Rodolfo Miguel Azevedo Ribeiro Marques

M. Sc. Engenharia Química e Bioquímica

Modeling and Optimization of Extracellular Polysaccharides Production by *Enterobacter* A47

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

Orientador: Rui Manuel Freitas Oliveira,
Professor Associado com Agregação,
FCT/UNL

Co-orientadora: Maria da Ascensão Miranda Reis,
Professora Catedrática, FCT/UNL

Júri:

Presidente:	Prof. Doutora Maria da Ascensão Miranda Reis
Arguentes:	Prof. Doutor Eugénio Manuel de Faria Campos Ferreira
	Prof. Doutora Petia Georgieva
Vogal:	Doutora Ana Maria Palma Teixeira

FACULDADE DE CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE NOVA DE LISBOA

January, 2016

Modeling and Optimization of Extracellular Polysaccharides Production by *Enterobacter* A47

Dedicado aos meus pais

Acknowledgements

Completing this thesis is a great personal and professional achievement. What I learned during these years will be part of my legacy for life.

I would like to first and foremost thank my supervisors Prof. Rui Oliveira and my co-supervisor Prof. Maria Ascensão Reis for the opportunity I was given to work on such a stimulating subject. I acknowledge in particular Prof. Rui Oliveira for the advice and support throughout my academic path. I have always felt inspired by him and am forever grateful for having me in the great SBEgroup.

A very, very special word of acknowledgment to my colleague Dr. Mortiz von Stosch for his guidance and specialized insights on Hybrid modeling and pretty much anything mathematical. I am deeply indebted for his insights, support and discussions. I am grateful to Dr. Filomena Freitas for being always receptive to discuss any matter concerning my work, and Dr. Cristiana Torres for the fruitful collaborations.

To Abel Folch-Fortuny I would like to thank his important contributions and collaboration.

I would like to address everyone in the SBE group for all the companionship and making this group great.To Dr. João Dias my special thanks for being a good friend and for his readiness to help; to Filipe Ataíde I owe my definitive conversion to Linux and other geeky tweaks on my cellphone. Thank you pal, you made a geek geekier. My praise to Rui Portela, the biologist in the middle of engineers I thank for being a great colleague, for the clever discussions and for reminding us, engineers that life is more than just a bunch of differential equations. To “Suecka”

my biggest sympathy for delighting everyone with her northern accent, and also for her nice poster selections for the office; to Mauro and his noteworthy talent for making scones and huge custard pies. I also wouldn't forget the itinerant Inês Isídro for her comradeship. For sure the group wouldn't be complete without her.

I also would like to extend my compliments to the BPEG group. This was the group where I started my research life and where I learned so many important skills. I would like to acknowledge the good friends I made, with whom I had the pleasure to share the working bench, discussions and companionship.

To Ana Lanham and Tiago my biggest sympathy for being such wonderful human beings. Our lives crossed at an important moment and I want to thank you both for your support and friendship.

To my big all time friends: Jorge, Ricardo, Sandro and Fábio. Without you life wouldn't be as colorful and meaningful. We all shared a past together, dwelling in a present and heading to a future which is most uncertain. I know that friendship always was, and always will be able to glue together all these temporal patches, more or less spaced in our lives. I would also like to gratefully acknowledge the financial support for this thesis, awarded from Fundação para Ciência e Tecnologia, with the reference SFRH/BD/67033/2009

Quero reservar aqui um espaço escrito na minha língua materna, que será sempre mais perto do coração do que qualquer outra. Dedico aqui os meus agradecimentos ao meu pai Jorge e à minha mãe Lúcia por todo o apoio incondicional. À Mara um obrigado pela amizade de irmã que sempre me ajudou. É também em português que faço um agradecimento muito especial à minha querida Margarida, onde tantas vezes foi a luz ao fundo do túnel...

Abstract

Polysaccharides are gaining increasing attention as potential environmental friendly and sustainable building blocks in many fields of the (bio)chemical industry. The microbial production of polysaccharides is envisioned as a promising path, since higher biomass growth rates are possible and therefore higher productivities may be achieved compared to vegetable or animal polysaccharides sources. This Ph.D. thesis focuses on the modeling and optimization of a particular microbial polysaccharide, namely the production of extracellular polysaccharides (EPS) by the bacterial strain *Enterobacter* A47. *Enterobacter* A47 was found to be a metabolically versatile organism in terms of its adaptability to complex media, notably capable of achieving high growth rates in media containing glycerol byproduct from the biodiesel industry. However, the industrial implementation of this production process is still hampered due to a largely unoptimized process. Kinetic rates from the bioreactor operation are heavily dependent on operational parameters such as temperature, pH, stirring and aeration rate. The increase of culture broth viscosity is a common feature of this culture and has a major impact on the overall performance. This fact complicates the mathematical modeling of the process, limiting the possibility to understand, control and optimize productivity. In order to tackle this difficulty, data-driven mathematical methodologies such as Artificial Neural Networks can be employed to incorporate additional process data to complement the known mathematical description of the fermentation kinetics. In this Ph.D. thesis, we have adopted such an hybrid

modeling framework that enabled the incorporation of temperature, pH and viscosity effects on the fermentation kinetics in order to improve the dynamical modeling and optimization of the process. A model-based optimization method was implemented that enabled to design bioreactor optimal control strategies in the sense of EPS productivity maximization. It is also critical to understand EPS synthesis at the level of the bacterial metabolism, since the production of EPS is a tightly regulated process. Methods of pathway analysis provide a means to unravel the fundamental pathways and their controls in bioprocesses. In the present Ph.D. thesis, a novel methodology called Principal Elementary Mode Analysis (PEMA) was developed and implemented that enabled to identify which cellular fluxes are activated under different conditions of temperature and pH. It is shown that differences in these two parameters affect the chemical composition of EPS, hence they are critical for the regulation of the product synthesis. In future studies, the knowledge provided by PEMA could foster the development of metabolically meaningful control strategies that target the EPS sugar content and oder product quality parameters.

Keywords: Exopolysaccharides, Enterobacter A47, Hybrid semi-parametric modeling, Model-based optimization

Resumo

Crescente atenção tem sido dada à utilização de polissacáridos como percursores químicos sustentáveis e ecológicos em vários sectores da indústria bio(química). A produção microbiana de polissacáridos é vista como uma tecnologia promissora, possibilitando elevadas taxas de crescimento de biomassa e, desta forma, maiores produtividades comparativamente a fontes de polissacáridos animal e vegetal. A presente tese de doutoramento foca-se na modelação e optimização de um processo de produção de polissacáridos microbianos extracelulares (exopolissacáridos), sintetizados pela bactéria *Enterobacter* A47. Este organismo é dotado de uma grande versatilidade metabólica em termos de adaptação a meios de crescimento complexos, tendo em particular sido verificado uma elevada taxa de crescimento de biomassa em subprodutos ricos em glicerol, provenientes da indústria do biodiesel. No entanto, as condições de operação deste processo ainda se encontram por optimizar, razão que limita a sua implementação industrial. As velocidades de reacção envolvidas na operação em bioreactor são fortemente dependentes de parâmetros operacionais tais como a temperatura, pH, taxas de agitação e arejamento. O aumento da viscosidade do meio de cultura é característico neste tipo de culturas e representa um obstáculo na performance global do processo. Este facto complica o tratamento matemático de modelação, limitando o grau de conhecimento e controlo do processo, bem como a optimização da produtividade. De forma a contornar esta dificuldade, algumas metodologias matemáticas tais como Redes Neuronais Artificiais permitem o emprego

de informação adicional acerca do processo nas leis cinéticas que descrevem a fermentação. No presente trabalho, foi adoptada uma estratégia da modelação híbrida que permitiu incorporar dados de temperatura, pH e viscosidade nas descrições cinéticas de reacção, melhorando o desempenho de modelação dinâmica e optimização do processo. O modelo híbrido desenvolvido foi usado para desenhar várias metodologias de controlo óptimo destinadas à maximização da produtividade e rendimento de EPS. Dado que os processos celulares ligados à produção de exopolssacáridos são altamente regulados, torna-se crítico o estudo e conhecimento do processo ao nível do metabolismo bacteriano. Desta forma, métodos de análise de vias metabólicas constituem ferramentas importantes para descobrir vias metabólicas fundamentais e a sua ligação com bioprocessos. Nesta tese de doutoramento foi desenvolvida uma nova metodologia de análise de redes metabólicas denominada Análise de Modos Elementares Principais, que permitiu identificar um subconjunto de vias metabólicas activadas para um grupo de experiências de diferentes condições de temperatura e pH. Variações nestes dois parâmetros afectam a composição química do exopolissacárido produzido, sendo críticos para a regulação e especificação do produto final. Em estudos futuros, o conhecimento preveniente desta metodologia pode fomentar o desenvolvimento de estratégias de controlo metabólico direccionadas à composição química de exopolissacáridos e outros parâmetros de qualidade do produto.

Palavras-chave: Polissacáridos extracelulares, *Enterobacter* A47, modelação híbrida semi-paramétrica, optimização de bioprocessos

Contents

List of Figures

List of Tables

Nomenclature

Abbreviations

ADP Adenosine diphosphate

ANN Artificial Neural Network

ATP Adenosine triphosphate

CCRD Central Composite Rotatable Design

CoA Coenzyme A

CTP Cardinal temperature and pH model

DE Differential Evolution

DNA Deoxyribonucleic acid

EM Elementary Mode

EPS Exopolysaccharide

FADH Flavine adenine dinucleotide (reduced form)

MSE Mean squared error

NAD^+ Nicotinamide adenine dinucleotide (oxidized form)

NADH Nicotinamide adenine dinucleotide (reduced form)

ODE Ordinary Differential Equation

ORF Open Reading Frame

PC Principal Component

PCA Principal Component Analysis

PDE Partial Differential Equation

PEMA Principal Elementary Mode Analysis

PPP Pentose-phosphate pathway

RSM Response Surface Methodology

SSE Sum of squared errors

TCA Tricarboxylic-acid

UDP-Gal Uridine diphosphate galactose

UDP-Glc Uridine diphosphate glucose

YA Yield Analysis

Greek letters

$\boldsymbol{\lambda}$ Weights vector for the elementary modes in $\mathbf{Z_y}$

$\boldsymbol{\Lambda}$ Matrix of weight vectors

μ Specific biomass growth rate (h^{-1})

Variables and parameters

$\mathbf{C_{in}}$ Vector of concentrations in the bioreactor inlet (kg m^{-3})

$\mathbf{C}$ Vector of concentrations species (kg m^{-3})

$\mathbf{G}$ Residual matrix

$\mathbf{N}$	Stoichiometric matrix
$\mathbf{Q_C}$	Vector of gas to liquid volumetric trasnfer rates (kg $m^{-3}\,h^{-1}$)
$\mathbf{v}$	Specific reaction rate vector (kg [kg CDW]$^{-1}\,h^{-1}$)
$\mathbf{Z_{pem}}$	Matrix of Principal Elementary Modes
$\mathbf{Z_y}$	Normalized matrix of elementary modes
$\mathbf{Z}$	Matrix of Elementary Modes
D	Dilution rate (h^{-1})
F	Volumetric feed (kg h^{-1})
K_N	Monod half saturation constant for ammonia (kg m^{-3})
K_P	Half specific EPS productivity rate
K_S	Monod half saturation constant for glycerol (kg m^{-3})
m_S	Maintenance coefficient for glycerol (h^{-1})
N	Ammonia concentration (kg m^{-3}))
P	EPS concentration (kg m^{-3})
P_i	Concentration of product i (kg m^{-3})
r_i^{ANN}	ith output from the ANN
S	Glycerol concentration (kg m^{-3})
S_i	Concentration of substrate i (kg m^{-3})
V	Volume (m^3)
v_N	Specific ammonia consumption rate (kg [kg N]$^{-1}\,h^{-1}$)

v_P Specific EPS production rate (kg [kg P]$^{-1}$ h^{-1})

v_S Specific glycerol consumption rate (kg [kg S]$^{-1}$ h^{-1})

X Biomass concentration (kg m^{-3}

Y_{ij} Yield coefficient of product i from product j (kg of i per kg of j)

CHAPTER 1

State of the art

Natural polysaccharides have shown an increased interest in recent years as a viable alternative to many petroleum-derived products. This interest stems from the important physical and chemical properties related to polysaccharides, but also from a growing awareness of the negative impact of non-renewable resources, and an urge to explore new and sustainable products.

Several properties are attributed to polysaccharides, which may suit them well to many industrial applications. Therefore, polysaccharides can be used as thickeners, bioadhesives, stabilizers, probiotics and gelling agents in food and cosmetic industries, and also as emulsifiers, bioabsorbents and bioflocculants in the environmental sector (Freitas et al. 2011a; Kumar et al. 2007). Polysaccharides also exhibit specific properties, related to their biological activity. Health promoting effects in the treatment of inflammatory diseases and cancer can be attributed to some polysaccharides, due to a proven immunomodular activity in them. This characteristic extends the range of possible uses from polysaccharides to health and biotechnological applications (Ramberg et al. 2010).

Most of the polysaccharides available in the market are obtained from plant

sources (e.g. Guar gum, Arabic gum or pectins), algae (*e.g.* alginate, carrageenan or agar) and crustacean (*e.g.* chitin), with polysaccharides from microbial sources (e.g. xanthan gum, gellan, pullulan and bacterial alginate) representing only a small fraction (Freitas et al. 2009). Industrial production of polysaccharides from microbial sources is particularly interesting, since they are generally better suited for large scale production than plant or algae, demonstrating higher growth rates and being more amenable to manipulation of growing/production conditions. However, some bottlenecks still limit the implementation and scaling of industrial processes for polysaccharide production, such that despite the large numbers of bacterial polysaccharides available, relatively few of them have been commercially developed. A direct cause for increased production costs are related to the commonly used carbon sources for microbial fermentations, which include sugars such as glucose, sucrose and fructose. Moreover, some polysaccharide producing bacteria are pathogenic and product quality may be difficult to maintain (Freitas et al. 2009; Kumar et al. 2007; Sutherland 2001).In order to make many of the current microbial polysaccharides available in the market, new and renewable resources should be implemented as alternative carbon sources from, for example, agro-industrial byproducts. Also, modeling studies represent an important aspect for the process design, scale-up, optimization and control of the fermentation system.

1.1 Microbial Polysaccharides

Polysaccharides play a very important role in the cell biology. These macromolecules can function as a storage polymer that is formed whenever the cell has a surplus of carbohydrates available; be part of the cell's structure, forming the cell wall, and also responsible for immunogenic properties exhibited in prokaryotes (Madigan et al. 2003). Chemically, polysaccharides can be divided in homopolysaccharides or heteropolysaccharides (Lehninger et al. 2000). In the former case, a single type of

monosaccharide makes up the chemical composition of the polymer (*e.g.* glycogen and cellulose), whereas in the latter case, different repeating units occur in the polymer (*e.g.* xanthan gum, dextran, etc.). The formation of polysaccharides with such varied structures requires the involvement of different enzymes and proteins, which is reflected in the varied organizations of the biosynthesis gene clusters (Rehm 2010).

Depending on the location of polysaccharide synthesis, three subdivisions are possible: exopolysaccharides (EPS), intracellular polysaccharides and capsular polysaccharides. Generally, exopolysaccharides are assembled in small repetitive oligosaccharides and exported to an outer membrane site, where polymerization to a larger molecule occurs (Kumar et al. 2007). Capsular polysaccharides are also secreted to the extracellular media, but remain attached to the cell, often functioning as major surface antigens and virulence factors. Details of biosynthesis and assembly were extensively studied for *Escherichia coli* (Whitfield and Paiment 2003).

Intracellular polysaccharides are produced in the cytoplasm and are mainly associated with storage functions. The only intracellular polysaccharide found in bacteria with this function is glycogen, which is synthesized as a response to starvation conditions in the environment. This way the cell accumulates carbon and energy in the form of glucose subunits, as part of a survival strategy (Wilson et al. 2010).

Exopolysaccharides represent the most relevant class of microbial polysaccharides for commercial purposes, due to their ability to change physical and chemical properties of aqueous solutions (*e.g.* thickening, emulsifying and stabilizing capacity), allowing their application on several products and processes, such as food, pharmaceutical, cosmetic, paint and oil drilling sectors. Also, the capacity of forming biodegradable films enables their use in packaging, pharmaceuticals and other industrial applications.

1.1.1 Overview of Polysaccharide synthesis

With the exception of levans, alternans and dextrans, the majority of bacterial EPS involves the synthesis of repetitive polymeric subunits, which are exported to the extracellular medium, where the polysaccharide chain is assembled (Vanhooren and Vandamme 1998). These repetitive subunits are composed of several sugar residues linked together by glycosidic bonds, in a process occurring in the cytoplasm. One of two possible mechanisms known to take place in Gram-negative bacteria for EPS biosynthesis are the Wzx-Wzy-dependent pathway, in which the polymeric repetitive subunit is assembled at the inner face of the cytoplasmic membrane and polymerized at the periplasm. Another mechanism involves the ABC transporter-dependent system, characterized by a polymerization at the cytoplasmic face of the inner membrane (Cuthbertson et al. 2009; Sutherland 2001).

The enzymes involved in EPS biosynthesis can be divided in four main groups, depending on the cell location (Kumar et al. 2007). A first group of intracellular enzymes is transversal to many other metabolic pathways, and is composed by hexokinase and phosphoglucomutase. These enzymes play an important role in the central carbon metabolism, the former responsible to the phosphorylation of glucose to glucose-6-phosphate, while the latter is involved in the displacing of the phosphoric group from glucose-6-phosphate to glucose-1-phosphate. A second group of reactions, also believed to be intracellular, catalyzes the conversion of glucose-1-phosphate to uridine diphosphate glucose (UDP-Glc), a nucleoside diphosphate sugar which is key to polysaccharide synthesis. This activated-sugar serves itself as a building block to polysaccharide composition, as well as a precursor to the formation of other nucleoside sugars, for example, the conversion of UDP-Glc to UDP-Gal (uridine diphosphate galactose). These high-energy molecules, provide a source of monosaccharide residues for exopolysaccharide synthesis (Kumar et al. 2007; Rehm 2010).

The third group of enzymes is composed of glycosyl transferases and is located at the periplasmic membrane. This enzymes deal with the transfer of sugar nucleotides

to a repeating unit, attached to an isoprenoid alcohol. The EPS synthesis mechanism is completed by the existence of a fourth group of enzymes, located at the external side of the cell wall, taking part in the polysaccharide polymerization, assembling all the repetitive units transported by the isoprenoid alcohol. The EPS is therefore extruded from the cell surface, forming a loose slime (Kumar et al. 2007; Rehm 2010).

Despite the sugar monomers, non-sugar molecules may also be incorporated in EPS structure, typically acyl groups. Such groups have been identified in *Enterobacter* A47, namely acetyl, pyruvyl and succinyl substituents (Freitas et al. 2011b). The incorporation of these substituents involves a mechanism that requires activated forms of acetate, pyruvate and succinate. Previous studies in other organisms revealed that acetyl-CoA and phosphoenolpyruvate are two precursor molecules involved in the inclusion of acetyl and pyruvyl groups respectively (Kumar et al. 2007; Sutherland 1993). As for the presence succinyl groups, no current study exists for the elucidation of a mechanism for inclusion in EPS, but it is likely that succinyl-CoA is involved in such process (Sutherland 1993).

1.2 *Enterobacter* A47: an EPS producing strain

The *Enterobacter* genus belongs to a group of enteric Gram-Negative bacteria, representative of the *Enterobacteriaceae* family. This prokaryote branch is a relatively homogeneous phylogenetic group, characterized by bacteria with rod morphology, nonsporulating facultative aerobes with relatively simple nutritional requirements, capable of fermenting sugars to a variety of end products. Even though enteric bacteria have similar morphology and metabolic characteristics, two main groups are found, setting apart bacteria by the type of products generated in the anaerobic fermentation of glucose, designated by mixed-acid and 2,3-butanediol fermentation. Species from the *Enterobacter* genus fall in the 2,3-butanediol group, which also

Table 1.1: Overview of the most relevant bacterial EPS [adapted from Freitas et al. (2011a)]

EPS	**Monomers**	**Charge**	**Organism**	**Properties**	**Applications**	**References**
Alginate	• Glucuronic acid • Mannuronic acid • Acetate	Anionic	*Pseudomonas aeruginosa* *Azotobacter vinelandii*	• Hydocolloid • Gelling capacity • Film-forming	• Food hydrocolloid • Filmforming	Rehm (2010) Hay et al. (2009) Peña et al. (2008)
Cellulose	• Cellulose	Neutral	*Gluconacetobacter xylinus*	• High Crystallinity • Insolubility in most solvents • High tensile strength • Moldability	• Foods • Biomedical applications	Nguyen et al. (2008) Rehm (2010)
Curdlan	• Glucose	Neutral	*Alcaligenes faecalis*	• Gel-forming ability • Water insolubility • Edible/non-toxic • Biological activity	• Foods • Pharmaceutical Industry • Heavy metal industry • Concrete aditive	Oliveira et al. (2007) Matsushita (1990)
Dextran	• Glucose	Neutral	*Leuconostoc mesenteroides*	• Non-ionic • Good stability • Newtonian fluid behaviour	• Foods • Pharmaceutical industry • Chromatographic media	Rehm (2009)
Gellan	• Glucose • Rhamnose • Glucuronic acid • Acetate • Glycerate	Anionic	*Sphingomonas paucimobilis*	• Hydrocolloid • Gelling capacity • Thermoreversible gels	• Foods • Pet food • Pharmaceuticals • Research	Fialho et al. (2008) Bajaj et al. (2007)
Hyaluronan	• Glucoronic acid • Acetylglucosamine	Anionic	*Pseudomonas aeruginosa* *Pasteurella multocida*	• Biological activity • Highly hydrophilic • Biocompatible	• Medicine • Solid culture media	Rehm (2009) Bonde et al. (1957) DeAngelis et al. (1998)
Levan	• Fructose	Neutral	*Bacillus subtilis* *Zymomonas mobilis* *Halomonas* sp.	• Low viscosity • High water solubility • Biological activity • Adhesive strength • Film-forming capacity	• Food • Animal feed • Pharmaceuticals • Cosmetics • Industry	Rehm (2009) Oliveira et al. (2007)
Xanthan	• Glucose • Mannose • Glucuronic acid • Acetate • Pyruvate	Anionic	*Xanthomonas campestris*	• Hydrocolloid	• Foods • Petroleum industry • Pharmaceuticals • Cosmetics • Agriculture	Rehm (2009) García-Ochoa et al. (2000)
GalactoPol	• Galactose • Mannose • Rhamnose • Acetate • Succinate • Pyruvate	Anionic	*Pseudomonas oleovorans*	• Viscous shear-thinning solutions in aqueous media • Film-forming • Emulsifying capacity • Flocculating capacity	• Hidrocolloid • Coatings • Packages	Freitas et al. (2009) Freitas et al. (2010) Alves et al. (2010a)
FucoPol	• Fucose • Galactose • Glucose • Acetate • Pyruvate • Succinate	Anionic	*Enterobacter* A47	• Viscous shear-thinning solutions in aqueous media • Film-forming • Emulsifying capacity • Flocculating capacity • Biological activity	• Hydrocolloid • Source of fucose and fuco-oligosaccharides	Freitas et al. (2011b) Alves et al. (2010b)

show the ability to produce CO_2, H_2, ethanol and small amounts of organic acids (Madigan et al. 2003).

The synthesis of EPS by *Enterobacter* species has also been reported in a number of academic journals, as well as in some patents that describe processes for EPS production by specific strains. Some examples include *Enterobacter sp.* CNCM 1-2744 (Philbe 2002), which was shown to produce EPS composed of fucose, galactose, glucose and glucuronic acid monomers, present in a 2:2:1:1 ratio. *Enterobacter sp.* SSYL (KCTC 0687BP) produces an exopolysaccharide with a composition of 8-10% of the total sugar content, being glucuronic acid the main component (40-70%) (Yang 2002). Some strains of *Enterobacter sakazakii*, namely ATCC 53017, ATCC 29004 and ATCC 12868 produce an EPS in which fucose represent 13-22% of the total polysaccharide sugar weight. *Enterobacter amnigenus* produces a heteropolymer containing glucose, galactose, fucose, mannose, glucuronic acid and pyruvil groups (Cescutti et al. 2005).

More recently, the bacterial strain *Enterobacter* A47 has shown the capacity to produce an exopolysaccharide composed of glucose, galactose and fucose, with the occurrence of the acylated substituents acetyl, pyruvyl and succynil Figure 1.1 (Freitas et al. 2011b). It was also show that differences in environmental conditions such as temperature and pH, trigger the synthesis of EPS with different compositions, with the inclusion of additional sugar monomers like rhamnose, glucoronic acid and glucosamine, while maintaining the same acylated substituents (Torres et al. 2012).

Exopolysaccharides produced by *Enterobacter* A47 share some interesting properties with other commercially successful microbial polysaccharides. Like xanthan gum, this EPS has a high molecular weight and develops highly viscous aqueous solutions with shear-thinning characteristics (*i.e.* pseudoplastic behavior). Despite its rheological properties, it exhibits film-forming, emulsion forming and stabilizing capacity, which is suitable for several applications, such as a hydrocolloid for food industry, cosmetics, pharmaceuticals and also oil-industry (Freitas et al. 2011b).

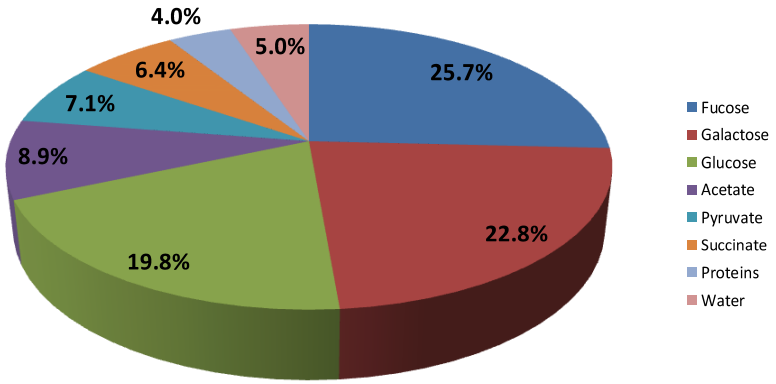

Figure 1.1: Composition of FucoPol produced by *Enterobacter* A47.

Moreover, the presence of fucose, a rare monosaccharide with important biological activity, adds value to this EPS, making it an interesting, marketable product.

1.3 Production of microbial EPS

Despite the large number of bacterial EPS available, only a few have been commercially developed, such as xanthan gum, dextran, gellan and curdlan. Typically, EPS production occurs in stirred-tank bioreactors capable of monitoring and controlling several process parameters. Although most of EPS chemical characteristics are determined by the genetics of the cell, they are largely influenced by media components and cultivation conditions. Temperature, pH, dissolved oxygen and agitation are some operational parameters critical to the control and optimization EPS production.

The synthesis is generally favored by the presence of carbon source in excess, with limitation of other nutrients, generally nitrogen, halting biomass growth and shifting the carbon source uptake to EPS synthesis (Freitas et al. 2011a). This strategy may be accomplished by a batch followed by a feeding rate that replenishes the bioreactor with fresh media for EPS conversion (Freitas et al. 2010). Aeration and stirring are also important aspects to take into account since most EPS producing cultures are aerobic. It has been observed that increased aeration rates have led to an optimal

EPS production (Lee et al. 2001; Rosalam and England 2006), whereas some cases have also demonstrated the opposite, with optimal production under microaerophilic conditions, such as the case of bacterial alginate (Freitas et al. 2011a).

EPS production is also sensitive to pH and temperature. The optimal temperature and pH conditions for biomass growth may differ, so that these fermentations are characterized by a different set of parameters for biomass growth and EPS production. It has been observed that in most cases, maximal EPS occurs at suboptimal growth temperatures (Cerning et al. 1992; Gancel and Novel 1994). This difference may be the result of increased enzyme activities of enzymes related to the synthesis of exopolysaccharide productions. For example, in the case of alginate, GDP-mannuronic acid is converted at an optimum rate below the optimum growth temperature, enhancing the polymer production (Kumar et al. 2007). In the case of pH, it has been found that most EPS are enhanced at neutral values, with some exceptions for acidic pH found on some meat starters and *Propionibacterium acidi-propionici* (Gorret et al. 2001; Kumar et al. 2007).

Improvement of EPS production can also be done at the level of metabolic engineering, either by manipulation of genes encoding enzymes, or by modifying regulatory pathways that affect gene expression. Biosynthetic pathways may either be controlled at the level of sugar nucleotide synthesis, assembly of EPS repeating unit and/or polymerization and export (Freitas et al. 2011a). The advantages reflect the increase of EPS productivities and yield, but also in the manipulation of EPS chemical composition, allowing the design of polymers of desired characteristics, paving the way for tailor-made biopolymers. Some applications of metabolic engineering were attempted with successful results. However, in many cases significant improvements have not been observed due to lack of knowledge of metabolic regulation mechanisms (Freitas et al. 2011a; Rehm 2010).

1.3.1 Utilization of alternative carbon sources for fermentation media

One of the main aspects impeding the economic viability of the industrial exploration of microbial exopolysaccharides is the high prices of carbon sources and it is estimated that culture media represent approximately 30% of the total fermentation cost, due to the use of expensive materials (Öner 2013). Most substrates used for exopolysaccharide production include sucrose, glucose, lactose, maltose, mannitol, sorbitol, whey, starch, sugar concentrates, methanol and C9 to C12 alkanes (Kumar et al. 2007). These materials of choice are selected for defined culture compositions because of a greater control over the final product, with minimum batch-to-batch variation and free of impurities that would interfere with their chemical and biological categorization. Nevertheless, the search for alternative renewable and cheaper materials is a key aspect to industrial implementation success (Sutherland 2001). For this reason, a wide range of agro-industrial by-products are being considered for industrial applications such as syrups, molasses, sugar beet pulp, olive mill wastewater, cheese whey, pomace and lignocellulosic biomass (Öner 2013). Glycerol by-product from biodiesel industry is another alternative which have been used for exopolysaccharide production (Freitas et al. 2009, 2011b).

The choice of the adequate carbon source is determined by the metabolic needs of the microorganism for EPS production. If the culture is adapted to glucose and sucrose, syrups and molasses may be adequate substitutes. In other cases, some byproducts require a pretreatment in order to extract important precursors for polysaccharide synthesis (Öner 2013). Despite the attractiveness of such resources, some problems may arise from their utilization. The presence of contaminants and different nutrient compositions may have an unpredictable effect in the cell's metabolic pathways (Freitas et al. 2011a). For this reason, different polymers and/or unwanted products might be synthesized by bacteria, lowering the product yield or be potential cell growth inhibitors. Therefore, to produce high quality polymers for

specific applications, the utilization of renewable feedstock may not be adequate, since there's a high risk of impurity carryover to the final product (Freitas et al. 2011a).

1.4 Modeling and optimization strategies employed in exopolysaccharide production

1.4.1 Basic modeling principles

The development of kinetic models is a crucial step to understand, control and optimize fermentation processes. Many factors can have an impact on bioreaction systems, such as the strain used for conversion of substrates to fermentation products, the type of reactor used and mode of operation employed, medium composition, and operation parameters. There are many possible modeling strategies that can be applied to model cellular systems and bioreactors systems, eventually integrated with each other. Here we overview the key modeling concepts employed in this thesis.

There are two prevalent approaches to develop a bioprocess dynamical model based on mechanistic and/or phenomenological knowledge:

- Empirical stoichiometric models. Cells are treated as black-boxes.

- Models based on a detailed metabolic network of the cells. The list of metabolites and metabolic reactions are considered in the model.

Empirical stoichiometric models are mostly useful when the purpose of the study is to check overall balances of metabolites in and out of the cell, reducing all the chemical reactions to a single lumped reaction (Nielsen et al. 2003). This approach assumes a constant stoichiometry for the conversion of substrates to the several

products of the cell, including biomass (Duboc and Stockar 1998):

$$X + \sum_{i=1}^{m} Y_{xp_i} P_i - \sum_{i=1}^{n} Y_{xs_i} S_i = 0 \tag{1.1}$$

Equation 1.1 represents the general biomass mass balance formulation for a black box stoichiometry, where X represents the biomass concentration, P_i the cell products concentration and S_i the substrates concentration consumed by the cell. The values of Y represent the yield coefficients of each of S_i substrates and P_i products per biomass produced.

When a detailed metabolic network is used for model development, the complexity of the problem increases but, because of the constant balancing of formation and consumption of intracellular metabolites, the degrees of freedom do not necessarily increase. This is an important property of metabolic networks, and is defined by a stoichiometric constraint, based on the mass balance of intracellular metabolites, establishing the basic structure for the network (Nielsen et al. 2003; Price et al. 2004).

The mass balance can be defined mathematically as a system of linear equations (Reder 1988). For a metabolic network consisting of m metabolites and r reactions, the system can be described by:

$$\frac{dC_i}{dt} = \sum_{j=1}^{r} n_{ij} v_j \quad for \quad i = 1, ..., m \tag{1.2}$$

This equation describes the change rate for each metabolite concentration (C_i) involved in the network, with the quantities n_{ij} representing the stoichiometric coefficients of the ith metabolite in the jth reaction. The system of linear equations thus represented can be conveniently rewritten in matrix notation, which is a compact form better suited for several types of analysis:

$$\frac{d\mathbf{C}}{dt} = \mathbf{N}\,\mathbf{v} \tag{1.3}$$

where $\frac{d\mathbf{C}}{dt}$ represents the vector of time derivatives of the metabolite concentration vector $\mathbf{C} = [C_1, C_2, \ldots, C_m]^T$, $\mathbf{v}$ the vector of intracellular reaction fluxes $[v_1, v_2, \ldots, v_r]^T$, and $\mathbf{N}$ a matrix of m rows and r columns holding the stoichiometric coefficients, defining the stoichiometric matrix (Reder 1988).

The structure of the metabolic network is a representation of the information contained in the stoichiometric matrix. With this information, it is possible to compute several important network properties, such as the determination of all admissible fluxes in steady-state, finding out conservation relations for included reactants as well as figuring out dead ends and unbranched reaction pathways. If a mass balance is assumed to be in steady-state, there is no net accumulation of each metabolite inside the cell, and consequently the concentration time derivatives are zero (Price et al. 2004). Thus in steady-state, we can write:

$$\mathbf{N}\,\mathbf{v} = \mathbf{0} \tag{1.4}$$

In most cases, intracellular reactions outnumber the total sum of metabolites and therefore, stoichiometric matrices are overdetermined, that is, its rank is inferior to the number of columns (intracellular fluxes). In such cases, Equation (1.4) has nontrivial solutions, implying that there are at least $r - rank(\mathbf{N})$ vectors belonging to and spanning the nullspace of $\mathbf{N}$. This way, a kernel matrix $\mathbf{K}$ fulfilling:

$$\mathbf{N}\,\mathbf{K} = \mathbf{0} \tag{1.5}$$

shows the respective linear dependencies (Heinrich and Schuster 1996). The column space of the kernel matrix has a very important meaning in the analysis of metabolic networks, mainly because any linear combination between these vectors can generate a steady state flux vector $\mathbf{v}$ that is a solution to Equation (1.4).

While all previous equations apply to intracellular metabolites, some metabolites and large molecules are transported across the cellular membrane and accumulate

in the extracellular phase. The material balancing of such compounds builds the backbone of a bioreactor dynamic model. When the bioreactor is well-mixed with a single inlet and single outlet streams, such material balances take the following general form:

$$\frac{d\mathbf{C}}{dt} = \mathbf{N}\,\mathbf{v}\,X + D\,\mathbf{C_{in}} - D\,\mathbf{C} + \mathbf{Q_C} \tag{1.6}$$

with $\mathbf{C} = [C_1, C_2, \cdots, C_m, X]^T$ a vector of extracellular compounds, one of them being biomass (X), D the dilution rate ($D = F/V$ with F the volumetric flow rate of the bioreactor outlet stream and V the liquid volume inside the reactor), $\mathbf{C_{in}}$ the vector of concentration in the bioreactor inlet stream and $\mathbf{Q_C}$ the vector of volumetric mass transfer rates.

The above models are only fully established when the reaction kinetics term, $\mathbf{v}$, is defined with appropriate kinetic laws in the sense of describing faithfully microbial kinetics, which is probably the most demanding task when developing a bioprocess model. Microbial kinetic models are typically divided between structured and unstructured models. The latter regards the cell as an undifferentiated mass, where all the conversion processes of substrates into products are lumped into a single reaction (Nielsen et al. 2003). In structured models, several levels of structure can be differentiated over the unstructured model, such as metabolic network of intracellular reactions, the consideration of cellular compartments where specific reactions occur, or even structure at the gene level (García-Ochoa et al. 2004). The application of unstructured models is much easier to apply, since it simplifies the cellular processes into fewer parameters than the structured models. Nonetheless, this simplification suffers from the fact that certain tendencies may not be successfully described by experimental data, particularly when some operational variables are modified (García-Ochoa et al. 1996). Unstructured models can be classified in two groups: those which take into account dependencies of growth and production on nutrients, and those where the growth and production is a function of biomass only and its function with time (García-Ochoa et al. 1995). Monod-type models are an

example of the first class of unstructured models, while the latter class is normally exemplified by the logistic equation for biomass growth prediction.

Due to the difficulties in developing accurate kinetic models based on mechanistic principles, hybrid semi-parametric modeling has been proposed as a cost-effective alternative to purely mechanistic or phenomenological modeling. Hybrid semi-parametric models bundle nonparametric models such as Artificial Neural Networks (ANN) with mechanistic descriptions of kinetic expressions in the same mathematical structure (Chen et al. 2000; Thompson and Kramer 1994). There are several successful case studies of the application of this technique in bioprocesses, such as in the production of inulinase (Menküc et al. 2008), recombinant proteins from BHK-21 cultures (Teixeira et al. 2005), and also in the production of xanthan gum from *Xanthomonas campestris* (Zabot et al. 2011).

1.4.2 Overview of EPS modeling and optimization studies

Currently, most industrial optimization problems are addressed empirically using statistical models. Within this class of methods, the Response Surface Methodology (RSM) takes a prominent position in the bioprocess industry panorama. It has aslo been previously reported for microbial EPS production. In short, this method consists on a statistic approach that is employed to identify relationships between response and independent variables. The data, generally obtained from a design of experiments (Lundstedt et al. 1998), is typically fit to a second-order polynomial equation in order to capture the effects of the operational factors under test (independent variables). Since second-order polynomial equations have one extreme point, it is possible, in theory, to identify optimum response values for a given set of independent variables. The popularity of this technique arises from its ease of use to perform optimization studies, and for being less labor intensive compared to the classical experimentation and optimization procedures, in which a one variable at a time technique is used. Hence the RSM provides a large amount of information from a smaller set of

experiments. Moreover, the individual effects of independent variables as well as their interactions can also be assessed with this approach (Baş and Boyacı 2007). Due to these properties, many studies are found in the literature for the optimization of culture media compositions and other critical variables in the production of exopolysaccharides, such as production of EPS from *Enterobacter* A47 (Torres et al. 2012), xanthan gum (Ben Salah et al. 2010; Psomas et al. 2007), curdlan gum (Cui and Qiu 2012), gellan gum (Bajaj et al. 2007; Banik et al. 2007), dextran (Naessens et al. 2004), wellan gum (Li et al. 2012), cellulose (Panesar et al. 2012; Zeng et al. 2011), hyaluronic acid (Chen et al. 2012) and levan (Silbir et al. 2014).

Despite the advantages in the application of RSM models, there are several limitations that must be taken into account. Fitting the data to a second-order polynomial may not be the best practice for all systems. For instance, temperature effects in biochemical processes are typically described by symmetrical or nonsymmetrical bell-shaped curves, which may be difficult to explain when the data is fit to a quadratic equation, especially in the case of nonsymmetrical temperature profiles. This fact restricts the usability of RSM models to changes which are described by quadratic functions (Baş and Boyacı 2007).

A more comprehensive approach for modeling the production of EPS is the development of dynamic models based on material balances and kinetic laws. Unlike the RSM, kinetic models take into account the temporal dimension and are more realistic in terms of the description of the mechanisms underlying the process. The EPS production is often modeled by means of the Luedeking-Piret equation, and the consumption of carbon and/or nitrogen sources expressed by stoichiometric coefficients (García-Ochoa et al. 1995). Xanthan gum is the polysaccharide with the highest industrial production and, since the early sixties, numerous publications have been available for xanthan batch and continuous production (García-Ochoa et al. 1995; García-Ochoa et al. 2004; Rosalam and England 2006). Most of the published works deal with the development of unstructured models, with some variations on

the nature of the kinetic laws applied. Structured models have also been described for xanthan gum, incorporating information about biochemical reactions and also process parameters such as temperature (García-Ochoa et al. 1996, 1998; García-Ochoa et al. 2004).

1.5 Thesis motivation and objectives

The present Ph.D. thesis focuses on the production of an EPS composed mainly by galactose, glucose and fucose, by *Enterobacter* A47 using glycerol as the main carbon source. Bench-scale bioreactor experiments have shown that EPS is produced in concentrations as high as 8 g/L, but also that the production rate and composition may vary along cultivation time due to changes in environmental conditions, such as pH, temperature, nutrients depletion, carbon and nitrogen feeding strategy, dissolved oxygen concentration and viscosity increase. The *Enterobacter* A47 genome has been sequenced during the course of the present thesis thus the mechanisms that control EPS synthesis are only scarcely known.

The main objective of this Ph.D. thesis is to develop an *in silico Enterobacter* A47 cellular model integrated with a bioreactor process model and to use these models to optimize EPS productivity with consistent quality. The model should comprise the following main blocks:

- Central carbon metabolic reactions
- Lumped biomass synthesis reaction
- Pathways for EPS synthesis
- Effect of environmental factors including pH and temperature, which have been identified as key factors in previous studies.
- Dynamical description of the key state variables

The model will be developed step wise, with incremental improvements by the inclusion of additional mechanisms as experimental data becomes available to validate such mechanisms. Once model validation with experimental data is concluded, if model validation and prediction power criteria are satisfied, optimal control scenarios will be investigated with the objective to further optimize EPS productivity under certain quality constrains. In particular the following control degrees of freedom will be investigated:

- Temperature
- pH
- Glycerol dynamic feeding
- Nitrogen to carbon ratio

Last but not the least it is expected with this thesis to generate new knowledge and tools that will help to engineer *Enterobacter* A47 either at and process level eventually leading to a robust, consistent and highly productive EPS production process.

1.6 Thesis outline

This Ph.D. dissertation comprises the following chapters:

- **Chapter 1** introduces the key scientific topics pertinent to the work developed, followed by the motivation, objectives and thesis outline.
- **Chapter 2** covers the construction of a metabolic network for *Enterobacter* A47. The theoretical background for the construction of structured metabolic models will be discussed, highlighting the contrast with black-box metabolic models. The metabolism of the *Enterobacteriaceae* group is revised, providing

the necessary information for the main catabolic and anabolic reactions related to biomass growth and exopolysaccharide synthesis.

- **Chapter 3** describes two methodologies that allow the interpretation of stoichiometric models based on cellular flux data. The first method is termed Principal Elementary Mode Analysis (PEMA) and is based on Principal Component Analysis (PCA) to capture the maximum variance of a fluxome data set, using Elementary Modes (EM) as Principal Components (PC). This way a given set of EMs was used to reconstruct the experimental data while providing a clearer picture of the metabolic pathways involved in the fluxome. The second strategy involves the methodology of Yield Analysis (YA) to determine a minimal set of EMs to describe each flux distribution on a data set. Both methodologies were applied to a data set of extracellular metabolite measurements in an *Enterobacter* A47 culture experiments. The chapter compares the results and discusses the pros and cons of PEMA and YA methodologies.

- **Chapter 4** covers the dynamic modeling of EPS production. In this section, a new modeling approach was applied to the fed-batch production of EPS from *Enterobacter* A47. The strategy applied here was based on the combination of a parametric model containing mechanistic expressions about the process, with a nonparametric model composed of an Artificial Neural Network that incorporates additional data from the process (temperature, pH and broth viscosity), to compensate for the parametric model inaccuracies.

- in **Chapter 5** a model-based optimization is described supported by the modeling studies developed in chapter 4. Several scenarios were designed for the process improvement following an optimal-control methodology. In each case, a different objective function was defined: maximization of EPS final mass, minimization of residual glycerol quantity, maximization of EPS productivity and maximization of the yield production of EPS per glycerol consumed. A

clustering method based on the kmeans algorithm was applied to assess the nonparametric model reliability.

- **Chapter 6** presents the final remarks and main conclusions of this thesis. Questions about future work are also addressed in this section.

Chapter 2

Metabolic Network for *Enterobacter* A47

Living cells are organized in a highly complex network of chemical reactions that permeate all their functions. The study of these networks is fundamental from the viewpoint of bioprocess engineering, since it contains all the information about the processes that are responsible for cell reproduction and viability, to the synthesis of products with commercial value. In this chapter we construct a metabolic model of *Enterobacter* A47 based on data provided by its genome reconstruction, online databases and literature surveys. This information allowed the identification of central carbon metabolism for *Enterobacter* A47, as well as the biosynthetic pathways leading to the key sugars involved in the synthesis of EPS. As such, it was possible to construct a core metabolic network for this organism, containing the glycolysis/gluconeogenesis pathways, glycerol dissimilation reactions, the pentose-phosphate pathway, TCA, respiratory chain and oxidative phosphorylation and also fermentative pathways. Biosynthetic pathways for the generation of EPS building blocks were also added to this model. This work represents the first step towards a systems biology approach to the study of the metabolism of *Enterobacter* A47.

2.1 Introduction

The field of systems biology has been detrimental to shape the way biotechnological research is performed. This systems level approach to the study of microbial cells has allowed the construction of predictive models for genetic and/or environmental perturbations, providing the means for the rational engineering of microbial strains with improved capabilities (Lee et al. 2005). Although a great variety of metabolic modeling approaches exists, all of them require a stoichiometric matrix based on a reconstructed metabolic network (Terzer et al. 2009). This matrix translates the network's information into a mathematical format, with the metabolites organized row-wise and the biochemical reactions displayed column-wise. Each matrix value corresponds to the respective stoichiometric coefficient for a given metabolite in a given reactions. This mathematical representation codifies the network's topological properties, providing the structural basis to determine all the observable cellular functions.

The success in the reconstruction of metabolic models is indebted to the advances in high-throughput and computational technologies, resulting in the genome sequencing of a growing number of organisms across all three domains of life (Janssen et al. 2005). Despite applications in metabolic engineering, genome-scale models have also been used in the contextualization of high-throughput data, to direct hypothesis-driven discovery, interrogation of multi-species relationships and network property discovery (Oberhardt et al. 2009).

Reconstructed metabolic networks attempt to represent the cellular metabolism occurring in biological systems. As such, metabolic reactions fall into two types of chemical transformations: catabolic and anabolic pathways. The former consist on a subset of reactions that break down molecules with a high reduction potential, such as sugar hexoses (*i.e.* glucose and fructose), ensuring a supply of Gibbs free energy (generally in the form of high energy phosphate bonds) and precursor metabolites. Conversely, anabolic reactions use the energy and metabolite pools generated by

catabolism to the synthesis of biomolecules with higher complexity (Lehninger et al. 2000; Nielsen et al. 2003).

There are several levels of hierarchical detail in the conceptualization of reconstructed networks (Palsson 2006). In a broader sense, the metabolism involves the uptake of substrates as inputs to the network and the synthesis of biomass and by-products as outputs. For many industrial fermentation processes, this simplified view of cells suffices. These descriptions comprise a simple set of coupled mass and energy balances, with various empirically determined yield coefficients, describing the partitioning of consumed substrate, with many biological processes lumped into single reactions (*e.g.* biomass synthesis equation) (Nielsen et al. 2003; Palsson 2006). Models of this type are only useful for a limited set of specific conditions.

Higher level of detail can be obtained by dividing the intermediary metabolism into two basic sectors: synthesis of biosynthetic precursors and synthesis of building blocks. The former are generated by the catabolism to integrate the various anabolic pathways, which originate the building blocks for macromolecular biosynthesis. It may be useful to describe the metabolism at this level to engineer bacteria in bioprocessing (Palsson 2006).

A third level of resolution involves detail about the metabolic pathways in networks. At this level of description, the kinetic regulation of biochemical reactions in the pathway and also key metabolic pools such as the energy charge dominate this characterization (Palsson 2006). Finally a fourth level of detail is focused in individual reactions as the simplest unit of chemical conversion in metabolic networks.

Various sources of information can be used to assist the reconstruction of metabolic networks, including biochemical information regarding metabolic reactions, genomics information containing functional assignments to open reading frames (ORF), based on DNA sequence homology, physiology and indirect information, such as empirical evidence of a strain's ability to produce certain by-products. A final source of information comes from *in silico* modeling data which often aid to infer metabolic

reactions that close the gaps in reconstructed networks (Palsson 2006; Reed et al. 2006). Most of the referred sources of information are contained in online databases such as MetaCyc (Caspi et al. 2006), EcoCyc (Keseler et al. 2005) and KEGG (Kanehisa and Goto 2000).

2.2 Construction of a core metabolic model for *Enterobacter* A47

2.2.1 Genome reconstruction of *Enterobacter* A47

The genome sequence of *Enterobacter* A47 was recently obtained by the Federal University of Pará, Brazil, representing the first step to generate a genome-scale reconstruction of the metabolism. The sequenced data was annotated and reconstructed using the RAST algorithm (Aziz et al. 2008), available as an on-line service (www.rast.nmpdr.org). This reconstructed genome is the main resource for the construction of a core metabolic network for *Enterobacter* A47. The whole genome has a size of 4,992,161 bp and the resulting reconstruction identified 4705 coding sequences, with 25 possibly missing genes. Figure 2.1 sumarizes the gene groups identified in the reconstructed genome.

2.2.2 Overview of *Enterobacter* A47 metabolism

The genome reconstruction represents the first step to uncover the details of *Enterobacter* A47's metabolism. Although a complete genome-scale metabolic reconstruction is still unavailable, the main features relative to the carbon metabolism and synthesis of EPS building blocks can be elucidated, based on the current knowledge obtained from the genome reconstruction. Moreover, the reliance on online databases and literature surveys contributed to the development of a small-scale metabolic network for *Enterobacter* A47.

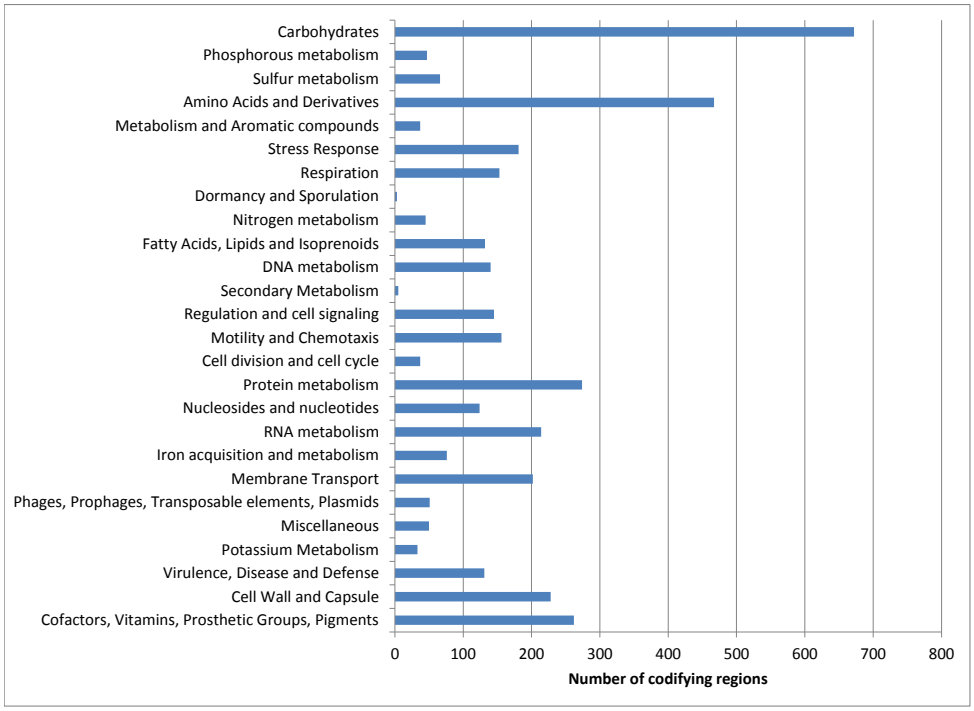

Figure 2.1: Summary of the genome reconstruction of *Enterobacter* A47

The group referred to as carbohydrates represented by the genome reconstruction (Figure 2.1) contains most of the catabolic reactions in *Enterobacter* A47. Two major pathways for sugar degradation are found in this strain's metabolism: the Embden-Meyerhof-Parnas and Entner-Doudoroff pathways. These pathways are responsible responsible for the generation of ATP and NADH pools, as well as the synthesis of precursor metabolites for biosynthesis such as pyruvate, phosphoenolpyrutvate and glyceraldehyde-3-phosphate. A third possibility for the sugar catabolism was found to be the methylglyoxal pathway, which is activated in conditions of overflow in the uptake of certain substrates such as glucose, glycerol or lactate (Weber et al. 2005).

The experimental evidence for the uptake of glycerol by *Enterobacter* A47 is also backed-up by the presence of coding regions in the genome for the glycerol transport system (glpT) and the glycerol uptake facilitator protein (glpF). The

transport system takes part in the phosphorylation of glycerol which is converted to dihydroxiacetate-phosphate, entering the glycolytic or gluconeogenic pathways.

The pyruvate is the final metabolite in both Embden-Meyerhof-Parnas and Entner-Doudoroff pathways. The tricarboxylic acid (TCA) cycle is one of the possible routes that can be taken by pyruvate. Following a decarboxylation reaction, pyruvate is converted to acetyl-CoA, which can enter TCA by reaction with oxaloacetate. The successive sequence of reactions results in the generation of four molecules of NADH, one molecule of $FADH_2$ and two CO_2 molecules. The glyoxylate pathway is also evidenced in the genome reconstruction of *Enterobacter* A47. This pathway is relevant in biosynthetic pathways, particularly in the biosynthesis of polysaccharides.

The reconstructed genome shows evidence for the group of respiratory complex reactions, which include the NADH (and NADPH) ubiquinone oxidoreductase chain, as well as a series of NADH dehydrogenases which intervene in the so-called respiratory chain. These reactions are the major sink for all the redox potential generated in the cellular metabolism, transferring electrons from NADH to O_2 in a series of redox reactions. The coupled activity of the ATP synthase takes advantage of the proton gradient generated in the respiratory chain to promote the generation of ATP molecules in the process of oxidative phosphorylation (Lehninger et al. 2000; Madigan et al. 2003).

The strain of *Enterobacter* A47 also possesses some fermentative pathways that convert pyruvate into organic acids whenever oxygen is unavailable. The genome reconstruction has shown evidence mixed-acid fermentation pathways, which include the conversion of pyruvate into ethanol, acetate and lactate. Moreover, the butanediol fermentation is also evidenced in the genome data, which is also a typical by-product among the group of enteric bacteria (Madigan et al. 2003).

As described in Section 1.2, *Enterobacter* A47 is able to synthesize heteropolysaccharides incorporating neutral sugars such as glucose, galactose, fucose, rhamnose, glucuronic acid and glucosamine (Torres et al. 2012). Acyl group substituents such

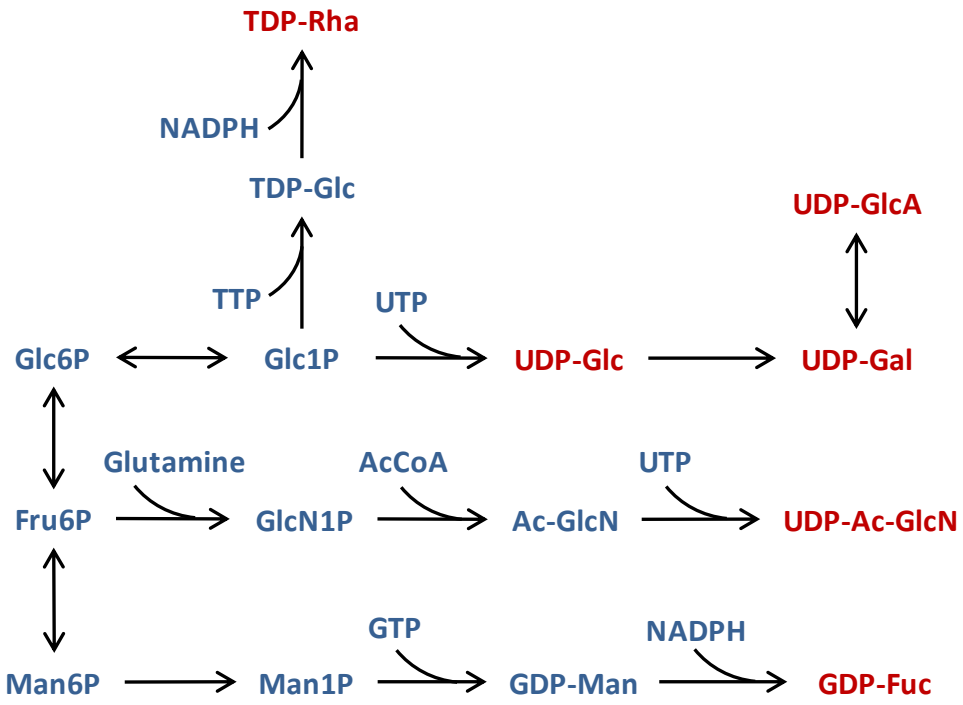

Figure 2.2: Biosynthetic reactions leading to the sugar building blocks for EPS synthesis (metabolites in red). Legend: UTP, uridine triphosphate; TTP, thymidine triphosphate; GTP, Guanosine triphosphate; AcCoA, acetyl coenzyme A; NADPH, nicotinamide adenine dinucleotide phosphate; Glc6P, glucose-6-phosphate; G1P, glucose-1-phosphate; Fru6P, fructose-6-phosphate; Man6P, mannose-6-phosphate; Man1P, mannose-1-phosphate; GlcN1P, glucosamine-1-phosphate; Ac-GlcN, acetyl glucosamine; UDP-Glc, UDP glucose; UDP-Gal, UDP galactose; UDP-GlcA, UDP glucuronic acid; TDP-Rha, TDP rhamnose; UDP-Ac-GlcN, UDP acetyl glucosamine; GDP-Fuc, GDP fucose; GDP-Man, GDP mannose.

as acetyl, pyruvyl and succinyl have also been shown to be incorporated into EPS composition (Torres et al. 2012). The synthesis of all major building blocks for EPS synthesis has been confirmed by the genome reconstruction of *Enterobacter* A47. The lack of details about these reactions was complemented with knowledge contained in online data bases and literature sources from closely related organisms such as *Escherichia coli*.

For *Enterobacter* A47, the synthesis of EPS building blocks was assumed to derive mainly from the conversion of Glucose-6-phosphate to uridine diphosphate glucose

(UDP-Glc). This sugar nucleoside is than converted into other sugar species such as UDP-Gal and uridine diphosphate glucuronic acid (UDP-GlcA) (Kumar et al. 2007; Madigan et al. 2003). For the synthesis of fucose, fructose-6-phosphate (Fru6P) is the starting point, involving the formation of mannose and the inclusion of a guanosine residue in this sugar (GDP-Man), which is therefore converted to guanosine diphosphate fucose (GDP-Fuc). For glucosamine, Fru6P is converted to glucosamine-6-phosphate by reaction with glutamine. The biosynthetic pathway for rhamnose starts with the conversion of Glc6P to Glc1P, followed by the addition of thymidine diphosphate, forming thymidine diphosphate glucose. This sugar nucleoside will than be the precursor of thymidine diphosphate rhamnose. The pathways for the synthesis of fucose, rhamnose and glucosamine were adopted from EcoCyc (Keseler et al. 2005). The diagram in Figure 2.2 ilustrates the biosynthesis reactions for EPS synthesis.

2.2.3 Representation of the metabolic network model

A core metabolic network model for *Enterobacter* A47 was constructed, based on its reconstructed genome. This model attempts to represent the metabolism of glycerol uptake and synthesis of EPS (Figure 2.3). The model contains 18 extracellular reactions, reflecting all the fluxes that exchange material with the environment: glycerol, ammonia, oxygen, carbon dioxide, acetate, formate, lactate, ethanol, EPS components (glucose, galactose, fucose, rhamnose, glucuronic acid, glucosamine, acetyl, pyruvyl and succinyl), and a general ATP maintenance term. All the external fluxes were defined as irreversible, with the uptake substrates entering the network and organic acids and EPS components exiting the system. The production of organic acids is linked to the fermentative pathways, which may take place whenever oxygen becomes scarce in the culture media.

The model also includes 37 intracellular metabolites and 58 intracellular reactions, forming the glycolysis/gluconeogenesis, penthose-phospates, tricarboxylic acid cycle and fermentative pathways. Since glycerol is the only carbon source used by the

cell, dissimilation reactions were considered, linking to the glycolysis pathway. A comprehensive list of the metabolic reactions and metabolites included in the model is available in Appendices A.1 and A.2.

The respiratory chain and phosphorilative oxidation reactions were lumped into a single overall reaction, based in the value of the P/O ratio, that is, the ratio between the number of ATP moles produced per oxygen consumed. This value is a variable quantity, depending on the relative formation of NADH and $FADH_2$. If only NADH molecules were used in the respiratory chain, the P/O ratio would have a value of 3 but, since $FADH_2$ is also formed in the TCA cycle, this value would likely be less than this value. Some microorganisms also lack one or more proton pumping sites, resulting in a substantially lower value of P/O ratio. The overall stoichiometry for the oxidative phosphorylation can be summarized in the following equation:

$$\begin{aligned} &NAD^+ + (1 + P/O)H_2O \\ &+ (P/O)ATP - NADH - 0.5O_2 - (P/O)ADP - H^+ - (P/O)P_i = 0 \end{aligned} \tag{2.1}$$

Exopolysacccharide biosynthesis was represented by the reactions that lead to the required sugar and nonsugar building-blocks. Each reaction was lumped on a single flux, linking a precursor metabolite from the central carbon metabolism, to a single building-block, that is, for each nucleoside sugar or activated acyl group that integrates the polysaccharide chain. The energetic and redox potential requirements were all taken into account, based on the literature and database information (Byun et al. 2007; Jarman and Pace 1984; Keseler et al. 2005; Troy et al. 1971).

2.3 Conclusions

In the present chapter we describe the construction of a metabolic model of *Enterobacter* A47, based on the genome sequence obtained from this organism, online databases and literature sources. The genome annotation and reconstruction was performed by the RAST algorithm, which is provided as an on-line service. The

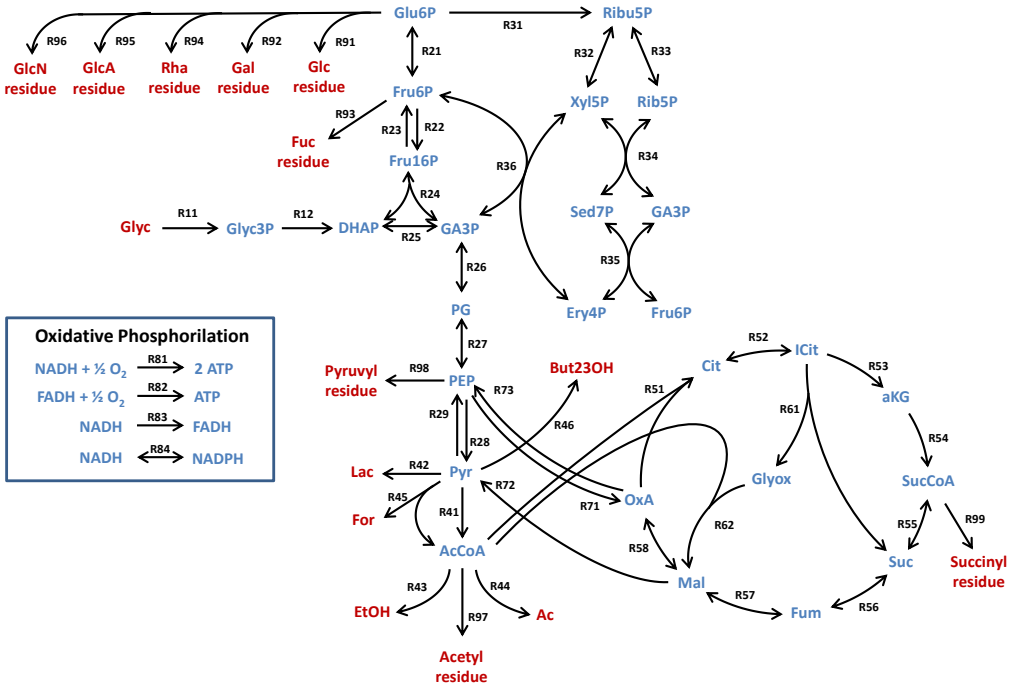

Figure 2.3: Metabolic network for the central carbon metabolism based on reconstructed genome of *Enterobacter* A47. Detail about reactions involved in the conversion to the EPS building blocks are lumped into single reactions. The metabolites in blue represent intracellular species, while the metabolites in red are exchanged with the extracellular medium.

genome has a dimension of 4,992,161 base pairs and 4705 coding sequences identified, with 25 possible missing genes. With this information it was possible to identify the core metabolic pathways present in the metabolism of *Enterobacter* A47, as well biosynthetic pathways leading to the formation of key sugars present in EPS.

The proposed metabolic network has 37 intracellular metabolites, 58 intracellular reactions and 18 exchange fluxes, establishing the material transfer between the cellular metabolism and the environment. These include the uptake of ammonia, glycerol and oxygen, as well as the excretion of metabolic products including EPS. Due to the current lack of information regarding the mechanisms of EPS assembly, the model was built to account only to the level of the synthesis of its building blocks. The intracellular reactions represent the reactions of glycolysis/gluconeogenesis pathway, glycerol dissimilation, the pentose-phosphate pathway, TCA cycle, respiratory chain

and oxidative phosphorylation and also fermentative pathways.

The construction of this metabolic model is a first step in the direction of a systems biology approach to the study of the metabolism of *Enterobacter* A47. This can potentially lead to the possibility of metabolic engineering for improving bioprocesses destined to the production of EPS.

Chapter 3

Identification of Elementary Flux Modes by experimental yield data: the case of EPS production by *Enterobacter* A47

An elementary flux mode may be defined as a minimal set of enzymes able to operate at steady state, encoding a particular metabolic state of the cell. The universe of elementary flux modes defines the full set of non-decomposable steady-state flux distributions that the underlying metabolic network can support. However, the vast majority of metabolic states encoded by elementary flux modes are not physiological or thermodynamically feasible under a given set of environmental conditions. For this reason, data reconciliation methods are needed to identify a subset of "active" elementary flux modes of the real system. In this work, two methodologies were studied to identify a subset of elementary flux modes constrained by experimental yield data: principle elementary modes analysis (PEMA) and yield analysis (YA). The EPS synthesis by *Enterobacter* A47 was used

We use as example EPS synthesis by *Enterobacter* A47 represented by a network of 58 metabolic reactions and 1066 elementary flux modes using a data set of 11×9

independent observations of yield values. Both PEMA and YA methods seek for the minimal set of active elementary modes compatible with the idea that a highly evolved organisms uses the most efficient pathways adapted to a given environmental condition. We have observed that PEMA identifies 5 active EMs explaining 98.5% of measured data variance while YA requires 9 active EMs to explain roughly the same amount of variance (99%). Moreover, the degree of orthogonality of EMs identified by YA is higher than those selected by PEMA. All in all we conclude that the PEMA is more efficient in identifying a minimal set of EMs then YA.

3.1 Introduction

Metabolic networks can be mathematically described by a system of linear equations (*i.e.* stoichiometric matrix), which represents an interconnected network of material relationships between metabolites assuming that the cells are in steady state. Such stoichiometric linear modes are generally undetermined, which means there are more variables (metabolic reaction rates) than system equations (metabolite balances). From these considerations, the stoichiometric matrix null space can be defined with a finite number of basis vectors, whose linear combination generate all the possible fluxes operating in steady-state. However, not all steady-state fluxes are biologically feasible and for that reason some constraints must be imposed in the null space (*e.g.* reversibility of biochemical reactions). Thus, the concept of elementary modes (EM) can be formulated as the minimal set of cellular reactions able to operate in steady-state (Schuster et al. 2000).

The set of elementary modes is obtained from convex analysis (Schuster et al. 2002) and is unique for a given metabolic network. Three basic conditions are fundamental to define the EM space: a pseudo-steady state condition, a feasibility condition and a non-decomposability condition (Klamt and Stelling 2003). As already stated, the first condition ensures the conservation of the metabolite levels in the network. The feasibility condition demands that only thermodynamically realizable fluxes are contained in a given elementary mode. The latter condition implies that there is no elementary mode (excluding the zero vector) that can be further decomposed in some linear combination of other elementary modes.

Moreover, because the set of elementary modes can function as a convex basis, any particular steady-state flux distribution can be obtained as a non-negative linear combination of elementary modes. Motivated by these properties, elementary mode analysis has become a widespread technique to assess robustness of cellular functions (Klamt 2006; Stelling et al. 2002; Wilhelm et al. 2004), interpreting metabolic functions (Carlson and Srienc 2004; Gayen and Venkatesh 2006), assessing aspects

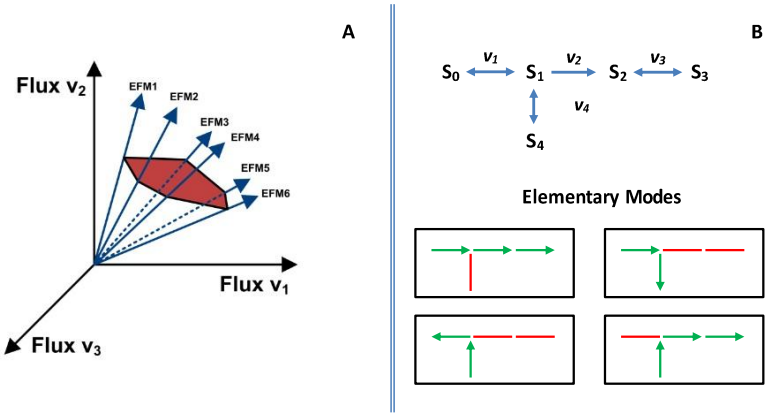

Figure 3.1: Illustration of the concept of Elementary Modes: (A) Representation of the flux cone as a subspace of the stoichiometric matrix null space. Each elementary mode consist on a convex basis for the set of feasible fluxes operating in steady-state. (B) schematic representation of Elementary Modes in a metabolic network.

of functionality and regulation in the network structure (Stelling et al. 2002) and improving strain performance (Trinh et al. 2006).

Current algorithms for the computation of elementary modes face a common problem when dealing with highly interconnected metabolic networks (Klamt and Stelling 2002). In such cases, the combinatorial calculation of the elementary modes number becomes increasingly complex, rendering the analysis of large networks difficult. However, there constantly appear new methods for computing EMs of large networks in an efficient and fast way, *e.g.* Badsha et al. (2014) and Quek and Nielsen (2014).

Most of the pathways described by elementary modes are not physiological or thermodynamically feasible under a given set of environmental conditions. For this reason, some methods have been proposed to select a set of most representative or active elementary modes. Among the many possible methods (for a review see Trinh et al. (2009)) the identification of a subset of elementary modes, constrained by measured yield data, is particularly interesting. This kind of data is frequently available in cell culture experiments. This problem has been studied before by Song et al. (2009), who have proposed the yield analysis (YA) method to discriminate a subset of elementary modes that best represent the measured extracellular yield data.

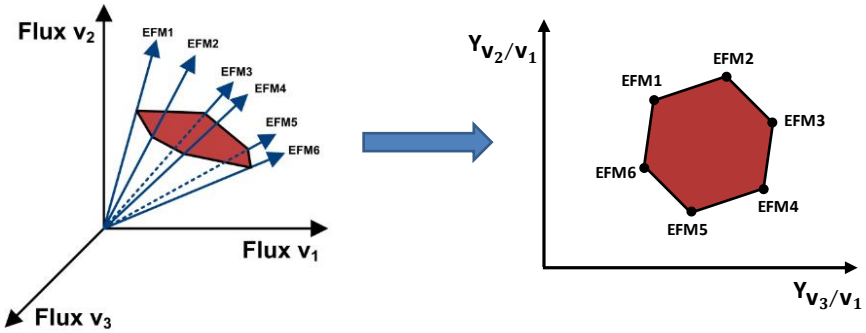

Figure 3.2: Transformation of the flux space into yield space, defined by a bounded convex hull.

In this method, the space defined by external fluxes is regarded as an important aspect of EM selection, since the internal fluxes of a metabolic network are strictly constrained by the status of external fluxes (Song et al. 2009). This way, yield analysis posits as a rational way for selecting a set of EMs with meaningful solution space. Thus the first step in YA consists in transforming the entire flux space defined by the EM set into yield space. This is done by dividing each of the fluxes contained in the elementary modes by a flux of reference, generally the carbon source (glycerol flux in the present work).

Figure 3.2 represents the mapping of a flux cone defined by three fluxes, into the yield space. By taking flux v_1 as the flux of reference, the yields of the remainder fluxes are calculated. Thus, the space defined by the edges of the flux cone (*i.e.* elementary modes) is recreated in yield space as vertices of the bounded convex hull. Every feasible metabolic pathway can thus be represented as a convex combination of the EMs at the vertices:

$$\mathbf{y} = \mathbf{Z_y} \cdot \boldsymbol{\lambda}, \quad \boldsymbol{\lambda} \geq \mathbf{0}, \quad \|\boldsymbol{\lambda}\|_1 = \mathbf{1} \tag{3.1}$$

The vector $\mathbf{y}$ represents the yield vector, $\mathbf{Z_y}$ the matrix of elementary modes converted

to yield values and $\boldsymbol{\lambda}$ the weights vector. In order to identify a subset of EMs that best represent the data some additional methods are needed to identify columns of $\mathbf{Z_y}$ and to estimate the respective weights $\boldsymbol{\lambda}$. Here we compare two approaches. In the original yield analysis method active EMs are chosen as the minimal number to enclose the measured yield point with the weighting coefficients determined by quadratic programming. As alternative we apply the recently published Principal Elementary Modes Analysis (PEMA) ("Principal Elementary Mode Analysis"), which consists of a principle components like decomposition methodology where principle components are substituted by elementary modes. Both methods are illustrated with a yield data set for EPS production from glycerol by *Enterobacter* A47.

3.2 Methods

3.2.1 Experimental data set

The experimental dataset used in this work was provided by Torres et al. (2012). The data was produced in eleven fed-batch experiments with EPS production under different values of temperature and pH. Measurements of the final EPS concentration and respective chemical composition in terms of its sugar monomers and acyl groups, were converted to yields by dividing by total amount of glycerol spent in the fed-batch experiment. The EPS components considered were fucose, glucose, galactose, glucuronic acid, rhamnose, glucosamine, pyruvate, succinate and acetate. The rows in the data matrix reflect the eleven fed-batch runs, viewed as the number of variable observations. Details of the experimental method are provided elsewhere (Torres et al. 2012).

3.2.2 Elementary Flux Modes

The metabolic network used in this study comprehends a subset of 58 reactions describing the synthesis of the EPS building blocks (Fuc, Gal, Glc, GlcA, Rha, GlcN,

Table 3.1: Central composite design with two independent variables: temperature, pH. The rates were converted to yields.

Temp	**pH**	**Yields of EPS monomers per glycerol consumed (mmol mol^{-1})**								
(°C)		**Fuc**	**Gal**	**Glc**	**GlcA**	**Rha**	**GlcN**	**Pyr**	**Succ**	**Acet**
20	6.0	1.400	2.545	6.109	1.145	1.145	0.255	0.538	0.000	0.793
40	6.0	1.957	3.478	12.83	1.304	1.739	0.435	0.938	0.349	1.383
20	8.0	9.734	7.138	11.68	3.569	0.324	0.000	5.428	1.730	5.718
40	8.0	0.000	12.80	28.80	4.267	5.333	2.133	0.000	0.000	1.640
15.9	7.0	6.839	5.524	9.469	3.156	0.526	0.526	1.731	0.858	2.552
44.1	7.0	0.000	3.393	9.856	1.616	0.323	0.808	0.000	0.774	0.512
30	5.6	14.42	14.42	64.33	8.873	6.654	2.218	7.329	0.000	14.41
30	8.4	0.000	0.995	3.067	0.912	2.404	0.912	0.176	0.000	0.777
30	7.0	47.08	34.01	36.62	13.08	0.000	0.000	29.01	4.793	33.27
30	7.0	52.54	36.50	40.87	16.06	0.000	0.000	35.56	5.289	26.22
30	7.0	70.74	49.71	51.62	19.12	0.000	0.000	66.02	11.33	59.92

Pyr, Succ and Acet) from glycerol. Biomass synthesis was disregarded form this analysis. The stoichiometric matrix and elementary modes were computed using the software METATOOL, version 5.1 (Pfeiffer et al. 1999). A total of 1066 EMs were computed and organized in matricial form with rows representing metabolic reactions and elementary modes in columns. The rate values were converted to yield values by dividing each value for the consumption rate of glycerol. The metabolic network and respective elementary modes are provided in Appendix A.1.

3.2.3 Data normalization

Both measured yield data and EMs have been normalized dividing by the measurement standard deviation of each yield value. More specifically, the data matrix is scaled by dividing by yields standard deviation (column-wise), the EM matrix is normalized by dividing row-wise by the same standard deviation values. To note that mean centering should not be applied, since changing the sign of the fluxes alters their direction, violating thermodynamic constraints. For details see Folch-Fortuny

et al. ("Principal Elementary Mode Analysis").

3.2.4 Principal Elementary Mode Analysis (PEMA) method

The PEMA model can be formulated as follows:

$$\mathbf{Y} = \mathbf{\Lambda} \cdot \mathbf{Z_{pem}^T} + \mathbf{G} \tag{3.2}$$

where $\mathbf{Z_{pem}}$ is the matrix of Principal Elementary Modes, which corresponds to a subset of the entire EMs set; $\mathbf{\Lambda}$ is the weights matrix, containing the coefficients to fit the original yield data contained in matrix $\mathbf{Y}$ using the PEMs. The term $\mathbf{G}$ represents the residuals matrix. It is worth noting that $\mathbf{\Lambda}$ values are forced to be positive, since only non-negative scaling factors are admitted in the subspace defined by EMs (Schuster and Hilgetag 1994).

The first step of the method consists on calculating a $\boldsymbol{\lambda}$ (*i.e.* weight vector) for each EM, solving the equation:

$$\boldsymbol{\lambda_k} = \mathbf{Y} \cdot \mathbf{Z_k} \cdot (\mathbf{Z_k^T} \cdot \mathbf{Z_k})^{-1} \tag{3.3}$$

where k denotes the k-th EM and, unlike the loadings in principal component analysis, the EMs are not orthonormal, so Equation 3.3 requires the pseudo-inverse computation of $\mathbf{Z_k^T} \cdot \mathbf{Z_k}$. This way the matrix $\mathbf{\Lambda}$ is constructed with each column represented by $\boldsymbol{\lambda_k}$. For each iteration running through all the elementary modes, the predicted data matrix can be computed with the following equation:

$$\mathbf{Y_k^{pred}} = \boldsymbol{\lambda_k} \cdot \mathbf{Z_k} \tag{3.4}$$

with $\boldsymbol{\lambda_k}$ the vector of weigths and $\mathbf{Z_k}$ the k-th elementary mode under evaluation. This information is used to calculate the explained variance attributed to each

elementary mode.

$$Var(\%) = 100 \cdot \left[1 - \frac{\sum_i \sum_j (y_{ij} - y_{ij}^{pred})^2}{\sum_i \sum j(y_{ij})^2}\right] \tag{3.5}$$

This equation measures the amount of variance explained by the k-th EM relative to the total amount of variance contained in the experimental data, with y_{ij} representing the element i row and j column of the data matrix $\mathbf{Y}$, and y_{ij}^{pred} the *i,j* element of the matrix of predicted data $\mathbf{Y^{pred}}$. The EM explaining more variance in the data is classified as the first EM, with this procedure repeated for the subsequent EMs, calculating for each new EM also the $\boldsymbol{\lambda}$ values of the previous EMs. This strategy is a greedy solution that maximizes the amount of variance explained by the current set of EMs.

3.2.5 Yield Analysis (YA) method

Yield analysis (YA) is a concept introduced by Song et al. (2009) as a method of extracting a subset of elementary modes essential for describing metabolic behaviors. This method deals with yield space, taking advantage of an important feature, that is, all feasible yield distributions are contained within a bounded convex hull, opposed to the unbounded flux cone that characterizes the flux space. This implies that a solution in yield space is bounded and therefore does not require the imposition of additional constraints (Song et al. 2009).

The first step in YA is the reduction of the whole set of EMs to a set of generating modes, defined as the minimal set of EMs that span the entire yield space (Wagner and Urbanczik 2005). This procedure was identified by convex hull analysis (*convhull* function in MATLAB). A further reduction can be performed by eliminating generating modes which have a negligible contribution to the overall volume (or area) delimited by the bounded convex hull (*i.e.* yield space). In the case data is not available for some species in the metabolic network, partial coordinates can be

considered for YA (Song et al. 2009). In the present work, only yield data for EPS monomers is available and thus, the yield space has a dimension of 9.

After the first reduction step, selection of active EMs falls within two possible classes of problems. If the measured yield point is inside the convex hull, the active set of EMs is chosen as the minimal number to enclose the experimental point, which is at most $n+1$ EMs in n-dimensional yield space (Song et al. 2009). The solution can be obtained by solving the following quadratic programming problem:

$$\max_{\boldsymbol{\lambda}} \frac{1}{2}\|\boldsymbol{\lambda}\|^{\mathbf{2}} \tag{3.6}$$

such that

$$\mathbf{Z_y} \cdot \boldsymbol{\lambda} - \mathbf{y_m} = \mathbf{0}, \quad \boldsymbol{\lambda} \geq \mathbf{0}, \quad \sum_{i=1}^{m} \lambda_i = 1 \tag{3.7}$$

where $\mathbf{Z_y}$ is the matrix of normalized EMs, $\mathbf{y_m}$ is the vector of measured yield data and $\boldsymbol{\lambda}$ the vector of weights. When the squared norm of $\boldsymbol{\lambda}$ is maximized, $n+1$ nonzero weights are obtained, corresponding to the minimal amount of EMs able generate a datum point inside the convex hull.

When the experimental yield data is located outside the convex hull, it cannot be exactly represented by a convex combination of any set of EMs. The best available option is to select the subset of modes best fitting the data. For this purpose, a least-squares problem can be solved:

$$\min_{\boldsymbol{\lambda}} \frac{1}{2}\|\mathbf{Z_y} - \mathbf{y_m}\|^{\mathbf{2}} \tag{3.8}$$

such that

$$\boldsymbol{\lambda} \geq \mathbf{0}, \quad \sum_{i=1}^{m} \lambda_i = 1 \tag{3.9}$$

After the EMs identification, they were ranked from high-to-low explaining variance. To accomplish this, an algorithm similar to the greedy search described in the Section 3.2.4 was constructed, by applying Equation 3.4 to generate the data predicted

by the k-th EM, followed by the calculation of explained variance (Equation 3.5).

3.3 Results and discussion

3.3.1 PEMA results

The application of PEMA requires the prior computation of EMs, which was performed by METATOOL. Following this procedure, 1066 EMs were calculated for the metabolic network considered in this study.

The performance of PEMA can be judged by the total amount of PEMs that are able to reconstruct the data set. The scree plot is graphical display which is used to assess the appropriate amount of Principal Components to extract from PCA (Bro and Smilde 2014). A similar analysis was performed for the EMs which is showed in 3.3A. This plot shows that 5 PEMs are able to explain 96.7% of the scaled data variance and 98.1% of the real variance. The selected PEMs are EM414, EM241, EM329, EM227 and EM559. It was opted not to consider EMs that explain less than 1% of variance since these modes would likely be describing noise.

3.3.2 YA results

The method of YA was applied to the set of 1066 EMs to reduce it to a smaller set of active elementary modes. Firstly, a pre-selection of EM candidates by convex hull analysis was performed, where the subset of 58 generating EMs was identified. The experimental data was then used to determine 18 active EMs that best characterize the metabolic model. These EMs were then ranked from higher to lower explained variance. The scree plot analysis (Figure 3.3B) shows that 9 active EMs explain 97% of the scaled data variance and 99% of the real variance. The consideration of more than ten active EM would extract less than 1% of variance and, therefore no more than 9 EMs were considered in the model. Hence, the selected EMs are: EM1021, EM103, EM1019, EM1020, EM589, EM524, EM811, EM792 and EM219.

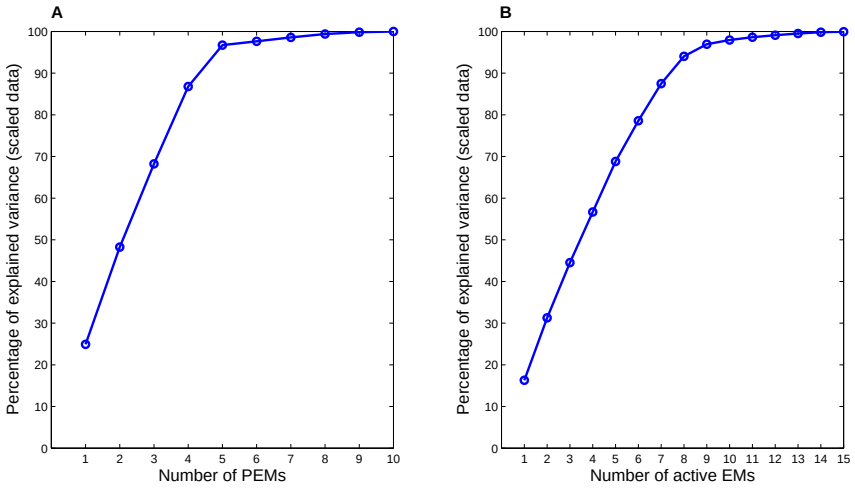

Figure 3.3: Scree plots for PEMA method (A) and YA method (B).

3.3.3 Orthogonality

In orthogonal principle component analysis (PCA), principal components are orthogonal, ensuring that the sum of variances determined by each component equal the totality of variance in the data. This does not occur with neither the PEMA nor the YA methods. As we can see in Figure 3.4 A and B, the sum of explained variances by the EMs with PEMA is 112%, while YA explains a totality of 104% of variance. This stems from the fact that EMs are not orthogonal to each other and, as a consequence they explain common sources of variability. Nevertheless, the EMs explaining more variance can be considered more significant in the model.

The EMs degree of orthogonality can also be assessed by dividing the explained variance explained by the selected EM set, with the sum of the variances explained by each individual EM. If all the EMs under this calculation were orthogonal, a 100% value would be obtained, such as the case with principal components in PCA. The degrees of orthogonality differ in both methods, namely 88% for PEMA and 96% for YA. This result shows that the EMs chosen from the YA exhibit a higher degree of orthogonality (close to 100%) and, due to that fact, each EM would be more efficient in describing unique sources of variation. However, EMs derived from the convex hull

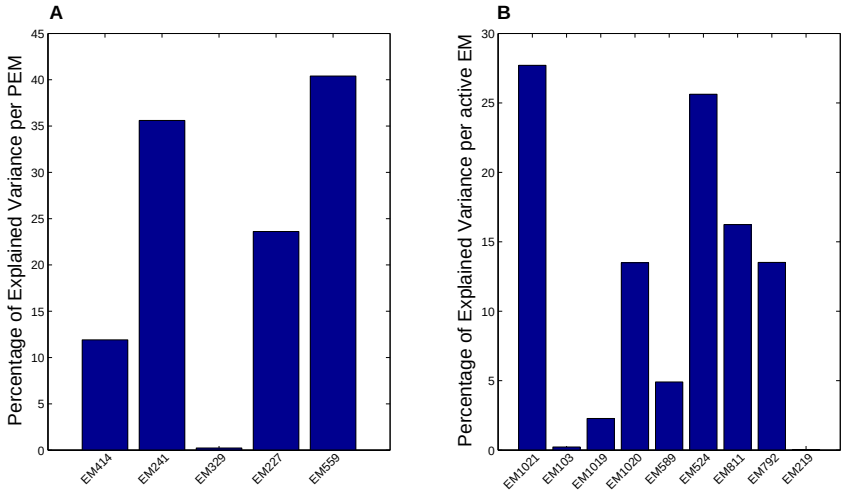

Figure 3.4: Bar plots representing the explained variance for the PEMs in PEMA (plot A) and the active EMs in YA (plot B).

do not extract much variance independently, requiring 9 EMs to reconstruct 98% of the data variance. This is not the case with PEMA, where the five EMs capture 97% of the data variance.

3.3.4 Metabolic interpretability

Despite having a different nomenclature, PEMs and active EMs have an analogous meaning regarding its metabolic significance, representing the most relevant pathways to reconstruct a given data set. Figure 3.5 shows a map of the selected EMs in both PEMA and YA. Both maps are a binary representation with the blue squares showing forward reactions, red squares reverse reactions and the white squares reactions with null value. The yield values for the selected EMs are provided in Appendices A.4 and A.5. Noting that the experimental data used correspond only to the yield values of EPS monomers, it is clear that by analyzing the respective region in the selected EMs (between yield 44 and 52), the YA method requires more EMs to reconstruct the data. Conversely, PEMA needed only 5 PEMs to reconstruct 98% of the data. It is interesting to note that most active EMs in YA predict only the formation of single EPS residues, with the exception of EM103 and EM589 which predict the formation

of two EPS residues. This result is not surprising if we take into account that active EMs are generating vectors for the yield space, that is, they cannot be obtained from any convex combination of other yield vectors. This suggests that the EMs defining the convex hull are closer to the original coordinate system representing the variables, which may explain the higher orthogonality of EMs selected by YA. Since PEMA is not restricted to generating vectors only, it has more freedom to search for PEM candidates to better represent the data. As consequence, PEMA selected PEMs predicting two and three EPS residues.

The EM representation in Figure 3.5 is useful to interpret the metabolic reactions involved in a given cellular state. Similar patterns can be observed in both methods applied. Since the data reflect the conversion of glycerol into EPS, it is expected that most pathways lead to the synthesis of glucose-6P, which is the principal sugar precursor in the metabolism. In fact, the PEMs and active EMs that explain most variance represent routes to the formation of fucose, glucose and galactose residues (EM227, EM241 and EM559 in PEMA and EM1021, EM524 and EM1020 in YA).

The pentose-phosphate pathway (PPP) seems to have a low impact in the EPS metabolism, judging by the result obtained by both methodologies. While no PPP reactions are active in YA solution, two PEMs show positive fluxes through this pathway (EM414 and EM329). However these PEMs explain a very low amount of data variance (EM414 with 11.9% and EM329 with 0.23%) and therefore have a low contribution.

Most of the energetic metabolism is derived from an activation of the respiratory reactions, which are quite pronounced in the network. This occurrence is paralleled with the high energetic demands in other EPS producing strains such as *Xanthomonas campestris*, also favoring the aerobic metabolism (Jarman and Pace 1984). No fermentative pathways were detected in all the PEMs, with the exception of EM227, EM792 and EM219 producing formate. One of the co-products of formate synthesis is acetyl-CoA which in these pathways is incorporated into EPS as an

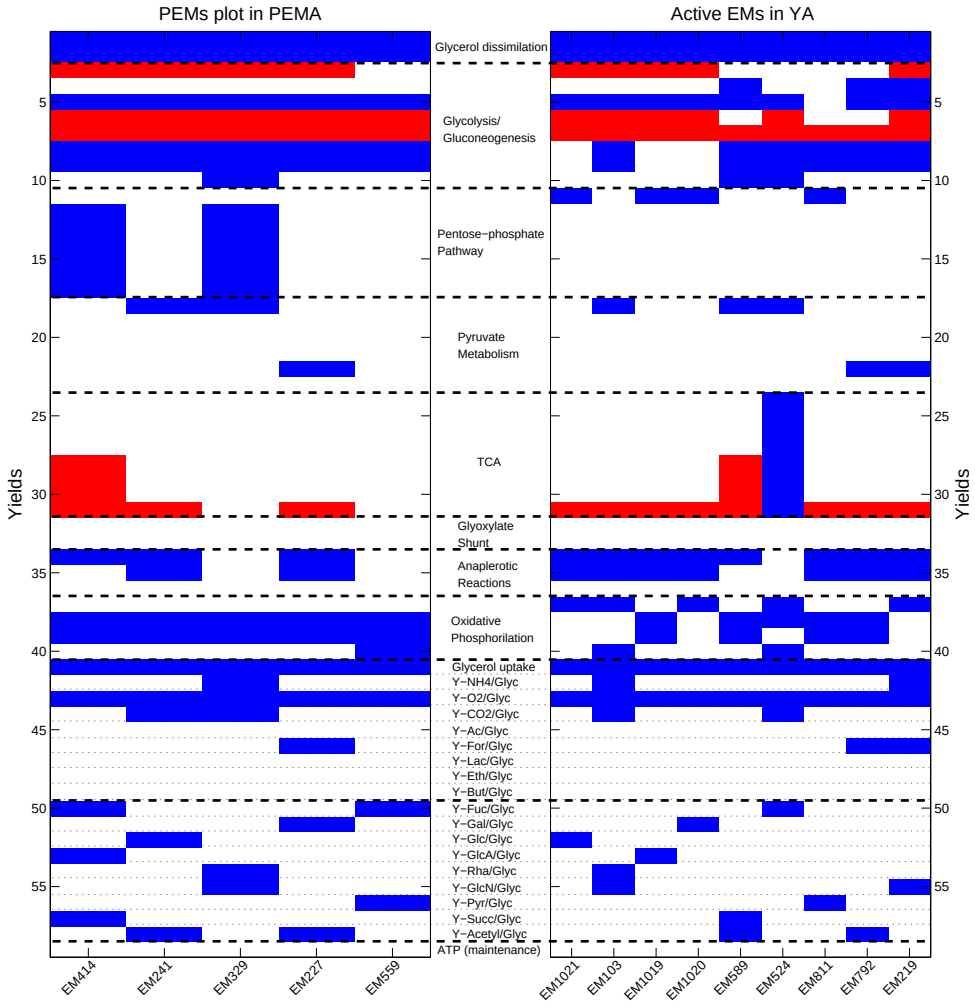

Figure 3.5: Binary plots for the PEMs in PEMA method and active EMs in YA. The blue rectangles represent a yield in the forward direction while the red color represents a yield in the reverse direction. The metabolic pathways are grouped within the dashed lines.

acetyl group.

Some EMs also show TCA activity, in particular EM414 in PEMA and EM524 and 589 in YA. In the case of EM414 and EM589, TCA is only partially activated to generate succCoA, which is the succinyl donor for EPS. EM524 has all the TCA reactions activated, indicating a catabolic activity to supply ATP for EPS synthesis.

3.4 Conclusions

In this work, the identification of a minimal set of active EMs from measured yield data was studied. Yield data requires only extracellular measurements, which are easily available in a standard experimental setting. One can argue if EMs can be discriminated on the basis of extracellular data only, but since the internal fluxes of a metabolic network are strictly constrained by the status of external fluxes (Song et al. 2009) the strategy is meaningful.

The first step in consists in transforming the entire flux space defined by the full set of EMs into the yield space. This space transformation is done by dividing each of the fluxes contained in the elementary modes by a flux of reference, generally the carbon source (glycerol flux in the present work) or alternatively the biomass production flux. The resulting yield space forms a convex hull that should enclose all measured yield values.

Two different methods have been compared to discriminate the minimal set of active EMs. PEMA is a principle component like method that maximizes explained variance. In YA, the EMs are chosen as the minimal number to enclose a given measured yield point, which is at most $n + 1$ EMs in n-dimensional yield space. This subset of EMs is determined by quadratic programing.

The overall results show that PEMA was able to explain 98.1% of measured yield variance with a minimal set of 5 EMs, while YA required 9 EMs to explain 99% of variance. The degree of orthogonality of EMs identified by YA is considerably higher than those identified by PEMA. This is reflected in the structure of EMs selected with the YA method promoting simpler single-substrate single-product EMs while PEMA selects more complex multi-substrate and multi-product EMs.

Chapter 4

Dynamic Modeling of EPS fed-batch production

The process dynamics of exopolysaccharide (EPS) production by the microbial species *Enterobacter* A47 is very complex. Previous studies have shown that the choices of temperature and pH are critical for both biomass growth and EPS synthesis. However, the complex kinetics and transport phenomena make the construction of first-principles models for bioprocess optimization difficult and laborious. An alternative, more cost-effective methodology was explored here for the modeling of fed-batch experiments with varying temperature and pH levels. Empirical and first-principles knowledge (parametric) were combined with artificial neural networks (nonparametric) into hybrid semi-parametric models, in various configurations. The performances of the hybrid models were assessed relative to their effectiveness in fitting a calibration data set of fed-batch experiments, while also being capable to describe data of an independent set. The results show that the hybrid-model configuration has a high impact in the model performance. It was found that the hybrid-models composed by the most detailed mechanistic descriptions coupled to an artificial neural network receiving inputs of temperature, pH, biomass and

substrate concentrations, achieved the best results. Although the incorporation of viscosity into this model improved the calibration error by 9%, the identification predictive capabilities of an independent data set were worse. It was concluded that this approach presented a significant improvement over previous modeling studies based on unstructured kinetics, fostering model-based EPS productivity optimization.

4.1 Introduction

The production of natural polysaccharides have registered an increased demand over recent years due to novel industrial applications, motivated by their unique physicochemical properties, and their potential for substituting petroleum-derived materials (Rehm 2010). Among the various sources of polysaccharides found in nature, exopolysaccharides (EPS) produced by microbial organisms are preferred for industrial exploitation due to several advantages inherent to microbial cultures, such as high growth rates, high yield, productivity, and the possibility to control polymer properties. These properties, which include water retention capacity, emulsifying properties, thickening, gelling and film-forming capacity open up several interesting applications in food, cosmetic and pharmaceutical industries (Freitas et al. 2011a; Kumar et al. 2007). However, a major limiting factor for the exploitation of industrial microbial production is the high cost of commonly used carbon sources (*e.g.* glucose, fructose, sucrose) (Freitas et al. 2010; Kumar et al. 2007; Moreno et al. 1998).

In this work, we focus on EPS production by *Enterobacter* A47, which is a Gram-negative microorganism with high affinity for glycerol. It achieves high growth rates in media containing glycerol byproduct from the biodiesel industry, a low cost carbon source (Freitas et al. 2011b). Typically, EPS production is performed in fed-batch mode, with *Enterobacter* A47 growing in a solution media containing glycerol and ammonia as carbon and nitrogen sources respectively fed along time. This organism was shown to be metabolically versatile, producing EPS composed of fucose, galactose and glucose sugar residues, with acetyl, succinyl and pyruvil substituents (Freitas et al. 2011b). The presence of fucose, a rare sugar, difficult to obtain in nature, confers the polysaccharide a high market value. Despite of the promising characteristics and value of this product, the process still remains largely unoptimized, owing to the success of modeling and optimization studies to design an efficient process.

In a previous study by Torres et al. (2012) a central composite rotatable design

(CCRD) of experiments was employed in combination with a response-surface method (RSM) to optimize the process. The quadratic RSM described the temperature and pH dependency of several response variables (specific growth rate, specific EPS productivity, maximum EPS concentration and relative fractions of EPS monomers). Although the results are compelling, there are some limitations in this modeling approach regarding the process optimization. The temporal dimension is not taken into account, thus excluding any attempt for dynamic optimization of nutrients feeding, pH or temperature dynamic control. Moreover, in some cases the quadratic model may not be adequate to explain the impact of some parameters such as temperature, especially when their curve displays a nonsymmetrical curvature (Baş and Boyacı 2007).

Dynamic modeling with unstructured kinetics is probably the most widely adopted method for bioprocess optimization. In a previous study (Torres et al. 2011), we have used simple Monod-type kinetics to describe biomass growth and the Luedeking-Piret equation to describe EPS synthesis. A similar approach has been used for modeling of the production of xanthan gum, an EPS with similar physico-chemical properties as *Enterobacter* A47's EPS (Faria et al. 2010; García-Ochoa et al. 1995). In the present study, we further pursued the goal of developing a simple unstructured kinetic model, but we failed to obtain a model with acceptable global performance. For this reason, we attempted to improve the unstructured model further using hybrid modeling techniques. In hybrid modeling, prior available process knowledge (*e.g.* knowledge from first-principles) is combined with data-driven approaches that account for process phenomena, which are difficult to model mechanistically (Chen et al. 2000; Psichogios and Ungar 1992). Hybrid modeling approaches have already been applied to model the production of exopolysaccharides, xanthan gum in specific (Zabot et al. 2011), of pharmaceutical important products like the human fusion glycoprotein IgG1-IL2, produced by Baby Hamster Kidney cell lines (Teixeira et al. 2005), or the production of the enzyme inolinase by *Kluyveromyces marxianus* NRRL

Y-7571, which has applications in the food industry (Menküc et al. 2008).

4.2 Methods

4.2.1 Exopolysaccharide production and experimental data

The data set used in this work was obtained from a design of experiments preformed by Torres et al. (2012) to assess the impact of temperature and pH on several performance measurements such as the biomass specific growth rate and EPS productivity. Hence, thirteen fed-batch experiments were performed with temperature values ranging from 15.9 ^{o}C to 44.0^{o}C, and pH values ranging from 5.6 to 8.4. Each experiment contains off-line measurements for the whole bioreactor run. Table 4.1 contains the record of the temperature and pH values used on each experiment. The experimental setup for exopolysaccharide production by *Enterobacter* A47 consisted on 2L bioreactors containing a modified medium E* supplemented with glycerol (Freitas et al. 2009). The bioreactor was operated in batch mode until the exhaustion of the nitrogen source, followed by a fed-batch operation with fresh medium being fed at a constant volumetric rate (4.5 mL h^{-1}). Further details about bioreactor operation and the design of experiments can be found elsewhere Torres et al. (2012).

4.2.2 Unstructured dynamic model

The dynamics of biomass (X), glycerol (S), ammonium (N) and EPS concentrations (P) are described by the following material balance equations assuming a perfectly mixed stirred tank reactor operated in fed-batch mode:

$$\frac{dX}{dt} = (\mu - K_d - D)\ X \tag{4.1}$$

$$\frac{dS}{dt} = -v_S\, X + D\,(S_f - S) \tag{4.2}$$

$$\frac{dN}{dt} = -v_N\, X + D\,(N_f - N) \tag{4.3}$$

$$\frac{dP}{dt} = v_P\, X - D\, P \tag{4.4}$$

$$\frac{dV}{dt} = F \tag{4.5}$$

The dilution rate (D) is represented by the quotient of the volumetric feed (F) given by Equation 4.5 with the reactor volume. The value of F is zero during the batch phase, changing to a fixed rate of 4.5 mL h^{-1} in the fed-batch phase. The glycerol and ammonia feeding concentrations are, respectively $S_f = 200\text{g L}^{-1}$ and $N_f = 0.90\text{g L}^{-1}$. The specific rates μ, K_d, v_S, v_N and v_P refer to biomass growth, biomass death, glycerol consumption, ammonium consumption and EPS synthesis, respectively. The specific cell growth rate is given by a Monod model with glycerol (S) and ammonia (N) limitation:

$$\mu = \mu_{max}\, \frac{S}{S + K_S}\, \frac{N}{N + K_N} \tag{4.6}$$

with μ_{max}, K_S and K_N the maximum specific growth rate and half saturation constants for glycerol and ammonium, respectively. In this work, the product formation is partly uncoupled from biomass growth and, for this reason the overall glycerol consumption, comprehends a term for the glycerol taken up for biomass synthesis, EPS synthesis and maintenance (m_S):

$$v_S = \frac{\mu}{Y_{xs}} + \frac{v_P}{Y_{ps}} + m_S \tag{4.7}$$

where v_S is the specific consumption rate of glycerol, v_P is the specific EPS production rate, $Y_{X/S}$ is the true yield of biomass formation from substrate, $Y_{P/S}$ the true yield

of product formation from substrate and m_S the maintenance coefficient. The yield constants are described in mass units.

The ammonium uptake was assumed to be associated with biomass synthesis only, given that nitrogen is found in EPS chemical composition only in residual amounts:

$$v_S = \frac{\mu}{Y_{xn}} \tag{4.8}$$

where v_N represents the specific ammonia consumption rate and $Y_{X/N}$ is the yield of biomass formation from ammonia consumption (mass units).

To describe the EPS specific synthesis rate, v_P, a Monod-type equation was derived, such that EPS production is enabled after the ammonia concentration decreases to values close to zero, which is a common behavior in this culture:

$$v_P = v_{P,max} \frac{K_P}{N + K_P} \tag{4.9}$$

A maximum specific productivity is defined by $v_{P,max}$, with K_P representing the concentration of ammonia when v_P is half the value of $v_{P,max}$.

As generally assumed and shown in Torres et al. (2012), biomass growth rate and EPS productivity are dependent on temperature and pH. Most of the models in the literature that relate μ_{max} with temperature and pH, lack obvious biological significance of the models' parameters and have significant structural correlations between parameters, which incur in estimation problems (Rosso et al. 1995). For this reason the cardinal temperature and pH (CTP) model proposed by Rosso et al. (1995) was adopted in this work, whose formulation avoids the aforementioned problems. The model was successfully applied to describe changes of specific growth rate related to changes in temperature and pH, in exopolysaccharide production by *Pseudomonas oleovorans* (Freitas et al. 2010). The CTP model is based upon the assumption that

both temperature and pH have independent effects in μ_{max}, following the equation:

$$\mu_{max}(T, pH) = \mu_{opt}\, \tau_{\mu}(T)\, \rho_{\mu}(pH) \tag{4.10}$$

where $\tau_{\mu}(T)$ is a function of temperature and $\rho_{\mu}(pH)$ a function of pH with μ_{opt} the optimum specific growth rate. The application of the CTP model was also extended to the prediction of $v_{P,max}$ with temperature and pH, due to the mathematical similarity between the EPS specific productivity rate and Equation (4.6). This way, an expression similar to Equation (4.10) can be derived:

$$v_{P,max}(T, pH) = v_{P,opt}\, \tau_{v_P}(T)\, \rho_{v_P}(pH) \tag{4.11}$$

The functions $\tau_{v_P}(T)$ and $\rho_{v_P}(pH)$ are analogous and have a similar meaning to the terms specified in Equation (4.10), but with specific parameter values for either the specific growth rate and specific EPS production. The temperature and pH terms of the CTP model are defined in the following equations:

$$\tau(T) = \frac{(T - T_{max})(T - T_{min})^2}{(T_{opt} - T_{min})[(T_{opt} - T_{min})(T - T_{min}) - (T_{opt} - T_{max})(T_{opt} - T_{min} - 2T)]} \tag{4.12}$$

$$\rho(pH) = \frac{(pH - pH_{min})(pH - pH_{max})}{(pH - pH_{min})(pH - pH_{max}) - (pH - pH_{opt})^2} \tag{4.13}$$

The parameters T_{min} and T_{max} are, respectively, the minimum and maximum temperature values within which microbial growth and EPS production can exist, that is, μ_{max} and $v_{P,max}$ are positive and different from zero. The same applies for the interpretation of pH_{min} and pH_{max} which accordingly delimit the pH range where the calculated kinetic rates have values greater than zero. The parameters T_{opt} and pH_{opt} correspond to the optimum values of temperature and pH, respectively. When $T = T_{opt}$ and $pH = pH_{opt}$, it is verified that $\mu_{max} = \mu_{opt}$ (or $v_{P,max} = v_{P,opt}$). For

the pH function (Equation (4.13)), the symmetry hypothesis was assumed, that is:

$$pH_{min} = \alpha\, pH_{opt} - pH_{max} \tag{4.14}$$

with $\alpha = 2$ for symmetric curves. According to Rosso et al. (1995), in cases were no structural correlation occurs between the parameters for the temperature profiles in the CTP model (Equation (4.12)), the symmetry assumption can be adopted for the pH profile, since it is known to avoid structural correlations between the parameters in Equation (4.13).

The empirical parameters contained in Equations 4.12 and 4.13 (μ_{opt}, $v_{P,opt}$, T_{min}, T_{max}, T_{opt}, pH_{min}, pH_{max}, pH_{opt}) were estimated by non-linear regression, independently from the whole system of mass balances specified by Equations 4.1-4.4. The data used in this identification problem was obtained from the design of experiments implemented by Torres et al. (2012), where values for specific growth rate and EPS productivity was available for different values of temperature and pH. A MATLAB program was written in order to minimize the sum of squared errors (SSE), employing the Levenberg-Marquardt algorithm (lsqnonlin MATLAB function). The Jacobian matrix obtained at the optimal solution was used to calculate an approximation to the Hessian matrix, which enabled the determination of the covariance matrix and 95% confidence intervals.

The identified CTP model was integrated with the mechanistic equations for biomass growth and EPS synthesis rate to complete the mechanistic model. This model was then integrated in the material mass balances to represent the bioreactor dynamics.

The model of differential equations was integrated using a $4^{\text{th}}/5^{\text{th}}$ order Runge-Kutta solver (ode45 MATLAB function) and the concentrations predicted by the model were compared to the off-line measurements. The kinetic parameters' estimation and computation of confidence intervals was performed using the methodology

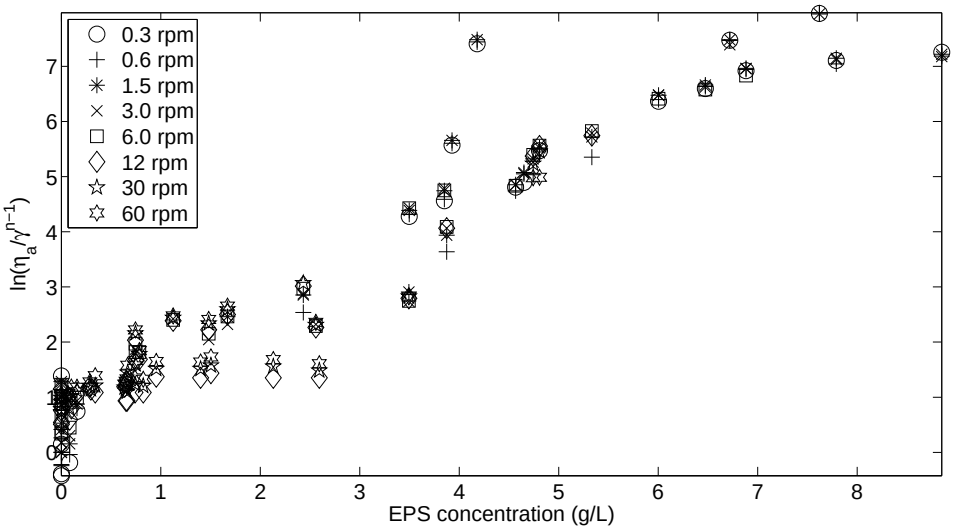

Figure 4.1: Correlation between EPS concentration and a viscosity dependent variable. Each symbol corresponds to a given shear rate value, representing the values 0.3, 0.6, 1.5, 3, 6, 30 and 60 rpm. The viscosity dependent variable represented in the y axis has a fixed n value of 0.81.

described above for the CTP model identification.

4.2.3 Correlation between viscosity and EPS concentration

Culture broth viscosity measurements were collected at discrete time instants during each bioreactor run. The measurements were performed at room temperature and shear rates ranging from 0.3 rpm to 60 rpm as described in (Freitas et al. 2010). The culture broth has a characteristic shear-thinning behavior which can be described by the power-law (Doran 1995):

$$\eta_a = K\,\gamma^{n-1} \tag{4.15}$$

where η_a represents the apparent broth viscosity, K the consistency index, $\gamma$the shear rate and n the fluid behavior index. The parameters K and n characterize the power-law fluids and in the case of non-Newtonian fluids $n < 1$.

Since the direct application of the power-law does not provide a time-dependent description of viscosity, a mathematical relationship was developed in order to

describe the dependence of viscosity on the EPS concentration. The assumption of this relationship is reinforced by empirical verification and also supported by other studies (Alves et al. 2010b), where EPS from *Enterobacter* A47 has been found to be one of the main contributors to viscosity increase in the culture broth. For this study, a dataset for all bioreactor experiments containing viscosity measurements for each shear rate was assembled. Since the fluid-flow behavior changes with time (*e.g.* from Newtonian profiles to non-Newtonian characteristics), each time sample can be described by the power-law in terms of its parameters K and n. It was verified that, in the case of fixed n values, a fairly linear relationship between EPS and the viscosity was observed, which can be described by the following equation:

$$\ln \frac{\eta_a}{\gamma^{n-1}} = a\,P + b \tag{4.16}$$

with a being a correlation variable between the EPS concentration (P) and the left-hand term of the equation, and b the point where the line crosses the y axis (Figure 4.1). The fixed value of n was chosen such that the standard deviations for all K determined for each time instant are minimized using MATLAB (*fmincon* function), at the minimum $n = 0.81$.

4.2.4 Hybrid model formulation

Hybrid modeling structures can generally be classified as parallel or serial (Oliveira 2004; Stosch et al. 2014). In the former case, a nonparametric model is used to compensate eventual inaccuracies of a complete mechanistic model. This strategy is typically employed when the mechanistic model is not sufficiently accurate for model-based applications. In serial structures, mechanistic parts are complemented with information from a nonparametric model. The hybrid model adopted in this work has an overall serial structure, but follows a flexible modular structure as depicted in Figure 4.2. The backbone of the model corresponds to the material

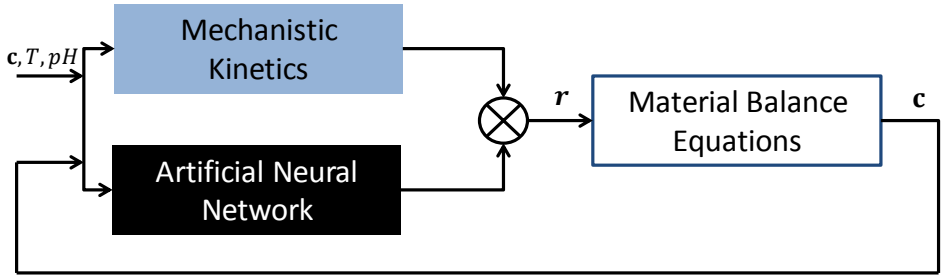

Figure 4.2: Hybrid model structure

balances Equations 4.1-4.4. The specific kinetic rates are represented by Equations 4.6-4.11 are modeled in parallel with a nonparametric part. These upper modules are an inherent part of the material balance equations, which represent known kinetic expressions. Each of the known mechanistic kinetic rates is multiplied by an unknown nonparametric function, which has the job to "correct" inaccuracies of the known kinetic term. This operation takes the following general form:

$$\mathbf{r}(\mathbf{c}) = \mathbf{r_{known}}(\mathbf{c})\,\mathbf{r_{correction}}(\mathbf{c}, \mathbf{w}) \tag{4.17}$$

with $\mathbf{r_{known}}(\mathbf{c})$ corresponding to the known kinetic expressions and $\mathbf{r_{correction}}(\mathbf{w})$ some unknown corrective function that needs to be identified from data, $\mathbf{w}$ representing a set of unknown parameters and $\mathbf{c}$ the vector of concentrations and/or other input variables (Oliveira 2004). The corrective function is a nonparametric model, which represents a class of models with loose structure and arbitrary number of parameters such as neural networks, wavelets, regression models, etc (Stosch et al. 2014). In this work, we have adopted a neural network with three layers, an input, hidden and output layer. The systematic investigation of network structures was pursued in order to define the optimal number of hidden nodes. The best structure was determined through cross validation, with the best performing topology corresponding to the lower modeling error.

For regression problems the activation functions of the input and output layer

Table 4.1: Segregation of experimental data among the training, validation and test sets.

Training set	**Validation set**	**Test set**
$T = 30.0^{o}$C, $pH = 7.0$	$T = 23.0^{o}$C, $pH = 7.5$	$T = 44.0^{o}$C, $pH = 7.0$
$T = 23.0^{o}$C, $pH = 6.5$	$T = 20.0^{o}$C, $pH = 6.0$	$T = 30.0^{o}$C, $pH = 5.6$
$T = 15.9^{o}$C, $pH = 7.0$	$T = 30.0^{o}$C, $pH = 7.0$	$T = 30.0^{o}$C, $pH = 8.4$
$T = 40.0^{o}$C, $pH = 8.0$		$T = 30.0^{o}$C, $pH = 7.0$
$T = 40.0^{o}$C, $pH = 6.0$		
$T = 20.0^{o}$C, $pH = 8.0$		

are linear while that of the hidden layer is usually the tangent hyperbolic function (Bishop 1995), *i.e.*:

$$\mathbf{r_{correction}}(\mathbf{c}, \mathbf{w}) = \mathbf{w_2} \tanh(\mathbf{w_1}\,\mathbf{c} + \mathbf{b_1}) + \mathbf{b_2} \tag{4.18}$$

In Equation (4.18), $\mathbf{c}$ is the vector of concentrations and/or other input variables, $\mathbf{w_1}$ represents the weight matrix that describes the connection of the input layer to the hidden layers, $\mathbf{w_2}$ the weight matrix with respect to the connection between the hidden layer and output layer. The vectors $\mathbf{b_1}$ and $\mathbf{b_2}$ represent the vectors of bias parameters.

Several variants of the hybrid structure depicted in Figure 4.2 were constructed with varying detail in the parametric model description. For the design of the nonparametric part, it was taken into consideration which factors impact most in the determination of the process kinetic rates. It was assumed, based on general knowledge of the mechanistic relationships that kinetic rates must depend on substrate concentrations and, for the case of specific growth rate and specific EPS production, temperature and pH were also chosen as inputs to the artificial neural network model.

The process data was split into three parts: a training, validation and test set (Table 4.1). The training data was used for parameter identification, while the validation data set was used for cross-validation. The test set is used to assess the generalization capabilities of the models. The hybrid models in this work were

developed with the intent to support model-based product titer optimization. With that in mind, we chose high performance batches (in terms of product synthesis) for the training and validation sets, such that the models can be used for process regions in which high product titers are observed.

The ANN parameter identification was performed by minimizing a weighted least squared residual function, coupled with the application of the sensitivities approach, whose details can be found elsewhere (Oliveira 2004). The training was at least 20 times initialized from random parameter values. The set of parameters that yielded the minimum weighted least squared error (MSE) for the validation set were chosen for each model structure. Different model structures, *i.e.* network topologies, were systematically investigated in order discriminate the optimal number of nodes in the hidden layer of the ANN of each hybrid model. The most simple but adequately performing network structure was chosen for each hybrid model, balancing the model performance against model complexity.

Parameter identification was performed by minimizing a weighted least squared residual function:

$$\min_{\mathbf{w}} \left\{ \frac{1}{N} \left[\sum_{i=1}^{M} \sum_{j=1}^{N} \left(\frac{C_{i,j}^{exp} - C_{i,j}}{s_i} \right)^2 \right] \right\} \tag{4.19}$$

with N representing the total number of observations for all variables, M the number of variables, C_i^{exp} the experimental concentration of variable i, C_i the model predicted concentration of i and s_i the standard deviation for the variable i. The sensitivities approach was employed for the minimization (Oliveira 2004). To avoid over-fitting, the training was stopped at the minimum weighted mean least squared error (MSE) value calculated for the validation set. The equation of the MSE reads:

$$MSE = \frac{1}{N} \left(\sum_{i=1}^{M} \sum_{j=1}^{N} \left[\frac{C_i^{exp} - C_i}{s_i} \right]^2 \right) \tag{4.20}$$

4.2.5 Hybrid model with CTP model kinetics

The mechanistic descriptions for biomass growth and EPS production can be coupled to the CTP model to account for the effects of temperature and pH in the bioreactor. This way a complete mechanistic model can be formulated and placed in parallel with an ANN in order to correct inaccurate descriptions. For the design of the nonparametric part, it was taken into consideration which factors impact the most on the determination of the process kinetic rates.

Two different mechanistic configurations were tested and named Hybrid Model 1 (HM1) and Hybrid Model 2 (HM2), which differ on the way the ANN interacts with the mechanistic part. In HM1 the ANN structures tested had two outputs for the description of the glycerol and ammonia consumption kinetics, while the biomass growth rate and EPS synthesis rate were estimated by the CTP model entirely. Regarding the inputs to the ANN, two different configurations were tested, namely with S and N in one case, and S, N, T and pH in another case.

The configuration of the mechanistic part in HM2 assumed the CTP model to estimate the biomass growth rate and EPS production rate, and Equations 4.7 and 4.8 to estimate the kinetics of glycerol and ammonia consumption. In this case, all the ANN configurations tested had 5 output rates: two ANN outputs to multiply to each CTP model expression (μ_{max} and $v_{P,max}$), and the remainder three ANN outputs as additive correction factors to each yield parameter in the mechanistic formulation ($Y_{X/S}$, $Y_{P/S}$ and $Y_{X/N}$). The ANN structure used in the HM2 had 5 input nodes corresponding to X, S, N, T and pH.

In all cases, the estimated states values of S, N (and X where applicable), were used as inputs to the networks since they are readily obtained when integrating the material balances. In case of T and pH, the measured inputs were provided at a sampling rate of 0.02 h, which is identical to the time differences used for the numerical integration.

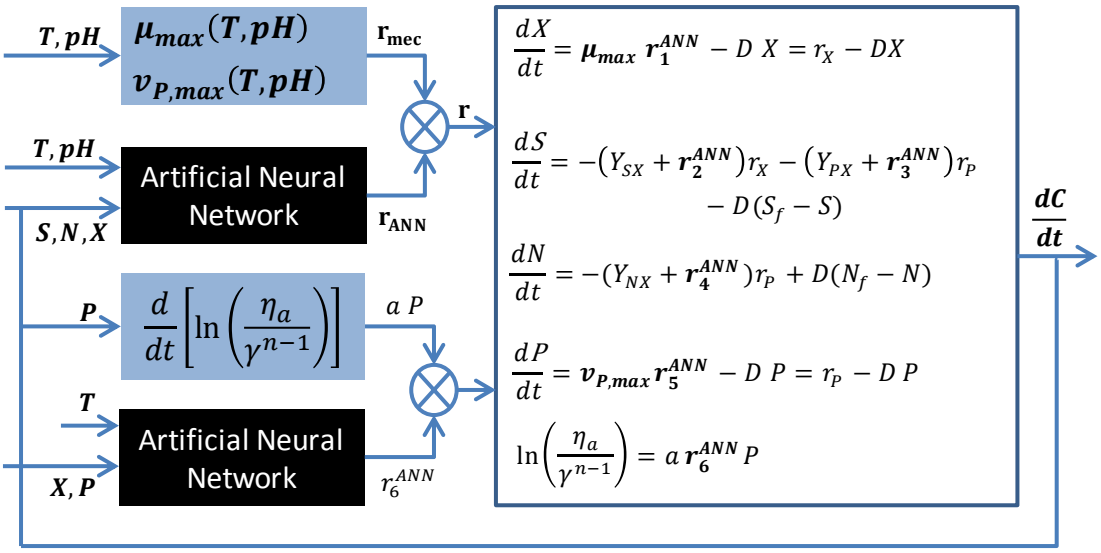

Figure 4.3: Representation of the HM3 modeling scheme, where the parametric module containing the correlation between viscosity and EPS concentration is added to the previously trained HM2 scheme. This added parametric equation is identified with the aid of a second ANN placed in parallel with temperature, biomass and EPS concentrations added as inputs.

4.2.6 Hybrid model with CTP model kinetics and viscosity knowledge

Knowledge from the culture broth viscosity was introduced into an upgraded form of the HM2 model by combining a second mechanistic module containing the parametric equation defined in Section 4.2.3 (Equation (4.16)). This module was connected to a second ANN destined to correct the inaccuracies of the parametric viscosity equation. Since temperature, biomass and EPS concentration are factors known to impact on the culture broth's viscosity, they are included in the input space of the second ANN. One output node was defined in this network arrangement. This model structure is termed HM3 and is illustrated in Figure 4.3.

The weights of the first ANN were pre-identified using the well-known back-propagation method (Oliveira 2004) and thereupon the parameters of the complete model were re-identified.

4.2.7 Hybrid model without CTP model

Several hybrid model configurations were constructed without the CTP model. Here, three hybrid model configurations termed HM4, HM5 and HM6 were proposed. The model HM4 lacks a mechanistic kinetic description and therefore, all the specific rates in the material balances are estimated by the ANN. This network was built with an input space represented by the state variables S, N and the off-line measurements of T and pH, and an output space of four kinetic rates for each state variable (X, S, N and P). Various hidden nodes were tested to assess the optimal network structure.

The models HM5 and HM6 were built with the same ANN composition but differed by the knowledge present in the mechanistic part. The kinetic rate expressions for biomass growth and EPS synthesis in HM5 were estimated with the multiplication of glycerol concentration by the respective ANN outputs, while glycerol and ammonia consumption rates were solely determined by the ANN. The model HM6 is an extension of the HM5 mechanistic part, with biomass concentration multiplied in all kinetic expressions, glycerol and ammonia concentrations multiplied in the biomass kinetic rate, and glycerol concentration multiplied by the product synthesis rate.

4.3 Results and discussion

4.3.1 CTP model identification

The parameters contained in the specific biomass growth rate and EPS productivities of the CTP model were identified by nonlinear regression as described in the methods section. The identified parameter values and their respective confidence intervals are organized in Table 4.2. The CTP model describes a characteristic bell-shaped curvature (Figure 4.4), with a maximum point which is defined by the parameter μ_{opt} (or $\upsilon_{P,opt}$). The remaining parameters define the model's extreme values, above which the specific rates have zero value. The CTP model was able to describe the

Table 4.2: CTP model kinetic parameters and respective 95% confidence intervals. Parameter T_{min} manually tuned in both CTP functions

Parameters relative to the μ_{max} function		
$T_{min} = 10^{o}C$	$T_{max} = 44.1 \pm 0.55^{o}C$	$T_{opt} = 37.2 \pm {}^{o}C1.92$
$pHmin = 4.50$	$pH_{max} = 9.03 \pm 0.66$	$pH_{opt} = 6.77 \pm 0.37$
$\mu_{opt} = 0.45 \pm 0.09\,h^{-1}$		
Parameters relative to the $v_{P,max}$ function		
$T_{min} = 10^{o}C$	$T_{max} = 47.5 \pm 10.6^{o}C$	$T_{opt} = 32.2 \pm 7.63^{o}C$
$pHmin = 4.99$	$pH_{max} = 10.40 \pm 5.70$	$pH_{opt} = 7.70 \pm 2.31$
$v_{P,opt} = 0.55 \pm 0.26\,h^{-1}$		

specific growth rate fairly well, exhibiting low confidence intervals. The identification for the EPS model yielded greater confidence intervals, which is likely attributed to the nature of experimental data for EPS specific productivities. In both cases T_{min} was found to be highly sensitive to the modeling error and therefore, this parameter was tuned manually.

4.3.2 Hybrid model with CTP model kinetics

Several hybrid model structures with varying degrees of mechanistic knowledge were investigated. Details on the performance and model structures are shown in Table 4.3. The structures HM1 and HM2 use the CTP model to describe the effects of temperature and pH in μ_{max} and $v_{P,max}$. The structures HM1 and HM2 use the CTP model to describe the effects of temperature and pH in μ and v_P, and two ANN outputs to estimate the volumetric rates of glycerol and ammonia concentration. Two scenarios were tested for this model: two ANN inputs (glycerol and ammonia concentration) and four ANN inputs (temperature, pH, glycerol and ammonia concentrations). The performance in terms of MSE seems distinct. A better fit and predictive performance were achieved when temperature and pH were included as inputs to the ANN, which suggests that the glycerol and ammonia consumption

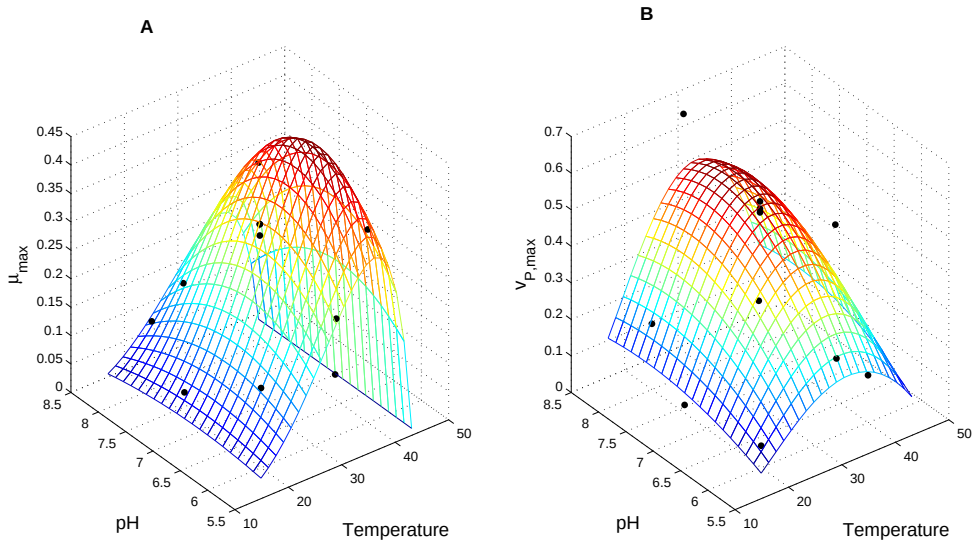

Figure 4.4: Hybrid model incorporating viscosity knowledge. A. Plot of μ_{max} dependency with T and pH. B. Plot of $v_{P,max}$ dependency with T and pH

rates are also pH and temperature dependent. This outcome makes sense, since these rates depend on biomass concentration, whose growth is determined by temperature and pH.

In the HM2, parametric detail, in the form of yield values and Monod constants from a previous study (Torres et al. 2011), was added to the glycerol and ammonia consumption rates. In this study, the kinetics were discriminated from data of a fed-batch experiment, in which EPS was produced under constant temperature and pH (30ºC and 7.0 respectively). The hybrid model was trained with the inclusion of yields for the quantities of glycerol and ammonia consumed per biomass produced (Y_{SX} and Y_{NX}), and of glycerol consumed per EPS produced (Y_{SP}). These values reflect the kinetic relationships for the consumption of glycerol and ammonia, assuming that only ammonia is consumed for biomass formation. In order to account for temperature and pH changes in the experiments, ANN estimates were added to the yield values. Also the estimated biomass concentration was included as an input to the nonparametric part of the model. Although the results show only a slight improvement in the calibration and validation error when compared to the model

HM1 with the four ANN inputs, a significant improvement was observed in the test set. The change in the structure of the hybrid model seems to be important to improve the predictive capability of the model. The latter seems to explain some variation that cannot be well described by the Monod kinetics, which essentially depend on glycerol and ammonia concentration.

4.3.3 Hybrid model with CTP model kinetics and viscosity dependency

Viscosity correlates positively with EPS concentration, as discussed above. Along with this factor, temperature and biomass concentration are known to impact on media broth viscosity (Al-Asheh et al. 2002; Alves et al. 2010a,b). The methodology exposed in Section 4.2.3 was integrated with the hybrid model HM2 in order to enable the prediction of viscosity during the fermentation time. Hence HM3 represents the model re-identification after the inclusion of Equation (4.16) to describe viscosity. The model training results is displayed graphically in Figure 4.5 and the validation and test results in Figure 4.6. When comparing to the HM2 case, the incorporation of viscosity seems to improve the model calibration, due to a lower training error, while also improving the validation MSE. The model did nonetheless perform worse in the description of the test set.

The central point in the experimental setup, characterized by $T = 30^{o}$C and $pH = 7.0$ is well described in every modeling partition, with the exception of EPS profile in the validation partition, where the points located between the 40 and 60 hours deviate from the predicted data. Except for the second training batch ($T = 23^{o}$C and $pH = 6.5$), it can be observed that the increase in viscosity is concomitant with EPS production, thus mimicking the process behavior effectively. Nevertheless, a particular case is observed for the experiment performed at $T = 40^{o}$C and $pH = 6.0$, where the model fails to accurately estimate the viscosity determined experimentally.

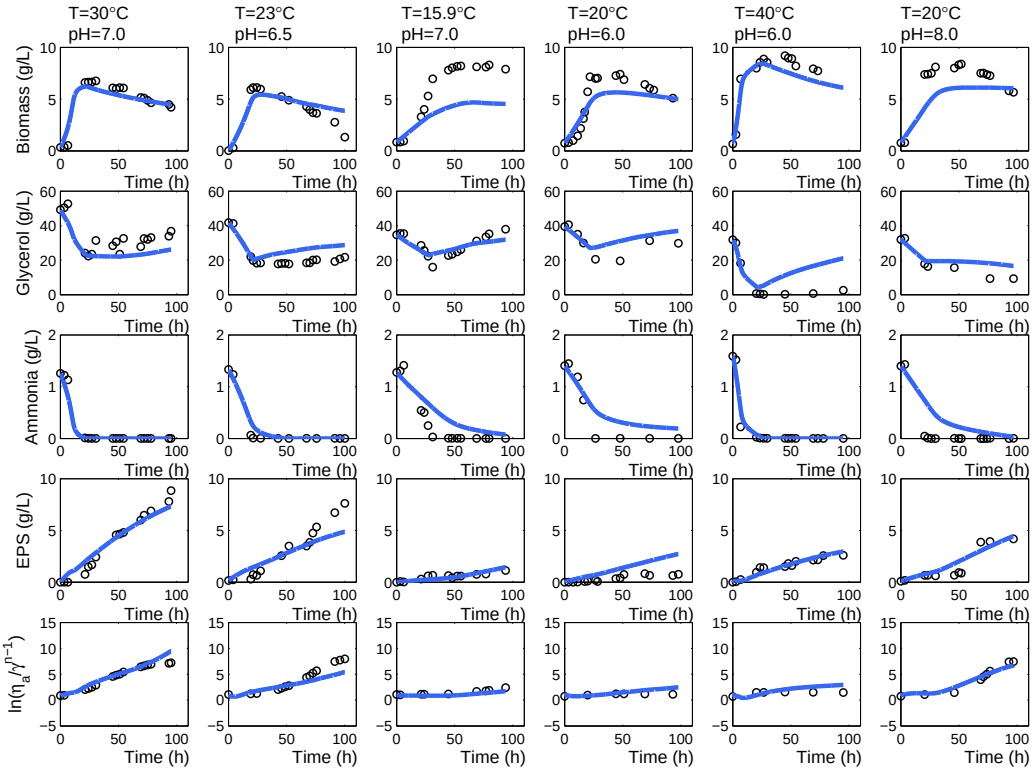

Figure 4.5: Hybrid modeling results for the training partition, validation and test sets for the hybrid model structure HM6. The rows correspond to the results of each state variable of the process as described in the picture. The circles represent experimental data while the solid line the model predictions.

The effects of temperature and pH are generally well described by the model. The maximum specific productivities for biomass and EPS, which are estimated by the CTP model, decrease when temperature and pH deviate from the optimum values. The latter are also the closest to the central point in the design of experiments. A special case is the training experiment represented by $T = 15.9^{o}$C, where the model underestimated the stationary biomass growth. Under these conditions, the CTP model estimates a specific growth rate for biomass of 0.034 h^{-1}, a fact that must be connected to the low growth rate of biomass predicted by the model.

It has to be noted that the generalization properties of the model should be assessed with respect to the limits of the input values of the calibration set. For this reason, it is not surprising that the model failed to describe the experiment with

$T = 44$ºC, displayed in the test partition.

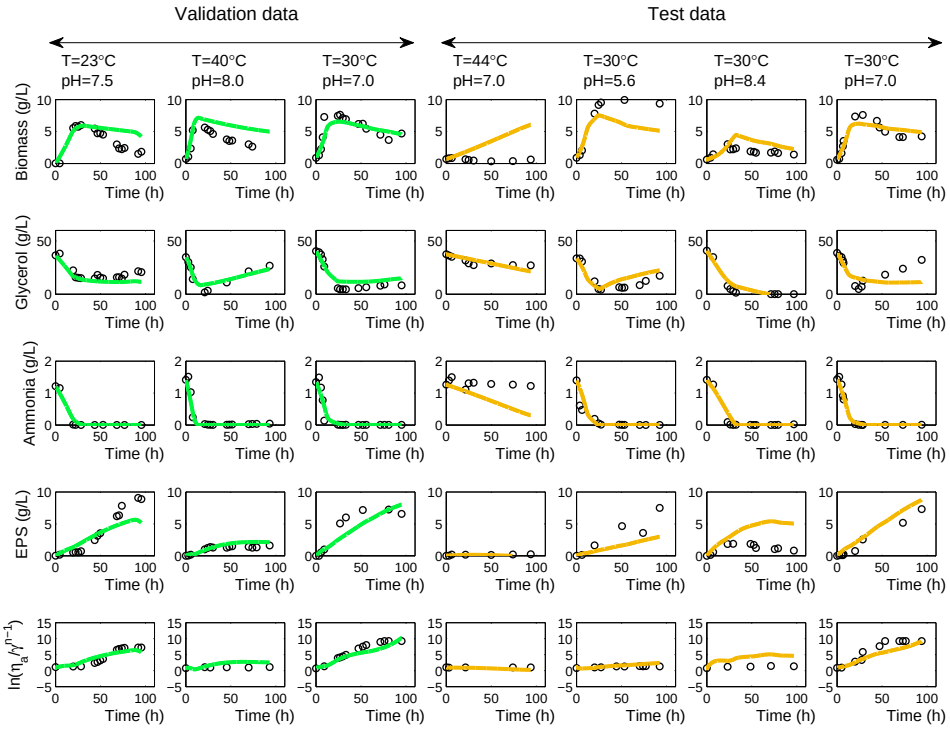

Figure 4.6: Hybrid modeling results for the training partition, validation and test sets for the hybrid model structure HM6. The rows correspond to the results of each state variable of the process as described in the picture. The circles represent experimental data while the solid line the model predictions.

Figure 4.7 shows the viscosity profiles of two experiments from the test set. The plot A represents an experiment with high EPS production ($T = 30$ºC, $pH = 7.0$) and an experiment with low EPS production ($T = 30$ºC, $pH = 5.6$). As expected, a higher viscosity was achieved in the experiment with higher EPS productivity. It can also be seen in both viscosity profiles that the model predictions become worse at later stages of the fermentation, in particular after 70h in Figure 4.7-A. This pattern can also be seen in most batches that achieved high EPS concentrations, as can be seen by Figures 4.5 and 4.6. This fact may be explained by the higher difficulty in the collection of viscous samples from the bioreactor. Moreover, it is known that at later stages of EPS production, cells form intricate chemical associations with EPS (Alves et al. 2010b), which have a negative impact on the efficacy of biomass and

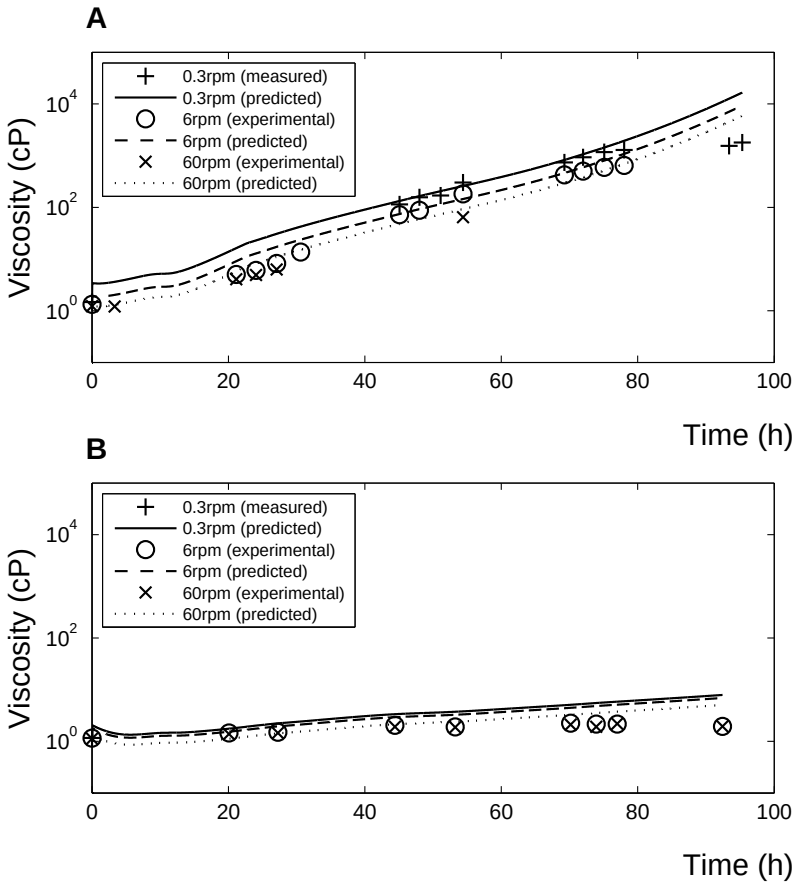

Figure 4.7: Viscosity profiles for two experiments for several shear rate values. The continuous lines represent the model predictions while the markers refer to experimental values. Plot A: Fed-batch experiment with T=30°C and pH=7.0 in the test partition. Plot B: Fed-batch experiment with T=30°C and pH=5.6 in the test partition.

EPS separation processes.

4.3.4 Hybrid model without CTP model kinetics

Three hybrid model structures were built without parametric kinetics, referred to as HM4, HM5 and HM6 (Table 4.3). In all cases, pH and temperature were included as inputs to the ANN, as well as estimated glycerol and ammonia concentrations, which are known to influence the kinetics. The models were trained with different numbers of hidden nodes and it was observed that generally the structures with four

nodes performed better.

The model HM4 has the simplest structure, with each kinetic rate being estimated solely by the nonparametric model. In HM5 the biomass rate comprises multiplications of glycerol and ammonia concentration, the EPS synthesis rate multiplications with glycerol, such that these rates are bound when no glycerol or ammonia is present. In the third case (HM6), it is in addition assumed that biomass is a catalyst for the reactions.

The lower calibration error ($MSE = 0.1823$) was obtained with model HM4 and 3 hidden nodes. However, in this case, the test set had a greater error value ($MSE = 2.5163$). Increasing the number of nodes in the hidden layer did however improve the fit of the test set while decreasing the training set accuracy. The models HM5 and HM6 performed worse than HM4, leading to the conclusion that increasing the detail in the kinetic rates significantly decreases the model performance.

4.3.5 Unstructured dynamic model

A dynamic unstructured model was built based on mass balance equations and Monod-type kinetic expressions including the coupled CTP model. Although single fed-batch experiments could be successfully described, it was verified that the addition of more experiments to the calibration set significantly degraded the results. The results of MSE values for the training and validation sets (according to Table 4.1) were 604 and 374 respectively. Thus, it might be hypothesized that the major source of error in the model description is a large amount of batch-to-batch variation that is not captured by the unstructured model.

4.3.6 Prediction power

The overall predictive power of the models can be assessed based on the modeling error obtained for the validation and test sets. This criterion is important since these data partitions reflect independent data sets relative to the calibration data. Models

Table 4.3: Synthesis of results obtained for each hybrid model structure trained. The number of parameters in HM3 model reflects only the parameters that were used to train the viscosity component of the model. NH and NP represent ANN number of hidden layers and parameters respectively.

Hybrid model structure	**ANN inputs**	**NH**	**NP**	**MSE train**	**MSE valid**	**MSE test**
Hybrid structures with CTP model kinetics						
HM1: $\mathbf{r(c)} = [\mu_{max}\, r_1^{ANN},\, r_2^{ANN},\, r_3^{ANN},\, v_{P,max}\, r_4^{ANN}]^T$	S, N, T, pH	4	40	0.2893	0.3231	0.8368
	S, N	4	32	0.5285	0.5659	4.8281
HM2: $\mathbf{r(c)} = [\mu_{max}\, r_1^{ANN}, -(Y_{sx} + r_2^{ANN})\,\mu - (Y_{sp} + r_3^{ANN})\, v_P, -(Y_{NX} + r_4^{ANN})\,\mu, v_{P,max}\, r_5^{ANN}]^T$	X, S, N, T, pH	5	60	0.2786	0.2961	0.4184
Hybrid structure with CTP model kinetics and viscosity correlation						
HM3: $\mathbf{r(c)} = [\mu_{max}\, r_1^{ANN}, -(Y_{sx} + r_2^{ANN})\,\mu - (Y_{sx} + r_3^{ANN})\, v_P, -(Y_{NX} + r_4^{ANN})\,\mu, v_{P,max}\, r_5^{ANN}, P\, r_6^{ANN}]^T$	X, S, N, T, pH	4	21	0.2530	0.2616	0.5258
Hybrid structures without CTP model kinetics						
HM4: $\mathbf{r(c)} = [r_1^{ANN},\, r_2^{ANN},\, r_3^{ANN},\, r_4^{ANN}]^T$	S, N, T, pH	3	31	0.1823	0.6465	2.5163
		4	40	0.2796	0.6041	0.9659
		5	49	0.5029	0.7623	1.2620
HM5: $\mathbf{r(c)} = [S\, N\, r_1^{ANN},\, r_2^{ANN},\, r_3^{ANN},\, S\, r_4^{ANN}]^T$	S, N, T, pH	4	40	1.1818	0.8710	1.1868
HM6: $\mathbf{r(c)} = [S\, N\, X\, r_1^{ANN},\, X\, r_2^{ANN},\, X\, r_3^{ANN},\, S\, X\, , r_4^{ANN}]^T$	S, N, T, pH	4	40	2.7248	2.3309	2.9076
Parametric model with unstructured kinetics	**MSE train**			**MSE valid**		
$\mathbf{r(c)} = [\mu\, X,\, v_S\, X,\, v_N\, X,\, v_P\, X]^T$	604			374		

with low validation and test error are thus suitable for application in optimization problems.

It can be observed that the hybrid structure choice greatly affects the predictive properties of the model. In hybrid serial structures (HM4, HM5 and HM6), it was observed that when increasing the knowledge detail of the parametric part, the predictive capabilities diminish. It is well known that parametric knowledge imposes an inductive bias on the model performance and the functional mismatch cannot always be accounted for by the nonparametric model (Psichogios and Ungar 1992; Stosch et al. 2014). Thus, the observation seems to indicate that the incorporated knowledge is not correct, *i.e.* the underlying assumptions do not hold.

Better generalization properties were attained with the serial/parallel hybrid structures (HM1, HM2 and HM3). It was verified that the parametric CTP model describes well the effects of temperature and pH on biomass and EPS synthesis rates. Nonetheless, the CTP model alone is not able to explain all the effects of temperature and pH on the process, and therefore the inclusion of these factors in the ANN input space is a requisite to improve the predictive capabilities of the hybrid-model. Although the inclusion of knowledge regarding the culture broth viscosity lead to an improvement of MSE values in the training and validation sets (HM3), the model error for the test partition was higher, decreasing the model's predictive power. For this reason, HM2 would be more suited for model-based predictions.

4.4 Conclusions

Several hybrid semi-parametric models were developed to describe the kinetics for EPS production, as an alternative to unstructured phenomenological modeling strategies. The process dynamics are very complex and non-linear involving *e.g.* the effects of temperature, pH and culture broth viscosity.

Serial and a combination of serial/parallel hybrid models were proposed to study

which setup is better to fit the process data and also has a good predictive capability to be applied in a model-based optimization.

The results show that the hybrid models that incorporated complete mechanistic kinetics resulted in better identification and generalization capabilities than model structures that incorporated only partial mechanistic knowledge. Thus, it can be concluded that the adoption of the CTP model for the description of biomass growth rate and EPS synthesis rate was successful to improve the process description. Regarding the structure of the ANN, temperature and pH were found to be important to the process description, since the results seem to reveal that the parametric CTP model alone cannot describe all the temperature and pH effects in the process.

Medium viscosity was also taken into account in the hybrid model to improve the process characterization. This was achieved by a correlation between viscosity and EPS concentration. The integration of this knowledge in the hybrid methodology resulted in an improvement of the model calibration and validation errors. However in this case, this model structure did not achieve the best generalization properties.

Given the complexity of the process under study, hybrid semi-parametric modeling was proven to be a useful, cost-effective technique to model the fed-batch production of EPS, opening perspectives for model-based optimizations and industrial implementation.

CHAPTER 5

Model-based optimization of EPS production by *Enterobacter* A47

Bioprocess economics is strongly dependent on the control of operational parameters. In this chapter we investigate optimal control scenarios of EPS production by *Enterobacter* A47. For this purpose, we have implemented a dynamic optimization optimal control method to optimize the time profiles of control variables that satisfy certain performance criteria. The best predictive hybrid model developed in Chapter 4 was adopted as dynamic equality constraints in the dynamic optimization problem. We have addressed several optimal control scenarios, namely maximization of endpoint EPS mass production, minimization of residual glycerol mass, maximization of EPS volumetric productivity and maximization of overall EPS/glycerol yield. In order to minimize risk of process-model mismatch, a prediction risk inequality constraint was implemented. In all cases studied, improvements of process operation in comparison to the best experimental batch were forecasted. The best result was obtained for the endpoint EPS maximization problem, where 22% improvement of EPS production over the experimental observation is forecasted. Improvements of 18% and 13% are also forecasted for the scenarios of residual glycerol minimization and EPS/glycerol

yield maximization. It was found that the optimal values of temperature and pH did not vary much across the various scenarios tested. The control of the feeding strategies for ammonia and glycerol have the highest impact in the EPS production performance.

5.1 Introduction

In many biotechnological processes, fed-batch operation is a preferred strategy because it provides better control and ultimately improved bioprocess economics. Fed-batch operation involves the feeding of a nutrients solution into the reactor along time, in the quantities that are needed at each time instant, thereby overcoming certain undesired effects such as substrate inhibition and catabolite repression (Lee et al. 1999). The opportunity to better control the process is however also a big challenge, as the design of optimal control schemes for fed-batch operation is a complex dynamic optimization problem, owing to mathematical modeling and analysis as important tools to design optimal fed-batch operation (Balsa-canto et al. 2000).

Dynamic optimization as received major attention in the process control community, as reviewed by Banga et al. (2003) and Rani and Rao (1999). Such problems involve the specification of a set of control variables such as the feeding rate, temperature and pH, which ensure the maximization of a pre-defined performance index (*e.g.* productivity, or another economical index derived from the operation time profiles and final concentrations). The definition of dynamic optimization comes from the fact that the optimization problem is constrained by a set of equality dynamic constraints, which are normally time dependent ordinary differential equations (ODEs) or partial differential equations (PDEs). In the case of bioreactor control problems, dynamic optimization is very challenging due to the complex nonlinear dynamical constraints, the presence of constraints on both the state and control variables, and the existence of time-varying parameters (Banga et al. 2003; Rani and Rao 1999).

In order to ensure a robust optimal control in bioprocesses, several methods are available to address the dynamical optimization problem. They can be classified in terms of optimization algorithms to solve the non-linear programming (NLP) problem as deterministic or stochastic. Due to the fact that NLP problems are often characterized by having multiple optima, deterministic methods (also known as gradient-based methods) may usually converge to local optima, especially when

the optimization problem starts far from the global solution. Conversely, stochastic methods are based on random search, consisting in global optimizers less prone to fall within local optima at the cost of higher computational power. For a review on optimal control problems based on stochastic methods see Banga et al. (1997, 2005).

The design of optimal control for the production of exopolysaccharides, more specifically xanthan gum has been previously addressed by Cacik et al. (2001). This study describes an application of a gradient-based open-loop optimal control problem applied to a batch production of xanthan gum. The method enabled to maximize the xanthan gum produced in a shorter time, using temperature as a time dependent decision variable. Another approach involved the application of an optimal control problem to a fed-batch reactor to produce xanthan gum by a multiple substrate feeding optimization (Chaitali et al. 2003), using as decision variables the feeding rates for the carbon, nitrogen and oxygen sources, resulting in a 148.7% increase in xanthan gum production. In the present work, a dynamic optimization study was made (optimal control problem) of EPS production in a fed-batch reactor using a dynamic hybrid model a hybrid as equality constraints. A stochastic optimization solver was adopted, namely the differential evolution (DE) algorithm, to find global optimal control scenarios (Storn and Price 1997). Several control objectives were covered, namely maximization of final EPS mass, product yield, productivity and minimization of residual substrate quantity.

5.2 Methodology

5.2.1 Dynamic optimization - optimal control problem

A dynamic optimization, optimal control problem, can be stated as to find the control inputs $\mathbf{u}(t)$ over the process $t \in [t_0, t_f]$ that maximize a given process performance index, J, subject to a set of constraints, some of which are expressed in the form of a dynamical model (Banga et al. 1998; Bryson and Ho 1975). In the present study,

the dynamic optimization problem takes the form:

$$\begin{aligned} \min_{\mathbf{u}} J &= \text{process performance index} \\ \text{Subject to} \quad & \text{Hybrid model equality constraints} \\ & \text{Prediction risk inequality constraint} \\ & \text{Physical boundaries inequality constraint} \end{aligned}$$

Several optimal control scenarios were studied summarized in Table 5.1. The control inputs are the temperature and pH for the fermentation batch phase (T_b and pH_b), the switch time from batch to fed-batch mode (t_{switch}), the glycerol feeding rate in feeding stream (F_S), the ammonia/glycerol mass ratio in the feeding stream (R_{NS}), and the temperature and pH in the fed-batch phase (T_{fb} and pH_{fb}). No time parametrization of control inputs within the batch and fed-batch phases was adopted.

In the first optimization scenario, the total mass of the EPS is maximized. In the second scenario, the residual glycerol mass is minimized with inclusion of an additional constraint to ensure a final EPS mass above the highest quantity achieved experimentally. The third scenario maximizes the average EPS production per unit time, *i.e.* productivity. In this case the total fermentation time was treated as control variable, since a higher productivity may be associated with a shorter fermentation.

Table 5.1: Summary of optimal control problems studied. Performance criteria are maximization of EPS mass production, minimization of residual glycerol mass, maximization of EPS productivity and maximization of EPS/glycerol yield. The control variables are: temperature of batch phase (T_b), batch phase pH (pH_b), fed-batch phase temperature (T_{fb}), fed-batch phase pH (pH_{fb}), glycerol mass feed (F_S), the ammonia/glycerol mass ratio in the feeding stream (R_{NS}), the batch to fed-batch switch instant (t_{switch}), final run time (t_f), initial volume (V_{init}).

Performance index (J)	Control inputs ($\mathbf{u}$)
$\max_{\mathbf{u}} J = P(t_f)\,V(t_f)$	T_b, pH_b, F_S, R_{NS}, t_{switch}, T_{fb}, pH_{fb}
$\min_{\mathbf{u}} J = S(t_f)\,V(t_f)$	T_b, pH_b, F_S, R_{NS}, t_{switch}, T_{fb}, pH_{fb}, V_{init}
$\max_{\mathbf{u}} J = \frac{P(t_f)\,V(t_f)}{t_f}$	T_b, pH_b, F_S, R_{NS}, t_{switch}, T_{fb}, pH_{fb}, t_f
$\max_{\mathbf{u}} J = Y_{P/S}$	T_b, pH_b, F_S, R_{NS}, t_{switch}, T_{fb}, pH_{fb}

The fourth scenario corresponds to the maximization of the overall product/glycerol yield calculated as:

$$Y_{P/S} = \frac{P(t_f)\,V(t_f)}{S_0\,V_0 - S(t_f)\,V(t_f) + F_S^{ac}} \tag{5.1}$$

with S_0 and V_0 the initial glycerol concentration and initial volume respectively and, $P(t_f)$ the final EPS concentration, $V(t_f)$ the final volume, $S(t_f)$ the final glycerol concentration and F_S^{ac} the total glycerol mass fed to the reactor.

To solve the optimization problem we adopted the direct method with control vector parameterization (see details of control vector parameterization in Table 5.1. This approach transforms the original dynamic optimization problem into a nonlinear programming (NLP) problem in which the dynamical equations constraints are numerically integrated for each objective function evaluation (*i.e.* inner initial value problem). To solve the NLP we have adopted the differential evolution (DE) algorithm (Storn and Price 1997). DE belongs to the class of stochastic optimization algorithms suitable to find global optima in complex constrained problems with discontinuities and complex nonlinear constraints (Chaitali et al. 2003). The DE algorithm works by parameterizing each decision variable into a set of discrete values limited by the upper and lower bounds. This in turn creates a quantization grid that defines the size of the population of candidates to the objective optimum. The higher the number of discretization, the bigger the population of candidates and therefore, the higher the computational demand. After a solution is reached by the DE algorithm, a gradient-based optimization routine based on the fmincon MATLAB function, using an interior-point method, was applied to fine tune the solution the optimum solution.

5.2.2 Dynamical equality constraints

A dynamic hybrid model sets the dynamical equality constraints to the optimization problem. All the details of the hybrid model structure are provided in Chapter 4. Briefly, the material balance equations take the form of time dependent ordinary

differential equations as follows:

$$\frac{dX}{dt} = (\mu - K_d - D)\ X \tag{5.2}$$

$$\frac{dS}{dt} = -v_S\, X + D\,(S_f - S) \tag{5.3}$$

$$\frac{dN}{dt} = -v_N\, X + D\,(N_f - N) \tag{5.4}$$

$$\frac{dP}{dt} = v_P\, X - D\, P \tag{5.5}$$

$$\frac{dV}{dt} = F \tag{5.6}$$

With X biomass concentration, S glycerol concentration, N ammonia concentration, P is EPS concentration, V the culture volume, F is the inlet volumetric feed rate, $D = F/V$ is the dilution rate, v_i are the specific reaction rates defined as follows:

$$\mu = \mu_{max}\, r_1^{ANN} \tag{5.7}$$

$$v_S = -(Y_{S/X} + r_2^{ANN})\,\mu - (Y_{S/P} + r_3^{ANN})\,v_P \tag{5.8}$$

$$v_N = -(Y_{N/X} + r_4^{ANN})\,\mu \tag{5.9}$$

$$v_P = v_{P,max}\, r_5 \tag{5.10}$$

The group of constraints defined by Equations 5.2 to 5.5 corresponds to the dynamic equality constraints. For further details see Chapter 4. As stated above, in the dynamic optimization direct method with control inputs parameterization, these equations must be numerically integrated within each objective function evaluation. The initial values for the state variables were set to be the same as in the experimental setup, that is with biomass, glycerol and ammonia concentrations of 0.34 g/L, 49.0

g/L and 1.26 g/L respectively and with initial reactor volume of 1.369 L. Moreover, in order to be able to compare the optimal control scenarios with the experimental data, we have implemented a sampling procedure along 13 discrete time instants where 23 mL samples are collected at each sampling time plus 50 mL at end of the fermentation. The integration time was fixed to 105 h except when cultivation time is also a decision variable.

5.2.3 Prediction risk inequality constraint

A prediction risk inequality constraint was included in order to constrain the NLP search space to those regions where the hybrid model has low prediction uncertainty. To this purpose we have adopted the method described in (Simutis et al. 1995), wherein the neural network input subspace covered by the training dataset is clustered by a set of NC Gaussian clusters of the general form:

$$r(\mathbf{c}, \mathbf{m_i}, \Sigma_i) = e^{-1/2(\mathbf{c}-\mathbf{m_i})^T \, \mathbf{\Sigma_i} \, (\mathbf{c}-\mathbf{m_i})} \tag{5.11}$$

with $\mathbf{m_i}$ the cluster center and $\mathbf{\Sigma_i}$ the (diagonal) variance matrices $\mathbf{\Sigma_i} = \text{diag}(\sigma_{ii})^2$. The neural network input space comprehends the state variables biomass, glycerol, ammonia and EPS concentrations, temperature and pH. We adopted the k-means algorithm to calculate de centers, $\mathbf{m_i}$, of the NC clusters using the minimal distance method (*e.g.* Leonard et al. (1992)). The number of clusters was set to $NC = P/3$, with P the number of data points (Leonard et al. 1992). The data points were auto-scaled prior to the clustering to ensure a uniform coverage over the hyperspace. Once the clusters centers $\mathbf{m_i}$ are established, the standard deviations in $\mathbf{\Sigma_i} = \text{diag}(\sigma_{ii})^2$ were estimated based on the distance of each cluster center j to the nearest $PDISP$ cluster centers, with $PDISP$ an heuristic parameter set by the user. $\mathbf{\Sigma_i} = \text{diag}(\sigma_{ii})^2$ is then determined as the norm of the distance of the $PDISP$ closest clusters to cluster j. The model experience measure for a given neural network input, $\mathbf{c}$, is

a scalar variable between $[0, 1]$ calculated as the maximum value possible over all clusters.

$$em(\mathbf{c}) = \max\, r(\mathbf{c}, \mathbf{m_i}, \Sigma_i) \tag{5.12}$$

Finally the risk of model unreliability is taken as the complement of the model experience measure:

$$RISK(\mathbf{c}) = 1 - em(\mathbf{c}) \tag{5.13}$$

When $RISK(\mathbf{c}) \to 0$ it means that the neural network input is close to a cluster center thus the input c belongs to the training subspace where the model is likely to be reliable. When $RISK(\mathbf{c}) \to 1$ the datum point is far away from the training data set and thus the model is likely to be unreliable. The average risk over the process time domain is constrained by a maximum risk value set by the user.

$$\frac{\int_{t_0}^{t_f} RISK[\mathbf{c}(dt)]\, dt}{t_f - t_0} \leq RISK_{max} \tag{5.14}$$

Equation (5.14) served as additional constraint to the optimization problem.

5.2.4 Physical bounds inequality constraints

Finally, the following set of physical boundaries inequality constraints were imposed to the NLP program:

$$V \leq 2.0 \tag{5.15}$$

$$15.9 \leq T \leq 44.0 \tag{5.16}$$

$$6.00 \leq pH \leq 8.00 \tag{5.17}$$

$$0.95 \leq Fs \leq 2.00 \tag{5.18}$$

$$0.003 \leq R_{NS} \leq 0.005 \tag{5.19}$$

$$20 \leq t_{switch} \leq 30 \tag{5.20}$$

5.3 Results and Discussion

5.3.1 Model prediction and prediction risk

The neural network input space was subject to clustering with $NC = 745$ Gaussian clusters as described previously. The calculation of the cluster centers was performed iteratively (k-means algorithm) in order to minimize the distance between the data points and the cluster centers. The clustering error can thus be assessed by the sum of the squared distances between the data and the cluster centers, divided by the number of data points. The final error of the clustering procedure was 0.0020. The cluster dispersion was determined based on the $PDISP = 50$ nearest points. The results of the clustering are shown in Figure 5.1 where the contour lines delimit regions of the input space with the same experience measure Equation (5.12).

5.3.2 Maximization of total EPS mass

The final EPS mass was maximized under the constraint of increasing levels of maximum risk of model unreliability (Table 5.2). As expected, the predicted final EPS mass increases as higher risk levels are undertaken. With 70% risk level the optimal EPS achieved was 16.6 g, which is already above the highest value obtained experimentally, which was 15.0 g of EPS. The maximum EPS mass increases up to 19.5 g for a maximum risk of 100%, *i.e.* giving complete freedom to the algorithm to explore the design space.

Table 5.2: Variation of the objective function result in terms of the maximum risk assumed.

Maximum risk (%)	**Objective function value**
70	16.6 g
80	18.4 g
90	19.3 g
100	19.5 g

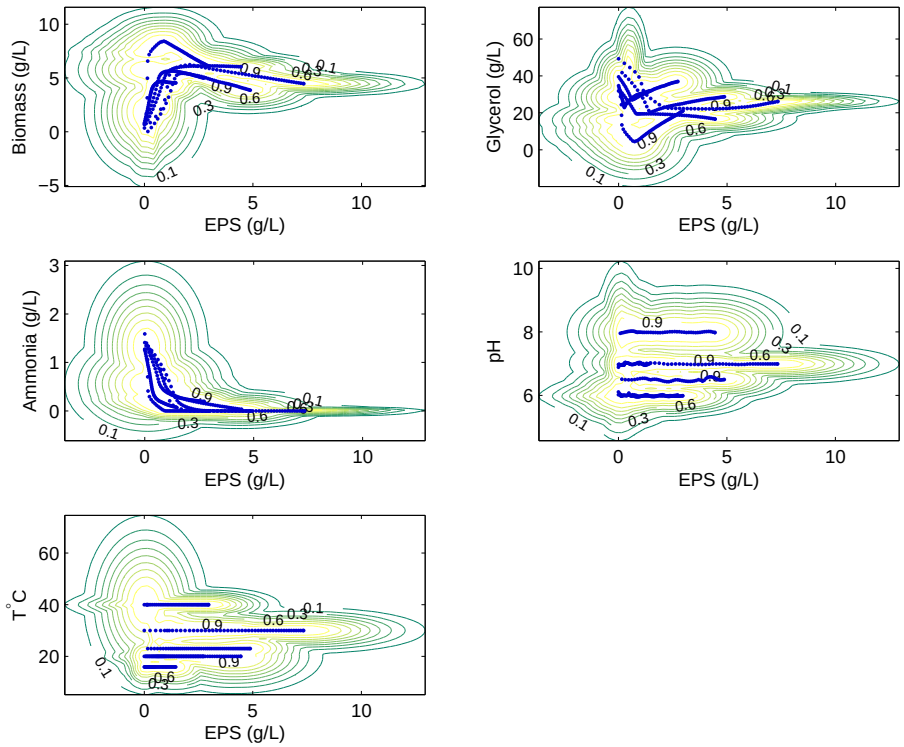

Figure 5.1: Trust level contours for the hybrid model. The line in blue corresponds to the training data that was used to determine the cluster centers.

A maximum risk of 80% was assumed for the rest of the optimization scenarios. The optimal fed-batch operation for risk of 80% is shown in Figure 5.3 compared to the best and worse experimental runs in the training partition. The optimal control variables are summarized in Table 5.3.

The results show an initial biomass growth coupled with the consumption of glycerol and ammonia until the end of the batch phase, at 20 h. After the start of the fed-batch phase, a slight increase on ammonia consumption is forecasted, reaching a peak at approximately 30 h, followed by its decrease to zero, where it remained constant until the end of the fermentation.

The ammonia mass balance (Equation (5.4)) can be rewritten to describe the

Table 5.3: Optimal control variables values for the EPS mass maximization scenario. Legend: T_b, temperature of the batch phase; pH_b, pH of the batch phase; F_S,, mass feeding rate of glycerol; R_{NS}, ammonia/glycerol concentration ratio in the feeding stream, t_{switch}, instant of time for batch to fed-batch switch; T_{fb}, temperature of the fed-batch phase; pH_{fb}, pH of the fed-batch phase; J_{opt}, value of the objective function.

Objective function	T_b	pH_b	F_S	R_{NS}	t_{switch}	T_{fb}	pH_{fb}	J_{opt}
$\max_{\mathbf{u}} J = P(t_f)\,V(t_f)$	29°C	7.00	1.55 g/h	0.005	20 h	26.8°C	7.32	18.4 g

ammonia mass dynamics in the fed-batch phase:

$$\frac{d(NV)}{dt} = -v_N\,X\,V + F_N \tag{5.21}$$

where F_N represents the ammonia mass flow (g/h) in the feeding stream, which is related to the glycerol mass flow in the feeding solution by the expression $F_N = R_{NS}\,F_S$. According to Equation (5.21), we can observe that the accumulation of ammonia mass is positive whenever the feeding rate outweighs the consumption rate. The optimization resulted in an increase of R_{NS} from 0.0041 (standard conditions) to 0.0050, which is equivalent to increase ammonia feeding F_N, from the standard 4.0 mg/h to 7.8 mg/h. This in turn seems to favor higher EPS production, since ammonia induces higher biomass growth, which in turn increases the volumetric EPS production.

The optimization results also suggest an increase in the mass flow rate of glycerol into the reactor (from 1 g/h in the standard conditions to 1.55 g/h), thus leading to a glycerol accumulation in the reactor. This can be further confirmed in Figure 5.2, where several model simulations with the fixed optimal parameters (Table 5.3) were performed except the glycerol feeding that was varied from 0.95 to 2 g/h. It may be observed that the total EPS produced has an optimum for $FS = 1,55$g/h, which is coincident with the optimization result. This could be however an indirect effect ruled by the need of increasing ammonia concentration in the reactor. Since the ratio of ammonia to glycerol in the feed R_{NS} is fixed and because the optimal R_{NS} coincided

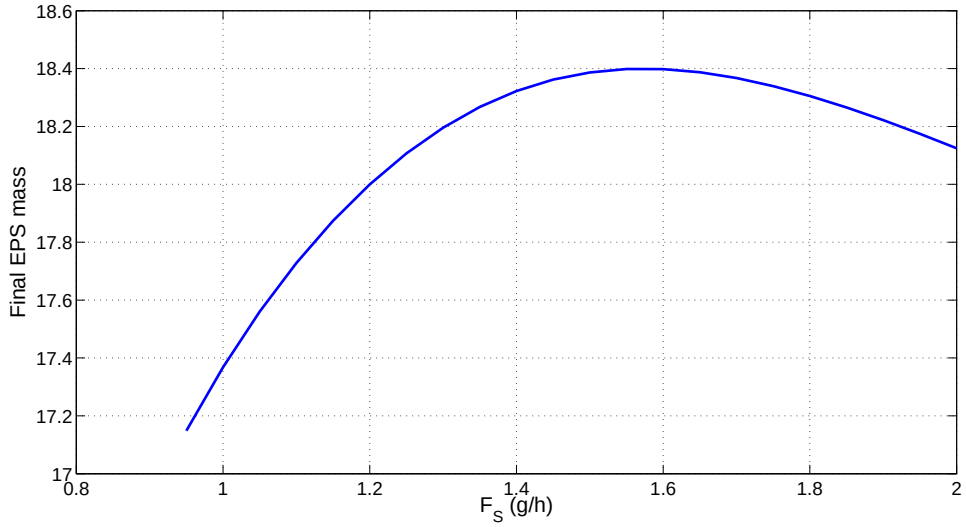

Figure 5.2: Local maximum found for the glycerol mass flow.

with the upper optimization bound, increasing F_S is the only way to increase F_N.

In this optimization scenario, due to the higher mass flows of ammonia and glycerol into the reactor, biomass growth occurs during the fed-batch phase, contributing to an increased EPS productivity. Previous studies for EPS production by *Enterbacter* A47 have also confirmed that an increased ammonia flow in a fed-batch operation induced a higher EPS productivity (Torres et al. 2014).

Regarding the control of temperature and pH, it was found that during the batch phase, these parameters were close to the standard conditions (T = 30°C, pH = 7.00), while the optimum temperature for the fed-batch phase was slightly lower (26.8 °C) and the pH higher (7.32). These values are in agreement with previous temperature and pH optimization studies for EPS production in *Enterobacter* A47 cultures (Torres et al. 2012).

5.3.3 Minimization of residual substrate mass

As seen in the previous section, maximization of EPS mass requires high mass flow rates of glycerol and ammonia. This might result in the accumulation of nutrients

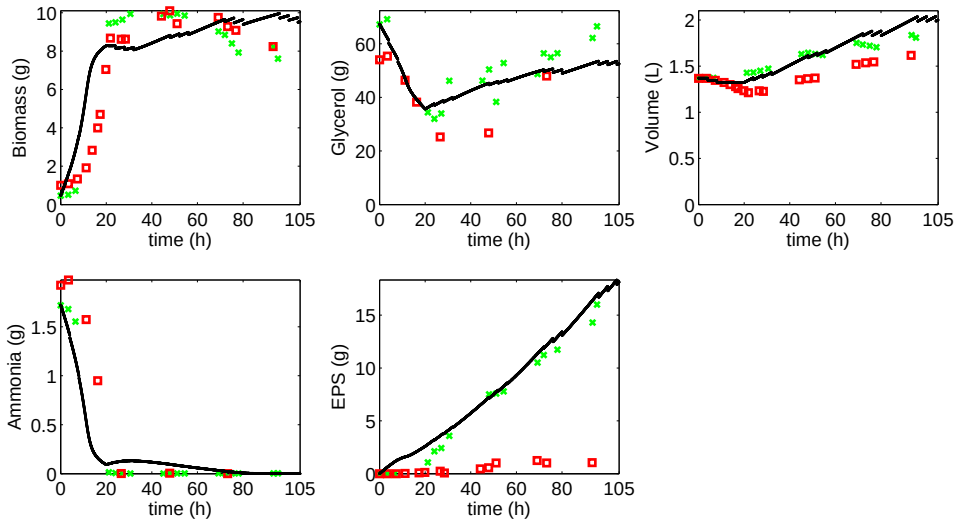

Figure 5.3: Simulation results for the maximization of EPS mass for a maximum risk value of 80%. The green crosses correspond to the best experimental batch and the red squares to the worst performed experiment.

that are wasted at the end of the fermentation. We have explored the optimal control scenario that minimizes residual glycerol mass while at the same time ensures a final production of EPS equal or higher than 15.0 g EPS, which was the highest value obtained experimentally. Moreover, the initial volume was set as a control variable, since it was found to be a critical parameter to find a feasible solution. The obtained optimum control variables are shown in Table 5.4. The simulation of the optimal control strategy is shown in Figure 5.4. This optimal control scenario predicts a significant drop of the residual glycerol mass from 66.5 g to 19.3 g, and a final EPS of 17.7 g. This was mainly achieved by a decrease in the glycerol mass flow to 0.95 g/h in the fed-batch phase. The optimal ammonia/glycerol feeding ratio was found to be 0.005 and therefore, the ammonia mass flow in the feeding was 4.7 mg/h. As discussed previously, maximizing EPS production requires an increase of feeding rate that induces high biomass growth in the fed-batch phase and also EPS production. In the case of residual glycerol minimization, the decrease in the feeding was compensated by a higher initial liquid volume in the reactor, implying an initial

Table 5.4: Optimal values for the control variables for the scenario of residual glycerol mass minimization. Legend: T_b, temperature of the batch phase; pH_b, pH of the batch phase; F_S„ mass feeding rate of glycerol; R_{NS}, ammonia/glycerol concentration ratio in the feeding stream, t_{switch}, instant of time for batch to fed-batch switch; T_{fb}, temperature of the fed-batch phase; pH_{fb}, pH of the fed-batch phase; V_{init}, initial reactor volume; J_{opt}, value of the objective function.

Objective function	T_b	pH_b	F_S	R_{NS}	t_{switch}	T_{fb}	pH_{fb}	V_{init}	J_{opt}
$\min_{\mathbf{u}} J = S(t_f)\,V(t_f)$	30.2°C	7.00	0.95 g/L	0.005	24 h	25.4°C	7.12	1.678 L	19.3 g

82.6 g of glycerol, contrasting with the 67.4 g of glycerol with standard conditions. This higher glycerol concentration allowed a higher biomass growth in the batch phase, reaching 11 g of biomass at the instant 24 h (t_{switch}), comparing with the previous optimization scenario which achieved a maximum of 8.3 g of biomass. The time instant when batch is switched to fed-batch mode seems to be linked to the exhaustion of ammonia, since it coincides in time.

The predicted values of temperature and pH didn't show practically any deviation from the previous optimization problem in the batch phase, meaning that biomass growth is favored by these values. Regarding the fed-batch phase, a similar pattern was also observed, with temperature achieving a lower value (25.4°C) and pH a slightly higher value (7.14). These results suggest EPS production is favored at lower temperatures and higher pH values.

5.3.4 Maximization of EPS productivity

We have investigated an optimal control scenario that maximizes EPS productivity taking the fermentation time as a control parameter. The results are shown in Table 5.5 and Figure 5.5. Although the EPS concentration profile seems to be growing at a constant rate, the highest productivity (0.18g/h) is obtained when the fermentation time is slightly decreased to 96 h. This optimal scenario is similar to the maximization of total EPS, but with the advantage of saving time, thus inversing the overall productivity.

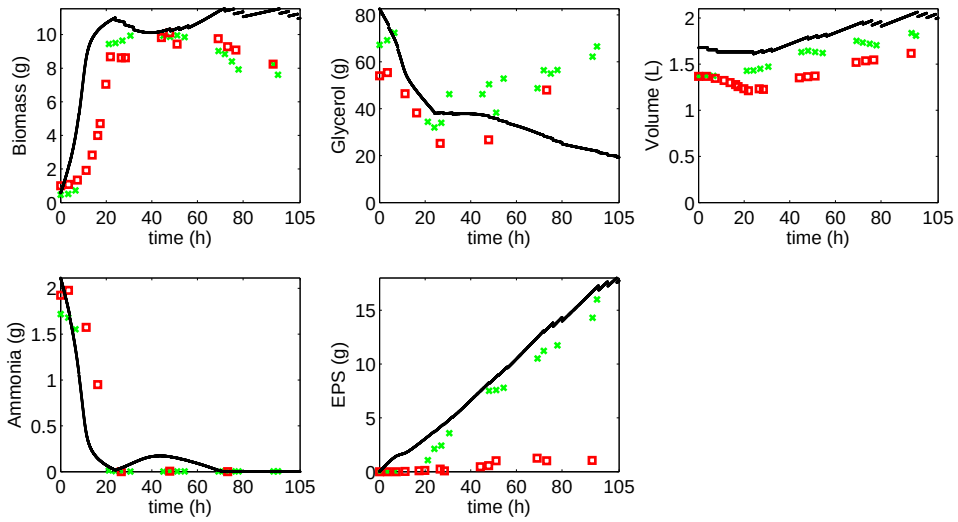

Figure 5.4: Simulation results for the minimization of residual glycerol mass for a maximum risk value of 80%. The green crosses correspond to the best experimental batch and the red squares to the worst performed experiment.

Table 5.5: Optimal values for the control variables for the scenario of EPS productivity maximization. Legend: T_b, temperature of the batch phase; pH_b, pH of the batch phase; F_S„ mass feeding rate of glycerol; R_{NS}, ammonia/glycerol concentration ratio in the feeding stream, t_{switch}, instant of time for batch to fed-batch switch; T_{fb}, temperature of the fed-batch phase; pH_{fb}, pH of the fed-batch phase; t_{final}, final reactor age; J_{opt}, value of the objective function.

Objective function	T_b	pH_b	F_S	R_{NS}	t_{switch}	T_{fb}	pH_{fb}	t_{final}	J_{opt}
$\min_{\mathbf{u}} J = S(t_f)\,V(t_f)$	29.0°C	7.06	1.55 g/L	0.005	22 h	26.5°C	7.60	96 h	0.18 g/h

5.3.5 Maximization of EPS/glycerol yield

Because glycerol is used for the formation of both EPS and biomass, this strategy attempts to balance the demands of substrate for these two fermentation products in favor of EPS synthesis. The optimal control results are shown in Table 5.6 and Figure 5.6. Ammonia feeding plays here an important role since it is a limiting substrate for biomass growth. The simulation of results show that ammonia is not entirely exhausted at the end of the batch phase and, when the fed-batch starts, the ammonia quantity registers a slight increase followed by its exhaustion until the end

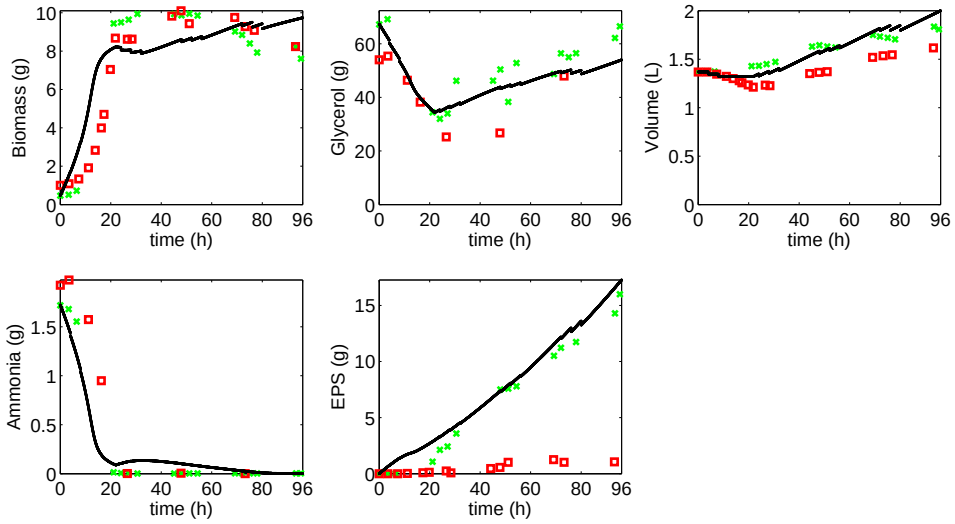

Figure 5.5: Simulation results for the minimization of residual glycerol mass for a maximum risk value of 80%. The green crosses correspond to the best experimental batch and the red squares to the worst performed experiment.

Table 5.6: Optimal control variables values for the EPS/glycerol yield maximization. Legend: T_b, temperature of the batch phase; pH_b, pH of the batch phase; F_S,, mass feeding rate of glycerol; R_{NS}, ammonia/glycerol concentration ratio in the feeding stream, t_{switch}, instant of time for batch to fed-batch switch; T_{fb}, temperature of the fed-batch phase; pH_{fb}, pH of the fed-batch phase; J_{opt}, value of the objective function.

Objective function	T_b	pH_b	F_S	R_{NS}	t_{switch}	T_{fb}	pH_{fb}	J_{opt}
$\max_{\mathbf{u}} J = P(t_f)\,V(t_f)$	28.9°C	7.01	0.95 g/h	0.005	24 h	25.8°C	7.50	0.16 g/g

of the fermentation. The feeding mass rates determined in this case were similar to the scenario of residual glycerol minimization, that is, a lower glycerol feeding rate (0.95 g/h), while the ammonia feeding was the maximum allowed by the R_{NS}, which was 4.7 mg/h. The analysis of the objective function (Equation (5.1)) helps to see that a lower glycerol feed (contrasting with the standard conditions) increases the value of $Y_{P/S}$, while the ammonia mass flow increase is important to sustain the biomass growth observed during the fed-batch phase, which in turn increases EPS productivity. A maximum of 17.0 g of EPS was produced in this scenario.

The three scenarios investigated have all registered a yield of 0.14 g/g, which is lower by a margin of 2 g of EPS per glycerol consumed. These previous cases

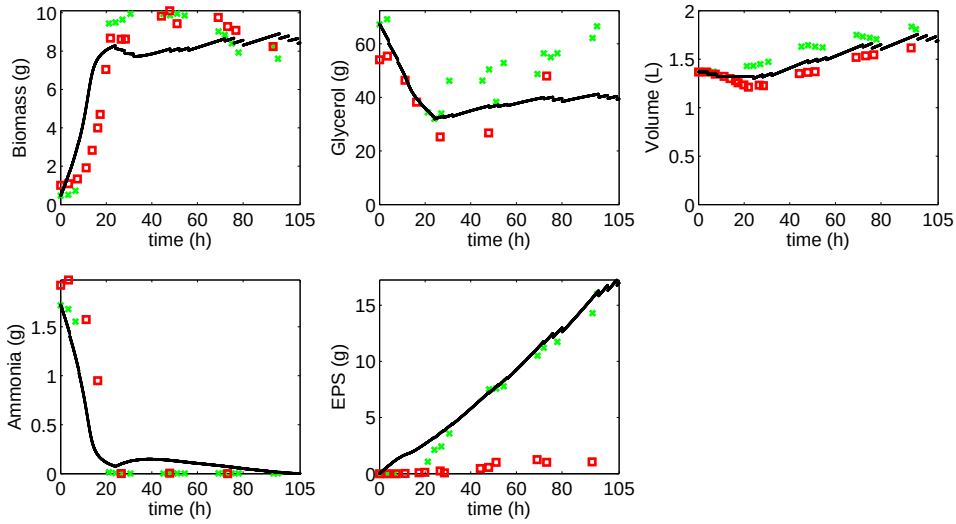

Figure 5.6: Simulation results for the maximization of EPS/glycerol yield for a maximum risk value of 80%. The green crosses correspond to the best experimental batch and the red squares to the worst performed experiment.

registered lower yields by different reasons. In the scenario of EPS maximization, the product quantity was improved by allowing higher nutrient feeding rates into the reactor, which consequently led to an accumulation of glycerol, thus lowering the yield value. Conversely, in the scenario where the residual glycerol was minimized, the initial quantity of glycerol was higher by virtue of a higher initial volume, which was used to produce biomass in higher quantity in the batch phase. For this reason a lower yield was achieved.

5.4 Conclusions

In this chapter, several model-based optimization strategies were investigated to improve the production of EPS in a fed-batch bioreactor. Different optimal control criteria were investigated, namely the maximization of total EPS produced, minimization of residual glycerol mass, maximization of EPS productivity and maximization of EPS/glycerol yield.

Since the hybrid model developed in the previous chapter was used to design the

optimal strategies, a clustering analysis was employed to assess the nonparametric model reliability to explore novel process trajectories. This method was effective to classify the input data in the model according to membership relationships defined by the cluster centers. A maximum risk parameter was defined to restrict the data into acceptable limits of extrapolation.

It was verified that the values of temperature and pH are consistent across all the performed studies, with values close to 30°C and pH = 7.0 in the batch phase, slightly lower temperature values and higher pH values for the fed-batch phase. These values reflect the optimum conditions for biomass growth in the batch phase and EPS synthesis in the fed-batch phase, respectively.

It was also verified that the feeding rates are important effectors of the overall process performance. To maximize EPS production, the optimal control consistently suggests an increase in the feeding rates of glycerol and ammonia. More specifically, an EPS increase from 15.0 to 18.4 g EPS is predicted, representing a 22% improvement over the best experimental data-set. Although a higher EPS quantity is desirable, this control policy led to an accumulation of residual glycerol in the reactor. Conversely, by lowering the glycerol feeding rate and increasing the initial volume for the same initial concentrations, the residual glycerol was lowered significantly while still achieving an 18% improvement over the best experimental run.

The results also show that the maximization of the overall EPS/glycerol yield could represent a more balanced approach between EPS productivity and accumulation of substrate in the reactor. With this strategy, it was observed an increased EPS production of 13% over the experimental batch, while avoiding accumulation residual glycerol.

CHAPTER 6

Conclusions and future work

The present Ph. D. thesis is focused on the model-based optimization of exopolysaccharide production by a novel microbial strain, *Enterobacter* A47, using glycerol as carbon source. The main objectives of the Ph. D. thesis are i) development of an *in silico* cellular model of *Enterobacter* A47 describing cellular growth and exopolysaccharide production from glycerol as major carbon source, ii) integrate the cellular model with a macroscopic bioreactor dynamical model that relates bioreactor operational parameters (effectors) with process performance criteria such as EPS productivity (consequence), and ii) perform optimal control studies using dynamic programing algorithms to maximize the process performance criteria by manipulating process control degrees of freedom.

The first task towards the development of an *in silico* model for *Enterobacter* A47 involved the genome sequencing and genome reconstruction, as well as the compilation of information available in the literature and public databases. Although the reconstruction of a genome-scale metabolic network is still an ongoing effort, a smaller core metabolic network was developed, describing the central carbon metabolism and biosynthetic pathways leading to all the intermediary sugars and

chemical substituents identified in the structure of EPS. The developed metabolic network comprised 58 intracellular reactions and 18 exchange fluxes. These include the uptake of ammonia, glycerol and oxygen, as well as the excretion of metabolic products including EPS. The intracellular reactions represent the glycolysis/gluconeogenesis pathway, glycerol dissimilation, the pentose-phosphate pathway, TCA cycle, anaplerotic reactions, respiratory chain and oxidative phosphorylation and also fermentative reactions. The exchange fluxes include the uptake of ammonia, glycerol and oxygen, as well as the excretion of metabolic products including EPS.

An elementary flux mode (EM) of a given metabolic network represents a minimal feasible set of metabolic reactions operating in steady-state. It may be viewed as a possible metabolic pathway for the cells to survive. A metabolic network has normally a very large number of EMs and not all of them are used simultaneously by the cells. The knowledge of which EMs are active among the whole set of EMs may be important to reduce models, thereby facilitating model-based optimization. In this thesis we have developed a methodology to identify a minimal set of active EMs from measured yield data. The advantage of yield data is that it requires only extracellular measurements and, since internal metabolic fluxes are strictly constrained by the status of external fluxes, the methodology is meaningful. After the transformation of the full EM set into yield space, two different methods were applied to discriminate the active set of EMs: PEMA and YA. The former is a principal component like method that maximizes the explained variance in the dataset, whereas in YA the EMs are chosen as the minimal number to enclose a given measured yield vector, using a quadratic programming algorithm. The results have shown that PEMA was able to explain 98.1% of the measured yield variance with a minimal set of 5 EMs, while YA required 9 EMs to explain 99% of variance. Both methods could be confronted in terms of the degree of orthogonality exhibited by their respective EM set. It was observed that EMs obtained from the YA method had a higher degree of orthogonality than the EMs obtained with PEMA, approximating the former to a

PCA solution. Moreover, it could be concluded that the YA a single-substrate single-product solution is obtained for most EMs, while PEMA favors a multi-substrate and multi-product solutions.

Regarding the formulation of a bioreactor model, several hybrid semi-parametric strategies were developed to describe the kinetics of EPS production as an alternative to unstructured phenomenological models. This proved to be a favorable strategy to cope with the complex and non-linear character of process dynamics involving temperature, pH and culture broth viscosity. The results have shown that, depending on the type of knowledge and size of the nonparametric part, the combination of parallel/serial structures seem to perform better, judging by the capacity of the model to fit a training data set, while describing an independent data set. Regarding the structure of the ANN, temperature and pH were important to the process description, since the results reveal that the parametric model cannot described the effects of temperature and pH alone. The medium viscosity was also incorporated into a hybrid model by means of an empirical correlation between EPS concentration and the apparent viscosity. Although not having the best test set description among all models compared, the incorporation of viscosity improved the model accuracy of the training and validation sets.

Following the hybrid-model results, the model with the best predictive power was chosen to design optimization strategies for process improvement. Different criteria were defined to quantify the performance index (*i.e.* objective function), namely the maximization of total EPS produced, minimization of residual glycerol mass, maximization of EPS productivity and maximization of EPS/glycerol yield. In order to avoid the hybrid-model to extrapolate into unreliable regions, a clustering technique was devised to classify the input data according to a membership function defined by the cluster centers.

The optimal-control results have shown temperature and pH to be consistent across all the studied scenarios. It was also verified that feeding rates have an

important impact in the overall process performance. To achieve higher quantities of EPS, an increase in the mass rates of glycerol and ammonia are suggested. A maximum EPS quantity of 18.4 g EPS was predicted with this feeding strategy, representing a 22% improvement over the best performing experimental data-set. Although a higher EPS quantity is desirable, this control policy led to an accumulation of residual glycerol in the reactor. Conversely, by lowering the glycerol feeding rate and increasing the initial volume for the same initial concentrations, the residual glycerol was lowered significantly while still achieving an 18% improvement over the best experimental run. The results have also shown that the maximization of the overall EPS/glycerol yield could result in a more balanced operation, with a compromise between maximal EPS productivity and minimal accumulation of substrate in the reactor. With this strategy, it is forecasted an increase of EPS production by 13% over the best experimental batch, while avoiding high amounts of residual glycerol.

The results achieved in this work may be improved and extended in diverse ways. First of all, the metabolic reconstruction of *Enterobacter* A47 is still an ongoing effort, which has the potential of unlocking interesting possibilities for genetic engineering and bioprocess design. Only recently the first studies for the genome reconstruction of this organism have emerged, shedding light into the biochemical details of this organism. This study would support the creation of a genome-wide metabolic reconstruction for *Enterobacter* A47, as well as helping to identify genetic traits that could have dramatic effects in the biotechnological applications of this organism, such as the existence of prophage genetic sequences or virulence factors encoded in the genome.

The inherent properties of a reconstructed metabolic network would provide an important insight into relevant features of the cell system regarding the biochemical details of polysaccharide synthesis. The possibility for strain improvement through metabolic engineering could impact the productivity or even allow the tailoring of

polysaccharide composition.

Although the application of hybrid semi-parametric models to the bioreactor process developed in this work represented a big improvement over previous modeling approaches, the underlying connections between bioreactor model and the cellular metabolism are still regarded as a black-box. The studies based on the identification of an active set of EMs could be accommodated into an updated hybrid model, where the structure of the metabolism is taken into account, detailing the parametric part of the model. This would improve the generalization capabilities of the model, potentially leading to better optimization scenarios.

Bibliography

Al-Asheh, S., B. Abu-Jdayil, N. Abunasser, and A. Barakat (2002). "Rheological characteristics of microbial suspensions of Pseudomonas aeruginosa and Bacillus cereus". In: *International Journal of Biological Macromolecules* 30, pp. 67–74.

Alves, V. D., F. Freitas, N. Costa, M. Carvalheira, R. Oliveira, M. P. Gonçalves, and M. A. M. Reis (2010a). "Effect of temperature on the dynamic and steady-shear rheology of a new microbial extracellular polysaccharide produced from glycerol byproduct". In: *Carbohydrate Polymers* 79.4, pp. 981–988.

Alves, V. D., F. Freitas, C. A. V. Torres, M. Cruz, R. Marques, C. Grandfils, M. P. Gonçalves, R. Oliveira, and M. A. M. Reis (2010b). "Rheological and morphological characterization of the culture broth during exopolysaccharide production by Enterobacter sp." In: *Carbohydrate Polymers* 81.4, pp. 758–764.

Aziz, R. K. et al. (2008). "The RAST Server: Rapid Annotations using Subsystems Technology". In: *BMC Genomics* 9.1, p. 75.

Badsha, M. B., R. Tsuboi, and H. Kurata (2014). "Complementary elementary modes for fast and efficient analysis of metabolic networks". In: *Biochemical Engineering Journal* 90, pp. 121–130.

Bajaj, I. B., S. A. Survase, P. S. Saudagar, and R. S. Singhal (2007). "Gellan gum: Fermentative production, downstream processing and applications". In: *Food Technology and Biotechnology* 45.4, pp. 341–354.

Balsa-canto, E., J. R. Banga, A. A. Alonso, and V. S. Vassiliadis (2000). "Efficient Optimal Control of Bioprocesses Using Second-Order Information". In: *Industrial & Engineering Chemistry Research* 39, pp. 4287–4295.

Banga, J. R., A. A. Alonso, and R. P. Singh (1997). "Stochastic dynamic optimization of batch and semicontinuous bioprocesses". In: *Biotechnology Progress* 13.3, pp. 326–335.

Banga, J. R., R. Irizarry-Rivera, and W. D. Seider (1998). "Stochastic optimization for optimal and model-predictive control". In: *Computers & Chemical Engineering* 22.4, pp. 603–612.

Banga, J. R., E. Balsa-canto, C. G. Moles, and A. A. Alonso (2003). "Dynamic Optimization of Bioreactors: A Review". In: *Proceedings of the National Academy of Sciences*, pp. 1–21.

Banga, J. R., E. Balsa-Canto, C. G. Moles, and A. a. Alonso (2005). "Dynamic optimization of bioprocesses: Efficient and robust numerical strategies". In: *Journal of Biotechnology* 117.4, pp. 407–419.

Banik, R. M., a. Santhiagu, and S. N. Upadhyay (2007). "Optimization of nutrients for gellan gum production by Sphingomonas paucimobilis ATCC-31461 in molasses based medium using response surface methodology". In: *Bioresource Technology* 98, pp. 792–797.

Baş, D. and I. H. Boyacı (2007). "Modeling and optimization I: Usability of response surface methodology". In: *Journal of Food Engineering* 78.3, pp. 836–845.

Ben Salah, R., K. Chaari, S. Besbes, N. Ktari, C. Blecker, C. Deroanne, and H. Attia (2010). "Optimisation of xanthan gum production by palm date (Phoenix dactylifera L.) juice by-products using response surface methodology". In: *Food Chemistry* 121.2, pp. 627–633.

Bishop, C. M. (1995). *Neural Networks for Pattern Recognition*. Vol. 92. 440. Birmingham: Oxford University Press.

Bonde, G. J., F. E. Carlsen, and C. E. Jensen (1957). "Production of hyaluronic acid by Pseudomonas aeruginosa". In: *Acta Pharmacologica et Toxicologica (Copenh)* 13.2, pp. 205–212.

Bro, R. and A. K. Smilde (2014). “Principal component analysis”. In: *Analytical Methods* 6.

Bryson, A. E. and Y.-C. Ho (1975). *Applied Optimal Control*. New York: Taylor and Francis Group.

Byun, S.-G., M.-D. Kim, W.-H. Lee, K.-J. Lee, N. S. Han, and J.-H. Seo (2007). “Production of GDP-L-fucose, L-fucose donor for fucosyloligosaccharide synthesis, in recombinant Escherichia coli.” In: *Applied microbiology and biotechnology* 74.4, pp. 768–775.

Cacik, F., R. G. Dondo, and D. Marqués (2001). “Optimal control of a batch bioreactor for the production of xanthan gum”. In: *Computers and Chemical Engineering* 25, pp. 409–418.

Carlson, R. and F. Srienc (2004). “Fundamental Escherichia coli biochemical pathways for biomass and energy production: identification of reactions.” In: *Biotechnology and bioengineering* 85.1, pp. 1–19.

Caspi, R., H. Foerster, C. A. Fulcher, R. Hopkinson, J. L. Ingraham, P. Kaipa, M. Krummenacker, S. Paley, J. Pick, S. Y. Rhee, C. Tissier, P. Zhang, and P. D. Karp (2006). “Metacyc: a multiorganism database of metabolic pathways and enzymes”. In: *Nucleic Acids Research* 34, pp. 511–516.

Cerning, J., C. Bouillanne, M. Landon, and M. Desmazeaud (1992). “Isolation and Characterization of Exopolysaccharides from Slime-Forming Mesophilic Lactic Acid Bacteria”. In: *Journal of Dairy Science* 75, pp. 692–699.

Cescutti, P., A. Kallioinen, G. Impallomeni, R. Toffanin, P. Pollesello, M. Leisola, and T. Eerikäinen (2005). “Structure of the exopolysaccharide produced by Enterobacter amnigenus.” In: *Carbohydrate research* 340.3, pp. 439–47.

Chaitali, M., M. Kapadi, G. K. Suraishkumar, and R. D. Gudi (2003). “Productivity improvement in Xanthan gum fermentation using multiple substrate optimization”. In: *Biotechnology Progress* 19.4, pp. 1190–1198.

Chen, L., O. Bernard, G. Bastin, and P. Angelov (2000). "Hybrid modelling of biotechnological processes using neural networks". In: *Control Engineering Practice* 8.7, pp. 821–827.

Chen, Y. H., J. Li, L. Liu, H. Z. Liu, and Q. Wang (2012). "Optimization of flask culture medium and conditions for hyaluronic acid production by a streptococcus equisimilis mutant NC2168". In: *Brazilian Journal of Microbiology* 43.4, pp. 1553–1561.

Cui, J. D. and J. Q. Qiu (2012). "Production of extracellular water-insoluble polysaccharide from Pseudomonas sp." In: *Journal of Agricultural and Food Chemistry* 60, pp. 4865–4871.

Cuthbertson, L., I. L. Mainprize, J. H. Naismith, and C. Whitfield (2009). "Pivotal roles of the outer membrane polysaccharide export and polysaccharide copolymerase protein families in export of extracellular polysaccharides in gram-negative bacteria." In: *Microbiology and molecular biology reviews : MMBR* 73.1, pp. 155–177.

DeAngelis, P. L., W. Jing, R. R. Drake, and A. M. Achyuthan (1998). "Identification and molecular cloning of a unique hyaluronan synthase from Pasteurella multocida". In: *Journal of Biological Chemistry* 273.14, pp. 8454–8458.

Doran, P. M. (1995). *Bioprocess Engineering Principles*. Elsevier Science & Technology Books.

Duboc, P and U von Stockar (1998). "Systematic errors in data evaluation due to ethanol stripping and water vaporization." In: *Biotechnology and bioengineering* 58.4, pp. 428–39.

Faria, S., P. A. Vieira, M. M. Resende, E. J. Ribeiro, and V. L. Cardoso (2010). "Application of a model using the phenomenological approach for prediction of growth and xanthan gum production with sugar cane broth in a batch process". In: *LWT - Food Science and Technology* 43.3, pp. 498–506.

Fialho, A. M., L. M. Moreira, A. T. Granja, A. O. Popescu, K. Hoffmann, and I. Sá-Correia (2008). "Occurrence, production, and applications of gellan: Current state and perspectives". In: *Applied Microbiology and Biotechnology* 79, pp. 889–900.

Folch-Fortuny, A., R. Marques, R. Oliveira, and A. Ferrer. "Principal Elementary Mode Analysis". In: *Molecular Biosystems* (submitted paper).

Freitas, F., V. D. Alves, J. Pais, N. Costa, C. Oliveira, L. Mafra, L. Hilliou, R. Oliveira, and M. A. M. Reis (2009). "Characterization of an extracellular polysaccharide produced by a Pseudomonas strain grown on glycerol". In: *Bioresource technology* 100.2, pp. 859–65.

Freitas, F., V. D. Alves, J. Pais, M. Carvalheira, N. Costa, R. Oliveira, and M. A. M. Reis (2010). "Production of a new exopolysaccharide (EPS) by Pseudomonas oleovorans NRRL B-14682 grown on glycerol". In: *Process Biochemistry* 45.3, pp. 297–305.

Freitas, F., V. D. Alves, and M. A. M. Reis (2011a). "Advances in bacterial exopolysaccharides: from production to biotechnological applications." In: *Trends in biotechnology* 29.8, pp. 388–98.

Freitas, F., V. D. Alves, C. A. V. Torres, M. Cruz, I. Sousa, M. J. Melo, A. M. Ramos, and M. A. M. Reis (2011b). "Fucose-containing exopolysaccharide produced by the newly isolated Enterobacter strain A47 DSM 23139". In: *Carbohydrate Polymers* 83.1, pp. 159–165.

Gancel, F. and G. Novel (1994). "Exopolysaccharide Production by Streptococcus salivarius ssp. thermophilus Cultures. 1. Conditions of Production". In: *Journal of Dairy Science* 77, pp. 685–688.

García-Ochoa, F., V. Santos, and A. Alcón (1995). "Xanthan gum production: An unstructured kinetic model". In: *Enzyme and Microbial Technology* 17.3, pp. 206–217.

García-Ochoa, F., V. E. Santos, and A. Alcón (1996). "Simulation of xanthan gum production by a chemically structured kinetic model". In: *Mathematics and Computers in Simulation* 42, pp. 187–195.

García-Ochoa, F., V. E. Santos, and A. Alcon (1998). "Metabolic structured kinetic model for xanthan production". In: *Enzyme and Microbial Technology* 23, pp. 75–82.

García-Ochoa, F., V. E. Santos, J. A. Casas, E. Gómez, and E. Go (2000). "Xanthan gum: production, recovery, and properties." In: *Biotechnology Advances* 18.7, pp. 549–79.

García-Ochoa, F., V. E. Santos, and A. Alcon (2004). "Chemical structured kinetic model for xanthan production". In: *Enzyme and Microbial Technology* 35, pp. 284–292.

Gayen, K. and K. V. Venkatesh (2006). "Analysis of optimal phenotypic space using elementary modes as applied to Corynebacterium glutamicum." In: *BMC bioinformatics* 7, p. 445.

Gorret, N, J. L. Maubois, J. M. Engasser, and M Ghoul (2001). "Study of the effects of temperature , pH and yeast extract on growth and exopolysaccharides production by Propionibacterium acidi-propionici on milk microfiltrate using a response surface methodology". In: *Journal of Applied Microbiology*, pp. 788–796.

Hay, I. D., K. Gatland, A. Campisano, J. Z. Jordens, and B. H. A. Rehm (2009). "Impact of alginate overproduction on attachment and biofilm architecture of a supermucoid Pseudomonas aeruginosa strain". In: *Applied and Environmental Microbiology* 75.18, pp. 6022–6025.

Heinrich, R. and S. Schuster (1996). *The Regulation of Cellular Systems*. New York: Chapman & Hall.

Janssen, P., L. Goldovsky, V. Kunin, N. Darzentas, and C. A. Ouzounis (2005). "Genome coverage, literally speaking". In: 6.5, pp. 9–11.

Jarman, T. R. and G. W. Pace (1984). “Energy requirements for microbial exopolysaccharide synthesis”. In: *Archives of Microbiology* 137, pp. 231–235.

Kanehisa, M. and S. Goto (2000). “KEGG: Kyoto Encyclopedia of Genes and Genomes”. In: *Nucleic Acids Research* 28.1, pp. 27–30.

Keseler, I. M., J. Collado-Vides, S. Gama-Castro, J. Ingraham, S. Paley, I. T. Paulsen, M. Peralta-Gil, and P. D. Karp (2005). “EcoCyc: a comprehensive database resource for Escherichia coli”. In: *Nucleic Acids Research* 33, pp. 334–337.

Klamt, S. (2006). “Generalized concept of minimal cut sets in biochemical networks.” In: *Bio Systems* 83.2-3, pp. 233–47.

Klamt, S. and J. Stelling (2002). “Combinatorial Complexity of Pathway Analysis in Metabolic Networks”. In: *Molecular Biology Reports* 29, pp. 233–236.

— (2003). “Two approaches for metabolic pathway analysis?” In: *Trends in biotechnology* 21.2, pp. 64–9.

Kumar, A. S., K. Mody, and B. Jha (2007). “Bacterial exopolysaccharides - a perception.” In: *Journal of Basic Microbiology* 47.2, pp. 103–17.

Lee, H. S., S. H. Park, J. H. Lee, and H. K. Lee (2001). “Effect of aeration rates on production of extracellular polysaccharide, EPS-R, by marine bacteriumHahella chejuensis”. In: *Biotechnology and Bioprocess Engineering* 6, pp. 359–362.

Lee, J., S. Y. Lee, S. Park, and A. P. J. Middelberg (1999). “Control of fed-batch fermentations”. In: *Biotechnology Advances* 17.1, pp. 29–48.

Lee, S. Y., D.-Y. Lee, and T. Y. Kim (2005). “Systems biotechnology for strain improvement”. In: *Trends in Biotechnology* 23.7, pp. 349–358.

Lehninger, A. L., D. L. Nelson, and M. M. Cox (2000). *Lehninger Principles of Biochemistry*. Third Edit. New York: W. H. Freeman.

Leonard, J. A., M. A. Kramer, and L. H. Ungar (1992). “A Neural Network Architecture that computes its Own Reliability”. In: *Computers & Chemical Engineering* 16.9, pp. 819–835.

Li, H., H. Xu, S. Li, X. Feng, and P. Ouyang (2012). "Optimization of exopolysaccharide welan gum production by Alcaligenes sp. CGMCC2428 with Tween-40 using response surface methodology". In: *Carbohydrate Polymers* 87.2, pp. 1363–1368.

Lundstedt, T., E. Seifert, L. Abramo, B. Thelin, A. Nystrom, P. Jarle, and R. Bergman (1998). "Experimental design and optimization". In: *Chemometrics and Intelligent Laboratory Systems* 42.1-2, pp. 3–40.

Madigan, M. T., J. M. Martinko, and J. Parker (2003). *Brock Biology of Microorganisms*. Tenth Edit. New Jersey: Pearson Education, Inc.

Matsushita, M. (1990). "Curdlan, a (1—-3)-beta-D-glucan from Alcaligenes faecalis var. myxogenes IFO13140, activates the alternative complement pathway by heat treatment". In: *Immunology letters* 26.1, pp. 95–97.

Menküc, B. S., C. Gille, H.-G. Holzhütter, and H.-g. Holzhuetter (2008). "Computer aided optimization of carbon atom labeling for tracer experiments." In: *Genome informatics. International Conference on Genome Informatics* 20, pp. 270–6.

Moreno, J., M. Vargas, H. Olivares, J. Rivas, and M. G. Guerrero (1998). "Exopolysaccharide production by the cyanobacterium Anabaena sp. ATCC 33047 in batch and continuous culture". In: *Journal of Biotechnology* 60.3, pp. 175–182.

Naessens, M., R. Vercauteren, and E. J. Vandamme (2004). "Three-factor response surface optimization of the production of intracellular dextran dextrinase by Gluconobacter oxydans". In: *Process Biochemistry* 39, pp. 1299–1304.

Nguyen, V. T., B. Flanagan, M. J. Gidley, and G. a. Dykes (2008). "Characterization of cellulose production by a Gluconacetobacter xylinus strain from Kombucha". In: *Current Microbiology* 57, pp. 449–453.

Nielsen, J., J. Villadsen, and G. Lidén (2003). *Bioreaction Engineering Principles*. 2nd Editio. New York: Kluwer Academic / Plenum Publishers.

Oberhardt, M. A., B. Ø. Palsson, and J. A. Papin (2009). "Applications of genome-scale metabolic reconstructions". In: *Molecular Systems Biology* 5.320, pp. 1–15.

Oliveira, M. R. de, R. S. S. F. da Silva, J. B. Buzato, and M. A. P. C. Celligoi (2007). "Study of levan production by Zymomonas mobilis using regional low-cost carbohydrate sources". In: *Biochemical Engineering Journal* 37, pp. 177–183.

Oliveira, R. (2004). "Combining first principles modelling and artificial neural networks: a general framework". In: *Computers & Chemical Engineering* 28.5, pp. 755–766.

Öner, E. T. (2013). "Microbial Production of Extracellular Polysaccharides from Biomass". In: *Pretreatment Techniques for Biofuels and Biorefineries*. Ed. by F. Zhen. Springer. Chap. 2, pp. 35–56.

Palsson, B. Ø. (2006). *Systems Biology: properties of reconstructed networks*. Cambridge: Cambridge University Press.

Panesar, P. S., Y. Chavan, H. K. Chopra, and J. F. Kennedy (2012). "Production of microbial cellulose: Response surface methodology approach". In: *Carbohydrate Polymers* 87.1, pp. 930–934.

Peña, C., M. Millán, and E. Galindo (2008). "Production of alginate by Azotobacter vinelandii in a stirred fermentor simulating the evolution of power input observed in shake flasks". In: *Process Biochemistry* 43, pp. 775–778.

Pfeiffer, T., I. Sánchez-Valdenebro, J. C. Nuño, F. Montero, and S. Schuster (1999). "METATOOL: for studying metabolic networks." In: *Bioinformatics (Oxford, England)* 15.3, pp. 251–7.

Philbe, J. L. (2002). *Nouveau microorganisme de la famille des Enterobacteriaceae*.

Price, N. D., J. L. Reed, and B. Ø. Palsson (2004). "Genome-scale models of microbial cells: evaluating the consequences of constraints." In: *Nature reviews. Microbiology* 2.11, pp. 886–97.

Psichogios, D. C. and L. H. Ungar (1992). "A hybrid neural network-first principles approach to process modeling". In: *AIChE Journal* 38.10, pp. 1499–1511.

Psomas, S. K., M. Liakopoulou-Kyriakides, and D. a. Kyriakidis (2007). "Optimization study of xanthan gum production using response surface methodology". In: *Biochemical Engineering Journal* 35, pp. 273–280.

Quek, L.-E. and L. K. Nielsen (2014). "A depth-first search algorithm to compute elementary flux modes by linear programming." In: *BMC systems biology* 8, p. 94.

Ramberg, J. E., E. D. Nelson, and R. a. Sinnott (2010). "Immunomodulatory dietary polysaccharides: a systematic review of the literature." In: *Nutrition journal* 9.1, p. 54.

Rani, K. Y. and V. S. R. Rao (1999). "Control of fermenters - A review". In: *Bioprocess Engineering* 21.1, pp. 77–88.

Reder, C. (1988). "Metabolic Control Theory: a structural approach". In: *Journal of Theoretical Biology* 135.2, pp. 175–201.

Reed, J. L., I. Famili, I. Thiele, and B. O. Palsson (2006). "Towards multidimensional genome annotation". In: *Nature reviews. Genetics* 7.2, pp. 130–41.

Rehm, B. H. A., ed. (2009). *Microbial Production of Biopolymers and Polymer Precursors: applications and perspectives*. Caister Academic Press.

— (2010). "Bacterial polymers: biosynthesis, modifications and applications." In: *Nature reviews. Microbiology* 8.8, pp. 578–92.

Rosalam, S. and R. England (2006). "Review of xanthan gum production from unmodified starches by Xanthomonas comprestris sp." In: *Enzyme and Microbial Technology* 39.2, pp. 197–207.

Rosso, L., J. R. Lobry, S. Bajard, and J. P. Flandrois (1995). "Convenient Model To Describe the Combined Effects of Temperature and pH on Microbial Growth". In: *Applied and Envirnomental Microbiology* 61.2, pp. 610–616.

Schuster, S., T. Pfeiffer, F. Moldenhauer, I. Koch, and T. Dandekar (2002). "Exploring the pathway structure of metabolism: decomposition into subnetworks and application to Mycoplasma pneumoniae." In: *Bioinformatics (Oxford, England)* 18.2, pp. 351–61.

Schuster, S. and C. Hilgetag (1994). “On Elementary Flux Modes in Biochemical Reaction Systems at Steady State”. In: *Journal of Biological Systems* 2.2, pp. 165–182.

Schuster, S., D. A. Fell, and T. Dandekar (2000). “A general definition of metabolic pathways useful for systematic organization and analysis of complex metabolic networks.” In: *Nature biotechnology* 18.3, pp. 326–32.

Silbir, S., S. Dagbagli, S. Yegin, T. Baysal, and Y. Goksungur (2014). “Levan production by Zymomonas mobilis in batch and continuous fermentation systems”. In: *Carbohydrate Polymers* 99, pp. 454–461.

Simutis, R., I. Havlik, F. Schneider, M. Dors, and A. Lubbert (1995). “Artificial Neural Networks of improved reliability for industrial process supervision”. In: *Preprints of Sixth International Conference on Computer Applications in Biotechnology*. Ed. by A. Munack and K. Schugerl, pp. 59–65.

Song, H.-S., J. A. Morgan, and D. Ramkrishna (2009). “Systematic Development of Hybrid Cybernetic Models : Application to Recombinant Yeast Co-Consuming Glucose and Xylose”. In: 103.5, pp. 984–1002.

Stelling, J., S. Klamt, K. Bettenbrock, S. Schuster, and E. D. Gilles (2002). “Metabolic network structure determines key aspects of functionality and regulation.” In: *Nature* 420.6912, pp. 190–3.

Storn, R. and K Price (1997). “Differential evolution–a simple and efficient heuristic for global optimization over continuous spaces”. In: *Journal of global optimization*, pp. 341–359.

Stosch, M. von, R. Oliveira, J. Peres, and S. Feyo de Azevedo (2014). “Hybrid semi-parametric modeling in process systems engineering: Past, present and future”. In: *Computers & Chemical Engineering* 60, pp. 86–101.

Sutherland, I. W. (1993). “Biosynthesis of Extracellular Polysaccharides (Exopolysaccharides)”. In: *Industrial Gums - Polysaccharides and their Derivatives*. Ed. by R. L. Whistler and J. N. Bemiller. 3rd. Elsevier Inc. Chap. 4.

Sutherland, I. W. (2001). “Microbial polysaccharides from Gram-negative bacteria”. In: *International Dairy Journal* 11.9, pp. 663–674.

Teixeira, A., A. E. Cunha, J. J. Clemente, J. L. Moreira, H. J. Cruz, P. M. Alves, M. J. T. Carrondo, and R. Oliveira (2005). “Modelling and optimization of a recombinant BHK-21 cultivation process using hybrid grey-box systems”. In: *Journal of Biotechnology* 118, pp. 290–303.

Terzer, M., N. D. Maynard, M. W. Covert, and J. Stelling (2009). “Genome-scale metabolic networks”. In: *Science* 1.December, pp. 285–297.

Thompson, M. M. and M. A. Kramer (1994). “Modeling chemical processes using prior knowledge and neural networks”. In: *AlChE Journal* 40.8, pp. 1328–1340.

Torres, C. A. V., R. Marques, S. Antunes, V. D. Alves, I. Sousa, A. Maria, R. Oliveira, F. Freitas, and M. A. M. Reis (2011). “Kinetics of production and characterization of the fucose-containing exopolysaccharide from Enterobacter A47”. In: *Journal of Biotechnology* 156.4, pp. 261–267.

Torres, C. A. V., S. Antunes, A. R. Ricardo, C. Grandfils, V. D. Alves, F. Freitas, and M. A. M. Reis (2012). “Study of the interactive effect of temperature and pH on exopolysaccharide production by Enterobacter A47 using multivariate statistical analysis.” In: *Bioresource technology* 119, pp. 148–56.

Torres, C. A. V., R. Marques, A. R. V. Ferreira, S. Antunes, C. Grandfils, F. Freitas, and M. A. M. Reis (2014). “Impact of glycerol and nitrogen concentration on Enterobacter A47 growth and exopolysaccharide production.” In: *International journal of biological macromolecules*, pp. 4–9.

Trinh, C. T., R. Carlson, A. Wlaschin, and F. Srienc (2006). “Design, construction and performance of the most efficient biomass producing E. coli bacterium.” In: *Metabolic engineering* 8.6, pp. 628–38.

Trinh, C. T., A. Wlaschin, and F. Srienc (2009). “Elementary mode analysis: A useful metabolic pathway analysis tool for characterizing cellular metabolism”. In: *Applied Microbiology and Biotechnology* 81.5, pp. 813–826.

Troy, F. A., F. E. Frerman, and E. C. Heath (1971). "The Biosynthesis of Capsular Polysaccharide in Aerobacter aerogenes". In: *The Journal of Biological Chemistry* 246.1, pp. 118–133.

Vanhooren, P. and E. J. Vandamme (1998). "Biosynthesis, physiological role, use and fermentation process characteristics of bacterial exopolysaccharides". In: *Recent Res. Devel. Fermen. Bioeng.* 1, pp. 253–299.

Wagner, C. and R. Urbanczik (2005). "The geometry of the flux cone of a metabolic network." In: *Biophysical journal* 89.6, pp. 3837–45.

Weber, J., A. Kayser, and U. Rinas (2005). "Metabolic flux analysis of Escherichia coli in glucose-limited continuous culture. II . Dynamic response to famine and feast, activation of the methylglyoxal pathway and oscillatory behaviour". In: *Microbiology* 151, pp. 707–716.

Whitfield, C. and A. Paiment (2003). "Biosynthesis and assembly of Group 1 capsular polysaccharides in Escherichia coli and related extracellular polysaccharides in other bacteria". In: *Carbohydrate Research* 338, pp. 2491–2502.

Wilhelm, T., J. Behre, and S. Schuster (2004). "Analysis of structural robustness of metabolic networks". In: *Systems Biology* 1.1, pp. 114–120.

Wilson, W. a., P. J. Roach, M. Montero, E. Baroja-Fernández, F. J. Muñoz, G. Eydallin, A. M. Viale, and J. Pozueta-Romero (2010). "Regulation of glycogen metabolism in yeast and bacteria". In: *FEMS Microbiology Reviews* 34, pp. 952–985.

Yang, Y. L. (2002). *Novel microorganism isolated from Chinese elm (Ulmus sp.) and process for preparing exopolysaccharides by employing the microorganism.*

Zabot, G. L., J. Mecca, M. Mesomo, M. F. Silva, V. D. Prá, D. de Oliveira, J. V. Oliveira, F. Castilhos, H. Treichel, and M. A. Mazutti (2011). "Hybrid modeling of xanthan gum bioproduction in batch bioreactor." In: *Bioprocess and biosystems engineering* 34.8, pp. 975–86.

Zeng, X., D. P. Small, and W. Wan (2011). “Statistical optimization of culture conditions for bacterial cellulose production by Acetobacter xylinum BPR 2001 from maple syrup”. In: *Carbohydrate Polymers* 85.3, pp. 506–513.

APPENDIX A

Appendix

A.1 List of metabolic reactions

The following list contains all the metabolic reactions in the metabolic network for *Enterobacter* A47. The reversibility of the reactions are defined by the symbol "=>" for irreversible reactions and "=" for reversible reactions.

Glycerol Dissimilation

R11: Glyc + ATP => Glyc3P

R12: Glyc3P => DHAP + NADH

Glycolysis/gluconeogenesis

R21: Glu6P = Fru6P

R22: Fru6P + ATP => Fru16P

R23: Fru16P => Fru6P

R24: Fru16P = DHAP + GA3P

R25: GA3P = DHAP

R26: GA3P = PG + ATP + NADH

R27: PG = PEP

R28: PEP => Pyr + ATP

R29: Pyr + 2 ATP => PEP

Pentose/phosphate pathway

R31: Glu6P => Ribu5P + 2 NADPH + CO_2

R32: Ribu5P = Xyl5P

R33: Ribu5P = Rib5P

R34: Rib5P + Xyl5P = Sed7P + GA3P

R35: GA3P + Sed7P = Ery4P + Fru6P

R36: Ery4P + Xyl5P = GA3P + Fru6P

Pyruvate metabolism

R41: Pyr => AcCoA + CO2 + NADH

R42: Pyr + NADH = Lac

R43: AcCoA + 2 NADH = EtOH

R44: AcCoA => Ac + ATP

R45: Pyr => AcCoA + For

R46: Pyr => But12OH

Tricarboxylic acid cycle

R51: OxA + AcCoA => Cit

R52: Cit = IsoCit

R53: IsoCit => aKG + NADH + CO2

R54: aKG => NADH + SuccCoA + CO2

R55: SuccCoA = Succ + ATP

R56: Succ = Fum + FADH

R57: Fum = Mal

R58: Mal = OxA + NADH

Glyoxylate shunt

R61: IsoCit => Succ + Glyox

R62: AcCoA + Glyox => Mal

Anaplerotic reactions

R71: PEP + CO2 => OxA

R72: Mal => Pyr + NADH + CO2

R73: OxA + ATP => PEP + CO2

Oxidative phosphorilation

R81: NADH + 0.5 O2 => 2 ATP

R82: FADH + 0.5 O2 => ATP

R83: NADH => FADH

R84: NADH = NADPH

EPS building block synthesis

R91: Glu6P + ATP => Glu_residue

R92: Glu6P + ATP => Gal_residue

R93: Fru6P + NADPH + ATP => Fuc_residue

R94: Glu6P + NADPH + ATP=>Rha_residue

R95: Glu6P + ATP => GlcA_residue +
+ 2 NADH

R96: Glu6P + 2 ATP + AcCoA + NH3 =>
=> GlcN_residue

R97: AcCoA => Acetyl_residue

R98: PEP => Pyruvyl_residue

R99: SuccCoA => Succinyl_residue

Transport reactions

TR1: Glyc_extracellular => Glyc

TR2: NH3_extracellular => NH3

TR3: O2_extracellular => O2

TR4: CO2 => CO2_extracellular

TR5: Ac => Ac_extracellular

TR6: For => For_extracellular

TR7: Lac => Lac_extracellular

TR8: EtOH => EtOH_extracellular

TR9: But23OH => But23OH_extracellular

A.2 List of metabolites

ATP	Adenosine triphosphate
Ac	Acetate
AcCoA	Acetyl-coenzyme-A
But23dOH	2,3-Butanediol
CO2	Carbon Dioxide
Cit	Citrate
DHAP	Dihidroxyacetone phosphate
Ery4P	Erythrose-4-phosphate
EtOH	Ethanol
FADH	Flavin adenine dinucleotide
For	Formate
Fru6P	Fructose-6-phosphate
Fru16P	Fructose-1,6-biphosphate
Fum	Fumarate
GA3P	Glyceraldehyde-3-phosphate
Glc6P	Glucose-6-phosphate
Glyc	Glycerol
Glyc3P	Glycerol-3-Phosphate
Glyox	Glyoxylate
IsoCit	Isocitrate
Lac	Lactate
Mal	Malate
NADH	Nicotinamide adenine dinucleotide
NADPH	Nicotinamide adenine dinucleotide phosphate
NH3	Ammonia
O2	Oxygen
OxA	Oxaloacetate
PEP	Phosphoenolpyruvate
PG	Phosphoglycerate
Pyr	Pyruvate
Rib5P	Ribose-5-phosphate
Ribu5P	Ribulose-5-phosphate
Sed7P	Sedoheptulose-7-phosphate
Succ	Succinate
SuccCoA	Succinyl-coenzyme-A
Xyl5P	Xylulose-5-phosphate
aKG	Alpha-ketoglutarate

A.3 Stoichiometric matrix

	R11	R12	R21	R22	R23	R24	R25	R26	R27	R28	R29	R31	R32
ATP	-1	0	0	-1	0	0	0	1	0	1	-2	0	0
Ac	0	0	0	0	0	0	0	0	0	0	0	0	0
AcCoA	0	0	0	0	0	0	0	0	0	0	0	0	0
But23dOH	0	0	0	0	0	0	0	0	0	0	0	0	0
CO2	0	0	0	0	0	0	0	0	0	0	0	1	0
Cit	0	0	0	0	0	0	0	0	0	0	0	0	0
DHAP	0	1	0	0	0	1	1	0	0	0	0	0	0
Ery4P	0	0	0	0	0	0	0	0	0	0	0	0	0
EtOH	0	0	0	0	0	0	0	0	0	0	0	0	0
FADH	0	0	0	0	0	0	0	0	0	0	0	0	0
For	0	0	0	0	0	0	0	0	0	0	0	0	0
Fru6P	0	0	1	-1	1	0	0	0	0	0	0	0	0
Fru16P	0	0	0	1	-1	-1	0	0	0	0	0	0	0
Fum	0	0	0	0	0	0	0	0	0	0	0	0	0
GA3P	0	0	0	0	0	1	-1	-1	0	0	0	0	0
Glu6P	0	0	-1	0	0	0	0	0	0	0	0	-1	0
Glyc	-1	0	0	0	0	0	0	0	0	0	0	0	0
Glyc3P	1	-1	0	0	0	0	0	0	0	0	0	0	0
Glyox	0	0	0	0	0	0	0	0	0	0	0	0	0
IsoCit	0	0	0	0	0	0	0	0	0	0	0	0	0
Lac	0	0	0	0	0	0	0	0	0	0	0	0	0
Mal	0	0	0	0	0	0	0	0	0	0	0	0	0
NADH	0	1	0	0	0	0	0	1	0	0	0	0	0
NADPH	0	0	0	0	0	0	0	0	0	0	0	2	0
NH3	0	0	0	0	0	0	0	0	0	0	0	0	0
O2	0	0	0	0	0	0	0	0	0	0	0	0	0
OxA	0	0	0	0	0	0	0	0	0	0	0	0	0
PEP	0	0	0	0	0	0	0	0	1	-1	1	0	0
PG	0	0	0	0	0	0	0	1	-1	0	0	0	0
Pyr	0	0	0	0	0	0	0	0	0	1	-1	0	0
Rib5P	0	0	0	0	0	0	0	0	0	0	0	0	0
Ribu5P	0	0	0	0	0	0	0	0	0	0	0	1	-1
Sed7P	0	0	0	0	0	0	0	0	0	0	0	0	0
Succ	0	0	0	0	0	0	0	0	0	0	0	0	0
SuccCoA	0	0	0	0	0	0	0	0	0	0	0	0	0
Xyl5P	0	0	0	0	0	0	0	0	0	0	0	0	1
aKG	0	0	0	0	0	0	0	0	0	0	0	0	0

...continued

	R33	R34	R35	R36	R41	R42	R43	R44	R45	R46	R51	R52
ATP	0	0	0	0	0	0	0	1	0	0	0	0
Ac	0	0	0	0	0	0	0	1	0	0	0	0
AcCoA	0	0	0	0	1	0	-1	-1	1	0	-1	0
But23dOH	0	0	0	0	0	0	0	0	0	1	0	0
CO2	0	0	0	0	1	0	0	0	0	0	0	0
Cit	0	0	0	0	0	0	0	0	0	0	1	-1
DHAP	0	0	0	0	0	0	0	0	0	0	0	0
Ery4P	0	0	1	-1	0	0	0	0	0	0	0	0
EtOH	0	0	0	0	0	0	1	0	0	0	0	0
FADH	0	0	0	0	0	0	0	0	0	0	0	0
For	0	0	0	0	0	0	0	0	1	0	0	0
Fru6P	0	0	1	1	0	0	0	0	0	0	0	0
Fru16P	0	0	0	0	0	0	0	0	0	0	0	0
Fum	0	0	0	0	0	0	0	0	0	0	0	0
GA3P	0	1	-1	1	0	0	0	0	0	0	0	0
Glu6P	0	0	0	0	0	0	0	0	0	0	0	0
Glyc	0	0	0	0	0	0	0	0	0	0	0	0
Glyc3P	0	0	0	0	0	0	0	0	0	0	0	0
Glyox	0	0	0	0	0	0	0	0	0	0	0	0
IsoCit	0	0	0	0	0	0	0	0	0	0	0	1
Lac	0	0	0	0	0	1	0	0	0	0	0	0
Mal	0	0	0	0	0	0	0	0	0	0	0	0
NADH	0	0	0	0	1	-1	-2	0	0	0	0	0
NADPH	0	0	0	0	0	0	0	0	0	0	0	0
NH3	0	0	0	0	0	0	0	0	0	0	0	0
O2	0	0	0	0	0	0	0	0	0	0	0	0
OxA	0	0	0	0	0	0	0	0	0	0	-1	0
PEP	0	0	0	0	0	0	0	0	0	0	0	0
PG	0	0	0	0	0	0	0	0	0	0	0	0
Pyr	0	0	0	0	-1	-1	0	0	-1	-1	0	0
Rib5P	1	-1	0	0	0	0	0	0	0	0	0	0
Ribu5P	-1	0	0	0	0	0	0	0	0	0	0	0
Sed7P	0	1	-1	0	0	0	0	0	0	0	0	0
Succ	0	0	0	0	0	0	0	0	0	0	0	0
SuccCoA	0	0	0	0	0	0	0	0	0	0	0	0
Xyl5P	0	-1	0	-1	0	0	0	0	0	0	0	0
aKG	0	0	0	0	0	0	0	0	0	0	0	0

...continued

	R53	R54	R55	R56	R57	R58	R61	R62	R71	R72	R73	R81
ATP	0	0	1	0	0	0	0	0	0	0	-1	2
Ac	0	0	0	0	0	0	0	0	0	0	0	0
AcCoA	0	0	0	0	0	0	0	-1	0	0	0	0
But23dOH	0	0	0	0	0	0	0	0	0	0	0	0
CO2	1	1	0	0	0	0	0	0	-1	1	1	0
Cit	0	0	0	0	0	0	0	0	0	0	0	0
DHAP	0	0	0	0	0	0	0	0	0	0	0	0
Ery4P	0	0	0	0	0	0	0	0	0	0	0	0
EtOH	0	0	0	0	0	0	0	0	0	0	0	0
FADH	0	0	0	1	0	0	0	0	0	0	0	0
For	0	0	0	0	0	0	0	0	0	0	0	0
Fru6P	0	0	0	0	0	0	0	0	0	0	0	0
Fru16P	0	0	0	0	0	0	0	0	0	0	0	0
Fum	0	0	0	1	-1	0	0	0	0	0	0	0
GA3P	0	0	0	0	0	0	0	0	0	0	0	0
Glu6P	0	0	0	0	0	0	0	0	0	0	0	0
Glyc	0	0	0	0	0	0	0	0	0	0	0	0
Glyc3P	0	0	0	0	0	0	0	0	0	0	0	0
Glyox	0	0	0	0	0	0	1	-1	0	0	0	0
IsoCit	-1	0	0	0	0	0	-1	0	0	0	0	0
Lac	0	0	0	0	0	0	0	0	0	0	0	0
Mal	0	0	0	0	1	-1	0	1	0	-1	0	0
NADH	1	1	0	0	0	1	0	0	0	1	0	-1
NADPH	0	0	0	0	0	0	0	0	0	0	0	0
NH3	0	0	0	0	0	0	0	0	0	0	0	0
O2	0	0	0	0	0	0	0	0	0	0	0	-0.5
OxA	0	0	0	0	0	1	0	0	1	0	-1	0
PEP	0	0	0	0	0	0	0	0	-1	0	1	0
PG	0	0	0	0	0	0	0	0	0	0	0	0
Pyr	0	0	0	0	0	0	0	0	0	1	0	0
Rib5P	0	0	0	0	0	0	0	0	0	0	0	0
Ribu5P	0	0	0	0	0	0	0	0	0	0	0	0
Sed7P	0	0	0	0	0	0	0	0	0	0	0	0
Succ	0	0	1	-1	0	0	1	0	0	0	0	0
SuccCoA	0	1	-1	0	0	0	0	0	0	0	0	0
Xyl5P	0	0	0	0	0	0	0	0	0	0	0	0
aKG	1	-1	0	0	0	0	0	0	0	0	0	0

...continued

	R82	R83	R84	TR1	TR2	TR3	TR4	TR5	TR6	TR7	TR8
ATP	1	0	0	0	0	0	0	0	0	0	0
Ac	0	0	0	0	0	0	0	-1	0	0	0
AcCoA	0	0	0	0	0	0	0	0	0	0	0
But23dOH	0	0	0	0	0	0	0	0	0	0	0
CO2	0	0	0	0	0	0	-1	0	0	0	0
Cit	0	0	0	0	0	0	0	0	0	0	0
DHAP	0	0	0	0	0	0	0	0	0	0	0
Ery4P	0	0	0	0	0	0	0	0	0	0	0
EtOH	0	0	0	0	0	0	0	0	0	0	-1
FADH	-1	1	0	0	0	0	0	0	0	0	0
For	0	0	0	0	0	0	0	0	-1	0	0
Fru6P	0	0	0	0	0	0	0	0	0	0	0
Fru16P	0	0	0	0	0	0	0	0	0	0	0
Fum	0	0	0	0	0	0	0	0	0	0	0
GA3P	0	0	0	0	0	0	0	0	0	0	0
Glu6P	0	0	0	0	0	0	0	0	0	0	0
Glyc	0	0	0	1	0	0	0	0	0	0	0
Glyc3P	0	0	0	0	0	0	0	0	0	0	0
Glyox	0	0	0	0	0	0	0	0	0	0	0
IsoCit	0	0	0	0	0	0	0	0	0	0	0
Lac	0	0	0	0	0	0	0	0	0	-1	0
Mal	0	0	0	0	0	0	0	0	0	0	0
NADH	0	-1	-1	0	0	0	0	0	0	0	0
NADPH	0	0	1	0	0	0	0	0	0	0	0
NH3	0	0	0	0	1	0	0	0	0	0	0
O2	-0.5	0	0	0	0	1	0	0	0	0	0
OxA	0	0	0	0	0	0	0	0	0	0	0
PEP	0	0	0	0	0	0	0	0	0	0	0
PG	0	0	0	0	0	0	0	0	0	0	0
Pyr	0	0	0	0	0	0	0	0	0	0	0
Rib5P	0	0	0	0	0	0	0	0	0	0	0
Ribu5P	0	0	0	0	0	0	0	0	0	0	0
Sed7P	0	0	0	0	0	0	0	0	0	0	0
Succ	0	0	0	0	0	0	0	0	0	0	0
SuccCoA	0	0	0	0	0	0	0	0	0	0	0
Xyl5P	0	0	0	0	0	0	0	0	0	0	0
aKG	0	0	0	0	0	0	0	0	0	0	0

...continued

	TR9	R91	R92	R93	R94	R95	R96	R97	R98	R99	ATP maint.
ATP	0	-1	-1	-1	-1	-1	-2	0	0	0	-1
Ac	0	0	0	0	0	0	0	0	0	0	0
AcCoA	0	0	0	0	0	0	-1	0	0	-1	0
But23dOH	-1	0	0	0	0	0	0	0	0	0	0
CO2	0	0	0	0	0	0	0	0	0	0	0
Cit	0	0	0	0	0	0	0	0	0	0	0
DHAP	0	0	0	0	0	0	0	0	0	0	0
Ery4P	0	0	0	0	0	0	0	0	0	0	0
EtOH	0	0	0	0	0	0	0	0	0	0	0
FADH	0	0	0	0	0	0	0	0	0	0	0
For	0	0	0	0	0	0	0	0	0	0	0
Fru6P	0	-1	0	0	0	0	0	0	0	0	0
Fru16P	0	0	0	0	0	0	0	0	0	0	0
Fum	0	0	0	0	0	0	0	0	0	0	0
GA3P	0	0	0	0	0	0	0	0	0	0	0
Glu6P	0	0	-1	-1	-1	-1	-1	0	0	0	0
Glyc	0	0	0	0	0	0	0	0	0	0	0
Glyc3P	0	0	0	0	0	0	0	0	0	0	0
Glyox	0	0	0	0	0	0	0	0	0	0	0
IsoCit	0	0	0	0	0	0	0	0	0	0	0
Lac	0	0	0	0	0	0	0	0	0	0	0
Mal	0	0	0	0	0	0	0	0	0	0	0
NADH	0	0	0	0	2	0	0	0	0	0	0
NADPH	0	-1	0	0	0	-1	0	0	0	0	0
NH3	0	0	0	0	0	0	-1	0	0	0	0
O2	0	0	0	0	0	0	0	0	0	0	0
OxA	0	0	0	0	0	0	0	0	0	0	0
PEP	0	0	0	0	0	0	0	-1	0	0	0
PG	0	0	0	0	0	0	0	0	0	0	0
Pyr	0	0	0	0	0	0	0	0	0	0	0
Rib5P	0	0	0	0	0	0	0	0	0	0	0
Ribu5P	0	0	0	0	0	0	0	0	0	0	0
Sed7P	0	0	0	0	0	0	0	0	0	0	0
Succ	0	0	0	0	0	0	0	0	0	0	0
SuccCoA	0	0	0	0	0	0	0	0	-1	0	0
Xyl5P	0	0	0	0	0	0	0	0	0	0	0
aKG	0	0	0	0	0	0	0	0	0	0	0

A.4 List of selected EMs in the PEMA solution

	EM227	EM241	EM329	EM414	EM559
R11	1.000	1.000	1.000	1.000	1.000
R12	1.000	1.000	1.000	1.000	1.000
R21	-0.400	-0.429	-0.545	-0.353	0.000
R22	0.000	0.000	0.000	0.000	0.000
R23	0.400	0.429	0.455	0.471	0.333
R24	-0.400	-0.429	-0.455	-0.471	-0.333
R25	-0.600	-0.571	-0.545	-0.529	-0.667
R26	0.200	0.143	0.136	0.088	0.333
R27	0.200	0.143	0.136	0.088	0.333
R28	0.000	0.000	0.136	0.000	0.000
R29	0.000	0.000	0.000	0.000	0.000
R31	0.000	0.000	0.136	0.088	0.000
R32	0.000	0.000	0.091	0.059	0.000
R33	0.000	0.000	0.045	0.029	0.000
R34	0.000	0.000	0.045	0.029	0.000
R35	0.000	0.000	0.045	0.029	0.000
R36	0.000	0.000	0.045	0.029	0.000
R41	0.000	0.143	0.136	0.000	0.000
R42	0.000	0.000	0.000	0.000	0.000
R43	0.000	0.000	0.000	0.000	0.000
R44	0.000	0.000	0.000	0.000	0.000
R45	0.200	0.000	0.000	0.000	0.000
R46	0.000	0.000	0.000	0.000	0.000
R51	0.000	0.000	0.000	0.000	0.000
R52	0.000	0.000	0.000	0.000	0.000
R53	0.000	0.000	0.000	0.000	0.000
R54	0.000	0.000	0.000	0.000	0.000
R55	0.000	0.000	0.000	-0.088	0.000
R56	0.000	0.000	0.000	-0.088	0.000
R57	0.000	0.000	0.000	-0.088	0.000
R58	-0.200	-0.143	0.000	-0.088	0.000
R61	0.000	0.000	0.000	0.000	0.000
R62	0.000	0.000	0.000	0.000	0.000
R71	0.200	0.143	0.000	0.088	0.000
R72	0.200	0.143	0.000	0.000	0.000
R73	0.000	0.000	0.000	0.000	0.000

...continued

	EM227	EM241	EM329	EM414	EM559
R81	0.000	0.000	0.000	0.000	0.000
R82	1.200	1.286	1.273	1.441	1.000
R83	1.200	1.286	1.273	1.529	1.000
R84	0.000	0.000	0.000	0.000	0.333
TR1	1.000	1.000	1.000	1.000	1.000
TR2	0.000	0.000	0.136	0.000	0.000
TR3	0.600	0.643	0.636	0.721	0.500
TR4	0.000	0.143	0.273	0.000	0.000
TR5	0.000	0.000	0.000	0.000	0.000
TR6	0.200	0.000	0.000	0.000	0.000
TR7	0.000	0.000	0.000	0.000	0.000
TR8	0.000	0.000	0.000	0.000	0.000
TR9	0.000	0.000	0.000	0.000	0.000
R91	0.000	0.000	0.000	0.176	0.333
R92	0.400	0.000	0.000	0.000	0.000
R93	0.000	0.429	0.000	0.000	0.000
R94	0.000	0.000	0.000	0.265	0.000
R95	0.000	0.000	0.273	0.000	0.000
R96	0.000	0.000	0.136	0.000	0.000
R97	0.000	0.000	0.000	0.000	0.333
R98	0.000	0.000	0.000	0.088	0.000
R99	0.200	0.143	0.000	0.000	0.000

A.5 List of selected EMs in the YA solution

	EM103	EM219	EM524	EM589	EM792	EM811	EM1019	EM1020	EM1021
R11	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
R12	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
R21	-0.467	-0.333	0.000	0.000	0.000	0.000	-0.500	-0.500	-0.500
R22	0.000	1.333	0.000	1.500	2.000	0.000	0.000	0.000	0.000
R23	0.467	1.667	0.484	1.500	2.000	0.000	0.500	0.500	0.500
R24	-0.467	-0.333	-0.484	0.000	0.000	0.000	-0.500	-0.500	-0.500
R25	-0.533	-0.667	-0.516	-1.000	-1.000	-1.000	-0.500	-0.500	-0.500
R26	0.067	0.333	0.032	1.000	1.000	1.000	0.000	0.000	0.000
R27	0.067	0.333	0.032	1.000	1.000	1.000	0.000	0.000	0.000
R28	0.000	0.000	0.032	0.500	0.000	0.000	0.000	0.000	0.000
R29	0.000	0.000	0.000	0.000	0.000	1.000	0.250	0.250	0.250
R31	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
R32	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
R33	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
R34	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
R35	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
R36	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
R41	0.067	0.000	0.032	0.500	0.000	0.000	0.000	0.000	0.000
R42	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
R43	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
R44	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
R45	0.000	0.333	0.000	0.000	1.000	0.000	0.000	0.000	0.000
R46	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
R51	0.000	0.000	0.032	0.000	0.000	0.000	0.000	0.000	0.000
R52	0.000	0.000	0.032	0.000	0.000	0.000	0.000	0.000	0.000
R53	0.000	0.000	0.032	0.000	0.000	0.000	0.000	0.000	0.000
R54	0.000	0.000	0.032	0.000	0.000	0.000	0.000	0.000	0.000
R55	0.000	0.000	0.032	-0.500	0.000	0.000	0.000	0.000	0.000
R56	0.000	0.000	0.032	-0.500	0.000	0.000	0.000	0.000	0.000
R57	0.000	0.000	0.032	-0.500	0.000	0.000	0.000	0.000	0.000
R58	-0.067	-0.333	0.032	-0.500	-1.000	-1.000	-0.250	-0.250	-0.250
R61	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
R62	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
R71	0.067	0.333	0.000	0.500	1.000	1.000	0.250	0.250	0.250
R72	0.067	0.333	0.000	0.000	1.000	1.000	0.250	0.250	0.250
R73	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

...continued

	EM103	EM219	EM524	EM589	EM792	EM811	EM1019	EM1020	EM1021
R81	0.733	1.333	0.677	0.000	0.000	0.000	0.000	1.000	1.000
R82	0.000	0.000	0.032	1.500	2.000	2.000	2.000	0.000	0.000
R83	0.000	0.000	0.000	2.000	2.000	2.000	2.000	0.000	0.000
R84	0.400	0.000	0.484	0.000	0.000	0.000	0.000	0.000	0.000
TR1	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
TR2	0.067	0.333	0.000	0.000	0.000	0.000	0.000	0.000	0.000
TR3	0.367	0.667	0.355	0.750	1.000	1.000	1.000	0.500	0.500
TR4	0.067	0.000	0.097	0.000	0.000	0.000	0.000	0.000	0.000
TR5	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
TR6	0.000	0.333	0.000	0.000	1.000	0.000	0.000	0.000	0.000
TR7	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
TR8	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
TR9	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
R91	0.000	0.000	0.484	0.000	0.000	0.000	0.000	0.000	0.000
R92	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.500	0.000
R93	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.500
R94	0.000	0.000	0.000	0.000	0.000	0.000	0.500	0.000	0.000
R95	0.400	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
R96	0.067	0.333	0.000	0.000	0.000	0.000	0.000	0.000	0.000
R97	0.000	0.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000
R98	0.000	0.000	0.000	0.500	0.000	0.000	0.000	0.000	0.000
R99	0.000	0.000	0.000	0.500	1.000	0.000	0.000	0.000	0.000