Valter Renato Silva Frade

Bachelor of Science in Biochemistry

Development of Principle Culture Medium Formulations for *Saccharomyces cerevisiae*

Dissertation to obtain the Master degree in Biotechnology

Supervisor: Rui Manuel Freitas Oliveira, Associate Professor, FCT-UNL

Jury:

Examiner: Afonso Miguel dos Santos Duarte, PhD;

President: Prof. Pedro Miguel Calado Simões;

Supervisor : Prof. Rui Manuel Freitas Oliveira

FACULDADE DE CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE NOVA DE LISBOA

September of 2014

Valter Renato Silva Frade

Bachelor of Science in Biochemistry

Development of Principle Culture Medium Formulations for *Saccharomyces cerevisiae*

Dissertation to obtain the Master degree in Biotechnology

Supervisor: Rui Manuel Freitas Oliveira, Associate Professor,
FCT-UNL

Jury:

Examiner: Afonso Miguel dos Santos Duarte, PhD;

President: Prof. Pedro Miguel Calado Simões;

Supervisor : Prof. Rui Manuel Freitas Oliveira

September of 2014

Development of Principle Culture Medium Formulations for *Saccharomyces cerevisiae*.

Aknowledgements

Gostaria de agradecer ao meu orientador Prof. Rui Oliveira por me permitir desenvolver a minha tese de mestrado.

Os meus agradecimentos estendem-se também aos meus colegas do SBE Group Dr. Moritz von Stoch, M.Sc Cristiana Azevedo, M.Sc Mauro Luís, M.Sc Rodolfo Marques, M.Sc Filipe Ataíde, M.Sc. Rui Portela, Dr. Nuno Carinhas, e Dr. Ana Ferreira por todo o apoio, companheirismo e hospitalidade com que me receberam.

Em último lugar, mas não menos importante, quero expressar toda a minha gratidão à minha família e amigos.

PS: Malta da Costa da Caparica, muito obrigado pelo tecto que me deram nos últimos dias

Abstract

Saccharomyces cerevisiae as well as other microorganisms are frequently used in industry with the purpose of obtain different kind of products that can be applied in several areas (research investigation, pharmaceutical compounds, etc.). In order to obtain high yields for the desired product, it is necessary to make an adequate medium supplementation during the growth of the microorganisms. The higher yields are typically reached by using complex media, however the exact formulation of these media is not known. Moreover, it is difficult to control the exact composition of complex media, leading to batch-to-batch variations. So, to overcome this problem, some industries choose to use defined media, with a defined and known chemical composition. However these kind of media, many times, do not reach the same high yields that are obtained by using complex media. In order to obtain similar yield with defined media the addition of many different compounds has to be tested experimentally. Therefore, the industries use a set of empirical methods with which it is tried to formulate defined media that can reach the same high yields as complex media.

In this thesis, a defined medium for *Saccharomyces cerevisiae* was developed using a rational design approach. In this approach a given metabolic network of *Saccharomyces cerevisiae* is divided into a several unique and not further decomposable sub networks of metabolic reactions that work coherently in steady state, so called elementary flux modes.

The EFMtool algorithm was used in order to calculate the EFM's for two *Saccharomyces cerevisiae* metabolic networks (amino acids supplemented metabolic network; amino acids non-supplemented metabolic network). For the supplemented metabolic network 1352172 EFM's were calculated and then divided into: 1306854 EFM's producing biomass, and 18582 EFM's exclusively producing CO2 (cellular respiration). For the non-supplemented network 635 EFM's were calculated and then divided into: 215 EFM's producing biomass; 420 EFM's producing exclusively CO2. The EFM's of each group were normalized by the respective glucose consumption value. After that, the EFMs' of the supplemented network were grouped again into: 30 clusters for the 1306854 EFMs producing biomass and, 20 clusters for the 18582 EFM's producing CO2. For the non-supplemented metabolic network the respective EFM's of each metabolic function were grouped into 10 clusters. After the clustering step, the concentrations of the other medium compounds were calculated by considering a reasonable glucose amount and by accounting for the proportionality between the compounds concentrations and the glucose ratios.

The approach adopted/developed in this thesis may allow a faster and more economical way for media development.

Keywords: EFM (Elementary Modes Fluxes); *Saccharomyces cerevisiae*; Chemical Defined Medium; Functional enviromics

Resumo

Hoje em dia a *Saccharomyces cerevisiae*, bem como outros microorganismos, são largamente utilizados em várias actividades indústriais com o intuito de a obter diferentes tipos de produtos biológicos que podem ser aplicados em várias áreas de investigação (pesquisa científica, produtos farmacêuticos, cosméticos, etc). Assim, por forma a obter altos rendimentos de produção é necessário fazer uma suplementação adequada do meio de cultura no qual os microorganismos estão inseridos. Rendimentos elevados podem ser facilmente alcançados através da utilização de um meio complexo, no entanto não se pode dizer exatamente qual a composição exacta deste meio. Além disso, a composição do meio complexo nunca é a mesma, levando a ocorrer certos problemas, tais como variações batch-to batch. Assim, a fim de superar este problema, as indústrias perferem usar um meio quimicamente definido, contudo este tipo de meio dificilmente atinge os rendimentos elevados que são obtidos usando meio complexo. Por isso, as indústrias utilizam um conjunto de métodos empíricos, pelos quais tentam formular um meio definido com a capacidade de alcançar os mesmos altos rendimentos do meio complexo. Nesta tese, teve –se por objectivo fazer a formulação de um meio quimicamente definido sem que fosse necessário o uso de qualquer método empírico. Este objectivo foi conseguido, tendo em conta a forma como os diferentes estados metabólicos (matematicamente descritos como MFE's) podem afectar a composição do meio extracelular.

Portanto, por forma a calcular todo o conjunto de MFE's de foi utilizado um algoritmo designado EFMtool através do qual foram calculados os MEF's para duas redes metabólicas de *Saccharomyces cerevisiae* (uma rede metabólica com suplementação de aminoácidos, e uma rede metabólica sem suplementação de aminoácidos). Para a rede metabólica suplementada com aminoácidos foram calculados 1352172 MFE's tendo sido divididos em: 1306854 de MFE's nos quais ocorre produção de biomassa, e 18582 da MFE's nos quais ocorre exclusivamente produção de CO2 (respiração celular). Para a rede metabólica sem suplementos, foram calculados 635 MFE's tendo sido divididos em: 215 do MFE's nos quais ocorre produção de biomassa; 420 MFE's nos quais ocorre exclusivamente produção de CO2. Assim, cada MFE de cada grupo foi normalizado pelo respectivo valor de consumo de glucose, sendo que em seguida, os MFE's resultantes foram agrupados da seguinte forma: para a rede metabólica na qual se considera a suplementação com aminoácidos foram obtidos 30 aglomerados, nos quais se inserem os 1.306.854 MFE's que levam à produção de biomassa, e 20 aglomerados nos quais se inserem os 18582 MFE's que levam à produção de CO2. Para a rede metabólica na

qual não se consideraram suplementos, os MFE's de cada função metabólica foram agrupados em 10 aglomerados. Após o passo de agrupamento, foi considerada uma determinada quantidade de glucose, sendo que a concentração de cada um dos restantes elementos foi obtida tendo em conta os fatores de proporcionalidade, que advêm dos MFE's agrupados. A aplicação deste métodos pode assim permitir o desenvolvimento de um meio de cultura de uma forma mais rápida e mais econômica.

Palavras Chave: MFE (Modos de Fluxos Elementares); *Saccharomyces cerevisiae*; Meio Quimicamente Definido; Envirómica Funcional.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

AC	Acetate	**PRPP**	5 Phosporibosyl pyrophosphate	**Glum**	Glutamine
Acet	Acetaldehyde	**Pro**	Proline	**Fum**	Fumarate
ADP	Adenosine 5- diphosphate	**MYTHF**	Methyl Tetrahydrofolate	**FAD**	Flavin adenine dinucleotide
AKG	alpha- Ketoglutarate	**PSACCH**	Polysaccharides	**Val**	Valine
AKI	alpha- Ketoisovalerate	**Rib5P**	Ribose 5- Phosphate	**Glyo**	Glyoxilate
Ala	Alanine	**Ribu5P**	Ribulose 5 Phosphate	**GMP**	Guanosine 5 monophosphate
AM	Amino Acids	**RNA**	Ribonucleic acid	**His**	Histidine
AMP	Adenosine 5 monophosphate	**Sed7P**	Sedoheptulose 7 phosphate	**HOM**	Homoserine
Arg	Arginine	**Ser**	Serine	**Ileu**	Isoleucine
ASN	Asparagine	**SO4**	Sulfate	**IMP**	Inosine 5 monophosphate
Asp	Aspartate	**Suc**	Succinate	**Isocit**	Isocitrate
ATP	Adenosine 5- triphosphate	**Met**	Methionine	**E4P**	Erythrose – 4 Phosphate
CARP	Carbanyl Phosphate	**SucCoA**	Succinyl coenzyme A	**Lac**	Lactate
CHO	Chorismate	**THF**	Tetrahydrofolate	**Leu**	Leucine
Cit	Citrate	**THR**	Threonine	**Lys**	Lysine
CMP	Cytidine 5 monophosphate	**Tryp**	Tryptophan	**FBA**	Flux Balance Analysis
CO2	Carbon dioxide	**Tyr**	Tyrosine	**CSM**	Complete supplement mixture
CoA	Coenzyme A	**UMP**	Uridine 5 monophosphate	**BSM**	Brent supplement mixture
CTP	Cytidine 5 triphosphate	**UTP**	Uridine 5 triphosphate	**HSM**	Hollenberg supplement mixture
Fruc6P	Fructose 6 Phosphate	**O2**	Oxygen	**GRAS**	Generally Recognized as Safe
G3P	3 – Phospho Glycerate	**OL**	Oleate	**Gly**	Glycine
GAP	Glyceraldehyde 3-Phosphate	**NADH**	Nicotinamide adenine dinucleotide reduced	**FADH2**	Flavin adenine dinucleotide reduced
Gluc	Glucose	**Mal**	Malate	**Cys**	Cysteine
Gluc6P	Glucose 6-Phosphate	**METHF**	Methylene Tetrahydrofolate	**EtOH**	Ethanol
GOH	Glycerol	**Oax**	Oxaloacetate	**MFA**	Metabolic Flux Analysis
H2O	Water	**NADPH**	Nicotinamide adenine dinucleotide phosphate reduced	**Glut**	Glutamate
NAD	Nicotinamide adenine dinucleotide	**NH4**	Ammonia	**EFM**	Elementary Flux Mode
PEP	Phospho- enolpyruvate	**Pal**	Palmitoleate	**DoE**	Design of Experiments
Pi	Phosphate	**NADP**	Nicotinamide adenine dinucleotide phosphate	**Xyl5P**	Xylose 5 - monophosphate
Pyr	Pyruvate	**Phen**	Phenylalanine	**FTHF**	Formyl - Tetrahydrofolate

1. Introduction

"We are living in Exponential times". The explosion in the number of new technologies that have become available in the post genomic era allows the study of biological systems at various levels (genomic, transcriptomic, proteomic, and endo-exo metabolomic) **[1]**. The field of biology evolved to become more than a discipline that so far was mostly descriptive **[2]**. Nowadays, biology embraces multidisciplinarity, including mathematics, physics and programming methods **[3]** giving rise to Systems Biology.

The vision of Systems Biology is to better address the diversity of causes and effects in biological networks, considering simultaneously quantitative measures of multiple components and their integration into mathematical models, which were developed in a biological context **[4]**. The application of these methods allows us to understand the different types of interactions that occur among the different levels of biological systems, giving us a predictive vision of such biological system **[5]**.

Constraint based modelling has emerged as a valuable tool, it makes use of constrains which are e.g. given by the metabolic network (i.e. a stoichiometric matrix) **[6, 7]**.

Figure 1.1. Schematization of Constraint Based Modelling steps.

Constraint Based Modeling comprises several methodologies such as Elementary Flux Modes (EFM) **[8]**, Metabolic Flux Analysis (MFA) **[9]** and Flux Balance Analysis (FBA) **[10, 11].** These methods enable different ways of analyzing the steady state flux region. In this way the potential phenotypes of several microorganisms can be exploited, such as *Escherichia coli* and *Saccharomyces cerevisiae* **[12]**.

Saccharomyces cerevisiae is an ascomycete organism, belonging to the budding yeast group (group of unicellular fungi). The application of this organism is not new for humankind. For centuries *Saccahromyces cerevisiae* was used in fermentation processes, especially for wine and beer production. Nowadays, this yeast is seen as something more than a 'tool' for fermentation processes. *Saccharomyces cerevisiae* was the first eukaryote for which the

complete genome was sequenced. It has been observed that several cellular mechanisms such as: DNA and chromosomal replication, gene expression, cell division, translation, and other ones remain still conserved between yeast and higher eukaryotes, which allowed *Saccharomyces cerevisiae* to become an eukaryote model organism **[13]**. Favorable characteristics of the yeast, such as the low biohazard (GRAS, Generally Recognized as Safe to work with) and its rapid growth in simple cultivation conditions have lead to many biological studies of this organism, and make it an interesting target not only for scientific research, but also for industrial application.

In fact, *Saccharomyces cerevisiae* upholds an important role in the industries. For instance, the pharmaceutical or petrochemical industries **[14, 15]** are researching and applying mainly yeast organisms in order to obtain products of high value. In this context, systems biology methods are of vital importance. They allow the biological quantification and integration of theoretical and experimental data, for models development, thereby enabeling the analysis of the system, calculating and predicting the yields and optimizing the process.

1.1 Metabolic Engineering

Metabolic engineering is defined as a set of strategies that can be applied to a given organism with the objective of improve the production of a particular compound or a particular cellular aspect by modifying or creating new biochemical reactions, making use of recombinant DNA technology [**16, 17**]. Thus, the application of metabolic engineering for a given organism comprises three phases: the design, the construction, and finally, the analysis.

During the design step, the pathways of a microorganism are (re-)designed to produce or improve the production of a desired compound. The main approaches that must be taken in order to accomplish this step are based on genetic modification [**18, 19**]. Therefore, a comparative analysis is performed, in order to identify other organisms that i) might have the desired metabolic pathways, and ii) whose phenotype resembles the organism to which the changes are intended to be made. This also presents an opportunity to check genetic differences between different organisms, and to identify the responsible genes for the desired phenotype, referred to as inverse metabolic coops, which also is a widely-used approach.

The construction phase comprises the implementation of the mechanisms that are involved in the expression of the genes, which might impact on the desired phenotype. These genes can be modified by overexpression, deletion or attenuation, according to the effect that these changes have on the phenotype [**18**].

Thus, in order to understand and analyze quantitatively the impact of the applied modifications in a particular microorganism, a set of integrative approaches can be applied, allowing to explore how the changes (previously done) made to the level of a given cell sector, affect the overall state of the organism [**18**].

1.2 Methodologies for culture media development

1.2.1 Traditional culture media development by statistical design of experiments

There are today more advanced Systems Biology tools that can be applied for less empirical and tendentially mechanistic culture medium design. Kell and co-workers showed that there is a tight link between the exometabolome (the concentration of all extracellular metabolites) and the intracellular state **[25]**. They showed that exometabolome dynamics provides an informative and accurate "footprint" of cellular metabolic activity and indirectly of genomic and proteomic states. Indeed, the medium transports not only the essential nutrients but also small molecules and proteins involved in gene expression regulation (e.g. [26]). The use of exometabolomics to improve the composition of culture media has however never been described before.

An alternative to the modulation of particular biochemical transformations, is function oriented engineering, which is the approach explored in this thesis. The metabolism of a cell can be decomposed into elementary cellular functions using the elementary flux modes analysis or extreme pathways analysis **[27]**. An elementary flux mode represents an unique and non-decomposable sub network of metabolic reactions that works coherently in steady state (see below). The universe of elementary flux modes is primarily determined by the genome of the cells. Examples of elementary modes determination from genome wide reconstructed metabolic networks can be found in **[28, 29]**.

In **[25, 30]** a method for culture media development based on cell functional enviromics is proposed. This method is also based on the concept of elementary mode. Contrary to metabolic engineering, cell functional enviromics is less invasive. It is based on the analysis and exploration of the different physiological states of an organism [**25**, **30**]. This strategy consists essentially in the analysis of all metabolites that are excreted from the intracellular environment to the extracellular environment, as well, the analysis of the group of metabolites which are absorbed from the extracellular environment into the intracellular environment [**31**, **32**].

This type of approach is relevant because it complements metabolic engineering, in that modifications in the extracellular environment influence the expression of genes, which

determine the synthesis of all proteins that again regulate the metabolic phenomena, among others [**33**, **34**, **35**]. Of course, an interplay exists between the extracellular environment and the metabolism, in that the environment impacts on the metabolism and the cellular metabolism on the surrounding environment, leaving a metabolic footprint [**31**, **36**]. The metabolic footprint corresponds to the effects of the genetic and metabolic regulation in a more global context.

Thus, by analyzing the information that is left in the extracellular level with statistical methods, it is possible to unravel and distinguish different physiological states that may occur in an organism. In this work, specifically, the different physiological states will correspond to different sets of 'active' elementary modes (described later). The activation of a certain set of elementary modes comes with a specific demand in the concentrations of the medium compounds that compose the extracellular environment [**37**].

So, functional enviromics is a valuable tool that very well complements metabolic engineering, allowing to understand the physiological state of a given organism by linking the footprint left in the extracellular environment to the set of active metabolic reactions [**37**].

1.3 *In silico* culture media design

The elementary mode concept is central in in this thesis. Herein we revise the mathematical modeling methods for the determination of network elementary modes.

1.3.1 Stoichiometric Modelling

When a given metabolic network is considered, we can define it as a set of chemical reactions, where ***m*** compounds (metabolites) convert in each other, through ***n*** chemical reactions **[38]**. So, if we consider the stoichiometric information embedded in the set of chemical reactions, and translate it into a mathematical representation, we will obtain a stoichiometric matrix with m × n dimensions **(where m corresponds to the number of metabolites and n is number of reactions)**. Once the stoichiometric matrix has been determined, the mass balances of intracellular metabolites can be mathematically represented by a set of ordinary differential equations:

$$\frac{d\boldsymbol{c}}{dt} = \boldsymbol{S}.\boldsymbol{v} - \mu.\boldsymbol{c} \qquad \text{Equation 1}$$

In the above equation $\boldsymbol{c}$ represents the concentration of the given intracellular metabolites $\boldsymbol{m}$, $\boldsymbol{v}$ is the vector that contains the reaction rates (fluxes), $\boldsymbol{S}$ is the stoichiometric matrix (stoichiometric coefficients), and μ is the specific growth rate of cells. This equation considers that concentration $\boldsymbol{c}$ of the metabolites $\boldsymbol{m}$ vary **[39]**.

However, in stoichiometric modeling the dynamic behavior of a given intracellular metabolite is disregarded. Pseudo-steady state for the intracellular metabolites is assumed, in other words, the concentration values of intracellular metabolites are not subject to any kind of variation. The homogeneous system of linear equations with the following general equation reads:

$$\boldsymbol{S}.\boldsymbol{v} = \boldsymbol{0} \qquad \text{Equation 2}$$

The Equation 2 defines the space for every feasible flux distribution $\boldsymbol{v}$ that respects the steady-state condition [**39**, **40**].

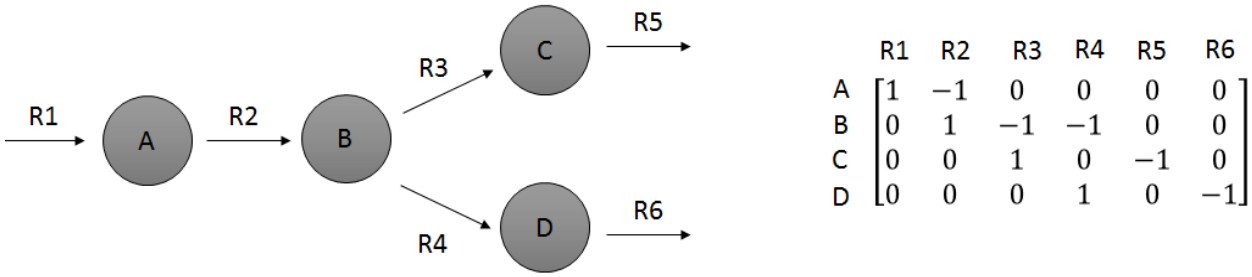

Figure 1.2. Schematic representation of a small metabolic network (left side), and respective stoichiometric matrix (right-side) [38].

In Figure 1.2 a small metabolic network is presented that constitutes the metabolites A, B, C, and D (nodes) which convert each other, through the reactions R1, R2, R3, R4, R5, and R6 (arrows). In the right side of this figure the stoichiometric matrix is presented that corresponds to the metabolic network on the left. The numeric values that can be observed in the stoichiometric matrix belong to the stoichiometric coefficient that a given metabolite has when it is subject to a given reaction. So, when the coefficient is 1 it means that a given metabolite is produced, for the other side when the coefficient is -1 it means that a given metabolite is consumed. In case that the coefficient has value 0 it means that a given metabolite does not participate in the corresponding reaction **[38]**.

1.3.2 Elementary Flux Modes

The metabolism of a cell can be decomposed into elementary cellular functions using elementary flux modes analysis [27]. An elementary flux mode represents an unique and non-decomposable sub network of metabolic reactions that works coherently in steady state (see below). The main difference between these metabolic sub-sets is the unique combination of metabolic fluxes that are active in each one of those sub-sets [**41, 42**].

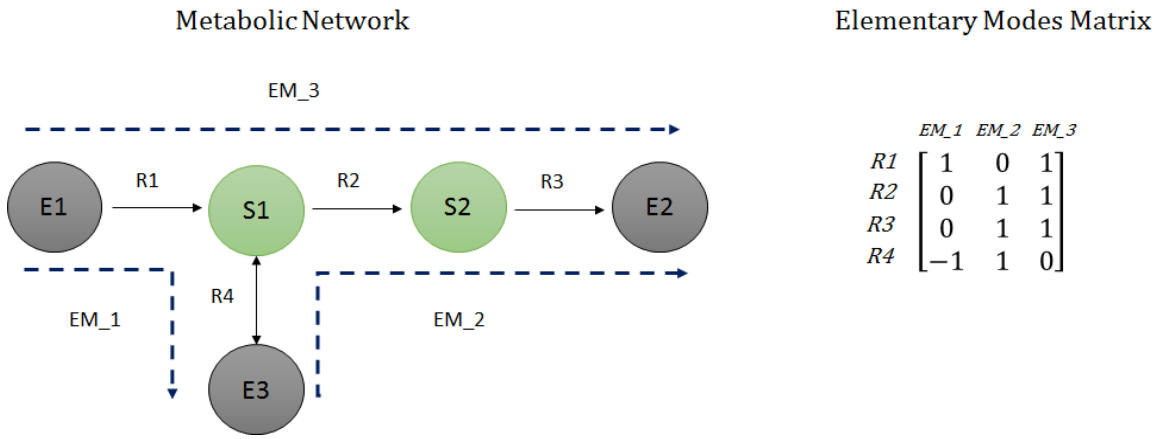

Figure 1.3. Schematic representation of metabolic network and respective elementary modes matrix. The green nodes represent the internal metabolites and grey nodes represent the external metabolites. The solid arrows represent the reactions that occur in the considered metabolic network. The solid arrows R1, R2 and R3 are irreversible reactions, the solid arrow R4 represents a reversible reaction. The dash arrows represent the all metabolic routes that can occur in the represented metabolic network [43].

Therefore, one elementary mode can be mathematically defined as being column vector ***E*** with **n×1** dimensions, where ***n*** corresponds to the number of reactions considered in the metabolic network. The set of all elementary modes is contained in a matrix, where each column represents a vector, which in turn represents an elementary mode, see figure above. When one of the rows in these vectors has a value of zero, means that the corresponding reaction v_i does not occur in the given elementary mode. On the other side, if the rows have a value different from zero, it means that reaction v_i occurs in elementary mode ***E*** [**8**, **42**]**.** If the values is positive then the reactions occurs in forward direction else in reverse direction.

1.3.2.1 EFM-Mathematical definitions

An elementary mode can be defined as a column vector. The calculation of elementary mode vectors must respect certain constraints. The first constraint tells us that the set of all elementary modes are comprised in the null space, imposed by Equation 2 ($\mathbf{S}.\mathbf{v} = \mathbf{0}$). The null space corresponds to the *Kernel* matrix of stoichiometric matrix [**44, 45**]. This first condition must be adhered, in order to preserve the assumed pseudo-steady state, i.e. no variation of metabolites concentrations occur.

The second constraint is due to the fact that in metabolic networks there exist irreversible fluxes (v^{irr}) and reversible fluxes (v^{rev}). This constraint restricts the sign of irreversible fluxes as being **non-negative**.

$$\boldsymbol{v^{irr} \geq 0} \qquad \text{Equation 3}$$

The v^{irr} represents the reaction rates that occur in one direction only, for the other side the v^{rev} represents the reactions that can proceed in either direction under physiological conditions, such as the reactions shared by glycolysis and gluconeogenesis. Therefore this reactions v^{rev} have no sign restriction **[45]**.

The last constraint is related to the non-decomposability criteria (simplicity). This criteria tells us that a given elementary mode **E** cannot be represented as a positive linear combination of two other elementary modes ***E'*** and ***E''*** ($\mathbf{E} \neq \boldsymbol{\lambda 1}\,\mathrm{E}' + \boldsymbol{\lambda 2}\,\mathrm{E}'',\ \lambda 1,\ \lambda 2 > 0$) when the last two have the following properties:

1. ***E'*** and ***E''*** obey to the first two constraints;
2. ***E'*** and ***E''*** contain at least the same number of zero elements as ***E***, and at least one of them contains more zero elements than ***E***;
3. For all indices ***i*** corresponding to boundary reactions, the components are not of opposite sign, sign (vi') ≠ - sign (vi'') **[46]**.

However, there exist elementary modes that do not obey to these conditions and are the result of positive linear combinations of two other elementary modes (as can be seen in the next figure).

$$EM\_1 \quad EM\_2 \quad EM\_3$$
$$\begin{bmatrix} 1 \\ 0 \\ 0 \\ -1 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 0 \end{bmatrix}$$

***F*igure 1.4. Positive linear combination of EM_1 and EM_2 resulting in EM_3 [43].**

As can be seen in Figure 4, the EM_3 (elementary mode that belongs to the metabolic network from Figure 2) results (mathematically) from the positive linear combination of EM_1 and EM_2 **[43]**.

The conditions under which a given elementary mode ***E*** can be the result of positive linear combination of two other elementary modes are described in **[43]**. The application of the non-decomposability criteria implies that in a biological system a given EFM will not be able to carry a steady-state flux, if any of its contributing reactions is deleted **[42]**.

The result of the intersection of all subspaces generated by all of the previous constraints, is a convex polyhedral cone ***F***, that corresponds to the space where all elementary flux modes exist.

$$\boldsymbol{F} = \{\mathbf{V} \in \mathbb{R}^{\boldsymbol{n}} : \mathbf{V} = \sum \boldsymbol{\lambda}'.\boldsymbol{f}' + \sum \boldsymbol{\beta i}.\boldsymbol{bi}, \quad \boldsymbol{\lambda}', \boldsymbol{\beta i} \in \mathbb{R}, \boldsymbol{\lambda}' \geq \mathbf{0}\} \qquad \text{Equation 4}$$

In the above equation two types of vectors can be distinguished: the vectors $\boldsymbol{f}'$ (irreversible flux vector) for which the negative, $\boldsymbol{-f}'$, is not contained in the flux cone, and the vectors $\boldsymbol{bi}$ (reversible flux vectors) for which the negative, $\boldsymbol{-bi}$, is contained in *F*. The $\mathbb{R}^n$ corresponds to the space with n dimensions (n is the number of reactions considered in the metabolic network) of flux cone *F*. $\boldsymbol{\beta i}$ and $\boldsymbol{\lambda}'$ are scalars, and $\boldsymbol{v}$ is the set of all vectors that span and are contained in flux cone *F* **[47]**.

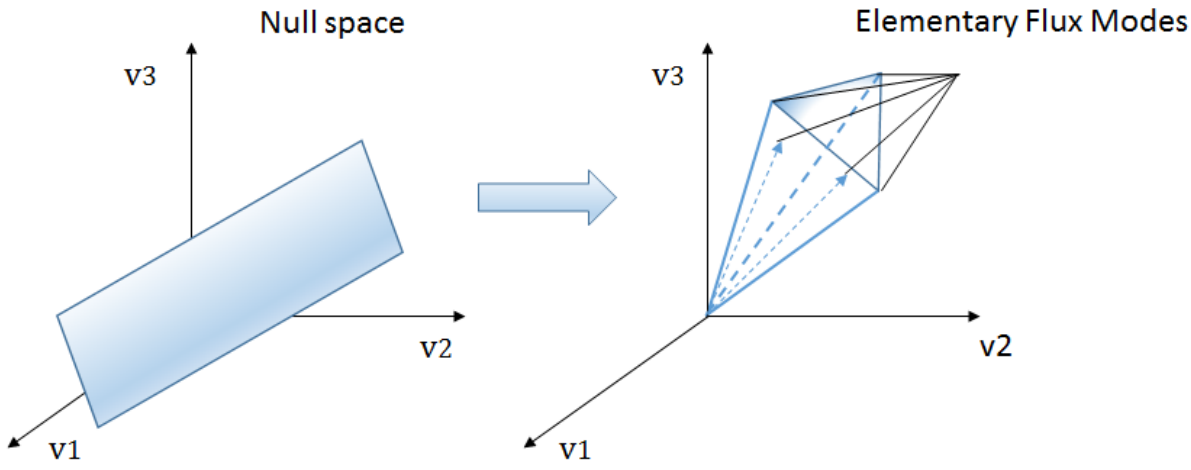

Figure 1.5. Schematic representation of a null space (space defined by Equation 2), and the corresponding convex polyhedral cone (Flux Cone) containing the elementary modes inside it. The number of axis that are considered when Flux Cone is spanned corresponds to the number of chemical reactions that occur in a given metabolic network.

Considering the elementary mode definition, it is then mathematically possible to describe the fluxome (the entire set of cellular fluxes) that occurs at the cellular level. The *in vivo* active cellular functions constituting the fluxome F' can thus be described by linear combinations of all elementary modes, En, whose corresponding scalar value γn will be the relative intensity of the respective elementary mode occurrence [**48**].

$$\boldsymbol{F' = \gamma 1\, E1 + \gamma 2\, E2 + \gamma 3\, E3 + \cdots + \gamma n\, En} \qquad \text{Equation 5}$$

1.4 Yeast Media

1.4.1 Complex Medium

By definition, a complex medium (a rich medium) is a type of culture medium where the most part of needed metabolites, such as amino acids, vitamins, nucleotides and other metabolic intermediates are present in excess **[49]**. Thus, complex media allow to obtain high cellular growth rates, because a great part of anabolic precursors are present in the medium, and can be channeled directly into anabolic pathways, reducing the need to produce biosynthetic precursors and thus saving metabolic energy **[50]**.

In the case of *Saccharomyces cerevisiae*, the most commonly used complex medium is YPD (Yeast Extract Peptone-Dextrose) that allows the yeast cells to divide every 90 minutes during the exponential growth phase of the growth cycle **[49]**.

Table 1.1 Composition of YPD medium [49].

Components	Final Concentration (g/L)
Yeast Extract	10
Dextrose (D-Glucose)	20
Peptone	20

1.4.2 Defined Medium

A chemical defined medium can be described as being a medium where all compounds and respective quantities are known **[49, 51]**. Usually, some of these media are known as 'minimal' media, because it contains only the nutrients that satisfy no more than minimal requirements of growth **[52, 53]**.

The most common and commercial available defined medium for yeast culture is the Yeast SD medium (Yeast Synthetic Defined Medium). The Yeast SD medium consists of a defined mixture of vitamins, salts, and a nitrogen source that is known as YNB medium (Yeast

Nitrogen Based). To this YNB medium a carbon source is added, usually dextrose (D-glucose) and a mixture of supplements that contain amino acids and nucleotide precursors **[49]**.

There are several mixtures of supplement formulations (see appendix Table 6.3) commercially available for *Saccharomyces cerevisiae* (CSM, Complete supplement mixture; BSM, Brent supplement mixture; HSM, Hollenberg supplement mixture) **[49, 54, 55]**. Although, the choice of which supplements can be added is made by each researcher/customer.

1.4.3 Defined Medium vs Complex Medium

The main difference between defined and complex medium is that in defined medium all compounds and respective quantities are known, whereas in complex medium the composition of some compounds remains unknown, and even the known compounds do not have constant amounts **(batch-to-batch variations)**. The use of a defined medium has been proven to bring advantages, for instance it is possible to determine the role that each medium compound has in growth/metabolism of a given organism. The defined medium also offers the opportunity for maximal process control, so problems such as batch-to-batch variations can be avoided (this kind of problem is commonly associated to the use of complex medium), in this way ensuring reproducible results at different times and at different laboratories **[52]**. Another advantage of defined mediums is that they are much less sensitive to sterilization conditions than complex medium **[56]**.

However, due to the fact that higher fermentation productivities are obtained with complex mediums **[50, 53]**, they are still the dominant medium in industry scale **[53]**. In spite of all the defined medium's advantages, their major problem is the development of their formulation. Since the formulation typically contains many different compounds, large DoE's become necessary in order to evaluate which impact certain compounds have on the metabolism. Therefore, the development of defined medium formulations is expensive and time consuming, and thus defined mediums are more expensive than complex mediums used for the same purpose.

To conclude, defined medium has revealed valuable properties for the industrial field, portending that it will be a great mean for production of biological products, such as proteins and polysaccharides **[53]**.

1.5 Objectives

As was previously mentioned, the use of a defined medium brings advantages as compared to complex medium, but currently one of the major barriers that are associated to the use of defined medium in relation to complex medium is the development of their formulation. In other words, the development of defined medium requires statistical data that are obtained empirically, which makes this process highly expensive and time consuming.

The focus of this work is to rationally develop a culture medium formulation, based on the assumption that extracellular components and the internal metabolism influence each other. Thus, by calculating all possible EFM's and considering simultaneously the entry and exit of certain metabolites (boundary reactions), it is possible infers the extracellular conditions (medium factors) that activate a certain metabolic function by the analysis of the metabolic footprint left at the extracellular level. This renders the deduction of a possible medium formulation, which from a metabolic point of view, it makes sense. This also offers the possibility to adjust the medium formulation accordingly to the desired product for a given microorganism. In other words, for the same microorganism, it is possible to deduce different medium formulations in order to obtain different products. Therefore, the application of this methodology allows us to understand, which are the medium requirements that will promote a particular cellular function, without requiring the execution of extensive numbers of experiments, such reducing the time and cost for the development of the medium formulation.

The main objective of this thesis is the rational development of principle culture medium formulations for *Saccharomyces cerevisiae*, based on the knowledge about the interplay of internal metabolism and external environment, such taking into account different metabolic states. In order to accomplish this work the following tasks were fulfilled:

1. Consideration of a particular metabolic network of *Saccharomyces cerevisiae* and calculation of the elementary modes with the EFMtool algorithm.
2. Identification of elementary modes that lead to a desired physiological state.
3. Clustering of the previously selected elementary modes, in this way obtaining a smaller but higher significant sample of the selected EFMs.
4. Develop medium formulations that cover the physiological states, previously selected.

2. Materials and Methods

2.1 Metabolic Network

In this work, the considered metabolic network was adapted from [**57**]. The original network consists of 99 reactions, from which 39 are reversible and have been found 98 metabolites. This network includes the following metabolic pathways:

- Glycolysis;
- TCA cycle;
- Amino acid metabolism;
- Nucleotide metabolism;
- Glyoxylate cycle;
- Phosphate Pentose pathway;
- Oxidative Phosphorylation;
- Polysaccharides synthesis;
- Fatty Acids Synthesis reactions
- RNA synthesis;
- General Protein Synthesis;
- Biomass Synthesis.

The main changes that were made to the original network consisted in not consider the proton (H^+), the omission of pyruvate, citrate, succinate, acetate and gluconic acid uptakes in order to make the calculation of EFM's, otherwise would be not possible to calculating them, because it there no exists enough memory. Moreover, it was added the uptake of nine amino acids (glutamate, glutamine, lysine, asparagine, alanine, arginine, proline, aspartate and glycine). According to *Görgens et al 2005* **[58]**, each one of these amino acids has a positive contribution for cell growth.

2.1 Calculating Elementary Modes

As previously said, this work had as main objective the medium formulation according to the different metabolic states that occur at intracellular level. Thus, knowing that each one of these metabolic states is characterized by the occurrence of specific elementary modes, there was the need of calculating them.

So, in order to make the elementary modes calculation, it was applied the EFMtool 4.7.1 algorithm, available in open source at http://www.csb.ethz.ch/tools/efmtool.

The EFMtool algorithm **[59]** was interfaced with MATLAB, version 8.1.0.604 (R 2013a) 64-bit in a 2.4 GHz PC with 8 GB RAM.

Among the main results obtained once the EFMtool algorithm is applied are a list of all metabolic reactions and an 'efms' matrix (elementary modes matrix), where the rows correspond to the metabolic reactions, and the columns correspond to the elementary modes. Since in this last matrix are listed all the resulting elementary modes, in order to obtain the matrix with the corresponding metabolic footprint, it was created a new matrix only considering the exchange reactions, which corresponding dimensions were 17x99 (17 extracellular metabolites). Therefore, by multiplying the exchange reactions matrix by efms matrix it is obtained the metabolic footprint matrix; the corresponding dimensions of this matrix are: the number of rows corresponds to the number of external metabolites, and the number of columns corresponds to the number of EFM's.

2.2 Determination of metabolic footprint

In order to capture the influence that intracellular metabolism has in the extracellular medium components, a matrix only containing the EFM's ratios that correspond to Boundary Reactions (Exchange Reactions) was acquired

2.3 Data normalization and Clustering

As can be seen later, the number of calculated elementary modes usually is really high, from thousands to millions. Theoretically, each one of these elementary modes corresponds to a metabolic state, and so, to a medium formulation. However, it is not reasonable to test thousands or millions of medium formulations, thus, knowing that there are EFM's that are very similar among them, it means that can be grouped in the same set.

Note: The data normalization and clustering were applied to the metabolic footprint matrix.

However, before preforming the elementary modes clustering, some data normalization steps were implemented. The data normalization was performed in two steps: in the first step all the corresponding values of each elementary mode were divided by the respective absolute value of glucose input (as can be seen in next figure).

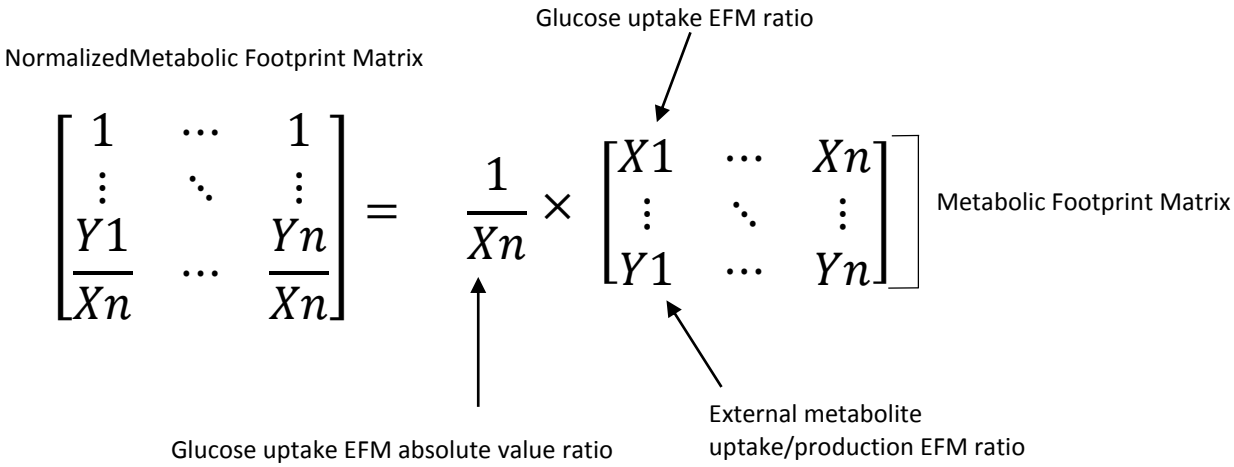

Figure 2.1 Schematic representation of how first the normalization was made.

The objective of the first step normalization was to prevent that linearly dependent EFM's were grouped into different clusters.

The second normalization was performed by row. The coefficients of each metabolite were subtracted by the mean value correspondent to each metabolite, and then was divided by the respective standard deviation value. The second normalization was performed in order to ensure all metabolites have coefficients in the same range and their contribution to the cluster analysis will not be biased. It is noteworthy that second normalization can only be considered in the clustering step, because each coefficient of the same EFM was divided for a different value, resulting this way in a different EFM. Therefore, the second normalization was reversed after the clustering, by multiplying the coefficients of each metabolite for the respective standard deviation value and then by summing, this last obtained value, for the metabolite corresponding mean value.

Before data normalization and clustering the resulting EFM's of each metabolic network were divided into two groups: EFM's that lead to Biomass formation; EFM's that lead only to CO2 production/Cellular Respiration. Therefore, the clustering was applied to each one of these two EFM's groups, separately.

The clustering was performed in MATLAB using the *K-means* algorithm. The K-means algorithm is a partitional technique that by iterative methods minimize the sum of distances between the points (clustering data; in this case EFM's) and centroids.

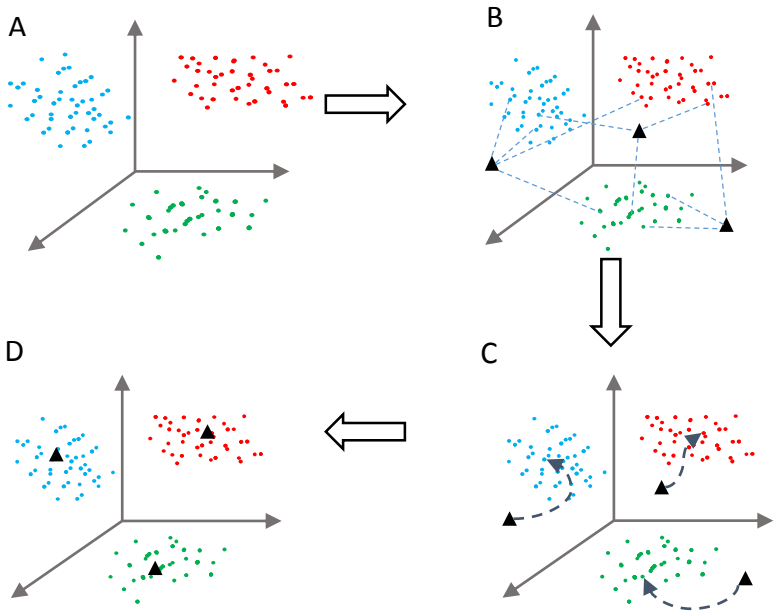

Figure 2.2 Schematic representation how K-means algorithm works. The data (colored points) are initially dispose in a multi-dimensional space (A). The centroids (black triangles) are randomly dispose, and the distances between them and data points are calculated (B). The black triangles coordinates are recalculated (C). The black triangles represent each group of data-points (D).

For using K-means it is needed to define the *k* number of intended clusters, and then the algorithm randomly determines the *k* clusters centers that are recalculated in order to minimize the distance between them and EFM's surrounding points, like it is schemed in the previous figure (Figure 2.2).

The choice of the number of clusters was made considering an acceptable explained variance of the data:

$$Var = 1 - \frac{SQres}{SQtot} \qquad \text{Equation 6}$$

Where Var is the explained variance, $SQres$ is the square of the sum of all the distances from each point (representing each EFM) to the respective cluster center; and $SQtot$ is the square of the sum of all points analyzed. The randomness of the initial choice can be a limitation when it is not done properly. To avoid it, the algorithm was set to repeat the operation 50 times, each with a new initial set of centers, and to chose the set with the lowest $SQres$ value. The maximum number of iterations allowed was increased to 5000.

2.4 Principle culture Medium Formulation

After K-means results were obtained (each cluster corresponds to a medium formulation), in order to calculate the medium formulation, the cluster results were multiplied by glucose amount (mol). The value of glucose amount by which the cluster coefficients were multiplied was 0.11 that correspond to a glucose concentration of 20 g/L, 2% (w/v).

Then, the resulting values were multiplied by the respective molecular weights (g/mol), obtaining in this way the medium formulations where each supplementing metabolite concentration is presented in (g/L). In Biomass case, according to *Lange et al 2001* **[60]** the considered molecular weight was **26.171 (g/C-mol).**

3. Results and Discussion

Since this work is based on the formulation of culture media by considering different metabolic states that may occur at intracellular level, the initial step consisted in analyzing a *Saccharomyces cerevisiae* metabolic network, and using the EFMtool algorithm the calculation of the different metabolic states that this metabolic network could achieve (which mathematical representation translates into elementary modes) was made.

In this work, it was also analyzed the importance / influence that certain amino acids may have at metabolic level, and how these ones may lead to the occurrence of different metabolic states, which can be advantageous, in relation to those ones that are obtained when the supplementation with amino acids is not made. In this study, the Synthesis of Biomass and Cellular Respiration represent the metabolic activities that were examined. Therefore, in order to understand how this group of amino acids influence the occurrence of such metabolic states, as well the yields associated with these same metabolic states, a comparative study was conducted. This study had consisted in the calculation of elementary modes for two distinct metabolic networks. A metabolic network that considered the amino acids supplementation, and another one that did not consider any kind of supplementation, except the minimal medium compounds (phosphate, ammonia, glucose and sulfate).

3.1 *Saccharomyces cerevisiae* metabolic network

As previously mentioned, the metabolic network considered in this work was adapted from *Vanrolleghem et al* (1996) **[57]**, however, it was needed to make some changes to the original metabolic network that are descripted above, section 2.1. The metabolic pathways that were previously enounced (Amino acid Synthesis; Glycolysis; TCA cycle; etc.) were maintained.

The resultant metabolic network (supplemented with nine amino acids) has 82 metabolites, and 99 metabolic reactions: 25 are reversible and there are 17 boundary reactions (in the supplemented metabolic network). In the non-supplemented metabolic network there is a total of 90 metabolic reactions: 25 are reversible and there are 8 boundary reactions that include consumption of oxygen (O_2), sulfate (SO_4), ammonia (NH_4), inorganic phosphate (Pi), and glucose (Gluc) and Biomass and carbon dioxide (CO_2) production. The boundary reactions (Exchange Reactions) of supplemented metabolic network take into account the consumption of the nine amino acids (listed above), oxygen (O_2), sulfate (SO_4),

ammonia (NH_4), inorganic phosphate (Pi), and glucose (Gluc). The supplemented metabolic network boundary reactions also include Biomass and carbon dioxide (CO_2) production.

The metabolic reaction for biomass synthesis is the following **(reaction 81 from table 6.1):** 47003 Protein + 35376 PSACCH + 5234 RNA + 344 Pal + 344 OL + 226 GOH + 162100 ATP --> 100000 Biomass + 162100 Pi + 162100 ADP

Note: This reaction obeys to the macromolecular composition published in reference [57].

So, and as previously said, in order to understand how the amino acids supplementation affects, and influence the occurrence of certain metabolic states, we will determine the principle culture media formulations for different culture media supplementations scenarios:

Case 1. Minimal media without amino acids supplementation;
Case 2. Media with amino acids supplementation.

So, in Case 1 will be studied by considering the non-supplemented *Saccharomyces cerevisiae* metabolic network.

On the other side, Case 2 will be studied by considering the supplemented *Saccharomyces cerevisiae* metabolic network, in which are considered glutamate (Glut), glutamine (Glum), proline (Pro), arginine (Arg), aspartic acid (Asp), asparagine (ASN), glycine (Gly), lysine (Lys), and alanine (Ala) uptakes.

Although, the necessity of choosing a smaller group of amino acids was caused by the limitation of not being possible to perform the elementary modes calculation when the number of reactions in the metabolic network increases. The elementary modes calculation requires some memory (java heap memory), and as higher the number of calculated elementary modes is, more memory is needed. So, in order to keep the number of calculated elementary modes at a reasonable value, the number of metabolic network reactions must not to be very high. Therefore, the reduction of number of 'Exchange Reactions' was the most appropriate measure.

3.2 Case 1: No Amino acids supplementation

In this section will be determined how much elementary modes can occur in *Saccharomyces cerevisiae* metabolic network in a growth medium without supplementing metabolites. Also, it will be analyzed how certain metabolic functions, such as Cellular Respiration and Cell growth are affected by the extracellular minimal medium components.

3.2.1 Determination of elementary modes

By the analysis of the *Saccharomyces cerevisiae* metabolic network without amino acids supplementation (Case 1), a total of 635 EFM's were calculated by EFMtool algorithm, and divided as follows:

• 215 EFM's correspond to the set of elementary modes that lead to Biomass formation;

• 420 EFM's correspond to the set of elementary modes that lead exclusively to Cellular Respiration.

In next table it is presented one example of one EFM where biomass is produced in non – supplemented medium.

Table 3.1 Example of an elementary mode (EFM) in which there is biomass production in non-supplemented medium, R# corresponds to reaction number of the considered metabolic network.

R1	9,02E+05	**R19**	4,60E+05	**R37**	-1,62E+04	**R55**	1,84E+04	**R73**	5,34E+03
R2	-1,55E+06	**R20**	2,35E+06	**R38**	0,00E+00	**R56**	2,43E+03	**R74**	2,28E+03
R3	0,00E+00	**R21**	7,84E+05	**R39**	9,41E+05	**R57**	9,23E+03	**R75**	2,28E+03
R4	0,00E+00	**R22**	1,55E+06	**R40**	2,41E+04	**R58**	2,20E+04	**R76**	2,28E+03
R5	4,09E+03	**R23**	7,84E+05	**R41**	0,00E+00	**R59**	2,68E+04	**R77**	1,00E+00
R6	7,68E+05	**R24**	7,84E+05	**R42**	7,89E+03	**R60**	1,27E+04	**R78**	6,22E+03
R7	2,53E+04	**R25**	7,71E+05	**R43**	7,70E+03	**R61**	1,41E+04	**R79**	6,22E+03
R8	0,00E+00	**R26**	9,39E+04	**R44**	7,70E+03	**R62**	1,26E+04	**R80**	1,07E+05
R9	6,33E+05	**R27**	9,39E+04	**R45**	0,00E+00	**R63**	6,41E+03	**R81**	1,81E+01
R10	0,00E+00	**R28**	0,00E+00	**R46**	1,37E+04	**R64**	4,88E+03	**R82**	3,52E+06
R11	0,00E+00	**R29**	1,63E+06	**R47**	7,43E+05	**R65**	1,34E+03	**R83**	3,63E+06
R12	4,05E+05	**R30**	1,78E+05	**R48**	1,62E+04	**R66**	1,22E+04	**R84**	-7,27E+06
R13	0,00E+00	**R31**	0,00E+00	**R49**	7,14E+05	**R67**	3,15E+03	**R85**	1,81E+06
R14	3,11E+05	**R32**	1,72E+06	**R50**	2,77E+03	**R68**	1,76E+01	**R86**	9,02E+05
R15	2,62E+05	**R33**	0,00E+00	**R51**	5,52E+04	**R69**	1,76E+01	**R87**	7,67E+03
R16	2,60E+05	**R34**	1,05E+04	**R52**	4,88E+03	**R70**	2,33E+03	**R88**	2,55E+05
R17	3,56E+05	**R35**	4,66E+03	**R53**	0,00E+00	**R71**	0,00E+00	**R89**	2,77E+03
R18	3,66E+05	**R36**	2,43E+03	**R54**	2,08E+04	**R72**	2,33E+03		

3.2.2 Analysis of Cell Growth Footprint

Before data clustering were obtained the all set Biomass producing EFM's metabolic footprint in which it is reflected the influence that all Biomass producing EFM's have in extracellular medium.

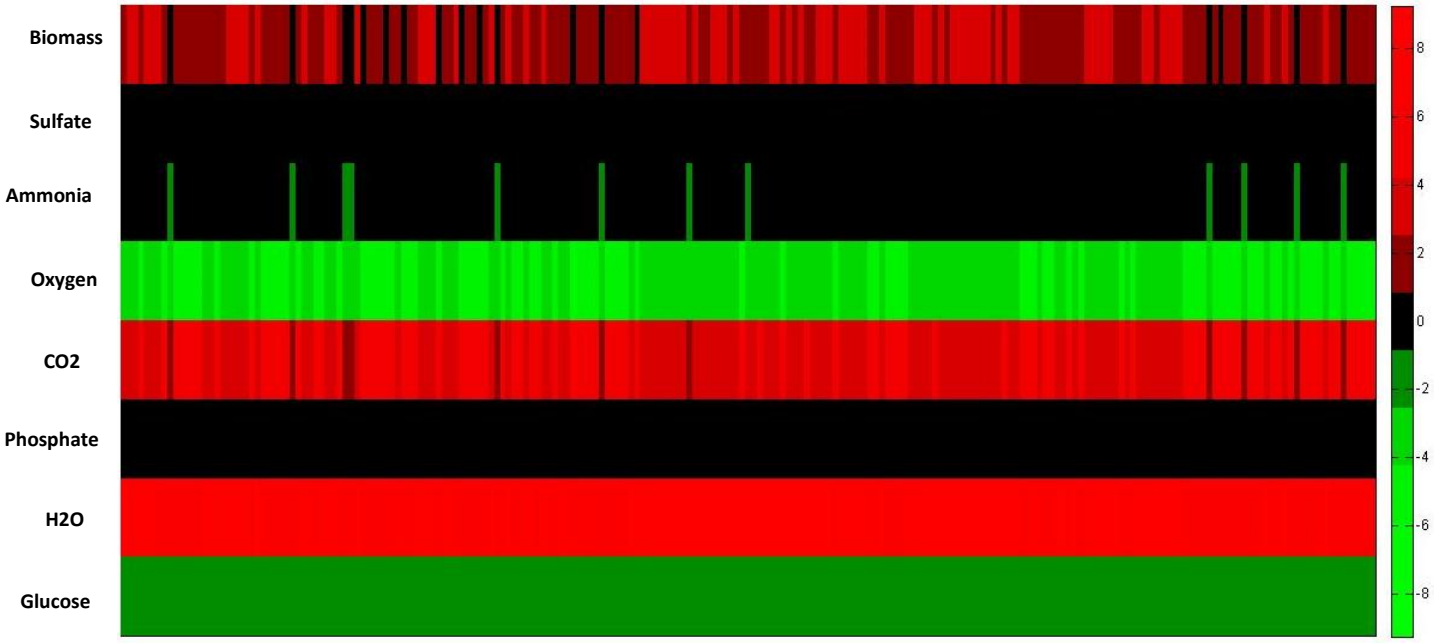

Figure 3.1 Heat map of all EFM's that lead to Biomass production. These data is normalized by glucose uptake value. The green color means that corresponding metabolite is consumed; the red color means that corresponding metabolite is produced; the black color means that corresponding metabolite is poorly consumed or produced.

In order to group the different 215 EFM's into smaller and more significant sets, the clustering algorithm *K-means* was applied. However, before it will be presented a graph (Figure 3.2), where it is possible to observe how the explained variance values vary according to the number of clusters (*k*) that are considered when *K-means* algorithm is applied.

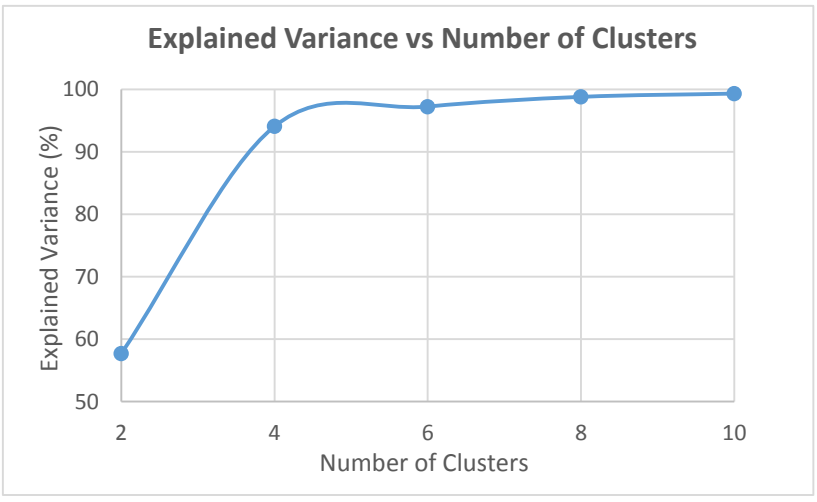

Figure 3.2 Evolution of explained variance of the results from cluster analysis (of Cell Growth EFM's in minimal medium) as function of the number of clusters considered in K-means clustering.

In this case the EFM's were grouped into 10 clusters, with an explained variance of **99.32%.**

Through the graphics presented below, where the overlap of EFM's and corresponding clusters is made, it is possible to have an idea how certain compounds that constitute the non – supplemented medium are influenced by cell growth.

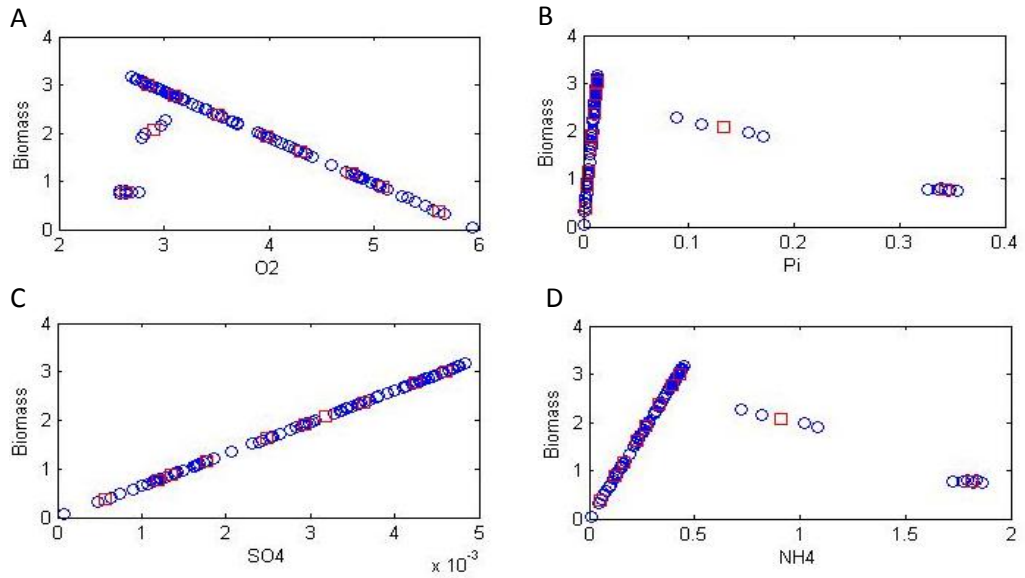

Figure 3.3 Set of graphics where is represented the Biomass formation in function of some external metabolites. The blue circles are the points that correspond to each EFM. The red squares are the points that correspond to each cluster (these values are normalized for glucose consumption ratio value).

By the observation of Figure 3.1 graphics it is possible to conclude that all minimal medium components have a positive effect in biomass formation. In graphic **3.1 A** it is possible to observe a typical behavior, in which initially the consumed oxygen (**around 2.7 to 3.1**) is associated with oxidative phosphorylation (cellular respiration) that will be translate into ATP energy that will be used for anabolic activities, contributing this way for biomass formation increase. But as oxygen consumption increases the carbon source degradation increases as well. This leads to a decrease of biomass formation, since there is less carbon to be incorporated in biomass formation. This behavior is described in **[61]**.

Moreover, it is possible to observe that Pi and NH_4 consumptions (graphics B and D, respectively) have a direct relation to biomass formation. Since that, Pi is involved in reactions that lead to ATP synthesis (Oxidative Phosphorylation R28-R32; ATPase R38, see table 6.1) that is posteriorly used in anabolic functions. On the other hand, NH_4 is directly involved in amino acids synthesis, which integrates all the protein content in a cell. However, for a higher consumption values of Pi and NH_4 this relation is dismissed. Probably NH4 and Pi metabolites are diverted for other metabolic functions, such as Cellular Respiration.

Finally, by observing graphic C we can conclude that SO_4 consumption is completely related to biomass formation, and by the later results is possible to observe that SO_4 metabolites does not have any kind of contribution for cellular respiration. It is important also taking into account that sulfate is the less consumed minimal medium compound. Biologically, this makes sense, since that sulfur compounds in cellular medium are very few, probably the amino acids methionine and cysteine are the biological compounds where the most part of sulfur element is present in a biological system. However, sulfate is very used to role the medium osmolality and electrical charge.

Therefore, by the results obtained using *K-means* algorithm, it was made a Heat map (Figure 3.4), where are represented, using a color code, the 10 clusters with cell growth (biomass production).

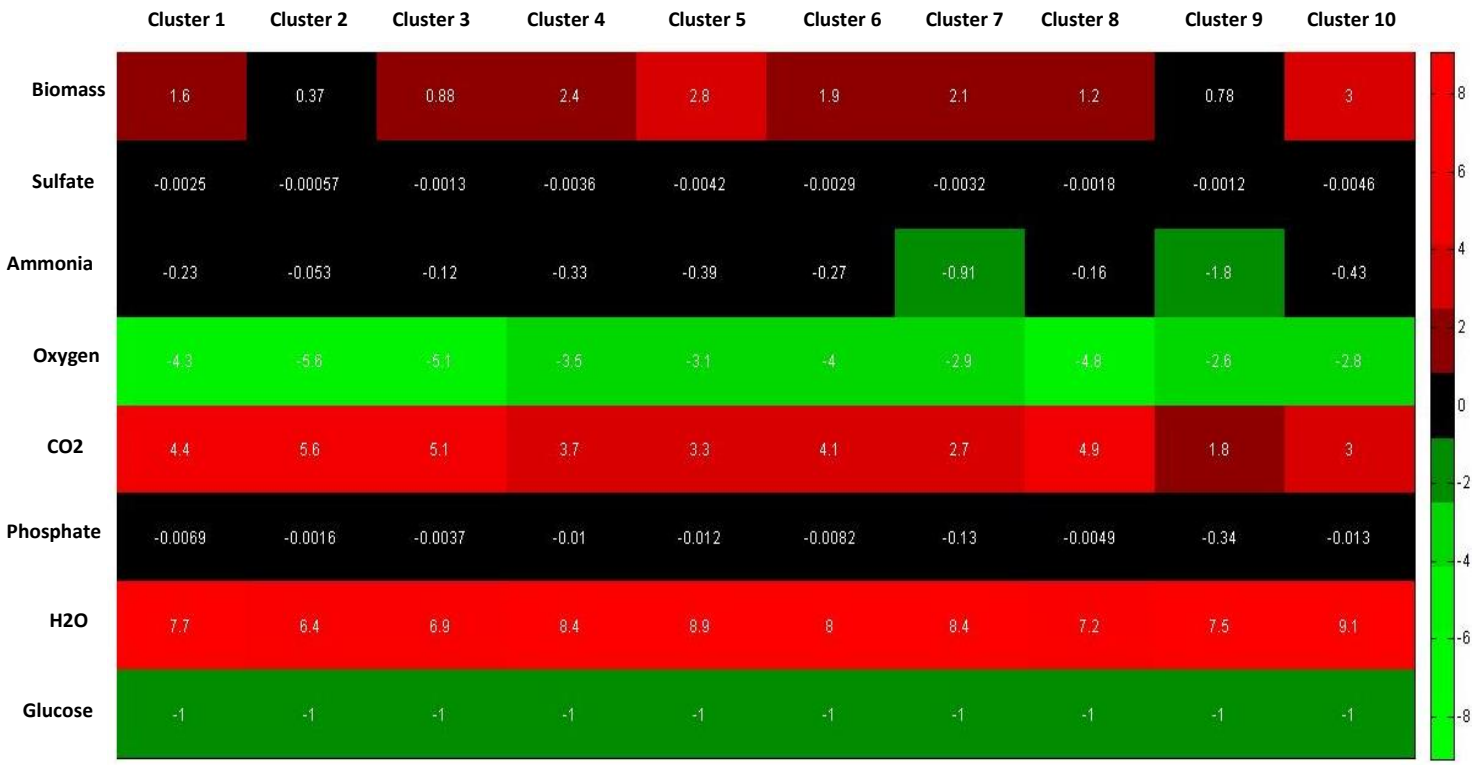

Figure 3.4 Heat map in which are presented the numerical values of grouped EFM's (10 clusters) that lead to biomass formation in non-supplemented metabolic network. The green metabolites are consumed, red metabolites are produced, and the black metabolites are poorly produced or consumed. (These values are normalized for glucose consumption ratio value).

So, by taking into account the clusters values shown above, it is possible to determine a medium formulation (one for each cluster) maintaining the proportions among the stoichiometric coefficients within each cluster. The resultant medium formulation is presented in next table.

Table 3.2 Respective medium formulations for each cluster obtained for non-supplemented medium and resulting Biomass content. All the values are in units of g/L.

External Metabolites	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7	Cluster 8	Cluster 9	Cluster 10
Glucose	20	20	20	20	20	20	20	20	20	20
Phosphate	0.07	0.02	0.04	0.11	0.12	0.09	1.40	0.05	3.63	0.13
Ammonia	0.43	0.1	0.23	0.63	0.74	0.51	1.72	0.31	3.43	0.8
Sulphate	0.03	0.01	0.01	0.04	0.05	0.03	0.03	0.02	0.01	0.05
Biomass	4.72	1.08	2.55	6.89	8.06	5.59	6.03	3.37	2.25	8.76

3.2.3 Analysis of Cellular Respiration Footprint

In next figure it is represented the metabolic footprint of all 420 Cellular Respiration EFM's, showing its influence in extracellular medium components.

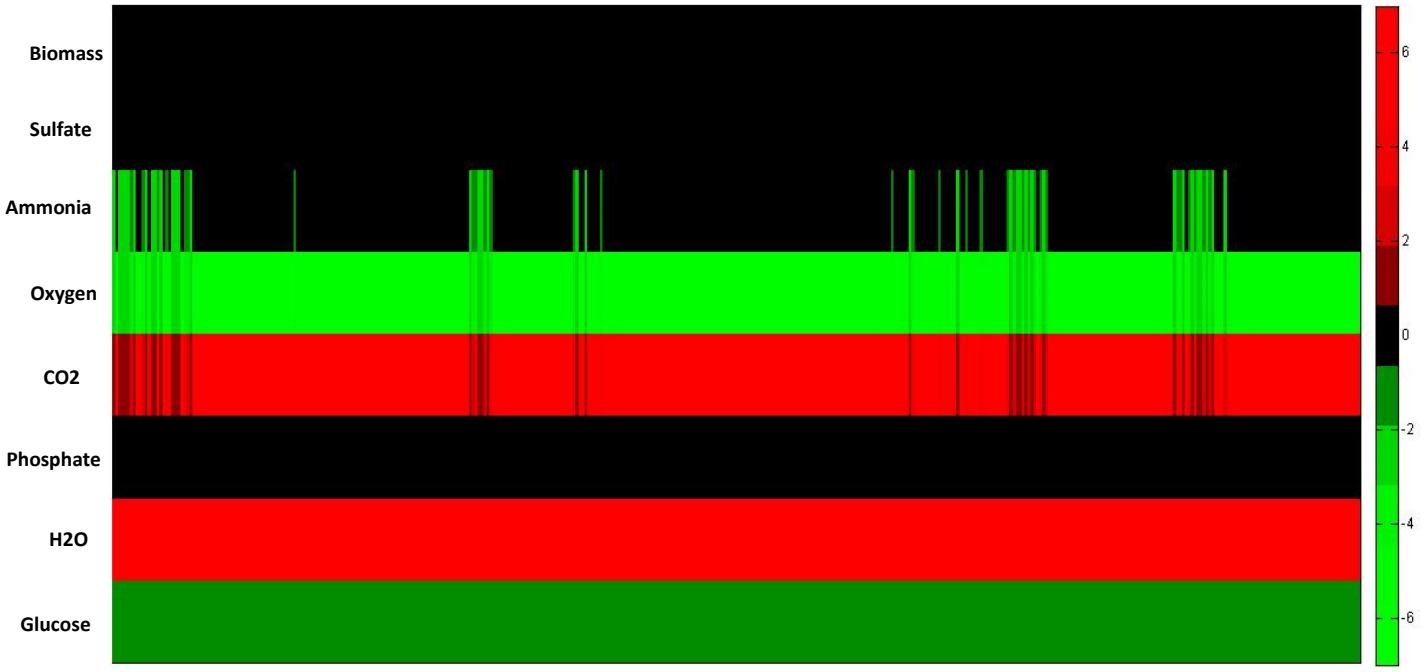

Figure 3.5 Heat map that corresponds to all EFM's that lead exclusively to Cellular Respiration. These data is normalized by glucose uptake value. The green color means that corresponding metabolite is consumed; the red color means that corresponding metabolite is produced; the black color means that corresponding metabolite is poorly consumed or produced.

By using the *K-means* clustering algorithm the 420 EFM's obtained for Cellular Respiration metabolic function were also grouped into clusters. For this case it was also made the study of how the explained variance values vary according to the considered number of clusters (k).

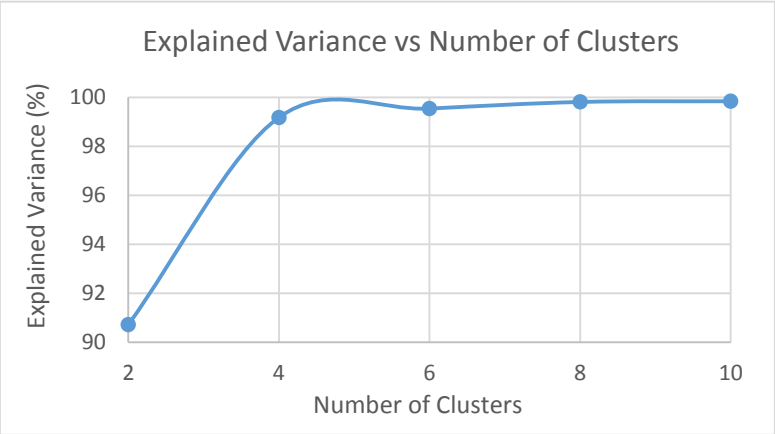

Figure 3.6 Evolution of explained variance of the results from cluster analysis (of Cellular Respiration EFM's in non-supplemented medium) as function of the number of clusters considered in K-means clustering.

Thus, the 420 EFM's, responsible for Cellular Respiration, were grouped into 10 clusters, with an explained variance of **99.84%.** From the graphics presented below, where the overlap of EFM's and corresponding clusters is made, it is possible to have an idea how certain medium compounds that constitute the non – supplemented medium are influenced by Cellular Respiration in *Saccharomyces cerevisiae*.

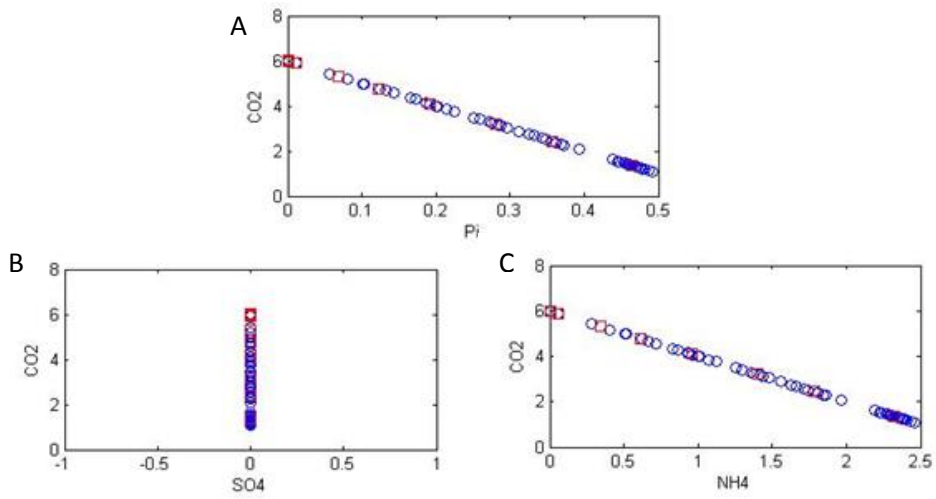

Figure 3.7 Set of graphics where is represented Cellular Respiration (CO2 production) in function of some external metabolites. The blue circles are the points that correspond to each EFM. The red squares are the points that correspond to each cluster (these values are normalized).

By the observation of graphics presented in the figure above, it is possible to conclude that Pi and NH_4 uptakes (graphs A and C, respectively) are inversely proportional to CO_2 production (Cellular Respiration). However the Pi uptake behavior is unexpected and contradictive, because CO_2 production is highly associated to ATP synthesis, which request Pi consumption, so it would be expected the increasing of Pi uptake when CO2 production increases. Relatively to NH_4 uptake, this result makes sense, since NH_4 is associated to anabolic functions. Thus, since Cellular Respiration consists in a set of catabolic reactions, it is expected that ammonia uptake decreases with the increasing of catabolic functions. It is also possible to confirm that when metabolic functions exclusively related with Cellular Respiration occur, the uptake of SO_4 metabolite does not happen (Graph B).

With the results obtained using *K-means* algorithm, a Heat Map was made, representing (through a color code) the 10 clusters where Cellular Respiration (CO2 production) occur, exclusively. (Figure 3.8 – next figure)

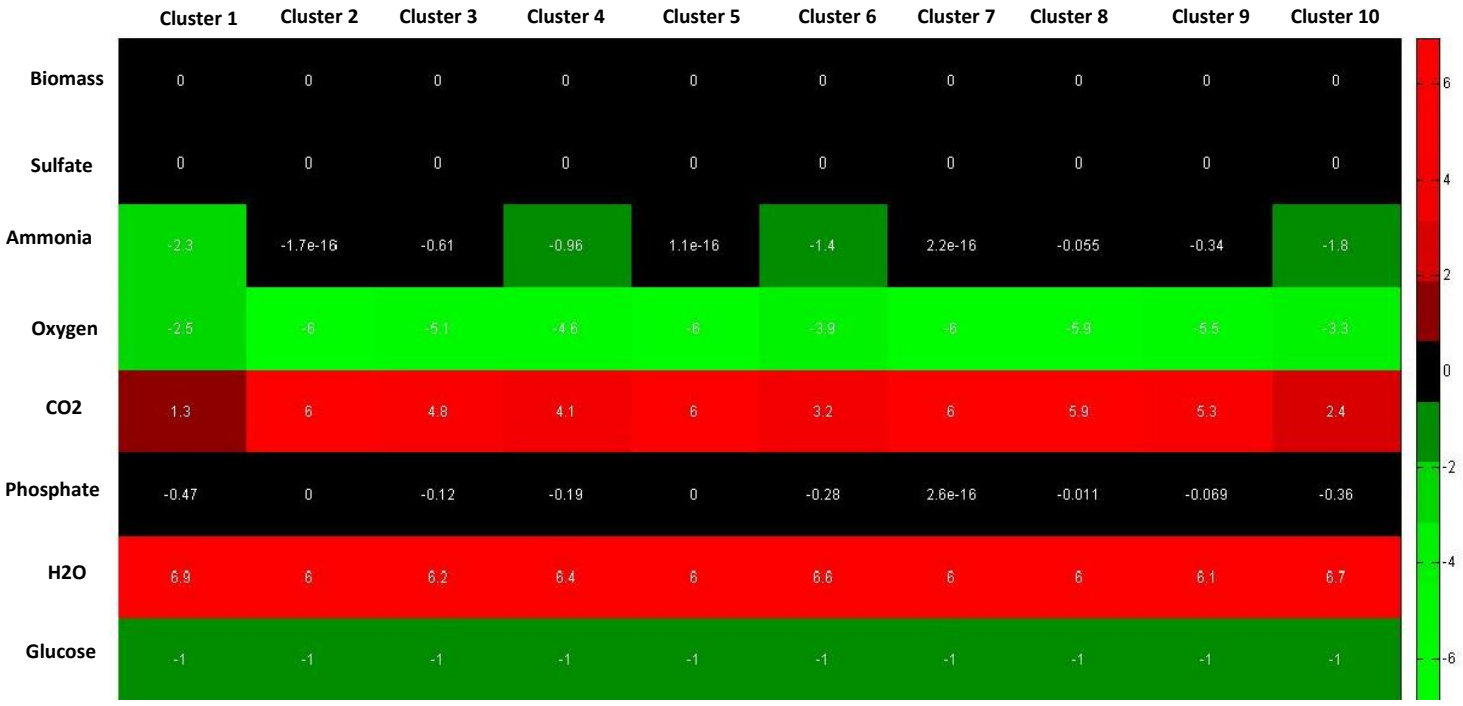

Figure 3.8 Heat map where are presented the numerical values of grouped EFM's (10 clusters) that lead exclusively to cellular respiration in the non-supplemented metabolic network. The green metabolites are consumed, red metabolites are produced, and the black metabolites are poorly produced or consumed. (These values are normalized for glucose consumption ratio value).

3.3 Case 2: Amino acids Supplementation

In this section it will be determined how many elementary modes can occur in *Saccharomyces cerevisiae* metabolic network, in which it is considered amino acids supplementation, and how certain metabolic functions, such as Cellular Respiration and Cell growth are affected by the extracellular supplementing medium components.

3.3.1 Determination of Elementary Modes

By the analysis of the *Saccharomyces cerevisiae* metabolic network where the amino acids supplementation was considered (Case 2), a total of 1352172 EFM's were calculated using EFM tool, which are divided as follows:

• 1306854 EFM's correspond to the set of elementary modes that lead to Biomass formation;

• 18582 EFM's correspond to the set of elementary modes that lead exclusively to Cellular Respiration.

In next table it is presented an example of one EFM where biomass is produced within amino acid supplementing medium.

Table 3.3 Example of an elementary mode (EFM) in which there is biomass production in supplemented medium. R# corresponds to reaction number of the considered metabolic network.

R1	5,69E+05	**R21**	4,50E+05	**R41**	0,00E+00	**R61**	1,41E+04	**R81**	1,81E+01
R2	-8,88E+05	**R22**	8,88E+05	**R42**	0,00E+00	**R62**	1,26E+04	**R82**	2,11E+06
R3	0,00E+00	**R23**	4,50E+05	**R43**	0,00E+00	**R63**	6,41E+03	**R83**	2,29E+06
R4	0,00E+00	**R24**	4,50E+05	**R44**	0,00E+00	**R64**	4,88E+03	**R84**	-5,56E+06
R5	4,09E+03	**R25**	4,38E+05	**R45**	0,00E+00	**R65**	1,34E+03	**R85**	1,81E+06
R6	4,35E+05	**R26**	2,32E+05	**R46**	0,00E+00	**R66**	1,22E+04	**R86**	5,69E+05
R7	-2,44E+05	**R27**	2,32E+05	**R47**	6,79E+05	**R67**	3,15E+03	**R87**	7,67E+03
R8	0,00E+00	**R28**	1,13E+06	**R48**	0,00E+00	**R68**	1,76E+01	**R88**	0,00E+00
R9	5,85E+05	**R29**	0,00E+00	**R49**	6,66E+05	**R69**	1,76E+01	**R89**	2,77E+03
R10	0,00E+00	**R30**	0,00E+00	**R50**	2,77E+03	**R70**	2,33E+03	**R90**	0,00E+00
R11	0,00E+00	**R31**	1,33E+05	**R51**	0,00E+00	**R71**	0,00E+00	**R91**	7,70E+03
R12	2,32E+05	**R32**	8,53E+05	**R52**	0,00E+00	**R72**	2,33E+03	**R92**	4,75E+04
R13	2,69E+05	**R33**	0,00E+00	**R53**	4,27E+04	**R73**	5,34E+03	**R93**	1,37E+04
R14	0,00E+00	**R34**	2,77E+03	**R54**	2,08E+04	**R74**	2,28E+03	**R94**	7,89E+03
R15	3,42E+04	**R35**	4,66E+03	**R55**	1,84E+04	**R75**	2,28E+03	**R95**	2,96E+04
R16	3,18E+04	**R36**	2,43E+03	**R56**	2,43E+03	**R76**	2,28E+03	**R96**	2,41E+04
R17	2,67E+05	**R37**	0,00E+00	**R57**	9,23E+03	**R77**	1,00E+00	**R97**	1,62E+04
R18	2,69E+05	**R38**	0,00E+00	**R58**	2,20E+04	**R78**	6,22E+03	**R98**	0,00E+00
R19	5,01E+05	**R39**	7,17E+05	**R59**	2,68E+04	**R79**	6,22E+03		
R20	1,35E+06	**R40**	0,00E+00	**R60**	1,27E+04	**R80**	1,07E+05		

3.3.2 Analysis of Cell Growth Footprint

For this case the corresponding Heat map representing the metabolic footprint is not presented due to high discrepancies among the EFM's values. Thus, the resulting heat map would be occupied by a large black area, not allowing any kind of analysis.

In the metabolic network where the amino acids supplementation was considered 1306854 EFM's were calculated, and posteriorly grouped into 30 clusters with an explained variance of 85.75%.

According to what was previously said, one of the approaches of this work was to understand which amino acids have a major impact on the biomass formation.

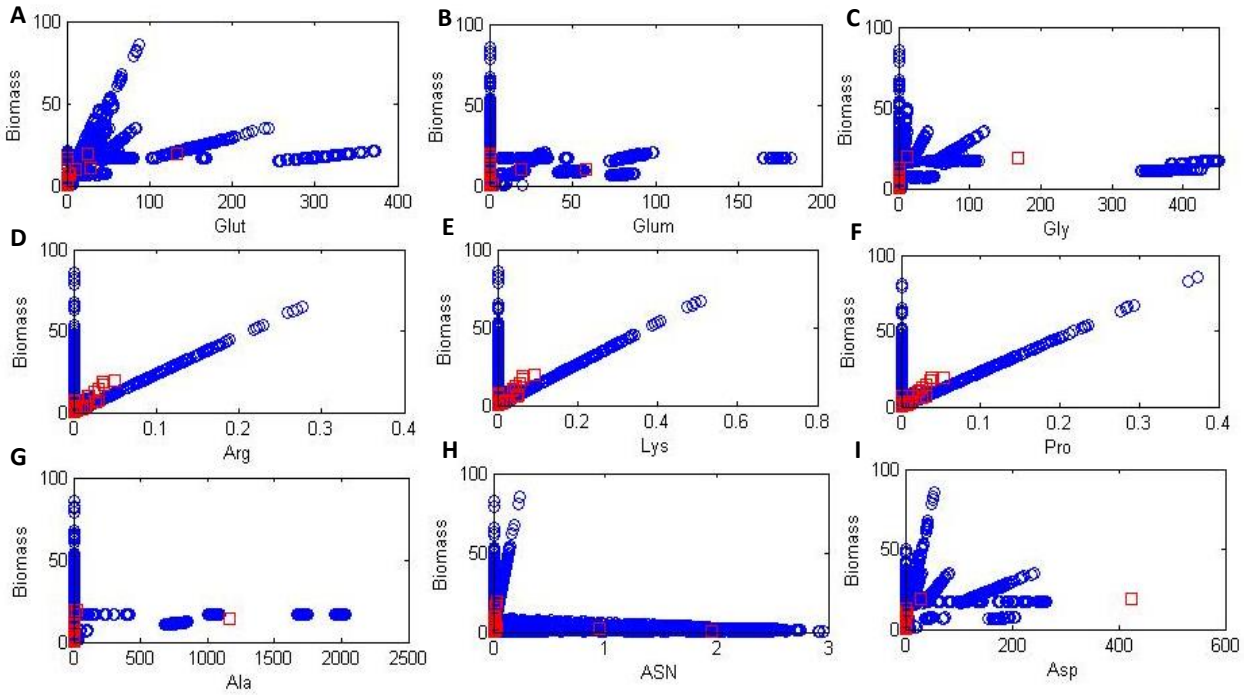

Figure 3.9 Set of graphics where is represented the Biomass formation in function of supplementing amino acids. The blue circles are the points that correspond to each EFM. The red squares are the points that correspond to each cluster. (These values are normalized).

By the observation of graphics presented above, it is possible to show that glutamate, glutamine, aspartate, alanine and glycine (Graphs A, B, I, G and C, respectively) are the most consumed amino acids (Group 1 amino acids). These results are not totally unexpected, since those amino acids are precursor compounds in the synthesis of many other compounds, such as other amino acids. Through a deeper observation of metabolic network, it is also possible to note that Group 1 amino acids are 'metabolically close' to TCA compounds. Thus, the metabolism of those amino acids can result into compounds that constitute the TCA cycle, increasing in this way the cellular respiration process, something that will be shown later (Figure 3.12).

By observing the elementary modes in which the consumption of Group 1 amino acids is higher, but biomass formation does increases as well, it would mean that those amino acids follow other metabolic pathways that do not lead only in biomass formation.

On the other hand, proline, lysine and arginine (Group 2 amino acids) are directly linked to biomass synthesis, i.e. when Group 2 amino acids are consumed, it always occurs biomass formation. Although, there are elementary modes that lead to biomass formation, but do not include the consumption of Group 2 amino acids, which means that those amino acids do not show the same metabolic versatility of Group 1 amino acids. However, it should be noted that in the considered metabolic network the catabolism reactions of Group 2 amino acids are not considered, limiting in this way the impact that Group 2 amino acids can have in other metabolic functions.

For this study it was also obtained the Heat map where the corresponding 30 clusters values are presented.

.

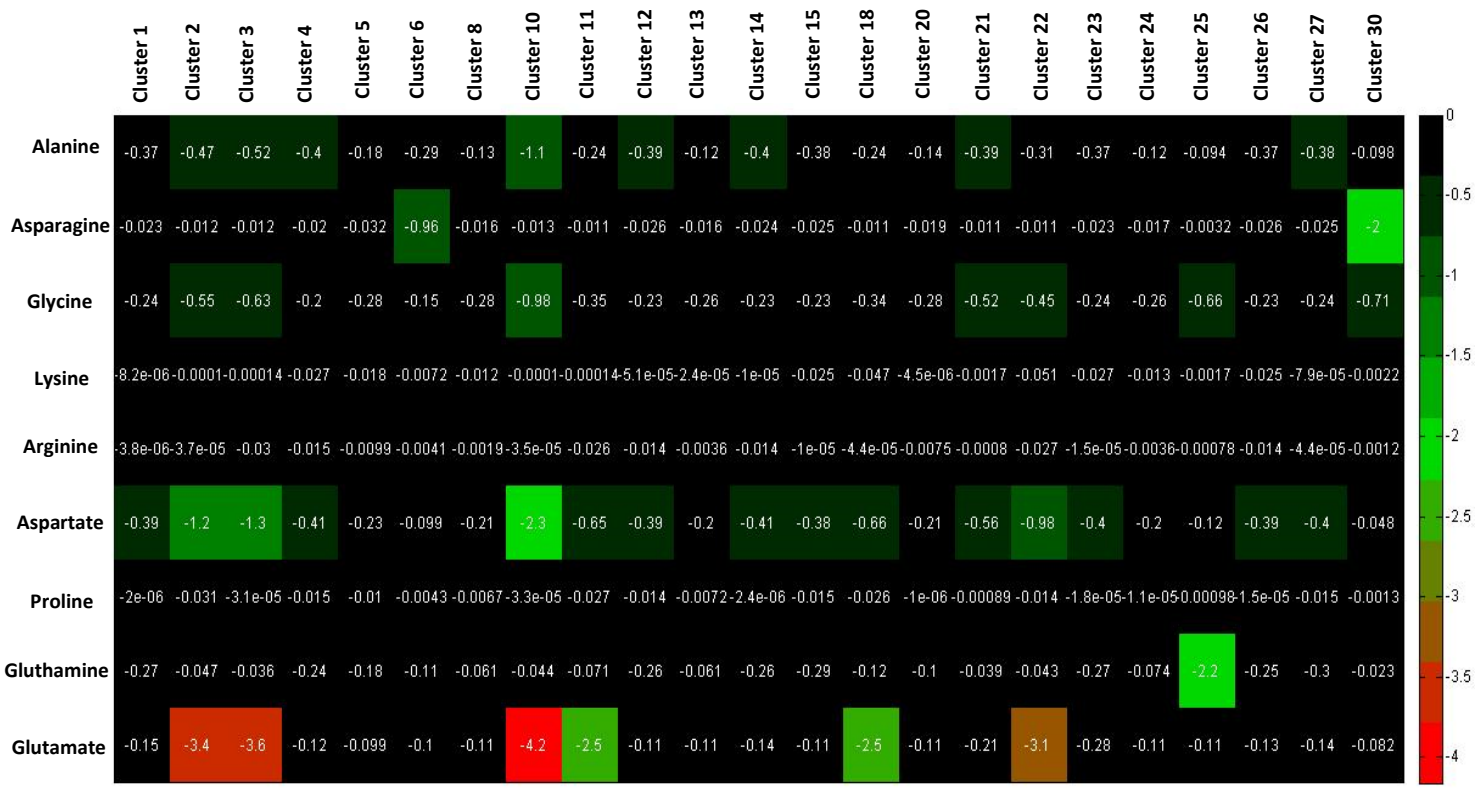

Figure 3.10 Heat map where are presented the numerical values of grouped EFM's (30 clusters) that lead to biomass formation in the supplemented metabolic network. The green metabolites are consumed, red metabolites are highly consumed, and the black metabolites are poorly consumed. These values are normalized.

The Heat map presented above correspond to the one obtained after the exclusion of certain clusters, whose corresponding final medium formulations are unrealistic and non - executable. Therefore, the excluded clusters were: **7, 9, 16, 17, 19, 28 and 29**, however these clusters can be found in the appendix. The main factor that contribute for the exclusion of previous listed clusters was the high concentrations values of certain amino acids.

Finally, knowing that heat map data correspond to values that were normalized for glucose consumption, so by assuming a given amount of glucose supplementation it is possible, for extrapolation, deduce the corresponding amounts for all other medium components. For this case **2 % (w/v) (20 g/l)** was the considered glucose amount.

Table 3.4 Respective cluster's medium formulations (g/L) obtained for amino acids supplemented medium and resulting Biomass content (g/L).

	External Metabolites										
Formulations	**Gluc**	**Glut**	**Glum**	**Pro**	**Asp**	**Arg**	**Lys**	**Gly**	**ASN**	**Ala**	**Biomass**
Cluster 1	20	2.46	4.40	0	5.81	0	0	1.99	0.33	3.65	9.40
Cluster 2	20	56.34	0.76	0.39	18.26	0	0	4.59	0.18	4.61	20.59
Cluster 3	20	58.61	0.59	0	19.36	0.58	0	5.21	0.17	5.14	20.56
Cluster 4	20	1.99	3.90	0.2	6.11	0.29	0.43	1.69	0.3	3.93	10.25
Cluster 5	20	1.61	2.95	0.13	3.34	0.19	0.29	2.34	0.46	1.81	6.94
Cluster 6	20	1.70	1.75	0.05	1.46	0.08	0.12	1.24	14.01	2.88	6.05
Cluster 8	20	1.75	1	0.09	3.07	0.04	0.19	2.35	0.24	1.26	4.50
Cluster 10	20	67.95	0.72	0	33.34	0	0	8.14	0.19	10.74	21.52
Cluster 11	20	41.05	1.16	0.34	9.67	0.51	0	2.94	0.16	2.36	17.96
Cluster 12	20	1.81	4.19	0.18	5.82	0.27	0	1.93	0.38	3.86	9.56
Cluster 13	20	1.87	0.99	0.09	2.92	0.07	0	2.20	0.23	1.14	4.82
Cluster 14	20	2.30	4.26	0	6.01	0.27	0	1.90	0.35	3.96	9.76
Cluster 15	20	1.75	4.62	0.19	5.67	0	0.41	1.96	0.37	3.73	9.74
Cluster 18	20	40.41	2.02	0.33	9.73	0	0.77	2.86	0.16	2.37	18.23
Cluster 20	20	1.80	1.69	0	3.13	0.14	0	2.33	0.28	1.35	5.21
Cluster 21	20	3.49	0.64	0.01	8.31	0.02	0.03	4.30	0.16	3.86	2.23
Cluster 22	20	50.13	0.70	0.18	14.49	0.52	0.83	3.75	0.15	3.09	19.72
Cluster 23	20	4.57	4.42	0	5.97	0	0.44	1.96	0.34	3.62	10.39
Cluster 24	20	1.79	1.21	0	2.99	0.07	0.22	2.19	0.25	1.21	5.13

Table 3.4 (Continued)

	External Metabolites										
Formulations	**Gluc**	**Glut**	**Glum**	**Pro**	**Asp**	**Arg**	**Lys**	**Gly**	**ASN**	**Ala**	**Biomass**
Cluster 25	20	1.76	36.46	0.01	1.78	0.02	0.03	5.48	0.05	0.93	1.67
Cluster 26	20	2.17	4.09	0	5.74	0.27	0.41	1.93	0.37	3.71	9.73
Cluster 27	20	2.24	4.83	0.19	5.84	0	0	1.99	0.36	3.80	9.72
Cluster 30	20	1.34	0.38	0.02	0.71	0.02	0.04	5.94	28.73	0.97	1.73

3.3.3 Analysis Cellular Respiration Footprint

For this cellular function, by considering amino acids supplementation were calculated 18582 EFM's that were grouped into 20 clusters, with an explained variance of 93.85%.

Similarly to it was done previously, it was made the study of how the explained variance values vary according to the considered number of clusters.

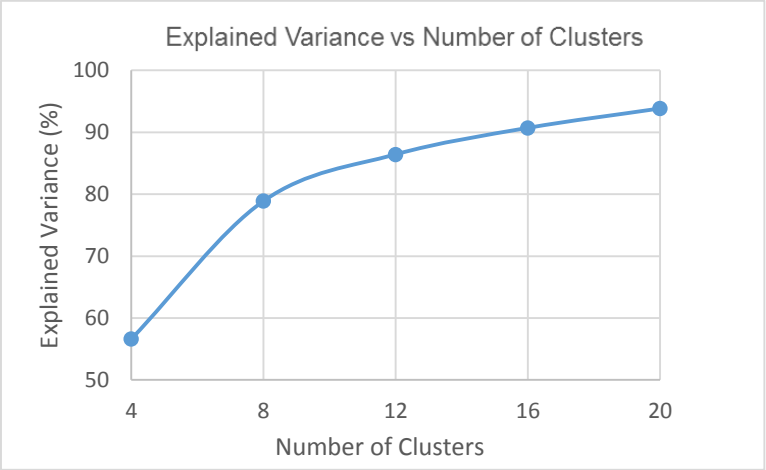

Figure 3.11 Evolution of explained variance of the results from cluster analysis (of Cellular Respiration EFM's in supplemented medium) as function of the number of clusters considered in K-means clustering.

For this case it was also obtained a set of graphics that help to figure out, which amino acids have a major contribution to cellular respiration (Figure 3.12).

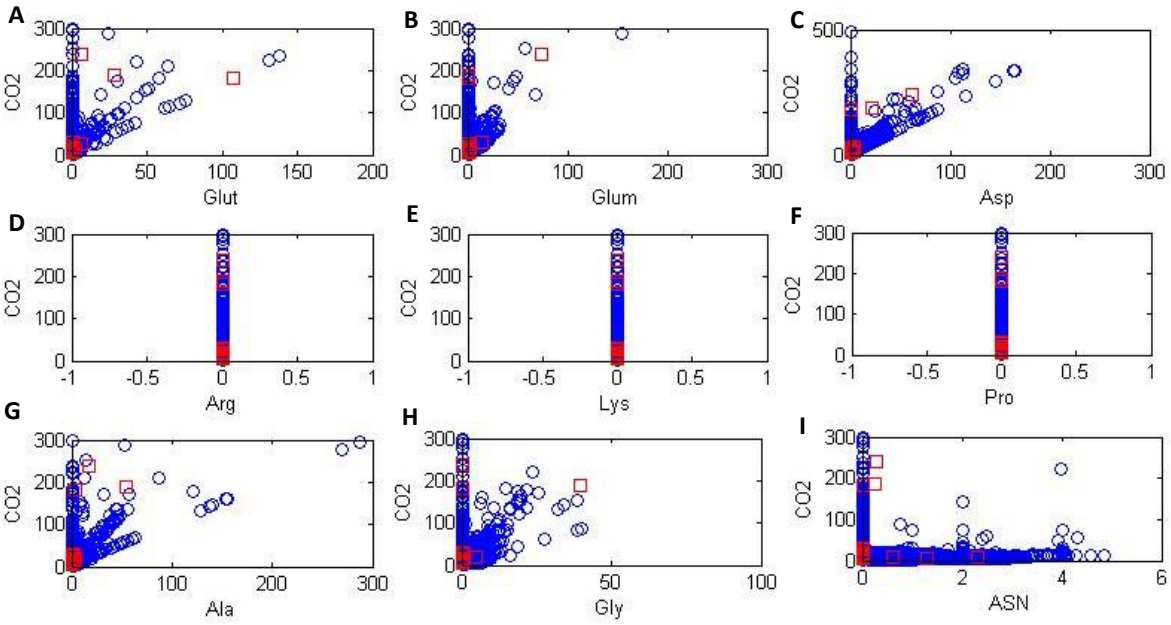

Figure 3.12 Set of graphics where is represented the Cellular respiration (CO2 production) in function of supplementing amino acids. The blue circles are the points that correspond to each EFM. The red squares are the points that correspond to each cluster. These values are normalized.

Since Cellular Respiration is highly associated with the oxygen consumption and CO2 production, it is possible to show, by the graphics presented above, that glutamate, glutamine, aspartate, alanine and glycine are the amino acids that most contribute to cellular respiration. This observation confirms the relevance of Group 1 amino acids in the cellular respiration process. However, it is shown that the set consisting of Group 2 amino acids has no contribution to cellular respiration, but we cannot forget that corresponding catabolism reactions of Group 2 amino acids are not present in the considered metabolic network.

3.4 Analysis of Amino acids supplementation

In order to understand in more detail what is the impact of amino acids supplementation in biomass formation a comparative study was made, making a comparison between the obtained results in minimal medium and those obtained in supplemented medium.

Taking into account the results presented in section 3.3, it is easily conclude that amino acids supplementation increases the metabolic range of a given organism. This is proven by the large increase of the EFM's number when was considered the amino acids supplementation (comparing to the EFM's number of non-supplemented metabolic network To be more specific, the EFM's percentage in which occurred biomass formation increased from **33,86%** to **96,65%** with amino acid supplementation, reflecting in this way the amino

acids positive contribution to biomass formation. The study of explained variance represent other proof of how the amino acids supplementation increases the metabolic diversity. Because if we want to capture all the EFM groups with *K-means* clustering it was needed to consider a larger number of k clusters. Other proof of how the amino acids supplementation has a positive effect in cell growth can be seen in tables 3.1 and 3.3 where the EFM's that lead to biomass production in non-supplemented and supplemented medium are represented, respectively. Thus, by consulting these tables it is possible to observe that in non-supplemented medium a great part of anabolic reactions **R39-R67** is active in non-supplemented medium. On the other hand, in table 3.3 it is observed that there is a large number of anabolic reactions that are not active in cell growth, this could mean that supplementing amino acids are directly incorporated into cellular metabolism, without to need of synthetize them at intracellular level.

However, a more detailed analysis of the results was performed. This analysis consisted in C-mol yields determination for both supplemented network, as well non-supplemented. Therefore, the non-supplemented network yield calculation consisted in dividing the amount of obtained biomass (C-mol) by the considered glucose amount (C-mol).

Table 3.5 Yield values from non-supplemented metabolic network.

Formulations	Yield Biomass/Glucose (C-mol/C-mol)
Cluster 1	0.271
Cluster 2	0.062
Cluster 3	0.146
Cluster 4	0.395
Cluster 5	0.462
Cluster 6	0.321
Cluster 7	0.346
Cluster 8	0.193
Cluster 9	0.129
Cluster 10	0.502

On the other hand, in the case of metabolic network in which the amino acids supplementation was considered the yields calculation consisted in two steps: in first step the yield calculation was made by dividing the obtained biomass amount (C-mol) by the glucose

amount (C-mol). The second step have consisted in calculating the resultant yield of the total Biomass amount (C-mol) by the total glucose and amino acids amount (C-mol).

Table 3.6 Yield values from supplemented metabolic network.

Formulations	Yield Biomass/Glucose (C-mol/C-mol)	Yield Biomass/total carbon content[A] (C-mol/C-mol)
Cluster 1	0.539	0.285
Cluster 2	1.181	0.228
Cluster 3	1.179	0.218
Cluster 4	0.588	0.308
Cluster 5	0.398	0.245
Cluster 6	0.347	0.166
Cluster 8	0.258	0.176
Cluster 10	1.234	0.179
Cluster 11	1.030	0.265
Cluster 12	0.548	0.291
Cluster 13	0.276	0.191
Cluster 14	0.560	0.292
Cluster 15	0.559	0.293
Cluster 18	1.046	0.267
Cluster 20	0.299	0.199
Cluster 21	0.128	0.065
Cluster 22	1.131	0.243
Cluster 23	0.596	0.293
Cluster 24	0.294	0.201
Cluster 25	0.096	0.029
Cluster 26	0.558	0.294
Cluster 27	0.558	0.290
Cluster 30	0.099	0.036

A – The total carbon content consists in the sum of glucose carbon content (C-mol) with amino acids carbon content (C-mol).

So, by the analysis of the previous tables it has been conclude that amino acids supplementation increases biomass formation. This conclusion can be taken comparing the obtained yields of **Biomass (C-mol)/Glucose(C-mol)** of respective supplemented and non – supplemented metabolic networks, where it is observed an increase of this yield from the non-supplemented to supplemented metabolic network.

However, when it is made a comparison between the yields of **Biomass (C-mol)/Glucose (C-mol)**, corresponding to non-supplemented network, and **Biomass (C-mol)/[Glucose (C-mol) + AM (total of amino acids)(C-mol)]**, corresponding to the supplemented network, it can be observed that yields corresponding to non-supplemented network are bigger than yields corresponding to supplemented network. So, this take on to conclude that a great part of the total carbon source (supplementing amino acids and glucose) instead of being incorporated in biomass formation, it is deviated for other metabolic functions that do not lead to biomass synthesis. Probably the metabolic functions for which this carbon content is diverted are related to Cellular Respiration, as it is implicit in graphics presented earlier (Figure 3.12).

3.5. Comparing obtained results with bibliographic sources

In order to have an idea of how these results are credible, a comparison between the results obtained in this thesis and experimental results found in literature was made.

3.5.1. Non Supplemented Medium

With the purpose of make a comparison between this thesis results and experimental ones, it is presented below a table with the YNB medium formulation.

Table 3.7 Table with YNB medium components based in reference [49].

Main supplements of YNB medium	Amount of each YNB component (g/L)
Glucose	20
Phosphate	0.7
Ammonia	1.36
Sulfate	4.03

Comparing the table 3.7 and table 3.2 we can find some discrepancies. However it should be noted that in some clusters, the concentration of some compounds (for example Ammonia concentration in cluster 7) are very similar to the concentrations that can be found in table 3.7.

It was found a reference value for the resulting biomass amount (g/L) when YNB medium is used with a glucose concentration of 20 (g/L). Accordingly to *Kilonzo et al (2008)* **[62]**, the resulting biomass amount, using YNB medium is about 3.7 g/L. So, making a

comparison with table 3.2 where the resulting biomass ranges between 1.08 and 8.76 g/L, it is concluded that cluster 8 (Resulting Biomass – 3.37 g/L) is the closest one from this experimental value.

It should be noted that in table 3.7, the sulfate concentration is higher than all others supplements concentrations (except glucose), however this supplement is highly used for medium buffering, so this higher concentration does not reflect a higher uptake of this compound in *Saccharomyces cerevisiae* metabolism.

3.5.2. Supplemented Medium

It was also made a comparison between this work results and other experimental results including amino acids supplementation.

Table 3.8. Amino acids supplementation according to *Wittrup et al (1994)* [63].

	Supplementing Amino Acids Amounts	
Amino acids	**SD-SCAA (mg/L) [A]**	**SD-2xSCAA (mg/L) [B]**
Arginine	95	190
Aspartate	200	400
Glutamate	630	1260
Glycine	65	130
Histidine	70	140
Isoleucine	145	290
Leucine	200	400
Lysine	220	440
Methionine	54	108
Phenylalanine	100	200
Serine	0	0
Threonine	110	220
Tyrosine	26	52
Valine	190	380

A – Synthetic dextrose (D-glucose; 20 g/L) medium supplemented with casaminoacids, which consists in acid hydrolysed casein; B – The same components of SD-SCAA but 2 times more concentrated.

Firstly, by looking the previous table and comparing it with table 3.4 it is concluded that the resulting amounts in this work, for some cases, correspond to overrated values.

However, for the previous table (Table 3.8) it is not possible to make a relation between supplementing amounts and the resulting yields of Biomass and Carbon dioxide (CO_2).

Thus, in order to have a more concrete idea if the resulting yields for Biomass and CO2 obtained in this work are credible, a comparison between our results and the results obtained in *Gorgens et al* **[58]** was made.

Table 3.9 Amino acids supplementation, according with *Görgen et al (2005)* [58], and respective yields of metabolic products (Biomass Yield: Biomass [C-mol]/Glucose [C-mol]; The feed medium contains 10 g/L of Glucose [49].

Amino acid Supplementation	Total amino-N (mM)[A]	Biomass Yield
SD-SCAA	20	0.74
Arginine, Asparagine	40	0.66
Arg (99.4), Asn (99.8), Gly (95.1), Ala (82.7), Gln (99.1), Glut (99.1) [B]	20	0.68
Arginine (59,1), Glutamine (97,9)	20	0.60

A: Total content of Nitrogen that is present in amino acid supplementing mixture.
B: Percentage of the amino acid supply in feed utilized by *Saccharomyces cerevisiae*

Table 3.10 Total content of Nitrogen (total amino-N [mM] that is present in amino acid supplementing mixture in each formulation that was obtained in this work.

Formulations	Total amino-N (mM)	Formulations	Total amino-N (mM)
Cluster 1	193,04	**Cluster 15**	198,63
Cluster 2	649,69	**Cluster 18**	455,96
Cluster 3	695,02	**Cluster 20**	112,60
Cluster 4	198,12	**Cluster 21**	198,93
Cluster 5	144,49	**Cluster 22**	571,01
Cluster 6	311,39	**Cluster 23**	214,33
Cluster 8	101,77	**Cluster 24**	102,35
Cluster 10	954,12	**Cluster 25**	609,33
Cluster 11	450,47	**Cluster 26**	198,59
Cluster 12	196,02	**Cluster 27**	201,45
Cluster 13	96,15	**Cluster 30**	545,89
Cluster 14	200,46		

By making a comparative analysis between the table 3.9 and 3.10 it is confirmed that the total amino-N content that is determined in this work is overrated when compared with experimental values. However it should be noted that these experimental values correspond to a feed medium with 10 g/L of Glucose, and in this work it was considered that the feed medium has 20 g/L of Glucose. Thus, taking into account the proportion between the carbon source (Glucose) and supplementing amino acids, the discrepancy between experimental values and values obtained in this work decreases.

By the results found in *Görgens et al 2005* [58] it is observed that the most consumed amino acids are Arginine and Asparagine, however it is also observed that amino acids such as Glycine, Alanine, Glutamine and Glutamate are also very consumed, coming into agreement with this work results.

On the other hand, by comparing the Biomass Yields presented in table 3.9 with Biomass Yields presented in table 3.6 it is possible to find some close values, such as: Cluster 1 - 0.539 ; Cluster 4 - 0.588; Cluster 12 - 0.548 ; Cluster 14 - 0.560; Cluster 15 - 0.559; Cluster 23 - 0.596; Cluster 26 0.558- ; Cluster 27 -0.558 . The discrepancies detected between experimental and theoretical yields (obtained in this work) probably can be the result of some mistake in Biomass Synthesis chemical reaction.

Watson **[64]** had confirmed the existence of amino acid pools (accumulation of amino acids), and in this article was concluded that histidine, isoleucine, methionine, phenylalanine, threonine, tryptophan, tyrosine and valine, during a nitrogen source starvation, were poorly accumulated; revealing in this way the low metabolic activity of these amino acids. In this article it was also concluded that the most accumulated amino acid in nitrogen starvation was Glutamate, and the Aspartate was associated to a high growth rates, confirming the results obtained in this work, where the most consumed amino acids were in fact Glutamate and Aspartate (see table 3.4).

4. Conclusion and Future Works

In this work were obtained some culture medium formulations for *Saccharomyces cerevisiae*, that were rationally designed, considering the different metabolic states that can occur at an intracellular level and by considering as well the metabolic footprint that is left in the extracellular environment, when the medium components are consumed. The combined analysis of the EFM's of the *Saccharomyces cerevisiae* metabolic network and the metabolic footprint, gave rise to the formulation of the mediums.

From the results we can conclude that amino acids supplementation increases the amount of resultant biomass. It can also be concluded that the most important amino acids for cell growth and cellular respiration are Glutamate and Aspartate. These results are supported by findings of *Watson 1979* **[64]** and *Görgens et al 2005* **[49].** Also in accordance with the findings of *Görgens et al 2005* **[49]** are the results that the amino acids glutamate, alanine, glycine are highly consumed. However, *Görgens et al 2005* **[49]** also concluded that Arginine and Asparagine are the most consumed amino acids, which was not found in this work. So, overall, the results obtained in this work seem to be credible when compared with bibliographic sources.

It was proven that the method used in this work enables the rational development of medium formulations, by taking the different metabolic states into account that can occur into a given organism. This method seems to be less expansive and resource consuming than traditional more empirical methods. However, there is a need to confirm the performance of the developed medium formulations experimentally. Therefore, as future work, it is proposed to test the medium formulations at first by the application of PLS regression methods to determine which metabolic pathways (EFM's) are being activated when the amino acids supplementation is made and to subsequently test the best performing medium formulations experimentally.

Also the utilization of a more complete metabolic network might better describe and predict the formulation medium and respective biomass formation, however a computer with more memory (RAM) would be required when calculating the EFMs.

5. References:

1. Castrillo, J., Oliver, S. Yeast Systems Biology Methods and Protocols. United Kingdom: Humana Press, 2011.

2. Wingreen, N. Botstein, D. Back to the future: education for systems-level biologists. *Nature Reviews Molecular Cell Biology*. **7,** 829–832 (2006).

3. Cain, C., Conte D., García-Ojeda, M., Daglio, L., Johnson, L., Lau, E., Manilay, J., Phillips, J., Rogers, N., Stolberg, S., Swift, H., Dawson, M. What Systems Biology Is (Not , Yet). 320**,** 1013-1014 (2008).

4. Sauer, U., Heinemann, M. & Zamboni, N. Getting Closer to the Whole Picture. *Science.* **316,** 550–551 (2007).

5. Knuf, C. & Nielsen, J. Aspergilli: systems biology and industrial applications. *Biotechnol. J.* 7**,** 1147–1155 (2012).

6. Wiback, S. J., Famili, I., Greenberg, H. J. & Palsson, B. Ø. Monte Carlo sampling can be used to determine the size and shape of the steady-state flux space. *J. Theor. Biol.* 228**,** 437–447 (2004).

7. Famili, I., Forster, J., Nielsen, J. & Palsson, B. O. *Saccharomyces cerevisiae* phenotypes can be predicted by using constraint-based analysis of a genome-scale reconstructed metabolic network. *Proc. Natl. Acad. Sci. U. S. A.* 100**,** 13134–13139 (2003).

8. Schuster, S., Fell, D. A. & Dandekar, T. A general definition of metabolic pathways useful for systematic organization and analysis of complex metabolic networks. *Nature Biotechnology* 18**,** 326-332 (2000).

9. Toya, Y., Kono, N., Arakawa, K. & Tomita, M. Metabolic flux analysis and visualization. *J. Proteome Res.* 10**,** 3313–3323 (2011).

10. Kauffman, K. J., Prakash, P. & Edwards, J. S. Advances in flux balance analysis. *Curr. Opin. Biotechnol.* **14,** 491–496 (2003).

11. Varma, A & Palsson, B. O. Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type Escherichia coli W3110. *Appl. Environ. Microbiol.* **60,** 3724–3731 (1994).

12. Park, J. H., Lee, S. Y., Kim, T. Y. & Kim, H. U. Application of systems biology for bioprocess development. *Trends Biotechnol.* **26,** 404–412 (2008).

13. Otero, J., Cimini, D., Patil, K., Poulsen, S., Olsson, L., Nielsen, J. Industrial systems biology of *Saccharomyces cerevisiae* enables novel succinic acid cell factory. *PLoS One* **8,** 1-10 (2013).

14. Nielsen, J., Larsson, C., van Maris, A. & Pronk, J. Metabolic engineering of yeast for production of fuels and chemicals. *Curr. Opin. Biotechnol.* **24,** 398–404 (2013).

15. Nielsen, J. Production of biopharmaceutical proteins by yeast: advances through metabolic engineering. *Bioengineered* **4,** 207–211 (2013).

16. Patil, K. R., Akesson, M. & Nielsen, J. Use of genome-scale microbial models for metabolic engineering. *Curr. Opin. Biotechnol.* **15,** 64–69 (2004).

17. Park, J. M., Kim, T. Y. & Lee, S. Y. Constraints-based genome-scale metabolic simulation for systems metabolic engineering. *Biotechnol. Adv.* **27,** 979–988 (2009).

18. Kim, I.-K., Roldão, A., Siewers, V. & Nielsen, J. A systems-level approach for metabolic engineering of yeast cell factories. *FEMS Yeast Res.* **12,** 228–248 (2012).

19. Ostergaard, S., Olsson, L. & Nielsen, J. Metabolic engineering of *Saccharomyces cerevisiae*. *Microbiol. Mol. Biol. Rev.* **64,** 34–50 (2000).

20. Mandenius, C.F. and A. Brundin, Bioprocess Optimization Using Design-of-Experiments Methodology. Biotechnology Progress. 24(6): p. 1191-1203 (2008).

21. Deshpande, R.R., C. Wittmann, and E. Heinzle, Microplates with integrated oxygen sensing for medium optimization in animal cell culture. Cytotechnology. 46(1): p. 1-8. (2004)

22. Chang, K.H. and P.W. Zandstra, Quantitative screening of embryonic stem cell differentiation: Endoderm formation as a model. Biotechnology and Bioengineering. 88(3): p. 287-298. (2004)

23. Ghosalkar, A., V. Sahai, and A. Srivastava, Optimization of chemically defined medium for recombinant Pichia pastoris for biomass production. Bioresource Technology. 99(16): p. 7906-7910 (2008).

24. Montgomery D.C., 2005, Design and Analysis of Experiments, John Wiley & Sons Inc, 6th edition.

25. Kell, D.B., et al., Metabolic footprinting and systems biology: The medium is the message. Nature Reviews Microbiology. 3(7): p. 557-565 (2005)

26. Hunter, D.J., Gene-environment interactions in human diseases. Nature Reviews Genetics. 6(4): p. 287-298 (2005).

27. Schuster, S. and H. Claus, On Elementary Flux Modes in Biochemical Reaction Systems at Steady State. Journal of Biological Systems. 2(2): p. 165-182 (1994).

28. Schilling, C.H. and B.O. Palsson, Assessment of the metabolic capabilities of Haemophilus influenzae Rd through a genome-scale pathway analysis. Journal of Theoretical Biology. 203(3): p. 249-283 (2000).

29. Stelling, J., et al., Metabolic network structure determines key aspects of functionality and regulation. Nature. 420(6912): p. 190-193 (2002).

30. Fukushima, A., Kusano, M., Redestig, H., Arita, M. & Saito, K. Integrated omics approaches in plant systems biology.*Chemical Biology*. **13**, 532-538 (2009).

31. Ellis, D. I., Dunn, W. B., Griffin, J. L., Allwood, J. W. & Ellis, E. D. Metabolic fingerprinting as a diagnostic tool. *Pharmacogenomics*. **8,** 1243–1266 (2007).

32. Mapelli, V., Olsson, L. & Nielsen, J. Metabolic footprinting in microbiology : methods and applications in functional genomics and biotechnology. *Trends in Biotechnol*. **26,** 490-497 (2008).

33. Chumnanpuen, P., Hansen, M. A. E., Smedsgaard, J. & Nielsen, J. Dynamic Metabolic Footprinting Reveals the Key Components of Metabolic Network in Yeast *Saccharomyces cerevisiae*. *Int. J. Genomics* **2014,** 1-14 (2014).

34. Fiehn, O. Combining genomics, metabolome analysis, and biochemical modelling to understand metabolic networks. *Comp. Funct. Genomics* **2,** 155–168 (2001).

35. Villas-Bôas, S. G., Rasmussen, S. & Lane, G. a. Metabolomics or metabolite profiles? *Trends Biotechnol.* **23,** 385–386 (2005).

36. Hollywood, K., Brison, D. R. & Goodacre, R. Metabolomics: current technologies and future trends. *Proteomics* **6,** 4716–4723 (2006).

37. Teixeira, A., Dias, J., Carinhas, N., Sousa, M., Clemente, J., Cunha, A., von Stosch, M., Alves, P., Carrondo M., Oliveira, R. Cell functional enviromics: unravelling the function of environmental factors. *BMC Syst. Biol.* **5,** 92 (2011).

38. Palsson, Bernhard Ø. Systems Biology Properties of Reconstructed Networks. United States of America: Cambridge University Press, 2006.

39. Llaneras, F. & Picó, J. Stoichiometric modelling of cell metabolism. *J. Biosci. Bioeng.* **105,** 1–11 (2008).

40. Poolman, M. G., Venakatesh, K. V, Pidcock, M. K. & Fell, D. A. A Method for the Determination of Flux in Elementary Modes , and its Application to Lactobacillus rhamnosus. *Biotechnology and Bioengineering.* 1–12 (2004). doi:10.1002/bit.20273

41. Reed, J. L. & Palsson, B. Ø. MINIREVIEW Thirteen Years of Building Constraint-Based In Silico Models of Escherichia coli. *Journal of Bacteriology*. **185,** 2692–2699 (2003).

42. Zanghellini, J., Ruckerbauer, D. E., Hanscho, M. & Jungreuthmayer, C. Elementary flux modes in a nutshell: properties, calculation and applications. *Biotechnol. J.* **8,** 1009–1016 (2013).

43. Hilgetag, S. S. C. & Fell, J. H. W. D. A. Reaction routes in biochemical reaction systems : Algebraic properties , validated calculation procedure and example from nucleotide metabolism. *Journal of Mathematical Biology*. **181,** 153–181 (2002).

44. Urbanczik, R. & Wagner, C. An improved algorithm for stoichiometric network analysis: theory and applications. *Bioinformatics* **21,** 1203–10 (2005).

45. Ballerstein, K., von Kamp, A., Klamt, S. & Haus, U.-U. Minimal cut sets in a metabolic network are elementary modes in a dual network. *Bioinformatics* **28,** 381–387 (2012).

46. Schuster, S., Hilgetag, C. On Elementary Flux Modes In Biochemical Reactions Systems at Steady State. *Journal biological systems*. 2, 165-182 (1994)

47. Pfeiffer, T., Sánchezćvaldenebro, I., Nuño, J. C., Montero, F. & Schuster, S. METATOOL: for studying metabolic networks. **15,** 251–257 (1999).

48. Ferreira, A. R. *et al.* Projection to latent pathways (PLP): a constrained projection to latent variables (PLS) method for elementary flux modes discrimination. *BMC Syst. Biol.* **5,** 181 (2011).

49. (2001), *Methods in Molecular Biology, Vol. 177, Two-Hybrid Systems: Methods and Protocols,* Humana Press Inc, Totowa, NJ, P. N. MacDonald.

50. Hahn-Hagerdal, B., Karhumaa, K., Larsson, C., Gorwa-Grauslund, M., Gorgens, J., van Zyl, W. **Role of cultivation media in the development of yeast strains for large scale industrial use.** *Microbial Cell Factories 2005*, 4:31.

51. van der Valk, J., Brunner, D., De Smet, K., Svenningsen, A., Honegger, P., Knudsen, L., Lindl, T., Noraberg, J., Price, A., Scarino, M., Gstraunthaler, G. **Optimization of chemically defined cell culture media – Replacing fetal bovine serum in mammalian in vitro methods**. *Toxicology in Vitro 2010*, 24:1053-1063.

52. Summers, M., Biggers, J. **Chemically defined media and the culture of mammalian preimplantation embryos: historical perspective and current issues**. *Human Reproduction Update 2003*, 9:557-582.

53. Zhang, J., Greasham R. **Chemically defined media for commercial fermentations**. *Appl Microbiol Biotechnol 1999*, 51: 407-421.

54. www.formedium.com.

55. www.mpbio.com.

56. Jain, D., Nielsen, J., Buckland, B. **Kinetics of efrotomycin synthesis in a synthetic fermentation medium**. *Bioprocess Engineering*. 7:257-263 (1992)

57. Vanrolleghem, P., Jong-Gubbels,P., van Gulik, W., Pronk, J., van Dijken, J., Heijen, S. Validation of a Metabolic Network for *Saccharomyces cerevisiae*. *Biotechnol. Prog* .**12,** 434–448 (1996).

58. Gorgëns, J., van Zyl, W., Knoetze, J., Hahn-Hägerdal, B. Amino acid supplementation improves heterologous protein production by *Saccharomyces cerevisiae* in defined medium. *Appl. Microbiol.Biotechnol*. 67, 684-691 (2005).

59. Terzer, M. & Stelling, J. Large-scale computation of elementary flux modes with bit pattern trees. *Bioinformatics* **24,** 2229–2235 (2008).

60. Lange, H. C. & Heijnen, J. J. Statistical reconciliation of the elemental and molecular biomass composition of *Saccharomyces cerevisiae*. *Biotechnol. Bioeng.* **75,** 334–344 (2001).

61. Chen, Z.., Hongjuan,L., Zhang, J., Liu, D. Elementary Mode Analysis for the Rational Design of Efficient Succinate Conversion from Glycerol by Escherichia coli. *Journal of Biomedicine and Biotechnology*. 2010, 1-14 (2010).

62. Kilonzo, P., Margaritis, A., Bergougnou, M. Effect of medium composition on Glucoamylase production during batch fermentation of recombinant *Saccharomyces cerevisiae*. *J. Inst. Brew.* 114, 83-96 (2008).

63. Witturp, K., Benig, V. Optimization of Amino acid supplements for heterologous protein secretion in *Saccharomyces cerevisiae*. *Biotechnology techniques.* 8, 161-166 (1994).

64. Watson, T. Amino-acid Pool Composition of *Saccharomyces cerevisiae* as a Function of Growth Rate and Amino-acid Nitrogen Source. *Journal of General Microbiology*. 96, 263-268 (1976).

6. Appendix

Table 6.1. Table with metabolic reactions considered in this work, and respective metabolic pathways [48].

Glycolysis and TCA cycle	
'R1'	'Gluc + ATP --> Gluc6P + ADP'
'R2'	'Gluc6P <--> Fruc6P'
'R3'	'Fruc6P + ATP --> 2 GAP + ADP'
'R4'	'2 GAP + H2O --> Fruc6P + Pi'
'R5'	'GAP + NADH + H2O --> GOH + Pi + NAD'
'R6'	'GAP + Pi + ADP + NAD <--> G3P + ATP + NADH'
'R7'	'G3P <--> PEP + H2O'
'R8'	'PEP + ADP --> Pyr + ATP'
'R9'	'Pyr + NAD + CoA --> AcCoA + NADH + CO2'
'R10'	'Pyr + ATP + CO2 + H2O --> Oax + Pi + ADP'
'R11'	'Acet + NAD + H2O <--> AC + NADH'
'R12'	'Oax + AcCoA + H2O --> Isocit + CoA'
'R13'	'Oax + ATP --> PEP + CO2 + ADP'
'R14'	'Isocit + NAD --> AKG + NADH + CO2'
'R15'	'AKG + CoA + NAD --> SucCoA + NADH + CO2'
'R16'	'SucCoA + Pi + ADP --> Suc + ATP + CoA'
'R17'	'Suc + FAD <--> Fum + FADH2'
'R18'	'Fum + H2O <--> Mal'
'R19'	'Mal + NAD --> Oax + NADH'
Phosphate Pentose Pathway	
'R20'	'Gluc6P + 2 NADP + H2O --> Ribu5P + 2 NADPH + CO2'
'R21'	'Ribu5P <--> Rib5P'
'R22'	'Ribu5P <--> Xyl5P'
'R23'	'Rib5P + Xyl5P <--> Sed7P + GAP'
'R24'	'Sed7P + GAP <--> Fruc6P + E4P'
'R25'	'Xyl5P + E4P <--> Fruc6P + GAP'
Glyoxilate Shunt	
'R26'	'Isocit <--> Glyo + Suc'
'R27'	'Glyo + AcCoA + H2O --> Mal + CoA'
Oxidative Phosphorilation	
'R28'	'2 NADH + O2 + 2.18 ADP + 2.18 Pi --> 2.18 ATP + 2 NAD + 4.18 H2O'
'R29'	'2 NADH + O2 --> 2 NAD + 2 H2O'
'R30'	'2 FADH2 + O2 + 2.18 Pi + 2.18 ADP --> 2.18 ATP + 4.18 H2O + 2 FAD'
'R31'	'2 FADH2 + O2 + 1.308 ADP + 1.308 Pi --> 1.308 ATP + 2 FAD + 3.308 H2O'
'R32'	'2 NADPH + O2 + 2.18 ADP + 2.18 Pi --> 2 NADP + 2.18 ATP + 4.18 H2O'

Other carbon sources than Glucose	
'R33'	'EtOH + NAD <--> Acet + NADH'
'R34'	'AC + 2 ATP + CoA + H2O --> AcCoA + 2 Pi + 2 ADP'
Transference of C1 compuns	
'R35'	'THF + ATP + NADH + CO2 <--> FTHF + Pi + ADP + NAD'
'R36'	'THF + CO2 + 3 NADH <--> MYTHF + 3 NAD + 2 H2O'
'R37'	'THF + CO2 + 2 NADH <--> METHF + 2 NAD + 2 H2O'
ATPase	
'R38'	'ATP + H2O --> Pi + ADP'
Amino acids Synthesis	
'R39'	'AKG + NH4 + NADPH <--> Glut + NADP + H2O'
'R40'	'Glut + NH4 + ATP --> Glum + Pi + ADP'
'R41'	'1 Glum + 1 H2O --> 1 Glut + 1 NH4'
'R42'	'Glut + ATP + 2 NADPH --> Pro + Pi + 2 NADP + H2O + ADP'
'R43'	'ATP + NH4 + CO2 --> CARP + ADP'
'R44'	'2 Glut + AcCoA + 4 ATP + NADPH + CARP + Asp + 3 H2O --> Arg + AKG + AC + Fum + 5 Pi + NADP + CoA + 4 ADP'
'R45'	'1 Arg + 1 AKG + 1 H2O --> 2 Glut + 1 NADPH'
'R46'	'2 Glut + AcCoA + 3 ATP + 2 NADPH + 2 NAD + 3 H2O --> Lys + CoA + AKG + CO2 + 3 Pi + 3 ADP + 2 NADP + 2 NADH'
'R47'	'G3P + Glut + NAD + H2O --> Ser + AKG + Pi + NADH'
'R48'	'Ser + THF <--> Gly + METHF + H2O'
'R49'	'1 Ser --> 1 Pyr + 1 NH4'
'R50'	'Ser + AcCoA + SO4 + 4 NADPH + ATP --> Cys + AC + CoA + 4 NADP + Pi + ADP + 3 H2O'
'R51'	'Oax + Glut <--> Asp + AKG'
'R52'	'Asp + NH4 + 2 ATP + H2O --> ASN + 2 Pi + 2 ADP'
'R53'	'1 ASN + 1 H2O --> 1 Asp + 1 NH4'
'R54'	'Asp + ATP + 2 NADPH --> HOM + Pi + 2 NADP + ADP'
'R55'	'HOM + ATP + H2O --> THR + Pi + ADP'
'R56'	'HOM + SucCoA + Cys + MYTHF + ATP + 2 H2O --> Met + CoA + Suc + Pyr + NH4 + ADP + Pi + THF'
'R57'	'THR + Pyr + NADPH + Glut --> Ileu + NH4 + NADP + H2O + CO2 + AKG'
'R58'	'Pyr + Glut <--> Ala + AKG'
'R59'	'2 Pyr + NADPH --> AKI + CO2 + NADP + H2O'
'R60'	'AKI + Glut <--> Val + AKG'
'R61'	'AKI + AcCoA + Glut + ATP + NAD + 2 H2O --> Leu + AKG + CoA + CO2 + Pi + NADH + ADP'
'R62'	'2 PEP + E4P + NADPH + ATP --> CHO + ADP + 4 Pi + NADP'
'R63'	'CHO + Glut --> Phen + AKG + CO2'
'R64'	'CHO + Glut + NAD --> Tyr + AKG + CO2 + NADH'
'R65'	'CHO + Glum + PRPP + Ser --> Tryp + 2 Pi + CO2 + GAP + Glut + Pyr + H2O'
'R66'	'Ribu5P + 2 ATP --> PRPP + 2 ADP'
'R67'	'PRPP + 3 ATP + 3 H2O + NH4 + Glum + NADPH + 2 NAD + CO2 --> His + 6 Pi + 2 NADH + NADP + 3 ADP + AKG'

Amino acids Polimerization	
'R68'	'820 Glut + 285 Glum + 448 Pro + 437 Arg + 776 Lys + 502 Ser + 787 Gly + 19 Cys + 806 Asp + 277 ASN + 518 THR + 138 Met + 524 Ileu + 1246 Ala + 719 Val + 803 Leu + 364 Phen + 277 Tyr + 76 Tryp + 179 His <--> 10000 AM'
'R69'	'10000 AM + 40000 ATP + 30000 H2O --> 48248 Protein + 40000 Pi + 40000 ADP'
Nucleotide Synthesis	
'R70'	'PRPP + 2 Glum + Gly + 4 ATP + Asp + 2 FTHF + CO2 + 2 H2O --> IMP + 4 ADP + 6 Pi + 2 Glut + 2 THF + Fum'
'R71'	'IMP + Asp + ATP --> Pi + Fum + ADP'
'R72'	'IMP + 2 ATP + Glum + NAD + 3 H2O --> GMP + 2 ADP + 2 Pi + Glut + NADH'
'R73'	'Glum + PRPP + 2 ATP + Asp + NAD + 2 H2O --> UMP + 2 ADP + 4 Pi + Glut + NADH'
'R74'	'UMP + 2 ATP <--> UTP + 2 ADP'
'R75'	'UTP + Glum + ATP + H2O --> CTP + ADP + Pi + Glut'
'R76'	'CTP + 2 ADP <--> CMP + 2 ATP'
RNA Synthesis	
'R77'	'2330 GMP + 3060 UMP + 2279 CMP + 32279 ATP + 22279 H2O --> 32279 ADP + 94660 RNA + 32279 Pi'
Synthesis of Fatty Acids	
'R78'	'8 AcCoA + 15 ATP + 13 NADPH + 9 H2O --> Pal + 13 NADP + 15 Pi + 8 CoA + 15 ADP'
'R79'	'9 AcCoA + 17 ATP + 15 NADPH + 10 H2O --> OL + 15 NADP + 17 Pi + 9 CoA + 17 ADP'
Synthesis of Polyssacharides	
'R80'	'Gluc6P + ATP + H2O --> 6 PSACCH + 2 Pi + ADP'
Biomass Formation	
'R81'	'47003 Protein + 35376 PSACCH + 5234 RNA + 344 Pal + 344 OL + 226 GOH + 162100 ATP --> 100000 Biomass + 162100 Pi + 162100 ADP'
Exchange Reactions/Boundary Reactions	
'R82'	'--> O2'
'R83'	'CO2 -->'
'R84'	'<--> H2O'
'R85'	'Biomass -->'
'R86'	'--> Gluc'
'R87'	'--> Pi'
'R88'	'--> NH4'
'R89'	'--> SO4'
'R90'	'--> 1 Ala'
'R91'	'--> 1 Arg'
'R92'	'--> 1 ASN'
'R93'	'--> 1 Lys'
'R94'	'--> 1 Pro'
'R95'	'--> 1 Glut'
'R96'	'--> 1 Glum'

'R97'	'--> 1 Gly'
'R98'	'--> 1 Asp'

Table 6.2. Respective medium formulations of excluded clusters (g/L).

Medium Components	Cluster 7	Cluster 9	Cluster 16	Cluster 17	Cluster 19	Cluster 28	Cluster 29
Glucose	20,00	20,00	20,00	20,00	20,00	20,00	20,00
Glutamato	166,66	14096,55	2160,68	406,73	459,19	0,00	0,43
Glutamina	307,17	0,00	0,00	2,03	937,49	0,00	0,43
Prolina	0,27	0,35	0,49	0,69	0,29	0,47	0,40
Aspartato	14,65	0,00	6268,99	440,99	45,71	50437,88	12308,31
Arginina	0,34	0,50	0,69	0,97	0,36	0,70	0,59
Lisina	0,61	0,75	1,04	1,51	0,63	1,04	0,89
Glicina	0,00	0,00	1397,01	95,10	0,00	7110,20	3803,84
Asparagina	0,17	0,25	0,37	0,38	0,18	0,34	0,28
Alanina	17,92	2,66	30,33	236,85	44,73	1,02	11515,68
Biomassa	28,32	36,16	55,22	56,46	30,22	49,28	41,76

Table 6.3. Commercial supplementing formulations.

Nutrinional Supplements	CSM (mg/L)	HSM (mg/L)	BSM (mg/L)
Aspartate	**80**	**50**	**100**
Serine	**-**	**50**	**375**
Glutamate	**-**	**-**	**100**
Histidine	**20**	**50**	**20**
Arginine	**50**	**100**	**20**
Threonine	**100**	**100**	**200**
Proline	**-**	**50**	**-**
Cysteine	**-**	**50**	**-**
Tyrosine	**50**	**50**	**30**
Valine	**140**	**100**	**150**
Methionine	**20**	**50**	**20**
Lysine	**50**	**100**	**30**
Isoleucine	**50**	**50**	**30**
Leucine	**100**	**100**	**60**
Phenylalanine	**50**	**50**	**50**
Tryptophan	**50**	**100**	**40**
Adenine	**10**	**100**	**40**
Uracil	**20**	**100**	**20**

Table 6.4. Clustering values for Cellular Respiration in Supplemented metabolic network

	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7	Cluster 8	Cluster 9	Cluster 10
Glucose	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
Glutamate	-5,09E+02	-2,33E-01	-2,01E-01	0,00E+00	-6,22E+00	0,00E+00	-2,83E+01	0,00E+00	-1,07E+02	-2,77E-01
Glutamine	0,00E+00	-1,78E-01	-5,32E-01	0,00E+00	-7,42E+01	-4,79E+02	-2,50E-01	0,00E+00	0,00E+00	-1,36E+01
Proline	0,00E+00	0,00E+00	0,00E+00	0,00E+00	0,00E+00	0,00E+00	0,00E+00	0,00E+00	0,00E+00	0,00E+00
Aspartate	0,00E+00	-2,06E-01	-2,59E-01	-1,12E+05	-6,17E+01	-6,39E+02	-2,18E+01	-2,62E+04	3,55E-15	-4,18E+00
Arginine	0,00E+00	0,00E+00	0,00E+00	0,00E+00	0,00E+00	0,00E+00	0,00E+00	0,00E+00	0,00E+00	0,00E+00
Lisyne	0,00E+00	0,00E+00	0,00E+00	0,00E+00	0,00E+00	0,00E+00	0,00E+00	0,00E+00	0,00E+00	0,00E+00
Glycine	1,10E-14	-3,20E-01	-1,31E-01	0,00E+00	1,11E-16	0,00E+00	-3,97E+01	0,00E+00	1,11E-16	-3,51E-01
Asparagine	0,00E+00	-2,33E+00	-6,24E-01	0,00E+00	-2,86E-01	0,00E+00	-2,50E-01	0,00E+00	0,00E+00	-2,94E-02
Alanine	0,00E+00	-7,03E-01	-3,11E-01	0,00E+00	-1,69E+01	0,00E+00	-5,37E+01	-8,59E+04	-3,23E+00	-2,97E+00

Table 6.4. Clustering values for Cellular Respiration in Supplemented metabolic network

	Cluster 11	Cluster 12	Cluster 13	Cluster 14	Cluster 15	Cluster 16	Cluster 17	Cluster 18	Cluster 19	Cluster 20
Glucose	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
Glutamate	-1,84E+02	-1,19E-01	-1,74E+02	-6,57E+00	-1,31E-01	-1,21E-01	-4,87E-01	-6,71E-01	-3,31E-01	-2,70E+00
Glutamine	0,00E+00	-2,93E-01	0,00E+00	-3,00E-03	-1,02E-01	-1,79E+00	-5,37E-01	-1,18E-01	-4,16E-01	-1,47E-02
Proline	0,00E+00	0,00E+00	0,00E+00	0,00E+00	0,00E+00	0,00E+00	0,00E+00	0,00E+00	0,00E+00	0,00E+00
Aspartate	0,00E+00	-1,96E-01	3,55E-15	-2,90E-01	-1,80E+00	-6,93E-01	-3,87E+00	-1,14E+00	-8,22E-01	-1,04E+00
Arginine	0,00E+00	0,00E+00	0,00E+00	0,00E+00	0,00E+00	0,00E+00	0,00E+00	0,00E+00	0,00E+00	0,00E+00
Lisyne	0,00E+00	0,00E+00	0,00E+00	0,00E+00	0,00E+00	0,00E+00	0,00E+00	0,00E+00	0,00E+00	0,00E+00
Glycine	-3,35E+02	-1,87E-01	-1,29E+01	-6,05E-01	-3,47E-01	-2,22E-01	-4,92E+00	-4,06E-01	-1,34E-01	-4,42E-01
Asparagine	0,00E+00	-1,29E+00	0,00E+00	-2,12E-03	-3,15E-03	-1,36E-02	-1,21E-02	-5,06E-03	-2,35E-03	-3,09E-03
Alanine	-3,91E+01	-4,09E-01	3,55E-15	-8,09E-02	-2,20E+00	-3,34E-01	-3,58E+00	-8,65E-01	-4,28E-01	-2,29E-01

Table 6.5. Clustering values for Biomass Production in Supplemented metabolic network

	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7	Cluster 8	Cluster 9	Cluster 10
Glucose	-1,00E+00	-1,00E+00	-1,00E+00	-1,00E+00	-1,00E+00	-1,00E+00	-1,00E+00	-1,00E+00	-1,00E+00	-1,00E+00
Glutamate	-1,50E-01	-3,45E+00	-3,59E+00	-1,22E-01	-9,89E-02	-1,04E-01	-1,02E+01	-1,07E-01	-8,63E+02	-4,16E+00
Glutamine	-2,71E-01	-4,67E-02	-3,62E-02	-2,41E-01	-1,82E-01	-1,08E-01	-1,89E+01	-6,14E-02	1,11E-16	-4,43E-02
Proline	-2,04E-06	-3,09E-02	-3,08E-05	-1,53E-02	-1,03E-02	-4,28E-03	-2,14E-02	-6,73E-03	-2,72E-02	-3,25E-05
Aspartate	-3,93E-01	-1,24E+00	-1,31E+00	-4,13E-01	-2,26E-01	-9,87E-02	-9,92E-01	-2,08E-01	4,00E-15	-2,26E+00
Arginine	-3,84E-06	-3,69E-05	-3,01E-02	-1,48E-02	-9,90E-03	-4,11E-03	-1,75E-02	-1,87E-03	-2,57E-02	-3,51E-05
Lisyne	-8,15E-06	-1,01E-04	-1,36E-04	-2,65E-02	-1,79E-02	-7,23E-03	-3,74E-02	-1,16E-02	-4,62E-02	-1,04E-04
Glycine	-2,38E-01	-5,51E-01	-6,25E-01	-2,03E-01	-2,81E-01	-1,49E-01	-4,83E-15	-2,82E-01	2,39E-15	-9,76E-01
Asparagine	-2,26E-02	-1,20E-02	-1,19E-02	-2,01E-02	-3,16E-02	-9,55E-01	-1,17E-02	-1,64E-02	-1,68E-02	-1,30E-02
Alanine	-3,69E-01	-4,67E-01	-5,20E-01	-3,98E-01	-1,83E-01	-2,91E-01	-1,81E+00	-1,27E-01	-2,69E-01	-1,09E+00
Biomass	3,24E+00	7,09E+00	7,08E+00	3,53E+00	2,39E+00	2,08E+00	9,75E+00	1,55E+00	1,24E+01	7,41E+00

Table 6.5. (Continued)

	Cluster11	Cluster12	Cluster13	Cluster14	Cluster15	Cluster16	Cluster17	Cluster18	Cluster19	Cluster20
Glucose	-2,51E+00	-1,11E-01	-1,15E-01	-1,41E-01	-1,07E-01	-1,32E+02	-2,49E+01	-2,47E+00	-2,81E+01	-1,10E-01
Glutamate	-7,15E-02	-2,58E-01	-6,07E-02	-2,62E-01	-2,85E-01	1,11E-16	-1,25E-01	-1,24E-01	-5,78E+01	-1,04E-01
Glutamine	-2,70E-02	-1,44E-02	-7,24E-03	-2,38E-06	-1,46E-02	-3,86E-02	-5,36E-02	-2,62E-02	-2,27E-02	-1,01E-06
Proline	-6,55E-01	-3,94E-01	-1,98E-01	-4,07E-01	-3,84E-01	-4,24E+02	-2,98E+01	-6,59E-01	-3,09E+00	-2,12E-01
Aspartate	-2,63E-02	-1,38E-02	-3,59E-03	-1,41E-02	-1,01E-05	-3,59E-02	-5,00E-02	-4,40E-05	-1,85E-02	-7,46E-03
Arginine	-1,42E-04	-5,06E-05	-2,35E-05	-1,01E-05	-2,53E-02	-6,41E-02	-9,29E-02	-4,73E-02	-3,91E-02	-4,49E-06
Lisyne	-3,53E-01	-2,32E-01	-2,63E-01	-2,28E-01	-2,35E-01	-1,68E+02	-1,14E+01	-3,43E-01	-2,94E-15	-2,80E-01
Glycine	-1,10E-02	-2,61E-02	-1,58E-02	-2,41E-02	-2,53E-02	-2,55E-02	-2,62E-02	-1,09E-02	-1,24E-02	-1,90E-02
Asparagine	-2,39E-01	-3,90E-01	-1,15E-01	-4,00E-01	-3,78E-01	-3,07E+00	-2,40E+01	-2,40E-01	-4,52E+00	-1,36E-01
Alanine	6,18E+00	3,29E+00	1,66E+00	3,36E+00	3,35E+00	1,90E+01	1,94E+01	6,27E+00	1,04E+01	1,79E+00
Biomass	1,08E+01	5,33E+00	5,54E+00	5,39E+00	5,35E+00	1,24E+03	1,50E+02	1,08E+01	1,52E+02	5,51E+00

Table 6.5 (Continued)

	Cluster21	Cluster22	Cluster23	Cluster24	Cluster25	Cluster26	Cluster27	Cluster28	Cluster29	Cluster30
Glucose	-1,00E+00	-1,00E+00	-1,00E+00	-1,00E+00	-1,00E+00	-1,00E+00	-1,00E+00	-1,00E+00	-1,00E+00	-1,00E+00
Glutamate	-2,14E-01	-3,07E+00	-2,80E-01	-1,09E-01	-1,08E-01	-1,33E-01	-1,37E-01	1,11E-16	-2,64E-02	-8,23E-02
Glutamine	-3,93E-02	-4,32E-02	-2,72E-01	-7,43E-02	-2,25E+00	-2,52E-01	-2,98E-01	1,11E-16	-2,64E-02	-2,34E-02
Proline	-8,86E-04	-1,43E-02	-1,80E-05	-1,09E-05	-9,78E-04	-1,54E-05	-1,46E-02	-3,70E-02	-3,15E-02	-1,28E-03
Aspartate	-5,63E-01	-9,81E-01	-4,04E-01	-2,03E-01	-1,21E-01	-3,88E-01	-3,95E-01	-3,41E+03	-8,33E+02	-4,79E-02
Arginine	-8,04E-04	-2,67E-02	-1,48E-05	-3,62E-03	-7,77E-04	-1,41E-02	-4,42E-05	-3,61E-02	-3,06E-02	-1,24E-03
Lisyne	-1,69E-03	-5,12E-02	-2,70E-02	-1,34E-02	-1,67E-03	-2,53E-02	-7,86E-05	-6,41E-02	-5,48E-02	-2,22E-03
Glycine	-5,16E-01	-4,50E-01	-2,35E-01	-2,63E-01	-6,58E-01	-2,31E-01	-2,38E-01	-8,53E+02	-4,56E+02	-7,13E-01
Asparagine	-1,11E-02	-1,05E-02	-2,31E-02	-1,72E-02	-3,17E-03	-2,55E-02	-2,46E-02	-2,29E-02	-1,94E-02	-1,96E+00
Alanine	-3,91E-01	-3,13E-01	-3,66E-01	-1,23E-01	-9,41E-02	-3,75E-01	-3,84E-01	-1,03E-01	-1,16E+03	-9,78E-02
Biomass	7,68E-01	6,79E+00	3,58E+00	1,77E+00	5,76E-01	3,35E+00	3,35E+00	1,70E+01	1,44E+01	5,94E-01

Table 6.6. Clustering values for Biomass Production in non-Supplemented metabolic network

	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7	Cluster 8	Cluster 9	Cluster 10
Glucose	-1,00E+00	-1,00E+00	-1,00E+00	-1,00E+00	-1,00E+00	-1,00E+00	-1,00E+00	-1,00E+00	-1,00E+00	-1,00E+00
Fosfato	-6,89E-03	-1,58E-03	-3,72E-03	-1,01E-02	-1,18E-02	-8,16E-03	-1,32E-01	-4,92E-03	-3,44E-01	-1,28E-02
Amonia	-2,29E-01	-5,26E-02	-1,24E-01	-3,34E-01	-3,91E-01	-2,71E-01	-9,11E-01	-1,64E-01	-1,81E+00	-4,25E-01
Sulfato	-2,49E-03	-5,70E-04	-1,34E-03	-3,63E-03	-4,24E-03	-2,94E-03	-3,18E-03	-1,77E-03	-1,19E-03	-4,61E-03
Biomassa	1,63E+00	3,73E-01	8,77E-01	2,37E+00	2,77E+00	1,92E+00	2,08E+00	1,16E+00	7,75E-01	3,01E+00