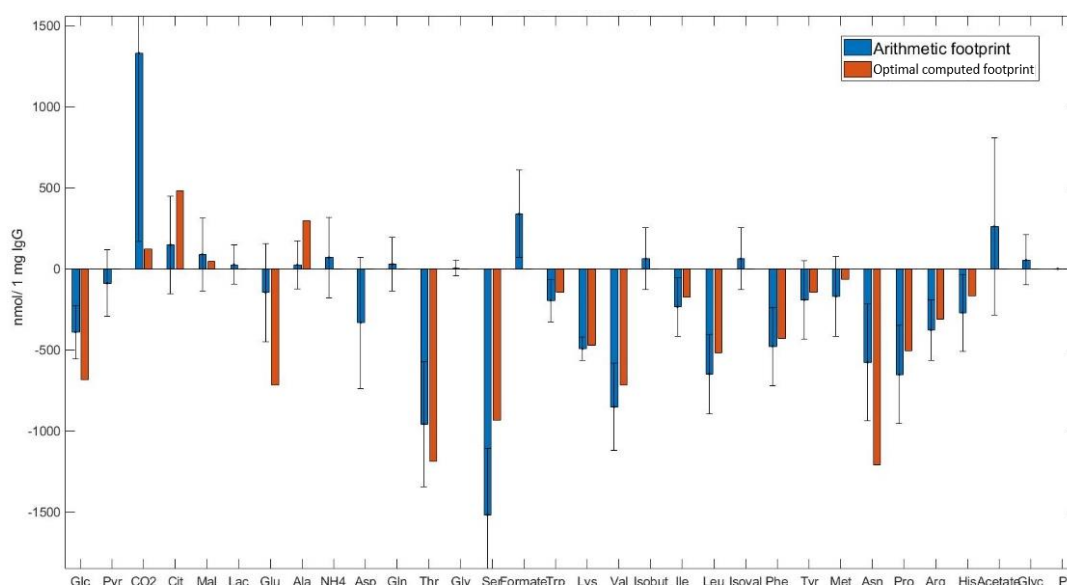


Figure 16 - Comparison between the optimal EM computed (orange bars) for the IgG producing network and the respective arithmetic footprint (blue bars).

The effects of minimizing osmolarity and CO₂ had similar results for the product producing network, with substrates having lower consumption (lysine, valine, proline, histidine etc.), the non-consumption of aspartate and lack of by-product formation (isobut, isoval, glycine, formate) and an increase in specific by-products productions (citrate and alanine). The CO₂ production also dropped sharply compared with the arithmetic footprint from 1330,80 nmol/10⁶ cells to 120,92 nmol/10⁶ cells, interestingly the glucose has a higher consumption yield in the optimal EM with 681,53 nmol/10⁶ cells compared with 390,55 nmol/10⁶ cells. The glutamate, serine and asparagine yields show similar behaviour to the biomass optimal EM. It should however be noted that the amount of glucose used for IgG production should be much less than that for biomass production. The full list of results can be found in Appendix E, Table E.2

3.4.3 Final culture media concentrations

In order to formulate the final optimal culture media composition, the concentrations of each compound need to be specified. The footprints are not ready to be used as culture media formulas because they are expressed as yields. As such, they need to be multiplied by a concentration value of the target product (Eq.30).

The concentration used for the optimal IgG producing EM was 0,56 mg of IgG/ml, this being the mean value of the results from the work of Reinhart, D, et. al.⁸⁴ where they benchmarked several commercial CHO culture medias for IgG antibody production, measuring the IgG concentration in all tested medias.

The concentrations for the optimal biomass producing EM were obtained by initially multiplying the computed formula by a cell concentration of 10 10⁶ cell/ml the formula. It was then compared with a lab tested CHO culture media. The cell concentration was adjusted such as to minimize the MSE between the computed and experimental concentrations. The final cell concentration calculated was 17,55 10⁶ cell/ml (still in reasonable range) for a MSE value of 42,88.

The comparison between lab tested CHO culture media with the final computed formula for biomass synthesis is shown in Fig 18, these results are also shown in Table 8 alongside the computed formula for IgG synthesis.

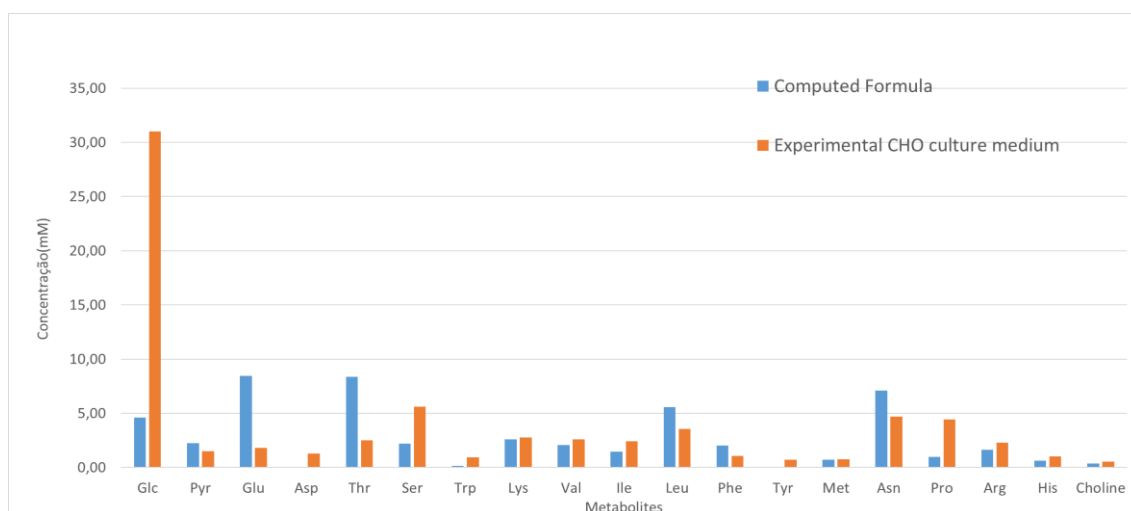


Figure 17 - Comparison between the computed formula for biomass synthesis (blue bars) with experimental CHO culture media (orange bars).

Fig.18 shows many similarities between the *In silico* formula for biomass synthesis (blue bars) and the lab tested CHO culture media (orange bars). The metabolites with more concordant concentrations (between *In silico* and experiment) were lysine, valine, histidine, arginine and choline. A significant difference is observed for glutamate and threonine concentrations, with higher *In silico* concentrations (8,46 mM and 8,35 mM respectively) than the experimental concentrations (1,79 mM and 2,50 mM respectively). The largest mismatch is a much lower *In silico* glucose concentration (4,62 mM) than the experimental concentration (31,01 mM). Glucose is the preferred carbon source by CHO cells. In batch runs, high amounts of glucose are usually formulated in the culture media. A lower glucose concentration in the *In silico* formula suggests that it needs to be complemented with a continuous feeding strategy of glucose.

Table 8 Computed formulas for IgG and biomass syntheses and Lab tested CHO culture media

<i>Metabolites</i>	<i>Computed formula for IgG synthesis (mM)</i>	<i>Computed formula for biomass synthesis (mM)</i>	<i>Lab tested CHO culture media (mM)</i>
<i>Glc</i>	0,38	4,62	31,01
<i>Pyr</i>	0,00	2,23	1,52
<i>Glu</i>	0,40	8,46	1,79
<i>Asp</i>	0,00	0,00	1,29
<i>Thr</i>	0,66	8,35	2,50
<i>Ser</i>	0,52	2,19	5,60
<i>Trp</i>	0,08	0,16	0,93
<i>Lys</i>	0,26	2,59	2,78
<i>Val</i>	0,40	2,08	2,58
<i>Ile</i>	0,10	1,45	2,41
<i>Leu</i>	0,29	5,57	3,54
<i>Phe</i>	0,24	2,02	1,05
<i>Tyr</i>	0,08	0,00	0,70
<i>Met</i>	0,04	0,70	0,74
<i>Asn</i>	0,67	7,10	4,71
<i>Pro</i>	0,28	0,97	4,42
<i>Arg</i>	0,17	1,61	2,28
<i>His</i>	0,09	0,64	1,00
<i>Choline</i>	-	0,37	0,52

4. Conclusions

In this work a method was developed capable of computing culture media *In silico* using only a metabolic network of a target cell line as input and elementary mode analysis (EMA). The Chinese Hamster Ovary (CHO) metabolic network was used as a case-study in order to validate our approach. The proposed tool is an alternative to empirical and wet lab intensive methods used in the industry, like RSM.

The main challenge in using EMA for *In silico* culture media design, is the elementary mode computation combinatorial explosion, the computational burden associated with computing all EM in the case study CHO metabolic network was too high. As a result, we used a more simplistic CHO metabolic network without the respective oxidative phosphorylation reactions, closing the balances for carbon and nitrogen but not for oxygen and hydrogen, which is sub-optimal. We want the computed footprints values displayed in the *footprintome* to be as representative of the different metabolisms of the targeted cell line as possible, the more complete the metabolic network in use is the better, a genome scale network would be ideal but it's not feasible due to the computation power needed to calculate all possible millions of EMs in a network of this scale.

Even with these drawbacks this study obtained optimistic results:

- The arithmetic and the weighted footprints when compared with the footprints calculated from the experimental literature data⁷⁷ (Fig. 13A), shows that the majority of metabolite yields values in the literature data, are on the variance bounds of the theoretical footprints computed (21 out of the 26 metabolites). Which indicates that a general predictions of the metabolism footprint of a target cell line can be done using this method
- The phenotype target design has shown to be a useful method for selecting optimal metabolic footprints, the differences between the optimal computed EMs for both networks and the arithmetic footprints of the respective networks were concordant with the selective criteria chosen.
- The similarities observed between the optimal EM for the biomass producing network with a lab teste CHO culture media (MSE value of 42,88 for a cell concentration of $17,55 \cdot 10^6$ cell/ml), shown that this tool is capable of computing reasonable culture media formulas that can be further tested in the lab.

The present M.Sc. thesis had the objective to develop a method in MATLAB capable of computing culture media *In silico* using only the targeted cell line metabolic network, this objective was reached, the positive results shown that further studies regarding the usage and development of the methods present in this thesis are worth pursuing.

4.1 Future work

The arithmetic and the weighted footprint shown similar results when compared with the data published by Duarte et. al.⁷⁷, as already discussed, but we could not conclude which of these methods is better and further research is needed. This methodology should be tested with different cell lines metabolic networks and the results compared with available literature data, to better understand the difference between the two methods and to confirm if the results observed in this thesis are also true for different cell lines.

There are several options for future work regarding the phenotype-targeted culture media phase:

- The culture media for growth and IgG production computed in this thesis, should be tested in the lab to quantify the differences between the computed formulas and the experimental results, to study if the computed culture medias can have a decisive effect in controlling the active EMs of the CHO cell metabolism.
- Different culture media formulas should be computed using different criteria (e. g. minimization of glucose consumption) to study which criteria or combination of criteria can be used to better match the experimental results, the number of criteria used can also be a decisive factor.
- Different metabolic networks can be tested to understand more the potential of the toolbox and to further develop it.

5. References

1. Plackett RL and Burman J . P. THE DESIGN OF OPTIMUM MULTIFACTORIAL EXPERIMENTS Design of optimum multifactorial experiments. *Biometrika*. 1946;33(4):305-325.
2. Box AGEP, Wilson KB, Journal S, Statistical R, Series S. On the Experimental Attainment of Optimum Conditions Published by : Wiley for the Royal Statistical Society Stable URL : <http://www.jstor.org/stable/2983966>. 1951;13(1):1-45.
3. Taguchi G, Rafanelli AJ. *Taguchi on Robust Technology Development: Bringing Quality Engineering Upstream*. Vol 116.; 1992. doi:10.1115/1.2905506
4. McCulloch WS, Pitts W. A logical calculus of the ideas immanent in nervous activity. *Bull Math Biophys*. 1943;5(4):115-133. doi:10.1007/BF02478259
5. Glassey J, Montague GA, Ward AC, Kara B V. Artificial neural network based experimental design procedure for enhancing fermentation development. *Biotechnol Bioeng*. 1994;44(4):397-405. doi:10.1002/bit.260440402
6. Mitchell M (Computer scientist). *An Introduction to Genetic Algorithms*. MIT Press; 1996.
7. Weuster-Botz D, Wandrey C. Medium Optimization by Genetic Algorithm for Continuous Production of Formate Dehydrogenase. *Process Biochem*. 1995;30(6):563-571. doi:10.1016/0032-9592(94)00036-0
8. Stephanopoulos G. Metabolic Fluxes and Metabolic Engineering. *Metab Eng*. 1999;1(1):1-11. doi:10.1006/mben.1998.0101
9. Reder C. Metabolic control theory: A structural approach. *J Theor Biol*. 1988;135(2):175-201. doi:10.1016/S0022-5193(88)80073-0
10. Stephanopoulos G, Aristidou AA NJ. *Metabolic Engineering PRINCIPLES AND METHODOLOGIES*. (Academic SD, ed.); 1998.
11. Kauffman KJ, Prakash P, Edwards JS. Advances in flux balance analysis. *Curr Opin Biotechnol*. 2003;14(5):491-496. doi:10.1016/j.copbio.2003.08.001
12. SCHUSTER, S., & HILGETAG C. ON ELEMENTARY FLUX MODES IN BIOCHEMICAL REACTION SYSTEMS AT STEADY STATE. *J Biol Syst*. 1994;02(02):165-182.
13. Im JH, Song JM, Kang JH, Kang DJ. Optimization of medium components for high-molecular-weight hyaluronic acid production by *Streptococcus* sp. ID9102 via a statistical approach. *J Ind Microbiol Biotechnol*. 2009;36(11):1337-1344. doi:10.1007/s10295-009-0618-8
14. Behera HT, Upadhyay AK, Raina V, Ray L. Optimization of media components for the production of N-acetylchitooligosaccharide from chitin by *Streptomyces chilikensis* through Taguchi experimental design. *J Microbiol Methods*. 2019;159(February):194-199. doi:10.1016/j.mimet.2019.03.014
15. Darvishi F, Moghaddami NA. Optimization of an industrial medium from molasses for bioethanol production using the Taguchi statistical experimental-design method. *Fermentation*. 2019;5(1). doi:10.3390/fermentation5010014
16. Khanna S, Ranjan A, Goyal A, Moholkar VS. Medium optimization for mixed alcohols

- production by glycerol utilizing immobilized *Clostridium pasteurianum* MTCC 116. *Chem Biochem Eng Q.* 2013;27(3):319-325.
17. Mehta PK, Bhatia SK, Bhatia RK, Bhalla TC. Enhanced production of thermostable amidase from *Geobacillus subterraneus* RL-2a MTCC 11502 via optimization of physicochemical parameters using Taguchi DOE methodology. *3 Biotech.* 2016;6(1):1-12. doi:10.1007/s13205-016-0390-1
 18. Rajendran A, Thirugnanam M, Thangavelu V. Statistical evaluation of medium components by Plackett-Burman experimental design and kinetic modeling of lipase production by *Pseudomonas fluorescens*. *Indian J Biotechnol.* 2007;6(4):469-478.
 19. Ashengroph M, Nahvi I, Amini J. Application of taguchi design and response surface methodology for improving conversion of isoeugenol into vanillin by resting cells of *Psychrobacter* sp. CSW4. *Iran J Pharm Res.* 2013;12(3):411-421. doi:10.22037/ijpr.2013.1337
 20. Teng Y, Xu Y. Culture condition improvement for whole-cell lipase production in submerged fermentation by *Rhizopus chinensis* using statistical method. *Bioresour Technol.* 2008;99(9):3900-3907. doi:10.1016/j.biortech.2007.07.057
 21. Hanchinal VM, Survase SA, Sawant SK, Annature US. Response surface methodology in media optimization for production of β -carotene from *Daucus carota*. *Plant Cell Tissue Organ Cult.* 2008;93(2):123-132. doi:10.1007/s11240-008-9350-8
 22. Isiaka Adetunji A, Olufolahan Olaniran A. Optimization of culture conditions for enhanced lipase production by an indigenous *Bacillus aryabhattai* SE3-PB using response surface methodology. *Biotechnol Biotechnol Equip.* 2018;32(6):1514-1526. doi:10.1080/13102818.2018.1514985
 23. Chen L, Alcazar J, Yang T, Lu Z, Lu Y. Optimized cultural conditions of functional yogurt for γ -aminobutyric acid augmentation using response surface methodology. *J Dairy Sci.* 2018;101(12):10685-10693. doi:10.3168/jds.2018-15391
 24. Ekpenyong MG, Antai SP, Asitok AD, Ekpo BO. Plackett-burman design and response surface optimization of medium trace nutrients for glycolipopeptide biosurfactant production. *Iran Biomed J.* 2017;21(4):249-260. doi:10.18869/acadpub.ijb.21.4.249
 25. Cui JD, Jia SR. Optimization of medium on exopolysaccharides production in submerged culture of *Cordyceps militaris*. *Food Sci Biotechnol.* 2010;19(6):1567-1571. doi:10.1007/s10068-010-0222-8
 26. Marteiijn RCL, Jurrius O, Dhont J, De Gooijer CD, Tramper J, Martens DE. Optimization of a feed medium for fed-batch culture of insect cells using a genetic algorithm. *Biotechnol Bioeng.* 2003;81(3):269-278. doi:10.1002/bit.10465
 27. Etschmann MMW, Sell D, Schrader J. Medium optimization for the production of the aroma compound 2-phenylethanol using a genetic algorithm. *J Mol Catal B Enzym.* 2004;29(1-6):187-193. doi:10.1016/j.molcatb.2003.10.014
 28. Camacho-Rodríguez J, Cerón-García MC, Fernández-Sevilla JM, Molina-Grima E. Genetic algorithm for the medium optimization of the microalga *Nannochloropsis gaditana* cultured to aquaculture. *Bioresour Technol.* 2015;177:102-109. doi:10.1016/j.biortech.2014.11.057
 29. García-Camacho F, Gallardo-Rodríguez JJ, Sánchez-Mirón A, Chisti Y, Molina-Grima E. Genetic algorithm-based medium optimization for a toxic dinoflagellate microalga. *Harmful Algae.* 2011;10(6):697-701. doi:10.1016/j.hal.2011.05.005
 30. Siddique S, Nelofer R, Syed Q, Adnan A, Qureshi FA. Optimization for the enhanced

- production of avermectin B1b from *Streptomyces avermitilis* DSM 41445 using artificial neural network. *J Korean Soc Appl Biol Chem*. 2014;57(5):677-683. doi:10.1007/s13765-014-4194-x
31. Pathak L, Singh V, Niwas R, et al. Artificial intelligence versus statistical modeling and optimization of cholesterol oxidase production by using *Streptomyces* sp. *PLoS One*. 2015;10(9):1-14. doi:10.1371/journal.pone.0137268
 32. Haider MA, Pakshirajan K, Singh A, Chaudhry S. Artificial neural network-genetic algorithm approach to optimize media constituents for enhancing lipase production by a soil microorganism. *Appl Biochem Biotechnol*. 2008;144(3):225-235. doi:10.1007/s12010-007-8017-y
 33. H.A. Shakir, R. Mahmood, M. Irfan, F. Deebe, I. Javed JIQ. PROTEASE PRODUCTION FROM *Bacillus safensis* IN SUBMERGED FERMENTATION USING RESPONSE SURFACE METHODOLOGY. *Rev Mex Ing Química*. 2019;18(1):375-385.
 34. Makowski K, Matusiak K, Borowski S, et al. Optimization of a culture medium using the Taguchi approach for the production of microorganisms active in odorous compound removal. *Appl Sci*. 2017;7(8). doi:10.3390/app7080756
 35. Response Surface Methods for Optimization - ReliaWiki. http://reliawiki.org/index.php/Response_Surface_Methods_for_Optimization. Accessed January 21, 2020.
 36. Morowvat MH, Ghasemi Y. Culture medium optimization for enhanced β -carotene and biomass production by *Dunaliella salina* in mixotrophic culture. *Biocatal Agric Biotechnol*. 2016;7:217-223. doi:10.1016/j.bcab.2016.06.008
 37. Kovárová-Kovar K, Egli T. Growth kinetics of suspended microbial cells: from single-substrate-controlled growth to mixed-substrate kinetics. *Microbiol Mol Biol Rev*. 1998;62(3):646-666. <http://www.ncbi.nlm.nih.gov/pubmed/9729604> <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC98929>.
 38. Hajmeer MN, Basheer IA, Najjar YM. Computational neural networks for predictive microbiology II. Application to microbial growth. *Int J Food Microbiol*. 1997;34(1):51-66. doi:10.1016/S0168-1605(96)01169-5
 39. A general model of a deep neural network. It consists of an input... | Download Scientific Diagram. https://www.researchgate.net/figure/A-general-model-of-a-deep-neural-network-It-consists-of-an-input-layer-some-here-two_fig1_308414212. Accessed June 26, 2020.
 40. Smaoui S, Ennouri K, Chakchouk-Mtibaa A, et al. Statistical versus artificial intelligence-based modeling for the optimization of antifungal activity against *Fusarium oxysporum* using *Streptomyces* sp. strain TN71. *J Mycol Med*. 2018;28(3):551-560. doi:10.1016/j.mycmed.2018.07.003
 41. Rafigh SM, Yazdi AV, Vossoughi M, Safekordi AA, Ardjmand M. Optimization of culture medium and modeling of curdlan production from *Paenibacillus polymyxa* by RSM and ANN. *Int J Biol Macromol*. 2014;70:463-473. doi:10.1016/j.ijbiomac.2014.07.034
 42. Talib NSR, Halimi MIE, Ghani SSA, Zaidan UH, Shukor MYA. Artificial Neural Networks (ANNs) and Response Surface Methodology (RSM) Approach for Modelling the Optimization of Chromium (VI) Reduction by Newly Isolated *Acinetobacter radioresistens* Strain NS-MIE from Agricultural Soil. *Biomed Res Int*. 2019;2019.

doi:10.1155/2019/5785387

43. Understanding the Genetic Algorithm - Noteworthy - The Journal Blog. <https://blog.usejournal.com/understanding-the-genetic-algorithm-4eac04a07a59>. Accessed April 10, 2020.
44. Genetic Algorithms Definition | DeepAI. <https://deepai.org/machine-learning-glossary-and-terms/genetic-algorithms>. Accessed January 24, 2020.
45. Özer A, Gürbüz G, Çalimli A, Körbahti BK. Biosorption of copper(II) ions on *Enteromorpha prolifera*: Application of response surface methodology (RSM). *Chem Eng J*. 2009;146(3):377-387. doi:10.1016/j.cej.2008.06.041
46. Tortajada M, Llaneras F, Picó J. 2010 - Tortajada, Llaneras, Picó - BMC systems biology.pdf. 2010.
47. Provost A, Bastin G, Agathos SN, Schneider YJ. Metabolic design of macroscopic bioreaction models: Application to Chinese hamster ovary cells. *Bioprocess Biosyst Eng*. 2006;29(5-6):349-366. doi:10.1007/s00449-006-0083-y
48. Noda S, Mori Y, Oyama S, Kondo A, Araki M, Shirai T. Reconstruction of metabolic pathway for isobutanol production in *Escherichia coli*. *Microb Cell Fact*. 2019;18(1):124. doi:10.1186/s12934-019-1171-4
49. BiGG Models. <http://bigg.ucsd.edu/>. Accessed April 10, 2020.
50. Trinh CT, Wlaschin A, Sreenc F. Elementary mode analysis: A useful metabolic pathway analysis tool for characterizing cellular metabolism. *Appl Microbiol Biotechnol*. 2009;81(5):813-826. doi:10.1007/s00253-008-1770-1
51. Gregoty N. Stephanopoulos, Aristos A. Aristidou JN. *Metabolic Engineering: Princeples and Methdologies.*; 1998.
52. Wilkens CA, Altamirano C, Gerdtsen ZP. Comparative metabolic analysis of lactate for CHO cells in glucose and galactose. *Biotechnol Bioprocess Eng*. 2011;16(4):714-724. doi:10.1007/s12257-010-0409-0
53. Xing Z, Kenty B, Koyrakh I, Borys M, Pan SH, Li ZJ. Optimizing amino acid composition of CHO cell culture media for a fusion protein production. *Process Biochem*. 2011;46(7):1423-1429. doi:10.1016/j.procbio.2011.03.014
54. McAtee Pereira AG, Walther JL, Hollenbach M, Young JD. 13C Flux Analysis Reveals that Rebalancing Medium Amino Acid Composition can Reduce Ammonia Production while Preserving Central Carbon Metabolism of CHO Cell Cultures. *Biotechnol J*. 2018;13(10):1-7. doi:10.1002/biot.201700518
55. Schuetz R, Kuepfer L, Sauer U. Systematic evaluation of objective functions for predicting intracellular fluxes in *Escherichia coli*. *Mol Syst Biol*. 2007;3(119). doi:10.1038/msb4100162
56. Costa RS, Nguyen S, Hartmann A, Vinga S. Exploring the cellular objective in flux balance constraint-based models. *Lect Notes Comput Sci (including Subser Lect Notes Artif Intell Lect Notes Bioinformatics)*. 2014;8859:211-224. doi:10.1007/978-3-319-12982-2_15
57. Chen Y, Dong F, Wang Y. Systematic development and optimization of chemically defined medium supporting high cell density growth of *Bacillus coagulans*. *Appl Microbiol Biotechnol*. 2016;100(18):8121-8134. doi:10.1007/s00253-016-7644-z
58. Rockafellar R. Book - Convex analysis. *Princet Univ Press*. 1997.

59. Çakir T, Kirdar B, Önsan ZI, Ülgen KÖ, Nielsen J. Effect of carbon source perturbations on transcriptional regulation of metabolic fluxes in *Saccharomyces cerevisiae*. *BMC Syst Biol*. 2007;1:1-10. doi:10.1186/1752-0509-1-18
60. R2019a - Updates to the MATLAB and Simulink product families - MATLAB & Simulink. https://www.mathworks.com/products/new_products/release2019a.html. Accessed April 10, 2020.
61. Terzer M, Stelling J. Accelerating the computation of elementary modes using pattern trees. *Lect Notes Comput Sci (including Subser Lect Notes Artif Intell Lect Notes Bioinformatics)*. 2006;4175 LNBI(1):333-343. doi:10.1007/11851561_31
62. Klamt S, Stelling J. Combinatorial complexity of pathway analysis in metabolic networks. *Mol Biol Rep*. 2002;29(1-2):233-236. doi:10.1023/A:1020390132244
63. Stephanopoulos GN, Aristidou AA, Nielsen J. Thermodynamics of Cellular Processes. *Metab Eng*. 1998;629-694. doi:10.1016/b978-012666260-3/50015-7
64. Gibbs (Free) Energy - Chemistry LibreTexts. [https://chem.libretexts.org/Bookshelves/Physical_and_Theoretical_Chemistry_Textbook_Maps/Supplemental_Modules_\(Physical_and_Theoretical_Chemistry\)/Thermodynamics/Energies_and_Potentials/Free_Energy/Gibbs_\(Free\)_Energy](https://chem.libretexts.org/Bookshelves/Physical_and_Theoretical_Chemistry_Textbook_Maps/Supplemental_Modules_(Physical_and_Theoretical_Chemistry)/Thermodynamics/Energies_and_Potentials/Free_Energy/Gibbs_(Free)_Energy). Accessed February 13, 2020.
65. eEquilibrator: The Biochemical Thermodynamics Calculator. <http://equilibrator.weizmann.ac.il/>. Accessed February 15, 2020.
66. Noor E, Haraldsdóttir HS, Milo R, Fleming RMT. Consistent Estimation of Gibbs Energy Using Component Contributions. *PLoS Comput Biol*. 2013;9(7). doi:10.1371/journal.pcbi.1003098
67. Laux H. Industry perspective on Chinese hamster ovary cell “omics” SPECIAL FOCUS y An 'omics approach to Chinese hamster ovary-based pharmaceutical bioprocessing. *Pharm Bioprocess*. 2014;2(5):377-381. doi:10.4155/PBP.14.48
68. Carinhas N, Duarte TM, Barreiro LC, Carrondo MJT, Alves PM, Teixeira AP. Metabolic signatures of GS-CHO cell clones associated with butyrate treatment and culture phase transition. *Biotechnol Bioeng*. 2013;110(12):3244-3257. doi:10.1002/bit.24983
69. Quek LE, Dietmair S, Krömer JO, Nielsen LK. Metabolic flux analysis in mammalian cell culture. *Metab Eng*. 2010;12(2):161-171. doi:10.1016/j.ymben.2009.09.002
70. Sengupta N, Rose ST, Morgan JA. Metabolic flux analysis of CHO cell metabolism in the late non-growth phase. *Biotechnol Bioeng*. 2011;108(1):82-92. doi:10.1002/bit.22890
71. Zamorano F, Wouwer A Vande, Bastin G. A detailed metabolic flux analysis of an underdetermined network of CHO cells. *J Biotechnol*. 2010;150(4):497-508. doi:10.1016/j.jbiotec.2010.09.944
72. efmtool – Computational Systems Biology | ETH Zurich. <https://csb.ethz.ch/tools/software/efmtool.html>. Accessed April 14, 2020.
73. Schuster S, Dandekar T, Fell DA. Detection of elementary flux modes in biochemical networks: A promising tool for pathway analysis and metabolic engineering. *Trends Biotechnol*. 1999;17(2):53-60. doi:10.1016/S0167-7799(98)01290-6
74. Song HS, Ramkrishna D. Reduction of a set of elementary modes using yield analysis. *Biotechnol Bioeng*. 2009;102(2):554-568. doi:10.1002/bit.22062
75. Folch-Fortuny A, Marques R, Isidro IA, Oliveira R, Ferrer A. Principal elementary mode analysis (PEMA). *Mol Biosyst*. 2016;12(3):737-746. doi:10.1039/c5mb00828j

76. Ferreira AR, Dias JML, Teixeira AP, et al. Projection to latent pathways (PLP): A constrained projection to latent variables (PLS) method for elementary flux modes discrimination. *BMC Syst Biol.* 2011;5. doi:10.1186/1752-0509-5-181
77. Duarte TM, Carinhas N, Barreiro LC, Carrondo MJT, Alves PM, Teixeira AP. Metabolic responses of CHO cells to limitation of key amino acids. *Biotechnol Bioeng.* 2014;111(10):2095-2106. doi:10.1002/bit.25266
78. Fan Y, Jimenez Del Val I, Müller C, et al. Amino acid and glucose metabolism in fed-batch CHO cell culture affects antibody production and glycosylation. *Biotechnol Bioeng.* 2015;112(3):521-535. doi:10.1002/bit.25450
79. Li F, Vijayasankaran N, Shen A, Kiss R, Amanullah A. Cell culture processes for monoclonal antibody production. *MAbs.* 2010;2(5):466-479. doi:10.4161/mabs.2.5.12720
80. Kurano N, Leist C, Messi F, Kurano S, Fiechter A. Growth behavior of Chinese hamster ovary cells in a compact loop bioreactor. 2. Effects of medium components and waste products. *J Biotechnol.* 1990;15(1-2):113-128. doi:10.1016/0168-1656(90)90055-G
81. Glacken MW, Fleischaker RJ, Sinskey AJ. Reduction of waste product excretion via nutrient control: Possible strategies for maximizing product and cell yields on serum in cultures of mammalian cells. *Biotechnol Bioeng.* 1986;28(9):1376-1389. doi:10.1002/bit.260280912
82. Hansen HA, Emborg C. Influence of Ammonium on Growth, Metabolism, and Productivity of a Continuous Suspension Chinese Hamster Ovary Cell Culture. *Biotechnol Prog.* 1994;10(1):121-124. doi:10.1021/bp00025a014
83. Zhu MM, Goyal A, Rank DL, Gupta SK, Vanden Boom T, Lee SS. Effects of elevated pCO₂ and osmolality on growth of CHO cells and production of antibody-fusion protein B1: A case study. *Biotechnol Prog.* 2005;21(1):70-77. doi:10.1021/bp049815s
84. Avid, Damjanovic L, Kaisermayer C, Kunert R. Benchmarking of commercially available CHO cell culture media for antibody production. *Appl Microbiol Biotechnol.* 2015;99(11):4645-4657. doi:10.1007/s00253-015-6514-4
85. Singh V, Haque S, Niwas R, Srivastava A, Pasupuleti M, Tripathi CKM. Strategies for fermentation medium optimization: An in-depth review. *Front Microbiol.* 2017;7(JAN). doi:10.3389/fmicb.2016.02087

Appendix

Appendix A – CHO Metabolic Networks

Table A.1 Metabolic Network for CHO cell line

Table A.1 Metabolic Network for CHO cell line

#	Reactions
	Glycolysis
1	1 Glc + 1 ATP --> 1 G6P
2	1 G6P --> 1 F6P
3	1 F6P + 1 ATP --> 1 GAP + 1 DHAP
4	1 DHAP --> 1 GAP
5	1 GAP --> 1 NADH + 1 ATP + 1 3PG
6	1 3PG --> 1 Pyr + 1 ATP
	TCA cycle
7	1 Pyr --> 1 CO ₂ + 1 AcCoA + 1 NADH
8	1 AcCoA + 1 Oxal --> 1 Cit
9	1 Cit --> 1 CO ₂ + 1 aKG + 1 NADH
10	1 aKG --> 1 CO ₂ + 1 SucCoA + 1 NADH
11	1 SucCoA --> 1 Succ + 1 ATP
12	1 Succ --> 1 Fum + 1 FADH ₂
13	1 Fum --> 1 Mal
14	1 Mal --> 1 Oxal + 1 NADH
	Pyruvate fates
15	1 Pyr + 1 NADH --> 1 Lac
16	1 Pyr + 1 Glu --> 1 Ala + 1 aKG
	Pentose Phosphate Pathway
17	3 G6P --> 3 CO ₂ + 3 R5P + 6 NADPH
	Anaplerotic Reaction
18	1 Mal --> 1 Pyr + 1 CO ₂ + 1 NADPH
	Amino Acid Metabolism
19	1 Glu <--> 1 aKG + 1 NH ₄ + 1 NADH
20	1 aKG + 1 Asp --> 1 Glu + 1 Oxal
21	1 Glu + 1 NH ₄ + 1 ATP --> Gln
22	1 Thr --> 1 AcCoA + 1 Gly + 1 NADH
23	1 Ser --> 1 Gly + 1 NADPH + 1 ATP + 1 Formate
24	1 Ser --> 1 Pyr + 1 NH ₄
25	1 Thr --> 1 NH ₄ + 1 aKb
26	1 ATP + 1 aKb --> 1 SucCoA + 1 NADH
27	1 Trp --> 2 CO ₂ + 1 Ala + 1 aKa
28	2 aKG + 1 Lys --> 2 Glu + 3 NADPH + 1 FADH ₂ + 1 aKa
29	1 aKa --> 2 CO ₂ + 2 AcCoA + 2 NADH
30	1 aKG + 1 Val --> 1 CO ₂ + 1 Glu + 1 NADH + 1 IsobutCoA

31	1 ATP + 1 IsobutCoA --> 1 SucCoA + 2 NADH + 1 FADH2
32	1 IsobutCoA --> 1 Isobut
33	1 aKG + 1 Ile + 1 ATP --> 1 Glu + 1 AcCoA + 1 SucCoA + 2 NADH + 1 FADH2
34	1 aKG + 1 Leu --> 1 CO2 + 1 Glu + 1 NADH + 1 IsovalCoA
35	1 CO2 + 1 SucCoA + 1 ATP + 1 IsovalCoA --> 3 AcCoA + 1 Succ + 1 FADH2
36	1 IsovalCoA --> 1 Isoval
37	1 Phe + 1 NADH --> 1 Tyr
38	1 aKG + 1 SucCoA + 1 Tyr --> 1 CO2 + 1 Glu + 2 AcCoA + 1 Fum + 1 Succ
39	1 Ser + 1 Met + 1 ATP --> 1 NH4 + 1 aKb
40	1 Asn --> 1 NH4 + 1 Asp
41	1 Pro --> 1 Glu + 1 NADH
42	1 aKG + 1 Arg --> 2 Glu + 1 NADH
43	1 His --> 1 Glu + 1 NH4
	Glycogen Synthesis
44	1 G6P --> 1 G1P
45	2 ATP + 1 G1P + 1 UMPN --> 1 UDPG
46	1 UDPG --> 1 Glycogen
	Nucleotide Synthesis
47	1 R5P + 1 ATP --> 1 PRPP
48	1 CO2 + 2 Gln + 1 Gly + 1 Asp + 5 ATP + 1 PRPP --> 2 Glu + 1 Fum + 1 IMP
49	1 Asp + 2 ATP + 1 GMPN + 1 IMP --> 1 Fum + 1 AMPN
50	1 Gln + 1 ATP + 1 IMP --> 1 Glu + 1 NADH + 1 GMPN
51	1 NH4 + 1 Asp + 2 ATP + 1 CO2 --> 1 NADH + 1 Orotate
52	1 Orotate + 1 PRPP --> 1 CO2 + 1 UMPN
53	1 Gln + 1 ATP + 1 UMPN --> 1 Glu + 1 CMPN
54	1 AMPN --> 1 dAMP
55	1 GMPN --> 1 dGMP
56	1 CMPN --> 1 dCMP
57	1 UMPN --> 1 dTMP
	Lipid Synthesis
58	1 ATP + 1 Choline --> 1 Pcholine
59	18 AcCoA + 33 NADH + 22 ATP + 1 Glyc3P + 1 Pcholine --> 1 PC
60	1 Ser + 1 PC --> 1 PS + 1 Choline
61	1 PS --> 1 CO2 + 1 PE
62	1 Glyc3P + 1 Choline --> 1 Glyc3PC
63	1 G6P --> 1 Inositol
64	18 AcCoA + 33 NADH + 22 ATP + 1 Glyc3P + 1 Inositol --> 1 PI
65	18 AcCoA + 33 NADH + 22 ATP + 2 Glyc3P --> 1 PG
66	2 PG --> 1 DPG + 1 Glyc
67	16 AcCoA + 1 Ser + 29 NADPH + 16 ATP + 1 Choline --> 2 CO2 + 1 SM
68	18 AcCoA + 14 NADPH + 18 ATP --> 9 CO2 + 1 Cholesterol
	Biomass Formation
69	160.1015 Ala + 235.2056 Glu + 70.3787 Gln + 174.6799 Gly + 114.9787 Ser + 147.4132 Lys + 157.4070 Leu + 82.6648 Ile + 91.8543 Arg + 169.492 Asp + 95.7754 Thr + 118.3569 Val + 40.0354 Met + 67.4027 Phe + 47.4956 Tyr + 36.2551 His + 55.559 Pro + 70.3787 Asn + 83943.063 ATP + 8.943 AMPN + 4.878 Cholesterol + 14.9321 CMPN + 4.0108 dAMP + 2.6829 dCMP + 2.6829 dGMP + 4.0108 dTMP + 0.813 DPG + 75.609 Glycogen + 16.9104 GMPN + 18.699 PC + 7.046 PE + 0.271 PG + 2.71 PI + 0.813 PS + 2.168 SM + 8.943 UMPN + 9.2297 Trp --> 1 X
	Other by-products

70	1 AcCoA --> 1 ATP + 1 Acetate
71	1 NADH + 1 DHAP --> 1 Glyc3P
72	1 Glyc3P --> 1 Glyc
	IgG Glycosylation
73	1 UDPG --> 1 UDPGal
74	1 Glc + 3 ATP + 1 GMPN --> 1 GDPMann
75	1 AcCoA + 1 Gln + 1 F6P + 1 UMPN + 2 ATP --> 1 Glu + 1 UDPNAG
76	3 ATP + 1 3PG + 1 UDPNAG + 1 CMPN --> 1 CMPSialic
77	1 NADPH + 1 GDPMann --> 1 GDPFuc
	IgG Formation
78	428.7 Ala + 362.75 Glu + 351.76 Gln + 516.64 Gly + 934.36 Ser + 472.67 Lys + 516.64 Leu + 175.88 Ile + 307.79 Arg + 296.8 Asp + 626.57 Thr + 714.51 Val + 65.954 Met + 285.8 Phe + 285.8 Tyr + 164.89 His + 505.65 Pro + 263.82 Asn + 142.9 Trp + 10.992 GDPFuc + 54.962 UDPNAG + 32.977 GDPMann + 21.985 UDPGal + 21.985 CMPSialic --> 1 P
	Transport Reactions
79	1 ATP -->
80	1 NADH -->
81	1 FADH2 -->
82	1 NADPH -->
83	1 X -->
84	1 P -->
85	--> 1 Glc
86	--> 1 His
87	1 Isobut -->
88	--> 1 Ile
89	1 Isoval -->
90	--> 1 Leu
91	--> 1 Lys
92	--> 1 Met
93	--> 1 Phe
94	--> 1 Thr
95	--> 1 Trp
96	--> 1 Val
97	1 CO2 -->
98	1 NH4 -->
99	1 Acetate -->
100	1 Ala -->
101	--> 1 Arg
102	--> 1 Asn
103	--> 1 Asp
104	1 Gln -->
105	1 Cit -->
106	--> 1 Choline
107	1 Formate -->
108	--> 1 Glu
109	1 Glyc -->
110	1 Gly -->
111	1 Lac -->
112	--> 1 Pro
113	--> 1 Pyr
114	--> 1 Ser

115	--> 1 Tyr
116	1 Mal -->
117	1 Glyc3PC -->
118	1 Glyc3PC -->

Table A.2 Biomass producing Network

#	Reactions
	Glycolysis
1	1 Glc --> 1 G6P
2	1 G6P --> 1 F6P
3	1 F6P --> 1 GAP + 1 DHAP
4	1 DHAP --> 1 GAP
5	1 GAP --> 1 3PG
6	1 3PG --> 1 Pyr
	TCA cycle
7	1 Pyr --> 1 CO ₂ + 1 AcCoA
8	1 AcCoA + 1 Oxal --> 1 Cit
9	1 Cit --> 1 CO ₂ + 1 aKG
10	1 aKG --> 1 CO ₂ + 1 SucCoA
11	1 SucCoA --> 1 Succ
12	1 Succ --> 1 Fum
13	1 Fum --> 1 Mal
14	1 Mal --> 1 Oxal
	Pyruvate fates
15	1 Pyr --> 1 Lac
16	1 Pyr + 1 Glu --> 1 Ala + 1 aKG
	Pentose Phosphate Pathway
17	3 G6P --> 3 CO ₂ + 3 R5P
	Anaplerotic Reaction
18	1 Mal --> 1 Pyr + 1 CO ₂
	Amino Acid Metabolism
19	1 Glu <--> 1 aKG + 1 NH ₄
20	1 aKG + 1 Asp --> 1 Glu + 1 Oxal
21	1 Glu + 1 NH ₄ --> Gln
22	1 Thr --> 1 AcCoA + 1 Gly
23	1 Ser --> 1 Gly + 1 Formate
24	1 Ser --> 1 Pyr + 1 NH ₄
25	1 Thr --> 1 NH ₄ + 1 aKb
26	1 aKb --> 1 SucCoA
27	1 Trp --> 2 CO ₂ + 1 Ala + 1 aKa
28	2 aKG + 1 Lys --> 2 Glu + 1 aKa
29	1 aKa --> 2 CO ₂ + 2 AcCoA
30	1 aKG + 1 Val --> 1 CO ₂ + 1 Glu + 1 IsobutCoA
31	1 IsobutCoA --> 1 SucCoA
32	1 IsobutCoA --> 1 Isobut
33	1 aKG + 1 Ile --> 1 Glu + 1 AcCoA + 1 SucCoA
34	1 aKG + 1 Leu --> 1 CO ₂ + 1 Glu + 1 IsovalCoA
35	1 CO ₂ + 1 SucCoA + 1 IsovalCoA --> 3 AcCoA + 1 Succ
36	1 IsovalCoA --> 1 Isoval
37	1 Phe --> 1 Tyr
38	1 aKG + 1 SucCoA + 1 Tyr --> 1 CO ₂ + 1 Glu + 2 AcCoA + 1 Fum + 1 Succ
39	1 Ser + 1 Met --> 1 NH ₄ + 1 aKb
40	1 Asn --> 1 NH ₄ + 1 Asp
41	1 Pro --> 1 Glu
42	1 aKG + 1 Arg --> 2 Glu

43	1 His --> 1 Glu + 1 NH ₄
	Glycogen Synthesis
44	1 G6P --> 1 G1P
45	1 G1P + 1 UMP _{PRN} --> 1 UDPG
46	1 UDPG --> 1 Glycogen
	Nucleotide Synthesis
47	1 R5P --> 1 PRPP
48	1 CO ₂ + 2 Gln + 1 Gly + 1 Asp + 1 PRPP --> 2 Glu + 1 Fum + 1 IMP
49	1 Asp + 1 GMP _{PRN} + 1 IMP --> 1 Fum + 1 AMP _{PRN}
50	1 Gln + 1 IMP --> 1 Glu + 1 GMP _{PRN}
51	1 NH ₄ + 1 Asp + 1 CO ₂ --> 1 Orotate
52	1 Orotate + 1 PRPP --> 1 CO ₂ + 1 UMP _{PRN}
53	1 Gln + 1 UMP _{PRN} --> 1 Glu + 1 CMP _{PRN}
54	1 AMP _{PRN} --> 1 dAMP
55	1 GMP _{PRN} --> 1 dGMP
56	1 CMP _{PRN} --> 1 dCMP
57	1 UMP _{PRN} --> 1 dTMP
	Lipid Synthesis
58	1 Choline --> 1 Pcholine
59	18 AcCoA + 1 Glyc3P + 1 Pcholine --> 1 PC
60	1 Ser + 1 PC --> 1 PS + 1 Choline
61	1 PS --> 1 CO ₂ + 1 PE
62	1 Glyc3P + 1 Choline --> 1 Glyc3PC
63	1 G6P --> 1 Inositol
64	18 AcCoA + 1 Glyc3P + 1 Inositol --> 1 PI
65	18 AcCoA + 2 Glyc3P --> 1 PG
66	2 PG --> 1 DPG + 1 Glyc
67	16 AcCoA + 1 Ser + 1 Choline --> 2 CO ₂ + 1 SM
68	18 AcCoA --> 9 CO ₂ + 1 Cholesterol
	Biomass Formation
69	160.1015 Ala + 235.2056 Glu + 70.3787 Gln + 174.6799 Gly + 114.9787 Ser + 147.4132 Lys + 157.4070 Leu + 82.6648 Ile + 91.8543 Arg + 169.492 Asp + 95.7754 Thr + 118.3569 Val + 40.0354 Met + 67.4027 Phe + 47.4956 Tyr + 36.2551 His + 55.559 Pro + 70.3787 Asn + 8.943 AMP _{PRN} + 4.878 Cholesterol + 14.9321 CMP _{PRN} + 4.0108 dAMP + 2.6829 dCMP + 2.6829 dGMP + 4.0108 dTMP + 0.813 DPG + 75.609 Glycogen + 16.9104 GMP _{PRN} + 18.699 PC + 7.046 PE + 0.271 PG + 2.71 PI + 0.813 PS + 2.168 SM + 8.943 UMP _{PRN} + 9.2297 Trp --> 1 X
	Other by-products
70	1 AcCoA --> 1 Acetate
71	1 DHAP --> 1 Glyc3P
72	1 Glyc3P --> 1 Glyc
	Transport Reactions
73	1 X -->
74	--> 1 Glc
75	--> 1 His
76	1 Isobut -->
77	--> 1 Ile
78	1 Isoval -->
79	--> 1 Leu
80	--> 1 Lys
81	--> 1 Met

82	--> 1 Phe
83	--> 1 Thr
84	--> 1 Trp
85	--> 1 Val
86	1 CO2 -->
87	1 NH4 -->
88	1 Acetate -->
89	1 Ala -->
90	--> 1 Arg
91	--> 1 Asn
92	--> 1 Asp
93	1 Gln -->
94	1 Cit -->
95	--> 1 Choline
96	1 Formate -->
97	--> 1 Glu
98	1 Glyc -->
99	1 Gly -->
100	1 Lac -->
101	--> 1 Pro
102	--> 1 Pyr
103	--> 1 Ser
104	--> 1 Tyr
105	1 Mal -->
106	1 Glyc3PC -->

Table A.3 Product Producing Network

#	Reactions
	Glycolysis
1	1 Glc --> 1 G6P
2	1 G6P --> 1 F6P
3	1 F6P --> 1 GAP + 1 DHAP
4	1 DHAP --> 1 GAP
5	1 GAP --> 1 3PG
6	1 3PG --> 1 Pyr
	TCA cycle
7	1 Pyr --> 1 CO ₂ + 1 AcCoA
8	1 AcCoA + 1 Oxal --> 1 Cit
9	1 Cit --> 1 CO ₂ + 1 aKG
10	1 aKG --> 1 CO ₂ + 1 SucCoA
11	1 SucCoA --> 1 Succ
12	1 Succ --> 1 Fum
13	1 Fum --> 1 Mal
14	1 Mal --> 1 Oxal
	Pyruvate fates
15	1 Pyr --> 1 Lac
16	1 Pyr + 1 Glu --> 1 Ala + 1 aKG
	Pentose Phosphate Pathway
17	3 G6P --> 3 CO ₂ + 3 R5P
	Anaplerotic Reaction
18	1 Mal --> 1 Pyr + 1 CO ₂
	Amino Acid Metabolism
19	1 Glu <--> 1 aKG + 1 NH ₄
20	1 aKG + 1 Asp --> 1 Glu + 1 Oxal
21	1 Glu + 1 NH ₄ --> Gln
22	1 Thr --> 1 AcCoA + 1 Gly
23	1 Ser --> 1 Gly + 1 Formate
24	1 Ser --> 1 Pyr + 1 NH ₄
25	1 Thr --> 1 NH ₄ + 1 aKb
26	1 aKb --> 1 SucCoA
27	1 Trp --> 2 CO ₂ + 1 Ala + 1 aKa
28	2 aKG + 1 Lys --> 2 Glu + 1 aKa
29	1 aKa --> 2 CO ₂ + 2 AcCoA
30	1 aKG + 1 Val --> 1 CO ₂ + 1 Glu + 1 IsobutCoA
31	1 IsobutCoA --> 1 SucCoA
32	1 IsobutCoA --> 1 Isobut
33	1 aKG + 1 Ile --> 1 Glu + 1 AcCoA + 1 SucCoA
34	1 aKG + 1 Leu --> 1 CO ₂ + 1 Glu + 1 IsovalCoA
35	1 CO ₂ + 1 SucCoA + 1 IsovalCoA --> 3 AcCoA + 1 Succ
36	1 IsovalCoA --> 1 Isoval
37	1 Phe --> 1 Tyr
38	1 aKG + 1 SucCoA + 1 Tyr --> 1 CO ₂ + 1 Glu + 2 AcCoA + 1 Fum + 1 Succ
39	1 Ser + 1 Met --> 1 NH ₄ + 1 aKb
40	1 Asn --> 1 NH ₄ + 1 Asp

41	1 Pro --> 1 Glu
42	1 aKG + 1 Arg --> 2 Glu
43	1 His --> 1 Glu + 1 NH4
	Glycogen Synthesis
44	1 G6P --> 1 G1P
45	1 G1P + 1 UMPN --> 1 UDPG
4	Nucleotide Synthesis
46	1 R5P --> 1 PRPP
47	1 CO2 + 2 Gln + 1 Gly + 1 Asp + 1 PRPP --> 2 Glu + 1 Fum + 1 IMP
48	1 Gln + 1 IMP --> 1 Glu + 1 GMPN
49	1 NH4 + 1 Asp + 1 CO2 --> 1 Orotate
50	1 Orotate + 1 PRPP --> 1 CO2 + 1 UMPN
51	1 Gln + 1 UMPN --> 1 Glu + 1 CMPN
	Other by-products
52	1 AcCoA --> 1 Acetate
53	1 DHAP --> 1 Glyc3P
54	1 Glyc3P --> 1 Glyc
	IgG Glycosylation
55	1 UDPG --> 1 UDPGal
56	1 Glc + 1 GMPN --> 1 GDPMann
57	1 AcCoA + 1 Gln + 1 F6P + 1 UMPN --> 1 Glu + 1 UDPNAG
58	1 3PG + 1 UDPNAG + 1 CMPN --> 1 CMPSialic
59	1 GDPMann --> 1 GDPFuc
	IgG Formation
60	428.7 Ala + 362.75 Glu + 351.76 Gln + 516.64 Gly + 934.36 Ser + 472.67 Lys + 516.64 Leu + 175.88 Ile + 307.79 Arg + 296.8 Asp + 626.57 Thr + 714.51 Val + 65.954 Met + 285.8 Phe + 285.8 Tyr + 164.89 His + 505.65 Pro + 263.82 Asn + 142.9 Trp + 10.992 GDPFuc + 54.962 UDPNAG + 32.977 GDPMann + 21.985 UDPGal + 21.985 CMPSialic --> 1 P
	Transport Reactions
61	1 P -->
62	--> 1 Glc
63	--> 1 His
64	1 Isobut -->
65	--> 1 Ile
66	1 Isoval -->
67	--> 1 Leu
68	--> 1 Lys
69	--> 1 Met
70	--> 1 Phe
71	--> 1 Thr
72	--> 1 Trp
73	--> 1 Val
74	1 CO2 -->
75	1 NH4 -->
76	1 Acetate -->
77	1 Ala -->
78	--> 1 Arg
79	--> 1 Asn
80	--> 1 Asp
81	1 Gln -->

82	1 Cit -->
83	1 Formate -->
84	--> 1 Glu
85	1 Glyc -->
86	1 Gly -->
87	1 Lac -->
88	--> 1 Pro
89	--> 1 Pyr
90	--> 1 Ser
91	--> 1 Tyr
92	1 Mal -->

Appendix B - Elementary mode that uses the minimum number of metabolic reactions required to synthesize 1 unit of biomass.

REACTION NUMBER	EM COEFF	REACTIONS
R1	269,67	'1 Glc --> 1 G6P'
R2	33,92	'1 G6P --> 1 F6P'
R3	33,92	'1 F6P --> 1 GAP + 1 DHAP '
R4	0	'1 DHAP --> 1 GAP'
R5	33,92	'1 GAP --> 1 3PG'
R6	33,92	'1 3PG --> 1 Pyr '
R7	202,25	'1 Pyr --> 1 CO2 + 1 AcCoA '
R8	0	'1 AcCoA + 1 Oxal --> 1 Cit'
R9	0	'1 Cit --> 1 CO2 + 1 aKG '
R10	136,33	'1 aKG --> 1 CO2 + 1 SucCoA '
R11	0	'1 SucCoA --> 1 Succ'
R12	136,33	'1 Succ --> 1 Fum'
R13	332,50	'1 Fum --> 1 Mal'
R14	0	'1 Mal --> 1 Oxal'
R15	0	'1 Pyr --> 1 Lac'
R16	164,08	'1 Pyr + 1 Glu --> 1 Ala + 1 aKG'
R17	51,83	'3 G6P --> 3 CO2 + 3 R5P '
R18	332,50	'1 Mal --> 1 Pyr + 1 CO2'
R19	325,67	'1 Glu <--> 1 aKG + 1 NH4'
R20	0	'1 aKG + 1 Asp --> 1 Glu + 1 Oxal'
R21	216,83	'1 Glu + 1 NH4--> Gln'
R22	225,67	'1 Thr --> 1 AcCoA + 1 Gly '
R23	0	'1 Ser --> 1 Gly + 1 Formate'
R24	0	'1 Ser --> 1 Pyr + 1 NH4'
R25	0	'1 Thr --> 1 NH4 + 1 aKb'
R26	0	'1 aKb --> 1 SucCoA '
R27	0	'1 Trp --> 2 CO2 + 1 Ala + 1 aKa'
R28	0	'2 aKG + 1 Lys --> 2 Glu + 1 aKa'
R29	0	'1 aKa --> 2 CO2 + 2 AcCoA '
R30	0	'1 aKG + 1 Val --> 1 CO2 + 1 Glu + 1 IsobutCoA'
R31	0	'1 IsobutCoA --> 1 SucCoA '
R32	0	'1 IsobutCoA --> 1 Isobut'
R33	0	'1 aKG + 1 Ile --> 1 Glu + 1 AcCoA + 1 SucCoA '
R34	0	'1 aKG + 1 Leu --> 1 CO2 + 1 Glu + 1 IsovalCoA'
R35	0	'1 CO2 + 1 SucCoA + 1 IsovalCoA --> 3 AcCoA + 1 Succ '
R36	0	'1 IsovalCoA --> 1 Isoval'
R37	185,00	'1 Phe --> 1 Tyr'
R38	136,33	'1 aKG + 1 SucCoA + 1 Tyr --> 1 CO2 + 1 Glu + 2 AcCoA + 1 Fum + 1 Succ'

R39	0	'1 Ser + 1 Met --> 1 NH4 + 1 aKb'
R40	0	'1 Asn --> 1 NH4 + 1 Asp'
R41	0	'1 Pro --> 1 Glu '
R42	217,17	'1 aKG + 1 Arg --> 2 Glu '
R43	0	'1 His --> 1 Glu + 1 NH4'
R44	77,50	'1 G6P --> 1 G1P'
R45	77,50	'1 G1P + 1 UMPN --> 1 UDPG'
R46	77,50	'1 UDPG --> 1 Glycogen'
R47	155,50	'1 R5P --> 1 PRPP'
R48	46,67	'1 CO2 + 2 Gln + 1 Gly + 1 Asp + 1 PRPP --> 2 Glu + 1 Fum + 1 IMP'
R49	13,25	'1 Asp + 1 GMPN + 1 IMP --> 1 Fum + 1 AMPN'
R50	33,33	'1 Gln + 1 IMP --> 1 Glu + 1 GMPN'
R51	108,83	'1 NH4 + 1 Asp + 1 CO2 --> 1 Orotate'
R52	108,83	'1 Orotate + 1 PRPP --> 1 CO2 + 1 UMPN'
R53	18,08	'1 Gln + 1 UMPN --> 1 Glu + 1 CMPN'
R54	4,08	'1 AMPN --> 1 dAMP'
R55	2,75	'1 GMPN --> 1 dGMP'
R56	2,75	'1 CMPN --> 1 dCMP'
R57	4,08	'1 UMPN --> 1 dTMP'
R58	27,25	'1 Choline --> 1 Pcholine'
R59	27,25	'18 AcCoA + 1 Glyc3P + 1 Pcholine --> 1 PC'
R60	8,08	'1 Ser + 1 PC --> 1 PS + 1 Choline'
R61	7,25	'1 PS --> 1 CO2 + 1 PE'
R62	0	'1 Glyc3P + 1 Choline --> 1 Glyc3PC'
R63	2,75	'1 G6P --> 1 Inositol'
R64	2,75	'18 AcCoA + 1 Glyc3P + 1 Inositol --> 1 PI'
R65	1,92	'18 AcCoA + 2 Glyc3P --> 1 PG'
R66	0,83	'2 PG --> 1 DPG + 1 Glyc'
R67	2,25	'16 AcCoA + 1 Ser + 1 Choline --> 2 CO2 + 1 SM'
R68	5,00	'18 AcCoA --> 9 CO2 + 1 Cholesterol'
R69	1,00	Biomass reaction synthesis (Reaction 2)
R70	0	'1 AcCoA --> 1 Acetate'
R71	33,92	'1 DHAP --> 1 Glyc3P'
R72	0	'1 Glyc3P --> 1 Glyc'
R73	1,00	'1 X -->'
R74	269,67	'--> 1 Glc'
R75	37,17	'--> 1 His'
R76	0	'1 Isobut -->'
R77	84,75	'--> 1 Ile'
R78	0	'1 Isoval -->'
R79	161,33	'--> 1 Leu'
R80	151,08	'--> 1 Lys'
R81	41,00	'--> 1 Met'
R82	254,08	'--> 1 Phe'
R83	323,83	'--> 1 Thr'

R84	9,50	'--> 1 Trp'
R85	121,33	'--> 1 Val'
R86	972,83	'1 CO2 -->'
R87	0	'1 NH4 -->'
R88	0	'1 Acetate -->'
R89	0	'1 Ala -->'
R90	311,33	'--> 1 Arg'
R91	72,17	'--> 1 Asn'
R92	342,50	'--> 1 Asp'
R93	0	'1 Gln -->'
R94	0	'1 Cit -->'
R95	21,42	'--> 1 Choline'
R96	0	'1 Formate -->'
R97	232,33	'--> 1 Glu'
R98	0,83	'1 Glyc -->'
R99	0	'1 Gly -->'
R100	0	'1 Lac -->'
R101	56,92	'--> 1 Pro'
R102	0	'--> 1 Pyr'
R103	128,17	'--> 1 Ser'
R104	0	'--> 1 Tyr'
R105	0	'1 Mal -->'
R106	0	'1 Glyc3PC -->'

Appendix C - Standard Gibbs free energy of formation of metabolites values

<i>METABOLITE</i>	<i>GIBBS</i>
ALANINE(ALA)	-366,7
GLUTAMATE(GLU)	-716,4
GLYCINE(GLY)	-379,1
SERINE(SER)	-522
LYSINE(LYS)	-303,8
LEUCINE(LEU)	-348,2
ISOLEUCINE(ILE)	-343,9
ARGININE(ARG)	-229,2
ASPARTATE(ASP)	-726,4
THREONINE(THR)	-529,3
VALINE(VAL)	-358,7
METHIONINE(MET)	-318,8
PHENYLALANINE(PHE)	-207,1
TYROSINE(TYR)	-370,7
HISTIDINE(HIS)	-179,8
PROLINE(PRO)	-285,6
ASPARAGINE(ASN)	-526
TRYPTOPHAN(TRP)	-112
GLUCOSE(GLC)	-916,3
ISOBUT	-368,6
ISOVAL	-362,5
PYRUVATE(PYR)	-483,6
CO ₂	-386
CITRATE(CIT)	-1238
NH ₄	-75,7
CHOLINE	-31,8
GLYCEROL(GLYC)	-493,6
ACETATE	-369,3
LACTATE(LAC)	-532,9
MALATE(MAL)	-887,9
FORMATE	-372,1
GLY3PC	-1334
GLUTAMINE(GLN)	-525,8
ATP	-26,2
NADH	-25,4
FADH₂	52,1
NADPH	-28
X	-3359687,065

P	-3479886,061
----------	--------------

Appendix D - Literature data cell growth rate and metabolic rates ⁷⁷

(The cell growth shown in (h⁻¹) and metabolic rates in (nmol/10⁶ cells/h) respectively)

<i>Growth rate</i>	0.021
<i>Ammonia</i>	5.52
<i>Acetate</i>	-0.08
<i>Alanine</i>	2.72
<i>Arginine</i>	-2.40
<i>Asparagine</i>	-11.6
<i>Aspartate</i>	-2.60
<i>Citrate</i>	0.51
<i>Choline</i>	-0.58
<i>Formate</i>	2.77
<i>Fumarate</i>	0.02
<i>Glucose</i>	-60.2
<i>Glutamate</i>	-4.57
<i>Glutamine</i>	0.18
<i>Glycerol</i>	2.91
<i>Glycine</i>	1.77
<i>Histidine</i>	-1.17
<i>Isobutyrate</i>	0.27
<i>Isoleucine</i>	-2.02
<i>Isovalerate</i>	0.85
<i>Lactate</i>	43.3
<i>Leucine</i>	-3.82
<i>Lysine</i>	-2.22
<i>Methionine</i>	-0.96
<i>Phenylalanine</i>	-1.33
<i>Proline</i>	-2.71
<i>Pyruvate</i>	-3.07
<i>Serine</i>	-7.20

Appendix E - Arithmetic footprint and optimal EM for biomass and IgG producing networks.

Table E.1 Arithmetic footprint and computed formula for the biomass producing network.

<i>Metabolites</i>	<i>Arithmetic Footprint (nmol/10⁶ cells)</i>	<i>Optimal EM (nmol/10⁶ cells)</i>
Glc	-356,20	-263,06
Pyr	-45,18	-127,04
CO2	983,73	161,46
Cit	51,46	0
Mal	82,81	218,56
Lac	14,90	0
Glu	-56,52	-482,09
Ala	76,71	0
NH4	61,41	0
Asp	-224,78	0
Gln	21,17	176,51
Thr	-336,50	-476,06
Gly	54,55	0
Ser	-312,07	-125,01
Formate	89,06	0
Trp	-67,21	-9,23
Lys	-168,29	-147,41
Val	-164,33	-118,36
Isobut	16,06	0
Ile	-128,27	-82,66
Leu	-255,46	-317,51
Isoval	16,07	0
Phe	-128,13	-114,90
Tyr	-60,61	0
Met	-92,96	-40,04
Asn	-215,80	-404,50
Pro	-117,97	-55,56
Arg	-120,90	-91,85
His	-80,80	-36,26
Choline	-46,17	-20,87
Glyc3PC	25,30	0
Glyc	46,05	0,81
Biomass(X)	1,00	1,00
Acetate	8,98	17,02

Table E.2 Arithmetic footprint and optimal EM for the IgG producing network metabolite values in

Table E.2 Arithmetic footprint and optimal EM for the IgG producing network metabolite values in

<i>Metabolites</i>	<i>Arithmetic Footprint (nmol/1mg of IgG)</i>	<i>Optimal EM (nmol/1mg of IgG)</i>
Glc	-390,55	-681,53
Pyr	-88,36	0
CO2	1330,80	120,92
Cit	148,34	483,66
Mal	89,48	43,97
Lac	26,75	0
Glu	-146,05	-714,51
Ala	25,11	296,79
NH4	70,61	0
Asp	-331,43	0
Gln	29,48	0
Thr	-956,37	-1187,18
Gly	4,17	0
Ser	-1516,11	-934,36
Formate	339,65	0
Trp	-194,71	-142,90
Lys	-493,11	-472,67
Val	-850,61	-714,51
Isobut	63,20	0
Ile	-234,69	-175,88
Leu	-647,85	-516,64
Isoval	63,20	0
Phe	-477,51	-428,70
Tyr	-191,71	-142,90
Met	-170,62	-65,95
Asn	-576,30	-1209,17
Pro	-651,70	-505,65
Arg	-377,61	-307,79
His	-270,76	-164,89
Acetate	261,76	0
Glyc	55,17	0
P	1,00	1,00

Appendix F – Results from the IgG producing network *footprintome* automatic reduction

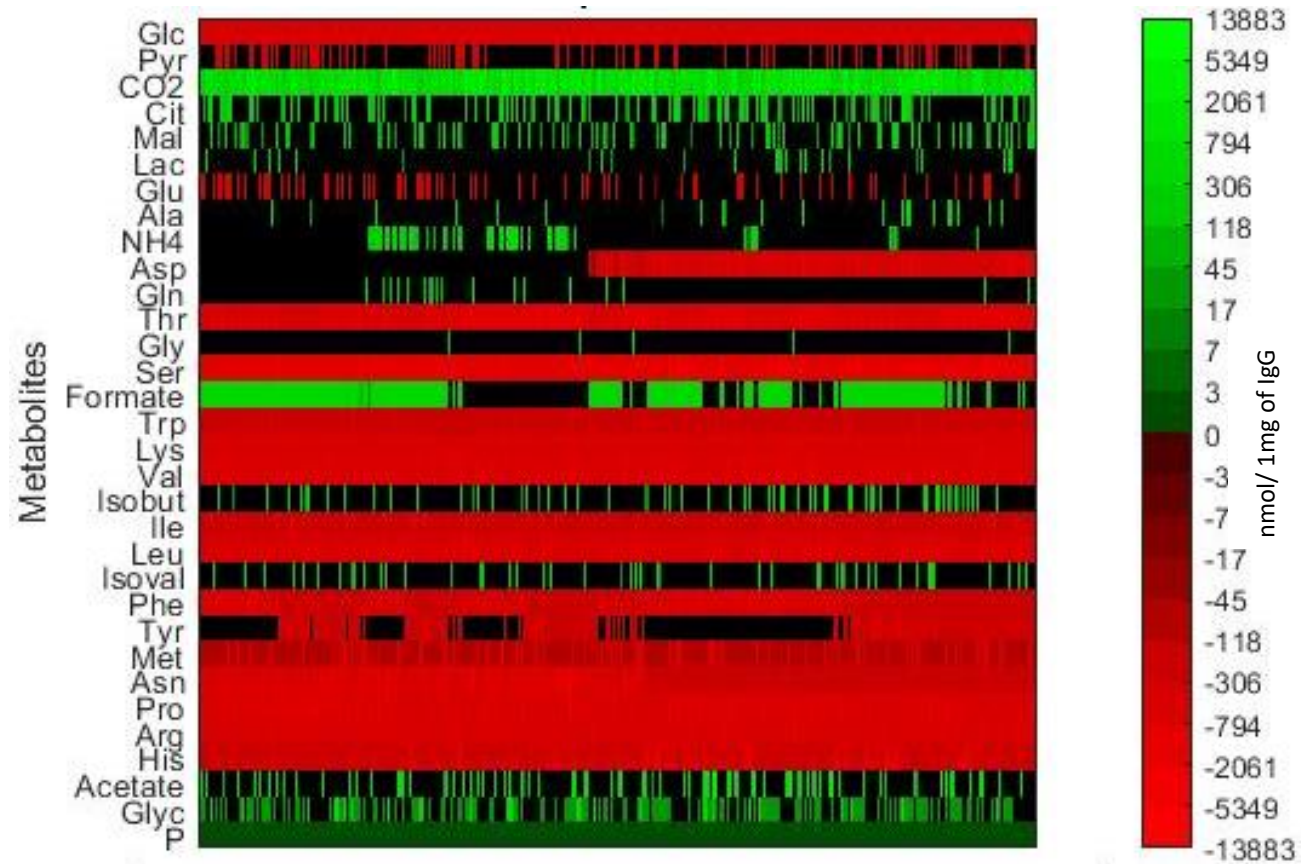


Figure F.1 Reduced *footprintome* for IgG(P) production (normalized to unit) after 2 steps of reduction: Step 1- IgG(P) production, Step 2-thermodynamic reduction. The columns represent the reduced set of EMs (307402). The colour green means that the compound is being produced by the EM. The colour red means it's being consumed by the EM. The black means it is neither consumed nor produced by the EM.

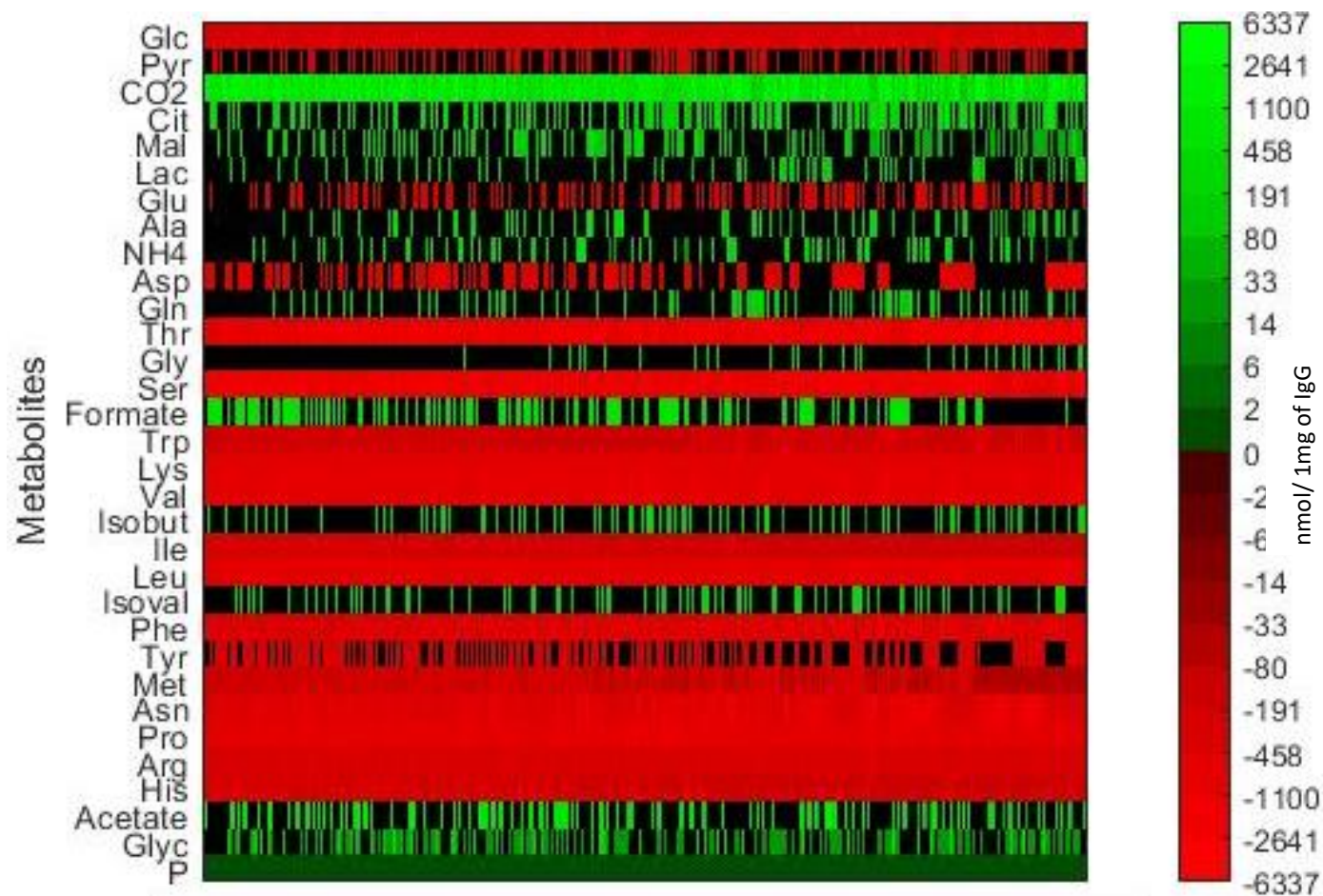


Figure F.2 Reduced footprint for IgG (P) production (normalized to unit) after 3 steps of reduction: Step 1- IgG(P) production, Step 2-Thermodynamic reduction, Step 3- Pattern clustering with arithmetic averaging. The rows represent extracellular media components. The columns represent average EM clusters (4704). The green gradient represents compound production associated with the production of 1 biomass unit. The red gradient represents compound consumption to generate 1 IgG unit. The black colour means that the compound is neither consumed nor produced by the EM cluster.