PONTIFICIA UNIVERSIDAD CATOLICA DE CHILE

SCHOOL OF ENGINEERING

BIOPROCESS ENGINEERING AND METABOLIC MODELING AS A ROADMAP FOR ENHANCED RECOMBINANT PROTEIN PRODUCTION IN *PICHIA PASTORIS*

PAULINA MACARENA TORRES PLAZA

Thesis submitted to the Office of Research and Graduate Studies in partial fulfillment of the requirements for the Degree of Doctor in Engineering Sciences

Advisor:

EDUARDO AGOSIN TRUMPER

Santiago de Chile, January, 2019

PONTIFICIA UNIVERSIDAD CATOLICA DE CHILE

SCHOOL OF ENGINEERING

BIOPROCESS ENGINEERING AND METABOLIC MODELING AS A ROADMAP FOR ENHANCED RECOMBINANT PROTEIN PRODUCTION IN *PICHIA PASTORIS*

PAULINA MACARENA TORRES PLAZA

Members of the Committee:

EDUARDO AGOSIN TRUMPER

J. RICARDO PEREZ CORREA

LUIS LARRONDO

PAU FERRER ALEGRE

MARCEL OTAVIO CERRI

JUAN de DIOS ORTÚZAR

Thesis submitted to the Office of Research and Graduate Studies in partial fulfillment of the requirements for the Degree of Doctor in Engineering Sciences

Santiago de Chile, January, 2019

"If you feel lost, disappointed, hesitant, or weak,
return to yourself, to who you are, here and now
and when you get there, you will discover yourself,
like a lotus flower in full bloom,
even in a muddy pond, beautiful and strong."
— Masaru Emoto, Secret Life of Water

To my family and friends

ACKNOWLEDGEMENTS

I want to express my gratitude to my supervisor, Prof. Eduardo Agosín, for his support and guidance throughout these professional and personal processes. Moreover, I would like to extend this to Professors Ricardo Pérez, Pedro Saa, Marcel Otavio Cerri, and Joan Albiol who provided technical support and guidance throughout many tasks.

I would like to give special thanks to Professor Pau Ferrer, who provided technical support and for accepting me as her visitor student in the Department of Chemical Engineering, Universitat Autònoma de Barcelona. In addition, I would also like to thank the rest of the members of the committee, Professors Luis Larrondo and Juan de Dios Ortuzar, for their time to correct the final manuscript and the thoughtful comments on it.

I would like to give special thanks to my *P. pastoris* teamwork, Alexandra, Francisco and Vasni, and my k_La teamwork, Bastian and Diego, without you nothing will be possible. Moreover, this thesis would not have been possible without the help of my Lab's friends; M. Carolina, Alexandra, Isabel, Francisco, Sebastian, Ingrid, Nicol, Rodrigo and Gabriela. Each one of you helped me in some way to fulfill this task, not only with professional support, if not that with your cheers, energy, time and love. Moreover, I wants to thanks my University's and Live's friends, Maria Ines, Teresa, Ingrid, La Sra Irma, Martin, Javiera, Veronica, Paulina, Patricio, Rocio, Marcela, Camila, Camila and Mireia who also played a relevant role in all my process.

I would like to thank my family, Roxana, Sebastian, Camila and my eight dogs, for their love and unconditional support. Moreover, I would like to thank to my guides, my angels, God and the Universe for their unconditional love and support.

Financial support: I really appreciate the financial support of the CONICYT through the Human Capital Formation Advanced Program National Doctoral Fellowship 2014 (Folio 21140759), Vice-rectory of Research UC (VRI), and the School of Engineering of Pontificia Universidad Católica de Chile to carry out my PhD studies.

LIST OF PAPERS

This thesis is based on the following papers, referred in the text by respective chapters:

Chapter I: **Torres P.,** Saa P., Albiol J., Ferrer P., Agosin E. **Revealing the metabolic effects of specific growth rate and oxygenation level on recombinant *Pichia pastoris* performance in glucose-limited continuous culture.** (Paper in revision in Microbial Cell Factories*)*.

Chapter II: Saitua F., **Torres P.,** Pérez-Correa J. R. & Agosin E.. (2017). **Dynamic Genome-Scale Metabolic Modeling of the Yeast *Pichia Pastoris***. *BMC* Systems Biology , 11:27.

Chapter III: **Torres P.,** Cerri M., Pérez-Correa J. R, Agosin E. (2016). **Automated algorithm to determine kLa considering system delay**. *Journal of Chemical Technology and Biotechnology* 92:7, 1630-1737

Chapter IV: **Torres P.,** de la Vega D., Cerri M., Pérez-Correa J. R., Agosin E. **Optimization of oxygen transfer in 1L Stirred Tank Bioreactors**. (Paper in preparation)

PROCEEDINGS

Parts of this work have also been presented at international congresses under the following references:

Francisco Saitua, **Paulina Torres**, Alexandra Lobos, J. Ricardo Pérez-Correa, Eduardo Agosin.
Dynamic Genome-Scale Metabolic Modeling: a Tool to Analyze and Design Aerobic Cultures of Pichia pastoris.
In: *Pichia* 2016.
Antalya, Turkey.
April, 2016 **(Oral Presentation).**

Francisco Saitua, **Paulina Torres**, Alexandra Lobos, J. Ricardo Pérez-Correa, Eduardo Agosín.
Genome-Scale Dynamic Flux Balance Analysis of Aerobic Batch Cultures of Recombinant Pichia pastoris Strains. I
In: 4th Conference on Constraint-Based Reconstruction and Analysis (COBRA 2015).
Heidelberg, Germany.
September, 2015 **(Poster).**

Alexandra Lobos, **Paulina Torres**, Francisco Saitua, Eduardo Agosin.
Sustainable production of the sweet protein thaumatin in Pichia pastoris.
In: 8th Conference on Recombinant Protein Production.
Mallorca, Spain.
April, 2015 **(Poster).**

FIGURE INDEX

TABLE INDEX

ABBREVIATION INDEX

AICc: Akaike's Information Criterion for low number of data points.

b0: Mean value of the response.

bi: Linear effects of i.

bii: Quadratic effects of i.

bij: Interaction effect between i and j.

C_e : Measured dissolved oxygen concentration[ppm] or [%].

$Ce_s(t)$: Simulated DO concentration [ppm] or [%].

C: Correlation coefficient matrix.

C: Dissolved oxygen concentration in the liquid phase of the bioreactor.

C*: Oxygen equilibrium concentration.

CC: Coefficients of confidence.

CF: Microbial Cell Factory.

CI: Confidence interval.

Cx : Biomass concentration.

DO: Dissolved oxygen [ppm] or [%].

E: Oxygen-excess conditions.

EE: Ear-Elephant impeller.

EEDP: Elephant Ear impellers operated in Down-Pumping mode.

EEUP: Elephant Ear impellers operated in Up-Pumping mode.

eSS: enhanced scatter search algorithm.

$F(t)$: Feed function for the fed-batch phase in [L/h].

FBA: Flux Balance Analysis.

FIM: Fisher Information Matrix.

F_{IN}: Feeding policy used in fed-batch cultures,

G: Glucose concentration in the medium [g/L].

GAME: Growth Associated Maintenance Energy.

G_F :Feed's glucose concentration [g/L].

g_{ik}: Relative sensitivity of parameter k on the state variable i.

GS-dFBA: Genome-Scale dynamic Flux Balance Analysis.

GSMM: Genome Scale Metabolic Models.

HIPPO: Heuristic Iterative Procedure for Parameter Optimization.

HPLC: High-Performance Liquid Chromatography.

HSA: Human Serum Albumin.

J: Objective function.

$J_{i,Original}$: Calibration objective function

$J_{i,Reduced}$: Calibration objective function achieved in dataset i using a reduced, *a priori* robust, modeling structure.

L: Oxygen-limited conditions.

lb: Lower bounds of the exchange reactions.

k_e : Sensitivity of the electrode, which is the inverse value of τ_e

K_G: Uptake half activity constant of this substrate [g/L].

$k_L a$: Volumetric mass transfer coefficient[1/s].

m: Slope of a part of the curve of instantaneous kLa v/s t.

m: Number of measured variables.

$m_{\text{ATP:}}$ ATP maintenance.

$\max_j\left(X_{ij}^{exp}\right)$: Maximum value measured for variable i.

MCF: Microbial Cell Factory.

MFA: Metabolic Fluxes Analysis

MNE: Mean Normalized Error.

MOMA: Minimization of Metabolic Adjustment algorithm.

MW: Molecular weight [g/mmol].

n: Number of measurements per variable.

N: Normoxic conditions.

N: Number of data.

N_{max}: Maximum number of data, which is necessary for use.

N_{min}: Minimum number of data, which is necessary for use.

NCDs: Non-Communicable Diseases.

NGAME: Non-Growth Associated Maintenance Energy.

OD: Optical density.

ODEs: Ordinary differential equations.

OTR: Oxygen transfer rate[mmol/h].

OUR: Oxygen Uptake Rate[mmol/h].

P: Product concentration in [g/L]

PAOX1: Methanol-inducible alcohol oxidase 1 promoter.

PG1: High-affinity glucose transporter gene GTH1.

PGAP: Glyceraldehyde-3-phosphate dehydrogenase promoter.

$P_{K:}$ k-th extracellular product concentration in [g/L]

PPP: Pentose Phosphate Pathway

q_{ATP} *: total* ATP production[mmol/g_{DCW}·h].

$q_{co2:}$ specific CO_2 production rate[mmol/g_{DCW}·h].

$Q_{j:}$ A weighting matrix given by the inverse of the measurement error covariance matrix assuming white and uncorrelated noise.

$q_{o2:}$ specific O_2 consumption rate[mmol/g_{DCW}·h].

QP: Quadratic Programming.

qs: specific glucose consumption rate[mmol/g_{DCW}·h].

q_{thau} specific production rate of secreted thaumatin[mmol/g_{DCW}·h].

R^2: Quality of the fitting.

RPP: Recombinant Protein Production.

RSS: Sum of the squared residuals.

RT: Rushton Turbine impeller.

S: Limiting substrate concentration in [g/L].

S: Stoichiometric Matrix (m x n).

SD: Simple Diffuser.

SR : Constant sampling rate [L/h].

t: Time [h].

TCA: Tricarboxylic acid cycle.

$t_{i::}$ Time at which the feed started for a given cultivation [h]

TSD: Triple Spin Diffuser

ub: Upper bounds of the exchange reactions.

V: Culture volume [L].

v: Vector of metabolic fluxes in [mmol/g_{DCW}·h].

v_{P_k}: Production rate of k [mmol/g_{DCW}·h].

$v_{G,Max}$: Maximum glucose uptake rate [mmol/g_{DCW}·h^{-1}].

v_G: Glucose uptake rate [mmol/g_{DCW}·h^{-1}].

V_i : Volume value at t_i [L].

X: Biomass concentration in [g/L].

X_i: Biomass value at t_i [g/L].

Xi: Independent variables.

$X_i(t)$: *i*-th state variable in time t.

X_{ij}^{exp}: Experimental value .

X_{ij}^{mod} : dFBA output of variable i and measurement j.

Y: Predicted response.

Y_{SX} : Experimental glucose-biomass yield [g_{DCW}/g].

Greek Letter:

α: Suboptimal growth coefficient.

$\varepsilon^2{}_i(t)$: Mean variance of the sensor.

$\varepsilon_i(t)$: Zero mean random value.

θ: Parameter space.

θ_k: *k*th parameter.

τ_e: Response time of the electrode's probe [s].

μ: Growth rate in [1/ h].

$\mu_{SET}(t)$: Time-dependent user-defined growth rate [1/ h].

.

LIST OF CONTENT

PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE
SCHOOL OF ENGINEERING

BIOPROCESS ENGINEERING AND METABOLIC MODELING AS A ROADMAP FOR ENHANCED RECOMBINANT PROTEIN PRODUCTION IN *PICHIA PASTORIS*

Thesis submitted to the Office of Research and Graduate Studies in partial fulfillment of the requirements for the Degree of Doctor in Engineering Sciences by

PAULINA MACARENA TORRES PLAZA

ABSTRACT

Pichia pastoris is recognized as a biotechnological workhorse for recombinant protein expression. Based on past achievements and novel developments, systems biotechnology of *P. pastoris* has significantly progressed over the last two decades.

In this doctoral thesis, a systematic analysis of operational conditions in conjunction with the development of computational toolboxes and optimal mechanical bioreactor configurations were developed as a roadmap to enhance the production of recombinant proteins through bioprocess engineering and metabolic modeling approaches. Although we used a *P. pastoris* strain expressing constitutively the sweet protein thaumatin, as case study, the developments achieved in this work are also applicable for the expression of other recombinant proteins of interest.

First, we present an integrated framework for revealing the metabolic effects of two operational parameters – specific growth rate and dissolved oxygen concentration - in glucose-limited continuous cultures. More specifically, we employed a rational experimental design to calculate the significant statistical effects from multiple chemostat data, which was later contextualized using a curated genome-scale metabolic model. Our

results revealed a negative effect of the oxygenation on the specific product formation rate, and a positive effect on the biomass yield. Notably, we identified a novel synergistic effect of both parameters on the specific product formation rate. Finally, model predictions indicated an opposite relationship between the oxygenation level and growth-associated ATP requirement, suggesting higher metabolic growth costs under low oxygenation.

We then assembled a robust dynamic genome-scale metabolic model for glucose-limited, aerobic cultures of *Pichia pastoris*. The model was employed to analyze the metabolic flux distribution of a fed-batch culture and to unravel genetic and process engineering strategies to improve the production of the recombinant Human Serum Albumin (HSA). Simulations of single knock-outs indicated that carbon deviations towards cysteine and tryptophan formation could improve 63 fold HSA production. Moreover, the model suggested that implementation of a decreasing specific growth rate during the feed phase of a fed-batch culture results in a 25% increase of the volumetric productivity of the protein.

Finally, optimization of Oxygen Transfer Rate (OTR) in 1-L reactors was carried out. For this purpose, we first formulated an automatic algorithm able to achieve reliable k_La estimations under different hydrodynamic conditions. Then, we presented a road map to optimize oxygen transfer rate in 1-L bioreactor, using different impeller-sparger configurations. The relative importance of aeration and agitation under several configurations and hydrodynamic conditions was assessed. Finally, we proposed a decision tree for the selection of the best configuration to improve oxygen transfer rates in bioreactors, according to viscosity and the range of operational parameters.

Overall, this thesis attempted to analyze and understand metabolic effects using rational design of experiments and metabolic modeling. Moreover, this systematic analysis in conjunction with the optimization of the oxygen transfer rate in bioreactors will allow to improve recombinant protein production in *Pichia pastoris*.

Members of the Doctoral Thesis Committee:

Eduardo Agosin Trumper

Jose Ricardo Perez Correa

Luis Larrondo Castro

Pau Ferrer Alegre

Marcel Otavio Cerri

Juan de Dios Ortuzar

Santiago, January 2019

PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE
ESCUELA DE INGENIERÍA

INGENERÍA DE BIOPROCESOS Y MODELAMIENTO METABÓLICO COMO UNA HOJA DE RUTA PARA MEJORAR LA PRODUCCIÓN DE PROTEÍNAS RECOMBINANTE EN *PICHIA PASTORIS*

Tesis enviada a la Dirección de Investigación y Postgrado en cumplimiento parcial de los requisitos para el grado de Doctor en Ciencias de la Ingeniería.

PAULINA MACARENA TORRES PLAZA

RESUMEN

Pichia pastoris es reconocida como un *workhorse* biotecnológico para la expresión de proteínas recombinantes. Basándose en sus logros pasados y novedosos desarrollos, la biotecnología de sistemas de *P. pastoris* ha progresado significativamente en las últimas dos décadas.

En esta tesis doctoral, el análisis sistemático de las condiciones operativas en conjunto con el desarrollo de herramientas computacionales y la selección de configuraciones mecánicas óptimas de bioreactores se llevaron a cabo como una hoja de ruta para mejorar la producción de proteínas recombinantes a través de la ingeniería de bioprocesos y el modelamiento metabólico. A pesar de que nosotros utilizamos una cepa de *P. pastoris* que expresaba constitutivamente la proteína endulzante taumatina, como caso de estudio, los desarrollos logrados también son aplicables para la expresión de otras proteínas recombinantes de interés.

Primero, presentamos un marco integrado para revelar los efectos metabólicos de dos parámetros operacionales – la tasa específica de crecimiento y la concentración de

oxígeno disuelto - en cultivos continuos limitados en glucosa. Más específicamente, empleamos un diseño experimental racional para calcular los efectos estadísticos significativos de los datos de múltiples quimiostatos, que luego fueron contextualizados utilizando un modelo metabólico a escala del genoma manualmente curado. Nuestros resultados revelaron un efecto negativo de la oxigenación en la tasa especifica de producción del producto y un efecto positivo en el rendimiento de la biomasa. En particular, identificamos un efecto sinérgico de ambos parámetros en la tasa específica de producción del producto. Finalmente, las predicciones del modelo indicaron una relación opuesta entre el nivel de oxigenación y el requerimiento de ATP asociado al crecimiento, sugiriendo un mayor gasto metabólico de crecimiento a bajo nivel de oxigenación.

En segundo lugar, construimos un modelo metabólico a escala genómica dinámico para cultivos aeróbicos de *Pichia pastoris* limitados en glucosa. El modelo se empleó para analizar la distribución del flujo metabólicos de un cultivo alimentado por lotes y para diseñar estrategias genéticas y de ingeniería de procesos para mejorar la producción de albúmina sérica humana recombinante. Simulaciones de *knock-outs* individuales indicaron que la desviación del carbono hacia la formación de cisteína y triptófano podría mejorar en 63 veces la producción de HSA. Además, el modelo sugirió que la implementación de una tasa de crecimiento específica decreciente durante la fase de alimentación de un cultivo alimentado por lotes produce un aumento del 25% de la productividad volumétrica de la proteína.

Finalmente, se llevó a cabo una optimización de la tasa de transferencia de oxígeno en reactores de 1 L. Para este propósito, primero discutimos la formulación de un algoritmo automático para la estimación de k_La bajo diferentes condiciones hidrodinámicas. Luego, presentamos una hoja de ruta para optimizar la tasa de transferencia de oxígeno en bioreactor de 1-L, utilizando diferentes configuraciones de aspas-difusor. La importancia relativa de la aireación y agitación bajo las diferentes configuraciones y condiciones hidrodinámicas fueron llevadas a cabo. Finalmente, propusimos un árbol de decisión para la selección de la mejor configuración con el fin de mejorar la transferencia de oxígeno en bioreactores según el rango de los parámetros operacionales y la viscosidad.

En general, esta tesis es un intento exitoso de analizar y comprender los efectos metabólicos utilizando un diseño racional de experimentos y modelos metabólicos. Además, este análisis sistemático en cooperación con una optimización de la transferencia de oxígeno en bioreactores permitirá mejorar la producción de proteínas recombinantes en *Pichia pastoris.*

Miembros de la Comisión de Tesis Doctoral:

Eduardo Agosin Trumper
José Ricardo Pérez Correa
Luis Larrondo Castro
Pau Ferrer Alegre
Marcel Otavio Cerri
Juan de Dios Ortuzar

Santiago, Enero, 2019

1. INTRODUCTION

1.1 Recombinant protein

Proteins are synthesized as part of their natural metabolism by all living forms. They are essential components and have a role in virtually every cellular process, e.g., some proteins, such as enzymes, serve as biocatalyst, others have structure function, or play a significant role in cell signaling, immune responses, and cell adhesion (Bill, 2014; Demain & Vaishnav, 2009). Their multiple-functionalities are used in biopharmaceutical industries, enzyme industries, and agricultural industries, specially in the fields of medicine, diagnostics, food, nutrition, detergents, textiles, leather, paper, pulp, polymers and plastics (BCC Research, 2018; Demain & Vaishnav, 2009; Jozala *et al*., 2016; Sanchez-Garcia *et al*., 2016).

While some proteins can be isolated from native source, in some cases, protein extraction can be difficult and expensive since stability problems and low productivities (Bill, 2014). Moreover, other proteins are not in the nature, such as antibody fragments (Ma, Drake, & Christou, 2003). Nowadays, most proteins are commercially produced with the aid of genetic and protein engineering using recombinant DNA (rDNA) technology, which was developed in the early 70's (Adrio & Demain, 2010; Demain & Vaishnav, 2011). rDNA technologies allow the production of a wide range of peptides, proteins and metabolites from naturally non-producing cells (Porro *et al*., 2011). Since human insulin was the first heterologous compound produced in a laboratory in 1977, rDNA technology has become one of the most important technologies developed in the 20th century (Lagassé *et al*., 2017; Ma *et al*., 2003; Porro *et al*., 2011; Sanchez-Garcia *et al*., 2016).

Recombinant Protein Production (RPP), employed in the industry to solve source availability problems, is considered a bio-safe and green process, and confers the ability to modify amino acid sequences and therefore protein function, to improve product features to a desired function (Sanchez-Garcia *et al*., 2016). The wide range of applications and the increasing demand of recombinant proteins are leading to an important growth of its market. Actually, RPP is considered a multibillion-dollar business nowadays (BCC Research, 2018; Oliveira & Domingues, 2018; Walsh, 2014), reaching $50 billion market

sales in 2016 only for biopharmaceutical industries (Jozala *et al*., 2016). Moreover, BCC Research estimates that global industrial enzymes market should reach $7.0 billion by 2023 with an annual growth rate of 4.9% for the period 2018-2023 (BCC Research, 2018).

1.2 Cell factories for recombinant protein production

Considering that more than 30 percent of the global sales of the bioindustry corresponds to the manufacture of recombinant proteins, there is a strong pressure to find cost-effective alternatives to participate in new synthesis systems (Corchero *et al*., 2013; Jozala *et al*., 2016; Lagassé *et al*., 2017; Sanchez-Garcia *et al*., 2016). In this context, the selection of the most appropriate Microbial Cell Factory (MCF) as a biological platform for expression, is crucial to be able to compete in an increasingly competitive market (Houdebine, 2009; Lagassé *et al*., 2017; Sanchez-Garcia *et al*., 2016; Vermasvuori, 2009).

In recent years, different microorganisms (such as yeast, filamentous fungi, bacteria, insect cells and mammalian cells) have been employed to produce foreign proteins of interest. The selection of the best MCF depends not only on reaching a high concentration of the product of interest in the shortest time possible, but it is also fundamental that the expressed recombinant protein retains its functionality (Cha *et al*., 2005; Ma *et al*., 2003). For proteins, its biological activity is linked to various aspects, such as its physicochemical properties, levels of expression, post-translational modifications, folding and secretion. This makes the choice of the most appropriate system not trivial (Table 1-1).

Despite the diversity of MCF used for the biosynthesis of pharmaceutical products and recombinant proteins, *E. coli* continues to be the most employed, since it allows fast and economical production (Bill, 2014; Vermasvuori, 2009). Nevertheless, this bacteria is preferred for producing simple proteins (Bill, 2014; Carter, 2011), since it is not capable to naturally achieve post-translational modifications, such as proper glycosylation, adequate folding, and disulfide bridge formation (Ferrer-Miralles *et al*., 2009). In addition, *E. coli* generates inclusion bodies, which significantly increases the purification costs of the protein (Vermasvuori, 2009).

Table 1- 1 Comparison of recombinant expression systems by Cha et al., 2005

	E. coli	**Yeast**	**Insects**	**Mammalian**
Growth rate	++	+	-	-
Expression yield (based on dry weight)	+ (1-5%)	+ (>1%)	++ (30%)	- (<1%)
Productivity	++	+	+	-
Media cost	--	-	+	++
Culture techniques	++	+	-	--
Production cost	--	-	+	++
Protein folding	--	+	++	++
Simple glycosylation	No	Yes	Yes	Yes
Complex glycosylation	No	No	Yes	Yes
Secretion	-	++	++	++
Functionality of expressed eukaryotic protein	-	+	++	++
Availability of genetic systems	++	+	-	-
Pyrogenic problem	Possible	No	No	No

On the other hand, although insect cells or mammalian cells are the best option for the expression of functional complex proteins, the difficulty to cultivate them, their slow growth rate, sophisticated nutritional requirements as well as high sensitivity to mechanical stresses and environmental variations limit their use to high value-added specialties (Bill, 2014; Cha *et al*., 2005; Houdebine, 2009; Porro *et al*., 2011; Vermasvuori, 2009).

Yeasts combine eukaryotic ability to achieve post-translational modifications with the bacterial capacity to grow at high cell densities (Cha *et al*., 2005; Houdebine, 2009; Porro *et al*., 2011). Similarly, yeasts present yields of recombinant proteins that are higher than mammalian cells in several cases (Corchero *et al*., 2013; Idiris *et al*, 2010; Porro *et al*., 2011) and are easier to scale at industrial level. *Saccharomyces cerevisiae* is the yeast species most widely used to express recombinant proteins. Nonetheless, there are other species, less conventional, that have certain advantages for the production of extracellular proteins, such as the methylotrophic yeast *Pichia pastoris* (Cereghino *et al*., 2002; Corchero *et al*., 2013).

1.3 *Pichia pastoris*, an emerging microbial cell factory

Within non-conventional MCFs, the methylotrophic yeast *Pichia pastoris* is currently one of the most effective and versatile MCF for RPP (Peña *et al*., 2018; Porro *et al*., 2011; Prielhofer *et al*., 2018; Theron *et al*., 2018). *P. pastoris* was developed as a host system for the first time in 1985 and since then, researchers have been increasingly recognizing its capacity and advantages for RPP (Bill, 2014; Peña *et al*., 2018; Porro *et al*., 2011). Actually, *P. pastoris* applications have steadily increased from 1995 to 2013, in contrast to all other host cells. Moreover, its increase was coupled with the decline in the usage of *E. coli* (Table 1- *2*). Indeed, *P. pastoris* can be easily manipulated at the genetic level (Balamurugan *et al*., 2007; Cereghino *et al*., 2002) and expresses foreign proteins, intracellular and extracellular, at high levels (Balamurugan *et al*., 2007; Cereghino *et al*., 2002; Jordà *et al*., 2012). Unlike *E. coli*, *P. pastoris* naturally performs post-translational modifications, such as glycosylations, disulfide bonds formation and proteolytic processing (Charoenrat *et al*., 2005; Cregg *et al*., 2000; Lopes *et al*., 2013), which are essential for most eukaryotic protein functionality (Ciofalo *et al*., 2006; Corchero *et al*., 2013; Masuda *et al*., 2010). Moreover, it possesses secretion pathways more similar to higher eukaryotic cells (Corchero *et al*., 2013; Delic *et al*., 2012); and secretes a lower number of native proteins, which makes downstream processing easier (Delic *et al*., 2013; Mattanovich *et al*., 2009). In contrast to *S. cerevisiae*, *P. pastoris* exhibits a Crabtree-negative phenotype, showing a reduced synthesis of undesirable products, like ethanol, in glucose-limited conditions (Çalık *et al*., 2015; Mattanovich *et al*., 2009). Finally, *P. pastoris* can be efficiently cultivated up to high cell densities, using fed-batch cultivation strategies (Daly & Hearn, 2005), achieving high titers and productivities.

Table 1- 2 Recombinant gene expression in the most commonly used cells until 2013. Taken from (Bill, 2014)

Year	All host cells	*E.coli*	*S. cervisiae*	*P. pastoris*	Insects cells	Mammalian cell-line
1980	0	0	0	0	0	0
1985	0	0	0	0	0	0
1990	12	75%(9;4E)	8%(1)	0	17%(2)	0
1995	37	70%(26;17E)	5%(2)	5%(2)	5%(2)	8%(3)
2000	50	70%(35;17E)	0	4%(2)	12%(6)	12%(6)
2005	121	85%(103;53E)	0	5%(6)	6%(7)	2%(2)
2010	172	76%(131;67E)	0	9%(15)	5%(6)	5%(9)
2013	128	73%(94;54E)	2%(2)	11%(14)	4%(5)	4%(5)

As a consequence, *P. pastoris* has been widely used for the expression of recombinant proteins, reaching titers of several grams per liter in several cases (Cereghino & Cregg, 2000; Čiplys *et al*., 2015; Hasslacher *et al*., 1997; Heyland *et al*., 2010; Yan Wang *et al*., 2001). Most remarkably, and as proof of its technical feasibility and adequacy, two recombinant proteins produced in this cell factory have already been approved by the Food and Drug Administration of the United States, FDA, for medical purposes (Ciofalo *et al*., 2006; Thompson, 2010).

As a methylotrophic yeast, *P. pastoris* is able to grow on methanol as the sole carbon and energy source. Its methanol metabolism was utilized to develop an efficient protein production system, using the extraordinarily high and methanol-inducible alcohol oxidase promoter (PAOX1) (Zahrl *et al*., 2017). AOX1 promoter is one of the most common strategies of the heterologous expression in this yeast since it has proven valuable for a growing list of therapeutic proteins and industrial enzymes (Weinacker *et al*. 2013; Spadiut *et al*. 2014). Nevertheless, the oxidation of methanol adds technological challenges for industrial scale production, such as a high oxygen consumption, high heat production and the additional demand for safety precautions when working with this flammable substrate (Mattanovich *et al*., 2014). To circumvent these disadvantages, variants of AOX1 promoter or other strong promoters which allow either constitutive or inducible gene expression are highly recommended (Çalık *et al*., 2015; Porro *et al*., 2005). Production processes utilizing *P. pastoris* favorably apply carbon source-dependent promoters, such as glyceraldehyde-3-phosphate dehydrogenase promoter (pGAP) or, more recently, the high-

affinity glucose transporter gene GTH1 (pGTH1) (Çalık et al., 2015; Cereghino & Cregg, 2000; Prielhofer et al., 2013). pGAP offers an alternative to pAOX1 when the toxicity of methanol is of concern, like in the food industry (Yang & Zhang, 2018). It is featured by a high-level constitutive expression and is more suitable for continuous cultivation due to simpler process control (Garcia-Ortega, *et al*., 2017; Looser *et al*., 2015; Mattanovich *et al*., 2014). Actually, the GAP promoter has been perceived as an efficient alternative production strategy with similar performance comparing with pAOX1 (Garcia-Ortega, *et al*., 2017; Looser *et al*., 2015; Mattanovich *et al*., 2014). Moreover, several alternative fermentation strategies have been extensively studied for this expression system (Garcia-Ortega *et al*, 2015; Heyland *et al*, 2011; Looser *et al*., 2015; Peña *et al*., 2018; Yang & Zhang, 2018).

Besides strong promoters, the high efficiency of protein production by *P. pastoris* is attributed to an efficient secretory pathway (Zahrl *et al*., 2017). Evidence shows that the secretion pathway of *P. pastoris* is more similar to higher eukaryotes than that of *S. cerevisiae* (Papanikou & Glick, 2009). Organelle structure and proliferation are different (Figure 1-1), and the regulation pattern of Unfolded Protein Response – UPR - shows significant differences between these two yeast species, favoring *P. pastoris* in response to one of the main limitations in RPP (Graf *et al*., 2008).

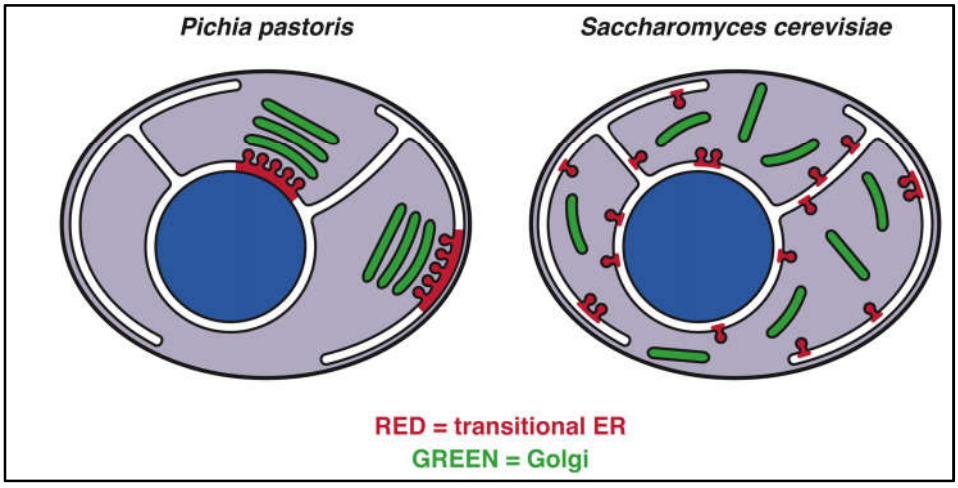


Figure 1-1 Diagram of transitional ER and Golgi organization in two budding yeasts; *Pichia pastoris* and *Saccharomyces cerevisiae* (taken from Papanikou & Glick, 2009).

Based on past achievements and novel developments, systems biotechnology of *P. pastoris* has seen significant progress over the last two decades (Figure *1-2*) (Theron *et al*., 2018; Yang & Zhang, 2018; Zahrl *et al*., 2017). Furthermore, driven by the advances in *omics* technologies, the understanding of *P. pastoris* as a host for protein production has significantly improved, although further systems biology studies are necessary (Theron *et al*., 2018; Yang & Zhang, 2018; Zahrl *et al*., 2017).

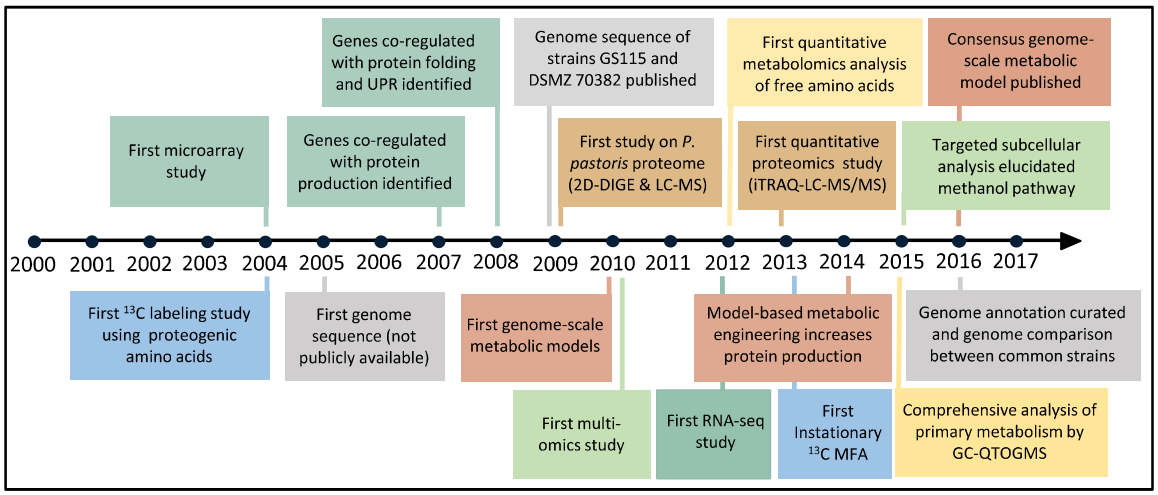


Figure 1-2 Hallmarks of systems biology development in *Pichia pastoris*. Genomics (gray), transcriptomics (dark green), proteomics (brown), metabolomics (yellow), fluxomics (blue), genome-scale metabolic model and model-based engineering (red), and combined omics studies (light green) (taken from Zahrl *et al*., 2017).

1.4 Optimization of recombinant protein production in *P. pastoris*.

Despite its growing acceptance and successful applications, recombinant protein production in *P. pastoris* can be undermined by several cellular processes, where protein folding and secretion are the most recurrent bottlenecks (Delic *et al*., 2013; Delic *et al*., 2014; Gasser *et al*., 2013). In fact, different genetic and physiological factors determine the productivity of recombinant systems. The most well-known limitations are codon usage of the recombinant protein (Wang *et al*., 2015), promoter selection (Prielhofer *et al*., 2013), carbon and oxygen availability in the culture (Baumann *et al*., 2008; Heyland *et al*., 2011) and fed-batch operational parameters (Maurer *et al*., 2006), seriously hampering protein yield, productivity and the economic feasibility of the process. In response to these,

strategies are divided into strain engineering and process engineering wise (summarized in Figure 1-3). In fact, several strategies have been proposed in recent years to optimize the production of recombinant proteins in *P. pastoris* (Arruda *et al*., 2016; Aw & Polizzi, 2016; Çalık *et al*., 2015; Garcia-Ortega *et al*., 2015; Looser *et al*., 2015; Nocon *et al*., 2016; Puxbaum *et al*, 2015; Tomàs-Gamisans *et al*, 2016; Vogl *et al*., 2016).

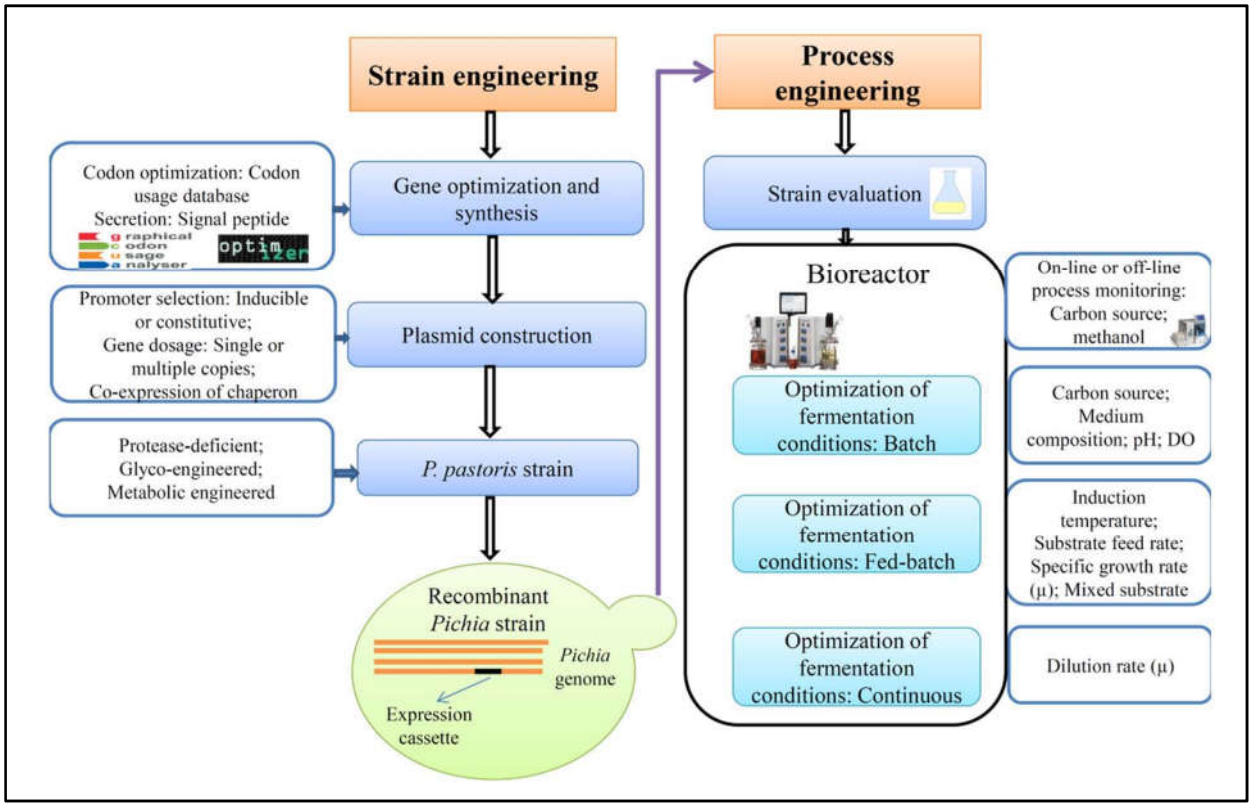


Figure 1-3 General strategies for enhanced production of recombinant proteins in *P. pastoris*. (taken from Yang & Zhang, 2018).

Within process engineering optimization, three cultivation methods can be considered for the production of heterologous proteins in yeasts, *i.e.* batch, fed-batch and continuous cultivation. Nevertheless, fed-batch and continuous processes are most commonly used for manufacturing and optimizing a targeted recombinant protein production (Hensing *et al*, 1995).

During fed-batch cultivations, nutrients are continuously fed to a culture, resulting in a continuous volume increase, with biomass and products being retained in the reactor. This cultivation ends by volume, oxygen or heat transfer limitation (Hensing *et al*., 1995;

Potvin *et al*., 2012; Saitua *et al*., 2017) (Figure *1-4* a). On the other hand, continuous cultivation is an open thermodynamic system that requires feeding and extraction pumps to be permanently activated, maintaining constant volume (Figure *1-4* b). When steady state is reached, the specific growth rate of the cells equal the dilution rate (Garcia-Ortega, 2016). During both cultivations, the growth rate can be manipulated by controlled addition of feed, allowing to minimize byproduct formation, oxygen uptake and heat generation, which can be adapted to the limitations set by the reactor specifications (Hensing *et al*., 1995).

Industrially, *P. pastoris* is commonly grown in fed-batch cultures in order to maximize the titer and volumetric productivity of a desired compound, often a recombinant protein (Looser *et al*., 2015; Riesenberg & Guthke, 1999). Nevertheless, it is essential to previously characterize and optimize operational parameters in continuous culture, to set conditions which maximizes the synthesis of the target product and to limit the formation of inhibitory compounds (Villadsen, Nielsen, & Lidén, 2011). In fact, operational parameters can introduce a significant variability that influence the production and secretion rate of recombinant proteins (Gasser *et al*., 2007, 2008; Love *et al*., 2012). It is well known that temperature, pH, osmolality, growth rate and dissolved oxygen can affect not only the metabolic pathways, but also the specific productivity of a target recombinant protein (Baumann *et al*., 2008; Charoenrat *et al*., 2005; Resina *et al*., 2007). In the literature, several relationships between production and growth rate have been reported (reviewed by Looser *et al*., 2015), as well as how they affect the biomass titer in the culture (Adelantado *et al*., 2017; Baumann *et al*., 2008; Garcia-Ortega *et al*., 2017). Therefore, it is important to generate a systematic analysis of these two factors.

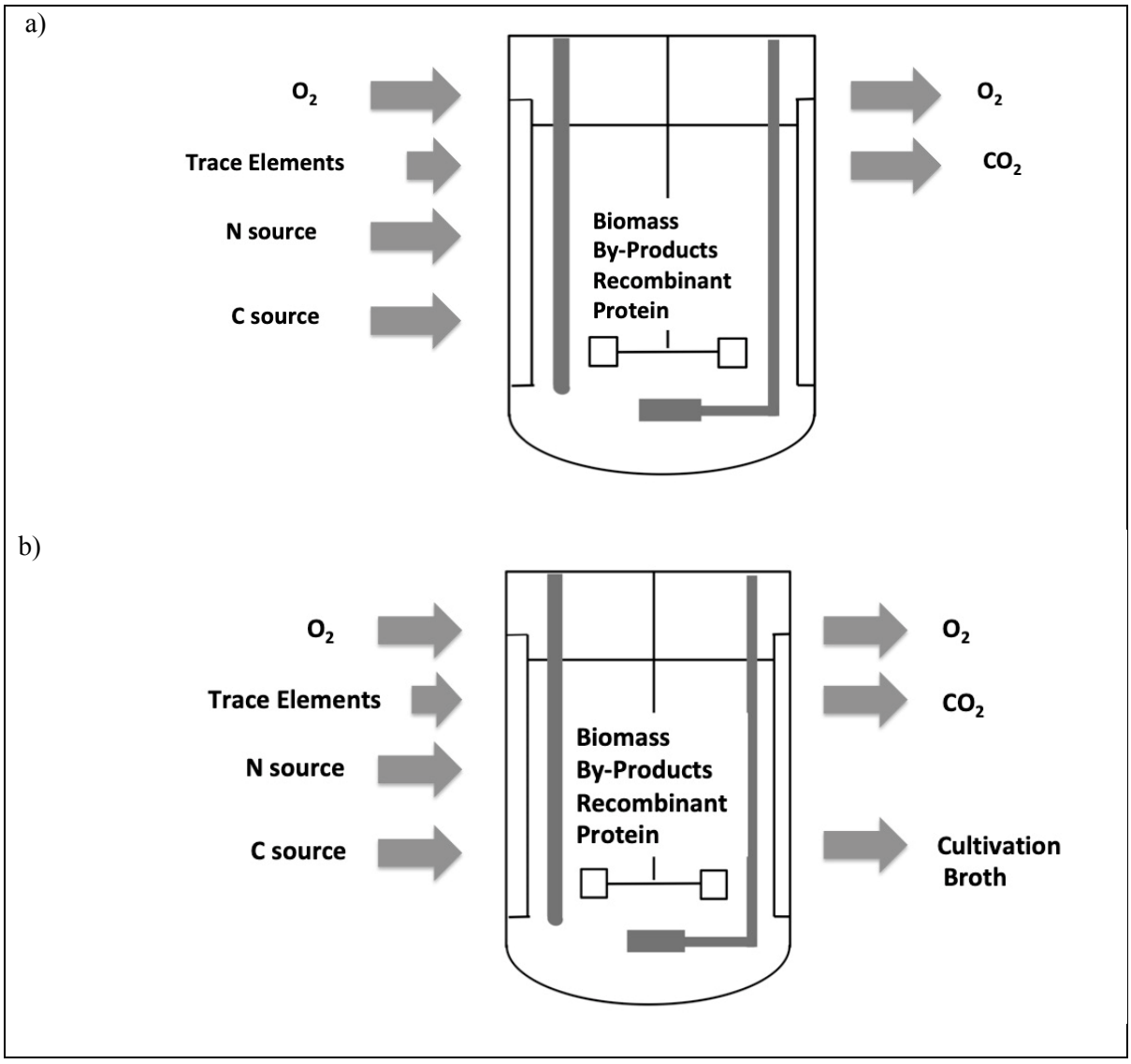


Figure 1-4 General cultivation scheme of a) fed-batch culture and b) continuous culture.

1.5 Genome Scale Metabolic Model

During cultivation, the cells adapt constantly to the changing extracellular environment and to the limited mass transfer conditions observed at high densities (Landi *et al*, 2015; Vargas *et al*., 2011). Therefore, it is critical to understand how the cell metabolism copes with nutritional and environmental stresses exerted by process conditions to improve bioreactor performance (Graf *et al*., 2009). This is a complex task, however, since the characterization and fine-tuning of strain's features and process variables often require significant amounts of time and money (Çalık *et al*., 2015). Therefore, it is desirable to

have several platforms to integrate different levels of information from continuous and dynamic cultivations of *P. pastoris* that can be used to elaborate rational hypotheses to increase process productivity.

Systems biology offers a quantitative and comprehensive approach to address this task (Kitano, 2002). In particular, Metabolic Flux Analysis (MFA) is an advantageous tool to determine the pathways of the global cellular metabolic process from a reductionist approach (Stephanopoulos, 1998). Metabolic pathway fluxes are obtained from the rates of substratum consumption and product formation, which is very useful when making comparisons under different experimental conditions (Varela *et al*, 2003). One type of MFA based on genomic scale, *i.e*. where a high number of reactions and cellular metabolites are used, allows to know the metabolic capacities of a cell, analyzing the connectivity of the metabolites or predicting genotypic-phenotypic relationships (Patil *et al*., 2005). Genomic-Scale Metabolic Models (GSMs) have been used in *P. pastoris* to optimize yeast with *in silico* modifications (Chung *et al*., 2010), to identify essential reactions of the yeast cells and to modulate the production of heterologous proteins, including the flow variation of amino acids (Caspeta *et al*, 2012; Sohn *et al*., 2010) and the enhanced recombinant protein production (Nocon *et al*., 2014), validating these frameworks as strain engineering tools for this particular yeast. In fact, in the last years GSMs have become one of the most useful and widely employed tools in systems biology to understand cellular behavior under different environmental conditions, to map over *omics* data, and to define a metabolic engineering targets (Sohn *et al*. 2012; Nocon *et al*., 2014; Saitua *et al*., 2017).

To the best of our knowledge, six GSMs have been reported so far *for P. pastoris* (Caspeta *et al*., 2012; Irani *et al*., 2016; Sohn *et al*., 2010; Tomàs-Gamisans *et al*., 2017; Ye *et al*., 2017), which have been developed to drive strain optimization process with a special emphasis on recombinant protein production. Nevertheless, the validation of these GSMs is usually performed using extracellular flux predictions in steady-state, without analysis of the intracellular flux distributions (Sánchez & Nielsen, 2015). Intracellular flux distributions are essential to understand cell metabolism and, therefore, it is highly recommended to manually curate (Pereira *et al* , 2016; Saitua *et al*., 2017) and to include thermodynamic restrictions (Saa & Nielsen, 2016) to the resulting GSM.

Once a robust GSM is developed, it allows to gain knowledge of the yeast's steady-state metabolism and to make accurate predictions for the overexpression of high-value metabolites in mutant strains (Oberhardt *et al*., 2009) though, the kinetics and physiology of yeast cells can be better understood in a dynamic setting (Sánchez *et al*, 2014). In particular, dynamic Flux Balance Analysis at Genome-Scale (GS-dFBA) (Höffner *et al*, 2013; Mahadevan *et al*, 2002; Varma & Palsson, 1994) is a modeling framework that allows to simulate the metabolism during non-stationary (batch or fed-batch) cultivation, using GSM coupled with dynamic mass balances from the extracellular environment of the bioreactor. We recently constructed a robust, dynamic GSM of glucose-limited aerobic cultivations of *P. pastoris,* which, if properly trained, can be used to predict bioreactor dynamics (Saitua *et al* 2017). Moreover, the model could also be employed to obtain realistic flux distributions through-out dynamic cultivations and to determine metabolic and process engineering strategies to improve the production.

Finally, GSM and GS-dFBA allow understanding and quantifying the metabolic burden caused by this production in steady state as well as dynamic culture systems. Moreover, they allow to simulate different environmental conditions, to map over *omics* data, and to define metabolic engineering targets in different carbon sources and feeding strategies. Also, they could be used to study perturbations such as oxygen limitation, which is a critical problem in industrial *P. pastoris* cultivations (Porro *et al*., 2005).

1.5 Optimization of Oxygen Transfer Rate in bioreactors.

Fed-batch cultivation is commonly used for microbial growth in order to reach high-cell density cultures. The main limitations of this system are volume, oxygen and heat transfer. In the case of *P. pastoris*, oxygen limitation is a critical operation variable in industrial cultivations (Porro *et al*., 2005), independent of the carbon source used (Looser *et al*., 2015).

Oxygen is an essential substrate in aerobic bioprocesses, since it allows the oxidation of the carbon source, increasing the energy efficiency of the microorganism (Felix Garcia-Ochoa & Gomez, 2009; Sonnleitner, 2016; Yong Wang *et al*., 2009). This substrate is commonly supplied from the gas phase, where it is transferred to the liquid

and, once dissolved, it is consumed by the cells. Nevertheless, in high density cultures, oxygen solubility is low and oxygen demand increase over time since biomass does (Garcia-Ochoa *et al*, 2010). Once Oxygen Uptake Rate (OUR) is higher than the Oxygen Transfer Rate (OTR), dissolved oxygen decrease fast and begin to be limiting. Oxygen limitation in microbial cultures promotes the formation of inhibitory by-products, which prevent reaching the maximum permissible biomass (Adelantado *et al*., 2017; Baumann *et al*., 2010; Enfors, 2011; Garcia-Ortega *et al*., 2017; Villadsen & Patil, 2007). As a consequence, the bioreactor type and its mechanical design, together with an adequate control of dissolved oxygen in cell cultures is crucial to guarantee a maximum OTR, to achieve high cell densities and volumetric productivities.

OTR is typically described as the product of a volumetric mass transfer coefficient, $k_L a$, and the average concentration gradient based on the two-film theory of Whitman (Whitman, 1924) (Equation 1-1).

$$\frac{dC}{dt} = k_L a \cdot (C^* - C) \qquad (1-1)$$

Where, C* is the oxygen equilibrium concentration and C is the dissolved oxygen concentration in the liquid phase of the bioreactor.

OTR mainly depends on the $k_L a$ coefficient in a specific saturation condition (Equation 1-1) and for this reason, $k_L a$, is a crucial factor to characterize the oxygen transfer capabilities of gas-liquid bioreactors (Garcia-Ochoa & Gomez, 2009; Labík *et al*., 2017; Suresh *et al*, 2009). Actually, its measurement and/or prediction is used to design, operate and scale stirred tank bioreactors (Linek *et al*, 1993; Suresh *et al*., 2009).

$k_L a$ is strongly influenced by the hydrodynamic conditions in the bioreactors, which are affected by operational conditions – such as agitation and airflow rate –, geometrical configurations – such as type of impeller or sparger– and physical properties. – such viscosity (Garcia-Ochoa & Gomez, 2009; Merchuk *et al*., 1990). Moreover, for Newtonian fluids in a conventional bioreactor, where the stirring has a strong influence on oxygen transfer, $k_L a$ can be correlated to the impeller speed (N), specific gas flow rate (∅), and dynamic viscosity (μ) (Buffo *et al*., 2016), as follows:

$$k_L a \ = \ b \cdot N^c \cdot \emptyset^d \cdot \mu^e \qquad (1-2)$$

Where b represents all other parameters not incorporated in the equation – such as geometric parameters.

Different strategies to increase OTR have been used (Felix Garcia-Ochoa & Gomez, 2009; Maclennan & Pirt, 1966; Siegell & Gaden, 1962). Some researchers have modified the bioreactor configuration; impellers, sparger and geometrical dimension in order to increase $k_L a$ (Arjunwadkar *et al*., 1998; Buffo *et al*., 2016; Vitae & Vitae, 2011; Yang *et al*., 2012); others have focused on increasing saturation oxygen concentrations (Equation 1-1), increasing partial pressure or enriching gaseous phase with pure oxygen (Flickinger & Perlman, 1977; Nielsen, Villadsen, & Lidén, 2002; Yamada *et al*, 1978) or pressuring the bioreactor (Onken & Liefke, 1989; Yang & Wang, 1992). Nevertheless, when strategies using maximum capable saturation oxygen concentration still limit biomass formation, it is necessary to focus on $k_L a$ optimization. Buffo *et al*., (2016) demonstrated that the combination of two types of impellers; Rushton and Ear-Elephant, provided considerable differences in oxygen transfer (Figure *1-5*). Moreover, some configurations were less harmful to microbial cells, have more efficiency and less shear rate, demonstrating the potential of this research. Otherwise, simulation from Gelves (2013), demonstrated rotating-sparger could increase up to 6-fold the oxygen transfer in large-scale bioreactors and improved air homogenization in the bioreactor. Nevertheless, both techniques have not been simultaneously evaluated, so far.

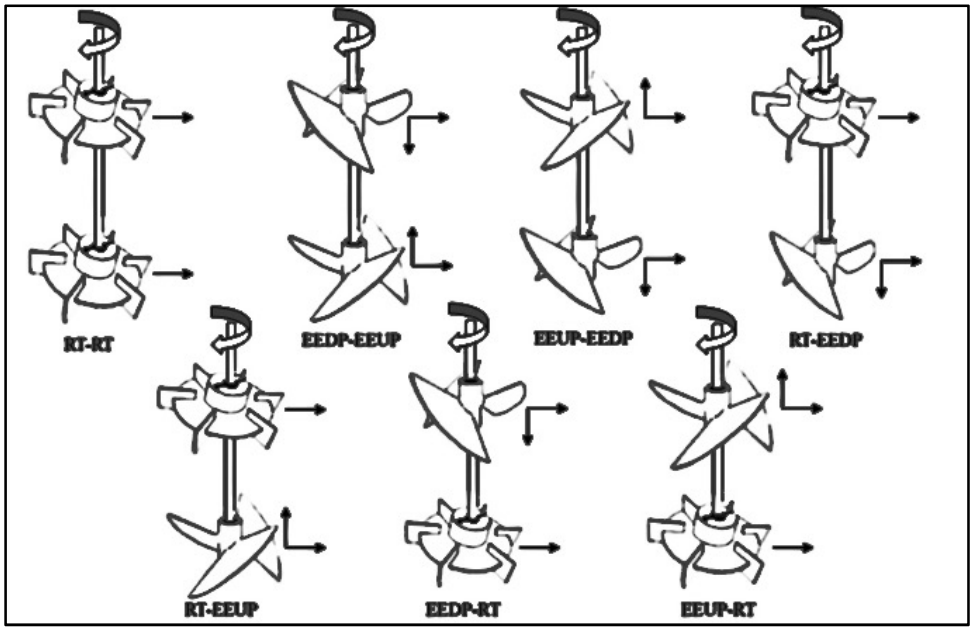


Figure 1-5 Dual-impeller configuration (Figure taken from Buffo *et al*., 2016)

1.6 Case of study: Thaumatin a sweetness protein

As a case study, we analyzed the metabolic behavior of a recombinant strain producing the sweet-tasting low-calorie protein thaumatin (van der Wel & Loeve, 1972). Thaumatin is a sweet protein (22 kDa), rich in cysteine (has 8 disulfide bounds), which is collected from the fruit of African bush *Thaumatococcus danielli* (van der Wel & Loeve, 1972). Thaumatin, as well as other sweet proteins, have became relevant as a sweet complement to improve the organoleptic and functional properties of natural sweeteners (Behrens, Meyerhof, Hellfritsch, & Hofmann, 2011). In total, five sweet proteins have been described: thaumatin, monellin, mabinlin, brazzein and miraculin (Wintjens *et al*, 2011) and their main properties are summarized in Table 1-3.

Table 1- 3 Properties of different sweet proteins

	Thaumatin	**Monellin**	**Mabinlin**	**Brazzein**	**Miraculin**
Origin	*Thaumatococcus danielli* (Africa)	*Pioscorephyllum cumminisii* (Africa)	*Capparis masaikai* (China)	*Pentadiplandra brazzeana* (Africa)	*Richadella dilcifica* (Africa)
Sweetness factor (weight)	1.600	3000	400	500	--
Types	I,II,a,b,c	2 chains	I,II,II,IV	--	--
M. W. [kDa]	22	10,7	12,4	6,4	50
Amino acids	207	45/50	33/72	54	191
Disulphide bounds	8	--	4	4	4
P. I.	12	9,3	11,3	5,4	7,7
Threshold [nM]	48	270	11,3	5,4	--
Stability	Resistant	Dissociation	Resistant	Resistant	Resistant

*Data taken from the following papers: (Assadi-Porter *et al*., 2010; Caldwell *et al*., 1998; de Vos *et al*., 1985; Guan *et al.*, 2000; Igeta, *et al.*, 1991; Ito *et al*., 2007; Kaneko & Kitabatake, 1999; Kaneko & Kitabatake, 2001; Kim *et al*., 1989; Ming & Hellekant, 1994; Monellin *et al.*, 1973; NIRASAWA *et al*., 1994; Theerasilp & Kurihara, 1988; van der Wel, *et al.*, 1984; Henrik van der Wel & Loeve, 1972; Wintjens *et al*., 2011)

During the last years, non-caloric sweeteners, artificial and natural, have shown a sustained increase in the market (IANSA, 2018; ISO, 2012; ODEPA, 2014), in order to reduce sugar consumption (FAO, 2018; ODEPA, 2014). Nevertheless, global sugar consumption per capita is still elevated (OMS, 2018). That is why, a countless of governmental and international organizations have alerted about an excessive consumption of high-sugar contented food, which have a high caloric content and down – or null- nutritional contribution (MINSAL, 2010, 2017b, WHO, 2013, 2015). Massive consumption of these foods has caused a considerable increase in the incidence of Non-Communicable Diseases (NCDs), like obesity, diabetes and cardiovascular diseases among others (MINSAL, 2017a; Morenga *et al*., 2014). NCDs are the main causes of premature death worldwide,

adjudicating a 68 percent of death in 2012 (WHO, 2015). Chile is within the countries with more overweight, with an incidence of 74 percent, as registered in 2016-2017 (MINSAL, 2017).

In response to this kind of illnesses, World Health Organization (WHO) strongly recommends the reduction of free-sugar consumption to less than 10 per cent of total daily calorie intake, *i.e.* 25 grams per day for an average-weight person (WHO, 2015). Nonetheless, current statistics indicate that the world average consumption is 56 grams per day in 2013 (FAO, 2018). The same year, Chile reached a consumption per capita of 123 grams per day, one of the highest in the world (FAO, 2018).

The most consumed artificial sweeteners are saccharine, aspartame, cyclamate and acesulfame K (ISO, 2012; Kant, 2005). However, consumers are prioritizing natural alternatives, since multiple adverse effects of artificial sweeteners have been reported, like mental disorders, heart failure and brain tumors (Behrens *et al*., 2011; Kant, 2005; Temussi, 2006). Moreover, Suez *et al.* (2014) demonstrated than saccharine consumption could cause insulin resistance due to an alteration of intestinal microbiota (Suez *et al*., 2014).

Among the most well known natural sweeteners are the glycosylated steviols, produced by the plant Stevia rebaudiana (IANSA, 2014); and mogrosides from the Lu Han Guo fruit (Dubois & Prakash, 2012). Unfortunately, both sweeteners show some disadvantages: in the case of Stevia, sweetness often brings with a sensation of bitterness and licorice; and in the case of mogrosides, they have a low solubility in water (Dubois & Prakash, 2012). Therefore, to improve the organoleptic and functional properties of natural sweeteners, the use of sweet proteins have become relevant as a complement (Behrens *et al*., 2011). Within sweet proteins, thaumatin is a perfect candidate to be employed together with other sweeteners, projected in a growing market. Since this protein not only has low calorie content, but is also functional and stable over a wide temperature range (until 80°C) and pH (2-10) (Kaneko & Kitabatake, 2001). Moreover, it has a triple functionality:

- A high sweetness potency, 3,000 fold that of sucrose in molar basis (Stoger, 2012; van der Wel & Loeve, 1972);
- A masking capacity for taste sensations, such as the bitterness of *Stevia* (Auger, 2013)*;*

- A flavor-enhancing capacity (Auger, 2013; I Faus, 2000), such as salty flavor.

Consequently, its high functional range makes thaumatin unique and its stability to extreme conditions allows its wide use in the food industry, including heat-processed foods.

Thaumatin is naturally found at very low concentrations - *circa* of 400 μg/g - in the fruits of *Thaumatococcus danielli* (van der Wel & Loeve, 1972), which hampers its industrial commercialization. In fact, in 2013 the extraction of one gram of thaumatin requires the use of at least one ton of fruits, resulting in a high extraction cost and a sales price of US$ 6,000 /kg of thaumatin (Auger, 2013). This high sale price opens opportunities to the heterologous synthesis of thaumatin with the aim of increasing its concentration and productivity, as well as ensuring a sustainable, environmentally friendly and economically feasible biosynthesis. The recombinant production of this sweet protein has been widely studied in different organisms, although high productions have not been reached so far (Table 1- *4*). For example, functional thaumatin has been produced in the yeast *Pichia pastoris*, with final titers of 100 mg/L after 160 hour cultivation (Masuda *et al*, 2010); and in *Aspergillus awamori*, reaching 150 mg/L in 120 hour cultivation (Moralejo *et al*., 2000). The main reason of its low productivity is associated with its eight disulfide bounds in most of the studies (Daniell *et al*., 2000; Edens *et al*., 1984; Faus *et al*., 1996; Illingworth *et al*., 1989; Masuda *et al*., 2016; Masuda *et al*., 2010; Moralejo *et al*., 2000; Moralejo *et al*., 2001). Moreover, efficient production has not yet been achieved, given that it is estimated that Y_{SP} = 74 mg of thaumatin per gram of glucose could be achieved, which would allow reaching up to 18 g/L in 80 h of culture (Supplementary material S1-1). However, a systematic analysis is necessary to better understand the effect in the microorganism of the recombinant production in this protein.

.

Table 1- 4 Summary of recombinant production of thaumatin in different organisms.

Organism	**Promoter**	**Titer**	**Sweetness**	**Reference**
Microorganism				
E. coli	*lac / tac*	Very low	No	(Edens *et al*., 1982; Faus *et al*., 1996)
	lac	40 mg/L	Yes with refolding process	(Daniell *et al*., 2000)
B. subtilis	α-amilase	1 mg/L	n. d.	(Iilingworth *et al.*, 1988)
S. lividans	β-galactosidase	0.2 mg/L	n.d.	(Illingworth *et al.*, 1989)
S. cerevisiae	GAPDH/PGK	low	No	(Edens *et al*., 1984; Lee *et al*., 1988)
	PGK	140 mg/L	Yes with refolding process	(Weickmann, 1994)
K. lactis	GAPDH	low	No	(Edens & van der Wel, 1985)
A. awamori	gdhA	150 mg/L	Yes	(Moralejo *et al*., 2000; Moralejo *et al*,1999; Moralejo *et al*., 2001)
P. pastoris	AOX1	100 mg/L	Yes	(Masuda *et al*., 2004; Masuda *et al*,2010)
Plants				
potato	CaMV35S	0.2-2 mg/Kg	Yes	(Witty, 1990)

pear	CaMV35S	Very low	n.d.	(Lebedev *et al.*, 2002)
tomato	CaMV35S	Very low	Yes	(Bartoszewski *et al.*, 2003; Firsov *et al.*, 2012)
cucumber	CaMV35S	Very low	Yes	(Szwacka *et al.*, 2002; Zawirska-Wojtasiak *et al.*, 2009)
raspberry	CaMV35S	Very low	n.d.	(Schestibratov & Dolgov, 2005)
apple	CaMV 35S	Very low	Yes	(Dolgov *et al.*, 2004)
barley	Hordein-D	2 mg/g	Yes	(Stahl *et al.*, 2009)

1.7 Hypothesis and Objectives

The hypothesis of this thesis is that the systematic characterization of *Pichia pastoris* in conjunction with the development of computational tools will allow to find genetic targets, optimum operation parameters, and optimum mechanic bioreactor configurations in order to enhance the production of recombinant proteins in the yeast *Pichia pastoris.*

General Objective

The general objective of this thesis is to build, characterize, and analyze recombinant *P. pastoris* strains in different environmental conditions using improved computational tools. These exhaustive characterizations together with the optimization of oxygen transfer rate could be used to improve recombinant protein production in *P. pastoris.*

The specific objectives are:

1. To construct recombinant *P. pastoris* strains producing thaumatin.
2. To characterize and analyze thaumatin producing *P. pastoris* in different environmental conditions.
3. To develop computational tools to analyze *P. pastoris* strain in different fermentation processes.
4. To optimize oxygen transfer rate in 1-L bioreactor to improve high density culture.

1.8 Approach of this thesis

In this work, the central theme is to develop computational tools and to analyze different type of strategies to enhance the production of recombinant protein production in the yeast *Pichia pastoris*

- Chapter 1 discusses the importance of developing a systematic analysis of two main operational variables: specific growth rate and dissolved oxygen. Moreover, we formulated a Flux-Distribution-Improved GSM of *Pichia pastoris* to understand the individual and synergistic effects of two main operational parameters.
- Chapter 2 deals with the formulation of a dynamic Genome Scale Metabolic Model of *Pichia pastoris* that yields realistic metabolic flux distributions throughout dynamic cultivations. The model was calibrated with experimental data and used to propose genetic and process engineering strategies to improve the performance of a *P. pastoris* strain of interest.
- Chapter 3 discusses the formulation of an automatic algorithm, to reliably estimate k_La under different hydrodynamic conditions. The latter allows to design high throughput k_La assessment systems to optimize oxygen delivery in bioreactors. The algorithm proved to be robust and highly efficient.
- Chapter 4 presents a road map to optimize oxygen transfer rate in 1L bioreactor, using different impeller-sparger configurations. First, we characterized all configurations in different hydrodynamics conditions to analysis level curves and parameter sensitivities. Then, we proposed a general decision tree for the selection of the best configuration in order to improve oxygen transfer in bioreactors according operational parameters' domain.

2. CHAPTER I: SYSTEMIC METABOLIC EFFECTS OF SPECIFIC GROWTH RATE AND OXYGENATION LEVEL IN GLUCOSE-LIMITED CONTINUOUS CULTURE OF RECOMBINANT *PICHIA PASTORIS*.

Paulina Torres[1], Pedro Saa[1], Joan Albiol[2], Pau Ferrer[2,3] and Eduardo Agosin[1]

[1]Department of Chemical and Bioprocess Engineering, School of Engineering, Pontificia Universidad Católica de Chile, Vicuña Mackenna 4860, Santiago, Chile

[2]Department of Chemical Engineering, Universitat Autònoma de Barcelona, 08193 Bellaterra (Cerdanyola del Vallès), Barcelona, Spain

[3]Environmental Research and Innovation, Luxembourg Institute of Science and Technology, 41 rue du Brill, L-4422 Belvaux, Luxembourg

Paper in revision in Microbial Cell Factories.

2.1 INTRODUCTION

The methylotrophic yeast *Pichia pastoris* (syn. *Komagataella phaffii*) is a popular cell factory for heterologous protein production due to its convenient physiology and the availability of genetic tools for its manipulation (Gasser and Mattanovich 2018; Peña et al. 2018; Theron et al. 2018). Amongst its features, *P. pastoris* displays a Crabtree-negative growth phenotype leading to reduced synthesis of by-products under aerobic conditions (Çalık et al. 2015). In addition, the availability of strong promoters for heterologous expression added to its natural capacity to perform post-translational modifications has made this yeast the platform of choice for producing a range of recombinant proteins up to grams per liter titers (Ciplys et al. 2015; Hasslacher et al. 1997; Heyland et al. 2010; Wang et al. 2001). Despite these achievements, efficient heterologous protein production in *P. pastoris* remains challenging as poorly-tuned protein overexpression can affect relevant cellular processes, such as protein folding and secretion (Delic et al. 2014; Gasser et al. 2007; Love et al. 2012). Moreover, codon usage level (Hu et al. 2013; Xiang et al. 2016), promoter selection (Prielhofer et al. 2013) as well as culture medium composition (Heyland et al. 2011) and operational conditions (Maurer et al. 2006) may also play major

roles on process performance. In particular, the operational conditions have gained increasing attention as they are known to introduce substantial variability in the process, significantly affecting the recombinant protein secretion (Looser et al. 2015).

High recombinant protein expression in *P. pastoris* relies on the use of strong promoters, like pAOX1 (promoter from alcohol oxidase I) and pGAP (promoter from glyceraldehyde-3-phosphate dehydrogenase). While pAOX1 offers strong inducible expression with methanol – thereby enabling uncoupling growth from production –, pGAP provides comparable constitutive expression (Peña et al. 2018). *P. pastoris* cultures incur in high oxygen consumption and heat production during methanol oxidation, and hence, its use is ill-advised for large-scale protein production (Mattanovich et al. 2014). Once a suitable expression system has been chosen, the next step is to optimize culture conditions to achieve the target productivity. Factors such as temperature, pH, osmolality, specific growth rate (μ) and dissolved oxygen (DO) are critical for the efficient operation of the culture, and they have been individually assessed on their influence on protein production and culture performance (Baumann et al. 2008; Charoenrat et al. 2005; Dragosits et al. 2009; Dragosits et al. 2010; Garcia-Ortega et al. 2017; Heyland et al. 2010; Maurer et al. 2006). Although there have been studies reviewing the relationships between protein production and growth (refer to Looser et al. (2015) for a comprehensive review), and how DO impacts the yeast's physiology (Adelantado et al. 2017; Baumann et al. 2008; Garcia-Ortega et al. 2017), current studies fail to evaluate both the *individual* and *combined* effects of these operational parameters on the metabolic performance of *P. pastoris*. Notably, such evaluation is not trivial as both, appropriate experimental design and modelling approach, are required to interpret the experimental results.

In this work, we present an integrated framework for dissecting the metabolic effects of key operational parameters – here μ and DO – during recombinant protein production by *P. pastoris* under glucose-limited conditions in continuous cultures. As a case study, we analyzed the metabolic behavior of a recombinant strain producing the sweet-tasting low-calorie protein thaumatin. This protein has 207 amino acid residues and 8 disulfide bonds (Illingworth et al. 1989), which are critical for its sweet taste (Masuda et al. 2016) and are considered the main reason behind the low titers achieved so far (Moralejo et al. 2001) (~ 100 mg L^{-1} in high-density cell cultures (Masuda et al. 2010)).

Folding of recombinant proteins with many disulfide bounds is both difficult and costly and demands high endoplasmic reticulum (ER) folding capacity to avoid triggering unfolded protein response (UPR) and ER associated degradation (ERAD) pathways (Gasser et al. 2007; Puxbaum et al. 2015). Thus, understanding the effects of μ and DO that have a strong impact on metabolism is critical for optimizing heterologous protein production in *P. pastoris*. To capture the individual and combined effects of μ and DO on the metabolic performance of this strain, we implemented a fractional experimental design to determine the statistically significant effects with a reduced dataset, and later integrated these with a tailored Genome-Scale Metabolic Model (GSMM) to interpret and further analyze the metabolic and energetic consequences. Briefly, GSMMs describe the set of biochemical reactions that represent the entire metabolic potential of an organism, and as such, are excellent tools to understand the physiological state of the cell under different environmental conditions (Price et al. 2004). Our results revealed opposite operational regions for efficient growth and protein production, as well as a tighter energetic constraint on *P. pastoris* growth under glucose-limited, low DO continuous culture conditions.

2.2 MATERIALS AND METHODS

2.2.1 Plasmid construction and strain transformation

The thaumatin gene – including its natural pre-region secretion signal – was synthesized by Genscript (Piscataway, NJ, USA) and was codon-optimized for expression in *P. pastoris*. The commercial recombinant vectors pGAPZB (Invitrogen, Carlsbad, CA, USA) were employed for the construction of the transformation plasmids using Gibson assembly (Gibson et al. 2009), and later used to integrate the thaumatin gene into the yeast´s genome under the constitutive GAP promoter. Amplification of DNA fragments for Gibson assembly was carried out in 35 PCR cycles using Phusion High-Fidelity DNA Polymerase (ThermoFisher, Waltham, MA, USA). All PCR products were treated with DpnI restriction enzyme to remove original vector residues and later were purified by gel extraction, using the Qiaquick Gel Extraction kit (Qiagen, Hilden, Germany). Finally, a 10-μL reaction mix was used to transform chemically competent *E. coli* TOP 10 cells. These cells were grown at 37 °C in low salt-LB medium, containing 25 μg mL^{-1} zeocin for selection of clones transformed with pGAPZB-TAU vector.

P. pastoris wild-type strain GS115 (Invitrogen, Carlsbad, CA, USA) was used as a host strain throughout this study, which was transformed using an in-house-built vector to revert its histidine auxotrophy (see Supplementary Material S2-1). AvrII was employed to linearize the transformation vector, which was introduced by electroporation into the competent the cells as described by Gasser et al. (2006). Both plasmids and transformations were verified by DNA sequencing (Macrogen Inc., Seoul, Korea).

2.2.2 Cell cultivation

Continuous cultures were started from pre-inocula grown overnight at 30 °C and 150 rpm in 200-mL shake flasks, containing YPG medium with 100 μg mL^{-1} zeocin. Prior to the inoculation of the bioreactors, each inoculum was centrifuged at 5,000 rpm for 5 min and resuspended in fresh culture medium without trace elements. Chemostat cultures were performed in 2-L benchtop Biostat B bioreactors (Sartorius AG, Göttingen, Germany) using 1 L of working volume.

P. pastoris was grown in glucose-limited chemostats under the culture conditions described by Adelantado et al. (2017) with only minor differences (see Supplementary Material S2-2). For DO control, a mixture of air, pure nitrogen and pure oxygen was used. Each gas was independently controlled using three mass flow meters (Bronkhorst High-Tech, Ruurlo, Netherlands) until the desired DO levels were reached (see below). The pH, agitation, and temperature were set respectively to 5.0, 700 rpm, and 25 °C in all conditions. Samples were taken after four and five residence times when the culture reached steady state.

2.2.3 Experimental design

Experimental design is a statistical tool that is used to systematically examine the effects of two or more factors that shape a determined response function. Among the different possible designs, Doehlert designs display higher efficiency and flexibility as they allow fitting more complex response models with less data and without requiring extreme points (Doehlert 1970; Hanrahan and Lu 2006).

In this study, a Doehlert matrix (Doehlert 1970) was constructed to evaluate the *individual* and *combined* effects of μ and DO on the macroscopic growth and production

parameters of the thaumatin-producing *P. pastoris* strain. The experimental domain was defined over 0.05-0.15 h^{-1} for μ with five dilution rate levels; 0.05, 0.075, 0.1, 0.125 and 0.15 h^{-1} and over 0-60% DO saturation with three oxygenation levels; low (4%, L condition), medium (30%, M condition) and excess (56%, E condition) (Figure 2-1A) (Additional studies at 108% DO saturation are in Supplementary material S2-4). The latter DO levels correspond to approximately 10.4 (L), 77.7 (M) and 145.1 μM (E) DO concentrations assuming a Henry's constant of 957.7 atm M^{-1} at 25 °C for a fermentation broth (Saa et al. 2012). This design samples seven experimental conditions in total: six external conditions performed in duplicate and in random order, and a central point performed in triplicate. Finally, a second-degree polynomial (Equation 2-1) was fitted to determine the response of the dependent variable as a function of the different factors, i.e., μ and DO.

$$Z = b_0 + b_1 X_1 + b_2 X_2 + b_{11} X_1^2 + b_{12} X_1 X_2 + b_{22} X_2^2 \quad (2\text{-}1)$$

In Equation 2-1, Z denotes the predicted response, X_1 and X_2 represent the independent variables, b_0 is the mean value of the response, b_1 and b_2 are linear (main) effects, b_{11} and b_{22} represent squared effects, and b_{12} describes an interaction term (combined effect). Statistical calculations and significant tests (ANOVA) were performed using STATGRAPHICS Plus 5.1 (Statpoint Technologies, Inc., VA, USA).

2.2.4 Extracellular metabolite and biomass quantification

Extracellular glucose, arabitol and ethanol were analyzed by HPLC. Samples of 2 mL each were taken and centrifuged at 10,000 rpm and 4 ºC for 5 min. The supernatant was collected and filtered using 0.45 μm-pore nitrocellulose filters (Merck Millipore, Carrigtwohill, Ireland). HPLC analysis was performed in duplicate in a HP 1050 liquid chromatography system (Dionex Corporation, Sunnyvale, USA) as described by Tomàs-Gamisans et al. (2016). Extracellular thaumatin was quantified in triplicate using Thaumatin ELISA Kit (CellTrend, Luckenwalde, Germany), according to manufacturer instructions. Finally, Dry Cell Weight (DCW) was quantified in triplicate, following the method of Jordà et al. (2012).

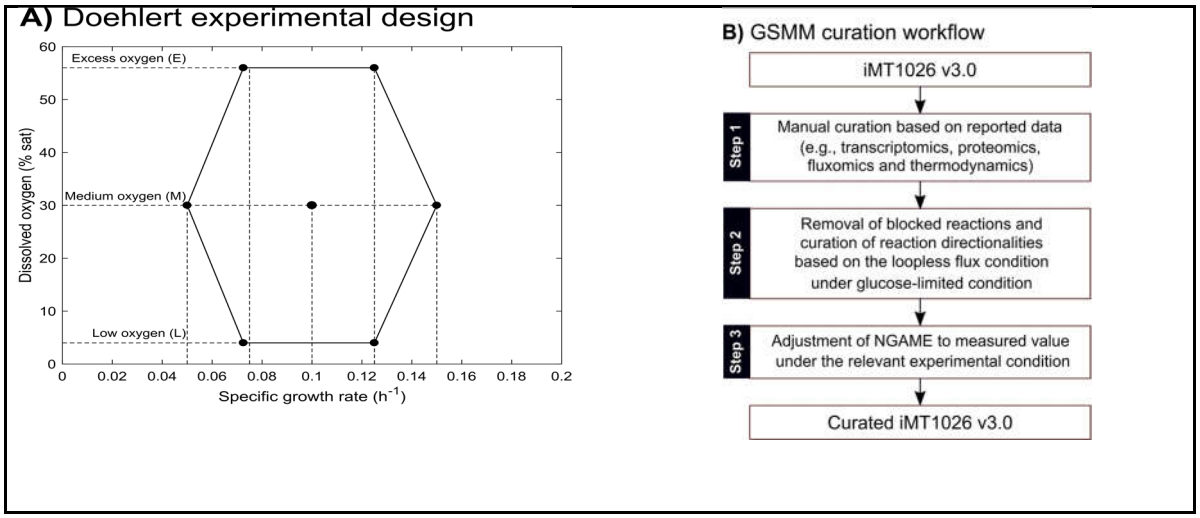


Figure 2-1 Overview of the experimental design and GSMM curation workflow employed in this study. A) Doehlert fractional experimental design for evaluating the metabolic impact of the specific growth rate (μ) and dissolved oxygen (DO) level on the metabolic performance of *P. pastoris*. This design considers 1 central point and 6 extreme points over 3 DO levels (4%, 30% and 56% oxygen saturation), and 5 μ levels (0.05, 0.075, 0.1, 0.125 and 0.15 h^{-1}). B) Manual GSMM curation workflow for flux simulations. The latest *P. pastoris* GSMM (*i*MT1026 v3.0) was sequentially tailored for accurately describing metabolic fluxes under glucose-limited conditions using reported experimental data and flux feasibility criteria.

2.2.5 Contextualization of the genome-scale metabolic model

Several GSMMs have been made available for modelling *P. pastoris* metabolism under different conditions and optimizing heterologous protein production (Caspeta et al. 2012; Irani et al. 2016; Sohn et al. 2010; Tomas-Gamisans et al. 2018; Ye et al. 2017). Typically, these models are validated by comparing experimental and predicted specific growth and exchange rates, obtained using Flux Balance Analysis (FBA) (Orth et al. 2010). However, this is often insufficient to validate GSMMs, as identical FBA predictions may be underpinned by completely different intracellular flux distributions. As such, consideration of available intracellular flux data (e.g., measured by ^{13}C fluxomics), is an important (often neglected) curation step for accurately capturing the metabolic mechanisms behind the simulated phenotype. An excellent example of the above was recently presented by Pereira et al. (2016).

In order to obtain accurate metabolic predictions underpinned by realistic flux distributions, we set up a sequential curation workflow starting from the latest and most curated *P. pastoris* GSMM – the *i*MT1026 v3.0 (Tomas-Gamisans et al. 2018) –, where

intracellular reactions were incrementally constrained – both activities and directionalities – in accordance with fluxomics, proteomics and transcriptomics data under appropriate experimental conditions (Clasquin et al. 2011; Gasser et al. 2007; Rußmayer et al. 2015; Zhang et al. 2017), and thermodynamically feasible directionalities derived from the loopless flux condition imposition (Saa and Nielsen 2016) (Figure 2-1B). In this way, the tailored metabolic model generated realistic intracellular fluxes and predicted accurately observed specific growth and exchange rates. Additional details about the contextualization workflow can be found in Supplementary Material S2-3.

2.2.6 Estimation of energetic requirements for growth

GSMMs have cellular ATP requirement divided into Growth-Associated Maintenance Energy (GAME) and Non-Growth-Associated Maintenance Energy (NGAME). The former represents the (milli)moles of ATP required for growing 1 gram of biomass ($Y_{X,ATP}$) – represented by the corresponding stoichiometric coefficient in the biomass equation –; whereas the latter (NGAME) describes the ATP consumption (m_{ATP}) required for cellular maintenance (Pirt 1982). These parameters together quantify the requirement of total ATP per gram of cell dry weight and per hour (q_{ATP}) for growing under a defined condition and follow the relation shown below.

$$q_{ATP} = Y_{X,ATP} \cdot \mu + m_{ATP} \tag{2-2}$$

So far, the energetic impact of the oxygenation on *P. pastoris* metabolism has not been fully determined, as reported experimental designs are inadequate to properly assess its effects, e.g., chemostat data has only been reported at extreme DO conditions and at a single μ of 0.1 h^{-1} (Adelantado et al. 2017; Baumann et al. 2008; Garcia-Ortega et al. 2017). In contrast, the proposed experimental design enables examination of the energetic parameters that best fit the experimental data (growth and production rates at different DO and μ conditions), shedding light on the energetic impact of the DO level. To determine the latter, the q_{ATP} of the contextualized GSMM was maximized in all the experimental conditions using the loopless FBA (ll-FBA) algorithm (Saa and Nielsen 2016) – i.e., a flux estimator that avoids thermodynamically infeasible loops. For this task, the m_{ATP} reaction was chosen as the objective, the GAME was set to zero, and the growth and exchange rates were fixed to the observed values. Once q_{ATP} was computed, the GAME was fitted to the

different DO conditions, using Eq. (2) after setting NGAME to the most recent measured value (Rebnegger et al. 2016). At this point, two models were hypothesized and fitted: one where $Y_{X,ATP}$ was the same for *all* DO conditions; and another, where $Y_{X,ATP}$ was fitted to *each* DO condition. Finally, the Akaike's Information Criterion for low number of data points (AIC_c) was applied to compare the goodness-of-fit of each model (Hurvich and Tsai 1989). The model with the smallest AIC_c was considered the best choice.

2.3 RESULTS

2.3.1 Effects of the specific growth rate and oxygenation level on the metabolic performance of *P. pastoris* under glucose-limited conditions.

Growth and metabolic production parameters of the recombinant *P. pastoris* strain were measured in seven chemostats, under different μ and oxygenation conditions. In each case, biomass and thaumatin yields ($Y_{S,X}$ and $Y_{S,P}$, respectively), specific thaumatin and carbon dioxide production rates (q_P and q_{CO2}, respectively), and specific glucose and oxygen consumption rates (q_S and q_{O2}, respectively) were estimated (Table 2-1). In all conditions, ethanol and arabitol were not detected, pointing to a purely respiratory metabolism which was later confirmed with measured RQ values (Supplementary Material S2-5).

Oxygenation exerted the greatest impact on both, $Y_{S,X}$ and q_P, whereas μ had the main effect on $Y_{S,P}$. In the case of $Y_{S,X}$, there were notable differences between different oxygenation levels at a given μ (e.g., L and E conditions showed a 14% relative difference for $\mu = 0.075\ h^{-1}$), but not vice versa (< 4.7% relative difference for a fixed oxygenation level and different μ). Overall, $Y_{S,X}$ varied from 0.507 to 0.589 $g_{DCW}\ g^{-1}$, attaining the lowest yield at $\mu = 0.075\ h^{-1}$ in the L condition. Statistical analysis showed that DO had the main positive and significant effect (*p-value*<0.001, Supplementary Material S2-6), although the interactive and quadratic terms displayed also significant (both negative) influences (respectively *p-value*<0.01 and *p-value*<0.05, refer to Supplementary Material S2-6 for details). The $Y_{S,X}$ regression explained a substantial part of the variability in the data ($R^2 = 0.948$) and, more importantly, it suggested high oxygenation and low μ as the most likely conditions for efficient growth (Figure 2-2A).

In the case of the thaumatin production parameters q_P and $Y_{S,P}$, they showed different behaviors. While $Y_{S,P}$ was almost unaffected by DO and displayed a negative

correlation with μ (e.g., varying from 0.129 to 0.399 mg thaumatin g^{-1} in the high and low μ conditions, respectively), q_P was largely influenced by DO (negatively); and to a lesser extent by μ (positively) (Table 2-1 and Figure 2-2B). In this case, the regression analysis explained a lower fraction of the data variability (R^2 = 0.68) and indicated a significant negative effect of DO (*p-value*<0.05, Supplementary Material S2-6) with a positive interaction term (*p-value*<0.05, Supplementary Material S2-6). Here, the magnitude of both individual and combined effects was quite similar (see Supplementary Material S2-6). In the case of the $Y_{S,P}$, significant individual (negative) and quadratic (positive) effects of μ were suggested by the regression analysis (in both cases *p-value*<0.001, and R^2 = 0.959, see Supplementary Material S2-6). Altogether, regions of both, low μ and oxygenation, favor thaumatin production and yield (Figure 2-2B-C).

Table 2- 1 Metabolic performance parameters of recombinant *P. pastoris* in glucose-limited chemostats under different μ and DO conditions[a]

Condition[b]	μ (h^{-1})	q_S (mmol g_{DCW}^{-1} h^{-1})	q_{CO2} (mmol g_{DCW}^{-1} h^{-1})	q_{O2} (mmol g_{DCW}^{-1} h^{-1})	q_P (mg g_{DCW}^{-1} h^{-1})	$Y_{S,X}$ (g_{DCW} g^{-1})	$Y_{S,P}$ (mg g^{-1})
1 (L)	0.072±0.003	0.787±0.038	1.838±0.101	1.842±0.097	0.047±0.005	0.507±0.002	0.267±0.042
2 (L)	0.119±0.006	1.247±0.031	2.766±0.017	2.791±0.317	0.036±0.003	0.531±0.011	0.159±0.016
3 (M)	0.048±0.002	0.480±0.029	1.028±0.051	1.052±0.068	0.034±0.001	0.551±0.008	0.399±0.019
4 (M)	0.101±0.001	0.989±0.009	1.980±0.016	1.973±0.076	0.033±0.001	0.566±0.002	0.186±0.001
5 (M)	0.150±0.002	1.495±0.021	3.016±0.052	2.955±0.099	0.035±0.001	0.566±0.002	0.129±0.002
6 (E)	0.072±0.001	0.682±0.007	1.213±0.004	1.300±0.005	0.031±0.001	0.589±0.004	0.256±0.007
7 (E)	0.120±0.001	1.143±0.014	2.182±0.147	2.309±0.129	0.032±0.001	0.583±0.004	0.175±0.009

[a]Results represent the average and the standard error of two or three independent experiments in the case of extreme points and the central point, respectively. In each case, the carbon balance closed satisfactorily and reconciled no less than 94.7% of the consumed carbon source (see Supplementary Material S2-5).

[b]The experimental conditions are abbreviated L for low oxygenation (4% dissolved oxygen saturation), M for normal oxygenation (30% dissolved oxygen saturation), and E for excess oxygenation (56% dissolved oxygen saturation).

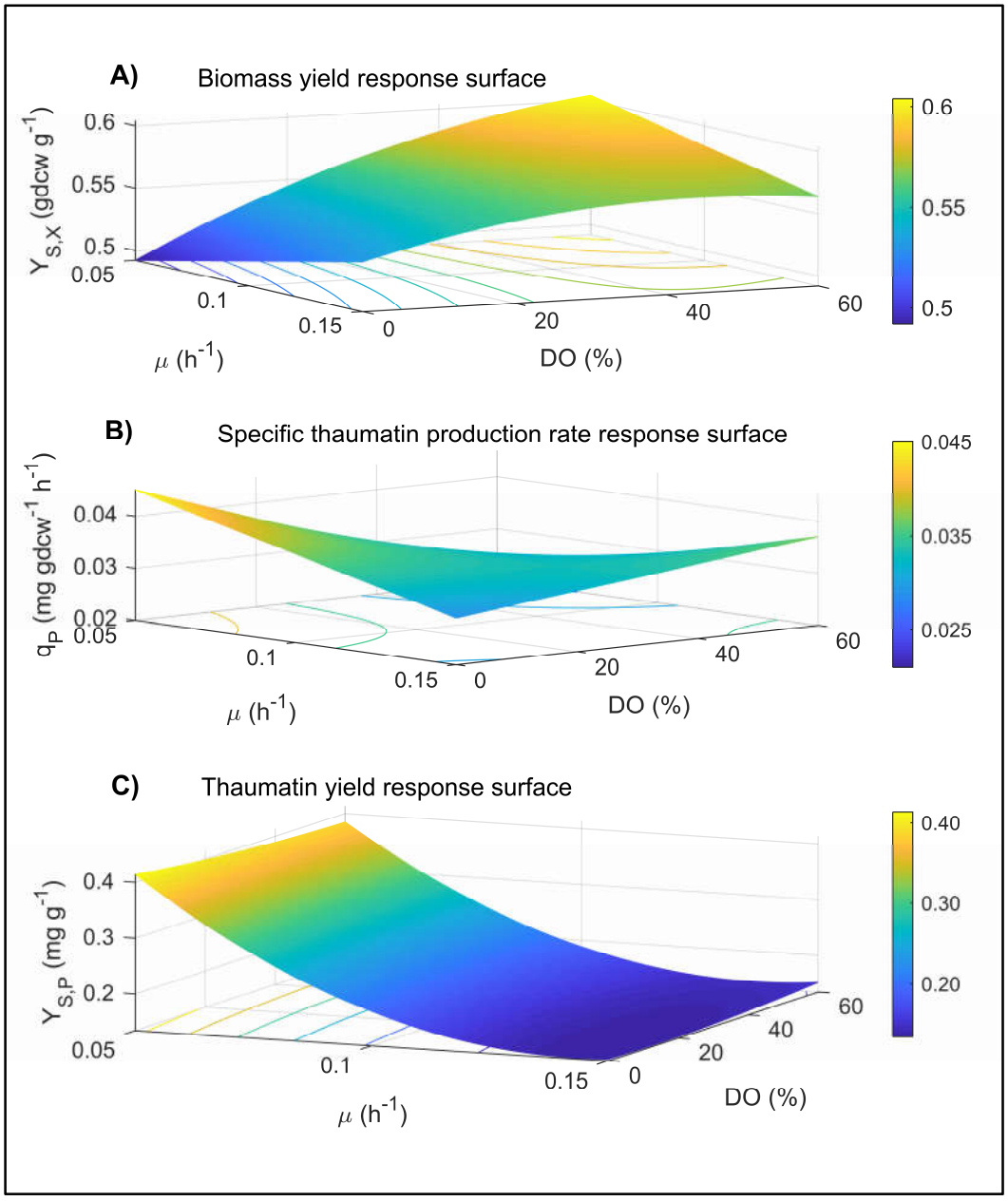


Figure 2-2 Response surfaces of *P. pastoris* macroscopic culture parameters under different DO and μ conditions. The fitted response surfaces were for the (encoded) biomass yield (**A**), specific thaumatin production rate (**B**) and thaumatin yield (**C**) are, respectively: $Y_{S,X} = 0.565 + 0.036 \cdot \hat{\mu} - 0.02 \cdot \hat{\mu} \cdot \hat{DO} - 0.017 \cdot \hat{DO}^2$ ($R^2 = 0.948$), $q_P = 0.033 - 0.004 \cdot \hat{DO} + 0.008 \cdot \hat{\mu} \cdot \hat{DO}$ ($R^2 = 0.68$), and $Y_{S,P} = 0.186 - 0.127 \cdot \hat{\mu} + 0.08 \cdot \hat{\mu}^2$ ($R^2 = 0.959$), where $\mu = 0.05 + 0.05 \cdot (\hat{\mu} + 1)$, $\hat{\mu} \in [-1, 1]$ and $DO = 4 + 30 \cdot (\hat{DO} + 0.8666)$, $\hat{DO} \in [-0.8666, 0.8666]$. Further details about the variable encoding, statistical effects and fitting can be found in Supplementary Material S 2-6.

2.3.2 GSMM curation and GAME evaluation under aerobic, glucose-limited conditions

The prediction power of *i*MT1026 v3.0 was improved through a sequential manual curation workflow. *i*MT1026 v3.0 supports growth on glycerol and methanol through the addition of accessory and degradation pathways in different compartments (e.g., peroxisome) (Tomas-Gamisans et al. 2018). The resulting model, however, consistently overestimated the maximum specific growth rate under aerobiosis with glucose as the sole carbon source, due to unrealistic flux rerouting through the peroxisome and unlikely activation of some redox reactions (Supplementary Material S2-7). Imposing appropriate reaction constraints, based on experimental and thermodynamic information, largely solved this issue.

Inclusion of transcriptomic and proteomic data under appropriate experimental conditions (Clasquin et al. 2011; Gasser et al. 2007; Rußmayer et al. 2015; Zhang et al. 2017), and directionalities based on reaction thermodynamics, reduced the number of reversible reactions by 25 and blocked 3 initially active reactions (Figure 2-3A). Furthermore, subsequent thermodynamic feasibility analysis, based on imposition of the loopless flux condition, yielded irreversible 215 reactions previously deemed reversible (Figure 2-3A). Altogether, these measures disabled the infeasible cycles responsible for the flux prediction inaccuracies. Finally, adjustment of the NGAME to 0.55 mmol ATP g_{DCW}^{-1} h^{-1} (assuming a yield of 32 mol of ATP per mol of glucose (Tomàs-Gamisans et al. 2016), and using the recent experimental maintenance value of 3.1 mg glucose g_{DCW}^{-1} measured in glucose-limited retentostats (Rebnegger et al. 2016)), increased the prediction fidelity of *i*MT1026 v3.0 by ca. 86% for the observed μ – from approx. 7.0% average relative error to 4.2% (Figure 2-3B). Moreover, the agreement between predicted and measured intracellular flux data improved greatly – explained flux variability increased from 78.3% to 97.2% in the cytosol (Figure 2-3C) and from 7.6% to 86.7% in the mitochondria (Figure 2-3D). We note that *i*MT1026 v3.0 uses a biomass macromolecular composition with ashes, although correction by the estimated ashes content yielded almost identical results (Supplementary Material S2-7). The final list with constraints employed can be found in Supplementary Material S2-7.

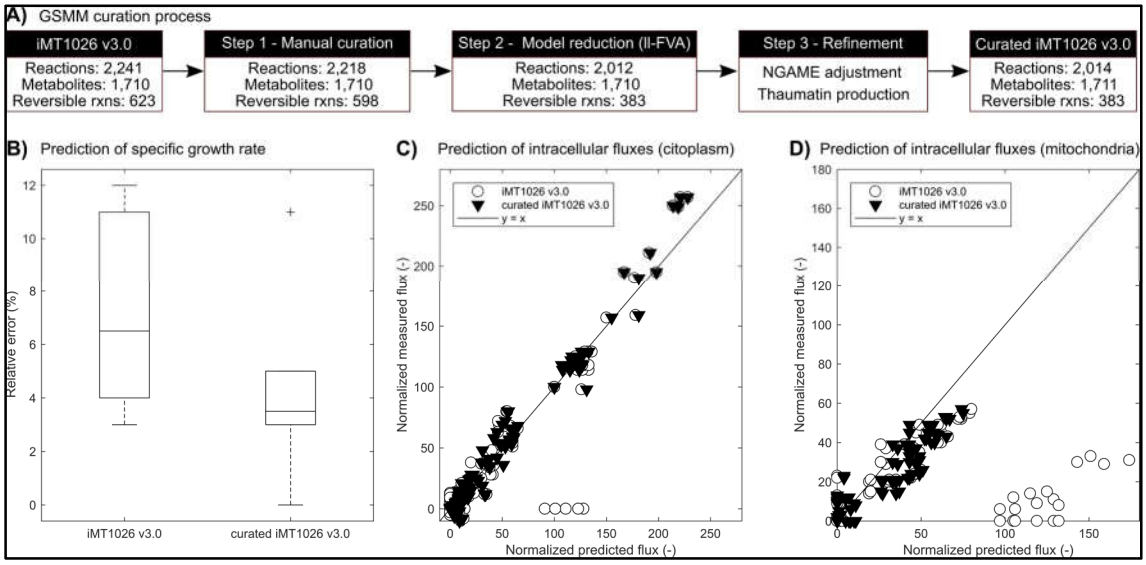


Figure 2-3 Evaluation of curated *P. pastoris* GSMM for metabolic flux prediction under glucose-limited conditions. **A)** Step-by-step curation process of the *P. pastoris* iMT1026 v3.0 GSMM for describing growth under glucose-limited conditions. **B)** Evaluation of maximum specific growth rate predictive performance of the initial versus curated GSMM. **C)** Comparison of the predicted intracellular flux through cytoplasmatic reactions of the initial (slope = 0.911, R^2 = 0.783) and curated GSMM (slope = 1.047, R^2 = 0.972) against experimental data under appropriate growth conditions. **D)** Comparison of the predicted intracellular flux through mitochondrial reactions of the initial (slope = 0.112, R^2 = 0.076) and curated GSMM (slope = 0.721, R^2 = 0.867) against experimental data under the same conditions used in **C)**. For more information about the employed experimental conditions refer to Methods.

The tailored *i*MT1026 v3.0 was employed to evaluate the energetic effects of the DO level on the *P. pastoris* GAME. To this end, the GAME parameter was refitted to the experimental data of this study and allowed to vary depending on the DO level (refer to *Estimation of energetic requirements for growth* for details). Our results showed that the model with a variable GAME as a function of the DO level explains a higher amount of the data variability (94.9% compared to 87.1%), and has a stronger statistical support, compared to the alternative (Figure 2-4A). Notably, this model displayed a decreasing GAME with DO level, starting from 95 mmol ATP g_{DCW}^{-1} for the L condition, and descending to 75 and 66 mmol ATP g_{DCW}^{-1} for the M and E conditions, respectively (Figure 2-4B). These results indicate a decreasing growth cost with increasing DO level.

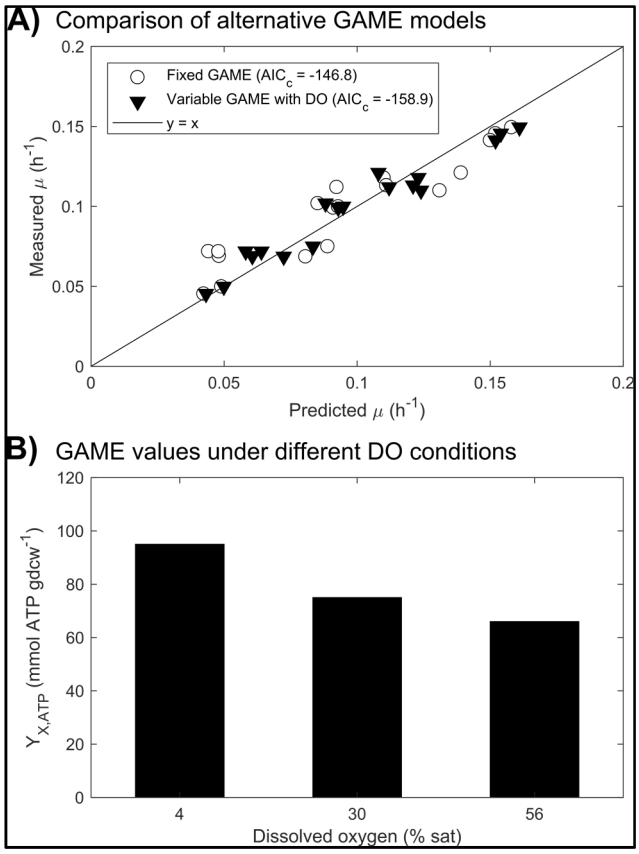


Figure 2-4 Evaluation of the DO influence on the growth-associated maintenance energy (GAME) requirement of *P. pastoris* under glucose-limited conditions. **A)** Comparison of the specific growth rate predictions of the model with a fixed GAME for all DO conditions (slope = 0.752, R^2 = 0.871) and the model with a variable GAME for each DO condition (slope = 0.844, R^2 = 0.949) against the experimental growth data of this study. The model with a variable GAME for each DO condition has stronger statistical support as shown by the lower AIC_c value (-158.9 versus -146.8). **B)** Fitted GAME values for different DO conditions.

2.3.3 Simulation of intracellular fluxes under extreme DO levels

As the DO level had the greatest effect on the metabolic parameters of *P. pastoris*, the intracellular flux distributions under extreme DO conditions were simulated at a fixed μ level (μ = 0.075 h^{-1}). The relative glycolytic flux increased by approximately 15% in the L condition, although no substantial carbon redistributions were predicted as observed by the almost identical glycolysis, PPP and biosynthetic relative fluxes (Figure 2-5A). Notably, both the TCA cycle and ATP transport relative fluxes increased as expected by the increased q_{CO2} and q_{O2} (Figure 2-5A and Table 2-1). Although there is also an important

difference between the pyruvate export between both conditions, there is no compensation by other by-products, i.e., ethanol (Figure 2-5A). Surprisingly, analysis of the feasible flux ranges revealed slightly tighter metabolic constraints under the L condition, despite the increased energetic growth cost (i.e., higher $Y_{X,ATP}$), which is consistent with the overall increased glucose uptake (Figure 2-5B).

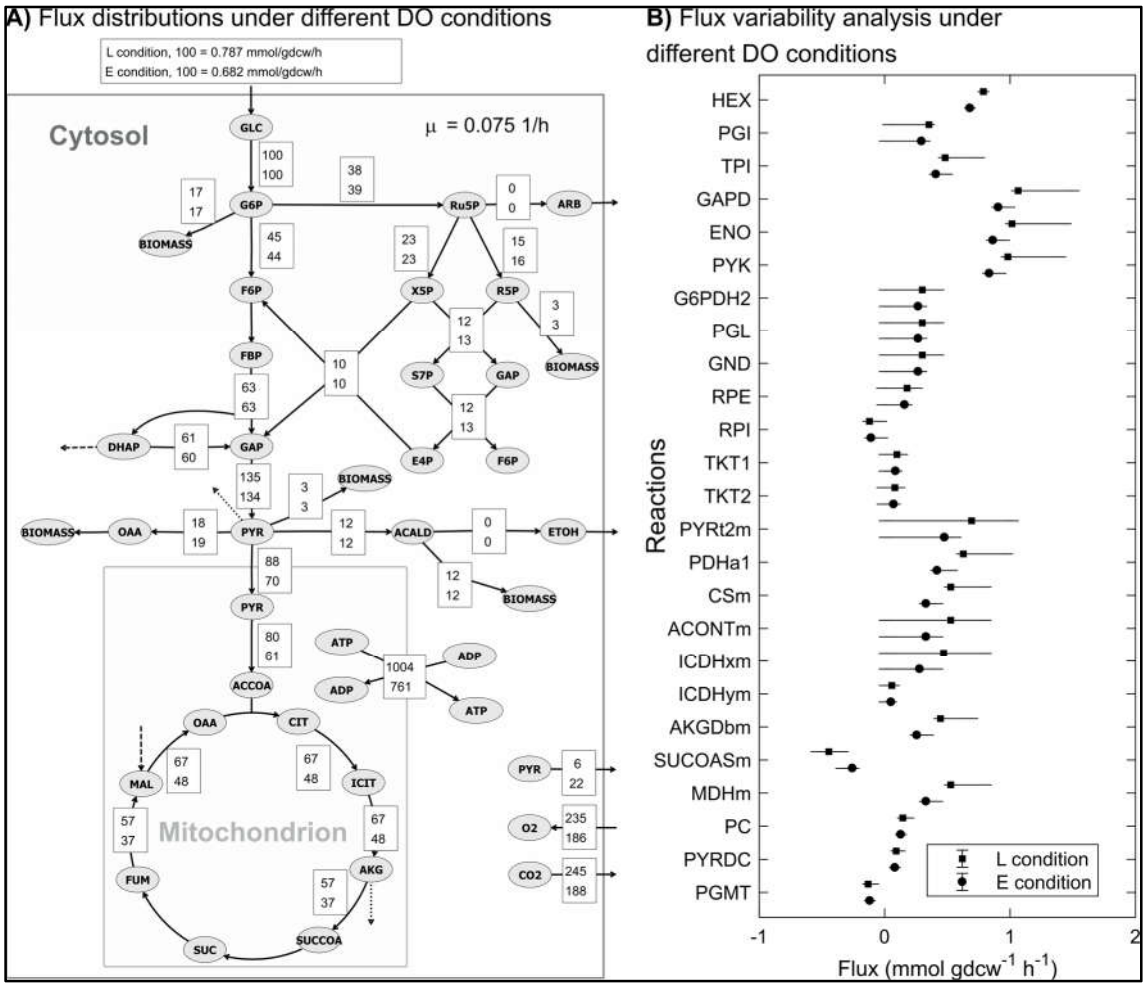


Figure 2-5 Flux distribution analysis of P. *pastoris* metabolism under extreme DO conditions. **A)** Predicted metabolic flux distributions under extreme DO conditions (4% and 56% oxygen saturation) at $\mu = 0.075\ h^{-1}$. Parsimonious FBA (Lewis *et al*. 2010) was employed to obtain a representative flux distribution under each condition. **B)** Feasible flux ranges for main central carbon reactions of *P. pastoris* under extreme oxygenation conditions. Flux ranges were calculated using loopless FVA (Saa and Nielsen 2016) as described in the Methods section.

2.4 DISCUSSION

2.4.1 Oxygenation exerts an *individual* negative effect and a *synergistic* positive effect with the specific growth rate on thaumatin production

The specific thaumatin production rate was influenced negatively by the DO level, and positively by μ through the interaction with DO (Figure 2-2). The latter effect is consistent with previous reports of a positive correlation between μ and protein secretion rates at fixed medium DO level, although the estimated effect was not so strong as previously reported (Buchetics et al. 2011; Maurer et al. 2006). Recent studies have shown that specific protein production in *P. pastoris* also depends more on the nature of the produced protein and the oxidative conditions than on μ (Gasser et al. 2006; Liu et al. 2013; Uchiyama and Shioya 1999). Indeed, despite having constitutive protein expression, non-growth-associated protein production patterns may arise due to global stress responses, high protein turnover and limitations in endoplasmic reticulum processing, among others (Liu et al. 2013). The latter points to protein expression, processing and/or secretion as likely – and increasingly more recognized – bottlenecks (Looser et al. 2015).

The calculated individual DO effect on protein production agrees well with previous studies using the PGAP-based expression system, reporting higher protein secretion under hypoxic conditions at $\mu = 0.1\ h^{-1}$ (Adelantado et al., 2017; Baumann et al., 2008; Garcia-Ortega, Valero, & Montesinos-Seguí, 2017). Likely causes of the latter considered stronger transcriptional induction of glycolysis, modification of membrane composition, and upregulation of Unfolded Protein Response (UPR) genes (Adelantado et al. 2017; Baumann et al. 2010). The first cause is particularly supported by our data and flux simulations as the absolute glycolytic fluxes were higher under the L condition at a fixed μ of $0.075\ h^{-1}$ (Table 2-1 and Figure 2-5A). In fact, increased transcriptional levels of glycolytic genes under low DO levels have been suggested as a plausible cause for higher protein production under the (glycolytic) GAP promoter (Baumann et al. 2010). Importantly, our analysis revealed that the DO negative effect can be potentially counteracted at high μ, yielding a more complex picture of the optimal protein production frontier (see below).

2.4.2 Optimal growth and thaumatin production lie in opposite operational regions

Identification of the response surface of the macroscopic growth (i.e., $Y_{S,X}$) and production (i.e., q_P) parameters enables determination of the *optimal operational* regions for achieving efficient growth and increased productivity, respectively. While maximum q_P is predicted to be at either low or high DO and μ, optimal $Y_{S,X}$ is strictly confined to the high DO and low μ region (Figure 2-6). The former case is somewhat different from what has been reported (Hensing et al. 1995), specifically in the case of high μ (> 0.2 h^{-1}) and DO (> 40%). A closer analysis of the statistical effects of μ and DO revealed, however, that μ only has an important positive influence at very high DO conditions, and not vice versa (Figure 2-2B). Reported metabolic studies are typically performed at low-medium oxygenation (< 30%), which may explain why this synergistic effect has gone unnoticed for other proteins. We note, however, that further experimentation is required to validate this particular region for optimal thaumatin production. In the case of $Y_{S,X}$, our results indicated a positive correlation with DO that is consistent with the available data at constant μ (Adelantado et al. 2017; Baumann et al. 2008; Garcia-Ortega et al. 2017). Overall, our model indicates that optimal growth and protein production lie in opposite operational regions, recapitulating the known natural trade-off between these metabolic tasks (Buchetics et al. 2011).

2.4.3 Predictive GSMM reveals higher energetic growth requirements of *P. pastoris* at lower DO levels

The introduction of proper constrains to GSMMs based on experimental data and feasibility criteria is critical to achieve reliable flux predictions under the simulated conditions. Here, we have employed a rigorous step-by-step curation process where the predictive capabilities of the GSMM were evaluated and improved (Figure 2-3), which later enabled us to evaluate the energetic impact of DO on *P. pastoris* growth. Adjustment of the GAME for different DO improved significantly the growth predictions of the GSMM (Figure 2-4A), revealing an important energetic effect of this variable on *P. pastoris* metabolism. Notably, the fitted GAME values showed excellent agreement with available data (e.g., the predicted GAME under the M condition was 75 mmol ATP g_{DCW}^{-1},

similar to the reported values of 70.5 and 72 mmol ATP g_{DCW}^{-1} under comparable DO conditions (Caspeta et al. 2012; Tomàs-Gamisans et al. 2016), which prompted us to analyze the DO impact on the metabolism of *P. pastoris*. Our results showed a linear GAME decrease of 0.56 mmol ATP g_{DCW}^{-1} per 1% increase in DO (Figure 2-4B), which describes an important increase in the growth efficiency at high DO levels (up to approx. 16% $Y_{S,X}$ decrease between L and E, Table 2-1). Our results also indicate that the extra metabolic capacity is not transferred to thaumatin production (Table 2-1) or associated to a substantial carbon redistribution (Figure 2-5A), which suggests DO plays a more complex regulatory role likely related to changes in cell composition and/or transcriptional regulation (Baumann et al. 2010; Gasser et al. 2007). More notably, our results assert that the DO level is a key operational variable for optimizing protein production in *P. pastoris*.

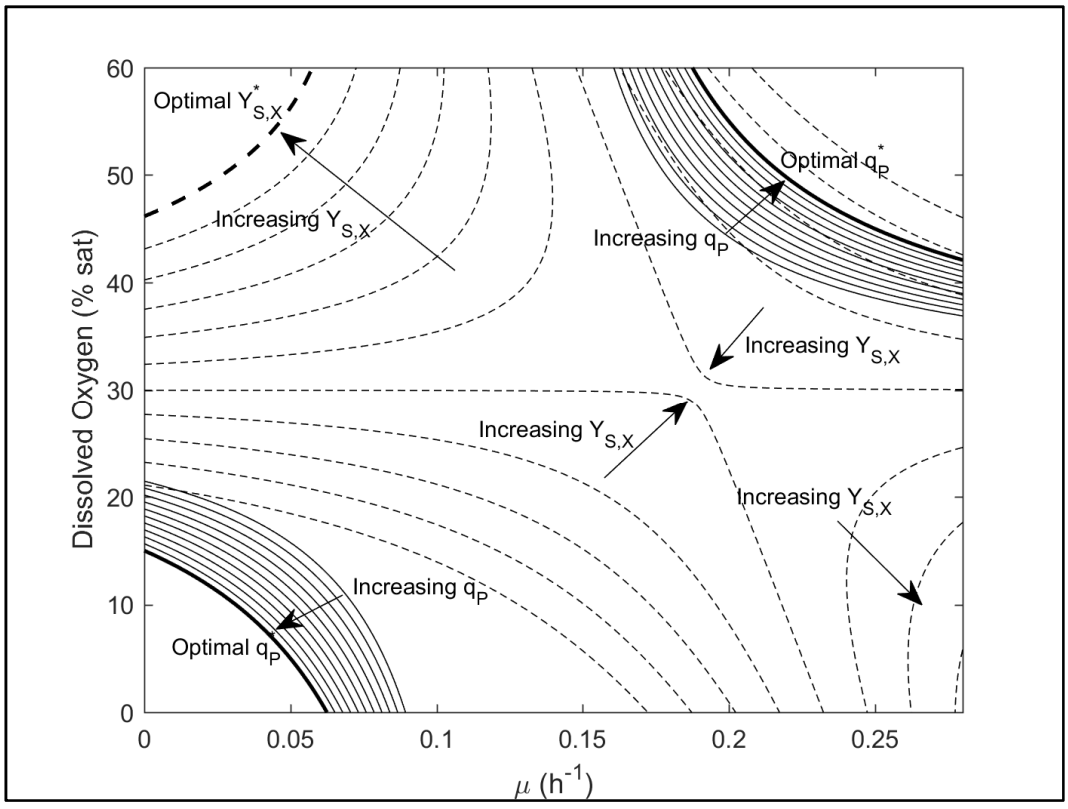


Figure 2-6 Phase plot of optimal *P. pastoris* growth and thaumatin production as a function the specific growth rate and dissolved oxygen. The optimal conditions for efficient growth (dashed line) and increased thaumatin production (continuous line) are shown in thick black lines, whereas suboptimal conditions are represented with thin lines using the same symbols as before. As shown in the figure, optimal growth conditions lie opposite to optimal production conditions.

3. CHAPTER II: DYNAMIC GENOME-SCALE METABOLIC MODELING OF THE YEAST *PICHIA PASTORIS*

Francisco Saitua[1], **Paulina Torres**[1], José Ricardo Pérez-Correa[1] and Eduardo Agosin[1].

[1] Department of Chemical and Bioprocess Engineering, School of Engineering, Pontificia Universidad Católica de Chile, Vicuña Mackenna 4860, Santiago, Chile.

Published in *BMC* Systems Biology (2017). Vol. 11, Issue 27

3.1 INTRODUCTION

Recombinant protein production is a multibillion-dollar business, mainly comprised by therapeutic agents (i.e. recombinant biologic drugs) and industrial enzymes (BCC Research, 2014; Markets and Markets, 2015; Walsh, 2014). These compounds are commonly synthesized in *Escherichia coli*, *Saccharomyces cerevisiae* and Chinese Hamster Ovary cells (CHO) (Ferrer-Miralles *et al*., 2009; Maccani *et al*., 2014; Overton, 2014; Walsh, 2014); however, there is strong pressure to find cost-effective alternatives to overcome technical and economic disadvantages of the aforementioned cell factories, especially in downstream processing (Corchero *et al*., 2013).

Among the unconventional cell factories used for recombinant protein production, the methylotrophic yeast *Pichia pastoris* (syn. *Komagataella phaffii*) has received special attention thanks to its convenient physiology and easy handling (Daly & Hearn, 2005). There are strong promoters for this cell factory which are commercially available and that allow for the controlled expression of heterologous proteins (Daly & Hearn, 2005). Unlike *E. coli*, *P. pastoris* naturally performs post-translational modifications (Cereghino & Cregg, 2000; Ferrer-Miralles *et al*., 2009), which are essential for most eukaryotic protein functionality (Ciofalo *et al*., 2006; Corchero *et al*., 2013; Masuda *et al*., 2010). In contrast to *S. cerevisiae*, *P. pastoris* exhibits a Crabtree-negative phenotype, showing a reduced synthesis of undesirable products, like ethanol, in glucose-limited conditions (Çalık *et al*., 2015; Mattanovich *et al*., 2009). It also shows a lower basal secretion of proteins when

compared to other yeasts, which makes downstream processing easier (Delic *et al*., 2013; Mattanovich *et al*., 2009). Finally, *P. pastoris* can be efficiently cultivated up to high cell densities using fed-batch technology (Daly & Hearn, 2005), achieving high titers and productivities. For these desirable features, *P. pastoris* has been widely used for the expression of recombinant proteins, reaching grams per liter concentrations in several cases (Cereghino & Cregg, 2000; Čiplys *et al*., 2015; Hasslacher *et al*., 1997; Heyland *et al*., 2010; Wang *et al*., 2001). Most remarkably, and as proof of its technical feasibility and adequacy, two recombinant proteins produced in this cell factory have already been approved by the FDA for medical purposes (Ciofalo *et al*., 2006; Thompson, 2010).

Despite its growing acceptance and actual successful applications, recombinant protein production in *P. pastoris* can be undermined by several cellular processes, where protein folding and secretion are the most recurrent bottlenecks (Delic *et al*., 2013, 2014; Gasser *et al*., 2013). In addition, limitations may also be caused by the codon usage of the recombinant protein (Wang *et al*., 2015), promoter selection (Prielhofer *et al*., 2013), carbon and oxygen availability in the culture (Baumann *et al*., 2008; Heyland *et al*., 2011) and fed-batch operational parameters (Maurer *et al*., 2006), seriously hampering protein yield, productivity and the economic feasibility of the process.

Industrially, *P. pastoris* is commonly grown in fed-batch cultures in order to maximize the titer and volumetric productivity of a desired compound, often a recombinant protein (Looser *et al*., 2015; Riesenberg & Guthke, 1999). This is achieved by adding a culture medium in such a way that the microorganism grows at a desired specific growth rate, which is chosen to maximize the synthesis of the target product and to limit the formation of inhibitory compounds (Villadsen *et al*., 2011). During this and other cultivation systems, the cells adapt constantly to the changing extracellular environment and to the limited mass transfer conditions observed at high densities (Landi *et al*., 2015; Vargas *et al*., 2011). Therefore, it is critical to understand how the cell metabolism interacts with the nutritional and environmental stresses exerted by process conditions to improve bioreactor performance (Graf *et al*., 2009). This is a complex task, however, since the strain's characteristics and process variables often require significant amounts of time and money for characterization and fine-tuning (Çalık *et al*., 2015). Therefore, it is desirable to have a platform to integrate different levels of information from dynamic

cultivations of *P. pastoris* that can be used to elaborate rational hypotheses to increase process productivity.

Systems biology offers a quantitative and comprehensive approach to address this task (Kitano, 2002). In particular, Genome-Scale dynamic Flux Balance Analysis (GS-dFBA) (Höffner *et al*., 2013; Mahadevan *et al*., 2002; Varma & Palsson, 1994) is a modeling framework that allows the simulation of metabolism during non-stationary (batch or fed-batch) cultures. GS-dFBA models couple the dynamic mass balances of the extracellular environment of the bioreactor with comprehensive mathematical representations of cellular metabolism called Genome Scale Metabolic Models (GSMs). These structures represent the cell's entire metabolism as a set of underdetermined constrained mass-balances (Palsson, 2015; Thiele & Palsson, 2010; Vargas *et al*., 2011). GSMs have been employed to understand cellular behavior under different environmental conditions, to map over omics data, and to define a metabolic engineering targets (Asadollahi *et al*., 2009; Park *et al*, 2007). There are currently five published GSMs of *P. pastoris* (Caspeta *et al*., 2012; Chung *et al*., 2010; Irani *et al*., 2016; Sohn *et al*., 2010; Tomàs-Gamisans *et al*., 2016) which have been developed to help the strain optimization process with a special emphasis on recombinant protein production. Moreover, one of these models has been successfully employed to improve recombinant protein production in *P. pastoris* (Nocon *et al*., 2014), validating these frameworks as strain engineering tools for tis particular yeast.

GS-dFBA models usually contain several parameters, whose values can be obtained by regression of experimental data. These parameters are used as inputs to obtain flux distributions throughout cultivations, so their values need to be reliable. To ensure this, pre- and post-regression diagnostics have been employed to determine if a certain parameter is supported by the observed data or not (Jaqaman & Danuser, 2006; Sánchez *et al*., 2014). These analyses consist in verifying the model's capacity to explain the behavior of a system (goodness-of-fit) and the presence of the following parametric limitations: (i) low or no impact on the state variables (sensitivity), (ii) strong correlations with other parameters of the model (identifiability) and (iii) lack of statistical significance (significance). A model is considered robust if it has the capacity to explain different conditions, while containing only sensitive, identifiable and significant parameters.

Here, we present a robust dynamic genome-scale metabolic model of *P. pastoris* in glucose-limited, aerobic batch and fed-batch cultivations. To assemble the dynamic modeling framework, we started by selecting one of the available genome-scale metabolic models (Chung *et al*., 2010) and manually curated it to yield realistic flux distributions. Then, we included it in a set of mass balances representing the main compounds present in culture supernatant. Once assembled, the model was calibrated using experimental data from eight batch and three fed-batch cultivations. Next, we employed pre/post regression diagnostics to determine sensitivity, significance and identifiability problems in the model. In order to avoid the aforementioned statistical limitations, problematic parameters were fixed (i.e. removed from the adjustable parameter set) based on the pre/post regression diagnostics, yielding reduced and potentially robust model structures. Potentially robust model structures consisted in the original model formulation with less adjustable parameters. After evaluating these reduced models for each type of cultivation, we chose the one that presented fewer parametric limitations after being re-calibrated with the available data. These reduced models yielded no (or just a few) significance, sensitivity or identifiability problems when calibrating new data and they could predict bioreactor dynamics in conditions like the ones used for their determination. Finally, we carried out simulations to assess the potential of the model to study *P. pastoris* metabolism under industrially relevant conditions, and to select molecular and process engineering strategies to improve recombinant protein production.

3.2 METHODS

3.2.1 Model construction

The structure of the model was based on an existing dFBA framework developed by Sanchez *et al* for *S. cerevisiae* (Sánchez *et al*., 2014), which divides the fermentation time into short integration periods where a metabolic steady state could be assumed (Mahadevan *et al*., 2002; Stephanopoulos *et al.*, 1998). The model considers the evolution of seven state variables throughout batch and fed-batch glucose-limited aerobic cultivations: culture volume as well as the concentrations of glucose, biomass, ethanol, arabitol, citrate and pyruvate. It consists of three linked blocks that are solved iteratively; (i) the kinetic block, (ii) the metabolic block and (iii) the dynamic block (Figure *3-1*).

First, the initial conditions of the system enter into the kinetic block to determine the specific consumption and production rates of the species involved in the analysis according to kinetic expressions. These rates are included as constraints to the corresponding exchange reactions of the metabolic model. The constrained model is then passed to the metabolic block of the framework, where the flux distribution inside the cell is determined. This procedure includes the calculation of the specific growth rate, which is passed along with the other exchange rates to the dynamic block as consumption and production terms in the mass balances. Here, the concentration of the state variables is updated and then incorporated into the kinetic block for the calculation of instantaneous exchange rates. This cycle iterates throughout the cultivation yielding the culture profile and instantaneous flux distributions that can be saved for further analysis. The latest version of model can be found online at:

https://github.com/fjsaitua/RY-dFBA/tree/master/main%20P_pastoris%20dFBA.

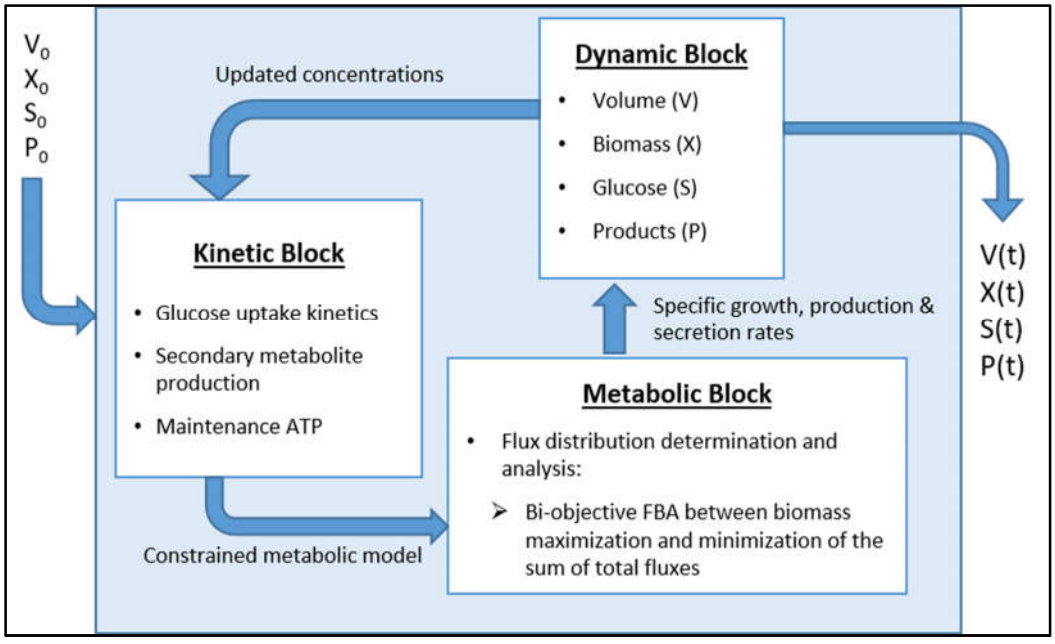


Figure 3-1 - Iterative structure of the model. V refers to culture volume [L], F_{IN} is the feeding policy used in fed-batch cultures, X, S and P are biomass, limiting substrate and Product concentration in [g/L] respectively.

3.2.1.1 Kinetic block

The kinetic block sets the uptake and production rates for all the compounds in the model. First, the glucose uptake rate (v_G) is determined using Michaelis-Menten kinetics (Postma, *et al.*, 1989).

$$v_G = \frac{v_{G,Max} \cdot G}{K_G + G} \tag{3-1}$$

Here, G is the glucose concentration in the medium [g/L], $v_{G,Max}$ is the maximum glucose uptake rate [mmol/g_{DCW}·h^{-1}] and K_G is the uptake half activity constant of this substrate [g/L]. Once determined, $-v_G$ [mmol/g_{DCW}·h^{-1}] is included as the lower bound of the corresponding exchange reaction in the model since substrate consumption is represented with a negative flux through this reaction.

Then, the lower bounds of the exchange reactions (lb) associated with the remaining k compounds (lb_k) are fixed. We considered ethanol, pyruvate, arabitol and citrate dynamics, besides glucose consumption and biomass formation.

$$lb_k = v_{P_k} \qquad k = 1 \dots 4 \tag{3-2}$$

These parameters are redefined during the fed-batch phase; therefore, they have two values during this type of cultivation.

Finally, the kinetic block fixes the non-growth associated maintenance ATP (m_{ATP}, a flux through the cytosolic ATP hydrolysis reaction in the model), which accounts for the energy drain caused by cellular processes not related with the generation of new cell material, such as osmoregulation, shifts in metabolic pathways, cell motility, etc.

3.2.1.2 Metabolic block

The metabolic block receives a constrained GSM from the kinetic block and solves an optimization problem to determine specific growth rate and the flux distribution in the cell. The GSM consists of a set of *m* metabolites and *n* reactions grouped in a Stoichiometric Matrix, S (m x n), that represents the cell's entire metabolism. If accumulation of metabolites is neglected, a mass balance can be stated according to equation (3-3):

$$S \cdot v = 0 \qquad (3\text{-}3)$$
$$s.t.$$
$$lb < v < ub$$

Where v is a vector of metabolic fluxes in [mmol/g_{DCW}·h], and lb and ub are the lower and upper bounds for each component of the flux vector.

The metabolic block solves a bi-objective Quadratic Programming (QP) problem between maximization of growth rate and minimization of the total absolute sum of fluxes (Holzhütter, 2004), subjected to the constraints imposed by the stoichiometric matrix mentioned above (Feng *et al.,* 2012):

$$Min\ \alpha \cdot \sum_{i=1}^{n} v_i^2 - (1-\alpha) \cdot \mu$$
$$s.t. \qquad (3\text{-}4)$$
$$S \cdot v = 0$$
$$lb_i \leq v_i \leq ub_i \qquad i = 1 \ldots n$$

In this formulation, α, the suboptimal growth coefficient, is an adjustable parameter from the model used to modulate the importance of the two – biologically relevant – competing objectives (Sánchez *et al.*, 2014; Schuetz *et al.*, 2007; Schuetz *et al.*, 2012). In our analysis, "optimal growth" occurs when the objective function of the cell is biomass maximization (α = 0). However, when α > 0, the calculated growth rate is lower than the theoretical maximum derived from biomass maximization, at the same glucose uptake rate. In this sense α is considered as a "suboptimal growth coefficient"; it is worthy to note that we do not refer to the optimality of the flux distribution vector, which is actually optimal, given the convexity of the problem in the metabolic block (Equation 3-4 - See Supplementary material S 3-1).

The minimization of total fluxes adds a quadratic term to the objective function, which has the practical benefit of eliminating Type III pathways *et al.*, 2002) from the flux distribution, which arise from the multiplicity of solutions of a LP problem. These

pathways appear as high fluxes (often taking the value of the upper bound of a particular flux) through closed cycles of reactions. This misleads pathway analysis because despite the mass balance around each participating metabolite is satisfied, the fluxes are thermodynamically infeasible (Price *et al.*, 2002). The use of Quadratic Programing makes pathway analysis easier since these large cycling fluxes undermine the minimization of the total fluxes term in the objective function (Equation 3-4), so they will be forced to a minimum by the optimization software and the flux distribution will be "cleaned" from these unrealistic fluxes. This is especially significant in large networks because these cycles are more recurrent.

In this study, we employed a curated version of the iPP668 model developed by Chung and collaborators (Chung *et al*., 2010), called iFS670 (Supplementary material S 3-2 and S3-7). In this updated version, we incorporated the arabitol biosynthesis pathway and the stoichiometric reactions for the production of three recombinant proteins (FAB fragment, Human Serum Albumin and Thaumatin). The arabitol synthesis pathway was included because it was a major compound in the culture supernatant of our experiments. Moreover, the reversibility of cytosolic reactions involving redox cofactors and mitochondrial symporters was checked according to Pereira *et al*. (Pereira *et al*., 2016) in order to obtain a more realistic flux distribution through the central metabolism. This was done because the initial flux distributions obtained with the un-modified iFS670 model presented the exact same problems as the iMM904 model of *Saccharomyces cerevi*siae on Pereira's work, suggesting that the central metabolism structure of the iPP668 model was based upon the aforementioned *S. cerevisiae* model. These problems were caused by: (i) the lack of a flux through the oxidative branch of the Pentose Phosphate Pathway; (ii) the presence of a flux of a cytosolic NAPDH dependent isocitrate dehydrogenase (which was the responsible of producing cytosolic NADPH); (iii) an unrealistic flux through mitochondrial symporters; and (iv) almost no mitochondrial formation of α-ketoglutarate. These model limitations are inconsistent with previous *P. pastoris* fluxomic studies in glucose-limited aerobic conditions (Baumann *et al*., 2010; Dragosits *et al*., 2009; Heyland *et al*., 2011).

FBA problems were solved using the Constraint-Based Reconstruction and Analysis (COBRA) toolbox (Becker *et al*., 2007; Hyduke *et al*., 2011), which employs the

programming library libSBML (Bornstein *et al.*, 2008) and the SBML toolbox (Keating *et al.*, 2006). Finally, we used Gurobi 6.0.2 as an optimization solver.

3.2.1.3 Dynamic block

The dynamic block consists of a set of ordinary differential equations (ODEs) that account for the volume change of the culture and the mass balances of biomass and the species considered by the model:

$$\frac{dV}{dt} = F(t) - SR \tag{3-5}$$

$$\frac{d(V \cdot X)}{dt} = \mu \cdot (V \cdot X) - SR \cdot X \tag{3-6}$$

$$\frac{d(V \cdot G)}{dt} = F(t) \cdot G_F - v_G \cdot MW_G \cdot (V \cdot X) \ - SR \cdot G \tag{3-7}$$

$$\frac{d(V \cdot P_k)}{dt} = v_{P_k} \cdot MW_{P_k} \cdot (V \cdot X) - SR \cdot P_k \tag{3-8}$$

Where V is volume [L], t is time [h], $F(t)$ is the feed function for the fed-batch phase in [L/h]. SR is a constant sampling rate [L/h] determined from each cultivation to emulate the remaining volume of the culture considering sampling, since this value is used for the calculation of the feeding profile during the feed phase. During the batch phase of the fed-batch cultures, we collected between 15 and 20% of the reactor volume in samples. For batch cultivations, $F(t)$ was eliminated from the mass balances. X is the biomass concentration [g/L], μ is the specific growth rate [h^{-1}] (obtained from equation 3-4), G is the extracellular concentration glucose [g/L], G_F is the feed's glucose concentration [g/L], P_K is the k-th extracellular product concentration in [g/L], v_{P_k} is the corresponding production rate [mmol/g_{DCW}·h] and MW accounts for the corresponding molecular weight [g/mmol].

The set of equations was solved in Matlab 2013a (Mathworks, USA) using the solvers ode113 and ode15s for batch and fed-batch cultures respectively.

3.2.1.4 Model parameters

The lower, upper and initial values of the parameters of the model used in all the calibrations are presented Table 3 - *1*. The lower and upper bounds of $v_{G,max}$, K_s, and m_{ATP} were chosen according to literature (Chung *et al*., 2010; van Urk *et al.*, 1989; Villadsen *et al*., 2011) while the rest of the bounds were selected to ensure that the algorithm had enough search space. To do this, the upper bounds of the rest of the parameters were set at higher values than the observed experimental rates, also taking into account reported values (Baumann *et al*., 2010; Dragosits *et al*., 2009; Heyland *et al*., 2011). In addition, initial estimated parameter values were chosen to attain a feasible simulation.

Table 3 - 1- Parameters of the model.

Symbol	**Name**	**Units**	**LB**	**Initial value**	**UB**
$v_{G,max}$	Maximum glucose uptake rate	$mmol/g_{DCW}h$	0	2.5	10
K_G	Half saturation constant for glucose uptake	g/L	0	10^{-4}	10^{-3}
$v_{EtOH,B}$	Ethanol minimum secretion rate (batch)	$mmol/g_{DCW}h$	0	0.5	3
$v_{Pyr,B}$	Pyruvate minimum secretion rate (batch)	$mmol/g_{DCW}h$	0	0.1	2
$v_{Arab,B}$	Arabitol minimum secretion rate (batch)	$mmol/g_{DCW}h$	0	0.2	2
$v_{Cit,B}$	Citrate minimum consumption rate (batch)	$mmol/g_{DCW}h$	0	0	2
$v_{EtOH,FB}$	Ethanol minimum consumption rate (fed-batch)	$mmol/g_{DCW}h$	0	0	2
$v_{Pyr,FB}$	Pyruvate minimum consumption rate (fed-batch)	$mmol/g_{DCW}h$	0	0	2
$v_{Arab,FB}$	Arabitol minimum consumption rate (fed-batch)	$mmol/g_{DCW}h$	0	0	2
$v_{Cit,FB}$	Citrate minimum consumption rate (fed-batch)	$mmol/g_{DCW}h$	0	0	2
α_B	Sub-optimal growth coefficient (batch)	$[-]$	0	0	10^{-3}
α_{FB}	Sub-optimal growth coefficient (fed-batch)	$[-]$	0	0	10^{-3}
m_{ATP}	Non-growth associated ATP	$mmol/g_{DCW}h$	0	2	10
T_{Fed}	Time when secondary metabolite consumption starts in fed-batch cultures	h	20	25	32

3.2.2 Model Calibration with experimental data

3.2.2.1 Strains

Four *P. pastoris* strains were employed in this study: a parental GS115 strain (Invitrogen) and three recombinant strains constructed according to the instructions of the manufacturers harboring respectively one, five and eight copies of the gene encoding for the sweet protein thaumatin. Even though the strains were transformed, thaumatin was not detected at concentrations higher than 100 μg/L in the cultivations. Therefore, due to its small contribution to the overall mass balance, thaumatin production was left out of the analysis and none of the parameters of the model were associated with it. Nevertheless, a mass balance for a recombinant protein can be easily added to the framework.

3.2.2.2 Experiments

The batch model was calibrated with aerobic glucose limited cultivations of the four strains available; each cultivation was performed twice. On the other hand, the fed-batch model was calibrated with data from three cultures of the strain with one copy the recombinant gene, under the same environmental conditions of the batch cultivations.

3.2.2.3 Cultivation Conditions

Each batch or fed-batch culture started from a 2 [mL] cryotube of the corresponding strain kept at -80 °C. A pre-culture was grown overnight at 30 °C in shake flasks with 50 [mL] of the inoculum medium. After reaching 1 OD_{600}, the whole broth was added to 450 [mL] of fresh medium to reach an initial volume of 500 [mL] in 1L bioreactors. Culture conditions were kept at 30 °C and pH = 6.0. Dissolved Oxygen was maintained above 40% saturation during all the cultivation period. Aerobiosis was achieved by a triple split-range control action, including agitation (200–800 [RPM]), air flow (0.25–1.0 [L/min]) and pure oxygen flow (0–1.0 [L/min]) (Cárcamo *et al*., 2014). pH was controlled using phosphoric acid 20% [v/v] and sodium hydroxide 20% [v/v]. The temperature was controlled with a mixture of hot and cold water, using the glass jacket of the reactors. Lastly, foam was controlled manually using silicone antifoam 10% [v/v]. Glucose starvation was detected when a sudden decrease of the CO_2 composition in the off-gas occurred, and it was confirmed each

time using Benedict's reagent. For fed-batch experiments, the feed F(t) was designed to track a variable growth rate for a predefined time. This feed can be calculated from the reactor's glucose and biomass mass balances, as detailed in the literature (Villadsen & Patil, 2007):

$$F(t) = \frac{\mu_{set(t)}}{G_F \cdot Y_{SX}} \cdot V_i X_i \cdot \exp\left(\int_{t_i}^{t} \mu_{set}(t)dt\right) \quad (3\text{-}9)$$

with G_F the glucose feed concentration [g/L], Y_{SX} the experimental glucose-biomass yield [g_{DCW}/g] calculated using the genome-scale model, t_i the time at which the feed started for a given cultivation [h], V_i and X_i the volume [L] and biomass [g/L] values at t_i, respectively, and $\mu_{SET}(t)$ is the time-dependent user-defined growth rate at which the fed-batch culture is grown. The latter was defined as follows:

$$\mu_{set(t)} = (\mu_{max} - \mu_{min}) \cdot e^{-Ct} + \mu_{min} \quad (3\text{-}10)$$

Where μ_{MAX} = 0.1 [1/h], μ_{MIN} = 0.07 [1/h] and C = 0.07 [1/h]. Therefore, $\mu_{SET}(t)$ decays exponentially from 0.1 to 0.07 [1/h], which has been found to increase (in contrast to constant growth rates in the feed phase) the final biomass concentration in fed-batch cultivations of *E. coli* and *S. cerevisiae* performed in our laboratory (Cárcamo, 2013).

3.2.2.4 Culture media

The culture media employed in these studies were based on Tolner *et al.* (Tolner *et al.* , 2006). Inoculum: Glucose 10 [g/L], $(NH_4)_2SO_4$ 1.8 [g/L], $MgSO_4 \cdot 7H2O$ 2.3 [g/L], K_2SO_4 2.9 [g/L], trace elements solution 0.8 [ml/L], histidine 0.08 [g/L], sodium hexametaphosphate 5 [g/L] and biotin 0.32 [mg/L]. Batch cultures: Glucose 50 [g/L], $(NH_4)_2SO_4$ 9 [g/L], $MgSO_4 \cdot 7H_2O$ 11.7 [g/L], K_2SO_4 14.7 [g/L], trace elements solution 4 [ml/L], histidine 0.4 [g/L], sodium hexametaphosphate 25.1 [g/L] and biotin 1.6 [mg/L] and sodium hydroxide NaOH 1 [g/L]. Feeding medium: Glucose 500 [g/L], $MgSO_4 \cdot 7H2O$ 9 [g/L], trace solution 12.5 [g/L], histidine 4 [g/L] and biotin 0.1 [g/L]. Sodium hydroxide was added to all the media until a pH of 6 was reached.

3.2.2.5 Analytical procedures

3.2.2.5.1 Sampling and biomass determination

Samples of ~6 mL were periodically collected (every 2-3 hours) from all fermentations. Biomass was measured by optical density (OD) at 600 nm using an UV-160 UV-visible spectrophotometer (Shimadzu, Japan). Biomass concentration was determined using the linear relationship: 1 OD_{600} = 0.72 [g/L] using the methodology from (Marx, Mecklenbr, Gasser, Sauer, & Mattanovich, 2009). Then, samples were centrifuged at 10.000 rpm for 3 min and the supernatant stored at -80°C for further analysis.

3.2.2.5.2 Extracellular metabolite concentration analyses

Glucose, ethanol, arabitol, citrate and pyruvate extracellular concentrations were quantified in duplicate by High-Performance Liquid Chromatography (HPLC), as detailed in Sánchez *et al*. (Sánchez *et al.*, 2014), with the exception of the working temperature of the Anion-Exchange Column (Bio-Rad, USA), which was lowered from 55°C to 35°C for better resolution.

3.2.2.6 Objective Function

For model calibration, we minimized the sum of square errors between the experimental data and the simulation output by searching the parameter space, with the enhanced scatter search algorithm (eSS) (Egea & Balsa-Canto, 2009), which has been successfully used to solve complex bioprocess optimization problems (Balsa-Canto *et al.*, 2007; Sacher *et al.*, 2011; Sriram *et al.*, 2012). The objective function J used in the minimization was normalized by the maximum corresponding measured variable to give all data a similar weight:

$$J = \min_{\theta} \sum_{i=1}^{m} \sum_{j=1}^{n} \left(\frac{X_{ij}^{mod} - X_{ij}^{exp}}{\max_{j} \left(X_{ij}^{exp} \right)} \right)^2 \quad (3\text{-}11)$$

With θ representing the parameter space, m the number of measured variables, n the number of measurements per variable, X_{ij}^{mod} the dFBA output of variable i and

measurement j, X_{ij}^{exp} the corresponding experimental value and $\max_j(X_{ij}^{exp})$ the maximum value measured for variable i.

3.2.3 Pre/Post regression analysis

Once the initial calibration of the model was completed, statistical tests were performed in order to determine if the initial model formulation had sensitivity, identifiability or significance problems (Jaqaman & Danuser, 2006).

Sensitivity corresponds to the impact that model parameters have on the state variables or process output. The relative sensitivity of parameter k on the state variable i (g_{ik}) was calculated according to the following formula

$$g_{ik}(t, \theta_k) = \frac{\theta_k}{X_i(t)} \cdot \frac{dX_i(t)}{d\theta_k} \quad (3\text{-}12)$$

Where $X_i(t)$ is the *i*th state variable in time t and θ_k is the *k*th parameter. With all g_{ik} values, we formed a sensitivity matrix g(t) for each experimental time, in which the *k*th column denotes the sensitivity of the kth parameter on the state variables. These matrices were averaged to obtain a single normalized score of the sensitivity of parameter k on the state variable i during the cultivation. Furthermore, if the score of each variable was under 0.01 for a given parameter, this parameter was considered insensitive and a candidate to be fixed (or left out of the adjustable parameter set) in the reparametrization stage.

Identifiability refers to the possibility of unambiguously determining the parameter values by fitting a model to experimental data. If parameter identifiability is not properly assessed, misleading parameter values can be obtained after model calibration. To calculate identifiability, we determined the correlation between the columns of the sensitivity matrix using the *corrcoef* function from Matlab, which yielded a correlation coefficient matrix (C). A pair of parameters j and k was considered to be correlated (therefore not-identifiable) if the absolute value of the number at the (j, k) position in the correlation coefficient matrix was higher than 0.95 ($(|C_{jk}| \geq 0.95)$).

To determine parameter significance, we started by calculating the Fisher Information Matrix (FIM) (Petersen *et al.*, 2001)

$$FIM = \sum_{j=1}^{n} g_j^T Q_j g_j \tag{3-13}$$

Here, g_j is the sensitivity matrix for measurement j, n is the number of samples, and Q_j is a weighting matrix given by the inverse of the measurement error covariance matrix assuming white and uncorrelated noise. Hence, the variances for each estimated parameter were calculated as in (Landaw & DiStefano 3rd, 1984; Petersen *et al.*, 2001)

$$\sigma_k^2 = FIM_{kk}^{-1} \tag{14}$$

which was used to determine the confidence interval (CI) with 5% significance for the kth parameter as follows:

$$CI_k = \left[\hat{\theta}_k \pm 1.96\sigma_k\right] \tag{15}$$

Here, $\hat{\theta}_k$ is the estimated value of the corresponding parameter. Finally, coefficients of confidence (CC) were calculated as follows:

$$CC_k = \frac{\Delta(CI_k)}{\hat{\theta}_k} = \frac{2 \cdot 1.96\sigma}{\hat{\theta}_k} \tag{3-16}$$

$\Delta(CI_k)$ is the CI's length. A parameter was not significant if the confidence interval contained zero, i.e. if the absolute value of the CC was equal or larger than 2.

3.2.4 Reparametrization

A reparametrization procedure called HIPPO (Sánchez *et al.*, 2014) (Heuristic Iterative Procedure for Parameter Optimization, http://www.systemsbiology.cl/tools/) was applied to overcome parametric statistical limitations in the model.

First, HIPPO performed sensitivity and identifiability tests on the initial calibration results for each dataset. Then, model parameters were fixed one by one until the non-fixed subset presented none of the statistical limitations. Finally, significance was determined for the remaining parameter set, also called the reduced model structure. If all the remaining parameters were significantly different from zero, the resulting structure is considered to be an *a priori* robust candidate for cross calibration with the available data.

3.2.5 Cross Calibration of robust structure candidates derived from the reparametrization stage using the available datasets.

After reparametrization of the model derived from each dataset, a potentially robust structure was generated. This structure was recalibrated with the rest of the datasets to assess its robustness. It is worthy to note that the parameters left out of the calibration were either fixed according to values reported in literature, assumed to be zero or fixed at the mean value achieved in the calibrations. This was done to avoid assuming a minimum production of compounds in batch cultivations and to ensure model convergence for parameters that had no reported values in literature (Table 3 - *2*). For example, fixing feed phase consumption rates at zero does not allow consumption of batch by-products and yielded poor fed-batch fittings (data not shown).

Table 3 - 2 Values at which problematic parameters were fixed in the cross-calibration stage.

Parameter	Fixation Value	Units	Reference
$v_{G,max}$	6	$mmol/g_{DCW}h$	(van Urk *et al*., 1989)
K_G	0.0027	g/L	(van Urk *et al*., 1989)
$v_{EtOH,B}$	0	$mmol/g_{DCW}h$	-
$v_{Pyr,B}$	0	$mmol/g_{DCW}h$	-
$v_{Arab,B}$	0	$mmol/g_{DCW}h$	-
$v_{Cit,B}$	0	$mmol/g_{DCW}h$	-
$v_{EtOH,FB}$	1.21	$mmol/g_{DCW}h$	*
$v_{Pyr,FB}$	0.14	$mmol/g_{DCW}h$	*
$v_{Arab,FB}$	0.15	$mmol/g_{DCW}h$	*
$v_{Cit,FB}$	0.008	$mmol/g_{DCW}h$	*
α_B	0	$[-]$	(Morales *et al*., 2014)
α_{FB}	0	$[-]$	(Morales *et al*., 2014)
m_{ATP}	2.18	$mmol/g_{DCW}h$	(Chung *et al*., 2010)
T_{Fed}	22	h	*

Parameters marked with '-' in the reference column indicate that no *a priori* value was assumed for that particular parameter, which is the case for the batch minimum secretion rates. '*' means that the value of a particular parameter was fixed at the mean value achieved in the calibrations, because no information about them could be found in the literature.

The reduced modeling structures were evaluated according to four parameters:

I. **Relative difference between calibration objective functions (J_{DIFF}):**

$$J_{DIFF} = \frac{1}{n} \cdot \sum_{i=1}^{n} \frac{J_{i,Reduced} - J_{i,Original}}{J_{i,Original}} \quad (3\text{-}17)$$

Where n corresponds to the number of cultures of each type, $J_{i,Original}$ is the calibration objective function (Equation 3-11) achieved for dataset i using the original model structure and $J_{i,Reduced}$ is the calibration objective function achieved in dataset i using a reduced, *a priori* robust, modeling structure.

II. **Percentage of Significance issues**; refers to the number of times a parameter is found to be non-significant out of the total of significance determinations performed for a structure. For instance, if a model structure had 6 parameters and 8 datasets were used to calibrate it, a total of 48 significance determinations were performed for that particular model.

III. **Percentage of Sensitivity issues;** refers to the number of times one of the estimated parameters shows low or no impact over state variables (average relative sensitivity ≤ 0.01) out of the total sensitivity determinations performed.

IV. **Percentage of Identifiability issues**; corresponds to the number of times a pair of parameters presents a strong correlation (≥ 0.95), out of the total parameter pairs of a modeling structure. If p is the number of parameters of the model and n is the number of datasets used for its calibration, the total of parameter pairs for which identifiability was determined is:

$$Total\ pairs = \frac{p \cdot (p-1)}{2} \cdot n \quad (3\text{-}18)$$

Finally, the modeling structure that presented the lowest J_{DIFF} and fewest statistical limitations was used as a robust structure candidate for the corresponding type of culture.

3.2.6 Robustness check of the chosen modeling structure

Once a candidate for a robust structure was determined for the batch and fed-batch configurations, we tested its robustness (absence of parametric problems) by calibrating it with new experimental data. For the batch model, we employed fermentation data from *P. pastoris* GS115 strain grown with 40 [g/L] of glucose as carbon source at T° = 25°C and pH = 6. The robustness of the fed-batch model was assessed with a glucose-limited cultivation consisting of a 60 [g/L] glucose batch phase and an exponential feed using 500 [g/L] of glucose. The medium was added in the feeding phase in order to achieve an exponentially decreasing growth rate from 0.1 to 0.07 [1/h].

3.2.7 Model validation

Finally, the predicting capability of the model was evaluated for conditions similar to the ones used in the initial calibrations (training set).

The robust batch model was first calibrated with the two cultivations of the strain harboring one copy of the thaumatin gene, obtaining a characteristic parameter set for that strain. Then, these parameters were used to predict the course of a different batch cultivation performed in the same conditions (30°C and pH 6).

This procedure was also applied for the fed-batch model. Here, the bioreactor dynamics was simulated using the parameters obtained in the best calibration within the training dataset (the one in which the calibration objective function was minimal compared to the rest of the calibrations) using the robust modeling structure obtained previously. This prediction was compared with experimental data of different fed-batch cultivation.

3.2.8 Goodness of fit

For both the robustness check and validation datasets, the goodness of fit was determined by two scores: the mean normalized error (MNE) and the Anderson-Darling test (Stephens, 1974). The MNE quantifies the difference between model simulations and experimental data; the closer the difference is to zero, the better the fit. In addition, the sign of MNE shows whether the model over (+) or underestimates (-) the observed data (equation 3-19).

$$MNE_i = \frac{\sum_{j=1}^{n}\left(X_{ij}^{mod} - X_{ij}^{exp}\right)}{n \cdot \max_j\left(X_{ij}^{exp}\right)} \quad (3\text{-}19)$$

with n the number of time points measured for variable i.

The Anderson-Darling test was used to verify if the residuals between simulations and experimental data $\left(X_{ij}^{mod} - X_{ij}^{exp}\right)$ were normally distributed. If they were, the differences between them can be attributed to measurement noise and not to model inadequacy. The failure of this test by one of the model's state variables (p-value < 0.05) indicates that a different mathematical relation than the one used in the model may underlie its dynamics. Therefore, the results of this test may be used to confirm or update the kinetic expressions associated with the consumption and production of compounds.

3.2.9 Simulation

3.2.9.1 Analysis of the metabolic flux distribution during key stages of a dynamic cultivation

After the calibration of the fed-batch model with the dataset used for checking its robustness, we evaluated the central metabolic flux distributions at three different stages of the cultivation: exponential growth during the batch phase (~20 hours), ethanol and arabitol consumption during glucose starvation phase (~27.5 hours) and controlled growth during the feeding phase (~45 hours).

3.2.9.2 Discovery of beneficial knock-out targets for the overproduction of recombinant Human Serum Albumin (HSA)

To show the potential applications of the model, gene targets for the overproduction of the recombinant Human Serum Albumin (HSA) were determined by simulating the growth and protein secretion of single knock-out strains of *P. pastoris* in batch cultivations. To do this, we included in the Metabolic Block a second quadratic programing problem consisting in the Minimization of Metabolic Adjustment (MOMA) algorithm (Segrè *et al.,*2002), which states that, after a genetic perturbation, the cell will attempt to redistribute its metabolic fluxes as similar as possible to the parental strain. Mathematically, equation

3-4 of the metabolic block is employed in order to obtain the parental flux distribution v_0 at a given instant.

$$Min\ \alpha \cdot \sum_{i=1}^{n} v_{0,i}^2 - (1-\alpha) \cdot \mu_0$$

$$s.t. \qquad (3\text{-}20)$$

$$S \cdot v_0 = 0$$

$$lb_{0,i} \leq v_{0,i} \leq ub_{0,i} \qquad i = 1 \ldots n$$

Then, the k reactions associated with gene j are blocked:

$$lb_{l,j} = ub_{l,j} = 0 \qquad l = 1 \ldots k \qquad (3\text{-}21)$$

Finally, the MOMA algorithm was applied using the flux distribution of the parental strain v_0 to calculate the knockout distribution v_{KO} as the Euclidean distance between them, considering that the actual model has the corresponding deletion.

MOMA:

$$Min\ \left(v_0 - v_{KO,j}\right)^2$$

$$s.t. \qquad (3\text{-}22)$$

$$S \cdot v_{KO,j} = 0$$

$$lb_i \leq v_{i,KO,j} \leq ub_i \qquad i = 1 \ldots n$$

The hypothetical parental strain was characterized using the parameters obtained above plus the growth rate dependent specific HSA productivity (q_P) of *P. pastoris* strain SMD1168H grown on glucose, as reported by Rebnegger *et al*. (Rebnegger *et al*., 2014), (Figure *3-2*). In each iteration of the model, the minimum HSA production was fixed according to this relationship, which was fitted with a third degree polynomial. Other kinetic expressions could be employed to represent the q_P vs μ relationship, depending on the strain and protein being produced (Maurer *et al*., 2006).

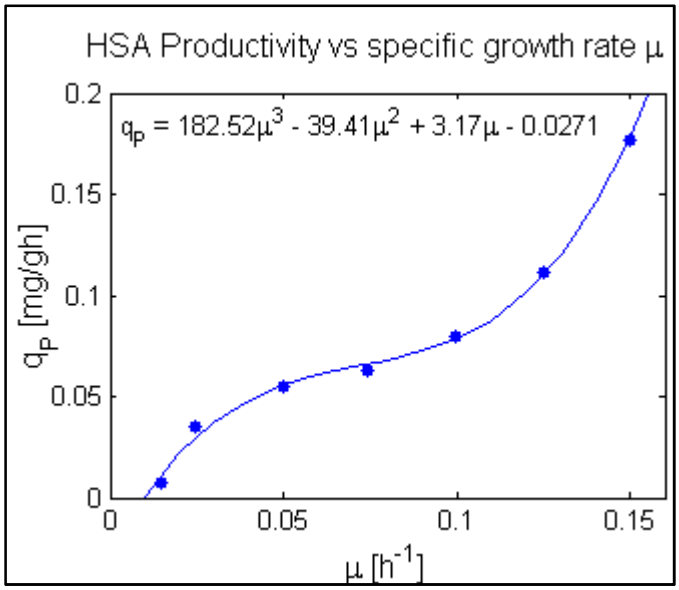


Figure 3-2 - Relation between Human Serum Albumin specific production rate (q_P) and growth rate (μ) in glucose limited chemostats, taken from Rebnegger *et al*. (Rebnegger *et al*., 2014). This relation was included to simulate the specific protein productivity for a given growth rate, allowing the assessment of the impact of different feeding profiles on process productivity.

We simulated one batch cultivation for each gene in the model and compared their final protein and biomass concentrations with those of the parental strain. The candidates that reached a higher HSA concentration than the parental strain were manually analyzed and some of them were proposed as candidates to improve HSA production. It is important to mention that we used a set of parameters derived in this study to characterize the growth kinetics of the HSA producing strain used in the simulations. Therefore, the predictions derived from this work should be assessed carefully and considered only as an example of the applicability of our modeling framework.

3.2.9.3 Evaluation of different feeding policies *in silico* to improve recombinant protein production considering specific information about the strain and process setup

Simulations were run using the parameters obtained in the calibration used for intracellular flux analysis and adding the q_P vs μ relation for HSA biosynthesis in the mass balances. The process limitations (based on our setup) were a maximum reactor volume of 1 L, and a maximum oxygen transfer rate of 10.9 [g/L·h]. If any of these limits were violated by

either the feeding rate of medium or the oxygen uptake rate (extracted from the model), the integration stopped.

We assessed 13 exponential feeding policies. Five of them maintained a constant growth rate during the feeding phase and the rest considered a decreasing growth rate throughout the culture (Supplementary material S 3-3). After the simulation, we ranked the strategies according to the volumetric productivity of recombinant HSA and chose the best one as a cultivation strategy that could potentially improve bioreactor performance.

3.3 RESULTS AND DISCUSSION

The batch and fed-batch models were developed in four steps: (i) determination of initial parametric problems, (ii) reparametrization and cross calibration, (iii) robustness evaluation and (iv) validation of predictive potential under the studied conditions.

Once the models were developed, three applications were proposed to improve recombinant |protein production using Human Serum Albumin as a case study.

3.3.1 Initial Parametric Problems

3.3.1.1 Batch model

The initial structure of the batch model comprised eight parameters (Table 3 - *3*). The model was able to successfully accommodate different cellular dynamics from eight glucose-limited aerobic cultivations. In these calibrations, several statistical parametric limitations were found (Supplementary material S 3-4). m_{ATP} was the parameter that presented the strongest correlation with other parameters, such as maximum specific glucose uptake rate ($v_{G,Max}$), ethanol and arabitol specific secretion rates ($v_{EtOH,B}$ and $v_{Arab,B}$), and with the sub-optimal growth coefficient (α_B). This might result from the fact that a change in m_{ATP} directly impacts the ATP-producing pathways in the metabolic model, affecting the biomass and product yields, which are also influenced by other parameters of the model. In addition, the glucose uptake saturation constant K_G was the only parameter with frequent sensitivity and significance problems, making it a potential candidate to be left out of the adjustable parameter set.

3.3.1.2 Fed batch model

Data from three aerobic, glucose-limited fed-batch cultivations was successfully calibrated with the initial model of fourteen parameters. As in the batch model, several statistical parametric limitations arose (Supplementary material S 3-4). The most frequent correlation (in two out of the three calibrations) was between $v_{G,Max}$ and the $v_{EtOH,B}$ during the batch phase. Also, $v_{EtOH,B}$ and $v_{Arab,FB}$ showed 5 and 6 strong correlations with other parameters of the model, respectively.

Finally, the citrate minimum secretion rate during the fed-batch phase and the suboptimal growth during the feeding phase (α_{FB}) were the parameters that presented more sensitivity and significance limitations.

3.3.2 Reparametrization and Cross Calibration

After model calibration and the subsequent determination of the parametric problems for each dataset, the non-relevant parameters were fixed (left out of the adjustable set) using HIPPO (Sánchez *et al*., 2014) to achieve robust modeling structures.

3.3.2.1 Batch model

The reduced batch models derived from the initial calibrations (Table 3 - *3*) were recalibrated with the available data (eight batch cultivations) to determine if they could reproduce *P. pastoris* behavior appropriately. The persistence of parametric problems in the reduced models was compared to the original model.

Table 3 - 3 - Potential Robust Structures Tested in the Cross-Calibration Stage for the batch model.

Structure	Parameters included
Original	$v_{G,Max}, K_G\ v_{EtOH,B}, v_{Pyr,B}, v_{Arab,B}, v_{Cit,B}, m_{ATP}\ and\ \alpha_B$
1	$v_{G,Max}, v_{EtOH,B}, v_{Pyr,B}, v_{Arab,B}, v_{Cit,B}\ and\ \alpha_B$
2	$V_{G,Max}, v_{Cit,B}\ and\ \alpha_B$
3	$K_G, v_{EtOH,B}, v_{Pyr,B}, v_{Arab,B}\ and\ v_{Cit,B}$
4	$v_{EtOH,B}\ and\ v_{Cit,B}$
5	$v_{G,Max}, v_{Pyr,B}, v_{Arab,B}$
6	$v_{G,Max}, v_{EtOH,B}, v_{Pyr,B}, v_{Arab,B}, v_{Cit,B}$
7	$v_{G,Max}, K_G, v_{EtOH,B}, v_{Pyr,B}, v_{Cit,B}$
8	$K_G, v_{Pyr,B}, v_{Arab,B}, \alpha_B\ and\ m_{ATP}$

Each one of these structures was derived using HIPPO after model calibration using each dataset

Structures 1 and 6 were the only parameter sets whose fitting capabilities were similar to the original eight parameters model (Table 3 - 4), showing the importance of including the specific uptake and production rates of the compounds considered in the model. On the contrary, m_{ATP} and K_G were left out of these structures because of the frequent identifiability and sensitivity-associated problems.

Structure 6 lacks the sub-optimal growth parameter α_B, which forces the solution of a linear programming (LP) problem of specific growth rate maximization in the metabolic block. This is because this parameter was assumed to be zero if it was left out of the adjustable parameter set (Table 3 - *2*), which eliminates the total flux minimization term from the objective function. This structure showed a significant increase in significance and sensitivity compared to the original model; however, identifiability was a major problem (Table 3 - 4). Probably, the multiple solutions associated with an underdetermined LP problem may hamper the possibility to unambiguously infer parameter values from the data.

Therefore, due to the recurrent identifiability issues found in Structure 6, it was preferable to apply Structure 1 to fit a different dataset to check its robustness in aerobic, glucose-limited batch cultures of *P. pastoris*.

Table 3 - 4 – Batch Cross Calibration summary

Structure	N° parameters	J_{DIFF}	Significance Issues	Sensitivity Issues	Identifiability Issues
Original	**8**	**0**	**23.6**	**16.7**	**17.4**
1	6	-0.10	22.9	18.8	15.0
2	3	2.77	29.1	8	25
3	5	0.18	90.0	23	68
4	2	0.56	62.5	13	0
5	3	4.10	29.2	0	54
6	5	0.18	0	2.5	61.3
7	5	2.93	18.8	20.8	60.0
8	5	2.82	22.5	25.0	33.8

Structures that reduced de frequency of parametric problems with respect to the original model are highlighted.

3.3.2.2 Fed-batch model

In the fed-batch model, three potentially robust model structures were found after its calibration with three datasets (Table 3 - 5).

Table 3 - 5 – Potential robust structures for a fed-batch model

Structure	Parameters included
Original	$v_{G,Max},\ K_G, v_{EtOH,B}, v_{Pyr,B}, v_{Arab,B}, v_{Cit,B},$ $v_{EtOH,FB}, v_{Pyr,FB}, v_{Arab,FB}, v_{Cit,FB}, \alpha_B, \alpha_{FB}, m_{ATP}, T_{Cons}$
1	$v_{G,Max},\ K_G, v_{Pyr,B}, v_{Cit,B}, v_{EtOH,FB}, v_{Pyr,FB}, v_{Arab,FB}, v_{Cit,FB}, \alpha_B, m_{ATP}, T_{Cons}$
2	$K_G, v_{EtOH,B}, v_{Pyr,B}, v_{Arab,B}, v_{Cit,B}, v_{EtOH,FB}, v_{Pyr,FB}, \alpha_B, m_{ATP}$
3	$v_{G,Max},\ K_G, v_{Pyr,B}, v_{Arab,B}, v_{Cit,B}, v_{Pyr,FB}\ \alpha_B, \alpha_{FB}, m_{ATP}, T_{Cons}$

All the candidate structures considered the following parameters: K_G, $v_{Pyr,B}$, $v_{Cit,B}$, α_B, $v_{Pyr,FB}$ and m_{ATP}. Contrary to the batch model, K_G plays an important role in this cultivation system. This parameter, which usually lies in the micromolar range (Boles & Hollenberg, 1997), can directly modulate substrate uptake under glucose-limited conditions. Therefore, when glucose concentration is close to zero (like in the feeding phase), slight variations in the value of K_G can change glucose uptake significantly, which has a direct impact in the specific growth rate. Also, m_{ATP} appears to have a relevant role since it might act as an energy sink when the glucose from the batch phase is depleted. Here, secondary product consumption occurs with a slower or null biomass formation prior to the addition of glucose (Figure 3 in Supplementary material S 3-4). This indicates that the substrates were consumed to maintain basic cellular functions to survive, instead of being used for cell division.

The three reduced structures improved the initial fittings (lower J_{DIFF}) and reduced the frequency of fitting problems observed in the initial model of 14 parameters (Table 3 - 6). Among these, Structure 3 performed the best in the cross-calibration in terms of fitting capability compared to the original model. On average, this structure improved in initial calibrations by 25%. It is worthy to note that, even though Structure 3 did not include the minimum production rate of ethanol during the batch phase, it could adequately reproduce the profiles of this compound by adjusting the objective function and the maintenance ATP. Finally, we chose to apply Structure 3 to fit new fed-batch data to check its robustness for modeling glucose-limited aerobic fed-batch cultivations of *P. pastoris*.

Table 3 - 6 - Summary of the Cross Calibration of the fed-batch datasets.

Structure	**N° parameters**	**J_{DIFF}**	**Significance Issues (%)**	**Sensitivity Issues (%)**	**Identifiability Issues (%)**
Original	**16**	**0**	**33**	**18.8**	**3.9**
1	11	-2%	27.2	2.6	1.8
2	9	-15%	25.1	7.4	0.9
3	10	-25%	26.7	3.7	0.9

Structures that reduced de frequency of parametric problems with respect to the original model are highlighted.

3.3.3 Robustness Check

3.3.3.1 Batch model

On this new dataset, Structure 1 showed a good fit to the data and did not yield identifiability nor significance problems. However, $v_{G,Max}$ had no impact on the state variables. Therefore, after the initial calibration (data not shown), we fixed this parameter at 6 [mmol/g_{DCW}h] (van Urk *et al*., 1989). Figure *3-3* illustrates the model fit and Table 3 - 7 presents the parameter values with their 95% confidence intervals achieved in the second calibration, which also had no identifiability, significance or sensitivity limitations. This calibration also yielded mean normalized errors close to zero and normally distributed residuals for all the state variables except for glucose (Supplementary material S 3-5).

Despite the sensitivity problem associated with $v_{G,Max}$ for this particular dataset, we included this parameter in the proposed robust modeling structure. This is because for some calibrations, e.g. the batch cultivations of strains harboring 8 copies of the thaumatin gene, the state variables were very sensitive to this parameter (average sensitivity > 0.7, recall that the sensitivity threshold is 0.01); hence, it should be included to achieve a close fit to the data. Therefore, if this parameter is found insensitive in future calibrations, it could be easily fixed at reported values.

We achieved a robust modeling structure for glucose-limited, aerobic batch cultivations of *Pichia pastoris*, composed of six parameters that estimate specific consumption and production rates of all the species involved in the mass balances. The modeling structure also allows us to determine the specific growth rate by solving a bi-objective optimization problem, which reduces the identifiability issues arising between parameters (comparison between candidate batch model robust structures 1 and 6).

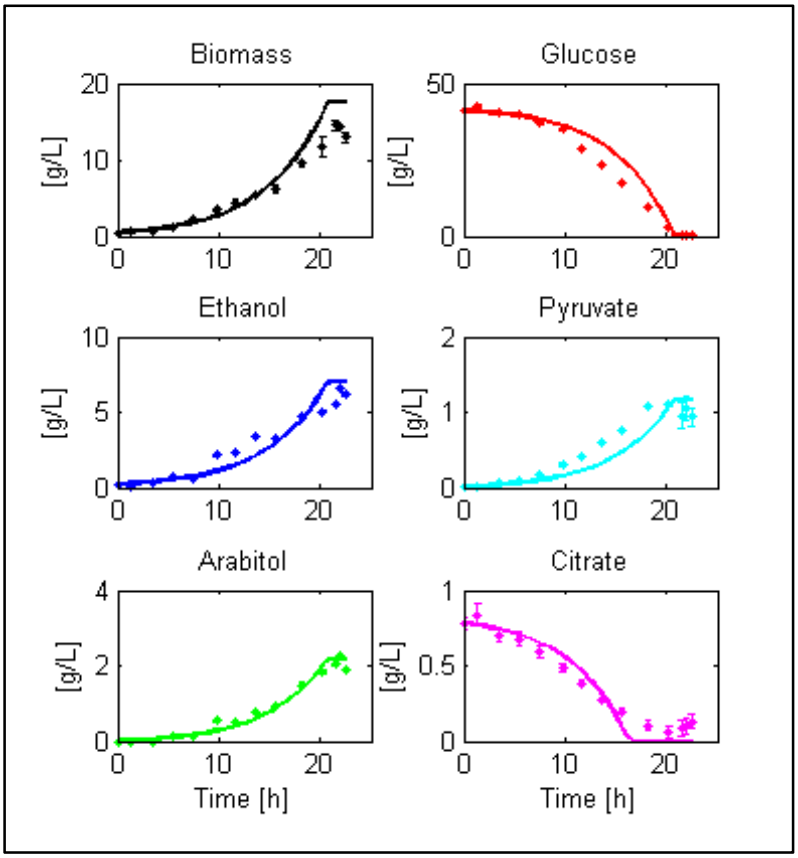


Figure 3-3 – Robustness check of Structure 1 as modeling framework for aerobic, glucose-limited batch cultures of *Pichia pastoris*. The figure shows the capacity of the reduced model structure to be calibrated with new data despite having fewer parameters than the original model structure (6 instead of 8 parameters). Points with whiskers represent experimental data and continuous lines correspond to the model approximation

Table 3 - 7 - Parameter values achieved in the validation of the batch model structure.

Parameter	Value	Units
$v_{G,Max}$	6	mmol/g_{DCW}·h
$v_{EtOH,B}$	1.47 ± 0.07	mmol/g_{DCW}·h
$v_{Pyr,B}$	0.13 ± 0.05	mmol/g_{DCW}·h
$v_{Arab,B}$	0.14 ± 0.06	mmol/g_{DCW}·h
$v_{Cit,B}$	0.09 ± 0.04	mmol/g_{DCW}·h
α_B	4.1 ± 0.9 ·10^{-4}	[-]

Values of the parameters are presented together with their 95% confidence intervals. In this calibration, $v_{G,Max}$ was fixed at a known value to avoid sensitivity issues. Finally, the calibration yielded no parametric problems.

3.3.3.2 Fed-batch model

Structure 3 shows a good fit to new experimental fed-batch data (Figure *3-4*) and did not yield identifiability or significance problems (Table 3 - 8 and Supplementary material S 3-5). The profile of some of the state variables still depends on the fixed values assigned. For example, arabitol was consumed at a slower rate than the profile observed in the experiment because the parameter representing this consumption ($v_{Arab,FB}$) was fixed as the mean of the training datasets (not included in the adjustable parameter set). Thus, the model assumed a faster consumption rate than observed in the cultivations. Also, pyruvate was found at such low concentrations that the parameters associated to its production ($v_{Pyr,B}$ and $v_{Pyr,FB}$) were ignored in this analysis.

Table 3 - 8 - Parameter values achieved in the calibration to check the robustness of the fed-batch model. The confidence interval on the time where the consumption of secondary metabolites started T_{CONS}, could not be determined due to the stiffness of the solution caused by a sudden consumption of arabitol and ethanol.

Parameter	Value	Units
$\boldsymbol{v_{MAX}}$	2.09 ± 0.46	mmol/g_{DCW}·h
$\boldsymbol{K_S}$	$5.55 \cdot 10^{-2} \pm 0.0000004 \cdot 10^{-2}$	g/L
$\boldsymbol{v_{Pyr,B}}$	0	mmol/g_{DCW}·h
$\boldsymbol{v_{Arab,B}}$	0.42 ± 0.17	mmol/g_{DCW}·h
$\boldsymbol{v_{Cit,B}}$	0.04 ± 0.00	mmol/g_{DCW}·h
$\boldsymbol{v_{Pyr,FB}}$	0	mmol/g_{DCW}·h
$\boldsymbol{\alpha_B}$	$2.6 \cdot 10^{-4} \pm 0.4 \cdot 10^{-4}$	[-]
$\boldsymbol{\alpha_{FB}}$	$2.455 \cdot 10^{-5} \pm 0.003 \cdot 10^{-5}$	[-]
$\boldsymbol{m_{ATP}}$	7.0 ± 1.4	mmol/g_{DCW}·h
$\boldsymbol{T_{Cons}}$	25.73	H

The chosen model structure showed a strong fitting capacity and a limited occurrence of parametric identifiability, sensitivity and significance problems. Therefore, we selected it as the most robust model structure for fed-batch cultivations of *P. pastoris*.

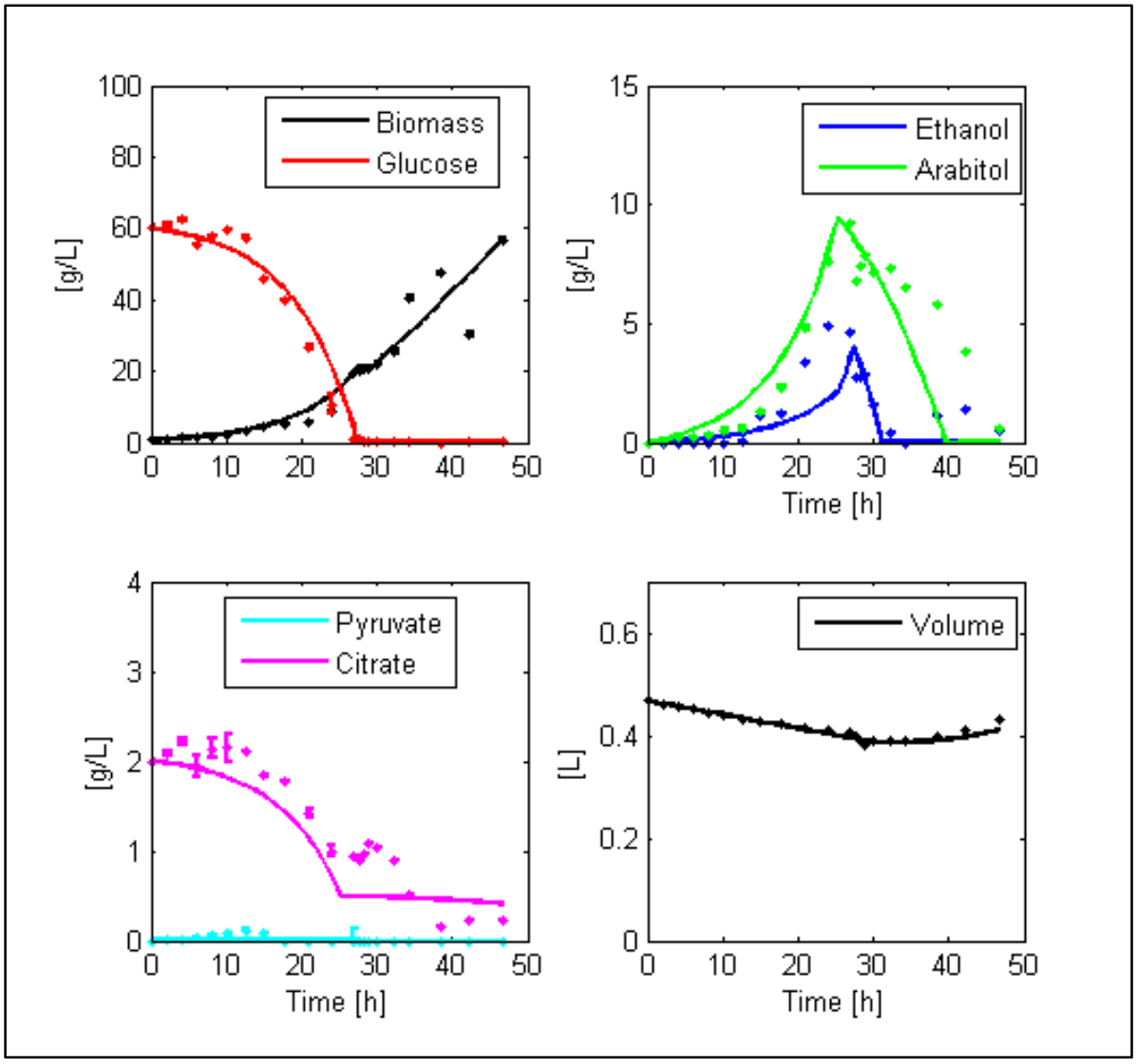


Figure 3-4 – Robustness check of Structure 3 as a modeling framework of aerobic glucose-limited fed-batch cultures of *Pichia pastoris*. The figure shows the capacity of the reduced model structure to be calibrated with new data, despite having fewer parameters than the original model structure (10 instead of 14 parameters). Points with whiskers represent experimental data and continuous lines correspond to the model approximation.

3.3.4 Model Validation

3.3.4.1 Batch model

The parameters found for the strain harboring one copy of the thaumatin gene were used to predict the dynamics of a different batch cultivation using the same strain (Figure *3-5*). Biomass and glucose profiles were correctly predicted by the model (MNEs close to zero and p-values of the Anderson-Darling test > 0.05, see Supplementary material S 3-6). Ethanol, pyruvate, citrate and arabitol dynamics also showed an overall concordance with the data, however the simulated profiles overestimated their final concentrations (see associated MNEs in Supplementary material S 3-7). These differences occurred probably because in the training datasets the initial concentration of glucose was higher than the one used in the validation experiment (~60 g/L vs. ~40 g/L), which might have increased the formation of secondary products (Cheng *et al.*, 2014). Therefore, future versions of the model may consider more elaborate kinetic expressions for the secretion of secondary products in order to accurately predict their formation in different circumstances.

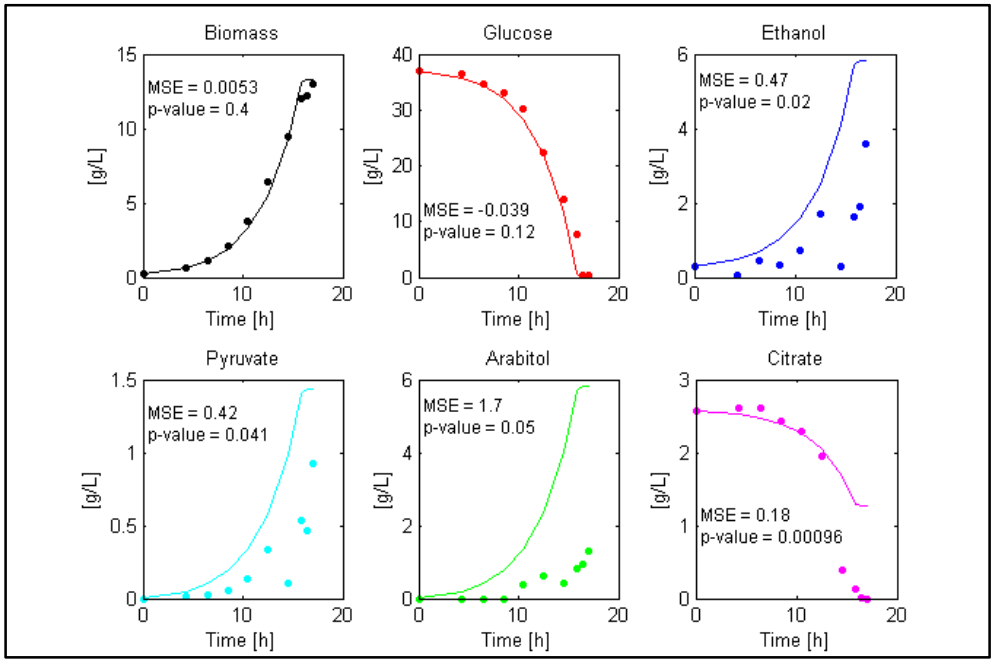


Figure 3-5 - Batch model preliminary validation. This figure shows how well the model predicts the course of a batch cultivation. To do this, we used the derived robust model structure to determine the characteristic parameters of a recombinant strain. Then, we simulated a batch culture (continuous line) and compared it with the experimental data (filled circles).

3.3.4.2 Fed-batch model

The prediction of biomass, glucose, ethanol and arabitol concentrations during the culture agreed with experimental data, whereas pyruvate and citrate dynamics were inaccurate (Figure *3-6*). Specifically, the simulation predicted that pyruvate was generated during the batch phase but experimental data did not show pyruvate production. In the experimental culture we saw that there was no generation of citrate in the feed phase, contrary to what the simulated predicted. These differences arose because in the culture from where the parameters were derived (Fed-batch culture 1, see Supplementary material S 3-4), pyruvate formation occurred in the batch phase and citrate was formed during the feed phase; therefore, the model assumed that these compounds were generated in the respective phase of the culture. Nevertheless, for the major compounds found in the culture, the model had a low mean normalized error.

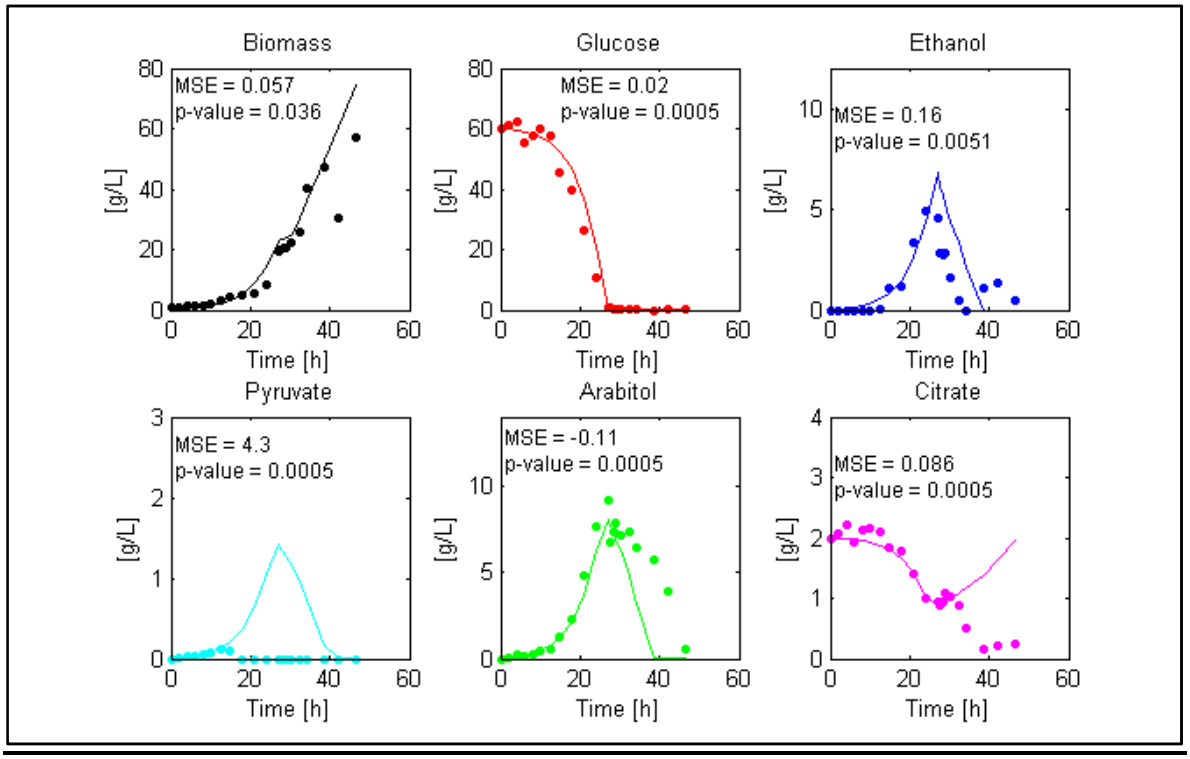


Figure 3-6 - Fed-batch model validation. This figure shows how well the model predicts the course of a fed-batch cultivation. To do this, we used the derived robust model structure to determine the characteristic parameters of a recombinant strain. Then, we simulated a fed-batch culture (continuous line) and compared it with the experimental data (filled circles).

3.3.5 Potential applications of the model

3.3.5.1 Analysis of the metabolic flux distribution at different stages of a dynamic cultivation.

Once we confirmed the robustness of the fed-batch model, we analyzed the redistribution of central carbon metabolic fluxes at three different stages of the cultivation (Figure *3-7*), i.e. exponential growth during the batch phase (~20 h, μ = 0.12 h^{-1}); co–consumption of arabitol and ethanol during the glucose starvation phase (~27.5 h, μ = 0.02 h^{-1}); and controlled exponential growth during the feeding phase (~45 h, μ = 0.06 h^{-1}) (Figure *3-7*). During exponential growth in the batch phase, the carbon reaching the glucose-6-phosphate node is split between carbohydrate production (11%), glycolysis (63%) and the oxidative branch of the PPP (24%). Furthermore, the latter is the main source of cytosolic NADPH. Cytosolic ATP is formed by the activity of the ATP synthase and substrate-level phosphorylation (glycolysis and synthesis of arabitol and ethanol) (data not shown). In the iPP618 model, which is the basis of the iFS670, cytosolic NADPH was produced by a NADP dependent isocitrate dehydrogenase, and no flux appeared through the oxidative branch of the PPP. Using the proposals from Pereira *et al*. (Pereira *et al*., 2016), the flux through this pathway was restored and overall agreement in directionality to fluxomic studies performed in similar conditions was achieved (Supplementary material S 3-2).

During the starvation phase, ethanol and arabitol are co-consumed with limited formation of biomass (μ=0.02 h^{-1}). As indicated by the negative fluxes, both compounds are directed towards the TCA cycle in order to synthesize the necessary reducing equivalents to fuel oxidative phosphorylation. The ATP formed in this pathway - ~ 7 mmol/g_{DCW}h -, is mostly employed for maintenance. Even though this m_{ATP} is high compared to other reported values for *P. pastoris* (2.2 – 5 mmol/g_{DCW}h) (Chung *et al*., 2010), it is required to account for the fast consumption of both secondary metabolites under limited cellular growth. The use of a recombinant strain for model calibration, which might have higher maintenance requirements, could further explain this result.

Finally, during controlled growth at the feed phase, neither ethanol nor arabitol are produced. All the carbon is directed towards biomass formation and the energy necessary for its synthesis and maintenance. This result agrees with previous fluxomic studies carried

out in aerobic, glucose-limited chemostats (Baumann *et al*., 2010; Dragosits *et al*., 2009), where significant carbon fluxes through the oxidative and non-oxidative branches of the PPP were found, without arabitol formation. Furthermore, the model shows significant oxaloacetate transport from the cytosol to the mitochondria, which was also observed in the cited studies. The most distinguishable feature of this phase is the high activity of the TCA cycle, which almost doubles the flux through this pathway reported under glucose limited conditions in chemostats (Baumann *et al*., 2010; Dragosits *et al*., 2009; Heyland *et al*., 2011). This higher activity in the TCA is probably associated with the need to cope with maintenance and growth-associated energy requirements under stressful conditions, such as high cell density, especially when no significant substrate level phosphorylation besides glycolysis occurs.

This analysis could have been performed using the genome-scale model in static conditions by deriving instantaneous exchange rates from contiguous samples and determining the flux distributions by specific growth rate maximization. Nevertheless, the inspection of flux distributions after model calibration has the advantage of considering the overall behavior of the cells during the cultivations. This provides more experimental support for the determination of parameters such as m_{ATP}, K_G, that cannot be directly estimated but that have a strong impact on the model output.

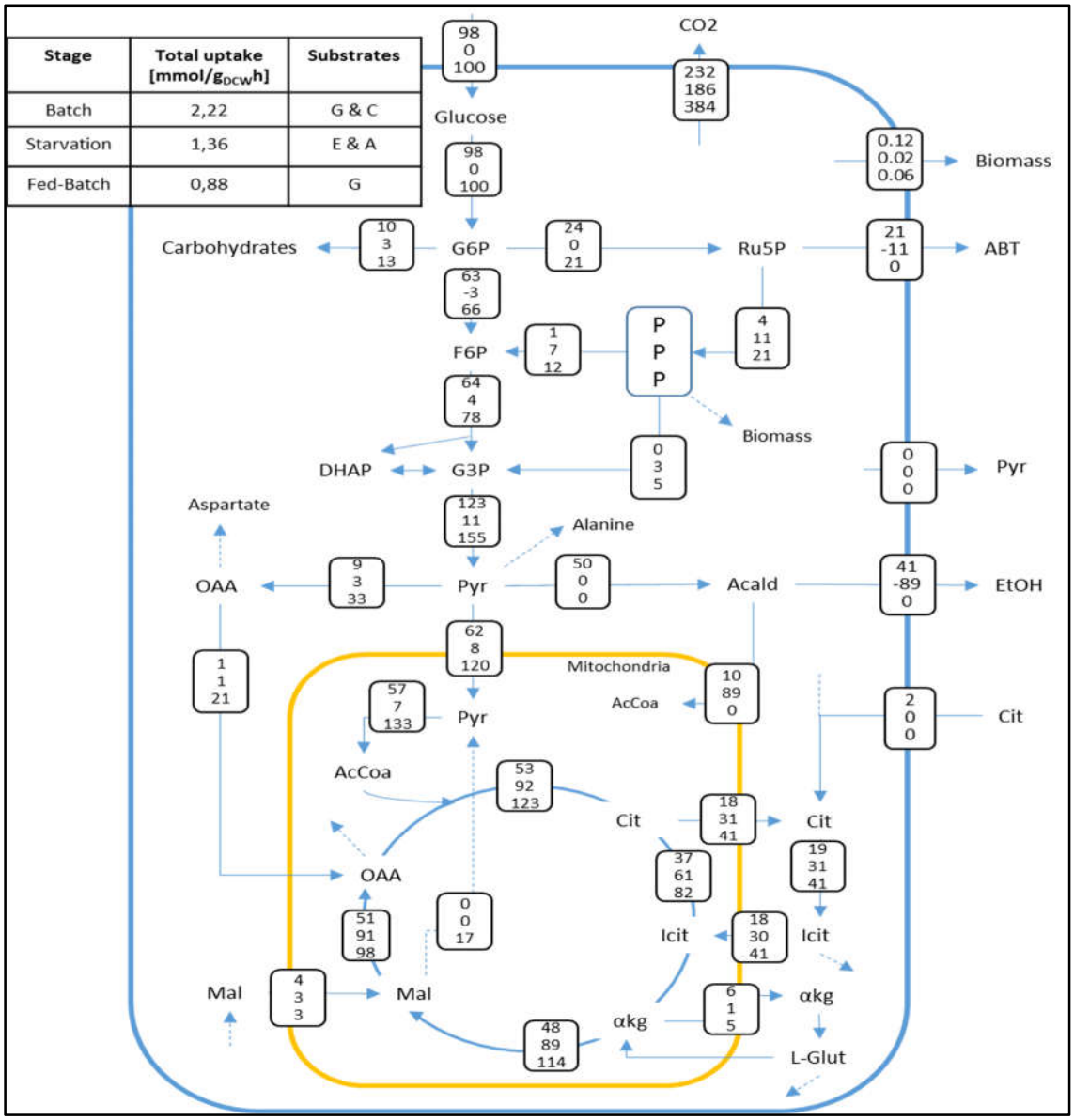


Figure 3-7 – Metabolic flux distribution in the Central metabolism for three different stages of the cultivation. Carbon uptake is detailed in the box of the upper left corner in mmol/g_{DCW}h and the fluxes are presented relative to this uptake. In each box between metabolites there are three numbers; which correspond, from top to bottom, to the relative flux during batch, starvation and feeding phases. Depending on the time analyzed, the cell consumes Glucose (G), Citrate (C), Arabitol (A) or Ethanol (E). The biomass flux corresponds to the specific growth rate of the cell in h^{-1} and the negative fluxes refer to a change in the reaction directionality. Nomenclature: G6P = Glucose 6 Phosphate, Ru5P = Ribulose 5 Phosphate, ABT = Arabitol, PPP = Non-oxidative phase of the Pentose Phosphate Pathway, F6P = Fructose 6 Phosphate, G3P = Glyceraldehyde 3 Phosphate, DHAP = Dihydroxyacetone Phosphate, Pyr = Pyruvate, OAA = Oxaloacetate, Acald = Acetaldehyde, EtOH = Ethanol, AcCoA = Acetyl Coenzyme A, Cit = Citrate, Icit = Isocitrate, αkg = Alpha-keto glutarate, Mal = Malate and L- glut = Glutamate.

3.3.5.2 Discovery of single knock-outs to improve recombinant Human Serum Albumin production using Minimization of Metabolic Adjustment (MOMA) as the objective function to simulate mutant behavior.

We performed 670 (number of genes in the model) batch simulations of single knock-out strains to discover beneficial deletions for the production of recombinant Human Serum Albumin (HSA), a 66 kDa protein with 16 disulfide bridges, that comprises about one half of the total blood serum protein (Verney, 1926).

The two main clusters (Figure *3-8*) show the relation between the final HSA and the final biomass concentration of the 130 mutations that improved HSA production (>30 mg/L at the end of the batch). The first cluster consists of strains that privilege HSA production over biomass formation; whereas the second one presents a trade-off between both.

We decided to leave Cluster I out of the analysis because of the impaired growth observed in the simulations, mainly due to the deletion of reactions associated to lipid biosynthesis. However, candidates from Cluster II (32 in total) were manually analyzed to identify the cause of HSA overproduction (Supplementary material S 3-7).

A relative increase in the formation of cysteine and tryptophan was found for most of the candidates for Cluster II when compared to the parental strain, a trend that was not observed for the rest of the amino acids (Figure *3-9*). These energetically costly residues (Raiford *et al*., 2008) are formed from serine. Therefore, re-routing carbon through this pathway could be beneficial to improve HSA production.

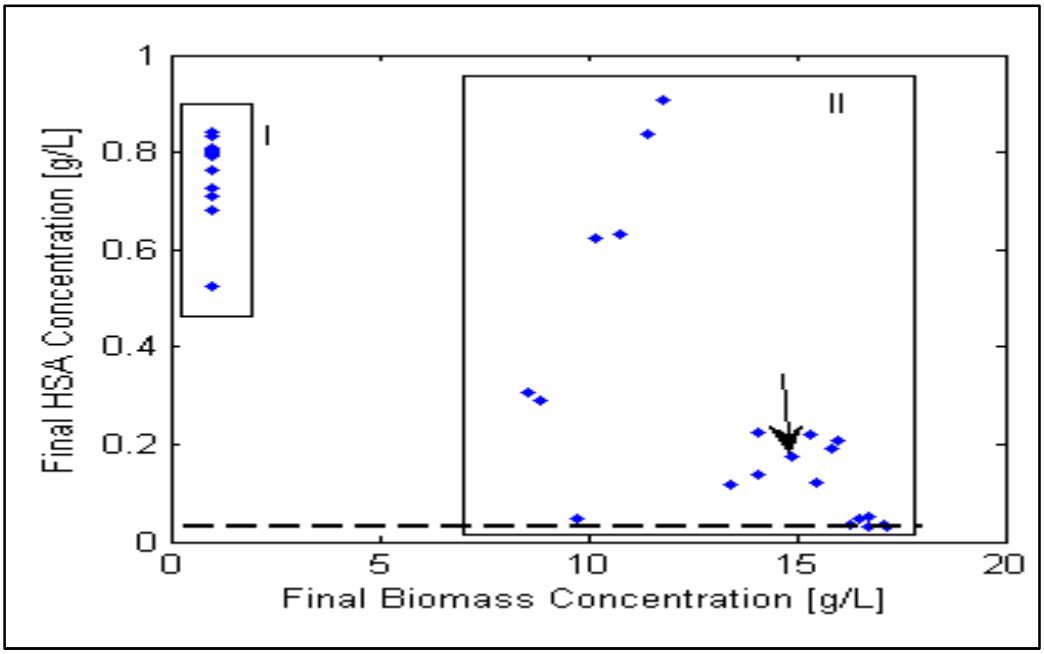


Figure 3-8 - Final HSA vs. final biomass concentrations of simulated batch cultivations of single knock-out-strains. Blue dots correspond to the output of strains that improved the initial final HSA concentration (30 mg/L). Candidates out of Cluster II were manually analyzed. The red circle indicates the performance of the parental strain and the black arrow points to the methylene tetrahydrofolate dehydrogenase knock-out strain.

After manually analyzing the candidates, we found that one possible strategy could be the deletion of the cytosolic NAD-dependent methylene tetrahydrofolate dehydrogenase (Figure *3-10*). When compared to the parental strain, the knock-out results in a 6.3 fold improvement of the final concentration of the recombinant protein with a 5.8-fold increase in protein volumetric productivity (arrow in Figure *3-8*). This deletion eliminates the transformation of serine to 5-10 methylene tetrahydrofolate; hence, serine can be re-routed to two cysteine reactions. This gene is non-essential in *S. cerevisiae* (West *et al.*, 1996) and, to the best of our knowledge, its essentiality has not been determined in *P. pastoris*. Therefore, it constitutes an interesting knock-out candidate to improve recombinant HSA production.

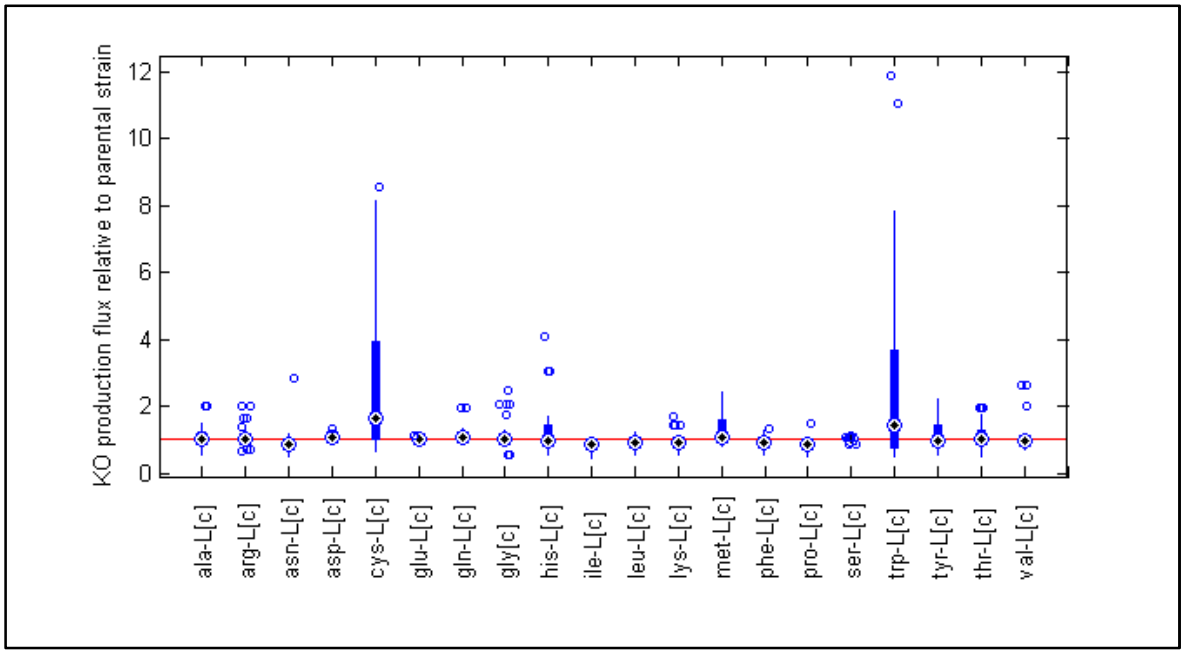


Figure 3-9 – Turnover of key amino acids in knock-out strains relative to the parental strain. Each box summarizes how the production of each amino acid changed in the 32 knock out strains of Cluster II relative to the production in the parental strain (Red Line). Black dots correspond to the sample median, the extreme of the boxes to the 25th and 75th percentiles, the whiskers extend to the most extreme data points and circles mark outliers.

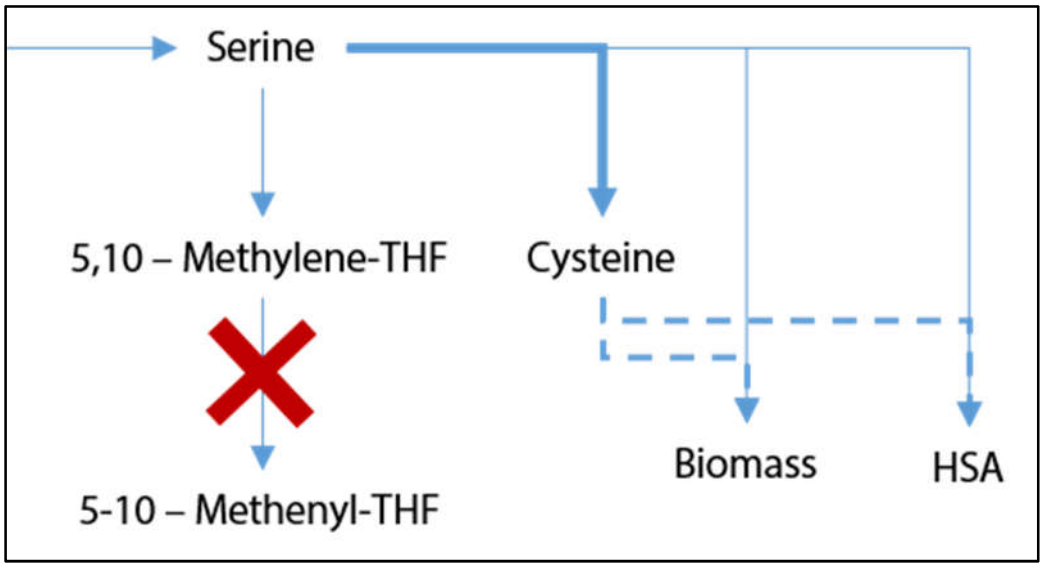


Figure 3-10 Rationale behind the knockout of the Methylene tetrahydrofolate (THF) dehydrogenase. By deleting this enzyme, the flux from Serine to 5-10-Methylene THF is blocked and redirected towards cysteine formation, whose availability increases the productivity of HSA.

3.3.5.3 Bioprocess optimization for HSA overproduction.

Here, we evaluated 13 feeding strategies of a fed-batch cultivation to improve the production of recombinant HSA. After the simulations, we selected a strategy that considered a slow decrease in the growth rate from μ=0.14 h^{-1} to μ=0.08 h^{-1} during the feeding phase (Table 3 - 9). The selected policy allows a 25% increase in volumetric productivity and reaches almost the same final HSA concentration as the constant growth rate strategy that reached the highest concentration (μ=0.06 h^{-1}).

Table 3 - 9 - Feeding policies evaluated to improve the production of Serum Albumin in a particular bioreactor setup.

Strategy	**μ_{MAX}**	**Rate**	**μ_{MIN}**	**q_P [mg/g·h]**	**X_{FINAL} [g/L]**	**P_{FINAL} [mg/L]**	**Limitation**
1	0.14	-	-	2.85	164.8	138	Oxygen
2	0.12	-	-	2.59	187.8	135	Oxygen
3	0.1	-	-	2.32	195.3	130	Volume
4	0.08	-	-	2.29	191.3	138	Volume
5	0.06	-	-	2.28	184.7	154	Volume
Best	0.14	0.1	0.08	2.83	197.5	150	Volume

This table shows the process indicators for the constant feeding Strategies (1-5) plus the best decreasing growth rate strategy. μ_{MAX} is the maximum growth rate in the feeding police. μ_{MIN} is the minimum growth rate in the feeding police. Rate is the rate of decreasing of set growth rate in feeding police. q_P is the protein productivity. X_{FINAL} and P_{FINAL} refer to the final concentration of biomass and serum albumin in the reactor when the simulation stops, which happened by either violating user-defined volume or Oxygen Transfer thresholds.

The improvement in process productivity by modifying substrate addition during the feed phase is less efficient than the one attained by genetic modifications. However, other process variables such as reactor volume and oxygen transfer may be modified to further improve HSA production.

3.4 CONCLUSIONS

Current GSMs of *P. pastoris* have been employed to address cellular behavior in stationary conditions. They have been successfully used for predicting production and consumption

rates of different compounds and even achieving a 40% improvement of recombinant protein production by model-discovered knock-outs. However, little attention has been given to the actual metabolic flux distribution that these reconstructions yield and how they evolve in a dynamic environment. Resulting flux distributions are important for two reasons: (i) they help to understand the cellular response to the different stresses to which the cell is subjected to and (ii) they can serve as input for several algorithms whose aim is to find metabolic engineering targets to improve the production of a certain compound.

In this work, we developed a robust dynamic GSM of glucose-limited aerobic cultivations of *P. pastoris*, linking and showing the impact that the model formulation process has over flux balance analysis. The assembled platform can fit several datasets with minimum significance, sensitivity and identifiability problems in its parameters. Moreover, if properly trained, it can be used to predict bioreactor dynamics. The model could also be employed to obtain realistic flux distributions throughout dynamic cultivations and to determine metabolic and process engineering strategies to improve the production of a target compound.

To broaden its applications to other relevant conditions for *P. pastoris*, the model could be calibrated with data from cultures with different carbon sources and feeding strategies, such as glycerol batch phase followed by a methanol induction phase. Also, the model could be used to study perturbations such as oxygen limitation, which is a common problem in industrial *P. pastoris* cultivations (Porro *et al*., 2005). Moreover, it would be desirable to calibrate the model with data from a strain capable of producing high concentrations of a recombinant protein to understand and quantify the metabolic burden caused by this production.

Finally, it is expected that the incorporation of more curated metabolic reconstructions (Tomàs-Gamisans *et al*., 2016), gas mass balances and the knowledge derived from testing the hypotheses proposed using the model would improve its accuracy and broaden its applicability.

4. CHAPTER III: AUTOMATED ALGORITHM TO DETERMINE KLA CONSIDERING SYSTEM DELAY.

Paulina Torres[1], Marcel Otavio Cerri[2], Marcelo Perencin de Arruda Ribeiro[3], J. Ricardo Perez-Correa[1] and Eduardo Agosin[1].

[1] Department of Chemical and Bioprocess Engineering, School of Engineering, Pontificia Universidad Católica de Chile, Vicuña Mackenna 4860, Santiago, Chile.

[2] Department of Bioprocesses and Biotechnology, Faculty of Pharmaceutical Sciences, Univ Estadual Paulista "Júlio de Mesquita Filho", Araraquara, SP, Brazil, CEP 14801-902.

[3] Department of Chemical Engineering, Federal University of São Carlos, C.P. 676, São Carlos, SP, Brazil, CEP 13565-905.

Published in Journal of Chemical Technology and Biotechnology (2016) 92:7, 1630-1737

4.1 INTRODUCTION

Oxygen is an essential substrate in aerobic bioprocesses. It is considered one of the main limiting factors to reach high-density cell cultures (Garcia-Ochoa & Gomez, 2009; Sonnleitner, 2016; Wang *et al*., 2009). Given its low solubility in high density cultures, and the varying demand in batch and fed-batch cultures, an increasing oxygen transfer rate (OTR) over time is required (Garcia-Ochoa *et al*., 2010). OTR is typically described as the product of a volumetric mass transfer coefficient, $k_L a$, and the average concentration gradient based on the two-film theory of Whitman (Whitman, 1924) (Equation 4-1).

$$\frac{dC}{dt} = k_L a \cdot (C^* - C) \qquad (4-1)$$

Where, C* is the oxygen equilibrium concentration and C is the dissolved oxygen concentration in the liquid phase of the bioreactor.

OTR mainly depends on the $k_L a$ coefficient in a specific saturation condition (Equation 4-1). This coefficient is fundamental to measure and characterize the oxygen

transfer capabilities of gas-liquid bioreactors (Linek *et al*., 1993). Although different mechanistic models have emerged for prediction of $k_L a$ (Buffo *et al*., 2016; Garcia-Ochoa *et al*., 2010; Kapic & Heindel, 2006; Moucha, Linek, & Prokopová, 2003; Yawalkar, Heesink, Versteeg, & Vishwas, 2002), they are not robust enough to yield an accurate quantitative determination of $k_L a$ (Villadsen *et al.*, 2011).

Physical methods for $k_L a$ estimation are based on dissolved oxygen (DO) concentration changes with time (Garcia-Ochoa & Gomez, 2009; Villadsen *et al*., 2011), where the dynamic gassing-in and gassing-out procedures are still the most commonly used (Cerri *et al.*, 2016; Garcia-Ochoa & Gomez, 2010; Sonnleitner, 2016). However, these approaches could become inaccurate if the delay of the measuring system is neglected (Badino *et al.* 2000; Garcia-Ochoa & Gomez, 2009; Gourich *et al.*, 2008; Merchuk *et al*., 1990). The latter can be mainly attributed to the response time, τ_e, of the electrode's probe, whose dynamic response can be approximated by the following first order equation (Cerri *et al*., 2016; Garcia-Ochoa & Gomez, 2009; Gourich *et al*., 2008; Tribe *et al.*, 1995):

$$\frac{dC_e}{dt} = k_e \cdot (C - C_e) \qquad (4-2)$$

Where, C_e is the measured DO concentration, C is the real DO concentration, and k_e is the sensitivity of the electrode, which is the inverse value of τ_e, i.e. $k_e = 1/\tau_e$.

The delay of the probe can be neglected only if the sensitivity of the electrode, k_e, is at least ten fold higher than $k_L a$ values, *i.e.* the ratio $k_e/k_L a$ > 10 (Garcia-Ochoa & Gomez, 2009; Gourich *et al*., 2008; Merchuk *et al*., 1990). For $k_e/k_L a$ ratios lower than 10, a second-order dynamic model should be considered for an accurate estimation of $k_L a$ (Equations 4-1 and 4-2) (Merchuk *et al*., 1990). In this range, the estimation of $k_L a$ is considered accurate when the $k_e/k_L a$ ratio is higher than 2 (Cerri *et al*., 2016). However, the quantification of the associated error for $k_L a$ estimation has not been formally studied yet. The characteristic τ_e of commercial, fast oxygen electrochemical sensors ranges between 5 and 90 s. However, τ_e depends also on the fluid dynamics in the reactor and the condition of the sensor (Badino *et al*., 2000; Cerri *et al*., 2016; Gourich *et al*., 2008).

Therefore, an accurate estimation of $k_L a$ requires to verify the $k_e/k_L a$ ratio in all the conditions of interest.

A new second-order method was recently proposed by Cerri *et al.* (2016) to overcome the effect of probe delay on $k_L a$ estimations. This method enables the simultaneous determination of $k_L a$ and k_e using the traditional gassing-out method, without any additional assays. Nevertheless, this new methodology requires the manual selection of a time range to estimate $k_L a$ values (Cerri *et al*., 2016). Manual selection and analysis are cumbersome and time consuming, and reproducible results are difficult to attain.

In this work, we developed an automated algorithm to estimate $k_L a$ and k_e values considering a second-order dynamic process. Moreover, the impact of the $k_e/k_L a$ ratio on the estimation of $k_L a$ and k_e is also discussed and it was proposed the values of $k_e/k_L a$ ratio to obtain reliable dynamic curves for accurate estimation of $k_L a$.

4.2 MATERIAL AND METHODS

4.2.1 Bioreactor configuration.

The fermenters employed in this study were in-house built 1 L bioreactors with 0.6 L of operational volume, equipped with a condenser, a simple diffuser, a stirrer and two Rushton turbines. Agitation was driven by brushless VEXTAs DC motors, AXH Series (Oriental Motor, Japan). Each bioreactor consists of a double-jacketed cylindrical vessel with an internal diameter of 9 cm. The vessel was equipped with four equally spaced vertical baffles of width 0.9 cm and thickness 1 mm. Both impellers were separated by 3 cm and the lower impeller was at 3 cm from de bottom of the vessel (Figure *4-1*). The 2 μm pores diameter diffuser was located below the lower. For monitoring and control, a SIMATIC PCS7 distributed control system (Siemens, Germany) was used. DO concentration and temperature were measured with fast response probes Oxymax COS22D (< 60 s) (Endress Hauser, Switzerland). Air and N_2 additions were measured and controlled with mass flow meters (FMA-A2407, Omega Engineering, USA).

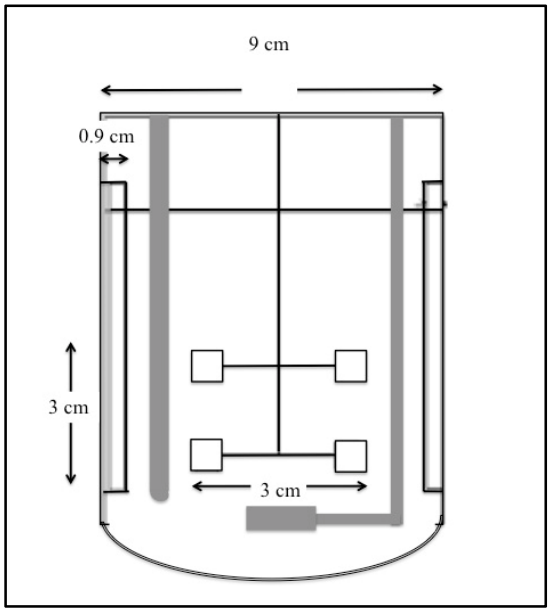


Figure 4-1 Scheme of the conventional bioreactor

4.2.2 Fluids

Three liquid mixtures were employed: distilled water (W: μ = 0.87 mPa.s) and two glycerol solutions: GS1, 30 % glycerol (v/v) with μ = 3.23 m·Pa s; and GS2, 50 % glycerol (v/v), with μ = 6.60 mPa s. The gas phase contained compressed air and 99.9 % nitrogen (Indura, Santiago, Chile). Dynamic viscosities of the solutions were determined from rheograms obtained at 25ºC, using a digital concentric-cylinder rheometer (DVIII Ultra, Brookfield Engineering Laboratories, MA, USA).

4.2.3 Methods for estimation of k_La considering system delay

Experimental absorption curves were generated by the gassing-out method (Garcia-Ochoa & Gomez, 2009; Linek *et al.*, 1987). The DO concentration in the fluid was measured continuously after a step change in the inlet gas. The oxygen sensor was installed in the bioreactor at the same height of the bottom impeller and close to the vessel wall (1 cm). The sensor was calibrated and then it was stabilized for 4 hours under air saturation conditions. Desorption was carried out by bubbling pure nitrogen and absorption by bubbling air; in both cases gas flow rate and the agitation were kept constant. In total, 15 conditions were tested varying gas flow rate (0.5 to 2 VVM), agitation (200 to 1000 rpm), and viscosity (W, GS1, GS2).

The system delay was quantified by the method of Cerri *et al*, 2016. This method considers that the measured DO concentration evolves in parallel to the real DO concentration in a specific time range (Figure *4-2* a-b). In this time range, both the measured DO concentration and the real DO concentration yield the same value of the estimated $k_L a$ (Figure *4-2* c), using a first order equation (Equation 4-1).

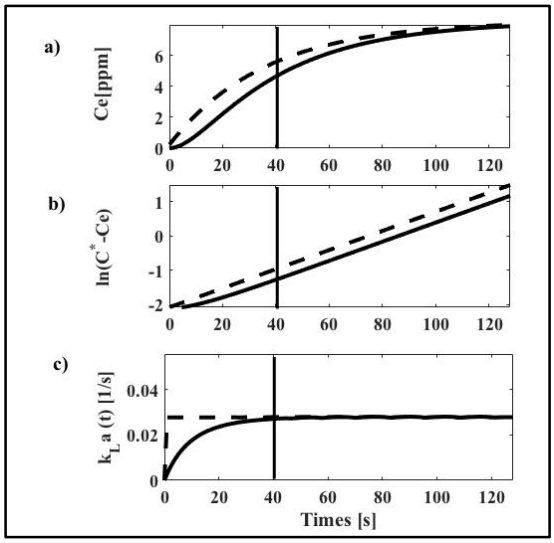


Figure 4-2 Comparison between measured oxygen concentration (solid line) and real oxygen concentration (dashed line) using simulated curves of $k_L a$ of $2.78 \cdot 10^{-2}$ s^{-1} and k_e of $12.95 \cdot 10^{-2}$ s^{-1} (Equations 4-1 and 4-2). The time range at which $k_L a$ can be estimated using a first–order equation is over 40 s. a) Dynamic absorption of oxygen concentration. b) – ln(C*-C) versus time graph to estimate $k_L a$ using first order equation. c) Instantaneous $k_L a$.

4.2.4 Modeling

The algorithm was coded in MATLAB 2016a (MATLAB, 2016) and implemented in a Windows 7 PC with an Intel® Core™ i5-6600K processor. The average computational time required for estimating $k_L a$ and k_e in a typical absorption curve was approximately 0.4 s. The identification of the $k_L a$ constant zone was carried out by the heuristic search algorithm *fminsearchbnd*, which is a variation of *fminsearch* of the MATLAB optimization toolbox. The fitting procedures were carried out by least squares minimization and the significance of the regression (p = 0.01) was assessed by applying the Fisher Test (Seber *et al.*, 2012) to the quotient between regression and residual variances; 99% Confidence Intervals (CIs) were inferred, following a t-student distribution.

Eight representative curves of different conditions were selected for pre- and post-regression analyses. Post regression analysis was carried out to assess the fitting performance and robustness (Jaqaman & Danuser, 2006). The algorithm parameters were tuned by trial and error to optimize the automatic selection of the $k_L a$ constant zones and the quality of the fitting. The tuning parameters of the algorithm were subjected to a sensitivity analysis considering +/- 50% variation of the selected value.

To verify the usefulness of the proposed procedure, the manual and the automated methods were applied to estimate the $k_L a$ of 80 absorption experiments obtained under different conditions of agitation, aeration and viscosity.

4.2.5 Simulation and $k_e/k_L a$ ratio

To quantify the errors associated with the $k_e/k_L a$ ratio, 640 curves were simulated with different $k_L a$ values (0.278 $\cdot 10^{-2}$ to 33.3 $\cdot 10^{-2}$ s^{-1}) and $k_e/k_L a$ ratios (0.25 to 25). The large amount of different combinations of $k_L a$ and k_e values represent the high variability of physical and geometrical conditions of bioreactors. In order to emulate real conditions, all absorption curves were simulated considering white Gaussian noise of real measurements (Equation 4-3).

$$Ce_s(t) = Ce(t) + \varepsilon_i(t) \qquad (4-3)$$

Where $Ce_s(t)$ is a simulated DO concentration, $\varepsilon_i(t)$ is a zero mean random value, where the expected value of $\varepsilon^2{}_i(t)$ is the mean variance of the sensor, and $Ce(t)$ is the deterministic simulated DO concentration, which is estimated using a system of ordinary differential equations (Equations 4-1 and 4-2).

Only curves with $k_e$values ranging from 0.556 $\cdot 10^{-2}$ to 27.8$\cdot 10^{-2}$ s^{-1} were considered, since these values were within an operating range of a real sensor. The accuracy of the model was evaluated by the absolute percentage of the error (Equation 4-4).

$$Error = \frac{|\theta_{estimated} - \theta_{real}|}{\theta_{real}} \cdot 100 \qquad (4-4)$$

Where, θ is the variable, which can be either $k_L a$ or k_e.

4.3 RESULTS AND DISCUSSION

For $k_L a$ estimations using a second-order dynamic model, three different zones could be identified in an absorption curve (Figure *4-3*):

- A first zone, at the beginning of the C_e curve, where the delay of the system considerably affects the measured oxygen concentration.
- A second zone, in the middle of the curve, where C_e can be represented as a first order kinetics, i.e. the system delay does not affect the dynamic of the curve. Here, $k_L a$ can be directly estimated from the C_e curve, since the slope of the natural logarithm of the difference between C^* and C_e is equivalent to the slope of the natural logarithm of the difference between C^* and C (Equation 4-5).

$$\frac{\mathrm{dln}(\mathrm{C}^* - C)}{\mathrm{dt}} = \frac{\mathrm{dln}(\mathrm{C}^* - C_e)}{\mathrm{dt}} = -k_L a \qquad (4-5)$$

- Finally, the last part of the curve, where C_e is close to the dissolved oxygen saturation concentration (C^*). Here, the OTR is close to zero (Equation 4-1), implicating that in this zone the estimation of $k_L a$ could have parametric sensitivity problems and C_e dynamics could be more sensitive to the sensor noise than to the real $k_L a$ value.

Although these zones are not clearly delimited, the second one can be easily identified by following four consecutive steps: data smoothing; selection of a zone with high oxygen transfer rate; identification within this zone of the area where instantaneous $k_L a$ is constant; and estimation of $k_L a$ and k_e. These different steps are described in detail below.

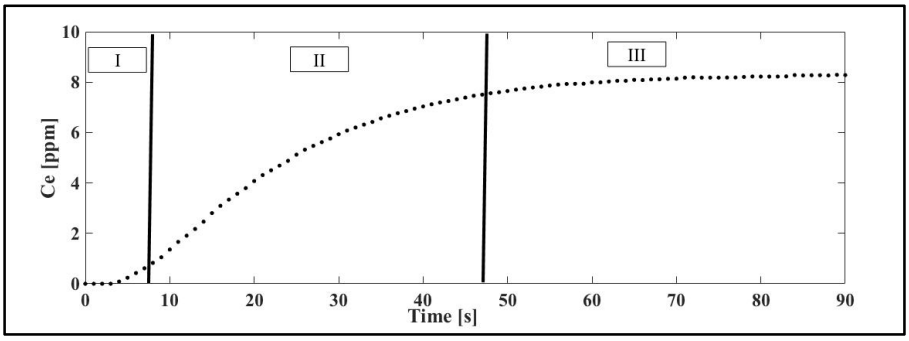


Figure 4-3 Identification of the three main zones of an experimental absorption curve. The absorption curve was obtained with water at 500 rpm and 0.75 VVM. Zone I is where the delay of the system considerably affects the measured oxygen concentration. Zone II is where measured dissolved oxygen concentration can be represented as a first order kinetics. Zone III is where measured dissolved oxygen concentration is close to the dissolved oxygen saturation concentration.

4.3.1 Smoothing data: Simple Moving Average Filtering

The smoothing process of raw data is carried out to reduce the impact of white noise and to identify more easily the curve trends (Hamilton, 1994). We employed a simple moving average filter to smooth the data of measured oxygen concentration (Ce), considering α data points, where α is a tuning parameter of the automated algorithm (Equation 4-6).

$$Ce(t) = \frac{\sum_{i=t-\alpha/2}^{i=t+\alpha/2} Ce(i)}{\alpha} \qquad (4-6)$$

4.3.2 First Filter: selection of data with high oxygen transfer rate

High instantaneous OTR values ($OTR(t)$) were employed to identify the second zone. *OTR(t)* were estimated using the second-order approximation of the Ce derivative (Equation 4-7).

$$OTR(t) = \frac{dCe(t)}{dt} = \frac{Ce(t+1) - Ce(t-1)}{(t+1) - (t-1)} \qquad (4-7)$$

This first filter allows to reduce the search space, selecting the data that have $OTR(t)$ values higher than a threshold (Figure *4-4*). This threshold is a tuning parameter of the algorithm, which is defined by a percentage of the maximum $OTR(t)$. Selection of the boundaries of this reduced search space is more difficult when a large number of outliers

exceed the OTR threshold. These outliers result from a high variability in the *OTR(t)* curve, and to identify them, we analyzed a predefined neighborhood, denoted as:

$$[t - \beta \cdot t_{total}, t + \beta \cdot t_{total}\,] \qquad (4-8)$$

Where t is the time, t_{total} is the total experimental time of the curve and β is the percentage of the total experimental time; an additional tuning parameter of the algorithm. An *OTR (t)* value is considered an outlier if the data in its close neighborhood is lower than the OTR threshold.

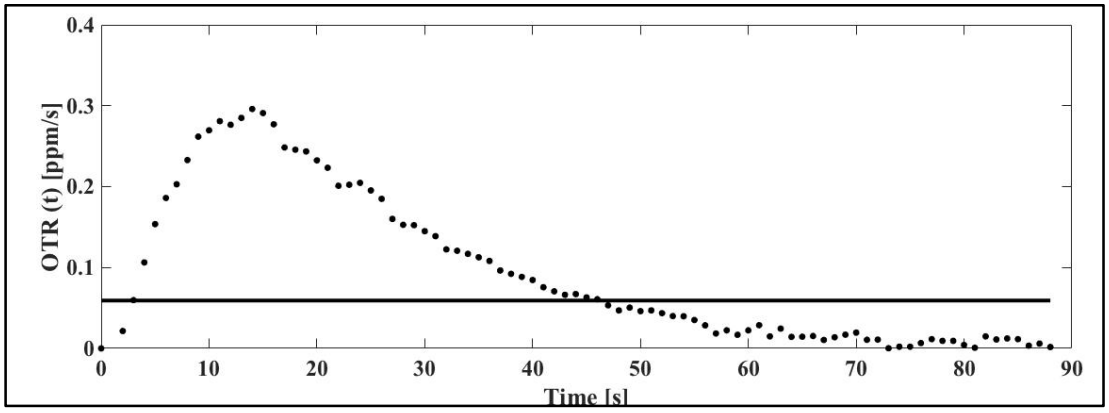


Figure 4-4 Representation of the first filter. The instantaneous oxygen transfer rates of an experimental absorption curve is represented by solid circles. The absorption curve was obtained with water at 500 rpm and 0.75 VVM. Instantaneous oxygen transfer rates were estimated using Equation 4-7 and the threshold instantaneous oxygen transfer rate was estimated as 20% of maximum instantaneous oxygen transfer rate (solid line).

4.3.3 Search algorithm: selection of the zone where instantaneous $k_L a$ is constant

After reducing the search space, instantaneous $k_L a$ ($k_L a(t)$) can be estimated using Equation 4-9 to identify the second zone.

$$k_L a(t) = \frac{OTR(t)}{(C^* - Ce(t))} \qquad (4-9)$$

In this zone, instantaneous inference of $k_L a$ using measured dissolved oxygen concentrations has a constant value (Figure *4-2*). Thus, in this zone, we consider that: 1) the variation of $k_L a(t)$ can be neglected, 2) this zone does not have a high variability, and

3) this zone includes as many data points as possible. Applying linear regression to the target region, these features can be mathematically expressed as:

1. The slope (m) of this part of the curve must be near zero, *i.e.* minimize the square of the slope.
2. The quality of the fitting (R^2) must be the best possible, i.e. maximize R^2.
3. The number of data (N) must be as much as possible, i.e. maximize N.

This zone can be denoted as $[C_e(Pi - n);\ C_e(Pi + n)]$, where Pi is the central point and $N = (2n + 1)$ is the size. An objective function was implemented to find this zone, considering the features mentioned above. Given that it is a multiobjective optimization, we chose the desirability approach (Costa *et al.*, 2011), considering higher weights for features 1 and 2 (Equation 4-10). It is worthy to mention that it is necessary to normalize the number of data, N, considering the minimum (N_{min}) and maximum number (N_{max}) of data.

$$f(Pi, n) = \begin{cases} 1 & N < N_{min} \\ \left(\dfrac{(m^2)}{\left(\dfrac{N - N_{min}}{N_{max} - N_{min}}\right) \cdot (R^2)^2} \right)^{\frac{1}{5}} & N_{min} < N < N_{max} \\ 1 & N > N_{max} \end{cases} \qquad (4-10)$$

$$s.t.$$

$$0 < Pi < N_{total}$$

$$0 < n < \frac{N_{total}}{2}$$

Using this search algorithm, the zone with the minimum value of the objective function was chosen as the second zone (Figure *4-5*). The optimization of this algorithm was carried out using three initial points within the reduced search zone, taken at the beginning, mid and final part of the curve.

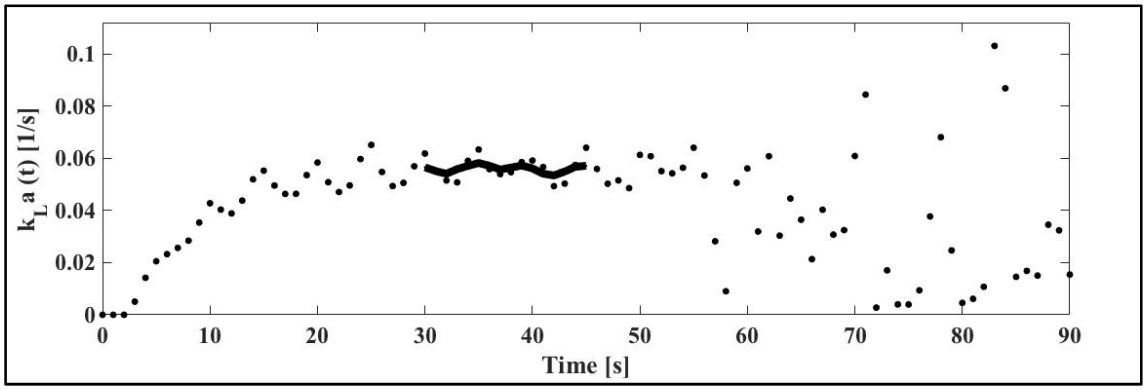


Figure 4-5 Representation of the search algorithm. Instantaneous $k_L a$ of an experimental absorption curve is represented in solid circle. The absorption curve was obtained with water at 500 rpm and 0.75 VVM. Instantaneous $k_L a$ was estimated using Equation 4-9. The selected zone, where instantaneous $k_L a$ is constant, is represented in solid line.

4.3.4 $k_L a$ and k_e estimation using the chosen range

Once the search algorithm selects the second zone, $k_L a$ is estimated by linear regression between $-\ln(C^* - Ce)$ and time, employing the analytic integration of Eq. 4-1 (Figure *4-6*). k_e is then estimated (Figure *4-7*) for the same interval, since the derivative of Ce is equal to the derivative of C (Equation 4-11),

$$k_e(t) = \frac{k_L a}{1 - \left(\frac{C^* - C(t)}{C^* - Ce(t)}\right)} \qquad (4-11)$$

Where C is the real value of oxygen concentration, which could be estimated by:

$$C(t) = \left(C^* - Ce(t0)\right) \cdot e^{k_L a \cdot (t - t0)} \qquad (4-12)$$

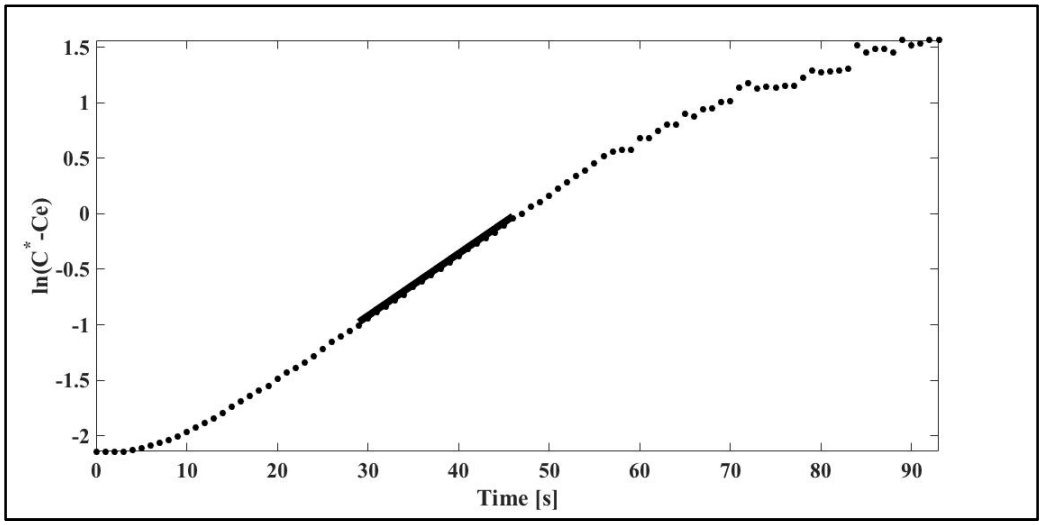


Figure 4-6 Estimation of $k_L a$ in the second zone. The natural logarithm of the difference between C^* and C of an experimental absorption curve is illustrated in solid circle. The absorption curve was obtained with water at 500 rpm and 0.75 VVM. The selected zone to estimate $k_L a$ is highlighted (solid line).

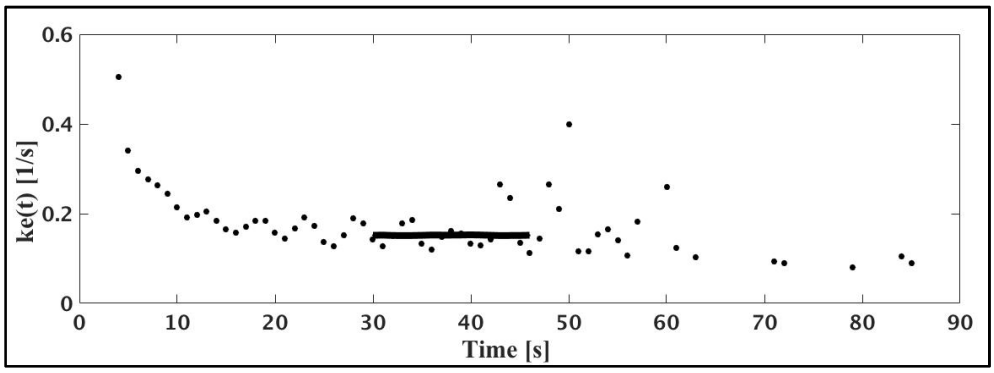


Figure 4-7 Estimation of k_e in the second zone. k_e of an experimental absorption curve is represented by solid circles. The absorption curve was obtained with water at 500 rpm and 0.75 VVM. k_e was estimated using Equation 4-11. The selected zone to estimate k_e is highlighted (solid line).

4.3.5 Tuning algorithm parameters

Different fluids (W, GS1 and GS2) and $k_L a$ values ($1.1 \cdot 10^{-2}$ to $7.7 \cdot 10^{-2}$ s^{-1}) were considered to tune the algorithm (Figure *4-8*). Parameter β showed low sensitivity and was fixed at 2%. The other parameters, i.e. α, OTR threshold, $Nmin$ and $Nmax$, were manually fixed by trial and error until the difference between manual and automatic selection was minimized (Table 4- *1*).

Table 4- 1 Sensitivity of the tuning parameters of the automated method.

	Parameters			k_La estimate variation (%)		
	Notation	**Value**	**Variation**	**Minimum**	**Mean**	**Maximum**
Smoothing			2	-9.79	-1.02	+6.05
	α	4	6	-2.83	-0.10	+1.80
1º Filter	OTR threshold	0.2·OTRmax	0.1·OTRmax	-0.53	+5.74	+20.37
			0.3·OTRmax	-3.44	-0.88	+1.09
Search algorithm	Nmin	0.3·Ntotal	0.15·Ntotal	-1.16	+2.82	+12.86
			0.45·Ntotal	-3.95	+0.40	+6.88
	Nmax	0.7·Ntotal	0.35·Ntotal	-5.36	-0.91	+2.84
			Ntotal	+0.01	+0.04	+0.07

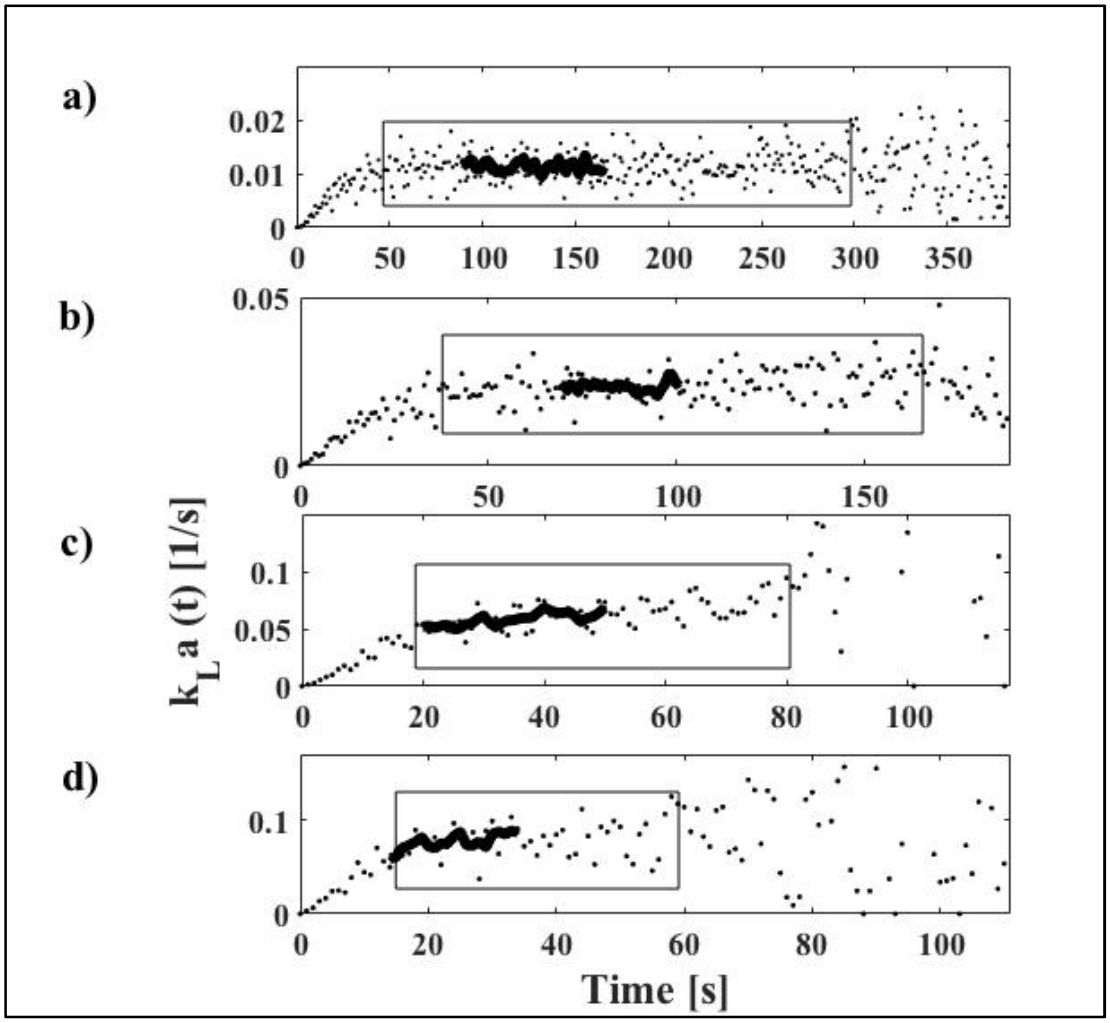


Figure 4-8 Instantaneous $k_L a$ curves of experimental absorption curves at different operational conditions in water solution. Solid circles represent instantaneous $k_L a$. In solid lines, the zone selected by the algorithm is highlighted. The manual selection is highlighted in a box. The different experimental absorption curves were carried out as follows: a) 200 rpm and 0.3 VVM, corresponding to a $k_L a$ of 1.129 $\cdot 10^{-2}$ s^{-1} (CI: 1.126-1.132$\cdot 10^{-2}$ s^{-1}). b) 200 rpm and 1.2 VVM, i.e. $k_L a$ = 2.330 $\cdot 10^{-2}$ s^{-1} (CI: 2.317-2.343$\cdot 10^{-2}$ s^{-1}). c) 500 rpm and 0.75 VVM, $k_L a$ = 5.855 $\cdot 10^{-2}$ s^{-1} (CI: 5.753-5.957$\cdot 10^{-2}$ s^{-1}). d) 800 rpm and 1.2 VVM, $k_L a$ = 7.686 $\cdot 10^{-2}$ s^{-1} (CI: 7.538-7.833$\cdot 10^{-2}$ s^{-1}).

4.3.6 Sensitivity analysis using different curves

Our analyses indicate that the tuning parameters were insensitive. Variation on $k_L a$ estimates ranged between -1.02% and +5.74% in average, although the parameters varied by 50 % (Table 4- *1*). Nevertheless, each parameter was analyzed separately:

- Parameter $\propto$ is involved in data smoothing and is more important for very noisy experimental curves (e.g. Figure *4-8* c), since it is difficult to identify curve trends in these curves. When the degree of smoothing is decreased, the algorithm could not distinguish the boundary between the first and the second zone, resulting in an underestimation of the $k_L a$ value.
- The OTR threshold, involved in the first filter, allows reducing the search space. This parameter is crucial for curves with low k_e values (e.g. glycerol curve in Figure *4-9*). However, a trade-off is recommended, since increasing this parameter excessively reduces the search space, which could lead to the identification of an erroneous zone.
- The parameter $Nmin$, which is involved in the search algorithm, avoids the selection of local minima. Notwithstanding, it is important to emphasize that it should not be higher than 30 %, since it might include other zones, affecting the fitting.
- Finally, the parameter $Nmax$, which is involved in the search algorithm, is used to normalize the number of data in the objective function. This parameter has a low sensitivity. However, values higher than 70 % of the total number of data are recommended since if the algorithm selects a too small data set, the selected zone will not be representative.

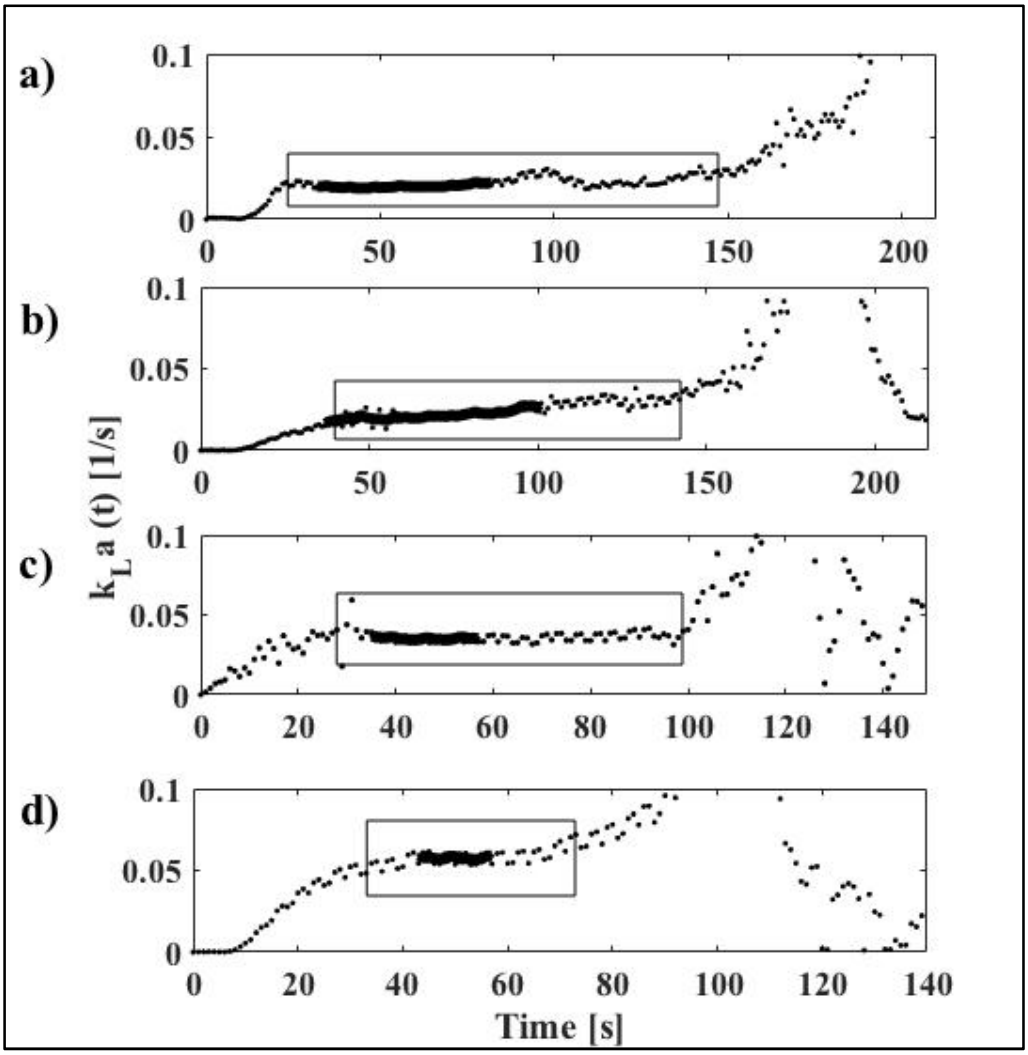


Figure 4-9 Instantaneous $k_L a$ curves of experimental absorption curves at different operational conditions in glycerol solutions. Instantaneous $k_L a$ are represented with solid circles. The zone selected by the algorithm is shown in solid lines. The manual selection is highlighted in a box. The different experimental absorption curves were a) GS2 solution at 500 rpm and 0.75 VVM, $k_L a$ = 1.969 $\cdot 10^{-2}$ s^{-1} (CI: 1.959-1.979$\cdot 10^{-2}$ s^{-1}). b) GS2 solution at 800 rpm and 1.2 VVM, $k_L a$ =2.069 $\cdot 10^{-2}$ s^{-1} (CI: 2.048-2.090$\cdot 10^{-2}$ s^{-1}). c) GS1 solution at 500 rpm and 0.75 VVM, $k_L a$ = 3.517 $\cdot 10^{-2}$ s^{-1} (CI: 3.505-3.528$\cdot 10^{-2}$ s^{-1}). d) GS1 solution at 1000 rpm and 0.75 VVM, $k_L a$ =5.598 $\cdot 10^{-2}$ s^{-1} (CI: 5.539-5.656$\cdot 10^{-2}$ s^{-1}).

4.3.7 Validation

The automatic procedure was compared to manual selection in 80 new experiments performed at 25°C, with different conditions of agitation, aeration and viscosity (Figure *4-10*). The average difference between the estimated value of both methods was 5.0% ±

3.4% (Min: 0.1% and Max: 14.2%). Manual selection was less reproducible since the users tend to choose different calculation data points (Zone 2 above). In turn, the automatic algorithm ensures a consistent selection of the calculation data points. Additionally, the automatic algorithm takes a fraction of the time required by the manual procedure to estimate the $k_L a$ values.

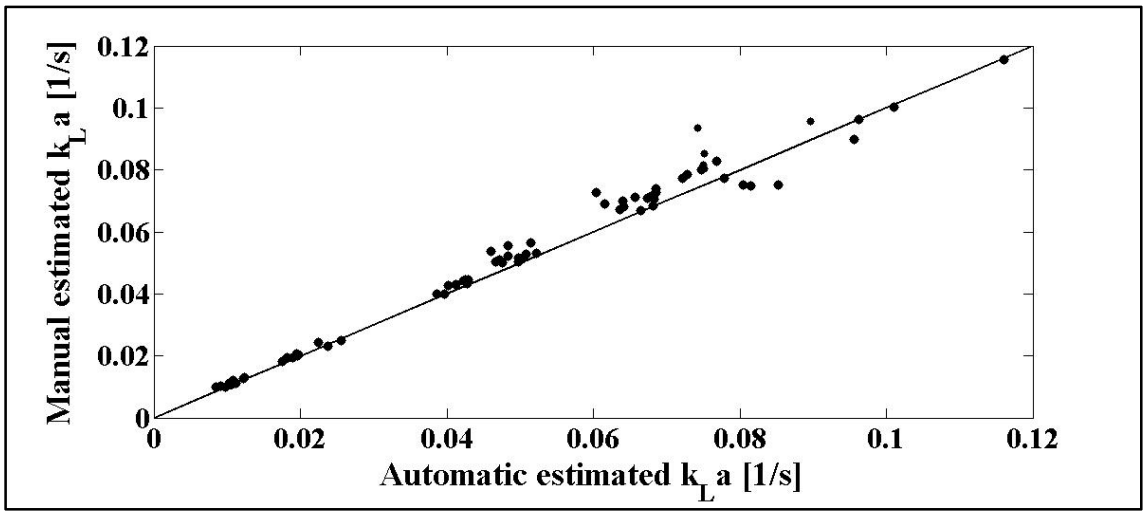


Figure 4-10 Comparison of manual and automatic strategies of 80 absorption experimental curves. The experimental data is represented by solid circle and the equation x=y is represented in solid line.

4.3.8 Simulation and $k_e/k_L a$ ratio

Once the model was validated, the impact of the $k_e/k_L a$ ratio on the $k_L a$ estimation, as well as the algorithm´s efficiency were evaluated using 640 simulated absorption curves (Figure *4-11*). The $k_L a$ estimation error decreases to around 1% for $k_e/k_L a$ ratios higher than 2.5 (Figure *4-11*). This ratio must be higher than 2 to ensure errors lower than 10 % (Cerri *et al*., 2016). For $k_e/k_L a$ ratios lower than 2, the error of $k_L a$ estimation increases exponentially and the algorithm tends to under estimated the real value of $k_L a$ (Data not show). Verification of the $k_e/k_L a$ ratios is crucial to analyze and design bioreactor since the $k_L a$ estimation would be limited by the system and would be under estimated. Actually, although most absorption curves fall in the accuracy zone of validation curves (Vertical line Figure *4-11*), 2 out of 80 estimated $k_L a$ were not considered accurate and

they should be repeated using faster sensor. In turn, accurate k_e estimations require k_e/k_La ratios higher than 4 (Figure *4-11*).

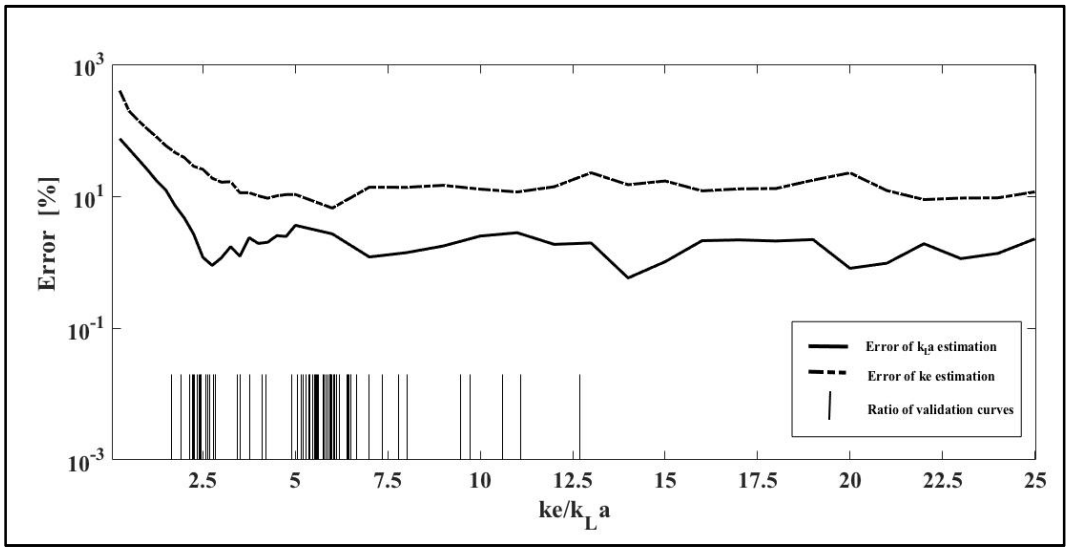


Figure 4-11 Effect of the k_e/k_La ratio on the estimations of k_La and k_e. The average error of k_La estimation is represented by solid lines and the average error of k_e estimation is represented by dashed lines. The k_e/k_La ratio of 80 absorption experimental curves, which were used to validate the algorithm, are represented by vertical lines.

4.4 CONCLUSIONS

A second-order automated method for data treatment is presented here to estimate reliably k_La and k_e values in conventional bioreactors operating in a wide range of hydrodynamic conditions. The algorithm achieves accurate estimations of k_La for k_e/k_La ratios larger than 2, while accurate estimations of k_e require k_e/k_La ratios higher than 4. Care must be taken when k_e/k_La ratios are smaller than 2, since k_La estimation errors increase exponentially. The proposed method can analyze hundreds of curves in few seconds, consequently it can be used to design a high throughput k_La estimation system to optimize oxygen delivery in highly demanding bioreactors.

5. CHAPTER IV: OPTIMIZATION OF OXYGEN TRANSFER IN LABORATORY STIRRED TANK BIOREACTOR

Paulina Torres[1], Diego de la Vega[1], Marcel Otavio Cerri[2], J. Ricardo Pérez-Correa[1] and Eduardo Agosin[1].

[1] Department of Chemical and Bioprocess Engineering, School of Engineering, Pontificia Universidad Católica de Chile, Vicuña Mackenna 4860, Santiago, Chile

[2] Department of Bioprocesses and Biotechnology, Faculty of Pharmaceutical Sciences, Univ Estadual Paulista "Júlio de Mesquita Filho", Araraquara, SP, Brazil, CEP 14801-902.

Paper in preparation

5.1 INTRODUCTION

Oxygen is an essential substrate for aerobic bioprocesses. It allows the complete oxidation of the carbon source, maximizing the energy efficiency of the microorganism (Garcia-Ochoa & Gomez, 2009; Sonnleitner, 2016; Wang et al., 2009). Oxygen is commonly supplied from the gas phase, where it is transferred to the liquid and, once dissolved, it is consumed by the cells. Nevertheless, oxygen solubility is low and oxygen demand increases over time in high cell density cultures (Garcia-Ochoa et al, 2010). Oxygen limitation in microbial cultures promotes the formation of inhibitory by-products, which prevents reaching the maximum permissible biomass (Adelantado et al., 2017; Baumann et al., 2010; Enfors, 2011; Garcia-Ortega et al., 2017; Villadsen & Patil, 2007). In fact, the oxygen transfer rate (OTR) is considered one of the main limiting factors in cultures of different microorganisms, e.g. *Escherichia coli* (Henes & Sonnleitner, 2007), *Pichia pastoris* (Henes & Sonnleitner, 2007; Looser et al., 2015), and filamentous microorganisms, to name a few (Garcia-Ochoa et al., 2010). As a consequence, the bioreactor type and its mechanical design, together with an adequate control of dissolved oxygen in cell cultures is crucial to achieve high cell densities and volumetric productivities (Garcia-Ochoa et al., 2010; Yoshida, 2017).

OTR is typically described as the product of a volumetric mass transfer coefficient, k_La, and the concentration gradient, based on the two-film theory of Whitman (Whitman, 1924) (Equation 5-1).

$$OTR = \frac{dO_2}{dt} = k_L a \cdot (O_2{}^* - O_2) \qquad (5-1)$$

Where, O_2* is the oxygen equilibrium concentration and O_2 is the actual dissolved oxygen concentration in the liquid phase of the bioreactor.

OTR mainly depends on the k_La coefficient at a specific oxygen saturation condition (Equation 5-1) and for this reason, it is a crucial factor to characterize the oxygen transfer capability of gas-liquid bioreactors (Garcia-Ochoa & Gomez, 2009; Labík et al., 2017; Suresh et al., 2009). Actually, its measurement and/or prediction is used to design, operate and scale stirred tank bioreactors (Linek et al., 1993; Suresh et al., 2009).

Different strategies to increase OTR have been used (Garcia-Ochoa & Gomez, 2009; Maclennan & Pirt, 1966; Siegell & Gaden, 1962). Some researchers have modified the bioreactor configuration - impellers, sparger and geometrical dimensions - in order to increase k_La (Arjunwadkar et al., 1998; Buffo et al., 2016; Vitae, 2011; Yang et al., 2012); others have focused on increasing saturation oxygen concentrations (Equation 5-1) by increasing oxygen partial pressure, either enriching the gaseous phase with pure oxygen (Flickinger & Perlman, 1977; Nielsen et al., 2002; Yamada et al., 1978) or pressurizing the bioreactor (Onken & Liefke, 1989; Yang & Wang, 1992). However, when strategies using maximum capable saturation oxygen concentration are not sufficient to reach the desired biomass concentration, it is necessary to focus on k_La optimization. For this purpose, several impeller configurations have been evaluated. For example, Buffo et al., 2016 demonstrated that the combination of two types of impellers, *i.e.* Rushton (RT) and Ear-Elephant (EE), provided several benefits to OTR with respect to RT alone. In fact, EE generates a differential axial/radial flow – contrary to only radial flow for RT –, which results in a circular path that increases gas hold-up and enhance bubble fragmentation, resulting in higher mass transfer area. On the other hand, Gelves, 2013 demonstrated that the Triple Spin Diffuser (TSD) could increase up to 6-fold the OTR in a pilot stirred tank

bioreactor (0,18 m^3 working volume) thanks to the axial/radial forces generated, which improves air homogenization throughout the whole volume and allows to reduce poor mixed lower regions of the bioreactor. Nevertheless, the combinations of both tools, impeller and sparger configurations, have not been simultaneously evaluated, so far.

In this work, we have optimized the oxygen transfer rate in 1-L bioreactors, using different impeller-sparger configurations and compared them with conventional ones. Finally, we proposed a decision tree to select the best configurations under different culture conditions to maximize the oxygen transfer rate in bioreactors.

5.2 MATERIAL AND METHODS

5.2.1 Bioreactor configuration.

In-house built 1-L bioreactors with 0.6 L of operational volume, equipped with a condenser, two types of diffusers, a stirrer, and interchangeable dual-impeller configurations were employed throughout this work. Agitation was driven by brushless VEXTAs DC motors, AXH Series (Oriental Motor, Japan). For monitoring and control, a SIMATIC PCS7 distributed control system (Siemens, Germany) was used. DO concentration and temperature were measured with fast response probes Oxymax COS22D (< 60 s) (Endress+Hauser, Switzerland). Air and N_2 additions were measured and controlled with mass flowmeters (FMA-A2407, Omega Engineering, USA).

Each bioreactor consists of a double-jacketed cylindrical vessel with an internal diameter of 9 cm and a height of 20 cm. The vessel was equipped with four equally spaced vertical baffles of 0.9 cm width and 1 mm thickness. Two diffuser configurations (4.5 μm) were evaluated; a classic Simple Diffuser (SD) and a Triple Spin Diffuser (TSD). Both were located below the lowest impeller. In total, three types of impeller were tested: two three-blade Elephant Ear impellers (EE) operated in Down-Pumping mode (EEDP) and Up-Pumping mode (EEUP), respectively; and a Rushton Turbine (RT). With these three impellers, four different dual-impeller combinations were tested for both diffusers (Buffo *et al*; 2016): RT-RT, RT-EEDP, EEDP-RT and EEDP-EEUP. The diameter of all impellers was 3 cm and the latter were separated by 3 cm, with the lowest impeller situated 3 cm above the bottom of the vessel.

5.2.2 Fluids

Three liquid mixtures were employed to simulate the effect of the increasing biomass during the cultivation time course: distilled water (W: μ = 0.87 mPa·s) and two glycerol solutions: GS1, 30 % glycerol (v/v) with μ = 3.23 mPa·s; and GS2, 50 % glycerol (v/v), with μ = 6.60 mPa·s. Compressed air and 99.9 % nitrogen (Indura, Santiago, Chile) were used to determine k_La. Dynamic viscosities of the solutions were determined from rheograms obtained at 25ºC, using a digital, concentric-cylinder rheometer (DVIII Ultra, Brookfield Engineering Laboratories, MA, USA).

5.2.3 Experimental Design

In total, eight different dual impeller-diffuser configurations were tested, with gas flow rate, agitation and viscosity, as operational variables. To assess the effect of gas flow rate and agitation, a 2^2 full factorial design with a central point was carried out, following the methodology described in Montgomery & Runger, 2003. The experimental range of operational parameters was as follows: agitation, between 200 to 800 rpm, with three different levels (200, 500 and 800 rpm); and aeration between 0.5 to 2.0 VVM, also with three levels (0.5, 1.25 and 2.0 VVM) (Table 5-1). To assess the effect of the viscosity, three liquid mixtures were prepared: W, GS1 and GS2, as previously described. These three mixtures were selected as they mimic the viscosity of a fed-batch culture of *Pichia pastoris* from 0 to 170 OD, approximately. All the experimental combinations are summarized in Figure 5-1.

Table 5- 1 Original values and corresponding coded values of the experimental design.

Independent variables		Coded and natural values		
		-1	0	+1
Agitation (RPM)	X_1	200	500	800
Aeration (VVM)	X_2	0.5	1.25	2.0

This design required a total of fifteen experiments, which were carried out in triplicate, with a random order, and the central point was performed in quintuplicate. To estimate the response of the dependent variable, a second-degree polynomial (Equation 5-2) was estimated, using the software STATGRAPHICS Plus 5.1 (Statpoint Technologies, Inc., VA, USA).

$$Y = b_0 + b_1 \cdot X_1 + b_2 \cdot X_2 + b_{12} \cdot X_1 \cdot X_2 \qquad (5-2)$$

where Y is the predicted response; X_1 and X_2 are independent variables; b_0 is the mean value of the response; b_1 and b_2 are linear effects; and b_{12} is an interaction term.

5.2.4 Methods for estimation of k_La considering system delay

Experimental absorption curves were generated by the gassing-out method (Garcia-Ochoa & Gomez, 2009; Linek et al., 1987). The DO concentration in the fluid was measured continuously, after a step change in the inlet gas. The oxygen sensor was installed in the bioreactor at the same height of the bottom impeller and close to the vessel wall (1 cm). The sensor was calibrated and then stabilized for 4 hours under air saturation conditions. Desorption was carried out by bubbling pure nitrogen, and absorption by bubbling air; in both cases, the gas flow rate and the agitation were kept constant. Estimation of k_La considering system delay was quantified using an automated algorithm, described by Torres et al., 2016.

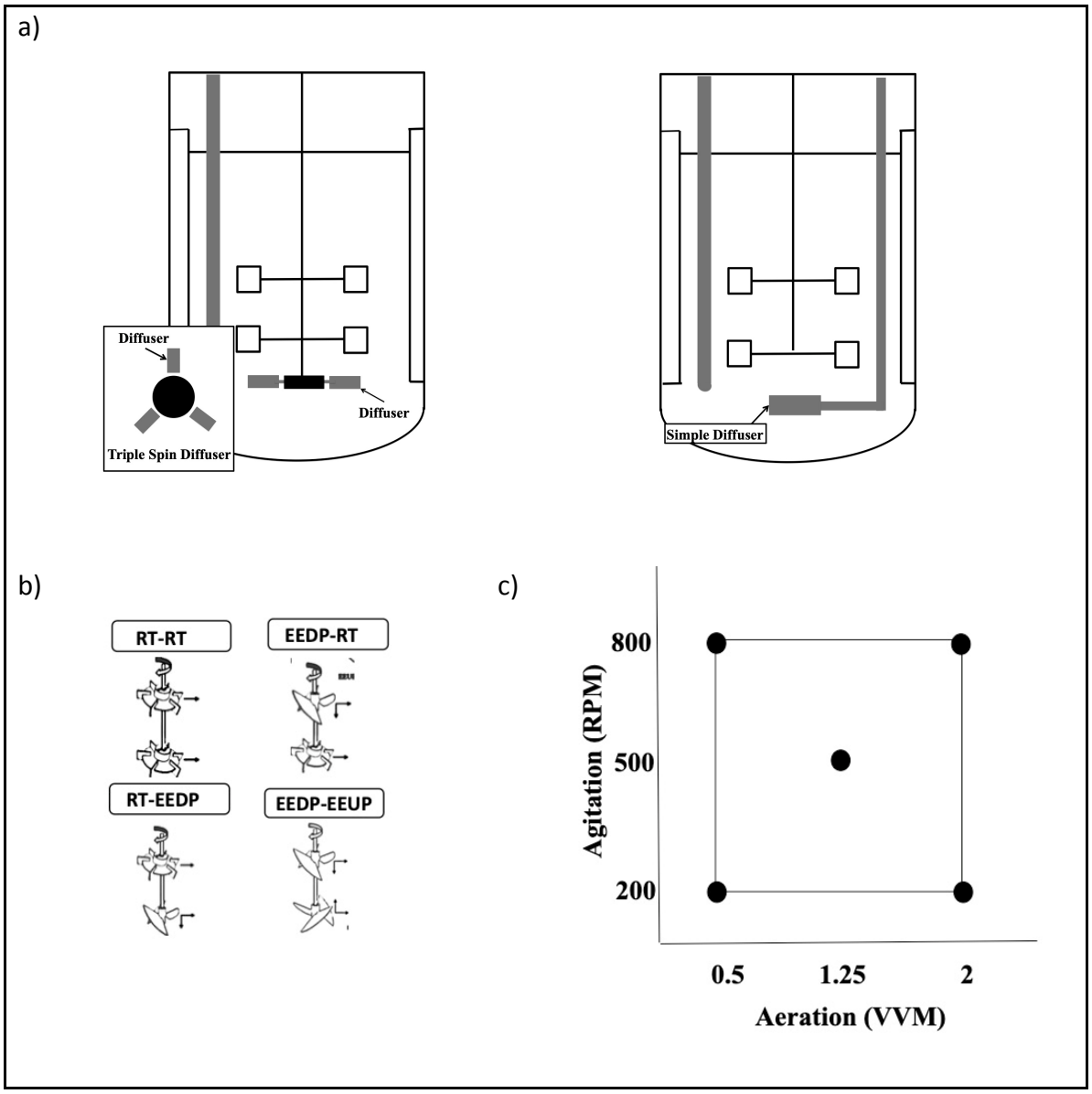


Figure 5-1 Summary of evaluated mechanical conditions - Difusser a) and impeller b), which were compared in different hydrodynamic conditions based on experimental design c).

5.2.5 Sensitivity analysis and decision tree.

In order to quantify the relative impact of agitation and aeration on the different dual-impeller-diffuser configurations, a sensitivity analysis was carried out in the central point of the experimental design, *i.e.* 500 rpm and 1.25 VVM. The relative sensitivity was estimated for agitation and aeration in the three fluids, with Equation 5-3:

$$g_i(k_{La}) = \frac{X_i}{k_{La}} \cdot \frac{dk_{La}}{dX_i} \qquad (5-3)$$

Where, g_i is the relative sensitivity and X_i is the normalized value of agitation or aeration.

Once all information was collected, a decision tree was constructed to select the best configuration, based on the viscosity and operational parameters' range. Since viscosity was the more influential parameter, it was the first criterion to consider. The second criterion dealt with the range of operational parameters (aeration and agitation). The third criterion was related with which of the operational parameters was limiting.

In the case that none of the operational parameters was limiting, the best configuration was the one achieving the highest k_La. If one or both parameters were limiting, the decision to select the best configuration considers the central point k_La value and the sensitivity analysis.

5.3 RESULTS AND DISCUSSIONS

All tested configurations showed reasonable values for the oxygen transfer coefficient (k_La) in a stirred tank reactor, ranging from 0.004 s^{-1} to 0.085 s^{-1} (Buffo et al., 2016; Bustamante et al, 2013; Cárcamo et al., 2014; Moucha et al, 2012) (Figure *5-2*). Our results confirmed previous reports (Buffo et al., 2016; Garcia-Ochoa & Gomez, 2009; Gelves, 2013) on the critical influence of hydrodynamics and mechanical conditions on k_La (Figure 5-2).

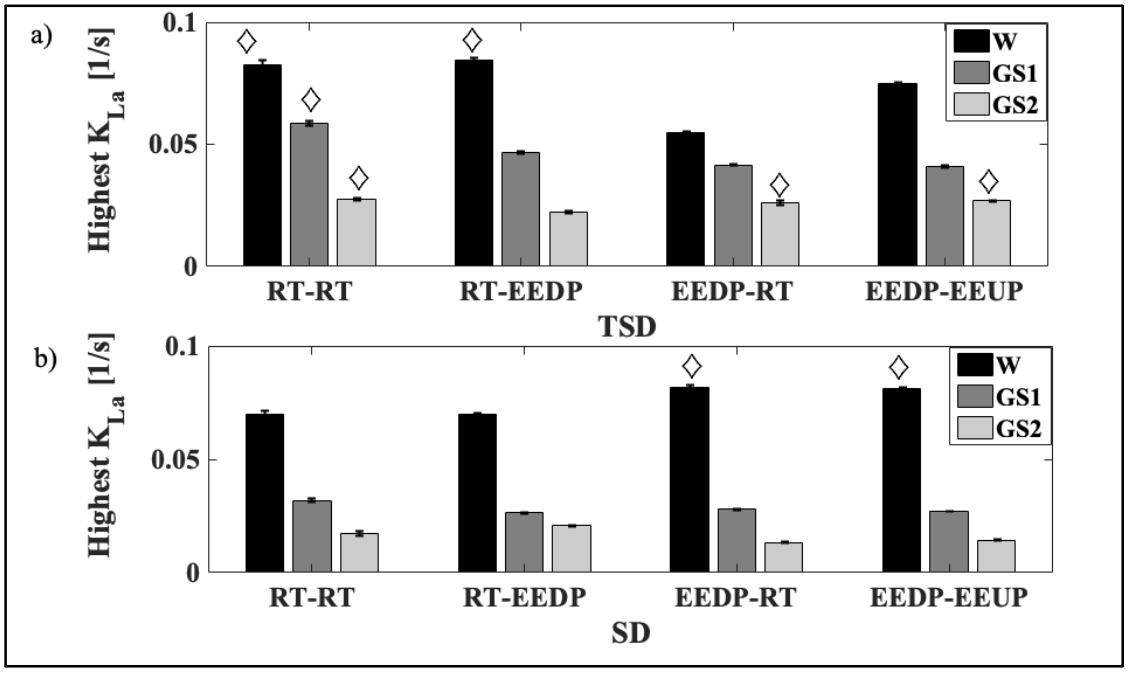


Figure 5-2. Impact of viscosity on the k_La reached at 800 RPM and 2 VVM according to the impeller-sparger configurations –TSD (a) and SD (b)-. Three increasing viscosities were evaluated: 0.87 cP (water, black), 3.23 cP (GS1, dark gray) and 6.60 cP (GS2, light gray).

◊, was the best configurations in relation with viscosity of the culture medium.

5.3.1 Effect of diffusers on k_{La}.

All these experiments were conducted with an RT-RT impeller configuration. Triple Spin Diffuser (TSD) increased k_La up to 8-fold in 1-L bioreactor, depending on the operational conditions and viscosity (Figure 5-3). TSD was particularly advantageous compared to a Simple Diffuser (SD) for slow agitation and high viscosity systems, suggesting that axial/radial flow could counteract the fluid resistance to gas dispersion and homogenization. Nevertheless, k_La increased less than 2 fold in a 1-L bioreactor with TSD at high agitation and aeration rates. Under these conditions, the axial force is counterbalanced by the corresponding radial force, reducing the differences of both diffusers (Figure 5-3). Poorly mixed regions were significantly reduced with TSD in the 1L bioreactor, as suggested by preliminary results of *E. coli* fermentations, where the formation of secondary products, such as acetate or ethanol, were reduced by 68% (data not shown).

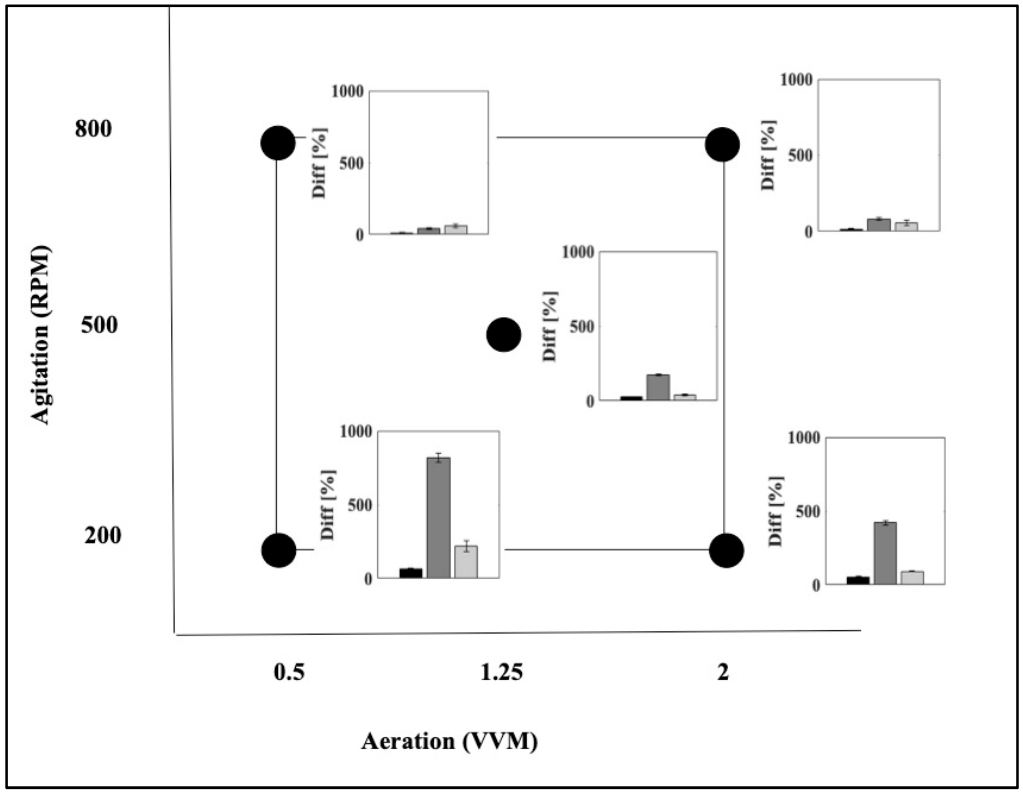


Figure 5-3 Impact of TSD sparger configuration using RT-RT impeller on experimental k_La values reached within the experimental range of operational parameters. Three increasing viscosities were evaluated: 0.87 cP (water, black), 3.23 cP (GS1, dark gray) and 6.60 cP (GS2, light gray).

5.3.2 Effect of impellers on k_La.

The impact of impellers on k_La was strongly dependent on the viscosity and operational parameters' range when a Simple Diffuser (SD) sparger was employed (Figure *5-4* a). Conventional impeller configuration (RT-RT) was the most efficient for average viscosity cultures (Figure 4a). For high viscosity cultures, we confirmed the data from Buffo et al, 2016, about the advantage of RT-EEDP configuration on k_La. However, we could not reproduce the positive impact of EEDP-EEUP reported in the same study for high agitation (800 rpm) and aeration levels (2VVM). We found the highest k_La for 500 rpm and 1.25 VVM operation, the central point of our experimental design (Figure 2). Nevertheless, EEDP-EEUP configuration might be optimal for a system operating with different

operational parameters than ours, *e.g.* agitation up to 1000 rpm and aeration lower than 1.25 VVM (Buffo et al., 2016; Bustamante et al., 2013; Zhu et al., 2009).

On the other hand, when a TSD sparger was employed with RT-RT, k_La was 40% higher than any other impeller configurations (Figure 4b). This could be attributed to the axial/radial flow resulting from TSD, since the circular path increased, decreasing the benefits of an EE impeller, like bubble fragmentation.

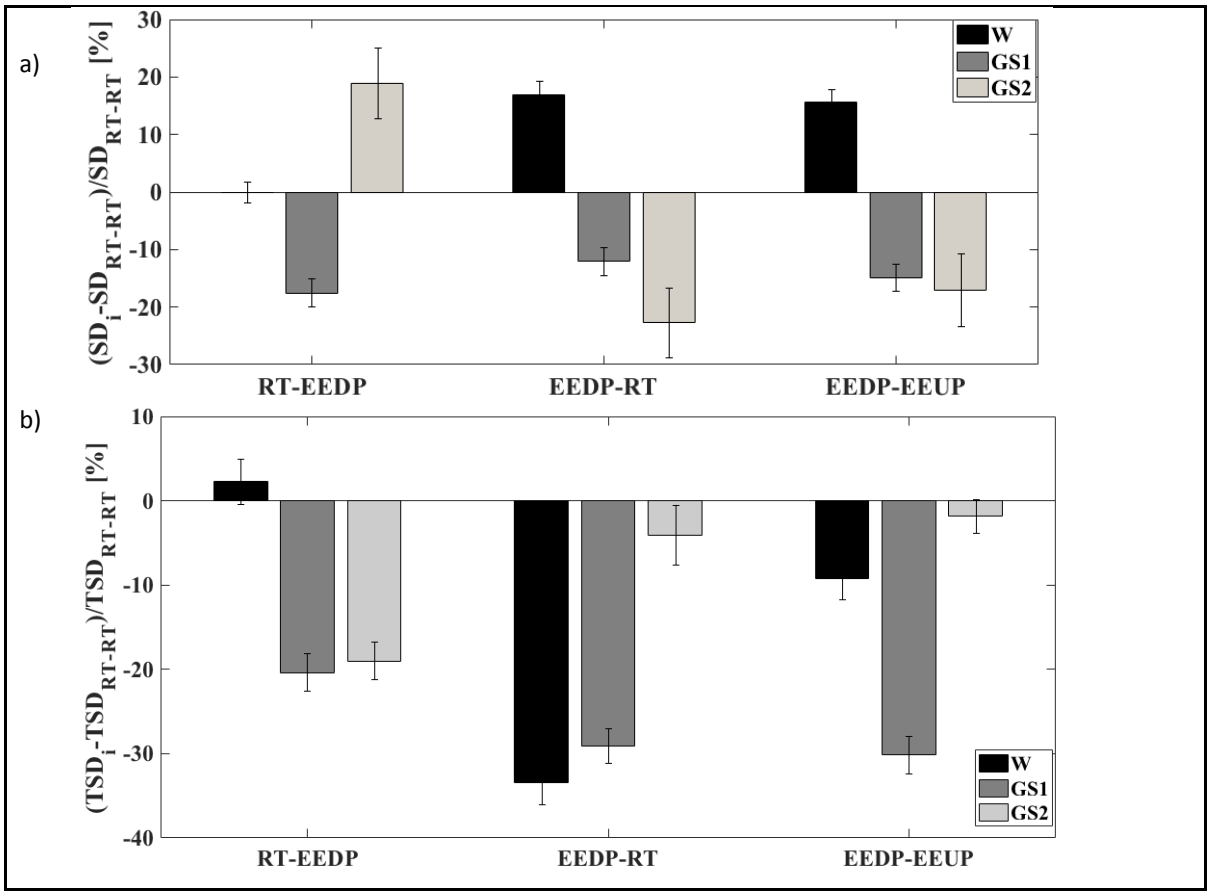


Figure 5-4 Impact of impeller configurations with Simple Diffuser (SD) a) and Triple Spin Diffuser (TSD) on the highest k_La value values reached at 800 RPM and 2 VVM. Differences (in %) were determined with RT-RT impeller in both cases. Three increasing viscosities were evaluated: 0.87 cP (water, black), 3.23 cP (GS1, dark gray) and 6.60 cP (GS2, light gray).

5.3.3 Effect of operational parameters on k_La.

The role of agitation and aeration on k_La of the different dual-impeller-diffuser configurations was assessed (see Supplementary material S5-1 for experimental design), and their relative importance was quantified by sensitivity analysis (Figure 5-5). Agitation was the most sensitive parameter affecting k_La for almost all impeller combinations and viscosities. However, if agitation is much higher than aeration, or vice versa, *kLa* enhancing strategies should consider only one of them. For example, under high viscosity conditions, TSD RT-RT will enhance OTR only with higher agitation; or in low viscosity cultures, TSD EEDP-RT will enhance OTR only with higher aeration (Figure 5-5). Since TSD is more sensitive to aeration than SD, it is more appropriate for cultures where agitation is limiting, like large-scale reactors.

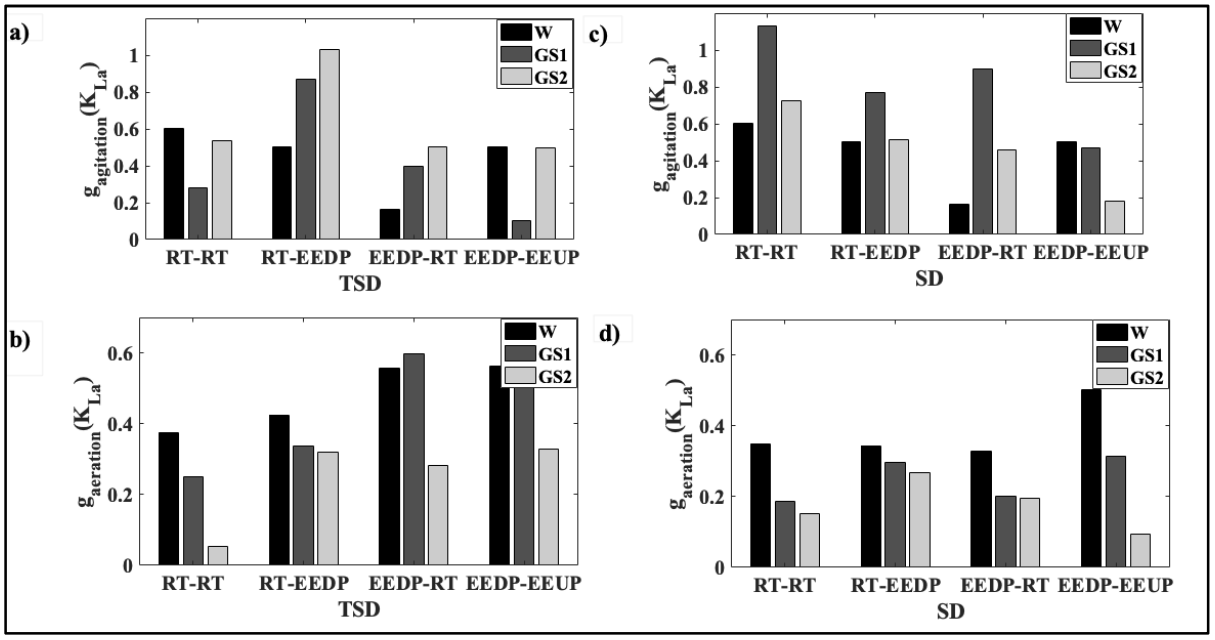


Figure 5-5 Sensitivity analysis for agitation and aeration in 500 rpm and 1.25 VVM. Three increasing viscosities were evaluated: 0.87 cP (water, black), 3.23 cP (GS1, dark gray) and 6.60 cP (GS2, light gray). a) Agitation sensitivity for all impeller configurations using Triple Spin Diffuser (TSD) b) Agitation sensitivity for all impeller configurations using Simple Diffuser (SD) c) Aeration sensitivity for all impeller configurations using TSD d) Aeration sensitivity for all impeller configurations using SD.

5.4 OUTLOOK

Oxygen transfer rate has been widely studied to improve high cell density cultures in bioreactors (Abbas et al, 2007; Abbott et al., 2013; Buffo et al., 2016; Garcia-Ochoa & Gomez, 2009). In this work, we showed that the optimal mechanical configuration, *i.e.* the best impeller-sparger combination to achieve the highest k_La, was strongly dependent on the viscosity and operational parameters' range.

Culture viscosity was the most influential parameter in the selection of the best possible configuration (Figure 5-2). Our results showed that TSD is more advantageous when high viscosities are reached by the culture medium; for low viscosities, however, some SD configurations have a similar performance (Figure 5-2). A decision tree is proposed for the selection of the optimal configuration, based on viscosity and operational parameters (Figure 5-6).

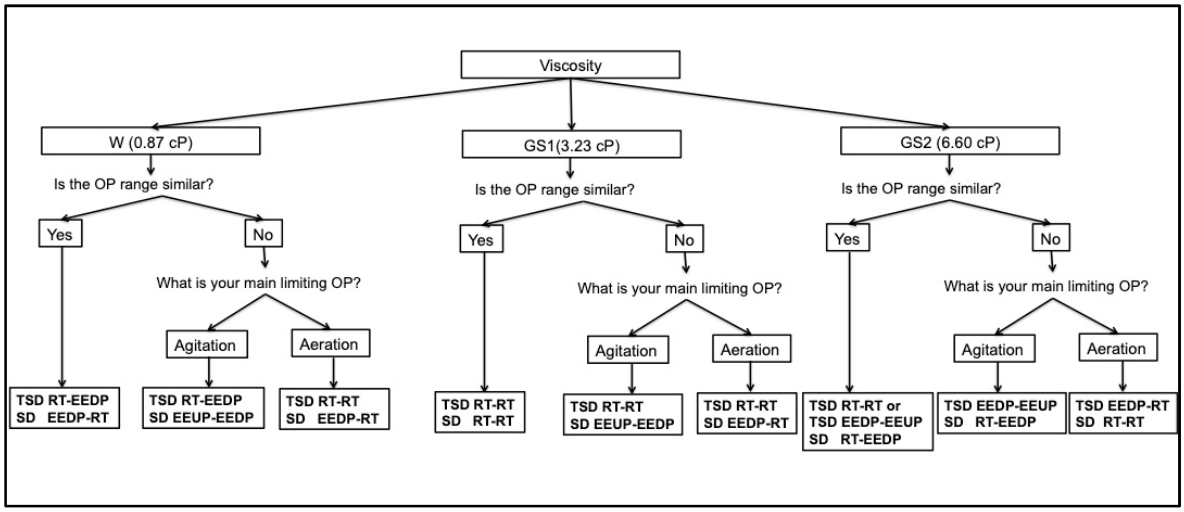


Figure 5-6 Decision tree to select best configuration based on viscosity and operational parameter (OP) range of agitation and aeration. OP range similar : 200 to 800 RPM and 0.5 to 2 VVM.

6. CONCLUSIONS AND FUTURE PERSPECTIVES

The results and knowledge obtained in this thesis can be used as a roadmap to improve recombinant protein production in *Pichia pastoris* through bioprocess engineering and metabolic modeling. Although we used a *P. pastoris* strain expressing constitutively thaumatin, a sweet protein, as a model, the developments achieved in this thesis are also applicable for the expression of other recombinant proteins of interest.

First, we developed an integrated framework for dissecting the metabolic effects of key operational parameters – μ and DO – in a recombinant thaumatin-producing *P. pastoris* strain. For this purpose, we implemented an experimental design to calculate the statistically significant effects from multiple chemostat data, which was later contextualized using a curated genome-scale metabolic model. Our results revealed a negative effect of the oxygenation on the specific product formation rate, and a positive effect on the biomass yield. Notably, we identified a novel synergistic effect of both parameters on the specific product formation rate. Finally, model predictions indicated an opposite relationship between the oxygenation level and growth-associated ATP requirement, suggesting tighter metabolic constraints under low oxygenation. However, our results also indicated that the extra metabolic capacity was not directly transferred to thaumatin production or associated to substantial a carbon redistribution, which suggests DO plays a more complex regulatory role likely related to changes in cell composition and/or transcriptional regulation (Gasser et al. 2007).

Furthermore, we discussed the importance of carefully analyzing the underlying intracellular flux distributions, which is essential to properly understand the metabolic mechanisms behind a determined phenotype. Nonetheless, this manually-curated version should be expanded to improve metabolic predictability of *P. pastoris* in other different carbon sources, like glycerol and methanol. In addition, future versions will need to take into account other processes, e.g. protein folding, secretion efficiencies and glycosylations.

Second, we assembled a robust dynamic genome-scale metabolic model for glucose-limited, aerobic cultivations of *P. pastoris* to simulated dynamic conditions – batch and fed-batch systems. The assembled platform can fit several datasets with minimum significance, sensitivity and identifiability problems in its parameters. Moreover,

if properly trained, it can be used to predict bioreactor dynamics and to determine metabolic and process engineering strategies to improve the production of a target compound.

Two limitations of our approach are worthy to mention. First, the model does not consider the calibration of gaseous mass balances of CO_2 and O_2. It predicts their specific production - and consumption – rates, using the genome-scale model instead. Despite the good predicting capability of the model for these compounds, it is desirable to include them in future versions of the platform, since CO_2 formation can adjust the flux through reactions that use or produce this compound and dissolved oxygen concentration has a strong impact on process performance, as demonstrated in the first part of this thesis. Moreover, in order to broaden its applications in other conditions relevant for *P. pastoris*, the model could be calibrated with other carbon sources and feeding strategies such as glycerol batch phase followed by methanol induction during the feed phase.

Finally, an optimization of Oxygen Transfer Rate (OTR) in 1-L reactors was carried out to improve high-density cultures in bioreactors. For this OTR's optimization, we firstly discussed the formulation of an automatic algorithm to estimate $k_L a$. Then, once the algorithm was validated, we presented a road map to optimize oxygen transfer rate in 1-L bioreactor using different impeller-sparger configurations. First, we characterized all configurations in different hydrodynamics conditions, and analyzed the relative importance of operational parameters. Our results showed that the optimal mechanical configuration, *i.e.* the best impeller-sparger combination to achieve the highest k_La, was strongly dependent on the viscosity and operational parameters' range. In fact, culture viscosity was the most influential parameter in the selection of the best possible configuration, where TSD was more advantageous. Then, a decision tree is proposed for the selection of the optimal configuration, based on viscosity and operational parameters. Nevertheless, it is a first approach in 1-L bioreactors and it is recommended to increase operational parameters' range, and analyze higher viscosities and Non-Newtonians fluids. Moreover, in order to broaden its applications in other reactors, this characterization could be carried out in other bioreactors, changing the geometry and scale.

Overall, this thesis attempts to analyze and understand metabolic effects using rational design of experiments and metabolic modeling. It provides consensus genome-

scale metabolic models for glucose-limited media in dynamic and continuous culture conditions of *Pichia pastoris.* Furthermore, we presented an automatic algorithm to estimated k_La values and a road map to optimize k_La in bioreactor. Finally, this systematic analysis in cooperation with an optimization of oxygen transfer in bioreactor could be used to improved recombinant protein production in *Pichia pastoris*.

7. REFERENCES

Abbas, S.; Marjan, S. . (2007). Recent advances in high cell density cultivation for production of recombinant protein, *31*(September 2016), 577–581.

Abbott, M. S. R., Harvey, A. P., Perez, G. V., & Theodorou, M. K. (2013). Biological processing in oscillatory baffled reactors: operation, advantages and potential. *Interface Focus*, *3*(1), 20120036. https://doi.org/10.1098/rsfs.2012.0036

Adelantado, N., Tarazona, P., Grillitsch, K., García-Ortega, X., Monforte, S., Valero, F., … Ferrer, P. (2017). The effect of hypoxia on the lipidome of recombinant *Pichia pastoris*. *Microbial Cell Factories*, *16*(1). https://doi.org/10.1186/s12934-017-0699-4

Adrio, J. L., & Demain, A. L. (2010). Recombinant organisms for production of industrial products. *Bioengineered Bugs*, *1*(2), 116–131. https://doi.org/10.4161/bbug.1.2.10484

Arjunwadkar, S. J., Saravanan, K., Pandit, A. B., & Kulkarni, P. R. (1998). Optimizing the impeller combination for maximum hold-up with minimum power consumption. *Biochemical Engineering Journal*, *1*(1), 25–30. https://doi.org/10.1016/S1369-703X(97)00005-3

Arjunwadkar, S. J., Sarvanan, K., Kulkarni, P. R., & Pandit, A. B. (1998). Gas-liquid mass transfer in dual impeller bioreactor. *Biochemical Engineering Journal*, *1*(2), 99–106. https://doi.org/10.1016/S1385-8947(97)00083-1

Arruda, A., Reis, V. C. B., Batista, V. D. F., Daher, B. S., Piva, L. C., De Marco, J. L., … Torres, F. A. G. (2016). A constitutive expression system for *Pichia pastoris* based on the PGK1 promoter. *Biotechnology Letters*, *38*(3), 509–517. https://doi.org/10.1007/s10529-015-2002-2

Asadollahi, M. a., Maury, J., Patil, K. R., Schalk, M., Clark, A., & Nielsen, J. (2009). Enhancing sesquiterpene production in Saccharomyces cerevisiae through in silico driven metabolic engineering. *Metabolic Engineering*, *11*(6), 328–334. https://doi.org/10.1016/j.ymben.2009.07.001

Assadi-Porter, F. M., Maillet, E. L., Radek, J. T., Quijada, J., Markley, J. L., & Max, M. (2010). Key amino acid residues involved in multi-point binding interactions between brazzein, a sweet protein, and the T1R2-T1R3 human sweet receptor. *Journal of Molecular Biology*, *398*(4), 584–599. https://doi.org/10.1016/j.jmb.2010.03.017

Auger, A. (2013). *Conversación personal*.

Aw, R., & Polizzi, K. M. (2016). Liquid PTVA: a faster and cheaper alternative for generating multi-copy clones in *Pichia pastoris*. *Microbial Cell Factories*, *15*(1), 29. https://doi.org/10.1186/s12934-016-0432-8

Badino, A. C., Cândida, M., Facciotti, R., & Schmidell, W. (2000). Improving k(L)a determination in fungal fermentation, taking into account electrode response time. *Journal of Chemical Technology and Biotechnology*, *75*(6), 469–474.

Balamurugan, V., Reddy, G. R., & Suryanarayana, V. V. S. (2007). *Pichia pastoris* : A notable heterologous expression system for the production of foreign proteins — Vaccines. *Indian Journal of Biotechnology*, *6*(April), 175–186.

Balsa-Canto, E., Rodriguez-Fernandez, M., & Banga, J. R. (2007). Optimal design of dynamic experiments for improved estimation of kinetic parameters of thermal degradation. *Journal of Food Engineering*, *82*(2), 178–188. https://doi.org/10.1016/j.jfoodeng.2007.02.006

Bartoszewski, G., Niedziela, a., Szwacka, M., & Niemirowicz-Szczytt, K. (2003). Modification of tomato taste in transgenic plants carrying a thaumatin gene from Thaumatococcus daniellii Benth. *Plant Breeding*, *122*(4), 347–351. https://doi.org/10.1046/j.1439-0523.2003.00864.x

, K., Carnicer, M., Dragosits, M., Graf, A. B., Stadlmann, J., Jouhten, P., … Ferrer, P. (2010). A multi-level study of recombinant *Pichia pastoris* in different oxygen conditions. *BMC Systems Biology*, *4*(1), 141. https://doi.org/10.1186/1752-0509-4-141

Baumann, K., Dato, L., Graf, A. B., Frascotti, G., Dragosits, M., Porro, D., … Branduardi, P. (2011). The impact of oxygen on the transcriptome of recombinant S. cerevisiae and *P. pastoris* - a comparative analysis. *BMC Genomics*, *12*(1), 218. https://doi.org/10.1186/1471-2164-12-218

Baumann, K., Maurer, M., Dragosits, M., Cos, O., Ferrer, P., & Mattanovich, D. (2008). Hypoxic fed-batch cultivation of *Pichia pastoris* increases specific and volumetric productivity of recombinant proteins. *Biotechnology and Bioengineering*, *100*(1), 177–183. https://doi.org/10.1002/bit.21763

BCC Research. (2014). Global Markets for Enzymes in Industrial Applications - Report Overview.

BCC Research. (2018). *Global markets for enzymes in industrial applications.*

Becker, S. a, Feist, A. M., Mo, M. L., Hannum, G., Palsson, B. Ø., & Herrgard, M. J. (2007). Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox. *Nature Protocols*, *2*(3), 727–738. https://doi.org/10.1038/nprot.2007.99

Behrens, M., Meyerhof, W., Hellfritsch, C., & Hofmann, T. (2011). Sweet and umami taste: natural products, their chemosensory targets, and beyond. *Angewandte Chemie (International Ed. in English)*, *50*(10), 2220–2242. https://doi.org/10.1002/anie.201002094

Bill, R. (2014). Playing catch-up with escherichia coli: Using yeast to increase success rates in recombinant protein production experiments. *Frontiers in Microbiology*, *5*(MAR), 1–5. https://doi.org/10.3389/fmicb.2014.00085

Boles, E., & Hollenberg, C. P. (1997). The molecular genetics of hexose transport in yeasts. *FEMS Microbiology Reviews*, *21*(1), 85–111. https://doi.org/10.1016/S0168-6445(97)00052-1

Bornstein, B. J., Keating, S. M., Jouraku, A., & Hucka, M. (2008). LibSBML: An API library for SBML. *Bioinformatics*, *24*(6), 880–881. https://doi.org/10.1093/bioinformatics/btn051

Buffo, M. M., Corrêa, L. J., Esperança, M. N., Cruz, A. J. G., Farinas, C. S., & Badino, A. C. (2016). Influence of dual-impeller type and configuration on oxygen transfer, power consumption, and shear rate in a stirred tank bioreactor. *Biochemical Engineering Journal*, *114*, 130–139.

Bustamante, M. C. C., Cerri, M. O., & Badino, A. C. (2013). Comparison between average shear rates in conventional bioreactor with Rushton and Elephant ear impellers. *Chemical Engineering Science*, *90*, 92–100. https://doi.org/10.1016/j.ces.2012.12.028

Caldwell, J. E., Abildgaard, F., Dzakula, Z., Ming, D., Hellekant, G., & Markley, J. L. (1998). Solution structure of the thermostable sweet-tasting protein brazzein. *Nature Structural Biology*, *5*(6), 427–431.

Çalık, P., Ata, Ö., Güneş, H., Massahi, A., Boy, E., Keskin, A., … Özdamar, T. H. (2015). Recombinant protein production in *Pichia pastoris* under glyceraldehyde-3-phosphate dehydrogenase promoter: From carbon source metabolism to bioreactor operation parameters. *Biochemical Engineering Journal*, *95*, 20–36. https://doi.org/10.1016/j.bej.2014.12.003

Cárcamo, M. (2013). *Producción de Proteínas Recombinantes en Cultivos Fed-Batch de Saccharomyces Cerevisiae y Escharichia Coli*. Pontificia Unversidad Católica de Chile.

Cárcamo, M., Saa, P. A., Torres, J., Torres, S., Mandujano, P., Correa, J. R. P., & Agosin, E. (2014). Effective Dissolved Oxygen Control Strategy for High-Cell-Density Cultures. *IEEE Latin America Transactions*, *12*(3), 389–394.

Carnicer, M., Baumann, K., Töplitz, I., Sánchez-Ferrando, F., Mattanovich, D., Ferrer, P., & Albiol, J. (2009). Macromolecular and elemental composition analysis and extracellular metabolite balances of *Pichia pastoris* growing at different oxygen levels. *Microbial Cell Factories*, *8*, 65. https://doi.org/10.1186/1475-2859-8-65

Carnicer, M., Ten Pierick, A., van Dam, J., Heijnen, J. J., Albiol, J., van Gulik, W., & Ferrer, P. (2012). Quantitative metabolomics analysis of amino acid metabolism in recombinant *Pichia pastoris* under different oxygen availability conditions. *Microb. Cell Fact.*, *11*(1), 83. https://doi.org/10.1186/1475-2859-11-83

Carter, P. J. (2011). Introduction to current and future protein therapeutics: a protein engineering perspective. *Experimental Cell Research*, *317*(9), 1261–1269. https://doi.org/10.1016/j.yexcr.2011.02.013

Caspeta, L., Shoaie, S., Agren, R., Nookaew, I., & Nielsen, J. (2012). Genome-scale metabolic reconstructions of Pichia stipitis and *Pichia pastoris* and in-silico evaluation of their potentials. *BMC Systems Biology*, *6*(1), 24. https://doi.org/10.1186/1752-0509-6-24

Cereghino, G. P. L., Cereghino, J. L., Ilgen, C., & Cregg, J. M. (2002). Production of recombinant proteins in fermenter cultures of the yeast *Pichia pastoris*. *Current Opinion in Biotechnology*, *13*, 329–332. https://doi.org/10.1016/S0958166902003300

Cereghino, J. L., & Cregg, J. M. (2000). Heterologous protein expression in the methylotrophic yeast *Pichia pastoris*. *FEMS Microbiology Reviews*, *24*(1), 45–66.

Cerri, M. O., Esperança, M. N., Badino, A. C., & Ribeiro, M. P. de A. (2016). A new approach for kLa determination by gassing-out method in pneumatic bioreactors. *Journal of Chemical Technology and Biotechnology*, *91*(12), 3061–3069.

Cha, H. J., Shin, H. S., Lim, H. J., Cho, H. S., Dalal, N. N., Pham, M. Q., & Bentley, W. E. (2005). Comparative production of human interleukin-2 fused with green fluorescent protein in several recombinant expression systems. *Biochemical Engineering Journal*, *24*(3), 225–233. https://doi.org/10.1016/j.bej.2005.03.002

Charoenrat, T., Ketudat-Cairns, M., Stendahl-Andersen, H., Jahic, M., & Enfors, S.-O. (2005). Oxygen-limited fed-batch process: an alternative control for *Pichia pastoris* recombinant protein processes. *Bioprocess and Biosystems Engineering*, *27*(6), 399–406. https://doi.org/10.1007/s00449-005-0005-4

Cheng, H., Lv, J., Wang, H., Wang, B., Li, Z., & Deng, Z. (2014). Genetically engineered *Pichia pastoris* yeast for conversion of glucose to xylitol by a single-fermentation process. *Applied Microbiology and Biotechnology*, *98*(8), 3539–3552. https://doi.org/10.1007/s00253-013-5501-x

Chung, B. K. S., Selvarasu, S., Andrea, C., Ryu, J., Lee, H., Ahn, J., … Lee, D. (2010). Genome-scale metabolic reconstruction and in silico analysis of methylotrophic yeast *Pichia pastoris* for strain improvement. *Microbial Cell Factories*, *9*, 1–15.

Ciofalo, V., Barton, N., Kreps, J., Coats, I., & Shanahan, D. (2006). Safety evaluation of a lipase enzyme preparation, expressed in *Pichia pastoris*, intended for use in the degumming of edible vegetable oil. *Regulatory Toxicology and Pharmacology*, *45*(1), 1–8. https://doi.org/10.1016/j.yrtph.2006.02.001

Čiplys, E., Žitkus, E., Gold, L. I., Daubriac, J., Pavlides, S. C., Højrup, P., … Slibinskas, R. (2015). High-level secretion of native recombinant human calreticulin in yeast. *Microbial Cell Factories*, *14*(1), 165. https://doi.org/10.1186/s12934-015-0356-8

Clasquin, M. F., Melamud, E., Singer, A., Gooding, J. R., Xu, X., Dong, A., … Caudy, A. A. (2011). Riboneogenesis in Yeast. *Cell*, *145*(6), 969–980. https://doi.org/10.1016/j.cell.2011.05.022

Corchero, J. L., Gasser, B., Resina, D., Smith, W., Parrilli, E., Vázquez, F., … Villaverde, A. (2013). Unconventional microbial systems for the cost-efficient production of high-quality protein therapeutics. *Biotechnology Advances*, *31*(2), 140–153. https://doi.org/10.1016/j.biotechadv.2012.09.001

Costa, N. R., Lourenço, J., & Pereira, Z. L. (2011). Desirability function approach: A review and performance evaluation in adverse conditions. *Chemometrics and Intelligent Laboratory Systems*, *107*(2), 234–244.

Cregg, J. M., Cereghino, J. L., Shi, J., & Higgins, D. R. (2000). Recombinant Protein Expression in *Pichia pastoris*, *16*.

Daly, R., & Hearn, M. T. W. (2005). Expression of heterologous proteins in *Pichia pastoris*: a useful experimental tool in protein engineering and production. *Journal of Molecular Recognition : JMR*, *18*(2), 119–138. https://doi.org/10.1002/jmr.687

Daniell, S., Mellits, K. H., Faus, I., & Connerton, I. (2000). Refolding the sweet-tasting protein thaumatin II from insoluble inclusion bodies synthesised in Escherichia coli. *Food Chemistry*, *71*, 105–110.

de Vos, A., Hatada, M., van der Wel, H., Krabbendam, H., Peerdeman, A., & Kim, S. H. (1985). Three-dimensional structure of thaumatin I , an intensely sweet protein. *PNAS Biochemistry*, *82*(March), 1406–1409.

Delic, M., Göngrich, R., Mattanovich, D., & Gasser, B. (2014). Engineering of protein folding and secretion-strategies to overcome bottlenecks for efficient production of recombinant proteins. *Antioxidants & Redox Signaling*, *21*(3), 414–437. https://doi.org/10.1089/ars.2014.5844

Delic, M., Rebnegger, C., Wanka, F., Puxbaum, V., Haberhauer-Troyer, C., Hann, S., … Gasser, B. (2012). Oxidative protein folding and unfolded protein response elicit differing redox regulation in endoplasmic reticulum and cytosol of yeast. *Free Radical Biology & Medicine*, *52*(9), 2000–2012. https://doi.org/10.1016/j.freeradbiomed.2012.02.048

Delic, M., Valli, M., Graf, A. B., Pfeffer, M., Mattanovich, D., & Gasser, B. (2013). The secretory pathway: Exploring yeast diversity. *FEMS Microbiology Reviews*, *37*(6), 872–914. https://doi.org/10.1111/1574-6976.12020

Demain, A. L., & Vaishnav, P. (2009). Production of recombinant proteins by microbes and higher organisms. *Biotechnology Advances*, *27*(3), 297–306. https://doi.org/10.1016/j.biotechadv.2009.01.008

Demain, A. L., & Vaishnav, P. (2011). *Production of Recombinant Proteins by Microbes and Higher Organisms*. *Comprehensive Biotechnology, Second Edition* (Second Edi, Vol. 3). Elsevier B.V. https://doi.org/10.1016/B978-0-08-088504-9.00542-0

Doehlert, D. H. (1970). Uniform Shell Designs. *Applied Statistics*, *19*(3), 231. https://doi.org/10.2307/2346327

Dolgov, S. V., Schestibratov, K. A., & Mikhailov, R. V. (2004). Apple transformation with the gene of supersweet protein thaumatin II. *Acta Hortic.*, *663*, 507–510.

Dragosits, M., Stadlmann, J., Albiol, J., Baumann, K., Maurer, M., Gasser, B., … Cerdanyola, B. (2009). The Effect of Temperature on the Proteome of Recombinant *Pichia pastoris* research articles. *Journal of Proteome Research*, (8), 1380–1392.

Edens, L., Bom, I., Ledeboer, A., & Maat, J. (1984). Synthesis and Processing of the Plant Protein Thaumatin in Yeast. *Cell*, *37*, 629–633. Retrieved from http://ukpmc.ac.uk/abstract/MED/6327079

Edens, L., Heslinga, L., Klolk, R., Ledeboer, A., Maat, J., Toonen, M. Y., … Verrips, C. T. (1982). Cloning of cDNA encoding the sweet-tasting plant protein thaumatin and its expression in Eschericha coli. *Gene*, *18*(1), 1–12.

Edens, L., & van der Wel, H. (1985). Microbial synthesis of the sweet-tasting plant protein thaumatin. *Cell*, *3*(3), 61–64.

Egea, J., & Balsa-Canto, E. (2009). Dynamic optimization of nonlinear processes with an enhanced scatter search method. *Industrial & Engineering Chemical Research*, *48*(9), 4388–4401.

Enfors, S.-O. (2011). *Fermentation Process Engineering*. Stockholm: School of Biotechnology Royal Institute of Technology.

FAO. (2018). FAO Statistics. Retrieved October 24, 2018, from http://faostat.fao.org/DesktopDefault.aspx?PageID=609&lang=es#ancor

Faus, I. (2000). Recent developments in the characterization and biotechnological production of sweet-tasting proteins. *Applied Microbiology and Biotechnology*, *53*(2), 145–151. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10709975

Faus, I., Patiño, C., Moral, C., Del Río, J. luis, Barroso, H. S., & Rubio, V. (1996). Expression of a Synthetic Gene Encoding the Sweet-Tasting Protein Thaumatin in Escherichia coli ´ Luis del Rı (upgrading or improving other flavours) of food ; in industry they are currently extracted from Ref . 1). Thaumatins can be isolated from thes. *Biochemical and Biophysical Research Communications*, *127*, 121–127.

Feng, X., Xu, Y., Chen, Y., & Tang, Y. J. (2012). Integrating flux balance analysis into kinetic models to decipher the dynamic metabolism of shewanella oneidensis MR-1. *PLoS Computational Biology*, *8*(2). https://doi.org/10.1371/journal.pcbi.1002376

Ferrer-Miralles, N., Domingo-Espín, J., Corchero, J. L., Vázquez, E., & Villaverde, A. (2009). Microbial factories for recombinant pharmaceuticals. *Microbial Cell Factories*, *8*, 17. https://doi.org/10.1186/1475-2859-8-17

Firsov, a. P., Pushin, a. S., Korneeva, I. V., & Dolgov, S. V. (2012). Transgenic tomato plants as supersweet protein thaumatin II producers. *Applied Biochemistry and Microbiology*, *48*(9), 746–751. https://doi.org/10.1134/S0003683812090025

Flickinger, M. C., & Perlman, D. (1977). Application of Oxygen-Enriched Aeration in the Conversion of Glycerol to Dihydroxyacetone by Gluconobacter melanogenus IFO 3293. *Applied and Environmental Microbiology*, *33*(3), 706–712.

Garcia-Ochoa, F., & Gomez, E. (2009). Bioreactor scale-up and oxygen transfer rate in

microbial processes: an overview. *Biotechnology Advances*, *27*(2), 153–176.

Garcia-Ochoa, F., & Gomez, E. (2010). Oxygen Transfer Rate Determination: Chemical, Physical and Biological Methods. In *Encyclopedia of Industrial Biotechnology*. New York, NY: Wiley.

Garcia-Ochoa, F., Gomez, E., Santos, V. E., & Merchuk, J. C. (2010). Oxygen uptake rate in microbial processes: An overview. *Biochemical Engineering Journal*, *49*(3), 289–307.

Garcia-Ortega, X. (2016). *A step forward in recombinant protein production regulated by the constitutive GAP promoter in Pichia pastoris through bioprocess engineering approaches.* Universitat Autonoma de Barcelona.

Garcia-Ortega, X., Adelantado, N., Ferrer, P., Montesinos, J. L., & Valero, F. (2015). A step forward to improve recombinant protein production in *Pichia pastoris*: From specific growth rate effect on protein secretion to carbon-starving conditions as advanced strategy. *Process Biochemistry*, *51*(6), 681–691. https://doi.org/10.1016/j.procbio.2016.02.018

Garcia-Ortega, X., Valero, F., & Montesinos-Seguí, J. L. (2017). Physiological state as transferable operating criterion to improve recombinant protein production in *Pichia pastoris* through oxygen limitation. *Journal of Chemical Technology & Biotechnology*, (March). https://doi.org/10.1002/jctb.5272

Gasser, B., Maurer, M., Rautio, J., Sauer, M., Bhattacharyya, A., Saloheimo, M., … Mattanovich, D. (2007). Monitoring of transcriptional regulation in *Pichia pastoris* under protein production conditions. *BMC Genomics*, *8*, 179. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1919374&tool=pmcentrez&rendertype=abstract

Gasser, B., Prielhofer, R., Marx, H., Maurer, M., Nocon, J., Steiger, M., … Mattanovich, D. (2013). *Pichia pastoris* : protein production host and model organism for biomedical research. *Future Microbiology*, *8*(2), 191–208.

Gasser, B., Saloheimo, M., Rinas, U., Dragosits, M., Rodríguez-Carmona, E., Baumann, K., … Villaverde, A. (2008). Protein folding and conformational stress in microbial cells producing recombinant proteins: a host comparative overview. *Microbial Cell Factories*, *7*, 11. https://doi.org/10.1186/1475-2859-7-11

Gelves, G. (2013). *Modelado computacional de la hidrodinámica en un bioreactor tipo planta piloto para fermentaciones aerobias*. *Universidad de antioquia*. https://doi.org/10.1007/s13398-014-0173-7.2

Gourich, B., Vial, C., El Azher, N., Belhaj Soulami, M., & Ziyad, M. (2008). Influence of hydrodynamics and probe response on oxygen mass transfer measurements in a high aspect ratio bubble column reactor: Effect of the coalescence behaviour of the liquid phase. *Biochemical Engineering Journal*, *39*(1), 1–14.

Graf, A., Dragosits, M., Gasser, B., & Mattanovich, D. (2009). Yeast systems biotechnology for the production of heterologous proteins. *FEMS Yeast Research*, *9*(3), 335–348. https://doi.org/10.1111/j.1567-1364.2009.00507.x

Graf, A., Gasser, B., Dragosits, M., Sauer, M., Leparc, G. G., Tüchler, T., … Mattanovich, D. (2008). Novel insights into the unfolded protein response using *Pichia pastoris* specific DNA microarrays. *BMC Genomics*, *9*, 390. https://doi.org/10.1186/1471-2164-9-390

Guan, R. J., Zheng, J. M., Hu, Z., & Wang, D. C. (2000). Crystallization and preliminary

X-ray analysis of the thermostable sweet protein mabinlin II. *Acta Crystallographica Section D: Biological Crystallography*, *56*(7), 918–919. https://doi.org/10.1107/S0907444900005850

Gurobi Optimization, I. (2016). Gurobi Optimizer Reference Manual. Retrieved from http://www.gurobi.com

Hamilton, J. D. (1994). *Time Series Analysis*. *Book* (Vol. 39). New Jersey, NJ: Princeton University.

Hasslacher, M., Schall, M., Hayn, M., Bona, R., Rumbold, K., Lückl, J., … Schwab, H. (1997). High-Level Intracellular Expression of Hydroxynitrile Lyase from the Tropical Rubber TreeHevea brasiliensisin Microbial Hosts. *Protein Expression and Purification*, *11*(1), 61–71. https://doi.org/http://dx.doi.org/10.1006/prep.1997.0765

Henes, B., & Sonnleitner, B. (2007). Controlled fed-batch by tracking the maximal culture capacity. *Journal of Biotechnology*, *132*(2), 118–126. https://doi.org/10.1016/j.jbiotec.2007.04.021

Hensing, M. C. M., Rouwenhorst, R. J., Heijnen, J. J., van Dijken, J. P., & Pronk, J. T. (1995). Physiological and technological aspects of large-scale heterologous-protein production with yeasts. *Antonie van Leeuwenhoek*, *67*(3), 261–279. https://doi.org/10.1007/BF00873690

Heyland, J., Fu, J., Blank, L. M., & Schmid, A. (2010). Quantitative physiology of *Pichia pastoris* during glucose-limited high-cell density fed-batch cultivation for recombinant protein production. *Biotechnology and Bioengineering*, *107*(2), 357–368. https://doi.org/10.1002/bit.22836

Heyland, J., Fu, J., Blank, L. M., & Schmid, A. (2011). Carbon metabolism limits recombinant protein production in *Pichia pastoris*. *Biotechnology and Bioengineering*, *108*(8), 1942–1953. https://doi.org/10.1002/bit.23114

Höffner, K., Harwood, S. M., & Barton, P. I. (2013). A reliable simulator for dynamic flux balance analysis. *Biotechnology and Bioengineering*, *110*(3), 792–802. https://doi.org/10.1002/bit.24748

Holzhütter, H. G. (2004). The principle of flux minimization and its application to estimate stationary fluxes in metabolic networks. *European Journal of Biochemistry*, *271*(14), 2905–2922. https://doi.org/10.1111/j.1432-1033.2004.04213.x

Houdebine, L. M. (2009). Production of pharmaceutical proteins by transgenic animals. *Comparative Immunology, Microbiology and Infectious Diseases*, *32*(2), 107–121. https://doi.org/10.1016/j.cimid.2007.11.005

Hyduke, D., Schellenberger, J., Que, R., Fleming, R., Thiele, I., Orth, J., … Palsson, B. (2011). COBRA Toolbox 2.0. *Protocol Exchange*. https://doi.org/10.1038/protex.2011.234

IANSA. (2018). *Memoria Anual de IANSA 2017*.

Idiris, A., Tohda, H., & Kumagai, H. (2010). Engineering of protein secretion in yeast : strategies and impact on protein production. *Applied Microbiology and Biotechnology*, *86*, 403–417.

Igeta, H., Tamura, Y., Nakaya, K., Nakamura, Y., & Kurihara, Y. (1991). Determination of disulfide array and subunit structure of taste-modifying protein, miraculin. *Biochimica et Biophysica Acta (BBA)/Protein Structure and Molecular*, *1079*(3), 303–307. https://doi.org/10.1016/0167-4838(91)90073-9

Iilingworth, C., Larson, G., & Heliekant, G. (1988). Secretion of the sweet-tasting plant

protein thaumatin by Bacillus subtilis. *Biotechnology Letters*, *10*(8), 587–592.

Illingworth, C., Larson, G., & Hellekant, G. (1989). Secretion of the sweet-tasting plant protein thaumatin by Streptomyces lividans. *Journal of Industrial Microbiology*, *4*(1), 37–42. https://doi.org/10.1007/BF01569691

Irani, Z. A., Kerkhoven, E. J., Shojaosadati, S. A., & Nielsen, J. (2016). Genome-scale metabolic model of *Pichia pastoris* with native and humanized glycosylation of recombinant proteins. *Biotechnology and Bioengineering*, *113*(5), 961–969. https://doi.org/10.1002/bit.25863

ISO. (2012). *Edulcorantes alternativos en un contexto de altos precios del azúcar*.

Ito, K., Asakura, T., Morita, Y., Nakajima, K., Koizumi, A., Shimizu-Ibuka, A., … Abe, K. (2007). Microbial production of sensory-active miraculin. *Biochemical and Biophysical Research Communications*, *360*(2), 407–411. https://doi.org/10.1016/j.bbrc.2007.06.064

Jaqaman, K., & Danuser, G. (2006). Linking data to models: data regression. *Nat Rev Mol Cell Biol*, *7*(11), 813–819.

Jordà, J., Jouhten, P., Cámara, E., Maaheimo, H., Albiol, J., & Ferrer, P. (2012). Metabolic flux profiling of recombinant protein secreting *Pichia pastoris* growing on glucose:methanol mixtures. *Microbial Cell Factories*, *11*, 57. https://doi.org/10.1186/1475-2859-11-57

Jozala, A. F., Geraldes, D. C., Tundisi, L. L., Feitosa, V. de A., Breyer, C. A., Cardoso, S. L., … Pessoa, A. (2016). Biopharmaceuticals from microorganisms: from production to purification. *Brazilian Journal of Microbiology*, *47*, 51–63. https://doi.org/10.1016/j.bjm.2016.10.007

Kaneko, R., & Kitabatake, N. (1999). Heat-induced formation of intermolecular disulfide linkages between thaumatin molecules that do not contain cysteine residues. *Journal of Agricultural and Food Chemistry*, *47*(12), 4950–4955. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10606557

Kaneko, R., & Kitabatake, N. (2001). Sweetness of Sweet Protein Thaumatin is more termoresistant under Acid Conditions than under neutral or alkaline condition. *Biosci. Biotech. Biochem.*, *65*(2), 409–413.

Kant, R. (2005). Sweet proteins-Potential replacement for artificial low calorie sweeteners. *Nutrition Journal*, *4*, 5. https://doi.org/10.1186/1475-2891-4-5

Kapic, A., & Heindel, T. J. (2006). Correlating Gas-Liquid Mass Transfer in a Stirred-Tank Reactor. *Chemical Engineering Research and Design*, *84*(3), 239–245.

Keating, S. M., Bornstein, B. J., Finney, A., & Hucka, M. (2006). SBMLToolbox: An SBML toolbox for MATLAB users. *Bioinformatics*, *22*(10), 1275–1277. https://doi.org/10.1093/bioinformatics/btl111

Kim, S. H., Kang, C. H., Kim, R., Cho, J. M., Lee, Y. B., & Lee, T. K. (1989). Redesigning a sweet protein: Increased stability and renaturability. *Protein Engineering, Design and Selection*, *2*(8), 571–575. https://doi.org/10.1093/protein/2.8.571

Kim, T. Y., Sohn, S. B., Kim, Y. Bin, Kim, W. J., & Lee, S. Y. (2012). Recent advances in reconstruction and applications of genome-scale metabolic models. *Current Opinion in Biotechnology*, *23*(4), 617–623. https://doi.org/10.1016/j.copbio.2011.10.007

Kitano, H. (2002). Systems biology: a brief overview. *Science (New York, N.Y.)*, *295*(5560), 1662–1664. https://doi.org/10.1126/science.1069492

Krivoruchko, A., Zhang, Y., Siewers, V., Chen, Y., & Nielsen, J. (2014). Microbial acetyl-CoA metabolism and metabolic engineering Microbial acetyl-CoA metabolism and metabolic engineering. *Metabolic Engineering*, *28*(December), 28–42. https://doi.org/10.1016/j.ymben.2014.11.009

Labík, L., Moucha, T., Petříček, R., Rejl, J. F., Valenz, L., & Haidl, J. (2017). Volumetric mass transfer coefficient in viscous liquid in mechanically agitated fermenters. Measurement and correlation. *Chemical Engineering Science*. https://doi.org/10.1016/j.ces.2017.04.006

Lagassé, H. A. D., Alexaki, A., Simhadri, V. L., Katagiri, N. H., Jankowski, W., Sauna, Z. E., & Kimchi-Sarfaty, C. (2017). Recent advances in (therapeutic protein) drug development. *F1000Research*, *6*, 113. https://doi.org/10.12688/f1000research.9970.1

Landaw, E. M., & DiStefano 3rd, J. J. (1984). Multiexponential, multicompartmental, and noncompartmental modeling. II. Data analysis and statistical considerations. *American Journal of Physiology - Regulatory, Integrative and Comparative Physiology*, *246*(5).

Landi, C., Paciello, L., De Alteriis, E., Brambilla, L., & Parascandola, P. (2015). High cell density culture with S. cerevisiae CEN.PK113-5D for IL-1β production: Optimization, modeling, and physiological aspects. *Bioprocess and Biosystems Engineering*, *38*(2), 251–261. https://doi.org/10.1007/s00449-014-1264-8

Lebedev, V. G. ., Taran, S. A. ., Shmatchenko, V. V. ., & Dolgov, S. V. (2002). Pear transformation with the gene for super- sweet protein thaumatin II. *Acta Hortic.*, *596*, 199–202.

Lee, J. H., Weickmann, J. L., Koduri, R. K., Ghosh-Dastidar, P., Saito, K., Blair, L. C., … Kendall, R. L. (1988). Expression of synthetic thaumatin genes in yeast. *Biochemistry*, *27*(14), 5101–5107.

Linek, V., Benes, P., Sinkule, J., & Moucha, T. (1993). Non-ideal pressure step method for kLa measurement. *Chemical Engineering Science*, *48*(9), 1593–1599.

Linek, V., Vacek, V., & Benes, P. (1987). A critical review and experimental verification of the correct use of the dynamic method for the determination of oxygen transfer in aerated agitated vessels to water, electrolyte solutions and viscous liquids. *The Chemical Engineering Journal*, *34*(1), 11–34.

Looser, V., Bruhlmann, B., Bumbak, F., Stenger, C., Costa, M., Camattari, A., … Kovar, K. (2015). Cultivation strategies to enhance productivity of *Pichia pastoris*: A review. *Biotechnology Advances*, *33*(6), 1177–1193.

Lopes, M., Mota, M., & Belo, I. (2013). Comparison of Yarrowia lipolytica and *Pichia pastoris* cellular response to different agents of oxidative stress. *Applied Biochemistry and Biotechnology*, *170*(2), 448–458. https://doi.org/10.1007/s12010-013-0205-3

Love, K. R., Politano, T. J., Panagiotou, V., Jiang, B., Stadheim, T. A., & Love, C. (2012). Systematic Single-Cell Analysis of *Pichia pastoris* Reveals Secretory Capacity Limits Productivity. *PloS One*, *7*(6), 1–11. https://doi.org/10.1371/journal.pone.0037915

Ma, J., Drake, P., & Christou, P. (2003). The production of recombinant pharmaceutical proteins in plants. *Nature Reviews Genetics*, *4*(10), 794–805. https://doi.org/10.1038/nrg1177

Maccani, A., Landes, N., Stadlmayr, G., Maresch, D., Leitner, C., Maurer, M., … Mattanovich, D. (2014). *Pichia pastoris* secretes recombinant proteins less efficiently than Chinese hamster ovary cells but allows higher space-time yields for less complex

proteins. *Biotechnology Journal*, *9*(4), 526–537. https://doi.org/10.1002/biot.201300305

Maclennan, D. G., & Pirt, S. J. (1966). Automatic Control of Dissolved Oxygen Concentration in Stirred Microbial Cultures. *Journal of General Microbiology*, *45*(2), 289–302. https://doi.org/10.1099/00221287-45-2-289

Mahadevan, R., Edwards, J. S., & Doyle, F. J. (2002). Dynamic flux balance analysis of diauxic growth in Escherichia coli. *Biophysical Journal*, *83*(3), 1331–1340. https://doi.org/10.1016/S0006-3495(02)73903-9

Markets and Markets. (2015). Industrial Enzymes Market by Type (Carbohydrases, Proteases, Non-starch Polysaccharides & Others), Application (Food & Beverage, Cleaning Agents, Animal Feed & Others), Brands & by Region - Globlar Trends and Forecasts to 2020.

Marx, H., Mecklenbr, A., Gasser, B., Sauer, M., & Mattanovich, D. (2009). Directed gene copy number amplication in *Pichia pastoris* by vector integration into the ribosomal DNA locus, *9*, 1260–1270. https://doi.org/10.1111/j.1567-1364.2009.00561.x

Masuda, T., Ide, N., Ohta, K., & Kitabatake, N. (2010). High-yield Secretion of the Recombinant Sweet-Tasting Protein Thaumatin I. *Food Science and Technology Research*, *16*(6), 585–592. https://doi.org/10.3136/fstr.16.585

Masuda, T., Ohta, K., Ojiro, N., Murata, K., Mikami, B., Tani, F., … Kitabatake, N. (2016). A Hypersweet Protein: Removal of The Specific Negative Charge at Asp21 Enhances Thaumatin Sweetness. *Scientific Reports*, *6*(October 2015), 20255. https://doi.org/10.1038/srep20255

Masuda, T., Tamaki, S., Kaneko, R., Wada, R., Fujita, Y., Mehta, A., & Kitabatake, N. (2004). Cloning, expression and characterization of recombinant sweet-protein thaumatin II using the methylotrophic yeast *Pichia pastoris*. *Biotechnology and Bioengineering*, *85*(7), 761–769. https://doi.org/10.1002/bit.10786

Mattanovich, D., Graf, A., Stadlmann, J., Dragosits, M., Redl, A., Maurer, M., … Gasser, B. (2009). Genome, secretome and glucose transport highlight unique features of the protein production host *Pichia pastoris*. *Microbial Cell Factories*, *8*, 29. https://doi.org/10.1186/1475-2859-8-29

Mattanovich, D., Jungo, C., Wenger, J., Dabros, M., & Maurer, M. (2014). 2 Yeast Suspension Culture. *Industrial Scale Suspension Culture of Living Cells*, 94–129.

Maurer, M., Kühleitner, M., Gasser, B., & Mattanovich, D. (2006). Versatile modeling and optimization of fed batch processes for the production of secreted heterologous proteins with *Pichia pastoris*. *Microbial Cell Factories*, *5*, 37. https://doi.org/10.1186/1475-2859-5-37

Merchuk, J. C., Yona, S., Siegel, M. H., & Ben Zvi, A. (1990). On the First-Order Approximation to the Response of Dissolved Oxygen Electrodes for Dynamic K,a Estimation. *Biotechnology and Bioengineering*, *34*(11), 1391–1392.

Ming, D., & Hellekant, G. (1994). Brazzein , a new high-potency thermostable sweet protein from Pentadiplandra brazzeana B. *FEBS Letters*, *355*, 106–108.

MINSAL. (2010). *Encuesta nacional de salud de Chile 2009-2010*. Santiago, Chile.

MINSAL. (2017a). *Encuesta Nacional de Salud 2016-2017 Primeros resultados*. *Departamento de Epidemiología, División de Planificación Sanitaria, Subsecretaría de Salud Pública*. https://doi.org/10.1139/O05-159

MINSAL. (2017b). *Política Nacional De Alimentación Y Nutrición*. *Politica Nacional De*

Alimentacion Y Nutricion. https://doi.org/A-286132

Monellin, C., Morris, J. a, Martenson, R., & Cagan, R. H. (1973). ARTICLE : That Tastes Sweet Characterization Tastes Sweet * of Monellin , a Protein That, *248*(2).

Montgomery, D. C., & Runger, G. C. (2003). Design of Experiments with Several Factors. In *Applied Statistics and Prbability for Engineers* (third, pp. 505–567). United States of America: John Wiley & Sons.

Moralejo, F. J., Cardoza, R. E., Gutierrez, S., & Martin, J. F. (1999). Thaumatin production in Aspergillus awamori by use of expression cassettes with strong fungal promoters and high gene dosage. *Applied and Environmental Microbiology*, *65*(3), 1168–1174.

Moralejo, F. J., Cardoza, R. E., Gutiérrez, S., Sisniega, H., Faus, I., & Martín, J. F. (2000). Overexpression and lack of degradation of thaumatin in an aspergillopepsin A-defective mutant of Aspergillus awamori containing an insertion in the pepA gene. *Applied Microbiology and Biotechnology*, *54*(6), 772–777. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11152068

Moralejo, F., Watson, A., Jeenes, D., Archer, D., & Martín, J. (2001). A defined level of protein disulfide isomerase expression is required for optimal secretion of thaumatin by Aspergillus awamori. *Molecular Genetics and Genomics*, *266*(2), 246–253. https://doi.org/10.1007/s004380100550

Morales, Y., Tortajada, M., Picó, J., Vehí, J., & Llaneras, F. (2014). Validation of an FBA model for *Pichia pastoris* in chemostat cultures. *BMC Systems Biology*, *8*(1), 142. https://doi.org/10.1186/s12918-014-0142-y

Moucha, T., Linek, V., & Prokopová, E. (2003). Gas hold-up, mixing time and gas–liquid volumetric mass transfer coefficient of various multiple-impeller configurations: Rushton turbine, pitched blade and techmix impeller and their combinations. *Chemical Engineering Science*, *58*(9), 1839–1846.

Moucha, T., Rejl, F. J., Kordač, M., & Labík, L. (2012). Mass transfer characteristics of multiple-impeller fermenters for their design and scale-up. *Biochemical Engineering Journal*, *69*, 17–27. https://doi.org/10.1016/j.bej.2012.08.007

Nelson, D. L., & Cox, M. M. (2008). *LEHNINGER, PRINCIPLES OF BIOCHEMISTRY* (Fifth Edit). New York: W. H. Freeman and company.

Nielsen, J., Villadsen, J., & Lidén, G. (2002). *Bioreaction Engineering Principles* (Second). Kluwe Academic/Plenum Publishers.

NIRASAWA, S., NISHINO, T., KATAHIRA, M., UESUGI, S., HU, Z., & KURIHARA, Y. (1994). Structures of heat-stable and unstable homologues of the sweet protein mabinlin. The difference in the heat stability is due to replacement of a single amino acid residue. *European Journal of Biochemistry*, *223*(3), 989–995. https://doi.org/10.1111/j.1432-1033.1994.tb19077.x

Nocon, J., Steiger, M. G., Pfeffer, M., Sohn, S. B., Kim, T. Y., Maurer, M., … Mattanovich, D. (2014). Model based engineering of *Pichia pastoris* central metabolism enhances recombinant protein production. *Metabolic Engineering*, *24*, 129–138. https://doi.org/10.1016/j.ymben.2014.05.011

Nocon, J., Steiger, M., Mairinger, T., Hohlweg, J., Rußmayer, H., Hann, S., … Mattanovich, D. (2016). Increasing pentose phosphate pathway flux enhances recombinant protein production in *Pichia pastoris*. *Applied Microbiology and Biotechnology*, 1–9. https://doi.org/10.1007/s00253-016-7363-5

ODEPA. (2014). *Evolución del consumo aparente de los principales alimentos en Chile:*

2003-2013.

Oliveira, C., & Domingues, L. (2018). Guidelines to reach high-quality purified recombinant proteins. *Applied Microbiology and Biotechnology*, *102*(1), 81–92. https://doi.org/10.1007/s00253-017-8623-8

OMS. (2015). *Sugars intake for adult and children*.

Onken, U., & Liefke, E. (1989). Effect of total and partial pressure (oxygen and carbon dioxide) on aerobic microbial processes. *BIOPROCESSES AND ENGINEERING*, *40*. https://doi.org/10.1007/BFb0009825

Orth, J. D., Thiele, I., & Palsson, B. Ø. (2010). What is flux balance analysis? *Nature Biotechnology*, *28*(3), 245–248. https://doi.org/10.1038/nbt.1614

Overton, T. W. (2014). Recombinant protein production in bacterial hosts. *Drug Discovery Today*, *19*(5), 590–601. https://doi.org/10.1016/j.drudis.2013.11.008

Palsson, B. O. (2015). *Systems Biology: Constraint-based Reconstruction and Analysis*. Cambridge: Cambridge University Press.

Papanikou, E., & Glick, B. S. (2009). The yeast Golgi apparatus: insights and mysteries Effrosyni. *FEBS Letters*, *583*(23), 3746–3751. https://doi.org/10.1016/j.febslet.2009.10.072

Park, J. H., Lee, K. H., Kim, T. Y., & Lee, S. Y. (2007). Metabolic engineering of Escherichia coli for the production of L-valine based on transcriptome analysis and in silico gene knockout simulation. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(19), 7797–7802. https://doi.org/10.1073/pnas.0702609104

Patil, K. R., Rocha, I., Förster, J., & Nielsen, J. (2005). Evolutionary programming as a platform for in silico metabolic engineering. *BMC Bioinformatics*, *6*, 308. https://doi.org/10.1186/1471-2105-6-308

Peña, D. A., Gasser, B., Zanghellini, J., & Steiger, M. G. (2018). Metabolic engineering of *Pichia pastoris*. *Metabolic Engineering*, (April), 1–14. https://doi.org/10.1016/j.ymben.2018.04.017

Pereira, R., Nielsen, J., & Rocha, I. (2016). Improving the flux distributions simulated with genome-scale metabolic models of Saccharomyces cerevisiae. *Metabolic Engineering Communications*, *3*, 153–163. https://doi.org/10.1016/j.meteno.2016.05.002

Petersen, B., Gernaey, K., & Vanrolleghem, P. A. (2001). Practical identifiability of model parameters by combined respirometric-titrimetric measurements. *Water Science and Technology*, *43*(7), 347–355.

Porro, D., Gasser, B., Fossati, T., Maurer, M., Branduardi, P., Sauer, M., & Mattanovich, D. (2011). Production of recombinant proteins and metabolites in yeasts. *Applied Microbiology and Biotechnology*, *89*(4), 939–948. https://doi.org/10.1007/s00253-010-3019-z

Porro, D., Sauer, M., Branduardi, P., & Mattanovich, D. (2005). Recombinant protein production in yeasts. *Molecular Biotechnology*, *31*(3), 245–259. https://doi.org/10.1385/MB:31:3:245

Postma, E., Verduyn, C., Scheffers, W. a., & Van Dijken, J. P. (1989). Enzymic analysis of the crabtree effect in glucose-limited chemostat cultures of Saccharomyces cerevisiae. *Applied and Environmental Microbiology*, *55*(2), 468–477.

Potvin, G., Ahmad, A., & Zhang, Z. (2012). Bioprocess engineering aspects of heterologous protein production in *Pichia pastoris*: A review. *Biochemical*

Engineering Journal, *64*, 91–105. https://doi.org/10.1016/j.bej.2010.07.017

Price, N. D., Famili, I., Beard, D. a, & Palsson, B. Ø. (2002). Extreme pathways and Kirchhoff's second law. *Biophysical Journal*, *83*(5), 2879–2882. https://doi.org/10.1016/S0006-3495(02)75297-1

Prielhofer, R., Maurer, M., Klein, J., Wenger, J., Kiziak, C., Gasser, B., & Mattanovich, D. (2013). Induction without methanol: novel regulated promoters enable high-level expression in *Pichia pastoris*. *Microbial Cell Factories*, *12*(1), 5. https://doi.org/10.1186/1475-2859-12-5

Prielhofer, R., Reichinger, M., Wagner, N., Claes, K., Kiziak, C., Gasser, B., & Mattanovich, D. (2018). Superior proteina titers in half the fermentation time: Promoter and process engineering for the glucose-regulated *GTH1* promoter of *Pichia pastoris*. *Biotechnology and Bioengineering*, (April). https://doi.org/10.1002/bit.26800

Puxbaum, V., Mattanovich, D., & Gasser, B. (2015). Quo vadis? The challenges of recombinant protein folding and secretion in *Pichia pastoris*. *Applied Microbiology and Biotechnology*, *99*(7), 2925–2938. https://doi.org/10.1007/s00253-015-6470-z

Raiford, D. W., Heizer, E. M., Miller, R. V, Akashi, H., Raymer, M. L., & Krane, D. E. (2008). Do amino acid biosynthetic costs constrain protein evolution in Saccharomyces cerevisiae? *Journal of Molecular Evolution*, *67*(6), 621–630. https://doi.org/10.1007/s00239-008-9162-9

Rebnegger, C., Graf, A. B., Valli, M., Steiger, M. G., Gasser, B., Maurer, M., & Mattanovich, D. (2014). In *Pichia pastoris*, growth rate regulates protein synthesis and secretion, mating and stress response. *Biotechnology Journal*, *9*(4), 511–525. https://doi.org/10.1002/biot.201300334

Resina, D., Bollók, M., Khatri, N. K., Valero, F., Neubauer, P., & Ferrer, P. (2007). Transcriptional response of *P. pastoris* in fed-batch cultivations to Rhizopus oryzae lipase production reveals UPR induction. *Microbial Cell Factories*, *6*, 21. https://doi.org/10.1186/1475-2859-6-21

Riesenberg, D., & Guthke, R. (1999). High-cell-density cultivation of microorganisms. *Applied Microbiology and Biotechnology*, *51*(4), 422–430. https://doi.org/10.1007/s002530051412

Rußmayer, H., Buchetics, M., Gruber, C., Valli, M., Grillitsch, K., Modarres, G., … Gasser, B. (2015). Systems-level organization of yeast methylotrophic lifestyle. *BMC Biology*, *13*(1), 80. https://doi.org/10.1186/s12915-015-0186-5

Saa, P. A., & Nielsen, L. K. (2016). Fast-SNP: A fast matrix pre-processing algorithm for efficient loopless flux optimization of metabolic models. *Bioinformatics*, *32*(24), 3807–3814. https://doi.org/10.1093/bioinformatics/btw555

Sacher, J., Saa, P., Cárcamo, M., López, J., Gelmi, C. a., & Pérez-Correa, R. (2011). Improved calibration of a solid substrate fermentation model. *Electronic Journal of Biotechnology*, *14*(5). https://doi.org/10.2225/vol14-issue5-fulltext-7

Saitua, F., Torres, P., Pérez-Correa, J. R., & Agosin, E. (2017). Dynamic genome-scale metabolic modeling of the yeast *Pichia pastoris*. *BMC Systems Biology*, *11*(27).

Sanchez-Garcia, L., Martín, L., Mangues, R., Ferrer-Miralles, N., Vázquez, E., & Villaverde, A. (2016). Recombinant pharmaceuticals from microbial cells: A 2015 update. *Microbial Cell Factories*, *15*(1), 1–7. https://doi.org/10.1186/s12934-016-0437-3

Sánchez, B. J., & Nielsen, J. (2015). Genome scale models of yeast: towards standardized evaluation and consistent omic integration. *Integrative Biology (United Kingdom)*, *7*(8), 846–858. https://doi.org/10.1039/c5ib00083a

Sánchez, B. J., Pérez-Correa, J. R., & Agosin, E. (2014). Construction of robust dynamic genome-scale metabolic model structures of Saccharomyces cerevisiae through iterative re-parameterization. *Metabolic Engineering*, *25*, 159–173. https://doi.org/10.1016/j.ymben.2014.07.004

Sánchez, B. J., Soto, D. C., Jorquera, H., Gelmi, C. A., & Pérez-Correa, J. R. (2014). HIPPO: An iterative reparametrization method for identification and calibration of dynamic bioreactor models of complex processes. *Industrial and Engineering Chemistry Research*, *53*(48), 18514–18525. https://doi.org/10.1021/ie501298b

Satrustegui, J., Bautista, J., & Machado, A. (1983). NADPH/NADP+ Ratio - Regulatory Implications in Yeast Glyoxylic-Acid Cycle. *Molecular and Cellular Biochemistry*, *51*(2), 123–127.

Schellenberger, J., Que, R., Fleming, R. M. T., Thiele, I., Orth, J. D., Feist, A. M., … Palsson, B. Ø. (2011). Quantitative prediction of cellular metabolism with constraint-based models : the COBRA Toolbox v2 . 0. https://doi.org/10.1038/nprot.2011.308

Schestibratov, K. a., & Dolgov, S. V. (2005). Transgenic strawberry plants expressing a thaumatin II gene demonstrate enhanced resistance to Botrytis cinerea. *Scientia Horticulturae*, *106*(2), 177–189. https://doi.org/10.1016/j.scienta.2005.03.016

Schuetz, R., Kuepfer, L., & Sauer, U. (2007). Systematic evaluation of objective functions for predicting intracellular fluxes in Escherichia coli. *Molecular Systems Biology*, *3*(119), 119. https://doi.org/10.1038/msb4100162

Schuetz, R., Zamboni, N., Zampieri, M., Heinemann, M., & Sauer, U. (2012). Multidimensional optimality of microbial metabolism. *Science (New York, N.Y.)*, *336*(6081), 601–604. https://doi.org/10.1126/science.1216882

Schutter, K. De, Lin, Y., Tiels, P., Hecke, A. Van, Glinka, S., Peer, Y. Van De, … Rouze, P. (2009). Genome sequence of the recombinant protein production host *Pichia pastoris*. *Nature Biotechnology*, *27*(6), 561–569. https://doi.org/10.1038/nbt.1544

Seber, George, & Lee, A. (2012). *Linear regression analysis* (Second Edi, Vol. 936). NJ: John Wiley & Sons.

Segrè, D., Vitkup, D., & Church, G. M. (2002). Analysis of optimality in natural and perturbed metabolic networks. *PNAS*, *99*(23), 15112–15117.

Siegell, S. D., & Gaden, E. L. (1962). Automatic Control of Dissolved Oxygen Levels in Fermentations, *IV*, 345–356.

Sohn, S. B., Graf, A. B., Kim, T. Y., Gasser, B., Maurer, M., Ferrer, P., … Lee, S. Y. (2010). Genome-scale metabolic model of methylotrophic yeast *Pichia pastoris* and its use for in silico analysis of heterologous protein production. *Biotechnology Journal*, *5*(7), 705–715. https://doi.org/10.1002/biot.201000078

Solà, A., Jouhten, P., Maaheimo, H., Sánchez-Ferrando, F., Szyperski, T., & Ferrer, P. (2007). Metabolic flux profiling of *Pichia pastoris* grown on glycerol/methanol mixtures in chemostat cultures at low and high dilution rates. *Microbiology (Reading, England)*, *153*(Pt 1), 281–290. https://doi.org/10.1099/mic.0.29263-0

Solà, A., Maaheimo, H., Ylönen, K., Ferrer, P., & Szyperski, T. (2004). Amino acid biosynthesis and metabolic flux profiling of *Pichia pastoris*. *European Journal of Biochemistry / FEBS*, *271*(12), 2462–2470. https://doi.org/10.1111/j.1432-

1033.2004.04176.x

Sonnleitner, B. (2016). Real-Time Measurement and Monitoring of Bioprocesses. In *Fundamental Bioengineering* (John Villa, pp. 409–437).

Sriram, K., Rodriguez-Fernandez, M., & Doyle, F. J. (2012). Modeling cortisol dynamics in the neuro-endocrine axis distinguishes normal, depression, and post-traumatic stress disorder (PTSD) in humans. *PLoS Computational Biology*, *8*(2). https://doi.org/10.1371/journal.pcbi.1002379

Stahl, R., Luhrs, R., & Dargatz, H. (2009). Thaumatin from transgenic barley.

Stephanopoulos, G. M. (1998). Metabolic Engineering. *Biotechnology and Bioengineering*, *58*, 301–359. https://doi.org/10.1007/1-4020-5252-9_10

Stephanopoulos, G. M., Aristidou, A. A., & Nielsen, J. (1998). *Metabolic Engineering Principles and Methodologies*. (Press San Diego: Academic, Ed.) (3rd Editio). San Diego.

Stephens, M. A. (1974). EDF Statistics for Goodness of Fit and Some Comparisons. *Journal of the American Statistical Association*, *69*(347), 730–737.

Stoger, E. (2012). Plant bioreactors - the taste of sweet success. *Biotechnology Journal*, *7*(4), 475–476. https://doi.org/10.1002/biot.201100472

Suez, J., Korem, T., Zeevi, D., Zilberman-Schapira, G., Thaiss, C. A., Maza, O., … Elinav, E. (2014). Artificial sweeteners induce glucose intolerance by altering the gut microbiota. *Nature*, *514*(7521), 181–186. https://doi.org/10.1038/nature13793

Suresh, S., Srivastava, V. C., & Mishra, I. M. (2009). Techniques for oxygen transfer measurement in bioreactors: A review. *Journal of Chemical Technology and Biotechnology*, *84*(8), 1091–1103. https://doi.org/10.1002/jctb.2154

Szwacka, M., Krzymowska, M., Osuch, A., Kowalczyk, M. E., & Malepszy, S. (2002). Variable properties of transgenic cucumber plants containing the thaumatin II gene from Thaumatococcus daniellii. *Acta Physiologiae Plantarum*, *24*(2), 173–185. https://doi.org/10.1007/s11738-002-0009-5

Te Morenga, L. A., Howatson, A. J., Jones, R. M., & Mann, J. (2014). Dietary sugars and cardiometabolic risk: Systematic review and meta-analyses of randomized controlled trials of the effects on blood pressure and lipids. *American Journal of Clinical Nutrition*, *100*(1), 65–79. https://doi.org/10.3945/ajcn.113.081521

Temussi, P. a. (2006). Natural sweet macromolecules: how sweet proteins work. *Cellular and Molecular Life Sciences : CMLS*, *63*(16), 1876–1888. https://doi.org/10.1007/s00018-006-6077-8

Theerasilp, S., & Kurihara, Y. (1988). Complete purification and characterization of the taste-modifying protein, miraculin, from miracle fruit. *Journal of Biological Chemistry*, *263*(23), 11536–11539.

Theron, C. W., Berrios, J., Delvigne, F., & Fickers, P. (2018). Integrating metabolic modeling and population heterogeneity analysis into optimizing recombinant protein production by Komagataella (Pichia) pastoris. *Applied Microbiology and Biotechnology*, *102*(1), 63–80. https://doi.org/10.1007/s00253-017-8612-y

Thiele, I., & Palsson, B. Ø. (2010). A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nature Protocols*, *5*(1), 93–121. https://doi.org/10.1038/nprot.2009.203

Thompson, C. A. (2010). FDA approves kallikrein inhibitor to treat hereditary angioedema. *American Journal of Health-System Pharmacy*, *67*, 93.

https://doi.org/10.2146/news100005

Tolner, B., Smith, L., Begent, R. H. J., & Chester, K. a. (2006). Production of recombinant protein in *Pichia pastoris* by fermentation. *Nature Protocols*, *1*(2), 1006–1021. https://doi.org/10.1038/nprot.2006.126

Tomàs-Gamisans, M., Ferrer, P., & Albiol, J. (2016). Integration and validation of the genome-scale metabolic models of *Pichia pastoris*: A comprehensive update of protein glycosylation pathways, lipid and energy metabolism. *PLoS ONE*, *11*(1), 1–24. https://doi.org/10.1371/journal.pone.0148031

Tomàs-Gamisans, M., Ferrer, P., & Albiol, J. (2017). Fine-tuning the *P. pastoris* iMT1026 genome-scale metabolic model for improved prediction of growth on methanol or glycerol as sole carbon sources. *Microbial Biotechnology*, *11*(1), 224–237.

Torres, P., Cerri, M. O., de Arruda Ribeiro, M. P., Ricardo Perez-Correa, J., & Agosin, E. (2016). Automated algorithm to determine kLa considering system delay. *Journal of Chemical Technology & Biotechnology*, (September 2016). https://doi.org/10.1002/jctb.5157

Tribe, L. a, Briens, C. L., & Margaritis, A. (1995). Determination of the volumetric mass transfer coefficient (k(L)a) using the dynamic "gas out-gas in" method: Analysis of errors caused by dissolved oxygen probes. *Biotechnology and Bioengineering*, *46*(4), 388–392.

van der Wel, H., Iyengar, R. B., van Brouwershaven, J., & van Wassenaar, P. D. (1984). Assignment of the disulphide bonds in the sweet-tasting protein thaumatin I. *European Journal of Biochemistry*, *144*, 41–45.

van der Wel, H., & Loeve, K. (1972). Isolation and Characterization of Thaumatin I and II , the Sweet -Tasting Proteins from Thaumatococcus daniellii Benth. *European Journal of Biochemistry*, *31*, 221–225.

van Urk, H., Postma, E., Scheffers, W. a, & van Dijken, J. P. (1989). Glucose transport in crabtree-positive and crabtree-negative yeasts. *Journal of General Microbiology*, *135*(1987), 2399–2406. https://doi.org/10.1099/00221287-135-9-2399

Varela, C., Agosin, E., Baez, M., Klapa, M., & Stephanopoulos, G. (2003). Metabolic flux redistribution in Corynebacterium glutamicum in response to osmotic stress. *Applied Microbiology and Biotechnology*, *60*(5), 547–555. https://doi.org/10.1007/s00253-002-1120-7

Vargas, F. a, Pizarro, F., Pérez-Correa, J. R., & Agosin, E. (2011). Expanding a dynamic flux balance model of yeast fermentation to genome-scale. *BMC Systems Biology*, *5*(1), 75. https://doi.org/10.1186/1752-0509-5-75

Varma, A., & Palsson, B. (1994). Stoichiometric Flux Balance Models Quantitatively Predict Growth and Metabolic By-Product Secretion in Wild-Type Escherichia coli W3110, *60*(10), 3724–3731.

Vermasvuori, R. (2009). *Production of recombinant proteins and monoclonal antibodies – Techno-economical evaluation of the production methods*. Helsinky University of Technology.

Verney, E. B. (1926). The osmotic pressure of the proteins of human serum and plasma. *The Journal of Physiology*, *61*(3), 319–328.

Villadsen, J., Nielsen, J., & Lidén, G. (2011). *Bioreaction Engineering Principles* (3rd ed.). Nueva York: Springer.

Villadsen, J., Nielsen, J., & Lidén, G. (2011). Gas–Liquid Mass Transfer. In *Bioreaction*

engineering principles (Thrid Edit, Vol. 1, pp. 459–495). New York, NY.

Villadsen, J., & Patil, K. R. (2007). Optimal Fed-Batch Cultivation When Mass Transfer Becomes Limiting, *98*(3), 706–710. https://doi.org/10.1002/bit

Vitae, K. van 't R., & Vitae, R. G. J. M. van der L. (2011). *Comprehensive Biotechnology*. *null* (Vol. null). Elsevier. https://doi.org/10.1016/B978-0-08-088504-9.00083-0

Vogl, T., Sturmberger, L., Kickenweiz, T., Wasmayer, R., Schmid, C., Hatzl, A. M., … Glieder, A. (2016). A Toolbox of Diverse Promoters Related to Methanol Utilization: Functionally Verified Parts for Heterologous Pathway Expression in *Pichia pastoris*. *ACS Synthetic Biology*, *5*(2), 172–186. https://doi.org/10.1021/acssynbio.5b00199

Walsh, G. (2014). Biopharmaceutical benchmarks 2014. *Nature Biotechnology*, *32*(7), 992–1000. https://doi.org/10.1038/nbt0910-917

Wang, J.-R., Li, Y.-Y., Liu, D.-N., Liu, J.-S., Li, P., Chen, L.-Z., & Xu, S.-D. (2015). Codon Optimization Significantly Improves the Expression Level of *α* -Amylase Gene from *Bacillus licheniformis* in *Pichia pastoris*. *BioMed Research International*, *2015*, 1–9. https://doi.org/10.1155/2015/248680

Wang, Y., Chu, J., Zhuang, Y., Wang, Y., Xia, J., & Zhang, S. (2009). Industrial bioprocess control and optimization in the context of systems biotechnology. *Biotechnology Advances*, *27*(6), 989–995.

Wang, Y., Liang, Z. H., Zhang, Y. S., Yao, S. Y., Xu, Y. G., Tang, Y. H., … Feng, Y. M. (2001). Human insulin from a precursor overexpressed in the methylotrophic yeast *Pichia pastoris* and a simple procedure for purifying the expression product. *Biotechnology and Bioengineering*, *73*, 74–79. https://doi.org/10.1002/1097-0290(20010405)73:1<74::AID-BIT1038>3.0.CO;2-V

Weickmann J.L, B. L. . and W. G. L. (1994). High level expression of thaumatin in Saccharomyces cerevisiae. In W. M & H. J.D (Eds.), *Thaumatin* (pp. 151–169). CRC Press.

West, M. G., Horne, D. W., & Appling, D. R. (1996). Metabolic role of cytoplasmic isozymes of 5,10-methylenetetrahydrofolate dehydrogenase in Saccharomyces cerevisiae. *Biochemistry*, *35*(9), 3122–3132. https://doi.org/10.1021/bi952713d

Whitman, W. G. (1924). The Two-Film Theory of Gas Absorption. *Chemical and Metallurgical Engineering*, *29*(4), 146–148.

WHO. (2013). Global action plan for the prevention and control of noncommunicable diseases 2013-2020., 102. https://doi.org/978 92 4 1506236

WHO. (2015). *Guideline : Sugars intake for adults and children*.

Wintjens, R., Viet, T. M. V. N., Mbosso, E., & Huet, J. (2011a). Hypothesis/review: The structural basis of sweetness perception of sweet-tasting plant proteins can be deduced from sequence analysis. *Plant Science*, *181*(4), 347–354. https://doi.org/10.1016/j.plantsci.2011.06.009

Witty, M. (1990). Preprothaumatin II is processed to biological activity in Solanum tuberosum. *Biotechnology Letters*, *12*, 131–136.

Yamada, S., Wada, M., & Chibata, I. (1978). Effect of High Oxygen Partial Pressure on the Conversion of Sorbitol to Sorbose by Acetobacter suboxydans : Studies on Aerobic Fermentation(II). *Journal of Fermentation Technology*, *56*(1), 29–34.

Yang, J. D., & Wang, N. S. (1992). Oxygen mass transfer enhancement via fermentor headspace pressurization. *Biotechnology Progress*, *8*(3), 244–251. https://doi.org/10.1021/bp00015a010

Yang, Y., Xia, J., Li, J., Chu, J., Li, L., Wang, Y., … Zhang, S. (2012). A novel impeller configuration to improve fungal physiology performance and energy conservation for cephalosporin C production. *Journal of Biotechnology*, *161*(3), 250–256. https://doi.org/10.1016/j.jbiotec.2012.07.007

Yang, Z., & Zhang, Z. (2018). Engineering strategies for enhanced production of protein and bio-products in *Pichia pastoris*: A review. *Biotechnology Advances*, *36*(1), 182–195. https://doi.org/10.1016/j.biotechadv.2017.11.002

Yawalkar, A. a, Heesink, A. B. M., Versteeg, G. F., & Vishwas, G. (2002). Gas – Liquid Mass Transfer Coefficient in Stirred Tank Reactors, *80*(October), 840–848.

Ye, R., Huang, M., Lu, H., Qian, J., Lin, W., Chu, J., … Zhang, S. (2017). Comprehensive reconstruction and evaluation of *Pichia pastoris* genome-scale metabolic model that accounts for 1243 ORFs. *Bioresources and Bioprocessing*, *4*(1), 22. https://doi.org/10.1186/s40643-017-0152-x

Yoshida, T. (2017). Bioreactor Development and Process Analytical Technology. In T. Yoshida (Ed.), *Applied Bioengineering: Innovations and Future Directions* (First, pp. 169–211). Weinheim, Germany: Wiley-VCH Verlag GmbH & Co. KGaA. https://doi.org/0.1002/9783527800599.ch6

Zahrl, R. J., Peña, D. A., Mattanovich, D., & Gasser, B. (2017). Systems biotechnology for protein production in *Pichia pastoris*. *FEMS Yeast Research*, *17*(7), 1–15. https://doi.org/10.1093/femsyr/fox068

Zawirska-Wojtasiak, R., Gośliński, M., Szwacka, M., Gajc-Wolska, J., & Mildner-Szkudlarz, S. (2009). Aroma evaluation of transgenic, thaumatin II-producing cucumber fruits. *Journal of Food Science*, *74*(3), 204–210. https://doi.org/10.1111/j.1750-3841.2009.01082.x

Zhang, M., Yu, X. W., Xu, Y., Jouhten, P., Swapna, G. V. T., Glaser, R. W., … Szyperski, T. (2017). 13C metabolic flux profiling of *Pichia pastoris* grown in aerobic batch cultures on glucose revealed high relative anabolic use of TCA cycle and limited incorporation of provided precursors of branched-chain amino acids. *FEBS Journal*, *284*(18), 3100–3113. https://doi.org/10.1111/febs.14180

Zhu, H., Nienow, A. W., Bujalski, W., & Simmons, M. J. H. (2009). Mixing studies in a model aerated bioreactor equipped with an up- or a down-pumping "Elephant Ear" agitator: Power, hold-up and aerated flow field measurements. *Chemical Engineering Research and Design*, *87*(3), 307–317. https://doi.org/10.1016/j.cherd.2008.08.013

8. SUPPLEMENTARY MATERIAL

S 1-1 Estimation of thaumatin production in *P. Pastoris*

We carried out an estimation of thaumatin production ($CH_{1.89}\,N_{0.62}\,O_{0.85}; PM = 36{,}17\ (C-mol)\ /\ gr$) in *P. pastoris* ($CH_{1.76}\,N_{0.143}O_{0.636}; PM = 25{,}94\ (C-mol)\ /\ gr$)(Jordà *et al.*, 2012) in aerobic conditions. For this, we used following assumptions:

- A constitutive promoter is used, so protein production is parallel to biomass generation. Thaumatin yield and biomass yield as a whole can not be higher than the production of maximum thaumatin and maximum biomass; i.e. when it is a single product.
- CO_2 production in base of glucose consumption, can be lower or equal to 4 $(mol\ CO_2)\ /\ (mol_{glu}\ )$(0,67 $(C-mol\ CO_2)\ /\ (C-mol_{glu})$). This value corresponds to the amount of CO2 produced, if glucose is only consumed via glycolysis, TCA and oxidative phosphorylation.
- All yield have to be higher or equal to zero.

Given this, the first thing that we estimated would be the maximum yields of biomass and thaumatin, that is, as unique products. For this, an elemental balance of C, H, N and O was used and, additionally, the use of redox potential.

1. Maximum biomass yield ($Y^{O}{}_{sx}$).

$$CH_2O + Y_{SN}\cdot NH_3 + Y_{SO}\cdot O_2 \rightarrow Y_{sx}\cdot CH_{1.76}N_{0.143}O_{0.636} + Y_{sc}\cdot CO_2 + Y_{sa}\cdot H_2O$$

$$Y_{SN} = 0{,}015\ {}^{C-mol\ NH_3}/_{C-mol\ glu}$$

$$Y_{SO} = 0{,}53\ {}^{C-mol\ O_2}/_{C-mol\ glu}$$

$$Y^{O}{}_{sx} = 0{,}46\ {}^{C-mol\ biomasa}/_{C-mol\ glu}$$

$$Y_{sc} = 0{,}58\ {}^{C-mol\ CO_2}/_{C-mol\ glu}$$

$$Y_{sa} = 0{,}62\ {}^{C-mol\ H_2O}/_{C-mol\ glu}$$

2. Maximum thaumatin yield ($Y^{O}{}_{sp}$)

$$CH_2O + Y_{SN} \cdot NH_3 + Y_{SO} \cdot O_2 \rightarrow Y_{sp} \cdot CH_{1.89}N_{0.62}O_{0.85} + Y_{sc} \cdot CO_2 + Y_{sa} \cdot H_2O$$

$$Y_{SN} = 0{,}44 \; {}^{C-mol\ NH_3}/_{C-mol\ glu}$$

$$Y_{SO} = 0{,}59 \; {}^{C-mol\ O_2}/_{C-mol\ glu}$$

$$Y^{O}{}_{sp} = 0{,}70 \; {}^{C-mol\ taumatina}/_{C-mol\ glu}$$

$$Y_{sc} = 0{,}30 \; {}^{C-mol\ CO_2}/_{C-mol\ glu}$$

$$Y_{sa} = 0{,}99 \; {}^{C-mol\ H_2O}/_{C-mol\ glu}$$

After this, the following restrictions are assumed for the case of yield estimation for thaumatin and biomass as a whole:

General equation:

$$CH_2O + Y_{SN} \cdot NH_3 + Y_{SO} \cdot O_2 \rightarrow Y_{sx} \cdot CH_{1.76}N_{0.143}O_{0.636} + Y_{sp} \cdot CH_{1.89}N_{0.62}O_{0.85} + Y_{sc} \cdot CO_2 + Y_{sa} \cdot H_2O$$

- Restrictions:

$$Y_{sp} \leq 0{,}70 \; {}^{C-mol\ taumatina}/_{C-mol\ glu}$$

$$Y_{sx} \leq 0{,}46 \; {}^{C-mol\ biomasa}/_{C-mol\ glu}$$

$$Y_{sc} \leq 0{,}67 \; {}^{C-mol\ CO_2}/_{C-mol\ glu}$$

Then if it is assumed that the carbon flow towards the biomass is not affected (worse situation for the production of protein), the following results are obtained:

$$Y_{SN} = 0{,}12 \; {}^{C-mol\ NH_3}/_{C-mol\ glu}$$

$$Y_{SO} = 0{,}49 \; {}^{C-mol\ O_2}/_{C-mol\ glu}$$

$$Y_{sx} = 0{,}46 \ {}^{C - mol\ biomasa}/_{C - mol\ glu}$$

$$Y_{sp} = 0{,}09 \, {}^{C - mol\ taumatina}/_{C - mol\ glu}$$

$$Y_{sc} = 0{,}46 \, {}^{C - mol\ CO_2}/_{C - mol\ glu}$$

$$Y_{sa} = 0{,}71 \ {}^{C - mol\ H_2O}/_{C - mol\ glu}$$

From this we can extract that thaumatin yield by glucose will be:

$$Y_{sp} = 0{,}09 \, {}^{C - mol\ taumatina}/_{C - mol\ glu} = 74 \, {}^{mg_{taumatina}}/_{gr_{glu}}$$

and biomass yield:

$$Y_{sx} = 0{,}46 \ {}^{C - mol\ biomasa}/_{C - mol\ glu} = 0{,}53 \, {}^{gr_{biomasa}}/_{gr_{glu}}$$

If this is coupled to a dynamic model coupled with assumptions based on experimentation already carried out:

- It is assumed that the growth rate may be affected, which means that

$$\mu_{max} = 0{,}2 \ \ 1 \ / \ h$$

- It is assumed a fed batch culture, with decreasing μ between 0.1 to 0.05 1/h.

- It is assumed a batch culture, with 20 g/L of glucose.

- 80 hours of simulation are considered, given the useful volume of reactors and biomass reached by the microorganism.

The simulation was carried out in Matlab software, which yields a productivity $225 \, {}^{\mathrm{mg_{tau}}}/_{\mathrm{L \cdot h}}$, which corresponds to 250 times more than what Masuda *et al*. (Masuda *et al*., 2010) and 180 times more than what Moralejo *et al.(Moralejo et al.*, 2000), which corresponds to a very low efficiency.

S 2-1. Plasmids construction and *Pichia pastoris* transformation

Plasmid construction

We constructed two plasmids using the commercial recombinant vectors pGAPZB (Invitrogen, Carlsbad, CA, USA): one for integrating the thaumatin gene (pGAPZB-TAU) and the other to revert the auxotrophy of the strain *P. pastoris* GS115 (PGAPZB-HIS4) (Invitrogen, Carlsbad, CA, USA).

PCR amplification of the *his4* fragment was performed using genomic DNA from strain *P. pastoris* X-33 (Invitrogen, Carlsbad, CA, USA). The genomic DNA was extracted using Wizard Genomic DNA purification kit according to the manufacturer's instructions (Promega, Madison, WI, USA). PCR amplification of the DNA fragments was performed by Gibson assembly in 35 PCR cycles using Phusion High-Fidelity DNA Polymerase (ThermoFisher, Waltham, MA, USA). All PCR products were treated with DpnI enzyme to eliminate original vector residues and purified by gel extraction using the Qiaquick Gel Extraction kit (Qiagen, Hilden, Germany) following the manufacturer's instructions. The purified genes fragments and vectors were mixed based on their molar ratios in a final volume of 5 μL containing 100 ng of total DNA. This DNA mix was added to 15 μL of 1.33X master mix (5X isothermal mix buffer, T5 exonuclease 1 U μL^{-1}, Phusion DNA polymerase 2U/μL, Taq DNA ligase 40 U μL^{-1} and milli-Q purified water) and the reaction mixture was incubated at 50 °C for 1 h. Finally, 10 μL of the reaction mix were used directly to transform chemically competent *E. coli TOP 10* cells. The cells were grown at 37 °C in low salt-LB medium containing 25 μg mL^{-1} zeocin for selection of clones transformed with both vectors.

Transformation of P. pastoris

P. pastoris GS115 wild type strain was used as host. Transformation was performed in two steps. The first transformation was performed using the vector pGAPZB-TAU and the AvrII restriction enzyme to linearize and introduce the latter vector into the cells by electroporation (Gasser et al. 2006). In this case, zeocin (100 μg mL^{-1}) was used for the selection of positively transformed clones. Once positive clones were obtained, the second

transformation was carried out using the empty vector PGAPZB-HIS4. In this case, the XbaI restriction enzyme was used to linearize the vector. Zeocin-supplemented YNB plates (100 μg mL^{-1} zeocin) were employed for the selection of positively transformed clones. All transformations were verified by genomic sequencing (Macrogen Inc., Seoul, Korea).

S 2-2. Bioreactor and culture medium.

The strain was cultivated in glucose-limited chemostat cultures at different dilution rates and dissolved oxygen. Overall, seven different experimental conditions were evaluated following a Doehrlet experimental design. Continuous cultures were performed at a working volume of 1 L in 2-L benchtop bioreactors (Sartorius AG, Göttingen, Germany) using peristaltic pumps (Ismatec, IDEX Health & Science, Germany) to control the feeding and at 1.2 bar total pressure. Once the batch phase was concluded, the continuous phase was initiated at the appropriate specific growth rate and dissolved oxygen (DO) level. DO was maintained using a mixture of three gases; air, pure nitrogen and pure oxygen depending on the DO set point. In all experimental conditions, the stirring rate was set to 700 rpm, aeration to 1 VVM (1 L min^{-1}), and the temperature and pH were respectively kept at 25 °C and 5.0. Off-gases were cooled in a condenser at 4 °C and desiccated in two silica gel columns. CO_2 and O_2 abundances were analyzed through BCP-CO_2 and BCP-O_2 Sensors (BlueSens gas sensor GmbH, Herten, Germany). Each dilution rate was kept for at least five residence times.

Batch medium was composed of (per liter): 39.9 g glycerol, 1.8 g citric acid, 12.6 g $(NH_4)_2HPO_4$, 0.022 g $CaCl_2 \cdot H_2O$, 0.9 g KCl, 0.5 g $MgSO_4 \cdot 7H_2O$, 2 mL biotin (0.2 g L^{-1}), 4 mL PTM1 trace salts stock solution. Chemostat medium composition was composed of (per liter): 50 g glucose, 0.9 g citric acid, 3.45 g $(NH_4)_2HPO_4$, 0.01 g $CaCl_2 \cdot H_2O$, 1.7 g KCl, 0.65 g $MgSO_4 \cdot 7H_2O$, 1 mL biotin (0.2 g L^{-1}), 1.6 mL PTM1 trace salts stock solution, and 0.2 mL of antifoam Glanapon 2000 (Bussetti & Co GmbH, Vienna, Austria). HCl 25% was used to reach pH 5.0.

S 2-3. GSMM tailoring for describing aerobic growth on glucose

The *i*MT1026 v3.0 genome-scale metabolic reconstruction from Tomàs-Gamisans et al. (2017) was employed to describe the metabolism *Pichia pastoris*. In order to accurately represent the central carbon metabolism behavior of this yeast, we systematically adapted the reconstruction to represent and capture known metabolic features. This process was composed of two main steps: 1) Blocking and modification of reaction directionalities based on published data (e.g., transcriptomics, proteomics, thermodynamics, etc.), and 2) Modification of reaction directionalities based on topological considerations, i.e., removal of thermodynamically infeasible internal loops (Saa & Nielsen, 2016). The details of each step are detailed below.

The first step leveraged the abundant experimental omics data available for this yeast under glucose-limited conditions (Clasquin et al., 2011; Gasser et al., 2007; Krivoruchko et al., 2014; Rußmayer et al., 2015; Zhang et al., 2017). Application of this data greatly constrained the genome-scale metabolic model (see Table 8-1 for the applied constraints), significantly impacting the flux distribution predictions (see supplementary material 2-6 for main impacts). The second step involved the modification of reaction directionalities based on flux simulations imposing the loopless flux condition. To this task, the Fast-SNP algorithm (Saa & Nielsen, 2016) was used using experimental training data (i.e., fluxomic constraints) under glucose-limited conditions. Briefly, three flux scenarios were simulated: 1) aerobic growth in glucose with q_S values between 0.025 to 0.275 mmol g_{DWC}^{-1} h^{-1} and conventional exchange directions, 2) 11 aerobic growth conditions on glucose without secondary product formation (i.e., arabitol and ethanol) (Adelantado et al., 2017; Baumann et al., 2011; Carnicer et al., 2009, 2012; Garcia-Ortega, Valero, & Montesinos-Seguí, 2017), and 3) 15 aerobic growth conditions on glucose with formation of secondary products (Adelantado et al., 2017; Baumann et al., 2011; Carnicer et al., 2009, 2012; Garcia-Ortega et al., 2017) (Table 8-2). Reactions that had the same directionality across all the conditions were set to irreversible. The definitive list with the constraints on the GSMM is shown in Table 8-3.

Flux simulations

Simulations were performed using the Constraint-Based Reconstruction and Analysis (COBRA) toolbox (Schellenberger et al., 2011) in MATLAB 2013 (Mathworks, Natick, MA, USA). Flux predictions were performed using parsimonious Flux Balance Analysis (pFBA). Topological modification of the reaction directionalities was performed using the Fast-SNP algorithm (Saa & Nielsen, 2016). Finally, Gurobi 2016 was employed for the required linear and mixed-integer linear optimizations.

Table 8- 1 Summary of reaction modifications based on experimental and literature data.

Reaction ID	Stoichiometric equation	Gene	Comment	Action
Exchange reactions				
Ex_glyc	gly =>		Not supplied in the medium	Blocked
Ex_etoh	etoh =>		Not supplied in the medium	Blocked
Ex_abt	abt =>		Not detected in the medium	Blocked
Ex_fab	fab =>		Protein not expressed	Blocked
Ex_pyr	pyr =>		Present in the medium	Open
Ex_cit	cit =>		Present in the medium	Open
Cytoplasm				
GLUK	glc + atp => h + adp + g6p	PAS_chr4_0624	Duplicated reaction	Blocked
MDH	mal + nad + h <=> nadh + oaa	PAS_chr4_0815	Part of the glyoxylate cycle (not needed for glucose growth) (Rußmayer et al., 2015b)	Reversibility constrained
FBA3	s17bp <=> dhap+e4p	PAS_chr1-1_0072	Used in the direction S17BP in ribogenesis (Clasquin et al., 2011; Rußmayer et al., 2015b)	Reversibility constrained

SULR	3 h2o + h2s + 3 nadp <=> 5 h + 3 nadph + so3	(PAS_chr4_0369 or PAS_chr3_1084)	Thermodynamically favorable in the reverse direction (dG=103 KJ/mol)	Reversibility constrained
MMSAD3	nad + coa + msa => co2 + nadh + accoa		Not considered for lipid synthesis (Krivoruchko et al., 2014)	Blocked
		Mitochondrion		
MDHm	mal + nad + h <=> nadh + oaa	PAS_chr2-1_0238	Part of the glyoxylate cycle (not needed for glucose growth) (Rußmayer et al., 2015b)	Reversibility constrained
ME1m	nad + mal => pyr + co2 + nadh	PAS_chr3_0181	Inactive during glucose growth (Zhang et al., 2017)	Blocked
ME2m	nadp + mal => pyr + co2 + nadph	PAS_chr3_0181	Inactive during glucose growth (Zhang et al., 2017)	Blocked
		Peroxisome		
AOD	meoh + o2 => h2o2 + fald	(PAS_chr4_0821 or PAS_chr4_0152)	Absent in peroxisomal fraction during glucose growth (Rußmayer et al., 2015b)	Blocked
DAS	fald + xu5p => g3p + dha	(PAS_chr3_0832 or PAS_chr3_0834)	Absent in peroxisomal fraction during glucose growth (Rußmayer et al., 2015b)	Blocked
DHAKx	atp + dha => h + adp + dhap	PAS_chr3_0841	Absent in peroxisomal fraction during glucose growth (Rußmayer et al., 2015b)	Blocked

FBAx	fdp => g3p + dhap	(PAS_chr1-1_0072 or PAS_chr1-1_0319)	Absent in peroxisomal fraction during glucose growth (Rußmayer et al., 2015b)	Blocked
FBPx	h2o + fdp => pi + f6p	PAS_chr3_0868	Absent in peroxisomal fraction during glucose growth (Rußmayer et al., 2015b)	Blocked
SHBPH	h2o + s17bp => pi + s7p	PAS_chr2-2_0177	Absent in peroxisomal fraction during glucose growth (Rußmayer et al., 2015b)	Blocked
CATp	2 h2o2 => o2 + h2o	PAS_chr2-2_0131	Absent in peroxisomal fraction during glucose growth (Rußmayer et al., 2015b)	Blocked
FALDH2	nad + hmgth => ndh + sfglutth	PAS_chr3_1028	Absent in peroxisomal fraction during glucose growth (Rußmayer et al., 2015b)	Blocked
SFGTH	h2o + sfglutth => h + gthrd + for	PAS_chr3_0867	Absent in peroxisomal fraction during glucose growth (Rußmayer et al., 2015b)	Blocked
FDH	nad + h + for => co2 + nadh	PAS_chr3_0932	Absent in peroxisomal fraction during glucose growth (Rußmayer et al., 2015b)	Blocked
CSp	h2o + accoa + oaa => coa + h + cit	PAS_chr1-1_0475	Absent in peroxisomal fraction during glucose growth (Rußmayer et al., 2015b)	Blocked

ICLx	icit => glx + succ	PAS_chr1-4_0338	Absent in peroxisomal fraction during glucose growth (Rußmayer et al., 2015b)	Blocked
MDHp	nad + mal => oaa + h + nad	PAS_chr4_0815	Absent in peroxisomal fraction during glucose growth (Rußmayer et al., 2015b)	Blocked
ASPTAp	ak + asp <=> oaa + glu	PAS_chr4_0974	Absent in peroxisomal fraction during glucose growth (Rußmayer et al., 2015b)	Blocked
TKT1x	xu5p + r5p => g3p + s7p	(PAS_chr1-4_0150 or PAS_chr3_0834 or PAS_chr3_0832)	Absent in peroxisomal fraction during glucose growth (Rußmayer et al., 2015b)	Blocked
CATp	2 h2o2 => o2 + h2o	PAS_chr2-2_0131	Absent in peroxisomal fraction during glucose growth (Rußmayer et al., 2015b)	Blocked
RPIx	r5p => ru5p	PAS_chr4_0212	Absent in peroxisomal fraction during glucose growth (Rußmayer et al., 2015b)	Blocked

Table 8- 2 Constraints used to simulate the different scenarios for the modification of reaction directionalities.

Conditions		Flux bounds	Exchanges						References
			Glucose	Arabitol	Ethanol	CO2	O2	Biomass	
Scenario 1	1	LB	-0.025	0	0	0	-1000	0	
		UB	-0.025	0	0	1000	0	1000	
	2	LB	-0.05	0	0	0	-1000	0	
		UB	-0.05	0	0	1000	0	1000	
	3	LB	-0.075	0	0	0	-1000	0	
		UB	-0.075	0	0	1000	0	1000	
	4	LB	-0.1	0	0	0	-1000	0	
		UB	-0.1	0	0	1000	0	1000	
	5	LB	-0.125	0	0	0	-1000	0	
		UB	-0.125	0	0	1000	0	1000	
	6	LB	-0.15	0	0	0	-1000	0	
		UB	-0.15	0	0	1000	0	1000	
	7	LB	-0.175	0	0	0	-1000	0	
		UB	-0.175	0	0	1000	0	1000	
	8	LB	-0.2	0	0	0	-1000	0	
		UB	-0.2	0	0	1000	0	1000	
	9	LB	-0.225	0	0	0	-1000	0	
		UB	-0.225	0	0	1000	0	1000	
	10	LB	-0.25	0	0	0	-1000	0	
		UB	-0.25	0	0	1000	0	1000	
	11	LB	-0.275	0	0	0	-1000	0	
		UB	-0.275	0	0	1000	0	1000	
Scenario 2	12	LB	-0.97	0	0	2.73	-2.84	0.1	(Carnicer et al., 2012)
		UB	-0.97	0	0	3.01	-2.57	0.1	
	13	LB	-1.04	0	0	2.29	-2.47	0.095	(Tomàs-Gamisans et al. 2016)
		UB	-0.96	0	0	2.55	-2.23	0.105	
	14	LB	-1.05	0	0	2.38	-2.58	0.095	(Tomàs-Gamisans et al. 2016)
		UB	-0.97	0	0	2.66	-2.3	0.105	
	15	LB	-1.55	0	0	3.02	-2.87	0.16	(Solà et al., 2004)
		UB	-1.55	0	0	3.34	-2.59	0.16	
	16	LB	-0.94	0	0	2.3	-2.26	0.1	(Baumann et al., 2010)
		UB	-0.94	0	0	2.38	-2.2	0.1	
	17	LB	-1.04	0	0	2.18	-2.29	0.1	(Carnicer et al., 2009)
		UB	-0.94	0	0	2.56	-1.95	0.1	
	18	LB	-0.957	0	0	1.99	-2.05	0.1	(Adelantado et al., 2017)
		UB	-0.957	0	0	2.01	-2.03	0.1	
	19	LB	-1.11	0	0	2.52	-2.3	0.1	(Adelantado et al., 2017)
		UB	-1.11	0	0	2.62	-2.2	0.1	

	20	LB	-1.2	0	0	2.22	-2.81	0.1	(Garcia-Ortega, et al., 2017)
		UB	-1.2	0	0	2.46	-2.55	0.1	
	21	LB	-1.2	0	0	2.06	-2.98	0.1	(Garcia-Ortega, et al., 2017)
		UB	-1.2	0	0	2.29	-2.7	0.1	
	22	LB	-1.2	0	0	2.04	-2.97	0.1	(Garcia-Ortega, et al., 2017)
		UB	-1.2	0	0	2.26	-2.69	0.1	
Scenario 3	23	LB	-1.31	0.02	0.069	2.41	-2.42	0.1	(Baumann et al., 2010)
		UB	-1.31	0.033	0.099	2.73	-2.1	0.1	
	24	LB	-1.34	0.11	0.27	2.43	-2.67	0.095	(Tomàs-Gamisans et al. 2016)
		UB	-1.22	0.15	0.35	2.67	-2.43	0.105	
	25	LB	-1.82	0.31	0.72	2.93	-2.15	0.095	(Tomàs-Gamisans et al. 2016)
		UB	-1.62	0.35	0.96	3.49	-1.87	0.105	
	26	LB	-1.43	0.17	0.35	2.54	-2.15	0.095	(Tomàs-Gamisans et al. 2016)
		UB	-1.31	0.21	0.47	2.82	-1.83	0.105	
	27	LB	-1.64	0.18	0.71	2.7	-2.07	0.095	(Tomàs-Gamisans et al. 2016)
		UB	-1.48	0.26	0.95	3.18	-1.55	0.105	
	28	LB	-1.69	0.13	1.1	5.3	-4.35	0.1	(Baumann et al., 2010)
		UB	-1.69	0.167	1.22	5.94	-3.93	0.1	
	29	LB	-1.4	0.2	0.31	1.93	-1.65	0.1	(Baumann et al., 2010)
		UB	-1.26	0.22	0.35	2.13	-1.49	0.1	
	30	LB	-1.83	0.4	0.95	1.57	-0.64	0.1	(Baumann et al., 2010)
		UB	-1.65	0.44	1.05	1.73	-0.44	0.1	
	31	LB	-1.32	0.035	0.38	2.55	-1.94	0.1	(Adelantado et al., 2017)
		UB	-1.32	0.061	0.49	2.65	-1.84	0.1	
	32	LB	-1.98	0.45	1.164	5.5	-3.81	0.1	(Adelantado et al., 2017)
		UB	-1.98	0.47	1.198	5.7	-3.67	0.1	
	33	LB	-1.28	0.0095	0.0665	2.06	-3.08	0.1	(Garcia-Ortega. et al.. 2017)
		UB	-1.28	0.0105	0.0735	2.28	-2.78	0.1	
	34	LB	-1.6	0.026	0.665	2.49	-2.99	0.1	(Garcia-Ortega. et al.. 2017)
		UB	-1.6	0.028	0.735	2.75	-2.71	0.1	
	35	LB	-2	0.038	1.33	3.06	-2.84	0.1	(Garcia-Ortega. et al.. 2017)
		UB	-2	0.042	1.47	3.38	-2.57	0.1	
	36	LB	-2.4	0.05	1.995	3.53	-2.89	0.1	(Garcia-Ortega, et al., 2017)
		UB	-2.4	0.056	2.205	3.91	-2.61	0.1	
	37	LB	-3.5	0.252	3.325	4.42	-2.81	0.1	(Garcia-Ortega, et al., 2017)
		UB	-3.5	0.278	3.675	4.88	-2.55	0.1	

Table 8- 3 Final restriction of manual-curated GSM

Reaction ID	lb	ub
Ex_glyc	0	0
Ex_fab	0	0
Ex_rol	0	0
BIOMASS_glyc	0	0
Ex_pyr	0	1000
Ex_cit	-1000	1000
BIOMASS	0	1000
Ex_thau	0	1000
thaut	0	1000
pThau	0	1000
thauAA	0	1000
thauRNA	0	1000
thauDNA	0	1000
GLUK	0	0
MDHm	0	1000
MDH	0	1000
SULR	-1000	0
AOD	0	0
DAS	0	0
DHAKx	0	0
FBAx	0	0
FBPx	0	0
SHBPH	0	0
RPIx	0	0
CATp	0	0
FALDH2	0	0
SFGTH	0	0
FDH	0	0
TKT1x	0	0
CSp	0	0
ICLx	0	0
MDHp	0	0
ASPTAp	0	0
ME1m	0	0
ME2m	0	0
PFK_3	-1000	0
Ex_pyr	0	1000
MMSAD3	0	0
Ex_h2o	0	1000
H2Ot	-1000	0
O2t	0	1000
GLCter	-1000	0
M4MPDOLter	0	1000
M7MASNBterg	0	1000
HCO3DH	-1000	0
FAcoaRavge	0	1000
HXDCEALR	0	1000
OGLYCOStg	0	1000
CHITINtg	0	1000
MANNANtg	-1000	0
FARCOAtm	0	1000
FACOAtm	-1000	0
CLPNtm	0	1000
NADHter	0	1000
NADter	-1000	0
NADPHter	0	1000
NADPter	-1000	0
12DGRter	-1000	0
PAter	0	1000
DOLter	-1000	0
34HPPt2p	0	1000
3C4MOPtm	0	1000
4ABUTNtm	0	1000
4ABUTtm	-1000	0
4H2OGLTtm	-1000	0
4H2OGLTtp	-1000	0
4HPRO_LTtm	0	1000
ACtp	0	1000
AKGtp	-1000	0
CBPtn	-1000	0
CITtap	-1000	0
CITtcp	0	1000
CYSTtp	0	1000
E4HGLUtm	-1000	0
E4HGLUtp	0	1000
E4Ptm	0	1000
GLNt2n	0	1000
GLUt2n	-1000	0
GLXtp	0	1000
HCYSt2p	-1000	0

MANNANter	-1000	0
PAN4Ptm	-1000	0
PAPtm	-1000	0
PHCHGS	0	1000
PIt2n	-1000	0
PNP	-1000	0
PTD1INOtn_SC	0	1000
PYRt2p	-1000	0
SHSL4r	-1000	0
MPK	-1000	0
TYRt2m	-1000	0
ASPt5n	-1000	0
PI4Ptn	-1000	0
2DDA7Ptm	-1000	0
4HGLSDm	0	1000
ADK1	0	1000
ADK3m	0	1000
AKGDam	0	1000
ALATA_Lm	-1000	0
ASNt2r	-1000	0
ASPt2n	-1000	0
ASPt2r	0	1000
CITtbm	-1000	0
ERGSTter	0	1000
ERGTETROLter	0	1000
FBA2	-1000	0
FBA3	-1000	0
GCC2am	0	1000
GCC2cm	0	1000
GK2	0	1000
GLCt1	0	1000
HCO3E	0	1000
HCO3Em	0	1000
HICITDm	0	1000
HMGCOASm	-1000	0
ILETAm	-1000	0
NDPK9	0	1000
O2ter	0	1000
OXAGm	0	1000
PANTtm	-1000	0
PAtm_PP	0	1000
PDX5PO	0	1000

PUNP7	0	1000
SQ23EPXter	0	1000
SQLter	0	1000
DASYNm_PP	0	1000
THIORDXm	0	1000
AACOAT	0	1000
ACONTx	0	1000
ALCD19	0	1000
COAtp	-1000	0
FRDO	-1000	0
FTCD	0	1000
GluForTx	-1000	0
HACD9m	0	1000
HIBDm	0	1000
MGCHrm	0	1000
MI14PP	0	1000
MI4PP	0	1000
MICITDm	-1000	0
MTRK	0	1000
NP1	-1000	0
PPItx	0	1000
PROAKGOX1r	0	1000
SBTPD	0	1000
TRPS3r	0	1000
ASNtm	-1000	0
GLCISO	0	1000
DHORDm	0	1000
OLIGOPK	0	1000
FOLR2m	0	1000
GGLUCTC	0	1000
ALLTNISOR	-1000	0
OHCUREIMDZLNCAR	0	1000
OGLYCOS	0	1000
G15LACH	0	1000
PIt2er	0	1000
ATPter	0	1000
ADPter	-1000	0
ATPter_H	0	1000
H2Stm	-1000	0
OROTtm	-1000	0
CYStm	-1000	0

ACSERtm	0	1000
ACACtm	0	1000
DHORtm	0	1000
CITtp	-1000	0
AKGMALtp	0	1000
DHAtv	0	1000
GTPGDPtv	-1000	0
CO2tg	-1000	0
DHFtm	0	1000
NH4tp	-1000	0
PEtg_SC	-1000	0
PROtm	0	1000
PStg_SC	0	1000
PStv_PP	0	1000
PEtv_PP	-1000	0
CDPDGPm_PP	0	1000
2DHPtm	0	1000
3C3HMPtm	-1000	0
3MOBtm	-1000	0
5AOPtm	-1000	0
AATA	0	1000
ACALDtm	-1000	0
ACCOACr	0	1000
ACGAM6PS	0	1000
ACGAMPM	0	1000
ACOATA	0	1000
ACONT	0	1000
ACONTm	0	1000
ADSL1r	0	1000
ADSL2r	0	1000
AHCYStm	-1000	0
AICART	0	1000
AIRCr	0	1000
ALCD2x	0	1000
ALLTAHr	0	1000
ALLTNr	0	1000
AMETtm	0	1000
AP4AHr	0	1000
ARGSL	0	1000
ARGSSr	0	1000
CO2t	-1000	0
CO2tm	-1000	0

CO2tv	-1000	0
CSNATm	0	1000
CSNATp	0	1000
DADK	-1000	0
DHORTS	-1000	0
DTMPK	0	1000
DURIPP	-1000	0
ETHAPT_PP	0	1000
ETOHtm	-1000	0
G5SADr	0	1000
GALU	0	1000
GCCam	0	1000
GCCcm	0	1000
GLCtv	-1000	0
H2Ot	-1000	0
H2Otp	0	1000
HACNHm	0	1000
HMGCOAR	-1000	0
HMGCOAS	-1000	0
HMGCOAtm	-1000	0
GLYCLTDy	-1000	0
IMPC	-1000	0
IPDDI	0	1000
IPPMIa	-1000	0
IPPMIb	-1000	0
LEUTA	-1000	0
MAN6PI	-1000	0
MCITDm	0	1000
MCOATA	0	1000
MTRI	0	1000
NDPK1	0	1000
NDPK2	0	1000
NDPK4	0	1000
NH4t	0	1000
NH4tm	-1000	0
NNAMr	0	1000
O2t	0	1000
O2tm	0	1000
ORPT	-1000	0
PGMT	-1000	0
PMANM	-1000	0
PRAGSr	0	1000

PRASCS	0	1000
PRPPS	0	1000
PUNP4	0	1000
PUNP6	0	1000
SACCD1	0	1000
SACCD2	0	1000
TREt2v	0	1000
UDPACGLP	0	1000
UGLYCHr	0	1000
UMPK	0	1000
UREASE	0	1000
DASYN_PP	0	1000
PEtm_PP	-1000	0
Ex_h2o	0	1000
Ex_o2	-1000	0
Ex_co2	0	1000
RBK_D	-1000	0
Htr	0	1000
PIt2m	0	1000
H2Otm	-1000	0
RPI	-1000	0
ETOHt	-1000	0
ETOHtm	-1000	0
FBA	0	1000
GAPD	0	1000
PFK	0	1000
PGK	-1000	0
Ex_etoh	0	1000
D_ABTt	0	1000

S 2-4. Raw *P. pastoris* chemostat data under different μ and DO of 108% measured in this study

Table 8- 4 Raw *P. pastoris* chemostat data under different μ and DO of 108% measured in this study

Condition	μ (h^{-1})		DOsat	$Y_{S,X}$ ($g_{DCW}\,g^{-1}$)		q_S (mmol $g_{DCW}^{-1}\,h^{-1}$)		q_{CO2} (mmol $g_{DCW}^{-1}\,h^{-1}$)		q_{O2} (mmol $g_{DCW}^{-1}\,h^{-1}$)		RQ (-)	
	Mean[a]	S.E.[b]	(%)	Mean[a]	S.E.[b]	Mean[a]	S.E.[b]	Mean[a]	S.E.[b]	Mean[a]	S.E.[b]	Mean[a]	S.E.[b]
8.1	7.14 E-02	-	108	5.72 E-01	1.11 E-02	6.94 E-01	5.38 E-05	1.37 E+00	6.24 E-05	1.15 E+00	2.03 E-04	1.18 E+00	1.35 E-01
8.2	7.44 E-02	-	108	5.40 E-01	9.77 E-03	7.64 E-01	1.34 E-05	1.52 E+00	2.13 E-05	1.51 E+00	1.99 E-05	1.00 E+00	7.47 E-03
8[c]	7.29 E-02	1.46 E-03	108	5.56 E-01	3.71 E-03	7.29 E-01	3.51 E-02	1.44 E+00	7.67 E-02	1.33 E+00	1.79 E-01	1.09 E+00	8.97 E-02
9.1	1.19 E-01	-	108	5.57 E-01	1.11 E-02	1.19 E+00	1.82 E-04	1.99 E+00	2.93 E-04	2.13 E+00	3.14 E-04	9.34 E-01	2.68 E-03
9.2	1.26 E-01	-	108	5.64 E-01	9.77 E-03	1.24 E+00	1.66 E-04	2.42 E+00	3.20 E-04	2.26 E+00	2.99 E-04	1.07 E+00	5.71 E-03
9[c]	1.22 E-01	2.63 E-03	108	5.60 E-01	3.71 E-03	1.21 E+00	2.79 E-02	2.20 E+00	2.18 E-01	2.19 E+00	6.82 E-02	1.00 E+00	6.83 E-02

S 2-5. Raw *P. pastoris* chemostat data under different μ and DO conditions measured in this study

Table 8- 5 Raw *P. pastoris* chemostat data under different μ and DO conditions measured in this study

Condition	μ (h^{-1})		DOsat	$Y_{S,X}$ (g_{DCW} g^{-1})		q_S (mmol g_{DCW}^{-1} h^{-1})		q_{CO2} (mmol g_{DCW}^{-1} h^{-1})		q_{O2} (mmol g_{DCW}^{-1} h^{-1})		RQ (-)	
	Mean[a]	S.E.[b]	(%)	Mean[a]	S.E.[b]	Mean[a]	S.E.[b]	Mean[a]	S.E.[b]	Mean[a]	S.E.[b]	Mean[a]	S.E.[b]
1.1	6.86 E-02	-	4	5.09 E-01	9.71 E-03	7.49 E-01	2.25 E-05	1.74 E+00	5.70 E-05	1.74 E+00	9.12 E-05	9.95 E-01	3.31 E-02
1.2	7.50 E-02	-	4	5.05 E-01	9.22 E-03	8.25 E-01	4.09 E-05	1.94 E+00	4.86 E-05	1.94 E+00	4.86 E-05	1.01 E+00	5.88 E-03
1[c]	7.18 E-02	3.16 E-03	4	5.07 E-01	1.87 E-03	7.87 E-01	3.76 E-02	1.84 E+00	1.01 E-01	1.84 E+00	9.70 E-02	1.00 E+00	5.50 E-03
2.1	1.14 E-01	-	4	5.19 E-01	1.02 E-02	1.22 E+00	4.20 E-05	2.78 E+00	9.32 E-05	2.80 E+00	9.28 E-05	9.94 E-01	4.10 E-03
2.2	1.25 E-01	-	4	5.42 E-01	1.28 E-02	1.28 E+00	4.04 E-05	2.75 E+00	5.10 E-05	2.78 E+00	4.91 E-05	9.88 E-01	4.56 E-03
2[c]	1.19 E-01	5.51 E-03	4	5.31 E-01	1.13 E-02	1.25 E+00	3.10 E-02	2.77 E+00	1.68 E-02	2.79 E+00	3.17 E-01	9.91 E-01	2.97 E-03
3.1	4.54 E-02	-	30	5.59 E-01	1.15 E-02	4.50 E-01	1.63 E-05	9.78 E-01	7.33 E-05	9.84 E-01	6.38 E-05	9.93 E-01	6.00 E-02
3.2	4.98 E-02	-	30	5.43 E-01	8.70 E-03	5.09 E-01	3.45 E-05	1.08 E+00	2.59 E-05	1.12 E+00	2.67 E-05	9.64 E-01	8.28 E-03
3[c]	4.76 E-02	2.24 E-03	30	5.51 E-01	7.88 E-03	4.80 E-01	2.94 E-02	1.03 E+00	5.06 E-02	1.05 E+00	6.77 E-02	9.78 E-01	1.48 E-02
4.1	9.92 E-02	-	30	5.65 E-01	1.04 E-02	9.75 E-01	2.77 E-05	1.97 E+00	4.98 E-05	2.03 E+00	5.13 E-05	9.72 E-01	2.29 E-16
4.2	1.02 E-01	-	30	5.64 E-01	1.46 E-02	1.00 E+00	6.00 E-05	1.95 E+00	9.50 E-05	1.84 E+00	3.52 E-04	1.06 E+00	1.39 E-01
4.3	9.99 E-02	-	30	5.69 E-01	1.20 E-02	9.75 E-01	3.43 E-05	2.01 E+00	6.38 E-05	2.10 E+00	6.66 E-05	9.56 E-01	3.97 E-03
4[c]	1.01 E-01	8.20 E-04	30	5.66 E-01	1.58 E-03	9.89 E-01	9.43 E-03	1.98 E+00	1.61 E-02	1.97 E+00	7.64 E-02	1.01 E+00	3.22 E-02
5.1	1.53 E-01	-	30	5.57 E-01	8.04 E-03	1.53 E+00	3.22 E-05	3.12 E+00	6.07 E-05	3.13 E+00	6.08 E-05	9.98 E-01	2.29 E-16
5.2	1.52 E-01	-	30	5.61 E-01	4.41 E-03	1.51 E+00	6.44 E-05	2.99 E+00	3.18 E-05	2.96 E+00	3.58 E-05	1.01 E+00	3.89 E-03
5.3	1.45 E-01	-	30	5.55 E-01	1.78 E-02	1.45 E+00	7.00 E-05	2.94 E+00	2.21 E-04	2.78 E+00	2.01 E-04	1.06 E+00	5.76 E-02
5[c]	1.50 E-01	2.44 E-03	30	5.58 E-01	1.77 E-03	1.50 E+00	2.10 E-02	3.02 E+00	5.24 E-02	2.96 E+00	9.90 E-02	1.02 E+00	1.85 E-02
6.1	7.19 E-02	-	56	5.89 E-01	1.11 E-02	6.78 E-01	5.58 E-05	1.22 E+00	3.39 E-05	1.29 E+00	3.30 E-05	9.40 E-01	7.52 E-03
6.2	7.17 E-02	-	56	5.90 E-01	1.76 E-02	6.75 E-01	2.58 E-05	1.21 E+00	3.72 E-05	1.30 E+00	3.99 E-05	9.27 E-01	3.54 E-03
6.3	7.18 E-02	-	56	5.79 E-01	8.74 E-03	6.89 E-01	4.42 E-05	1.28 E+00	3.66 E-05	1.29 E+00	5.05 E-05	9.98 E-01	2.43 E-02
6[c]	7.18 E-02	6.62 E-05	56	5.89 E-01	4.12 E-03	6.82 E-01	6.79 E-03	1.21 E+00	4.19 E-03	1.30 E+00	4.59 E-03	9.63 E-01	3.53 E-02
7.1	1.22 E-01	-	56	5.79 E-01	3.61 E-03	1.17 E+00	5.60 E-05	2.33 E+00	2.50 E-05	2.44 E+00	2.18 E-05	9.55 E-01	4.64 E-03
7.2	1.19 E-01	-	56	5.87 E-01	2.08 E-02	1.13 E+00	7.55 E-05	2.04 E+00	9.02 E-05	2.18 E+00	9.78 E-05	9.34 E-01	5.00 E-03
7.3	1.19 E-01	-	56	5.69 E-01	8.47 E-03	1.16 E+00	3.02 E-05	2.27 E+00	6.46 E-05	2.40 E+00	5.46 E-05	9.46 E-01	1.16 E-02
7[c]	1.20 E-01	1.04 E-03	56	5.83 E-01	3.94 E-03	1.14 E+00	1.41 E-02	2.18 E+00	1.47 E-01	2.31 E+00	1.29 E-01	9.40 E-01	6.31 E-03

[a,b] Measured mean and standard error calculated for 3 technical replicates.

[c] Mean and standard error calculated for 2 or 3 biological replicates.

Table 8- 5 Raw *P. pastoris* chemostat data under different μ and DO conditions measured in this study (continuation)

Condition	Carbon Balance (-)		Electronic balance (-)		C_X (g_{DCW})		$Y_{S,P}$ (mg g^{-1})		q_P (mg $g_{DCW}^{-1} \cdot h^{-1}$)	
	Mean[a]	S.E.[b]	Mean[a]	S.E.[b]	Mean[a]	S.E.[b]	Mean[a]	S.E.[b]	Mean[a]	S.E.[b]
1.1	9.46 E-01	1.07 E-02	9.59 E-01	1.09 E-02	2.52 E+01	4.82 E-01	3.09 E-01	1.91 E-02	4.16 E-02	2.70 E-03
1.2	9.48 E-01	1.02 E-02	9.57 E-01	1.04 E-02	2.50 E+01	4.57 E-01	2.25 E-01	3.66 E-03	5.25 E-02	1.35 E-06
1[c]	9.47 E-01	7.60 E-04	9.58 E-01	1.42 E-03	2.51 E+01	9.34 E-02	2.67 E-01	4.16 E-02	4.70 E-02	5.41 E-03
2.1	9.54 E-01	1.13 E-02	9.67 E-01	1.15 E-02	2.58 E+01	5.07 E-01	1.75 E-01	1.36 E-02	3.82 E-02	3.12 E-03
2.2	9.56 E-01	1.41 E-02	9.72 E-01	1.43 E-02	2.69 E+01	6.33 E-01	1.43 E-01	4.24 E-03	3.29 E-02	1.08 E-03
2[c]	9.55 E-01	1.04 E-03	9.70 E-01	2.31 E-03	2.63 E+01	5.62 E-01	1.59 E-01	1.58 E-02	3.56 E-02	2.66 E-03
3.1	9.78 E-01	1.27 E-02	9.92 E-01	1.29 E-02	2.83 E+01	5.81 E-01	4.18 E-01	1.17 E-02	3.39 E-02	1.18 E-03
3.2	9.52 E-01	9.58 E-03	9.77 E-01	9.77 E-03	2.70 E+01	4.31 E-01	3.81 E-01	5.56 E-03	3.49 E-02	7.65 E-04
3[c]	9.65 E-01	1.30 E-02	9.85 E-01	7.76 E-03	2.76 E+01	6.71 E-01	3.99 E-01	1.87 E-02	3.44 E-02	5.03 E-04
4.1	9.58 E-01	1.15 E-02	9.81 E-01	1.17 E-02	2.80 E+01	5.18 E-01	1.83 E-01	7.75 E-03	3.22 E-02	1.49 E-03
4.2	9.46 E-01	1.61 E-02	9.40 E-01	1.64 E-02	2.76 E+01	7.15 E-01	1.87 E-01	4.94 E-03	3.38 E-02	1.09 E-03
4.3	9.70 E-01	1.32 E-02	9.98 E-01	1.35 E-02	2.85 E+01	5.99 E-01	1.85 E-01	4.60 E-03	3.25 E-02	1.11 E-03
4[c]	9.58 E-01	7.00 E-03	9.69 E-01	1.73 E-02	2.81 E+01	2.44 E-01	1.86 E-01	1.19 E-03	3.32 E-02	5.13 E-04
5.1	9.54 E-01	8.85 E-03	9.68 E-01	9.03 E-03	2.76 E+01	3.99 E-01	1.34 E-01	3.19 E-03	3.67 E-02	1.02 E-03
5.2	9.48 E-01	4.86 E-03	9.57 E-01	4.96 E-03	2.82 E+01	2.22 E-01	1.25 E-01	5.28 E-03	3.39 E-02	1.46 E-03
5.3	9.48 E-01	1.96 E-02	9.42 E-01	2.00 E-02	2.78 E+01	8.91 E-01	1.29 E-01	3.68 E-03	3.38 E-02	1.52 E-03
5[c]	9.50 E-01	2.07 E-03	9.56 E-01	7.45 E-03	2.79 E+01	1.61 E-01	1.29 E-01	2.47 E-03	3.48 E-02	9.56 E-04
6.1	9.48 E-01	1.22 E-02	9.80 E-01	1.25 E-02	2.95 E+01	5.56 E-01	-	-	-	-
6.2	9.48 E-01	1.94 E-02	9.84 E-01	1.98 E-02	2.93 E+01	8.75 E-01	2.49 E-01	1.75 E-02	3.03 E-02	2.22 E-03
6.3	9.49 E-01	9.63 E-03	9.62 E-01	9.82 E-03	2.91 E+01	4.40 E-01	2.63 E-01	6.40 E-03	3.26 E-02	9.53 E-04
6[c]	9.48 E-01	3.87 E-04	9.73 E-01	1.11 E-02	2.92 E+01	6.01 E-02	2.56 E-01	6.85 E-03	3.14 E-02	1.15 E-03
7.1	9.69 E-01	3.98 E-03	9.97 E-01	4.06 E-03	2.89 E+01	1.81 E-01	1.85 E-01	1.20 E-02	3.89 E-02	2.53 E-03
7.2	9.46 E-01	2.30 E-02	9.81 E-01	2.34 E-02	2.95 E+01	1.05 E+00	1.66 E-01	1.12 E-02	3.37 E-02	2.52 E-03
7.3	9.53 E-01	9.33 E-03	9.84 E-01	9.52 E-03	2.85 E+01	4.24 E-01	2.61 E-01	4.02 E-03	3.10 E-02	1.50 E-03
7[c]	9.50 E-01	3.29 E-03	9.82 E-01	1.69 E-03	2.90 E+01	5.01 E-01	1.75 E-01	9.49 E-03	3.23 E-02	1.36 E-03

[a,b] Measured mean and standard error calculated for 3 technical replicates.

[c] Mean and standard error calculated for 2 or 3 biological replicates.

S 2-6. Regression analysis and Effects' estimation with statics values for second- degree equation

Table 8- 6 Regression analysis and Effects' estimation with statics values for second- degree equation

		Coefficient	Yxs[gDCW/g]		qs [mmol/gDCW·h]		q_{CO2} [mmol/gDCW·h]		q_{O2} [mmol/gDCW·h]		q_{tau} [mg/gDCW·h]		Y_{SP}[mg/g]	
		value	p-value	value	p-value	value	p-value	value	p-value	value	p-value	value	p-value
Response means	b0	0.565		0.984		1.983		1.993		0.033		0.186	
μ	b1	0.005	0.1311	0.492	0	0.975	0	0.963	0	-0.001	0.5509	-0.127	0
pO2	b2	0.036	0	-0.062	0	-0.348	0	-0.292	0	-0.004	0.0256	-0.004	0.625
μ□	b11	-0.010	0.0753	0.015	0.1513	0.060	0.2717	0.022	0.7724	0.002	0.5222	0.080	0.0004
μ·pO2	b12	-0.020	0.0071	0.004	0.713	0.032	0.5938	0.070	0.4176	0.008	0.0205	0.011	0.52
$pO2^2$	b22	-0.017	0.0153	0.032	0.0143	0.149	0.0265	0.207	0.0298	0.005	0.125	0.006	0.7349
R^2		94.8		93.8		94.4		94.2		68.3		95.9	
	Effects												
		value	DS	value	DS	value	DS	value	DS	value	DS	value	DS
Average		0.565	0.005	0.984	0.008	1.983	0.043	1.994	0.061	0.033	0.002	0.186	0.0127
μ		0.009	0.006	0.985	0.011	1.950	0.056	1.926	0.081	-0.002	0.003	-0.254	0.0164
pO2		0.072	0.006	-0.124	0.011	-0.696	0.060	-0.583	0.085	-0.008	0.003	-0.008	0.0164
μ□		-0.021	0.010	0.030	0.019	0.121	0.104	0.044	0.148	0.003	0.005	0.160	0.0324
μ·pO2		-0.039	0.012	0.008	0.022	0.064	0.116	0.140	0.165	0.016	0.006	0.022	0.0324
$pO2^2$		-0.034	0.012	0.063	0.021	0.298	0.115	0.414	0.164	0.010	0.006	0.011	0.0324

S 2-7. Principal effects of the manual curation of GSMM.

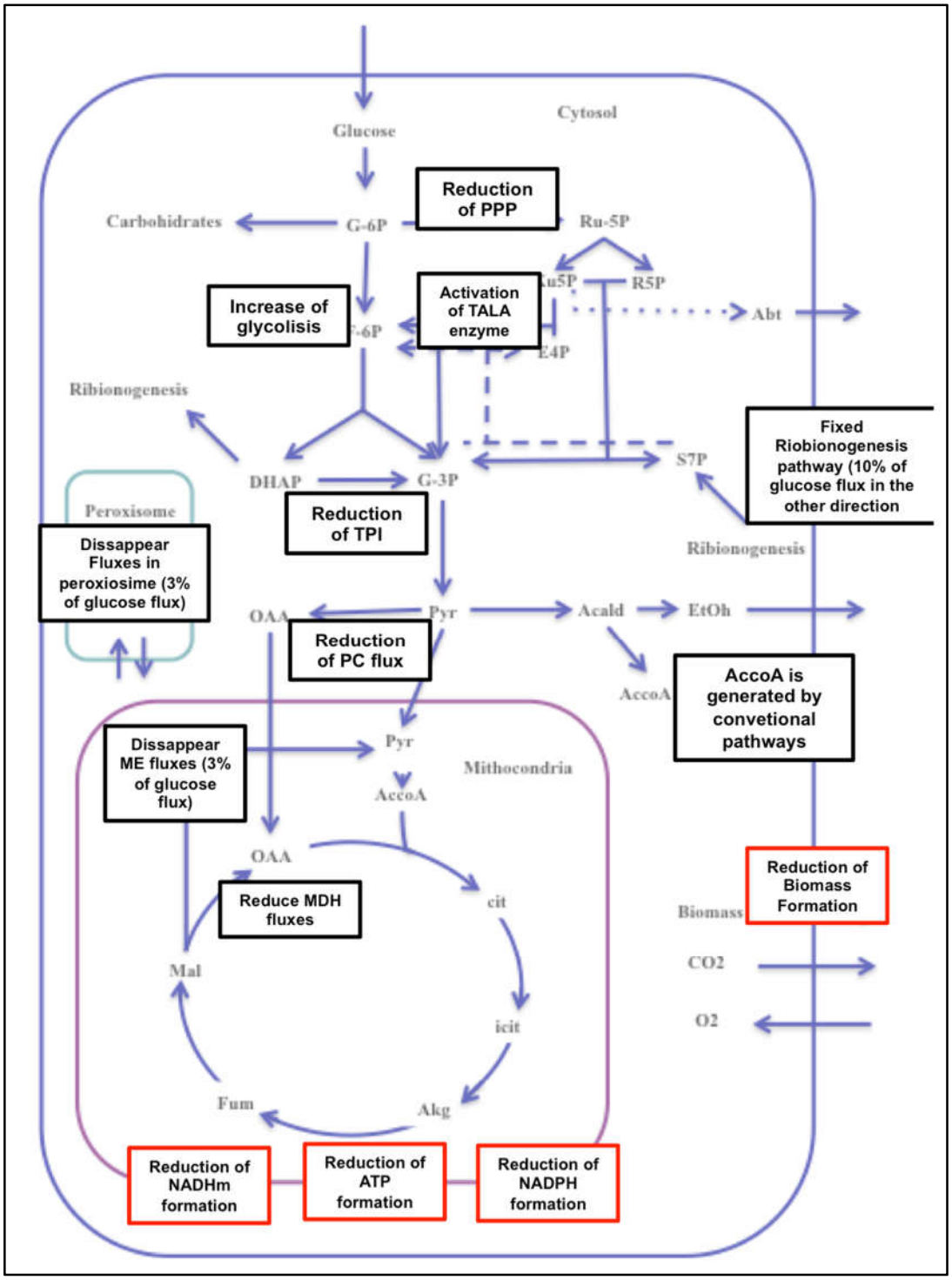


Figure 8-1 Principal effects of the manual curation of GSMM. In black, it is highlight the effects of manual curation and it is highlight the main effects of SNP algorithm in red

S 2-8. Effect of ashes in the prediction of intracellular fluxes.

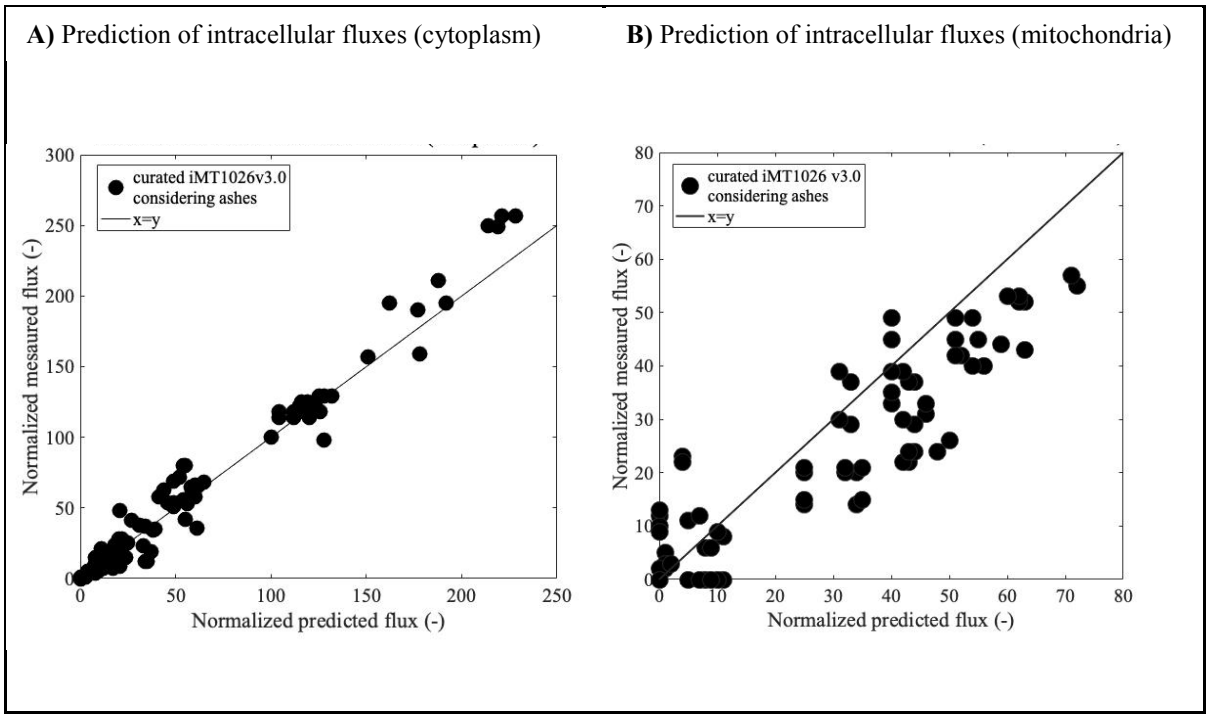


Figure 8-2**.** Evaluation of the tailored *P. pastoris* GSMM for metabolic flux prediction under glucose-limited conditions correcting by the estimated ashes content. A) Comparison of the predicted intracellular flux through cytoplasmic reactions of the GSMM considering (slope = 1.033, R^2 = 0.970) against experimental data under appropriate growth conditions. B) Comparison of the predicted intracellular flux through mitochondrial reactions of the GSMM (slope = 0.787, R^2 = 0.853) against experimental data under the same conditions used in A). For more information about the employed experimental conditions refer to Supplementary material S2-8.

S 2-9. Macromolecular composition of *P. pastoris* GS115.

Table 8- 7. Macromolecular composition of *P. pastoris* GS115

Component	Experimental composition (g g_{DCW}^{-1})[a]	Estimated composition without ashes (g/g_{DCW}^{-1})[b]
Proteins	0.3700	0.3983
Carbohydrates	0.3690	0.3972
Lipids	0.0620	0.0667
RNA	0.0660	0.0710
DNA	0.0013	0.0014

[a] Data taken from Carnicer et al (2009)

[b] Assuming 7.1% ashes content based on Carnicer (2012)

S 3-1. Demonstration of the convexity of the solution space of the QP problem in the metabolic block

A convex optimization problem is one of the form:

$$\begin{aligned} &Min && f_0(x) \\ &subject\ to && f_i(x) \le b_i \quad i = 1, \dots, m \end{aligned}$$

Where the functions $f_0, \dots, f_m: R^n \to R$ are convex, i.e., satisfy:

$$f_i(\alpha x + \beta y) \le \alpha f_i(x) + \beta f_i(y)$$

For all x, y $\in R^n$ and all $\alpha, \beta \in R$ with $\alpha + \beta = 1, \alpha \ge 0, \beta \ge 0$.

Now, for the problem solved in the metabolic block we have:

$$\begin{aligned} &Min && f_0(v):\ \alpha \cdot \sum v^2 - (1-\alpha)\cdot\mu \\ &subject\ to && f(v):\ S \cdot v = b \end{aligned}$$

Here, v is an R^n flux distribution vector (also solution to the system), α is the suboptimal growth coefficient, μ is the specific growth rate of the cell (also a component of v), S is the stoichiometric matrix (m metabolites x n reactions) and b is the overall balance for each of the m metabolites of the network.

A) Demonstration of the convexity of $f_0(v):\ \alpha \cdot \sum v^2 - (1-\alpha)\cdot\mu$

In order to determine the convexity of the problem, we have to demonstrate that:

$$f_0(\beta \cdot v_1 + \gamma \cdot v_2) \le \beta \cdot f_0(v_1) + \gamma \cdot f_0(v_2)$$

$$\underbrace{\alpha \cdot \sum(\beta \cdot v_1 + \gamma \cdot v_2)^2 - (1-\alpha)\cdot\left(\beta \cdot \mu_{v_1} + \gamma \cdot \mu_{v_2}\right)}_{I} \le \underbrace{\beta \cdot f_0(v_1) + \gamma \cdot f_0(v_2)}_{II}$$

Here, $v_1, v_2 \in R^n$ are any flux distribution vectors, $\beta + \gamma = 1, \beta \ge 0, \gamma \ge 1$. Now Expanding the expression above:

$$\begin{aligned} &\alpha \cdot \sum(\beta^2 \cdot v_1^2 + 2 \cdot \beta \cdot \gamma \cdot v_1 \cdot v_2 + \gamma^2 \cdot v_2^2) - (1-\alpha)\cdot\left(\beta \cdot \mu_{v_1} + \gamma \cdot \mu_{v_2}\right) \\ &\quad \le \beta \cdot \underbrace{\left(\alpha \cdot \sum v_1^2 - (1-\alpha)\cdot\mu_{v_1}\right)}_{f_0(v_1)} + \gamma \cdot \underbrace{\left(\alpha \cdot \sum v_2^2 - (1-\alpha)\cdot\mu_{v_2}\right)}_{f_0(v_2)} \end{aligned}$$

Which can be rearranged to:

$$\begin{aligned} &\alpha \cdot \sum(\beta^2 \cdot v_1^2 + 2 \cdot \beta \cdot \gamma \cdot v_1 \cdot v_2 + \gamma^2 \cdot v_2^2) - (1-\alpha)\cdot\left(\beta \cdot \mu_{v_1} + \gamma \cdot \mu_{v_2}\right) \\ &\quad \le \alpha \cdot (\beta \cdot \sum v_1^2 + \gamma \cdot \sum v_2^2\ ) - (1-\alpha)\cdot\left(\beta \cdot \mu_{v_1} + \gamma \cdot \mu_{v_2}\right) \end{aligned}$$

Eliminating equal terms

$$\alpha \cdot \sum(\beta^2 \cdot v_1^2 + 2 \cdot \beta \cdot \gamma \cdot v_1 \cdot v_2 + \gamma^2 \cdot v_2^2) \leq \alpha \cdot (\beta \cdot \sum v_1^2 + \gamma \cdot \sum v_2^2)$$

In our problem, $\alpha \geq 0$. If α = 0, i.e. the metabolic block solves a LP problem, the inequality above is $0 \leq 0$, which is true, and therefore the problem is convex. If $\alpha > 0$, it can be eliminated from the expression above:

$$\beta^2 \cdot \sum v_1^2 + 2 \cdot \beta \cdot \gamma \cdot \sum v_1 \cdot v_2 + \gamma^2 \cdot \sum v_2^2 \leq \beta \cdot \sum v_1^2 + \gamma \cdot \sum v_2^2$$

Now, $\gamma = 1 - \beta$

$$\beta^2 \cdot \sum v_1^2 + 2 \cdot \beta \cdot (1-\beta) \cdot \sum v_1 \cdot v_2 + (1-\beta)^2 \cdot \sum v_2^2 \leq \beta \cdot \sum v_1^2 + (1-\beta) \cdot \sum v_2^2$$

Rearranging the terms:

$$-\beta \cdot (1-\beta) \cdot \sum v_1^2 + 2 \cdot \beta \cdot (1-\beta) \cdot \sum v_1 \cdot v_2 - \beta \cdot (1-\beta) \cdot \sum v_2^2 \leq 0$$

Which can be reduced to:

$$\beta \cdot (1-\beta) \cdot \sum (v_1 - v_2)^2 \geq 0$$

According to the problem specifications, $\beta \geq 0$ and $\beta + \gamma = 1$, which constrains $\beta \in [0,1]$. Also the term $\sum (v_1 - v_2)^2$ is always positive or zero (if $v_1 = v_2 = \vec{0}$). Therefore, the inequality from above is always equal or larger than zero for the domain of the problem, confirming that the objective function is convex.

B) Demonstration of the convexity of $f(v)$: $S \cdot v = b$

To demonstrate that the problem is convex, the restrictions for the optimization should be convex as well. Therefore, imitating the previous procedure:

$$f(\beta \cdot v_1 + \gamma \cdot v_2) \leq \beta \cdot f(v_1) + \gamma \cdot f(v_2)$$

$$S \cdot (\beta \cdot v_1 + \gamma \cdot v_2) \leq \beta \cdot S \cdot v_1 + \gamma \cdot S \cdot v_2$$

$$\beta \cdot S \cdot v_1 + \gamma \cdot S \cdot v_2 \leq \beta \cdot S \cdot v_1 + \gamma \cdot S \cdot v_2$$

$$0 \leq 0$$

The last statement is true; therefore, the restrictions are also convex.

Finally, since the bi-objective function in the metabolic block and its restrictions are convex, the solution space of the problem is also convex. This ensures the optimality of the obtained flux distribution.

S 3-2. Construction and evaluation of the iFS670 model

The iFS670 model modified the iPP668 model from Chung *et al*., 2010 in three aspects:

1. Contains stoichiometric equations for the production of recombinant Thaumatin, Human Serum Albumin (HSA) and FAB fragment.
2. Includes an NAD-dependent arabitol biosynthesis pathway
3. Reversibility from mitochondrial symporters and cytosolic reactions involving NAD/NADP was updated according to the suggestions from Pereira, Nielsen, & Rocha, 2016.

At the end of this Supplementary material we compare the performance of the iFS670 with two other models from *Pichia pastoris*.

1. Stoichiometric reactions for the production of three recombinant proteins

Thaumatin, HSA and FAB synthesis pathways were also included in the model according to the DNA, RNA and amino acid requirements employed in the iLC915 model (Caspeta *et al*., 2012) to form the primary structure of the protein:

$$0.997 \cdot AA_{Prot} + 0.0029 \cdot RNA_{Prot} + 0.000028 \cdot DNA_{Prot} \rightarrow Protein\,[c] \quad (23)$$

$$\left(\sum_{1}^{20} \beta_i \cdot aa_i\right) + \gamma \cdot ATP[c] + \gamma \cdot H_2O[c] \rightarrow AA_{Prot} \quad (24)$$

$$\left(\sum_{1}^{4} \alpha_i \cdot NMP_i\right) + \delta \cdot ATP[c] + \delta \cdot H_2O[c] \rightarrow RNA_{Prot} \quad (25)$$

$$\left(\sum_{1}^{4} \varepsilon_i \cdot dNMP_i\right) + \theta \cdot ATP[c] + \theta \cdot H_2O[c] \rightarrow RNA_{Prot} \quad (26)$$

Coefficients for the different components of the proteins are detailed in Table 8-8, Table 8-9 and Table 8-10. Codon usage was taken from (Schutter *et al*., 2009) and was used as

input for the calculation of RNA and DNA sequences online (http://www.bioinformatics.org/sms2/rev_trans.html).

Table 8- 8 Amino acid requirements to form 1 gram of thaumatin, HSA and FAB fragment in the iPP669 model. The coefficients here reported were used in equation 2, a cost of 4,3 mole of ATP was assumed per mole of amino acid assembles in the protein. All coefficients have mmol/gram of protein units.

Substrate	**Thaumatin**	**HSA**	**Fab Fragment**
L-Alanine	0.721	0.909	0.559
L-Arginine	0.541	0.390	0.430
L-Asparagine	0.451	0.245	0.215
L-Aspartate	0.541	0.519	0.387
L-Cysteine	0.721	0.505	0.215
L-Glutamate	0.271	0.895	0.473
L-Glutamine	0.180	0.289	0.559
Glycine	1.082	0.188	0.602
L-Histidine	0.000	0.231	0.172
L-Isoleucine	0.361	0.130	0.215
L-Leucine	0.406	0.923	0.774
L-Lysine	0.496	0.866	0.387
L-Methionine	0.045	0.101	0.043
L-Phenylalanine	0.496	0.505	0.387
L-Proline	0.541	0.346	0.516
L-Serine	0.631	0.404	1.376
L-Threonine	0.902	0.418	0.731
L-Tryptophan	0.135	0.029	0.086
L-Tyrosine	0.361	0.274	0.387
L-Valine	0.451	0.620	0.645
ATP (γ)	40.1	37.8	39.4

Table 8- 9 RNA requirements for the production of 1 gram of Thaumatin, HSA and Fab fragment codifying RNA. A cost of 2,4 mol of ATP per gram of RNA was assumed.

Substrate	Thaumatin	HSA	Fab Fragment
AMP	0.73	0.86	0.74
UMP	1.14	1.09	1.18
GMP	0.73	0.72	0.58
CMP	0.48	0.40	0.57
ATP (δ)	7.38	7.38	7.38

Table 8- 10 DNA requirement for the formation of 1 gram of codifying DNA for Thaumatin, HSA and Fab Fragment. A cost of 3,4 mol ATP per gram of DNA produced was assumed.

Substrate	Thaumatin	HSA	Fab Fragment
dAMP	1.14	0.86	1.18
dTMP	0.73	1.09	0.74
dGMP	0.48	0.72	0.57
dCMP	0.73	0.40	0.58
ATP(θ)	10.45	10.45	10.45

2. Arabitol Biosynthesis Pathway

In total, we added four reactions associated to this pathway (Cheng *et al*., 2014). First, Ribulose-5P is converted into D-Ribose by a kinase with the formation of ATP. Then, D-ribose is converted into D-arabitol by a dehydrogenase with the formation of NAD^+ from NADH. After D-arabitol is synthesized, it is transported to the extracellular medium and then "consumed" by an exchange reaction (Figure 8-3).

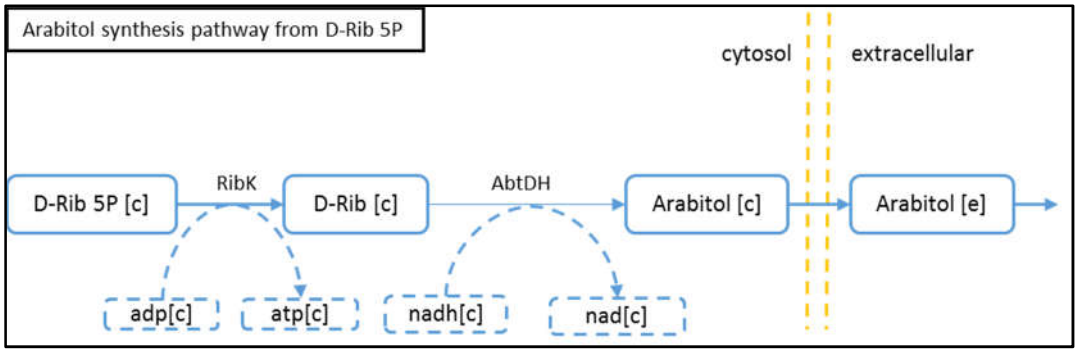


Figure 8-3 – D-Arabitol synthesis pathway from D-Ribulose-5-phosphate in *Pichia pastoris.*

3. Model manual curation

Initially, the main problem of the model is that it did not carry flux through the oxidative part of the Pentose Phosphate Pathway (PPP), the main source of the reducing cofactor NADPH. Instead, this cofactor was synthesized by a cytosolic NADP-dependent isocitrate dehydrogenase, which was also the main source of α-ketoglutarate in the cytosol (data not shown). Fluxomic studies in aerobic glucose-limited conditions in *P. pastoris* (Baumann *et al*., 2010; Dragosits *et al*., 2009; Heyland *et al*., 2010), have shown that about 40% of the carbon that reaches glucose 6 phosphate is carried through the oxidative branch of the PPP, which is also thermodynamically favorable (Nelson & Cox, 2008). Moreover, α-ketoglutarate is considered to be produced in the mitochondria and then exported to the cytosol for nitrogen fixation and anabolic reactions.

These inconsistencies have been recently addressed for several genome scale metabolic models of *Saccharomyces cerevisiae* (Pereira *et al*., 2016). Therefore, we performed the following changes to our reconstruction according to the indication of the authors:

i. We enabled the transport of α-ketoglutarate from the mitochondria to the cytosol using transporters present in *P. pastoris* (Rußmayer *et al*., 2015; Tomàs-Gamisans *et al*., 2016).

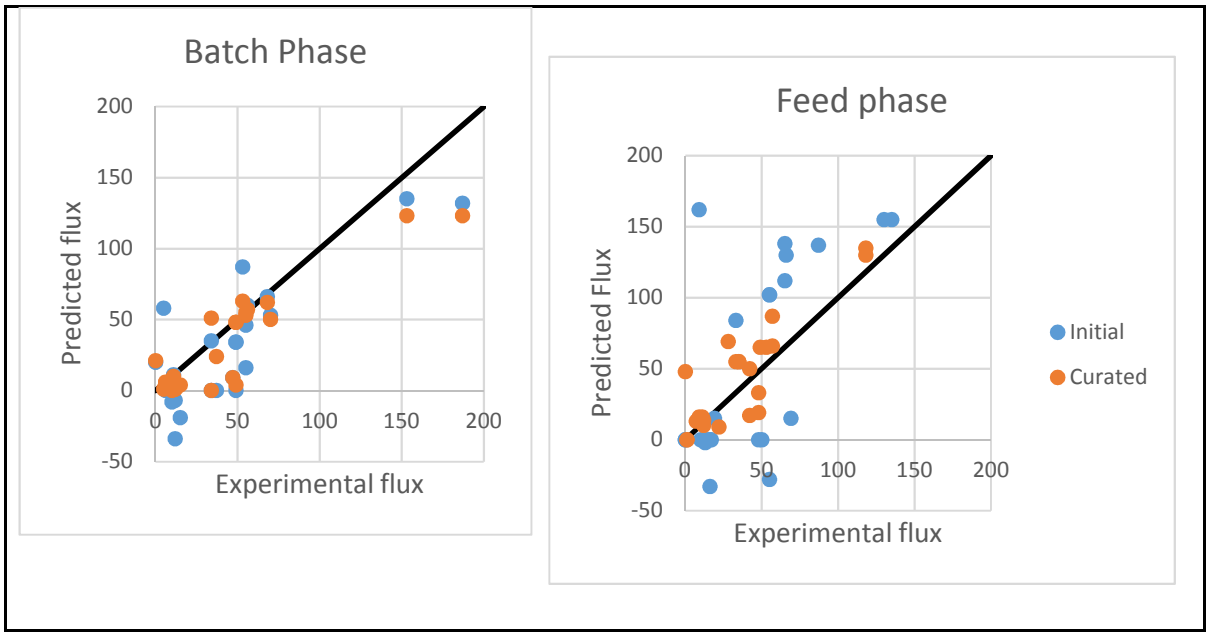
ii. The flux through three symporters that passively carried protons against the electrochemical gradient in the mitochondria was blocked in the direction of export to the cytosol.

iii. Based on the assumption postulated by Satrustegui *et al* (Satrustegui, Bautista, & Machado, 1983) for *Saccharomyces cerevisiae*, we considered that the NAD^+/NADH and NADPH/$NADP^+$ ratios in aerobic glucose-limited conditions are high enough to block the flux towards the formation of NAD^+ and NADPH. Therefore, we blocked 30 cytosolic reactions in the direction of either NAD^+ or NADPH formation. The only reactions left to produce cytosolic NADPH were the ones from the PPP and the acetate-forming acetaldehyde dehydrogenase, whose presence has been experimentally determined in aerobic glucose-limited cultivations of *P. pastoris* (Heyland *et al*., 2011).

Applying these modifications resulted in a spontaneous flux through the PPP, a mitochondrial formation of α-ketoglutarate with its subsequent secretion to the cytosol and an overall concordance in the direction of the fluxes with respect to experimental data. This makes the model a reasonable approximation of *Pichia pastoris* central carbon metabolism.

When the flux distribution derived from the fed-batch robustness check dataset was compared with the fluxomic data obtained by Heyland *et al* (Heyland *et al*., 2010) (equivalent conditions), the average error in the prediction of 23 fluxes of the central metabolism dropped three times with respect to the predictions made by the not curated model - from 128% to 39% for the exponential batch phase and from 160% to 63% for the controlled feed phase (Figure 8-4).This drop was mainly caused by the change in the direction (from negative to positive flux) of the non-oxidative part of the PPP, the spontaneous flux through the oxidative branch of this pathway and the reduction in the predicted influx of oxaloacetate to the cell. The overall agreement in directionality can be seen in Figure 8-4 by the elimination of negative predicted fluxes in the curated model.

Figure 8-4 - Predicted versus experimental fluxes of the central metabolism. The flux distributions determined by Heyland *et al* (Heyland *et al*., 2010) during an aerobic glucose limited fermentation were compared to the output of the model in equivalent stages of a cultivation (exponential and controlled growth phases) during the experiment used for checking fed-batch model robustness. Values are presented as normalized to carbon uptake and the black line represents the unit.



4. Model Performance

Usability and similarity to experimental chemostat data were used as criteria to select the most appropriate genome-scale model for building the dynamic framework. In terms of usability, we verified that the models had an adequate annotation, *i.e.* balanced equations, intuitive metabolite and reaction names, compartmentalization, functional gene-reaction associations and adequate representation of the central metabolism, among others. We then evaluated model similarity to experimental data from two chemostats (Table 8-11) using the normalized square differences between experimental and simulated rates (Equation 5):

$$F_i = \sum_{j=1}^{2} \frac{1}{n_j} \cdot \sum_{k=1}^{n_j} \frac{\sqrt{\left(v_{exp_{k,j}} - v_{mod_{k,j}}\right)^2}}{v_{exp_{k,j}}} \quad (27)$$

Here, F is the overall fitting relative error of model i, n_J corresponds to the number of predicted rates determined in each dataset (12 in dataset 1 and 30 in dataset 2). Also, v_{exp_k} corresponds to the vector of experimental rates of condition k in dataset j and $v_{mod_{k,j}}$ is the model's estimation of the experimental rates of condition k in dataset j.

For each prediction, we first constrained each model with n_J-1 experimental rates. Then, Flux Balance Analysis (FBA) (Orth, Thiele, & Palsson, 2010) was performed using biomass maximization as objective function to predict the remaining one.

It is worthy to note that whenever a model yielded an infeasible solution (due to carbon imbalance) or erroneously predicted the production of a compound under certain experimental condition, an error of 100% was assumed for that particular rate.

The model that gave best predictions compared to experimental data was chosen as the basis for the dynamic model. We tested the iFS670 model against three genome-scale metabolic models of *Pichia pastoris* that were available at the beginning of this study: the iPP668(Chung *et al*., 2010), the iLC915 (Caspeta *et al*., 2012) and the PpaMBEL1254 (Sohn *et al*., 2010)

Table 8- 11 Chemostat data used for model selection

Set	Type of data	Rates	Conditions	Reference
1	Glycerol- and/or methanol-limited chemostats	5	4	(Solà *et al*., 2007)
2	Glucose-limited chemostats at different oxygen levels	7	6	(Carnicer *et al*., 2009)

The main components and the relevant usability features of published GSMs of *Pichia pastoris* are detailed in Table 8-12. The PpaMBEL1254 model was discarded due to the lack of intuitive reaction and metabolite names in the online version, as well as the absence of gene-protein relations (at least in the online version), hampering the analysis of knock-

out strains. All the models share the same structure of the central metabolism, which carries most of the flux entering the cell.

Table 8- 12 Main components and usability features of available genome-scale metabolic models of *Pichia pastoris*

	iPP668	**iFS670**	**PpaMBEL1254**	**iLC915**
Number of genes	669	670	540	915
Reactions	1354	1383	1254	1426
Metabolites	1177	1195	1058	1302
Compartments	8	8	8	6
Platform used for analysis	Cobra	Cobra	Cobra	Raven
Intuitive nomenclature for reactions and metabolites	Yes	Yes	No	No
Capable of performing Single Gene deletions	Yes	Yes	No	Yes
Capable of automatically checking mass balance	No	No	No	Yes

After the determination of the average relative error between model predictions and experimental data (Carnicer *et al*., 2009; Solà *et al*., 2007) (Table 8-13), we selected the iFS670 model since it has a desirable structure and better reproduces experimental data from *P. pastoris* chemostats. It is worth mentioning that the inclusion of the arabitol biosynthesis pathway into Chung's (iPP668) model – resulting in the iFS670 model - greatly improved the predictions of specific growth rate, Oxygen Uptake Rate (OUR) and Carbon Dioxide Evolution Rate (CER) in hypoxic glucose-limited chemostats (Figure 8-5 and Figure 8-6).Specifically, the deviation of carbon towards arabitol reduced the predicted growth rate in those conditions when compared to the iPP668 model, resulting in a reduction of the difference with the corresponding experimental value.

Table 8- 13 Average error of model predictions using two datasets from carbon-limited chemostats. In glycerol- and/or methanol (MetOH) – limited chemostats, the models were employed to predict specific growth rate μ, Oxygen Uptake Rate (OUR) and carbon dioxide evolution rate (CER) in four different conditions, which gives a total of 12 predictions. In the glucose limited chemostats, the models were used to estimate μ, OUR, CER, ethanol secretion rate and arabitol secretion rate in six conditions, which gives a total of 30 model predictions. Experimental data was taken from (Carnicer *et al*., 2009; Solà *et al*., 2007)

Carbon Source	**iLC915**	**iFS670**	**iPP669**	**Number of predictions**
Glycerol/MetOH	78%	37%	38%	12
Glucose	85%	36%	52%	30
Overall Error (F)	83%	36%	48%	42

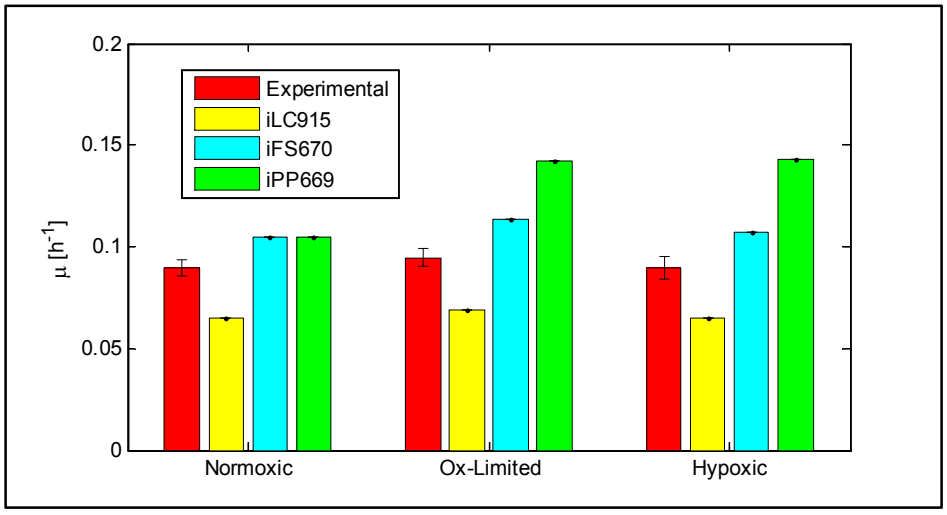


Figure 8-5 - Experimental and model-predicted specific growth rates using glucose as the only carbon source at different oxygen levels for a *P. pastoris* wild type strain. Data taken from (Carnicer *et al*., 2009)

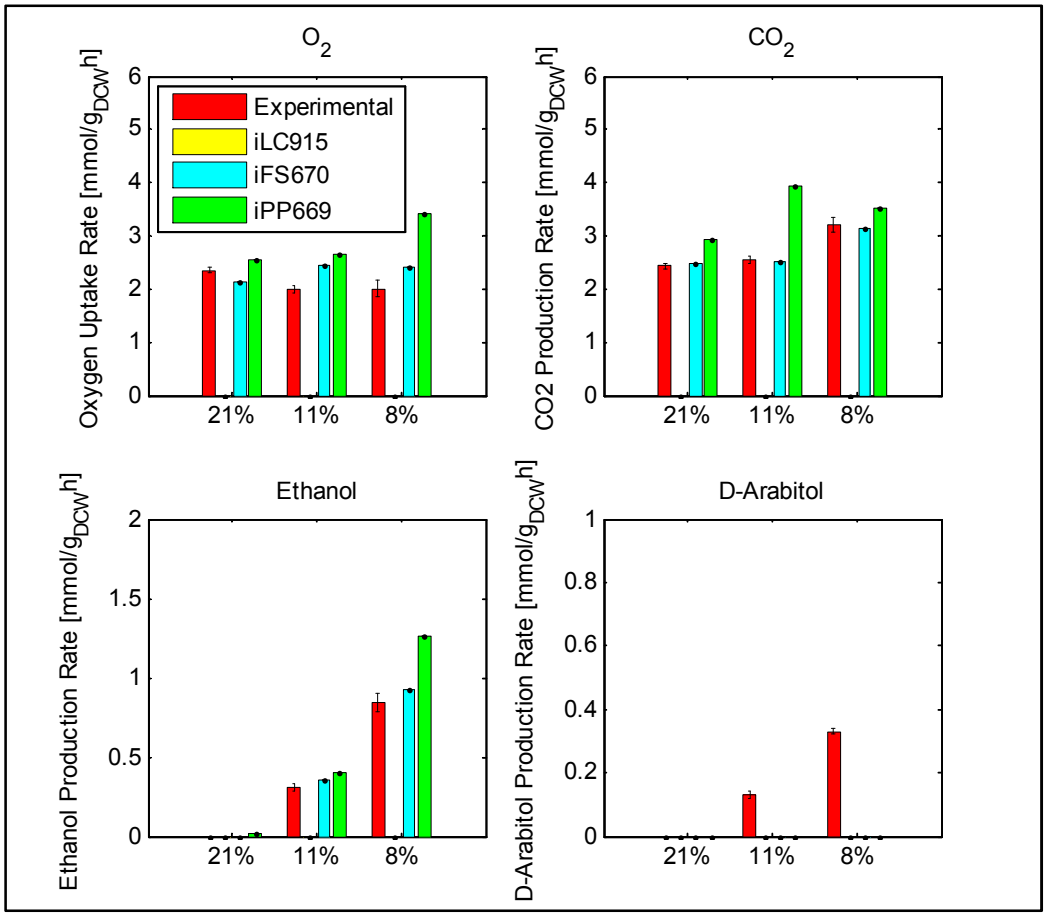


Figure 8-6 – Prediction of Gas exchange and secondary metabolite production by the tested models. The percentage in the x axis correspond to oxygen fraction in the gas inlet of the bioreactor used to perform the study (21% → normoxic, 11% → oxygen limited, 8% → hypoxic). Data taken from

S 3-3 Evaluation of feeding policies

We tested 13 different feeding strategies, which yielded a constant or decreasing growth rate (Figure 8-7).The details of the strategies are presented in Table 8-14.

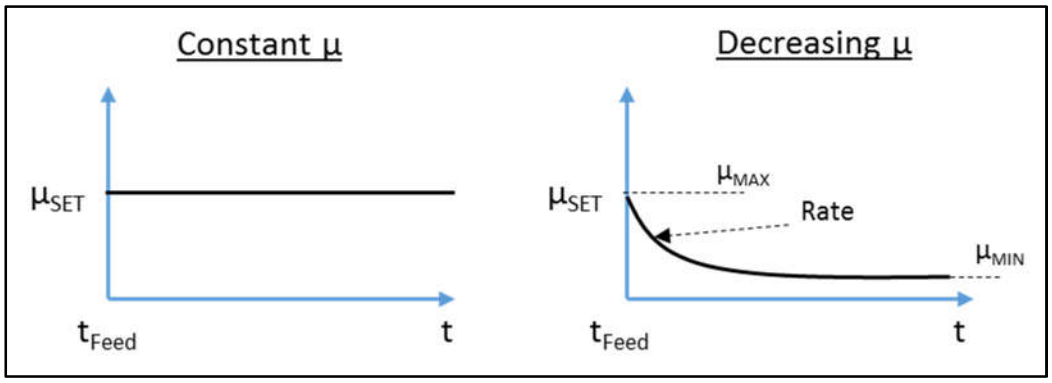


Figure 8-7 – Constant (left) versus decreasing (right) growth rates during fed-batch culture. Here, t_{FEED} corresponds to the time when the feed of the culture starts after batch cultivation. μ_{MAX}, Rate and μ_{MIN} refer to the parameters used to describe the decreasing growth rate profile of each culture. We evaluated two values for each one of these parameters, which yielded eight dynamic feeding strategies (6-13 in Table 8-14)

Table 8- 14 Feeding strategies evaluated and productivity indicators. The first five strategies attempt to make the culture grow at a constant growth rate while the rest produce a decreasing growth rate profile.

Strategy	$\mu_{SET.MAX}$ [h^{-1}]	Rate	$\mu_{SET.MIN}$ [h^{-1}]	q_P [$mg/g_{DCW}h$]	X_{FINAL} [g/L]	P_{FINAL} [mg/L]	Limitation
1	0.14	-	-	2.85	164.8	138	Oxygen
2	0.12	-	-	2.59	187.8	135	Oxygen
3	0.1	-	-	2.32	195.3	130	Volume
4	0.08	-	-	2.29	191.3	138	Volume
5	0.06	-	-	2.28	184.7	154	Volume
6	0.14	0.07	0.08	2.13	193.1	121.6	Volume
7	0.14	0.07	0.04	1.33	176.6	92.3	Volume
8	0.14	0.01	0.08	2.83	197.5	150.0	Volume
9	0.14	0.01	0.04	2.34	195.1	128.0	Volume
10	0.1	0.07	0.08	1.88	191.0	111.6	Volume
11	0.1	0.07	0.04	0.89	172.8	66.9	Volume
12	0.1	0.01	0.08	1.41	193.7	81.3	Volume
13	0.1	0.01	0.04	2.30	188.8	140.5	Volume

S3-4 Initial Calibrations and parametric limitations

Batch model

The model was used to calibrate data from eight aerobic, glucose-limited batch cultivations. The parameter values achieved in the calibrations is presented in Table 8-15, while the time of calibration and objective function value of the calibration are presented in Table 8-16. Datasets correspond to the following strains:

- Datasets 1 and 2: parental GS115 strain
- Datasets 3 and 4: recombinant strain harboring one copy of the thaumatin gene
- Datasets 5 and 6: recombinant strain harboring five copies of the thaumatin gene
- Datasets 7 and 8: recombinant strain harboring eight copies of the thaumatin gene

Table 8- 15 Parameter values achieved in the calibration of data from eight batch cultivations using the initial batch model structure.

	Dataset							
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**
V_{MAX}	7.75	3.34	7.95	2.74	3.07	6.67	1.51	1.27
K_S	9.80E-04	9.60E-05	7e -4	3.00E-04	1.00E-05	1.5 e-4	1.00E-05	1.03E-05
$v_{EtOH,B}$	1.98	1.78	1.76	1.48	0.89	2.97	0.03	0.02
$v_{Pyr,B}$	0.21	0.18	0.25	0.19	0.12	0.2	0.01	0.003
$v_{Arab,B}$	0.54	0.32	0.5	0.51	0.4	0.48	0.16	0.09
$v_{Cit,B}$	0.06	0.04	0.08	1.1	0	0.05	0.03	0.05
α_B	4.05E-04	2.90E-04	4.20E-04	2.30E-04	1.45E-06	3.0E-04	7.19E-06	7.17E-05
m_{atp}	0.52	0.68	0.001	4.09	9.99	4.61	3.29	1.32

Table 8- 16 General features of initial batch model calibration.

Dataset	N	Min. squares difference	Time of calibration [h]
1	8	0.26	1.92
2	8	1.55	5.29
3	8	0.47	3.46
4	8	0.88	4.40
5	6	1.69	4.11
6	9	7.27	4.41
7	12	1.17	4.32
8	13	3.30	4.30

We also provide two examples of how the model fitted two of these cultivations.

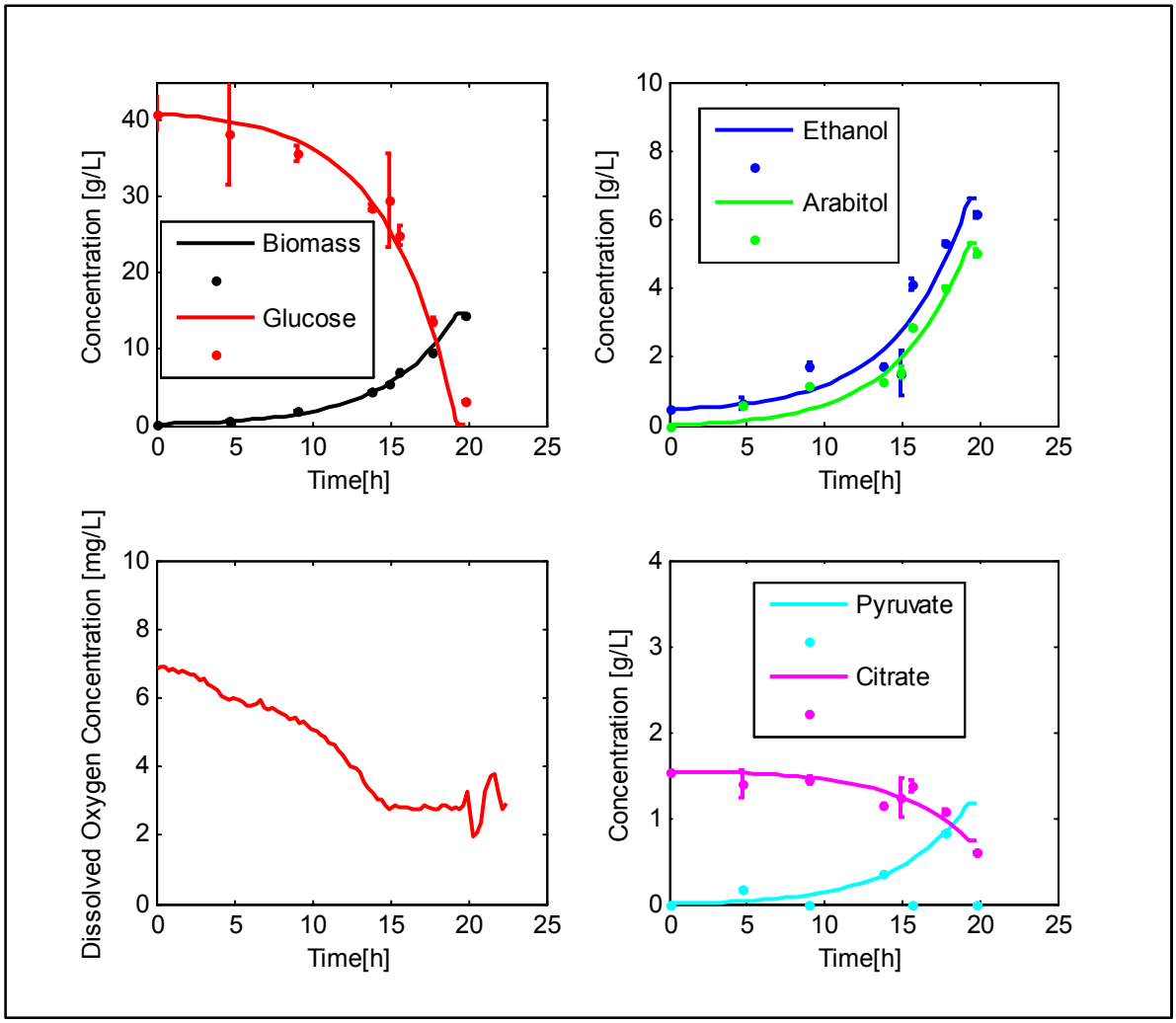


Figure 8-8 - Batch model calibration of GS115 culture 1

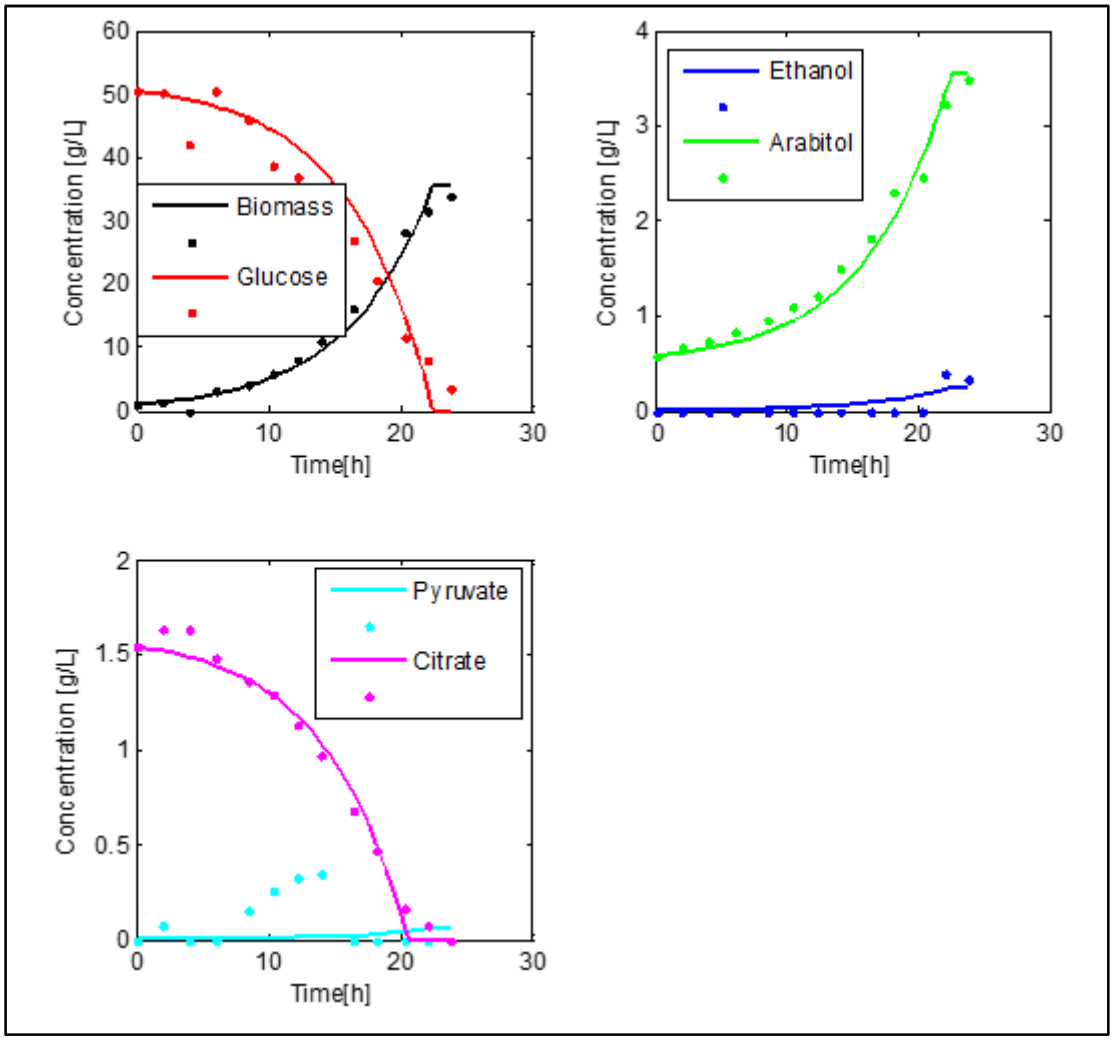


Figure 8-9 - Batch model calibration of GS115 culture 8

The recurrence of significance, sensitivity and identifiability issues found amongst the initial batch model calibrations is presented in Table 8-17 and Table 18.

Table 8- 17 Percentage (o Frequency) of calibrations (8 in total) where a parameter presented sensitivity or significance issues. Parameters with recurrent problems are highlighted.

	V_{MAX}	K_S	$v_{EtOH,B}$	$v_{Pyr,B}$	$v_{Arab,B}$	$v_{Cit,B}$	α_B	m_{atp}
Sensitivity	0	38	0	0	0	13	25	13
Significance	25	50	0	13	0	25	25	25

Table 8- 18 Percentage of calibrations (8 in total) where pairs of parameters show identifiability issues (correlation ≥ 0.95). Parameters with recurrent identifiability issues are highlighted.

	V_{MAX}	K_S	$v_{EtOH,B}$	$v_{Pyr,B}$	$v_{Arab,B}$	$v_{Cit,B}$	α_B	m_{atp}
V_{MAX}	-							
K_S	13	-						
$v_{EtOH,B}$	50	0	-					
$v_{Pyr,B}$	25	0	38	-				
$v_{Arab,B}$	25	0	25	13	-			
$v_{Cit,B}$	13	0	0	13	0	-		
α_B	25	25	0	0	0	13	-	
m_{atp}	63	25	38	25	38	13	50	-

Fed-batch model

Table 8-19 indicates the parameter values achieved for the three fed-batch cultivations used in the initial calibration of the fed-batch model. Figure 8-10, Figure-11 and Figure 8-12, show the model fits to the experimental data.

Table 8- 19 Parameter Values of the Initial calibrations performed with the complete fed-batch model (14 parameters)

Parameter	Dataset 1	Dataset 2	Dataset 3	Mean	Units
$v_{S,max}$	2.74	3.29	2.59	2.94	$mmol/g_{DCW}h$
K_S	0.05	0.03	0.07	0.05	g/L
$v_{EtOH.B}$	1.95	2.18	0.98	1.58	$mmol/g_{DCW}h$
$v_{Pyr.B}$	0.18	0.18	0.13	0.15	$mmol/g_{DCW}h$
$v_{Arab.B}$	0.5	0.24	0.11	0.18	$mmol/g_{DCW}h$
$v_{Cit.B}$	0.12	0.11	0.22	0.16	$mmol/g_{DCW}h$
$v_{EtOH.FB}$	1.13	1.20	1.22	1.21	$mmol/g_{DCW}h$
$v_{Pyr.FB}$	0.10	0.02	0.26	0.14	$mmol/g_{DCW}h$
$v_{Arab.FB}$	0.07	0.13	0.17	0.15	$mmol/g_{DCW}h$
$v_{Cit.FB}$	0.005	0.00	0.01	0.008	$mmol/g_{DCW}h$
α_B	3.03E-04	4.26E-05	1.49E-04	9.6E-05	[–]
α_{FB}	1.28E-04	2.22E-14	1.44E-04	7.2E-05	[–]
m_{ATP}	4.38	9.00	8.13	8.6	$mmol/g_{DCW}h$
T_{Fed}	23	22.02	22.94	22.5	h

Table 8- 20 General features of initial batch model calibration.

Dataset	N	Min. squares difference	Time of calibration [h]
1	22	3.84	29.05
2	21	4.35	30.75
3	22	4.17	29.21

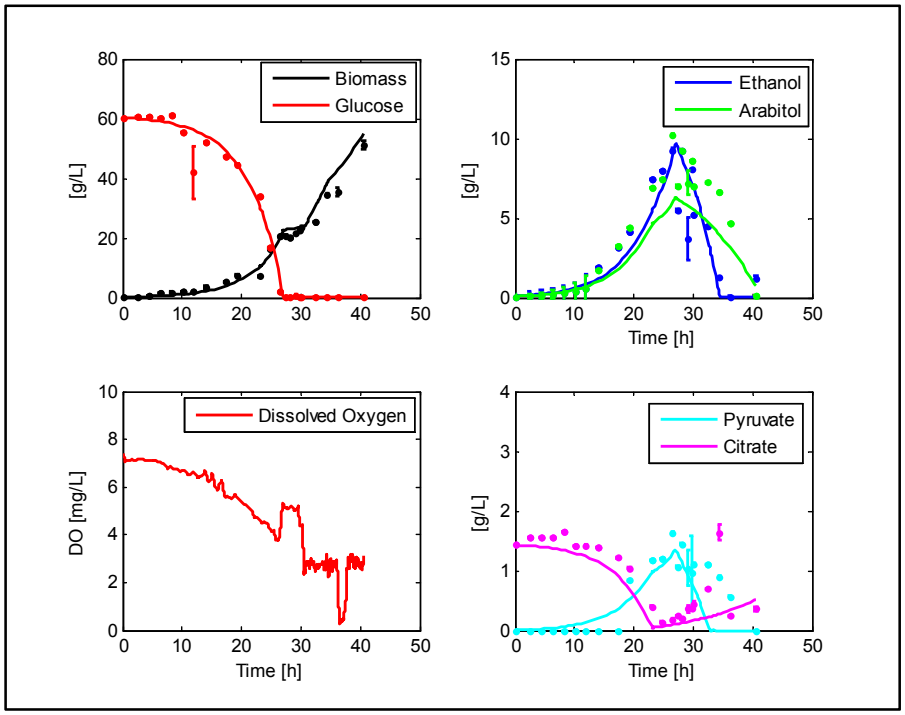


Figure 8-10 - Calibration of fed-batch dataset 1 using the original model structure

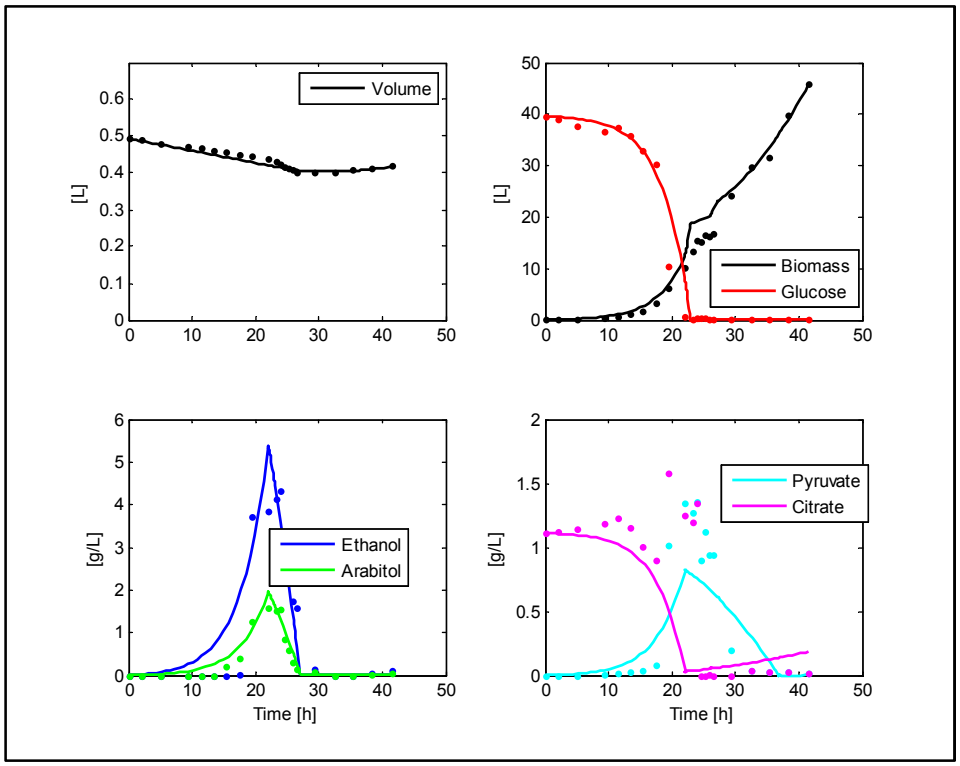


Figure 8-11 - Calibration of fed-batch dataset 2 using the original model structure

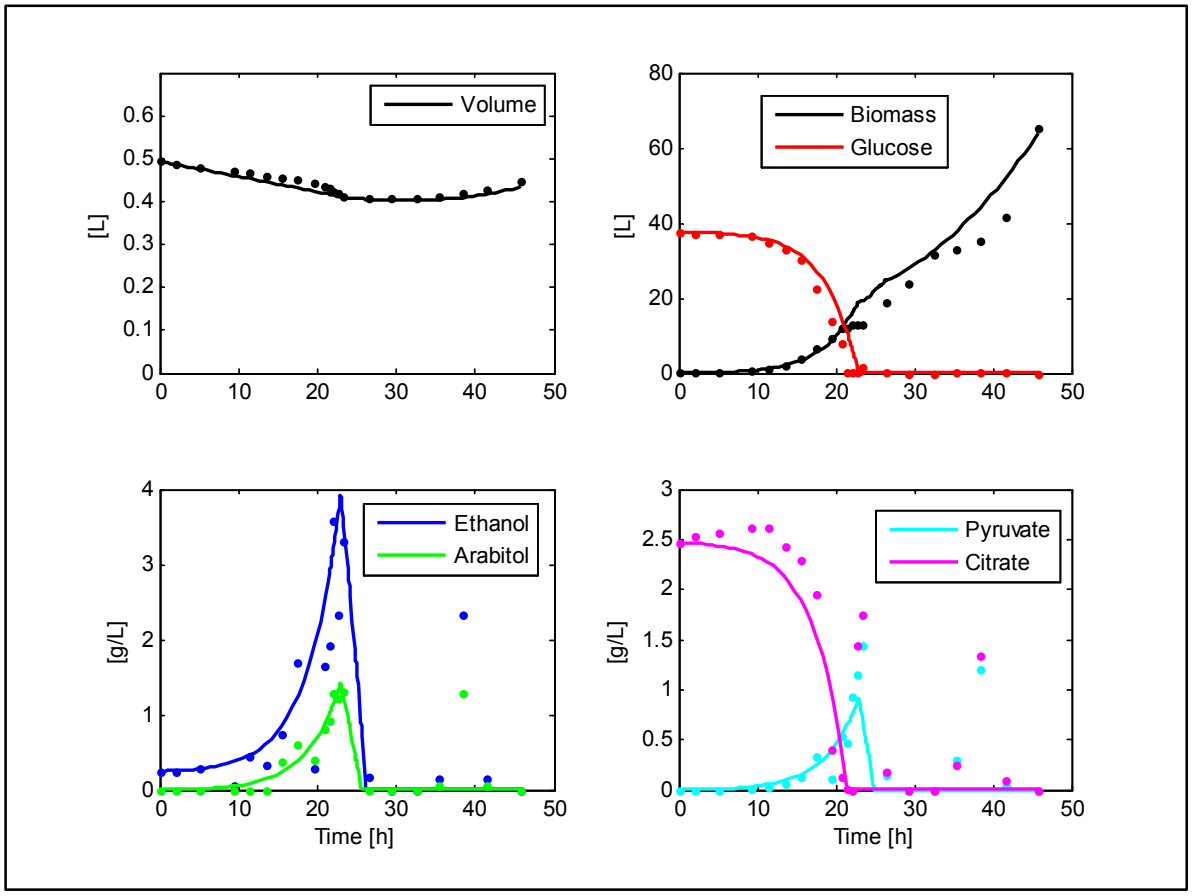


Figure 8-12 - Calibration of fed-batch dataset 3 using the original model structure

The recurrence of significance, sensitivity and identifiability issues found amongst the initial fed-batch model calibrations is presented in Table 8-21 and Table 8-22.

Table 8- 21 Percentage of times a parameter of the model presented sensitivity or significance problems out of a total of three model calibrations. Parameters with sensitivity or significance issues are highlighted.

	$v_{S,max}$	K_S	$v_{Et,B}$	$v_{Py,B}$	$v_{Ar,B}$	$v_{Ci,B}$	$v_{Et,FB}$	$v_{Py,FB}$	$v_{Ar,FB}$	$v_{Ci,FB}$	α_B	α_{FB}	m_{ATP}	T_{Cons}
Sensitivity	0	0	0	0	0	0	0	0	0	33	0	67	0	0
Significance	0	33	0	0	0	0	0	33	0	100	67	100	0	0

Table 8- 22 Frequency (in %) with which a pair of parameters presented identifiability issues in the initial modeling structure of fed-batch cultures of *Pichia pastoris* (3 datasets). Parameters with recurrent identifiability issues are highlighted.

	$v_{S,max}$	K_S	$v_{EtOH,B}$	$v_{Pyr,B}$	$v_{Arab,B}$	$v_{Cit,B}$	$v_{EtOH,FB}$	$v_{Pyr,FB}$	$v_{Arab,FB}$	$v_{Cit,FB}$	α_B	α_{FB}	m_{ATP}	T_{Cons}
$v_{S,max}$														
K_S	0													
$v_{EtOH,B}$	67	0												
$v_{Pyr,B}$	0	0	0											
$v_{Arab,B}$	33	0	33	0										
$v_{Cit,B}$	0	0	33	0	0									
$v_{EtOH,Fb}$	0	0	33	0	0	0								
$v_{Pyr,FB}$	0	0	0	0	0	0	0							
$v_{Arab,FB}$	0	0	33	0	0	33	33	0						
$v_{Cit,FB}$	0	0	0	0	0	0	0	0	33					
α_B	0	0	0	0	0	0	0	0	33	0				
α_{FB}	0	0	0	0	0	0	0	0	0	0	0			
m_{ATP}	0	0	0	0	0	0	0	0	0	0	0	0		
T_{Cons}	0	0	0	0	0	0	0	0	33	0	33	0	0	

S 3-5 Absence of parametric problems in the Robustness Check Datasets and Goodness of Fit

Batch Model

Table 8-23 and Table 8-24 show that the calibration of the Robustness Check dataset yielded no identifiability, sensitivity (correlation > 0.95 between parameters) and sensitivity issues. Figure 8-13 shows the results of the goodness of fit analysis for this dataset

Table 8- 23 Correlation Matrix of the robust parameter set used to calibrate the batch validation dataset. Each cell contains the correlation between the two corresponding parameters.

	$v_{EtOH,B}$	$v_{Pyr,B}$	$v_{Arab,B}$	$v_{Cit,B}$	α_B
$v_{EtOH,B}$	1	-0.45	0.43	0.78	-0.86
$v_{Pyr,B}$	-0.45	1	-0.17	-0.56	0.50
$v_{Arab,B}$	0.43	-0.17	1	0.38	-0.43
$v_{Cit,B}$	0.78	-0.56	0.38	1	-0.88
α_B	-0.86	0.50	-0.43	-0.88	1

Table 8- 24 Sensitivity Matrix of the robust Parameter set used to calibrate the batch validation dataset. Each cell contains the average sensitivity of a particular parameter over the state variables.

	Volume	**Biomass**	**Glucose**	**Ethanol**	**Pyruvate**	**Arabitol**	**Citrate**
$v_{EtOH,B}$	0	0.15	0.12	0.91	0.14	0.14	0.04
$v_{Pyr,B}$	0	0.02	0.00	0.02	0.95	0.02	0.00
$v_{Arab,B}$	0	0.03	0.01	0.02	0.02	0.97	0.00
$v_{Cit,B}$	0	0.03	0.00	0.02	0.03	0.03	0.45
α_B	0	2.62	1.09	1.80	1.93	2.00	0.73

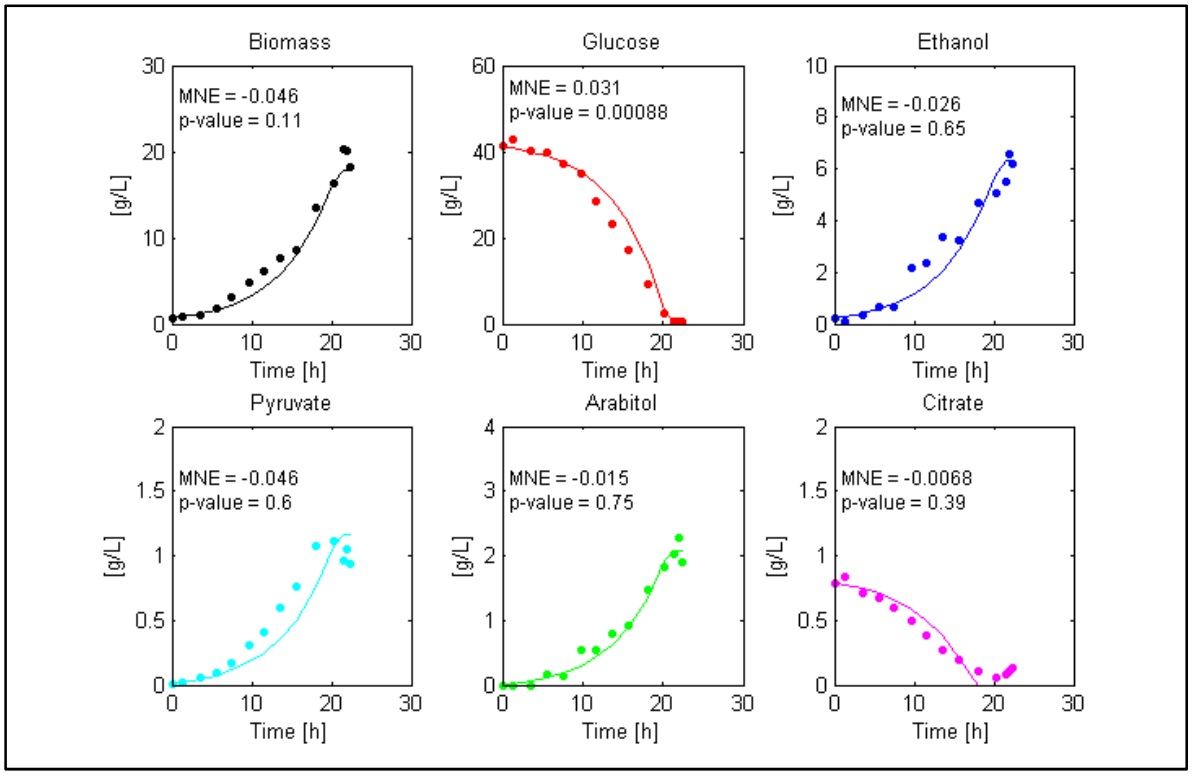


Figure 8-13 - Goodness of fit of the batch model to the Robustness Check dataset. This figure shows the model fit to experimental data along with the mean normalized error and the p-value of the Anderson-Darling test associated to each variable.

Fed-Batch Model

Table 8-25 and Table 8-26 indicate that the calibration using Structure 3 of new fed-batch data yielded no parametric problems. Figure 8-14 shows the results of the goodness of fit analysis for this dataset.

Table 8- 25 Correlation Matrix of the calibration of the fed-batch validation dataset. Recall that we left the parameters associated to Pyruvate dynamics out of the analysis since it was not produced during the cultivation

	v_{MAX}	K_S	$v_{Pyr,B}$	$v_{Arab,B}$	$v_{Cit,B}$	$v_{Pyr,FB}$	α_B	α_{FB}	m_{ATP}	T_{Cons}
v_{MAX}	1	0.15	-	-0.47	0.45	-	-0.92	0.14	-0.85	0.04
K_S	0.15	1	-	-0.85	0.62	-	-0.20	0.68	0.24	-0.32
$v_{Pyr,B}$	-	-	-	-	-	-	-	-	-	-
$v_{Arab,B}$	-0.47	-0.85	-	1	-0.66	-	0.44	-0.70	0.07	0.55
$v_{Cit.B}$	0.45	0.62	-	-0.66	1	-	-0.51	0.60	-0.20	-0.10
$v_{Pyr.FB}$	-	-	-	-	-	-	-	-	-	-
α_B	-0.92	-0.20	-	0.44	-0.51	-	1	-0.18	0.88	-0.08
α_{FB}	0.14	0.68	-	-0.70	0.60	-	-0.18	1	0.21	-0.42
m_{ATP}	-0.85	0.24	-	0.07	-0.20	-	0.88	0.21	1	-0.25
T_{Cons}	0.04	-0.32	-	0.55	-0.10	-	-0.08	-0.42	-0.25	1

Table 8- 26 Sensitivity Matrix of the calibration of the fed-batch validation dataset. All of the included parameters have a significant impact in at least one of the state variables.

	Volume	**Biomass**	**Glucose**	**Ethanol**	**Pyruvate**	**Arabitol**	**Citrate**
v_{MAX}	0.00	0.65	0.93	4.32	3.67	1.57	0.61
K_S	0.00	0.22	0.00	0.01	1.24	0.02	0.00
$v_{Pyr,B}$	-	-	-	-	-	-	-
$v_{Arab,B}$	0.00	0.10	0.09	0.69	2.74	0.91	0.22
$v_{Cit,B}$	0.00	0.03	0.00	0.09	0.96	0.04	0.62
$v_{Pyr,FB}$	-	-	-	-	-	-	-
α_B	0.00	1.03	0.88	5.82	6.27	2.91	1.24
α_{FB}	0.00	0.03	0.00	0.01	0.00	0.02	0.00
m_{ATP}	0.00	0.79	0.65	3.24	3.66	2.13	0.95
T_{Cons}	0.00	0.26	0.00	0.22	3.66	3.06	0.88

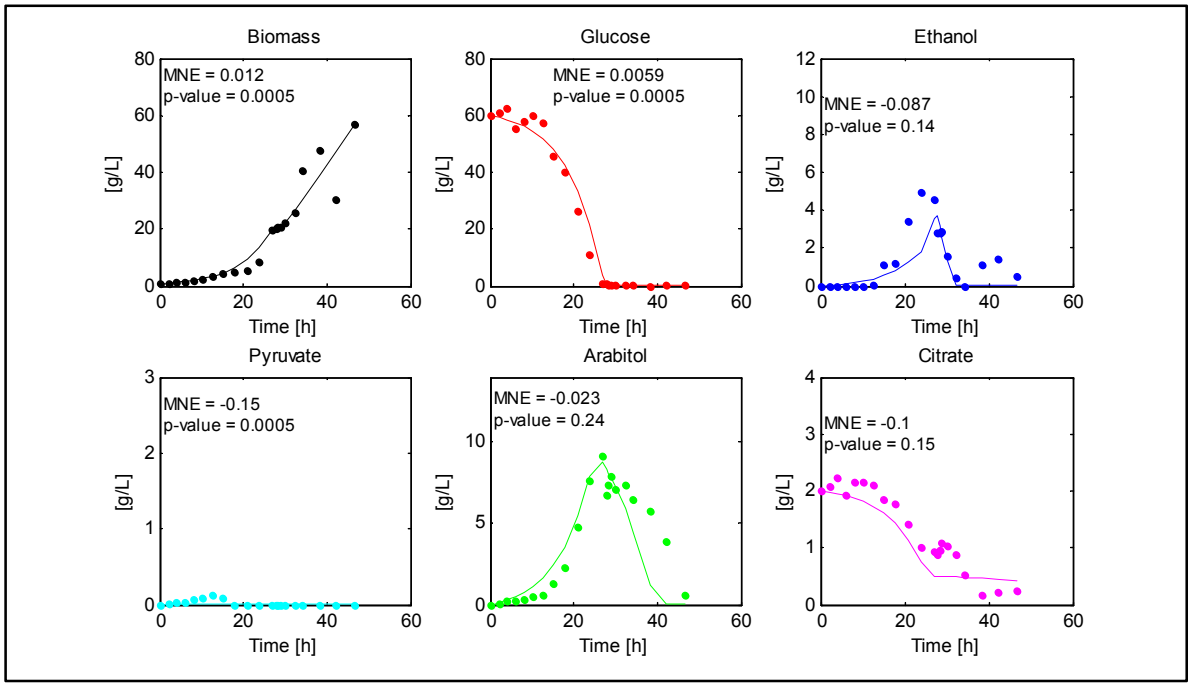


Figure 8-14 - Goodness of fit of the batch model to the Robustness Check dataset. This figure shows the model fit to experimental data along with the mean normalized error and the p-value of the Anderson-Darling test associated to each variable.

S 3-6: Goodness of Fit analysis of the Validation datasets

Figure 8-15 and Figure 8-16 present the mean normalized error and the p-value of the Anderson-Darling test for each state variable in the validation cultures for the batch and fed-batch models, respectively.

Batch model

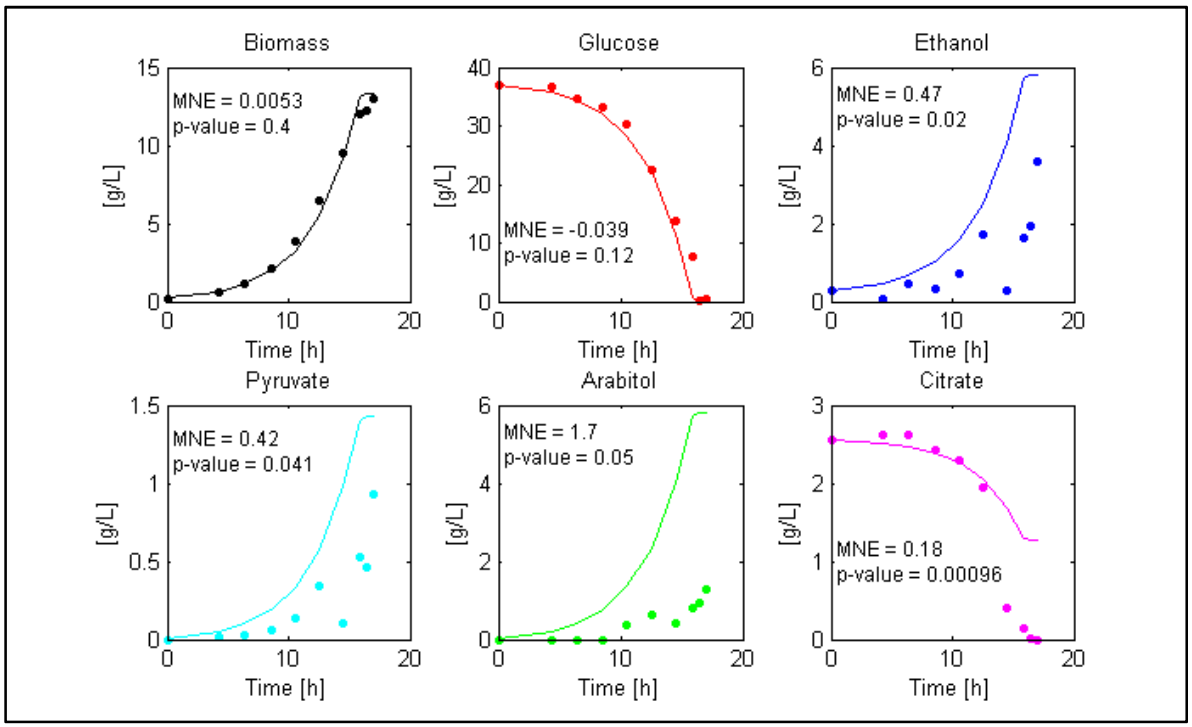


Figure 8-15 - Goodness of fit analysis of the batch model prediction of the Validation dataset.

Fed-batch model

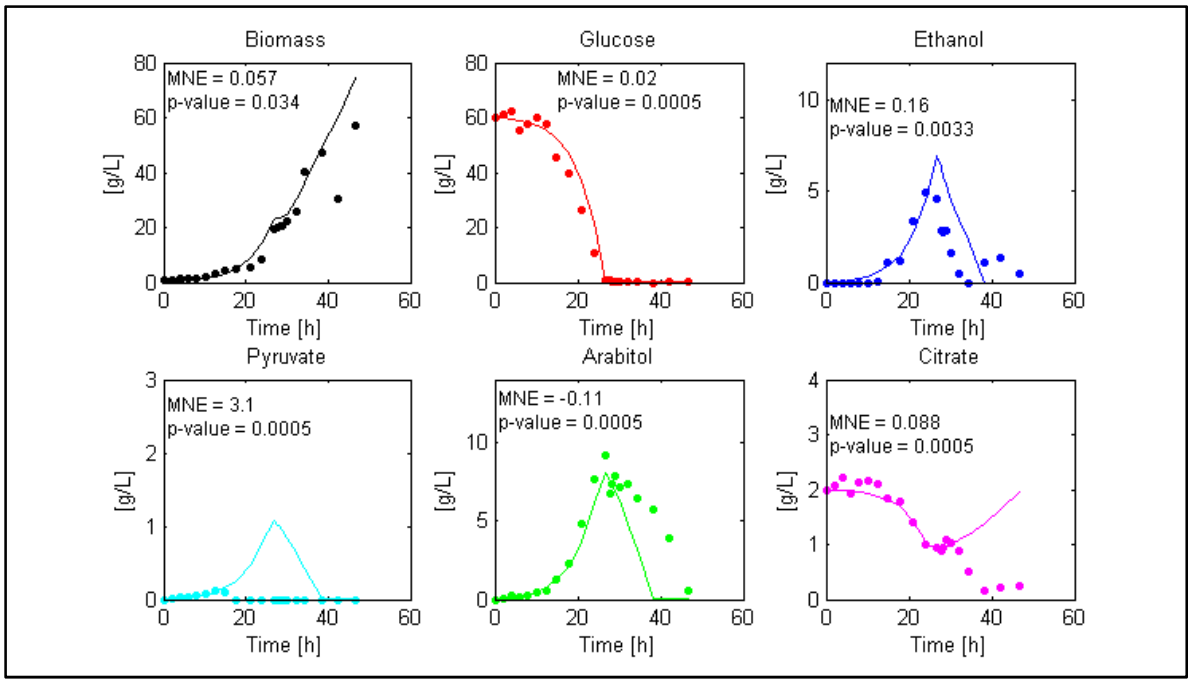


Figure 8-16 - Goodness of fit analysis of the fed-batch model prediction of the Validation dataset.

S 3-7 – Knockout Candidates derived using MOMA

Table 8- 27 Knockout candidates for HSA overproduction

	Deleted Gene	Final Biomass [g/L]	Final HSA [g/L]	Reaction(s) Name(s)
1	PAS_chr2-2_0094	11.81	0.91	Chitin synthase
2	PAS_chr1-4_0194	11.41	0.84	Putrescine and Spermidine transport
3	PAS_chr4_0836	11.41	0.84	Putrescine and Spermidine transport
4	RPPA11109	10.74	0.63	ribulose 5-phosphate 3-epimerase
5	RPPA11110	10.74	0.63	ribulose 5-phosphate 3-epimerase
6	PAS_chr2-2_0330	10.20	0.62	Phosphoryl ceramide syntase
7	PAS_chr2-2_0044	8.59	0.30	CDP-Diacylglycerol synthetase, yeast-specific
8	PAS_chr4_0210	8.84	0.29	ADP/ATP transporter, mitochondrial
9	PAS_chr4_0212	8.84	0.29	ribose-5-phosphate isomerase
10	PAS_chr3_0604	14.05	0.23	Deoxyribokinase and ribokinase
11	PAS_chr4_0408	15.33	0.22	phosphoethanolamine cytidyltransferase
12	PAS_chr1-1_0418	16.02	0.21	Acetate transporter
13	PAS_chr1-3_0220	15.88	0.19	Methylenetetrahydrofolate dehydrogenase NAD
14	PAS_chr1-4_0487	14.85	0.17	Succinate Dehydrogenase
15	PAS_chr2-2_0278	14.85	0.17	Peptide alpha-N-acetyltransferase
16	PAS_chr3_1110	14.85	0.17	Tyrosyl-tRNA synthetase, mitochondrial
17	PAS_chr4_0733	14.85	0.17	Succinate Dehydrogenase
18	PAS_chr3_0646	14.04	0.14	Phospholipase D, yeast-specific
19	PAS_chr3_0471	15.44	0.12	aspartate-semialdehyde dehydrogenase, irreversible
20	PAS_chr2-1_0657	13.40	0.12	phosphoglycerate dehydrogenase
21	PAS_chr4_0284	13.40	0.12	ribonucleoside-diphosphate reductase
22	PAS_chr4_0877	16.73	0.05	malate, succinate and fumarate transport, mitochondrial
23	PAS_chr3_0176	16.53	0.05	N-acteylglutamate synthase and ornithine transacetylase , mitochondrial
24	PAS_chr1-1_0050	9.78	0.05	Pyruvate dehydrogenase
25	PAS_chr1-4_0254	9.78	0.05	Ppyruvate dehydrogenase
26	PAS_chr1-4_0593	9.78	0.05	Pyruvate dehydrogenase, tetrahydrofolate aminomethyltransferase
27	PAS_chr2-2_0288	9.78	0.05	Arginase
28	PAS_chr3_0649	16.26	0.03	Thiamine transport in via proton symport
29	PAS_chr2-2_0127	17.10	0.03	Cytochrome c peroxidase, mitochondrial
30	PAS_chr1-4_0659	17.10	0.03	Hydrogen peroxide reductase thioredoxin, peroxisomal
31	PAS_chr2-1_0547	16.70	0.03	3',5'-bisphosphate nucleotidase
32	PAS_chr3_0462	17.14	0.03	Alanyl-tRNA synthetase

Table 8- 28 Reactions and pathways associated to the deletion candidates

	Deleted Gene	Reactions	Pathway
1	PAS_chr2-2_0094	udpacgam[c] => h[c] + udp[c] + chitin[c]	Glutamate metabolism
2	PAS_chr1-4_0194	h[c] + ptrc[e] => h[e] + ptrc[c] h[c] + spmd[e] => h[e] + spmd[c] h[c] + sprm[e] => h[e] + sprm[c]	Transport, Extracellular
3	PAS_chr4_0836	h[c] + ptrc[e] => h[e] + ptrc[c] h[c] + spmd[e] => h[e] + spmd[c] h[c] + sprm[e] => h[e] + sprm[c]	Transport, Extracellular
4	RPPA11109	ru5p-D[c] <=> xu5p-D[c]	Pentose Phosphate Pathway
5	RPPA11110	ru5p-D[c] <=> xu5p-D[c]	Pentose Phosphate Pathway
6	PAS_chr2-2_0330	ptd1ino_PP[c] + cer1_24[c] => 12dgr_PP[c] + ipc124_PP[c]	Sphingolipid Metabolism
7	PAS_chr2-2_0044	h[c] + pa_PP[c] + ctp[c] <=> ppi[c] + cdpdag_PP[c] h[m] + ctp[m] + pa_PP[m] <=> ppi[m] + cdpdag_PP[m]	Phospholipid Biosynthesis
8	PAS_chr4_0210	h[c] + adp[c] + atp[m] => h[m] + atp[c] + adp[m]	Transport, Mitochondrial
9	PAS_chr4_0212	r5p[c] <=> ru5p-D[c]	Pentose Phosphate Pathway
10	PAS_chr3_0604	atp[c] + rib-D[c] => h[c] + adp[c] + r5p[c]	Pentose Phosphate Pathway
11	PAS_chr4_0408	h[c] + ctp[c] + ethamp[c] => ppi[c] + cdpea[c]	Phospholipid Biosynthesis
12	PAS_chr1-1_0418	ac[e] <=> ac[c]	Transport, Extracellular
13	PAS_chr1-3_0220	nad[c] + mlthf[c] => nadh[c] + methf[c]	Folate Metabolism
14	PAS_chr1-4_0487	fad[m] + succ[m] <=> fadh2[m] + fum[m] q6[m] + succ[m] <=> q6h2[m] + fum[m] q6[m] + fadh2[m] <=> q6h2[m] + fad[m]	Citric Acid Cycle/Oxydative Phosphorilation
15	PAS_chr2-2_0278	accoa[c] + pepd[c] => h[c] + coa[c] + apep[c]	Other Amino Acid Metabolism
16	PAS_chr3_1110	atp[m] + tyr-L[m] + trnatyr[m] => amp[m] + ppi[m] + tyrtrna[m]	tRNA charging
17	PAS_chr4_0733	fad[m] + succ[m] <=> fadh2[m] + fum[m] q6[m] + succ[m] <=> q6h2[m] + fum[m] q6[m] + fadh2[m] <=> q6h2[m] + fad[m]	Citric Acid Cycle/Oxydative Phosphorilation
18	PAS_chr3_0646	h2o[c] + pc_PP[c] => h[c] + pa_PP[c] + chol[c]	Phospholipid Metabolism
19	PAS_chr3_0471	h[c] + nadph[c] + 4pasp[c] => pi[c] + nadp[c] + aspsa[c]	Alanine and Aspartate Metabolism
20	PAS_chr2-1_0657	nad[c] + 3pg[c] => h[c] + nadh[c] + 3php[c]	Glycine and Serine Metabolism
21	PAS_chr4_0284	19 Reactions	Nucleotide Salvage Pathway
22	PAS_chr4_0877	pi[m] + mal-L[c] <=> pi[c] + mal-L[m]	Transport, Mitochondrial
23	PAS_chr3_0176	accoa[m] + glu-L[m] => h[m] + coa[m] + acglu[m] glu-L[m] + acorn[m] => acglu[m] + orn[m]	Arginine and Proline Metabolism
24	PAS_chr1-1_0050	nad[m] + coa[m] + pyr[m] => nadh[m] + co2[m] + accoa[m]	Glycolysis/Gluconeogenesis
25	PAS_chr1-4_0254	nad[m] + coa[m] + pyr[m] => nadh[m] + co2[m]	Glycolysis/Gluconeogenesis

		+ accoa[m]	
26	PAS_chr1-4_0593	udpacgam[c] => h[c] + udp[c] + chitin[c]	Glycolysis/Gluconeogenesis
27	PAS_chr2-2_0288	h[c] + ptrc[e] => h[e] + ptrc[c] h[c] + spmd[e] => h[e] + spmd[c] h[c] + sprm[e] => h[e] + sprm[c]	Arginine and Proline Metabolism
28	PAS_chr3_0649	h[c] + ptrc[e] => h[e] + ptrc[c] h[c] + spmd[e] => h[e] + spmd[c] h[c] + sprm[e] => h[e] + sprm[c]	Transport, Extracellular
29	PAS_chr2-2_0127	ru5p-D[c] <=> xu5p-D[c]	Oxidative Phosphorylation
30	PAS_chr1-4_0659	ru5p-D[c] <=> xu5p-D[c]	Other
31	PAS_chr2-1_0547	ptd1ino_PP[c] + cer1_24[c] => 12dgr_PP[c] + ipc124_PP[c]	Cysteine Metabolism
32	PAS_chr3_0462	h[c] + pa_PP[c] + ctp[c] <=> ppi[c] + cdpdag_PP[c] h[m] + ctp[m] + pa_PP[m] <=> ppi[m] + cdpdag_PP[m]	tRNA charging

S 5-1 – Regression analysis of all mechanical combinations using experimental design

Table 8- 29 Regression analysis of all mechanic combinantions using experimental design

		TSD RT-RT			TSD RT-EEDP			TSD EEDP-RT			TSD EEDP-EEUP		
Coefficient		W	GS1	GS2	W	GS1	GS2	W	GS1	GS2	W	GS1	GS2
Response means	b0	5.59E-2	4.67E-2	1.97E-2	6.00E-2	2.75E-2	1.58E-2	4.27E-2	2.93E-2	1.91E-2	5.05E-2	3.07E-2	1.92E-2
RPM	b1	2.03E-2***	7.78E-3***	6.32E-3***	1.81E-2***	1.43E-2***	9.77E-3***	4.16E-3^{ns}	6.95E-3**	5.75E-3***	1.52E-2**	1.89E-3^{ns}	5.74E-3***
VVM	b2	1.26E-2**	7.00E-3**	6.17E-4^{n}s	1.53E-2**	5.54E-3***	3.03E-3***	1.43E-2**	1.05E-2***	3.24E-3***	1.71E-2**	1.01E-2***	3.79E-3**
RPM·VVM	b12	3.65E-3^{ns}	2.8E-4^{ns}	4.17E-4^{ns}	3.63E-3^{ns}	4.96E-4^{ns}	-1.93E-3***	3.08E-4^{ns}	-5.17E-5^{ns}	-8.53E-4*	5.52E-3^{ns}	2.56E-3^{ns}	8.15E-4^{ns}
R^2		87.9	84.4	97.8	85.0	99.3	99.2	68.1	87.4	97.7	82.2	87.0	88.3

		SD RT-RT			SD RT-EEDP			SD EEDP-RT			SD EEDP-EEUP		
Coefficient		W	GS1	GS2	W	GS1	GS2	W	GS1	GS2	W	GS1	GS2
Response means	b0	4.47E-2	1.82E-2	1.25E-2	3.67E-2	1.71E-2	1.57E-2	5.46E-2	1.84E-2	1.10E-2	5.49E-2	2.11E-2	1.21E-2
RPM	b1	2.05E-2***	1.23E-2***	5.43E-3***	2.34E-2***	7.92E-3***	4.86E-3***	2.30E-2***	9.94E-3***	3.01E-3***	1.82E-2**	5.94E-3**	1.29E-3^{ns}
VVM	b2	9.4E-3**	2.03E-3***	1.13E-3*	7.57E-3***	3.05E-3*	2.53E-3**	1.07E-2*	2.22E-3*	1.28E-3^{ns}	1.65E-2**	3.99E-3*	6.92E-4^{ns}
RPM·VVM	b12	2.98E-3^{ns}	3.58E-5^{ns}	-3.79E-4^{ns}	3.52E-3***	6.66E-4^{ns}	-3.78E-4^{ns}	3.77E-3^{ns}	-1.41E-4^{ns}	-5.45E-4^{ns}	6.71E-3^{ns}	1.90E-3^{ns}	-6.58E-4^{ns}
R^2		92.4	99.4	95.0	99.3	88.9	87.0	87.4	95.2	88.1	82.0	70.8	46.7

*** Significant($p<0.001$)

** Significant($p<0.01$)

* Significant($p<0.05$)

^{ns} Not Significant